Network Working Group H. Chan (Ed.)
Internet-Draft Huawei Technologies
Intended status: Informational October 19, 2010
Expires: April 22, 2011

Problem statement for distributed and dynamic mobility management
draft-chan-distributed-mobility-ps-00

Abstract

Mobility solutions deployed with centralized mobility anchoring in existing hierarchical mobile networks are more prone to the following problems or limitations compared with distributed and dynamic mobility management: (1) Routing via a centralized anchor is often longer, so that those mobility protocol deployments that lack optimization extensions results in non-optimal routes, affecting performance; whereas routing optimization may be an integral part of a distributed design. (2) As mobile network becomes more flattened centralized mobility management can become more non-optimal, especially as the content servers in a content delivery network (CDN) are moving closer to the access network; in contrast, distributed mobility management can support both hierarchical network and more flattened network as it also supports CDN networks. (3) Centralized route maintenance and context maintenance for a large number of mobile hosts is more difficult to scale. (4) Scalability may worsen when lacking mechanism to distinguish whether there are real need for mobility support; dynamic mobility management, i.e., to selectively provide mobility support, is needed and may be better implemented with distributed mobility management. (5) Deployment is complicated with numerous variants and extensions of mobile IP; these variants and extensions may be better integrated in a distributed and dynamic design which can selectively adapt to the needs. (6) Excessive signaling overhead should be avoided when end nodes are able to communicate end-to-end; capability to selectively turn off signaling that are not needed by the end hosts will reduce the handover delay. (7) Centralized approach is generally more vulnerable to a single point of failure and attack often requiring duplication and backups, whereas a distributed approach intrinsically mitigates the problem to a local network so that the needed protection can be simpler.

Status of this Memo

Table of Contents

1. Introduction

In the past decade a fair number of mobility protocols have been standardized. Although the protocols differ in terms of functions and associated message format, we can identify a few key common features:

- presence of a centralized mobility anchor providing global reachability and an always-on experience;
- extensions to optimize handover performance while users roam across wireless cells;
- extensions to enable the use of heterogeneous wireless interfaces for multi-mode terminals (e.g. cellular phones).

The presence of the centralized mobility anchor allows a mobile device to be reachable when it is not connected to its home domain. The anchor, among other tasks, ensures forwarding of packets destined to or sent from the mobile device. As such, most of the deployed architectures today have a small number of centralized anchors managing the traffic of millions of mobile subscribers. Coompared with a distributed approach, a centralized approach have several issues or limitations affecting performance and scalability, which require costly network dimensioning and engineering to fix them.

To optimize handovers for mobile users, the base protocols have been extended to efficiently handle packet forwarding between the previous and new points of attachment. These extensions are necessary when applications impose stringent requirements in terms of delay. Notions of localization and distribution of local agents have been introduced to reduce signalling overhead. Unfortunately today we witness difficulties in getting such protocols deployed, often leading to sub-optimal choices.

Moreover, all the availability of multi-mode devices and the possibility to use several network interfaces simultaneously have motivated the development of more new protocol extensions. Deployment will be further complicated with so many extensions.

Mobile users are, more than ever, consuming Internet content, and impose new requirements on mobile core networks for data traffic delivery. When this traffic demand exceeds available capacity, service providers need to implement new strategies such as selective traffic offload (e.g. 3GPP work items LIPA/SIPTO) through alternative access networks (e.g. WLAN). Moreover, the localization of content providers closer to the Mobile/Fixed Internet Service Providers network requires taking into account local Content Delivery Networks (CDNs) while providing mobility services.

As long as demand exceeds capactity, both offloading and CDN techniques could benefit from the development of more flat mobile architectures (i.e., fewer levels of routing hierarchy introduced into the data path by the mobility management system). This view is reinforced by the shift in users!_ traffic behavior, aimed at increasing direct communications among peers in the same geographical area. The development of truly flat mobile architectures would result in anchoring the traffic closer to point of attachment of the user and overcoming the suboptimal routing issues of a centralized mobility scheme.

