

Min-Cost Max-Flow Characterization of Shared-FDL Optical Switches

M. Rodelgo-Lacruz, C. López-Bravo, F. J. González-Castaño, F. Gil-Castiñeira, and H. J. Chao

Abstract—We characterize the assignment of shared fiber delay loops (FDLs) in optical switches as a min-cost max-flow problem to obtain a bound on their optimum performance.

Index Terms—OPS, FDLs, scheduling algorithms, max-flow.

I. INTRODUCTION

THE main architectures for time-slotted all-optical packet switching are [1]: output buffering, whose optimum scheduling is trivial; input buffering, whose scheduling is equivalent to a matching problem in bipartite graphs [2] (optimum output-buffering scheduling algorithms are the basis for feasible maximal size matching schedulers that approximate the optimum solution); and shared feedback buffering (SFB), whose scheduling problem we characterize in this paper as a min-cost max-flow problem.

The input ports of an SFB switch [3] (Figure 1(a)) share a pool of feedback FDLs. Each FDL delays (buffers) cells for a fixed number of time slots to avoid contention. Let Z and N respectively be the number of FDLs and input/output ports. The outputs (inputs) of the FDLs and the inputs (outputs) of the switch are collectively the *inlets* (*outlets*) of the switch fabric. The main purpose of the reservation scheduler is to maximize the number of nonblocking FDL routes for incoming cells to minimize cell losses in each time slot. The switch module maintains a configuration table $C(s, t)$, for $s \in \{\text{outlets}\}, 0 \leq t < F$, where $F - 1$ is the maximum aggregated cell delay. An entry $C(s, t)$ acquires a value of $i \in \{\text{inlets}\}$ if outlet s (output or FDL) is connected to inlet i , and a value of 0 if outlet s is idle at time slot t . This can be logically represented by a slot transition diagram. Figure 1(b) shows an example with $N = 2$, $Z = 2$ (FDL lengths 1 and 2), $F = 4$, and a maximum number of circulations $K = 2$, where some routes have been previously scheduled. Node $T(t)$ represents time slot t . If a FDL with length l is available at time slot t (i.e. $C(\text{FDL}_l, t) = 0$), arc $(T(t), T(t + l))$ exists. In this representation, finding a valid FDL route to a time slot when output p is available is equivalent to finding a path from $T(0)$ to any $T(t)$ where $C(\text{output } p, t) = 0$. The sequential FDL assignment algorithm (SEFA) [3] performs a breadth-first search in the slot transition diagram to sequentially find nonblocking FDL routes for every incoming cell. When there are multiple choices, SEFA first selects the FDL route with the fewest circulations and, if there are multiple such routes, it picks the one with the shortest delay. Since SEFA is sequential, it performs suboptimally. In the example in Figure 1(b), if two cells requesting outputs 1 and 2 are scheduled in output

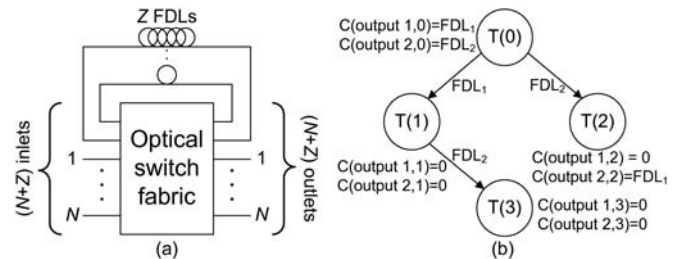


Fig. 1. (a) Shared-FDL optical switch. (b) Slot transition diagram.

index order, the first is assigned route $T(0) \xrightarrow{\text{FDL}_1} T(1)$ and the second cannot be assigned. Nevertheless, both cells could be scheduled by assigning routes $T(0) \xrightarrow{\text{FDL}_2} T(2)$ and $T(0) \xrightarrow{\text{FDL}_1} T(1)$, respectively.

