

LISP-HIMS: 一种基于 LISP 的层次化标识映射系统

徐 畅^{1,2} 任勇毛¹ 李 俊¹

(中国科学院计算机网络信息中心 北京 100190)¹ (中国科学院研究生院 北京 100049)²

摘 要 针对目前全球路由表规模高速增长的问题, IETF 提出可以通过分离主机的身份标识和路由标识地址空间来重新考量互联网寻址体系结构。思科公司提出的 LISP(Locator Identity Separation Protocol) 是基于该思想提出的主流方案之一。首先提出一种新的基于 LISP 的层次化标识映射系统, 以克服目前其他 LISP 映射系统的不足, 然后详细阐述该系统的模型、工作模式以及主机身份标识的分配方案; 最后与其他 LISP 映射系统进行了分析与比较, 结果表明, 该系统具有良好的可扩展性以及更低的映射信息查询时延。

关键词 位置与标识分离, LISP 协议, 映射系统, 身份标识, 路由标识

中图法分类号 TP393 文献标识码 A

LISP-HIMS: A Hierarchical Identifier Mapping System for LISP

XU Chang^{1,2} REN Yong-mao¹ LI Jun¹

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)¹

(Graduate University, Chinese Academy of Sciences, Beijing 100049, China)²

Abstract Aiming at the problem of currently rapid growth of global routing table size in DFZ, IETF has proposed to reconsider the Internet addressing architecture by splitting host identifier and its routing locator. LISP is considered to be one of the most promising solutions based on such idea, which was proposed by Cisco. This paper presented a novel mapping system for LISP proposal, elaborated the model and modes of this mapping system and also the allocation strategy of the hierarchical host identifier. Compared with other LISP-based mapping systems, the proposed mapping system has better scalability and lower mapping querying latency.

Keywords ID/Locator split, LISP, Mapping system, Identifier, Routing locator

1 引言

随着全球路由表的高速膨胀, 互联网路由系统的可扩展性受到了严峻的挑战。全球路由表的规模已经连续几年以超线性速度增长^[1]。国际 IAB 组织将该问题列为当今互联网面临的最亟需解决的问题之一^[2]。产生该问题的原因, 一是源于互联网接入的快速普及, 然而更重要的是随着 Multi-homing、流量工程和网络移动等技术的广泛使用, 大量不可聚合 IP 地址前缀被通告至全球路由表, 导致路由表项急剧增多, 路由更新频率增大, 路由收敛变慢, 从而严重影响了路由的可扩展性。目前, 针对互联网路由可扩展性的实现主要依赖于大量可聚合的 IP 地址前缀, 当用户网络广泛使用诸如 Multi-homing、流量工程和移动性等技术来满足自身业务需求时, 便向核心路由表中注入了大量非可聚合的 IP 地址前缀。有研究表明, 互联网全球路由表中, 仅有 11% 的地址前缀来源于核心网络本身的路由通告, 其余的都源自于大量边缘网络^[3]。

全球路由表可扩展性问题的根源在于当前互联网路由体

系结构将所有 AS(Autonomous System) 同等对待, 没有在商业模式、业务需求和增长趋势等方面合理区分用户网络和 ISP 网络, 致使域间路由呈现扁平结构^[3]。为了解决该问题, 学术界普遍认为必须重新审视和设计互联网路由体系结构。2007 年 IETF 提出可以通过分离主机或网络的身份标识和路由标识的地址空间来解决上述问题^[2]。随后基于该思想, 人们提出了各种解决方案 (HIP^[4], ILNP^[5], GLI-split^[6], LISP^[7], Shim6^[8], APT^[9]), 其中 LISP 方案以其易于实现和部署的优势, 加以全球最大的网络设备提供商思科公司的支持, 被认为是众多方案中最具发展前途的一个。

2 LISP 及其映射系统

LISP 是一种基于位置和标识信息分离的系统化解决方案。其具体实现是将边缘网络地址空间与核心网络地址空间进行分离, 进而引入两类地址: EID(End identifier) 和 RLOC(Routable Locator)。EID 用于全局唯一标识站点内的主机, 属于 PI 地址, 不参与全局路由, 不会被通告至全球路由表, 主机之间基于 EID 建立通信会话。RLOC 用于标识站点的接

