

文章编号 1004-924X(2009)05-1134-07

# 突发交换网边缘节点汇聚模块的设计

付明磊<sup>1</sup>, 乐孜纯<sup>2</sup>

(1. 浙江工业大学 信息工程学院, 浙江 杭州 310014; 2. 浙江工业大学 理学院, 浙江 杭州 310014)

**摘要:** 为了提高光突发交换(OBS)网络的汇聚性能,以 OBS 边缘节点的汇聚模块为研究对象,提出了一种新型的 OBS 汇聚模块设计方案。与以往的汇聚模块采用开环控制的方式不同,本文设计的汇聚模块通过引入突发数据包(BDP)长度反馈,将一个基于时间门限的汇聚模块变成了一个闭环控制系统。汇聚模块通过测量网络业务流量,根据控制曲线,为定时器选择合适的汇集时间门限,同时根据 BDP 长度反馈值修正控制曲线。仿真结果表明:与基于时间门限的汇聚模块相比,本文设计的汇聚模块在控制 BDP 长度性能上改进了 4.95%;汇聚时延降低了 15%。因此,该汇聚模块不仅能够使汇聚时间门限随网络流量的变化而动态改变,而且能够有效控制输出 BDP 长度的分布状况。

**关键词:** 汇聚模块;光突发交换;自适应控制;边缘节点

**中图分类号:** TP393 **文献标识码:** A

## Design of assembly modular for OBS edge node

FU Ming-lei<sup>1</sup>, LE Zi-chun<sup>2</sup>

(1. Zhejiang University of Technology, College of Information Engineering, Hangzhou 310014, China;

2. Zhejiang University of Technology, College of Science, Hangzhou 310014, China)

**Abstract:** With the purpose of improving the assembly performance of an Optical Burst Switching (OBS) network, the assembly modular for the edge node of the OBS is studied in this paper and a novel design scheme for Assembly Controllers (AC) is proposed. By introducing the feedback of BDP sizes, a time-threshold based modular is changed into a closed-loop system different from former assembly modules using an open-loop system. By measuring the network traffic flow, a suitable time-threshold for a timer is chosen according to the control curve of the novel modular, meanwhile, the control curve is modified according to the BDP size feedback. Simulation results show that the novel modular can improve the distribution of BDP size by 4.95%, and decrease the assembly delay by 15% as compared with that of the time-threshold based modular. Therefore, the novel modular can not only adjust the time threshold according to the traffic flow, but also can control the distribution of BDP sizes.

**Key words:** assembly modular; optical burst switching; self-tuning control; edge node

收稿日期:2008-08-12;修订日期:2008-08-27.

基金项目:浙江省自然科学基金资助项目(No. Y1080172)

## 1 引言

光突发交换(OBS)技术是实现未来高速、大容量传输网络的一项富有前景的技术<sup>[1-2]</sup>。它融合了光电路交换(Optical Circuit Switching, OCS)技术和光分组交换(Optical Packet Switching, OPS)技术的优点,同时避免了OCS技术交换粒度大,无法实现统计复用,以及OPS技术中光RAM等技术不成熟的缺点,因此,近年来得到光通信领域的广泛关注。

在OBS网络中,网络节点分为边缘节点与核心节点两类。边缘节点负责将来自子网的数据包汇聚为突发数据分组(Burst Data Packet, BDP)及相应的突发控制分组(Burst Header Packet, BHP),并完成数据信道调度等功能。核心节点负责在电域中处理提前发送的BHP,并通过解析BHP携带的消息信令,完成对光交叉矩阵的配置,以实现为BDP进行路由选择和波长分配。汇聚模块是OBS边缘节点的一个重要组成部分,它主要完成数据包的分类、缓存和汇聚功能。汇聚模块对OBS网络输入业务流量变化、BDP丢包率和传输时延都具有重要的影响<sup>[3-4]</sup>。

在汇聚模块中,汇聚规则决定了BDP的产生条件。在文献中,汇聚规则可以分为基于时间门限的汇聚规则和基于长度门限的汇聚规则两类<sup>[1,3-4]</sup>。而按照门限的性质划分,汇聚规则又可以分为固定门限汇聚规则和自适应门限汇聚规则<sup>[5-11]</sup>。但是,从控制系统的角度分析,现有的大多数汇聚模块属于开环控制方式,即汇聚模块一般通过控制定时器或者计数器来完成对输入数据包的组装,而缺乏对组装完成后的BDP的相关信息反馈处理。这可能导致基于时间门限的汇聚规则无法有效控制BDP的字节长度分布,从而导致BDP字节长度变化范围过大而影响BDP的调度环节,造成BDP丢包率增加、传输时延加大等不良情况<sup>[3-4]</sup>。