II. MIN-COST MAX-FLOW CHARACTERIZATION

Since the cells share the FDLs, a global scheduling decision seems necessary to achieve optimum performance. We modified the slot transition diagram in Figure 1(b) to that shown in Figure 2. Let $T_k(t)$ be a *time slot node* of level k , $1 \leq k \leq K$, such that any cell arriving at $T_0(0)$ takes k circulations to reach it, where t is the aggregated delay of the traversed FDLs. An FDL arc from $T_k(t)$ to $T_{k+1}(t + l)$ represents the availability constraint for all FDLs with length l , and its capacity is the number of available FDLs with delay l in time slot t . To represent the output availability constraints, we introduced the *delay nodes* $D(t)$ and the *output nodes* $O(p)$, $1 \leq p \leq N$. Node $D(t)$ represents an aggregated delay t , which is achievable in K circulations at most. A *delay arc* with capacity N connects every $T_k(t)$ to $D(t)$. The capacity of the *output arcs* connecting every $D(t)$ to an $O(p)$ represents the availability of switch output p at time slot t , $(C(\text{output } p, t))$. Finding a valid FDL route from time slot 0 to a time slot when p is available is equivalent to finding a path from $T_0(0)$ to $O(p)$. The capacities of the FDL arcs and the output arcs respectively represent FDL availability constraints $C(\text{FDL } l, t)$ and output availability constraints $C(\text{output } p, t)$. Since at most N cells can arrive at a particular time slot simultaneously, the delay arcs impose no restrictions. If a flow unit represents a cell, the number of lost cells is minimum for a maximum flow from $T_0(0)$ to the $O(p)$ nodes. The sink node S , which completes the model, is connected to every $O(p)$ by the requesting arcs, whose capacity is the number of cells demanding a connection to output p at the current time slot. In a particular solution, each flow unit through a delay arc $(T_k(t), D(t))$ means that a cell is delayed for an aggregated delay t resulting from k circulations. To account for the cost of a given assignment, and to introduce the

Manuscript received November 17, 2008. The associate editor coordinating the review of this letter and approving it for publication was N. Ghani.

The authors are with Gradiant, Spain (e-mail: mrodelgo@gradient.org).

Digital Object Identifier 10.1109/LCOMM.2009.081942

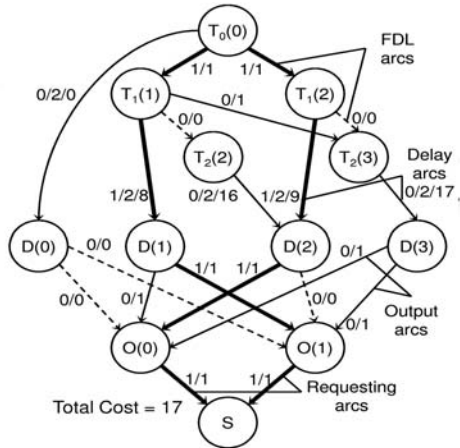


Fig. 2. Modified slot transition diagram (flow/capacity/[cost]).

number of circulations and the cell delay in the optimization process, each delay arc $(T_k(t), D(t))$ has an appropriate cost $cost(T_k(t), D(t))$. If we denote the flow through a delay arc $(T_k(t), D(t))$ of a given assignment as $flow(k, t)$, the number of circulations and the total delay for that solution will be $\sum_{\forall k,t} flow(k, t) \cdot k$ and $\sum_{\forall k,t} flow(k, t) \cdot t$, with respective upper bounds $U_c = N \cdot K$ and $U_d = N \cdot (F - 1)$. If we assign a weight $k \cdot (U_d + 1) + t$ to output arc $(T_k(t), D(t))$, the total cost becomes $(U_d + 1) \cdot \sum_{\forall k,t} flow(k, t) \cdot k + \sum_{\forall k,t} flow(k, t) \cdot t$. According to this expression, saving a single cell circulation is better than decreasing the cell delay for its maximum value. Therefore, a minimum cost assignment has the lowest average delay for a minimum number of circulations. In the example in Figure 2 the cost of delay arc $(T_1(1), D(1))$ is $k \cdot (U_d + 1) + t = 1(2 \cdot 3 + 1) + 1 = 8$. Analogously, if the cost of the output arcs is $k + t \cdot (U_c + 1)$, cell delay is more important for minimization than the number of circulations.

Given this model, the well-known min-cost max-flow optimization from the source $T_0(0)$ to the sink S seeks the maximum flow between them that fulfills the capacity constraints (FDL/output availability and cell requesting demands), at a minimum cost, as shown in Figure 2. This maximizes throughput and minimizes the number of circulations and cell delays. Each flow unit through an output node $O(p)$ represents a scheduled cell to output p and a route that is given by the flow through the FDL arcs.

There is an inconsistency in both the proposed characterization and the PIFA model [3]. If two FDL arcs of the same length l depart from two different time slot nodes associated to the same delay $T_k(t)$ and $T_{k' \neq k}(t)$, there are two representations of the availability of the FDLs with length l at time slot t , and the characterization is not valid. nevertheless, it is valid in practical scenarios where $K = 2$, since the FDL arcs from $T_0(0)$ and $T_1(t)$ have different associated delays. Moreover, since the integer decomposition as a sum of powers of two is unique, with common power-of-two FDL delays [3], and if the cells are not allowed to traverse an FDL with a given delay twice, the time slot nodes for a given time slot t are unique, $T_k(t) = T(t)$ (the delay nodes $D(t)$ are not needed at all) and there are no inconsistencies for any K . In fact, since only a few inefficient routes (longer than necessary) are

discarded, the impact on performance is negligible.

