

文章编号:1000-582x(2000)03-0066-04

18

66-69

基于层次控制树和分组的可靠多点广播协议

TP393

符云清¹, 吴中福¹, 郑运忠²

(1 重庆大学 计算机学院, 重庆 400044; 2. 北方航空公司 昆明飞行训练中心, 昆明 650200)

摘要: 论文提出了适用于实时交互式协作环境的可靠多点广播协议。协议利用分组和层次控制树机制使差错恢复完全局部化, 并有效地减少主干网上的报文流量和减轻主机的处理负担, 分析表明协议具有良好的吞吐性能和扩展性。

关键词: 协议; 多点广播; 可靠多点广播; 流控

中图分类号: TP 393.04

文献标识码: A

计算机网络

实时交互式远程教育等协同应用中, 发送方要将报文信息以较小的端到端延迟传送给远程的多个接收者, 以保证协作的实时性和交互性^[1]。传统点对点通信方式和广播通信方式都难以满足大规模协作应用要求, 而多点广播方式(multicast)综合了点点对通信和广播通信的优点。发送节点一次便可将报文高效地传送给指定的多个接收者, 即多点广播分组, 且多点广播路由协议能保证任何给定的链路仅使用一次。因此, 多点广播方式能充分利用通信网络的传输能力, 降低报文传输延迟。

对于要求实时传输和处理的音频和视频等媒体数据流来说, 由于这些数据流能够容忍一定程度的丢失和出错, 适合采用多点广播方式传送。而另一些应用中, 如远程教学中授课内容、学习资料、交互中的共享信息及控制信息等, 既要求较小的端到端延迟, 又要保证传输的可靠性, 传输过程中出现的任何差错必须纠正, 以保证协作的一致性^[2]和有效性, 这便要求在多点广播的基础上提供可靠报文传输服务。

传统点对点通信方式中的可靠传送机制并不适合一个发送节点和多个接收点的应用场合。首先, 多个接收节点在很短的时间内几乎同时发送答报文(ACK/NAK), 大量突发的应答报文将导致“应答爆炸”^[3], 从而造成网络拥塞。其次, 发送节点要负责处理所有接收点的应答报文、维护和管理大量接收者的状态信息并负责报文重传, 容易导致主机性能瓶颈。可靠多点

广播协议大多采用NAK抑制技术^[4], 即接收节点利用随机延迟和多点广播应答报文方法保证发送方在最好情形下最多只收到一个NAK报文, 达到抑制重复NAK报文的目, 以免造成网络拥塞。目前, 世界上许多研究小组致力于可靠多点广播协议的研究, 并提出了多种适合不同应用要求的可靠多点广播协议, 如基于登记服务器的可靠多点广播协议^[5]和基于ACK树的可靠多点广播协议^[6]等。但是, 大多数协议中小部分接收节点因链路拥塞、本地缓冲容量或接收速度限制等原因造成报文丢失或出错时, 其应答和重传报文都涉及整个网络或大量已正确接收报文的节点, 且大多是固定节点(如发送节点)来处理请求和进行重传。因此, 协议的吞吐性能和扩展性难以满足大规模协作的要求。

1 基于层次控制树和分组的可靠多点广播协议

当网络范围很大时(如大型广域网, 甚至Internet), 各协作对象分散在网络的各个部分。可靠多点广播协议的性能直接影响协作过程中的交互性、实时性和一致性。高效和可扩展的可靠多点广播协议必须将差错恢复尽可能地控制在网络的局部, 而不应涉及整个网络或大量已正确接收报文的节点, 且应在节点间均衡处理负担, 而不应由某个固定节点来集中处理。

