Computer Networks 72 (2014) 113–126

Contents lists available at ScienceDirect

Computer Networks

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

Review Article

Considerations of IP multicast for load balancing in Proxy Mobile IPv6 networks

Tien-Thinh Nguyen *, Christian Bonnet

Department of Mobile Communications, EURECOM, 450 Route des Chappes, 06410 Biot, France

ARTICLE INFO

Article history:
Received 10 January 2014
Received in revised form 2 June 2014
Accepted 23 July 2014
Available online 30 July 2014

Keywords:
Load balancing
Multicast-based load balancing
Proxy mobile IPv6
IP multicast
Multicast listener mobility

ABSTRACT

Proxy Mobile IPv6 (PMIPv6), taking advantages of the network-based mobility management, enables mobility support for the mobile nodes (MNs) without requiring their involvement in mobility signaling. However, as a centralized mobility management, PMIPv6 relies on a central mobility entity, i.e., Local Mobility Anchor (LMA) to provide the mobility support. The LMA is responsible for maintaining the mobile node's (MN) reachability state and forwarding the traffic from/to the current location of the MN. As mobile traffic demand rapidly increases, it is easy to make the LMA a bottleneck and a single point of failure. Therefore, load balancing (LB) mechanism among LMAs is a promising solution for these issues. Although previous studies proposed several solutions for distributing the load among the LMAs, none of them considers the multicast service. From the fact that the multicast service is expected to be widely used for delivering multimedia traffic (which will account for the majority of mobile traffic), it can also be considered as a crucial load factor. As a result, the efficiency of the existing solutions may be degraded when considering multicast. Furthermore, applying the existing LB mechanisms can raise several issues for not only the ongoing unicast sessions but also the multicast ones. To tackle these issues, this paper proposes a new LB solution which mainly focuses on the multicast service. The experiments and the numerical results show that this solution helps to better distribute the load among the LMAs while greatly reducing the multicast service disruption as well as avoiding the influence on the ongoing unicast sessions. In addition, the proposed solution can co-operate with the existing proposals to improve the performance of the network.

1. Introduction

Nowadays, the mobile data services have become an essential part of many consumers' life [1,2]. So far, users are using their mobile devices (e.g., smartphones and tablets) not only for personal life (e.g., making voice/video calls, sending email, watching video/TV, playing online games, and so on) but also for work (general and job-specific work applications such as multimedia conferencing, and distance learning, etc.) on a regular basis [3–5]. As a result, the mobile data traffic has been nearly doubled each year during the last few years [6]. This trend is expected to continue in upcoming years, especially with the deployment of 4G networks. The increase in traffic is mainly driven by mobile video traffic: estimates say that the mobile video traffic will account for 66.5 percent of total data traffic by 2017 [1]. The wide usage of mobile data services has been driven by the variety of different reasons such as: the increasing number of mobile devices which become more and more powerful and intelligent,

* Corresponding author. Tel.: +33 4 93 00 82 15.
E-mail addresses: tien-thinh.nguyen@eurecom.fr (T.-T. Nguyen), christian.bonnet@eurecom.fr (C. Bonnet).

http://dx.doi.org/10.1016/j.comnet.2014.07.008
1389-1286/

the enhancement of wireless access technology in terms of coverage, speed and quality, as well as the explosion of mobile applications [6]. The mobility of the devices puts a new requirement on mobile operators to provide connectivity anywhere and at anytime. Moreover, providing consistent and continuous seamless services is required for satisfying user's expectations and fulfilling even the high application requirements in terms of service disruption on the move [7].