III. MIN-COST MAX-FLOW ALGORITHM

The well-known Ford-Fulkerson algorithm [4] is used to solve the min-cost max-flow problem. If there are paths from the source to the sink with excess capacity, the algorithm sends flow along the cheapest path, which is called an *augmenting path*. After each iteration, the cost of the flow through the graph is minimum, and the algorithm stops at maximum flow. In order to find augmenting paths, the algorithm looks for the minimum cost path in the residual network, namely the same network with different capacities, such that the capacity in arc (u, v) from node u to node v is $capacity(u, v) - flow(u, v)$, and no flows. Note that sending a flow $flow(u, v)$ through any arc (u, v) with cost $cost(u, v)$ may saturate (u, v) , but it opens a new arc (v, u) in the opposite direction with capacity $flow(u, v)$ and cost $-cost(u, v)$. The residual graph may have negative costs and the Bellman-Ford algorithm has to be used instead Dijkstra's to find the min-cost paths. Since every augmenting path has to traverse an output arc with maximum capacity 1, it will have unitary flow and each algorithm step will assign a single cell. At each time slot, there may be many solutions fulfilling the given assumptions. For the sake of fairness, the priority of the cells demanding each output is balanced along different time slots in a round-robin scheme. To simplify this process, the direction of every arc is reversed: node $T_0(0)$ becomes the sink, both S and the requesting arcs are deleted, and each $O(p)$ node acts as a source until all the cells demanding the output (the capacity of the original $(D(t), O(p))$ requesting arc) are assigned. The algorithm then considers node $O((p + 1) \bmod N)$, and so forth. At the next time slot, the algorithm first considers output $O((p + 1) \bmod N)$, and so forth.

Figure 3 shows the evolution of the min-cost max-flow algorithm for a given configuration if two cells requesting outputs 1 and 2 are scheduled and the number of cell circulations is prioritized over cell delay. In Figure 3(a), the min-cost augmenting path (in bold dashed line) from $O(1)$ is calculated in order to assign the cell that demands output 1, and its cost is 8 since it traverses just one delay arc with cost 8. The cell is temporarily scheduled with delay 1, since the corresponding FDL route has minimum cost. In Figure 3(b), the min-cost augmenting path with cost 9 from $O(2)$ is calculated in order to assign the cell that demands output 2, and the previous cell assignment changes. Unlike the SEFA solution, both cells become scheduled in the optimum in Figure 3(c). The min-cost max-flow algorithm has N Bellman-Ford steps at most (one per incoming cell) and each one has $O(V \cdot E)$ complexity (V and E respectively stand for the number of nodes and arcs). Thus, the proposed algorithm has $O(N \cdot V \cdot E)$ complexity. This may be high for a real-time implementation, but it allows the generation of theoretical bounds to evaluate suboptimal methods. The large simulations of section IV took a few minutes on a desktop computer.

IV. RESULTS

We evaluated the practical SEFA algorithm [3] with our optimum bound on a $N = 32$ optical switch for i.i.d. Bernoulli

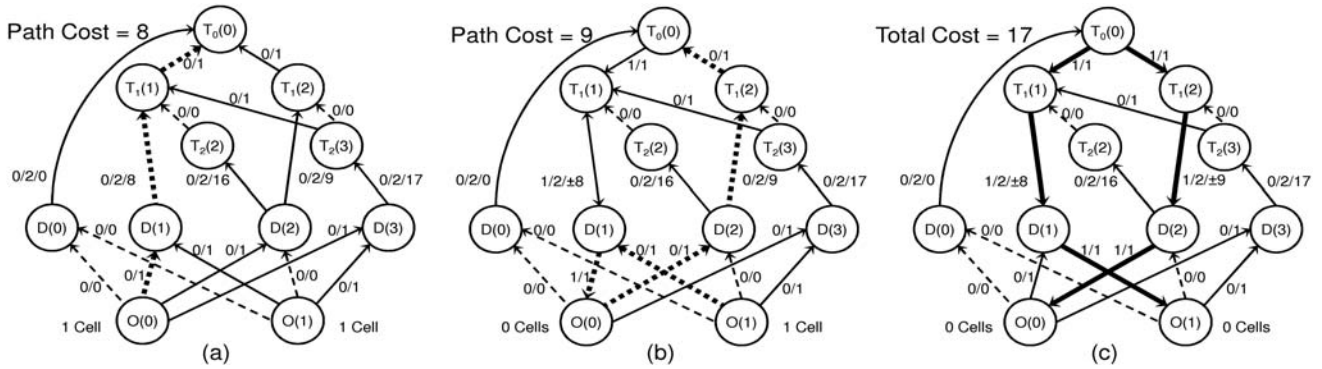


Fig. 3. Algorithm example (flow/capacity/[cost]): (a) Augmenting path for output 1 cell. (b) Augmenting path for output 2 cell. (c) Assignment result.

