

## 两种新型的 OBS 网络数据信道调度算法<sup>1</sup>

袁巍 洪佩林 李津生 孙卫强 马明

(中国科学技术大学信息网络实验室 合肥 230027)

**摘要:** 光突发交换 (OBS) 是实现 IP/DWDM 网络的一种极有前途的方案。该文提出了两种新型的 OBS 网络数据信道调度算法: 最少波长转换 (MWCT) 调度算法和区分服务 (SD) 调度算法。通过仿真, 证明了这两种算法在减少 OBS 网络中的光信号的波长转换和保证高优先级数据可靠传输方面比已有的 FF 和 LAUC 调度算法具有更好的性能。

**关键词:** 光突发交换, 光核心路由器, 数据净荷, 自相似特性

**中图分类号:** TN919.3 **文献标识码:** A **文章编号:** 1009-5896(2004)02-0272-06

## New Data Channel Scheduling Algorithms in OBS Networks

Yuan Wei Hong Pei-lin Li Jin-sheng Sun Wei-qiang Ma Ming

(InfoNet Lab, University of Science and Technology of China, Hefei 230027, China)

**Abstract** Optical Burst Switching (OBS) is a promising solution for realizing IP/DWDM. In this paper, two new data channel scheduling algorithms in OBS networks, Minimum Wavelength Conversion Times (MWCT) algorithm and Service Differentiation (SD) algorithm, are proposed. Computer simulation has demonstrated that MWCT algorithm leads to much fewer wavelength conversion times and the SD algorithm can provide better transmission reliability in OBS networks for high priority data comparing with the existing FF and LAUC algorithms.

**Key words** OBS(Optical Burst Switching), Optical core router, Data burst, Self-similarity

### 1 引言

急剧增长的因特网业务要求未来的全光网络 (AON) 具有极高的传输容量并使用高速的路由设备。密集波分复用 (DWDM) 技术被认为是满足骨干网络巨大传输带宽要求的最好方法, 而新一代的高速路由器则为从现有因特网过渡到未来的全光网奠定了基础。未来的光网络去掉了中间层 (如 ATM, SDH 或 SONET 等层), IP 层直接映射到 DWDM 层, 而寻路和光交换平台则被综合成光路由器。

光路由器<sup>[1,2]</sup>有两个主要部件: 光交换单元和 IP 路由单元。光交换单元对通过光路由器的数据在光域内进行交换; IP 路由单元对数据分组进行寻路等处理, 并对光交换单元进行控制 (主要是配置光交换矩阵)。由于 IP 路由单元对数据的处理在电域中进行, 这大大限制了光路由器的转发速率。

解决这一问题的一种方法是提高光路由器的交换粒度。基于这种想法, 人们提出了光突发交换 (Optical Burst Switching, OBS)<sup>[1,3-5]</sup>, 将多个具有相同目的地址和特性 (比如 QoS 参数等) 的 IP 分组作为一个整体, 即光突发单元 (Optical burst) 进行交换。

<sup>1</sup> 2002-08-14 收到, 2003-03-17 改回  
国家自然科学基金项目资助课题 (90104011)

OBS 网络由光核心路由器、边缘路由器及光链路组成。在骨干网络边缘, 来自接入网的 IP 分组在边缘路由器中被汇聚 (Assemble) 成光突发单元, 通过核心路由器的转发在 OBS 骨干网络中传输, 再在目的端的边缘路由器中拆分 (Disassemble) 恢复成一个个的 IP 分组进入对方接入网。光突发单元分为两个部分: 数据净荷 (Data burst) 和控制分组 (Burst Head Packet, BHP)。一般来说, 控制分组比数据净荷提前一段时间 (偏移时间) 从边缘路由器发出, 在控制信道上传输。光核心路由器收到控制分组后, 首先根据其携带的地址信息进行寻路, 然后进行数据信道调度以选择其对应的数据净荷的输出数据信道, 最后对光交换矩阵进行配置, 以对数据净荷进行转发。每个信道对应一个波长, 数据信道和控制信道占用不同波长。

在 OBS 网络研究中, 数据信道调度算法<sup>[3,4]</sup>是一项很重要的研究课题。已有的算法<sup>[3,4]</sup>功能比较简单, 对网络性能的改善大多局限于数据净荷的丢失率, 很少考虑波长转换、QoS 性能等因素。事实上, 波长转换次数和 QoS 性能是影响网络性能的关键因素。波长转换增加了操作的复杂性并影响信号质量, 故应当尽量避免。因此, 本文提出了一种最少波长转换 (Minimum Wavelength Conversion Times, MWCT) 调度算法。仿真证明, 这种算法能够有效地减少网络中的波长转换, 而且不会对网络的其他性能造成明显影响。如何在 OBS 网络中支持 QoS 是一个富有挑战性的课题。OBS 网络应提供类似于 IP 网络的区分服务功能以保证关键业务的服务质量。本文提出了一种在网络拥塞发生时按数据优先级分配网络资源来实现区分服务的调度算法。

