Network Working Group
Request for Comments: 4081
Category: Informational

H. Tschofenig
D. Kroeselberg
Siemens

June 2005

Security Threats for Next Steps in Signaling (NSIS)

Status of This Memo

Copyright Notice

Abstract

This threats document provides a detailed analysis of the security threats relevant to the Next Steps in Signaling (NSIS) protocol suite. It calls attention to, and helps with the understanding of, various security considerations in the NSIS Requirements, Framework, and Protocol proposals. This document does not describe vulnerabilities of specific parts of the NSIS protocol suite.

Table of Contents

1. Introduction 2
2. Communications Models 3
3. Generic Threats 7
   3.1. Man-in-the-Middle Attacks 8
   3.2. Replay of Signaling Messages 11
   3.3. Injecting or Modifying Messages 11
   3.4. Insecure Parameter Exchange and Negotiation 12
4. NSIS-Specific Threat Scenarios 12
   4.1. Threats during NSIS SA Usage 13
   4.2. Flooding 13
   4.3. Eavesdropping and Traffic Analysis 15
   4.4. Identity Spoofing 15
   4.5. Unprotected Authorization Information 17
   4.6. Missing Non-Repudiation 18
   4.7. Malicious NSIS Entity 19
   4.8. Denial of Service Attacks 20
   4.9. Disclosing the Network Topology 21
   4.10. Unprotected Session or Reservation Ownership 21
   4.11. Attacks against the NTLP 23

5. Security Considerations 23
6. Contributors 24
7. Acknowledgements 24
8. References 25
8.1. Normative References 25
8.2. Informative References 25

1. Introduction

Whenever a new protocol is developed or existing protocols are modified, threats to their security should be evaluated. To address security in the NSIS working group, a number of steps have been taken:

NSIS Analysis Activities (see [RSVP-SEC] and [SIG-ANAL])

Security Threats for NSIS

NSIS Requirements (see [RFC3726])

NSIS Framework (see [RFC4080])

NSIS Protocol Suite (see GIMPS [GIMPS], NAT/Firewall NSLP [NATFW-NSLP] and QoS NSLP [QOS-NSLP])

This document identifies the basic security threats that need to be addressed during the design of the NSIS protocol suite. Even if the base protocol is secure, certain extensions may cause problems when used in a particular environment.

This document cannot provide detailed threats for all possible NSIS Signaling Layer Protocols (NSLPs). QoS [QOS-NSLP], NAT/Firewall [NATFW-NSLP], and other NSLP documents need to provide a description of their trust models and a threat assessment for their specific application domain. This document aims to provide some help for the subsequent design of the NSIS protocol suite. Investigations of security threats in a specific architecture or context are outside the scope of this document.

We use the NSIS terms defined in [RFC3726] and in [RFC4080].

2. Communications Models

The NSIS suite of protocols is envisioned to support various signaling applications that need to install and/or manipulate state at nodes along the data flow path through the network. As such, the NSIS protocol suite involves the communication between different entities.

This section offers terminology for common communication models that are relevant to securing the NSIS protocol suite.

An abstract network topology with its administrative domains is shown in Figure 1, and in Figure 2 the relationship between NSIS entities along the path is shown. For illustrative reasons, only end-to-end NSIS signaling is depicted, yet it might be used in other variations as well. Signaling can start at any place and might terminate at any other place within the network. Depending on the trust relationship between NSIS entities and the traversed network parts, different security problems arise.

The notion of trust and trust relationship used in this document is informal and can best be captured by the definition provided in Section 1.1 of [RFC3756]. For completeness we include the definition of a trust relationship, which denotes a mutual a priori relationship between the involved organizations or parties wherein the parties believe that the other parties will behave correctly even in the future.

An important observation for NSIS is that a certain degree of trust has to be placed into intermediate NSIS nodes along the path between an NSIS Initiator and an NSIS Responder, specifically so that they perform message processing and take the necessary actions. A complete lack of trust between any of the participating entities will cause NSIS signaling to fail.

Note that it is not possible to describe a trust model completely without considering the details and behavior of the NTLP, the NSLP (e.g., QoS NSLP), and the deployment environment. For example, securing the communication between an end host (which acts as the NSIS Initiator) and the first NSIS node (which might be in the attached network or even a number of networks away) is impacted by the trust relationships between these entities. In a corporate network environment, a stronger degree of trust typically exists than in an unmanaged network.

Figure 1 introduces convenient abbreviations for network parts with similar properties: first-peer, last-peer, intra-domain, or inter-domain.