本文给出了一种新型的汇聚模块设计方案,它采用基于自适应时间门限的汇聚规则,即时间门限能够根据网络流量自动调节。同时,汇聚模块引入了BDP长度反馈信息,通过反馈信息与历史均值的比较来调整时间门限控制曲线,使得时间门限控制曲线适合当前的网络流量特征。

## 2 OBS边缘节点的基本结构及功能

### 2.1 OBS边缘节点的基本结构

图1给出一种OBS边缘节点的基本结构。它主要由汇聚模块、调度模块、信号处理模块、队列矩阵模块以及输入、输出接口组成,其中汇聚模块和调度模块是边缘节点的核心部分。汇聚模块首先把来自输入接口的数据包按照目的地址和服务等级(QoS)要求,进行分类并发送到所对应的缓存队列中。然后,按照事先设定的汇聚规则将这些数据包组装为BDP,并产生相应的BHP。调度模块则根据BDP的类型和等级,根据一定的调度算法为BDP分配传输时隙和光通道。最后,BDP从数据信道组(Data Channel Group, DCG)输出,而BHP从控制信道组(Control Channel Group, CCG)输出。

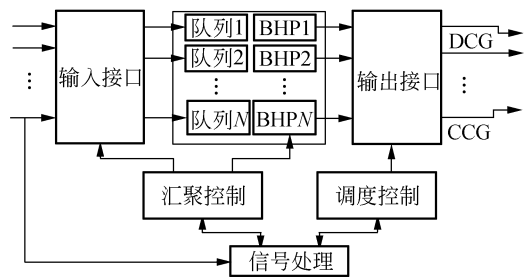


图1 OBS边缘节点的基本结构

Fig. 1 Structure of OBS edge node

### 2.2 OBS边缘节点的功能细化

为了使OBS边缘节点的基本功能更明确,本文给出了节点的发送部分和接收部分的功能模块,如图2所示。

由图2的功能划分可以看出:汇聚模块无论

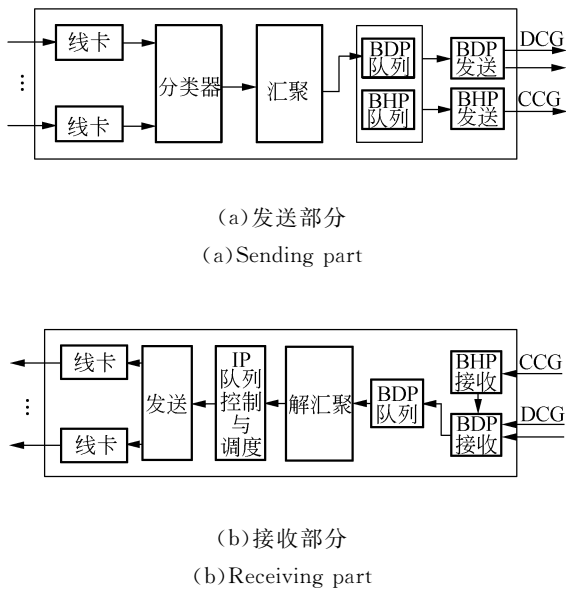


图 2 OBS 边缘节点的功能模块

Fig. 2 Function modules in OBS edge node

在发送部分还是接收部分都具有重要的作用,它决定了 BDP 和 BHP 的产生条件以及其长度分布。

### 3 汇聚模块的主要设计参数

在汇聚模块的设计过程中,主要的设计参数有两个:一个是汇聚时延  $T_a$ ,另一个是 BDP 长度  $L_B$ 。

#### 3.1 汇聚时延要求

BDP 在边缘模块的总时延可分为 3 个部分:BDP 汇聚时间  $T_a$ 、处理时延  $T_q$ 、偏置时间  $T_{off}$ 。如果综合考虑 BDP 在 OBS 网络中传输时延  $T_p$ 、波长适配时间  $T_{adp}$  以及反向拆卸时延  $T_{rec}$ ,设  $T_d$  为实时业务可允许的最大端到端时延,那么汇聚时延应满足:

