

一种新颖的全光分组交换节点结构及其性能分析

张治中, 雒江涛, 曾庆济, 王建新, 叶 通

(上海交通大学区域光纤通信网与新型光通信系统国家重点实验室, 上海 200030)

摘 要: 本文提出了一种共享波长转换器和光纤延迟线的全光分组交换结构, 与传统的设计方法和实验报道相比, 该结构提供了更灵活的冲突解决措施. 为了获得与在交换系统输入级分别为每个波长信道提供波长转换器相等价的性能, 本文给出了交换结构最多所需的转换器数量. 仿真结果表明, 本文结构是兼顾交换系统体积、成本和性能等三方面较理想的折衷方案. 研究还表明, 光纤数与每纤的波长数量之积愈大, 本文结构所节约的转换器和延迟线数量将愈加可观.

关键词: 全光分组交换; 光缓存; 波长转换器; 共享; 分组丢弃率

中图分类号: TN9291.11 **文献标识码:** A **文章编号:** 0372-2112 (2004) 05-0713-05

On the Design and Performance Analysis of a Novel All-Optical Packet Switching Node Architecture

ZHANG Zhizhong, LUO Jiangtao, ZENG Qingji, WANG Jianxin, YE Tong

(R&D Center for Broadband Optical Networking Technology, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: A novel all-optical packet switching node architecture is proposed, which is equipped with a set of fiber delay lines (FDLs) and tunable optical wavelength converters (TOWCs) shared among the input lines. As far as contention resolution is concerned, the developed architecture is more flexible and cost effective than the traditional design and the methods introduced in the literatures. To obtain an equivalent packet loss probability to the method that provides TOWCs for each input wavelength channel, we give the maximum number of converters required by the system. Simulation results show that the developed architecture can guarantee a good tradeoff in terms of system size, cost and performance. We also get the conclusion that with the increase of the product of the number of input/output fibers and the number of wavelengths, the number of TOWCs and FDLs needed will, being different from the traditional design, gradually decrease.

Key words: all-optical packet switching; optical buffering; wavelength converter; sharing; packet loss probability

1 引言

光信号处理可以是线路级的、分组级的或比特级的. WDM光传输网属于线路级的光信号处理, 是粗粒度的信道分割. 光时分复用(OTDM)是比特级的光信号处理, 由于对光器件的工作速度要求很高, 离实用化还有相当距离. 光分组交换(OPS)属于分组级的光信号处理, 和 OTDM 技术相比, 对光器件工作速度的要求大大降低, 和 WDM 技术相比, 能以更细的粒度快速分配信道, 更加灵活、有效地利用带宽. 因此, 全光分组交换网应用前景广阔^[1-7].

在全光分组交换节点中, 当两个或两个以上的分组试图同时离开某个输出端口时, 即会产生冲突, 若没有采取光缓存措施, 将会造成数据包的丢失, 全光分组交换实用化的一大障碍即在于此. 目前, 不管是实验报道结果, 还是大多文献的研

究, 除了光纤延迟线(FDL, Fiber Delay Line)外, 尚没有更好的办法解决光分组的缓存问题. 对于 FDL 光缓存, 也许可以忽略其引入的光功率损失^[2], 但是, 因实现较大缓存容量而额外引入交换设备和连接器所带来的以下三方面不足是必须要考虑的: (1) 成本和尺寸的显著增加; (2) 大量交换设备所引入的插入损耗; (3) 串扰影响. 同时, FDL 是类似于 FIFO (First in first out) 的缓存结构, 还可能带来另两方面不利之处: (a) 一旦光分组进入 FDL, 即使其目的端口已经空闲, 但分组 (或许) 仍不能使用其目的端口, 待该分组离开 FDL 而进入目的端口时, 可能又会产生新的冲突, 无论如何都会影响系统的性能; (b) 由于 FDL 的影响, 接收节点所收到的数据分组可能发生乱序, 网络互连节点较多时情况会更严重. 因此, 接收端需配置一定容量的电缓存, 以便正确还原数据.

