ATM网络基于队列长度阈值的传输调度*

林闯

(国家信息中心信息经济与技术研究所 北京 100045)

摘要 本文提出了ATM网络的一种实时传输调度和信元丢失控制的综合方案.这种方案是基于队列长度阈值而设计的,它适应于ATM网络面向连接的特性.本文给出了这种方案的随机Petri网性能模型,并给出模型分解和迭代的近似求解方法.

关键词 ATM网络,实时传输调度,信元丢失控制,随机Petri网模型,性能分析.

中图法分类号 TP393

基于异步传输模式(ATM)的宽带网络设计支持各种类型信息的传输,例如视频、声音和数据等的传输.ATM网络要保证用户要求的服务质量(QOS).不同用户可能有不同的传输要求.例如,视频和声音的传输有实时的要求,超时的信息不能使用,同时可以容忍某种程度的信息丢失;而数据的传输则不容许信息的丢失,但传输的延时则不成问题.在目前的很多应用中,一个用户传输要求中包括多种类型信息的一同传输,既有视频或声音信息也有数据信息.在同一个用户传输连接中,这些不同类型的信息有相同的实时传输要求.因此,要保证信息传输的实时性和丢失的综合要求是ATM网络传输控制的一个重要问题.

现在已建立了一些传输调度方案,例如:最简单的先入先出方案FIFO(first-in-first-out),没有考虑网络性能的要求.静态优先级SP(static priority)方案不能充分地利用网络资源,而降低了网络的吞吐量.最早到时首先发送(Earliest-deadline-first)方案在不考虑信息丢失优先级情况下是一个最佳方案,但没有考虑信息丢失控制要求.在已有的信元(Cell)丢失优先级控制方案中,例如最简单的分路(Separate Route)方案,部分缓冲共享PBS(partial buffer sharing)方案和复杂控制的推出(Push-out)方案[1],仅考虑了信元丢失优先级的控制,而不能保证信元实时传输的调度要求.Chiopalkatti等人的队列长度阈值QLT(queue length threshold)调度方案[2]给出了在性能和复杂性方面的最好折衷,但是这个方案没有考虑同一用户呼叫中可能包含着不同类型的信息,它们有不相同的丢失优先级,但它们有相同的实时传输要求,亦即,它们要求同步.因此,在这个方案中,由于同一个用户呼叫的信元可能放入不同实时优先级队列而改变发送次序,一个重新排序的设置必须加入到ATM的适配层,从而破坏了ATM面向连接的特性.Ling等人的实时调度方案[3]给出信元加权丢失率最小的算法,但是这个方案实现复杂,而且也同样不能保证同一个呼叫中的信元能够得到FIFO的发送服务.

本文建议一个基于队列长度阈值的实时传输调度和信元丢失控制的综合方案.这个新方案能保证同一个用户呼叫中的信元能够得到FIFO的发送服务,而且具有实现简单等优点.我们将提供这种方案的随机Petri网模型,并且提供这种模型的分析方法.

1 建议方案的描述

我们将研究在ATM网络一个节点上实时传输的调度问题,这个节点包括一个发送服务器和多个缓冲队列,如图1所示.每一个用户呼叫(也可以是用户连接,在不混淆的情况下,本文将二者看成是相同的)或几个具有相同实时传送要求的用户呼叫使用一个缓冲队列,这一点同QLT方案不同.在QLT中按信元的优先级分别进入不同队列,而没有考虑同一个呼叫信元发送的次序.在我们的方案中,按呼叫和传输实时要求区分队列,从而可以保证呼叫的FIFO服务次序.另外,同一呼叫中不同优先级信元的丢失控制采用部分缓冲共享(PBS)方案,在同一缓冲队列中为不同丢失优先级信元设置不同的控制缓冲阈值.任一缓冲队列的最大空间为B_i,队列是从头向尾排序,它是一个FIFO队列

---

* 本文研究得到国家自然科学基金和国家科技攻关项目基金资助.作者林闯,1948年生,博士,研究员,主要研究领域为ATM网络,系统性能评价和Petri网理论和应用.

