超大容量光交换机中器件约束的 Clos 网络路径设计

周宇萌 邱 昆 许 渤 崔展齐 凌 云 陈锡莲
(电子科技大学光纤传感与通信教育部重点实验室, 四川 成都 611731)

摘要 基于快速可调激光器和阵列波导光栅的光交换结构是实现超大容量光交换机的理想选择,但是快速可调激光器的快速调谐要求使得空闲时隙的激光器必须保持发光状态,并对有效业务光信号造成干扰。为了解决此问题,提出了采用无效业务填充空闲时隙的方法,从而使得系统中的有效业务不再受到无效业务的干扰。通过两种不同的方法生成无效业务即随机策略和顺序策略,仿真研究和对比了两种策略在不同业务量下各自的优势。同时,为了降低包含无效业务时的调度算法重排次数要求,还进一步提出了阈值策略计算无效业务路径的方法,并通过仿真验证了该方法可以大大降低重排次数,提高调度算法的计算效率。

关键词 光通信;光交换;调度算法;Clos 结构;阵列波导光栅;快速可调激光器
中图分类号 TN913.7 **文献标识码** A **doi**: 10.3788/AOS201333.0806003

Path Design of Clos Network Based on Device Constraint in Large Capacity Optical Switches

Zhou Yumeng Qiu Kun Xu Bo Cui Zhanqi Ling Yun Chen Xilian
(*Key Laboratory of Optical Fiber Sensing and Communications, Ministry of Education, University of Electronic Science and Technology of China, Chengdu, Sichuan* 611731, *China*)

Abstract Optical switching architecture based on fast tunable laser and arrayed waveguide grating is an ideal choice to realize optical switches with large capacity. However, the fast tuning requirement of the fast tunable laser makes it necessary for the fast tunable laser to keep seeding light even under idle state without data, and the optical light from the idle fast tunable laser may introduce significant interference to a valid optical signal if they are routed to the same receiver. In order to solve the influence of the fast tunable laser constraint, a method is proposed to use the invalid demands to fill idle time slots so that the valid demands and invalid demands have different output ports and no longer interfere with each other. Two different methods to generate the invalid demands are proposed, namely the random strategy and the sequential strategy, and the performances of the two strategies with different traffic loads are studied and compared through simulations. Meanwhile, in order to reduce the number of rearrangement in the scheduling algorithm, a threshold-based strategy to find the path for an invalid demand is proposed. The simulation results show that this threshold-based method can greatly reduce the number of rearrangement, and improve the computation efficiency of the scheduling algorithm.

Key words optical communications; optical switch; scheduling algorithm; Clos structure; arrayed waveguide grating; fast tunable laser
OCIS codes 060.1155; 060.1810; 060.4251

1 引 言

随着互联网中多媒体会议、远程教育、视频点播、网络游戏等应用的爆炸式增长,网络的数据流量大幅增加,因此人们极希望电交换机的容量能够急剧增长。但随着线路传输速率和交换机容量的增长,电交换机的交换速率以及它的热耗散和功率损

收稿日期: 2013-02-08; **收到修改稿日期**: 2013-04-15
基金项目: 国家 863 计划(2012AA011304)、中央高校基本科研业务费(ZYGX2011J009)、国家自然科学基金青年科学基金(61101095)
作者简介: 周宇萌(1987—),女,硕士研究生,主要从事光通信方面的研究。E-mail: meng_zym@yahoo.com.cn
导师简介: 邱 昆(1964—),男,博士,教授,主要从事光通信方面的研究。E-mail: kqiu@uestc.edu.cn

耗等出现了瓶颈，使得交换容量的增加变得非常困难[1]。因而人们迫切希望设计出下一代大容量低功耗的交换机[2-3]。

光纤链路具有低损耗、高速率的信号传输的特点，因此 Pb/s(1Pb＝10^3Tb)及以上超大容量的交换机中使用光纤链路来实现电交换背板间的互连和交换，有望大幅度降低需要的功耗[4]。在超大规模的电交换背板间的光交换和互连中，其关键是实现超大规模的光交换结构，相应的研究已经得到了越来越多的关注[5-6]。

