

通用分组无线业务中基于测量的接纳控制算法^{*}刘衍珩¹, 苑森森¹, 孙惠平²¹(吉林大学 计算机科学与技术学院, 吉林 长春 130012);²(北京邮电大学 电子工程系, 北京 100876)

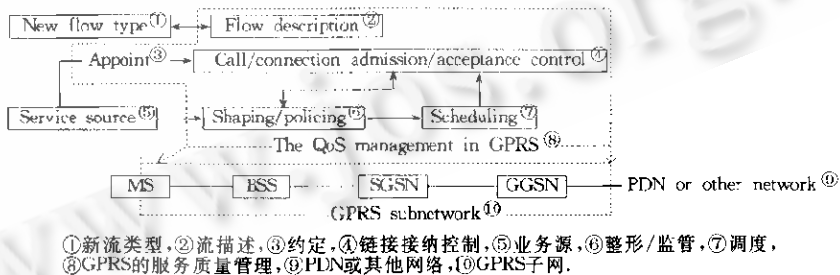
E-mail: lyh_lb_lk@yahoo.com.cn

摘要: GPRS(general packet radio service)通过使用类似于 ATM 中采用的 QoS 管理算法来为用户提供预留服务. 许多学者的研究表明: 基于测量的链接接纳控制算法与其他链接接纳控制算法相比, 对 GPRS 网络具有更好的适应性和更高的效率. 利用优先级队列的最坏延迟计算模型、等效漏桶近似和延迟标准化, 在充分分析 GPRS 特性的基础上, 提出了一种适合于 GPRS 的基于测量的接纳控制算法. 通过仿真发现, 该算法能在满足 GPRS 的 QoS 要求的同时获得 80% 的资源利用率.

关键词: GPRS(general packet radio service); 预留服务; QoS 管理机制; 链接接纳控制算法; 基于测量的链接接纳控制

中图法分类号: TP393 文献标识码: A

通用分组无线业务(general packet radio service, 简称 GPRS)^[1~3]是在 GSM 网上开发的针对突发数据分组传输的一种新业务, 是 GSM 向第三代移动通信系统过渡的一种可提供数据和多媒体业务的方案. 为了给网络用户提供端到端的分组传输模式, GPRS 在原有 GSM 系统的基础上增加了两个关键的网络部件: GPRS 网关支持节点(GGSN)和服务支持节点(SGSN). 为了在维持较高资源利用率的基础上满足不同应用的 QoS 要求, 就必须对网络进行相应的 QoS 管理. GPRS 网络中的 QoS 管理机制如图 1 所示. 其中, CAC 算法决定网络是否有足够的资源, 满足链接的 QoS 要求是 QoS 管理机制中的重要部分.



①新流类型, ②流描述, ③约定, ④链接接纳控制, ⑤业务源, ⑥整形/监管, ⑦调度, ⑧GPRS的服务质量管理, ⑨PDN或其他网络, ⑩GPRS子网.

Fig. 1 The QoS management mechanism of GPRS
图1 GPRS中的QoS管理机制

基于测量的接纳控制算法(MBAC)^[4~6]不仅仅依赖于用户的资源要求描述, 而且通过对系统的实时测量决定是否接纳一个新链接, 因而对不断变化的网络环境具有更好的适应性. 本文利用优

* 收稿日期: 2000-05-08; 修改日期: 2000-10-23

基金项目: 国家自然科学基金资助项目(69873013); 吉林省自然科学基金资助项目(20000545)

作者简介: 刘衍珩(1958—), 男, 吉林松原人, 教授, 主要研究领域为计算机网络与通信, 多媒体技术; 苑森森(1939—), 男, 河北石家庄人, 教授, 博士生导师, 主要研究领域为数据库, 智能网络系统; 孙惠平(1975—), 男, 江苏徐州人, 博士生, 主要研究领域为计算机网络与通信.