本文接下来的部分将首先介绍两种已有的调度算法, 然后介绍最少波长转换调度算法和区分服务 (Service Differentiation, SD) 调度算法, 最后, 给出使用网络仿真软件 (NS) 进行仿真的结果, 并从数据净荷的丢失率、波长转换次数、网络延时等方面对这几种数据信道调度算法的性能进行评价。

## 2 数据信道调度算法

所谓数据信道调度就是指为到达的数据净荷选择一条输出数据信道, 即当控制分组到达核心路由器后, 选择一条在数据净荷到达光交换矩阵时或再经过光纤延迟线 (Fiber Delay Line, FDL) 缓存<sup>[1]</sup>的延时时间之后可用的数据信道作为输出信道; 当没有可用信道时, 必须将此数据净荷和控制分组丢弃。调度算法需要考虑的因素包括数据净荷的丢失率、波长转换次数和 QoS 性能等。

这一节先讨论最基本的 FF (First Fit) 调度算法和最近可用优先 (Latest Available Unscheduled Channel, LAUC) 调度算法<sup>[3]</sup>, 然后在此基础上提出 MWCT 调度算法和 SD 调度算法。在以下的讨论中假定光核心路由器中有  $B$  个 FDL 缓存, 第  $i$  个缓存的延时时间为  $D_i$ ,  $1 \leq i \leq B$ 。

### 2.1 FF 调度算法和 LAUC 调度算法

FF 调度算法按一定顺序搜索数据信道, 首先找到的可用信道作为输出信道。如图 1, 长度为  $L$  的新数据净荷将于  $T$  时刻到达光交换矩阵, 调度程序从信道 0 开始搜索在  $T$  时刻可用的数据信道。由于数据净荷 4 在  $T$  时刻已经占据数据信道 0, 故数据信道 0 阻塞。数据净荷 5 已经预定数据信道 1, 且其使用信道的时间与新数据净荷的使用时间相冲突, 因此数据信道 1 也不可用。信道 2 在  $T$  时刻和  $T + L$  时刻之间一直是空闲的, 因此调度程序不必再查询其他信道的状态, 选中数据信道 2 作为输出信道。若在  $T$  时刻没有可用数据信道, 调度算法继续搜索在  $T + D_i$  ( $1 \leq i \leq B$ ) 时刻可用的数据信道。如果仍然没有可用信道, 则丢弃此数据净荷和相应的控制分组。对于 LAUC 调度算法, 不同之处在于此算法要从所有可用信道中选择一条距离前一个数据净荷尾端最近的信道。从图 1 可见, 数据信道 2, 3 均可用, 但由于  $T - t_2$  小于  $T - t_1$ , 因此选择数据信道 3 作为输出信道。

### 2.2 最少波长转换调度算法

以上介绍的两种调度算法都没有考虑数据净荷在进行交换时的波长转换问题。如前所述,

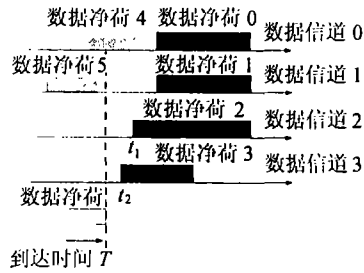


图 1 OBS 网络中数据信道调度算法示意图

波长转换会影响信号的传输质量,并增加了操作的复杂性。因此,在实际应用中应尽可能少地对光信号进行波长转换操作。本算法正是基于这一考虑而提出的。这种算法的基本思想是优先选择与待调度数据净荷的波长相同的数据信道作为输出信道,仅此信道不可用时才使用 FF 算法或 LAUC 算法从其他信道中搜索可用信道。图 2 给出了使用 FF 算法查找可用信道的 MWCT 调度算法的流程图。

### 2.3 区分服务调度算法

在 OBS 网络中,如果网络资源不足以满足突发性的高业务量就会导致拥塞。为保证关键业务的服务质量,要求 OBS 网络提供类似 IP 网络的 SD 功能。文献 [6] 提出了一种利用偏移时间和延时资源预留 (Delayed reservation) 来实现 SD 的方法。但这种方法在网络负载不大时会导致高优先级数据不必要的延时。