While deploying [Paper-Locating.User] today!_s mobile networks, service providers face new challenges. More often than not, mobile devices remain attached to the same point of attachment, in which case specific IP mobility management support is not required for applications that launch and complete while connected to the same point of attachment. However, the mobility support has been designed to be always on and to maintain the context for each mobile subscriber as long as they are connected to the network. This can result in a waste of resources and ever-increasing costs for the service provider. Infrequent mobility and intelligence of many applications suggest that mobility can be provided dynamically, thus simplifying the context maintained in the different nodes of the mobile network.

The proposed work will address two complementary aspects of mobility management procedures: the distribution of mobility anchors to achieve a more flat design and the dynamic activation/deactivation of mobility protocol support as an enabler to distributed mobility management. The former has the goal of positioning mobility anchors (HA, LMA) closer to the user; ideally, these mobility agents could be collocated with the first hop router. The latter, facilitated by the distribution of mobility anchors, aims at identifying when mobility must be activated and identifying sessions that do not impose mobility management -- thus reducing the amount of state information to be maintained in the various mobility agents of the mobile network. The key idea is that dynamic mobility management relaxes some constraints while also repositioning mobility anchors; it avoids the establishment of non optimal tunnels between two anchors topologically distant.

This document discusses the issues with centralized IP mobility management compared with distributed and dynamic mobility management. A companion document [dmm-senario] discusses the use case senarios.

2. Conventions used in this document

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

3. Centralized versus distributed mobility management

Mobility management functions may be implemented at different layers of the OSI stack. At the IP layer, they may reside in the network or in the mobile node. In particular, network-based solution resides in the network only. It therefore enables mobility for hosts and network applications which lack mobility support in them but are already in deployment.

At the IP layer, a mobility management protocol to achieve session continuity are typically based on the principle of distinguishing between session identifier and routing address and maintaining a mapping between them. With Mobile IP, the home address takes the role of session identifier whereas the care-of-address takes the role of routing address, and the binding between them is maintained at the mobility anchor, i.e., the home agent.

Mobility management functions in the network may be centralized or distributed, as is explained in the next two subsections.

3.1. Centralized mobility management

With centralized mobility management, the mapping information is kept at a centralized mobility anchor, and packets destined to the mobile node are routed via this anchor. That is, such mobility management systems are centralized in both the data plane and the control plane.

Existing mobility solutions leverage on centralized mobility anchoring in a hierarchical architecture. Examples of such centralized mobility anchors are the home agent (HA) and local mobility anchor (LMA) in mobile IP [RFC3775] and proxy mobile IP [RFC5213] respectively. Current mobile networks such as UMTS network, CDMA network, and 3GPP SAE network also use centralized mobility management.

```
          UMTS                  3GPP SAE              MIP/PMIP
        +------+              +------+              +------+
        | GGSN |              | P-GW |              |HA/LMA|
        +------+              +------+              +------+
           /\                    /\                    /\
          /  \                  /  \                  /  \
         /    \                /    \                /    \
        /      \              /      \              /      \
       /        \            /        \            /        \
   +------+  +------+    +------+  +------+    +------+  +------+
   | SGSN |  | SGSN |    | S-GW |  | S-GW |    |FA/MAG|  |FA/MAG|
   +------+  +------+    +------+  +------+    +------+  +------+
```

Figure 1. Centralized mobility management.

3.2. Distributed mobility management

The mobility management functions may be distributed to multiple locations in different networks, so that a mobile node in any of these networks may be served by a close by mobility function (MF).

```
   +------+  +------+  +------+  +------+
   |  MF  |  |  MF  |  |  MF  |  |  MF  |
   +------+  +------+  +------+  +------+
                          |
                        ----
                       | MN |
                        ----
```

Figure 2. Distributed mobility management.

Distributed mobility management may be partially distributed, i.e., only the data plane is distributed, or fully distributed where both data plane and control plane are distributed. These different approaches are described in [I-D.dmm-scenario].

4. Problem statement

This section describes the problems or limitations in a centralized mobility approach and compares it against the distributed approach.

