

基于效用EPON分布式动态带宽分配实现机制^{*}

张晋豫¹⁺, 刘 犁²

¹(北京交通大学 计算机与信息技术学院, 北京 100044)

²(Computer Department, New Jersey's Science and Technology University, Newark, N.J. 07102, USA)

Implement Mechanism on Distributed EPON DBA Based on Utility

ZHANG Jin-Yu¹⁺, LIU Li²

¹(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

²(Computer Department, New Jersey's Science and Technology University, Newark, N.J. 07102, USA)

+ Corresponding author: E-mail: zjy@bjtu.edu.cn

Zhang JY, Liu L. Implement mechanism on distributed EPON DBA based on utility. *Journal of Software*, 2008,19(7):1693–1706. <http://www.jos.org.cn/1000-9825/19/1693.htm>

Abstract: EPON (Ethernet-based passive optical network) is optical-fiber-based access technology and is becoming a primary one for the next generation access network. But the Polling mechanism of IEEE 802.3ah EPON has low bandwidth efficiency, and many available bandwidths are wasted through UWR (unused window remainder), USR (unused slot remainder), UQR (unused queue remainder), and UPR (unused package remainder) produced by its DBA (dynamic bandwidth allocation) algorithm. IPACT (interleaved Polling with adaptive cycle time) EPON presents a new Polling mechanism with high bandwidth efficiency, but its DBA algorithm has the same disadvantages as IEEE 802.3ah. Based on IPACT, this paper presents an implement mechanism on distributed EPON utility DBA, it can provide user service different treatment according to his SLA (service level agreement). By a concentrated and a distributed recursive utility algorithms, it effectively diminishes UWR-reproduced and UQR. By a distributed UPR diminishing mechanism, it reduces UPR. By putting forward a USR diminish mechanism based on interleaved baton, it increases the success probability of diminishing USR, and by adding the USR of baton-sender to the grant of his baton-receiver, it improves the bandwidth efficiency. The simulation results verify it merits well.

Key words: Ethernet-based passive optical network (EPON); dynamic bandwidth allocation (DBA); quality of service (QoS); service level agreement (SLA)

摘 要: EPON(Ethernet-based passive optical network)作为基于光纤的宽带网络接入技术,已经成为下一代接入网络的关键技术之一,但是 IEEE 802.3ah EPON 的 Polling 机制存在带宽使用率不高的问题,其 DBA(dynamic bandwidth allocation)算法产生 UWR(unused window remainder),USR(unused slot remainder),UQR(unused queue remainder)和 UPR(unused package remainder),浪费了许多带宽资源.IPACT(interleaved polling with adaptive cycle

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2007AA01Z203 (国家高技术研究发展计划(863)); the National Basic Research Program of China under Grant No.2007CB307101-4 (国家重点基础研究发展计划(973))

Received 2007-07-28; Accepted 2008-01-29

time) EPON 提出了一个带宽使用率较高的新的 Poling 机制,但没有解决 DBA 算法存在的缺点.在 IPACT 基础上,提出了一个基于效用的分布式 EPON DBA 实现机制,实现了对不同 SLA(service level agreement)用户的相同业务类应用的差分处理.通过一种集中递归效用算法,有效地消除了再生 UWR,通过一种分布式递归效用算法,有效地消除了 UQR,通过一个分布式 UPR 消除机制,减少了 UPR.提出了一个基于交织接力棒的 USR 消除机制,提高了消除 USR 条件满足的概率,通过将交棒者的 USR 追加到接棒者的授权中,提高了带宽的使用效率.仿真结果很好地验证了该机制的优点.

关键词: 以太无源光网络;动态带宽分配;服务质量;服务等级协议

中图法分类号: TP393 文献标识码: A

EFM(Ethernet in first mile)任务组通过IEEE 802.3ah^[1-3]定义的EPON DBA(Ethernet-based passive optical network dynamic bandwidth allocation)机制存在以下效率不高的问题:1) 对IEEE 802.3ah EPON采用Polling机制对上行带宽的分配进行控制:OLT(optical line termination)向每个ONU(optical network unit)发出Request控制消息,询问它们有多少数据要上传($(1/2)RTT$ (round trip time));ONU通过Report控制消息向OLT反馈当前存储在所有队列中等待上传的数据的数量($(1/2)RTT$);OLT收到所有ONU的Report消息后由DBA算法算出每一个ONU该周期的上传的窗口体积和相位,并通过Grant控制消息授权它($(1/2)RTT$);每一个ONU在自己的窗口打开时打开激光器发送和窗口体积相当的数据.因此,每一个Polling周期,需要 $3/2 RTT$ 的时间用于交换Report/Grant,带宽的利用率不高.2) Inter-ONUs DBA和Intra-ONUs DBA算法都是以字节为单位的,但是队列调度是以数据包为单位.在一个Polling周期,在总授权量小于总需求量的情况下,一个ONU获得的授权不一定包含整数个数据包,于是有部分发送时隙被浪费掉,会产生USR(unused slot remainder)^[4,5].同样,在一个Polling周期,在一个ONU的不同缓存队列之间,每一个队列获得的子授权不一定包含整数个缓存包,于是有长度小于一个数据包长度的调度资源被浪费掉,会产生UPR(unused package remainder).USR和UPR都降低带宽的使用效率.3) Inter-ONUs DBA算法给一个ONU授权的窗口大小可能大于它的Report,此时,部分窗口将空载,会产生UWR(unused window remainder)^[6].4) Intra-ONUs DBA算法给每一个优先队列分配的授权也可能大于它的需求,部分队列调度时间闲置,会产生UQR(unused queue remainder).

在 IPACT(interleaved polling with adaptive cycle time)机制基础上,本文提出了一个基于效用的分布式 EPON DBA 实现机制,包括以下内容:1) 对 OLT,提出了一个基于效用的 Inter-ONUs DBA 算法,在网络 congestion 的时候,以 ONU 的动态平均 SLA(service level agreement)为权重来进行窗口指派,并通过集中的基于效用的递归算法来分配 UWR 产生的剩余窗口,从而消除 UWR.2) 对 ONU,提出了一种基于效用的 Intra-ONUs DBA 算法,在网络 congestion 的情况下,以每一个队列的动态平均 SLA 为权重来分配带宽资源,并通过分布式的基于效用的递归机制来分配 UQR 产生的剩余带宽,从而消除 UQR.3) 对 ONU,提出了一个分布式的基于效用的 UPR 消除机制,通过再分配所有队列产生的 UPR,实现对 UPR 的深度消除,减少 UPR.4) 在文献[5]提出的 USR 消除机制基础上,提出了一个交织接力机制来增加消除 USR 条件满足的概率,提高 USR 的消除效果.通过将满足 USR 消除条件的交棒者的 USR 追加到接棒者的授权中,维持周期的大小,提高带宽的使用效率.

本文第 1 节论述相关的研究.第 2 节论述基于效用的 EPON 分布式 DBA 实现机制.第 3 节论述仿真并进行讨论.最后给出结论.

