维普资讯 http://www.cqvip.com

计算机科学2000Vol. 27№. 1

组播错误控制技术

The Error Control Technology for Multicasting

李方敏[1,2] 姚 砺[1] 叶澄清[1]

(浙江大学计算机系 杭州310027)[1] (湘潭工学院信电系 湘潭411201)[2]

Abstract With the development of Internet and the increasing bandwidth, multicasting has been widely used in many applications, such as videoconferencing, dissemination of stock quotes, Computer Supported Collaboraed Work (CSCW) and so on. When we do research about reliable multicast protocol, we place our focus on the issues, including error control, flow control and scalability. This paper overviews the error control and introduces an integrated error control mechanism in detail.

Keywords Reliable multicast, Error control, Layered transmission, Forward error correction

1. 引言

传统的点点通信(unicast)和广播通信(broadcast)是两种主要的通信方式,但随着 Internet 的日益普及、带宽的不断提高,利用 IP 组播技术在 Internet 上建立了虚拟的 multicast backbone(Mbone)[1],其上的点到多点通信的应用和各种工具也相继产生。人们可方便地在 Mbone 上举行远程视频会议,股票实时行情的分发,参与远程教育等。但可靠的组播协议的研究相对传统的点点协议而言,在错误控制、路由、流量控制、可伸缩性都具有其独有的特点,其复杂度也大得多。就错误控制和流量控制而言,传统的点点通信只涉及收发双方,而在组播环境下,一个发送者对应的接收者是动态变化的,从几个到几十甚至上千个,而且接收者的处理能力、可用的带宽可能都不一样,因而使得数据的可靠传送变得更加复杂。

就传输数据的可靠性而言,组播应用可分成两类:一类不需要完全的可靠传送,可容忍适量的错误,如视频应用和音频应用,这一类应用可通过人本身对上下文环境的理解来弥补适量的错误,但这类应用对延迟比较敏感;另一类应用需要可靠的数据传输,不允许传输错误。诸如软件的分发,分布式的交互仿真,股票实时行情的分发等,这类应用对延迟不是非常敏感,对这两类应用其错误控制机制是不一样的,在采用层次传输机制的情况下,前者可将音频、视频信号通过编码转换成不同的信号层,接收者可根据自己的可用带宽选择订购不同的层。

根据错误控制的负责者不一样可分成两类:发送者主动(sender-initated)和接收者主动(receiver-initated),在文[2]中对这两类机制进行了比较,后者有更好的可伸缩性,较好地适应组播环境的异构性。在所有的组播协议中,发送者对每个发送的包都进行顺序编号,发送者主动的协议要求每个接收者对每个接收的包都发送确认信息(ACK),发送者需保存所有接收者的状态信息,然后由发送者确定是否所有的接收者都正确接收了所有的包,没有的话由发送者决定采用组播或点点投递给接受者。在发送者主动的可靠组播协议中,错误的检测和重传都由发送者负责,发送者将被确认信息所淹没,并且每个接收者对每个接收的包都发送确认信息(ACK)将严重地加重网络的负荷。而在基于接受者主动的协议中,由接收者检测错误或丢失的包,然后通知发送者(NAK)没有正确接收的包,这样接收者的处理负荷将大量地减少,有较好的可伸缩性。

2. 相关工作

在可靠组播技术的错误恢复处理中,有两种常用技术可用即自动重传(ARQ:Automatic Repeat Request)和前向纠错技术(FEC:Forward Error Correction)。ARQ 根据接收者的请求重传丢失的数据包,FEC 通过传送冗余数据来达到错误恢复的目的。在文[3]中提出了一种可应用在可靠组播(RM)环境中的 FEC 技术,文[4]经过仿真证明了软件 FEC 技术的可

李方敏 博士生,讲师,研究方向为网络协议、组播技术、网络多媒体技术。姚 励 博士生,研究方向为计算机网络。叶澄清 教授、博导,研究领域为高性能体系结构,计算机网络。

维普资讯 http://www.cqvip.com

行性，在文[5]中进一步提出了一种混合 ARQ 和 FEC，并对其作为 RM 的下层来恢复丢失的包的技术。文[3,4]说明将 FEC 用于 RM 的差错控制可有效地减少传输时间和网络带宽需求。

文[6]提出了一种基于层次传输的可靠组播的流控机制，文[7]进一步将 FEC 技术与层次传输技术结合起来用于流控，较好地解决了异构环境下的 RM 的可伸缩性和发送速率的可选择性。