到稿日期: 2011-12-03 返修日期: 2012-03-11 本文受国家发改委项目(GC-HG100452), 中国科学院院长优秀奖科研启动基金项目资助。

徐 畅(1987-), 男, 硕士生, 主要研究方向为未来互联网体系结构, E-mail: nuaaxc@gmail.com; 任勇毛(1981-), 男, 博士, 助理研究员, 主要研究方向为网络协议及体系结构; 李 俊(1968-), 男, 博士, 研究员, 博士生导师, 主要研究方向为下一代互联网。

入点和骨干网络节点,属于 PA 地址,其严格按照网络拓扑进行分配,具有层次结构,便于聚合,进而减少了核心路由表的大小。为了保证主机之间能够进行通信,报文在进入 ISP 网时必须完成从 EID 到 RLOC 的映射和转发,LISP 采用“映射+封装”的方式完成报文的转发,在发送端由 ITR(入口隧道路由器)为报文添加含有 RLOC 的外部报头,内部报头包含由主机创建的源和目的 EID 地址,在接收端由 ETR(出口隧道路由器)完成报文的解封封装和最终的投递。目前 LISP 引入 6 种映射机制(NERD^[10],ALT^[11],CONS^[12],EMACS^[13],DHT^[14],TREE^[15])来完成地址的转换。NERD 利用网络数据库存放原始映射信息,ITR 通过 HTTP 协议进行访问和获取,每个 ITR 拥有整个映射数据库的副本。然而随着映射数据库规模的扩大,映射信息的查询和更新时延逐渐上升,因此存在严重的可扩展性问题。ALT 通过构建基于 BGP 和 GRE^[16]的位于互联网之上的层叠网络来完成映射服务。映射数据库分布式存储在 ETR 中,层叠网仅负责映射信息查询和响应消息的转发。层叠网中 ALT 路由器之间通过 GRE 隧道转发报文,利用 BGP 承载 EID 前缀路由信息,完成 EID 前缀路由信息的通告和聚合。虽然通过层叠网络的复杂性换来了映射系统的可扩展性,但引入的间接层无疑增大了映射信息解析的处理时延。EMACS 采用一系列默认转发路由器负责报文的转发,这些默认转发路由拥有更多的映射信息,可完成 EID 前缀的聚合。这样虽然降低了每个默认转发路由器所负责的映射数据库规模,却增大了整个转发路径的长度。CONS 与 ALT 类似,同时结合了 NERD 和 EMACS 的若干机制,采用了“Push”和“Pull”相结合的映射方式。LISP-DHT 是基于 DHT 的分布式映射系统,其假定 EID 为扁平标识,映射信息存放在 LISP-DHT 节点上,使用 DHT 发布映射信息,ITR 和 ETR 通过查询获取映射信息。该系统充分利用了 DHT 结构所具有的良好可扩展性和可靠性的优势,但随着网络规模的增大,映射信息查询路径的长度无法确定,导致查询时延成为阻碍系统扩展的瓶颈。LISP-TREE 是一种基于 DNS 的映射服务,提供了两种基于 DNS 的工作模式进行映射信息查询,即递归查询和迭代查询。其中,迭代查询模式增加了查询过程中实体的交互次数,而递归查询模式无法利用 DNS 的缓存机制,因此为了保证端到端的高速通信,整个系统的查询时延仍有待改进。

3 LISP-HIMS

本文提出了一种新的基于 LISP 的映射系统 HIMS(Hierarchical Identifier Mapping System),从而克服了上述 LISP 映射系统存在的不足。其核心思想是构建一个层次化结构的映射系统并对 EID 标识进行层次划分,使之成为具有层次特性的 HID 标识(Hierarchical ID),以此实现 EID 的聚合,提高了主机标识地址的可扩展性。HIMS 的映射数据库分布存储在每个站点的 ETR 中。当需要查询目标主机的“HID-RLOC”映射信息时,ITR 首先向 HIMS 发送目的地址为目标主机 HID 的 Map-Request 映射请求消息,随后 HIMS 直接利用 Map-Request 中的源和目的 HID 进行转发,最终送至目的

