

# 基于树环 Chord 的大规模覆盖网的拓扑结构

徐 玉,程春玲,周 芸

(南京邮电大学 计算机学院,江苏 南京 210003)

**摘 要:**提出一种基于树环 Chord 的网络拓扑结构,并设计了与该拓扑结构对应的节点的路由表结构。基于树环 Chord 的网络拓扑结构有效利用了 IPv6 地址协议的地址聚类特性把网络中的节点划分到相应的自治域,有效解决了物理网络和逻辑网络不匹配时路由绕路产生的搜索延迟问题;新的路由表消除了冗余信息并增加了目标资源列表,增加了启发信息,缩短了重复搜索的延迟。仿真实验结果表明,基于树环 Chord 的网络拓扑结构的搜索平均延迟和平均跳数优于 Chord 和 DChord,有效提高了资源搜索的性能。

**关键词:**网络拓扑;IPv6;树环 Chord;路由表

**中图分类号:**TP393

**文献标识码:**A

**文章编号:**1673-629X(2011)08-0097-04

## Topology of Large-Scale Overlay Network Based on Tree Ring Chord

XU Yu, CHENG Chun-ling, ZHOU Yun

(Computer School, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** A new network topology based on tree ring Chord has been proposed, and the corresponding routing table structure has also been designed. The network topology based tree ring Chord uses the address cluster properties of IPv6 to divide the nodes into corresponding domains, solving the problem of detention that physical network and logical network does not match. The new routing table eliminates redundant information and increases objective resource table, increasing the useful information, reducing search detention. The simulation shows that the the average detention and average hop count of search model based on tree ring Chord are better than Chord and DChord, improves the efficiency of resource searching effectively.

**Key words:** network topology; IPv6; tree ring Chord; routing table

### 0 引 言

近年来,网络规模不断增长,网络拓扑结构处于持续的演化状态之中,针对大规模网络的研究也处于探索与发展阶段。大规模网络的特点是节点众多,分支复杂,并且包含多个网段<sup>[1,2]</sup>。传统的集中式网络拓扑结构在处理大规模网络问题时越来越显得无能为力,因此分布式网络系统应运而生。分布式网络是一种大规模覆盖网,分布式网络系统对大规模网络具有很好的适应性,它已经被证明具有良好的可伸缩性和可靠性<sup>[3,4]</sup>,成功应用于科学计算系统、即时消息传递系统等文件共享服务,大规模分布式网络的搜索问题也成为研究的热点问题<sup>[5]</sup>。

基于 DHT (distributed hash table, 分布式哈希表)

的搜索模型是目前扩展性最好的分布式网络路由方式之一,它具有精确定位的优良特性<sup>[6,7]</sup>。最具有代表性的基于 DHT 的协议有 Chord、CAN、Pastry 和 Tapestry 等<sup>[8,9]</sup>,然而,在基于 DHT 的路由协议中,大都存在物理网络和逻辑网络不匹配、没有考虑实际网络中节点的异构性和路由表冗余等问题,当网络规模较大时,这些问题严重制约着网络中资源搜索的性能。而这些原因大都和网络的拓扑结构以及路由表的结构密切相关。因此,如何解决大规模网络搜索效率低下的问题归根结底是如何设计有效的大规模网络拓扑结构和路由表结构的问题,使网络拓扑结构和路由表结构能够更好地处理大规模网络的搜索问题。

### 1 相关工作

国内外学者做了一些不同程度的研究,最具代表性的研究成果包括以下几个方面:文献[10]中提出把大规模分布式网络划分成两层结构,把网络的节点按能力分成超级节点和普通节点,双层结构虽然能在一

收稿日期:2011-01-10;修回日期:2011-03-29

基金项目:江苏省科学技术支持项目(BE2009158)

作者简介:徐 玉(1985-),女,安徽舒城人,硕士,主要研究领域为 P2P 网络;程春玲,副教授,主要研究领域为分布式网络,数据挖掘等。

定程度上减少网络中消息的流量,但当网络规模较大时,双层结构也显得无能为力,同样会产生网络拥塞和搜索延迟。Kim C S 等提出增加逆向 Finger 表 (DChord)<sup>[11,12]</sup>,把节点掌握的正向和逆向信息全部放在节点的 Finger 表里,这在一定程度上增加了搜索的范围,但增加一个逆向的 Finger 表浪费了宝贵的空间资源,并且当网络规模较大时,同样会出现搜索延迟的问题。

