Internet Research Task Force (IRTF)
Request for Comments: 6252
Category: Informational
ISSN: 2070-1721

A. Dutta, Ed.
V. Fajardo
NIKSUN
Y. Ohba
K. Taniuchi
Toshiba
H. Schulzrinne
Columbia Univ.

June 2011

A Framework of Media-Independent Pre-Authentication (MPA) for Inter-Domain Handover Optimization

Abstract

This document describes Media-independent Pre-Authentication (MPA), a new handover optimization mechanism that addresses the issues on existing mobility management protocols and mobility optimization mechanisms to support inter-domain handover. MPA is a mobile-assisted, secure handover optimization scheme that works over any link layer and with any mobility management protocol, and is most applicable to supporting optimization during inter-domain handover. MPA’s pre-authentication, pre-configuration, and proactive handover techniques allow many of the handoff-related operations to take place before the mobile node has moved to the new network. We describe the details of all the associated techniques and their applicability for different scenarios involving various mobility protocols during inter-domain handover. We have implemented the MPA mechanism for various network-layer and application-layer mobility protocols, and we report a summary of experimental performance results in this document.

This document is a product of the IP Mobility Optimizations (MOBOPTS) Research Group.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Research Task Force (IRTF). The IRTF publishes the results of Internet-related research and development activities. These results might not be suitable for deployment. This RFC represents the consensus of the MOBOPTS Research Group of the Internet Research Task Force (IRTF). Documents approved for publication by the IRSG are not a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Copyright Notice

Table of Contents

1. Introduction

As wireless technologies, including cellular and wireless LANs, are becoming popular, supporting terminal handovers across different types of access networks, such as from a wireless LAN to CDMA or to General Packet Radio Service (GPRS), is considered a clear challenge. On the other hand, supporting seamless terminal handovers between access networks of the same type is still more challenging, especially when the handovers are across IP subnets or administrative domains. To address those challenges, it is important to provide terminal mobility that is agnostic to link-layer technologies in an optimized and secure fashion without incurring unreasonable complexity. In this document, we discuss a framework to support terminal mobility that provides seamless handovers with low latency and low loss. Seamless handovers are characterized in terms of performance requirements as described in Section 1.2. [MPA-WIRELESS] is an accompanying document that describes implementation of a few MPA-based systems, including performance results to show how existing protocols could be leveraged to realize the functionalities of MPA.

Terminal mobility is accomplished by a mobility management protocol that maintains a binding between a locator and an identifier of a mobile node, where the binding is referred to as the mobility binding. The locator of the mobile node may dynamically change when there is a movement of the mobile node. The movement that causes a change of the locator may occur when there is a change in attachment point due to physical movement or network change. A mobility management protocol may be defined at any layer. In the rest of this document, the term "mobility management protocol" refers to a mobility management protocol that operates at the network layer or higher.

There are several mobility management protocols at different layers. Mobile IP [RFC5944] and Mobile IPv6 [RFC3775] are mobility management protocols that operate at the network layer. Similarly, MOBIKE (IKEv2 Mobility and Multihoming) [RFC4555] is an extension to the Internet Key Exchange Protocol (IKEv2) that provides the ability to deal with a change of an IP address of an IKEv2 end-point. There are several ongoing activities in the IETF to define mobility management protocols at layers higher than the network layer. HIP (Host Identity Protocol) [RFC5201] defines a new protocol layer between the network layer and transport layer to provide terminal mobility in a way that is transparent to both the network layer and transport layer. Also, SIP-based mobility is an extension to SIP to maintain the mobility binding of a SIP user agent [SIPMM].

While mobility management protocols maintain mobility bindings, these cannot provide seamless handover if used in their current form. An additional optimization mechanism is needed to prevent the loss of in-flight packets transmitted during the mobile node’s binding update procedure and to achieve seamless handovers. Such a mechanism is referred to as a mobility optimization mechanism. For example, mobility optimization mechanisms for Mobile IPv4 [RFC4881] and Mobile IPv6 [RFC5568] are defined to allow neighboring access routers to communicate and carry information about mobile terminals. There are protocols that are considered as "helpers" of mobility optimization mechanisms. The CARD (Candidate Access Router Discovery) protocol [RFC4066] is designed to discover neighboring access routers. CXTP (Context Transfer Protocol) [RFC4067] is designed to carry state that is associated with the services provided for the mobile node, or context, among access routers. In Section 4, we describe some of the fast-handover schemes that attempt to reduce the handover delay.

There are several issues in existing mobility optimization mechanisms. First, existing mobility optimization mechanisms are tightly coupled with specific mobility management protocols. For example, it is not possible to use mobility optimization mechanisms designed for Mobile IPv4 or Mobile IPv6 with MOBIKE. What is strongly desired is a single, unified mobility optimization mechanism that works with any mobility management protocol. Second, there is no existing mobility optimization mechanism that easily supports handovers across administrative domains without assuming a pre-established security association between administrative domains.

A mobility optimization mechanism should work across administrative domains in a secure manner only based on a trust relationship between a mobile node and each administrative domain. Third, a mobility optimization mechanism needs to support not only terminals with multiple interfaces where simultaneous connectivity through multiple interfaces or connectivity through a single interface can be expected, but also terminals with a single interface.

This document describes a framework of Media-independent Pre-Authentication (MPA), a new handover optimization mechanism that addresses all those issues. MPA is a mobile-assisted, secure handover optimization scheme that works over any link layer and with any mobility management protocol, including Mobile IPv4, Mobile IPv6, MOBIKE, HIP, and SIP mobility. In cases of multiple operators without a roaming relationship or without an agreement to participate in a key management scheme, MPA provides a framework that can perform pre-authentication to establish the security mechanisms without assuming a common source of trust. In MPA, the notion of IEEE 802.11i pre-authentication is extended to work at a higher layer, with additional mechanisms to perform early acquisition of an IP address from a network where the mobile node may move, as well as proactive handover to the network while the mobile node is still attached to the current network. Since this document focuses on the MPA framework, it is left to future work to choose the protocols for MPA and define detailed operations. The accompanying document [MPA-WIRELESS] provides one method that describes usage and interactions between existing protocols to accomplish MPA functionality.

This document represents the consensus of the IP Mobility Optimizations (MOBOPTS) Research Group. It has been reviewed by Research Group members active in the specific area of work.

1.1. Specification of Requirements

In this document, several words are used to signify the requirements of the specification. These words are often capitalized. 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].

1.2. Performance Requirements

In order to provide desirable quality of service for interactive Voice over IP (VoIP) and streaming traffic, one needs to limit the value of end-to-end delay, jitter, and packet loss to a certain threshold level. ITU-T and ITU-E standards define the acceptable values for these parameters. For example, for one-way delay, ITU-T

G.114 [RG98] recommends 150 ms as the upper limit for most of the applications, and 400 ms as generally unacceptable delay. One-way delay tolerance for video conferencing is in the range of 200 to 300 ms [ITU98]. Also, if an out-of-order packet is received after a certain threshold, it is considered lost. According to ETSI TR 101 [ETSI], a normal voice conversation can tolerate up to 2% packet loss. But this is the mean packet loss probability and may be applicable to a scenario when the mobile node is subjected to repeated handoff during a normal conversation. Measurement techniques for delay and jitter are described in [RFC2679], [RFC2680], and [RFC2681].

In the case of interactive VoIP traffic, end-to-end delay affects the jitter value, and thus is an important issue to consider. An end-to-end delay consists of several components, such as network delay, operating system (OS) delay, codec delay, and application delay. A complete analysis of these delays can be found in [WENYU]. During a mobile node’s handover, in-flight transient traffic cannot reach the mobile node because of the associated handover delay. These in-flight packets could either be lost or buffered. If the in-flight packets are lost, this packet loss will contribute to jitter between the last packet before handoff and the first packet after handoff. If these packets are buffered, packet loss is minimized, but there is additional jitter for the in-flight packets when these are flushed after the handoff. Buffering during handoff avoids the packet loss, but at the cost of additional one-way delay. A tradeoff between one-way delay and packet loss is desired based on the type of application. For example, for a streaming application, packet loss can be reduced by increasing the playout buffer, resulting in longer one-way packet delay.

The handover delay is attributed to several factors, such as discovery, configuration, authentication, binding update, and media delivery. Many of the security-related procedures, such as handover keying and re-authentication procedures, deal with cases where there is a single source of trust at the top, and the underlying Authentication, Authorization, and Accounting (AAA) domain elements trust the top source of trust and the keys it generates and distributes. In this scenario, there is an appreciable delay in re-establishing link-security-related parameters, such as authentication, link key management, and access authorization during inter-domain handover. The focus of this document is the design of a framework that can reduce the delay due to authentication and other handoff-related operations such as configuration and binding update.

2. Terminology

Mobility Binding: A binding between a locator and an identifier of a mobile terminal.

Mobility Management Protocol (MMP): A protocol that operates at the network layer or above to maintain a binding between a locator and an identifier of a mobile node.

Binding Update (BU): A procedure to update a mobility binding.

Media-independent Pre-Authentication Mobile Node (MN): A mobile node using Media-independent Pre-Authentication (MPA). MPA is a mobile-assisted, secure handover optimization scheme that works over any link layer and with any mobility management protocol. An MPA mobile node is an IP node. In this document, the term "mobile node" or "MN" without a modifier refers to "MPA mobile node". An MPA mobile node usually has a functionality of a mobile node of a mobility management protocol as well.

Candidate Target Network (CTN): A network to which the mobile node may move in the near future.

Target Network (TN): The network to which the mobile node has decided to move. The target network is selected from one or more candidate target networks.

Proactive Handover Tunnel (PHT): A bidirectional IP tunnel [RFC2003] [RFC2473] that is established between the MPA mobile node and an access router of a candidate target network. In this document, the term "tunnel" without a modifier refers to "proactive handover tunnel".

Point of Attachment (PoA): A link-layer device (e.g., a switch, an access point, or a base station) that functions as a link-layer attachment point for the MPA mobile node to a network.

Care-of Address (CoA): An IP address used by a mobility management protocol as a locator of the MPA mobile node.

3. Handover Taxonomy

Based on the type of movement, type of access network, and underlying mobility support, one can primarily define the handover as inter-technology, intra-technology, inter-domain, and intra-domain. We describe briefly each of these handover processes. However, our focus of the discussion is on inter-domain handover.

Inter-technology: A mobile node may be equipped with multiple interfaces, where each interface can support a different access technology (e.g., 802.11, CDMA). A mobile node may communicate with one interface at any time in order to conserve power. During the handover, the mobile node may move out of the footprint of one access technology (e.g., 802.11) and move into the footprint of a different access technology (e.g., CDMA). This will warrant switching of the communicating interface on the mobile node as well. This type of inter-technology handover is often called "vertical handover", since the mobile node moves between two different cell sizes.

Intra-technology: An intra-technology handover is defined as when a mobile node moves within the same type of access technology, such as between 802.11[a,b,n] and 802.11 [a,b,n] or between CDMA1XRTT and CDMA1EVDO. In this scenario, a mobile node may be equipped with a single interface (with multiple PHY types of the same technology) or with multiple interfaces. An intra-technology handover may involve intra-subnet or inter-subnet movement and thus may need to change its L3 locator, depending upon the type of movement.

Inter-domain: A domain can be defined in several ways. But for the purposes of roaming, we define "domain" as an administrative domain that consists of networks managed by a single administrative entity that authenticates and authorizes a mobile node for accessing the networks. An administrative entity may be a service provider, an enterprise, or any organization. Thus, an inter-domain handover will by default be subjected to inter-subnet handover, and in addition it may be subjected to either inter-technology or intra-technology handover. A mobile node is subjected to inter-subnet handover when it moves from one subnet (broadcast domain) to another subnet (broadcast domain). Inter-domain handover will be subjected to all the transition steps a subnet handover goes through, and it will be subjected to authentication and authorization processes as well. It is also likely that the type of mobility support in each administrative domain will be different. For example, administrative domain A may have Mobile IP version 6 (MIPv6) support, while administrative domain B may use Proxy MIPv6 [RFC5213].

Intra-domain: When a mobile node’s movement is confined to movement within an administrative domain, it is called "intra-domain movement". An intra-domain movement may involve intra-subnet, inter-subnet, intra-technology, and inter-technology as well.