主机所在 LISP 站点的 ETR 获取相应的映射信息。此外,由于 HIMS 具有层次化特性,在转发映射查询请求的过程中不再需要路由信息,只需比较 Map-Request 中源和目的 HID 以及转发节点的可见层次地址字段就能决定转发路径,因而从根本上解决了映射系统的路由可扩展性问题。同时由于不再需要查表,减小了系统的查询时延和系统中每个转发节点的存储负载,提高了系统的查询效率。

3.1 HID 标识结构

HID 用来标识 LISP 站点内的主机,是层次化的 EID 标识。HID 属于 PI 地址,不可路由,不会被通告至全球路由表。为了在 HIMS 中直接利用 HID 转发映射查询请求,在 HIMS 中,HID 必须是可寻址的。实际中可直接采用 IPv6 地址作为 HID 来标识站点主机,以兼容现有主机的网络协议栈。HID 的地址结构如图 1 所示。

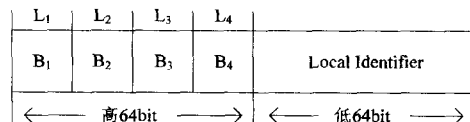


图 1 HID 标识结构

这里借助 IPv6 的可聚类全局单播地址格式来进行阐述。在 HID 的高 64bit 中, B_i 表示 HIMS 中第 i 层节点的可见层次地址块, L_i 表示 B_i 的长度。将从 B_1 到 B_{i-1} ($i \geq 2$) 构成的比特序列作为 HIMS 中第 i 层节点的唯一标识。对于不同的 i 和 j , L_i 和 L_j 可以相同也可以不同。对于所有的 L_i ($1 \leq i \leq N$, N 为 HIMS 的层数),有 $\sum_{i=1}^N L_i = 64$ 。低 64bit 的 Local Identifier 字段最终分配给站点内的主机,作为主机在站点内部的唯一标识。Local Identifier 字段连同高 64bit 的层次地址块一起来全局唯一标识一个 LISP 主机。

3.2 LISP-HIMS 模型

LISP-HIMS 是一种层次化的映射系统,其体系结构如图 2 所示。

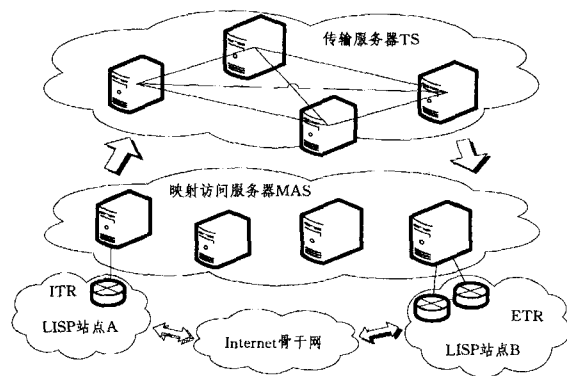


图 2 LISP-HIMS 模型

LISP-HIMS 从功能上主要划分为两层:传输层和终端层。传输层又可划分为多层,其基本组成部件是传输服务器 TS(Transmitting Server),它主要负责 LISP 映射查询请求 Map-Request 的转发。终端层的基本组成部件是映射访问服务器 MAS(Mapping Access Server),其位于 HIMS 边界,与其管辖范围内 LISP 站点的 XTR(ITR 或 ETR)相连,主要负

责两方面的工作,即映射查询和映射登记。映射查询是指 MAS 负责代理 ITR 查询所需的“HID 前缀-RLOC”映射,并负责随后的查询结果返回,这样带来的好处是实现了整个映射系统对 ITR 的完全透明,降低了 ITR 的复杂性;映射登记是指 MAS 负责接收 ETR 发送的映射注册消息 Map-Register^[17],完成对 ETR 及相应的 HID 前缀的登记,从而获得一张其管辖范围内站点的所有 ETR 列表,称为 ETR 注册表。表中每条记录包括两个字段,即 ETR 在本地的链路端口号 and 该 ETR 所拥有的 HID 前缀。在转发 Map-Request 的过程中,MAS 查询 ETR 注册表,匹配 Map-Request 中的目标 HID 和表中的 HID 前缀,找到目标 ETR 进行投递,最终由目标 ETR 负责发送 Map-Reply 进行回复。为区分 MAS 的上述两种职责,称负责映射查询的 MAS 为源 MAS,负责映射登记的 MAS 为目的 MAS。此外,为方便概念阐述,将 TS 和 MAS 统称为层次标识映射服务器 HIMS(Hierarchical Identifier Mapping Server)。