为了避免 FDL 作为光缓存时所带来的问题, 文献[4, 5]提

出了纯粹利用波长转换器来解决冲突,但是,该方案却会带来以下两方面的问题:(1)波长转换器较 FDL 而言,成本上昂贵得多,特别是要求转换器在整个波长信道内实现快速可调时;(2)波长转换器解决冲突的能力未必如 FDL 有效^[7].

本文基于上述考虑提出了一种新颖的、将波长转换器和 FDL 同时设计成反馈共享方式的全光分组交换节点结构.研究表明,本文结构可以同时兼顾系统体积、成本和分组丢弃率等方面的性能.

2 研究背景

光分组交换网络中常常基于以下方式解决分组冲突问题:(1)从时间域上解决:如采用 FDL 缓存冲突数据包,直到所争用的波长通道空闲为止;(2)从波长域上解决:即通过波长变换将冲突数据包的承载波长转换到另一空闲波长上,使数据包能在这一空闲波长通道上传送到下一节点;(3)从空间域上解决:即偏转路由方式,它利用网络中的闲置路由,使数据包绕过冲突节点迂回去往目的节点.偏转路由是从网络级别来考虑解决冲突的,它所引起的接收分组乱序会很严重,另外还带来了较长的延时.当网络负载较重或网络连接度较低时,偏转路由方式不再有效.因此,偏转路由方式并没有从根本上解决好交换节点的冲突问题.

通常,可以将光分组交换网络分成两大类^[1]:具有定长分组的时隙网络和可变长分组的无时隙网络.大量研究已经表明,时隙网络能大大降低分组冲突机率,因此,在本文的研究中,我们采用了定长的数据包,当然,对硬件实现而言,需要额外设计同步电路.

作为缓存的 FDL 根据其连接方式,也可分为两类,一为前馈(Feed forward)式连接方式,被缓存的分组离开 FDL 时必须立即进行处理.另一种为反馈(Feedback)式连接方式,离开了 FDL 的分组可以通过交换矩阵重新进入 FDL.显然,同样长度的 FDL 被连接成反馈形式时,其缓存能力将大有提高,解决冲突的能力也就大大增强了.这两种连接方式各有优缺点,本文不作过多比较.

尽管有大量文献研究了将 FDL 设计成反馈形式的交换结构,但将波长转换器也设计成反馈形式的研究很少(只有文献[5,7]对此进行了研究),亦未见有将波长转换器和 FDL 同时设计成为反馈形式的交换结构.在下一节里,我们将对此进行深入研究.

3 交换节点模型及分析

交换节点具有 N 个基于 WDM 技术的光纤输入/输出端口,每根光纤支持 W 个波长,记为{K₁, , K_W}.如图 1 所示,本文提出的交换结构由三部分构成:(1)输入级,由一组光波分解复用器组成.在这里,解复用器将各个输入光纤上的分组解复用,并送入下一级(交换级)进行处理;(2)交换级,由分光器和光门组成的空分交换阵列组成.在这里,空分交换将输入分组交换到其对应的目的端口.如果有冲突发生,则在 M(R) 个共享的 FDL(波长转换器)中选择空闲的 FDL(转换器),将冲突分组缓存后,在将来的时隙进行发送,或者将输入分组的

承载波长进行波长变换,立即进行冲突解决.图 1 中用于冲突解决的 FDL 和波长转换器同时被设计成了反馈共享的形式,据我们所知,该结构还未在国内外进行过研究;(3)输出级,由一组无缘耦合器组成,它们将输入分组耦合到对应的输出光纤端口,进而去往下一个节点.

