

ATM 网络的流量控制

数字网, 通信网

方震龙 浩 吴时霖

(复旦大学计算机与通信实验室 上海200433)

TN/13.24

摘要 After a brief introduction to the four service categories defined by ATM Forum, traffic management of ABR service is discussed in detail, including bandwidth allocation fairness, deadlock, and credit-based and rate-based flow control. Finally a proposal of integration of the two flow control schemes is given.

关键词 ATM, Traffic management, ABR, Credit-based/rate-based flow control, I/O automata model

一、前言

在任何一个通信网中,为保证网络的有效工作,流量控制都是非常重要的。电路交换网用简单的接受或拒绝来控制呼叫的人网与否,分组交换网采用窗口式流量控制方案,它们均不能适应高速 ATM (异步传输模式)网络的流量控制要求。ATM 网络的流量控制有两种策略,一是依赖上层协议的流控机制,如 TCP 和 DECnet 的运输层滑动窗口算法,计算和模拟结果表明^[1],这种策略不能有效地利用 ATM 网络的带宽,当端站点以 100Mbps 的速率发送数据包时,40%的带宽将浪费在信元丢失所引起的重传上;二是专门在 ATM 层设计流量控制机制,根据 ATM 的传输特点进行发送速率、信元缓冲、丢失重传等的控制。本文讨论这种策略。

ATM Forum 定义了四种 ATM 网络流量与服务控制方式:固定位速率(CBR)、可变位速率(VBR)、不定位速率(UBR)和可用位速率(ABR)。以下分别对上述四种服务方式与流量控制策略进行简要的介绍与比较,阐述重点将放在 ATM Forum 重点推出的 ABR 反馈环控制上。

二、ATM 的四种服务方式

固定位速率(CBR)用于对时延敏感的实时数据的传送,典型的如话音和动态图象,ATM 网络必须为该类业务提供一恒定的带宽和足够低的信元延迟

方差(抖动)特性。当用户提出 ATM 连接建立请求时,如果 ATM 交换机的连接接受控制模块(CAC)认为能够提供符合要求的服务,就按照用户所请求的带宽建立虚连接(VC)并分配网络资源,否则拒绝此 CBR 服务请求。在 CBR 数据传输过程中,为防止发送方违反合约规定并因此妨碍其他用户的业务,ATM 交换机必须监视各 ABR 连接的传输速率,为此常采用 GCRA 算法,即漏桶算法。

可变位速率(VBR)主要是针对对时延和抖动要求比较严格,而带有明显的突发特性的数据,这类业务的峰值速率与平均速率的比值远大于1,如实时的事务处理信号和高度压缩的影像数据。用户提出 VBR 服务请求时,给出峰值信元速率(PCR)和平均信元速率(SCR)两个参数,如果 CAC 接受了此请求,用户突发数据可以在短时间内超过 SCR,甚至达到 PCR,但 VBR 必须随之降低传输速率以满足 SCR 的要求。监控 VBR 的主要算法是双漏桶算法,对 PCR 和 SCR 分别进行动态监视,特别是要严格限制 PCR 时间,以便给 VBR 留出带宽。

不定位速率(UBR)是为对传输延迟和信元抖动都没有严格要求的应用设计的,无需进行连接接受控制,或者也可以认为,如果网路通畅而且有可用的 VCI 空间,CAC 对 UBR 请求总是予以满足的。但是 ATM 网络对 UBR 实际所能获得的带宽没有保证,当网络负载过重时,UBR 业务将首当其冲陷入阻塞状态。

方震、龙浩 硕士研究生,专业方向:计算机网络与数据通信。吴时霖 教授,博士生导师,主要研究领域:计算机网络与数据通信,Petri 网理论

可用位速率 (ABR) 服务用于对时延不敏感,但可靠性要求比 UBR 更高的应用,如关键数据、网络服务器信息等。ABR 可用的带宽是经 CBR 和 VBR 分配以后剩余的部分。ABR 与 UBR 的主要区别,是 ATM 网络为 ABR 提供一种反馈控制机制,有效地提高了信道利用率,降低了信元丢失率,只要信元丢失率与可用带宽均落在约定的范围内,用户就能获得有保证的服务质量。当用户请求作 ABR 连接时,先借助 ATM 信令系统约定 PCR 值,再提出“最小信元速率”(MCR)参数。网络将信守约定提供不小于 MCR 的带宽,在此基础上为各个以 ABR 方式存在的 VC 公平地分配余下的带宽。在 LAN 环境中采用 ATM 技术主要是它能高效支持以太网、令牌环与 TCP/IP,而在这些应用场合中,ABR 服务将最适用于 LAN 仿真^[2],可以预计,多数数据传输将以 ABR 方式进行。