本文通讯联系人:林闯,北京 100045,国家信息中心信息经济与技术研究所

本文1997-03-26收到原稿,1997-05-16收到修改稿

(如图1所示).在ATM网络中,每一个信元发送服务时间是确定的,系统可以想象为分割成时间片段(Slot),每个时间片是一个ATM信元的发送时间.

在模型中用下列变量参数来描述传输的调度和信元丢弃的控制:

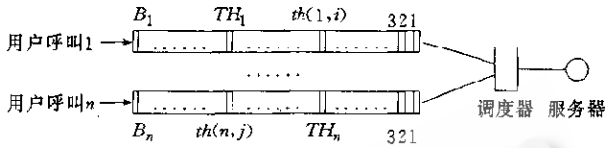

图1　调度方案的描述模型

TH_i:表示用户呼叫i的实时调度缓冲阈值.当用户呼叫i的信元在它自己的缓冲队列的占有值达到或超过TH_i时,并且更高实时优先级的用户呼叫的队列缓冲占有值都小于它们自己的调度缓冲阈值,那么将发送权调度给用户呼叫i的信元缓冲队列.一般情况下,实时优先级越高的呼叫,其调度缓冲阈值越小,反之亦然.

$th(i,j)$:表示用户呼叫i的丢弃优先级为j的信元丢弃缓冲阈值.当用户呼叫i具有丢弃优先级为j的信元到达缓冲队列时,如果缓冲队列的长度小于$th(i,j)$,则信元进入缓冲队列,否则,此到达信元被丢弃.丢弃优先级j越大,$th(i,j)$值就越大,信元丢失率就越小,$th(i,j)$取值范围为$1\leqslant th(i,j)\leqslant B_i$.

在这个方案中,信元丢弃不需要缓冲队列的扫描和到达信元的重新排序等复杂操作,因此方案实现简单.这个方案可以看作是QLT和PBS方案的结合,保持两者的优点,用户呼叫缓冲队列可以根据用户要求动态分配.由于要保持同一用户呼叫中信元的发送次序,而在不同实时传送要求的用户呼叫之间不能共享使用缓冲,有可能损失部分缓冲资源的利用率.

2　方案的随机Petri网模型

我们假设读者对随机Petri网(SPN)的理论和应用有一些基本的了解,有关详细的SPN描述可参阅文献[4,5].在我们的SPN模型中,信元的产生和发送可由时间变迁来表示;信元的丢弃和进入缓冲队列可由瞬时变迁表示,它们不占用时间;缓冲队列可由位置来表示,它们占有程度可由位置的标识(Marking)表示.另外,在模型中,允许变迁有实施的优先级,当几个变迁同时可实施时,优先级高的实施,优先级低的则不能实施,相同优先级的变迁则都有实施可能性.[5]瞬时变迁比时间变迁有更高的实施优先级.在模型中,允许变迁的实施条件用变迁的谓词规定,当变迁谓词条件不能满足时,变迁不能实施.模型中的标记(Token)可以着色区分,其着色变量可以表现在模型的弧上和变迁的谓词中.[6]

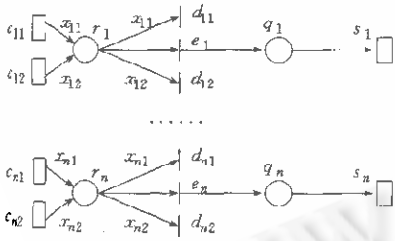

图2　调度方案的SPN模型

在图2中,我们给出了传输调度方案的SPN模型.为了便于说明,我们假定用户呼叫实时传输的优先级按缓冲队列的次序反向排列,亦即,用户呼叫i的实时传输的优先级高于用户呼叫$i+1$的优先级,用户呼叫1的实时传输的优先级最高,用户呼叫n的实时传输的优先级最低.在模型中,字母的下标的第1位表示实时传输的优先级,第2位表示信元丢弃的优先级,仅一位的下标表示实时传输的优先级.为了简洁,在模型中,仅描述了呼叫1和呼叫n的子模型,其它的子模型相同.每一个呼叫中,信元丢弃优先级可以是仅一级或多级,模型中仅描述为两级.

