基于活跃度的分级映射解析系统

伊 鹏 王 鹏* 申 涓 张校辉 兰巨龙
(国家数字交换系统工程技术研究中心 郑州 450002)

摘 要：针对当前映射解析系统存在的映射解析时延过高的问题，该文依据终端的活跃程度，提出一种基于活跃度的分级映射解析系统。该系统将通信对端的身份位置映射信息划分为活跃级、中性级和稳定级 3 个等级，并据此建立了一种 3 层的映射解析存储架构，映射副本可根据自身活跃度的变化在 3 层之间动态调整存储位置。为最小化映射解析时延，在系统构建过程中，针对传统 DHT 构建方式存在的非位置感知问题，将系统构建过程建模为马尔科夫决策过程，并提出一种马尔科夫决策构建算法用于求解该模型。仿真结果表明，该系统能够显著降低映射解析时延和提升路由性能，对网络结构的动态变化具有良好的适应性。

关键词：映射系统；身份与位置分离；移动性；分级映射
中图分类号：TP393 **文献标识码**：A **文章编号**：1009-5896(2017)04-0832-08
DOI: 10.11999/JEIT160526

A Hierarchical Mapping Resolution System Based on Active Degree

YI Peng WANG Peng SHEN Juan ZHANG Xiaohui LAN Julong
(*National Digital Switching System Engineering & Technological R&D Center*, *Zhengzhou* 450002, *China*)

Abstract: To deal with the high resolution latencies in current existing mapping system, a hierarchical mapping system is proposed based on active degree. In the system, the mappings between the identifiers and locators are divided into three levels: active level, neutral level, and constant level. Based on these, a three tiers system architecture for mapping entries storing and resolving is designed. Stored mapping entries in different levels vary with the different active degrees of the remote communication terminal，and flow from one level to another. In order to minimize the mapping resolution latency, the construction model is proposed, which models the system construction process as a Markov Decision Process (MDP). Moreover, a Markov decision construction algorithm is proposed, which improves reinforcement learning to get the global optimal or near-optimal construction strategy. The simulation results show that the system has low resolve latency and good adaptability for network topology dynamic changes.

Key words: Mapping system; Identifier/locator separation; Mobility; Hierarchical mapping

1 引言

随着智能设备和多样化网络业务的流行，当前互联网已经成为人类社会必不可少的基础设施之一。然而，它却面临着诸如可扩展性、移动性、安全性、多家乡等诸多挑战[1−3]。导致这些问题的一个主要原因是IP地址的语义过载[4]，即IP地址不仅标识通信主机，也标识网络位置。为了应对这些挑战，研究人员提出了身份与位置分离架构[5−8]，被公认为未来网络中一种非常有潜力的技术。

尽管身份位置分离架构从不同角度解决了当前互联网面临的可扩展性问题，但同时也引入了映射解析系统存储身份与位置之间的映射表项，而且映射解析系统的性能在一定程度上决定了身份与位置分离架构的性能。针对这一问题，研究人员基于分布式哈希表(Distributed Hash Table, DHT)提出了许多映射解析系统[3−5, 9−11]。这些系统都有各自的优点，但大部分都假设标识结构是分级可汇聚的。然而，越来越多的研究成果建议使用扁平标识如HIP[7], AIP[12], MobilityFirst[13]等，这就使得针对分级标识设计的映射解析系统无法直接支持扁平标识。最近，研究人员也提出了一些针对扁平标识的映射解析系统，如DHT-MAP[4], SLIMS[14], DMAP[15]等。但是，DHT-MAP和SLIMS或者引起较高的映射解析时延导致时延敏感业务不可用，或者具有较高的管理控

收稿日期：2016-05-23；改回日期：2016-12-29；网络出版：2017-02-24
*通信作者：王鹏 15803846349@163.com
基金项目：国家 863 计划项目(2015AA016102)，国家自然科学基金创新研究群体科学基金(61521003)
Foundation Items: The National 863 Program of China (2015AA-016102), The National Natural Science Foundation of China (61521003)

制开销限制了系统的可扩展性。DMAP是一种基于网络共享主机的直接映射解析系统，该共享主机的假设在未来网络中可能是适用的，但无法与当前网络兼容。

文献[3]提出了一种映射解析系统LISP-DHT，该系统应用Chord[16]机制注册和解析身份和位置之间的映射关系。文献[4]提出应用DHT机制将身份与位置标识之间的映射表项分布到多个解析节点上，不同于文献[3]，文献[4]将映射解析服务器构建成一个内容寻址网络(Content-Addressable Network, CAN)[17]。文献[5]进一步提出了一种增强型的基于DHT的映射解析系统，该系统允许端点自由选择映射表项的存储位置。文献[18]提出了一种应用于信息中心网络场景中名字解析的多层DHT架构，但是，在这种架构中，内容名称不仅要存储在本地最底层节点中，也要存储到本地和最高层之间的所有解析域中，因此，最高层节点将会成为一个热点。尽管这些基于传统DHT机制的映射解析系统具有良好的可扩展性、自组织性和健壮性，但传统DHT机制本身具有非位置感知的缺陷，即在构建覆盖网络时没有考虑物理网络中节点之间的邻近关系，导致逻辑节点和物理节点之间不匹配，即逻辑最短路径不一定是物理上最短的路径。这种非位置感知的缺陷随着底层物理网络规模的不断扩大容易引起查询性能急剧恶化[19−22]，导致在查询映射表项时具有较高的映射解析时延。另外，这些映射系统对所有终端的活跃性或移动特性不加区分进行存储，拖长了总体的映射解析查询时延。文献[23]中提出了映射信息活跃性的概念，但并没有给出活跃性的确切定义，并且根据传统Chord算法构建映射解析系统，存在非位置感知导致的映射解析时延较高的问题。

