HIP Working Group
Internet-Draft
Intended status: Experimental
Expires: April 30, 2009

J. Melen
Ericsson Research Nomadiclab
M. Komu
HIIT
M. Bagnulo
Universidad Carlos III de Madrid
T. Henderson
The Boeing Company
October 27, 2008

# HIP Mobility and Multihoming Extensions for the Traversal of Network Address Translators
# draft-melen-hip-nat-mm-00



Abstract

This document defines extensions for HIP mobility and multihoming mechanisms to operate in network environments with NAT devices. The

extensions are based on the ICE protocol that allows two communicating end-hosts to establish a direct communications path with each other even when residing in separate private address realms. The focus of the extensions in this document is on fault-tolerance with symmetric locator pairs, and load-balancing is also discussed. This document also updates RFC5206.

Table of Contents

## 1. Terminology

In the absence of better terms, this document uses the terms Mobile Node (MN) and Corresponding Node (CN) borrowed from Mobile IP terminology even though HIP allows both ends to be mobile even simultaneously.

## 2. Introduction

The protocol extensions defined in this document extend HIP mobility and multihoming to operate in NATted environments. The extensions use combine ICE with HIP to create end-to-end connectivity and global naming for end-hosts located in different private address realms. This document focuses on fault-tolerance with symmetric locator pairs, but also load-balancing is discussed. This document updates [RFC5206].

The extensions in this document assume that the two communicating end-hosts have run successfully the base exchange procedure through a HIP Relay as descibed in [I-D.ietf-hip-nat-traversal]. In other words this document excludes the mechanisms to solve the initial contact problem. This document specifies extensions that allow HIP end-hosts to support end-host mobility and multihoming in NATted environments. The handover procedure is similar to the base exchange with NAT extensions. First, both end-hosts exchange offer and answer, i.e, their locators, using UPDATE messages. Second, the hosts start ICE connectivity checks [I-D.ietf-mmusic-ice] to discover a working address pair. Third, the hosts can use the discovered address pair for data traffic.

End-host mobility usually involves a disconnectivity period while a host is moving from a network to another. The delay caused by the disconnectivity period can have negative effects on transport layer traffic. Further, ICE connectivity checks also amplify the delay but are necessary to restore the connectivity. This document proposes some optimizations to reduce the length of the disconnectivity periods.

In the case of multihoming, a host first gathers its host candidates from its local network interfaces. Then, it collects server reflexive addresses by varying the source interface in UPDATE exchanges with RELAY server(s) and STUN server(s).

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

## 3. Mobility and Multihoming Scenarios

This section discusses end-host and site multihoming use cases. We assume that there are two communicating end-hosts that are located behind separate NAT devices.

### 3.1. End Host detects mobility event

#### 3.1.1. Make-before-break

In the make-before-break scenario the mobile node has at least two interfaces that can be simultaneously connected to different networks and can have distinct addresses configured. In the make-before-break scenario the existing security association is updated after the new pair of IP addresses has been detected to be working. As an example, let's consider a 4G phone with multiaccess capabilities. First, the phone is already transmitting data over one active interface. Then, the phone starts to use the other interface, but only after the handover procedure has been completed over the other link. The phone can trigger the handover procedure simultaneously while sending data over the active interface. Figure 1 depicts the make-before-break scenario.