本文提出的 SD 调度算法可以保证高优先级数据比低优先级数据在 OBS 网络中更可靠地传输。它的基本思想如下:在网络正常运行时,所有数据净荷同等地占用网络资源;仅当拥塞发生时,使部分信道为高优先级数据净荷所独占,减少低优先级数据净荷所能使用的数据信道数,从而降低了高优先级数据净荷的丢失率。若网络中的数据具有多种优先级,在网络拥塞发生时,数据可使用的信道数与它的优先级大小成正比。数据信道的查找可使用 FF 算法或 LAUC 算法。为简单起见,我们只讨论网络中的数据具有两种优先级的情况。图 3 给出了使用 FF 算法查找可用信道的 SD 算法的流程图。其中,网络拥塞发生的判定条件为:  $P(k) > p_0$ ,  $P(k)$  为第  $k$  个时间间隔的数据净荷丢失率,其值为该时间间隔内成功转发的数据净荷数与此时间间隔内收到的数据净荷数之比,  $p_0$  为一指定的阈值。  $P(k)$  的值应不断更新以实时地反映网络负载情况。

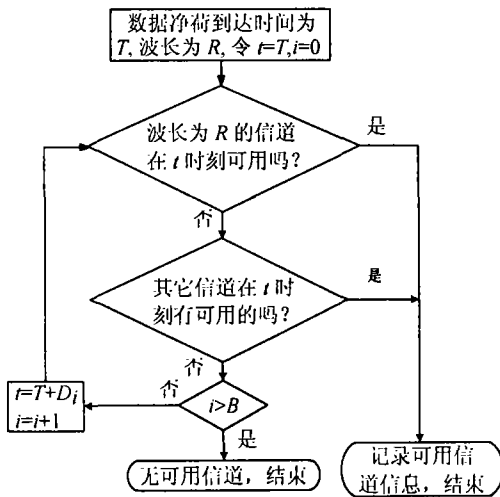


图 2 MWCT 调度算法的流程图

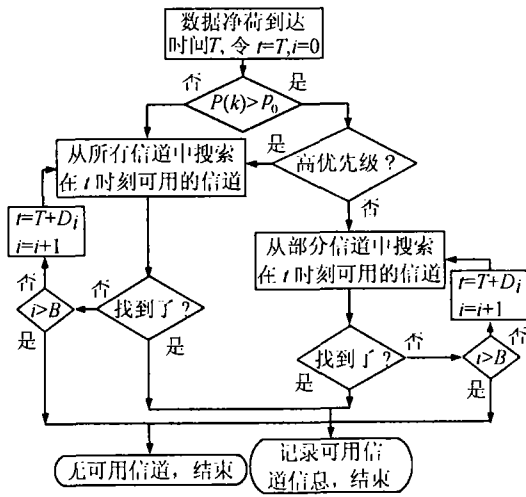


图 3 SD 调度算法的流程图

### 3 数据信道调度算法的仿真

在这一节,我们根据使用网络仿真软件 NS 仿真的结果比较以上几种数据调度算法的性能。

我们使用的仿真网络拓扑如图 4 所示。核心路由器均运行数据信道调度算法，而边缘路由器运行数据突发汇聚算法 (Burst assembly mechanism)<sup>[3]</sup>，将多个 IP 分组汇聚成光突发单元。

最新研究表明，因特网业务具有自相似 (self-similar) 特性<sup>[3,4]</sup>。为模拟具有自相似性的 IP 业务，仿真中将一组 ON/OFF 源模型 (on/off sources model) 的输出数据复合为边缘路由器的输入。这种模型交替处于两种状态：ON 状态和 OFF 状态。处于 ON 状态时，业务模型持续产生分组；处于 OFF 状态时，业务模型停止产生分组。两种状态的保持时间  $T_{ON}$  和  $T_{OFF}$  服从 Pareto heavy-tail 分布<sup>[7]</sup>。自相似特性用 Hurst 参数来衡量，最小值为 0.5。在仿真过程中，取  $H = 0.85$ 。