光交换的核心是光开关(OS)，虽然目前已能实现纳秒级的光开关，但其通常只能实现很少的端口数[7-8]。三维微机电系统(3D-MEMS)光交换结构虽然可以实现很大的端口数，但它降低了交换速度[9-10]。基于快速可调激光器(FTL)和阵列波导光栅(AWG)的光交叉连接技术，则可以充分利用 AWG 的波长选路特性和 FTL 的快速波长调谐特性，实现快速的光交叉连接的建立，因此成为了超大容量光交换机的理想选择[11]。在超大容量光交换机的实际研制中，华为公司在 2012 年国际光纤通信会议(OFC)上报道的容量为 1 Pb/s 的光交换机，使用的就是基于 FTL 和 AWG 的光交换结构[5]。

超大规模光交换结构的实际应用还需要高效的交换调度算法的支持，同时这种交换调度算法还必须要考虑光交换结构与传统的电交换结构的不同。因此，本文介绍了拟研究的基于 AWG 和 FTL 的光交换结构，以及该结构中存在的器件约束。针对 FTL 对调度的约束，提出了不同的解决方法。通过仿真对各种解决方法的性能进行了分析和对比，并对所做工作进行总结得出结论。

2　系统结构及约束条件

传统的交换机是读取每一时隙数据包的包头，根据目的地址进行路径分配，在每个时隙都会进行一次路径计算。而本文所研究的交换机是信号在进入交换机前便会对信号进行处理，将发往同一目的端口的包的数据拆分为不同时隙的突发包，因此在信号进入交换机前，交换机的连接状态已经计算结束，在一定时间内交换机的连接状态不会改变。

图 1 为最基本的三级 Clos 网络，它的特点是每个输入级交换单元和每个中间级交换单元之间有且仅有一个链接相连，同样每个中间级交换单元和每个输出级交换单元之间也是有且仅有一个链接相连。

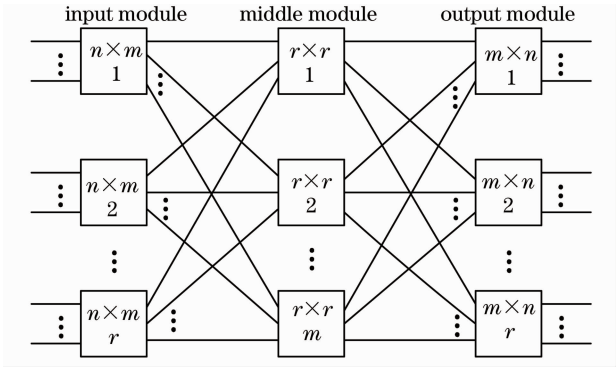

图 1　Clos 网络

Fig. 1　Clos network

该系统结构是一个七级 Clos 交换结构，具体如图 2 所示，由三级 Clos 网络扩展而得。它的第一级和第七级是由 M 个 $L\times L$ 的电交换模块组成，其中 $L=n$。第二级的输出端是 FTL，第六级模块的输入端是实发接收器(BMR)，其中第二级输出端的端口数与第六级中输入端的端口数相同，均为 k。第三级到第五级是一个由光开关和 AWG 组成的三级 Clos 网络，其中第三级和第五级均由 r 个 2×2 的光开关组成，AWG 的输入输出端口数目为 r。第二级到第六级组成五级 Clos 网络称为不同的平面。

图 2 所示的光交换机的核心是具有大端口数目的 AWG 和 FTL。AWG 可以根据每个输入端口的光信号的波长，将其路由到相应的输出端口实现交换的功能，目前能够实现的 AWG 的端口数目和规模可以达到 80×80[6]。为了使同一输入端口的光信号可以按照需要交换到不同的输出端口，在发射端需要使用 FTL 对拟发送的光包信号进行波长的设置，在接收端则通过 BMR 将光信号转换成电信号，再进行后续的电交换处理。由此可见，FTL 的波长调谐所需要的时间直接决定了光突发包发送的效率，在光突发包的间隔时间内，FTL 需要完成下一个光突发包发送所需要的波长调谐。三级 Clos 网络可重排无阻塞条件是 $m\geqslant n$[12]。由于系统中每一级的输入输出端口都是相等的，每一级都满足可重排无阻塞的条件，因此图 2 所示的光交换机满足可重排无阻塞条件。

为了提高 FTL 的波长调谐速度，FTL 需要保持发光的状态，因为若某个 FTL 被关断后再重新开启，波长调谐所需要的时间将远大于 FTL 保持发光状态下进行波长调谐所需要的时间。但是，如果 FTL 在空闲时保持发光状态会带来另一个问题：假设在某个时隙，某个 FTL 没有业务需要发送，那么为了保持发光的状态，该 FTL 如果不加判断地继续