1 相关研究

文献[6-8]提出了一个IPACT机制:OLT维持一个Polling表,并且发送GATE到ONU.每个ONU在接收到GATE消息后,在自己的窗口开启时打开激光器发送数据,并在发送数据的末尾绑定一个REPORT消息,告诉OLT它当前的队列中缓存的数据长度,也就是下一个周期请求发送的数据量.在一个ONU数据发送结束以前,下一个ONU被Polling,实现了多个Polling在时间上的重叠.与传统的机制(IEEE 802.3ah)相比,IPACT将交换Grant/Report的时间减少为 $1/2RTT$,从而提高了带宽的使用率;多个ONU的Polling周期相互交织,Grant/Report交

换不占用专门的上行窗口,从而增加了可用带宽的数量.IPACT优越的性能使它得到了学术界和工业界的广泛关注,基于它的研究有很多成果^[9,10],本文也在其基础上加以研究.

文献[11-13]研究了集中的DBA模式,OLT通过Inter-ONUs DBA算法对所有ONU资源进行集中管理.集中的模式比较适合只有无差别BE(best-effort)业务类的单一业务量环境,因为要动态维护多个ONU的多个业务类的状态将增加DBA算法的复杂性和OLT的负担.但随着多媒体业务的发展,EPON需要支持多个QoS类,要求OLT和ONU实现支持不同优先级别的队列,这样,除了集中的Inter-ONUs DBA算法以外,还需要在每个ONU实现Intra-ONUs DBA算法,也就是分布式DBA算法,来管理它的多个队列的带宽资源.当前,几乎所有的EPON都要求支持差分业务,如DiffServ(differentiated service)和IEEE 802.1d/1q/1p等,所以,当前的研究主要集中在分布DBA算法^[14,15].

文献[5,9]研究了分布式的DBA模式.文献[5]提出了一个分层的细粒度的调度器,与我们提出的机制非常相似.对于Inter-Class,它采用传统的SP(strict priority)机制,对于Intra-Class,它采用M-SFQ(modified start-time fair queuing)机制.机制通过给每个ONU、ONU的每个业务类、ONU的业务类的每一个用户队列赋予不同的SLA权重,并按照这些权重分别给它们授权,实现了对不同SLA的用户相同业务类应用的差分处理,提高了网络带宽的效益.但由于需要在每一个业务类里为每一个用户维持一个队列,限制了它的扩展性.在文献[9]提出的机制中,ONU除了向OLT上报它们的带宽需求外,还要与其他ONU交换它们的队列状态.所有ONU同步和独立地运行相同的DBA算法实例,输出相同的带宽分配结果.一旦算法被运行,ONU连续并且有次序地在没有碰撞的情况下发送它们的数据.OLT集中处理请求和带宽授权任务被取消,只负责同步任务.该机制简化了OLT的复杂性,但该机制也存在缺点:1) 要求ONU和OPS(optical passive splinter)具有较高的智能性,增加了它们的复杂性,提高了网络的成本.2) 由于要求所有的ONU都支持相同的算法,具有相近的处理能力,因此也限制了它的扩展性.文献[5,9]提出的机制都实现了ONU之间以及一个ONU内不同优先级队列之间的协商,使业务量大的ONU能够使用业务量小的ONU的窗口资源以及使大业务量业务类使用小业务量业务类的带宽资源,实现了资源的统计复用功能,消除了部分UWR和UQR,提高了资源的使用效率.但两种机制提出的窗口和带宽复用算法都是一次性的,由于复用算法的结果有可能使一个ONU或一个ONU的队列获得的窗口和带宽授权大于它的需求,产生再生的UWR和UQR,所以两者都没有根除UWR和UQR.

文献[4,5,16]研究了消除USR的机制.文献[4]提出D-CRED(dynamic credit distribution)机制:ONU向OLT上报可能授权的上一个帧边界和下一个帧边界,OLT通过它们来计算USR,并在实际的授权中直接消除它.由于ONU之间缺乏实时业务量协商机制,一个ONU在不知道其他ONU业务量状态的情况下,无法对授权实现准确评估,所以机制的可行性受到限制.文献[5]提出了接力棒机制:使用Report的冗余字节携带USR,把Report作为接力棒放在窗口的最前面进行传输.OLT在接到接力棒后,取出USR并把它传递给下一个ONU(进行授权).如果下一个ONU收到接力棒(授权)后,前一个ONU的数据还没有发送完毕,则交接棒成功,接棒者把自己窗口开启时间提前于交棒者的USR.否则,交接棒不成功,OLT不等接力棒到达,就按照传统的机制对接棒者进行授权,因此前者的USR被浪费掉.在周期的末尾,最后一个ONU把接力棒传给下一个周期的第一个ONU,同时,提前本周期结束时间(最后一个ONU的USR时间长度),也就是提前下一个周期开始时间.该机制在没有显著增加所有包的时延的前提下,提高了系统带宽的使用效率.但是每一次交接棒是否成功是不确定的.由于机制中缺乏提高交接棒成功概率的机制,因此,USR消除的效果不好.文献[16]提出二次授权的USR消除机制:OLT在收到所有ONU的Report后,通过DBA算法计算出它们的授权窗口的大小,并一次性地对它们统一授权:ONU收到授权后并不发送数据,而是计算USR,并把计算结果发送给OLT.OLT在收到所有ONU的USR后,重新计算消除了USR的授权窗口的大小,并重新对ONU进行授权.该机制可以完全消除所有USR,但增加一个RTT来传送USR消息,因此不但所有数据包的时延被延迟,而且也降低了带宽的使用率.

2 基于效用的 EPON 分布式 DBA 算法

本文提出一个基于效用的 EPON 分布式 DBA 机制,如图 1 所示.

与本文相关的机制包括:1) S-ATQ(SLA-based arrive time queue)队列管理机制;2) 效用 Inter-ONUs DBA; 3) 效用 Intra-ONUs DBA 算法;4) 公平调度机制;5) UPR 消除机制;6) USR 消除机制.

2.1 S-ATQ队列管理机制

按照经济学的观点,对运营商来说,提供相等的带宽应该得到相等的效益.当相等的带宽支持的用户数比较少小时,数据包遭遇拥塞的概率就比较低,相应得到的服务质量就比较高,当然,带宽的价格也比较昂贵^[17,18].

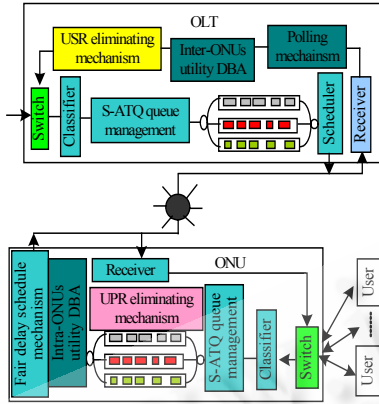


Fig.1 Distributed DBA mechanism based-utility
图 1 基于效用的分布式 DBA 机制

按照经济学观点,我们给一个 ONU 的一个业务类的一个 SLA 类定义一个 SLA 权重:

$$\rho^{ij}(x) = \frac{B_{MAX}^{ij}(x)}{N_{MAX}^{ij}(x)} \quad (1)$$

式中, $B_{MAX}^{ij}(x)$, $N_{MAX}^{ij}(x)$ 分别表示第 i 个 ONU 的第 j 个业务类为第 x 个 SLA 类分配的最大带宽和允许接纳的最大应用数,满足 $\sum_x B_{MAX}^{ij}(x) = B_{MAX}^{ij}$, $\sum_x N_{MAX}^{ij}(x) = N_{MAX}^{ij}$, 其中 B_{MAX}^{ij} 和 N_{MAX}^{ij} 分别表示第 i 个 ONU 为第 j 个业务类指配的最大带宽和允许接纳的最大应用数, X 为 EPON 支持的总 SLA 级别数.式(1)表示: $\rho^{ij}(x)$ 等于第 i 个 ONU 的第 j 个业务类为第 x 个 SLA 类的一个用户所指配的最大平均带宽数量.

