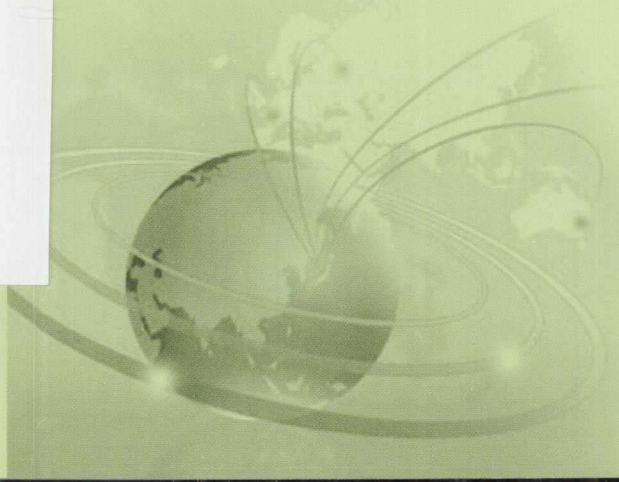
华中师范大学出版基金丛书
学 术 著 作 系 列

拓扑感知应用层组播模型构建及性能优化方法

崔建群 著

CBJJ



華中師範大學出版社

01306307O

TP393.02
43

华中师范大学出版基金丛书
学 术 著 作 系 列

本书由华中师范大学出版社提供的出版基金全额资助

拓扑感知应用层组播模型构建及性能优化方法

崔建群 著

北京航空航天大学
藏书
图书馆

TP393.02
43

北航 C1670269

華中師範大學出版社

新出图证(鄂)字 10 号

图书在版编目(CIP)数据

拓扑感知应用层组播模型构建及性能优化方法/崔建群著. —武汉:华中师范大学出版社,2013.5

ISBN 978-7-5622-6068-4

Ⅰ.①拓… Ⅱ.①崔… Ⅲ.①计算机网络—网络拓扑结构—研究 Ⅳ.①TP393.02

中国版本图书馆 CIP 数据核字(2013)第 094503 号

拓扑感知应用层组播模型构建及性能优化方法

© 崔建群 著

责任编辑:罗 挺 冯会平
责任校对:易 雯
编 辑 室:学术出版中心
出版发行:华中师范大学出版社有限责任公司
社址:湖北省武汉市珞喻路 152 号
传真:027—67863291
网址:http://www.ccnupress.com
印刷:武汉理工大印刷厂
字数:186 千字
开本:787mm×960mm 1/16
版次:2013 年 6 月第 1 版
定价:29.00 元

封面设计:罗明波
封面制作:胡 灿
电 话:027—67863220
电 话:027—67863040(发行部)
027—67861321(邮购)

电子邮箱:hscbs@public.wh.hb.cn
督印:章光琼

印张:11
印次:2013 年 6 月第 1 次印刷

欢迎上网查询、购书

敬告读者:欢迎举报盗版,请打举报电话 027—67861321

前　言

组播(multicast)是一种一对多的网络传输服务，它是互联网上单播、尽力传送模型的扩充，是许多 Internet 应用，如内容分发、文件共享、视频会议等服务的基础。由于组播服务的基础性和应用驱动的特殊性，如何提高组播的服务质量，即提高组播服务的高效性、健壮性(又称鲁棒性，robustness)和可扩展性等，一直是互联网的重要研究课题。

传统的 IP 组播运行在网络层，发送端发送的数据包由传输拓扑中的每一个中间路由器复制、转发到任意数量的接收端，在传送的物理链路上没有产生重复的数据包，可以获得良好的数据传播性能。但是由于 IP 组播服务需要依靠 Internet 基础结构中的路由器支持，因而无法得到广泛的应用。为此国内外的众多研究机构提出了将组播服务提升到应用层来实现的解决机制，即应用层组播(application layer multicast，ALM)。应用层组播将组播功能从路由器转移到端系统，由端系统完成所有组播组通讯的功能，从而摆脱了传统的 IP 组播对路由器的依赖，充分发掘端用户的计算资源，并且可以灵活地添加各种功能。但应用层组播在数据传输过程中会产生数据冗余，因此它们的数据传播性能比 IP 组播差。

