

组播策略的应用研究

刘思建, 陈元琰, 田立伟

(广西师范大学 计算机科学与信息工程学院, 广西 桂林 541004)

摘要:组播技术是一种针对点对点或多点对多点的数据传输的组通信模型,是适应未来网络发展的重要支撑技术。虽然组播技术的应用明显地提高了网络利用率,但组播网络本身存在的拥塞问题和组播路由协议在其扩展性和公平性方面存在固有的问题,严重阻碍组播技术的广泛应用。文中对组播进行了简单介绍,并着重针对组播拥塞控制技术以及组播的各种路由协议应用分别进行了对比,指出了各种算法及协议本身存在的优缺点,并指出了组播技术的研究方向。

关键词:组播树;组播协议;组播拥塞控制

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2009)08-0197-04

Application Study of Multicast Strategy

LIU Si-jian, CHEN Yuan-yan, TIAN Li-wei

(School of Computer Science and Information Engineering, Guangxi Normal University, Guilin 541004, China)

Abstract: Multicast technology is a group communication model of point-to-point or multipoint-to-multipoint data transmission, which is also an important supporting technology to adapt to the development of the future network. Although the application of multicast technology significantly improves the utilization ration of the network, the multicast networks exist the congestion problems in itself and multicast protocols exist the inherent problems in its scalability and fairness are seriously obstruct the wider application of the multicast technology. In this paper, introduce the multicast technology, then focus on the comparison of the multicast congestion control technology and the application of many existing multicast routing protocol respectively, and in the last, point out the further research trend of the multicast.

Key words: multicast tree; multicast protocol; multicast congestion control

0 引言

当今电子商务、远程教育以及娱乐(如VOD视频点播和网络游戏)等网络应用的不断出现,给网络提出了各种复杂的问题,如网络传输中的冗余、带宽不足、网络资源短缺等。而组播的出现有效地解决了以上问题。但由于组播在技术上(如组播树管理、组播协议等)、网络费用的收取等方面都存在一定问题,因此组播在Internet上的应用并不是想象的那样得到了广泛推广。

组播拥塞控制是实现网络一点对多点、多点对多点数据传输的重要前提和保障。不过组播服务的引入,也给网络传输带来了许多潜在的问题,其中怎样解决组播网络中的拥塞是最核心的问题。组播技术的应

用如果在网络拥塞时不能对其进行正确地响应,这将会给Internet带来比单播应用更为严重的影响。因为:(1)一个组播数据流可能沿着其组播分布树广泛分布于整个组播组网络;(2)组播组内成员具有其本身固有的异构性(软、硬件配置的不同);(3)组播发送端在一个大的网络中大多必须处理大量的反馈控制信息。

1 组播介绍

1.1 组播基础知识综述

单播、广播和组播是TCP/IP协议传送数据的三种通信类型,为了大家能更好地了解组播,先来简单介绍一下这三种传送方式。

单播(Unicast):在发送者和每一个接收者之间都需要建立单独的信息转发路径,如图1所示。这也就决定了只有当接收方数量较少时才能保证网络的QoS,而当有大量接收端渴望获得同一份数据时将导致发送端负载增加、时延加大,以致网络发生拥塞,甚至会导致拥塞崩溃现象的发生。

广播(Broadcast):是指发送端在整个网络内广播

收稿日期:2008-11-26;修回日期:2009-02-17

基金项目:广西自然科学基金资助项目(0728099)

作者简介:刘思建(1980-),男,硕士研究生,研究方向为计算机网络、计算机网络拥塞控制;陈元琰,教授,博士,研究方向为计算机网络、计算机图形学等。

数据包,这样整个网络内的主机都能接收到数据包,有利于网络资源的充分利用,但是网络内不需要该数据包的接收端也会接收到数据包,如图 2 所示。这样反而影响了网络资源的有效利用,而且网络越大带宽资源浪费就厉害。所以在当今网络规模越来越大的今天,广播不可能得到广泛使用,而且如果应用不当甚至会引起广播风暴。

