Problem Statement for Heterogeneous Handover
draft-ohba-mobopts-heterogeneous-requirement-01

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 30, 2006.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

This document describes the problem statement and a set of issues and requirement specific to heterogeneous handover. Careful consideration needs to be given to these set of requirements while
designing an optimized handover framework in a heterogeneous mobility scenario. These problem statements help determine the key functions that are responsible for delay during handover involving interdomain and heterogeneous access technology. Analysis of these issues and requirements can be applied to any existing mobility optimization framework.

Table of Contents

1. Introduction ................................................. 3
2. Handover Taxonomy .......................................... 4
3. Performance Requirements ................................. 8
4. Handover Delay Analysis ................................. 10
   4.1. Layer 2 delay .................................. 10
   4.2. Layer 3 Delay .................................. 11
   4.3. Application Layer Delay ......................... 12
5. Optimizing Handover Delay ............................... 13
6. Security Considerations ................................. 15
7. IANA Considerations ..................................... 16
8. Acknowledgments ........................................... 17
9. References .................................................. 18
   9.1. Normative References ............................. 18
   9.2. Information References ......................... 18
Authors’ Addresses ........................................... 20
Intellectual Property and Copyright Statements ........... 21
1. Introduction

Handover is a process by which a mobile node moves from one point of attachment to another point of attachment in a network. This handover can be primarily classified into homogeneous and heterogeneous handover. Heterogeneous handover includes movement between various types of wireless technologies and different administrative domains. Supporting terminal handovers across heterogeneous access networks such as CDMA, 802.11 and GPRS is a clear challenge, as each access network has different QoS, security and bandwidth characteristics. Similarly movement between two different kinds of domains poses a challenge since a mobile will need to re-establish authentication and authorization as well as in the new domain and each administrative domain may or may not have any prior security arrangement. In order to be able to provide an optimized handover solution for a heterogeneous handover case, it is essential to determine the key functions that contribute to the handover delay. Determining these key functions help define the right set of mechanisms needed to design an optimized handover framework. In this document, we introduce the concept of heterogeneous handover, illustrate the key performance characteristics requirement for a real-time communication, cite alternate scenarios that represent a heterogeneous handover involving different access technologies and different administrative domains. We discuss several challenges, problems and issues that give rise to handover delay, packet loss and jitter during the heterogeneous handover. These issues and problem sets can be used as guidelines while designing an optimized heterogeneous handover solution.
2. Handover Taxonomy

An access network is defined as the backhaul network that provides the first hop or last hop access to the mobile in an end-to-end communication system. When a mobile during an active communication session moves from one access network to another access network and changes its point of attachment it is subjected to disruption in the continuity of service. During the handover, as the mobile changes its point-of-attachment in the network a mobile terminal may end up communicating using its second interface in the new network, change its subnet or domain it is connected to. Based on the type of movement and access network we can primarily define the handover in the following ways. We can define following combination of handovers in our analysis. Primarily these could be categorized as A) Inter-subnet and B) Intra-subnet.

An Inter-subnet Handover can comprise the following combination of handover that fall into category A.

1) Inter-technology, Inter-domain
2) Inter-technology, Intra-domain
3) Intra-technology, Inter-domain
4) Intra-technology, Intra-domain

An intra-subnet handover can consist of the following types of handover that fall into category B.

1) Intra-Tech, Intra-domain
2) Inter-Tech, Intra-domain

Heterogeneous handover is a handover that requires authorization for acquisition or modification of resources assigned to a mobile and the authorization needs interaction with a central authority in a domain. In many cases an authorization procedure in a heterogeneous handover follows an authentication procedure that also requires interaction with a central authority in a domain. All the scenarios A-1, A-2, A-3, A-4, B-1 and B-2 described above could cover the scope of heterogeneous handover. These will be more clear as we describe different types of handoff scenarios in the following sections.
Figure 1: Handover Taxonomy

Intra-subnet: When a mobile moves between two radio access networks that are part of the same subnet and does not change broadcast domain, it is called Intra-subnet handover. During an intra-subnet handover, the mobile may be subjected to an inter-technology handover as well if both the access networks with different radio access technologies (e.g., CDMA, 802.11) belong to the same subnet. However, during an intra-subnet and inter-technology handover, the effective L3 identifier of the terminal may change if the terminal starts to communicate using an interface that is part of a different access network, since the IP address associated with each interface will be different although these addresses belong to the same subnet.

