

计算机网络群组通信研究*

白成杰(山东师范大学) 白成林**(山东聊城师范学院)

【摘要】 计算机网络的发展使通过网络相连的计算机集合成为一个相互协调的统一整体,并为人类群体成员间的协作配合提供了新的工具。目前计算机网络中有关群组通信的研究十分活跃,是计算机网络研究的关键技术之一。群组通信为分布式文件系统、分布式数据库系统、并行计算、容错系统和计算机协同工作系统等各种分布式应用系统提供所需的通信服务。文中给出了群组通信的体系结构,并对群组通信中的点到多点通信、群同步等问题进行了研究。

【关键词】 群组通信 多媒体通信 多目发送 群同步

The Study of Group Communication

Bai Chengjie

(Shandong Teachers University)

Bai Chenglin

(Liaocheng Teachers College)

Abstract: With the development of computer network, the set of computers connected with network is becoming a cooperative whole entity and providing a new kind of tools to coordinate human group. Currently, Group communication is an active area and key technique in the field of computer network. The object of group communication is to provide special communication service for distributed application systems, such as distributed file system, distributed database system, parallel computing, fault tolerance system and CSCW. This paper proposed an architecture for group communication, and discussed the multicast and group synchronization of the group communication.

Keywords: Group communication Multimedia communication Multicast Group synchronization

1 前言

随着通信技术、计算机技术和多媒体技术的高速发展,特别是随着具有高速和有效的多媒体通信能力的宽带综合业务数字网 B-ISDN 的逐步实用化,分布式多媒体多用户通信系统变得越来越重要。这些分布式多媒体多用户通信系统都要求利用计算机系统综合控制多媒体信息的采集、存储、通信、操作以及表现等一系列过程,其中群组通信是其关键技术之一。

群组通信有多种层次,从广义的角度来讲包括各种层次的通信群组内各种通信实体间的通信。如应用系统中各进程间的信息交流、操作系统层次各进程间的协调和数据传输层次的信息传送。在应用系统和操作系统层次上讨论的进程可能在一台机器中,也可能在不同的机器上,它们之间的群组通信主要体现为系统的控制策略。这里讨论的群组通信是对传统的计算机网络概念的发展,需要引入一些原来没有或不重要的新特征。这也是群组通信研究的出发点和目标。

2 群组通信的体系结构

群组通信的体系结构,可考虑划分为四个层次,如图 1 所示。

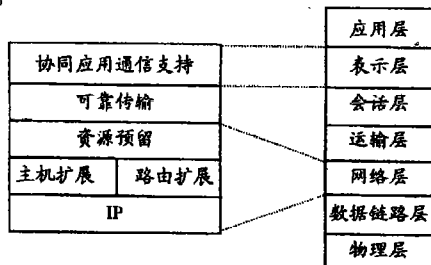


图 1 群组通信体系结构

(1) 主机及路由扩展子层:主机扩展的目的是使主机具有多目标分组发送(multicast)和接收功能,其内容包括:群组地址管理、群组成员管理、多目标分组的发送和接收。

通信群组使用 D 类 IP 地址作为标识。群组地址可分成永久组地址和临时组地址。永久组地址是由管理机构预先分配的,它所对应的永久通信群组的成员没有限制,甚至可为 0 个成员。临时组地址指永久组地址以外的群组地址,可动态地分配给只在有成员时才存在的临时通信群组。

对应用程序使用的临时群组地址的动态分配方法没有统一的协议。目前最常用的方法是由应用程序随机分配。这种方法有可能因出现群组地址冲突而重新分配,直到找到一个空闲的群组地址时才完成分配。其它可能的方法还有:由上层的应用协议分配和向网络中相互联系的群组地址分配服务器申请临时组地址。这方面的研究还在进行中,由群组地址分配服务器分配地址的方法是较理想的方案。

数据的收发,群组通信涉及的机器可分成发送方和接收群组成员。只有一台机器加入一个接收群组后和退出该群组前的这一段时间内,才能接收发送到该接收群组的多目标分组,而对多目标分组发送方,并不要求它一定是相应接收群组成员。

发送多目标分组与发送一般 IP 分组有一定区别,主要区别在于:目标是 D 类 IP 地址;需指定发送分组的生存期(TTL, Time - To - Live);当一台机器同时有多个网络接口时,需指定多发送目标分组的网络接口;当发送方同时是该群体成员时,应向本机高层回送该多目标分组的一个副本。