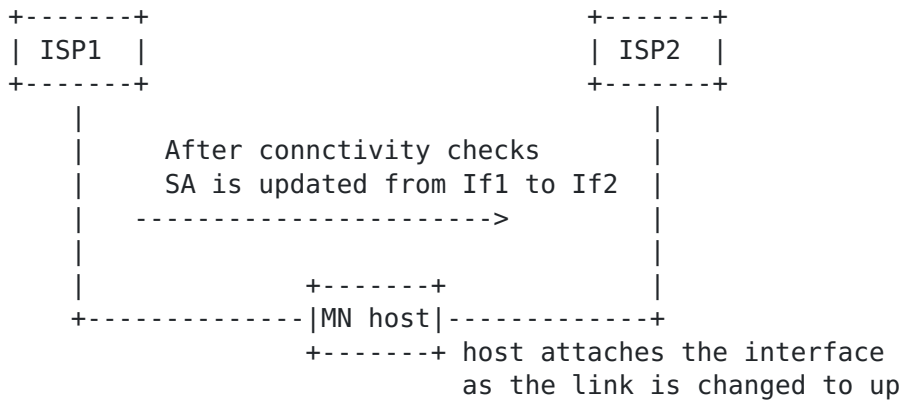


Figure 1: Basic make before break

#### 3.1.2. Break-before-make

In the break-before-make scenario, the connection to the peer is lost for a while when detaching from old access network and while attaching to new one. In this scenario, there is no data transmitted to the peer until the new attachment procedure has finished. A common example is that the host detaches from one access network and attaches to new one with the same interface. The detachment period may vary from a few milliseconds to hours. In this scenario, there is the possibility that the communication may have to be reinitated

after the detachment period depending on whether the peer has dropped the previous communication context or not.

## 3.2. End Host Detects a Failure in the End-to-end Path

### 3.2.1. Simultaneous End-host Multihoming

This section describes a scenario where a mobile host has two network interfaces which it uses simultaneously. The scenario is visualized in Figure 2.

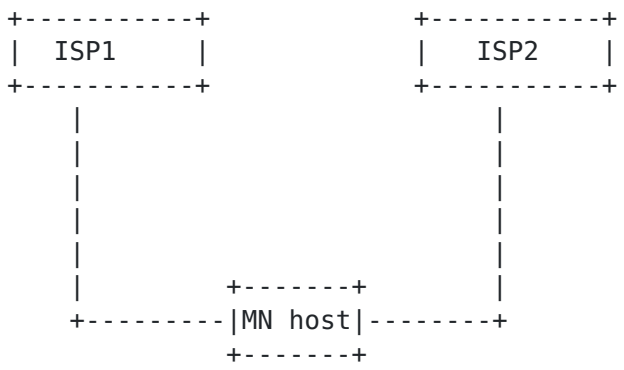


Figure 2: Simultaneous End-host Multihoming

Once the base exchange has been successfully completed as described in [I-D.ietf-hip-nat-traversal], the MN can gather the candidates for the other interface that was not used during the base exchange. For gathering the candidates, the host may use either an UPDATE exchange with the Relay server in Section 5.1.1 or a STUN server. After gathering the candidates, the MN MAY send an UPDATE packet containing an ICE offer, and the old SPI value in the ESP_INFO MUST be set to zero to denote that the MN creates a new multihoming SA pair that is parallel and independent from the SA pair that was previously created. The MN sends the UPDATE packet listing all candidates in the LOCATOR using a relay of the CN. The UPDATE exchange for setting up new SAs is same as in the case of mobility described in Section 5.1.2.

Another configuration would be to use multihoming for fault tolerance. In such a case, there is a primary path and a backup path. The backup path could be a "hot backup" or a "cold backup." In the hot backup case, the multihoming host knows the backup address beforehand and keeps the path up using keepalives as described in section Section 5.2. In the cold backup case, the host detects the failure and only then discover the candidates for the alternative path. The hot backup may cost more because the path needs to be kept alive. The cold backup requires just one SA pair which is then

changed similarly as in the case of mobility.

## 3.3. Routing System Detects a Failure in the End-to-end Path

In site multihoming, the end-host is not usually aware of the different paths the site has with the rest of the network. A typical configuration for site multihoming using multiple ISPs for outgoing traffic and for redundancy is in Figure 3. If one of the links fail, the traffic is handed over to run over a different link of an ISP.