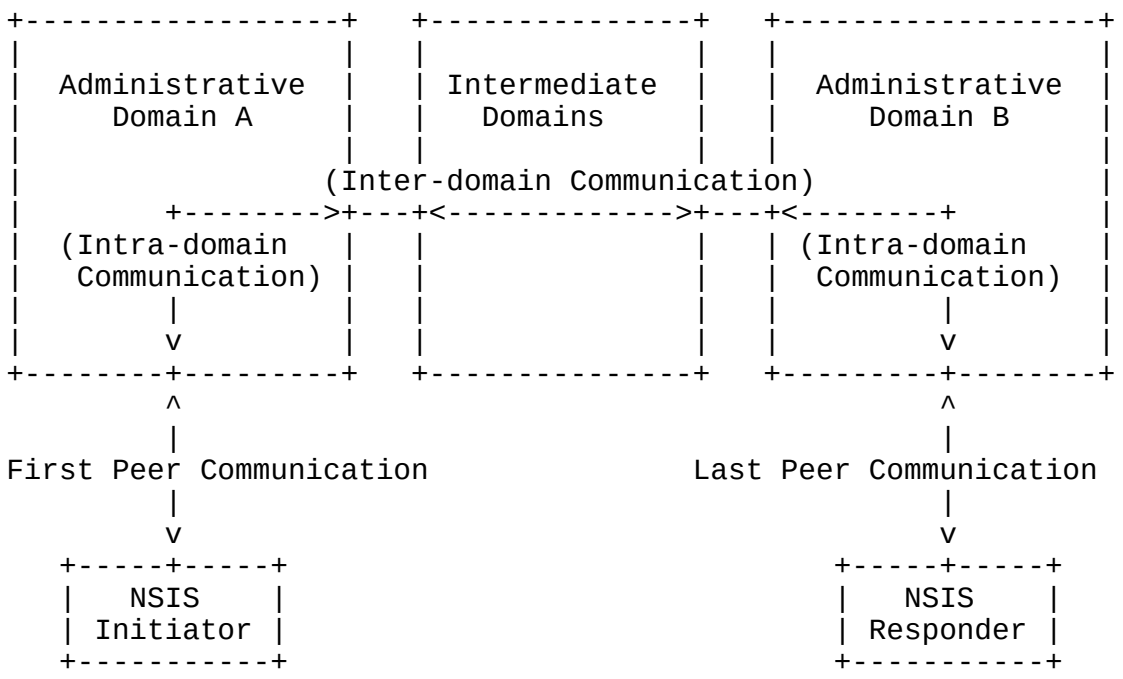


Figure 1: Communication patterns in NSIS

First-Peer/Last-Peer Communication:

The end-to-end communication scenario depicted in Figure 1 includes the communication between the end hosts and their nearest NSIS hops. "First-peer communications" refers to the peer-to-peer interaction between a signaling message originator, the NSIS Initiator (NI), and the first NSIS-aware entity along the path. This "first-peer communications" commonly comes with specific security requirements that are especially important for addressing security issues between the end host (and a user) and the network it is attached to.

To illustrate this, in roaming environments, it is difficult to assume the existence of a pre-established security association directly available for NSIS peers involved in first-peer communications, because these peers cannot be assumed to have any pre-existing relationship with each other. In contrast, in enterprise networks usually there is a fairly strong (pre-established) trust relationship between the peers. Enterprise network administrators usually have some degree of freedom to select the appropriate security protection and to enforce it. The choice of selecting a security mechanism is therefore often influenced by the infrastructure already

available, and per-session negotiation of security mechanisms is often not required (although, in contrast, it is required in a roaming environment).

Last-Peer communication is a variation of First-Peer communication in which the roles are reversed.

Intra-Domain Communication:

After verification of the NSIS signaling message at the border of an administrative domain, an NSIS signaling message traverses the network within the same administrative domain to which the first peer belongs. It might not be necessary to repeat the authorization procedure of the NSIS initiator again at every NSIS node within this domain. Key management within the administrative domain might also be simpler.

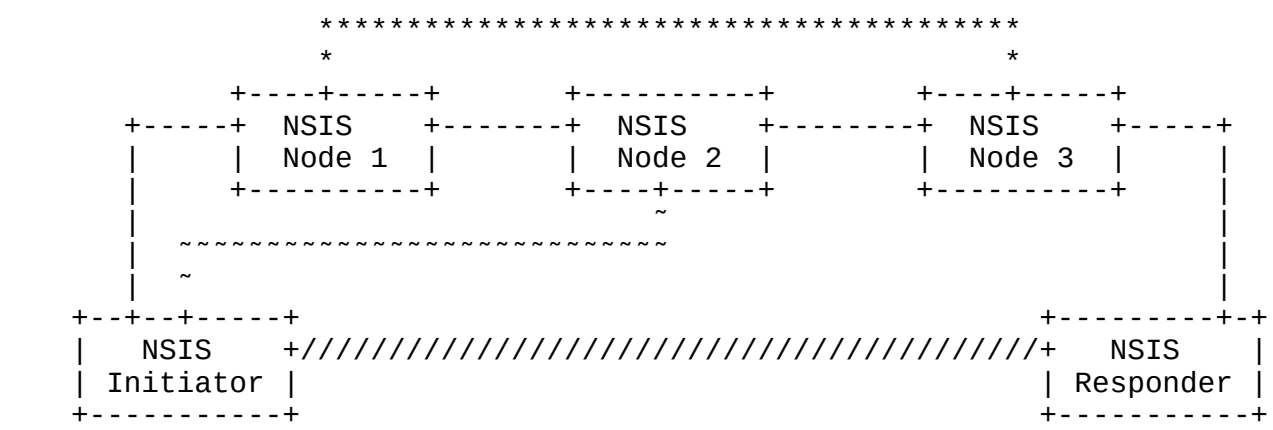
Security protection is still required to prevent threats by non-NSIS nodes in this network.

Inter-Domain Communication:

Inter-Domain communication deals with the interaction between administrative domains. For some NSLPs (for example, QoS NSLP), this interaction is likely to take place between neighboring domains, whereas in other NSLPs (such as the NAT/Firewall NSLP), the core network is usually not involved.

If signaling messages are conveyed transparently in the core network (i.e., if they are neither intercepted nor processed in the core network), then the signaling message communications effectively takes place between access networks. This might place a burden on authorization handling and on the key management infrastructure required between these access networks, which might not know of each other in advance.

To refine the above differentiation based on the network parts that NSIS signaling may traverse, we subsequently consider relationships between involved entities. Because a number of NSIS nodes might actively participate in a specific protocol exchange, a larger number of possible relationships need to be analyzed than in other protocols. Figure 2 illustrates possible relationships between the entities involved in the NSIS protocol suite.



Legend:
-----: Peer-to-Peer Relationship
/////: End-to-End Relationship
*****: Middle-to-Middle Relationship
~~~~~: End-to-Middle Relationship

Figure 2: Possible NSIS Relationships

End-to-Middle Communications:

The scenario in which one NSIS entity involved is an end-entity (Initiator or Responder) and the other entity is any intermediate hop other than the immediately adjacent peer is typically called the end-to-middle scenario (see Figure 2). A motivation for including this scenario can, for example, be found in SIP [RFC3261].

An example of end-to-middle interaction might be an explicit authorization from the NSIS Initiator to some intermediate node. Threats specific to this scenario may be introduced by some intermediate NSIS hops that are not allowed to eavesdrop or modify certain objects.

Middle-to-Middle Communications:

Middle-to-middle communication refers to the exchange of information between two non-neighboring NSIS nodes along the path. Intermediate NSIS hops may have to deal with specific security threats that do not involve the NSIS Initiator or the NSIS Responder directly.
~~~~~

End-to-End Communications:

NSIS aims to signal information from an Initiator to some NSIS nodes along the path to a data receiver. In the case of end-to-end NSIS signaling, the last node is the NSIS Responder, as it is the data receiver. The NSIS protocol suite is not an end-to-end protocol used to exchange information purely between end hosts.

