Computer Networks 94 (2016) 390–413

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A survey on handover management in mobility architectures

Stefano Ferretti*, Vittorio Ghini, Fabio Panzieri

Department of Computer Science and Engineering, University of Bologna Mura Anteo Zamboni 7, 40127 Bologna, Italy

ARTICLE INFO

Article history:
Received 9 February 2015
Revised 14 October 2015
Accepted 5 November 2015
Available online 23 November 2015

Keywords:
Mobility management
Handover
Mobile applications
Multi-homing
Cross-layer protocols

ABSTRACT

This work presents a comprehensive and structured taxonomy of available techniques for managing the handover process in mobility architectures. Representative works from the existing literature have been divided into appropriate categories, based on their ability to support horizontal handovers, vertical handovers and multihoming. We describe approaches designed to work on the current Internet (i.e. IPv4-based networks), as well as those that have been devised for the "future" Internet (e.g. IPv6-based networks and extensions). Quantitative measures and qualitative indicators are also presented and used to evaluate and compare the examined approaches. This critical review provides some valuable guidelines and suggestions for designing and developing mobility architectures, including some practical expedients (e.g. those required in the current Internet environment), aimed to cope with the presence of NAT/firewalls and to provide support to legacy systems and several communication protocols working at the application layer.

1. Introduction

Over the last few years, several architectural solutions have been proposed to support users that connect to the Internet through a Mobile Node (MN). The main objective is to provide seamless communications, i.e. ensuring that if a MN changes its point of attachment to the Internet, while in movement, no communication interruptions are perceived at the application level, and if such interruptions occur, they do not significantly degrade the Quality of Service (QoS) delivered at the application level. While throughput remains a major goal of system design, the main concern of mobility architectures is how to best manage situations where a MN changes network. This event is currently referred to as *handover* (or *handoff*).

By default, current operating systems installed on smartphones adopt the following strategy for data transmission: one Network Interface Card (NIC) at a time is configured and employed to send data. If a WiFi network is available, the terminal switches to WiFi; otherwise a cellular network is utilized, if the latter is available too. During the handover, communications are interrupted. While the widespread use of current smartphones confirms that in general such a simple approach may be a viable solution, in some cases this strategy has some severe limitations. Just as an example (which is actually a true story), let us consider the case of a researcher working in an university campus composed of several buildings, all covered by a WiFi network. He is a commuter and, just before leaving to go home, he receives an important Voice over IP (VoIP) phone call. Since he is leaving to take the last train home, he decides to answer the call using his mobile phone; today, there are plenty of smartphone apps that offer very efficient VoIP services. At that moment, the device is connected through a WiFi network, but as he comes out of the building, the WiFi signal is lost; thus, the smartphone automatically switches to 4G without any handover management at the application level; this causes a first communication interruption. While moving, he passes through other buildings (hence, within their WiFi coverage); as a consequence, the smartphone switches back to WiFi (i.e. a

* Corresponding author. Tel.: +39 051 2094845.
E-mail addresses: s.ferretti@unibo.it (S. Ferretti), vittorio.ghini@unibo.it (V. Ghini), fabio.panzieri@unibo.it (F. Panzieri).

http://dx.doi.org/10.1016/j.comnet.2015.11.016
1389-1286/

second communication interruption occurs), and then back to 4G (i.e. yet another communication interruption) and so on. One might suggest that the employee should turn off the WiFi NIC before leaving, thus using the cellular NIC only; yet, a full 3G/4G coverage may not be available in all the various buildings he goes through. Indeed, when moving, there are cases where one needs to change NIC without breaking the communication at the application layer.

Handovers may actually require that a MN changes its connection from its current AP to another one (cell or WiFi AP), thus causing a reconfiguration at the datalink layer of the NIC in use (horizontal handover). However, the MN can also change the network technology, switching from one NIC to another, which causes a reconfiguration at higher network layers (vertical handover). Changing network means that the IP address associated to the MN changes as well; this has repercussions on the application layer, since in the current Internet, the IP address of a node usually plays the twofold role of MN locator and MN identifier. This "change of identity" causes a service interruption that requires more time than a simple network reconfiguration at the operating system level.