先级队列的最坏延迟计算模型、等效漏桶近似和延迟标准化,提出了一种适合于 GPRS 的新的 MBAC 算法,并通过 3 组仿真实验验证表明,我们的算法在 GPRS 网络上有很好的执行效果。

1 模型假设与延迟标准化

我们首先给出研究的基本模型,本文仅关注单个可以提供 GPRS 服务的小区,不考虑小区之间的切换,小区内分组交换的数据服务可使用 8 个信道:信道 0 作为主信道,其余 7 个作为从信道,MS 可以根据需要使用多个信道,编码使用 1/2 速率,每一个信道的速率为 9.05Kbits,小区的带宽为 72.4Kbits. 端到端延迟是指从进入小区到离开小区的时间差,在 Um(空中接口)与 Gi(GGSN 和外部 PDN 之间)之间测量得到。

我们使用漏桶模型^[7,8]来描述每一个预留服务流,每一个漏桶均有两个参数:令牌产生速率和漏桶深度 (r^p, b^p) . 对于每一类预留服务,小区均保留带宽的叠加使用量 \hat{v}_i 、队列的实际延迟 \hat{D}_i 和预先定义的延迟约束 D_i . 调度原则采用基于延迟要求的优先级分配调度算法^[9].

1.1 延迟计算模型

具有多个不同优先级的系统,假设源的峰值速率无限大,每一类中的服务原则是先来先服务,已证明^[10]类 j 的最坏延迟为

$$D_j^* = \frac{\sum_{i=1}^j b_i}{\mu - \sum_{i=1}^j r_i} \tag{1}$$

现在,我们利用式(1)作为基本等式来计算接入一个新流对原有预留服务的影响,定义 D_k^* 为计算得到的延迟,把同一类的全部流近似为一个符合漏桶参数 (v_i, b_i) 的流,在接入一个漏桶参数为 (r_k^a, b_k^a) 的新流后,这类漏桶参数可认为是 $(v_i + r_k^a, b_i + b_k^a)$. 可以得到,接入一个新流对同一优先级流的影响,表示为

$$D_k^{*'} = \frac{\sum_{i=1}^{k-1} b_i}{\mu - \sum_{i=1}^{k-1} v_i} - \frac{b_k + b_k^a}{\mu - \sum_{i=1}^{k-1} v_i} = D_k^* + \frac{b_k^a}{\mu - \sum_{i=1}^{k-1} v_i} \tag{2}$$

接入一个新流对低优先级流的影响,可表示为

$$D_j^{*'} = \frac{\sum_{i=1}^{k-1} b_i + b_k + b_k^a + \sum_{i=k+1}^j b_i}{\mu - \sum_{i=1}^{k-1} v_i - v_k - r_k^a - \sum_{i=k+1}^{j-1} v_i} = D_j^* \frac{\mu - \sum_{i=1}^{j-1} v_i}{\mu - \sum_{i=1}^{j-1} v_i - r_k^a} + \frac{b_k^a}{\mu - \sum_{i=1}^{j-1} v_i - r_k^a} \tag{3}$$

1.2 延迟标准化

因为 GPRS 中分组是变长的,而分组的延迟又与分组的长度相关,所以,不同长度的分组的延迟之间不具有可比性和可加性,为了使测量时各服务类中不同长度的分组的延迟具有可比性和可加性,必须把延迟标准化(见表 1),具体方法如下:

Table 1 The delay requirement classes
表 1 延迟要求类

Service classes ^①		Predictive QoS ^②			Best-Effort transmission ^③
Delay classes ^④		1	2	3	4
Size of individual group ^⑤	128 Octet	Average ^⑥	0.5	5	50
		95%	1.5	2.5	150
	1 024 Octet	Average	2	15	75
		95%	7	75	375

①服务类,②预留服务,③尽力传送,④延迟类,⑤分组尺寸,⑥平均,⑦没有描述.