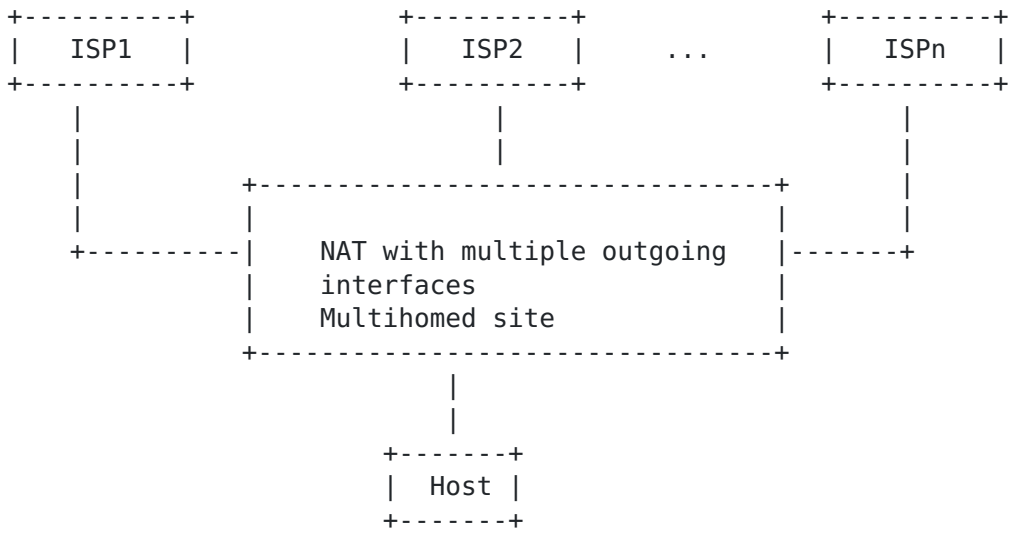


Figure 3: Site multihoming

In this scenario, the the host can discover only the public address of the NAT. When a failure occurs, the intrasite routing system will simply reroute to an alternative path without the host noticing it. The result is that the peer of the host starts receiving packets originating from the different transport address that belongs to a new NAT device. The peer learns a new route to the host and can start using it after successful HIP return routability or ICE connectivity checks.

To detect the disconnectivity, the host has to periodically send keep-alives through the active connection if no other data is being sent on the security association. The keep-alive interval SHOULD be configurable. When the host has not received a response to keep-alives for a configurable period, it should gather new ICE candidates and send a new ICE offer using an UPDATE packet to the peer. The peer responds to this with an UPDATE packet containing the ICE answer after which both of the end-hosts start the ICE connectivity checks.

## 4. Locator management

A multihomed HIP node has multiple locators that can have different reachability status. Some of them can be operational/reachable while other may be not. Fault tolerance is a preferred capability of such configuration. In order to provide basic fault tolerance support, a HIP node should be able to perform the following functions: First, the multihomed HIP nodes must be able to convey the locally available locator set to the peer. Second, the nodes should be able to monitor the communication and detect failures. In case that a failure is detected, they must be able to discover alternative working locator pairs and divert the communication through the alternative locator pair. It should be noted that for the provision of basic fault tolerance capabilities, the locators are only used sequentially and not in parallel. This is so, because as long a locator pair is working, the peers stick to that pair for exchanging data packets and they only change the locator pair used when there is a failure. This is different from the general multihoming scenario considered in [RFC5206] since locator pairs are not used in parallel. This particular constraint reduces considerably the possibility of packet reordering and hence the possibility of having problems with the reply protection window due to reordering of packets that travel through different paths.

In the general multihoming scenario defined in [RFC5206], a multihomed node is recommended to create different SAs and use different SPIs for interface available for the communication between two multihomed nodes to avoid problems with the anti-replay protection window resulting from reordering packets when using multiple paths simultaneously. While this is required for the general multihoming scenario, this is somewhat expensive approach, because it requires a high number of SAs to be created and it also requires some signaling overhead. Basically in a multihoming scenario where a multihomed node A that has m interfaces is communicating with another multihomed node B that has n interfaces, they need to exchange m+n UPDATE messages to convey all the locator information. This is so, because they need to convey SPI information for each of the interface pairs. Node A does so by sending an n UPDATE messages. While all the overhead and complexity is required when using multiple interface pairs in parallel, this is not the case for a fault tolerant configuration, where the locator pairs will be used sequentially.