Typically, it is not required to protect NSIS messages cryptographically between the NSIS Initiator and the NSIS Responder. Protecting the entire signaling message end-to-end might not be feasible since intermediate NSIS nodes need to add, inspect, modify, or delete objects from the signaling message.

3. Generic Threats

This section provides scenarios of threats that are applicable to signaling protocols in general. Note that some of these scenarios use the term "user" instead of "NSIS Initiator". This is mainly because security protocols allow differentiation between entities that are hosts and those that are users (based on the identifiers used).

For the following subsections, we use the general distinction in two cases in which attacks may occur. These are according to the separate steps, or phases, normally encountered when applying protocol security (with, e.g., IPsec, TLS, Kerberos, or SSH). Therefore, this section starts by briefly describing a motivation for this separation.

Security protection of protocols is often separated into two steps. The first step primarily provides entity authentication and key establishment (which result in a persistent state often called a security association), whereas the second step provides message protection (some combination of data origin authentication, data integrity, confidentiality, and replay protection) using the previously established security association. The first step tends to be more expensive than the second, which is the main reason for the separation. If messages are transmitted infrequently, then these two steps may be collapsed into a single and usually rather costly one. One such example is e-mail protection via S/MIME. The two steps may be tightly bound into a single protocol, as in TLS, or defined in separate protocols, as with IKE and IPsec. We use this separation to cover the different threats in more detail.

3.1. Man-in-the-Middle Attacks

This section describes both security threats that exist if two peers do not already share a security association or do not use security mechanisms at all, and threats that are applicable when a security association is already established.

Attacks during NSIS SA Establishment:

While establishing a security association, an adversary fools the signaling message Initiator with respect to the entity to which it has to authenticate. The Initiator authenticates to the man-in-the-middle adversary, who is then able to modify signaling messages to mount DoS attacks or to steal services that get billed to the Initiator. In addition, the adversary may be able to terminate the Initiator's NSIS messages and to inject messages to a peer itself, thereby acting as the peer to the Initiator and as the Initiator to the peer. As a result, the Initiator wrongly believes that it is talking to the "real" network, whereas it is actually attached to an adversary. For this attack to be successful, pre-conditions that are described in the following three cases have to hold:

Missing Authentication:

In the first case, this threat can be carried out because of missing authentication between neighboring peers: without authentication, an NI, NR, or NF is unable to detect an adversary. However, in some practical cases, authentication might be difficult to accomplish, either because the next peer is unknown, because there are misbelieved trust relationships in parts of the network, or because of the inability to establish proper security protection (inter-domain signaling messages, dynamic establishment of a security association, etc.). If one of the communicating endpoints is unknown, then for some security mechanisms it is either impossible or impractical to apply appropriate security protection. Sometimes network administrators use intra-domain signaling messages without proper security. This configuration allows an adversary on a compromised non-NSIS-aware node to interfere with nodes running an NSIS signaling protocol. Note that this type of threat goes beyond those caused by malicious NSIS nodes (described in Section 4.7).

Unilateral Authentication:

In the case of unilateral authentication, the NSIS entity that does not authenticate its peer is unable to discover a man-in-the-middle adversary. Although mutual authentication of signaling messages should take place between each peer participating in the protocol operation, special attention is given here to first-peer communications. Unilateral authentication between an end host and the first peer (just authenticating the end host) is still common today, but it opens up many possibilities for man-in-the-middle attackers impersonating either the end host or the (administrative domain represented by the) first peer.

Missing or unilateral authentication, as described above, is part of a general problem of network access with inadequate authentication, and it should not be considered something unique to the NSIS signaling protocol. Obviously, there is a strong need to address this correctly in a future NSIS protocol suite. The signaling protocols addressed by NSIS are different from other protocols in which only two entities are involved. Note that first-peer authentication is especially important because a security breach there could impact nodes beyond the entities directly involved (or even beyond a local network).

Finally, note that the signaling protocol should be considered a peer-to-peer protocol, wherein the roles of Initiator and Responder can be reversed at any time. Thus, unilateral authentication is not particularly useful for such a protocol. However, some form of asymmetry might be needed in the authentication process, whereby one entity uses an authentication mechanism different from that of the other one. As an example, the combination of symmetric and asymmetric cryptography should be mentioned.

Weak Authentication:

In the case of weak authentication, the threat can be carried out because information transmitted during the NSIS SA establishment process may leak passwords or allow offline dictionary attacks. This threat is applicable to NSIS for the process of selecting certain security mechanisms.

Finally, we conclude with a description of a man-in-the-middle (MITM) attack during the discovery phase. This attack benefits from the fact that NSIS nodes are likely to be unaware of the network

topology. Furthermore, an authorization problem might arise if an NSIS QoS NSLP node pretends to be an NSIS NAT/Firewall-specific node or vice versa.

An adversary might inject a bogus reply message, forcing the discovery message initiator to start a messaging association establishment with either an adversary or with another NSIS node that is not along the path. Figure 3 describes the attack in more detail for peer-to-peer addressed messages with a discovery mechanism. For end-to-end addressed messages, the attack is also applicable, particularly if the adversary is located along the path and able to intercept the discovery message that traverses the adversary. The man-in-the-middle adversary might redirect to another legitimate NSIS node. A malicious NSIS node can be detected with the corresponding security mechanisms, but a legitimate NSIS node that is not the next NSIS node along the path cannot be detected without topology knowledge.