$$T_a < T_d - T_q - T_{off} - T_p - T_{adp} - T_{rec}. \quad (1)$$

因此,由式(1)可知,BDP 的汇聚时延具有时延上限。

#### 3.2 BDP 长度要求

BDP 汇聚时延上限的存在使得 BDP 长度也具有最大突发长度的限制,记为  $L_{max}$ 。同时,由于每一个 BDP 均需产生一个对应的 BHP,在 CCG 容量小于 DCG 容量的情况下,过短的 BDP 长度将引起控制信道的拥塞,从而导致 OBS 网络丢包率增加。

假设当前边缘节点输出接口  $K$  个信道,包含了  $k$  个 CCG,  $K-k$  个 DCG。令  $L_b$  为 BDP 的平均长度,  $L_c$  为 BHP 的平均长度,考虑 CCG 和 DCG 满负荷的极限情况,此时 CCG 的最大发包速率为  $k/L_c$ ,DCG 的最大发包速率为  $(K-k)/L_b$ 。为保证汇聚模块不产生拥塞,必须使 CCG 的最大发包速率大于 DCG 的最大发包速率,即:

$$(K-k)/L_b < k/L_c. \quad (2)$$

由式(2)可得:  $L_b > \frac{K-k}{k} \cdot L_c$ , 记  $L_{min} = \lceil \frac{K-k}{k} \cdot L_c \rceil$ , 那么  $L_{min} < L_b < L_{max}$ , 即 BDP 长度存在一个变化范围。

## 4 汇聚模块中的汇聚规则

### 4.1 固定门限汇聚规则

固定门限汇聚规则是对汇聚模块的两个主要设计参数(汇聚时延和 BDP 长度)做静态设置的方式。可以分为:固定时间门限(Fixed-Assembly-Period, FAP)汇聚规则和固定长度门限(Fixed-Assembly-Size, FAS)汇聚规则。这两种汇聚规则的思想类似,即以单一的固定组装时延或者 BDP 长度作为门限值,对数据包进行组装。当第一个数据包进入队列后开始计时(或者计数),一旦达到或超过了所设定的时间门限值(或者长度门限值),BDP 被立即组装<sup>[2,3,6]</sup>。下面给出 FAS 汇聚规则的描述。

假设  $b$  表示当前达到的数据包的字节数,  $q(i)$  表示第  $i$  个队列当前的队列长度,  $L$  表示算法设定的 BDP 长度门限,则 FAS 汇聚规则的伪代码描述如下:

当一个字节数为  $b$  的数据包被分配到队列  $i$  后:

```

if( $q(i) = 0$ )
     $q(i) = b$ ; // 记录当前队列的长度为  $b$ 
else if ( $q(i) + b < L$ )
     $q(i) = q(i) + b$ ; // 将数据包插入队列  $i$ 
else
    { 立即产生一个长度为  $q(i)$  的 BDP 及对应的 BHP;
       $q(i) = b$ ; // 更新当前队列的长度
    }

```

固定门限汇聚规则步骤简单,容易实现,但是由于其门限值固定,无法随网络业务流量而动态调整,因此将导致较长的传输时延(对于FAS)或者形成较大的BDP长度(对于FAP)。

#### 4.2 混合门限汇聚规则

混合门限汇聚规则兼顾了汇聚模块对组装时延和BDP长度的敏感性,将最大组装时延 $T$ 设定为上限,最小BDP长度 $L$ 设定为下限。

$t(i)$ 表示队列 $i$ 的计时器,其它参量定义与4.1相同,这里不再重复定义。

当一个字节数为 $b$ 的数据包被分配到队列 $i$ 后:

```

if( $q(i) = 0$ )
     $q(i) = b$ ; //记录当前队列的长度为 $b$ 
else if ( $q(i) + b < L$ )
     $q(i) = q(i) + b$ ; //将数据包插入队列 $i$ 
else
{
    立即产生一个长度为 $q(i)$ 的BDP及相应的BHP;
     $q(i) = b$ ; //更新队列 $i$ 的长度
     $t(i) = 0$ ; //定时器清0
}
if( $t(i) = T$ )
{
    立即产生一个长度为 $q(i)$ 的BDP及相应的BHP;
     $q(i) = 0$ ; //队列长度清0
     $t(i) = 0$ ; //定时器清0
}