文中针对大规模网络的特点,提出的 V6TRC (IPv6 Tree Ring Chord) 模型利用 IPv6 地址协议特性和树环网特有的优点有效解决了大规模网络中物理网络和逻辑网络不匹配的问题,在路由表中利用节点的逆时针方向信息替换存在的冗余信息,并通过构建目标资源列表解决网络中重复搜索带来的网络延迟问题,为大规模覆盖网搜索算法的研究提供了一种新的思路。

## 2 模型的构造

### 2.1 设计思想

V6TRC 模型利用 IPv6 地址协议特有的地址聚类特性把网络中的节点划分到不同的自治域,采用树环网来构建具层次结构的覆盖网络结构,且树环中每个环的构造按照 Chord 协议组建。由于 IPv6 能够把节点地址划成三段,因此选用高度为 3 的树环网结构与此对应,在每个环中运用 Chord 协议。路由表设计时,在不增加额外的空间开销情况下,增加了逆时针节点的信息来替代冗余信息,此外增加了目标资源列表,有效解决重复搜索的问题,并能快速精确定位资源,提高了搜索性能。

### 2.2 基本概念

定义 1 树环 Chord: 是一棵广义的树  $T = \{N_1, N_2, \dots, N_n\}$ , 若任意的  $N_i \in T$ ,  $N_i$  是一个按照 Chord 协议组成的环,则称  $T$  为树环 Chord。

定义 2 节点顶级标识符 GID: 对节点的 IPv6 地址的前 24 比特的顶级聚类标识符进行哈希得到的函数值。

定义 3 节点次级标识符 SID: 对节点的 IPv6 地址的第 25 到 64 位的次级聚类标识符哈希得到的函数值。

定义 4 节点本地标识符 LID: 对节点的 IPv6 地址的最后 64 位哈希得到的函数值。

定义 5 本地域 (Loc\_domain): 具有相同的 GID 和 SID 的节点构成一个 Chord 环,称为本地域。

定义 6 次级域 (Sub\_domain): 具有相同 GID 的本地域代表节点 R\_LN 构成一个 Chord 环,称为次级域。

定义 7 顶级域 (Top\_domain): 由所有次级域代表

节点 R\_SN 构成一个 Chord 环,称为顶级域。

定义 8 本地域代表节点 R\_LN (Representative Loc\_domain node): 每个 Loc\_domain 中,与其所属的次级域的结合点称为本地域代表节点 R\_LN。R\_LN 是所属 Loc\_domain 中节点综合性能最高的节点。

定义 9 次级域代表节点 R\_SN (Representative Sub\_domain node): 每个 Sub\_domain 中,与顶级域的结合点称为次级域代表节点 R\_SN。R\_SN 是所属 Sub\_domain 中节点综合性能最高的节点。

### 2.3 拓扑构造

V6TRC 模型中节点标识符通过三元组 (GID, SID, LID) 表示。V6TRC 按照自底向上的动态原则组成,具体过程如下:首先具有相同的 GID 和 SID 的节点,构成一个本地域 Loc\_domain,再由每个 Loc\_domain 选择一个综合性能最高的节点为每个本地域代表节点 R\_LN。其中 GID 相同 R\_LN 组成一个个次级域 Sub\_domain,每个 Sub\_domain 选择出其代表节点 R\_SN,每个 R\_SN 组成一个顶级域 Top\_domain。Loc\_domain、Sub\_domain 和 Top\_domain 中的所有节点分别按照节点标识符的大小顺序顺时针排列构成一个 Chord 环,如图 1 所示。如此一来,可以达到将地址邻近的节点划分在同一个或相近的本地域中,以减少实际的传输时延,实现路由的本地性,从而提高了搜索性能。

在 V6TRC 中,所有的 R\_SN 和 R\_LN 不仅存放数据 (key, value),而且存放稀疏数据索引,负责相应域内节点的发布、查询、加入或离开网络请求,其它节点存放数据 (key, value),能在网络中进行数据查询与发布、请求加入或离开网络。

