

基于 IPv6 的分布式网络 搜索机制的研究

◎ 林玉香 著



山东大学出版社

基于 IPv6 的分布式网络 搜索机制的研究

© 林玉香 著



图书在版编目(CIP)数据

基于 IPv6 的分布式网络搜索机制的研究 / 林玉香著.
— 济南: 山东大学出版社, 2018.10
ISBN 978-7-5607-6205-0

I. ①基… II. ①林… III. ①数字电路—逻辑设计—
研究 IV. ①TN79

中国版本图书馆 CIP 数据核字 (2018) 第 229758 号

责任编辑: 宋亚卿

封面设计: 优盛文化

美术编辑: 张 荔

出版发行: 山东大学出版社

社 址 山东省济南市山大南路 20 号

邮 编 250100

电 话 市场部(0531)88363008

经 销: 新华书店

印 刷: 济南巨丰印刷有限公司

规 格: 710 毫米 × 1000 毫米 1/16

11.5 印张 206 千字

版 次: 2018 年 10 月第 1 版

印 次: 2018 年 10 月第 1 次印刷

定 价: 45.00 元

版权所有, 盗印必究

凡购本书, 如有缺页、倒页、脱页, 由本社营销部负责调换

内容提要

IPv6 作为国际下一代互联网发展的核心技术和创新平台，给中国互联网技术和产业的创新发展带来了难得的历史机遇。2018 年是政策推动下 IPv6 规模部署的重要一年，随着物联网等技术部署带来的海量连接需求，IPv6 的需求会越来越大。在中国移动互联网、云服务、物联网、区块链、5G 等技术创新浪潮中，IPv6 将成为支撑相关产业发展的重要基础资源。在技术和市场的推动下，我国 IPv6 产业的发展时机已经成熟，将迈入一个全新的全面服务时代。

加速 IPv6 升级部署的最好方式是与其他网络技术进行有效的协作融合，分布式网络技术有效整合了 Internet 的潜在资源（包括存储资源、计算资源和内存资源等）。借助其零成本、病毒式的传播模式，基于它的分布式应用得到了长足的发展，未来将与区块链、SDN/NFV（软件定义网络 / 网络功能虚拟化）、5G 网络等新技术融合，爆发出新的生命力。大部分分布式网络应用都有相应的网络模型支撑，但特定的分布式网络模型因自身的局限性而只适合特定的分布式网络应用，尤其对 IPv6 网络。因此，需要寻求更为通用的适用于 IPv6 的分布式网络模型，如何进一步提高基于 IPv6 的分布式网络路由效率，降低网络带宽占用，提高网络的可扩展性，提高海量数据的搜索性能以及增强网络流量检测的精准度，都是非常值得研究的问题。

绪 论 / 001

- 0.1 研究背景与意义 / 001
- 0.2 国内外研究现状 / 004
- 0.3 本书的主要研究内容 / 008
- 0.4 本书的组织结构 / 009

第 1 章 相关知识概述 / 011

- 1.1 分布式网络模型研究 / 011
- 1.2 分布式网络的特点 / 021
- 1.3 分布式网络应用的分类 / 022
- 1.4 分布式网络的关键技术 / 023
- 1.5 分布式网络搜索算法 / 025
- 1.6 分布式网络流量检测技术 / 030
- 1.7 IPv6 技术概述 / 034
- 1.8 JXTA 技术概述 / 037
- 本章小结 / 050

第 2 章 一种基于 IPv6 的混合分布式网络模型 / 051

- 2.1 设计思想 / 051
- 2.2 HMP2P6 的系统结构 / 052
- 2.3 节点的分类 / 053
- 2.4 上层网络结构 / 056
- 2.5 下层网络结构 / 065
- 2.6 HMP2P6 网络模型的资源定位 / 067
- 2.7 HMP2P6 网络模型的性能分析 / 069
- 本章小结 / 071