为了满足未来网络对映射解析系统的需求，根据终端的活跃度，本文提出了一种分级结构的映射解析系统。为最小化映射解析时延，该系统将位置感知 DHT 的构建过程建模为马尔科夫决策过程(Markov Decision Process, MDP)，并提出了一种马尔科夫决策构建算法 MDC(Markov Decision Construction)用于求解该模型，MDC 算法基于传统的增强学习求解最优或近似最优的构建策略。该分级映射解析系统实现了可扩展性、低映射解析时延以及递增部署之间的良好平衡，能够显著降低映射解析时延和提升路由性能。

2　分级映射解析系统

2.1 基本思想

在描述分级映射解析系统的基本思想之前，首先对活跃度在本系统中的定义进行界定。活跃度(Active Degree, AD)是指映射信息表项在映射解析系统中的活跃程度，也即一段时间内映射信息表项被查询的次数。活跃度可以定义为

$$\mathrm{AD} = \frac{QN}{T} \tag{1}$$

其中，QN表示在时间段T内某映射信息表项的查询次数。活跃度定义了映射信息表项的活跃程度，当活跃度达到或低于某一阈值时，映射信息表项将在不同级别的映射服务器之间动态流动，实现映射解析系统的自我调整和重构。

人类活动的需求导致通信的产生，从最初的点到点通信发展到了通信网络。因此，网络本质上是人类活动的反映。在设计未来网络时有必要参考人类活动的一些规律，比如“二八定律”等，这些规律直接反映在现有网络的访问特性上。从网络中的二八定律现象可以看出[24]，20%的访问产生了80%的访问流量，即很小比例的通信对端的访问占据了大部分的访问量。根据此现象，本文设计映射解析系统的基本思想是根据通信对端的活跃程度，将通信对端的身份位置映射信息划分为不同的等级，并根据等级不同建立分级的映射解析存储架构，映射副本可根据自身活跃度的变化动态调整存储位置，实现映射解析系统的自我重构。

针对当前映射解析系统存在的物理拓扑和逻辑拓扑不一致的问题，在构建映射解析系统时采用位置感知DHT构建方法，将位置感知DHT的构建过程建模为马尔科夫决策过程，并对该过程进行求解，构建一种物理拓扑与逻辑拓扑相一致的DHT存储系统，实现映射信息的快速解析。

2.2 总体框架

分级映射解析系统总体框架如图 1 所示。从图中可以看出，映射解析系统采用分层存储的方式存储映射信息，不同活跃度的映射信息存储在不同层级，根据活跃程度的不同分为活跃级、中性级和稳定级。其中，活跃程度较高的映射信息存储在映射解析系统的活跃级；活跃程度一般的映射信息存储在映射解析系统的中性级；稳定级则存储全部的映射信息。根据活跃程度将映射信息分别存储在三级之中并根据活跃度的变化实现映射信息在三级之间的动态流动。按照分级分层思想构建的映射解析系统包含活跃级映射服务器、中性级映射服务器和稳定级映射服务器。活跃级映射服务器能保证快速响应大多数终端的映射解析查询请求，并负责为终端分配路由标识；中性级映射服务器能够减少由于终端大量移动带来的映射信息更新频繁的问题；而存储所有映射信息的稳定级映射服务器则能够保证映射解析系统的鲁棒性。关于映射解析系统中稳定级

的构建方法将在下一小节中进行详细描述。

随着移动终端的不断增多，当前网络所面临的移动性问题越来越突出。为了应对大量移动终端所带来的映射信息更新问题，本文对网络中的路由器按照地理区域进行分组，在同一地理区域内的若干个接入路由器以及核心路由器分为一组。在每个组内，根据组规模的大小，分配活跃级映射服务器和中性级映射服务器，中性级映射服务器负责存储本组内不太活跃的通信对端的标识映射信息以及新注册的映射信息。为便于描述，本文只考虑组内有一个中性级映射服务器的情形。

2.3 映射解析

基于活跃度的映射解析系统主要包含映射信息注册与更新、查询和动态流动 3 个基本机制。当一个主机连接到一个 ITR 路由器上以后，首先需要从 ITR 路由器获得一个位置标识并向映射解析系统注册主机身份标识和该位置标识的映射。映射解析系统使用注册操作完成映射表项的注册与更新。相应地，当一个 ITR 路由器接收到含有目的标识的数据包后，首先需要查询本地映射表是否缓存该目的标识的位置信息，如果没有，就需要使用映射查询操作从映射解析系统中获得该目的标识的位置信息。当映射信息表项的活跃度出现一定程度的变化时，该映射信息表项就会在 3 级之间进行动态流动，以使整个系统性能保持在较优水平。