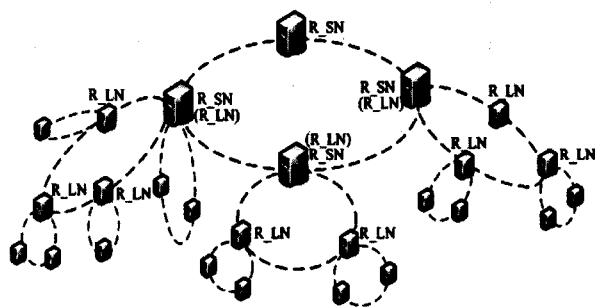


图 1 V6TRC 结构图

### 2.4 路由表设计

V6TRC 中本地域内各节点的路由表称为 Loc\_Finger, 位于次级域的节点还要维护本地域间通信的路由表,称为 Sub\_Finger, 位于顶级域的节点还要维护次级域间通信的路由表 Top\_Finger。为了提高路由效率, Loc\_Finger, Sub\_Finger 和 Top\_Finger 中用逆时针方向的信息替代冗余信息,具体的做法是:首先仍然从顺时针方向按照 Chord 协议中计算节点后继的方法,即利用  $[(ID + 2^{i-1})] \bmod 2^m$  逐个增加节点的后继节点,如

果当前后继节点和上一个后继节点相同,则利用逆时针方向的地址计算该节点的后继,逆时针方向的公式为 $[2^m + (ID - 2^{i-1}) \bmod 2^m]$ ,即 Finger 表逆时针方向的信息是小于或等于 $[2^m + (ID - 2^{i-1}) \bmod 2^m]$ ,且最靠近 $[2^m + (ID - 2^{i-1}) \bmod 2^m]$ 的节点。为解决重复搜索带来的路由延迟问题,增加了目标资源列表。

所谓目标资源列表即在源节点中存储目标资源的相关信息,每个表包含四个表项,分别是目标资源所在的节点标识符和其 IP 地址,目标资源所在节点的直接前驱的节点标识符和其 IP 地址,目标资源列表的格式见表 1。每个节点都有一个目标资源列表,其大小是 Finger 表的一半,目标资源列表中的表项利用 LRU (Least Recently Used,最近最久未使用)算法管理。

表 1 目标资源列表

符号	定义
Predecessor_id	目标资源所在的节点的直接前驱节点标识符
Predecessor_ip	前驱节点的 IP 地址
Object_id	目标资源所在的节点标识符
Object_ip	目标节点的 IP 地址

### 3 V6TRC 资源定位算法描述

V6TRC 模型中资源定位算法基于 Loc\_Finger、Sub\_Finger 和 Top\_Finger,即节点先通过 Loc\_Finger 在本地域查询,如果没有查询到,再通过 Sub\_Finger 到其他的本地域查询,如果仍没有查询到,则通过 Top\_Finger,到其他全局域内查询,直到查询到为止;如果在规定时间内没有查询到或者超时,则返回定位失败信息。V6TRC 的资源定位算法的主要步骤如下:

算法 1: V6TRC\_Location(N, Key)

输入:搜索键 Key(查询目标文件的关键字标识符)

输出:目标资源所在节点地址

```

if (ObjNode = n.find_in_objtab(N, Key)) //查询缓存中的资源索引
    return ObjNode;
//在本地环中搜索
if (ObjNode = n.local_closest_successor_node(N, Key))
    N' = n.find_LocNode(N); //取得节点 N 所在的本地域代表节点
//在次级域中查找
if (Chief_LocNode = n.sub_closest_successor_node(N', Key))
if (ObjNode = n.find_closest_successor_node(Chief_LocNode, Key))
    return ObjNode;
//在顶级域中获取节点 N 所在的次级域的代表节点
N'' = n.find_SubNode(N);
//在次级域中查找
if (Chief_SubNode = n.top_closest_successor_node(N'', Key))
if (Chief_LocNode = n.sub_closest_successor_node(Chief_SubNode, Key)) //在本地域中查找
    if (ObjNode = n.find_closest_successor_node(Chief_LocNode, Key))
        return ObjNode;

```

return NULL;