第 3 章	HMP2P6 网络模型的自适应机制 / 073
3.1	子网规模的自适应调整 / 073
3.2	节点的失效恢复机制 / 079
	本章小结 / 082
第 4 章	HMP2P6 网络模型的负载均衡机制 / 083
4.1	负载均衡思想 / 083
4.2	现有的分布式网络负载均衡算法 / 084
4.3	HMP2P6 网络模型的负载均衡机制 / 088
	本章小结 / 092
第 5 章	HMP2P6 网络模型的仿真及结果分析 / 093
5.1	P2PSim 的结构 / 093
5.2	P2PSim 基础上的仿真实现策略及仿真条件设置 / 096
5.3	仿真结果与分析 / 097
	本章小结 / 105
第 6 章	基于自适应对等组的搜索算法 / 106
6.1	JXTA 搜索算法存在的问题 / 106
6.2	自适应对等组 / 107
6.3	基于自适应对等组的选择性算法 (ASGnutella) / 111
6.4	自适应对等组相似度算法 / 115
6.5	仿真实验与结果分析 / 120
	本章小结 / 127
第 7 章	基于 JXTA 的 PAGES 模型 / 128
7.1	设计思想 / 128
7.2	PAGES 模型架构 / 129
7.3	PAGES 模型主要模块的设计与分析 / 132
7.4	基于 JXTA 的 PAGES 系统的实现 / 139
	本章小结 / 145

第 8 章 基于支持向量机的分布式网络流量检测系统模型 / 146

8.1 支持向量机技术 / 146

8.2 基于 SVM 的分布式网络流量检测模型 / 151

8.3 网络连接信息的提取及特征选取 / 154

8.4 实验及数据分析 / 160

本章小结 / 164

第 9 章 总结与展望 / 165

9.1 主要工作总结 / 165

9.2 未来工作展望 / 167

主要参考文献 / 168

绪 论

0.1 研究背景与意义

IPv6 是国际下一代互联网发展的核心技术和创新平台。我国于 2017 年 11 月 26 日印发的《推进互联网协议第六版 (IPv6) 规模部署行动计划》，明确提出“用 5~10 年建成全球最大规模的 IPv6 商用网络”。作为基础协议,IPv6 与 5G、区块链、SDN/NFV 等最新技术的结合,将成为 IPv6 下一代互联网发展的强大驱动力,推动产业发展。

2018 年是政策推动下 IPv6 规模部署的重要一年,随着物联网等技术部署带来的海量连接需求,IPv6 的需求会越来越大。工信部曾在 2018 年 5 月表示,到 2018 年年末,移动互联网 IPv6 用户规模应不少于 5 000 万户。其中,中国移动用户不少于 1 000 万户,中国电信用户不少于 1 000 万户,中国联通用户不少于 1 000 万户。

2018 年 6 月 20 日,阿里云和三大运营商联合宣布,将全面对外提供 IPv6 服务,并致力在 2025 年前助推中国互联网真正实现“IPv6 Only”(仅支持 IPv6)。对于亟待突破性发展的 IPv6 产业而言,这无疑是个“大动作”。在技术和市场的推动下,我国 IPv6 产业的发展时机已经成熟,云服务领军企业阿里云率先和三大运营商携手力推 IPv6 全面服务,并给出了具体的实施目标,在给产业界各方带来动力和信心的同时,更是宣告我国 IPv6 产业迈入了一个全新的全面服务时代。

IPv6 是数字化时代的网络基石,此次升级行动规格如此之高,不仅是因为 IPv6 的商业价值,更是因为 IPv6 将成为新技术革命和新产业升级的重要基石。

相比 IPv4 来说,IPv6 有三大优势可以为即将到来的数字化社会奠定基石。

(1) 近乎无限的地址数量将赋予每一个物体以专署 IP 地址。IPv4 的地址数量

极限为 43 亿个，无法为地球上的每人分配一个，而 IPv6 地址可以理解为是无穷的，用个形象的比喻就是连地球上的每一粒沙子都能分配一个 IPv6 地址。

(2) 面对海量连接管理更加方便。由于地址足够充足，因此不再需要人工配置或者动态主机配置协议 (DHCP) 这类地址分配服务，同样的道理，也无须网络地址转换 (NAT) 这样的地址转译。可以说，IPv6 节省了很多复杂的 IP 设计，让网络层次进一步扁平化。