接收多目标分组与接收一般 IP 分组基本类似。为了能收到多目标分组,高层实体必须要求 IP 层加入相应接收群组,IP 层应扩展相应的加入组和退出组操作。

路由扩展,即扩展群组通信的路由控制,确定从多目标分组的发送方到接收群体各成员的多目标分组发送树。为了提高分发效率,路由算法必须在保证接收群组所有成员都收到发送到该群组的多目标分组的条件下,尽量减少网络资源占用量和多目标分组的发送范围。

多目标分组发送范围的限制方法有修剪多目标分组分发树和限制多目标分组生存期两种。修剪多目标分组分发树,是通过路由算法来尽量使多目标分组只向接收群组成员分发,从而提高网络效率。限制多目标分组生存期,是通过设置每个转发端口(downstream interface)和转发节点(downstream neighbor)的 TTL 值,限制小生存期分组不向远的接收群组成员转发,从而避免浪费网络带宽。

现有的主要多目标路由算法包括:扩散(Flooding)算法、分发树(Spanning Tree)算法、逆向路径广播(RPB, Reverse Path Broadcasting)算法、修剪的逆向路径广播(TRPB, Truncated Reverse Path Broadcasting)算法、逆向路径多目标发送(RPM, Reverse Path Multicast)算法、核心分发树(BT, Core - Based Tree)算法。

目前主要的群组通信路由协议有:距离矢量多目标路由协议(DVMRP, Distance Vector Multicast Routing Protocol)、开放的最短路径优先多目标路由协议(MOSPF, Multicast extension to the Open Shortest Path First routing protocol)、与下层协议独立的多目标路由(PIM, Protocol - Independent Multicast)协议和 CBT 协议等。

上面几种群组通信路由算法实现的群组通信,还存在着如何提高转发效率、尽量减少人工参与、各多目标路由协议的互连问题。

(2) 资源预留子层:资源预留的目的是为了保证数据传送所需要的网络资源。对于群组通信,要在从多目标分组的发送方到接收群组所有成员的分发树经过的各主机和路由器上预留相应资源,保证多目标分组的有效传送。已经提出的资源预留协议(RSVP, Resource reSer Vation

Protocol)是主机及路由扩展子层上的一个子层,提供群组通信所需的服务质量,用于高层的数据传送。依据主机及路由扩展子层计算的多目标分组分发树,RSVP 通过网络把有关资源预留的 QoS 要求传送到分发树经过的各路由器节点及各收发方,在每个节点为多目标分组申请资源预留。

当从高速网络向低速网络转发多目标分组时,带宽无法满足最高要求。为了保证实时性,只能丢弃部分次要信息。信息过滤的目的就是为了在网络带宽限制条件下,针对信息特点选择最优的信息丢弃部分次要信息,保留尽可能多的有效信息,达到优化带宽利用、协调接收群组的不同环境、优化资源分配等。

信息过滤机制的研究目前集中在音频、视频等实时多媒体信息的压缩算法。已提出的主要信息过滤方法有:丢弃信息片断,如丢弃视频流中的一帧信息;丢弃信息流中的次要特征,如丢弃视频信息流中的彩色信息,仅保留灰度信息;信息压缩方式变换,如低压缩比方式到高压缩比方式的变换。

(3) 可靠传输子层:通过群组通信各方的缓冲机制,实现可靠的和有序的群组通信。这里的有序不仅要考虑同一发送方数据在各接收方的接收顺序,在某些应用中还要考虑多个发送方到同一接收群组中各成员的总体接收顺序。

这一子层采用的可靠多目标发送协议(Reliable Multicast Protocol, RMP)对应于运输层和会话层协议,提供多发送方到接收群体各成员的可靠、容错、有序的信息传送。RMP 的错误恢复对于应用程序是透明的。

(4) 协同应用通信支持层:这一子层针对各类信息的特点,对各类信息的群组通信给出相应的分组格式和各字段意义的约定,反映具体信息传送的需求。实时传输协议(RTP, Transport Protocol for Real - Time Application)就是针对实时视频和音频信息传输给出了一组分组格式,分组中不仅有压缩的视频和音频数据,还包括压缩格式、发送时间、分组序号、发送方信息(发送方标识、机器名、地理位置、电子邮件地址等)。这些约定对于不同类型系统间的群组通信是重要的。