LISP-HIMS 的映射数据库采用分布式存储方式,即所有的“HID 前缀-RLOC”映射信息被分布式地存储在每个 LISP 站点的 ETR 中。每个 ETR 向其站点隶属的 MAS 注册自己的唯一标识(例如 RLOC)。为了解析并转发源和目的地址字段为 HID 格式的 Map-Request 报文,需要为每个 HIMS-Server 配置两个参数 S (start) 和 L (length)(记作 S/L),用来指示待处理的 HID 中该 HIMS-Server 可见的层次地址块。其中, S 代表可见层次地址块在整个 HID 地址中的起始位置(从 0 开始), L 表示层次地址块的长度。在转发 Map-Request 的过程中,位于转发路径上的每个 TS 将根据 Map-Request 中包含的源和目的 HID 地址以及自身的可见层次地址块参数 S/L ,正确地计算出下一跳 HIMS-Server,最终将 Map-Request 投递到目的 MAS。

假设一个 LISP-HIMS 映射系统具有 N 层($N \geq 1$)结构,位于顶层(第 1 层)的 HIMS-Server 称为根 HIMS-Server,位于底层(第 N 层)的 HIMS-Server 称为叶子 HIMS-Server(即 MAS)。存在一个位于第 i 层($1 \leq i < N$)的 HIMS-Server 和一个位于第 $(i+1)$ 层的 HIMS-Server,分别记作 H_i 和 H_{i+1} ,HIMS 标识分别为 P_i 和 P_{i+1} ,层次地址块参数分别为 S_i/L_i 和 S_{i+1}/L_{i+1} 。称 H_i 是 H_{i+1} 的父节点及 H_{i+1} 是 H_i 的子节点,当且仅当 1) $P_i < P_{i+1}$; 2) $S_{i+1} = S_i + L_i$ 。其中 1) 中小于号的含义为 P_i 是 P_{i+1} 的前缀,且 P_i 的长度小于 P_{i+1} 的长度。LISP-HIMS 即通过上述关系形成层次结构。每一个 HIMS-Server 都维护两张关系表用于转发 Map-Request,即父节点关系表和子节点关系表。父节点关系表记录到父节点的链路端口号(根 HIMS-Server 的为空),子节点关系表包含子节点在本节点中的链路号到连接子节点的链路端口号的映射(叶子 HIMS-Server 的为空)。链路号是在子节点 H_{i+1} 的标识 P_{i+1} 中,从第 S_i 位开始的长度为 $L_{i+1} - L_i$ 的字段。显然,一个 HIMS-Server 最多有 $2^{(L_{i+1} - L_i)}$ 个子节点。注意到,父子关系表中涉及的链路号和链路端口号仅用于本地表示,具有局部性。父节点关系表可在其父节点为其分配标识时生成,子节点关系表可在为其子节点分配标识时生成。

3.3 LISP-HIMS 工作模式

LISP-HIMS 的工作模式主要分为两个过程,即映射注册和映射获取。

3.3.1 映射注册过程

映射注册过程是指 LISP 站点的 ETR 向 LISP-HIMS 中的 MAS 注册本站点的 HID 前缀的过程,如图 3 所示。具体步骤如下:

- (1) LISP 站点连接 ISP, 获取 RLOC 地址;
- (2) LISP 站点从 HIMS 下载 MAS 信息列表, 表中包含所有 MAS 的标识信息以及各自负责的 HID 地址块, 通过最长前缀匹配搜索 MAS 信息列表, 确定本站点的 HID 前缀所属的 MAS;
- (3) LISP 站点的 ETR 向 MAS 发送 Map-Register 消息进行注册, 其中包括本站点的 HID 前缀信息;
- (4) MAS 接收到 Map-Register 消息, 若不存在该 ETR 的注册信息, 则保存; 否则, 进行更新。