Both inter-domain and intra-domain handovers can be subjected to either inter-technology or intra-technology handover based on the network access characteristics. Inter-domain handover requires authorization for acquisition or modification of resources assigned to a mobile node, and the authorization needs interaction with a central authority in a domain. In many cases, an authorization procedure during inter-domain handover follows an authentication procedure that also requires interaction with a central authority in a domain. Thus, security associations between the network entities, such as routers in the neighboring administrative domains, need to be established before any interaction takes place between these entities. Similarly, an inter-domain mobility may involve different mobility protocols, such as MIPv6 and Proxy MIPv6, in each of its domains. In that case, one needs a generalized framework to achieve the optimization during inter-domain handover. Figure 1 shows a typical example of inter-domain mobility involving two domains, domain A and domain B. It illustrates several important components, such as a AAA Home server (AAAH); AAA visited servers (e.g., AAAV1 and AAAV2); an Authentication Agent (AA); a layer 3 point of attachment, such as an Access Router (AR); and a layer 2 point of attachment, such as an Access Point (AP). Any mobile node may be using a specific mobility protocol and associated mobility optimization technique during intra-domain movement in either domain. But the same optimization technique may not be suitable to support inter-domain handover, independent of whether it uses the same or a different mobility protocol in either domain.

```
                        +-----------------------------+
                        |      +--------+             |
                        |      |        |             |
                        |      | AAAH   ------------------|
                        |      |        |             |   |
                        |      +|-------+             |   |
                        |       |                     |   |
                        |       |  Home Domain        |   |
                        |       |                     |   |
                        +-------|---------------------+   |
                                |                         |
                                |                         |
                                |                         |
   +----------------------------|---------+ +-------------|------------+
   | Domain A                   |         | | Domain B    |            |
   |                            |         | |            +|-------+    |
   |                    +-------|+        | | +-----+    |        |    |
   |                    |        |        | | |     ------ AAAV2  |    |
   |                    | AAAV1  |        | | | AA  |    |        |    |
   |      +--------------        |        | | +|----+    +--------+    |
   |      |     |       +--------+        | |  |                       |
   |      |AA   |                         | |  |---         ----       |
   |      +--|--+                         | | /    \       /    \      |
   |         |              /----\        | || AR   |-----| AR   |     |
   |        -|--           /      \       | | \    /       \    /      |
   |       /    \         | AR     |      | |  -|--         --|-       |
   |      | AR   -----------      /       | |+--|---+  +------|------+ |
   |       \    /           \--|-/        | || AP4  |  |  L2 Switch  | |
   |        -/--         +-----|------+   | ||      |  +-|---------|-+ |
   |        /            |  L2 Switch |   | |+------+    |         |   |
   |       /             +-|-------|--+   | |        +---|--+ +----|-+ |
   | +----/-+         +----|-+   +-|----+ | |        |      | |      | |
   | |      |         |      |   |      | | |        | AP5  | |AP6   | |
   | | AP1  |         | AP2  |   | AP3  | | |        +----|-+ +------+ |
   | +------+         +------+   +--|---+ | |             |            |
   +--------------------------------|-----+ +------------ |------------+
                                  --|---------            |
                              ////            \\\\   -----|-----
                            //    +------+       ////  +------+ \\\\
                            |     | MN   ------------->|MN  |     \\\
                           |      |      |    |     |  |      |       |
                            |     +------+   |     |   +------+        |
                            \\                |   //                  |
                              \\\\            \\\/                  ///
                                  ------------   \\\\------------- ////
```

Figure 1: Inter-Domain Mobility

4. Related Work

While basic mobility management protocols such as Mobile IP [RFC5944], Mobile IPv6 [RFC3775], and SIP-Mobility [SIPMM] provide continuity to TCP and RTP traffic, these are not optimized to reduce the handover latency during a mobile node’s movement between subnets and domains. In general, these mobility management protocols introduce handover delays incurred at several layers, such as layer 3 and the application layer, for updating the mobile node’s mobility binding. These protocols are affected by underlying layer 2 delay as well. As a result, applications using these mobility protocols suffer from performance degradation.

There have been several optimization techniques that apply to current mobility management schemes that try to reduce handover delay and packet loss during a mobile node’s movement between cells, subnets, and domains. Micro-mobility management schemes such as [CELLIP] and [HAWAII], and intra-domain mobility management schemes such as [IDMP], [MOBIP-REG], and [RFC5380], provide fast handover by limiting the signaling updates within a domain. Fast Mobile IP protocols for IPv4 and IPv6 networks [RFC4881] [RFC5568] utilize mobility information made available by link-layer triggers. Yokota et al. [YOKOTA] propose the joint use of an access point and a dedicated Media Access Control (MAC) bridge to provide fast handover without altering the MIPv4 specification. Shin et al. [MACD] propose a scheme that reduces the delay due to MAC-layer handoff by providing a cache-based algorithm. In this scheme, the mobile node caches the neighboring channels that it has already visited and thus uses a selective scanning method. This helps to reduce the associated scanning time.

Some mobility management schemes use dual interfaces, thus providing make-before-break [SUM]. In a make-before-break situation, communication usually continues with one interface when the secondary interface is in the process of getting connected. The IEEE 802.21 working group is discussing these scenarios in detail [802.21]. Providing fast handover using a single interface needs more careful design than for a client with multiple interfaces. Dutta et al. [SIPFAST] provide an optimized handover scheme for SIP-based mobility management, where the transient traffic is forwarded from the old subnet to the new one by using an application-layer forwarding scheme. [MITH] provides a fast-handover scheme for the single-interface case that uses mobile-initiated tunneling between the old Foreign Agent and a new Foreign Agent. [MITH] defines two types of handover schemes: Pre-MIT (Mobile Initiated Tunneling) and Post-MIT (Media Initiated Tunneling). The proposed MPA scheme is very similar to Mobile Initiated Tunneling Handoff’s (MITH’s) predictive scheme, where the mobile node communicates with the

Foreign Agent before actually moving to the new network. However, the MPA scheme is not limited to MIP; this scheme takes care of movement between domains and performs pre-authentication in addition to proactive handover. Thus, MPA reduces the overall delay to a period close to that of link-layer handover delay. Most of the mobility optimization techniques developed so far are restricted to a specific type of mobility protocol only. While supporting optimization for inter-domain mobility, these protocols assume that there is a pre-established security arrangement between two administrative domains. But this assumption may not always be viable. Thus, there is a need to develop an optimization mechanism that can support inter-domain mobility without any underlying constraints or security-related assumptions.

Recently, the HOKEY working group within the IETF has been defining ways to expedite the authentication process. In particular, it has defined pre-authentication [RFC5836] and fast re-authentication [RFC5169] mechanisms to expedite the authentication and security association process.

5. Applicability of MPA

MPA is more applicable where an accurate prediction of movement can be easily made. For other environments, special care must be taken to deal with issues such as pre-authentication to multiple CTNs (Candidate Target Networks), and failed switching and switching back as described in [MPA-WIRELESS]. However, addressing those issues in actual deployments may not be easier. Some of the deployment issues are described in Appendix C.

The authors of the accompanying document [MPA-WIRELESS] have cited several use cases of how MPA can be used to optimize several network-layer and application-layer mobility protocols. The effectiveness of MPA may be relatively reduced if the network employs network-controlled localized mobility management in which the MN does not need to change its IP address while moving within the network. The effectiveness of MPA may also be relatively reduced if signaling for network access authentication is already optimized for movements within the network, e.g., when simultaneous use of multiple interfaces during handover is allowed. In other words, MPA is more viable as a solution for inter-administrative domain predictive handover without the simultaneous use of multiple interfaces. Since MPA is not tied to a specific mobility protocol, it is also applicable to support optimization for inter-domain handover where each domain may be equipped with a different mobility protocol.

Figure 1 shows an example of inter-domain mobility where MPA could be applied. For example, domain A may support just Proxy MIPv6, whereas domain B may support Client Mobile IPv6. MPA’s different functional components can provide the desired optimization techniques proactively.

6. MPA Framework

6.1. Overview

Media-independent Pre-Authentication (MPA) is a mobile-assisted, secure handover optimization scheme that works over any link layer and with any mobility management protocol. With MPA, a mobile node is not only able to securely obtain an IP address and other configuration parameters for a CTN, but also able to send and receive IP packets using the IP address obtained before it actually attaches to the CTN. This makes it possible for the mobile node to complete the binding update of any mobility management protocol and use the new CoA before performing a handover at the link layer.

MPA adopts the following basic procedures to provide this functionality. The first procedure is referred to as "pre-authentication", the second procedure is referred to as "pre-configuration", and the combination of the third and fourth procedures is referred to as "secure proactive handover". The security association established through pre-authentication is referred to as an "MPA-SA".

This functionality is provided by allowing a mobile node that has connectivity to the current network, but is not yet attached to a CTN, to

(i) establish a security association with the CTN to secure the subsequent protocol signaling, then

(ii) securely execute a configuration protocol to obtain an IP address and other parameters from the CTN as well as execute a tunnel management protocol to establish a Proactive Handover Tunnel (PHT) [RFC2003] between the mobile node and an access router of the CTN, then

(iii) send and receive IP packets, including signaling messages for the binding update of an MMP and data packets transmitted after completion of the binding update, over the PHT, using the obtained IP address as the tunnel inner address, and finally

(iv) delete or disable the PHT immediately before attaching to the CTN when it becomes the target network, and then re-assign the inner address of the deleted or disabled tunnel to its physical interface immediately after the mobile node is attached to the target network through the interface. Instead of deleting or disabling the tunnel before attaching to the target network, the tunnel may be deleted or disabled immediately after being attached to the target network.

Step (iii) above (i.e., the binding update procedure), in particular, makes it possible for the mobile node to complete the higher-layer handover before starting a link-layer handover. This means that the mobile node is able to send and receive data packets transmitted after completing the binding update over the tunnel, while data packets transmitted before completion of the binding update do not use the tunnel.

6.2. Functional Elements

In the MPA framework, the following functional elements are expected to reside in each CTN to communicate with a mobile node: an Authentication Agent (AA), a Configuration Agent (CA), and an Access Router (AR). These elements can reside in one or more network devices.

An authentication agent is responsible for pre-authentication. An authentication protocol is executed between the mobile node and the authentication agent to establish an MPA-SA. The authentication protocol MUST be able to establish a shared key between the mobile node and the authentication agent and SHOULD be able to provide mutual authentication. The authentication protocol SHOULD be able to interact with a AAA protocol, such as RADIUS or Diameter, to carry authentication credentials to an appropriate authentication server in the AAA infrastructure. This interaction happens through the authentication agent, such as the PANA Authentication Agent (PAA). In turn, the derived key is used to derive additional keys that will be applied to protecting message exchanges used for pre-configuration and secure proactive handover. Other keys that are used for bootstrapping link-layer and/or network-layer ciphers MAY also be derived from the MPA-SA. A protocol that can carry the Extensible Authentication Protocol (EAP) [RFC3748] would be suitable as an authentication protocol for MPA.

A configuration agent is responsible for one part of pre-configuration, namely securely executing a configuration protocol to deliver an IP address and other configuration parameters to the mobile node. The signaling messages of the configuration protocol (e.g., DHCP) MUST be protected using a key derived from the key corresponding to the MPA-SA.

An access router in the MPA framework is a router that is responsible for the other part of pre-configuration, i.e., securely executing a tunnel management protocol to establish a proactive handover tunnel to the mobile node. IP packets transmitted over the proactive handover tunnel SHOULD be protected using a key derived from the key corresponding to the MPA-SA. Details of this procedure are described in Section 6.3.