(3) 为不同的应用场景提供了更好的移动特性。相比 IPv4，IPv6 提供了嵌入在网络设备中，支持互操作和移动的能力，这对万物智联时代不同场景的智能硬件应用而言至关重要。

加速 IPv6 升级部署的最好方式是与其他网络技术进行有效的协作融合。区块链技术是一种打破传统互联网中心化特征的分布式网络平台，它更像是一个人人都能记录、查看和维护的公用账簿，任何记录都有永久的时间标记，无法篡改，如此重建信任机制。基于区块链的技术机制，确保了每个参与者的诚信行为，让信任在区块链平台上得以良好的维持。在分布式网络的技术支撑下，IPv6 会在区块链领域取得突破性进展。未来，IPv6 将与区块链、SDN/NFV、5G 网络等新技术融合，爆发出新的生命力。

随着互联网用户和数据量的急剧增长，用来提供数据集中存储和计算能力的服务器已不堪重负，而以视频为主的多媒体数据服务以及大型文件传输的流行更突显了这一问题。为解决此类问题，分布式网络技术应运而生。有别于传统的客户端 / 服务器 (模式) (Client/Server, 简称“C/S 模式网络”)，分布式网络中的每个节点在行为上是自由的，在功能上是平等的，在连接上是互联的，所有节点分布式地组成一个整体网络。因此，它能够极大地提高网络效率，充分利用网络带宽，开发每个网络节点的潜力。C/S 结构和分布式网络结构 (P2P 结构) 的对比如图 0-1 所示。

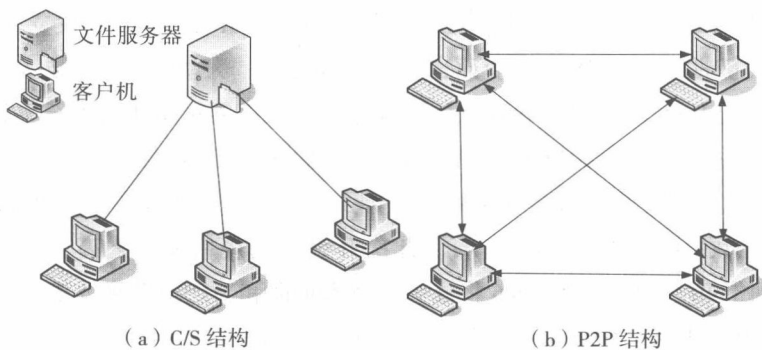


图 0-1 C/S 和分布式网络结构 (P2P 结构) 的对比

通过这种对等的服务，服务器的作用被淡化和减弱了，节点间通过分布、对等的算法实现协作和共享，所有数据可以分布存储在分布式网络内的各个节点上，网络的资源被最大化利用，整个网络的负载也可以达到均衡。

分布式网络应用的发展如火如荼、势不可挡。自 1999 年起，随着共享工具 Napster 的出现，分布式网络应用在短短几年之内就迅速发展并占领了多个网络应用领域。以 Napster、BitTorrent、eDonkey、KaZaA 为代表的文件共享应用，凭借自由开放的资源共享方式、高速的用户下载体验迅速成为主流的下载工具。以 Skype 为代表的网络电话（VoIP）应用凭借流畅清晰的通话质量、低廉的通话费用使网络电话迅速普及，对电信业务构成了强有力的冲击。以 PPLive、QQ 直播、PPStream 为代表的流媒体应用凭借清晰流畅的播放体验、丰富多彩的节目内容使网络电视摆脱了带宽瓶颈，成为视频、电视节目、电影等多媒体资源新的发布平台。

