# Understanding Multi-Layer Mobility

**Sasu Tarkoma**
*Helsinki Institute for Information Technology, Finland*

**Jouni Korhonen**
*TeliaSonera Corporation, Finland*

## INTRODUCTION

*Mobility* is an important requirement for many application domains, where entities change their physical or logical location. Physical location denotes the real-world location of a device, whereas logical location is not necessarily dependent on the physical environment. Mobility support may be divided into several technical layers and also categories depending on the nature of mobility. In this article, we consider mobility protocols starting from the network layer (layer 3 in the OSI stack) and ending at the application layer (layer 7), and focus on physical mobility.

The most fundamental network-level protocols for supporting mobile hosts are the Mobile-IP protocols standardized by the IETF (Perkins, 2002; Johnson, Perkins, & Arkko, 2004). Another related network-level solution is *network mobility* (NEMO) (Devarapalli et al., 2004), in which complete sub-networks may change location as well as single hosts. Mobility can also be handled on the transport layer. *Transport layer seamless handover* (TraSH) (Fu et al., 2004), *datagram congestion control protocol* (DCCP) (Kohler, 2006), and mSCTP (Xing, Karl, Wolisz, & Müller, 2002) are recent examples of such solutions. Yet another way of managing host mobility is with mobility-aware *virtual private networks* (VPNs) such as MOBIKE-based IPSec VPNs (Kivinen, 2006). Protocols such as wireless CORBA (WCORBA) (OMG, 2004) and the *session initiation protocol* (SIP) (Schulzrinne & Wedlund, 2000) provide more fine-grained mobility than host based, and they do not assume underlying transport- or network-level mobility support.

Middleware support for mobility is required in order to provide location transparency for objects, agents, and other components; support efficient and reliable communication in wireless environments; and buffer messages and other data for disconnected operation. In addition, the middleware may support scalability and availability of resources and services.

Mobility is inherently tied with the way nodes are addressed in a distributed network. In this article, we examine three different ways to address mobile nodes and components: addresses with *location and identity, locator/identity split,* and *content-based addressing*. The first addressing model is used by the IP protocol. The second model is an extension of the first and used, for example, in the *host identity protocol* (HIP) and the i3 overlay (Stoica, Adkins, Zhuang, Shenker, & Surana, 2002). The third model has been proposed for expressive communication in ubiquitous environments.

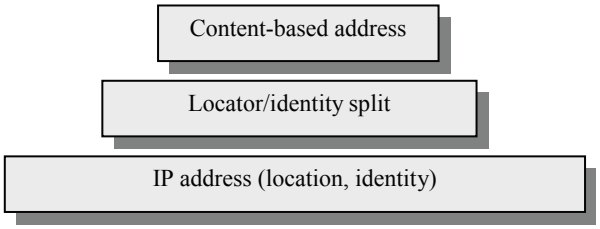
The aim of this article is to examine the addressing models and investigate cross-layer interactions of different mobility protocols. One of the interesting questions is how mobility should be handled and coordinated when there are multiple layers offering support for mobility. We also consider the case of the hop-by-hop routed layer-7 environment, implemented typically using SOAP (W3C, 2003), CORBA, or SIP in the telecommunications sector. These three technologies are the most frequently used, have differing characteristics and product bases, and contain the essence of middleware/application layer communication. SOAP is an abstract and generic messaging framework with extendable header system, allowing rich facilities for hop-by-hop propagation of messages.