3 群组通信的 multicast 结构模型与实现

multicast(多目标发送)是群组通信最基本的操作,其服务质量直接关系到整个群组通信服务。为了提供可靠的 multicast 服务,必须采用应答机制,传输应答的方式有很多种,如独立发送(每一个叶结点都建立一条到根结点的连接,并利用这个连接发送应答)、基于树的应答传输(所有叶结点形成一棵树, multicast 连接的根结点作为这棵树的根结点)。当叶结点数目很大时,独立发送显然不适合,因为根必须建立 n 条连接(假定有 n 个叶结点),将消耗大量的资源,但这种结构易于维护和处理,基于树的方法有比较好的可伸缩性和性能,但当有 multicast 成员发生改变时(加入或离开),比较难于维护。高级通信层的主要功能是提供可靠的群组通信操作,为了实现这些群组通信操作,我们提出了一种称为混合树的群组通信结构。

3.1 基本的群组通信结构模型

我们利用混合树(Hybrid - tree)结构来实现我们的群组通信操作。其结构如图 2 所示。

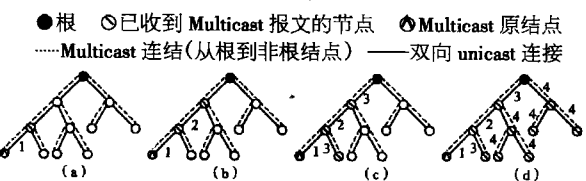


图 2 基于混合树的 multicast

(下转第 28 页)

* 本文于 1999 年 11 月 4 日定稿。

** 本文作者白成杰系山东师范大学现代教育技术研究所讲师,白成林系山东聊城师范学院通信工程系副教授。

snoop> finger johndoe@ @ @ @ @ ... @ @ victim.com
这会导致 victim.com 循环地“Finger”自己,如果 victim.com 循环“finger”足够多次,则内存、交换空间(swap space)和硬盘空间最终被充满,从而导致机器崩溃。阻止这类攻击的明显的方法就是让 finger 服务器失效。

5.2 报文泛滥(message flooding)

当在一个用户希望降低网络系统工作效率以阻止其处理正常工作时,就会发生 Message flooding。Message flooding 攻击经常都是直接指向认证服务器的。受到攻击的认证服务器会变得不堪重负,无法响应来自其客户的请求。这就可能让攻击者的机器轻易地注册到一特权帐户,伪装成合法的认证服务器(用假冒信息应答认证请求)。Message flooding 攻击经常是用成千上万的 ICMP echo 请求来袭击服务器实现的。

Message flooding 也可用来堵塞网络通道。一个简单的方法就是在两个著名的端口(用于提供标准 TCP/UDP 服务的端口)之间产生无限的序列(如发送更换分组)。当然,这种特定的 Message flooding 可以通过关闭可能用来产生无限序列的服务器来防范,如使用能阻止选定端口间重定向数据的过滤程序。

5.3 拥塞(clogging)

TCP 建立连接的三方握手的实现,可以被滥用而引发一个非常危险的 DoS 攻击——TCP SYN Flooding 攻击(或称为 SYN 攻击)。

攻击者通过伪造从不同 IP 地址源发出的大量 TCP 连

接请求到目标主机,却不完成 TCP 连接,即熟知的半开连接,从而导致目标主机的连接队列被占满而无法响应正常的 TCP 连接。

正常的 TCP 连接开始后,目标主机收到 TCP SYN 分组,发回 SYN ACK 确认报文,然后等待源主机对 SYN ACK 报文的确认报文。只有当 SYN ACK 的 ACK 到达后, TCP 连接才能完成。而正常情况下,目标主机在有限大小的连接队列中记录连接等待完成的状态, SYN/ACK 的 ACK 会在几毫秒内到达且很快被清空。但是如果攻击者冒充不同 IP 地址向目标主机发送大量的连接请求,目标主机不可能得到 SYN/ACK 的 ACK,从而无法完成连接,甚至超时才能将连接请求从队列中清除,这样目标主机的连接队列会很快被不能完成的 TCP 连接请求占满,导致正常服务请求得不到响应。

如果希望让路由器拒绝任何来自一指定 IP 地址的数据分组来阻止正在进行的攻击,显然不是最好的办法。因为通常被 Spoof 过的数据分组的源地址是由攻击者的软件随机选择的。一般可以选择安装操作系统补丁、增加 TCP 连接队列的大小,减少等待 SYN/ACK 的 ACK 超时时间等方法。

6 结束语

Internet 上的安全攻击与防范将会是一场持久战,而且这场战争必将随着 Internet 的繁荣和安全技术的发展而愈来愈激烈。目前还不能从根本上防范各类安全攻击。我们认为只有保持高度警惕,采取相应的积极手段,保