基于上述要求, 笔者提出基于层次控制树和分组的可靠多点广播协议(Tree and Group based Reliable Multicast

· 收稿日期: 1999-08-18

作者简介: 符云清(1969-), 男, 湖南衡南人, 讲师, 博士生, 主要从事计算机网络通信、远程教育、多媒体技术等方面的研究开发工作。

Protocol, TGRMP)。该协议对节点按一定原则进行分组,并引入层次控制机制,从而使差错恢复完全本地化,使协议在吞吐性能和扩展性等方面具有良好的性能。

1.1 多点广播分组思想

大型网络通常是由许多局域网(LAN)与骨干网连接而成的互连网络(internetwork),协作用户分布在各个局域网中。局域网一般具有较高的传输带宽和可靠性,如 Ethernet 传输速率可达 10 Mbps、100 Mbps 甚至 1000 Mbps,而骨干网的传输速率和可靠性相对要低得多,如 DDN 的传输速率通常为 64 kbps 或 128 kbps。差错恢复时若利用所有参与协作的节点组成的多点广播分组来进行请求和重传,短小时内大量的重复请求将集中地传给发送节点,使处理机负担过重而导致性能瓶颈,且这些重传请求都会蔓延到低速的骨干网上,容易造成骨干网拥塞和出现饱和,如图 1 所示。

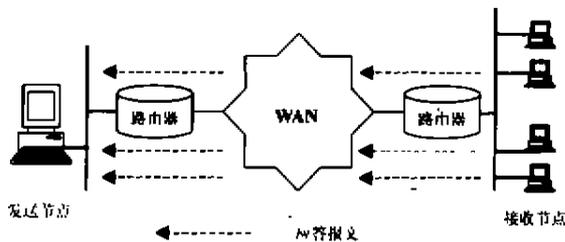


图 1 接收节点直接向发送节点传送应答过程

在协议 TGRMP 中,我们按一定的原则对参与节点进行分组。每个组包含若干成员节点,并用唯一的 D 类 IP 地址标识。每个分组中指定一个管理节点,由管理节点负责处理组内应答报文和进行组内重传。这样,分组成员的应答和重传报文都限制在网络的局部,而不会蔓延到骨干网上。即使分组管理节点未正确接收报文,也只有管理节点向外发送重传请求,重传报文也只传给管理节点,然后再由管理节点传给其它组内的成员。这样既分散差错恢复的处理负担,又大大减少骨干网链路的通信流量,如图 2 所示。

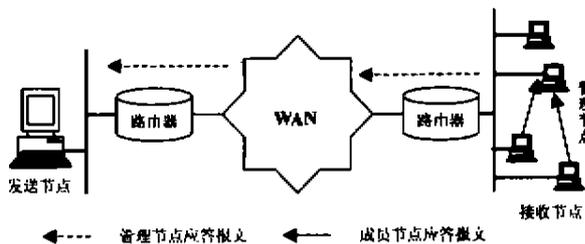


图 2 接收节点通过管理节点向发送节点传送应答过程

1.2 层次控制树思想

无论采用哪种方法进行分组,每个分组都不能太大,否则,分组管理节点或网络资源成为性能瓶颈。因此,分组管理节点数目随参与协作的用户增加而增大。若将所有管理节点组成一个平面结构的多点广播分组,难以实现管理节点间差错恢复的局部化。为此,我们在协议中引入了层次控制机制,即将所有分组管理节点组织成层次树状结构。以发送节点所在的分组管理节点作为根节点,根节点和中间节点最多有 K 个其它分组管理节点作为其子节点。树状结构中有子节点的管理节点称作中间节点,而无儿子节点的管理节点称作叶节点,由此形成一个最大度为 $K+1$ 的树。如图 3 所示。

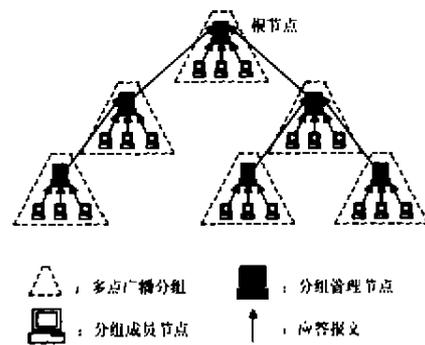


图 3 分组管理节点构成层次控制树示意图

根节点和中间节点负责处理其子节点请求,并在它和子节点组成的单层子树内进行报文重传。文献 [6] 提出的 TMTP 协议中,采用报文传输的跳数(hops, 也称作 TTL)来限制重传报文的传播范围,即以最近子节点的跳数为半径来多点广播重传报文。但该方法并不能限定只有子节点才接收重传报文,而且不能很好地适应网络物理结构的变化。为此,将层次树中每个父节点和其所有子节点组成一个多点广播分组。这样,父节点的重传报文将严格地限制在其子节点范围内,且与下层网络物理结构无关。