P2P 技术又称“对等互联网络技术”，是一种网络新技术。目前，P2P 技术将与新型网络技术融合，爆发出新的生命力。Ulord 是一个基于区块链技术构建的分布式 P2P 网络开源项目，这句话是对 Ulord 所使用技术的经典描述。Ulord 有别于其他区块链项目，它是颠覆性技术（区块链技术）与传统技术（分布式网络技术）的完美结合，这也是让人眼前一亮的地方。

P2P 通信在未来将是一种非常重要的通信方式，以网页实时通信（WebRTC）为代表的 P2P 通信将深刻影响人们的日常通信方式。

中国的 5G 网络有望在 2020 年正式商用，届时移动端网络的状况将大幅提高，P2P 通信的质量将大幅提高，将是 P2P 通信应用的大繁荣时期。中国基于 WebRTC 做实时通信云 PaaS（平台即服务）的厂商不胜枚举。野狗、声网等厂商已经打造出了成熟的 WebRTC 通信云平台，其中野狗通信云是国内第一家 P2P 通信免费的厂商。企业对实时视频的需求量大。社交、在线教育、金融、客服等都需要实时视频来给用户提供更好的实时互动服务。智能手机、电脑等硬件设备的性能不断提升，为 P2P 通信提供了良好的硬件基础。Google（谷歌）、Apple（苹果）等大厂商对 WebRTC 的更新和适配加速了 WebRTC 的普及，间接推进了 P2P 通信的普及。

P2P 技术的广泛应用也深刻影响了传统搜索技术，基于 P2P 的搜索系统大量出现。P2P 网络中的用户可以选择将本地的资源共享到网络，用户搜索时不通过中心服务器，节点间可以相互搜索。因此，P2P 搜索的效率与传统搜索系统相比具有很大优势，P2P 搜索的范围可以在短时间内以几何级数增长。同时，P2P 搜索技术在搜索到的信息的及时性、有效性等方面也有很大优势，并且能更快地定

位下载资源，缓解人们从搜索资源到下载资源之间的矛盾。随着 5G、物联网、云服务的加入，万物互联时代即将到来，海量连接数据的精准搜索定位将是一个研究课题。

P2P 应用已经是当今互联网上最广泛的应用之一。但 P2P 应用很多是基于 IPv4 网络，在 IPv6 网络上的应用仍存在制约其发展的障碍。其中，主要的障碍就是 P2P 系统的封闭及无标准性。另外，许多 P2P 系统虽然在拓扑结构上属于 P2P 结构，但许多关键服务是通过服务器来完成的。作为具有颠覆性的新一代网络技术，区块链在许多行业的应用层出不穷。IPv6 除了作为区块链的底层协议之外，与区块链结合将在安全性方面发挥巨大的效果。IPv6 通过地址的管理和路由机制，使 IP 层的溯源与可信验证成为可能。区块链技术去中心化的信任传递机制将为网络空间的发展注入强大的动力，在安全性上解决了信任的问题。二者结合，将极大地提高信息的可信性与真实性。大部分分布式网络应用都有相应的网络模型支撑，但特定的分布式网络模型因自身的局限性而只适合特定的分布式网络应用，尤其对 IPv6 网络。因此，需要寻求更为通用的适用于 IPv6 的分布式网络模型，如何进一步提高基于 IPv6 的分布式网络的路由效率，降低网络带宽占用，提高网络的可扩展性，提高海量数据的搜索性能以及增强网络流量检测的精准度，都是非常值得研究的问题。

0.2 国内外研究现状

分布式网络模型研究是分布式网络技术研究中的一个重要组成部分。网络模型的研究旨在为解决分布式网络的可扩展性、负载以及管理等方面的问题，提供可实现的方案。迄今为止，分布式网络技术研究已经取得了长足的发展，分布式网络模型也已经过了几代的更新，根据拓扑结构的不同，可以将分布式网络模型分为以下四种。