2.3.1 映射信息注册与更新 当一个主机连接到一个 ITR 路由器(新加入网络或移动到新的接入网络)上以后，必须让其它的主机知道自己并使自己能够被访问。所以，就必须向映射解析系统注册或更新信息。注册与更新的主要步骤为：

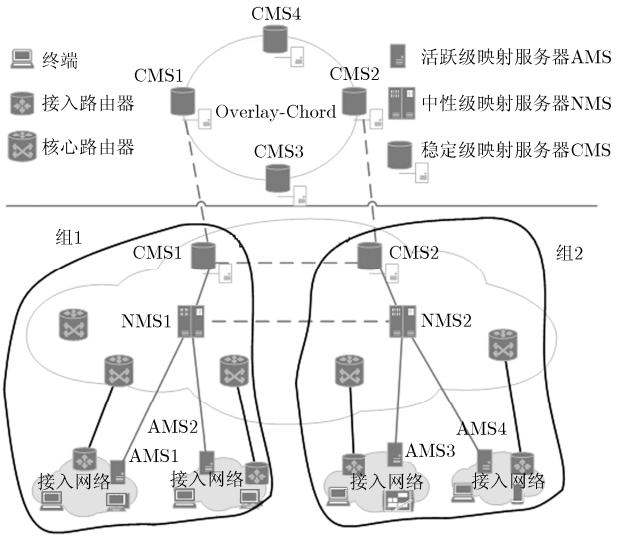


图 1 映射解析系统总体架构

(1)终端通过 ITR 路由器向映射解析系统发出注册请求；

(2)ITR 首先验证请求终端是否合法，如果合法，ITR 向对应的活跃级映射服务器 AMS 发送注册请求；

(3)AMS 首先查询自己是否存储相关表项，如果有，使用原映射表项并返回注册成功消息，否则首先为该终端分配映射对并存储，然后向中性级映射服务器 NMS 汇报；

(4)NMS 首先查询自身是否存储有关表项，如果没有，生成一个 NMS 表项并存储，向稳定级映射服务器 CMS 上报该表项；如果存在相关表项，首先判断原表项是否有效，有效则表明该终端为多宿终端，在此表项中新增一条映射信息；该表项无效则说明终端移动了位置，需用现有标识替换原标识。为了判断终端是频繁移动还是偶尔更换接入点，需要设定一个判定标准。假设 QN 表示映射查询的次数，T 是映射表项建立的时间，T_n 表示当前时间，如果 $QN/(T_n-T)<a$ (a 是预设的阈值)，也即单位时间内表项请求次数小于某一阈值，上报接入路由标识，如果大于阈值，为该终端分配中性路由标识并上报。基于此上报策略，可将映射更新数量降低到原先的 $1/N$ ，N 表示组内活跃映射服务器数量。

(5)稳定级映射服务器 CMS 收到上报的映射表项后，根据位置感知 DHT 算法首先查询是否已经存储，如果没有存储，则增加此表项。

2.3.2 映射信息查询 解析请求首先被发送到相应的 AMS 中，AMS 查询是否已经存储相关映射，如果有，则增加映射表项的查询次数 QN 项，返回查询结果；如果没有查中，AMS 向相应的 NMS 请求查询，NMS 如果有相关表项，增加该表项的查询次数 QN 项，启动映射表项的动态流动机制，返回查询结果；如果 NMS 中没有相关表项，NMS 需向 CMS 发送查询请求，如果 CMS 中存在该表项，返回查询结果，如果没有该表项，查询失败。需要特别指出的是，映射解析系统每过固定的时间，都会更新所有表项的查询次数，这样可以保证实现所有映射信息在 3 级之间的动态流动。

2.3.3 映射信息动态流动 映射信息在 3 级之间的动态流动主要有 4 种不同的情况：一是首次从稳定级映射服务器查询到的映射信息需要流动到中性级服务器；二是中性级映射信息活跃度达到预设阈值时要流动到活跃级；三是活跃级映射信息活跃程度下降到预设阈值时要流动到中性级；四是中性级映射信息活跃度下降到一定程度则直接在中性级映射服务器中删除该表项。具体的阈值设定可以根据映射解析系统的情况进行设置。

2.4 系统构建

2.4.1 模型构建　在原始的 Chord 中，物理节点组织成一个环，通过对节点地址进行哈希为每个节点分配一个节点标识。在本文设计的系统中，环路的构建要尽量考虑到物理网络中节点之间的邻近关系，保持物理拓扑和逻辑拓扑之间的一致性。在构建环路时从某一初始节点开始，沿着邻近节点的顺序进行构建，类似于一个顺序决策过程，而且只跟当前节点有关，与环路中已经选择的节点无关，也就是具有无记忆性，因此，本节把映射解析系统构建过程建模为一个马尔科夫决策过程(Markov Decision Process, MDP)[25]。