文[8]比较了两种本地恢复方案：基于服务器和基于接收者，通过分析前者在协议吞吐率、网络带宽、整体处理代价等方面能获得更高的性能增益。文[9]将 FEC/ARQ 在层次传输机制的模型下与本地恢复技术结合起来，并就其性能与未使用 FEC 的层次传输本地恢复方案进行了比较，前者只需要更少的网络带宽和缓冲容量，缓冲容量的减少是限制修复服务器保存少量的编码包而不是缓冲原始包来取得的，因为编码包可被任何接收者用以重建其原始数据。

本文将结合自动重传和前向纠错技术，研究基于层次传输[4,5,8,10]的本地错误恢复模型。文[9]没有论及发送者如何将原始数据源划分成 n 个层次的信号，本文将结合文[11]提出的 SCUBA 协议进一步完善文[9]的协议模型，并将讨论层次组播协议错误控制技术需进一步研究的问题。

3. 前向纠错技术

FEC 通过将采用(n,k)编码将 k 个原始包编码成 n 个包，产生 n－k 个冗余包，接收者只要成功地接收了 k 个包就可以重构原始数据。这种技术在电信通信领域已得到广泛的应用，但电信通信中每个包只包含少数位，而计算机通信的信息量大得多，因此其复杂度也大得多。随着微机以及工作站性能的不断提高，FEC 的软件编码解码速度在奔腾133上已可达到10Mb/s以上，因此软件 FEC 技术已能广泛应用于当前的大多数机器，而10Mb/s 的速率也可匹配计算机从 Internet 得到的可用带宽。

文[3]提出的块删除编码方法已在各种组播环境得到广泛的应用，该方法采用线性编码技术。其思想如下(详细内容参见文[3])：

假设$\underline{x}=x_0\cdots x_{k-1}$是源数据，G 是一个 nxk 的矩阵，则一个(n,k)线性编码表示为：$\underline{y}=G\underline{x}$。

如果编码数据包含源数据的一个拷贝，则这种编码叫做系统编码，对应于在 G 中包含一个单位距阵 I_k，采用系统编码可减轻接收者的解码负担。

4. 层次传输协议模型

4.1 网络模型

传统的 RM 协议为了修复接收者丢失的包，由发送者组播丢失的包给接收者，导致每个接收者即使已正确接收了该包，仍将接收重复的包，这样不仅导致网络带宽的无谓浪费，而且加重了已正确接收该包的接收者的处理负担。因此，通过将具有相同特性(如处理能力、相同的可用带宽等)的接收者划分成一个修复域，并在该域指定一个修复服务器，从而可以大大地减少网络带宽的使用，并且可以减轻发送者的负荷。

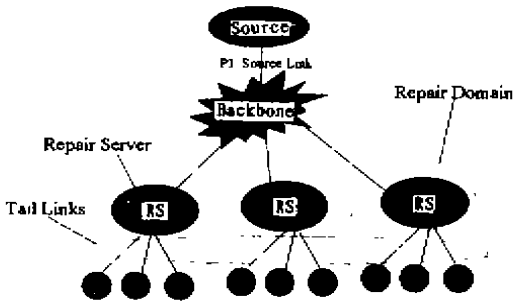

图1 网络模型

图1为采用了层次传输并设置了本地恢复服务器的网络模型，典型地我们把从源即发送者到骨干网的叫源链路，文[12]认为包错误的产生主要在源链路和从骨干网到单个接收站点之间的尾链路(Tail Links)，骨干网和接收者几乎不产生错误，并且尾链路很有可能就是瓶颈所在，因此可假设错误事件是时间上独立的。假使在源链路和尾链路的出错概率均为 p_1，则接受者的端端出错概率 p 为$1-(1-p_1)^2$。

RS 为恢复服务器，一般说来，RS 应当位于接入骨干网的边缘路由器的同一子网，也就是说与边缘路由器是通过局网连接，可认为同一修复域的所有接收者能正确接收的包 RS 都能正确接收。

4.2 层次传输编码

在对发送者的源数据进行编码时，对可容忍错误的实时音、视频数据和需完全可靠传输的其它数据应采用不同的编码方法。

实时音、视频数据层次编码通过层次编码算法[10,11]将源数据编码成一系列的层 l1,l2,…,ln，l1叫做基层，其它层是增强层，每一层都对应不同的网络通道，实际上也就是对应一个组播组。在实际实现时利用 IP 组播技术，接收者可根据其可用的带宽选择加入多个不同的组，即订阅不同速率的通道。

对需可靠的数据传输的应用来说，如软件分发，可将通过 FEC 编码的数据包分配在不同的层传输，同样

维普资讯 http://www.cqvip.com

接收者可根据其可用的带宽选择加入多个不同的组，即订阅不同速率的通道，接收者只要正确接收了任意k个包(假设FEC将k个包编码成n个)则可成功地重建原始包，这样的话，将不会出现可用带宽高的接收者吃不饱，而低带宽的用户却会丢失大量的包，加重发送者或恢复服务器的处理负担。