Figure 2 shows the basic functional components of MPA.

```
                                        +----+
                                        | CN |
                                        +----+
                                         /
                              (Core Network)
                             /              \
                            /                \
          +----------------/--------+    +----\-----------------+
          | +-----+                 |    |+-----+               |
          | |     |        +-----+  |    ||     |       +-----+ |
          | | AA  |        |CA   |  |    ||AA   |       | CA  | |
          | +--+--+        +--+--+  |    |+--+--+       +--+--+ |
          |    |   +------+   |     |    |   | +-----+    |     |
          |    |   | pAR  |   |     |    |   | |nAR  |    |     |
          | ---+---+      +---+-----+----+---+-+     +-----+    |
          |        +---+--+         |    |     +-----+          |
          |            |            |    |                      |
          |            |            |    |                      |
          |            |            |    |                      |
          +------------+------------+    +--------|-------------+
          Current      |                 Candidate| Target Network
          Network      |                          |
                    +------+                  +------+
                    | oPoA |                  | nPoA |
                    +--.---+                  +--.---+
                       .                         .
                       .                         .
                    +------+
                    |  MN  |  ---------->
                    +------+
```

Figure 2: MPA Functional Components

6.3. Basic Communication Flow

Assume that the mobile node is already connected to a point of attachment, say oPoA (old point of attachment), and assigned a care-of address, say oCoA (old care-of address). The communication flow of MPA is described as follows. Throughout the communication flow, data packet loss should not occur except for the period during the switching procedure in Step 5 below, and it is the responsibility of link-layer handover to minimize packet loss during this period.

Step 1 (pre-authentication phase): The mobile node finds a CTN through some discovery process, such as IEEE 802.21, and obtains the IP addresses of an authentication agent, a configuration agent, and an access router in the CTN (Candidate Target Network) by some means. Details about discovery mechanisms are discussed in Section 7.1. The mobile node performs pre-authentication with the authentication agent. As discussed in Section 7.2, the mobile node may need to pre-authenticate with multiple candidate target networks. The decision regarding with which candidate network the mobile node needs to pre-authenticate will depend upon several factors, such as signaling overhead, bandwidth requirement (Quality of Service (QoS)), the mobile node’s location, communication cost, handover robustness, etc. Determining the policy that decides the target network with which the mobile node should pre-authenticate is out of scope for this document.

If the pre-authentication is successful, an MPA-SA is created between the mobile node and the authentication agent. Two keys are derived from the MPA-SA, namely an MN-CA key and an MN-AR key, which are used to protect subsequent signaling messages of a configuration protocol and a tunnel management protocol, respectively. The MN-CA key and the MN-AR key are then securely delivered to the configuration agent and the access router, respectively.

Step 2 (pre-configuration phase): The mobile node realizes that its point of attachment is likely to change from the oPoA to a new one, say nPoA (new point of attachment). It then performs pre-configuration with the configuration agent, using the configuration protocol to obtain several configuration parameters such as an IP address, say nCoA (new care-of address), and a default router from the CTN. The mobile node then communicates with the access router using the tunnel management protocol to establish a proactive handover tunnel. In the tunnel management protocol, the mobile node registers the oCoA and the nCoA as the tunnel outer address and the tunnel inner address, respectively. The signaling messages of the pre-configuration protocol are protected using the MN-CA key and the MN-AR key. When the configuration agent and the access router are co-located in the same device, the two protocols may be integrated into a single protocol, such as IKEv2. After completion of the tunnel establishment, the mobile node is able to communicate using both the oCoA and the nCoA by the end of Step 4. A configuration protocol and a tunnel management protocol may be combined in a single protocol or executed in different orders depending on the actual protocol(s) used for configuration and tunnel management.

Step 3 (secure proactive handover main phase): The mobile node decides to switch to the new point of attachment by some means. Before the mobile node switches to the new point of attachment, it starts secure proactive handover by executing the binding update operation of a mobility management protocol and transmitting subsequent data traffic over the tunnel (main phase). This proactive binding update could be triggered based on certain local policy at the mobile node end, after the pre-configuration phase is over. This local policy could be Signal-to-Noise Ratio, location of the mobile node, etc. In some cases, it may cache multiple nCoA addresses and perform simultaneous binding with the Correspondent Node (CN) or Home Agent (HA).

Step 4 (secure proactive handover pre-switching phase): The mobile node completes the binding update and becomes ready to switch to the new point of attachment. The mobile node may execute the tunnel management protocol to delete or disable the proactive handover tunnel and cache the nCoA after deletion or disabling of the tunnel. This transient tunnel can be deleted prior to or after the handover. The buffering module at the next access router buffers the packets once the tunnel interface is deleted. The decision as to when the mobile node is ready to switch to the new point of attachment depends on the handover policy.

Step 5 (switching): It is expected that a link-layer handover occurs in this step.

Step 6 (secure proactive handover post-switching phase): The mobile node executes the switching procedure. Upon successful completion of the switching procedure, the mobile node immediately restores the cached nCoA and assigns it to the physical interface attached to the new point of attachment. If the proactive handover tunnel was not deleted or disabled in Step 4, the tunnel is deleted or disabled as well. After this, direct transmission of data packets using the nCoA is possible without using a proactive handover tunnel.

Call flow for MPA is shown in Figures 3 and 4.

```
                                                         IP address(es)
                                                          Available for
                                                             Use by MN
                                                                   |
                           +-----------------------------------+   |
                           |     Candidate Target Network      |   |
                           |     (Future Target Network)       |   |
             MN       oPoA | nPoA     AA        CA        AR   |   |
             |         |   |  |       |         |         |    |   |
             |         |   +-----------------------------------+   |
             |         |      |       |         |         |        .
    +---------------+  |      |       |         |         |        .
    |(1) Found a CTN|  |      |       |         |         |        .
    +---------------+  |      |       |         |         |        |
             |   Pre-authentication   |         |         |        |
             |   [authentication protocol]      |         |        |
             |<--------+------------->|MN-CA key|         |        |
             |         |      |       |-------->|MN-AR key|        |
   +-----------------+ |      |       |------------------>|        |
   |(2) Increased    | |      |       |         |         |     [oCoA]
   |chance to switch | |      |       |         |         |        |
   |     to CTN      | |      |       |         |         |        |
   +-----------------+ |      |       |         |         |        |
             |         |      |       |         |         |        |
             |   Pre-configuration    |         |         |        |
             |   [configuration protocol to get nCoA]     |        |
             |<--------+----------------------->|         |        |
             |   Pre-configuration    |         |         |        |
             |   [tunnel management protocol to establish PHT]     V
             |<--------+--------------------------------->|
             |         |      |       |         |         |        ^
   +-----------------+ |      |       |         |         |        |
   |(3) Determined   | |      |       |         |         |        |
   |to switch to CTN | |      |       |         |         |        |
   +-----------------+ |      |       |         |         |        |
             |         |      |       |         |         |        |
             |   Secure proactive handover main phase     |        |
             |   [execution of binding update of MMP and  |        |
             |    transmission of data packets through AR | [oCoA, nCoA]
             |    based on nCoA over the PHT]   |         |        |
             |<<=======+================================>+--->...  |
             .         .      .       .         .         .        .
             .         .      .       .         .         .        .
             .         .      .       .         .         .        .
```

Figure 3: Example Communication Flow (1/2)

```
             |         |      |       |         |         |        |
   +----------------+  |      |       |         |         |        |
   |(4) Completion  |  |      |       |         |         |        |
   |of MMP BU and   |  |      |       |         |         |        |
   |ready to switch |  |      |       |         |         |        |
   +----------------+  |      |       |         |         |        |
             |   Secure proactive handover pre-switching phase     |
             |   [tunnel management protocol to delete PHT]        V
             |<--------+--------------------------------->|
    +---------------+         |       |         |         |
    |(5)Switching   |         |       |         |         |
    +---------------+         |       |         |         |
             |                |       |         |         |
    +---------------+         |       |         |         |
    |(6) Completion |         |       |         |         |
    |of switching   |         |       |         |         |
    +---------------+         |       |         |         |
             o<- Secure proactive handover post-switching phase ^
             |   [Re-assignment of Tunnel Inner Address   |        |
             |                 to the physical I/F]       |        |
             |                |       |         |         |        |
             |   Transmission of data packets through AR  |     [nCoA]
             |   based on nCoA|       |         |         |        |
             |<---------------+---------------------------+-->...  |
             |                |       |         |         |        .
```

Figure 4: Example Communication Flow (2/2)

7. MPA Operations

In order to provide an optimized handover for a mobile node experiencing rapid movement between subnets and/or domains, one needs to look into several operations. These issues include:

i) discovery of neighboring networking elements,

ii) connecting to the right network based on certain policy,

iii) changing the layer 2 point of attachment,

iv) obtaining an IP address from a DHCP or PPP server,

v) confirming the uniqueness of the IP address,

vi) pre-authenticating with the authentication agent,

vii) sending the binding update to the Correspondent Host (CH),

viii) obtaining the redirected streaming traffic to the new point of attachment,

ix) ping-pong effect, and

x) probability of moving to more than one network and associating with multiple target networks.

We describe these issues in detail in the following paragraphs and describe how we have optimized these issues in the case of MPA-based secure proactive handover.

7.1. Discovery

Discovery of neighboring networking elements such as access points, access routers, and authentication servers helps expedite the handover process during a mobile node’s movement between networks. After discovering the network neighborhood with a desired set of coordinates, capabilities, and parameters, the mobile node can perform many of the operations, such as pre-authentication, proactive IP address acquisition, proactive address resolution, and binding update, while in the previous network.

There are several ways a mobile node can discover neighboring networks. The Candidate Access Router Discovery protocol [RFC4066] helps discover the candidate access routers in the neighboring networks. Given a certain network domain, SLP (Service Location Protocol) [RFC2608] and DNS help provide addresses of the networking components for a given set of services in the specific domain. In some cases, many of the network-layer and upper-layer parameters may be sent over link-layer management frames, such as beacons, when the mobile node approaches the vicinity of the neighboring networks. IEEE 802.11u is considering issues such as discovering the neighborhood using information contained in the link layer. However, if the link-layer management frames are encrypted by some link-layer security mechanism, then the mobile node may not be able to obtain the requisite information before establishing link-layer connectivity to the access point. In addition, this may add burden to the bandwidth-constrained wireless medium. In such cases, a higher-layer protocol is preferred to obtain the information regarding the neighboring elements. Some proposals, such as [802.21], help obtain information about the neighboring networks from a mobility server. When the movement is imminent, the mobile node starts the discovery process by querying a specific server and obtains the required parameters, such as the IP address of the access point, its characteristics, routers, SIP servers, or authentication servers of the neighboring networks. In the event of multiple networks, it may obtain the required parameters from more than one neighboring network and keep these in a cache. At some point, the mobile node finds several CTNs out of many probable networks and starts the pre-authentication process by communicating with the required entities in the CTNs. Further details of this scenario are in Section 7.2.

7.2. Pre-Authentication in Multiple-CTN Environment

In some cases, although a mobile node selects a specific network to be the target network, it may actually end up moving into a neighboring network other than the target network, due to factors that are beyond the mobile node's control. Thus, it may be useful to perform the pre-authentication with a few probable candidate target networks and establish time-bound transient tunnels with the respective access routers in those networks. Thus, in the event of a mobile node moving to a candidate target network other than that chosen as the target network, it will not be subjected to packet loss due to authentication and IP address acquisition delay that could occur if the mobile node did not pre-authenticate with that candidate target network. It may appear that by pre-authenticating with a number of candidate target networks and reserving the IP addresses, the mobile node is reserving resources that could be used otherwise. But since this happens for a time-limited period, it should not be a big problem; it depends upon the mobility pattern and duration. The mobile node uses a pre-authentication procedure to obtain an IP address proactively and to set up the time-bound tunnels with the access routers of the candidate target networks. Also, the MN may retain some or all of the nCoAs for future movement.

The mobile node may choose one of these addresses as the binding update address and send it to the CN (Correspondent Node) or HA (Home Agent), and will thus receive the tunneled traffic via the target network while in the previous network. But in some instances, the mobile node may eventually end up moving to a network that is other than the target network. Thus, there will be a disruption in traffic as the mobile node moves to the new network, since the mobile node has to go through the process of assigning the new IP address and sending the binding update again. There are two solutions to this problem. As one solution to the problem, the mobile node can take advantage of the simultaneous mobility binding and send multiple binding updates to the Correspondent Host or HA. Thus, the Correspondent Host or HA forwards the traffic to multiple IP addresses assigned to the virtual interfaces for a specific period of time. This binding update gets refreshed at the CH after the mobile node moves to the new network, thus stopping the flow to the other candidate networks. RFC 5648 [RFC5648] discusses different scenarios of mobility binding with multiple care-of-addresses. As the second solution, in case simultaneous binding is not supported in a specific mobility scheme, forwarding of traffic from the previous target network will help take care of the transient traffic until the new binding update is sent from the new network.