由表 1 可得:

$$D_{c,l} = \begin{cases} \frac{0.5 \times l}{128}, & c=1, l \leq 128, \\ 0.5 + \frac{1.5 \times l}{1024 - 128}, & c=1, l > 128, \\ \frac{5 \times l}{128}, & c=2, l \leq 128, \\ 5 + \frac{10 \times l}{1024 - 128}, & c=2, l > 128, \\ \frac{50 \times l}{128}, & c=3, l \leq 128, \\ 50 + \frac{25 \times l}{1024 - 128}, & c=3, l > 128. \end{cases} \quad (4)$$

在式(4)中 c 为分组的服务类; l 为分组的长度; $D_{c,l}$ 为服务类为 c , 长度为 l 的分组的延迟要求. 把 $D_{c,l}$ 作为一个标准, 所有的分组的测量延迟均与 $D_{c,l}$ 相比, 比值定义为标准化延迟. 定义测量的实际延迟为 D , 标准化延迟为 D^a , 则可知:

$$D^a = \frac{D}{D_{c,l}}. \quad (5)$$

每一类平均的标准化延迟为

$$\bar{D}_i^a = \frac{\sum_{i=1}^N D_i^a}{N}, \quad 1 \leq i \leq 3. \quad (6)$$

2 算法描述

2.1 基本接纳控制规则

(1) 如果当前使用带宽加上新流要求的带宽超过系统的可用带宽, 则拒绝流 p 的申请.

$$vu > r_p^p + \sum_{i=1}^3 \hat{v}_i. \quad (7)$$

其中 v 是带宽利用因子. 当系统负载较高时, 延迟值的方差会变得较大, 而此时 MBAC 算法是无法发挥作用的. 故在 MBAC 中不可以使用 100% 带宽, 所以设一个带宽利用目标因子, 它的取值由链接流的特性来决定. 每个流信源的速率与系统总的链接能力相比较小, 数据突发较短时可取的值较高; 反之取值较低. GPRS 中虽然数据突发较短, 但因为重传率较高并有切换的影响, 故 v 取值不可以太高, 本文中取值为 0.8.

(2) 如果接纳流 p 后可以使同一优先级类破坏延迟约束,则拒绝流 p 的申请:

$$D_k > \hat{D}_k + \frac{b_k^l}{\mu - \sum_{i=1}^{k-1} \hat{v}_i}, \tag{8}$$

或接纳流 p 后能引起低优先级类破坏延迟约束,也拒绝流 p 的申请:

$$D_j > \hat{D}_j - \frac{u - \sum_{i=1}^{j-1} \hat{v}_i}{u - \sum_{i=1}^{j-1} \hat{v}_i - b_k^l} \mid \frac{b_k^l}{u - \sum_{i=1}^{j-1} \hat{v}_i - r_k^l}, \quad k < j \leq 3. \tag{9}$$

经过延迟标准化后,式(8)、(9)变为

$$1 > \bar{D}_k + \frac{b_k^l}{u - \sum_{i=1}^{k-1} \hat{v}_i} \times \frac{1}{D_{k,t}}, \tag{10}$$

$$1 > \bar{D}_j - \frac{u - \sum_{i=1}^{j-1} \hat{v}_i}{u - \sum_{i=1}^{j-1} \hat{v}_i - b_k^l} + \frac{b_k^l}{u - \sum_{i=1}^{j-1} \hat{v}_i - r_k^l} \times \frac{1}{D_{j,t}}, \quad k < j \leq 3. \tag{11}$$

(3) 如果式(7)、(10)、(11)这 3 式均满足,则接纳这个流.

2.2 时间窗测量算法

2.2.1 测量带宽

为了估计使用带宽,我们把一个时间窗分为 10 个采样周期,在每个周期中计算每一类预留服务的平均使用带宽 \hat{v}_i^s (采样时间为 S),在下面 4 种情况下更新估计带宽的值:① 当时间窗结束时,取最大的 \hat{v}_i^s 作为估计带宽;② 当接入新流时,更新估计带宽,并重启时间窗;③ 当 \hat{v}_i^s 超过估计带宽时,以 \hat{v}_i^s 作为估计带宽;④ 当有一流离开时,更新估计带宽,并重启时间窗.估计带宽的更新公式如下:

$$\hat{v}' = \begin{cases} \text{MAX}(\hat{v}^s), & \text{测量窗口结束时,} \\ \hat{v}^s, & \text{如果 } \hat{v}^s > \hat{v}, \hat{v} \text{ 是 } S \text{ 中的平均使用带宽,} \\ \hat{v} + r^p, & \text{当接纳一新流 } p \text{ 时,} \\ \hat{v} - \beta r^p, & \text{当有一个流 } p \text{ 结束时.} \end{cases} \tag{12}$$