第一，中心化拓扑网络模型中存在中央索引服务器，用来存储资源的索引信息，资源的发现依赖该索引系统，而数据资源在对等节点间直接传递。第一代 MP3 共享软件 Napster 是一个中心化拓扑网络的典型应用。中心化结构最大的优点是维护简单，发现效率高。资源的发现依赖中心化的目录系统，发现算法灵活高效并能够实现复杂查询。最大的问题与传统客户机 / 服务器结构类似，容易造成单点故障、访问的“热点”现象和版权纠纷等相关问题。

第二，非结构化拓扑网络模型主要采用随机图的组织方式，没有中心服务器，

各节点完全平等。各节点间采用 TTL (Time-to-Live) 广播洪泛 (Flooding)、随机漫步或有选择地转发算法等机制建立通信, 提供资源发现方法, 典型的应用有 Gnutella (努特拉) 和 Freenet (自由网)。网络中的节点度数服从幂律 (Power-law) 分布规律, 能够较快地发现目的节点, 在面对网络的动态变化时, 体现了较好的稳定性和容错能力, 因此具有较好的可用性。同时, 可以支持复杂查询, 如带有规则表达式的多关键词查询、模糊查询等。

由于没有确定拓扑结构的支持, 非结构化网络无法保证资源发现的效率, 其扩展性较差, 即使需要查找的目的节点存在, 发现也有可能失败。由于采用洪泛等方式作为查找算法, 在查找的过程中网络的负载很重, 可能一个查询请求会被发放到整个网络, 造成网络资源的很大浪费, 并且可能产生网络的广播风暴, 造成网络瘫痪。

第三, 以基于分布式哈希表 (DHT) 的定位和路由机制为核心的分布式网络结构化拓扑网络模型是最新的研究热点。这种网络系统能够自适应节点的动态加入/退出, 有着良好的可扩展性、健壮性、负载均衡性和自组织能力。DHT 使用分布式哈希算法来解决结构化的分布式存储问题。分布式哈希算法的核心思想是将存储对象的特征 (关键字) 进行哈希运算, 得到键值, 对象的分布存储依据键值来进行。采用 DHT 的分布式网络路由协议主要有 Chord、CAN、Pastry、Tapestry。由于采用了确定性拓扑结构, DHT 可以提供精确的发现。只要目的节点存在于网络中, DHT 总能发现它, 发现的准确性得到了保证, 而且在每个节点维护少量的路由信息的情况下, 资源可以在有限的步骤内被定位到。以 Chord 协议为例, N 个节点的分布式网络中每个节点只需要维护 $O(\log N)$ 项路由信息, 即可在 $O(\log N)$ 跳内定位到资源。目前, 基于它们的应用主要集中在数据和文件共享系统上, 主要有协作文件系统 (Cooperative File System, CFS)、PAST、分布式存储系统 (OceanStore)。DHT 结构最大的问题是维护机制较为复杂, 尤其是节点频繁加入、退出造成的网络波动会极大地增加 DHT 的维护代价, 从而影响其稳定性。

第四, 如今网络上流行的分布式网络应用系统主要还是基于前两种拓扑结构, 或结合这两种结构而形成的混合分布式网络, 风靡全球的文件共享软件 KaZaA、BitTorrent、语音通信软件 Skype, 以及国内流行的分布式网络流媒体自播软件 PPLive, 都是混合分布式网络的应用。面对纷繁复杂的网络环境, 混合分布式网络能结合两种结构的优点, 很好地适应环境, 并可以根据具体环境适时调整部分结构。该结构的特点是在性能、可控性、扩展性方面都达到了一种折中。Considine (2002) 提出了基于 DHT 的群组 (Cluster) 概念, 在 DHT 结构网络上

把通信延时较短的节点分类组合成一个个群组，节点间的查找和通信可以通过该群组建立快速通信，DHT 的查询性能得以提高。Ganesan 等（2004）给出了一种层次化树形的分布式网络结构设计，把很多个小的 DHT 网络组织成多层的树状网络，从外面看来是这些小的 DHT 网络合并而成的一个大的 DHT 网络，有节点的层次化控制及网络节点的组织适应底层物理网络等优点。