本文给最低优先级定义了一个转发能力下限,给其他高优先级定义了一个转发能力上限,它们满足:

$$\sum_j L_{MAX}^{ij} + L_{MIN}^{ij} = \sum_j L_{ALL}^{ij} \quad (2)$$

其中, L_{ALL}^{ij} 和 L_{MAX}^{ij} 分别表示为第 i 个 ONU 的第 j 个业务类的构造的队列长度和可占有的最大队列长度, L_{MIN}^{ij} 为最低业务类保证的最小队列长度, J 为支持的最大优先级数目.

一个业务类保证的最大带宽 B_{MAX}^{ij} 由 L_{MAX}^{ij} 决定,因此式(1)可以表达为

$$\rho^{ij}(x) = \frac{B_{MAX}^{ij}(x)}{N_{MAX}^{ij}(x)} = \frac{L_{MAX}^{ij}(x)}{N_{MAX}^{ij}(x)} \quad (3)$$

$L_{MAX}^{ij}(x)$ 表示第 i 个 ONU 的第 j 个业务类为第 x 个 SLA 类分配的最大队列长度,满足 $\sum_x L_{MAX}^{ij}(x) = L_{MAX}^{ij}$.

为了提高公平性并保证用户业务流的时延要求,我们提出了 S-ATQ 队列管理机制,通过对 OLT 和 ONU 的优先级队列的准入控制,实现对可用优先级队列资源的基于 SLA 的差分分配.它包括:1) 基于 FAFI(first arrive first in)的数据包准入原则;2) 基于 SLA 权重的数据包准入数量.

S-ATQ 调度算法如图 2 所示.

对每一个新到达的用户 k 数据包,维持两个变量:1) t^{ik} , 表示到达的时间;2) $l^{ik}(t^{ik})$, 表示到达的数据包的长度.在 m 周期开始,第 i 个 ONU 的第 j 个业务类可用的队列长度为

$$\Delta L_{MAX}^{ij}(m) = L_{MAX}^{ij} - \Delta L_u^{ij}(m-1) \quad (4)$$

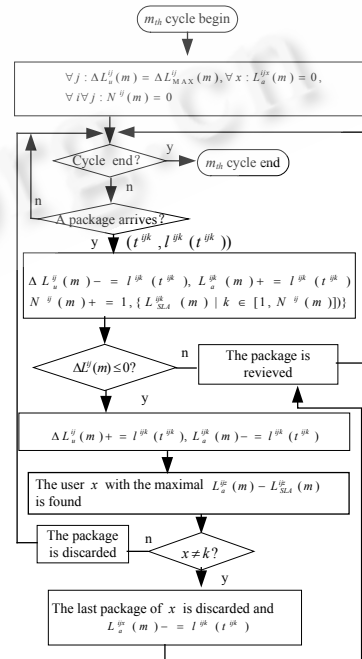


Fig.2 S-ATQ scheduling algorithm
图 2 S-ATQ 调度算法

式中, $\Delta L_u^j(m-1)$ 表示周期 $m-1$ 调度结束后剩余的队列长度.

在 m 周期开始,为每一个 ONU 的每一个业务类队列维持两个变量:1) $\Delta L_u^j(m)$,表示剩余队列的长度,每接收一个数据包,它都要减去接收到的收据包的长度.2) $N^{ij}(m)$ 表示在线应用数,每当有一个新应用的数据包被接纳,它的值就加 1. 为每一个 ONU 的每一个业务类的每一个应用维持两个变量:1) $L_a^{ik}(m)$,表示在 m 周期,应用 k 在第 i 个 ONU 的第 j 个业务类占用的队列的长度,每当该应用的一个新数据包到达,就加上它的长度.2) $L_{SLA}^{ik}(m)$,表示用户基于 SLA 权重的期望长度,可表示为

$$L_{SLA}^{ik}(m) = \Delta L_{MAX}^{ij}(m) \times \rho^{ij}(k^{SLA}) / \sum_x^{N^{ij}(m)} \rho^{ij}(x^{SLA}) \quad (5)$$

k^{SLA} 表示应用 k 属于的 SLA 业务类, $n^{ij}(m)$ 表示在周期 m 的在线用户数.

调度算法的基本思想是:对第 i 个 ONU,在周期 m 开始以前,首先计算 $m-1$ 周期数据发送结束后每一个业务类可用(剩余)队列的长度 $\Delta L_u^j(m)$,并把每个应用占用的队列长度 $L_a^{ik}(m)$ 清零.当长度为 $l^{ijk}(t^{ijk})$ 的一个数据包到达时,只要它属于的业务类的缓存队列不满,就填充进去, $L_a^{ik}(m)$ 加 $l^{ijk}(t^{ijk})$, $\Delta L_u^j(m)$ 减 $l^{ijk}(t^{ijk})$,并刷新当前所有在线应用基于 SLA 权重的期望长度 $\{L_{SLA}^{ik}(m) | k \in [1, N^{ij}(m)]\}$. 当一个业务类的队列满时,找到该类所有在线应用中 $L_a^{ik}(m)$ 和 $L_{SLA}^{ik}(m)$ 的差最大的用户 x :

$$L_a^{ix}(m) - L_{SLA}^{ix}(m) = \max \{L_a^{ik}(m) - L_{SLA}^{ik}(m) | k \in [1, N^{ij}(m)]\} \quad (6)$$

如果 x 是当前应用,就丢弃新到达的包;如果不是,就丢弃 x 的最后到达的包,并同时接收新到达的包.

在 $m-1$ 周期数据发送结束时,每个队列缓存的数据量为

$$L_R^{ij}(m) = \Delta L_u^j(m-1) + \sum_x^{N^{ij}(m)} L_a^{ix}(m) \quad (7)$$

在 m 周期,每个 ONU Report 的值为

$$R^i(m) = \sum_j L_R^{ij}(m) \quad (8)$$

2.2 效用 Inter-ONUs DBA 算法

对不同的 ONU,在一个周期内,由于其在线应用的平均 SLA 水平存在差异,为了提高网络资源的效益,在网络拥塞时,应根据它来分配带宽资源.本文提出了一种基于 ONU 的动态 SLA 权重的 Inter-ONU DBA 算法,即效用 Inter-ONUs DBA 算法.