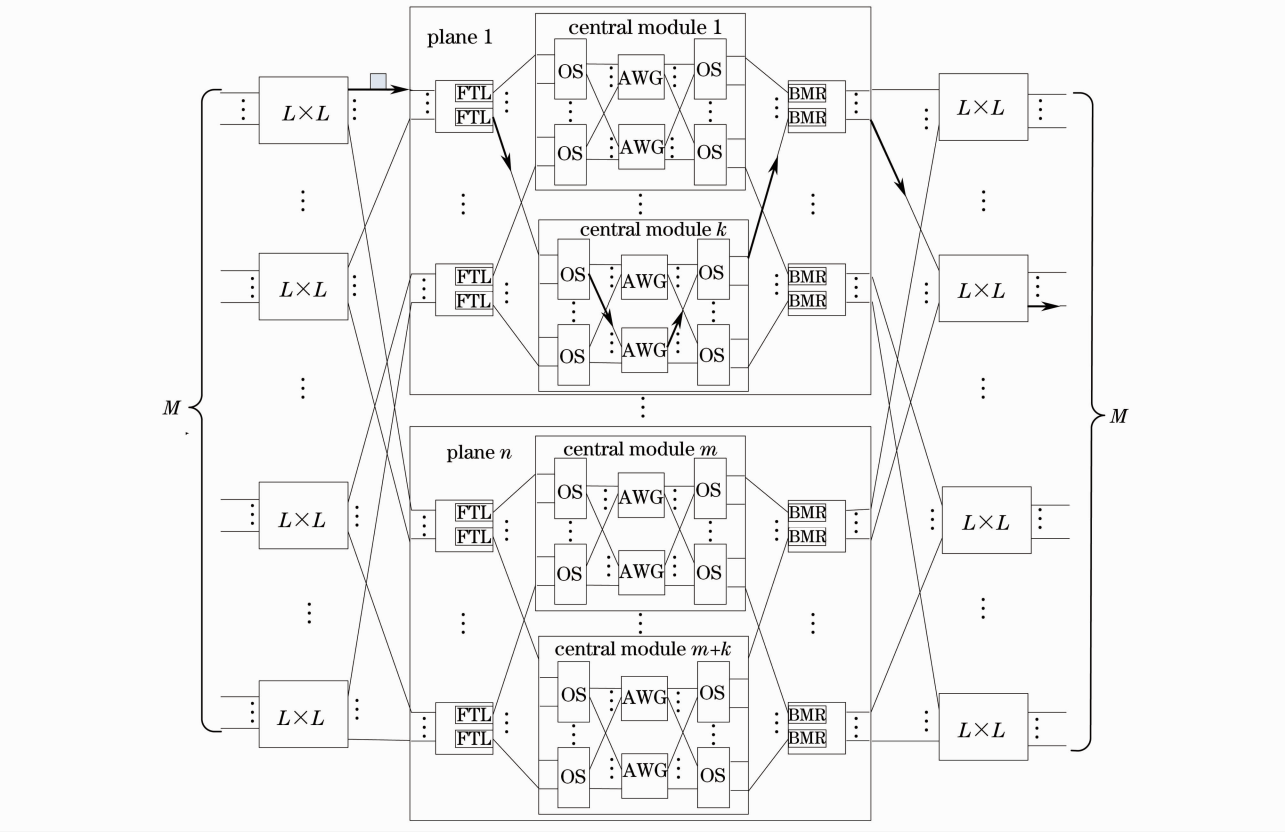

图 2 系统结构

Fig. 2 System structure

使用前一个时隙的波长发光，此光信号有可能会与某一个业务信号从同一个 AWG 输出端口输出，从而对该业务信号造成严重的干扰。

图 3 为 FTL 发出的无效业务光信号对交换结构中有效业务光信号的影响。其中模块 1、2 的输出端是 FTL，模块 A、B、C、D 为四个 2×2 的光开关，模块 3、4 为 AWG。每个 FTL 给出了前后两个时隙发出的光突发包，每个光突发包用一位字母和一位数字表示，其中字母 a、b、c、d 代表了该光突发包使用的波长，也同时对应了该突发包从 AWG 的输出的端口，数字 3、4 代表了该突发包将通过光开关选择模块 3 还是模块 4，如果某个光突发包上没有标注则表示该突发包承载的是无效业务。在时隙 1，模块 1 的输出端口 2 的业务标识为 d3（即在第 3 模块 d 端口输出），输出端口 1 的路径标识为 a3。在时隙 2，模块 1 的输出端 1 变为 d3，而端口 2 为空闲时隙。如果该端口 2 继续按照时隙 1 发送 d3 的光突发包，在模块 3 的输出端口 d 就会同时出现一个有效业务光信号和一个无效业务光信号，两者相互干扰，无法恢复有效业务信息。