拓扑感知(topology-aware)应用层组播由于采用事先探测端节点的拓扑信息方法，充分利用所获取的底层网络拓扑信息来构造覆盖网络，使组播树可以尽可能地与真实网络吻合，减小了因覆盖网络与真实网络不符而导致的最短路径计算误差，提高了数据传播性能，而成为目前应用层组播研究的一个热点。但是目前这种基于拓扑感知的应用层组播还未能大量应用于实际网络中，主要原因在于拓扑感知应用层组播在拓扑发现、覆盖网络构造及数据转发、组播生成树负载均衡以及故障恢复等方面还存在一定的模型构建和性能优化问题，本书针对上述问题展开了较深入的研究。

本书共分为 8 章，各章的主要内容组织如下：

第 1 章为绪论。该章介绍了本书的主要研究背景及研究意义，对当前相关研究现状进行了总结、归类与分析，给出了本书的研究目标和研究内容。

第 2 章对现有拓扑感知应用层组播进行了较详细的分析和讨论，对拓扑

感知应用层组播的原理、相关算法作了一定的介绍，并通过对网络代价、算法复杂度的计算和对应用层组播重要性能指标的实验模拟，将其与 IP 组播及典型应用层组播协议进行比较，从而对拓扑感知应用层组播的优缺点进行综合评价。

第 3 章针对拓扑感知应用层组播方案是否能够最大程度发挥其优势，源路径拓扑发现的效率和准确度至关重要这一核心问题，首先引入网关级拓扑图定义，并证明了对于相同的组播成员，最大前缀路径匹配算法根据网关级拓扑图与完整拓扑图所生成的组播树相同这一原理。根据上述原理，介绍了一种利用 p-tracert 路径信息的延时粗粒度匹配法来构造网关级拓扑图的方法，并对拓扑信息的获取、网关级拓扑图的构造、拓扑图及 p-tracert 路径信息的维护及其他优化策略进行了详细讨论，并通过性能分析和仿真实验证明上述方法不但简化了构造拓扑图所需的拓扑信息，而且加快了获取拓扑信息和构造拓扑图的速度，同时并未降低所生成的拓扑图与真实拓扑图的匹配度，有效优化了拓扑感知应用层组播中最耗费资源的源路径拓扑发现过程。

第 4 章对应用层组播覆盖网络构造方法进行了分析和研究。应用层组播中成员加入时延是用户衡量组播服务质量的重要性能指标，成员加入时延过长，可能直接导致用户还未加入组播组就因等待时间过长而退出组播服务。本章介绍了两种优化成员加入时延的拓扑感知覆盖网络构造方法：子网优先拓扑感知覆盖网络构造方法(STAG)和多域拓扑感知(MTAG)覆盖网络构造方法。前者可以在不影响正常组播的情况下，大大提高在同一子网中新节点加入的处理速度，节约系统资源；后者则可节省拓扑发现时间和缩短组播树的深度，在优化成员加入时延的同时，缩短组播数据转发时延。

第 5 章探讨了应用层组播负载均衡问题的解决方案。拓扑感知应用层组播在构造组播树时，考虑了底层物理网络的真实连接情况，其目的是尽量缩短端到端的数据转发延迟，但这可能会导致有些离组播源节点近、带宽高的端系统需要为很多子节点提供数据转发服务，而成为整个组播树的瓶颈。为平衡上述需求，本章介绍了一个更全面的求解应用层组播树的问题模型：具有度约束的最小时延负载均衡生成树模型 MDDLRB，希望在不超过节点度约束的前提下，达到整个组播树传播时延最小、负载最均衡的状态。同时为解决 MDDLRB 问题，给出了一种启发式算法时延递增均衡优化策略 DIB，并对其在不同环境和性能要求下进行了相应扩展。另外，为解决当拓扑感知组播生成树处于相对静止状态时(没有节点加入和退出)，并没有任何机制保证每个节点有足够的带宽来支持所有子节点的组播服务问题，本章还介绍了