## 4 仿真实验与性能分析

### 4.1 实验环境

文中对 Chord、DChord 和 V6TRC 三种模型在相同的网络条件下进行仿真,模拟系统运行在主频为 2.4G、内存为 1G 的 PC 机上,软件环境为 Windows XP、C++、网络拓扑生成器 BRUTE。在大规模覆盖网中,如果域的个数过少,实验效果不明显,如果域的个数过大,则网络结构变得比较复杂,因此构造的 DChord 的域的个数为 16 个,为考虑每个顶级域下的节点数目基本平衡,V6TRC 的顶级域的个数也为 16 个,V6TRC 的每个顶级域下面有 10 个次级域。并假定 DChord 模型的域间延迟为 20ms,V6TRC 模型的顶级域间的延迟也为 20ms,V6TRC 模型的次级域间的延迟为 10ms,在本地域时,节点之间的延迟为 0.1ms。针对不同的拓扑规模,共选取 10 组数据进行模拟,节点数目从 10 万到 100 万线性增加,基本符合大规模网络的特点。在每组实验中,系统中的所有节点共随机地发出 100 万次查询请求,每组数据共进行 10 次实验,统计平均性能,仿真实验及性能分析如下。

### 4.2 仿真实验

首先对相同网络拓扑规模的 Chord、DChord 和 V6TRC 在随机查询条件下网络节点与平均延迟的关系进行比较,实验结果见图 2。

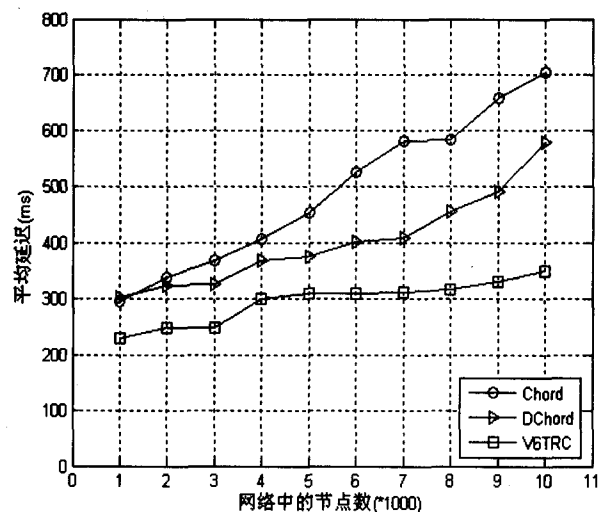


图 2 网络节点与平均延迟的关系

另一个能显示网络拓扑结构和路由表结构优越的参数是跳数,因此文中在相同网络拓扑规模下对 Chord、DChord、V6TRC 的跳数进行对比,实验结果如图 3 所示。

### 4.3 性能分析

仿真实验结果表明,在相同的网络拓扑规模条件下,V6TRC 在搜索的平均延迟和平均跳数方面都优于

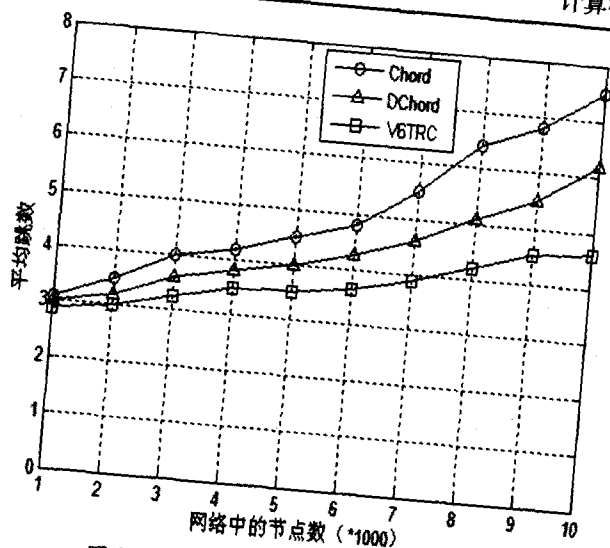


图 3 网络节点与路由跳数的关系

Chord 和 DChord。这主要有两方面的原因,一是 V6TRC 构建了合理的网络拓扑结构,基本实现了网络的逻辑结构和物理结构匹配,减少了搜索时“绕路”产生的延迟;二是由于 V6TRC 优化了路由表结构,消除了路由表的冗余信息,增加了节点逆时针方向的信息,为查询提供更多的启发信息。