Inter-subnet: A mobile is subjected to an Inter-subnet handover when it moves between two radio access networks that belong to two different subnets. As a result, its L3 identifier (e.g., IP address) is changed thus giving rise to a need for any mobility management protocol such as Mobile IP [RFC3344], Mobile IPv6 [RFC3775], SIP-Mobility [SIPMM], HIP [I-D.ietf-hip-base] that can take care of the continuity of the existing application. An Inter-subnet handover can be viewed as orthogonal to intra-subnet handoff scenarios and may include intra-domain, Inter-domain, Inter-technology or Intra-technology handover. All of these types of handovers are described in this document. Inter-subnet handover potentially gives rise to
packet loss and jitter because of delay associated with transition at layer 2 and layer 3.

Inter-technology: A mobile may be equipped with multiple interfaces, where each interface can support different access technology (802.11, CDMA). A mobile may like to communicate with one interface at any time in order to conserve the power. During the handover the mobile 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 as well. This type of Inter-technology handover is often called as Vertical Handover since the mobile makes movement between two different cell sizes. A vertical handover can be termed as upward vertical handover or downward vertical handover based on the direction of movement such as smaller cell to larger cell or vice versa [vertical]. A mobile moving from 802.11 network to cellular network can be viewed as upward vertical handover. An inter-technology handover may affect the quality-of-service of the multimedia communication, since each access network offers different bandwidth.

Intra-technology: An intra-technology handover is defined when a mobile moves between 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 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 identifier depending upon type of movement. An intra-technology handover may involve networks with different access characteristics and thus may very well belong to heterogeneous handover category. However a handover between 802.11b networks if these belong to same subnet or domain may not be termed as heterogeneous handover.

Inter-domain: A domain can be defined in several ways. But for the purposes of roaming we define domain as an administrative domain which consists of networks that are managed by a single administrative entity which authenticates and authorizes a mobile for accessing the networks. An administrative entity may be a service provider, an enterprise and any organization. As a mobile moves between two administrative domains, it is also subjected to inter-subnet handover, as two different domains exclusively have two different subnets. 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. Inter-domain handover will be subjected to all the transition steps a subnet handover goes through in addition to an authentication process. These extra steps contribute to additional delay on the top
of delays due to regular subnet handover.

Intra-domain: When a mobile’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.
3. Performance Requirements

In order to provide desirable quality of service for interactive 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-R standards define the acceptable values for these parameters. For example for one-way delay, ITU-T G.114 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. Also if an out-of-order packet is received after a certain threshold it is considered lost. References [RFC2679], [RFC2680] and 2681 [RFC2681] describe some of the measurement techniques for delay and jitter.

An end-to-end delay consists of several components, such as network delay, operating system (OS) delay, CODEC delay and application delay. Network delay consists of transmission delay, propagation delay, queueing delay in the intermediate routers. Operating System related delay consists of scheduling behavior of the operating system in the sender and receiver. CODEC delay is generally caused due to packetization and depacketization at the sender and receiver end. Different CODECs will give rise to different delay, for example G.729 encodes speech at 8 Kbps and adds one-way delay of about 25 ms. Application delay is mainly attributed to playout delay that helps compensate the delay variation within a network. End-to-end delay and jitter values can be adjusted using proper value of the playout buffer at the receiver end. In case of interactive VoIP traffic, end-to-end delay affects the jitter value and is an important issue to consider. During a mobile’s frequent handover, transient traffic cannot reach the mobile and this may contribute to jitter even if the packet loss is avoided by means of buffering or some other forwarding mechanism. If the end system has a playout buffer, then this jitter is subsumed by the playout buffer delay, but otherwise this adds to the delay for interactive traffic. Packet loss is typically caused by congestion, routing instability, link failure, lossy links such as wireless links. During a mobile’s handover a mobile is subjected to packet loss because of its change in attachment to the network. Thus for both streaming traffic and VoIP interactive traffic packet loss will contribute to the service quality of the real-time application. Number of packets lost is proportional to the delay during handover and rate of traffic the mobile is receiving. Lost packets contribute to congestion in case of TCP traffic because of re-transmission, but it does not add to any congestion in case of streaming traffic that is RTP/UDP based. Thus it is essential to reduce the packet loss and effect of handover delay in any mobility management scheme.