试读结束：需要全本请在线购买：www.ertongbook.com

一种自适应带宽调整机制 SAB 来动态调整子节点的数量，均衡系统负载，减少节点的断连率。

第 6 章针对应用层组播需要良好的故障恢复机制，才能确保组播服务的鲁棒性这一问题展开讨论。现有拓扑感知应用层组播中的故障恢复机制极其简单，如果节点在一定的时间间隔内发现自己无法获得父节点的组播数据，则必须以新节点的身份向源节点再次申请加入组播树。由于拓扑感知应用层组播的成员节点加入机制需要耗费较多的系统资源和时间，另外上游节点的故障可能会影响所有以该节点为根的子节点，因此这种被动的故障恢复机制大大削弱了拓扑感知应用层组播的系统性能。为此本章介绍了一种反向抑制主动告警故障检测与恢复机制 RAA，该机制通过提前预警避免因缓冲区没有数据而导致的服务中止，同时通过反向抑制告警策略，减少重复和错误警报数量，为故障定位排除干扰。实验证明这种机制可以获得理想的告警准确率，并能以较小的维护控制开销降低组播成员因故障而导致的断连延时。

第 7 章结合应用层组播的具体技术介绍了三种不同的应用层组播流媒体播放系统，并对三种系统具体的设计和实现过程进行了分析和描述。其中，可定制转发的流媒体直播系统在设计中为了减少不必要的数据丢失和保证系统播放的流畅性而实现了一种可定制的转发机制，这种机制通过消除节点之间接收转发数据因存在时延而造成的数据延迟丢失，有效解决了由于节点退出和组播树调整造成的端主机播放器频繁缓冲的问题。基于 Scribe 的流媒体直播系统在设计中应用了基于发布/订阅模式的应用层组播协议 Scribe，并对 Scribe 的组播树管理机制进行了改进，系统底层通过采用通用的、可扩展的、高效的 P2P 系统 Pastry 在 Internet 中组建了一个分布式的、健壮的、自组织的覆盖网络。基于 JMF 的应用层组播流媒体播放系统包括服务器和客户端，在实现了直播的同时还具有本地媒体播放和点播的功能，它主要采用了 Java 提供的 JMF 多媒体框架技术，应用了本书前面章节所提出的拓扑发现、覆盖网络拓扑构造技术等新的解决方案，能够适应在低带宽下为用户提供流畅的多媒体传输和播放服务。

第 8 章对本书进行总结，对拓扑感知应用层组播模型构建及性能优化方法中有待解决的问题进行了思考，并展望了后续工作。

本书所介绍的研究工作是经过华中师范大学计算机学院和武汉大学计算机学院众多科研人员多年学习、研究和工程实践沉淀的成果。特别感谢武汉大学计算机学院何炎祥教授和华中师范大学计算机学院何婷婷教授对本

书的指导。参与本研究工作的人员包括吴黎兵、贾珂铭、陈传河、赖敏财、叶咏佳、范静、高宽等，在此对他们表示衷心的感谢。

本书是国内第一部专门针对拓扑感知应用层组播的研究著作，对相关领域的研究人员具有一定的借鉴意义和参考价值。本书的出版得到国家自然科学基金(No. 61170017)、湖北省自然科学基金(2011CDB156)、武汉市青年科技晨光计划(No. 200950431183)、华中师范大学中央高校基本科研业务费专项资金(No. 09010064)等项目的资助，在此一并表示感谢。