下面给出 ONU 的动态平均 SLA 权重的定义:

$$\rho^i(m) = \sum_j \sum_k^{N^{ij}(m)} \rho^{ij}(k^{SLA}) \quad (9)$$

在网络负载较重的情况下,为保证数据包时延并降低丢包率,我们给周期规定一个最大值 T_{MAX}^{cycle} :

$$T_{MAX}^{cycle} \leq \min(\forall i \forall j : d_{max}^{ij}) - T_{OLT-ONU} \quad (10)$$

其中, $T_{OLT-ONU}$ 是数据包在 OLT 和 ONU 之间的最大传输时延,由它们之间的距离(10 公里或是 20 公里)决定. d_{max}^{ij} 是第 i 个 ONU 的第 j 个业务类数据包在 EPON 上的传输最大时延要求.式(10)表示:一个数据包在 EPON 中的最大传输时延(在队列中等待的最长时间 T_{MAX}^{cycle} 和传送时延 $T_{OLT-ONU}$ 之和)要不大于所有数据包的最大时延要求的最小值.

本文提出了一种基于效用的窗口分配算法,如图 3 所示.为了提高资源的利用率,我们的机制采用自适用周期大小,也就是在动态业务量环境中,周期的长度始终等于 Inter-ONUs DBA 算法给每一个 ONU 授权的数据窗口大小、数据窗口之间的保护缓冲带大小以及控制窗口大小之和.由于在每一个 Polling 周期每一个 ONU 获得的授权窗口动态变化,所以周期的长度也动态变化.自适应周期可以避免业务量比较小时的空载现象,提高资源利用率.

$t_g^i(m)$ 为在周期 m 为第 i 个 ONU 授权的窗口大小, t_p 为一个周期前缀, t^g 为 ONU 之间的保护窗口大小, t^r 为 Report 窗口的大小, C 为 EPON 的最大数据发送速率, $N^o(m)$ 表示所有请求不被满足的 ONU 集合.

算法的思想为: 在周期 m 内, 当所有 ONU 的总的请求 (Report 和) 小于最大窗口时, 按照需求给每个 ONU 授权, 反之, 进行 SLA 授权, 也就是以 ONU 的动态平均 SLA 权重和需求的积为新权重来分配带宽, SLA 授权可表示为

$$t_{SLA}^i(m) = \frac{\rho^i(m)R^i(m)}{\sum_i (\rho^i(m)R^i(m))} (T_{MAX}^{cycle} - t_p - I \times t^g - I \times t^r) \quad (11)$$

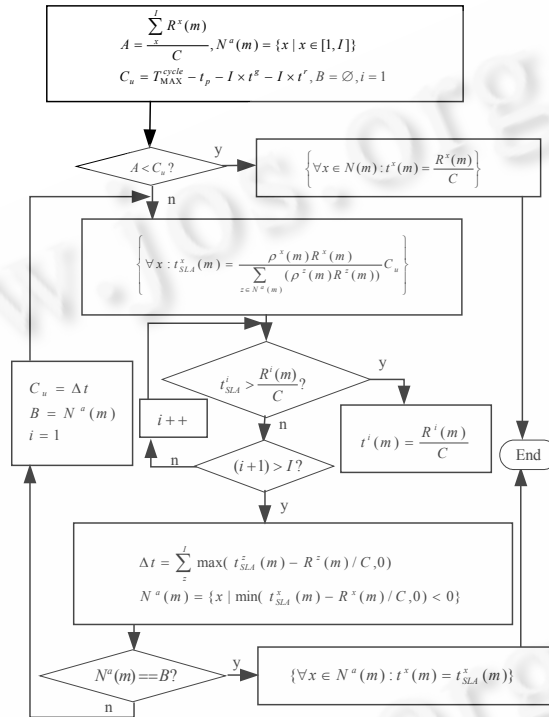


Fig.3 Inter-ONUs utility DBA algorithm

图 3 效用 Inter-ONUs DBA

如果一个 ONU 的 $t_{SLA}^i(m)$ 大于 $R^i(m)$, 则 $t_g^i(m)$ 等于 $R^i(m)$, 同时计算剩余窗口的大小, 即 UWR. 所有 ONU 剩余窗口 (总 UWR): $\Delta t = \sum_z \max(t_{SLA}^z(m) - R^z(m)/C, 0)$, 供不能满足条件的 ONU: $N^o(m) = \{x | \min(t_{SLA}^x(m) - R^x(m)/C, 0) < 0\}$ 分配. 分配仍然采用 SLA 授权的方法:

$$t_{SLA}^i(m) = t_{SLA}^i(m) + \frac{\rho^i(m)R^i(m)}{\sum_x (\rho^x(m)R^x(m))} \Delta t \quad (12)$$

这是一个递归的过程, 直到不满足条件的 ONU 数值不再发生变化, 表明没有多余资源可以分配, 则算法结束, UWR 被彻底清除. 由文献[11]和式(12), 给需求不能满足的 ONU 进行 SLA 授权, ONU 的动态 SLA 权重越大, 申请的字节越多, 则获得的授权窗口越大.

2.3 效用 Intra-ONUs DBA 算法

下面给出 ONU 的队列的动态 SLA 权重的定义, 表示为

$$\rho^{ij}(m) = \sum_k^{N^{ij}(m)} \rho^{ij}(k^{SLA}) \tag{13}$$

效用 Intra-ONUs DBA 算法的主要任务是,在一个周期中,当带宽需求不被满足时,按照每个队列的动态 SLA 权重来分摊分配的窗口.本文提出了一种基于效用的 ONU 的队列授权算法,如图 4 所示.

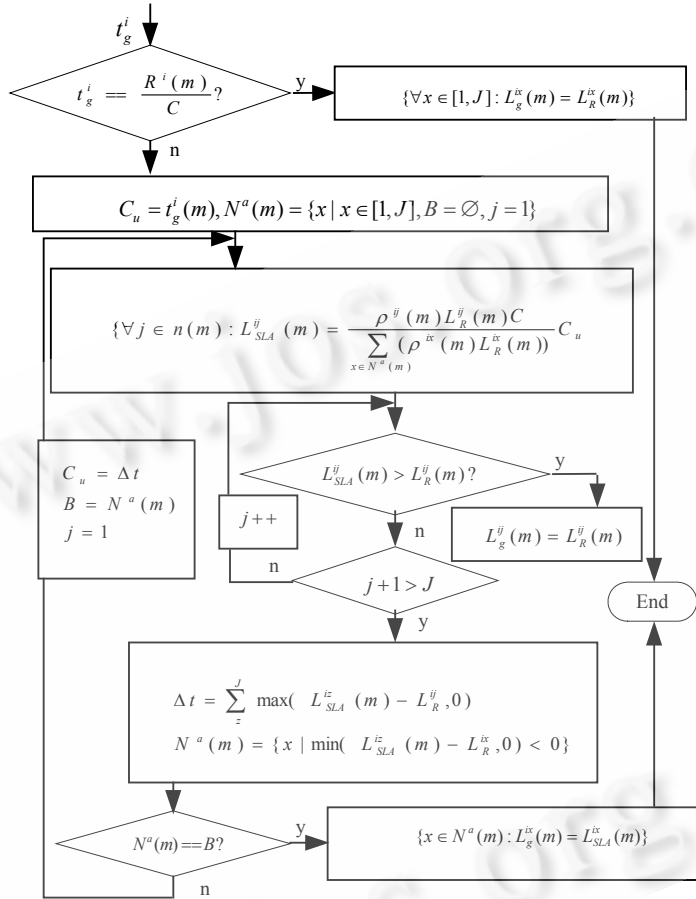


Fig.4 Intra-ONUs utility DBA algorithm

图 4 效用 Intra-ONUs DBA 算法

$L_g^{ij}(m)$ 表示第 m 周期,为第 i 个 ONU 的第 j 个业务类队列授权的队列长度.