以上所介绍的各种网络模型都有针对性地解决了分布式网络技术中某几方面的问题，但同时也存在其无法避免的先天缺陷。分布式网络技术的不断发展以及网络用户呈几何级数的增加，对当前的 Internet 架构产生了巨大的冲击，各种网络模型中存在的问题已经成为阻碍分布式网络技术进一步发展的一个障碍。分布式网络中主要存在以下问题：

第一，网络模型的构造与可扩展性。

可扩展性问题一直以来都是困扰分布式网络研究人员的大问题。在典型模型中，无论是集中式路由网络模型还是非结构化路由网络模型，都受到了严重的困扰。以目前应用最广泛的 Gnutella 网络模型为例，Gnutella 网络中的节点在搜寻数据时是以洪泛的方式将消息散布在网络上的，这会造成消息泛滥的问题，也使系统的可扩展性无法提升，并加重了网络的负荷。

结构化网络模型采用规则结构化的拓扑设计虽然在一定程度上解决了分布式网络的可扩展性问题，但同时也出现了新的问题：由于采用的是单向定位机制，所以支持的是精确匹配查询，当该机制用于文件共享等存在标识模糊性的网络中时，上述单向定位机制缺乏可选择性的弱点就暴露出来，精确匹配的限制导致结果唯一或结果集偏小（在存在多个同名副本的情况下）。此外，在 Chord 型的网络模型中，由于其查询机制带来了匿名性较弱的问题，所以，可扩展性所带来的问题并没有完全在结构化网络模型中得以解决。

第二，分布式网络模型中覆盖层网络与底层网络物理拓扑结构的差异问题。

同样，以 Gnutella 为例说明这个问题。在 Gnutella 路由网络模型中，经常会出现一条 Gnutella 消息多次在同一条链路上反复传递的情况，从而导致网络路由的效率低下，造成大量不必要的带宽损耗。这种情况产生的根本原因是 Gnutella 网络逻辑拓扑结构和网络底层实际拓扑结构存在严重的差异性。同样的问题也存在于其他的路由网络模型中。比如，结构化路由网络模型 Chord 中的绕路（Detouring）问题，就是由于该系统通过哈希算法所设计的映射方法，导致了在哈希环域中非常邻近的两个节点却有可能在实际网络拓扑中相距很远，而在实际网络中相距很远的节点经过哈希算法映射却可能成为哈希环中非常邻近的两点，从而导致路由效率的低下。

针对这些问题，国内外的研究者投入了大量的人力和物力进行研究，近年来产生了许多有针对性的解决方案。

分级集中式 P2P 网络路由模型（Hierarchical Aggregation P2P Network Routing Model, HARM）通过引入基于网络实际拓扑的网络分层和组群划分机制，对 Internet 中的所有节点进行区域性分组，配合分布式哈希散列算法（也称为“杂凑算法”）建立起一个具有完全分布式结构、高度可扩展性和健壮性的分布式网络模型。该网络模型避免了 Chord 等结构化网络普遍存在的绕路问题，并调整节点命名机制和路由表结构以降低路由存储开销，通过引入相关优化机制进一步提高网络模型的路由效率。但是该网络模型仍然具有服务质量不高、缺乏管理等问题。

Brocade（博科通讯系统有限公司）也提出了一个类似于 HARM 的基于 Internet 物理拓扑的分级覆盖层网络模型，它采用半集中式（Semi-Centralized）结构，通过在覆盖层网络中引入强节点，将物理距离较近的对等点聚集成组。一个强节点起的作用类似于一个网络区域的界标（Landmark），每个组群的强节点都是在 Brocade 组网之前预设的，并且通常由子网的网关或路由器担当。与 Brocade 提出的模型不同的是，HARM 是一个具有完全自组织机制的系统。这些相关的路由网络模型大部分只是针对某一问题进行详细的设计和探讨，对当前分布式网络模型存在的问题解决得不够彻底。