拓扑感知应用层组播模型构建及性能优化方法(Model Construction and Performance Optimization Methods for Topology-aware Application Layer Multicast)是当前处于科学前沿的论题，许多理论和思想还处于探索阶段，由于作者的水平和经验有限，错误和不妥之处在所难免，恳请读者给予批评指正，共同推进应用层组播研究的进步和发展。

作者

2013 年 2 月于桂子山

目 录

第1章 绪 论

IP组播需要Internet基础结构中的路由器支持，并由这些组播路由器负责与组播有关的路由、复制和转发数据包，这些要求导致IP组播从给出到现在将近30年的时间还是未能在Internet上得到广泛使用。针对IP组播的上述缺陷，应用层组播给出保持网络层单播而由应用层端主机来实现组播转发功能，解决了IP组播不能被广泛应用的问题，成为近年来的研究热点。

本章首先介绍应用层组播的研究背景和意义，然后分析了国内外在应用层组播领域的研究现状，并由此得出本书的研究目标和研究内容。

1.1 研究背景和意义

组播(multicast)是一种一对多的网络传输服务，它是互联网上单播(unicast)、尽力传送模型的扩充，是许多Internet应用，如内容分发、文件共享、视频会议等服务的基础。由于组播服务的基础性和应用驱动，如何提高组播的服务质量(quality of service，QoS)，即高效性(efficiency)、健壮性(robustness)和可扩展性(scalability)等，一直是互联网的重要研究课题。

1.1.1 IP组播

传统的组播运行在网络层，称为IP组播。20世纪80年代末至90年代初斯坦福大学的Steve Deering首先在ACM Transaction杂志上发表的学术论文[1]及博士论文[2]中提出IP组播。IP组播用于一对多、多对多的组通信。IP组播需要Internet基础结构中的路由器支持，并由这些组播路由器负责与组播有关的路由、复制和转发数据包。每一个IP组播会话，都存在一个传输

拓扑(或称路由拓扑)。IP 组播传输是通过传输拓扑实现的,即发送端发送的数据包由传输拓扑中的每一个中间路由器复制、转发到任意数量的接收端。IP 组播的传输拓扑结构主要是树状模型,其组成元素是组播路由器,IP 组播原理如图 1.1 所示。

图 1.1 中 A、B、C、D 分别代表端主机节点,R_1～R_5为组播路由器,图中显示了主机 A 利用 IP 组播将数据发送到 B、C、D 主机的过程。从图 1.1 中可以发现主机 A 只需要发送一个数据包,而由组播路由器负责将其复制分发到三个目的端节点,在传送的物理链路上没有产生重复的数据包。

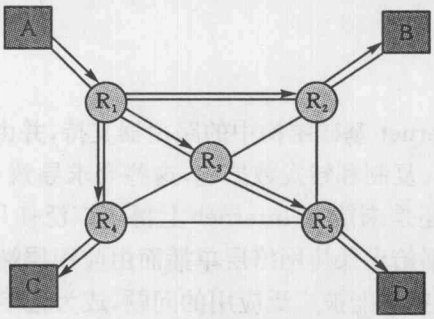


图 1.1　IP 组播原理示例

IP 组播通信必须依赖于 IP 组播地址,在 IPv4 中它是一个 D 类 IP 地址,范围从 224.0.0.0 到 239.255.255.255,并被划分为局部链接组播地址、预留组播地址和管理权限组播地址三类。其中局部链接组播地址范围为224.0.0.0～224.0.0.255,这是为路由协议和其他用途保留的地址,路由器并不转发属于此范围的 IP 包;预留组播地址为 224.0.1.0～238.255.255.255,可用于全球范围(如 Internet)或网络协议;管理权限组播地址为 239.0.0.0～239.255.255.255,可供组织内部使用,类似于私有 IP 地址,不能用于 Internet,可限制组播范围。