决策问题的求解就是要计算出一个策略使得用户设定的最优化标准的值最大化，称为最优策略。本节应用的马尔科夫决策过程最优化标准是有限界条件下的总期望奖赏。为了构建位置感知的 DHT，本节应用 MDP 模型刻画映射解析系统构建过程，通过求解马尔科夫决策问题获得一个最优的构建策略。

(1)决策点、状态、动作和转移概率：系统构建过程可以自然地划分为一些决策点，每个决策点对应构建过程中的一个节点选择。显然地，系统中映射解析服务器的数量是有限的，因此，构建过程属于一个有限界问题。在系统构建过程中，系统状态定义为所有可能选择的节点，用 i 表示节点标识，s_i 表示节点 i 当前的状态，m 表示节点数量。然后，状态空间可以表示为 $S=\{s_1,s_2,\cdots,s_m\}$, $1\le i\le m$。在每个决策点，基于当前状态根据转移概率做出决定，从动作空间中选择动作进行执行。为了避免路由环路，每个节点只能选择一次，即动作空间可以表述为剩余候选节点的集合，$A(s)=\{i_1,i_2,\cdots,i_k\}$, $1\le k\le m$。执行完动作以后，系统将从当前状态 s 转移动状态 s'。考虑到构建过程的特性，转移概率可以定义为

$$p(s'\mid s,a)=\frac{h_{\max}-h(i,j)}{h_{\max}-h_{\min}} \tag{2}$$

其中，$h_{\max}$ 和 $h_{\min}$ 分别表示从当前节点到候选节点的最大和最小跳数。$h(i,j)$ 表示从当前选择的节点 i 到所选择的节点 j 之间的跳数。

(2)奖赏函数：在奖赏函数的设计时，应该综合考虑当前节点和可选择的下一节点之间的物理距离和逻辑时延，从整体上构建一个最优或近似最优的映射解析系统。考虑到物理位置邻近的节点之间时延不一定最短的可能性，奖赏函数可以定义为当前节点和可选择下一节点之间的物理距离和逻辑时延的联合。以 $r_h(s,a)$ 表示系统从节点间物理距离获得的收益，$r_d(s,a)$ 表示系统从逻辑时延获得的收益，总的奖赏函数可以定义为

$$r(s,a)=\alpha r_h(s,a)+\beta r_d(s,a),\ \ \alpha+\beta=1 \tag{3}$$

其中，α，β >0，是权重因子。权重因子为系统提供了一个额外的特性，可以通过调整权重因子改变两个收益函数之间的比例，不同的权重因子值可以应用于不同的场景下。

在奖赏函数中，以跳数作为距离的测量标准，以 $h(i,i_m)$ 表示当前节点 i 和所有可选择的下一节点间距离最大值，距离收益函数可以定义为

$$r_h(s,a)=\begin{cases}\dfrac{1}{h(i,a)}, & \forall i_m\ \ h(a,i_m)<h(i,i_m)\\ 0, & \text{其他}\end{cases} \tag{4}$$

从式(4)中可以看出，所选择的节点离当前节点距离越近，系统获得的收益就越大。以 $r_d(s,a)$ 表示系统在某一状态 s 下通过选择动作 a 在时延方面获得的收益，时延收益函数可以定义为

$$r_d(s,a)=\min_{i_m\in A(s)}\{d_{i_m}\}\Big/d_a \tag{5}$$

其中，d_{i_m} 表示当前节点到所有可能选择的下一节点之间的时延值，d_a 表示当前节点与要选择的下一节点之间的时延。从式(5)可以看出，所选择的节点与当前节点之间时延越小，系统获得的收益就越大。

(3)最优化方程：决策问题的目标是计算一个使得某种最优化标准值最大化的策略，系统构建过程中应用的马尔科夫决策过程最优化标准是有限界的总期望奖赏值。在该系统构建过程中，总期望奖赏值是物理距离和逻辑时延的收益之和。给定一个策略 π 和一个初始状态 s，MDP 最优化问题可以表述为

$$V_{\pi^*}=\max\ E\left\{\sum_{k=0}^{N_{\pi^*}(s)} r\left(s_k,a(s_k)\right)\right\} \tag{6}$$

其中，π^* 表示最优策略，根据 MDP 的特点，最优化的值函数是独一无二的，并可以定义为式(7)的解。

$$V_{\pi^*}(s)=\max_{a\in A(s)}\left\{r(s,a)+\gamma\sum_{s'}p(s'\mid s,a)V^*(s')\right\},\quad \forall s\in S \tag{7}$$

式(7)表明某一状态的总奖赏值是期望的立即奖赏与采用最优动作所期望获的折扣奖赏之和。给定最优的值函数，最优策略可以表示为

$$\pi^*(s)=\arg\max_{a\in A(s)}\left\{r(s,a)+\gamma\sum_{s'}p\left(s'\mid s,a\right)V^*(s')\right\} \tag{8}$$

在该系统构建过程中，MDP 最优策略 π^* 表示

选择哪一个节点作为下一个节点的决策。根据最优策略能够构建一个在整体时延性能上最优或近似最优的映射解析系统。

2.4.2 构建算法 Q 学习[26]通常用于求解模型参数事先未知的 MDP 问题，Q 学习应用 Q 函数模拟累计奖赏，Q 函数代表了在状态 s 下执行动作 a 可能获得的最大奖赏。在学习的过程中，首先以某个初始状态 $Q(s,a)$ 开始，然后根据式(9)以不断试错的方式递归地更新 $Q(s,a)$ 。