国外比较成熟的 P2P 应用研究一般有两类：运算群组和数据传输。国内的研究机构由于 CPU 核心技术的研发能力较弱，所以很少研究运算群组相关技术。用户日常接触和使用到的 P2P 应用大都是数据传输方面的研究成果。大家较为熟悉的成果就是文件传输和资源共享系统。

JXTA 以协议栈的形式解决了各种 P2P 应用中所涉及的各类问题，同时为各类应用提供了泛型编程接口。JXTA 协议族是通过 XML（可扩展标记语言）来描述的，并且它没规定实现的具体语言以及平台。因此，JXTA 是目前 P2P 研究领域第一个结构完整的广义架构。

综合当前国内外的 P2P 的各种理论与应用研究成果，JXTA 的可扩展性最好，并且由于 JXTA 是开放源代码，方便了 P2P 爱好者参与开发 P2P 系统。所以，JXTA 技术已经成为 P2P 研究领域参与人数最多的技术。

JXTA 最大的特点在于它是由一系列网络协议构成的，没有规定具体的实现平台和语言，所以 JXTA 是跨平台的技术。因此，在 JXTA 的基础上，实现 P2P 的各种应用，可以解决当前分布计算应用中的许多困难与问题。

JXTA 通过使用 XML 技术和小型、简单且便于开发的构造模块，可以使开发者从建立各自框架的复杂工作中得以解放，从而集中精力研究各类分布式计算应

用。我国网民最熟悉的两家搜索引擎公司百度和谷歌也致力于将 P2P 技术与搜索技术相结合的研究，其他一些公司和研究机构也纷纷加入 P2P 搜索相关技术的研究行列。然而，到目前为止，市场上还没有一种成熟的被广泛使用的 P2P 搜索系统。可以说，虽然目前国内外已有许多公司和研究机构认识到 P2P 技术的光明未来，并且投入了大量资金研发 P2P 技术，但 P2P 搜索系统的相关研究也只是刚刚起步。因此，研究基于 JXTA 的 P2P 网络的搜索技术具有良好的发展前景与重要意义。

目前，对分布式网络技术的研究大多都是在 IPv4 的环境中进行的，由于 IPv4 本身的局限性，在一定程度上限制了分布式网络技术的进一步发展。

0.3 本书的主要研究内容

作为基础协议，IPv6 与 5G、区块链、云服务、物联网、SDN/NFV 等最新的技术的结合，将成为 IPv6 下一代互联网发展的强大驱动力，推动产业发展。在技术和市场的推动下，我国 IPv6 产业的发展时机已经成熟，将迈入一个全新的全面服务时代。加速 IPv6 升级部署的最好方式是与其他网络技术进行有效的协作融合。本书围绕 IPv6 与分布式网络的有效融合方式，就如何进一步提高基于 IPv6 的分布式网络路由效率，降低网络带宽占用，提高网络的可扩展性，提高海量数据的搜索性能以及增强网络流量检测的精准度展开研究。

(1) 对分布式网络技术进行了探讨，对各种分布式网络的网络模型进行了研究和分析，并着重研究了非结构化分布式网络和结构化分布式网络。两种网络各有优缺点：非结构化分布式网络组织简单，稳定性好，但查询效率不高且不支持负载平衡；结构化分布式网络查询效率高，支持负载平衡，但稳定性不高，容易产生网络动荡。

(2) IPv6 除了作为区块链的底层协议之外，与区块链结合将在安全性方面发挥巨大的效果。IPv6 通过地址的管理和路由机制，使 IP 层的溯源与可信验证成为可能；区块链技术去中心化的信任传递机制将为网络空间的发展注入强大的动力，在安全性上解决了信任的问题。二者结合，将极大地提高信息的可信性与真实性的结合。结合 IPv6 的地址聚类性，并考虑到节点的异构性，本书提出了 HMP2P6 (Hybrid Model of P2P network based on IPv6) 网络模型，将结构化的 Chord 网络与非结构化的 Gnutella 网络进行有效的融合。针对 Chord 所存在的绕路、单向定位缺陷、缺乏缓存机制等问题，模型中还提出了基于最近访问记录的双向定位策略，