```

经仿真:当网络处于低流量状态时,混合门限汇聚规则接近于FAP规则;而在高流量状态下,它则接近于FAS规则<sup>[6]</sup>。因此,混合门限汇聚规则仍然属于固定门限汇聚规则。

#### 4.3 自适应门限汇聚规则

为了能够适应网络流量的动态变化,汇聚模块采用的汇聚规则应该具有自适应调节汇聚参数的功能。在目前已经报道的文献中,自适应门限汇聚规则可以分为两类:一类是自适应调节单一汇聚参数<sup>[6-9]</sup>;另一类是同时自适应调节两个汇聚参数<sup>[10-11]</sup>。后者虽然从理论的角度出发,具有更好的性能,但是算法复杂度增加,参数设定困难。而且两个汇聚参数之间存在一定的约束关系,如

组装时延限制最大BDP长度等,因此本文只考虑调节单一参数的自适应汇聚规则。

在调节单一参数的自适应汇聚规则中,单一参数经常选择为BDP长度。如文献[7]是根据计量器所计划得的各相应队列的数据包的到达速率来动态改变BDP的大小;文献[8]则是引入BDP长度限制门限、调节步长和调节容限等,通过监测流量负荷的变化来调节汇聚门限值;文献[9]采用模糊控制的思想,引入了一种模糊变长门限,根据边缘节点数据流量的动态变化,来改变汇聚长度门限。

以上的汇聚规则偏重于通过检测或者估计数据包的流量变化来改变汇聚长度门限值,无法有效地控制汇聚时延。

## 5 本文给出的汇聚模块设计方案

与采用BDP长度作为自适应调整参数的思路不同,本文采用汇聚时延作为自适应调节参数。其优势在于有效地控制汇聚时延,以满足对时延敏感的实时业务的需求。同时,为了控制BDP的长度分布,在汇聚模块中引入了BDP长度反馈,以修正控制曲线。这样,汇聚模块从一个开环控制系统变为一个闭环控制系统,在根据数据包的流量变化自适应调节汇聚时延的同时,也能够控制BDP的长度分布。

### 5.1 汇聚模块的结构框图

本文给出一种新型的OBS汇聚模块设计方案。它由流量检测、缓存队列、汇聚队列和汇聚控制器4个部分组成,如图3所示。流量检测部分负责统计在测试周期 $T_s$ 内经过的数据包数目,并负责将检测结果 $P\_num$ 输出到汇聚控制器中。流量监测部分受汇聚队列控制,当控制信号Detector\_en信号有效时,内部计时器开始计时,直到计时时限到达 $T_s$ 时,停止记录数据包的数目,并将记录结果通知汇聚控制器。缓存队列是一个单输入单输出的FIFO,只有在使能信号Buffer\_en有效时才能输出缓存的IP分组。汇聚控制器通过 $P\_num$ 确定定时门限值Timer\_th,并以此门限值控制汇聚队列组装IP分组成为突发数据包Data Burst。在Data Burst组装完成后,将突发包长度值反馈给汇聚控制器进行控制曲线修正。

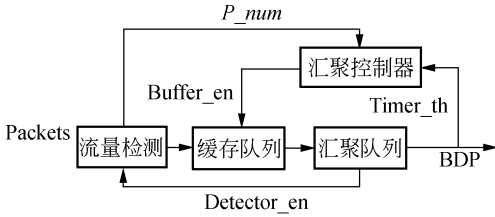


图 3 控制器结构框图

Fig. 3 Diagram of assembly controller

### 5.2 汇聚模块的主要参数

对于控制器中的定时器, 设  $T_{max}$  表示最大汇聚时限,  $T_{min}$  表示最小汇聚时限,  $T_{step} = (T_{max} - T_{min})/N$  表示汇聚时限调整步长,  $N$  表示最大调整步长数;  $T_s$  表示流量测试周期,  $0 < T_s \leq T_{min}$ ;

对于流量检测部分, 设  $P\_num$  表示在  $T_s$  内进入缓存队列的 IP 数据包的总数,  $L_b$  表示 BDP 的长度, 那么由 3.2 节可知,  $L_b$  满足下面的约束条件:  $L_{min} < L_b < L_{max}$ , 其中  $L_{min}$  表示 BDP 的长度下限,  $L_{max}$  表示 BDP 的长度上限。