使用同一个 IP 组播地址接收组播数据包的所有主机构成了一个主机组,也称为组播组。一个组播组的成员是随时变动的,一台主机可以随时加入或离开组播组,组播组成员的数目和所在的地理位置也不受限制,一台主机也可以属于几个组播组。此外,不属于某一个组播组的主机也可以向该组播组发送数据包。

为了向所有接收主机传送组播数据,用组播分布树来描述 IP 组播在网络中传输的路径。组播分布树有两个基本类型:有源树和共享树。有源树是以组播源作为有源树的根,有源树的分支形成通过网络到达接收主机的分布

试读结束:需要全本请在线购买:www.ertongbook.com

树,因为有源树以最短的路径贯穿网络,所以也常被称为最短路径树 SPT (spanning tree)。共享树以组播网中某些可选择的组播路由中的一个作为共享树的公共根,这个根被称为汇合点 RP(rendezvous point)。共享树又可分为单向共享树和双向共享树。单向共享树指组播数据流必须经过共享树从根发送到组播接收机。双向共享树指组播数据流可以不经过共享树。

IP 组播传输是高效的,因为它能避免在物理链路上传输重复的数据包,节省了网络带宽。但 IP 组播存在以下问题[3]:

(1)路由器必须为每个组播组保存状态,扩展性差;

(2)要求所有路由器都支持,不利于推广使用;

(3)用统一的模型来适应所有应用,算法设计困难;

(4)组播组加入、退出和管理等开销大;

(5)组播地址空间太小(针对 IPv4);

(6)打破了传统的根据进入流量计费的机制。

IP 组播在安全、拥塞控制等方面也存在问题。这些问题的存在导致 IP 组播没能在 Internet 上得到广泛使用。

1.1.2 应用层组播

针对 IP 组播不能被广泛应用的事实,应用层组播给出保持网络层单播而由应用层端主机来实现组播转发功能。

20 世纪 80 年代初总结出 Internet 边缘论原则的 MIT 教授 David Clark 给出了著名的"End-to-End Argument"理论[4],其主要思想是由于互联网是一种尽力而为(best effort)的不可靠网络,那么任何一种具体的应用功能只有在核心网络的边缘(应用层)才能被完全和正确地验证并实现,所以力求将给出的某种应用功能作为通信系统的本身性质是不可能的。这种让网络核心部分只做最通用的数据传输而不实现特殊应用的思想有不少优点:如降低核心网络复杂性,便于升级;提高网络通用性和灵活性,增加新应用不必改变核心网络;提高可靠性等。而 IP 组播恰恰试图将分组通信功能置于核心网络(路由器)中,所以造成 IP 组播发展陷入了困境。

基于这种"End-to-End Argument"思想,同时面对组播通信需求的不断增长同 IP 组播不能在 Internet 上广泛应用的矛盾,国内外的众多研究机构给出了将组播服务提升到应用层来实现的解决机制,即应用层组播 ALM (application layer multicast)。

1. 应用层组播原理

应用层组播中端主机负责与组播有关的路由、组成员管理、复制和转发数据包等功能。端主机自组织成一个覆盖网络(overlay network),并成为覆盖网络中的节点。覆盖链路(overlay link)由两个端主机之间的单播传输服务实现,数据包在覆盖链路上传输。在覆盖网络中,构成覆盖链路的两个端主机被认为是邻居,且每一个端主机只和邻居节点通讯。每一个应用层组播会话,也存在一个传输拓扑。应用层组播的传输拓扑结构主要是树状模型和网状(mesh)模型,其组成元素是端主机。应用层组播传输是通过构建于覆盖网络上的传输拓扑实现的,其工作模式类似于 IP 组播传输,如图 1.2 所示。

图 1.2 描绘了三种不同实现方式的应用层组播,它们共同的特点是网络中传递的数据包与单播数据包相同,数据的复制与转发不是在路由器中完成而是在端主机 A、B、C、D 中完成。