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(r(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\right) \tag{9}$$

Q 函数的递归定义是 Q 学习的根本。考虑到构建过程的特点，在改进 Q 学习的基础上，提出了马尔科夫决策构建算法 MDC(Markov Decision Construction)。该算法包含以下 3 个步骤：

步骤 1 **初始化** MDC 算法首先初始化 $\boldsymbol{Q}$ 矩阵为零矩阵，以拓扑矩阵 $\boldsymbol{TM}$ 和奖赏函数矩阵 $\boldsymbol{r}(s,a)$ 作为输入，并初始化所有参数；

步骤 2 **迭代** 当算法满足终止条件时，迭代过程终止。如果不满足终止条件，迭代过程继续并从给定的初始节点 s_0 开始计算一个完整的包含所有节点的构建路径。在构建路径计算的过程中，系统从初始状态 s_0 开始，遵循贪婪策略执行一系列动作。在执行完每个动作以后，系统根据式(9)更新 $Q(s,a)$ 。另外，在构建路径的计算过程中必须去除环路的影响。

步骤 3 **构建** 在迭代终止以后，将从 $\boldsymbol{Q}$ 矩阵中发现最优构建策略。然后，基于最优构建策略，就可以构建具有最优时延性能的映射解析系统。

2.4.3 算法复杂性分析 由以上算法的流程可以看出，步骤 1 是需要初始化 $\boldsymbol{Q}$ 矩阵，需要 $O(N_{s,a})$时间，$N_{s,a}$表示状态值的数量。步骤 2 中需要计算给定的初始节点 s_0，开始计算一个完整的包含所有节点的构建路径，在最坏的情况下需要迭代所有可能的状态，因此可以在 $O(N_{s,a})$时间内解决。最后，整个计算过程需要循环执行 $O(M)$, M 表示 Q 学习算法探索场景的数量。因此，MDC 算法的复杂度都是为 $O(MN_{s,a})$。关于算法复杂度的详细分析参见文献[26]。

3 性能仿真及分析

3.1 实验参数

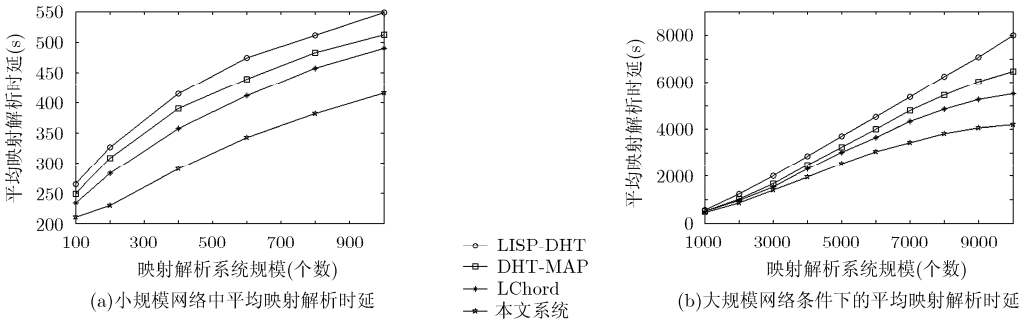
本节基于覆盖网仿真框架 OverSim[27]搭建仿真实验环境。OverSim 是一个运行于 OMNeT++ [28]之上的 P2P 覆盖网仿真模块。OMNeT++是一个离散事件驱动的仿真器，广泛应用于学术领域。为了模拟应用场景，网络拓扑按照典型的互联网模型 InetUnderlay 随机创建，创建的网络拓扑由 10 个骨干路由器和 100 个接入路由器组成。映射解析服务器的数量按照两种情况进行设置，一种是映射解析服务器数量小于 1000，在此情形下，映射解析服务器数量分别配置为 100, 200, 400, 600, 800 和 1000，主要用于评估映射解析系统在小规模网络下的性能；另外一种情况是映射解析服务器数量从 1000 递增到 10000，间隔为 1000，主要用于评估映射解析系统在大规模网络环境下的性能。每个接入路由器平均连接有 100 个终端，因此，网络中总共有近似 10000 个终端节点。在每次实验中，映射解析服务器的数量是恒定的，这是因为本文所提映射解析系统主要用于互联网基础骨干网中。在骨干网中，映射解析系统中节点的加入、离开或发生故障的概率很小，因此，在映射解析系统性能的评估过程中可以忽略节点扰动(churn)的影响。