三、ABR 流量控制

1 带宽分配公平性的定义

上面提到,ABR 方式服务的一个基本内容,是为各个需连接提供公平的带宽。这里“公平”是一个模糊的概念,一个 VC 获得的总的 ABR 带宽可以有多种方案,典型的如以下两种:(1)用户指明的 MCR 加上 ATM 网络剩余带宽的等分值;(2)等分总带宽,并且保证其中每一份都不小于用户提出的 MCR。举个例子。有两个应用程序需要在 1 秒内将 1M bytes (约 20,000 个信元)传送到目标节点,另两个应用程序需要在 2 秒内将 1M bytes 传送到目标节点,四个应用程序依次占用同一 ATM 交换机端口的虚连接 VC1~VC4,该端口尚有 48Mbps 带宽可用。容易计算,四个用户的 MCR 依次是 20000, 20000, 10000, 10000 信元/秒,分别等于 8, 8, 4, 4Mbps。那么,按照上述第一种“公平”的定义,四个 VC 分别获得 14, 14, 10, 10Mbps 的带宽,而按照第二种定义,四个 VC 各为 12Mbps。两种分配方法各有各的道理。进一步,如果此时本 ATM 端口有 48Mbps 可用于 ABR,VC1~VC4 的 MCR 均为 0,但由于上游网络严重阻塞,VC1 与 VC2 都只有 5Mbps 速率的数据到达,这时最合理的分配方法显然是 VC1、VC2 各 5Mbps,VC3、VC4 各 19Mbps,而不是等分为四个 12Mbps,或者 14, 14, 10, 10Mbps。可见,ABR 服务带宽分配的公平性定义亟需统一,以保证不同 ATM 交换机间的互操作性。目前 ATM Forum 的标准文本中尚未给出这个定义。

• 38 •

2 死锁

ATM 适配层(AAL)的设计一般采用面向包数据的 AAL5 规范。当网络上的一个节点用尽了信元缓冲区,并且还在等待一个 AAL5 帧余下的部分时,就可能发生死锁。在所有端口共用缓冲池的情况下,死锁发生的概率极小,但是如果资源管理采用的是每个端口有独立缓冲区的策略,则死锁发生的概率就大大增加,特别是当缓冲区容量较小时。必须仔细地设计反馈控制策略,必要时以牺牲缓存(因而可能牺牲传输速率)的代价保证信元的可靠传输。比如,在早期分组丢弃(Early Packet Discard)拥塞控制算法中,适当降低告警阈值,可以有效地保持 AAL5 到达分组的完整性^[3]。基于逐连接(参见下面“3.1 基于信用的流量控制”)方式的反馈控制,针对死锁问题,已有一些成果^[4-7],但它们多数存在公平性的缺陷。

3 ABR 的流量控制

主要有两种 ABR 流量控制方案:基于信用的流量控制和基于速率的流量控制,它们都是闭环反馈型。每个虚连接的物理路由不同,所以每个虚连接对应一个单独的控制环。概括地说,基于信用的方案的核心,是控制环中的 ATM 交换机与源端机交换端口空闲缓存(即信用)的信息;基于速率的方案的核心则是将信元沿途的拥塞状况反馈给源端,以使其提高或降低发送速率。虽然基于速率的流控方案最终战胜基于信用的方案被 ATM Forum 采纳为行业标准,但后者仍然有其优势,仍受到一些 ATM 交换机厂商的青睐,如著名的 DEC GIGAswitch 的流控算法 FLOWmaster 便是典型的基于信用的方案。笔者认为,将两种流控方案结合起来,根据特定的网络环境采用相应的集成方式,发挥两种流控方案各自的长处,是一个最佳的途径。

3.1 基于信用的流量控制 是一种逐连接(link-by-link)的窗口流控方式,见图1。网络的每个环节都独立地运行面向缓存(信用)的流控机制,每个环节的接收端都为各个虚连接预留了缓冲区,该缓冲区的大小必须满足来回一次所需的信元缓存量,这个缓存量由传播时延、交换时延和发送速率三者决定。在广域网(WAN)环境中,传播时延很大,需要支持逐连接信用反馈的缓存器就非常庞大,因而这种流量控制方式不适用于 WAN。



图1 基于信用的流量控制

在局域网(LAN)和城域网(MAN)环境中,由于基于信用的流量控制采用逐连接方式的反馈控制,与采用端对端方式的基于速率的流量控制策略(见下面3.2节)相比,它的一个最吸引人的特点就是快速。随着超高速主机(如800Mbps的HIPPI接口^[3])的出现以及具有高突发特性的声、像多媒体数据业务的需要,提供快速的反馈机制对减少信元传输时延方差、避免信元丢失具有重要意义。