2.2.2 测量延迟

为了估计延迟,我们测量每一个帧的队列延迟.在以下 4 种情况下更新估计延迟的值:① 当时间窗结束时,取最大的 D^a 作为估计延迟;② 当接入新流时,更新估计延迟,并重启时间窗;③ 当 D^a 超过估计延迟时,以 D^a 作为估计延迟;④ 当有流离开系统时,立即重启时间窗.估计延迟的更新公式如下:

$$\bar{D}_j^a = \begin{cases} \text{MAX}(D^a), & \text{测量窗口结束时,} \\ \lambda D^a, & \text{如果 } D^a > \bar{D}_j^a, \\ \text{等式(10)和(11)的右边,} & \text{接纳一个新流 } p \text{ 时,} \\ \text{不变} & \text{有一流 } p \text{ 结束时.} \end{cases} \tag{13}$$

时间窗测量过程中使用 4 个变量:

(1) 延迟增长因子(λ).当帧的队列延迟超过估计的最大延迟时,使用更大的最大延迟(用增长

因子乘以帧的队列延迟).本文中取值为2.

(2) 采样周期(S).采样周期控制速率测量的敏感性,其值越小对突发就越敏感.本文中一个采样周期包括54个TDMA循环,取值为996.84微秒.

(3) 时间窗(T).时间窗的大小控制着测量算法的适应性,时间窗越小则适应性越好,时间窗越大则稳定性越好,它应该是采样周期的倍数.接纳一个新流后,则重启一个时间窗来收集新流的信息.本文中一个时间窗包括10个采样周期,取值为9.9684秒.

(4) 带宽减少因子(β).当有流结束离开网络时,需要减少使用带宽的估计值,减少的带宽值大致等于该流的等效带宽,可以根据流的不同为 β 确定不同的值.本文为了简化算法,固定 β 的取值为1.

3 仿 真

3.1 数据源的3种模型

我们采用3种数据源以及它们的混合来仿真未来的GPRS数据流.3个模型均假设为帕松到达过程,间隔时间符合指数分布.分组长度分别由下面3个统计模型来描述.最大帧长为1600字节,大于这个值的分组必须被分段为多帧.

3.1.1 FUNET 模型

这个模型是基于FUNET中E-mail使用的统计数据来产生的,具体描述见“SMG1 238/93 Packet Data over GSM Network”文档. FUNET模型的一般描述如下,见表2.

Table 2 The distribution of FUNET model

表2 FUNET模式的分布

Length of individual group ^① (K)	<0.5	<1	<1.5	<2	<3	<5	<10
Percentage ^②	10	36	54	67	79	87	91

①分组长,②百分比.

这个模型的概率分布函数可以近似认为符合参数为(0.8,1)的柯西分布,其中 $0 < X < M$, M 表示分组尺寸的最大值(10K字节):

$$f(x) = \text{Cauchy}(0.8, 1) = \frac{1}{\pi(1 - (x - 0.8)^2)} \quad (14)$$

3.1.2 Mobitex 模型

Mobitex模型是基于瑞典在船只管理应用中Mobitex无线数据网络上使用的统计数据来产生的.上行信道:30+random(-15,15)字节;下行信道:115-random(-57,57)字节.

3.1.3 WWW 模型

我们选择了EPA-HTTP,它是一个繁忙的WWW服务器上一天的HTTP记录.其格式描述如下:主机号或Internet地址:[日:小时:分:秒]:服务要求:HTTP回应码:回应字节数.例如:141.243.1.172:[29:23:53:25]:GET/Software.html/1.0:200:497.我们通过一个程序把上述文件转化为两个我们需要的文件,一个存储上行信道的数据传输描述,一个存储下行信道的数据传输描述,形式均为“到达时间:分组长度”.