7.3. Proactive IP Address Acquisition

In general, a mobility management protocol works in conjunction with the Foreign Agent or in the co-located address mode. The MPA approach can use both the co-located address mode and the Foreign Agent address mode. We discuss here the address assignment component that is used in the co-located address mode. There are several ways a mobile node can obtain an IP address and configure itself. In some cases, a mobile node can configure itself statically in the absence of any configuration element such as a server or router in the network. In a LAN environment, the mobile node can obtain an IP address from DHCP servers. In the case of IPv6 networks, a mobile node has the option of obtaining the IP address using stateless autoconfiguration or DHCPv6. In some wide-area networking environments, the mobile node uses PPP (Point-to-Point Protocol) to obtain the IP address by communicating with a NAS (Network Access Server).

Each of these processes takes on the order of few hundred milliseconds to a few seconds, depending upon the type of IP address acquisition process and operating system of the clients and servers. Since IP address acquisition is part of the handover process, it adds to the handover delay, and thus it is desirable to reduce this delay as much as possible. There are a few optimized techniques available, such as DHCP Rapid Commit [RFC4039] and GPS-coordinate-based IP address [GPSIP], that attempt to reduce the handover delay due to IP address acquisition time. However, in all these cases, the mobile node also obtains the IP address after it moves to the new subnet and incurs some delay because of the signaling handshake between the mobile node and the DHCP server.

In Fast MIPv6 [RFC5568], through the RtSolPr and PrRtAdv messages, the MN also formulates a prospective new CoA (nCoA) when it is still present on the Previous Access Router’s (pAR’s) link. Hence, the latency due to new prefix discovery subsequent to handover is eliminated. However, in this case, both the pAR and the Next Access Router (nAR) need to cooperate with each other to be able to retrieve the prefix from the target network.

In the following paragraph, we describe a few ways that a mobile node can obtain the IP address proactively from the CTN, and the associated tunnel setup procedure. These can broadly be divided into four categories: PANA-assisted proactive IP address acquisition, IKE-assisted proactive IP address acquisition, proactive IP address acquisition using DHCP only, and stateless autoconfiguration. When DHCP is used for address configuration, a DHCP server is assumed to be serving one subnet.

7.3.1. PANA-Assisted Proactive IP Address Acquisition

In the case of PANA-assisted proactive IP address acquisition, the mobile node obtains an IP address proactively from a CTN. The mobile node makes use of PANA [RFC5191] messages to trigger the IP address acquisition process via a DHCP client that is co-located with the PANA authentication agent in the access router in the CTN acting on behalf of the mobile node. Upon receiving a PANA message from the mobile node, the DHCP client on the authentication agent performs normal DHCP message exchanges to obtain the IP address from the DHCP server in the CTN. This address is piggy-backed in a PANA message and is delivered to the mobile node. In the case of IPv6, a Router Advertisement (RA) is carried as part of the PANA message. In the case of stateless autoconfiguration, the mobile node uses the prefix(es) obtained as part of the RA and its MAC address to construct the unique IPv6 address(es) as it would have done in the new network. In the case of stateful address autoconfiguration, a procedure similar to DHCPv4 can be applied.

7.3.2. IKEv2-Assisted Proactive IP Address Acquisition

IKEv2-assisted proactive IP address acquisition works when an IPsec gateway and a DHCP relay agent [RFC3046] are resident within each access router in the CTN. In this case, the IPsec gateway and DHCP relay agent in a CTN help the mobile node acquire the IP address from the DHCP server in the CTN. The MN-AR key established during the pre-authentication phase is used as the IKEv2 pre-shared secret needed to run IKEv2 between the mobile node and the access router. The IP address from the CTN is obtained as part of the standard IKEv2 procedure, using the co-located DHCP relay agent for obtaining the IP address from the DHCP server in the target network using standard DHCP. The obtained IP address is sent back to the client in the IKEv2 Configuration Payload exchange. In this case, IKEv2 is also used as the tunnel management protocol for a proactive handover tunnel (see Section 7.4). Alternatively, a VPN gateway can dispense the IP address from its IP address pool.

7.3.3. Proactive IP Address Acquisition Using DHCPv4 Only

As another alternative, DHCP may be used for proactively obtaining an IP address from a CTN without relying on PANA or IKEv2-based approaches by allowing direct DHCP communication between the mobile node and the DHCP relay agent or DHCP server in the CTN. The mechanism described in this section is applicable to DHCPv4 only. The mobile node sends a unicast DHCP message to the DHCP relay agent or DHCP server in the CTN requesting an address, while using the address associated with the current physical interface as the source address of the request.

When the message is sent to the DHCP relay agent, the DHCP relay agent relays the DHCP messages back and forth between the mobile node and the DHCP server. In the absence of a DHCP relay agent, the mobile node can also directly communicate with the DHCP server in the target network. The broadcast option in the client’s unicast DISCOVER message should be set to 0 so that the relay agent or the DHCP server can send the reply directly back to the mobile node using the mobile node’s source address.

In order to prevent malicious nodes from obtaining an IP address from the DHCP server, DHCP authentication should be used, or the access router should be configured with a filter to block unicast DHCP messages sent to the remote DHCP server from mobile nodes that are not pre-authenticated. When DHCP authentication is used, the DHCP authentication key may be derived from the MPA-SA established between the mobile node and the authentication agent in the candidate target network.

The proactively obtained IP address is not assigned to the mobile node’s physical interface until the mobile node has moved to the new network. The IP address thus obtained proactively from the target network should not be assigned to the physical interface but rather to a virtual interface of the client. Thus, such a proactively acquired IP address via direct DHCP communication between the mobile node and the DHCP relay agent or the DHCP server in the CTN may be carried with additional information that is used to distinguish it from other addresses as assigned to the physical interface.

Upon the mobile node’s entry to the new network, the mobile node can perform DHCP over the physical interface to the new network to get other configuration parameters, such as the SIP server or DNS server, by using DHCP INFORM. This should not affect the ongoing communication between the mobile node and Correspondent Host. Also, the mobile node can perform DHCP over the physical interface to the new network to extend the lease of the address that was proactively obtained before entering the new network.

In order to maintain the DHCP binding for the mobile node and keep track of the dispensed IP address before and after the secure proactive handover, the same DHCP client identifier needs to be used for the mobile node for both DHCP for proactive IP address acquisition and for DHCP performed after the mobile node enters the target network. The DHCP client identifier may be the MAC address of the mobile node or some other identifier.

7.3.4. Proactive IP Address Acquisition Using Stateless Autoconfiguration

For IPv6, a network address is configured either using DHCPv6 or stateless autoconfiguration. In order to obtain the new IP address proactively, the router advertisement of the next-hop router can be sent over the established tunnel, and a new IPv6 address is generated based on the prefix and MAC address of the mobile node. Generating a CoA from the new network will avoid the time needed to obtain an IP address and perform Duplicate Address Detection.

Duplicate Address Detection and address resolution are part of the IP address acquisition process. As part of the proactive configuration, these two processes can be done ahead of time. Details of how these two processes can be done proactively are described in Appendix A and Appendix B, respectively.

In the case of stateless autoconfiguration, the mobile node checks to see the prefix of the router advertisement in the new network and matches it with the prefix of the newly assigned IP address. If these turn out to be the same, then the mobile node does not go through the IP address acquisition phase again.

7.4. Tunnel Management

After an IP address is proactively acquired from the DHCP server in a CTN, or via stateless autoconfiguration in the case of IPv6, a proactive handover tunnel is established between the mobile node and the access router in the CTN. The mobile node uses the acquired IP address as the tunnel’s inner address.

There are several reasons why this transient tunnel is established between the nAR and the mobile node in the old PoA, unlike the transient tunnel in FMIPv6 (Fast MIPv6) [RFC5568], where it is set up between the mobile node’s new point of attachment and the old access router.

In the case of inter-domain handoff, it is important that any signaling message between the nPoA and the mobile node needs to be secured. This transient secured tunnel provides the desired functionality, including securing the proactive binding update and transient data between the end-points before the handover has taken place. Unlike the proactive mode of FMIPv6, transient handover packets are not sent to the pAR, and thus a tunnel between the mobile node’s new point of attachment and the old access router is not needed.

In the case of inter-domain handoff, the pAR and nAR could logically be far from each other. Thus, the signaling and data during the pre-authentication period will take a longer route, and thus may be subjected to longer one-way delay. Hence, MPA provides a tradeoff between larger packet loss or larger one-way packet delay for a transient period, when the mobile node is preparing for handoff.

The proactive handover tunnel is established using a tunnel management protocol. When IKEv2 is used for proactive IP address acquisition, IKEv2 is also used as the tunnel management protocol. Alternatively, when PANA is used for proactive IP address acquisition, PANA may be used as the secure tunnel management protocol.

Once the proactive handover tunnel is established between the mobile node and the access router in the candidate target network, the access router also needs to perform proxy address resolution (Proxy ARP) on behalf of the mobile node so that it can capture any packets destined to the mobile node’s new address.

Since the mobile node needs to be able to communicate with the Correspondent Node while in the previous network, some or all parts of the binding update and data from the Correspondent Node to the mobile node need to be sent back to the mobile node over a proactive handover tunnel. Details of these binding update procedures are described in Section 7.5.

In order for the traffic to be directed to the mobile node after the mobile node attaches to the target network, the proactive handover tunnel needs to be deleted or disabled. The tunnel management protocol used for establishing the tunnel is used for this purpose. Alternatively, when PANA is used as the authentication protocol, the tunnel deletion or disabling at the access router can be triggered by means of the PANA update mechanism as soon as the mobile node moves to the target network. A link-layer trigger ensures that the mobile node is indeed connected to the target network and can also be used as the trigger to delete or disable the tunnel. A tunnel management protocol also triggers the router advertisement (RA) from the next access router to be sent over the tunnel, as soon as the tunnel creation is complete.

7.5. Binding Update

There are several kinds of binding update mechanisms for different mobility management schemes.

In the case of Mobile IPv4 and Mobile IPv6, the mobile node performs a binding update with the Home Agent only, if route optimization is not used. Otherwise, the mobile node performs the binding update with both the Home Agent (HA) and Correspondent Node (CN).

In the case of SIP-based terminal mobility, the mobile node sends a binding update using an INVITE to the Correspondent Node and a REGISTER message to the Registrar. Based on the distance between the mobile node and the Correspondent Node, the binding update may contribute to the handover delay. SIP-fast handover [SIPFAST] provides several ways of reducing the handover delay due to binding update. In the case of secure proactive handover using SIP-based mobility management, we do not encounter the delay due to the binding update at all, as it takes place in the previous network.

Thus, this proactive binding update scheme looks more attractive when the Correspondent Node is too far from the communicating mobile node. Similarly, in the case of Mobile IPv6, the mobile node sends the newly acquired CoA from the target network as the binding update to the HA and CN. Also, all signaling messages between the MN and HA and between the MN and CN are passed through this proactive tunnel that is set up. These messages include Binding Update (BU); Binding Acknowledgement (BA); and the associated return routability messages, such as Home Test Init (HoTI), Home Test (HoT), Care-of Test Init (CoTI), and Care-of Test (CoT). In Mobile IPv6, since the receipt of an on-link router advertisement is mandatory for the mobile node to detect the movement and trigger the binding update, a router advertisement from the next access router needs to be advertised over the tunnel. By proper configuration on the nAR, the router advertisement can be sent over the tunnel interface to trigger the proactive binding update. The mobile node also needs to make the tunnel interface the active interface, so that it can send the binding update using this interface as soon as it receives the router advertisement.

If the proactive handover tunnel is realized as an IPsec tunnel, it will also protect these signaling messages between the tunnel end-points and will make the return routability test secured as well. Any subsequent data will also be tunneled through, as long as the mobile node is in the previous network. The accompanying document [MPA-WIRELESS] talks about the details of how binding updates and signaling for return routability are sent over the secured tunnel.

7.6. Preventing Packet Loss

In the MPA case, packet loss due to IP address acquisition, secured authentication, and binding update does not occur. However, transient packets during link-layer handover can be lost. Possible scenarios of packet loss and its prevention are described below.