证使网络受攻击和受损的可能性降到最低程度。为此,首先是要加强安全性教育,让系统管理员和网络管理员能及时了解 TCP/IP 的安全缺陷和操作系统的漏洞;其次是尽可能采用防火墙机制,限制外部网络对内部网络的访问;再就是使用入侵检测系统,对可能的攻击迹象进行实时监测,并进行审记;最后就是应当对已经公开的服务进行安全性分析,及时修补潜在的漏洞,变被动为主动。希望下一代 Internet 的协议 IPv6 能给 Internet 的安全带来福音。

参考文献

- [1]周明天,汪文勇,TCP/IP 网络原理与技术,清华大学出版社,1993.12
- [2]Andrew S. Tanenbaum, Computer Networks(Third Edition), 清华大学出版社,1998.9
- [3]Othmar Kvas, Internet Security—Risk Analysis, Strategies and Firewalls, 王震,铁满霞,陈希南译,中国水利出版社,1998.6
- [4]王锐,陈靛,靳若明,周刚,网络最高安全技术指南,机械工业出版社,1993.5
- [5]Marco de Vivo, Gabriela O. de Vivo, Germinal Isern, Internet Security Attacks at the Basic levels, Operating System Reviews, Vol32, No02, p4—15, 1998.4
- [6]王贵驹,张莺,张欢,DoS 安全攻击的原理分析与对策研究,中国计算机协会信息保密专业委员会论文集,第八卷,1998.10

(上接第25页)

在图2中,所有结点形成一棵树(一般为二叉树),非根结点同它的父结点之间都是一条双向 unicast 连接,非叶结点同它的子结点之间也有一条双向 unicast 连接。从根结点到所有非根结点之间还存在一条硬件提供的 multicast 连接,这条一对多连接同时形成另一棵树,为了在由 unicast 连接形成的树上同时表示这棵树,用图2中的虚线来表示它,但从逻辑上看,这些非根结点均为硬件提供的 multicast 连接的根结点的叶结点。我们称这种树为混合树(Hybrid-tree)。这样,这棵混合树就形成了一条多点到多点的连接,任何一个结点均可给其它所有结点发送报文。当有结点加入或退出时,树结构发生改变,一个结点可能有两个以上的子结点,点到多点连接的叶结点的数目也发生改变。

当有结点失败或新结点加入或结点退出时树结构必须发生改变,改变方法如下:

当有非根非叶结点失败或退出时,该结点的所有子结点的父结点改为失败或退出结点的父结点,这一改变要求受影响的结点拆除原有的 unicast 连接,同新的父结点建立新的 unicast 连接,而从根到非根结点间的 multicast 连接只需从其叶结点中去掉失败或退出结点即可。最坏结果是除根结点外,其余结点都是叶结点。

当叶结点失败或退出时,失败或退出的叶结点的父结点只需拆除同它的 unicast 连接即可。multicast 连接的改变同非根非叶结点时一样,当有新结点加入时,如果先前有退出或失败的结点,则该新结点取代退出或失败的结点在原树中的位置。如果没有这样的结点则将新结点作为叶结点。对于 multicast 连接,根将新的结点加入到这个 multicast 连接的叶结点中去。

3.2 群组通信操作的实现

首先讨论可靠的 multicast 的实现,在混合树中,如果非根结点在它的 unicast 链路上收到报文,则它在所有除收到该报文的链路外的链路上转发报文。如果它是在 multicast 链路上收到报文,则不转发。如果根结点收到一个报文,则用 multicast 链路转发报文。在非 multicast 链路上转发报文是因为:其一,将要发送的报文传至根结点;其二,提高报文传输的可靠性,如果一个结点没有从 multicast 链路上收到报文则它还有机会从 unicast 上收到报文;其三,一些结点可以在更短的时间内收到报文,如与报文源同在根结点的同一侧的子树上的一些结点。因为结点既可从 unicast 连接上接收报文也可从 multicast 连接接收报文,所以一些结点可能会收到一些重复的报文,根据报文序号可以保证结点不会重复接收。尽管这样会消耗一些额外的网络带宽,但它增加了传输的可靠性,降低了时延。我