### 5.3 主要参数的函数关系

对于每一个  $P\_num$ , 控制器都可以找到一个时间门限值  $T_A$  与之对应。

#### 5.3.1 $P\_num$ 与 $T_A$ 的函数关系

$$T_A = \begin{cases} T_{max} & P\_num \leq L_{min} \\ T_{max} - T_{step} & L_{min} < P\_num \leq L_{min} + k \cdot P\_num_{step} \\ \vdots & \\ T_{min} + T_{step} & L_{max} - k \cdot P\_num_{step} \leq P\_num < L_{max} \\ T_{min} & P\_num \geq L_{max} \end{cases}$$

$P\_num_{step} = (L_{max} - L_{min})/M$  表示汇聚队列中数据包数量的调整步长,  $M$  表示最大调整步长数,  $M = k \cdot N$ ,  $k$  是一个正整数。

图 4 是  $P\_num$  与  $T_A$  的函数关系的直观表述。其中,  $T_{max} = 100$ ,  $T_{min} = 10$ ,  $N = 9$ ,  $L_{min} = 20$ ,  $L_{max} = 200$ ,  $k = 4$ 。

#### 5.3.2 $L_b$ 与 $T_A$ 的修正关系

$T_A'$  表示修正后的时间门限值,  $ave\_L_b$  表示当前时刻之前  $L$  个突发包长度的均值。

$$ave\_L_b = \frac{1}{L} \sum_{i=0}^L L_b(i),$$

$$T_A' = \begin{cases} T_A - \theta \cdot T_{step} & L_b - ave\_L_b > R \\ T_A & L_b - ave\_L_b \leq R \\ T_A + \theta \cdot T_{step} & L_b - ave\_L_b < -R \end{cases}$$

其中,  $\theta$  是一个与  $R$  有关的常量, 表示修正幅

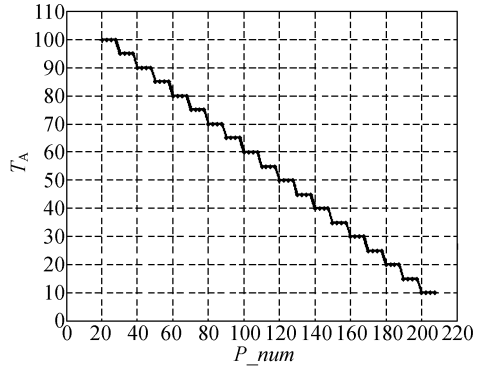


图 4  $P\_num$  与  $T_A$  的函数关系

Fig. 4 Relationship between  $P\_num$  and  $T_A$

度。

## 6 仿真结果

本文设计了一个简单的仿真场景模拟 OBS 汇聚节点, 并对引入 OBS 汇集节点控制器前后的突发包长度分布情况及汇聚时间门限的变化情况进行了比较。业务量部分采用了 10 组随机采样的 IP 数据流量, 并且 IP 数据包具有相同的长度。流量检测部分的采样周期为 1 s, 缓存队列和汇聚队列的队列长度为 500 个 IP 数据包的容量。

### 6.1 BDP 长度分布性能仿真

本文首先仿真的是引入 BDP 长度反馈前后, 汇聚模块输出的 BDP 长度的分布情况。图 5、图 6 为仿真结果, 其中图例中 AC 是汇聚控制器 (Assembly Controller) 的简写。仿真结果表明: 在采用了本文给出的汇聚模块后, BDP 长度的变化幅度被控制在了一个允许的变化范围 (本次仿真设定的范围为  $100 \pm 10$ )。因此, 与一般的基于

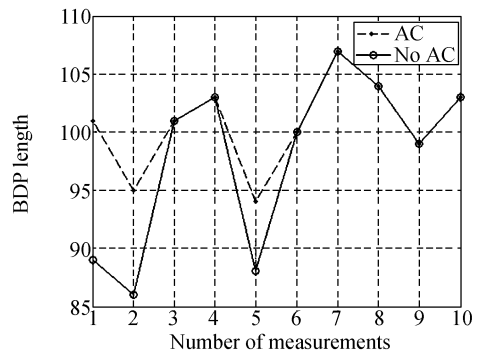


图 5 BDP 长度变化比较结果(1)