模型中的变迁和位置的描述含义如下:

c_{ij}:表示用户呼叫i的丢弃优先级j信元的产生,其产生速率是指数分布的,其值为λ_{ij}.在同一用户呼叫i中,具有优先级j信元的产生比例为$\frac{\lambda_{ij}}{\sum_j \lambda_{ij}}$.

d_{ij}:表示用户i丢弃优先级j信元的丢弃.

e_i:表示用户i到达的信元进入缓冲队列q_i.

s_i:表示用户i缓冲队列中的信元的发送服务,其服务速率为μ_i.所有的$\mu_i(1\leqslant i\leqslant n)$都有相同值.

r_i:作为用户i到达信元的临时保留场所,在此位置信元接受判断,以决定是否丢弃.其容量仅为1.

q_i:作为用户i到达信元的缓冲队列,其容量为B_i.

在模型中变迁实施的优先级规定如下:

d_{ij}的实施优先级为2;e_i的实施优先级为1;c_{ij}和s_i的实施优先级同为0.

在模型中变迁相关联的谓词描述如下:

d_{ij}的谓词为 $M(q_i)\geqslant th(i,j)$，如果 $t<k$，则有 $1\leqslant th(i,t)<th(i,k)\leqslant B_i$. 对于用户 i 的最大丢弃优先级 m，则有 $th(i,m)=B_i$.

s_i 的谓词为

$$[(M(q_i)\geqslant TH_i)\cap(\forall\ j,1\leqslant j<i,M(q_j)<TH_j)]\cup[\forall\ j,1\leqslant j<i,M(q_j)=0]$$

对于实时传输优先级最高的用户 1，s_1 的谓词则有

$$(M(q_1)\geqslant TH_1)\cup(\forall\ j,1<j\leqslant n,M(q_j)<TH_j)$$

在其他方案中，对于信元丢弃控制，我们采用的是静态 SPB 方案，$th(i,j)$值是静态规定的，在系统进行期间不变化. 为了提高系统资源利用率和传输效率，可采用动态 SPB 方案，使 $th(i,j)$值随着信元丢失率的变化而变化. 本文不讨论这个问题.

对于图 2 的模型，可以进一步精化设计，使之更便于系统性能分析，图 3 的模型是图 2 模型的等价模型. 在图 3 的模型中删除了所有的瞬时变迁，将 d_{ij}的谓词相应地描述在 c_{ij}的谓词中，其余的不变. c_{ij}的谓词为 $M(q_i)<th(i,j),1\leqslant th(i,j)\leqslant B_i$.

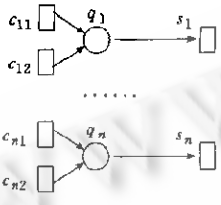

图3 调度方案的精化SPN模型

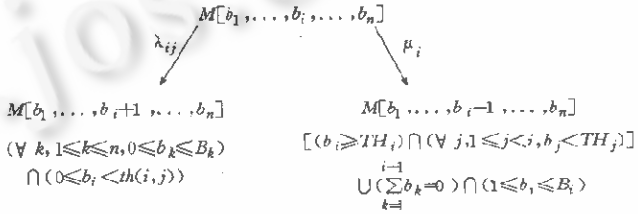

图4 图3模型的部分马尔可夫链

3 性能分析

对图 3 模型的性能分析可有两种方法：整个模型的直接求解和对模型分解、迭代近似求解.

在已有的直接求解方法中，根据模型可以构造其对应的马尔可夫链. 缓冲队列 q_i 的标识使用变量 b_i 表示. 模型的一个实存状态(标识)一般可由 $M[b_1,\ldots,b_i,\ldots,b_n]$表示. 图 4 给出图 3 模型马尔可夫链中任一实存状态和实存状态之间的转换及其条件.