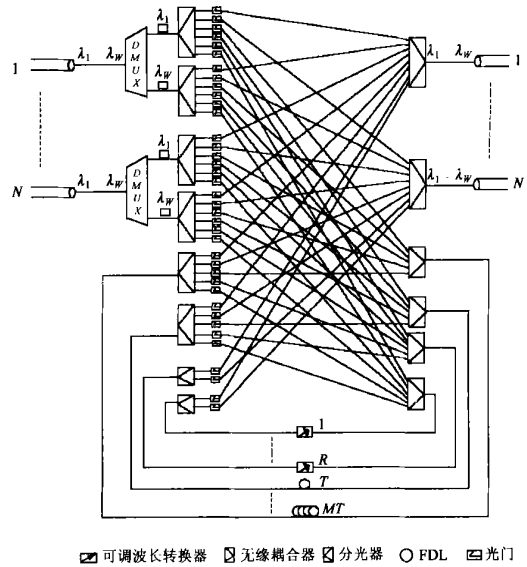


图 1 基于反馈共享波长转换器和光纤延迟线的全光分组交换结构

假设分组具有固定长度,且是基于时隙同步的.一个时隙的持续时间(T)表示一个时间单元.不失一般性,设每个 FDL 的长度为时隙长度 T 的整数倍,如图 1 所示,最短长度的 FDL 对分组提供的缓存时间(缓存容量)是 T,最长的 FDL 提供的缓存时间是 MT.注意图 1 中的缓存属于简并(degenerate)形式的结构,即延迟线所延迟的时长变化是连续的(步长为 T).

输入/输出信道中的分组可由变量组 I(i, K_j)/O(i, K_j) 进行描述,其中, i={1, , N} 表示输入/输出光纤, K_j{j=1, , W} 表示波长.为便于分析,首先定义下列算式:

$$I(i, K_j, t) = \begin{cases} 1, & \text{在第 } t \text{ 时隙时,第 } i \text{ 条输入光纤的波长} \\ & K_j \text{ 上有分组到达} \\ 0, & \text{其他} \end{cases}$$

$$O(i, K_j, t) = \begin{cases} 1, & \text{在第 } t \text{ 时隙时,第 } i \text{ 条输出光} \\ & \text{纤的波长 } K_j \text{ 上有分组离开} \\ 0, & \text{其他} \end{cases}$$

$$[I(i, K_j, t) \vee O(h, K_k, t)]$$

$$= \begin{cases} 1, & \text{在第 } t \text{ 时期时,第 } i \text{ 条输入光纤的波长 } K_j \text{ 上} \\ & \text{有分组到达,且以波长 } K_k \text{ 离开输出光纤 } h \\ 0, & \text{其他} \end{cases}$$

$$WC(k, t) = \begin{cases} 1, & \text{波长转换器 } k \text{ 在第 } t \text{ 时隙时被占用} \\ 0, & \text{其他} \end{cases}$$

$$FDL(k, t) = \begin{cases} 1, & \text{光纤延迟线 } k \text{ 在第 } t \text{ 时隙不可用} \\ 0, & \text{其他} \end{cases} \quad k = 1, \dots, M \quad (5)$$

假设当前时隙为 t , 且同时存在有 $I(1, K_i, t) = 1$ 及 $I(N, K_i, t) = 1$, 它们的目的端口均为第一根输出光纤. 假设交换控制调度器允许 $I(1, K_i)$ 直接进入 $O(1, K_i)$, 即 $[I(1, K_i, t) \text{ 与 } O(1, K_i, t)] = 1$, 则对于 $I(N, K_i)$, 以下任意一种情况满足时都可以解决冲突:

(1) 如果 $\forall k \in \{2, \dots, W\}$ 及 $k \in \{1, \dots, R\}$, 满足 $O(1, K_i, t) = 0$ 及 $WC(k, t) = 0$.

(2) 如果 $\forall k \in \{1, \dots, M\}$, 满足 $FDL(k, t) = 0$.