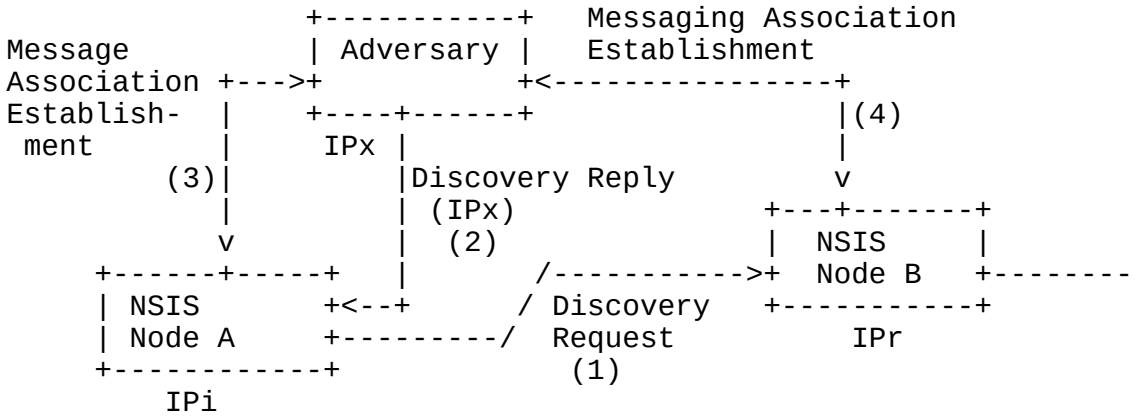


Figure 3: MITM Attack during the Discovery Exchange

This attack assumes that the adversary is able to eavesdrop on the initial discovery message sent by the sender of the discovery message. Furthermore, we assume that the discovery reply message by the adversary returns to the discovery message initiator faster than the real response. This represents some race condition characteristics if the next NSIS node is very close (in IP-hop terms) to the initiator. Note that the problem is self-healing since the discovery process is periodically repeated. If an adversary is unable to mount this attack with every discovery message, then the correct next NSIS node along the path will be discovered again. A ping-pong behavior might be the consequence.

As shown in message step (2) in Figure 3, the adversary returns a discovery reply message with its own IP address as the next NSIS-aware node along the path. Without any additional information, the discovery message initiator has to trust this information. Then a messaging association is established with an entity at a given IP address IPx (i.e., with the adversary) in step (3). The adversary then establishes a messaging association with a further NSIS node and forwards the signaling message. Note that the adversary might just modify the Discovery Reply message to force NSIS Node A to establish a messaging association with another NSIS node that is not along the path. This can then be exploited by the adversary. The interworking with NSIS-unaware NATs in particular might cause additional unexpected problems.

As a variant of this attack, an adversary not able to eavesdrop on transmitted discovery requests could flood a node with bogus discovery reply messages. If the discovery message sender accidentally accepts one of those bogus messages, then a MITM attack as described in Figure 3 is possible.

3.2. Replay of Signaling Messages

This threat scenario covers the case in which an adversary eavesdrops, collects signaling messages, and replays them at a later time (or at a different place, or uses parts of them at a different place or in a different way; e.g., cut-and-paste attacks). Without proper replay protection, an adversary might mount man-in-the-middle, denial of service, and theft of service attacks.

A more difficult attack (that may cause problems even if there is replay protection) requires that the adversary crash an NSIS-aware node, causing it to lose state information (sequence numbers, security associations, etc.), and then replay old signaling messages. This attack takes advantage of re-synchronization deficiencies.

3.3. Injecting or Modifying Messages

This type of threat involves integrity violations, whereby an adversary modifies signaling messages (e.g., by acting as a man-in-the-middle) in order to cause unexpected network behavior. Possible actions an adversary might consider for its attack are reordering, delaying, dropping, injecting, truncating, and otherwise modifying messages.

An adversary may inject a signaling message requesting a large amount of resources (possibly using a different user’s identity). Other resource requests may then be rejected. In combination with identity

spoofing, it is possible to carry out fraud. This attack is only feasible in the absence of authentication and signaling message protection.

Some threats directly related to these are described in Sections 4.4, 4.7, and 4.8.

3.4. Insecure Parameter Exchange and Negotiation

First, protocols may be useful in a variety of scenarios with different security requirements. Second, different users (e.g., a university, a hospital, a commercial enterprise, or a government ministry) have inherently different security requirements. Third, different parts of a network (e.g., within a building, across a public carrier's network, or over a private microwave link) may need different levels of protection. It is often difficult to meet these (sometimes conflicting) requirements with a single security mechanism or fixed set of security parameters, so often a selection of mechanisms and parameters is offered. Therefore, a protocol is required to agree on certain security mechanisms and parameters. An insecure parameter exchange or security negotiation protocol can help an adversary to mount a downgrading attack to force selection of mechanisms weaker than those mutually desired. Thus, without binding the negotiation process to the legitimate parties and protecting it, an NSIS protocol suite might only be as secure as the weakest mechanism provided (e.g., weak authentication), and the benefits of defining configuration parameters and a negotiation protocol are lost.

4. NSIS-Specific Threat Scenarios

This section describes eleven threat scenarios in terms of attacks on and security deficiencies in the NSIS signaling protocol. A number of security deficiencies might enable an attack. Fraud is an example of an attack that might be enabled by missing replay protection, missing protection of authorization tokens, identity spoofing, missing authentication, and other deficiencies that help an adversary steal resources. Different threat scenarios based on deficiencies that could enable an attack are addressed in this section.

The threat scenarios are not independent. Some of them (e.g., denial of service) are well-established security terms and, as such, need to be addressed, but they are often enabled by one or more deficiencies described under other scenarios.

4.1. Threats during NSIS SA Usage

Once a security association is established (and used) to protect signaling messages, many basic attacks are prevented. However, a malicious NSIS node is still able to perform various attacks as described in Section 4.7. Replay attacks may be possible when an NSIS node crashes, restarts, and performs state re-establishment. Proper re-synchronization of the security mechanism must therefore be provided to address this problem.

4.2. Flooding

This section describes attacks that allow an adversary to flood an NSIS node with bogus signaling messages to cause a denial of service attack.