基于上述马尔可夫链和状态转换速率，我们能构造状态转换矩阵并获得所有状态的稳定状态概率，从而我们可以讨论模型系统的性能参数. 状态 M 的稳定状态概率使用 $P[M]$表示.

用户呼叫 i 的丢弃优先级为 j 信元的丢失率 L_{ij}可以表达为

$$L_{ij}=P[M(q_i)\geqslant th(i,j)] \tag{1}$$

用户呼叫 i 缓冲队列 q_i 的平均利用率 $U(q_i)$可以表达为

$$U(q_i)=\frac{\sum_{k=1}^{B_i}k\times P[M(q_i)=k]}{B_i} \tag{2}$$

进一步整个系统的缓冲队列的平均利用率 $U(Q)$可以表达为

$$U(Q)=\frac{\sum_{i=1}^{n}\sum_{k=1}^{B_i}k\times P[M(q_i)=k]}{\sum_{i=1}^{n}B_i} \tag{3}$$

用户呼叫 i 的信元发送的吞吐量 $T(s_i)$可以表达为

$$T(s_i)=\mu_i\times\sum_{M\in E}P[M] \tag{4}$$

其中 E 是使 s_i 可实施的所有可达标识集合，实施条件如图 4 所示.

同样地，进一步可利用式(4)表达整个系统的吞吐量 $T(S)$

$$T(S)=\sum_{i=1}^{n}T(s_i) \tag{5}$$

当模型规模较小时，可以考虑采用上述直接求解方法. 一般情况下，上述模型是一个 n 维的马尔可夫链，随着 B_i

和 n 的增大,系统的状态呈指数增长.当状态数量超过一定限制后,当前的一般计算机的存储和计算能力无法忍受,而使问题成为不可实际求解.

对复杂、多维的马尔可夫链求解是一个具有挑战性的困难问题.一种可能的解决方法是基于模型分解和迭代求解子模型之间相互关系的近似求解方法,正如在文献[6]中所采用的方法一样,本文的模型也可采用类似的方法进行求解.在一般情况下,完成此方法需要如下的步骤:

(1) 模型精化设计:利用 SPN 模型的变迁实施谓词和变迁实施速率函数去简化 SPN 模型的结构和暴露子模型的独立性.

(2) 模型的分解:将给定的 SPN 模型 A,分解成 n 个子模型 $A_1,A_2,\ldots,A_n$.

(3) 确立子模型之间的输入和输出:对于每个子模型 A_i,设置它来自于 $A_j(1\leqslant j\leqslant n)$的输入参数.当 A_i 求解后,规定它的输出参数.

(4) 迭代求解:根据输入和输出参数之间的关系,确定子模型的求解次序.每个子模型被求解后,更多迭代需要执行,每次都使用最新的结果做为迭代执行的输入参数.当所有输入参数聚集到达,迭代停止.

在我们的实时调度模型中,图 3 的模型已经精化设计,每个结构独立部分就是一个子模型.子模型之间的相互影响的关系,亦即子模型之间输入和输出参数关系描述在变迁 s_i 的实施条件谓词中.每个变迁 s_i 可实施与否,不但与本子模型队列状态相关,而且与其它子模型队列状态相关.在模型中,队列状态(标识)不能直接做子模型间的输入、输出参数,它是时间函数,但是稳定状态下标识概率可作为输入、输出参数.每一个子模型 A_i 以其它所有子模型 $A_j(1\leqslant j<i)$的稳定状态下队列占有分布概率作为输入参数.在子模型 A_j 中,缓冲队列占有分布概率表达为

$$\forall\ r,0\leqslant r\leqslant B_j,G_j(r)=P[M(q_j)=r] \tag{6}$$

让我们表示子模型 A_j 中,缓冲队列占有小于 TH_j 的概率

$$F_j(TH_j)=\sum_{r=0}^{TH_j-1}G_j(r) \tag{7}$$

子系统的缓冲队列竞争发送的影响可以表示为每个缓冲队列发送速率的降低.对于子系统 $A_i(2\leqslant i\leqslant n)$,变迁 s_i 的实施速率可以表达为

$$\begin{cases}\mu_i\times\prod_{j=1}^{i-1}G_j(0) & M(q_i)<TH_i\\ \mu_i\times\prod_{j=1}^{i-1}F_j(TH_j) & M(q_i)\geqslant TH_i\end{cases} \tag{8}$$