第一种情况表明, 可以利用波长转换器, 将冲突分组的承载波长变换到输出光纤中空闲的波长上, 以解决冲突. 第二种情况则表明, 可以利用 FDL, 将冲突的分组调度到将来的时隙予以传送. 显然, 第一种措施对分组传输的时延改善有利, 也使交换矩阵的容量得到更充分利用, 且不存在经过 FDL 缓存的分组在离开缓存时产生新的冲突的问题. 因此, 当两种方法都可以采用时, 本文采用第一种策略.

令 $G = R + M$, 为 FDL 和转换器的总数量. 对于给定的 G , 我们来分析一种特殊配置 $(R, M) = (G, 0)$ 时, 系统的分组丢弃性能. 正如下文所示, 该配置为 (R, M) 的其余取值提供了分组丢弃率的上限.

假设波长信道的分组到达是独立同分布的, 每个到达分组指向各输出光纤端口的概率均为 $1/N$. 此时, 交换系统的分组丢弃率即等价于单个输出端口的分组丢弃率. 对于每纤负载为 p 的分组到达, 若令 $p_i = \Pr\{I(i, K_j, t) = 1\}$, 即第 t 时隙输入光纤 i 上的波长 K_j 有分组到达的概率, 则每个信道的负载可表示为:

$$p_w = \lim_{t \rightarrow \infty} p_i = \lim_{t \rightarrow \infty} \Pr\{I(i, K_j, t) = 1\} = p/W \quad (6)$$

显然, 所有输入光纤中到达的分组同时要求交换到同一输出端口的个数 (h) 服从二项分布, 为:

$$p_a(h) = \binom{NW}{h} (p_w/N)^h (1 - p_w/N)^{NW-h} \quad (7)$$

分组阻塞可能由两种情况引起, 一是由于转换器的数量不够, 另一种是由于超过 W 个分组要求同时从同一输出光纤端口输出. 假设转换器数量足够, 则分组丢弃率为:

$$P_{loss} = p_a(h > W) = \sum_{h=W+1}^{NW} \binom{NW}{h} (p_i/N)^h (1 - p_i/N)^{NW-h} \quad (8)$$

图 2 给出了基于式 (8) 的分组丢弃率, 可见, 对于每纤负载为 0.18 Erlang 的分组到达, 当每纤的波长数大于 6 时, 只要转换器数量足够, 分组丢弃率已经小于百万分之一, 系统可以不使用 FDL, 这得到了和文献 [4, 5] 中仿真结果相一致的结论.

目前, 绝大多数文献^[4]都是在交换系统输入级为每个波长信道配置一个可调波长转换器, 此时, 系统所需的波长转换器数量为 NW 个, 对于图 1 所示的反馈共享方案, 系统所需的转换器数量显然要少于 NW 个, 那么, 如何确定转换器数量以

使系统性能不受影响呢? 我们有如下定理:

定理 一个具有 N 输入/输出的, 每纤波长数量为 W 个的, 基于时隙的全光分组交换系统, 图 1 所示系统最多只需配置 $G_{opt} = W @ N \lfloor W/N \rfloor + \text{GCD}(W, N)$ 个波长转换器, 便可获得与在输入级为每个波长信道分别配置转换器时等价的性能. 其中, $\text{GCD}(x, y)$ 为 x 和 y 的最大公因子.

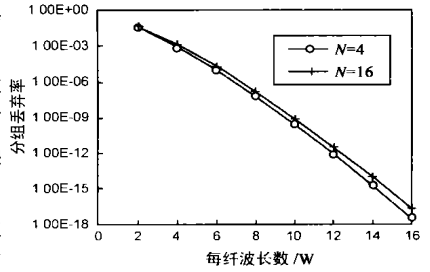


图 2 仅利用足够多的共享波长转换器解决冲突时, 波长数量和分组丢弃率的理论关系曲线 (每纤负载为 0.18)

要证明上述定理, 首先给出一个定义, 如下:

最小成本划分问题 给定 $W @ N$ 个方块 (W 种颜色, 每种颜色有 N 个方块), 找到最小成本划分以使得这些方块被分成 N 组, 每组有 W 个方块. 划分的成本是所有分组的成本之和, 组的成本为其内不同颜色方块的数量.