We will discuss this threat at different layers in the NSIS protocol suite:

Processing of Router Alert Options:

The processing of Router Alert Option (RAO) requires that a router do some additional processing by intercepting packets with IP options, which might lead to additional delay for legitimate requests, or even rejection of some of them. A router being flooded with a large number of bogus messages requires resources before finding out that these messages have to be dropped.

If the protocol is based on using interception for message delivery, this threat cannot be completely eliminated, but the protocol design should attempt to limit the processing that has to be done on the RAO-bearing packet so that it is as similar as possible to that for an arbitrary packet addressed directly to one of the router interfaces.

Attacks against the Transport Layer Protocol:

Certain attacks can be mounted against transport protocols by flooding a node with bogus requests, or even to finish the handshake phase to establish a transport layer association. These types of threats are also addressed in Section 4.11.

Force NTLP to Do More Processing:

Some protocol fields might allow an adversary to force an NTLP node to perform more processing. Additionally it might be possible to interfere with the flow control or the congestion control procedure. These types of threats are also addressed in Section 4.11.

Furthermore, it might be possible to force the NTLP node to perform some computations or signaling message exchanges by injecting "trigger" events (which are unprotected).

Force NSLP to Do More Processing:

An adversary might benefit from flooding an NSLP node with messages that must be stored (e.g., due to fragmentation handling) before verifying the correctness of signaling messages.

Furthermore, causing memory allocation and computational efforts might allow an adversary to harm NSIS entities. If a signaling message contains, for example, a digital signature, then some additional processing is required for the cryptographic verification. An adversary can easily create a random bit sequence instead of a digital signature to force an NSIS node into heavy computation.

Idempotent signaling messages are particularly vulnerable to this type of attack. The term "idempotent" refers to messages that contain the same amount of information as the original message. An example would be a refresh message that is equivalent to a create message. This property allows a refresh message to create state along a new path, where no previous state is available. For this to work, specific classes of cryptographic mechanisms supporting this behavior are needed. An example is a scheme based on digital signatures, which, however, should be used with care due to possible denial of service attacks.

Problems with the usage of public-key-based cryptosystems in protocols are described in [AN97] and in [ALN00].

In addition to the threat scenario described above, an incoming signaling message might trigger communication with third-party nodes such as policy servers, LDAP servers, or AAA servers. If an adversary is able to transmit a large number of signaling messages (for example, with QoS reservation requests) with invalid credentials, then the verifying node may not be able to process other reservation messages from legitimate users.

4.3. Eavesdropping and Traffic Analysis

This section covers threats whereby an adversary is able to eavesdrop on signaling messages. The signaling packets collected may allow traffic analysis or be used later to mount replay attacks, as described in Section 3.2. The eavesdropper might learn QoS parameters, communication patterns, policy rules for firewall traversal, policy information, application identifiers, user identities, NAT bindings, authorization objects, network configuration and performance information, and more.

An adversary's capability to eavesdrop on signaling messages might violate a user's preference for privacy, particularly if unprotected authentication or authorization information (including policies and profile information) is exchanged.

Because the NSIS protocol signals messages through a number of nodes, it is possible to differentiate between nodes actively participating in the NSIS protocol and those that do not. For certain objects or messages, it might be desirable to permit actively participating intermediate NSIS nodes to eavesdrop. On the other hand, it might be desirable that only the intended end points (NSIS Initiator and NSIS Responder) be able to read certain other objects.

4.4. Identity Spoofing

Identity spoofing relevant for NSIS occurs in three forms: First, identity spoofing can happen during the establishment of a security association based on a weak authentication mechanism. Second, an adversary can modify the flow identifier carried within a signaling message. Third, it can spoof data traffic.

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 consumed. This type of attack is possible if authentication is based on a simple username identifier (i.e., in absence of cryptographic authentication), or if authentication is provided for hosts, and multiple users have access to a single host. This attack could also be classified as theft of service.

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 (UDP, TCP), port numbers, and flow labels (see [RFC1809] and [RFC3697]). Modification of these flow identifiers allows adversaries to exploit or to render ineffective quality of service

reservations or policy rules at middleboxes. An adversary could mount an attack by modifying the flow identifier of a signaling message.

In the third case, an adversary may spoof data traffic. NSIS signaling messages contain some sort of flow identifier that is associated with a specified behavior (e.g., a particular flow experiences QoS treatment or allows packets to traverse a firewall). An adversary might, therefore, use IP spoofing and inject data packets to benefit from previously installed flow identifiers.

We will provide an example of the latter threat. After NSIS nodes along the path between the NSIS initiator and the NSIS receiver processes a properly protected reservation request, transmitted by the legitimate user Alice, a QoS reservation is installed at the corresponding NSIS nodes (for example, the edge router). The flow identifier is used for flow identification and allows data traffic originated from a given source to be assigned to this QoS reservation. The adversary Eve now spoofs Alice's IP address. In addition, Alice's host may be crashed by the adversary with a denial of service attack or may lose connectivity (for example, because of mobility). If Eve is able to perform address spoofing, then she is able to receive and transmit data (for example, RTP data traffic) that receives preferential QoS treatment based on the previous reservation. Depending on the installed flow identifier granularity, Eve might have more possibilities to exploit the QoS reservation or a pin-holed firewall. Assuming the soft state paradigm, whereby periodic refresh messages are required, Alice's absence will not be detected until a refresh message is required, forcing Eve to respond with a protected signaling message. Again, this attack is applicable not only to QoS traffic, but also to a Firewall control protocol, with a different consequence.

The ability for an adversary to inject data traffic that matches a certain flow identifier established by a legitimate user and to get some benefit from injecting that traffic often also requires the ability to receive the data traffic or to have one's correspondent receive it. For example, an adversary in an unmanaged network observes a NAT/Firewall signaling message towards a corporate network. After the signaling message exchange was successful, the user Alice is allowed to traverse the company firewall based on the establish packet filter in order to contact her internal mail server. Now, the adversary Eve, who was monitoring the signaling exchange, is able to build a data packet towards this mail server that will pass the company firewall. The packet will hit the mail server and cause some actions, and the mail server will reply with some response messages. Depending on the exact location of the adversary and the

degree of routing asymmetry, the adversary might even see the response messages. Note that for this attack to work, Alice does not need to participate in the exchange of signaling messages.