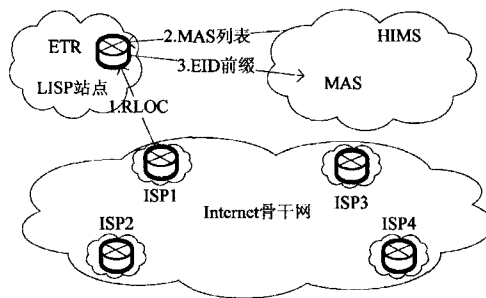


图 3 LISP-HIMS 的映射注册过程

3.3.2 映射获取过程

映射获取过程包括对映射信息的请求和响应,如图 4 所示。映射请求指的是 Map-Request 消息正确而快速地从 ITR 到达目的 ETR 的过程,进而请求目的 ETR 返回匹配的“HID 前缀-RLOC”映射信息。这一过程的主要步骤如下。

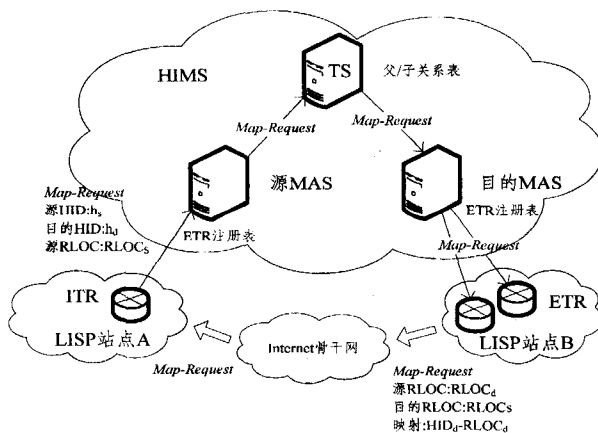


图 4 LISP-HIMS 的映射获取过程

- (1) 假设通信双方的 HID 分别为 h_s 和 h_d , RLOC 分别为 r_s 和 r_d 。ITR 首先产生源地址为 h_s 、目的地址为 h_d 的 Map-Request, 且附加 r_s , 并将 Map-Request 发送至与之相连的源 MAS。
- (2) 源 MAS 运用最长前缀匹配原则在 ETR 注册表中查找 h_d , 若匹配, 则将 Map-Request 发往匹配表项中所指定的

ETR;否则,将 Map-Request 发往 HIMS 传输层节点 T ,假设 T 的可见层次地址块为 S/L 。

(3)中间节点 T 在收到 Map-Request 后,通过如下方法将 Map-Request 发往下一跳 HIMS Server;

(3.1)从子节点接收到 Map-Request,则:

(3.1.1)若 h_s 和 h_d 中的前 S 位相同,则下一跳 HIMS Server 在 T 中的链路号为 h_d 中与 T 的可见层次地址块对应的字段的值;

(3.1.2)若 h_s 和 h_d 中前 S 位不同,则下一跳 HIMS Server 为 T 的父节点,链路号为 FFFF;

(3.2)从父节点接收到 Map-Request,则同(3.1.1)。

(4)根据上一步获得的下一跳 HIMS Server 链路号,查询本地父/子关系表,进而确定下一跳 HIMS Server 的链路端口号,随即将 Map-Request 发往下一跳 HIMS Server。

按照上述过程,经过一系列的转发,Map-Request 将被投递到目的 MAS。目的 MAS 再根据 Map-Request 中的 h_d 查询本地 ETR 注册表,进而获得目的 ETR 的 RLOC,最终将 Map-Request 发至目的 ETR。映射请求过程结束。

在映射响应过程中,目的 ETR 回复 Map-Reply 消息进行响应。具体过程是将请求的“HID 前缀—RLOC”映射信息填入 Map-Reply 的 EID-prefix 和 Locator 字段^[7],将 Map-Request 中的 r_s 填入 Map-Reply 的目的地址字段,最后通过 Internet 核心网络(而不是 LISP-HIMS)发送至源 ITR,在收到 Map-Reply 后,源 ITR 提取其中的“HID 前缀—RLOC”映射信息,映射响应过程结束。