在每次实验中，每个映射解析服务器平均注册有 1000 个映射解析表项，因此当映射解析服务器数量为 1000 时，整个映射解析系统整体上总共有近似 1000000 个映射解析表项。在实验测量过程中，每个终端节点上都运行一个测试程序周期性地发送映射解析请求，请求间隔为 1 s，请求模式遵循如文献[29]中提出的 web 页面浏览模型，映射解析表项的流行度遵循类似于 web 页面流行度的 Zipf 概率分布，均匀分布的 160 bits 长度的标识用于映射解析系统。活跃度阈值设为$(\mathrm{AD}_{\min} + \mathrm{AD}_{\max})/2$，其中 $\mathrm{AD}_{\min}$, $\mathrm{AD}_{\max}$ 分别表示本地活跃映射服务器中映射信息活跃度的最小值和最大值。为了反映一般情况，实验重复 10 次，取 10 次的平均值作为最终的结果。而且，每次仿真的时间设定为 1 h。

3.2 性能分析

3.2.1 平均映射解析时延 平均映射解析时延性能对映射解析系统十分重要，因为它定义了发送一个映射解析请求以后获得映射表项所需的时间。映射解析服务一个重要的设计目标就是保持映射解析时延处于可接受的水平，即使标识数量非常多。

在图 2 中，x 轴表示映射解析系统的规模大小，y 轴表示平均映射解析时延。图 2(a)给出了小规模网络中平均映射解析时延的仿真结果对比情况。从图中可以看出，本文所提分级映射解析系统相比于 LISP-DHT, DHT-MAP 和 LChord 具有较低的平均映射解析时延，分别平均降低了 18.53%, 13.72%和 8.65%。这证明了当网络规模较小时，本文所提分级映射解析系统在降低映射解析时延方面是有效的。图 2(b)给出了不同映射解析系统在大规模网络条件



(a)小规模网络中平均映射解析时延　(b)大规模网络条件下的平均映射解析时延

图 2 平均映射解析时延

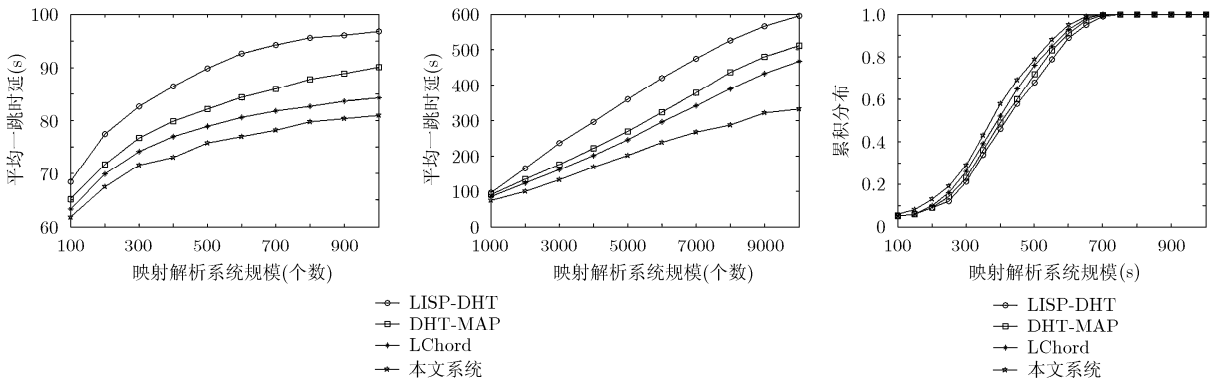
下的平均映射解析时延对比情况，从图中可以看出，本文所提分级映射解析系统相比 LISP-DHT, DHT-MAP 和 LChord 在映射解析时延方面平均分别降低了 32.43%, 23.04%和 13.78%。特别地，随着映射解析系统规模的不断增大，本文所提分级映射解析系统的优势更加明显。这表明当网络规模比较大时本文所提分级映射解析系统在降低映射解析时延方面更加有效。

3.2.2 平均一跳时延　本小节对映射解析系统的平均一跳时延性能进行评估。平均一跳时延是平均映射解析时延与平均路由跳数之比，它能够准确地反映物理网络结构和时延特性，在某种程度上，也能够反映物理拓扑和逻辑拓扑之间的匹配程度。

图 3 分别给出了几种映射解析系统在不同网络规模下的平均一跳时延性能的对比情况。图 3(a)给出的是小规模网络下的平均一跳时延性能对比情况，从图中可以看出，本文所提系统和 LChord 的平均一跳时延性能要优于 LISP-DHT 和 DHT-MAP。与 LISP-DHT 相比，本文所提系统降低了约 15.61%的平均一跳时延，这是因为本文所提系统和 LChord 都是基于位置感知 DHT 构建的映射解析系统，平均一跳映射解析时延明显低于基于非位置感知 DHT 构建的映射解析系统。图 3(b)给出了大规模网络条件下几种映射解析系统的平均一跳时延性能对比情况，从图中可以看出，随着映射解析系统规模的不断增大，本文所提系统与 LChord 之间的平均一跳时延性能差距越来越大，这表明在大规模网络中本文所提映射解析系统的优势更加明显。

3.2.3 时延累积分布　本小节评估了几种映射解析系统的时延累积分布情况。时延累积分布反映了映射解析系统的覆盖网络中数据包传递的时延分布情况。在本实验中，固定映射解析服务器的数量为 1000 个，实验结果如图 4 所示。