We could imagine using attributes of a flow identifier that is not related to source and destination addresses. For example, we could think of a flow identifier for which only the 21-bit Flow ID is used (without source and destination IP address). Identity spoofing and injecting traffic is much easier since a packet only needs to be marked and an adversary can use a nearly arbitrary endpoint identifier to achieve the desired result. Obviously, though, the endpoint identifiers are not irrelevant, because the messages have to hit some nodes in the network where NSIS signaling messages installed state (in the above example, they would have to hit the same firewall).

Data traffic marking based on DiffServ is such an example. Whenever an ingress router uses only marked incoming data traffic for admission control procedures, various attacks are possible. These problems have been known in the DiffServ community for a long time and have been documented in various DiffServ-related documents. The IPsec protection of DiffServ Code Points is described in Section 6.2 of [RFC2745]. Related security issues (for example denial of service attacks) are described in Section 6.1 of the same document.

4.5. Unprotected Authorization Information

Authorization is an important criterion for providing resources such as QoS reservations, NAT bindings, and pinholes through firewalls. Authorization information might be delivered to the NSIS-participating entities in a number of ways.

Typically, the authenticated identity is used to assist during the authorization procedure (as described in [RFC3182], for example). Depending on the chosen authentication protocol, certain threats may exist. Section 3 discusses a number of issues related to this approach when the authentication and key exchange protocol is used to establish session keys for signaling message protection.

Another approach is to use some sort of authorization token. The functionality and structure of such an authorization token for RSVP is described in [RFC3520] and [RFC3521].

Achieving secure interaction between different protocols based on authorization tokens, however, requires some care. By using such an authorization token, it is possible to link state information between different protocols. Returning an unprotected authorization token to the end host might allow an adversary (for example, an eavesdropper)

to steal resources. An adversary might also use the token to monitor communication patterns. Finally, an untrustworthy end host might also modify the token content.

The Session/Reservation Ownership problem can also be regarded as an authorization problem. Details are described in Section 4.10. In enterprise networks, authorization is often coupled with membership in a particular class of users or groups. This type of information either can be delivered as part of the authentication and key agreement procedure or has to be retrieved via separate protocols from other entities. If an adversary manages to modify information relevant to determining authorization or the outcome of the authorization process itself, then theft of service might be possible.

4.6. Missing Non-Repudiation

Signaling for QoS often involves three parties: the user, a network that offers QoS reservations (referred to as "service provider") and a third party that guarantees that the party making the reservation actually receives a financial compensation (referred to as "trusted third party").

In this context,"repudiation" refers to a problem where either the user or the service provider later deny the existence or some parameters (e.g., volume or price) of a QoS reservation towards the trusted third party. Problems stemming from a lack of non-repudiation appear in two forms:

Service provider’s point-of-view:
: A user may deny having issued a reservation request for which it was charged. The service provider may then want to be able to prove 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.

In today’s networks, non-repudiation is not provided. Therefore, it might be difficult to introduce with NSIS signaling. The user has to trust the network operator to meter the traffic correctly, to collect and merge accounting data, and to ensure that no unforeseen problems

occur. If a signaling protocol with the non-repudiation property is desired for establishing QoS reservations, then it certainly impacts the protocol design.

Non-repudiation functionality places additional requirements on the security mechanisms. Thus, a solution would normally increase the overhead of a security solution. Threats related to missing non-repudiation are only considered relevant in certain specific scenarios and for specific NSLPs.

4.7. Malicious NSIS Entity

Network elements within a domain (intra-domain) experience a different trust relationship with regard to the security protection of signaling messages from that of edge NSIS entities. It is assumed that edge NSIS entities are responsible for performing cryptographic processing (authentication, integrity and replay protection, authorization, and accounting) for signaling messages arriving from the outside. This prevents unprotected signaling messages from appearing within the internal network. If, however, an adversary manages to take over an edge router, then the security of the entire network is compromised. An adversary is then able to launch a number of attacks, including denial of service; integrity violations; replay and reordering of objects and messages; bundling of messages; deletion of data packets; and various others. A rogue firewall can harm other firewalls by modifying policy rules. The chain-of-trust principle applied in peer-to-peer security protection cannot protect against a malicious NSIS node. An adversary with access to an NSIS router is also able to get access to security associations and to transmit secured signaling messages. Note that even non-peer-to-peer security protection might not be able to prevent this problem fully. Because an NSIS node might issue signaling messages on behalf of someone else (by acting as a proxy), additional problems need to be considered.

An NSIS-aware edge router is a critical component that requires strong security protection. A strong security policy applied at the edge does not imply that other routers within an intra-domain network do not need to verify signaling messages cryptographically. If the chain-of-trust principle is deployed, then the security protection of the entire path (in this case, within the network of a single administrative domain) is only as strong as the weakest link. In the case under consideration, the edge router is the most critical component of this network, and it may also act as a security gateway or firewall for incoming and outgoing traffic. For outgoing traffic, this device has to implement the security policy of the local domain and to apply the appropriate security protection.

For an adversary to mount this attack, either an existing NSIS-aware node along the path has to be attacked successfully, or an adversary must succeed in convincing another NSIS node to make it the next NSIS peer (man-in-the-middle attack).

4.8. Denial of Service Attacks

A number of denial of service (DoS) attacks can cause NSIS nodes to malfunction. Other attacks that could lead to DoS, such as man-in-the-middle attacks, replay attacks, and injection or modification of signaling messages, etc., are mentioned throughout this document.

Path Finding:

Some signaling protocols establish state (e.g., routing state) and perform some actions (e.g., querying resources) at a number of NSIS nodes without requiring authorization (or even proper authentication) based on a single message (e.g., PATH message in RSVP).

An adversary can utilize this fact to transmit a large number of signaling messages to allocate state at nodes along the path and to cause resource consumption.

