Computer Communications 62 (2015) 13–22

Contents lists available at ScienceDirect

Computer Communications

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

H-Pastry: An inter-domain topology aware overlay for the support of name-resolution services in the future Internet

Nikos Fotiou [a,*], Konstantinos V. Katsaros [b], George Xylomenos [a], George C. Polyzos [a]

[a] Mobile Multimedia Laboratory, Department of Informatics, School of Information Sciences and Technology, Athens University of Economics and Business, 76 Patision Str., Athens 104 34, Greece
[b] Department of Electronic & Electrical Engineering, University College London, WC1E 7JE, Torrington Place, London, UK

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 2 February 2015
Accepted 7 February 2015
Available online 17 February 2015

Keywords:
Information-centric networking
Internet of Things
Name resolution
DHT

ABSTRACT

Overlay networks are widely used for locating and disseminating information by means of custom routing and forwarding on top of an underlying network. Distributed Hash Table (DHT) based overlays in particular, provide good scalability and load balancing properties. However, these come at the cost of inefficient routing, caused by the lack of adaptation to the underlying network, as DHTs often overlook physical network proximity, administrative boundaries and/or inter-domain routing policies. In this paper we show how to construct a DHT-based overlay network that takes all these aspects into account, so as to ease the global deployment of Future Internet architectures which require large-scale name resolution, such as Information-Centric Networking (ICN) and the Internet of Things (IoT). Based on the Pastry distributed object location and routing substrate and the Canon paradigm for multi-level DHTs, we developed H-Pastry, an overlay DHT scheme that harvests the scalability and load balancing features of DHTs, while also adapting to the underlying network topology, administrative structure and routing policies. We evaluate the performance characteristics of the proposed scheme through an extensive set of detailed simulations over realistic inter-network topologies. Our results show that H-Pastry substantially improves routing by reducing both overlay path stretch (by up to 55%) and routing policy violations (by up to 70%), compared to the Canonical (multi-level) Chord DHT. In addition, the design of H-Pastry keeps traffic within administrative boundaries as far as possible, reducing inter-domain hops by up to 27% compared to Pastry, while also creating excellent opportunities for the support of caching and multicast.

1. Introduction

Overlay networks are virtual networks formed by servers operating above the existing inter-networking infrastructure i.e., their functionality is based on higher layers of the protocol stack than those supported by the network layer, and they can be operated by third-parties [1]. Among several overlay network designs, *Distributed Hash Tables* (DHTs) have attracted considerable attention due to important structural advantages, such as their logarithmic scalability and their inherent load balancing capabilities. These characteristics have been considered especially advantageous for the support of emerging Future Internet architectures. In the context of the relatively recent Information-Centric Networking paradigm [2], the network is responsible for locating and delivering the information objects requested by end-hosts, thus necessitating the availability of a scalable name resolution service e.g., [3–8]. Considering that (a) the current number of unique web pages indexed by Google is greater than 1 trillion [9] and that (b) billions [10,11] of devices (mobile phones, sensors, home appliances, etc.) will be offering additional content to future networks, one should expect that in the context of ICN, any name resolution approach will have to handle unique information objects (IOs) in the order of 10^{13}. Some studies raise this estimate even further to 10^{15} [12]. But even beyond the ICN paradigm, the current expectations for even up to 50 billions of interconnected devices, i.e., sensors and actuators, in the envisioned Internet of Things/Everything (IoT/IoE), call for lookup services of corresponding scalability [11,13].

Even though DHTs present the desired scalability properties for such a task, the resulting overlay routing is, in principle, solely based on the logical organization of the overlay nodes, often neglecting their location in the physical network. This translates to routing schemes that ignore one or more of the following aspects: (i)

* Corresponding author. Tel.: +30 210 8203 693; fax: +30 210 8203 686.
E-mail addresses: fotiou@aueb.gr (N. Fotiou), k.katsaros@ucl.ac.uk (K.V. Katsaros), xgeorge@aueb.gr (G. Xylomenos), polyzos@aueb.gr (G.C. Polyzos).

http://dx.doi.org/10.1016/j.comcom.2015.02.003
0140-3664/

physical network proximity, (ii) administrative domain boundaries, (iii) inter-domain routing policies. Several approaches, such as Pastry [14], take physical network proximity into account by incorporating proximity metrics (e.g., hop count, RTT) in the overlay construction process. However, they do not consider administrative domain boundaries. The Canon paradigm takes DHT design one step further by enabling the construction of multi-level DHT-based overlay networks [15]. Canon enables the progressive merging of individual DHT constructions, assuming a hierarchical structure for the inter-domain topology. The resulting overlay networks, usually termed *Canonical*, not only achieve scalability and efficient load balancing, they also facilitate the deployment of fault isolation and security mechanisms, effective caching and bandwidth utilization, hierarchical storage, and hierarchical access control [15].