arrivals. The simulated configurations had $Z = 32$ FDLs, in groups of (5, 5, 5, 5, 4, 4, 4) FDLs with respective delay values of (1, 2, 4, 8, 16, 32, 64) cell times. Finally, we set $K = 2$ and $F = 128$. Figure 4 shows the simulation results for SEFA and the min-cost max-flow bound in terms of cell loss rate, average delay, and number of circulations. We checked that, for every configuration, the performance of SEFA was close to the optimum, as provided by our theoretical bound. It should also be noted that emphasizing delay over circulation minimization increases cell circulations and cell loss rate, and cell delay is even higher at medium loads, since using more FDLs than necessary to reduce cell delay at the current time slot yields an overall delay increase in subsequent time slots.

V. CONCLUSIONS

We have proposed a min-cost max-flow optimization model for shared-FDL optical switches that provides optimum performance bounds. These bounds allow the evaluation of practical schedulers. In this sense, we have shown that the performance of SEFA is close to optimum.

ACKNOWLEDGMENTS

The work described in this paper was carried out with the support of the BONE-project (“Building the Future Optical Network in Europe”), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme, and the support of grants TEC2007-67966-C03-02 CON-PARTE 2 (Spanish Ministry of Education and Science) and PGDIT08TIC010CT MIND-GAP-5 (Xunta de Galicia).

REFERENCES

- [1] D. K. Hunter, M. C. Chia, and I. Andonovic, “Buffering in optical packet switches,” *J. Lightwave Technol.*, vol. 16, pp. 2081-2094, Dec. 1998.
- [2] P. Pavon-Marín, J. García-Haro, J. Malgosa-Sanahuja, and F. Cerdan, “Maximal matching characterization of optical packet input-buffered wavelength routed switches,” in *Proc. HPSR 2003*, Torino, Italy, June 2003.
- [3] S. Y. Liew, H. J. Chao, and G. Hu, “Scheduling algorithms for shared fiber-delay-line optical packet switches—Part I: the single-stage case,” *J. Lightwave Technol.*, vol. 23, no. 4, pp. 1586-1600, Apr. 2005.
- [4] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

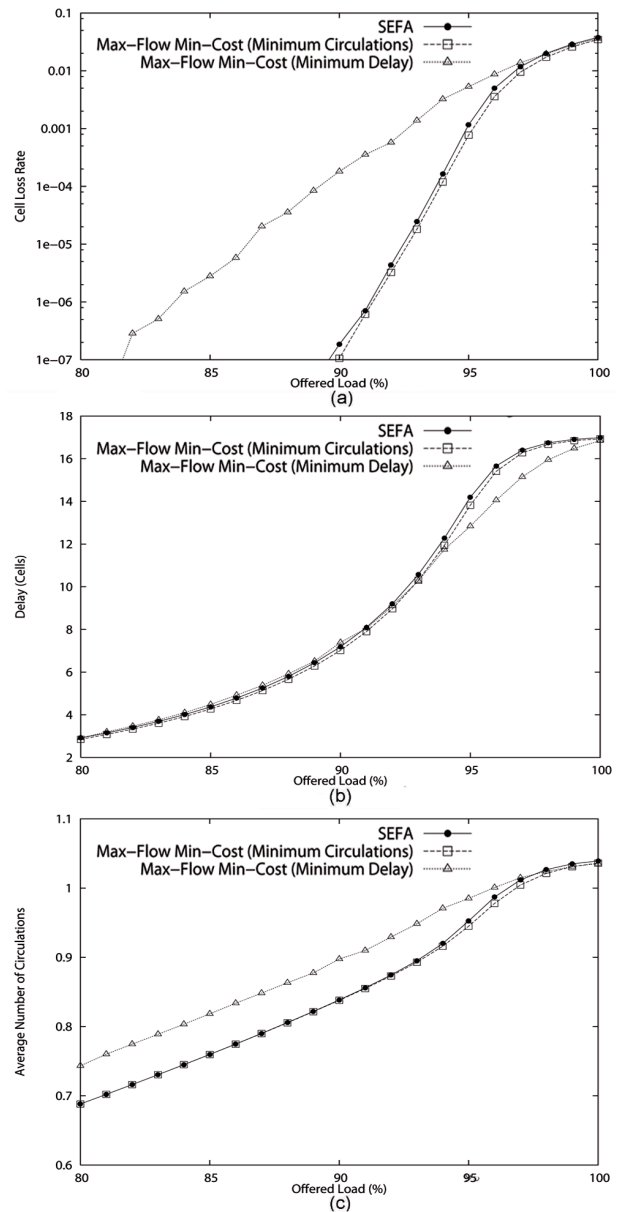


Fig. 4. Simulation results: (a) Cell loss rate. (b) Average cell delay. (c) Average number of circulations.