An NSIS responder might not be able to determine the NSIS initiator and might even tend to respond to such a signaling message with a corresponding reservation message.

Discovery Phase:

Conveying signaling information to a large number of entities along a data path requires some sort of discovery. This discovery process is vulnerable to a number of attacks because it is difficult to secure. An adversary can use the discovery mechanisms to convince one entity to signal information to another entity that is not along the data path, or to cause the discovery process to fail. In the first case, the signaling protocol could appear to continue correctly, except that policy rules are installed at the incorrect firewalls or QoS resource reservations take place at the wrong entities. For an end host, this means that the protocol failed for unknown reasons.

Faked Error or Response Messages:

An adversary may be able to inject false error or response messages as part of a DoS attack. This could be at the signaling message protocol layer (NTLP), the layer of each client layer protocol (e.g., QoS NSLP or NAT/Firewall NSLP), or the transport protocol layer. An adversary might cause unexpected protocol behavior or might succeed with a DoS attack. The discovery protocol, especially, exhibits vulnerabilities with regard to this threat scenario (see the above discussion on discovery). If no separate discovery protocol is used and signaling messages are addressed to end hosts only (with a Router Alert Option to intercept message as NSIS aware nodes), an error message might be used to indicate a path change. Such a design combines a discovery protocol with a signaling message exchange protocol.

4.9. Disclosing the Network Topology

In some organizations or enterprises there is a desire not to reveal internal network structure (or other related information) outside of a closed community. An adversary might be able to use NSIS messages for network mapping (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 [RFC2745] for a description of diagnostic message functionality for RSVP), and query messages, in addition to record route and route objects, provide potential assistance to an adversary. Thus, the requirement of not disclosing a network topology might conflict with other requirements to provide means for discovering NSIS-aware nodes automatically or to provide diagnostic facilities (used for network monitoring and administration).

4.10. Unprotected Session or Reservation Ownership

Figure 4 shows an NSIS Initiator that has established state information at NSIS nodes along a path as part of the signaling procedure. As a result, Access Router 1, Router 3, and Router 4 (and other nodes) have stored session-state information, including the Session Identifier SID-x.

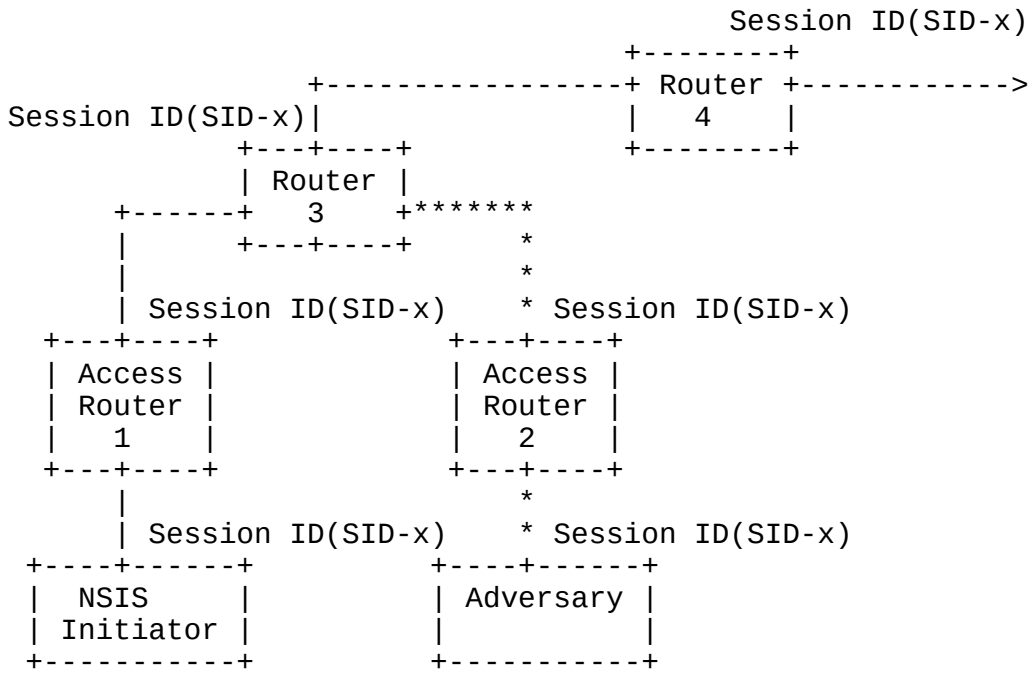


Figure 4: Session or Reservation Ownership

The Session Identifier is included in signaling messages to reference to the established state.

If an adversary were able to obtain the Session Identifier (for example, by eavesdropping on signaling messages), it would be able to add the same Session Identifier SID-x to a new signaling message. When the new signaling message hits Router 3 (as shown in Figure 4), existing state information can be modified. The adversary can then modify or delete the established reservation and cause unexpected behavior for the legitimate user.

The source of the problem is that Router 3 (a cross-over router) is unable to decide whether the new signaling message was initiated from the owner of the session or reservation.

In addition, nodes other than the initial signaling message originator are allowed to signal information during the lifetime of an established session. As part of the protocol, any NSIS-aware node along the path (and the path might change over time) could initiate a signaling message exchange. It might, for example, be necessary to provide mobility support or to trigger a local repair procedure. If only the initial signaling message originator were allowed to trigger signaling message exchanges, some protocol behavior would not be possible.

If this threat scenario is not addressed, an adversary can launch DoS, theft of service, and various other attacks.

4.11. Attacks against the NTLP

In [2LEVEL], a two-level architecture is proposed, that would split an NSIS protocol into layers: a signaling message transport-specific layer and an application-specific layer. This is further developed in the NSIS Framework [RFC4080]. Most of the threats described in this threat analysis are applicable to the NSLP application-specific part (e.g., QoS NSLP). There are, however, some threats that are applicable to the NTLP.

Network and transport layer protocols lacking protection mechanisms are vulnerable to certain attacks, such as header manipulation, DoS, spoofing of identities, session hijacking, unexpected aborts, etc. Malicious nodes can attack the congestion control mechanism to force NSIS nodes into a congestion avoidance state.