上述 K 的取值据参与节点规模和通信要求,在组创建和登记时动态确定。由于树的最大度为 $K+1$,限制了发送节点及每个分组管理节点的处理负荷,使协议开销随参与协作的总节点数而缓慢增加。为实现差错恢复的本地化及减少应答报文和重传报文的延迟,应尽可能在树状结构中将子节点与最近的父节点组织在一起。

1.3 层次树的构建和维护

协议采用自顶向下的方法来构建层次控制树。未连到层次树的管理节点周期性多点广播请求信息,已

在层次树上且其子数目小于 K 的管理节点进行响应。为找到较近的管理节点作为父节点,在请求报文中最初设置较小的 TTL 值(如 $TTL = 1$)。管理节点多点广播请求报文后,立即启动定时器。定时器超时后,首先检查在定时期内是否收到来自其它管理节点的响应报文。若没有,则认为请求报文中的 TTL 值太小,以致请求不能达到任何其它管理节点,于是增大请求报文中 TTL 值和定时值,然后再多点广播请求报文,并启动定时器。该过程一直重复,直到收到一个或多个响应报文为止,然后从响应中选择时间距离最近的节点作为父节点,从而连入层次树。当所有管理节点都连到层次树中后,便形成以发送节点所在分组的管理节点为根的层次控制树。

在远程教育等协作系统中,参与节点随时可能退出和加入协作系统。当一个新的分组产生后,应将该分组的管理节点连到层次树中。这便涉及对层次树的动态维护,即新管理节点加入和层次树中管理节点退出时如何来维护树状结构。

1.3.1 新管理节点的加入

新的管理节点产生后,跟最初构建层次树一样,该管理节点周期性多点广播请求信息,已在层次树上且子数目小于 K 的管理节点对该请求进行响应。管理节点然后从响应中选择时间距离最近的节点作为父节点,从而连入层次树。

1.3.2 层次树中管理节点的退出

层次树中管理节点的退出时相对新管理节点的加入要复杂些。退出的节点有两类:叶子节点和中间节点。因叶子节点无子节点,所以退出控制很简单,而中间节点退出时应将其下的子节点连到新的父节点上,控制和管理比较复杂。下面对这两种情形分别讨论。

叶子退出之前,应向其父节点发送退出层次树请求。父节点收到子节点(也是叶子节点)的请求后,则向发请求的子节点发送退出确认,并从子节点集合中删除请求的子节点。发送退出请求的节点收到确认后,退出相应的多点广播分组,并结束运行。

对中间节点的退出,有许多算法来维护层次树,如将退出节点 q 的子节点连到 q 的姊妹节点或父节点上等。但这些方法难以保证子节点尽可能与较近的父节点相连,从而使重建的层次树不是较好的逻辑层次组织。在本协议中使用一种简单的方法:中间节点 q 在退出之前,先强迫 q 所有的子节点去寻找新的父节点,直到所有子节点响应后,节点 q 才退出。要注意的是, q 节点的子节点要避免将下层已在层次树上的某个节点作为父节点。

1.4 数据传送和差错恢复

协议 TGRMP 中,数据传送和差错恢复算法很简单:发送节点将报文向所有节点组成的多点广播分组进行广播,层次树上的管理节点和各分组的成员节点都可能收到该多点广播报文。分组成员节点发现报文丢失后,采用前述的 NAK 抑制方法向其分组管理节点发送 NAK 报文,请求重传相应的报文。若管理节点能满足该请求,则在分组内以多点广播方式进行报文重传,否则不进行任何响应。如果分组管理节点也未正确接收某个报文,则使用 NAK 抑制方法向其层次树上的父节点发送 NAK 报文来请求重传,若父节点有请求的报文,则向它和所有子节点组成的多点广播分组重传相应的请求报文,否则不进行任何响应。

2 协议性能分析结果

笔者采用 B. N. Levine 和 J. J. Garcia - Luna - Aceves 在文[3]中提出的模型,对协议的最大吞吐性能进行分析。分析中假设协议 TGRMP 的层次树中树根节点和中间节点的子节点数最多为 10,每个分组平均节点数为 20。图 4 给出了协议 TGRMP 和基于接收方带 NAK 抑制的可靠多点广播协议 (N_2) 在报文丢失概率 $p = 0.15$ 时的吞吐性能,图 5 给出了协议 TGRMP 与 N_2 在 $p = 0.25$ 时的吞吐性能(图中横坐标为参与协作的节点数目,纵坐标为协议 TGRMP 与协议 N_2 的吞吐性能,单位为 Packets/ms)。