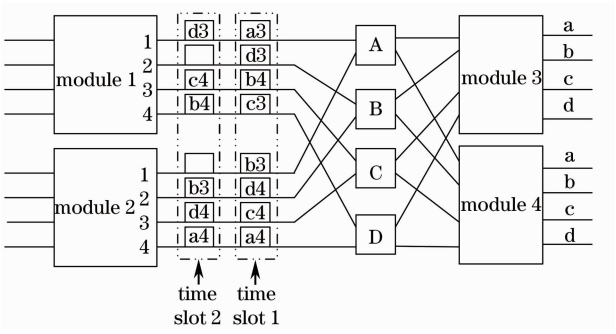

图 3 FTL 中无效业务对交换结构的影响

Fig. 3 Influence of the FTL′s invalid demand on the switch structure

由此可见，与传统的电交换结构不同，为了保证如图 2 所示的超大容量的光交换机能够正常工作，必须充分考虑 FTL 的工作特性及其对业务路由分配的约束，保证无效业务不对有效业务产生不利的干扰。

3 考虑无效业务的 Clos 网络路径分配算法

在 Clos 网络中，根据路由算法对新到的请求有两种不同的处理方式，一种称为统一调整，另一种称为逐条调整。统一调整是将已建立的请求和新到达

的请求进行统一处理，对整个网络进行一次输入输出匹配。逐条调整是对每一条新到达的请求分配路径，如果没有空闲的中间级模块时，则按照一定的规则将已建立的连接进行调整，以释放出空闲的中间级模块，来建立新的连接[13]。此处所研究的路由算法是采用逐条调整的方式，当没有空闲中间级模块可供选择时，采用 Paull 算法[14]进行重排。

对系统网络进行路径分配时，将七级网络分为三层，每条业务通过在三层交换结构中的三次选路得出完整的路径信息，每次选路都看作一个三级 Clos 选路：1）将第二级到第六级划分为不同的平面作为第一次选路的中间级模块，即第一次选路是选择平面；2）将第三级到第五级划分为不同的中间级模块，作为第二次选路的中间级，即第二次选路为选择中间级模块；3）直接将第三级光开关看作三级 Clos 网络的输入级，第四级 AWG 看作三级 Clos 网络的中间级，第五级光开关看作三级 Clos 网络的输出级进行选路，即第三次选路为选择最内层中间级模块。对于如图 2 所示的业务数据，源自 1 号输入模块，去往 2 号输出模块。对该业务进行选路时，平面选择的结果为 1 号平面；然后继续选择中间级模块，选择结果为第 k 个中间级模块；最后在第 k 个中间级模块选择最内层的中间级，即 AWG。通过以上三步，一个业务数据包的路径就确定了。如果在为某个业务数据进行选路时发现所选择的路径已经没有可用资源了，这时需要通过对已分配的业务及其路径进行重排。