In this context, various IP mobility management protocols have been introduced by the Internet Engineering Task Force (IETF)[1] ranging from the host-based (including Mobile IPv6 (MIPv6) [8] and its extensions e.g., Hierarchical Mobile IPv6 (HMIPv6) [9]) to the network-based mobility approach (e.g., Proxy Mobile IPv6 (PMIPv6) [10] and Fast Proxy Mobile IPv6 (FPMIPv6) [11]). On one hand, PMIPv6, as a network-based mobility management, provides mobility for the mobile nodes (MNs) without their involvement. This means the network handles the mobility management on behalf of the MN. As a result, PMIP helps to avoid the complexity of the protocol stack in the MN as well as to reduce tunneling overhead (over the air) and handover latency compared to the MIPv6. It is achieved by introducing two new network entities, namely the Local Mobility Anchor (LMA) and the Mobile Access Gateway (MAG). The former, similar to the home agent (HA) in MIPv6, is in charge of tracking the location of the MN and redirecting the MN's traffic towards its current topological location. While the latter is responsible for detecting and registering the movement of the MN. On the other hand, both PMIPv6 and MIPv6 are the centralized mobility approaches, which rely on the mobility anchor to enable mobility support (LMA in PMIPv6 and HA in MIPv6). In PMIPv6, it is common to have a huge number of devices associated with the LMA. As the traffic demand is increased rapidly [1], a traffic bottleneck can be formed at the LMA. Consequently, the quality of the ongoing sessions could be degraded (e.g., longer queuing delay, and increased packet loss). Also, in a heavy load condition, the LMA can drop the new sessions. In this circumstance, mobile network operators may need to deploy multiple LMAs in a large PMIPv6 domain, so that the traffic can be distributed among the LMAs [10]. Yet, it is highly possible that some LMAs become overloaded while the others are underutilized. Thus, load balancing (LB) among the LMAs is needed.

Several LB proposals [12–15] have been introduced to allow the LMA to be dynamically assigned and changed according to the load of all LMAs in the domain. When an MN initially attaches to the domain, the LB will be executed to select the appropriate LMA in terms of load to serve this MN (namely proactive-MN approach). However, the varying session rate (of the existing MNs) and data rate (of the existing sessions) may cause load-unbalanced situation between the LMAs. In order to address this issue, the LB can be triggered when the load of an LMA exceeds a specified threshold (called reactive-MN approach). In this case, an MN will be selected to move from the overloaded LMA to a less loaded one. Yet, changing LMA causes some issues for the ongoing sessions such as service disruption and packet loss.

As Internet is widely deployed and spread across a large area, it carries a variety of common information resources and services. In a sharing world, the group communication service, which refers to the ability to send data to several receivers at the same time, is naturally becoming more and more important especially in some areas like multimedia distribution, gaming, and financial services, etc [16]. In this context, the scalability and bandwidth efficiency from the multicast routing make the IP multicast a remarkable solution from the application point of view to allow the mobile networks to deal with a huge number of traffic, particularly, in mobile environments where users usually share frequency bands and limited capacity [17]. However, its role has been neglected in all existing LB proposals. As such, the consideration of multicast in the existing LB mechanisms can bring several issues from both load balancing (efficiency degradation) and multicast service perspective (e.g., tunnel convergence problem [18] and service disruption).

For these reasons, a LB mechanism which takes the multicast service into account is needed. In this paper, we will introduce such a LB mechanism, the so-called multicast-based mechanism. The key idea is that by separating the multicast LB from the unicast LB, the proposed solution helps better distribute the load among the LMAs in runtime, thus, improving the efficiency of resource utilization. In more details, when an LMA is overloaded, a multicast session will be selected to move to a less loaded one. The LB will also be executed when a listener starts a new multicast session to select the appropriate LMA to serve this session. As a result, the proposed solution does not influence the ongoing unicast/multicast sessions (except the selected session with which the multicast service disruption, in most cases, satisfies the requirements for the real-time services [19]).

As this article is an extension of [20], we will make a quick view on the issues caused by applying multicast in the existing proposals as well as the multicast-based solution. We will discuss in detail the criteria for the selection of the LMA and the multicast session. Next, the performance analysis will be done regarding the LMA load and the multicast service disruption. Finally, we will evaluate the multicast-based solution in terms of load distribution among LMAs using a near-to-real testbed. The testbed which is a combination of virtual machines and the network simulator NS-3 [21] has been deployed to reduce the hardware cost and to provide more flexible experiment while allowing to obtain the realistic results. It is noted that this paper mainly focuses on the multicast listener.

The rest of this paper is organized as follows. Section 2 presents the existing LB mechansims as well as the issues when considering multicast with these mechanisms. Section 3 introduces the multicast-based LB as well as the criteria for the LMA and multicast session selection. Section 4 presents the performance analysis regarding LMA load and multicast service disruption. Section 5 takes a look on the experiment testbed including the testbed description, the experiment scenarios and the collected results. Section 6 discusses the limitations of the proposed solution as well

[1] IETF, http://www.ietf.org.