In order to support fault tolerance, the following behaviour is defined for HIP nodes. Each node conveys the available locator set information to the peer in a single UPDATE message. The Old SPI value of the ESP_INFO parameter are equal to the current SPI value. Each node uses a single SA and a single SPI value for all the locator

pairs available for the configuration. Only a single locator pair is active, and all the traffic is sent using the active locator pair. Upon the reception of one UPDATE message containing multiple locators with a single SPI value for all the locators, the receiver verifies the locators contained in the UPDATE message. After that, the receiver identifies that it is in the fault tolerance scenario and creates locator pairs using all the received locators and all the locally available locators, irrespectively of the locator to which the UPDATE message was sent. The result is that each of the peers has all the locator pairs available for use in case that a failure occurs.

For simultaneous multihoming, an end-host should not assign locators that are assigned in different interfaces to a single SPI value. Instead, the host should acquire an SPI value value for each interface separately. Each end-host conveys the available locator set information to the peer in a separate UPDATE message. Upon the reception of UPDATE message containing multiple locators with a Old SPI value zero and the New SPI non-zero for all the locators, the receiver verifies the locators contained in the UPDATE message and acquire an SPI value value for these locators. The result is that each of the peers has multiple locator pairs available for use to transfer the traffic between the hosts.

## 5. Packet Processing

This section describes general packet sending and processing procedures in the different NAT traversal scenarios.

### 5.1. Handover Procedures

This section describes the handover procedures using NAT traversal techniques. In order to notify the peer nodes of changed locator(s), an end-host MUST execute following steps, summarized below at a high level:

1. UPDATE its location to the Relay Server(s)

2. Update bindings to TURN server(s)

3. Gather new unreflexive, reflexive and relayed-transport candidates

4. Exchange Offer and Answer with its peer nodes

5. Execute connectivity and optionally return-routability checks

6. Set Preferred bit to zero for all locators.

### 5.1.1. HIP Relay Server Update and Gathering New Candidates

The Relay Client communicates its changes in its locators to its Relay Server. Otherwise, other hosts trying to communicate with the Relay Client may fail to contact it.

[I-D.ietf-hip-nat-traversal] recommends that the HIP Relay does not include NAT traversal mode parameter in the base exchange. As a consequence, HIP control plane operates over UDP, but HIP Relay Client and Server do not use ICE for connectivity tests. Therefore, the Relay Client MUST use UPDATE to inform its Relay server(s) on its new locators as defined in [RFC5206] except that the Client follows the UDP encapsulation procedures for type 2 locators as described in [I-D.ietf-hip-nat-traversal].

As an alternative to STUN, host MAY use the UPDATE packet to gather the server reflexive addresses from the Relay server. The Mobile Node sends a UPDATE packet containing REG_REQUEST parameter registering to the Relay service. The Relay acknowledges registration with REG_RESPONSE and REG_FROM parameters. The same procedure is used to update the registration lifetime in [RFC5203]. Figure 4 illustrated address gathering procedure combined with location UPDATE.