7.6.1. Packet Loss Prevention in Single-Interface MPA

For single-interface MPA, there may be some transient packets during link-layer handover that are directed to the mobile node at the old point of attachment before the mobile node is able to attach to the target network. Those transient packets can be lost. Buffering these packets at the access router of the old point of attachment can eliminate packet loss. Dynamic buffering signals from the MN can temporarily hold transient traffic during handover, and then these packets can be forwarded to the MN once it attaches to the target network. A detailed analysis of the buffering technique can be found in [PIMRC06].

An alternative method is to use bicasting. Bicasting helps to forward the traffic to two destinations at the same time. However, it does not eliminate packet loss if link-layer handover is not seamlessly performed. On the other hand, buffering does not reduce packet delay. While packet delay can be compensated by a playout buffer at the receiver side for a streaming application, a playout buffer does not help much for interactive VoIP applications that cannot tolerate large delay jitters. Thus, it is still important to optimize the link-layer handover anyway.

7.6.2. Preventing Packet Losses for Multiple Interfaces

MPA usage in multi-interface handover scenarios involves preparing the second interface for use via the current active interface. This preparation involves pre-authentication and provisioning at a target network where the second interface would be the eventual active interface. For example, during inter-technology handover from a WiFi to a CDMA network, pre-authentication at the CDMA network can be performed via the WiFi interface. The actual handover occurs when the CDMA interface becomes the active interface for the MN.

In such scenarios, if handover occurs while both interfaces are active, there is generally no packet loss, since transient packets directed towards the old interface will still reach the MN. However, if sudden disconnection of the current active interface is used to initiate handover to the prepared interface, then transient packets for the disconnected interface will be lost while the MN attempts to be reachable at the prepared interface. In such cases, a specialized form of buffering can be used to eliminate packet loss where packets are merely copied at an access router in the current active network prior to disconnection. If sudden disconnection does occur, copied packets can be forwarded to the MN once the prepared interface becomes the active reachable interface. The copy-and-forward mechanism is not limited to multi-interface handover.

A notable side-effect of this process is the possible duplication of packets during forwarding to the new active interface. Several approaches can be employed to minimize this effect. Relying on upper-layer protocols such as TCP to detect and eliminate duplicates is the most common approach. Customized duplicate detection and handling techniques can also be used. In general, packet duplication is a well-known issue that can also be handled locally by the MN.

If the mobile node takes a longer amount of time to detect the disconnection event of the current active interface, this can also have an adverse effect on the length of the handover process. Thus, it becomes necessary to use an optimized scheme of detecting interface disconnection in such scenarios. Use of the current interface to perform pre-authentication instead of the new interface is desirable in certain circumstances, such as to save battery power, or in cases where the adjacent cells (e.g., WiFi or CDMA) are non-overlapping, or in cases when the carrier does not allow the simultaneous use of both interfaces. However, in certain circumstances, depending upon the type of target network, only parts of MPA operations can be performed (e.g., pre-authentication, pre-configuration, or proactive binding update). In a specific scenario involving handoff between WiFi and CDMA networks, some of the PPP context can be set up during the pre-authentication period, thus reducing the time for PPP activation.

7.6.3. Reachability Test

In addition to previous techniques, the MN may also want to ensure reachability of the new point of attachment before switching from the old one. This can be done by exchanging link-layer management frames with the new point of attachment. This reachability check should be performed as quickly as possible. In order to prevent packet loss during this reachability check, transmission of packets over the link between the MN and the old point of attachment should be suspended by buffering the packets at both ends of the link during the reachability check. How to perform this buffering is out of scope of this document. Some of the results of using this buffering scheme are explained in the accompanying document [MPA-WIRELESS].

7.7. Security and Mobility

This section describes how MPA can help establish layer 2 and layer 3 security association in the target networks while the mobile node is in the previous network.

7.7.1. Link-Layer Security and Mobility

Using the MPA-SA established between the mobile node and the authentication agent for a CTN, during the pre-authentication phase, it is possible to bootstrap link-layer security in the CTN while the mobile node is in the current network, as described in the following steps. Figure 5 shows the sequence of operation.

(1) The authentication agent and the mobile node derive a PMK (Pair-wise Master Key) [RFC5247] using the MPA-SA that is established as a result of successful pre-authentication. Successful operation of EAP and a AAA protocol may be involved during pre-authentication to establish the MPA-SA. From the PMK, distinct TSKs (Transient Session Keys) [RFC5247] for the mobile node are directly or indirectly derived for each point of attachment of the CTN.

(2) The authentication agent may install the keys derived from the PMK and used for secure association to points of attachment. The derived keys may be TSKs or intermediary keys from which TSKs are derived.

(3) After the mobile node chooses a CTN as the target network and switches to a point of attachment in the target network (which now becomes the new network for the mobile node), it executes a secure association protocol such as the IEEE 802.11i 4-way handshake [802.11], using the PMK in order to establish PTKs (Pair-wise Transient Keys) and group keys [RFC5247] used for protecting link-layer packets between the mobile node and the point of attachment. No additional execution of EAP authentication is needed here.

(4) While the mobile node is roaming in the new network, the mobile node only needs to perform a secure association protocol with its point of attachment, and no additional execution of EAP authentication is needed either. Integration of MPA with link-layer handover optimization mechanisms such as 802.11r can be archived this way.

The mobile node may need to know the link-layer identities of the points of attachment in the CTN to derive TSKs.

```
          _________________        ____________________________
         | Current Network |      |           CTN              |
         |   ____          |      |                 ____       |
         |  |    |      (1) pre-authentication     |    |      |
         |  | MN |<------------------------------->| AA |      |
         |  |____|         |      |                |____|      |
         |    .            |      |                  |         |
         |    .            |      |                  |         |
         |____.____________|      |                  |         |
              .movement           |                  |(2) Keys |
          ____.___________________|                  |         |
         |   _v__                      _____         |         |
         |  |    |(3) secure assoc.   |     |        |         |
         |  | MN |<------------------>| AP1 |<-------+         |
         |  |____|                    |_____|        |         |
         |    .                                      |         |
         |    .movement                              |         |
         |    .                                      |         |
         |    .                                      |         |
         |   _v__                      _____         |         |
         |  |    |(4) secure assoc.   |     |        |         |
         |  | MN |<------------------>| AP2 |<-------+         |
         |  |____|                    |_____|                  |
         |_____________________________________________________|
```

Figure 5: Bootstrapping Link-Layer Security

7.7.2. IP-Layer Security and Mobility

IP-layer security is typically maintained between the mobile node and the first-hop router, or any other network element such as SIP proxy by means of IPsec. This IPsec SA can be set up either in tunnel mode or in ESP mode. However, as the mobile node moves, the IP address of the router and outbound proxy will change in the new network. The mobile node’s IP address may or may not change, depending upon the mobility protocol being used. This will warrant re-establishing a new security association between the mobile node and the desired network entity. In some cases, such as in a 3GPP/3GPP2 IMS/MMD environment, data traffic is not allowed to pass through unless there is an IPsec SA established between the mobile node and outbound proxy. This will of course add unreasonable delay to the existing real-time communication during a mobile node’s movement. In this scenario, key exchange is done as part of a SIP registration that follows a key exchange procedure called AKA (Authentication and Key Agreement).

MPA can be used to bootstrap this security association as part of pre-authentication via the new outbound proxy. Prior to the movement, if the mobile node can pre-register via the new outbound proxy in the target network and completes the pre-authentication procedure, then the new SA state between the mobile node and new outbound proxy can be established prior to the movement to the new network. A similar approach can also be applied if a key exchange mechanism other than AKA is used or the network element with which the security association has to be established is different than an outbound proxy.

By having the security association established ahead of time, the mobile node does not need to be involved in any exchange to set up the new security association after the movement. Any further key exchange will be limited to renew the expiry time. This will reduce the delay for real-time communication as well.

7.8. Authentication in Initial Network Attachment

When the mobile node initially attaches to a network, network access authentication would occur regardless of the use of MPA. The protocol used for network access authentication when MPA is used for handover optimization can be a link-layer network access authentication protocol such as IEEE 802.1X, or a higher-layer network access authentication protocol such as PANA.

8. Security Considerations

This document describes a framework for a secure handover optimization mechanism based on performing handover-related signaling between a mobile node and one or more candidate target networks to which the mobile node may move in the future. This framework involves acquisition of the resources from the CTN as well as data packet redirection from the CTN to the mobile node in the current network before the mobile node physically connects to one of those CTNs.

Acquisition of the resources from the candidate target networks must be done with appropriate authentication and authorization procedures in order to prevent an unauthorized mobile node from obtaining the resources. For this reason, it is important for the MPA framework to perform pre-authentication between the mobile node and the candidate target networks. The MN-CA key and the MN-AR key generated as a result of successful pre-authentication can protect subsequent handover signaling packets and data packets exchanged between the mobile node and the MPA functional elements in the CTNs.

The MPA framework also addresses security issues when the handover is performed across multiple administrative domains. With MPA, it is possible for handover signaling to be performed based on direct communication between the mobile node and routers or mobility agents in the candidate target networks. This eliminates the need for a context transfer protocol [RFC5247] for which known limitations exist in terms of security and authorization. For this reason, the MPA framework does not require trust relationships among administrative domains or access routers, which makes the framework more deployable in the Internet without compromising the security in mobile environments.

9. Acknowledgments

We would like to thank Farooq Anjum and Raziq Yaqub for their review of this document, and Subir Das for standardization support in the IEEE 802.21 working group.

The authors would like to acknowledge Christian Vogt, Rajeev Koodli, Marco Liebsch, Juergen Schoenwaelder, and Charles Perkins for their thorough review of the document and useful feedback.

Author and Editor Ashutosh Dutta would like to thank Telcordia Technologies, and author Victor Fajardo would like to thank Toshiba America Research and Telcordia Technologies, for supporting the development of their document while they were employed in their respective organizations.

10. References

10.1. Normative References

[RFC5944] Perkins, C., Ed., "IP Mobility Support for IPv4, Revised", RFC 5944, November 2010.

[RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, Ed., "Extensible Authentication Protocol (EAP)", RFC 3748, June 2004.

[RFC3775] Johnson, D., Perkins, C., and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

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

[RFC5380] Soliman, H., Castelluccia, C., El Malki, K., and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management", RFC 5380, October 2008.

[RFC5568] Koodli, R., Ed., "Mobile IPv6 Fast Handovers", RFC 5568, July 2009.

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

[RFC4555] Eronen, P., "IKEv2 Mobility and Multihoming Protocol (MOBIKE)", RFC 4555, June 2006.

[RFC4881] El Malki, K., Ed., "Low-Latency Handoffs in Mobile IPv4", RFC 4881, June 2007.

[RFC4066] Liebsch, M., Ed., Singh, A., Ed., Chaskar, H., Funato, D., and E. Shim, "Candidate Access Router Discovery (CARD)", RFC 4066, July 2005.

[RFC4067] Loughney, J., Nakhjiri, M., Perkins, C., and R. Koodli, "Context Transfer Protocol (CXTP)", RFC 4067, July 2005.

[RFC5247] Aboba, B., Simon, D., and P. Eronen, "Extensible Authentication Protocol (EAP) Key Management Framework", RFC 5247, August 2008.

[RFC5191] Forsberg, D., Ohba, Y., Ed., Patil, B., Tschofenig, H., and A. Yegin, "Protocol for Carrying Authentication for Network Access (PANA)", RFC 5191, May 2008.

[RG98] ITU-T, "General Characteristics of International Telephone Connections and International Telephone Circuits: One-Way Transmission Time", ITU-T Recommendation G.114, 1998.

[ITU98] ITU-T, "The E-Model, a computational model for use in transmission planning", ITU-T Recommendation G.107, 1998.

[ETSI] ETSI, "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 1: General aspects of Quality of Service (QoS)", ETSI TR 101 329-1 V3.1.2, 2002.