4.1. Non-optimal routes

Routing via a centralized anchor often results in a longer route. Figure 3 shows two cases of non-optimized routes. In the first case, the mobile node and the correspondent node are close to each other but are both far from the mobility anchor. In the second case, the mobile node is getting content from a local content delivery network (CDN). In both cases, packets destined to the MN have to be routed via the mobility anchor, which is not in the shortest path.

```
         MIP/PMIP
         +------+
         |HA/LMA|
         +------+
           /\ \  \                   +---+
          /  \   \    \              |CDN|
         /    \     \      \         +---+
        /      \       \        \      |
       /        \         \          \ |
   +------+  +------+  +------+   +------+
   |FA/MAG|  |FA/MAG|  |FA/MAG|   |FA/MAG|
   +------+  +------+  +------+   +------+
                          |          |
                        ----       ----
                       | CN |     | MN |
                        ----       ----
```

Figure 3. Non-optimized route when communicating with CN and when accessing local content.

In a distributed mobility management design, there are numerous anchors in different networks so that packets may be routed via a nearby mobility function, as shown in Figure 4.

```
                                     +---+
                                     |CDN|
                                     +---+
                                       |
                                       |
   +------+  +------+  +------+   +------+
   |  MF  |  |  MF  |  |  MF  |   |  MF  |
   +------+  +------+  +------+   +------+
                          |          |
                        ----       ----
                       | CN |     | MN |
                        ----       ----
```

Figure 4. Mobile node in any network is served by a close by mobility function.

With the centralized mobility anchor design, route optimization extensions to mobility protocols are therefore needed, and those mobility protocol deployments that lack such optimization extensions will encounter non-optimal routes, which affect performance. In contrast, route optimization may be an integral part of a distributed design.

4.2. Network architecture evolution

The centralized mobility management is currently deployed to support the existing hierarchical networks. It leverages on the hierarchical architecture. However, the volume of wireless data traffic continues to increase exponentially. The data traffic increase would require too much costly capacity upgrade of centralized architectures. It is thus predictable that the data traffic increase will soon overload the centralized data anchor point, e.g., P-GW in 3GPP/SAE. In order to address this issue, a trend in the evolution of mobile networks is to distribute the network functions close to the access gateways. These network functions can be the content servers in a Content Delivery Network (CDN) but also the data anchor point.

As mobile network becomes more flattened the centralized mobility management can become non-optimal.

Mobility management deployment with distributed architecture is then needed to support the more flattened network and the CDN networks. In addition, such distributed design may likely be able to also support the hierarchical network.

4.3. Centralized route and mobility context maintenance

Routes are set up to enable session continuity when handover occurs. Mobility contexts are also maintained to route via a mobility anchor.

Setting up such routes and maintaining the mobility context is more difficult to scale in a centralized design with a large number of mobile hosts. Distributing the route maintenance function and the mobility context maintenance function among different networks can be more scalable.

4.4. Need versus no need for mobility support

The problem of centralized route and mobility context maintenance is aggravated when the via routes are set up for many more mobile nodes that are not going to physically move out of a radio cell. For example, the user may be communicating at home, in one's office, or at a cafe. Much more resources are also wasted to maintain the mobility context of these routes. Similarly, intelligent applications that do not need network-layer mobility support coexist with applications lacking such intelligence.

Dynamic mobility management, i.e., to dynamically set up the via routes only for mobile nodes that actually undergoes handover and lacks higher-layer mobility support, is needed. With distributed mobility anchors, the mechanism to support dynamic mobility may then also be distributed. Therefore, dynamic mobility and distributed mobility may complement each other and may be integrated.

4.5. Numerous variants and extensions of MIP

Mobile IP (MIP) already has numerous variants and extensions including PMIP, FMIP, HMIP, and there may be more to come. Deployment can then become complicated, especially if interoperability with existing deployments is an issue.

Because the distributed mobility management needs to work both with network architectures of hierarchical networks as well as flattened networks, our design needs to be flexible enough to support different networks. In addition, one goal of dynamic mobility management is the capability to selectively turn on and off mobility support and certain different mobility signaling. Such flexibility in the design provides a good foundation to integrate the different mobility variants. Some additional extensions to the base protocols may then be needed to improve the integration.