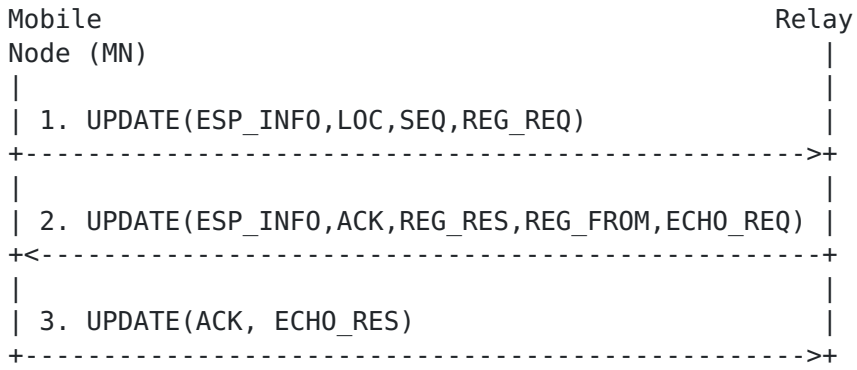


Figure 4: Updating Relay Server Combined with Gathering Addresses

Steps 2 and 3 are repeated for each locator contained in the LOCATOR parameter (LOC in the figure).

### 5.1.2. Handover Procedure with ICE

When there a changes in the locators of the MN, it communicates its new LOCATOR set to its CNs. To reduce the latency of the handover, the MN MAY do this in parallel with updating and gathering new

candidates from its Relay Server as described in Section 5.1.1. The MN learns its peer reflexive transport locator during the handover procedure and therefore gathering the server reflexive transport locator is not necessary. The details of the handover are

End-hosts that have negotiated successfully the ICE-STUN-UDP mode during the base exchange, use the HIP UPDATE packet to exchange the ICE offer and answer when a locator change is detected as illustrated in Figure 5. The UPDATE packet contains a LOCATOR parameter containing unreflexive, reflexive and relayed transport locators of the Mobile Node (MN). In steps 1 and 2, the MN sends the UPDATE packet through the relay server that was previously used for the base exchange. The Corresponding Node (CN) responds to the UPDATE with another UPDATE packet in steps 3 and 4. It contains a LOCATOR parameter listing unreflexive, reflexive and relayed transport locators of the CN. The MN completes the procedure by acknowledging the sequence number in steps 5 and 6. Finally, the end-hosts start the ICE connectivity checks direcly with each other in step 7 as described in [I-D.ietf-hip-nat-traversal]. The end-hosts set up a pair of IPsec SA for each successfully tested address pair. In the case of failure, the end-hosts send a NOTIFY through the relay to each other.

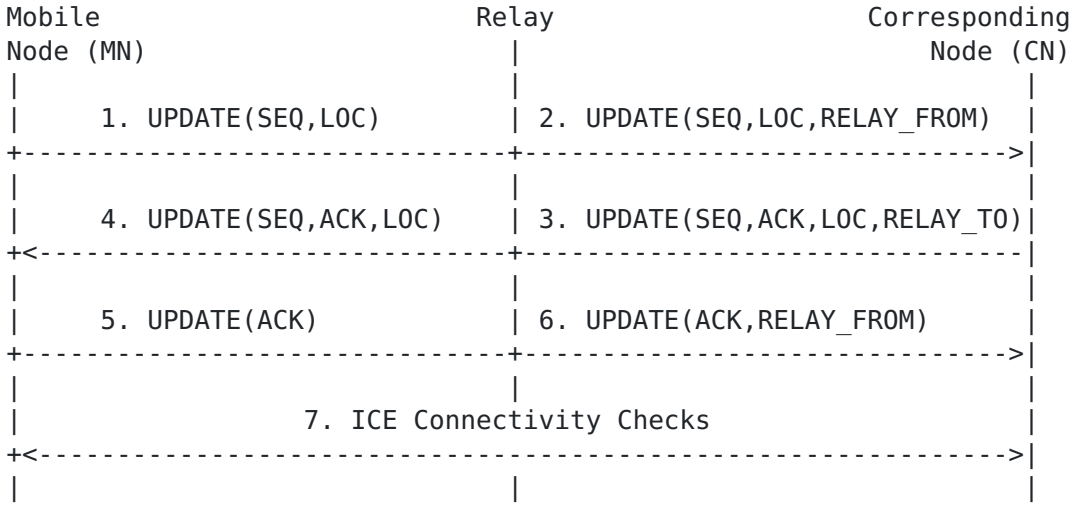


Figure 5: Handover with ICE

### 5.1.3. Handover Procedure without ICE

End-hosts that have negotiated UDP-ENCAPSULATION mode during the base exchange do not use ICE, but instead follow the procedures in [RFC5206]. However, the LOCATOR parameter may include type 2 locators and MUST be sent over transport layer. The return routability tests are established with or without transport layer