4.3 本地恢复

在选择恢复者时，文[8]对基于服务器和基于接收者两种本地恢复方案在SRM环境下进行了比较研究，基于本地恢复的机制在带宽使用、吞吐率、延时方面相对于不使用本地恢复机制来说能获得较好的性能增益，而基于服务器的方案，能产生更高的协议吞吐率和低的带宽使用。

上述两种方案均基于NAK，恢复服务器与恢复域的接入路由器位于相同地点，如图1所示，接收者从服务器恢复丢失的包，如果服务器没有其所要的包则该服务器从上一级的服务器或发送者恢复，因此，针对每个会晤服务器都需要一定的缓冲。基于接收者的机制，其错误恢复工作由任意的接收者和发送者共同完成。下面简单描述文[10]基于服务器机制的协议行为：

·发送者在组播地址A上组播包。

·一旦接收者检测到包错误，接收者等待一个随机时间后组播一个NAK消息给其同域的接收者和恢复服务器，然后启动NAK重传定时器。如果一个接收者在等待时从别的接收者那里收到对同一个包的NAK消息，它取消自己的NAK消息，并设置NAK重传定时器就象自己发送了NAK消息一样。

·当恢复服务器检测到包错误时，它等待一个随机时间后组播一个NAK消息给发送者(或上一层的恢复服务器)以及在组播树中的同层恢复服务器，然后启动NAK重传定时器。如果在等待发送NAK时从别的恢复服务器那里收到对同一个包的NAK消息，它取消自己的NAK消息，并设置NAK重传定时器就象自己发送了NAK消息一样。

·如果恢复服务器的缓冲区中有它负责的恢复域中某个成员请求的包，则在本域组播该包，否则按三处理。

·发送者在接收到某个包的NAK消息时，重新组播该包。

·重传定时器过期时，如果还没有收所请求重传的包，则触发NAK的重传。

该协议没有考虑将FEC集成，在将FEC/ARQ与本地恢复机制集成时，发送者在基层11发送原始数据，而在其它层发送冗余的FEC编码数据，这样可以减轻RS和接收者的解码负担。同时对上述协议的重传机制也应作适当修改，重传以前没有传送过的FEC编码数据包，这样任何RS和接收者均可以用来重构其数据。

5. 进一步研究的问题

由于组播技术相对点点通信来讲，是一个新的研究方向，很多方面尚待进一步的研究。就错误控制来说，需解决的问题有：

·网络拓扑动态改变：当网络拓扑动态改变时，如何重新设置协议参数，使尽快收敛。

·RS的缓冲管理：一个RS可能参与多个会晤，每个会晤都需一定的缓冲容量，如何检测缓冲中的过时数据即垃圾的收集问题。

·各种定时机制。

·采用层次组播时，会晤目录服务、组播地址的分配等。

·各种流量的公平性问题。

·性能评价：性能度量指标的选择，网络模型的选取，仿真环境的研制等都对性能评价有重要的影响。

参考文献

1 Mbone E H. The Multicast Backbone. CACM, 1994, 37(8), 54～60

2 Pingali S, et al. A comparision of sender-initiated and receiver-initiated reliable multicast protocols. IEEE Journal on Selected Areas in Comm., 1997, 15(3): 398～406

3 Rizzo I. Effective erasure codes for reliable computer communication protocols. ACM Computer Communication Review, 1997, 27(2): 24～26

4 Rizzo L, Vicisano L. A Reliable Multicast data Distribution Protocol based on software FEC techniques. In: Proc. of the The Fourth IEEE Workshop on the Architecture and Implementation of High Performance Communication System(HPCS' 97)Sani Beach, Chalkidiki, Greece, 1997

5 Nonnenmacher J, et al. Parity-Based loss recovery for reliable multicast transmission. IEEE/ACM Transaction on Networking, 1998, 6(4): 349～361

6 Bhattaacharyya S, et al. Efficient multicast flow control using multiple multicast groups. University of Massachusetts: [CMPSCI Technical Report TR97-15]. 1997

7 赵东升，卢锡城，周兴铭. 基于集成FEC和层次传输的可靠组播的流控技术. 计算机学报, 1999, 22(1): 86～91

8 Kasera S, et al. A Comparision of Server-based and Receiver-Based Local Recovery Approaches for Scalable Reliable Multicast. University of Massachusetts: [CMPSCI Technical Report TR97-69]. December 1997

