Computer Communications 66 (2015) 14–24

Contents lists available at ScienceDirect

Computer Communications

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

# Towards locality-aware DHT for fast mapping service in future Internet

Peng Wang*, Julong Lan, Yuxiang Hu, Shuqiao Chen

*National Digital Switching System Engineering and Technological Research and Development Centre (NDSC), Zhengzhou, Henan 450000, China*

ARTICLE INFO

*Article history:*
Received 26 May 2014
Revised 22 April 2015
Accepted 23 April 2015
Available online 12 May 2015

*Keywords:*
Future Internet
Mapping service
Identifier/locator separation
Locality-aware
MDP

ABSTRACT

The identifier/locator separation has been shown to be critical for the design of future Internet. A key aspect of the identifier/locator separation is to design an identifier-to-locator mapping service to map identifiers onto locators. Although several mapping services have been presented in previous works, they either are designed based on aggregable identifiers, or suffer from high resolution latency. That is, they hardly meet the demands of the future Internet, which is desired to support fast mapping and self-certifying flat identifiers. In this paper, we propose LMChord, a fast mapping service that is based on the idea of locality-aware and hierarchical Distributed Hash Table (DHT). To address the mismatch problem between overlay and physical network, we present the LMChord construction model, which models the LMChord construction process as a Markov decision process (MDP). Moreover, we present a Markov decision construction algorithm, which improves reinforcement learning to get the global optimal or near-optimal construction strategy. To further improve routing efficiency, we also modify the finger table to optimize the LMChord's routing hops. We show that, besides the capability to support incremental deployment and flat identifiers, the mapping scheme is more scalable and has lower resolution latency. The evaluation also demonstrates the performance of our approach.



## 1. Introduction

With the popularity of smart devices and diversified networking applications, the current Internet has become an indispensable infrastructure of our society. However, it has been widely recognized that the current Internet is facing some serious challenges such as scalability, mobility, security, and multi-homing [1–3]. There is a common consensus that the overloading of Internet protocol (IP) address semantics is one of the main causes [4]. That is, IP addresses are not only names that identify communication hosts, but also routing locators that are used in the network layer protocols to forward packets and locate the destination hosts. In order to address these challenges, the identifier/locator separation architectures have been proposed and generally considered as a promising technology for the future Internet [5,6]. The identifier/locator separation decouples the tight relationship between the identifier and locator, and bridges between them with a mapping service.

A critical challenge with the identifier/locator separation is how to perform the mapping service. Several proposals have been proposed to design the mapping system based on distributed hash table (DHT) [7–11]. In [7], the authors proposed LISP-DHT which uses a DHT mechanism to register/resolve identifier-to-locator mappings. LISP-DHT uses the highest ID in an announced ID prefix as ChordID and lets domains announcing more than one prefix enter the ring more than one instance, each having the ChordID associated to their respective prefix. This assures that a node will be the one managing the prefix it announces. The authors in [8] proposed to distribute identifier-to-locator mappings to many resolution nodes using DHTs. Different from [7,8] constructs these nodes into a content-addressable network [12]. In [9], the authors further proposed an enhanced DHT based mapping system to permit endpoints to choose freely where to store their mappings. Although DHT-based mapping systems exhibit many interesting properties such as scalability, self-configuration, and robustness which are desirable for a mapping system, many conventional DHTs such as CAN [12] and Chord [13] do not take into account the proximity of nodes in the physical network, which causes significant performance degradation with the size of underlying network increasing [14–17]. These non-locality-aware DHT schemes lead to the mismatch between logical nodes and physical nodes. That is, the adjacent nodes in overlay network often are not close to each other in physical network. Therefore, a logical shortest path may not be the physical shortest path. The mismatch problem is precisely the main reason of high mapping resolution latency.

To address this issue, we present a fast global mapping resolution service at Internet scale in this paper, and describe and evaluate a specific locality-aware Multi-level Chord (LMChord) scheme for achieving a good balance between scalability, low update/query

* Corresponding author.
*E-mail address:* wangpeng.ndsc@gmail.com (P. Wang).

http://dx.doi.org/10.1016/j.comcom.2015.04.003

latency, and incremental deployment. In particular, we propose a new mapping system based on LMChord. The framework of our mapping system is arranged in a nested, hierarchical structure which reflects the underlying network topology. The nested, hierarchical structure not only improves the scalability, but also minimizes the routing stretch inefficiencies and reduces the mapping resolution latency. We also present the LMChord construction model, which models the LMChord construction process as a Markov Decision Process (MDP), and define the reward function combining logical delay with physical distance. Then we present a Markov decision construction algorithm called MDC. The algorithm improves reinforcement learning to get the global optimal or near-optimal policy which constructs an overlay network whose topology is consistent with the physical network. The simulation results show that our approach can remarkably reduce the mapping resolution latency and improve the routing performance in identifier/locator separation routing. It is worthy of noting that LMChord can be incrementally deployed in the current Internet because of the hierarchical, domain-based structure.