引理 图 1 结构所需的转换器数量可通过最小成本划分问题予以解决, 所需的转换器数量最多为 $W @ N$ 与最小成本划分所对应的成本之差.

证明 显然, 当以下情况发生时, 所需要的转换器数量最多: (1) 每个输入信道都有分组到达; (2) 指向每个输出光纤的分组恰好为 W 个 (当多于 W 个时会发生分组丢弃, 但此时并不是由于转换器数量不够而引起的). 因此, 只要考虑如何将 $W @ N$ 个分组分配到 N 条输出光纤, 每纤有 W 个分组, 以使得所需的转换器数量最多即可. 将到达的分组与上述定义中的 $W @ N$ 个方块相对应, 将波长与颜色相对应, 显然这 $W @ N$ 个分组交换到 N 条输出光纤即等价于将 $W @ N$ 个方块划分成 N 组的问题, 交换到相同输出光纤的分组所需要的转换器数量即为 W 与那些具有不同波长的分组数量之差 (即与这些划分相联系的成本). 因此, 最多所需的转换器数量的求解即等价于最小成本划分问题, 需要的转换器数量最多为 $W @ N$ 与最小成本划分所对应的成本之差.

对于任意一个图, 最小成本划分问题的求解是一个穷举问题, 其成本为 $W + N - \text{GCD}(W, N)$ (由于其求解推导过程很冗长, 鉴于文章篇幅, 这里不再给出推导过程, 具体求解可参阅文献 [6]).

因此, 图 1 所示系统最多只需要 $W @ N - W - N + \text{GCD}(W, N)$ 个波长转换器. 可见, 即使是取此上界值, 也要远少于传统的、非共享的、为每个波长信道配置转换器的设计方案. 例如, 取 $(N, W) = (4, 11)$ 时, 非共享方案需要 44 个转换器, 而本文结构最多只需要 30 个便可获得与其等价的性能. 事实上, 远远少于 30 个便可逼近其性能.

4 仿真及数值分析

仿真程序由 C++ 实现, 仿真中的每一个数据点都是在

10^8 个分组到达后获得的. 通过研究 McKeown 等人在评估 Cisco 12000 系列路由器交换阵列时所采用的方法, 我们发现其信元的产生过程和本文基于时隙的光分组网络很相似, 因此在仿真研究中, 我们也假定: (1) 数据分组以贝努力(Bernoulli)过程独立同分布到达; (2) 到达分组的目的端口均匀分布于各个输出光纤端口. 仿真中还假设了以下条件:

(1) 输入/输出光纤端口数 $N = 16$.

(2) 负载为 $Q = 0.8$ Erlang/每纤(对于时隙网络来说, 若不考虑 WDM 技术, 则每根光纤的满负载为 1 Erlang, 即每个时隙都有分组到达. 本文类似文献 [2, 4, 5], 假设采用 WDM 技术的主要目的是对负载进行分担, 从而减少冲突. 因此, 每纤的满负载仍然为 1).

首先, 我们考查系统无任何冲突解决措施 ($G = R + M = 0$), 单独考虑 FDL 解决冲突, 以及单独考虑转换器解决冲突等三种不同情况下系统的分组丢弃率. 为便于对比衡量 FDL 和转换器的冲突解决能力, 实验中 FDL 和转换器的数量均取为 8 个, 分别对应 $(R, M) = (0, 8)$ 和 $(R, M) = (8, 0)$. 仿真结果如图 3 中所示. 可见: (1) 无冲突解决措施时, 交换系统的分组丢弃率太高, 远远不能满足实际应用的需要; (2) 单独考虑 FDL 和转换器解决冲突时, 同样数量的 FDL 比转换器有效得多, 但注意到当 $M = 8$ 时, 延迟线的总缓存容量为 $36T$, 若分组时隙长度为 100fs , 则总共需要约 1080km 长度的光纤, 其体积、成本和损耗已不能忽略. 因此, 单独利用 FDL 解决冲突未必是可取的; (3) 每纤波长数不少于 8 时, 单独利用 FDL 和转换

器均可获得较理想的分组丢弃率.