下面,在给出有关的符号说明之后,我们将用伪PASCAL语言对最基本的基于信用的流量控制算法进行形式的描述。

假定共有 N 个 VC, 缓冲区总量 $TotalCells$ 个信元,从逻辑上分为各 VC 公用和每个 VC 私用两种缓冲池,每个 VC 私用缓存设定为 Min 个信元,故公用区为 $TotalCells - N * Min$ 信元。记 Max 为缓冲两个 ATM 节点间来回一次所需的最少信元数, $MAX = 传播时延 * 发送速率 * 2$ 。在发送方,每个 $VC_i (1 \leq i \leq N)$ 对应一个变量 $sent_i$, 记录该 VC 可以拥有的“未担保信元”(已经发出但尚未收到 credit 的信元)限额, $sent_i = 已经发出的信元数 - VC_i$ 的信用数,并且令 $sent = \sum sent_i$ 。在任一时刻,两节点间的链路(或者说发送方)处于 congested 和 uncongested 两个状态之一,如果是 congested 状态,每个 VC 只能有 Min 个未担保信元;如果是 uncongested 状态,每个 VC 可以有 Max 个未担保信元。在当前总的可用缓存小于私用缓存之和时,即 $TotalCells - sent < N * Min$ 时,链路进入 congested 状态。定义上限 $UT = TotalCells - N * Min$, 上述条件为 $sent > UT$ 。考虑到信元滞后现象,定义下限 $LT \leq UT$, 当 $sent < LT$ 时链路进入 uncongested 状态。信元和信用的发送使得链路状态在 congested 和 uncongested 之间转换。引入二元整型变量 $mode$ 表示链路当前状态, $mode = 0$ 为 uncongested, $mode = 1$ 为 congested。定义数组 $limit[0..1]$, $limit[0] = Max$, $limit[1] =$

Min 。在状态为 $mode$ 时, VC_i 可以发送信元当且仅当 $sent_i < limit[mode]$ 。

可以认为发送方和接收方由两条单向的 FIFO 连接,协议的全局状态由发送方和接收方两者的状态以及两条 FIFO 上的信息序列组成。遵循前面的变量说明,我们运用 N. A. Lynch 和 M. R. Tuttle 的 I/O 自动机模型^[10],用伪 PASCAL 语言进行算法描述。当满足给定的条件(preconditions)时,动作执行(如发送一个信元或信用),动作完成后有关节点更新其状态(effects)。

Initial State, $sent_i = 0, i = 1, 2, \dots, N$ $mode = uncongested$
 $SEND_i(c)$ (* 发送方发送 VC_i 上的信元 c *)

Preconditions,
 $sent_i < limit[mode]$

Effects,
 $sent_i = sent_i + 1$
 $sent = sent + 1$

if $sent > UT$ then $mode = congested$

$RECEIVE_i(credit)$ (* 发送方接收 VC_i 的信用 *)

Effects,
 $sent_i = sent_i - 1$
 $sent = sent - 1$

if $sent < LT$ then $mode = uncongested$

$RECEIVE_i(c)$ (* 接收方接收 VC_i 的信元 c *)

Effects,

Add c to any free cell buffer

$SEND_i(credit)$ (* 接收方发送 VC_i 的信用 *)

Preconditions,

VC_i cell c dequeued from receiver buffer

Effects,

Mark buffer containing c as free

容易证明以下三个引理:

引理1 $\sum d(i) \geq sent - UT$, 其中 $d(i) = \min\{sent_i, Min\}, i = 1, 2, \dots, N$ 。

引理2 $sent \leq UT + N * Min = TotalCells$ 。

引理3 $sent = M + Q + C$, 其中 M 为正在发往接收方的信元数, Q 为接收方缓存中的信元数, C 为接收方积存和正在发送的信用数

进而可以证明:

定理1 算法不会产生信元丢失。

定理2 如果 VC_i 的发送方有一个信元待发送,它将在一有限的时间内被发送。

定理2阐明了,系统不会由于本算法产生死锁。但链路错误仍有可能产生信元和信用丢失(如 FDDI/SONET 的位差错率是 10^{-9}),使得 VC 速率降低,极端情况就是发生死锁。为此,需要为每个 VC 建立一种同步机制,以避免由线路物理条件引起的随机位差错导致系统死锁。

出于简化概念的目的,我们对链路状态只规定了 congested 和 uncongested 两个值,通过使 $mode$ 的取值超出 $\{0, 1\}$ 范围,可以建立比我们的 on/off 更加精细的速率等级控制模型。