算法的思想是:对一个 ONU,在一个周期,当授权的窗口大小 $t_g^i(m)$ 等于报告的数值时,按照每个队列的需求 $L_R^{ij}(m)$ 来给它授权.否则,计算每一个队列的 SLA 队列长度:

$$L_{SLA}^{ij}(m) = \frac{\rho^{ij}(m)L_R^{ij}(m)C}{\sum_x(\rho^{ix}(m)L_R^{ix}(m))}t_g^i(m) \tag{14}$$

并按照 SLA 队列长度来授权.如果 $L_{SLA}^{ij}(m) > L_R^{ij}(m)$,则 $L_g^{ij}(m) = L_R^{ij}(m)$,并计算剩余队列的长度,即 UQR.所有队列剩余的总队列长度之和(总的 UQR): $\Delta t = \sum_z^J \max(L_{SLA}^z(m) - L_R^z(m), 0)$ 给所有剩下的需求不能被满足的队列集合: $N^a(m) = \{x | \min(L_{SLA}^z(m) - L_R^z(m), 0) < 0\}$ 分配,分配仍然采用 SLA 队列长度的形式:

$$L_{SLA}^{ij}(m) = L_{SLA}^{ij}(m) + \frac{\rho^{ij}(m)L_R^{ij}(m)}{\sum_x \rho^{ix}(m)L_R^{ix}(m)} \Delta t \quad (15)$$

这是一个递归的过程,直到不能满足要求的队列集不发生变化,UQR 被消除干净为止.此时,给需求不能被满足的队列进行 SLA 授权.

2.4 公平调度机制

公平调度机制主要负责依据授权对缓存队列中的数据包的发送开始时间以及发送个数实施控制.本文提出一个 ONU 的公平的资源调度机制,如图 5 所示.

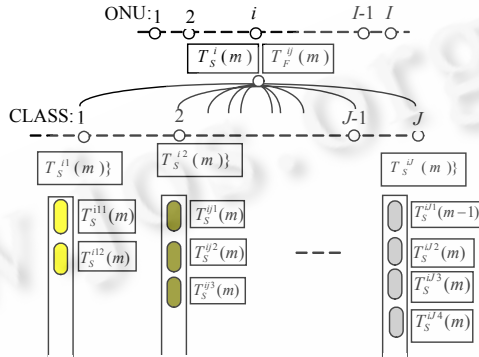


Fig.5 Fair resource scheduling mechanism

图 5 公平资源调度机制

S-ATQ机制为每个优先级队列每一个缓存的数据包维持一个到达时间标签 t^{ijk} 和长度标签 l^{ijk} (t^{ijk}), 调度机制为它们维持调度指针 $T_S^{ij}(m)$, 它初始指向 t^{ijk} . 调度机制为每一个优先级队列维护一个调度指针 T_S^j , 它始终指向该队列到达时间最早的数据包. 即: $T_S^j(m) = \min\{\forall k: T_S^{ijk}\}$. 调度机制为每一个 ONU 维持两个调度指针:

- 1) T_S^i , 它始终指向具有最小 $T_S^j(m)$ 值的队列, 即 $T_S^i = \min(\forall j: T_S^j)$.
- 2) T_F^{ij} , 表示 T_S^i 所指向队列的剩余授权窗口大小. 周期开始时为 $T_F^{ij} = L_g^{ij}(m)/C$.

调度算法的流程如图 6 所示. 调度算法的思想是:

- 1) 前一个周期发送完毕, 后一个周期才能发送;
- 2) FIFO (first in first out);
- 3) 总包数由授权确定.

对第 i 个 ONU, 在一个周期开始, 刷新 $T_S^i(m)$, 使它指向所有队列中到达时间最早的缓存包 $T_S^{ijk}(m)$. 在调度该缓存包之前, 都要检验 $T_F^{ij}(m) - L_{package} \geq 0$ 是否成立:

- 1) 若成立, 则说明该队列授权窗口没有用完, 数据包被发送, 同时刷新 $T_S^j(m)$, 指向该队列中下一个起始时间最早的缓存包. 若此时周期没有结束, 则刷新 $T_S^i(m)$, 使它指向 ONU 所有

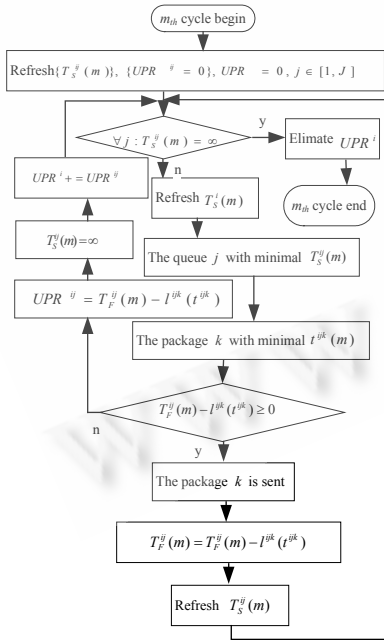


Fig.6 Scheduling algorithm

图 6 调度算法

具有最早起始时间缓存包的队列,然后重复以上过程.

2) 若不成立,则说明在该周期该业务类的授权窗口已经用完,不能再发送数据.于是将 $T_S^{ij}(m)$ 置为 ∞ . 由于业务量具有随机性,因此,在一个周期,分配给每一个 ONU 的每一个队列的授权长度不一定刚好到达一个包的边界,由于队列的调度只能以包为单位,所以产生:

$$UPR^{ij}: UPR^{ij} = T_F^{ij}(m) - l^{ijk}(t^{ijk}).$$

累计ONU的 $UPR^i: UPR^i += UPR^{ij}$. 如果该调度结束条件: $\exists j: T_S^{ij}(m) \neq \infty$ 不满足,则刷新 $T_S^i(m)$, 此时,窗口用尽的队列 $T_S^{ij}(m) = \infty$ 不可能被选中. 如果调度结束条件: $\forall j: T_S^{ij}(m) = \infty$ 满足,启动 UPR^i 消除机制, UPR^i 为本周期所有队列的 UPR^{ij} 之和.

2.5 UPR消除机制

基于 S-ATQ,本文提出了一个 UPR 消除机制,如图 7 所示.