其次, 我们特别考查单独利用转换器解决冲突时, 不同数量的转换器配置对分组丢弃率的影响, 仿真结果如图 4 所示. 图 4 中, 转换器数量 $R = G_{opt}$ 意味着在每个数据点, 所用的转换器数量 $R = W @ N - W - N + GCD(W, N)$, 注意对应不同的波长数, G_{opt} 也相应变化. 理论值则是依据式 (8) 计算所得的结果. 可见: (1) 8 个转换器已能较好的逼近理论分析值; (2) 当转换器数量为 G_{opt} 时, 对应的分组丢弃率与理论分析结果完全重合, 说明定理中所确定的转换器数量是正确的; (3) 对于给定的分组丢弃率要求, 例如 10^{-6} , 随着每纤波长数的增加, 系统所需的转换器数量可以逐步减少. 相对于传统设计而言(所需的转换器数量随着每纤波长数线性递增^[4]), 系统所节约的转换器数量是相当可观的.

最后, 我们考查对于总数量给定的 FDL 和转换器, 即 $G = R + M = 8$, FDL 和转换器的各种配置对分组丢弃率的影响. 仿真结果如图 5 所示. 图 5 中, 转换器数量 $R = G_{opt}$ 时, 与上述情况相同.

从图 5 的仿真结果可以看出: (1) 对于 $M + R = 8$, 随着 FDL 数量的逐渐增多, 分组丢弃率将逐渐下降, 但 FDL 的总数超过 4 个后, 改善就很有有限了, 有意思的是, 最理想的配置是取 $(M, R) = (7, 1)$; (2) 当每纤波长数不少于 8 个时, 不管如何分配转换器和 FDL 的个数, 分组丢弃率都将低于 10^{-6} ; (3) 只要有一根 FDL, 分组丢弃性能都将大大改善, 且丢弃率低于只考虑转换器解决冲突时最理想的分组丢弃率.

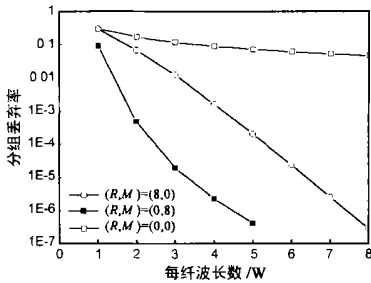


图 3 不同冲突解决策略下的分组丢弃率比较

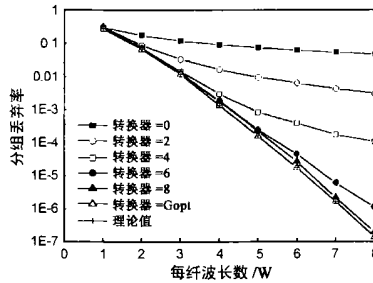


图 4 不同数量的转换器配置对分组丢弃率的影响 ($M = 0$)

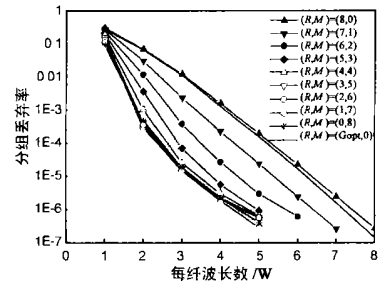


图 5 转换器和 FDL 组合配置下分组丢弃率比较 ($R + M = 8$)