提高了资源的查找效率,降低了查找延时。

(3) 针对 HMP2P6 网络模型中超级节点的承载力有限和 Gnutella 子网的洪泛效果不稳定问题,本书提出了子网规模的自适应调整策略,保证了系统模型的稳定性。针对 HMP2P6 网络模型中由“热点”访问导致的负载不平衡问题,本书提出了基于节点间物理距离的负载均衡方案。该方案利用 IPv6 的地址聚类性,使负载转移在物理位置相近的节点区域之间进行,减少了负载转移的开销(包括带宽、延迟等)。

(4) 在深入研究分析 JXTA 的系统架构和协议,以及现有的 JXTA 搜索机制的原理和优缺点的基础上,针对 HMP2P6 网络模型提出了改进的搜索算法,并设计了一种 PAGES 模型,且对此种模型在适用场景下进行了仿真测试。最后根据搜索机制的设计思想完成了整个 P2P 系统架构的设计以及系统关键部分的实现。

(5) 为了应对万物互联时代,海量连接数据的精准搜索定位,本书在支持向量机的基础上,提出了一个基于支持向量机的分布式网络流量检测模型,并对该模型的功能模块进行了优化,为提高基于 IPv6 网络的 P2P 搜索算法的精准度提供了参考。

0.4 本书的组织结构

本书共分为 10 部分。

绪论。对本书的研究背景、主要内容、问题的提出以及本书的主要结构进行了介绍。

第 1 章 相关知识概述。重点研究并分析了几种典型的分布式网络模型,分布式网络的关键技术,分布式网络搜索算法,分布式网络流量检测技术,IPv6 相关基础知识等。

第 2 章 一种基于 IPv6 的混合分布式网络模型。结合 IPv6 的地址聚类性,并考虑到节点的异构性,提出了一种基于 IPv6 的混合分布式网络模型 HMP2P6,将结构化的 Chord 网络与非结构化的 Gnutella 网络进行有效的融合。针对 Chord 所存在的绕路、单向定位缺陷、缺乏缓存机制等问题,模型中还提出了基于最近访问记录的双向定位策略,提高了资源的查找效率,降低了查找延时。

第 3 章 HMP2P6 网络模型的自适应机制。针对超级节点的承载力和 Gnutella 子网的洪泛效果问题,提出了子网规模的自适应调整策略,保证了系统模型的稳定性。

第4章 HMP2P6 网络模型的负载均衡机制。针对 HMP2P6 网络模型中由“热点”访问带来的负载均衡问题，提出了一种基于节点间物理距离的负载均衡方案。

第5章 HMP2P6 网络模型的仿真及结果分析。介绍了分布式网络模拟系统 P2PSim，且采用 P2PSim 进行了仿真实验并对结果进行了分析。

第6章 基于自适应对等组的搜索算法。在提出的基于 IPv6 的分布式网络模型 HMP2P6 的基础上，提出了改进的搜索算法，并设计了一种 PAGES 模型，并对此种模型在适用场景下进行了仿真测试。

第7章 基于 JXTA 的 PAGES 模型。在上一章的基础上提出 PAGES 模型，并对基于 PAGES 模型的系统进行总体设计和实现。

第8章 基于支持向量机的分布式网络流量检测系统模型。在支持向量机的基础上，提出了一个基于支持向量机的分布式网络流量检测模型，并对该模型的功能模块进行了优化和实验验证。这为提高基于 IPv6 网络的 P2P 搜索算法的精准度提供了参考。

第9章 总结与展望。对全书的主要工作进行了总结，对未来进行了展望。