在上一节调度过程的最后,产生了一个ONU的 UPR^i ,它等于所有队列的 UPR^{ij} 之和,可能大于某一个队列中某个数据包长度,因此,可以在ONU内消滅它.此时,所有队列的 $T_S^{ij}(m)$ 为 ∞ ,刷新它,使它指向调度结束后当前队列的最早到达的数据包,刷新 $T_S^i(m)$,使它指向具有最早到达的数据包的队列,此时判别条件 $T_F^{ij}(m) - l^{ijk}(t^{ijk}) \geq 0$ 是否成立:

- 1) 若不成立,则说明该队列的授权窗口不够用,数据包不能发送,设置 $T_S^{ij} = \infty$;
- 2) 若成立,则说明该队列的授权窗口没有用完,具备继续发送的条件.然后判断条件 $UPR^i - l^{ijk}(t^{ijk}) \geq 0$ 是否成立:
 - a) 若成立,则说明该包长度比 UPR^i 要小,需发送该数据包,同时减少 UPR^i 的值: $UPR^i - = l^{ijk}(t^{ijk})$;
 - b) 若不成立,则说明该包的长度比 UPR^i 要大,设置 $T_S^{ijk} = \infty$. 如果还有剩余的数据包:

$$\exists j: T_S^{ij}(m) \neq \infty \ \& \ \exists j \exists k: T_S^{ijk}(m) \neq \infty,$$

则重复上面的过程;如果没有,

$$\forall j: T_S^{ij} = \infty \ \parallel \ \forall j \forall k: T_S^{ijk} = \infty,$$

则 UPR^i 消除过程结束,剩余的 UPR^i 要小于本周期调度结束后所有剩余队列中最小的包,我们定义它为 UPR^i . 而文献 [4,5] 提出的方法产生的 UPR^i 都小于队列中可能存在的任何一个包,因此,本机制比以上两个机制更优越.

2.6 USR消除机制

在上一节UPR消除过程结束后,产生了 UPR^i ,本节我们提出一个机制来消除它.

文献[5]提出的接力棒消除 USR 方法中,在一个周期,比邻的两个 ONU 只有满足条件:

$$t^i(m) + t^g + t^r - RTT^i / 2 \geq RTT^{i+1} / 2 \tag{16}$$

交接棒才能成功,前一个 ONU 的 USR 才能被消除.为了提高交接成功的概率,我们提出了一个交织接力机制,其过程如图 8 所示.

每个 OLT 定义了一个全局的 UPR,在第 1 个周期, $UPR=0$. 我们给一个 ONU 定义一个交接棒能力函数:

$$w_{usr}^i = t^i(m) + t^g + t^r - RTT^i \tag{17}$$

在当前一个ONU数据的窗口 $t^i(m) + t^g + t^r$ 很大时,交棒的条件很容易满足,而在 RTT^i 很小时,接棒的条件很容易满足,因此 w_{usr}^i 可以衡量一个ONU交接棒能力.值越大,满足式(16)的概率就越高.在一个周期,计算每个

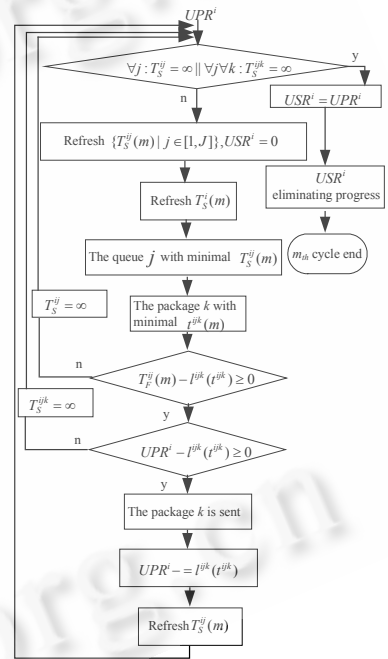


Fig.7 UPR eliminating mechanism
图 7 UPR 消除机制

ONU的 w_{USR}^i , 并按照其值从大到小给ONU重新排序(交织), 则前面比邻ONU之间满足式(16)条件的概率较高.

OLT 从交织后的 ONU 中取出第 1 个 ONU 作为交棒者, 在其授权窗口加上前一个周期残留的 USR, 并使 $USR=0$. 如果此时需求被满足, 则授权等于需求, 同时计算新的 $USR: USR = t_g^i(m) - R^i(m)$, 并开始授权: 使窗口开启时间提前 USR 的时间. 取出第 2 个 ONU 作为接棒者, 判断式(16)是否满足, 如果不满足, 则按照正常的机制对第 2 个 ONU 进行授权, USR 被浪费掉. 否则 OLT 要等到第 1 个 ONU 的接力棒(Report)到达后才给第 2 个 ONU 授权, 并把它作为交棒者, 重复以上过程, 直到周期结束.

与文献[5]的机制相比, 本机制有两大优点: (1) 通过基于交接棒能力对所有 ONU 交织, 增加了消除 USR 的次数; (2) 在满足 USR 消除条件时, 通过将前者的 USR 追加到后者的授权上, 保持了周期的大小.

以上两种方法都可以提高带宽的使用效率.

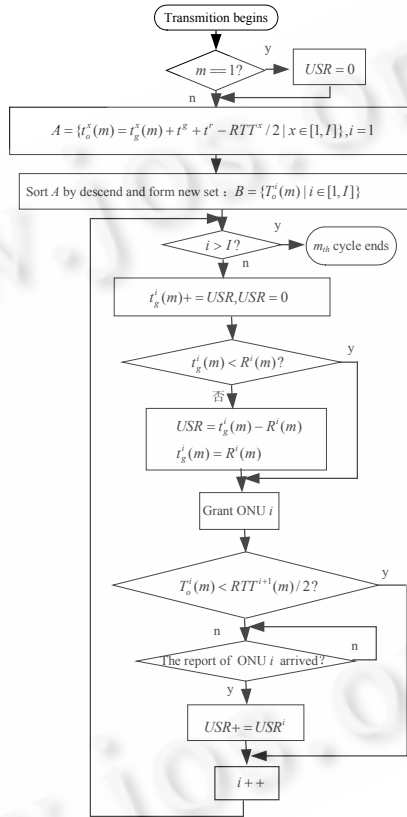


Fig.8 USR eliminating mechanism

图 8 USR 消除机制

3 仿真实验

我们建立了 1000BASE-PX10 EPON 仿真网络环境, 通过仿真实验, 对本机制和式(5)的机制进行了比较.

仿真网络 EPON 的线速率为 1Gbit/s, 支持 16 个 ONU, RTT 分别为 12.5~50us, 公差为 2.5us, 对应的距离为 2.5~10 公里, 公差为 0.5 公里. 支持 1 600 个用户, 每个 ONU 接 100 个用户, 用户和 ONU 之间的线速度为 100Mbit/s. $t_p = 64ns$, $t^s = 5us$, $t^r = 512ns$. OLT 和 ONU 传输时延为 $T_{OLT-ONU} = 50us$. 构建 3 个缓存队列, 每个队列的长度分别为 2, 18 和 6Mbytes.

业务级别. 在 OLT 和 ONU 中支持 3 类差分业务流: 1) CBR(constant bit rate)类, 映射为 DiffServ 的 EF 类, 主要对应语音业务, VoIP 在 EPON 中的时延要求为 10ms, 带宽需求为 18Kbit/s; 2) VBR(variant bit rate)类, 映射为

DiffServ 的 AF(assured forward)类,主要对应视频业务,本仿真实验中为 VoD(video on demand)业务:在 EPON 中最大时延要求为 20ms,要求带宽为 2Mbit/s.3) 非实时类,没有时延要求,要求带宽为 512Kbit/s.

根据 CBR 类的要求来定义最大周期: $T_{MAX}^{cycle} = 10 - 0.05 = 9.95(ms)$.

用户级别.实现了 3 个 SLA 类: 1) 金牌用户 160 个,分占 3 个业务类拥有带宽的各 1/3;2) 银牌用户 480 个,分占 3 个业务类拥有总带宽的各 1/3;3) 铜牌用户为 960 个,分占 3 个业务类拥有带宽的各 1/3.