3.4 HID 地址的分配

HID 地址的分配方案借鉴目前由正规地址分配机构分配 IP 地址^[18]的经验¹⁾,根据集中式层次授权机制逐层分配 HID 地址。从概念层面来说,需要两类组织的参与:一类是类似于 IANA 的国际权威地址分配机构 GAHA(Globe Assigned HID Authority),另一类是类似于 RIR 的区域地址分配机构 LHR(Local HID Registry)。GAHA 负责为各个地区的 LHR 分配 HID 地址前缀,再由 LHR 向各自管理的 TS 和 MAS 分配下级 HID 前缀,并发布相应的 HID 前缀分配信息,包含所有 MAS 的 RLOC 以及相应的 HID 前缀。LISP 站点可以根据上述 HID 前缀分配信息自行选择合适的 HID 前缀,并向该 HID 前缀所属的 MAS 进行注册,具体的注册步骤见 3.3.1 节。

4 与其他 LISP 映射系统的比较

LISP 架构的核心是映射系统,而映射系统最重要的特性是可扩展性,即当系统规模扩张时,其仍然保持良好的性能。具体到 LISP 环境下,映射系统的可扩展性主要包括两方面内容:映射系统节点负载和映射信息获取时延。

4.1 映射系统节点负载

这里主要讨论的是映射系统中节点的映射信息存储负载和映射查询请求处理负载。映射信息存储负载体现为分配一

定规模 HID 前缀时映射系统需要存储的映射信息的总量。考虑到 LISP-HIMS 中每个 TS 存储了父/子关系表,每个 MAS 存储了 ETR 注册表。简单起见,假设 HID 采用图 1 指示的结构,LISP-HIMS 映射系统采用两层结构,第一层为 TS 节点,形成 Mesh 结构,第二层为 MAS 节点,由 MAS 节点直接连接 LISP 站点中的 ETR,后者形成 ETR 层(该层不属于 LISP-HIMS),此外,每条映射信息中一个 HID 前缀仅包含一个 RLOC 地址,则 LISP-HIMS 中存储的映射信息总量(字节)为:

$$DB_{size} = E_{TS} \times (2^{L_1} + 2^{L_1+L_2} - 2^{L_1}) + E_{MAS} \times 2^{L_1+L_2+L_3} \quad (1)$$

其中, DB_{size} 表示 HIMS 存储的映射信息总大小, L_1, L_2 和 L_3 分别表示 TS 层、MAS 层以及 ETR 层的可见层次地址块长度。 E_{TS} 表示 TS 中父/子关系项的单位存储大小, E_{MAS} 表示 MAS 中 ETR 注册表项的单位存储大小。此外,假设 HID 前缀规模为 10^n , n 为正整数。则有:

$$2^{L_1+L_2+L_3+L_4} = 10^n \quad (2)$$

简单起见,假设 $L_1 = L_2 = L_3 = L_4 = L$,结合式(1)和式(2)可得:

$$DB_{size} = E_{TS} \times (2 \times 10^{\frac{n}{L}} - 10^{\frac{n}{L}}) + E_{MAS} \times 10^{\frac{3n}{L}} \quad (3)$$

由于 $\sum_{i=1}^4 L_i = 64 \text{bit}$,可以令 E_{TS} 为 8 字节,同时按照文献[10]中提出的度量方式, E_{MAS} 为 30 字节。将 LISP-HIMS 系统中存储的映射信息总量随 HID 前缀规模的变化而变化的趋势与文献[10]中的结果进行比较,结果如图 5 所示。