在图 4 中，x 轴表示平均映射解析时延，y 轴表示时延的累积分布函数。从图中可以看出，几乎 99% 查询请求的时延位于 100 ms 和 700 ms 之间。特别地，在相同时延下，本文所提映射解析系统具有较大的 CDF 值。这表明，与其他映射解析系统相比，本文所提映射解析系统中小于等于该时延的查询请求更多。此种现象的原因是本文所提映射解析系统的分层结构使得本地化的请求在本域内完成，域间请求通过高层的驿站完成，这不仅降低了每个请求的平均时延，而且也提高了路由的效率。总之，本文所提映射解析系统的查询性能更加有效，能够更快地完成映射解析。



(a)小规模网络下的平均一跳时延　(b)大规模网络下的平均一跳时延

图 3 平均一跳时延

图 4 时延累积分布

4　结束语

本文设计了一种基于活跃度的分级映射解析系统。该系统通信对端活跃程度的不同，将映射信息分为活跃级、中性级和稳定级3个等级，并据此设计了一种分层分级的映射信息存储架构，映射副本可根据自身活跃度的变化在3层之间动态调整存储位置。为最小化映射解析时延，在构建位置感知DHT系统时，将DHT系统的构建过程建模为一个有限界的马尔科夫决策过程，并提出了用于求解MDP的马尔科夫决策构建算法MDC。根据MDC算法求得的构建策略构建的位置感知DHT系统能够有效地避免路由无效性和冗余流量。该系统能够显著降低映射解析时延和提升路由性能，具有较小的映射解析代价，对网络结构的动态变化具有良好的适应性。由于本文设计的映射解析系统属于域内的移动解析映射机制，如何将其和域间的解析方法有效结合，设计全局性的映射解析方案，是下一步的重点研究工作。

参 考 文 献

[1] WANG Peng, LAN Julong, HU Yuxiang, *et al*. Towards locality-aware DHT for fast mapping service in future Internet[J]. *Computer Communications*, 2015, 66(C): 14–24. doi: 10.1016/j.comcom.2015.04.003.

[2] 兰巨龙, 熊钢, 胡宇翔, 等. 可重构基础网络体系研究与探索[J]. 电信科学, 2015, 31(4): 57–65. doi: 10.11959/j.issn.1000-0801.2015099.

LAN Julong, XIONG Gang, HU Yuxiang *et al*. Research on the architecture of reconfigurable fundamental information communication network[J]. *Telecommunications Science*, 2015, 31(4): 57–65. doi: 10.11959/j.issn.1000-0801.2015099.

[3] MATHY L and LANNONE L. LISP-DHT: Towards a DHT to map identifiers onto locators[C]. Proceedings of the ACM CoNEXT Conference, New York, 2008: 1–6. doi: 10.1145/1544012.1544073.

[4] LUO H, QIN Y, and ZHANG H K. A DHT-based identifier-to-locator mapping approach for a scalable internet [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2009, 20(12): 1790–1802. doi: 10.1109/TPDS.2009.30.

[5] LUO Hongbin, ZHANG Hongke, and MOSHE Zukerman. Decoupling the design of identifier-to-locator mapping services from identifiers[J]. *Computer Networks*, 2011, 55(4): 959–974. doi: 10.1016/j.comnet.2010.12.009.

[6] CONTI M, CHONG S, FDIDA S, *et al*. Research challenges towards the future internet[J]. *Computer Communications*, 2011, 34(18): 2115–2134. doi: 10.1016/j.comcom.2011.09.001.

[7] MUNGUR A and EDWARDS C. Performance of a tiered architecture to support end-host mobility in a locator identity split environment[C]. 2015 IEEE 40th Conference on Local Computer Networks (LCN 2015). IEEE Computer Society, Clearwater Beach, FL, USA, 2015: 446–449. doi: 10.1109/LCN.2015.7366349.

[8] RODRIGUEZ-NATAL A, PORTOLES-COMERAS M, ERMAGAN V, *et al*. LISP: A southbound SDN protocol?[J]. *IEEE Communications Magazine*, 2015, 53(7): 201–207. doi: 10.1109/MCOM.2015.7158286.

[9] LUO Hongbin, ZHANG Hongke, and QIAO Chunming. Optimal cache timeout for identifier- to-locator mappings with handovers[J]. *IEEE Transactions on Network and Service Management*, 2013, 10(2): 204–217. doi: 10.1109/TNSM.2012.122612.110221.

[10] LUO Hongbin, ZHANG Hongke, QIN Yajuan, *et al*. An approach for building scalable proxy mobile IPv6 domains [J]. *IEEE Transactions on Network and Service Management*, 2011, 8(3): 176–189. doi: 10.1109/TNSM.2011.071511.20100063.

[11] HOEFLING M, MENTH M, and HARTMANN M. A survey of mapping systems for locator/identifier split internet routing[J]. *IEEE Communications Surveys & Tutorials*, 2013, 15(4): 1842–1858. doi: 10.1109/SURV.2013.011413.00039.

[12] ANDERSEN D G, BALAKRISHNAN H, FEAMSTER N, *et al*. Accountable Internet Protocol (AIP)[C]. Proceedings of ACM SIGCOMM, Seattle, Washington, USA. 2008: 17–22. doi: 10.1145/1402958.1402997.