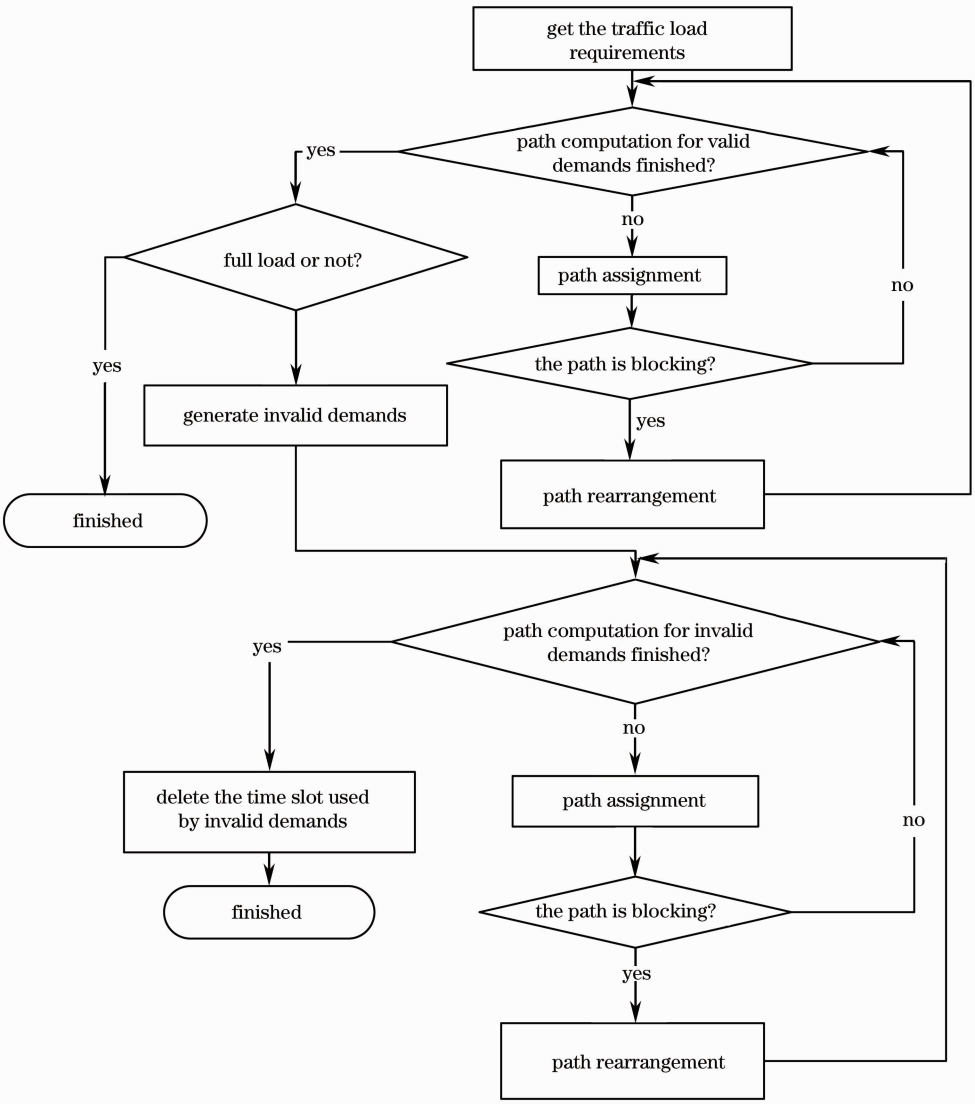

图 4　路径计算流程图

Fig. 4　Flow chart of path computation

整个调度过程中，如果待分配的业务量为100%，即整个网络中能够承载业务的最大量（称为满配业务）。由交换结构的可重排无阻塞特性可知，通过重排一定能够为每条业务找到一条不相互冲突的路径。

如果待分配的业务量不是满配的，则一定存在某些 FTL 在有些时隙不需要发送业务数据，此时可以为这些 FTL 分配一个无效业务。与有效业务相同，每个无效业务都将对应一个输入端口和一个输出端口，只是必须保证无效业务的输出端口不能与任何有效业务相同，以保证无效业务不会干扰有效业务。

整个系统的路径分配过程如图 4 所示，具体步骤如下：

1）获得待分配路径的有效业务需求。

2）对有效业务进行路径计算和分配，对存在阻塞的路径调用重排算法进行重排。

3）判断 Clos 网络是否为满配业务，如果是，调度过程结束；如果不是则进入步骤 4）。

4）为每个空闲的 FTL 生成一个无效业务，无效业务的输入端口和输出端口需要满足与有效业务不冲突的约束条件。无效业务的生成，可以采用随机策略或者顺序策略。

5）为每个无效业务进行路径计算和分配，如果出现阻塞，则调用重排算法进行重排。因为系统满足可重排无阻塞的条件，因此通过重排一定能够为所有的有效业务和无效业务找到对应的路径。

虽然通过空闲 FTL 引入满足约束条件的无效业务，可以保证所有的有效业务和无效业务都能找到一条不相互冲突的路径，但是对于调度算法来说，由于各无效业务的输入输出端口间也必须满足互不相同的条件，因此其路径计算的难度等同于一次满配业务的调度，可能需要很多次的重排之后才能找到最后的分配结果。