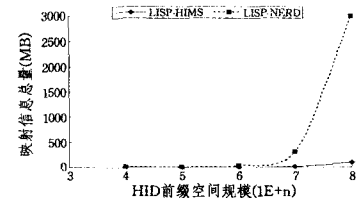


图 5 LISP-HIMS 与 LISP-NERD 的可扩展性比较

可以看到,随着网络中 HID 前缀规模的增大,LISP-NERD 需要存储的映射信息总量呈现指数级增长,而 LISP-HIMS 需要存储的映射信息总量增长幅度明显缓慢。这是因为在 LISP-HIMS 中,每个 HIMS Server 仅需配置自身的可见层次地址块,根据可见层次地址块和 Map-Request 中的源和目的 HID 前缀即可计算下一跳端口。因此,每个 HIMS Server 节点仅需存储规模为 $O(N)$ 的转发表,其中 N 为该节点的子节点个数。

节点的映射请求处理负载主要体现为映射系统的每个节点需要处理 Map-Request 的数量。假设来自 LISP 站点的 Map-Request 的数量为 R ,一个具有 L 层的 LISP-HIMS 系统中每一层节点转发 Map-Request 至其域内的概率为 p ,则第 i ($i \geq 2$) 层节点需要处理的 Map-Request 数目为 $R \times (1-p)^{i-1}$,这说明随着层级的增加,节点需要处理的 Map-Request 数目呈负指数增长,而在 LISP-ALT 中,每个 Map-Request 的处理都需要经过顶层节点,因此 LISP-HIMS 的上层节点相

¹⁾ 实际中,考虑到兼容性和成本最小化,HID 地址分配任务可考虑由当前的 IP 地址分配机构负责。

对 LISP-ALT 来说负载小得多,同时降低了单点失效发生的概率。

4.2 映射信息获取时延

LISP-DHT、LISP-ALT 和 LISP-CONS 都是通过通过在核心网上构建层叠网络完成映射服务。这样虽然提高了系统的可扩展性,但层叠网络的引入无疑增加了系统的复杂性和映射解析时延。LISP-TREE 沿袭了 DNS 系统的递归查询和迭代查询模式,没有引入层叠网络,与 LISP-HIMS 类似。下面比较 LISP-TREE 和 LISP-HIMS 的映射信息获取时延。假设 TR(包括 ITR 和 ETR)与映射系统之间的传输时延为 RTT_1 , ITR 与 ETR 之间的传输时延为 RTT_2 , LISP-TREE 映射系统内部的处理时延为 $Proc_{TREE}$, LISP-HIMS 映射系统内部的处理时延为 $Proc_{HIMS}$,根据 LISP-TREE 的映射查询模式^[15]可以得到 LISP-TREE 中映射信息查询时延 d_{TREE} 为:

$$d_{TREE} = 2(RTT_1 + RTT_2) + Proc_{TREE} \quad (4)$$

而根据 3.3 节阐述的 LISP-HIMS 映射查询过程,可以得到 LISP-HIMS 中映射信息查询时延 d_{HIMS} 为:

$$d_{HIMS} = RTT_1 + Proc_{HIMS} + RTT_2 \quad (5)$$

由于 LISP-TREE 和 LISP-HIMS 内部均为层次结构,查询路径相似,查询路径跳数约为 $O(\log N)$, N 为系统叶子节点的数目,因此有 $Proc_{TREE} \approx Proc_{HIMS}$,进而根据式(4)和式(5)有:

$$d_{TREE} = 2d_{HIMS} - Proc_{HIMS} \quad (6)$$

可见,相对于 LISP-TREE 来说,LISP-HIMS 具有更短的映射查询时延。

结束语 本文通过对基于位置与标识分离的命名和寻址体系结构的研究,提出了一种新的基于 LISP 的映射系统。其核心思想是构建一个层次化结构的映射系统并对 EID 标识进行层次划分,一方面实现 EID 的聚合,更主要是通过运用层次式的寻址策略,简化映射查询请求的转发过程,有效减小了映射查询时延,提高了系统的可扩展性。此外,本文对该映射系统的模型、工作模式以及 HID 地址的分配策略等方面进行了详细的阐述,并与其他基于 LISP 的映射系统进行了分析和比较。另一方面,如果 LISP-HIMS 在拓扑上完全严格遵循层次结构,则实际中网络很容易受到单点失效的影响,且由于是单路径传输,一旦网络中出现节点失效,就会导致以该节点为根的子树涵盖的所有 HID 前缀解析失败。因此,为了将实际中因网络失效带来的影响降至最低,需要考虑针对网络容错和自愈能力的设计,这也是下一步研究的重点。目前对于 LISP-HIMS 来说,该问题可以通过将 Fat-tree^[19]结构作为 HIMS 传输层的拓扑结构加以解决。在一个交换端口为 k 的 Fat-tree 结构中,不同域的任意两结点之间有 $(k/2)^2$ 条长度相等的路径,当出现节点失效时,可通过特定的容错路由机制绕过该失效节点,保证映射查询请求顺利到达目的 ETR。