图 5-7 示出了使用 MWCT 算法对网络性能的改善。图 5 表示使用 MWCT 算法 (使用 FF 算法搜索可用信道) 和 FF 调度算法时网络中发生的波长转换次数的比值。计算表明，与 FF 调度算法相比，使用 MWCT 调度算法可使波长转换次数减少 18%~42%。图 6 比较了 MWCT 调度算法和 FF 调度算法的数据净荷丢失率。可以看到，两条曲线基本吻合，这说明使用了 MWCT 调度算法不会对数据净荷丢失率造成明显影响。图 7 是对这两种算法的延时性能的比较，其中纵坐标表示 MWCT 调度算法和 FF 调度算法的 FDL 延时时间 (平均每个数据净荷在网络中经历的所有 FDL 缓存延时时间) 的比值。仿真结果表明，使用 MWCT 调度算法的 FDL 延时时间比 FF 调度算法平均增加了约 2.1%，这是 MWCT 调度算法尽量选用数据净荷波长所对应的数据信道的代价。由于数据净荷在 OBS 网络中所经历的延时包括 FDL 缓存延时，光纤链路中的传输延时和边缘路由器上的延时，因此，使用 MWCT 调度算法导致的延时的增加与整个端

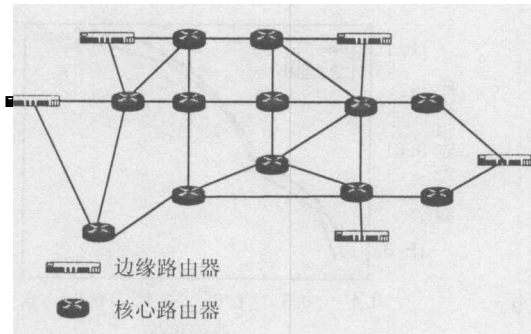


图 4 仿真网络拓扑示意图

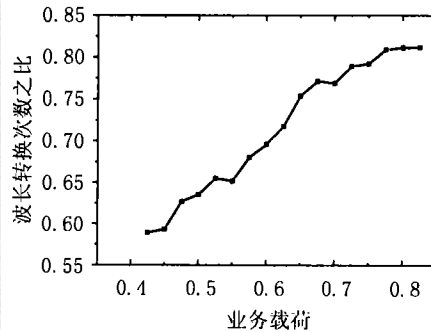


图 5 波长转换次数比率图

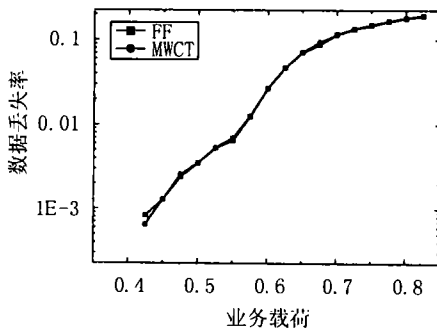


图 6 数据净荷丢失率对比图

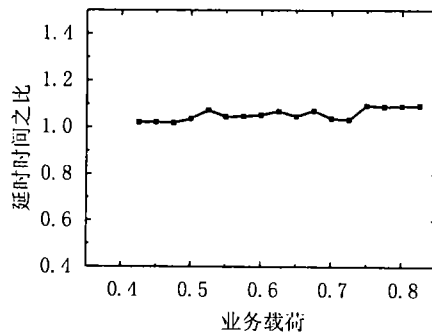


图 7 延时时间比率图

到端延时相比几乎可以忽略不计。仿真还显示使用这两种算法的信道利用率没有明显差异, MWCT 调度算法仅比 FF 调度算法平均低 1.5%。另外, 我们还对使用了 LAUC 算法搜索可用信道的 MWCT 算法和 LAUC 调度算法进行仿真, 由于篇幅限制, 仿真数据不再给出。分析仿真数据发现, 使用 MWCT 调度算法对网络性能的改善与图 5-7 显示的情况基本相同(数据净荷丢失率略有降低)。因此, 可以作出以下结论: MWCT 调度算法能有效降低 OBS 网络中光信号波长转换的次数, 而且不会对数据丢失率等其他性能造成明显影响。

图 8-10 是对 SD 调度算法(使用 FF 算法搜索可用信道)和 FF 调度算法的仿真结果的比较。仿真过程中, 设定每一根光纤具有 16 个波长(数据信道), 网络中数据有两种优先级。高优先级的数据净荷可使用全部的 16 个数据信道; 对于低优先级数据净荷, 当网络拥塞发生时, 只能使用其中的 12 个数据信道。图 8 比较了使用两种调度算法的低优先级数据净荷丢失率。很明显, SD 调度算法的低优先级数据的丢失率要高一些。这是因为, 此算法在网络拥塞发生时减少了低优先级数据可使用的数据信道数。图 9 则是高优先级数据净荷丢失率的比较。计算表明 SD 调度算法的高优先级数据净荷丢失率比 FF 调度算法减少了 10.6%~50%, 平均减少了 21.5%。图 10 比较了两种算法的总数据净荷的丢失率, 二者很接近。我们还对使用 SD 调度算法和 FF 调度算法的延时时间(平均每个数据净荷在网络中经历的所有 FDL 缓存延时时间)和信道利用率进行了比较, 两者无显著差别。同样地, 当与 LAUC 调度算法比较时, 可以得到类似结果。由此可见, SD 调度算法在网络资源一定的情况下, 能通过减少低优先级数据的可用资源来尽量保证高优先级业务的可靠传输, 而且, 使用 SD 调度算法也不会对网络其它性能造成明显影响。