为了降低非满配业务时无效业务的路由分配所需要的计算或者重排次数，进一步提出了一种多个无效业务可以使用相同的空闲输出端口的方法。这时，虽然多个无效业务光信号间存在相互干扰，但是接收端并不需要恢复业务数据。同时，由于降低了对无效业务输出端口选择的约束，无效业务的路由计算复杂度或者需要的重排次数减少。不过，为了避免同时到达某个光突发接收机的无效业务光信号过多造成对光突发接收机的损坏，可以为每个空闲输出端口能够同时接收的无效业务光信号的个数做一个阈值的限制。这种算法称为基于阈值限制的无效业务的计算和路由算法。

4 算法仿真与分析

通过系统仿真对算法进行对比和分析。其中仿真的系统结构如图 2 所示，设 L 为 20，M 为 168，平面数为 20，参数的变化只会按比例增加或减少重排次数。下面通过不同的有效业务量来进行对比。

针对 FTL 器件约束引入的无效业务，显然会增加调度算法的计算量，这里以需要重排的次数来表示。图 5 给出了两种不同的无效业务生成策略的重排次数对比。由于两种方法都是先对有效业务进行调度和路径分配，因此有效业务路径计算时需要的重排次数相同。但是，由于顺序和随机这两种无效业务生成策略存在本质的差异，随机策略是通过产生随机数来决定无效业务的输入输出端口，而顺序策略是为输入输出端口依次分配空闲的端口，因此有效业务量的多少对两种策略下的调度算法的性能有很大的影响。

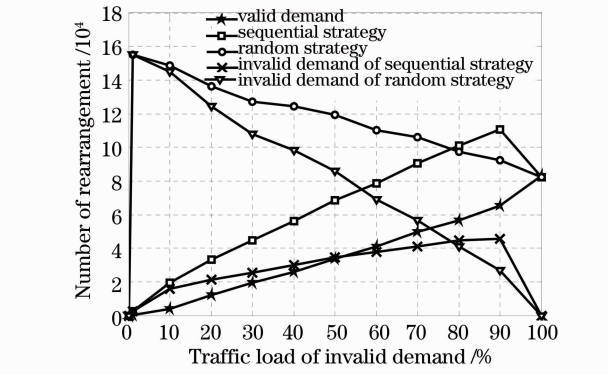

图 5 不同无效业务生成策略需要的调度算法重排次数的比较

Fig. 5 Comparison of the number of rearrangement in the scheduling algorithm for different invalid demand generation strategies

如图 5 所示，当业务量为 100%时，仿真中未添加无效业务，即此时为满配时最内层重排的次数。由方形曲线和圆形曲线对比可以看出，采用顺序策略有效业务量小于 80%时最内层重排次数都是小于随机策略的。有效业务量越小两种策略间重排次数的差异越大，因为顺序策略中在业务量很小时易建立连接，从而减少了重排的次数。因此在业务量小于 80%时采用顺序策略生成无效业务能够提高算法的性能，而当业务量大于 80%时采用随机策略生成无效业务更有优势。由三角形曲线和×形曲线

的对比可以看出，随机策略中无效业务的重排次数随着有效业务量的增加而降低，顺序策略则相反，随着有效业务量的增加，无效业务的重排次数也相应地增加。这是因为无效业务是在有效业务路径计算完成后重新建立计算的，在有效业务量越多时空闲的路径就越少，因而顺序策略不能够体现出它的优势。

对于第 3 节中提出的基于阈值限制的无效业务的计算和路由算法，图 6 给出了该策略与顺序策略生成无效业务的重排次数的比较，其中阈值设置为 3，其他仿真参数与前面相同。由图 6 的结果可以明显看出，基于阈值限制的无效业务的计算和路由算法能够有效地降低所需要的重排次数。继续增加阈值至 5 时，仿真结果显示对重排次数无影响，图中随机生成策略代表了阈值为 1 的仿真结果，阈值为 5 时已经无限接近于阈值为 3 的结果，因此其结果未在图中给出。