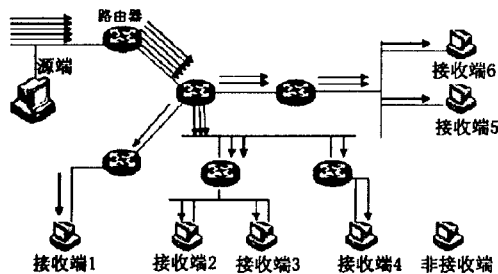


图 1 单播

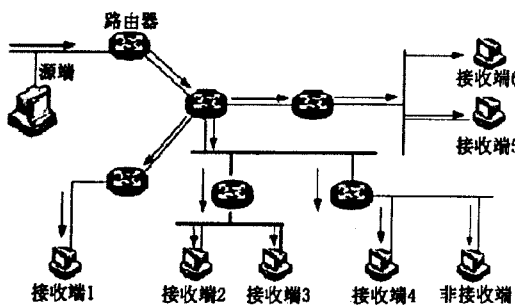


图 2 广播

组播 (Multicast): 组播是一种实现了从一点可以同时向多点发送数据传输方式,如图 3 所示。利用组播可以很大程度地节省带宽,因为在网络中无论有多少接收方,发送端只提供一份数据到组播组中的接收方。组播是相对于单播而言的,如果组播组中的接收方只有一个成员,那么组播也可以看作是单播的形式了。图 1、图 2、图 3 简单描绘了单播、广播、组播数据传输方式的不同。组播和单播传输的数据流只能是组内的成员才能接收。而广播组中的非接收端也能接收到组内的数据流。

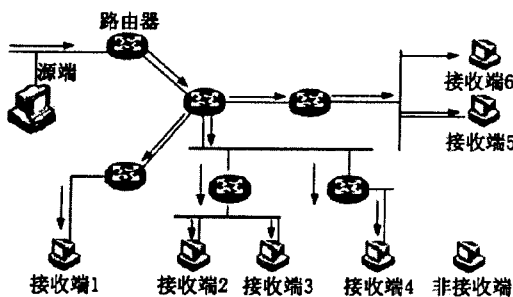


图 3 组播

组播的性质决定了它是能适应当前高传输量多媒体网络的最佳选择。但组播也存在本身固有的问题:

为了让组播组内的接收端能快速地接收到数据,组播使用了面向无连接的 UDP (User Datagram Protocol) 协议传输数据。由于 UDP 协议不能对传输的数据进行检验,也缺少对一定的流量控制和拥塞控制机制,这使得组播的可靠性得不到保证^[1]。组播传输的关键是建立一棵组播分布树,而网络中的任何终端都可以随时加入组播树,这就使得组播组的成员不能得到很好的控制,再加上组播对组内的数目没有限制,所以组播的安全性也受到了限制。所以组播在 ISP (Internet Service Provider) 中不能得到广泛应用。

可扩展性和 TCP 友好^[1]是研究组播拥塞控制的两个重要的评价目标。可扩展性是指随着组规模的增大,拥塞控制协议不会造成组播性能的下降。但是随着组播组规模的增大,不可避免地会带来队列延迟及 LPM^[2]问题,这都影响了组播的可扩展性。TCP 友好则要求组播流量和 TCP 流量能公平地竞争带宽,即保证 Internet 中原有的 TCP 流量和组播流量的公平性。这两个评价目标在一定程度上是对立的,组播拥塞控制协议需要根据实际需求在两者间作出权衡^[3,4]。

1.2 组播树

一个组播组的成员分布存在广泛性的特点,所以要在组播组内有效地传输数据选择合适的组播路由就显的格外重要。下面简单介绍几种分布树策略:

洪泛法 (Flooding): 其原理是网络中间节点收到一个分组后不是马上转发,而是从分组带的信息检查是不是原始分组。如果是,则复制数据包然后发送到本机的所有接口上,确保能够到达所有组播组内的目的主机。如果不是,则丢弃。它的优点是,只要网络中还有一个信道是连通的,就能将信息发达到目的网络,有效地增加了传输的可靠性。但是当接收端主机数量很大时,就会因在发送端大量复制数据包,而占用相当大的内存。而且有可能导致组播组内主机多次接收到同一个数据包。