业务量模型.当网络负载为 l 时,在 1~1600 之间随机产生一个整数 x 表示应用,直到应用数达到 1600 l 为止.产生的先后次序表示应用到达的先后顺序. x 模 10 的余数:如果为 0,表示是金牌用户;如果在 1~3,表示是银牌用户;如果在 4~9,表示是铜牌用户. x 模 16 的余数表示应用所属于的 ONU.随机产生一个 1~3 之间的整数表示它选择的业务类:0 表示 CBR,在 48~500 之间随机产生一个整数,表示它的 MTU;1 表示 VBR,在 48~1500 之间随机产生一个整数表示它的 MTU(maximum transfer unit);2 表示 BE,在 48~578 之间随机产生一个整数表示它的 MTU.为了简单起见,本仿真中不考虑实际网络环境中经常出现的残包现象,也就是不考虑小于 MTU 的包.由于本机制产生的 USR 小于 ONU 的所有队列中所有数据包,考虑残包的性能改善要比不考虑的优越.按照应用的数据包到达的先后次序来排列它们的数据包.在周期 m 中,用户终端 z 的第 n 个数据包的到达的时间为

$$t_z^i(m, n) = \frac{L_{package}^i \times 8 \times (n-1)}{w^i}, n \leq \frac{w^i \times cycle(m)}{L_{package}^i \times 8} \tag{18}$$

仿真结果及讨论

(1) 消除 UWQ 效果比较

在满负载情况下(1600 用户同时在线情况下),我们在周期长度为 8ms,9ms 和 9.95ms 的情况下,每小时统计一次平均的再生 UWR 值,仿真结果如图 9 所示.从图 9 可以看出:周期越小,再生 UWR 越大.主要原因为:在满负载和最大周期相等的情况下,对不同长度的周期,一个 ONU 获得的授权窗口的大小差别不大,但周期越短,在一个周期内,队列缓存数据的长度越短,因此,授权超过需求的概率增加,再生 UWR 增加.仿真结果显示:在周期长度为 8ms,9ms 和 9.95ms 的情况下,大约可以节省的带宽分别为 100Mbps,40Mbps 和 10Mbps.

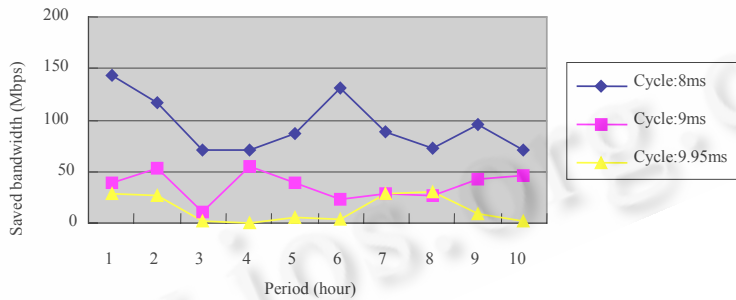


Fig.9 The saved bandwidth by Inter-ONU

图 9 Inter-ONU 节省带宽情况

在周期长度为 8ms,9ms 和 9.95ms 的情况下,每小时统计一次消除的 UWR 大小,仿真结果如图 10 所示.

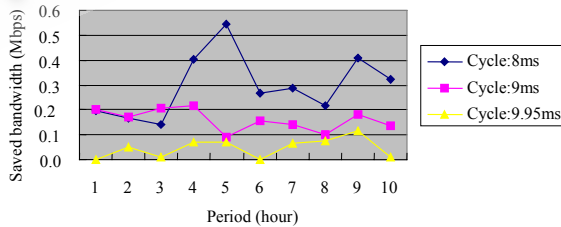


Fig.10 The saved bandwidth by eliminating UWR

图 10 机制消除 UWR 的效果

由图 10 可以看出:周期越小,消除的 UWR 越多,主要原因与图 9 相同.仿真结果显示:在周期长度为 8ms,9ms

和 9.95ms 的情况下,大约可以节省的带宽分别为 0.35Mbps,0.15Mbps 和 0.05Mbps,是本文提出机制的约 1/287,1/268 和 1/201.

(2) 消除 UQR 效果比较

在所有公开发表的文献中,我们还未发现关于消除 UQR 的机制的论述.在满负载情况下,我们在周期长度为 8ms,9ms 和 9.95ms 时,每小时统计一次平均的 UPR 的值,仿真结果如图 11 所示.从仿真结果我们看出,UQR 的大小和周期的大小之间没有关系,其主要原因为业务量是随机的.图 11 显示,效用 Intra-ONUs DBA 算法产生的平均 UQR 大小为 4Mbps.

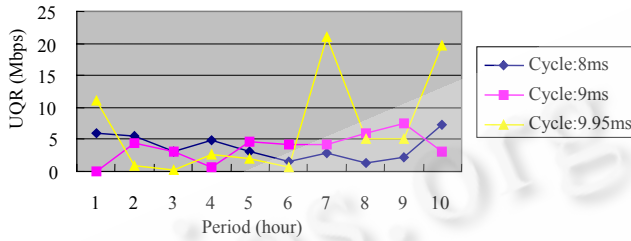


Fig.11 The saved bandwidth by Intra-ONU DB

图 11 Intra-ONU DBA 节省带宽

(3) UPR 消除效果比较

在已发表的文献中,我们还未发现消除 UPR 机制的相关论述,在满负荷和周期长度为 9.95ms 的情况下,我们统计了分布式 UPR 消除方法使用前后第 3 个 ONU 每小时平均 UPR 的大小,仿真结果如图 12 所示.

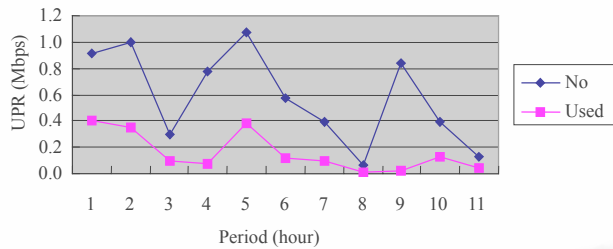


Fig.12 The saved bandwidth by eliminating UPR

图 12 消除 UPR 的效果

图 12 显示:在分布式 UPR 方法使用之前,总 UPR 的大小平均为 0.6Mbps;使用之后,值为 0.1Mbps.

(4) USR 消除效果比较

本文提出了基于交织接力棒机制来提高消除 USR 条件满足的概率,而文献[5]没有提出使用任何措施.在满负荷和周期长度为 9.95ms 时,统计了 1s 内本机制和文献[5]机制消除 USR 的次数,仿真结果见表 1.

从表 1 结果可以看出:采用交织接力机制后,满足 USR 消除条件的次数增加了.

Table 1 The eliminated USR number between our mechanism and the mechanism in Ref.[5]

表 1 本机制和文献[5]机制 1s 内 USR 消除次数比较

Time (s)	Our	In Ref.[5]	Time (s)	Our	In Ref.[5]
1	986	986	6	969	967
2	977	977	7	979	976
3	956	956	8	100 8	100 4
4	966	963	9	951	949
5	997	996	10	100 1	998

本机制把满足消除 USR 条件的交棒者的 USR 追加到接棒者的授权中,而文献[5]的机制则是把它删除,因此,本机制可维持周期的长度并提高带宽的使用效率.在满负荷和周期长度为 9.95ms 时,我们统计了 1 小时本机制和文献[5]机制平均节省的带宽,仿真结果如图 13 所示.