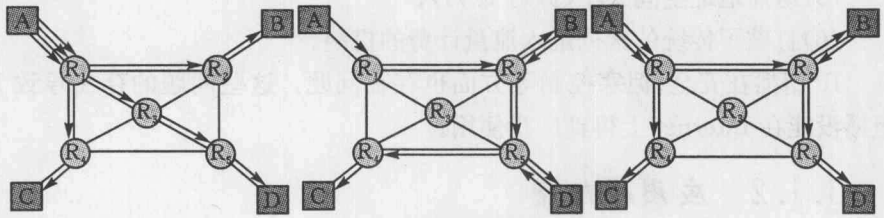


图 1.2　应用层组播示例

从网络设计的角度来看,应用层组播和传统 IP 组播在网络代价模型以及路由策略方面全然不同。这些主要的不同之处包括:

(1)网络可达性。端节点之间的应用层覆盖网络是一个全互联的网络,每个节点都可以通过单播连接到达其他节点,而 IP 组播中一个路由器到达另一个路由器的路径由物理链路定义,一个有 n 个节点的应用层组播覆盖网络将有 n^{n-2} 个不同的支撑树。

(2)网络代价。IP 组播中网络的代价通常由单个链路的代价和来决定的。但是从应用的角度来看,网络的代价是由从服务商获得的接入骨干网络的接入带宽所需要的代价来决定的,这种不同的代价尺度对设计和选择路由策略有很深的影响。

(3)路由限制。IP 组播路由通过最短路径树(SPT)最小化从源到每个目标节点的最小平均时延。构造应用层网络路由则需要适应不同的应用需要。例如对于流媒体应用或会议应用,一个满足任意两个节点之间的时延约束的路由策略就是一个高质量的策略。

与 IP 组播相比,应用层组播具有以下优点:

(1)应用层组播能够很快就进入应用,不需要改变现有网络路由器。

(2)接入控制更容易实现。由于单播技术在这方面比较成熟,而应用层组播是通过终端系统之间单播来实现的,所以差错控制、流控制、拥塞控制容易实现。

(3)不需要采用特殊有限的组播地址,直接使用单播地址就可以实现应用层组播服务。

(4)应用层组播易于针对特定应用进行优化,或者针对不同应用采用不同的应用层组播协议。

2. 应用层组播性能评价指标

评价应用层组播方案的性能好坏,可以从以下角度进行衡量。

(1)可扩展性。由于应用层组播中的部分成员节点取代了路由器的功能,负责数据的复制和转发,而处于应用层的主机的情况比较复杂,因此能够同时支持的组成员的数目越多,表明应用层组播方案的可扩展性越好。

(2)路径伸展率(path stretch)。路径伸展率也被称为相对延迟损耗RDP(relative delay penalty),它是指对于每个成员节点,数据从数据源发送,然后经过其他端节点的转发到达目的节点,所经过的路径的长度与直接由源节点通过单播方式发送到成员所走过的路径的长度的比值。平均路径伸展率APS(average path stretch)的计算公式如式1.1所示,其中S_r是某个成员节点r的伸展率,$|R|$是组播组中的成员数目。

$$\mathrm{APS}=\frac{\sum_{r\in R}S_r}{|R|} \tag{1.1}$$

纯单播模式由于选择的是最短路径,其伸展度最小,所以组播通信模型在这方面的性能都要与单播模式相比较,也就是单播模式的伸展度是标准1。IP组播报文的转发是由组播树转发,其伸展度比单播稍大。而应用层组播的每个报文都是依靠一些终端主机作为中继来转发,所以其伸展度比IP组播要大。一般情况下,单播模式的伸展度最小,应用层组播的伸展度最大,IP组播介于两者之间。