Various proposals have been put forward to deal with this problem. Some approaches described in these proposals suggest a decoupling of the node identifier from its address (locator), e.g. GLI-Split [1], HIP [2,3], Hi3 [4], ILNP [5], LISP [6], MILSA [7], NIIA [8,9], RANGI [10]. These approaches usually comply with future Internet visions, requiring some radical changes in the network architecture. Other approaches address the above mentioned "change of identity" issue by exploiting (and enhancing) protocols of the current Internet stack, e.g. ABPS [11], DCCP [12], SIP-IAPP [13,14], I-TCP [15], MMUSE [16], MPTCP [17,18], m-SCTP [19], MSOCKS [20], TCP-migrate [21]. These latter approaches can be classified based on their ability to support the use of a single NIC at a time, or the (possibly concurrent) use of multiple NICs. They may work at various levels in the network stack, or even use a cross-layer strategy employing different functionalities at different levels.

The plethora of available proposals reveals that there are many technical issues concerned with the main problem of mobility management, as well as different technologies that need to be supported, and several alternative ways of solving these issues and using the technologies available. There are solutions which are effective in principle, but that cannot be deployed in practice because, for example, (i) the protocols they use do not comply with the current Internet implementation, (ii) they cannot deal with the presence of Network Address Translation (NATs)/firewalls, and (iii) they are not able to cope with problems introduced by those (popular) existing applications that do not respect the Internet protocol stack stratification. Thus, there is a significant need to identify and state the main issues making up the whole problem, and to classify the possible approaches for mobility management. This paper illustrates these main issues, as well as experiences and lessons learned from systems and proposals available in the literature, and eventually provides a critical discussion that might help practitioners in devising a holistic solution for mobility management support.

In the rest of the paper, we give some background information on host mobility management services, review the main architectural solutions proposed in the literature, and come up with a classification that arranges solutions according to their design principles. We also discuss aspects that have an impact on their deployment in real scenarios and limit their applicability. It is worth mentioning that, while this paper was being written, new studies have been published on the same topic, e.g. [22]. However, these works mainly focus on the aforementioned future generation Internet and on locator-identifier separation mechanisms. Instead, our approach emphasizes multihoming, mobility, the possibility of easily switching from one NIC to another, the compatibility with the existing Internet and problems strictly concerned with the limitations of the current applications and the architectural solutions employed (e.g., presence of NATs, firewalls, violations of the protocol stratification). Moreover, in this paper we focus on host centric networks, i.e. the traditional host-based conversation model that is exploited in the current Internet. Thus, we do not consider neither information-centric networks [23] nor user-centric networks [24,25].

The main contributions of this work are the following.

1. We provide a comparative overview of the main architectural solutions for mobility support in wireless networks. All the considered systems are classified based on their main characteristics, offered features, and based on the level of network protocol stack they operate.
2. We provide some valuable guidelines for developing mobility architectures in the current Internet, summarised as follows. Firstly, proxies used in many applications (VoIP, Session Initiation Protocol (SIP)-based applications, and optionally HTTP-based applications [26]) should be upgraded/extended to cope with mobility issues. Secondly, NATs and firewalls have to be handled carefully. Thirdly, multihoming solutions should take into account that many widespread applications and their related protocols (e.g. applications based on (SIP)) do violate the layered structure of the Internet protocol stack. Solutions to this problem require the use of an external proxy and/or the modification of application messages.
3. We describe the main quantitative measures and qualitative indicators for evaluating mobility architecture, and classify all the presented approaches accordingly.

The remainder of this paper is structured as follows. Section 2 provides the background information and the basic definitions related to this topic. Section 3 presents the existing architectural solutions working with single NICs, while Section 4 discusses solutions that exploit multiple NICs. Section 5 gives a qualitative comparison of the host mobility architectures discussed in the paper. Finally, Section 6 provides some concluding remarks and the main guidelines for developing mobility architectures.

2. Main definitions and concepts

The aim of a host mobility architecture is to ensure that a MN can move seamlessly across different access networks, without any interruptions of the active network services. Before going into the details of such architectures, in this