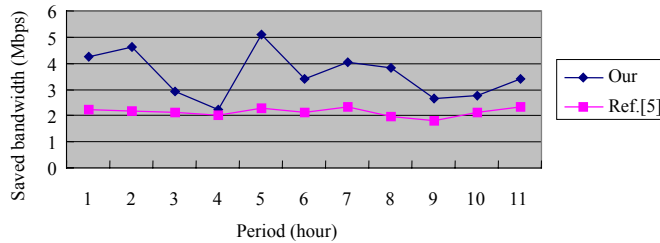


Fig.13 The saved bandwidth by eliminating USR

图 13 USR 消除效果

图 13 表示,本机制在消除 USR 后可以节省约 3.5Mbps 的带宽,而文献[5]的机制可以节省约 2Mbps 带宽。

综上所述,在满负荷情况下,本机制总共节省的带宽大约为 $10+16\times(4+(0.6-0.1))+3.5=85.5\text{Mbps}$,约是 EPON 线速率的 8.55%。而文献[5]论述的机制约为 0.2%。

4 结论

在 IPACT 的基础上,本文提出了一个基于效用的 EPON 分布式 DBA 实现机制,通过基于效用 Inter-ONUs DBA 算法,在网络拥塞的情况下,按照 ONU 的动态平均 SLA 来分配资源,同时采用递归效用算法,有效消除 UWQ,提高了带宽的使用率。通过基于效用的 Intra-ONUs DBA 算法,在网络发生拥塞时,在一个 ONU 内部按照队列的动态平均 SLA 来分配调度资源,同时采用递归效用算法,消除 UQR,提高了网络的使用效率。以上两个机制共同实现了对不同 SLA 类用户的相同业务类业务应用的差分处理。通过分布式 UPR 消除机制,将一个 ONU 总 UPR 的数值减少到一个周期所有队列的所有缓存包最小的包长度之内。在文献[5]提出的 USR 消除机制的基础上,通过建立交织接力机制来提高 USR 消除条件满足的概率。通过将满足交接棒条件的交棒者的 USR 追加到接棒者的授权中,维护了周期的大小,提高了带宽的使用效率。

仿真结果显示,本机制大约可以节约 8.55%带宽资源,高于其他任何机制。

References:

- [1] IEEE 802.3 Ethernet in the first mile study group. <http://www.ieee802.org/3/efm/public/index.html>
- [2] Chen J, Ao L, Liu Q, Zhang KB, Xie YP, He Y, Song XY. Technical requestments for access network-passive optical network based on Ethernet (EPON). YD/T 1475 2006 (in Chinese).
- [3] Wang Y, Wang ZL, Cheng L, Zhan ZQ, Zhai H, Zhang ZB. Requirements and analysis for the management interface of Ethernet passive optical network (EPON). ITU-T Recommendation Q.838.1, 2004.
- [4] Miyoshi H, Inoue T, Yamashita K. QoS-Aware dynamic bandwidth allocation scheme in Gigabit-Ethernet passive optical networks. In: Proc. of the IEEE ICC 2004. Paris: IEEE Communications Society Press, 2004. 90-94.
- [5] Chen B, Chen JJ, He SL. Efficient and fine scheduling algorithm for bandwidth allocation in Ethernet passive optical networks. IEEE Journal of Selected Topics In Quantum Electronics, 2006,12(4):653-660.
- [6] Kramer G, Mukherjee B. IPACT: A dynamic protocol for an Ethernet PON (EPON). IEEE Communication Magazine, 2002,40(2):74-80.
- [7] Zhu YQ, Ma M, Cheng TH. IPACT with grant estimation for EPON. In: Proc. of the IEEE Singapore Int'l Conf. on Communication Systems, ICCS 2006. Reading: IEEE Communications Society Press, 2006. 1-5.
- [8] Bhatia S, Garbuzov D, Bartos R. Analysis of the gated IPACT scheme for EPONs. In: Proc. of the IEEE Int'l Conf. on Communications, Vol.6. Istanbul: IEEE Communications Society Press, 2006. 2693-2698.
- [9] Sherif RS, Hadjiantonis A, Ellinas G, Assi C, Ali MA. A novel decentralized Ethernet-based PON access architecture for provisioning differentiated QoS. Journal of Lightwave Technology, 2004,22(11):2483-2497.
- [10] Chang KC, Liao WJ. On the throughput and fairness performance of TCP over Ethernet passive optical networks. IEEE Journal on Selected Areas in Communications, 2006,24(12):3-12.

- [11] Assi CM, Ye YH, Dixit S, Ali MA. Dynamic bandwidth allocation for quality-of-service over Ethernet PONs. IEEE Journal of Selected Topics in Communication, 2003,21(9):1467-1477.
- [12] Choi SI, Huh JD. Dynamic bandwidth allocation algorithm for multimedia services over Ethernet PONs. Journal of Electronics and Telecommunications Research Institute, 2002,24(6):465-468.
- [13] Kramer G, Banerjee A, Singhal NK, Mukherjee B, Dixit S, Ye YH. Fair queueing with service envelopes (FQSE): A cousin-fair hierarchical scheduler for subscriber access networks. IEEE Journal Selected Areas Communication, 2004,22(8):1497-1513.
- [14] Ghani N, Shami A, Assi C, Raja MYA. Intra-ONU bandwidth scheduling in Ethernet passive optical networks. IEEE Communication Letters, 2004,8(11):683-685.
- [15] Chen J, Chen B, He S. A novel algorithm for intra-ONU bandwidth allocation in Ethernet passive optical networks. IEEE Communication Letters, 2005,9(9):850-852.
- [16] Lallukka S, Raatikainen P. Link utilization and comparison of EPON and GPON access network cost. In: Proc. of the IEEE Globecom. St. Louis: IEEE Communications Society Press, 2005. 301-305.
- [17] Zhang JY, Meng LM, Qiu XS. A real time accounting management with constant price for IP-DiffServ network. Journal of Beijing University of Posts and Telecommunications, 2005,32(3):66-70 (in Chinese with English abstract).
- [18] Zhang JY, Meng LM, Qiu XS, Li HH, Liu F. An SLS based edge-to-edge QoS management mechanism in access network. Journal of Electronics and Information Technology, 2005,27(7):1141-1446 (in Chinese with English abstract).

附中文参考文献:

- [2] 陈杰,敖立,刘谦,张凯宾,谢云鹏,何岩,宋小燕.接入网络技术要求-基于以太网方式的无源光网络(EPON).YD/T 1475,2006.
- [17] 张晋豫,孟洛明,邱雪松.一个恒价的IP接入网实时计费管理机制.北京邮电大学学报,2005,32(3):66-70.
- [18] 张晋豫,孟洛明,邱雪松,李红辉,刘峰.一个基于SLS的接入网边到变QoS管理机制.电子与信息学报,2005,27(7):1141-1446.



张晋豫(1967-),男,河南洛阳人,博士,副教授,主要研究领域为宽带网络技术,IPTV.



刘犁(1982-),女,博士生,主要研究领域为计算机网络,通信协议.