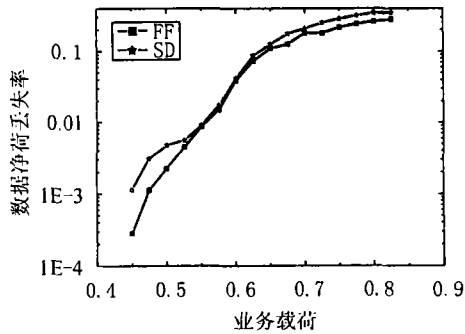


图 8 低优先级数据净荷丢失率对比图

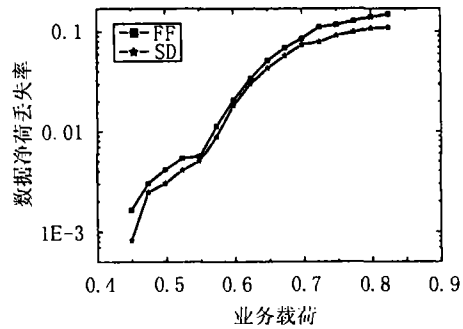


图 9 高优先级数据净荷丢失率对比图

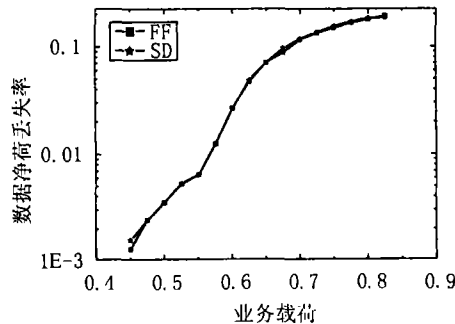


图 10 总数据净荷丢失率对比图

## 4 结论

本文简要介绍了 OBS 网络的基本概念, 并着重讨论了 OBS 网络中的数据信道调度算法。在调度算法部分, 首先介绍了已有的 FF 调度算法和 LAUC 调度算法。为尽量避免波长转换的发生和保证 OBS 网络中关键数据的传输质量, 本文分别提出了 MWCT 调度算法和 SD 调度算法。在文章的最后, 我们分析了以上几种算法的仿真结果, 通过与已有的调度算法进行比较, 证明了这两种新调度算法能显著地改善网络性能。

## 参 考 文 献

- [1] Yoo M, Qiao C, Ravikanth R. The effect of limited fiber delay lines on QoS performance of optical burst switched WDM networks. *IEEE ICC2000*, New Orleans, USA, 2000, 2: 974-979.
- [2] Cankaya H C, Xiong Y, Vandenhoute M. Design issues of optical IP routers for Internet backbone applications. *IEEE Communications Magazine*, 1999, 37(12): 124-128.
- [3] Xiong Y, Cankaya C. Control architecture in optical burst-switched WDM network. *IEEE J. on Selected Areas in Communications*, 2000, 18(10): 1838-1851.
- [4] Tancevski L, Yegnanarayanan S, Castanon G, Tamil L, Masetti F, McDermott T. Optical routing of asynchronous, variable length packets. *IEEE J. on Selected Areas in Communications*, 2000, 18(10): 2084-2093.
- [5] Verma S, Chaskar H, Ravikanth R. Optical burst switching: a viable solution for terabit IP backbone. *IEEE Network*, 2000, 14(6): 48-53.
- [6] Yoo M, Qiao C, Dixit S. Optical burst switching for service differentiation in the next-generation optical internet. *IEEE Communications Magazine*, 2001, 39(2): 98-104.
- [7] Ge A, Callegati F, Tamil L S. On optical burst switching and self-similar traffic. *IEEE Communications Letters*, 2000, 4(3): 98-100.

袁 巍: 男, 1978 年生, 硕士, 主要研究方向是下一代网络体系结构。

洪佩琳: 女, 1961 年生, 教授, 博士生导师, 主要研究方向是信息通信网。

李津生: 男, 1937 年生, 教授, 博士生导师, 主要研究方向是信息通信网。