10.2. Informative References

[RFC5201]	Moskowitz, R., Nikander, P., Jokela, P., Ed., and T. Henderson, "Host Identity Protocol", RFC 5201, April 2008.
[RFC2679]	Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.
[RFC2680]	Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Packet Loss Metric for IPPM", RFC 2680, September 1999.
[RFC2681]	Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, September 1999.
[RFC2003]	Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.
[RFC2608]	Guttman, E., Perkins, C., Veizades, J., and M. Day, "Service Location Protocol, Version 2", RFC 2608, June 1999.
[RFC2473]	Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, December 1998.
[RFC3046]	Patrick, M., "DHCP Relay Agent Information Option", RFC 3046, January 2001.
[RFC4039]	Park, S., Kim, P., and B. Volz, "Rapid Commit Option for the Dynamic Host Configuration Protocol version 4 (DHCPv4)", RFC 4039, March 2005.
[RFC5172]	Varada, S., Ed., "Negotiation for IPv6 Datagram Compression Using IPv6 Control Protocol", RFC 5172, March 2008.
[RFC5648]	Wakikawa, R., Ed., Devarapalli, V., Tsirtsis, G., Ernst, T., and K. Nagami, "Multiple Care-of Addresses Registration", RFC 5648, October 2009.
[RFC4429]	Moore, N., "Optimistic Duplicate Address Detection (DAD) for IPv6", RFC 4429, April 2006.

[RFC5836]	Ohba, Y., Ed., Wu, Q., Ed., and G. Zorn, Ed., "Extensible Authentication Protocol (EAP) Early Authentication Problem Statement", RFC 5836, April 2010.
[RFC5213]	Gundavelli, S., Ed., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.
[RFC5974]	Manner, J., Karagiannis, G., and A. McDonald, "NSIS Signaling Layer Protocol (NSLP) for Quality-of-Service Signaling", RFC 5974, October 2010.
[RFC5169]	Clancy, T., Nakhjiri, M., Narayanan, V., and L. Dondeti, "Handover Key Management and Re-Authentication Problem Statement", RFC 5169, March 2008.
[SIPMM]	Schulzrinne, H. and E. Wedlund, "Application-Layer Mobility Using SIP", ACM MC2R, July 2000.
[CELLIP]	Campbell, A., Gomez, J., Kim, S., Valko, A., Wan, C., and Z. Turanyi, "Design, Implementation, and Evaluation of Cellular IP", IEEE Personal Communications, August 2000.
[MOBIQUIT07]	Lopez, R., Dutta, A., Ohba, Y., Schulzrinne, H., and A. Skarmeta, "Network-layer assisted mechanism to optimize authentication delay during handoff in 802.11 networks", IEEE Mobiquitous, June 2007.
[MISHRA04]	Mishra, A., Shin, M., Petroni, N., Clancy, T., and W. Arbaugh, "Proactive key distribution using neighbor graphs", IEEE Wireless Communications Magazine, February 2004.
[SPRINGER07]	Dutta, A., Das, S., Famolari, D., Ohba, Y., Taniuchi, K., Fajardo, V., Lopez, R., Kodama, T., Schulzrinne, H., and A. Skarmeta, "Seamless proactive handover across heterogeneous access networks", Wireless Personal Communications, November 2007.
[HAWAII]	Ramjee, R., La Porta, T., Thuel, S., Varadhan, K., and S. Wang, "HAWAII: A Domain-based Approach for Supporting Mobility in Wide-area Wireless networks", International Conference on Network Protocols ICNP’99.

[IDMP] Das, S., McAuley, A., Dutta, A., Misra, A., Chakraborty, K., and S. Das, "IDMP: An Intra-Domain Mobility Management Protocol for Next Generation Wireless Networks", IEEE Wireless Communications Magazine, October 2000.

[MOBIP-REG] Gustafsson, E., Jonsson, A., and C. Perkins, "Mobile IPv4 Regional Registration", Work in Progress, June 2004.

[YOKOTA] Yokota, H., Idoue, A., Hasegawa, T., and T. Kato, "Link Layer Assisted Mobile IP Fast Handoff Method over Wireless LAN Networks", Proceedings of ACM MobiCom02, 2002.

[MACD] Shin, S., Forte, A., Rawat, A., and H. Schulzrinne, "Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs", MobiWac Workshop, 2004.

[SUM] Dutta, A., Zhang, T., Madhani, S., Taniuchi, K., Fujimoto, K., Katsube, Y., Ohba, Y., and H. Schulzrinne, "Secured Universal Mobility for Wireless Internet", WMASH’04, October 2004.

[SIPFAST] Dutta, A., Madhani, S., Chen, W., Altintas, O., and H. Schulzrinne, "Fast-handoff Schemes for Application Layer Mobility Management", PIMRC 2004.

[PIMRC06] Dutta, A., Berg, E., Famolari, D., Fajardo, V., Ohba, Y., Taniuchi, K., Kodama, T., and H. Schulzrinne, "Dynamic Buffering Control Scheme for Mobile Handoff", Proceedings of PIMRC 2006, 1-11.

[MITH] Gwon, Y., Fu, G., and R. Jain, "Fast Handoffs in Wireless LAN Networks using Mobile initiated Tunneling Handoff Protocol for IPv4 (MITHv4)", Wireless Communications and Networking 2003, January 2005.

[WENYU] Jiang, W. and H. Schulzrinne, "Modeling of Packet Loss and Delay and their Effect on Real-Time Multimedia Service Quality", NOSSDAV 2000, June 2000.

[802.21] "IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services, IEEE 802.21-2008", a contribution to IEEE 802.21 WG, January 2009.

[802.11]	"IEEE Wireless LAN Edition A compilation based on IEEE Std 802.11-1999(R2003)", Institute of Electrical and Electronics Engineers, September 2003.
[GPSIP]	Dutta, A., Madhani, S., Chen, W., Altintas, O., and H. Schulzrinne, "GPS-IP based fast-handoff approaches for Mobiles", IEEE Sarnoff Symposium 2006.
[MAGUIRE]	Vatn, J. and G. Maguire, "The effect of using co-located care-of addresses on macro handover latency", 14th Nordic Teletraffic Seminar 1998.
[MPA-MOBIKE]	El Mghazli, Y., Bournelle, J., and J. Laganier, "MPA using IKEv2 and MOBIKE", Work in Progress, June 2006.
[MPA-WIRELESS]	Dutta, A., Famolari, D., Das, S., Ohba, Y., Fajardo, V., Taniuchi, K., Lopez, R., and H. Schulzrinne, "Media- Independent Pre-authentication Supporting Secure Interdomain Handover Optimization", IEEE Wireless Communications Magazine, April 2008.

Appendix A. Proactive Duplicate Address Detection

When the DHCP server dispenses an IP address, it updates its lease table, so that this same address is not given to another client for that specific period of time. At the same time, the client also keeps a lease table locally so that it can renew when needed. In some cases where a network consists of both DHCP and non-DHCP-enabled clients, there is a probability that another client in the LAN may have been configured with an IP address from the DHCP address pool. In such a scenario, the server detects a duplicate address based on ARP (Address Resolution Protocol) or IPv6 Neighbor Discovery before assigning the IP address. This detection procedure may take from 4 sec to 15 sec [MAGUIRE] and will thus contribute to a larger handover delay. In the case of a proactive IP address acquisition process, this detection is performed ahead of time and thus does not affect the handover delay at all. By performing the Duplicate Address Detection (DAD) ahead of time, we reduce the IP address acquisition time.

The proactive DAD over the candidate target network should be performed by the nAR on behalf of the mobile node at the time of proactive handover tunnel establishment, since DAD over a tunnel is not always performed. For example, in the case of IPv6, DAD over an IP-IP tunnel interface is turned off in an existing implementation. In the case of IPv6 over PPP [RFC5172], the IP Control Protocol (IPCPv6) negotiates the link-local addresses, and hence DAD over the tunnel is not needed. After the mobile node has moved to the target network, a DAD procedure may be started because of reassignment of the nCoA to the physical interface to the target network. In that case, the mobile node should use optimistic DAD [RFC4429] over the physical interface so that the nCoA that was used inside the proactive handover tunnel before handover can be immediately used over that physical interface after handover. The schemes used for the proactive DAD and optimistic DAD are applicable to both stateless and stateful address autoconfiguration schemes used for obtaining a nCoA.

Appendix B. Address Resolution

Address resolution involves updating the next access router’s neighbor cache. We briefly describe these two operations below.

During the process of pre-configuration, the MAC address resolution mappings needed by the mobile node to communicate with nodes in the target network after attaching to the target network can also be known, where the communicating nodes may be the access router, authentication agent, configuration agent, or Correspondent Node. There are several possible ways of performing such proactive MAC address resolution.

- One can use an information service mechanism [802.21] to resolve the MAC addresses of the nodes. This might require each node in the target network to be involved in the information service so that the server of the information service can construct the database for proactive MAC address resolution.
- One can extend the authentication protocol used for pre-authentication or the configuration protocol used for pre-configuration to support proactive MAC address resolution. For example, if PANA is used as the authentication protocol for pre-authentication, PANA messages may carry attribute-value pairs (AVPs) used for proactive address resolution. In this case, the PANA authentication agent in the target network may perform address resolution on behalf of the mobile node.
- One can also make use of DNS to map the MAC address of the specific interface associated with a specific IP address of the network element in the target network. One may define a new DNS resource record (RR) to proactively resolve the MAC addresses of the nodes in the target network. But this approach may have its own limitations, since a MAC address is a resource that is bound to an IP address, and not directly to a domain name.

When the mobile node attaches to the target network, it installs the proactively obtained address resolution mappings without necessarily performing address resolution queries for the nodes in the target network.

On the other hand, the nodes that reside in the target network and that are communicating with the mobile node should also update their address resolution mappings for the mobile node as soon as the mobile node attaches to the target network. The above proactive address resolution methods could also be used for those nodes to proactively resolve the MAC address of the mobile node before the mobile node attaches to the target network. However, this is not useful, since those nodes need to detect the attachment of the mobile node to the target network before adopting the proactively resolved address resolution mapping. A better approach would be integration of attachment detection and address resolution mapping update. This is based on gratuitously performing address resolution [RFC5944], [RFC3775] in which the mobile node sends an ARP Request or an ARP Reply in the case of IPv4, or a Neighbor Advertisement in the case of IPv6, immediately after the mobile node attaches to the new network, so that the nodes in the target network can quickly update the address resolution mapping for the mobile node.

Appendix C. MPA Deployment Issues

In this section, we describe some of the deployment issues related to MPA.

C.1. Considerations for Failed Switching and Switch-Back

The ping-pong effect is one of the common problems found during handover. The ping-pong effect arises when a mobile node is located at the borderline of the cell or decision point and a handover procedure is frequently executed. This results in higher call drop probability, lower connection quality, increased signaling traffic, and waste of resources. All of these affect mobility optimization. Handoff algorithms are the deciding factors for performing the handoff between the networks. Traditionally, these algorithms employ a threshold to compare the values of different metrics to decide on the handoff. These metrics include signal strength, path loss, Carrier-to-Interference Ratio (CIR), Signal-to-Interference Ratio (SIR), Bit Error Rate (BER), and power budget. In order to avoid the ping-pong effect, some additional parameters are employed by the decision algorithm, such as hysteresis margin, dwell timers, and averaging window. For a vehicle moving at a high speed, other parameters, such as the distance between the mobile node and the point of attachment, velocity of the mobile node, location of the mobile node, traffic, and bandwidth characteristics are also taken into account to reduce the ping-pong effect. More recently, there are other handoff algorithms available that help reduce the ping-pong effect in a heterogeneous network environment and that are based on techniques such as hypothesis testing, dynamic programming, and pattern recognition techniques. While it is important to devise smart handoff algorithms to reduce the ping-pong effect, it is also important to devise methods to recover from this effect.

In the case of the MPA framework, the ping-pong effect will result in the back-and-forth movement of the mobile node between the current network and target network, and between the candidate target networks. MPA in its current form will be affected because of the number of tunnels set up between the mobile node and neighboring access routers, the number of binding updates, and associated handoff latency resulting from the ping-pong situation. The mobile node’s handoff rate may also contribute to delay and packet loss. We propose a few techniques that will help reduce the probability of the ping-pong effect and propose several methods for the MPA framework so that it can recover from the packet loss resulting from the ping-pong effect.

The MPA framework can take advantage of the mobile node’s geo-location with respect to APs in the neighboring networks using GPS. In order to avoid the oscillation between the networks, a location-aware algorithm can be derived by using a co-relation between the user’s location and cached data from the previous handover attempts. In some cases, location may not be the only indicator for a handoff decision. For example, in Manhattan-type grid networks, although a mobile node is close to an AP, it may not have enough SNR (Signal-to-Noise Ratio) to make a good connection. Thus, knowledge of the mobility pattern, dwell time in a call, and path identification will help avoid the ping-pong problem to a great extent.