We provide some instances of performance requirement here. The performance requirement will vary based on the type of application.
and its characteristics such as delay tolerance and loss tolerance limit. Interactive traffic such as VoIP and streaming traffic will have different tolerance for delay and packet loss. For example, according to ETSI TR 101 [ETSI] a normal voice conversation can tolerate up to 2% packet loss. If a mobile is subjected to frequent handover during a conversation, each handover will contribute to packet loss for the period of handover. Thus maximum loss during a conversation needs to be reduced to an acceptable level. There is no clear threshold value for packet loss for streaming application, but it needs to be reduced as much as possible to provide better quality of service to a specific application. Similarly there are other factors such as Transmission Rating Factor (R), End to End delay (one way mouth-to-ear) and Call blocking ratio that are some of the parameters that determine the QoS metrics. R factor is standardized by ITU-T G.107. R value is based on E Model. An R value of 50 is considered to be poor and a value of 90 can be considered as the best that provides most user satisfaction. As an example, a class B QoS which is equivalent to cellular telephony has a R factor that is greater than 70, E2E delay of less than 150 ms and call blocking ratio which is less than or equal to 0.15. Class A QoS that is the highest and is equivalent to fixed phone quality has an R value that is more than 80, E2E delay that is less than 100 ms. Similarly, 3GPP TS23.107 defines 4 classes: conversational, streaming, interactive and background. The streaming class has the packet (SDU) error rates ranging from 0.1 to 0.00001 and the transfer delay of less than 300ms. On the other hand, in ITU-T Y.1541, streaming is categorized as Class 4 out of existing 6 classes, where the upper bound of the transfer delay is 1 sec and the upper bound of packet loss ratio is 0.001, thus its delay tolerance is more than that of 3GPP’s. The delay variation is not specified in this class. On the other hand different vendor products suggest values up to 3 to 4 sec as tolerance value for the streaming traffic.
4. Handover Delay Analysis

This section analyzes the handover delay associated with heterogeneous handover and the factors that contribute to this. As a mobile goes through a heterogeneous handover process, it is subjected to handover delay because of the rebinding of properties at several layers of the protocol stack. There are several common properties that may contribute to the re-establishment or modification of these layers. These properties can mostly be attributed to things such as access characteristics (e.g., bandwidth, channel characteristics), access mechanism (e.g. CDMA, CSMA/CA, TDMA), configuration of layer 3 parameters, re-authentication, re-authorization, rebinding of security association at all layers.

During any handover procedure any multimedia application running within a client gets affected because of the delays incurred within each of the layers of the protocol stack. In the next subsections we provide analysis of several sub-components that give rise to the delay within each layer.

4.1. Layer 2 delay

Depending upon the access type (e.g., 802.11, CDMA), the mobile goes through several transition steps, before the new layer 2 link is re-established. As an example an 802.11 link goes through the process of scanning, authentication and association during the reattachment to the new 802.11 access point. Similarly other access networks such as CDMA and GPRS networks go through a series of state transitions during their reassociation to the new point of attachment. Each of these state transitions contributes to certain delays as the mobile goes through each of these intra-layer transitions. After a new layer 2 link is established, depending upon the type of handover scenario, other layers in the protocol need to go through the rebinding process. For example in case of intra-technology, intra-subnet handover, there is no other layer besides layer 2 that is affected and the handover delay is attributed to the delay due to layer 2 transitions only.

Even in the case of handoff consisting of intra-domain, intra-technology, intra-subnet handoff where there is layer 2 delay only, an authentication and authorization procedure needs to be performed when switching to the new access network. The procedure may take more steps if layer-2 characteristics such as QoS characteristics are different between the old and new access networks than the case where layer-2 characteristics are the same between the old and new access networks. So the difference of layer-2 access characteristics can qualify a intra-technology handover as a heterogeneous handover as well. Hence the scenario B-1 may also qualify to be a
heterogeneous handover as it may involve movement between 802.11b and 802.11n network or between CDMA1XRTT or CDMA1XEVDO networks. In case of movement between 802.11 type networks EAP authentication and 4-way handshake involving 802.11i can also add delay. Similarly in case of handoff between CDMA2000 family networks authentication procedure such as CHAP for PPP session may add to the layer 2 delay.

4.2. Layer 3 Delay

A layer 3 transition process goes through several steps such as new IP address acquisition, duplicate address detection, ARP update, and local authentication. Similarly a successful local authentication allows the mobile to send packets beyond the first hop router and may add delay to the handover process. For example in some cases there is some authorization process involved even before a new IP address is obtained. Methods of IP address acquisition are different based on if it is IPv4, IPv6 and access network is LAN or WAN. There are several protocols such as DHCP, DHCPv6, PPP or stateless autoconfiguration that can be used to take care of IP address acquisition. Each of these methods may take different amount of time for a layer 3 transition to complete. After a layer 3 transition is complete and layer 3 address is re-established, there are certain functions in the upper layer that need to be completed. These include functions such as sending the binding update from the mobile, media redirection at the sending host. Delays incurred due to these functions can be attributed to transport delay, processing delay at the end hosts etc. In the inter-subnet, intra-domain mobility category an authentication procedure is needed to obtain and use the IP address in the new subnet. But inter-subnet, inter-domain mobility may need additional authorization process. For example during an interdomain movement that includes MIPv4 as the binding protocol, AAA interaction will contribute to additional delay.