The rest of the paper is structured as follows. Section 2 outlines the background and motivation for LMChord. The framework of LMChord based mapping system is described in Section 3. The overview of LMChord based mapping system and the optimal design of LMChord is presented in Sections 4 and 5. In Section 6, we present detailed simulation evaluation results. Finally, we have the concluding remarks in Section 7.

## 2. Background and motivation

### 2.1. Requirements of the future Internet

The following requirements should be fulfilled by a mapping system. First, the mapping system should support the scalability of the future Internet. The Internet has been growing fast in the last three decades in terms of hosts and networks. We expect this process to continue due to the increasing ubiquity of the Internet applications and smart devices. Therefore, the number of identifiers will also increase in the future. As we could not foresee the tremendous growth of the Internet in the past, the mapping system must be able to handle a similar growth in the future. In addition, the flat identifiers would lead to substantially more number of identifier-to-locator mapping entries. The mapping scheme needs to scale to the order of billions of entries instead of thousands. Second, the mapping system should support structure-less, flat identifiers. With the research evolving, more and more works propose to use flat identifiers because the use of self-certifying flat identifiers can benefit for privacy and security. Third, the mapping resolution latency should be very low. On one hand, the real-time applications have stricter requirement for the mapping resolution latency. On the other hand, since mobility is directly handled using dynamic identifier to locator mapping, the latency requirement is much stricter in the future Internet. Finally, to support fast mobility, the mapping system requires that the identifier-to-locator mappings should update at a time-scale smaller than the inter-query time.

To meet the above requirements, the mapping system calls for a fundamental shift from traditional DHT mechanisms. The DHT-based schemes and optimized variations can perfectly solve the scalability problem, but they faced with severe challenge of high mapping resolution latency. In addition, most of the traditional DHT-based schemes assume that the identifiers are aggregable. Hence, we should redesign the mapping system to meet the above requirements.

### 2.2. Related work

As the research about the future Internet architectures goes deeply, there has been a flurry of work in the literature that is related to our research. Given that the paper space is limited, we cannot enumerate all the related work. The interested readers could refer to [5,18,19] and the references therein for more details. In this paper, we only focus on these most related works to our research.

To address the scaling issues in the Internet, the identifier/locator separation paradigm has been widely discussed and recognized as a promising solution. Since then, a number of proposals have been presented, and some have been already deployed in international test beds [20,21]. Although all the proposals improve the Internet architecture in a way or another by some form of identifier/locator separation, they also introduce mapping systems to distribute and store bindings between identifiers and locators which is the key critical component that can make the difference.

The design of mapping systems is not a new problem. There have been several proposals to address this issue [22]. While each of these proposals has its pros and cons, a common assumption of them is that the identifiers are aggregable. However, more and more works such as HIP [20], AIP [23], and MobilityFirst [24] propose to use self-certifying flat identifier which makes these mapping systems inapplicable. There are some recent mapping system design proposals that incorporate flat identifier space such as DHT-MAP [8], SLIMS [25], and DMAP [26]. However, DHT-MAP and SLIMS either incur high resolution latency, making it inapplicable to latency-sensitive applications, or prohibitive management overhead which limits scalability. DMAP is a direct mapping scheme based on the principle of shared hosting of the identifier to locator mappings among the entire ASs in the network. This assumption may be applicable for the future Internet, but restricted in the current Internet.

As stated above, most of the proposed mapping systems are based on flat, structured DHT such as CAN [12], Chord [13], Pastry [27], and Tapestry [28]. These DHT based mapping systems are scalable solutions for handling flat identifiers and provide several desirable properties such as robustness, self-configuration, and self-maintenance. However, the main problem with mapping systems based on flat DHT, as also concluded by Cox et al. [29], is that they suffer from high mapping resolution latencies when distributed globally. The cause of high latencies is that the DHT typically has no topological information and is not topologically embedded. In [30], the authors present a comprehensive study of the network locality properties of a DHT overlay network, and prove that it is possible to efficiently exploit network proximity in self-organizing DHT substrates. The importance of locality properties in overlay performance has been explored in [31]. To exploit network proximity, some modifications to CAN, Pastry, and Tapestry are made to provide locality to some extent [15]. However, these DHTs needs significantly more expensive overlay maintenance protocols compared with original DHTs. Wu et al. [32] proposed one hop DHT-based identifier-to-locator mapping service to achieve efficient flat identifier resolution. Each server in the proposed system maintains the full routing state. This will incur prohibitive storage overhead with the increasing of the size of system.

To decrease the storage overhead and communication cost between end hosts, the authors in [15] proposed the mOverlay, a locality-aware overlay that assigns nodes to groups depending on their distance based on dynamic landmark in the underlying network. mOverlay is a two-level hierarchical network. The top level consists of groups and the bottom level consists of hosts within groups. Similarly, Plethora, a two-level overlay P2P network was presented in [33]. A local overlay in Plethora acts as a locality-aware cache for the global overlay, grouping nodes close together in the underlying network. Shen and Xu [34] proposed locality aware randomized load balancing algorithms to take advantage of the hierarchical structure of Cycloid [35] to cope with