我们假设MS可以使用的时隙的数目 n 的取值范围为1、2、3、7,每一个时隙的传输速率为9.05Kb(CS1).给E-mail(业务类1)、Fleet管理(业务类2)和WWW浏览(业务类3)这3种不同的应用分配不同的延迟类和时隙,可得到11种数据分组的不同类型,详细描述见表3.在生成数据分组的时候,数据源生成部分按照表3所示的比率,随机选择11种数据分组类型中的一种来描述要

生成的数据分组.

Table 3 The different types and ratios of groups
表 3 分组的不同类型和比率

Semanteme ^①	1	2	3	4	5	6	7	8	9	10	11
Traffic type ^②	1	1	1	1	1	2	2	3	3	3	3
Delay classes ^③	1	1	2	3	4	1	1	1	1	2	2
Time slots ^④	1	2	1	1	2	1	2	3	7	3	7
Ratio ^⑤	1/24	1/24	1/12	1/12	1/12	1/16	1/16	1/12	1/12	1/12	1/12

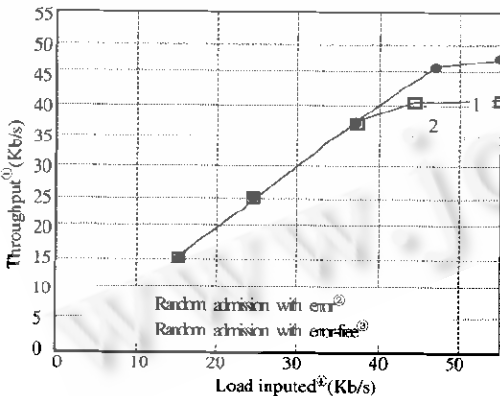
①语义,②业务类,③延迟类,④时隙,⑤比率.

3.2 仿真实验和仿真结果

3.2.1 错误和随机接入对系统性能的影响

在实验 1 中,我们构造了两种实验语义,且均假设在一个小区内,MS 的数目为 200 个,可用的分组传输信道为 7 个,总的带宽数量为 63.35Kb/s.一种方式为假设没有资源竞争机制,每个 MS 的资源申请均能得到满足,而且线路传输不会有错误发生;另一种方式为存在资源竞争和传输错误.在这两种实验语义中均使用峰值速率的 CAC 算法.

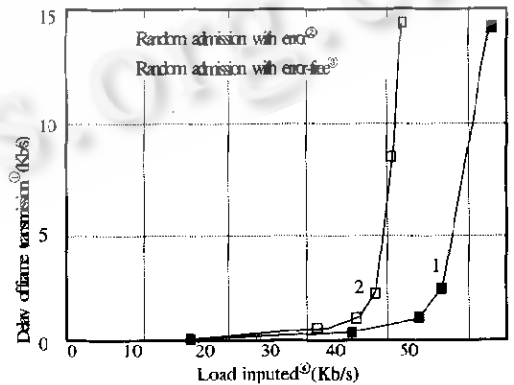
我们在实验 1 中记录在具有不同输入负载的情况下,两种实验语义中的系统通过量和每一帧的传输延迟,具体的仿真结果如图 2 和图 3 所示.从中可以发现,当系统负载量低于 32Kb/s 时,两种实验语义所得到的实验结果差别不大,在高于 32Kb/s 以后,实验语义 1 的执行效果明显优于实验语义 2.在两个实验语义中,当系统输入负载较低时,通过量和帧传输延迟均随着输入负载的增加而线性增加.当输入负载超过一定点时,通过量和帧传输延迟与输入负载的比例关系显著增大.在实验语义 1 中,这个点是 47Kb/s,而在实验语义 2 中,这个点提前到了 37Kb/s,大约提高了 20%,从以上分析可以得出,在设计 GPRS 网络中的 CAC 算法时,要考虑为随机接入和传输出错预留必要的资源量,才能保证 CAC 的有效性.



①通过量,②无错误、随机接入,③有错误、随机接入,④输入负载量.

Fig. 2 The effect on flow by error and random admission

图 2 错误和随机接入对通过量的影响



①帧传输延迟,②无错误、随机接入,③有错误、随机接入,④输入负载量.