对于子系统 A_1,变迁 s_1 的实施速率可以表达为

$$\begin{cases}\mu_1\times\prod_{j=2}^{n}F_j(TH_j) & M(q_1)<TH_1\\ \mu_1 & M(q_1)\geqslant TH_1\end{cases} \tag{9}$$

式(8)(9)表达了输入参数的对子系统 A_i 行为的影响.在迭代中,可按 $A_1,A_2,\ldots,A_n$ 子系统顺序求解.在初始迭代中,首先求解子系统 A_1,对所有 $F_j(TH_j)$的初值可以设置为 $0<F_j(TH_j)<1$ 之间的任一值.

每个子系统的性能参数的求解,同样可以根据公式(1)~(3)和(5)进行计算,但是公式(4)中的 μ_i 值要根据公式(8)和(9)确定,而且在不同的标识下,可能有不同的值.

限于篇幅,本文不再显示传输方案性能的具体数据例子.

4 结　论

本文描述了一种基于缓冲队列长度阈值进行实时调度和信元丢失控制相结合的方案.这种方案可以保证同一用户连接信元的 FIFO 服务特性,而且具有简单和易实现的特点.本文给出了这种方案的 SPN 模型以及这个模型的求解方法.SPN 模型的求解基本上有两种方法:直接整体模型的求解和模型的分解、迭代近似求解.后一种方法是更重要的,它可以求解实际系统的模型.本文所采用的调度模型的分解和迭代求解方法具有普遍意义,可以应用到多种系统的性能模型中.

其他的实时调度和信元丢失控制的统一结合方案及其模型和性能分析方法是作者进一步研究的课题.

参考文献

1　Kroner H, Hebuterne, Boyer P *et al*. Priority management in ATM switching nodes. IEEE Journal on Selected Areas in

Communications, 1991,9(3):418~427
2 Chlopalkatti R, Kurose J F, Towsley D. Scheduling policies for real-time and non-real-time traffic in a statistical multiplexer. In: Proceedings of the IEEE INFOCOM'89. Ottawa, Canada: IEEE Computer Society Press, April 1989. 774~783
3 Ling T L, Shroff N. Scheduling real-time traffic in ATM networks. In: Proceedings of the IEEE INFOCOM'96. San Francisco, California, USA: IEEE Computer Society Press, March 1996. 198~205
4 Lin C, Marinescu D C. Stochastic high-level Petri nets and applications. IEEE Transactions on Computers, 1988,37(7):815~825
5 Ciaodo G, Muppala J, Trivedi K S. SPNP: stochastic Petri net package. In: Proceedings of the Petri nets and Performance Models. Kyoto, Japan: IEEE Computer Society Press, December 1989. 142~151
6 林闯.一种资源共享系统的模型和近似性能分析.计算机学报,1997,20(10):865~871
(Lin Chuang. A model of systems with shared resources and analysis of approximate performance. Chinese Journal of Computers, 1997,20(10):865~871)

Traffic Scheduling Based on the Queue Length Threshold in ATM Networks

LIN Chuang

(*Institute of Information Economics and Technology State Information Center Beijing* 100045)

Abstract In this paper, a new integrated scheme for scheduling real-time traffic and cell loss control based on the queue length threshold is proposed and modeled using stochastic Petri nets. The connection-oriented nature of ATM networks can be kept and no resequencing device is needed in this new scheme. This large model of the scheme is decomposed into near-independent submodels to evaluate the model performance.

Key words ATM networks, scheduling real-time traffic, cell loss control, stochastic Petri nets, performance analysis.