In the absence of a good handoff algorithm that can avoid the ping-pong effect, it may be required to put in place a good recovery mechanism so as to mitigate the effect of ping-pong. It may be necessary to keep the established context in the current network for a period of time, so that it can be quickly recovered when the mobile node comes back to the network where the context was last used. This context may include security association, IP address used, and tunnels established. Bicasting the data to both the previous network and the new network for a predefined period will also help the mobile node to take care of the lost packets in case the mobile node moves back and forth between the networks. The mobile node can also take certain action, after it determines that it is in a stable state with respect to a ping-pong situation.

When the MPA framework takes advantage of a combination of IKEv2 and MOBIKE, the ping-pong effect can be reduced further [MPA-MOBIKE].

C.2. Authentication State Management

In the case of pre-authentication with multiple target networks, it is useful to maintain the state in the authentication agent of each of the neighboring networks for a certain time period. Thus, if the mobile node does move back and forth between neighboring networks, already-maintained authentication state can be helpful. We provide some highlights on multiple security association state management below.

A mobile node that has pre-authenticated with an authentication agent in a candidate target network and has an MPA-SA may need to continue to keep the MPA-SA while it continues to stay in the current network or even after it makes a handover to a network that is different from the candidate target network.

When an MN that has been authenticated and authorized by an authentication agent in the current network makes a handover to a target network, it may want to hold the SA that has been established between the MN and the authentication agent for a certain time period so that it does not have to go through the entire authentication signaling to create an SA from scratch, in case it returns to the previous network. Such an SA being held at the authentication agent after the MN’s handover to another network is considered as an MPA-SA. In this case, the authentication agent should change the fully authorized state for the MN to an unauthorized state. The unauthorized state can be changed to the fully authorized state only when the MN comes back to the network and provides proof of possession of a key associated with the MPA-SA.

While an MPA-SA is being held at an authentication agent, the MN will need to keep updating the authentication agent when an IP address of the MN changes due to a handover, to re-establish the new SA.

C.3. Pre-Allocation of QoS Resources

In the pre-configuration phase, it is also possible to pre-allocate QoS resources that may be used by the mobile node not only after handover but also before handover. When pre-allocated QoS resources are used before handover, they are used for application traffic carried over a proactive handover tunnel.

It is possible that QoS resources are pre-allocated in an end-to-end fashion. One method to achieve this proactive end-to-end QoS reservation is to execute the NSIS Signaling Layer Protocol (NSLP) [RFC5974] or the Resource Reservation Protocol (RSVP) [RFC2205] over a proactive handover tunnel where pre-authentication can be used for bootstrapping a security association for the proactive handover tunnel to protect the QoS signaling. In this case, QoS resources are pre-allocated on the path between the Correspondent Node and a target access router and can be used continuously before and after handover. On the other hand, duplicate pre-allocation of QoS resources between the target access router and the mobile node is necessary when using pre-allocated QoS resources before handover, due to differences in paths between the target access router and the mobile node before and after handover. QoS resources to be used for the path between the target access router and the mobile node after handover may be pre-allocated by extending NSLP to work for off-path signaling (Note: this path can be viewed as off-path before handover) or by media-specific QoS signaling at layer 2.

C.4. Resource Allocation Issue during Pre-Authentication

In the case of multiple CTNs, establishing multiple tunnels with the neighboring target networks provides some additional benefits. But it contributes to some resource utilization issues as well. A pre-authentication process with multiple candidate target networks can happen in several ways.

The very basic scheme involves authenticating the mobile node with the multiple authentication agents in the neighboring networks, but actual pre-configuration and binding update take place only after layer 2 movement to a specific network is complete.

Similarly, in addition to pre-authentication, the mobile node can also complete the pre-configuration while in the previous network, but can postpone the binding update until after the mobile node has moved. Like the previous case, in this case the mobile node also does not need to set up the pre-configured tunnels. While the pre-authentication process and part of the pre-configuration process are taken care of before the mobile node has moved to the new network, the binding update is actually done after the mobile node has moved.

The third type of multiple pre-authentication involves all the three steps while the mobile node is in the previous networks, such as authentication, configuration, and binding update. But, this specific process utilizes the highest amount of resources. Some of the resources that get used during this process are as follows:

(1) Additional signaling for pre-authentication in the neighboring networks

(2) Holding the IP address of the neighboring networks in the mobile node’s cache for a certain amount of time. Additional processing in the mobile node is needed for storing these IP addresses. In addition, this caching of addresses also uses up the temporary IP addresses from the neighboring routers.

(3) There is an additional cost associated with setting up additional transient tunnels with the target routers in the neighboring networks and the mobile node.

(4) In the case of a binding update with multiple IP addresses obtained from the neighboring networks, multiple transient streams flow between the CN and mobile node using these transient tunnels.

However, there are pros and cons related to sending the binding update after the handover. If the binding update is sent after the mobile node has moved to the new network, this will contribute to the delay if the CH or HA is far from the MN. Multiple binding updates can be taken care of in many different ways. We describe a few of these update mechanisms below.

When only pre-authentication and pre-configuration are done ahead of time with multiple networks, the mobile node sends one binding update to the CN. In this case, it is important to find out when to send the binding update after the layer 2 handoff.

In case a binding update with multiple contact addresses is sent, multiple media streams stem out of the CN, using the transient tunnels. But in that case, one needs to send another binding update after the handover, with the contact address set to the new address (only one address) where the mobile node has moved. This way, the mobile node stops sending media to other neighboring networks where the mobile node did not move.

The following is an illustration of this specific case that takes care of multiple binding streams, when the mobile node moves only to a specific network, but sends multiple binding updates in the previous network. The MN sends a binding update to the CH with multiple contact addresses, such as c1, c2, and c3, that were obtained from three neighboring networks. This allows the CN to send transient multiple streams to the mobile node over the pre-established tunnels. After the mobile node moves to the actual network, it sends another binding update to the CN with the care-of address of the mobile node in the network where the mobile node has moved. One issue with multiple streams is consumption of extra bandwidth for a small period of time.

Alternatively, one can apply the buffering technique at the target access router or at the Home Agent. Transient data can be forwarded to the mobile node after it has moved. Forwarding of data can be triggered by the mobile node either as part of Mobile IP registration or as a separate buffering protocol.

C.5. Systems Evaluation and Performance Results

In this section, we present some of the results from MPA implementation when applied to different handover scenarios. We present the summary of results from our experiments using MPA techniques for two types of handovers: i) intra-technology and intra-domain, and ii) inter-technology and inter-domain. We also present the results of how the MPA can bootstrap layer 2 security for both roaming and non-roaming cases. Detailed procedures and results are explained in [MOBIQUIT07] and [SPRINGER07].

C.5.1. Intra-Technology, Intra-Domain

The results for MIPv6 and SIP mobility involving intra-domain mobility are shown in Figures 6 and 7, respectively.

	Buffering (disabled) & RO (disabled)	Buffering (enabled) & RO (disabled)	Buffering (disabled) & RO (enabled)	Buffering (enabled) & RO (enabled)
L2 handoff (ms)	4.00	4.33	4.00	4.00
L3 handoff (ms)	1.00	1.00	1.00	1.00
Avg. packet loss	1.33	0	0.66	0
Avg. inter-packet arrival interval (ms)	16.00	16.00	16.00	16.00
Avg. inter-packet arrival time during handover (ms)	n/a	45.33	n/a	66.60
Avg. packet jitter (ms)	n/a	29.33	n/a	50.60
Buffering Period (ms)	n/a	50.00	n/a	50.00
Buffered Packets	n/a	2.00	n/a	3.00

RO = Router Optimization

Figure 6: Mobile IPv6 with MPA Results

	Buffering disabled	Buffering enabled
L2 handoff (ms)	4.00	5.00
L3 handoff (ms)	1.00	1.00
Avg. packet loss	1.50	0
Avg. inter-packet arrival interval (ms)	16.00	16.00
Avg. inter-packet arrival time during handover (ms)	n/a	29.00
Avg. packet jitter (ms)	n/a	13.00
Buffering Period (ms)	n/a	20.00
Buffered Packets	n/a	3.00

Figure 7: SIP Mobility with MPA Results

For all measurements, we did not experience any performance degradation during handover in terms of the audio quality of the voice traffic.

With the use of buffering during handover, packet loss during the actual L2 and L3 handover is eliminated with appropriate and reasonable settings of the buffering period for both MIP6 and SIP mobility. In the case of MIP6, there is not a significant difference in results with and without route optimization. It should be noted that results with more samples would be necessary for a more detailed analysis.

In the case of non-MPA-assisted handover, handover delay and associated packet loss occur from the moment the link-layer handover procedure begins, up to successful processing of the binding update. During this process, IP address acquisitions via DHCP incur the longest delay. This is due to the detection of duplicate IP addresses in the network before the DHCP request completes. The binding update exchange also experiences a long delay if the CN is too far from the MN. As a result, the non-MPA-assisted handover took an average of 4 seconds to complete, with an approximate packet loss of about 200 packets. The measurement is based on the same traffic rate and traffic source as the MPA-assisted handover.

C.5.2. Inter-Technology, Inter-Domain

Handoff involving heterogeneous access can take place in many different ways. We limit the experiment to two interfaces, and therefore results in several possible setup scenarios, depending upon the activity of the second interface. In one scenario, the second interface comes up when the link to the first interface goes down. This is a reactive scenario and usually gives rise to undesirable packet loss and handoff delay. In a second scenario, the second interface is being prepared while the mobile node still communicates using the old interface. Preparation of the second interface should include setup of all the required state and security associations (e.g., PPP state, the Link Control Protocol (LCP), the Challenge Handshake Authentication Protocol (CHAP)). If such a lengthy process is established ahead of time, it reduces the time taken for the secondary interface to be attached to the network. After preparation, the mobile node decides to use the second interface as the active interface. This results in less packet loss, as it uses make-before-break techniques. This is a proactive scenario and can have two "flavors". The first is where both interfaces are up; the second is when only the old interface is up and the prepared interface is brought up only when handoff is about to occur. This scenario may be beneficial from a battery management standpoint. Devices that operate two interfaces simultaneously can rapidly deplete their batteries. However, by activating the second interface only after an appropriate network has been selected, the client may utilize battery power effectively.

As compared to non-optimized handover that may result in a delay of up to 18 sec and loss of 1000 or more packets during the handover from the wireless LAN (WLAN) to CDMA, we observed 0 packet loss and a 50-ms handoff delay between the last pre-handoff packet and the first in-handoff packet. This handoff delay includes the time due to link down detection and time needed to delete the tunnel after the mobile node has moved. However, we observed about 10 duplicate packets because of the copy-and-forward mechanism at the access routers. But these duplicate packets are usually handled easily by the upper-layer protocols.