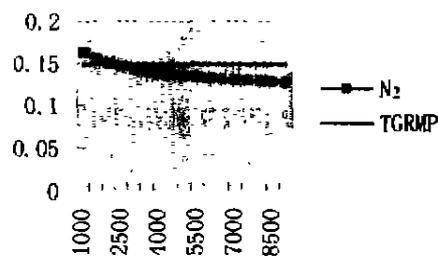


图 4 TGRMP 和 N_2 的吞吐性能($p = 0.15$)

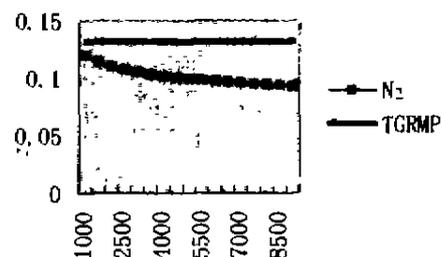


图 5 TGRMP 和 N_2 的吞吐性能($p = 0.25$)

从上面两个图示可知,协议 N_2 吞吐性能随节点总数增加而明显下降,而协议 TGRMP 的吞吐性能与系统节点总数无关,而是保持不变。随报文丢失概率 p 的增加,协议 TGRMP 和 N_2 的吞吐性能都下降,但协议 N_2 的吞吐性能下降幅度更明显。因此,协议 TGRMP 具有良好的扩展性,能适用于大规模的实时协作应用环境。

3 结论

协议 TGRMP 利用层次控制和分组机制能有效地使差错恢复局部化和减轻节点的处理负担,从而使协议具有很好的吞吐性能,能适用于大规模的分布式协同应用。目前,该协议只支持一对多应用,进一步的研究是使协议能支持多对多应用。

参 考 文 献

- [1] 尤晓苑,吴良芝. 多媒体通信协议与应用[M]. 北京:清华大学出版社,1998.
- [2] 郑庆华. CSC 的理论、实现方法与应用[J]. 计算机科学, 1996,23(5):18~25.
- [3] DON TOWSLEY. An Comparison of Sender - Initiated and Receiver - Initiated Reliable Multicast Protocol[J]. Proceedings of ACM Multimedia, 1996,11(2):282~291.
- [4] HOLBROOK H W. Log - base Receiver - Reliable Multicast for DIS[J]. Proceedings of SIGCOMM'95, 1995,25(4):328~341.
- [5] LEVINE B. The Case for Concurrent Multicast Using Shared ACK Trees[J]. Proceedings of ACM Multimedia, 1996,11(2):365~376.
- [6] YAVATKAR R. A reliable Dissemination Protocol for Interactive Collaborative Application[J]. University of Kentucky, 1995, (10):36~47.

Tree and Group Based Reliable Multicast Protocol

FU Yun-qing¹, WU Zhong-fu¹, ZHENG Yun-zhong²

(1. Computer College of Chongqing University, Chongqing 400044, China;

2. Kunming Flying Training Center of North Airline Corp, Kunming 650200, China)

ABSTRACT: The paper proposes a reliable multicast protocol which is appropriate for real-time and interactive collaborative environment. The control tree and grouping mechanism are introduced to make error recovery localized, which can efficiently decrease the amount of packet stream on the backbone of network and the workload of host. Analysis shows that the protocol has perfect throughput and can be extended to very wide applications.

KEYWORDS: protocol; multicast; reliable multicast; flow control

(责任编辑 吕赛英)

·下期论文摘要介绍·

并行工程在陶瓷产品开发制造中的应用

赵象元, 刘飞, 余力, 但斌

(重庆大学 制造工程研究所, 重庆 400044)

摘 要: 基于并行工程的思想, 通过对陶瓷制造业日用陶瓷产品开发制造过程的研究, 建立了产品并行开发制造过程模型及其实现的支撑环境, 并就关键技术进行了讨论。研究成果在企业得到初步应用, 缩短了产品开发周期、提高了产品质量。