生成树 (Spanning Tree, ST): ST 有效地解决了上面算法在大的组播组中重复接收同一个数据包的问题。ST 策略充分利用了树的结构特点,避免了在组播组内形成回路。而当组播组内成员很少时,生成树算法也要覆盖组播组内的所有节点,这就造成了资源的浪费。逆向路径转发 (Reverse Path Forward, RPF) 为每个组播源建立一个单独的树,并根据路由表查找逆向最短路由,利用最短路径的方法转发分组,可以使网络得到较好的利用。特别是有利于流量大、时延长的媒体应用,但是没有很好地解决路由器重复接收一个分组的问题。另外因为 RPF 要为每个分组建树,所以开销也相对较大。

共享树(RPT):是指组播组选定一个共用根 RT 建立组播树,RPT 只是把从源点发出的第一个分组转发到所有的网络结点,虽然只是执行了一次扩散传播,但却增加了路由器的负担,因为路由器要记录每个组及源的状态。典型的共享树有:

(1) Steiner 树,是一种延迟受限的树,理论上 Steiner 树是使用较少的资源,以减少拥塞的发生,但是由于是根据组中成员关系而动态地建立树,因此稳定性严重降低。经典的算法有 KMB、MPH 等算法。

(2) 有核树(Core Based Trees, CBT)^[5],是由根到所有组播组成员的最短路径合并而成的树,而且处理过程只占用少量的内存和带宽资源,有效地解决了 RPF 树和剪枝算法存在的缺点。有核树算法将组播的扩散范围限制在所有的接收者范围内,第一个分组也不需要在全网扩散,且只需记录组的状态信息,而不是源站点和组配对的状态信息。

2 组播协议

由于组播能同时发送大量的数据而不会造成网络负载的增大,因此受到了网卡生产商的欢迎。经过调查,目前生产的以太网卡大多都支持组播,这就说明了 IP 组播技术的应用环境已基本具备。但组播策略并没有在 Internet 上得到广泛应用,主要是因为还没有使组播路由选择协议能在各个中间节点上互相操作,以保证组播信息准确到达接收端网络。而当前网络的应用对组播的要求是多样化的(如:网络媒体希望变化尽量平滑,以保证连续性,对可靠性不做要求;而文件分发等就要求要对数据的可靠性做出保证),这导致了组播协议也必须是多样化的。这给组播路由协议的设计造成了很大的困难。IETF(Internet Engineering Task Force)和 IRTF 已经取得共识^[6]:一种组播拥塞控制协议不可能满足所有的需求,必须针对不同的应用设计不同的协议。

2.1 组管理协议

Internet Group Multicast Protocol(IGMP)是实现 IP 组播的基础组管理协议。它主要功能是,当一台主机请求加入组播组时,路由器就使用 IGMP 协议查询本地子网中是不是有属于某个组播组的主机。如果存在,就在主机和路由器之间建立一条通信信道。当主机退出时,自动删除信道。而要在现实网络中实现组播数据的正确传送,必须在路由器上建立相应地支持组播数据包的路由协议。

2.2 组播路由协议

组播路由协议^[7]可分为三类:密集模式协议、稀疏模式协议和链路状态协议。

2.2.1 密集模式协议

距离向量组播路由协议(Distance Vector Multicast Routing Protocol,DVMRP)^[8],在路由信息协议(Routing Information Protocol,RIP)之上的 DVMRP 是第一个在组播传输中得到广泛应用的协议。其中 RIP 是施乐公司 20 世纪 80 年代推出的小规模的内部网关协议(Interior Gateway Protocol,IGP)。DVMRP 采用面向无连接的机制定时向组播组中的主机发送数据包,这样,刚刚加入组播组的主机也能保证能收到数据组。但这种方式过多浪费了网络带宽;由于需要定时发送,这也就增加了路由器本身的负载。