9 Rubenstein D, et al. Improving Reliable Multicast Using Active Parity Encoding Services (APES). University of Massachusetts. [Technical Report 98-79]. 1998

10 McCanne S, Jacobson V. Receiver-driven Layered Multicast. In: Proc. of SIGCOMM' 96. Stanford, CA, ACM, 1996

11 Amir E, et al. Receiver-driver Bandwidth Adaptation for Light-Weight Sessions. ACM Multimedia, Seattle, WA, 1997

12 Yajnik M, et al. Packet Loss Correlation in the Mbone Multicast Network. In: Proc. of Global Internet Conf. 1996

维普资讯 http://www.cqvip.com

B-ISDN ATM 交换结构 交换机 通信网 (21)

计算机科学2000Vol.27№.1

72-74,79

典型ATM交换结构分析与交换机设计分类*)

Analysis of Typical ATM Switching Fabrics and Categories of Switch Design

陈永光 TN913.24

(解放军电子工程学院 合肥230037)

Abstract This paper introduces ATM switching functions and several indexes for evaluating the performance of ATM switch. Some typical switching fabrics and their influence on switch performance are also analyzed. Furthermore, the paper discusses categories of switch design in detail and proposes a scheme to realize small type of ATM switch by using MB866 family chips.

Keywords ATM, Switching fabric, Shared medium, Self routing

目前的通信网正逐步向宽带化、综合化和智能化方向发展，为了传送高清晰度电视(HDTV)图像及高速数据，要求通信网高速宽带化。高速宽带通信与现有的电话及数据通信相比，应具有极宽的通信频带(几十Mbps以上)。另一方面，随着科技的发展，人们对通信的需求不断增加，导致未来通信业务的不确定性。因此，人们致力于研究新的通信方式以适应上述情况，其中，ATM可能是实现宽带ISDN的最佳方式，可以满足B-ISDN的要求，因而受到各发达国家重视。ATM的核心是交换技术，从商用ATM交换机产品的设计中可以发现，目前已经形成了若干种典型ATM交换结构，因此，对交换结构的分析和对交换机设计方法的归纳具有实际意义。

1 ATM交换功能和交换机性能指标

ATM交换机需完成以下动作，即三个功能：(1)空分交换。信息从编号为N的入口通过交换通往编号为M的出口，这是一种路由选择功能。(2)时隙交换。时隙(信道)i中信息被交换到时隙(信道)j，这里牵涉到信元头的转译。(3)排队。两个来自不同链路的信元可能同时到达ATM交换机，并去向同一链路，这时交换机中缓冲存储器就用以暂放那些暂时不能送出的ATM信元，即有些ATM信元在排队等待送出，这样就有可能对信元统计复用，这是ATM交换机的典型功能。

ATM交换机完成这些功能的质量如何，其性能的评价主要由三个指标来衡量：

·连接阻塞率：ATM是面向连接的，连接阻塞率就是在入出口之间没有足够的资源以保证现有连接和新建连接质量的概率。这是由交换机配置、诸如交换机内部连接数以及这些连接上的负载决定的。

·信元丢失/信元插入概率：在ATM交换机中，当有很多信元的去向是同一链路，而缓冲存储器容量又不足以同时存入所有要排队的信元时，就会发生信元丢失。信元丢失概率应有限度以保证语意之透明性；ATM交换机信元丢失概率典型范围是10^{-8}～10^{-12}。另一种可能是ATM交换机内部选择路由有差错，导致某些信元被错误地送往另一逻辑连接。信元插入概率典型值应小于信元丢失率的1000倍或更多些。

·交换时延：ATM交换机的交换时延范围典型值为100～1000μs之间，其抖动为几百μs或更小一些。该指标通常以时延不超过某一值的概率来表示，如时延大于10μs的概率为$<10^{-10}$。这些性能指标受到ATM交换机排队方式的直接影响，根据缓冲存储器所处位置，排队方式有三种，即输入排队、输出排队和中央排队，有三种参数影响到排队系统的复杂性，即队列长短、存储器速度和存储器控制。

输入排队的有用负载最多只有58.6%[1]，即58.6%的来向信元可含有有用数据，其余皆为空的，因而性能较差。输出排队的有用负载为80%左右[2]。中央排队和它差不多，但可用较小规模缓冲存储器，例如当信元丢失率为10^{-8}时，输出排队系统队列规模为40个信元，而中央排队系统中每条输出链路队列可短于10个信元，只是中央排队系统需要较快的速度和存取时

*)安徽省自然科学基金资助项目98312518。陈永光 博士(后)，教授，研究方向为ATM交换技术、电子站仿真、电子对抗作战效能分析等。