## 5 结束语

提出了基于树环 Chord 的大规模覆盖网的拓扑结构。为适应 IPv6 地址协议,基于树环网建立了覆盖网络结构,并给出相应的资源搜索算法。V6TRC 模型中用 IPv6 进行较细粒度的地址聚类并利用树环网构造 P2P 网络的三层网络拓扑结构,这打破了用单层和双层结构的限制;运用 Chord 环中节点的逆向启发信息替换原 Chord 环中节点 Finger 表中的冗余信息,并增加目标资源列表,以提高资源搜索性能。最后通过仿真证明了 V6TRC 模型和其资源搜索算法的高效性。V6TRC 模型相比 Chord、DChord 大大提高了网络的搜索性能,同时也为基于大规模分布式网络搜索的研究提供了一个新的思路。

下一步的工作将在本文的基础上研究大规模分布

式网络下多维查询和模糊查询的问题。

## 参考文献:

- [1] 熊伟,谢冬青,焦炳旺,等.一种结构化 P2P 协议中的自适应负载均衡方法[J].软件学报,2009,20(3):660-670.
- [2] Mustafa R M, Butt, Ali R, et al. Designing accelerator-based distributed systems for high performance[C]//CCGrid 2010-10th IEEE/ACM International Conference on Cluster, Cloud, and Grid Computing. Melbourne, VIC, Australia: [s. n.], 2010: 165-174.
- [3] Werth D, Walter P, Emrich A, et al. Peer-to-peer business processes-conception, construction, tooling[J]. Journal of Software Maintenance and Evolution-Research and Practice, 2010, 22(6-7): 477-495.
- [4] Gao W L, Zhang G H, Wang M W, et al. An Agricultural Information Sharing System Based on P2P Network Technology[J]. Intelligent Automation and Soft Computing, 2010, 16(6): 945-951.
- [5] 牛新征,周明天,余堃.一种应用于移动 P2P 网络的资源协作共享策略[J].电子学报,2010,38(1):18-24.
- [6] 周鹏,鱼滨.基于 P2P 分布式数据库实时更新[J].计算机技术与发展,2007,17(5):144-147.
- [7] 黄烟波,周磊戈.基于流特征的 P2P 流量识别方法研究[J].计算机技术与发展,2009,19(9):46-49.
- [8] 宁多彪,陶中平,吕光宏.基于 PNS-PGrid 的 P2P 路由算法的设计与实现[J].计算机技术与发展,2009,19(12):47-50.
- [9] 杨文俊. P2P 网络系统中节点自组织管理机制[J].计算机技术与发展,2006,16(7):57-59.
- [10] Li Y, Chen F H, Sun X, et al. Self-Adaptive Resource Management for Large-Scale Shared Cluster[J]. Journal of Computer Science and Technology, 2010, 25(5):945-957.
- [11] Kim C S, Lee S, Han J I, et al. Dchord: an Efficient and Robust Peer to Peer Lookup System[J]. Malaysian Journal of Computer Science, 2010, 23(1):37-48.
- [12] Remondino M, Boella G. How users' participation affects reputation management systems: The case of P2P networks[J]. Simulation Modelling Practice and Theory, 2010, 18(10): 1493-1505.

上接第 96 页)

- (ICCCAS 2004). [s. l.]: IEEE press, 2004: 1057-1060.
- [1] 谢传泉,何晨.混沌神经网络模型中的模拟退火策略[J].上海交通大学学报,2003,37(3):36-39.
- [2] 刘守生.遗传算法与小波神经网络中若干问题的研究[D].南京:南京航空航天大学,2005.
- [3] 唐运虞,刘向东,修春波.一种新型暂态混沌神经网络及其在函数优化中的应用[J].计算机工程与科学,2006,28

(3):116-118.

- [12] 徐耀群,孙明,段广仁.暂态混沌神经网络中的激励函数[C]//WCICA2006.大连:[出版者不详],2006:2906-2911.
- [13] Hopfield J, Tank D. Neural computation of decisions in optimization problems[J]. Biology Cybernetics, 1985, 52:141-152.