Fig. 3 The effect on frame transmission delay by error and random admission

图 3 错误和随机接入对帧传输延迟的影响

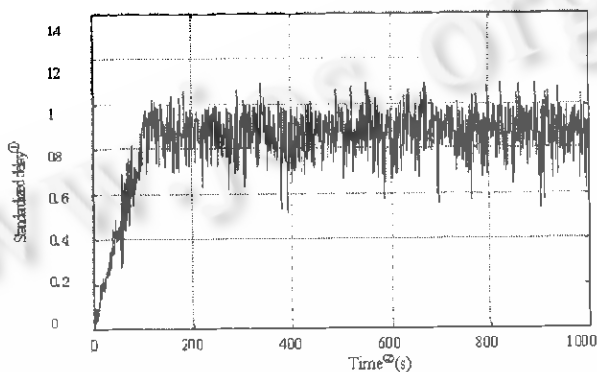
3.2.2 几种 CAC 算法的比较

在实验 2 中,我们比较了几种 CAC 算法在 GPRS 网络中的执行效果.3 种算法中均使用漏桶算法来描述资源要求,两种 MBAC 中均使用时间窗测量算法,时间窗大小为 9.968 4 秒,包括 10

个采样周期. 本次实验中, 我们记录了一些必要的系统参数, 并在仿真结束后, 根据这些记录来判断每一次的 CAC 算法是否准确, 并由此构造出一种完美的 MBAC 算法.

3 种 CAC 算法在 GPRS 网络中的执行效果比较如图 5 所示, 仅使用带宽的 MBAC 算法在同等情况下可以达到 70% 的资源利用率, 我们的算法的资源利用率则可以达到 80%, 与完美的 MBAC 算法在性能上大约相差 5% 左右.

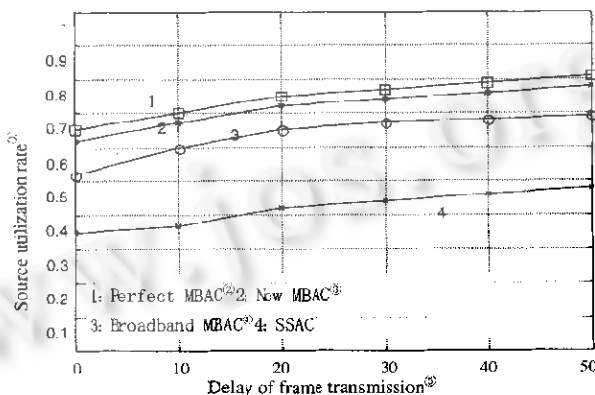
图 4 详细列出了我们的算法中 1 000 个采样周期中的平均标准化延迟的变化情况, 从该图可以发现, 我们的算法在获得 80% 的资源利用率的同时, 可以保证绝大多数的数据帧的传输延迟在延迟约束以内.



①标准化延迟, ②时间.

Fig. 4 The simulation result of our algorithm

图 4 我们的算法的仿真结果



①资源利用率, ②完美 MBAC, ③新 MBAC, ④带宽 MBAC, ⑤分组传输延迟.

Fig. 5 The simulation comparison of four algorithms

图 5 4 种 CAC 算法的比较

3.2.3 时间窗最佳长度的确定

如何确定测量过程中的最佳时间窗长度是 MBAC 算法中公认的难题. 在实验 3 中, 我们比较了我们的算法在 5 种不同的时间窗长度下的不同执行情况, 5 种时间窗长度分别为 1 倍、3 倍、5 倍、10 倍和 13 倍的采样周期 (采样周期为 996.84 微秒). 由图 6 可知, 当时间窗长度大于 10 倍采样周期时, 延迟冒犯率很低, 一般在万分之一左右, 而且资源的利用率可达 80% 以上. 而一旦小于