C.5.3. MPA-Assisted Layer 2 Pre-Authentication

In this section, we discuss the results obtained from MPA-assisted layer 2 pre-authentication and compare these with EAP authentication and IEEE 802.11i’s pre-authentication techniques. Figure 8 shows the experimental testbed where we have conducted the MPA-assisted pre-authentication experiment for bootstrapping layer 2 security as explained in Section 7. By pre-authenticating and pre-configuring the link, the security association procedure during handoff reduces to a 4-way handshake only. Then the MN moves to the AP and, after association, runs a 4-way handshake by using the PSKap (Pre-shared Key at AP) generated during PANA pre-authentication. At this point, the handoff is complete. Details of this experimental testbed can be found in [MOBIQUIT07].

```
   +----------------------------+-----------+ +-------------+----------+
   |                                        | |                        |
   |  Home Domain       +-------++          | |                        |
   |                    |        |          | |                        |
   |                    |AAAHome |          | |                        |
   |                    +        |          | |                        |
   |                    +-----+--+          | |                        |
   |                          |             | |  Network B             |
   |   Network A              |             | |                        |
   |                        /----\          | |            /---\       |
   |                       /nAR   \         | |           /     \      |
   |                      | PAA    |--------+-+----------+ pAR   |     |
   |                       \      /         | |           \     /      |
   |                        \----/          | |            \-+-/       |
   |                           |            | |              |         |
   |             +-------------------|      | |              |         |
   |             |       IEEE 802.11i|      | |              |         |
   |           +------+          +------+   | |          +---+--+      |
   |           |      |          |      |   | |          |      |      |
   |           |AP2   |          |AP1   |   | |          |AP0   |      |
   |           +------+          +------+   | |          +------+      |
   |           +------+            +-----+  | |           +-----+      |
   |           |      |            |     |  | |           |     |      |
   |           |MN    +----------->|MN   |<+------------- |MN   |      |
   |           +------+            +-----+  | |           ++----+      |
   |----------------------------------------+ +------------+-----------+
```

Figure 8: Experimental Testbed for MPA-Assisted
L2 Pre-Authentication (Non-Roaming)

```
                        +-----------------------------+
                        |      +--------+             |
                        |      |        |             |
                        |      | AAAH   +             |
                        |      |        |             |
                        |      ++-------+             |
                        |       |                     |
                        |       |  Home AAA Domain    |
                        |       |                     |
                        +-------+---------------------+
                                |
                                |
                                |
                       RADIUS/  |
                       Diameter |
                                |
                                |
   +----------------------------+-----------+ +-------------+----------+
   |                            |           | |                        |
   | Roaming            +-------++          | |                        |
   | AAA Domain A       |        |          | |                        |
   |                    | AAAV   |          | |                        |
   |                    +        |          | |                        |
   | Network A          +-----+--+          | |  Network B             |
   |                          |             | |                        |
   |                          |             | |                        |
   |                        /----\          | |            /---\       |
   |                       /nAR   \         | |           /     \      |
   |                      | PAA    |--------+-+----------+ pAR   |     |
   |                       \      /         | |           \     /      |
   |                        \----/          | |            \-+-/       |
   |                           |            | |              |         |
   |             +-------------------|      | |              |         |
   |             |       IEEE 802.11i|      | |              |         |
   |           +------+          +------+   | |          +---+--+      |
   |           |      |          |      |   | |          |      |      |
   |           |AP2   |          |AP1   |   | |          |AP0   |      |
   |           +------+          +------+   | |          +------+      |
   |           +------+            +-----+  | |           +-----+      |
   |           |      |            |     |  | |           |     |      |
   |           |MN    +----------->|MN   |<---------------| MN  |      |
   |           +------+            +-----+  | |           ++----+      |
   -----------------------------------------+ +------------+-----------+
```

Figure 9: Experimental Testbed for MPA-Assisted
L2 Pre-Authentication (Roaming)

We have experimented with three types of movement scenarios involving both non-roaming and roaming cases, using the testbeds shown in Figures 8 and 9, respectively. In the roaming case, the MN is visiting in a domain different than its home domain. Consequently, the MN needs to contact the AAA server in the home domain (AAAH) from its new domain. For the non-roaming case, we assume the MN is moving within its home domain, and only the local AAA server (AAAHome), which is the home AAA server for the mobile node, is contacted.

The first scenario does not involve any pre-authentication. The MN is initially connected to AP0 and moves to AP1. Because neither network-layer authentication nor IEEE 802.11i pre-authentication is used, the MN needs to engage in a full EAP authentication with AP1 to gain access to the network after the move (post-authentication). This experiment shows the effect of the absence of any kind of pre-authentication.

The second scenario involves 802.11i pre-authentication and involves movement between AP1 and AP2. In this scenario, the MN is initially connected to AP2, and starts IEEE 802.11i pre-authentication with AP1. This is an ideal scenario to compare the values obtained from 802.11i pre-authentication with that of network-layer assisted pre-authentication. Both scenarios use RADIUS as the AAA protocol (APs implement a RADIUS client). The third scenario takes advantage of network-layer assisted link-layer pre-authentication. It involves movement between two APs (e.g., between AP0 and AP1) that belong to two different subnets where 802.11i pre-authentication is not possible. Here, Diameter is used as the AAA protocol (PAA implements a Diameter client).

In the third movement scenario, the MN is initially connected to AP0. The MN starts PANA pre-authentication with the PAA, which is co-located on the AR in the new candidate target network (nAR in network A) from the current associated network (network B). After authentication, the PAA proactively installs two keys, PSKap1 and PSKap2, in AP1 and AP2, respectively. By doing the key installations proactively, the PAA preempts the process of communicating with the AAA server for the keys after the mobile node moves to the new network. Finally, because PSKap1 is already installed, AP1 immediately starts the 4-way handshake. We have used measurement tools such as ethereal and kismet to analyze the measurements for the 4-way handshake and PANA authentication. These measurements reflect different operations involved during network-layer pre-authentication.

In our experiment, as part of the discovery phase, we assume that the MN is able to retrieve the PAA’s IP address and all required information about AP1 and AP2 (e.g., channel, security-related parameters, etc.) at some point before the handover. This avoids the scanning during link-layer handoff. We have applied this assumption to all three scenarios. Because our focus is on reducing the time spent on the authentication phase during handoff, we do not discuss the details of how we avoid the scanning.

Types	802.11i Post-authentication		802.11i Pre-authentication		MPA-assisted Layer 2 Pre-authentication	
Operation	Non-Roaming	Roaming	Non-Roaming	Roaming	Non-Roaming	Roaming
Tauth	61 ms	599 ms	99 ms	638 ms	177 ms	831 ms
Tconf	--	--	--	--	16 ms	17ms
Tassoc+ 4way	18 ms	17 ms	16 ms	17 ms	16 ms	17 ms
Total	79 ms	616 ms	115 ms	655 ms	208 ms	865 ms
Time affecting handover	79 ms	616 ms	16 ms	17 ms	15 ms	17 ms

Figure 10: Results of MPA-Assisted Layer 2 Pre- and Post-Authentication

Figure 10 shows the timing (rounded off to the most significant number) associated with some of the handoff operations we have measured in the testbed. We describe each of the timing parameters below.

"Tauth" refers to the execution of EAP-Transport Layer Security (TLS) authentication. This time does not distinguish whether this authentication was performed during pre-authentication or a typical post-authentication.

"Tconf" refers to the time spent during PSK generation and installation after EAP authentication is complete. When network-layer pre-authentication is not used, this time is not considered.

"Tassoc+4way" refers to the time dedicated to the completion of the association and the 4-way handshake with the target AP after the handoff.

The first two columns in the figure show the results for non-roaming and roaming cases, respectively, when no pre-authentication is used at all. The second two columns depict the same cases when IEEE 802.11i pre-authentication is used. The last two columns show when we used network-layer-assisted layer 2 pre-authentication. When pre-authentication is used, only the factor Tassoc+4way affects the handoff time. When no pre-authentication is used, the time affecting the handoff includes Tauth (the complete EAP-TLS authentication) plus Tassoc+4way.

That is precisely the time affecting the handoff in the case where the MN moves from AP0 to AP1 in the absence of pre-authentication. As it is seen, these delays are not suitable for real-time applications. Indeed, for the non-roaming case, we obtained a ˜80-ms delay for re-establishing the connection with target AP1. It takes about 600 ms to complete the handoff when the MN moves to a visited domain and the home AAA server is located far away. However, network-layer pre-authentication is only affected by Tassoc+4way (˜17 ms) involving any kind of handoff authentication. As is evident, IEEE 802.11i pre-authentication provides a comparable benefit (˜16 ms) in terms of handoff but is limited to cases when APs are in the same Distribution System (DS). Additionally, network-layer pre-authentication leverages a single EAP authentication to bootstrap security in several target APs by allowing the MN to move among APs under the same PAA without running EAP and consequently without contacting the AAA server. In this sense, it extends IEEE 802.11r advantages over IEEE 802.11i by allowing inter-subnet and inter-domain and even inter-technology handoffs.

C.6. Guidelines for Handover Preparation

In this section, we provide some guidelines for the roaming clients that use pre-authentication mechanisms to reduce the handoff delay. These guidelines can help determine the extent of the pre-authentication operation that is needed based on a specific type of movement of the client. IEEE 802.11i and 802.11r take advantage of the pre-authentication mechanism at layer 2. Thus, many of the guidelines observed for 802.11i-based pre-authentication and 802.11r-based fast roaming could also be applicable to the clients that use MPA-based pre-authentication techniques. However, since MPA operations are not limited to a specific subnet and involve inter-subnet and inter-domain handover, the guidelines need to take into account other factors, such as movement pattern of the mobile node, cell size, etc.

The time needed to complete the pre-authentication mechanism is an important parameter, since the mobile node needs to determine how much ahead of time the mobile node needs to start the pre-authentication process so that it can finish the desired operations before the handover to the target network starts. The pre-authentication time will vary, depending upon the speed of the mobile node (e.g., pedestrian vs. vehicular) and cell sizes (e.g., WiFi, Cellular). Cell residence time is defined as the average time the mobile node stays in the cell before the next handoff takes place. Cell residence time is dependent upon the coverage area and velocity of the mobile node. Thus, cell residence time is an important factor in determining the desirable pre-authentication time that a mobile node should consider.

Since the pre-authentication operation involves six steps as described in Section 6.3 and each step takes some discrete amount of time, only part of these sub-operations may be completed before handoff, depending upon the available delay budget.

For example, a mobile node could complete only network discovery and the network-layer authentication process before the handoff and postpone the rest of the operations until after the handover is complete. On the other hand, if it is a slow-moving vehicle and the adjacent cells are sparsely spaced, a mobile node could complete all the desired MPA-related operations. Finishing all the MPA-related operations ahead of time reduces the handoff delay but adds other constraints, such as cell residence time.

We give a numerical example here, similar to [MISHRA04].

D = Coverage diameter

v = Mobile node's velocity

RTT = round trip time from AP to AAA server, including processing time for authentication (Tauth)

Tpsk = Time spent to install keys proactively on the target APs

If for a given value of D = 100 ft, Tpsk = 10 ms, and RTT = 100 ms, a mobile node needs to execute only the pre-authentication procedure associated with MPA, then the following can be calculated for a successful MPA procedure before the handoff is complete.

$$2RTT + Tpsk < D/v$$

$$v = 100\ ft/(200\ ms + 10\ ms) = \tilde{}500\ ft/sec$$

Similarly, for a similar cell size, if the mobile node is involved in both pre-authentication and pre-configuration operations as part of the MPA procedure, and it takes an amount of time Tconf = 190 ms to complete the layer 3 configuration including IP address configuration, then for a successful MPA operation,

$$2RTT + Tpsk + Tconf < D/v$$

$$v = 100 \text{ ft}/(200 \text{ ms} + 10 \text{ ms} + 190 \text{ ms}) = \tilde{}250 \text{ ft/sec}$$

Thus, compared to only the pre-authentication part of the MPA operation, in order to be able to complete both pre-authentication and pre-configuration operations successfully, either the mobile node needs to move at a slower pace or it needs to expedite these operations for this given cell size. Thus, types of MPA operations will be constrained by the velocity of the mobile node.

As an alternative, if a mobile node does complete all of the pre-authentication procedure well ahead of time, it uses up the resources accordingly by way of an extra IP address, tunnel, and extra bandwidth. Thus, there is always a tradeoff between the performance benefit obtained from the pre-authentication mechanism and network characteristics, such as movement speed, cell size, and resources utilized.

Authors' Addresses

Ashutosh Dutta (editor)
NIKSUN
100 Nassau Park Blvd.
Princeton, NJ 08540
USA

EMail: ashutosh.dutta@ieee.org

Victor Fajardo
NIKSUN
100 Nassau Park Blvd.
Princeton, NJ 08540
USA

EMail: vf0213@gmail.com

Yoshihiro Ohba
Corporate R&D Center, Toshiba Corporation
1 Komukai-Toshiba-cho, Saiwai-ku
Kawasaki, Kanagawa 212-0001
Japan

EMail: yoshihiro.ohba@toshiba.co.jp

Kenichi Taniuchi
Toshiba Corporation
2-9 Suehiro-cho
Ome, Tokyo 198-8710
Japan

EMail: kenichi.taniuchi@toshiba.co.jp

Henning Schulzrinne
Columbia University
Department of Computer Science
450 Computer Science Building
New York, NY 10027
USA

Phone: +1 212 939 7004
EMail: hgs@cs.columbia.edu