Threats that address parts of the NTLP that are not related to attacks against the use of transport layer protocols are covered in various sections throughout this document, such as Section 4.2.

If existing transport layer protocols are used for exchanging NSIS signaling messages, security vulnerabilities known for these protocols need to be considered. A detailed threat description of these protocols is outside the scope of this document.

5. Security Considerations

This entire memo discusses security issues relevant for NSIS protocol design. It begins by identifying the components of a network running NSIS (Initiator, Responder, and different Administrative Domains between them). It then considers five cases in which communications take place between these components, and it examines the trust relationships presumed to exist in each case: First-Peer Communications, End-to-Middle Communications, Intra-Domain Communications, Inter-Domain Communications, and End-to-End Communications. This analysis helps determine the security needs and the relative seriousness of different threats in the different cases.

The document points out the need for different protocol security measures: authentication, key exchange, message integrity, replay protection, confidentiality, authorization, and some precautions against denial of service. The threats are subdivided into generic ones (e.g., man-in-the-middle attacks, replay attacks, tampering and forgery, and attacks on security negotiation protocols) and eleven threat scenarios that are particularly applicable to the NSIS

protocol. Denial of service, for example, is covered in the NSIS-specific section, not because it cannot be carried out against other protocols, but because the methods used to carry out denial of service attacks tend to be protocol specific. Numerous illustrative examples provide insight into what can happen if these threats are not mitigated.

This document repeatedly points out that not all of the threats are equally serious in every context. It does attempt to identify the scenarios in which security failures may have the highest impact. However, it is difficult for the protocol designer to foresee all the ways in which NSIS protocols will be used or to anticipate the security concerns of a wide variety of likely users. Therefore, the protocol designer needs to offer a full range of security capabilities and ways for users to negotiate and select what they need, on a case-by-case basis. To counter these threats, security requirements have been listed in [RFC3726].

6. Contributors

We especially thank Richard Graveman, who provided text for the security considerations section, as well as a detailed review of the document.

7. Acknowledgements

We would like to thank (in alphabetical order) Marcus Brunner, Jorge Cuellar, Mehmet Ersue, Xiaoming Fu, and Robert Hancock for their comments on an initial version of this document. Jorge and Robert gave us an extensive list of comments and provided information on additional threats.

Jukka Manner, Martin Buechli, Roland Bless, Marcus Brunner, Michael Thomas, Cedric Aoun, John Loughney, Rene Soltwisch, Cornelia Kappler, Ted Wiederhold, Vishal Sankhla, Mohan Parthasarathy, and Andrew McDonald provided comments on more recent versions of this document. Their input helped improve the content of this document. Roland Bless, Michael Thomas, Joachim Kross, and Cornelia Kappler, in particular, provided good proposals for regrouping and restructuring the material.

A final review was given by Michael Richardson. We thank him for his detailed comments.

8. References

8.1. Normative References

[RFC4080] Hancock, R., Karagiannis, G., Loughney, J., and S. van den Bosch, "Next Steps in Signaling (NSIS): Framework", RFC 4080, June 2005.

[RFC3726] Brunner, M., "Requirements for Signaling Protocols", RFC 3726, April 2004.

8.2. Informative References

[ALN00] Aura, T., Leiwo, J., and P. Nikander, "Towards Network Denial of Service Resistant Protocols, In Proceedings of the 15th International Information Security Conference (IFIP/SEC 2000), Beijing, China", August 2000.

[AN97] Aura, T. and P. Nikander, "Stateless Connections", In Proceedings of the International Conference on Information and Communications Security (ICICS'97), Lecture Notes in Computer Science 1334, Springer", 1997.

[2LEVEL] Braden, R. and B. Lindell, "A Two-Level Architecture for Internet Signaling", Work in Progress, November 2002.

[RFC3697] Rajahalme, J., Conta, A., Carpenter, B., and S. Deering, "IPv6 Flow Label Specification", RFC 3697, March 2004.

[NATFW-NSLP] Stiemerling, M., "A NAT/Firewall NSIS Signaling Layer Protocol (NSLP)", Work in Progress, February 2005.

[GIMPS] Schulzrinne, H., "GIMPS: General Internet Messaging Protocol for Signaling", Work in Progress, February 2005.

[QOS-NSLP] Bosch, S., Karagiannis, G., and A. McDonald, "NSLP for Quality-of-Service signaling", Work in Progress, February 2005.

[RSVP-SEC] Tschofenig, H., "RSVP Security Properties", Work in Progress, February 2005.

[SIG-ANAL] Manner, J. and X. Fu, "Analysis of Existing Quality-of-Service Signaling Protocols", RFC 4094, May 2005.

[RFC1809] Partridge, C., "Using the Flow Label Field in IPv6", RFC 1809, June 1995.

[RFC2745] Terzis, A., Braden, B., Vincent, S., and L. Zhang, "RSVP Diagnostic Messages", RFC 2745, January 2000.

[RFC3182] Yadav, S., Yavatkar, R., Pabbati, R., Ford, P., Moore, T., Herzog, S., and R. Hess, "Identity Representation for RSVP", RFC 3182, October 2001.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[RFC3520] Hamer, L-N., Gage, B., Kosinski, B., and H. Shieh, "Session Authorization Policy Element", RFC 3520, April 2003.

[RFC3521] Hamer, L-N., Gage, B., and H. Shieh, "Framework for Session Set-up with Media Authorization", RFC 3521, April 2003.

[RFC3756] Nikander, P., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, May 2004.

Authors' Addresses

Hannes Tschofenig
Siemens
Otto-Hahn-Ring 6
Munich, Bavaria 81739
Germany

EMail: Hannes.Tschofenig@siemens.com

Dirk Kroeselberg
Siemens
Otto-Hahn-Ring 6
Munich, Bavaria 81739
Germany

EMail: Dirk.Kroeselberg@siemens.com

Full Copyright Statement

Intellectual Property

Acknowledgement

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