10 倍的采样周期,特别是小于 5 倍的采样周期时,虽然资源利用率可以提高一些,但延迟冒犯率上升速度更快,故我们选取 10 倍的采样周期为算法中的时间窗长度。

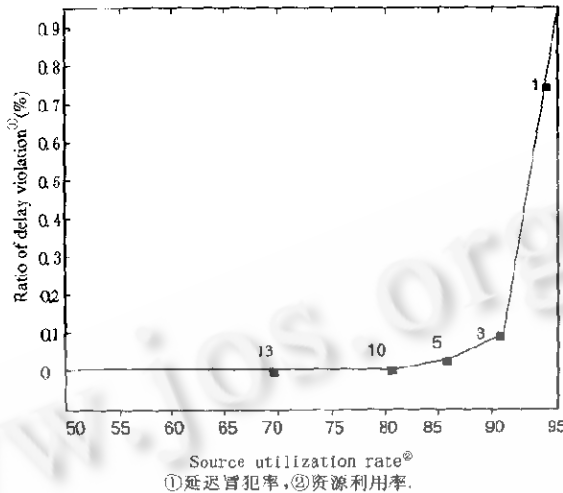


Fig. 6 The length comparison of five different time windows
图 6 5 种不同时间窗长度的比较

4 结论

本文在充分分析 GPRS 特性的基础上提出了一种适合于 GPRS 的 MABC 算法. 通过仿真发现, 该算法可在满足 GPRS QoS 要求的同时获得 80% 的资源利用率。

References:

- [1] GSM 03.60. General packet radio service (GPRS); stage 2 service description. ETSI, 1998.
- [2] GSM 04.60. General packet radio service (GPRS); overall description of the GPRS radio interface (Um). ETSI, 1998.
- [3] Cai, J., Goodman, D. J. General packet radio service in GSM. *IEEE Communication Magazine*, 1997, 35(10): 122~131.
- [4] Sugih, Jamun, Danzig, P. B., Shenker, S. J. Comparison of measurement-based admission control algorithms for controlled-load services. In: *Proceedings of the IEEE INFOCOM'97*. Kobe, Japan, CA: IEEE Computer Society Press, 1997. 973~980.
- [5] Tse, D., Grossglauser, M. Measurement-based call admission control: analysis and simulation. In: *Proceedings of the IEEE INFOCOM'97*. Kobe, Japan, CA: IEEE Computer Society Press, 1997. 981~989.
- [6] Jamun, S., Danzig, P. B., Shenker, S. J., et al. A measurement-based admission control algorithm for integrated services packet networks. *IEEE/ACM Transactions on Networking*, 1997, 5(1): 56~70.
- [7] Logothetis, D., Trivedi, K. S. The leaky bucket as a policing device: transient analysis and dimensioning. Technical Reports, Alberta, Canada: CACC of ECE, 1996.
- [8] Kulkarni, V. G., Gautam, N. Leaky bucket: sizing and admission control. Technical Report, TR-96-39, Chapel Hill; Department of Operations Research, University of North Carolina, 1993.
- [9] Harry, P. G., Khaled, M. E. Call admission control schemes: a review. *IEEE Communication Magazine*, 1996, 34(11): 82~91.
- [10] Parekh, A. K. A generalized processor sharing approach to flow control in integrated service network [Ph. D. Thesis]. Technical Report, LIDS-TR-2089. Laboratory for Information and Decision System, MIT, 1992.

A Measurement-Based Admission Control Algorithm of General Packet Radio Service *

LIU Yan-hang¹, YUAN Sen-miao¹, SUN Hui-ping²

¹(College of Computer Science and Technology, Jilin University, Changchun 130012, China);

²(Department of Electronics Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

E-mail: lyh_lb_lk@yahoo.com.cn

Abstract: The general packet radio service (GPRS) aims at providing predictive QoS to user by using QoS management mechanisms similar to those employed by ATM. It shows that the measurement-based admission control algorithm has better adaptability and higher efficiency than other call admission control algorithms in many literatures. In this paper, based on the enough analysis on characteristic of GPRS, a measurement-based admission control algorithm is presented fitting for GPRS using a model for the worst case delay of priority queue, equivalent approximate leaky bucket and delay normalized. By simulation, the proposed algorithm can satisfy the required GPRS QoS while achieving high sources utilization rates of about 80%.

Key words: general packet radio service (GPRS); predictive QoS; QoS management mechanism; call admission control algorithm; measurement-based admission control

* Received May 8, 2000; accepted October 23, 2000

Supported by the National Natural Science Foundation of China under Grant No. 69873019; the Natural Science Foundation of Jilin Province of China under Grant No. 20000545