## ADDRESSING MODELS

The way mobile and stationary nodes are addressed is crucial in how mobility is supported in a distributed system. We define three different addressing models for mobile systems (see Figure 1):

1. **Address with Both Location and Identity:** This form of addressing couples the communicating end-points to specific locations in a network. For example the IP address is used in both identifying a node and routing packets to it. This form of addressing typically uses a mediating stationary node to handle the mobility management and location updates for the mobile nodes.
2. **Address with Locator/Identity Split:** This way of addressing separates the identity of a node and the location of the node. This allows more flexible mobility support since the identity may be used to lookup the physical location of a node. For example the Internet Indirection Architecture (i3) and the HIP are based on this form of addressing.
3. **Content-Based Addressing:** This goes beyond locator/identity split, because it decouples the destination

*Figure 1. Three addressing models*




from both identity and location. The destination is no longer defined by a single identity, such as the IP address or a cryptographic public key, but rather it is defined by logical rules set by applications running on the destination host. The rules are applied on messages or packets in order to make forwarding decisions. This means that using content-based addressing, we have decoupled many-to-many communication. On the other hand, the realization of content-based communication is more complex and costly. The cost of mobility in content-based routing is high when compared with the other forms of addressing. Research systems such as Siena (Carzaniga, Rosenblum, & Wolf, 2000) and Rebeca (Mühl, Ulbrich, Herrmann, & Weis, 2004) use content-based addressing.

These addressing models are not orthogonal and may be applied on different layers of the communications stack. Since the current Internet is based on the IP protocol, it provides the baseline addressing with location and identity contained in the IP address. Above that, we may implement the locator/identity split using HIP or an overlay network such as i3. Content-based addressing is also implemented above IP using application-level routers.

Identity-based mechanisms may be extended to support anonymous communication and multicast. For example, i3 supports multicast using triggers and anonymity by chaining private and public triggers. Content-based routing may, on the other hand, be extended to support identity-based communication by subscribing public keys, for example.

The addressing models have differing notions of the addressing space, in which addresses are defined. These differences can be used to characterize the difference between identity-based addressing and content-based addressing. The identity vector (public key) is a point in the flat one-dimensional addressing space of an overlay system. The content-based address, which is defined using a logical rule, is a subspace of a multi-dimensional addressing space. This illustrates the main difference, which is the expressiveness of the communication. In essence, for IP mobility there is a single, fixed indirection point; for locator/identity split there is a single indirection point; and with content-based there are multiple indirection points.

## MOBILITY-ENABLING PROTOCOLS

U

### A Taxonomy

Host mobility happens when a host relocates to a new location in the network, thereby possibly causing a change of the underlying IP address. Since IP addressing is tied to the location, this may cause a fundamental change in routing for the relocated host. This host relocation is commonly referred to as a *handover.* Handovers are usually divided into two main categories: *horizontal handovers* and *vertical handovers*. A horizontal handover is commonly understood as a handover that takes place within the same access network technology. A vertical handover is handover that takes place across different access network technologies (and from the host's point of view, between different networking interfaces).

There are also two ways of doing the handover: *break-before-make* or *make-before-break.* The difference of these two approaches is whether the mobility-enabling protocol or the terminal implementation (in hardware, point of view) allows creating connectivity to the new access network or router before leaving the old access network or router.

The host may also have several active IP addresses, which is called *multi-addressing.* Multi-addressing may also be used to realize *multi-homing,* which generally means that the client is connected to two independent networks for increased reliability. Multi-homing is also needed when several different access network technologies are used simultaneously. Server-side resiliency is commonly realized by connecting services to multiple network providers. This is called site multi-homing.

User *mobility* happens when a user changes the host device or access host, which causes a change in the underlying physical address of the user. The device characteristics may also change, for example when the user changes from a PDA (personal digital assistant) to a laptop. An important subcategory of user mobility is *session mobility,* which allows the relocation of user sessions from one host to another. Session mobility is an important requirement for current and future mobile applications, in which instant messaging (IM), multimedia, and voice sessions, for example, are moved from one device to another.

Service or *application mobility* happens when a service relocates or resides on a mobile host that moves. Service mobility may be triggered by factors not related with a user, for example load balancing.

### Network Layer Solutions

The current solutions being standardized by IETF for network-layer mobility support are the Mobile IPv6 (MIP6) and Mobile IPv4 (MIP4) protocols. MIP is a layer-3 mobility protocol for supporting clients that roam between IP net-