协议无关组播-密集模式(Protocol Independent Multicast-Dense Mode,PIM-DM),PIM-DM 路由器收到第一个分组时,如果是原始分组就向所有接口转发组播数据;如果不是,刚丢弃,同时向上面的节点发送消息避免资源浪费。由以上特点知,PIM-DM 适用于小型的高速网络。如果是一个大的网络,刚在扩散时会耗费大量带宽。

2.2.2 稀疏模式协议

协议无关组播稀疏模式(Protocol Independent Multicast-Sparse Mode,PIM-SM)^[9],PIM-SM 通过组播共享树来在组播组内分发数据,这样就不会像 PIM-DM 那样保存过多的组播状态。它适用于网络中组播组内主机数目较少的环境。

有核树组播路由协议(Core Based Trees, CBT)^[10],CBT 是利用一个核心路由器而建立的双向共享树,它允许组播信息向两个方向流动,这就有效增强了组播的可扩展性。网络主机想加入组播组,都必须向核心路由器申请,而且信息也要经过它转发,这样就增加了核心路由器负载过重的可能性。

2.2.3 链路状态协议

链路状态协议(Link-State Routing Protocol)通过扩散链路使用 SPT 向网络中发送组播信息,可以迅速识别组播组内成员信息,并建立对应的最短路径树。因此可以快速响应网络发生的变化。

开放式组播最短路径优先协议(Multicast Open Shortest Path First,MOSPF)^[11],MOSPF 是基于单播协议 OSPF 的一种链路状态的路由协议。MOSPF 路由器利用 Dijkstra 算法构造最短路径树(SPT),明显提高了链路的利用率和网络变化的响应能力,但同时也耗费了大量资源。因此 MOSPF 比较适用网络变化比较小的专用网络。

3 组播拥塞控制探讨

拥塞是指网络长时间工作在一个持续过载的状

态。从图 4 中可以清晰地看出,在刚开始时吞吐量(Throughput)和负载(Load)基本成线性关系;而当 Load 超过膝点(Knee)后,Throughput 的增长开始变的缓慢;当 Load 超过崖点(Cliff)之后,Throughput 开始急剧下降。如果长时间处在这种状态就会导致拥塞现象的发生,再严重就可能产生拥塞崩溃(Congestion Collapse)^[12]现象。而拥塞控制就是在端节点或中间节点采用一定的策略或机制,让网络尽可能地工作在接近 Cliff 的区域。拥塞控制的直接目的是让 Internet 尽可能在重载的情况下也能提供良好的服务,保证网络的鲁棒性。

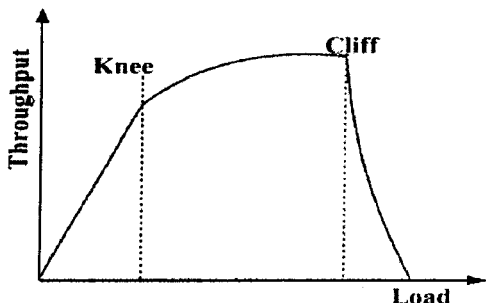


图 4 网络拥塞

3.1 组播拥塞控制遇到的问题

虽然与单播技术相比组播技术有效地减轻了网络负担,但同时也产生了许多潜在的问题,例如组播发送方要比单播发送方多处理很多反馈报文,而大量的反馈信息会导致组播性能下降,如果处理不当,会造成反馈爆炸;组播靠建议转发树来传输数据到接收端,但组播并不提供流量控制,所以网络资源可能会大量消耗,从而导致拥塞。组播技术有效地实现了一点对多点的通信,但接收端的主机固有的异构性要求组播拥塞控制策略能适应复杂的网络,所以组播拥塞控制已经成为一个迫切要解决的问题。

