On the Capacity of Optical Networks: A Framework for Comparing Different Transport Architectures

Guy Weichenberg[†⋆], Vincent W. S. Chan[†], and Muriel Médard[‡]
{gew, chan, medard}@mit.edu
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

***Abstract*— In this work, we compare three optical transport network architectures: optical packet switching (OPS), optical flow switching (OFS), and optical burst switching (OBS). Our comparison is based on a notion of network capacity as the set of exogenous traffic rates that can be stably supported by a network under its operational constraints. We characterize the capacity regions of the transport architectures, and show that the capacity region of OPS dominates that of OFS, and that the capacity region of OFS dominates that of OBS. Motivated by the incommensurate complexity/cost of comparable transport architectures, we investigate the dependence of their relative capacity performance on the number of switch ports per fiber at core nodes. We find that when OFS and OBS core nodes have significantly many more switch ports per fiber than OPS core nodes, then the capacity regions of OFS and OBS (in the absence of receiver collisions) dominate that of OPS; and when the number of switch ports per fiber is only moderately more in OFS and OBS than in OPS, then it is possible that OFS and OBS do not dominate OPS, but are not dominated by OPS either.**

I. Introduction

Optical networking is unfolding as a two-generation story. The first generation of optical networks employed optical fibers as replacements for copper or microwave radio links. These networks maintained traditional architectures that were tailored to the use of electronic network components. Second generation optical networks, which are presently emerging, employ other optical components for network functions in addition to transport. Most of these components either have no electronic analog or behave very differently from their electronic counterparts. As a result, the design of optical networks must be rethought at the most fundamental level.

Expansive optical networks are conventionally partitioned into three hierarchical tiers: the local-area, metropolitan-area, and wide-area. For simplicity, we refer to the local-area and metropolitan-area collectively as the access, and to the wide-area as the core. The networks at these different tiers, while often treated as decoupled systems, are actually highly interdependent. For example, access network architecture influences core network architecture through traffic which is aggregated in the access and fed to the core. Thus, the design and analysis of optical networks should employ a holistic approach which considers networks in their entirety.

In this work, we develop a framework for comparing transport network architectures on the basis of network capacity, which we define as the set of exogenous traffic rates that can be stably supported by a network under its operational constraints. We are able to suppress access network architecture in our capacity analyses, as they are independent of detailed traffic statistics. We examine three prominent candidate architectures for optical transport in the core: optical packet switching (OPS), optical flow switching (OFS), and optical burst switching (OBS). The relationship among these architectures is illustrated via a three-dimensional taxonomy in Figure 1. Each architecture's physical and operational properties (e.g. core buffering, scheduling) impose constraints on its capability for logical topology reconfiguration, and naturally lead to different performance regimes.

Our performance metric of network capacity is particularly relevant to core networks because, owing to the high cost of supporting a wavelength of traffic, capacity is a precious commodity in the core (but not necessarily in the access network). We recognize, however, that while the question of the capacity limits of a core network is important, it is not the only performance criterion by which the network should be assessed. Delay, for example, is another key performance metric that is not addressed in this work. Ultimately, a network should not be judged on performance alone, but rather on the performance-cost trade-off it presents to the end user. Thus, the most useful comparison would include a detailed complexity/cost model for each of the candidate networks. Indeed, this issue of complexity/cost motivates our discussion in Section V.

Our approach to characterizing the capacity of optical networks differs from those of preceding works in various respects. Other works have applied the matrix decomposition results of Birkhoff [1] and von Neumann [2] to networks [3]–[5]. Inherent in such approaches are two limiting assumptions. First, the network, when viewed as a switch, is nonblocking. Although the underlying physical topology of a network is rarely a complete graph (i.e. a graph in which each node shares

The research in this paper was supported by: [†]Defense Advanced Research Projects Agency, "Robust Architectures for Multi-Service, Multi-Level Reliability, Multi-Level Service and Multi-Priority WDM Local Area Networks", MDA972-02-1-0021; [‡]National Science Foundation ITR/SY, "High Speed Wavelength-Agile Optical Networks", 008963-001; and [⋆]Natural Sciences and Engineering Research Council of Canada.