参考文献

[1] Huston G. The growth of the BGP table: 1994 to present[EB/

OL]. <http://bgp.potaroo.net>, 2011-06-15

- [2] Meyer D, Zhang Li-xia, Fall K. Report from the IAB Workshop on Routing and Addressing[S]. RFC4984, 2007
- [3] Massey D, Wang L, Zhang B, et al. A scalable routing system design for future Internet[C]// Proc of ACM SIGCOMM Workshop on Large-Scale Attack Defense. New York: ACM Press, 2007: 790-796
- [4] Moskowitz R, Nikander R. Host identity protocol (HIP) architecture[S]. RFC5201, 2008
- [5] Atkinson R, Bhatti S, Hailes S. ILNP: Mobility, Multi-Homing, Localized Addressing and Security through Naming[J]. Telecommunication System, 2009, 42: 273-291
- [6] Menth M, Hartmann M, Klein D. Global Locator, Local Locator, and Identifier Split[R]. Würzburg: Institute of Computer Science of University of Würzburg, 2010
- [7] Farinacci D, Fuller V, Meyer D, et al. Locator/ID separation protocol (LISP) [R/OL]. <http://www.ietf.org/id/draft-ietf-lisp-16.txt>, 2011-10-31
- [8] Nordmark E, Bagnulo M, Shim6; Level 3 multihoming shim protocol for IPv6[S]. RFC5533, 2009
- [9] Jen D, Meisel M, Massey D, et al. APT: a practical tunneling architecture for routing scalability[R]. Los Angeles: University of California Los Angeles, 2008
- [10] Lear E. NERD: a not-so-novel EID to RLOC database[R/OL]. <http://www.ietf.org/id/draft-lear-lisp-nerd-08.txt>, 2011-03-06
- [11] Fuller V, Farinacci D, Meyer D, et al. LISP alternative topology (LISP + ALT) [R/OL]. <http://www.ietf.org/id/draft-ietf-lisp-alt-09.txt>, 2011-09-20
- [12] Brim S, Chiappan N, Farinacci D, et al. LISP-CONS: a content distribution overlay network service for LISP[R/OL]. <http://tools.ietf.org/id/draft-meyer-lisp-cons-04.txt>, 2008-04-09
- [13] Brim S, Farinacci D, Meyer D. EID mappings multicast across cooperating systems for LISP [R/OL]. <http://tools.ietf.org/id/draft-curran-lisp-emacs-00.txt>
- [14] Mathy L, Iannone L. LISP-DHT: Towards a DHT to map identifiers onto locator[C]// Proc of the Re-Architecting the Internet Conf (ReArch 2008). Madrid: ACM Press, 2008: 7-12
- [15] Jakab L, Cabellos-Aparicio A, Coras F, et al. LISP-TREE: a DNS hierarchy to support the LISP mapping system[J]. IEEE Journal on Selected Areas in Communications, Special Issue on Internet Routing Scalability, 2010, 28(8): 1332-1343
- [16] Farinacci D, Li T, Hanks S, et al. Generic routing encapsulation (GRE)[S]. RFC2784, 2000
- [17] Fuller V, Farinacci D. LISP Map Server Interface[R/OL]. <http://www.ietf.org/id/draft-ietf-lisp-ms-12.txt>, 2011-10-25
- [18] Hubbard K, Koster M, Conrad D, et al. Internet Registry IP Allocation Guidelines[S]. RFC2050, 1996
- [19] Al-Fares M, Loukissas A, Vahdat A. A scalable, commodity data center network architecture[C]// Proc of ACM SIGCOMM. Aug. 2008: 63-74