3.2 基于速率的流量控制 是以端到端(end-to-end)的反馈方式为基础的。端系统和网络使用一种特殊的专用信元—资源管理信元(RM-cell)来运载流量控制信息,以固定时间间隔或者信元间隔向信源告知网络拥塞情况。ATM Forum 规定的 RM-cell 净荷格式分为三个域:

- ER(Explicit Rate) 由源端初始化为最大速率,当网络交通情况要求源端降低发送速率时,远端系统将该域改写为一个合适的低数值。

- ACR(Advised Cell Rate) 源端把这个域按照正在发送的速率设定。

- CI(Congestion Indication) CI=0表示网路没有阻塞现象,接收方端系统可以将 CI 改写为1,源端在收到 CI=1的 RM-cell 后就会降低发送速率。

基于速率的流量控制方法有三种,前向显式阻塞告知(FECN),显式的速率建议(ER),和后向显式阻塞告知(BECN)。

FECN:按照 ITU 建议 I.371,利用每个 ATM 信元头部 Payload Type Identification 域中的一位作为“显式的前向阻塞标识”(EFCI)。如果在从源端到目的端的通路上某个虚连接发生阻塞,此 VC 上所有数据信元的 EFCI 位就被置为1,目的端交换机看到到达的数据信元 EFCI=1,就把下一个将要发送给源端交换机的 RM-cell 的 CI 位置为1,信源立即降低该 VC 的发送速率。由于要经过前向信息搜集-后向信息反馈的过程,而且信源要逐步调节达到最优的发送速率,所以 FECN 存在着速率调节太慢的缺陷,这在 WAN 环境中更加突出。

ER,显式的速率建议方法则由目的端交换机根据各数据信元的 EFCI 位,把一个最佳速率建议值显式地写入 RM-cell 头部的 ER 域,从而使得信源的速率调节可以一步到位,比 FECN 逐步调整方式的反应速度大大加快,但是它也要经过前向信息搜集-后向信息反馈的过程。

BECN:针对 FECN 和 ER 的弱点,一个改进方法是发生信元阻塞的交换机直接把该 VC 已阻塞的信息通过 RM-cell 的 CI 位反向通知信源,而不再进行前向的信息传递。这提高了信源发送速率调节的灵敏度,但是又额外产生了大量的反向信元,加重了网络负担,模拟结果表明^[1],100个满负荷发送的信元产生的反向 RM-cell 数量为前向数据信元的9%。

另一种改进的方法是,仍采用 FECN 或 ER,但是依靠源端机至目的端机通路中的 ATM 交换机把

端到端的控制环分成若个子环,中间子环始于虚源(Virtual Source, VS),止于虚目(Virtual Destination, VD),一个 VD 同时又是下一个子环的 VS,子环分割的极端情况就成为形式上与基于信用的流控相同的逐连接的方式。

上述几种基于速率的控制方法各有优劣,比如基本的 FECN 对交换机复杂度的要求最低,因而适合于工作组或 LAN 应用;引入虚源/虚目的子环分割方法对 WAN 来说是一种比较合适的方案。

我们前面采用的 I/O 自动机模型可以很方便地扩展到描述基于速率的流控方法,限于篇幅,此处不再赘述。

3.3 两种流控方案的结合 基于信用的流量控制和基于速率的流量控制两种方案各有所长,目前不同的 ATM 交换机生产厂商也分别采用不同的流控方案。我们认为,针对特定的应用环境,在网络的不同部分采用不同的流控方法,可以达到最佳的控制效果。考虑到基于信用和基于速率两种方法各自的特点,我们提出,LAN 部分的流控采用基于信用方式,WAN 部分采用基于速率的方式,见图2。LAN/WAN 接口处的交换机除了转发数据之外,应具备两种功能:作为基于信用的流控段的虚目,向 LAN 一侧回送信用信息;作为基于速率的流控段的虚源,接受 WAN 一侧的速率控制,按照合适的速率向 WAN 发送信元。

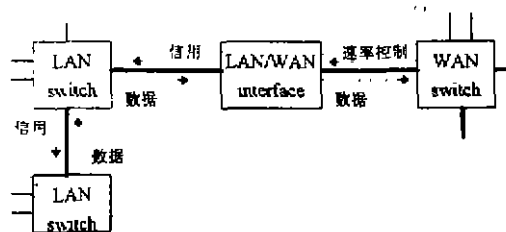


图2 基于信用与基于速率的流控的集成方案

结束语 以上我们主要讨论了 ATM ABR 服务的流量控制。需要指出,虽然采用了严格的流控算法,但是由于 ATM 是基于统计复用的,也由于价格等方面的考虑,ATM 交换机端口的缓存往往小于理论计算所要求的数量,所以信元溢出是可能的,这时要进行信元取舍,有关算法有 Early Packet Discard, Partial Packet Discard 等^[1]。(参考文献共14篇略)