Fig. 5 Simulation result of the change of BDP length (1)

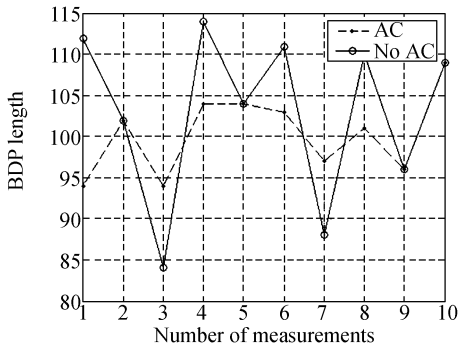


图 6 BDP 长度变化比较结果(2)

Fig. 6 Simulation result of change of BDP length (2)

时间门限的汇聚模块相比(图 5、图 6 中的实线)相比,本文设计的汇聚模块(图 5、图 6 中的虚线所示)改进了 BDP 长度的分布状况,为 OBS 突发包调度部分的设计降低了困难。

图 7 是对图 5、图 6 仿真结果的统计结果。统计后得出:本文设计的汇聚模块比基于时延的汇聚模块在控制 BDP 长度性能上改进了 4.95%。

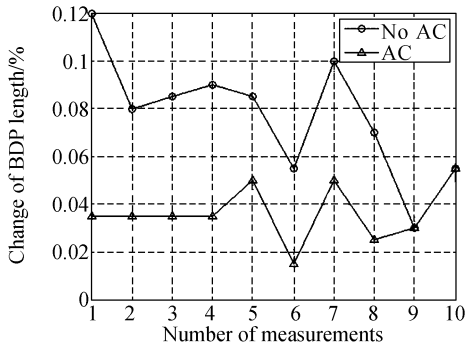


图 7 BDP 长度变化百分比统计结果

Fig. 7 Statistics of change of BDP length

### 6.2 汇聚时间门限性能仿真

基于时间门限的汇聚模块受到一个固定门限的定时器控制,因此它的时间门限不会随着网络业务流量的变化而改变,如图 8、图 9 实线部分所示。本文设计的汇聚模块能够根据采集的业务流量,动态改变汇聚时间门限(如图 8、图 9 虚线部分所示),因此能够使得汇聚模块适应网络流量的变化。同时,图 8、图 9 的汇聚时延统计结果表示:本文设计的汇聚模块的汇聚时延与基于时间门限的汇聚模块相比降低了 15%。

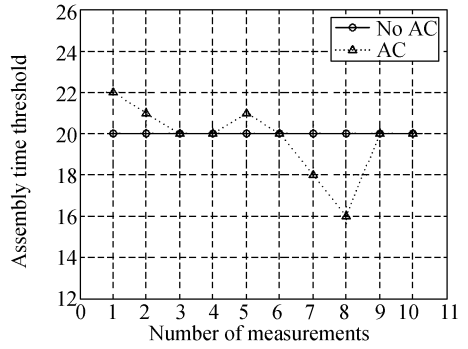


图 8 汇聚时间门限变化比较结果(1)

Fig. 8 Simulation result of the assembly time-threshold (1)

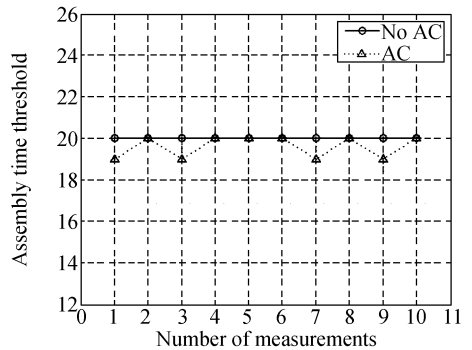


图 9 汇聚时间门限变化比较结果(2)

Fig. 9 Simulation result of assembly time-threshold (2)

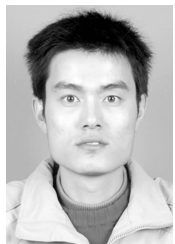
## 7 结 论

汇聚模块的性能优劣对整个 OBS 网络的数据传输性能具有重要的影响。本文设计的汇聚模块能够根据网络业务流量的变化,动态地改变定时器的时间门限,因此能够动态调节汇聚时间,有效降低网络时延,并且通过引入 BDP 长度反馈,汇聚模块能够将输出 BDP 长度控制在一个允许范围内(由调度部分决定),从而能够避免由于 BDP 长度变化过大而引起的丢包问题。仿真结果表示:与基于时间门限的汇聚模块相比,本文设计的汇聚模块在控制 BDP 长度性能上改进了 4.95%;汇聚时延降低了 15%。