3.2 组播拥塞控制算法分析

20 世纪 90 年代后期组播拥塞控制算法研究开始兴起,而且大部分算法都是针对某些特定问题提出的。

3.2.1 基于窗口的组播拥塞控制

基于窗口的组播控制机制与 TCP 端系统的窗口机制类似,是通过调整拥塞窗口来控制分组的数量。发送端每发送一个数据分组,窗口就相应减少一个分组的空间,如果窗口中有空闲位置时可以继续发送数据分组。直到收到一个分组的确认后才释放占用的部分。窗口的大小随着网络拥塞的情况变化。

基于窗口的组播拥塞控制中,发送端窗口有单窗口和多窗口之分。基于单窗口:也就是用一个窗口来维护整个组播组内的接收方信息。实现方法简单,而且可扩展性好。但数据发送速率会因组播组内部分成

员(处理速率慢的主机)不能及时返回确认信息而影响整个组播组的吞吐量。基于多窗口:数据发送端为组播组中的每个接收方建立一个独立的窗口,这样以来就有效地节约了网络资源。但随着组播组中主机数越多,发送端维护窗口的数目也越大,这样就严重地影响了组播的可扩展性。而且如果处理不当就会引发反馈风暴。

3.2.2 基于速率的组播拥塞控制算法

基于速率的组播拥塞控制算法可以分为单速率组播控制和多速率组播拥塞控制,它们是根据网络中的拥塞控制信息动态的调整发送速率 TR(Transmission Rate)的。

基于单速率:是指整个数据发送过程中,发送方只用一种速率在发送数据。这种方法实现比较简单,有较好的可扩展性,但受到了组播组内接收端部分主机(处理速度慢的主机)的影响。随着网络越来越复杂,这种缺点也就越来越明显。

基于多速率:与单速率不同,多速率算法中发送端可以采用多种速率发送数据。这样就可以根据网络的具体情况而灵活地改变速率,即当接收方处理速率快时就用相对快的速率,反之就用较小的速率发送数据。但是随着接收端不同类型主机的增加,发送端任务的复杂度也相对增大,大量的反馈信息也增加了反馈爆炸的危险,影响了组播拥塞控制的公平性。其中分层组播拥塞控制算法是一种典型的基于多速率的拥塞控制策略,分层组播对要传输的分组给出一定的编码,并把接收端要接收的分组分为一个层次,其余提高 QoS 的分组分若干个层。这样接收端除了一定要接收第一个层次以外,可以根据自己的接收能力有选择地利用其它层,也就是接收端接收的层数越多,质量就越高。分层组播策略有效解决了异构网络中视频传输的异常问题,提高了网络利用率。因此在网络(远程教育、远程会诊等)中得到广泛应用,但是却影响了组播组中接收方的公平性。

以上算法在可扩展性和公平性上都存在本身的问题,而怎么解决这些问题,是当今一个充满活力的研究领域。

4 结束语

从组播的定义出发,总结了组播中的协议以及组播中的拥塞控制算法,指出了各个策略的优缺点。组播技术的应用有效地节省了带宽,减轻了网络负载,因此一定有广阔的发展前景。组播技术之所以现在还没能在 Internet(特别是 P2P 网络)中得到更广泛的利用,

(下转第 204 页)

和更新操作。

3 结束语

Ajax 和 CORBA 技术都是当前流行的设计技术。Ajax 作为一种基于 JavaScript 和 XML 的客户端技术,不仅增进了客户端体验,同时也改进了软件体系设计特别是对服务器的调用设计。CORBA 作为开放、标准的规范体系使得各种中间件技术可以相互支持,特别是 CORBA 的语言无关性使得本系统中 EJB 和 C++ 实现的调用成为可能。

文中结合 Ajax 和 CORBA 技术设计和实现了基于 J2EE 的分布式订单管理系统。该系统利用 Ajax 和 CORBA 的优势把表示逻辑和业务逻辑向分离,同时利用 CORBA 的语言无关性和 RMI-IIOP 通信机制,实现了 Java 调用和 C++ 实现的编码透明性,有利于开发者分工和协同开发,提高了基于 J2EE 的分布式系统的开发效率。

参考文献:

- [1] OASIS SOA Reference Model TC. Reference model for Service Oriented Architecture 1.0[EB/OL]. 2006. <http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf>.
- [2] Stal M. Using architectural patterns and blueprints for service

-oriented architecture[J]. Software, IEEE, 2006, 23(2): 54-61.