encapsulation according to the type of the locator being tested. It should be noticed that this mode has limited applicability, i.e., the return routability checks succeed when only the mobile node is behind a NAT.

### 5.1.4. Connectivity checks

After the hosts have exchanged the candidate pairs, they will start the connectivity checks for each candidate pair one at a time in a specific priority order. The connectivity checks proceed sequentially with paces between the checks to avoid network flooding. The pacing of connectivity checks and the priority order are defined in [I-D.ietf-hip-nat-traversal].

In order to recover faster from the data plane disconnectivity, the mobile node MAY initiate a return routability test immediately through its TURN media relay. This allows the mobile node to restore data plane connectivity in parallel with ICE connectivity checks which may take a while to complete. Further, to facilitate faster recovery, successfully tested address pairs MAY be taken into use immediately instead of waiting for checks for all addresses to be completed in regular ICE nomination.

## 5.2. E2E Failure Detection Mechanism

As described in the [I-D.ietf-hip-nat-traversal], the keepalives between HIP end-host and TURN server are STUN Binding Indications. Similarly, the keepalives are STUN Binding Indications for two HIP hosts that have negotiated ICE-STUN-UDP as the nat traversal mode. Keepalives for two HIP hosts operating in UDP-ENCAPSULATION mode use HIP NOTIFY messages as keepalives. Keepalive are send in periods of 20 seconds, but MUST be omitted if some other traffic (e.g. ESP) occupies the corresponding transport-layer connection. The absence of keepalives and ESP packets are used to detect end-to-end or end-to-middle failures according to timeouts based on local policies.

# 6. Packet Formats

TBD.

# 7. Security Considerations

None yet.

## 8. Acknowlegements

Thanks for Ari Keraenen for comments.

## 9. Normative References

[I-D.ietf-hip-nat-traversal]
Komu, M., Henderson, T., Matthews, P., Tschofenig, H., and A. Keraenen, "Basic HIP Extensions for Traversal of Network Address Translators", draft-ietf-hip-nat-traversal-04 (work in progress), July 2008.

[I-D.ietf-mmusic-ice]
Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", draft-ietf-mmusic-ice-19 (work in progress), October 2007.

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

[RFC5203] Laganier, J., Koponen, T., and L. Eggert, "Host Identity Protocol (HIP) Registration Extension", RFC 5203, April 2008.

[RFC5206] Nikander, P., Henderson, T., Vogt, C., and J. Arkko, "End-Host Mobility and Multihoming with the Host Identity Protocol", RFC 5206, April 2008.

## Appendix A. Document Revision History

To be removed upon publication

| Revision | Comments |
|---|---|
| draft-00 | Initial version. |

Authors' Addresses

Jan Melen
Ericsson Research Nomadiclab
Hirsalantie 11
02420 Jorvas
Finland

Phone: +358 9 2991
Email: jan.melen@ericsson.com

Miika Komu
Helsinki Institute for Information Technology
Metsanneidonkuja 4
Espoo
Finland

Phone: +358503841531
Fax: +35896949768
Email: miika@iki.fi
URI: http://www.hiit.fi/

Marcelo Bagnulo
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: +34 91 6249500
Email: marcelo@it.uc3m.es

Thomas Henderson
The Boeing Company
P.O. Box 3707
Seattle, WA
USA

Email: thomas.r.henderson@boeing.com

## Full Copyright Statement



## Intellectual Property



## Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).