Nevertheless, even Canonical DHT designs present significant inefficiencies. The Canon DHT merge process was illustrated based on the Chord DHT [16], yielding Crescendo [15], along with sketches of the Canonical versions of several other DHT's, except Pastry. All these designs inherently neglect physical network proximity. To adress this issue, the Crescendo (Prox.) variant offers *proximity adaptation* based on the creation of groups of densely connected nodes sharing the same prefix [15,17]. However, proximity adaptation is not integrated in the overlay construction process and the corresponding overhead in terms of signaling and routing state has not been investigated. Furthermore, the Canon approach to multi-level DHTs was tailored to a strictly hierarchical domain-level network structure, which limits its applicability in the case of the Internet: as reported in [18], the increasing number of peering relationships as well as multihoming, result in an inter-domain graph that is not strictly hierarchical. These inter-domain relationships are not reflected in the overlay structure, leading to routing policy violations (as further illustrated in Section 4). This issue was addressed in [19] in a completely different networking context, where the DHT construction completely ignores the underlying routing substrate, resulting in highly stretched paths.

As a result, overlay paths in currently available DHT designs tend to be considerably longer than their underlay counterparts, often unnecessarily crossing administrative domain boundaries and/or violating established inter-domain routing policies. These facts actually constitute the basis of the polemic of ISPs against overlays and "justify" their efforts in cutting off overlay traffic whenever possible (e.g. through firewalls, deep packet inspection techniques, etc.) [20]. Such inefficiences become of paramount importance when the supported services represent core network functionalities. In the context of ICN architectures and/or the envisioned IoT/IoE, access to any type of information or device is expected to lead to increased volumes of resolution traffic[1] magnifying the impact of inefficient overlay paths on both the quality of experience (QoE) at the edge (e.g., name-resolution delays) and the overall utilization of resources inside the network. In this respect, the need for a DHT design that presents both the highly desired scalability and load balancing properties, and efficient overlay paths that adapt to the structure of the underlying inter-domain network topology, becomes more than apparent.

To this end, in this paper, we present *Hierarchical Pastry* (H-Pastry), a multi-level DHT scheme that aims to bring together the benefits of the Pastry DHT and the Canon approach, adding also support for multihoming and peering relationships. We explore the advantages of the proposed design, based on a series of detailed packet-level simulations, paying particular attention to the structure of the underlying inter-domain network graph and the established inter-domain routing policies. To the best of our knowledge, this is the first work to investigate the performance of DHT-based overlay routing in this context. Our results show that, by taking network proximity into account H-Pastry yields shorter overlay paths, reducing the perceived path stretch by approximately 50% (on average) compared to Crescendo. At the same time, H-Pastry achieves a 67% reduction of inter-domain routing policy violations compared to Crescendo (on average), and reduces inter-domain hops by an average of 23% compared to Pastry, thus better respecting the underlying administrative domain boundaries. Some of the performance benefits of H-Pastry were previously investigated in [6], in the context of a direct comparison between lookup-by-name and route-by-name inter-domain name resolution systems for ICN. In this paper, we present the detailed design of H-Pastry and further demonstrate the resulting performance benefits with a comprehensive performance evaluation of H-Pastry in comparison to other DHTs.

In the following, we first provide background information for Pastry and Canon, introducing their main features and highlighting the addressed inefficiencies (Section 2). We then proceed with a detailed description of the proposed design (Section 3). In order to investigate the performance characteristics of H-Pastry, we engaged in extensive simulations, comparing H-Pastry against alternative approaches i.e., Pastry, Chord and Crescendo. The results of the performance evaluation are presented in Section 4. Finally, we provide a discussion on related approaches in Section 5 and conclude in Section 6.

2. Background

2.1. Pastry

In the Pastry DHT, every node is assigned, uniformly and randomly, a unique identifier (ID). Each ID is a 128-bit number (though other lengths may be used), handled as a sequence of b-bit digits (b is a configuration parameter with a typical value of 4). Given a message and a key, Pastry routes the message to the node with the ID that is numerically closest to the key, in less than $log_{2^b}N$ steps, where N is the number of nodes in the DHT.

The routing state of each Pastry node is organized in two routing structures: the *Routing Table* and the *Leaf Set*. A Routing Table is organized in $\frac{128}{2^b}$ rows, with each row containing $(2^b - 1)$ entries. The entries at row i refer to nodes whose IDs share only the first i digits with the current node's ID (e.g., row 0 contains node IDs that do not share any digit with the current node's ID, row 1 contains node IDs whose first digit is the same as the current node's ID but the second is different, and so on). The Leaf Set contains $|L|$ entries for nodes whose IDs are the numerically closest to the present node's identifier. The set is split in two parts with $|L/2|$ entries for numerically smaller IDs and $|L/2|$ for numerically larger IDs. A typical value for $|L|$ is 2^b.

A Pastry node routes a message with a key K as follows: initially it examines if K falls within the range of the node IDs stored in the Leaf Set; if so, the message is forwarded to the node with the ID closest to K, which is the destination node. Otherwise, the Routing Table is used i.e., the message is forwarded to a node that shares a common prefix with K and this prefix is at least one digit longer than the common prefix of K and the current node's ID.

During routing state creation and maintenance, Pastry takes network locality into consideration i.e., among equally qualified candidate Routing Table entries, the one corresponding to the closest node in the network (with respect to the employed proximity metric, e.g., hop count, RTT, etc.) is selected. This feature yields Pastry's *short routes* property and constitutes one of the reasons

[1] This is due to both the high volume of information objects and devices, and the lack of a host-centric mode of communication where requests are directly targeted to specific end-hosts (e.g., file requests from a specific web server), which nevertheless allows for the seamless support of in-network caching, multicast forwarding and mobility [21].