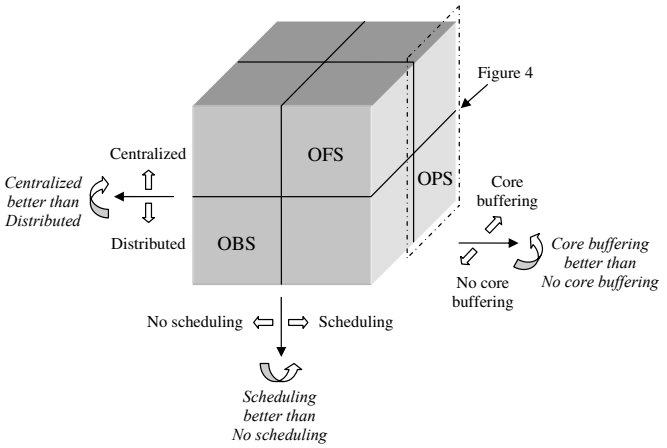


Fig. 1. Taxonomy of optical network architectures. Relative merit is based on capacity performance. Figure 4 further categorizes the indicated region according to the nature of the scheduling used.

an edge with every other node), these approaches achieve nonblocking logical topologies by assuming sufficiently many active wavelengths in a fiber. Second, switching of data in the network is cell-based, even though data transactions are naturally variable in length. These cell-based schemes thus either employ framing or segmentation and reassembly of transactions which result in additional overhead, and possibly larger transaction delays. Our work is general in that arbitrary numbers of wavelengths are considered, and data transactions are treated as indivisible entities in OFS and OBS[1]. Furthermore, we investigate the relationship among the capacity regions of the optical network architectures as a function of the number of switch ports per fiber in core network nodes. Finally, our approach to characterizing the capacity region is constructive in that online, capacity-achieving scheduling policies are outlined.

We further remark that our investigation of network capacity assumes uncoded data transmission. Coding over networks has recently been proposed as a method of altering the capacity regions of networks [6], [7]. Network coding is a particularly attractive technique in settings where channel bandwidth is precious and computation is inexpensive. We do not pursue network coding in this work, as in our present context of optical networks, the converse is true: computational costs at network nodes outstrip the costs of supporting data transmission in fiber.

This work invokes several results from switching and networking theory, the background for which will be presented as required in the following sections. In the next section, we describe the candidate optical network architectures. We capture the essence of the architectures with a simple example in Section III. In Section IV, we formally introduce the notion of network capacity and characterize capacity regions of the three architectures. In Section V, we investigate the relationship among the capacity regions of the architectures as a function of the number of switch ports per fiber in core nodes. We conclude this work in Section VI.

[1]These two transport mechanisms consume resources for set-up, and we therefore feel that entire transactions should be sent once set-up is complete.

II. Description of optical transport network architectures

In this section, we describe the three optical transport network architectures of interest to us. The detailed hardware implementations of these architectures are not described. We suggest that the reader refer to Figure 1 to maintain perspective of the relationship among these architectures.

A. OPS

We assume OPS to be functionally equivalent to its electronic analog, electronic packet switching, although we recognize that the necessary building blocks to realize OPS networks have yet to be developed, or are in their infancy.

An optical packet switch is a cell-based, input queued (IQ) switch[2] employing virtual output queueing at its input ports. That is, each input port keeps a separate queue for each output port, for a total of N^2 queues in an $N \times N$ IQ switch. In this model, each input and output port corresponds to a wavelength channel. In order to establish full connectivity between input and output ports, wavelength converters are necessary in optical packet switches. We assume that each virtual output queue (VOQ) has infinite buffering capability. An OPS network is a network of optical packet switches which make scheduling decisions in a distributed fashion.