根据以上仿真实验及分析, 可以得到以下结论:

- (1) 不管光纤数和波长数为多少, 必须要进行冲突解决.
- (2) 将 FDL 和转换器同时设计成反馈共享的形式是兼顾系统成本、体积和性能的较理想的方案. 例如, 对于前述仿真实验, 若取 $(M, R) = (4, 4)$, 系统所需的 FDL 总缓存容量则为 $10T$, 大大少于 8 根 FDL 所需要的长度 ($36T$). 同时, 相对于纯粹利用转换器解决冲突而言, 既取得了更好的分组丢弃率, 也少用了 4 个昂贵的转换器.
- (3) 与传统设计相比, 光纤数与波长数之积越大, 本文结构所节约的 FDL 和波长转换器数量将愈加可观.
- (4) 对于其它负载取值, 本文结构可以得到类似的结论, 只不过具体数值有些差异.

一般而言, 反馈式结构比前馈式结构需要更多的光开关, 这是反馈式结构不可避免的缺陷, 但值得提到的是, 相对于传

统设计中利用 FDL 解决冲突而言, 本文结构却未必比前馈 FDL 方式的结构需要更多的光开关. 例如, 若在每个交换输出端口配置一组简并方式的延迟线, 系统所需的光开关甚至可能多于本文结构所用的光开关数量. 因此, 本文结构是具有一定应用价值或设计参考价值的.

5 结束语

本文研究了基于时隙的全光分组交换网络的冲突解决问题, 提出了一种新颖的交换结构))) 将波长转换器和延迟线同时设计成为反馈共享式的交换结构. 研究表明, 本文结构相对于传统设计而言, 在系统体积、成本和性能方面取得了较好的折衷考虑.

参考文献:

[1] T S ElBawab, Jong Dug Shin. Optical packet switching in core net

works: between vision and reality[J]. IEEE Communications Magazine, 2002, 40(9): 60- 65.

- [2] D K Hunter, M C Chia, I Andonovic. Buffering in optical packet switches[J]. IEEE Journal of Lightwave Technology, 1998, 16(12): 2081- 2094.
- [3] W D Zhong, R S Tucker. A new wavelength2routed photonic packet buffer combining traveling delay lines with delay2line loops[J]. IEEE Journal of Lightwave Technology, 2001, 19(8): 1085- 1092.
- [4] S L Danielsen, et al. Optical packet switched network layer without optical buffers[J]. IEEE Photonics Technology Letters, 1998, 10(6): 896 - 898.
- [5] V Eramo, M Listanti, P Pacifici. A comparison study on the number of wavelength converters needed in synchronous and asynchronous all2optical switching architectures[J]. IEEE Journal of Lightwave Technology, 2003, 21(2): 340- 355.
- [6] K C Lee, et al. Optimization of a WDM optical packet switch with wavelength converters[A]. IEEE INFOCOM[C]. Boston, 1995. 2. 423 - 430.
- [7] Zhizhong Zhang, Qingji Zeng. Wavelength converter and fiber delay2line sharing in WDM optical packet switches: dimensioning and performance issues[A]. OptiComm 2003[C]. Dallas, TX, Oct. 2003. 5285. 267- 277.

作者简介:



张治中 男,1972 出生于湖北恩施,副教授,硕士生导师,于 2002 年毕业于电子科技大学通信电子学院,获工学博士学位,目前主要从事光分组、光标记交换技术及其实现,10G 城域以太网组网及节点关键技术研究及其实现技术等。
Email: zhangzz0078@vip. sina. com.



曾江涛 男,1971 出生于河南郑州,副教授,博士,目前主要从事光标记交换技术的研究开发工作。

曾庆济 男,1938 年出生于重庆,目前为上海交通大学电子工程系教授,博士生导师,主要研究领域涉及宽带光网,光交换技术与 IP/WDM 光网,光网络软件平台。