(3)链路压力(link stress)。除了服务器之外,应用层组播中的某些成员节点需要对数据进行复制转发,因此链路压力是指在某一条链路上同时要转发的相同的数据包的数目。显然IP组播进行转发的时候并未进行多余的复制,所以其链路压力是最优值1。平均链路压力ALS(average link stress)的计算公式如式(1.2)所示,其中S_l是某条链路l的压力,$|L|$是物理拓扑中链路的数目。

$$\mathrm{ALS}=\frac{\sum_{l\in L}S_l}{|L|} \tag{1.2}$$

IP 组播中由于分组不进行多余的复制，因此其链路压力最小（压力为1）；对于纯单播模式，组播组有多大，在每条链路上就有多少个同样的分组，这样一来每条链路的“劳动强度太高”或称“通信压力太大”，这也是纯单播模式不适合组播通信的原因；而对于应用层组播，可能在某些链路上有相同的分组，也就是存在一些“劳动强度大”的链路或路由器。一般情况下，应用层组播的链路压力指标介于纯单播模式和 IP 组播之间，IP 组播的强度最小，纯单播的强度最大。

（4）资源使用量（resource usage）。定义为参与数据传输的所有链路上 delay * stress 的和，这里认为延时大的链路代价高，给网络传输的过程提供了一个参考值。

（5）组播控制开销（control overhead）。这和“可扩展性”有一定的关系，但是两者并不等价。组播维护开销包括：保存的状态信息的数量；组播节点之间为维护组播组的交互信息数量；组播节点加入、退出、失效等信息在组播组内同步的时间开销等。例如首次加入时延定义了一个新成员从加入一个正在进行的会话到接收到该会话第一个数据包的时间。加入代价则定义了一个成员在加入组播树的过程中与相关的节点交互的报文数量。

（6）组播节点的“度”（degree）。在应用层组播中，很多转发由位于网络边界性能不太高的主机来完成，每个节点的带宽和处理能力都存在限制，所以在应用层组播算法的设计中，通常要限制每个节点的“度”。

（7）鲁棒性（robustness，即健壮性）。应用层组播系统的鲁棒性无法和使用路由器或者专用服务器的组播系统相比，但是可以通过一定的机制提供一定程度的鲁棒性。而且对于不同的应用，要求的鲁棒性也有所不同。鲁棒性通常用数据丢失率来描述。

3. 应用层组播存在的问题

目前应用层组播研究集中于视频会议系统、媒体流的分发系统（如视频广播）和订阅分发系统（Publish/Subscribe System）等。它主要用于实时的多媒体传输，这利用了多媒体信息的性质，即在传输链路质量下降时，用户仍可利用收到的低速率或者不完整的信息，也利用了组播“时间上集中、空间上分布”的特点。虽然应用层组播在部署、可定制性等方面有着 IP 组播无法比拟的优势，但它也存在以下一些问题：

（1）可靠性。终端系统的可靠性比路由器差，由于参与转发的端系统可

能不稳定，使组播转发的可靠性受到影响。

(2)可扩展性。底层的路由信息对应用层组播来说是隐藏起来的，可扩展性不好。

(3)延迟较大。IP 组播主要是链路上的延迟，而在应用层组播中，数据还要经过端系统，由于参与转发的端系统的性能无法保证，可能导致延迟、转发速率等性能的下降。

(4)传输效率不如 IP 组播。应用层组播在数据传输过程中会产生数据冗余，因此它们比 IP 组播的效率差。

其中影响应用层组播传输效率不高的主要因素有以下几点：

(1)端系统对 IP 网络的了解有限，节点参与组网时，只能通过探测获得一些网络性能参数，选取的逻辑链路难以优化。

(2)主机不了解 IP 网络的拓扑结构，只能通过带宽和时延等外在的特性参数，以启发式的方式建立覆盖网络，逻辑链路不能较好地利用质量较好的底层网络资源，覆盖网络的多条链路可能经过同一条物理链路。

(3)为了弥补 IP 组播在安全可靠方面的缺憾，应用层组播还要增加诸如安全可靠之类的服务，这将进一步消耗主机的资源。

综上所述，可以看出应用层组播有其广泛的应用前景，关键在于如何充分发挥应用层组播的优势，并对其进行性能优化，满足用户对组播服务质量的更高要求，这也正是本书所要研究的主要内容。