[13] BRONZINO F, RAYCHAUDHURI D, and SESKAR I, Experiences with testbed evaluation of the mobilityfirst future internet architecture[C]. Proceedings of European Conference on Networks and Communications 2015 (EUCNC 2015), Paris, France, 2015. doi: 10.1109/EuCNC.2015.7194127.

[14] HOU J, LIU Y, and GONG Z. Silms: A scalable and secure identifier-to-locator mapping service system design for future internet[C]. International Workshop on Computer Science and Engineering, Qingdao, China, 2009, 2: 54–58. doi: 10.1109/WCSE.2009.765.

[15] Vu T, Baid A, Zhang Y, *et al*. Dmap: A shared hosting scheme for dynamic identifier to locator mappings in the global internet[C]. 2012 IEEE 32nd International Conference on Distributed Computing Systems (ICDCS), Macau, China, 2012: 698–707. doi: 10.1109/WCSE.2009.765.

[16] STOICA I, MORRIS R, LIBEN-NOWELL D, *et al*. Chord: A scalable peer-to-peer lookup protocol for internet

applications[J]. *IEEE/ACM Transactions on Networking*, 2003, 11(1): 17–32, doi: 10.1109/TNET.2002.808407.

[17] RATNASAMY S, FRANCIS P, HANDLEY M, *et al*. A scalable content-addressable network[C]. Proceedings of ACM SIGCOMM'01, UC San Diego, USA, 2001: 161–172. doi: 10.1145/383059.383072.

[18] DANNEWITZ C, D'AMBROSIO M, and VERCELLONE V. Hierarchical DHT-based name resolution for information-centric networks[J]. *Computer Communications*, 2013, 36(7): 736–749. doi: 10.1016/j.comcom.2013.01.014.

[19] ZHOU S, GANGER G R, and STEENKISTE P A. Location-based Node IDs: Enabling Explicit Locality in DHTs[R]. Computer Science Department Carnegie Mellon University, 2003.

[20] ZHANG X Y, ZHANG Q, ZHANG Z, *et al*. A construction of locality-aware overlay network: Moverlay and Its Performance[J]. *IEEE Journal on Selected Areas in Communications*, 2004, 22(1): 18–28. doi: 10.1109/JSAC.2003.818780.

[21] ZHAO G, CUI R, and LIU Y. Lchord: Locality-aware chord for fast mapping in ID/locator split routing[J]. *Journal of Computational Information Systems*, 2013, 9(4): 1399–1406. doi: 10.1109/TNSM.2012.122612.110221.

[22] TAI Z, SHENG W, and DAN L. LISP-PCHORD: An enhanced pointer-based DHT to support LISP[J]. *China Communications*, 2013, 10(7): 134–147. doi: 10.1109/CIS.2007.62.

[23] 刘建强, 程东年, 邬江兴, 等. 一种扁平身份标志位置解析系统[J]. 计算机应用研究, 2010, 27(9): 3466–3469.
LIU Jianqiang, CHENG Dongnian, WU Jiangxing, *et al*. Locator resolving system for flat identity[J]. *Application Research of Computers*, 2010, 27(9): 3466–3469.

[24] 马卫东, 李幼平, 马建国, 等. 面向 Web 网页的区域用户行为实证研究[J]. 计算机学报, 2008. 31(6): 960–967.
MA Weidong, LI Youping, MA Jianguo, *et al*. Empirical study of region user behaviors for web[J]. *Chinese Journal of Computers*, 2008, 31(6): 960–967.

[25] PIROTTA M, RESTELLI M, and BASCETTA L. Policy gradient in Lipschitz Markov decision processes[J]. *Machine Learning*, 2015, 100(2-3): 255–283. doi: 10.1007/s10994-015-5484-1.

[26] WATKINGS J. B C. Learning from delayed rewards[D]. [Ph.D/Master dissertation], University of Cambridge, 1989.

[27] INGMAR B, BERNHARD H, and STEPHAN K. OverSim: A flexible overlay network simulation framework[C]. Proceedings of 10th IEEE Global Internet Symposium (GI '07) in Conjunction with IEEE INFOCOM, Alaska, USA 2007: 79–84. doi: 10.1109/GI.2007.4301435.

[28] ROCAMORA B and PEDRASA I. Evaluation of hierarchical DHTs to mitigate churn effects in mobile networks[J]. *Computer Communications*, 2016, 85: 41–57. doi: 10.1016/j.comcom.2016.02.003.

[29] JOHNSON T and SEELING P. Landing on the mobile web: From browsing to long term modeling[J]. *IEEE Communications Magazine*, 2016, 54(2): 146–151. doi: 10.1109/MCOM.2016.7402274.

伊　鹏：男，1977 年生，教授，研究方向为新型网络体系结构与网络安全.

王　鹏：男，1985 年生，助理研究员，研究方向为软件定义网络与网络安全.

申　涓：女，1977 年生，讲师，研究方向为内容中心网络与网络安全.

张校辉：男，1979 年生，讲师，研究方向为网络安全.

兰巨龙：男，1963 年生，教授，研究方向为新型网络体系结构、网络安全.