- [3] 童鑫,李军义. 面向 SOA 的企业服务总线研究与实现[J]. 计算机应用, 2008, 28(3): 819-822.
- [4] 彭武良,周丽,王雷. 企业应用集成技术综述[J]. 计算机应用研究, 2007, 24(9): 12-15.
- [5] 蔡萍,华庆一. 基于 SOA/EDA 电信企业应用集成技术研究[J]. 计算机技术与发展, 2007, 17(11): 560-561.
- [6] 赵永屹,宿红毅,胡韶辉. 基于 AJAX 与 J2EE 的新型 Web 应用的设计与实现[J]. 计算机工程与设计, 2007, 28(1): 189-192.
- [7] 田福生,张燕平. 用 Ajax 技术实现 B/S 模式下客户端间信息交互[J]. 计算机技术与发展, 2007, 17(10): 38-39.
- [8] 代霞,黄劲松. 基于 CORBA 综合网络配置管理的设计与实现[J]. 计算机技术与发展, 2008, 18(2): 91-93.
- [9] Oh Joo-Yong, Park Jun-Ho, Jung Gi-Hoon, et al. CORBA based core middleware architecture supporting seamless interoperability between standard home network middlewares[J]. IEEE Transactions on Consumer Electronics, 2003, 49(3): 581-586.
- [10] 韦兆文,刘波. 利用 RMI-IIOP 实现 EJB 容器之间的通信[J]. 计算机应用, 2003, 23(11): 128-133.
- [11] 夏丹. EJB/CORBA 集成应用研究与实现[D]. 武汉: 武汉理工大学, 2006.
- [12] Eichorn J. 深入理解 Ajax: 基于 JavaScript 的 RIA 开发[M]. 北京: 人民邮电出版社, 2007.

(上接第 200 页)

主要是因为对路由器的要求太高。面对这种情况近几年出现了应用层组播策略,这种策略将组播策略的实现转移到了各个终端,从而避免了改变网络配置,同时也减轻了路由器的负担。而怎样使应用层组播策略、IP 层组播策略和原始 TCP 流量控制合作更密切,是目前非常值得研究的一个问题。

参考文献:

- [1] Forouzan B A. TCP/IP 协议族[M]. 北京: 清华大学出版社, 2006: 372-375.
- [2] Bhattacharyya S, Towsleu D, Kurose J. The loss path multiplicity problem in multicast congestion control[C]//INFOCOM apos 99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. New York, NY, USA: [s. n.], 1999: 856-863.
- [3] 苏晓丽,郑明春,孟强. TCP 友好多播拥塞控制算法研究[J]. 计算机工程与科学, 2005, 25(4): 26-29.
- [4] 潘国庆,李陶深. 一种基于策略函数的应用层组播路由算

法[J]. 计算机技术与发展, 2008, 18(5): 138-140.

- [5] Ballardie A. Core Based Trees(CBT) Multicast Routing Architecture[S]. RFC 2201, 1997.
- [6] 苏晓丽,郑明春,孟强. 多播拥塞控制研究进展[J]. 通信学报, 2003, 24(5): 94-104.
- [7] 程传庆. IP 组播组管理协议及其在二层的实现[J]. 信息技术, 2003, 27(7): 50-52.
- [8] Waitzman D, Partridge C, Deering S. Distance vector multicast routing protocol[S]. RFC 1075, 1988.
- [9] Estrin D, Farinacci D, Helmy A, et al. Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification[S]. RFC 2362, 1998.
- [10] Ballardie A, Cain B, Zhang Z. Core Based Trees (CBT version 2) Multicast Routing: Protocol Specification[S]. RFC 2189, 1997.
- [11] Moy J. Multicast Extensions to OSPF[S]. RFC 1584, 1994.
- [12] Nagle J. Congestion Control in TCP/IP Internetworks[J]. Communications Review, 1984, 14(4): 11-17.