B. OFS

In an OFS network, transmission of data is coordinated on an electronic control plane in a scheduled manner between end users, akin to circuit-switching, albeit for shorter durations [8]. We assume that the smallest granularity of bandwidth that can be reserved across the core is a wavelength. We further assume that wavelength conversion is not used in the network core. Motivated by the minimization of network management and switch complexity in the network core, we require that flows be serviced as indivisible entities. In the event that several single users have transactions which are not sufficiently large to warrant their own wavelength channels, they may multiplex their data onto wavelength channels for transmission across the core. Note that, in OFS networks, unlike OPS networks, all queueing of data occurs at the end users, thereby obviating the need for buffering in the core. As in OPS, we assume infinite buffering capability at queues.

In reference to Figure 1, OFS is considered a centralized transport architecture in that coordination is required for logical topology reconfiguration. However, traffic in the core will likely be efficiently aggregated and sufficiently intense to warrant a quasi-static logical topology that changes on coarse time scales. Hence, the centralized management and control required for OFS is not expected to be onerous. The network

[2]While the IQ switch design is less general than that of the combined input and output queued (CIOQ) switch, these switches are optimal with respect to capacity performance assuming unicast flows.

management and control carried out on finer time-scales will be distributed in nature in that only the relevant ingress and egress access networks need to communicate.

C. OBS

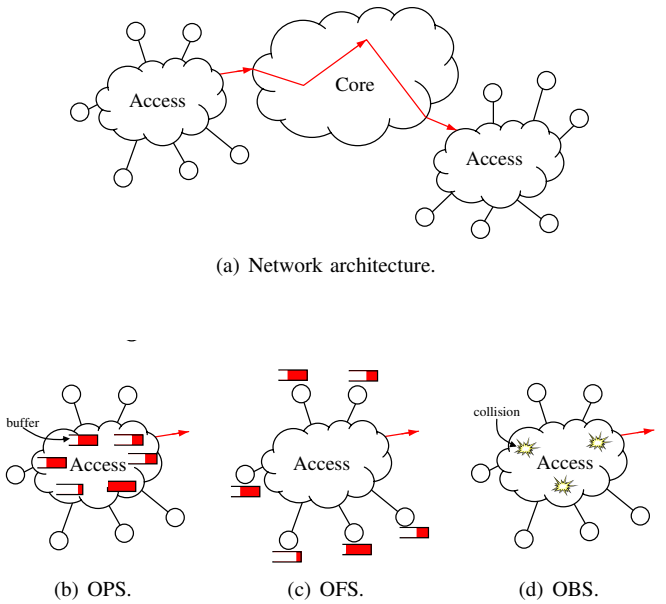
Like OFS, OBS is a scheme that uses an electronic control plane which is separated from an optical data plane in both time and space. One variation of OBS is based on the Tell-and-Go protocol for Asynchronous Transfer Mode (ATM) networks [9], in which end users act as sources for bursts of data [10]. In the more common implementation of OBS, however, packets are assembled at access nodes, according to destination and quality of service, to form collections of packets known as bursts. In either case, prior to transmission of a burst, a control packet is sent into the control plane, where it is processed electronically at intermediate nodes requesting that an all-optical path be set up for the ensuing burst. After a delay, the burst is transmitted into the core on a wavelength channel without acknowledgement, ideally reaching its destination access node transparently. Note, however, that there is a chance that the burst may be discarded and/or handled in another way owing to a lack of resources at one or more of the intermediate nodes.

While the above description captures the spirit of the OBS transport architecture, there is a great deal of variability in exactly how an OBS network can be implemented. For such details, we refer the reader to [11]–[14] and references therein. In OBS, all queueing of data occurs in infinite buffers at the access nodes or the end users, as there is no buffering capability in the core. We also assume, as in OFS, that wavelength conversion is not used in the network core.