4.6. Peer-to-peer communication

As MIPv6 Route Optimization (RO) mode enables a more efficient data packets exchange than the bidirectional tunneling (BT) mode, it is expected to be used whenever possible unless the mobile node is not interested in disclosing its topological location, i.e., care-of address, for the CN (e.g., for privacy reasons) or some other network constraints are put in place.

However, MIPv6 RO mode requires exchanging a significant amount of signaling messages in order to establish, and periodically refresh a bidirectional security association (BSA) between the mobile node and the correspondent node (CN). While the mobility signaling exchange impacts the overall handoff latency, the BSA is needed to authenticate the binding update and acknowledgment messages (note that the latter is not mandatory). In addition, the amount of mobility signaling messages increases further when both endpoints are mobile.

A dynamic mobility management capability to turn off these signaling when they are not needed will enable the RO mode between two mobile endpoints at minimum or no cost. It will reduce the handoff latency owing to the removal of the extra signaling. It will encourage its adoption and large scale deployment.

4.7. Single point of failure and attack

A centralized anchoring architecture is generally more vulnerable to a single point of failure or attack, requiring duplication and backups of the support functions.

A distributed mobility management architecture has intrinsically mitigated the problem to a local network which is then of a smaller scope. In addition, the availability of such functions in neighboring networks has already provided the needed architecture to support protection.

5. Security Considerations

TBD

6. IANA Considerations

None

7. Co-authors and Contributors

This problem statement document is a joint effort among the following participants in a design team. Each individual has made significant contributions to this work.

Dapeng Liu: liudapeng@chinamobile.com

Pierrick Seite: pierrick.seite@orange-ftgroup.com

Hidetoshi Yokota: yokota@kddilabs.jp

Charles E. Perkins: charles.perkins@tellabs.com

Melia Telemaco: telemaco.melia@alcatel-lucent.com

Hui Deng: denghui@chinamobile.com

Elena Demaria: elena.demaria@telecomitalia.it

Zhen Cao: caozhen@chinamobile.com

Wassim Michel Haddad: Wassam.Haddad@ericsson.com

8. References

8.1. Normative References

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

8.2. Informative References

[I-D.dmm-scenario]
Yokota, H., Seite, P., Demaria, E., and Z. Cao, "Use case scenarios for Distributed Mobility Management", draft-yokota-dmm-scenario-00 (work in progress), October 2010.

[Paper-Locating.User]
Kirby, G., "Locating the User", Communication International, 1995.

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

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

Author's Address

H Anthony Chan (editor)
Huawei Technologies
1700 Alma Dr. Plano, TX 75075, USA
Email: h.a.chan@ieee.org
-
Dapeng Liu
China Mobile
Unit2, 28 Xuanwumenxi Ave, Xuanwu District, Beijing 100053, China
Email: liudapeng@chinamobile.com
-
Pierrick Seite
France Telecom - Orange
4, rue du Clos Courtel, BP 91226, Cesson-Sevigne 35512, France
Email: pierrick.seite@orange-ftgroup.com
-
Hidetoshi Yokota
KDDI Lab

2-1-15 Ohara, Fujimino, Saitama, 356-8502 Japan
Email: yokota@kddilabs.jp
-
Charles E. Perkins
Tellabs Inc.
3590 N. 1st Street, Suite 300, San Jose, CA 95134, USA
Email: charles.perkins@tellabs.com
-
Melia Telemaco
Alcatel-Lucent Bell Labs
Email: telemaco.melia@alcatel-lucent.com
-
Wassim Michel Haddad
Ericsson
300 Holger Dr, San Jose, CA 95134, USA
Email: Wassam.Haddad@ericsson.com
-
Elena Demaria
Telecom Italia
via G. Reiss Romoli, 274, TORINO, 10148, Italy
Email: elena.demaria@telecomitalia.it
-
Hui Deng
China Mobile
Unit2, 28 Xuanwumenxi Ave, Xuanwu District, Beijing 100053, China
Email: denghui@chinamobile.com
-
Zhen Cao
China Mobile
Unit2, 28 Xuanwumenxi Ave, Xuanwu District, Beijing 100053, China
Email: caozhen@chinamobile.com
-