们以最左边的叶结点作为源结点来说明在这样的结构中是如何实现 multicast 通信的。在这里,我们假定一个结点一次只能发送一个报文。如图2(a),(b),(c)所示,报文经过了三步(logN)到达根结点,根结点收到报文后利用 multicast 连接在一个时间步内将报文送至所有非根结点。图中链路边上的数字代表着报文步数,在本例中,完成 multicast 需要4个报文步。

应答的传送也是通过树来完成的,即执行一个 multi-recv 操作,结果存放在报文源结点。一个结点只有在收到它的所有输出链路返回的应答后才向它的输入链路发送应答。注意,报文源不同,结点的输入输出链路是不同的。为了解决应答风暴问题,将输出链路返回的应答合并后再向它的输入链路发送应答,这样对一次 multicast 通信每个结点一般只向它的输入链路传送一个应答报文,提高了算法的可伸缩性。

就效率而言,在最坏情况下(深度最大的叶结点发送 multicast),能够在(logN+1)个报文步完成 multicast 操作。在资源需求方面,建立混合树对通信资源的需求量不大((N-1)条 unicast 连接和1条点到多点 multicast 连接),因而实现起来比较可行。

4 群同步

群同步是分布式多媒体群组通信中特有的同步问题,群同步的实质是保持群中各个用户之间相互同步,群中的各个用户同时采集媒体单元,然后实时发送给群中的其它用户,各个用户接收到同一个时刻的各个用户媒体单元后,同时合成播放,而不受地域、传输时间等因素的影响。

由于群同步技术是分布式多媒体通信的关键技术之一,所以许多研究人员从各个方面对此做了广泛的研究,并且已经取得了一定的进展。其中文献[3]提出的群同步算法虽然具有很高的精确性,但是对网络中间节点(交换机)提出了很高的要求,要求实时提供各个中间节点对传输媒体单元的延迟时间。而文献[4]则是基于统计的方法,通过确定每个用户分组的平均到达时间,确定如何将各个用户的分组数据混合在一起播出。这种方法十分简单,但适用范围十分有限,只适用于比较简单的多媒体会议。为此我们提出了基于群同步参考点的群同步算法。

考虑到每个用户通常都以固定的时钟周期中的其它用户发送连续媒体信息,我们定义这些发送连续媒体信息的时间点为该用户的同步参考点。群中各个用户的参考点具有周期性,称为各用户的同步参考点周期。为了简单起见。规定群中每个用户具有相同的同步参考点周期,为实现群同步,选择某一个用户(称为基准参考点用户)的参考

点为基准,并将基准参考点用户的参考点时间信息利用连续媒体单元传送给群中的各个用户,从而实现群中各个用户同步参考点的统一。基于群同步参考点的群同步算法是基于统计的方法确定和调整群中各个用户的同步参考点时间。基准参考点用户的初始同步参考点时间和序号由该用户自己确定,经过一个同步参考点周期,序号加1。基准参考点用户的同步参考点可以认为是理想的同步参考点,即基准同步参考点周期保持恒定不变,无须调整。群中其它各个用户的同步参考点以基准参考点用户的分组到达时间为依据计算初始同步参考点时间和当前同步参考点时间。根据各自的终端时钟,经过一个同步参考周期,同步参考点序号加1。群中各个用户相互对应的同步参考点要保持相同的同步参考序号,并且通过实时调整各个用户的同步参考时间实现群同步。

5 结论

群组通信研究的目标是提供分布式文件系统、分布式数据系统、容错系统、并行计算和 CSCW 系统等各种分布式应用系统所需的群组通信服务。目前的研究结果并未很好地解决群组通信问题,还有待进一步拓宽现有计算机网络概念,改进解决方案,以推动群组通信研究和应用的发展。

参考文献

- [1]Huang C C, Eric P K, Mckinley P K, Design and implementation of multicast for ATM-based high performance computing, Proc Supercomputing'94, Los Alamitos, Calif, 1994.164-173
- [2]Huang Y, Huang C C, Mckinley P K, Multicast virtual topologies for group communication in MPCs and ATM clusters, On-line Proc Supercomputing'95 at <http://www.supercom.org/sc95/proceedings/>, San Diego, CA, 1995
- [3]Akyildiz I F, Yen W, Multimedia group synchronization protocols for integrated services networks, IEEE J Selected Areas in Commun, 1996,14(1)
- [4]Zarros P N, Lee M J, Saadawi T N, Interparticipant synchronization in real-time multimedia conferencing using feedback, IEEE/ASM Transactions on network, 1996,4(2)
- [5]史美林等,协同科学—从“协同学”到 CSCW,清华大学学报,1997,37(1):85~88