In a typical inter-domain mobility scenario independent of inter-technology or intra-technology handoff, an authentication and authorization procedure need to be executed for establishing layer 2 and layer 3 properties from scratch in the new domain. This can be an extraordinarily undesired delay component during handoff. References [multinetwork], [hybrid], [interdomain] are few examples of inter-subnet handover involving inter-domain mobility that provide some case studies. If the inter-subnet handover involves inter-domain mobility, then the delay introduced due to authentication and authorization procedure adds to the handover latency and consequently affects the ongoing multimedia sessions. The authentication and authorization procedure may include EAP authentication where an AAA server may be involved in EAP messaging during the handover. Depending upon the type of architecture, in some cases the AAA signals traverse all the way to the AAA server in the home domain of
the mobile as well before the network service is granted to the mobile in the new network. Thus it is desirable that interactions between the mobile and AAA servers must be avoided or be reduced during the handover. By way of expedited registrations, these interaction can be reduced. But it is more desirable if these can be avoided altogether during the handover itself.

4.3. Application Layer Delay

Application layer delay is the delay needed to re-establish or modify application layer properties involving end-to-end applications such as SIP specific sessions. Application layer binding update or any other mid-session signaling thus can be regarded as application layer delay. By way of application layer signaling any mid-session info such as IP address, codec parameters can also be changed.
5. Optimizing Handover Delay

It is evident from the analysis provided in the previous section that a mobile is subjected to different amount of delay based on the types of handover it is subjected to. For example an intra-subnet, intra-technology handover will take much less time than inter-subnet and inter-technology handover, because the mobile goes through a less number of state transitions during the handover process. Thus there are certain common properties such as characteristics of the access network, security association at each layers of the protocol stack during re-attachment, ways of layer 3 binding (e.g., obtaining IP address, binding update, media redirection), authorization, authentication in the new network play an important role to determine the overall handover delay. But the required steps can be minimized if already established security contexts in the domain are utilized for the procedure. Thus without any optimization technique, any type of heterogeneous handover will be subjected to higher handover delay compared to complete homogeneous handover (e.g., a combination of intra-technology (802.11b - 802.11b), intra-domain, intra-subnet).

Thus it is desirable to devise a mechanism or framework that helps reduce these handover delays either by executing these state transitions in parallel or completing some of these proactively ahead of time. A simple example may be to reduce the handover delay by completing the authentication and authorization ahead of time prior to the movement into the new network. Additional level of optimizations for other layer 2 and layer 3 related operation may also be possible. In some cases lower layer hints may help to optimize upper layer delays as well. There are several optimized mobility management solutions at different layers, [RFC4068], [RFC4140], [I-D.ohba-mobopts-mpa-framework], [SIPFAST], currently proposed that attempt to reduce these handover delay and mitigate the effect of handover delay.

In case of intra-domain movement, it might be possible to give authorization to the mobile on the use of every access network within the domain at the time of mobile’s initial entry to the domain so that no additional authorization is required for each heterogeneous movement as well as homogeneous movement. This approach still needs authentication for each access network in the domain during mobile’s movement to make sure that the entity attaching to the access network is the authorized mobile. Pre-authentication will help minimize such post-movement authentication delay for such an approach as well. For example during a movement to a CDMA1xEVDO network, a PPP connection cannot be set up until user authentication is complete using procedures such as CHAP or PAP with all the way to a AAA server in the home or a visited domain. In order to make the handover to a new PPP network faster pre-authentication using CHAP and PAP can take
place ahead of time allowing part of this phase to be set up ahead of time.
6. Security Considerations

This document describes the problem statement and issues associated with heterogeneous handover. Any framework supporting optimized heterogeneous handover will need to make sure that the signaling and data are properly secured during the handover. This security mechanism can be provided at any layer of the protocol stack.
7. IANA Considerations

This document has no actions for IANA.
8. Acknowledgments

Authors would like to acknowledge Kenichi Taniuchi, Victor Fajardo and Subir Das for many helpful discussions.
9. References

9.1. Normative References


9.2. Information References


Authors’ Addresses

Ashutosh Dutta
Telcordia Technologies
1 Telcordia Drive
Piscataway, NJ 08854
USA

Phone: +1 732 699 3130
Email: adutta@research.telcordia.com

Yoshihiro Ohba
Toshiba America Research, Inc.
1 Telcordia Drive
Piscataway, NJ 08854
USA

Phone: +1 732 699 5305
Email: yohba@tari.toshiba.com

Hidetoshi Yokota
KDDI R&D Labs
2-1-15 Ohara Fujimino-Shi
Saitama, Japan 356-8502
Japan

Phone: +81 49 278 7894
Email: yokota@kddilabs.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
Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment

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