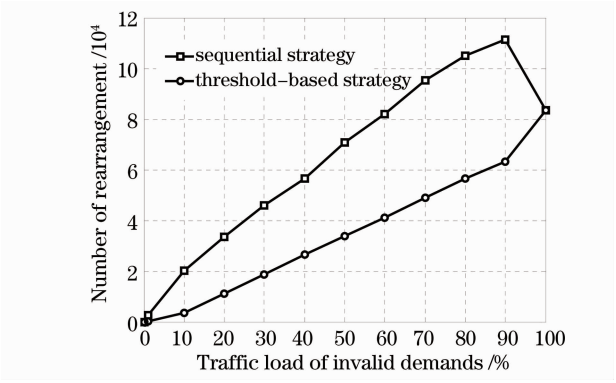

图 6　阈值策略的性能

Fig. 6　Performance of the threshold-based strategy

5　结　　论

通过对空闲 FTL 引入无效业务的方法，降低了空闲业务对有效业务光突发包出现干扰的可能。并且提出了两种不同的生成无效业务的策略，通过仿真数据可以看出，在有效业务量低于 80％的情况下采用顺序策略生成无效业务，业务量大于 80％的情况下采用随机策略生成无效业务，可以使得系统的性能达到最优。提出的阈值策略计算无效业务路径的方法，可以减少调度器的重排次数。通过仿真结果可以看出，与原有策略相比阈值策略大大降低了调度器的重排次数，在无效业务部分不再有重排出现，系统性能得到了大幅提升。

参　考　文　献

1 D T Neilson. Photonics for switching and routing [C]. IEEE J Sel Top Quantum Electron, 2006, 12(4): 669－678.

2 Le Zichun, Chen Jun, Fu Minglei, *et al*.. Optical cross connection: novel architecture and performance analysis [J]. Acta Optica Sinica, 2011, 31(3): 0306005.
乐孜纯，陈　君，付明磊，等. 一种新型结构光交叉连接节点及其联网性能分析[J]. 光学学报，2011，31(3)：0306005.

3 Zhao Zisen. Past, present and future of optical fiber communications [J]. Acta Optica Sinica, 2011, 31(9): 0900109.
赵梓森. 光纤通信的过去、现在和未来[J]. 光学学报，2011，31(9)：0900109.

4 Guo Aihuang, Feng Shengyi, Xue Lin, *et al*.. Research on power efficient routing algorithm in green optical networks [J]. Acta Optica Sinica, 2012, 32(4): 0406002.
郭爱煌，冯圣毅，薛　琳，等. 基于节能的绿色光网络路由算法的研究[J]. 光学学报，2012，32(4)：0406002.

5 Shiyi Cao, Shaofeng Qiu, Lun Wei, *et al*.. An optical burst switching fabric of multi-granularity for petabits/s multi-chassis switches and routers [C]. Optical Fiber Communication Conference, 2012.

6 Shaofeng Qiu, Shiyi Cao, Lun Wei, *et al*.. A cost-effective scheme of high-radix optical burst switch based on fast tunable lasers and cyclic AWG [C]. OFC/NFOEC, 2012.

7 H Wang, A Wonfor, K A Williams, *et al*.. Demonstration of a lossless monolithic 16 × 16 QW SOA switch [C]. ECOC′09, 2009.

8 K Nashimoto, D Kudzuma, H Han. Nano-second response, polarization insensitive and low-power consumption PLZT 4×4 matrix optical switch [C]. Optical Fiber Communication Conference, 2011.

9 Cao Zhonghui, Bao Junfeng, Yuan Ye, *et al*.. A non-silicon-based 1×4 MEMS optic switch [J]. Acta Optica Sinica, 2003, 23(9): 1041－1044.
曹钟慧，鲍俊峰，袁　野，等. 非硅基底 1×4 微机电系统光开关[J]. 光学学报，2003，23(9)：1041－1044.

10 Yong-Kee Yeo, Zhaowen Xu, Chi-Yi Liaw, *et al*.. A 448×448 optical cross-connect for high-performance computers and multi-terabit/s routers [C]. OFC/NFOEC, 2010.

11 H J Chao, Zhigang Jing, S Y Liew. Matching algorithms for three-stage bufferless Clos network switches [J]. IEEE Communication Magazine, 2003, 41(10): 46－54.

12 A M Duguid. Structural Properties of Switching Networks [R]. Providence: Brown University, Progress Report BTL-7, 1959. 1481－1492.

13 Shi Zengzeng. Research on Routing Algorithms in Clos Matrix [D]. Xi′an: Xidian University, 2008. 9－20.
石增增. Clos 交叉矩阵中的路由算法研究[D]. 西安：西安电子科技大学，2008. 9－20.

14 Yuan Liming. Studies of Key Technology for Clos Matrix [D]. Xi′an: Xidian University, 2008. 27－28.
袁立明. Clos 交叉矩阵关键技术研究[D]. 西安：西安电子科技大学，2008. 27－28.

栏目编辑：王晓琰