1.2 国内外相关研究现状

1.2.1 国外研究现状

2000 年 6 月，卡耐基梅隆大学的 Chu YH 在 ACM SIGMETRICS 上发表了一篇关于端系统组播[5]的论文，标志着应用层组播开始进入了热点研究。2001 年，Chu YH 又在 ACM SIGCOMM 上发表了关于在 Internet 上实现应用层组播[6]的论文。同年，Ratnasamy 在 ACM SIGCOMM 上发表了基于 Peer-to-Peer 网络的应用层组播论文 CAN Multicast[8]。Zhang SQ 也在 NOSSDAV 上发表了基于 Peer-to-Peer 网络的应用层组播的论文 Bayeux[9]。

2002 年，应用层组播的研究进入了更辉煌的一年。Suman Banerjee 在 ACM SIGCOMM 上发表了基于 NICE 应用层组播[10]的论文。Zhang B 在 IEEE INFOCOM 上发表了关于 Host Multicast[11]的论文，Castro 在 JSAC

上发表了关于 Scribe 的论文[12]，Liebeher 在 JSAC 上发表了关于 Delaunay Triangulation[13]的论文，Shi SY 也在 INFOCOM 上发表了关于应用层组播中路由问题的论文[14]。

2003 年，EL-Sayed 在 IEEE Network 上发表了一篇关于目前应用层组播方案的综述文章[15]，Suman Banerjee 也在 INFOCOM 上发表了新的论文[16]。近几年有关应用层组播的文章更是层出不穷，包括最近召开的 INFOCOM 2005 和 INFOCOM 2007 上也有与之相关的论文[94][144]。

应用层组播思想给出后的短短几年内，多个研究机构开展了应用层组播体系结构的研究项目。目前对应用层组播的研究成果主要有以下一些代表性方案。

1. 小规模的多源组播方案

小规模的多源组播方案代表是终端系统组播 ESM（end system multicast）[7] 和应用层组播体系结构 ALMI（application level multicast infrastructure）[17]，针对小规模、多数据源的情况，典型应用是视频会议系统。

终端系统组播 ESM 是 CMU（卡耐基梅隆大学）开展的一个端系统组播研究项目，是目前为止最成功的一个项目。终端系统组播给出 Narada 协议，运行完全分发协议，终端系统以自组织方式形成覆盖网络。终端系统通过适应网络中的动态性和考虑应用层的性能指标优化覆盖网的效率。Narada 给出以下目标：(1) 自组织（self-organize）。终端覆盖网的构造要以完全分发的方式，动态适应组成员变化时，需具有较强的鲁棒性。(2) 有效的覆盖网。覆盖网的构造必须使物理传输链路的冗余性最小化。(3) 自身优化的能力。终端系统要能广泛收集网络中的信息，并借此对 Mesh 网的结构进一步优化。终端系统组播的方案是：首先将组播组的成员组织成一个“网”，每个成员都维护所有组成员的列表，提高了组播组的可靠性；在 mesh 上以每个数据源为根各构造一个生成树，这样可针对每个数据源进行性能优化。其缺点是系统开销比较大，降低了系统的可扩展性，适合小规模组播组的情况。Narada 的组成员维护原理如图 1.3 所示。

ALMI 是美国华盛顿大学 St. Louis 分校计算机系从 2000 年开始进行的研究项目，给出了将应用层组播作为端系统基础服务功能的体系结构。ALMI 设计了在操作系统的套接口（socket）之上，以中间件（middleware）的形式向上层应用提供组播服务的结构，中间件实现自组织组网、组播复制和转发功能，在组播成员节点之间组成一个应用层组播网。ALMI 研究组以 Java 代码实现了中间件的原型。ALMI 的自组织协议在组成员节点之间建

试读结束：需要全本请在线购买：www.ertongbook.com