## 参考文献:

- [1] CHRISTODOULOPOUOS K, VARVARIGOS E, VLACHOS K. A new burst assembly scheme based on the average packet delay and its performance for TCP traffic [J]. *Optical Switching and Networking*, 2007, 4(3):200-212.
- [2] REVIRIEGO P, HERNÁNDEZ J A, ARACIL J. Analysis of average burst-assembly delay and applications in proportional service differentiation [J]. *Photon Netw Commun.*, 2007, 14:183-197.
- [3] KANTARCIA B, OKTUGA S F, ATMACA T. Performance of OBS techniques under self-similar traffic based on various burst assembly techniques [J]. *Computer Communications*, 2007, 30: 315-325B.
- [4] TACHIBANA T, KASAHARA S. Performance analysis of timer-based burst assembly with slotted scheduling for optical burst switching networks [J]. *Performance Evaluation*, 2006, 63: 1016-1031.
- [5] WANG R Y, WU D P, LONG K P, *et al.*. A novel optical burst assembly technique based on a dynamic threshold mechanism to support scalable QoS [J]. *Photonic Network Communications*, 2003, 6(2):161-168.
- [6] YANG X L, ZHANG M, LI L M. Burst assembly schemes and performance evaluation in optical burst switching networks [J]. *The Journal of China Universities of Posts and Telecommunications*, 2004, 11(3):27-31.
- [7] 汪纪锋, 蒋玉莲, 夏汉铸. OBS 网络中的自适应汇聚算法[J]. *重庆大学学报(自然科学版)*, 2005, 28(5): 90-93.
- WANG J F, JIANG Y L, XIA H ZH. An adaptive data burst assembly algorithm in OBS networks[J]. *Journal of Chongqing University(Natural Science Edition)*, 2005, 28(5):90-93. (in Chinese)
- [8] 李彦君, 洪小斌, 郭宏翔, 等. 光突发交换网中的智能组包研究[J]. *电子学报*, 2005, 33(11): 2040-2043.
- LI Y J, HONG X B, GUO H X, *et al.*. Study on intelligent burst assembly mechanism in optical burst switch networks[J]. *Acta Electronica Sinica*, 2005, 33(11):2040-2043. (in Chinese)
- [9] 杨九如, 贾世楼, 王钢. 基于 fuzzy-adaptive-threshold 的数据突发组装算法[J]. *哈尔滨工程大学学报*, 2007, 28(6):669-672.
- YANG J R, JIA SH L, WANG G. Burst assembly algorithm based on fuzzy-adaptive-threshold [J]. *Journal of Harbin Engineering University*, 2007, 28(6):669-672. (In Chinese)
- [10] 程红霞, 张玉兴, 吴援明. 改进的 OBS 长度自适应门限组装算法[J]. *电子科技大学学报*, 2006, 35(3):302-304.
- CHENG H X, ZHANG Y X, WU Y M, A modified OBS adaptive assembly algorithm based on length threshold[J]. *Journal of University of Electronic Science and Technology of China*, 2006, 35(3):302-304. (in Chinese)
- [11] 温立, 涂晓东, 王凯, 等. 一种用于 Optical Crossbar 的自适应分组汇聚算法[J]. *电子科技大学学报*, 2005, 34(6):951-954.
- WEN L, TU X D, WANG K, *et al.*. A new adaptive assembly algorithm in optical crossbar for switching networks[J]. *Journal of University of Electronic Science and Technology of China*, 2005, 34(6):951-954. (in Chinese)

## 作者简介:



付明磊(1981—),男,河北唐山人,博士研究生,主要从事光突发交换网络中拥塞控制技术、节点硬件实现及算法等方面的研究。E-mail: fuml\_zjut@yahoo.com.cn

## 导师简介:



乐孜纯(1965—),女,浙江杭州人,教授,博士生导师,主要从事光纤通信网络组网技术和微结构光电子器件等方面的研究。E-mail: lzc@zjut.edu.cn