III. An illustration: Example 1

In this section, we consider the performance of OPS, OFS, and OBS in the context of the simple network drawn in Figure 2: an access network with N users connected to another access network via a dedicated core wavelength channel.

One purpose of this example is to capture the essence of the aforementioned optical transport network architectures. In spite of its simplicity, this example also embodies some key ingredients of communication across the core. In particular, we envision communication between access networks to be carried out via quasi-static wavelength circuits, as we assume that core traffic will be efficiently aggregated and sufficiently intense to always warrant at least a wavelength of traffic between access network pairs. If traffic demands between access networks are sub-wavelength, however, then it is reasonable to assume that traditional electronic data networking solutions will suffice. Within the access, we envision broadcast local-area network (LAN) and routed metropolitan-area network (MAN) architectures. Our justification is that these will be lightly loaded optical networks in which utilization of fiber should be prudently traded for network architecture simplicity up to a maximum number of users dictated by power limitations, beyond which routing is required.



(a) Network architecture.

(b) OPS. (c) OFS. (d) OBS.

Fig. 2. Illustration of the network considered in Section III. The OPS and OFS implementations prevent collisions with buffering and scheduling, whereas the OBS implementation does not.

A. OPS

In the OPS implementation of our simple network, each user transmits on a dedicated channel, which is fed into an input port of an $N \times 1$ optical packet switch. Here, the packet switch merely multiplexes traffic from the N data streams onto the dedicated core wavelength channel. After the data is multiplexed, it is fed into a passive, optical, broadcast network which distributes copies of the data to each of the receiving users. Owing to the simplicity of the network, we do not break up transactions into cells, as this would have a minor effect on network performance. In analyzing this network, we may treat the switch as a queueing system employing first-in first-out (FIFO) scheduling. In more complex meshed networks, where output ports of switches ramify into the network, FIFO scheduling results in poor performance, owing to head-of-line blocking.

If we assume packet arrivals that occur according to a poisson process of rate λ and packet lengths that are exponentially distributed with mean $1/\mu$, then the performance of our system is that of an $M/M/1$ queueing system. The throughput is thus given by:

$$S_{ops} = \rho = \frac{\lambda}{\mu}.$$

Neglecting propagation delay, the mean delay is:

$$D_{ops} = \frac{1}{\mu - \lambda}.$$

The throughput-delay performance is plotted in Figure 3.

B. OFS

In OFS, an electronic scheduler and an optical cross-connect (OXC) are used to mediate access to the dedicated core

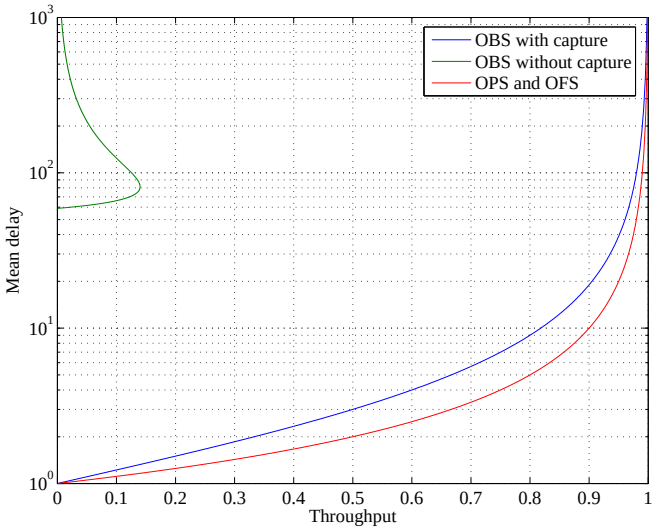


Fig. 3. Mean delay versus throughput for the network considered in Section III. We assume exponentially distributed transaction lengths with unit mean ($\mu = 1$), exponentially distributed back-off durations with unit mean ($\theta^{-1} = 1$), and 30 users ($N = 30$).

wavelength channel, akin to an optical packet switch in OPS. A critical difference, however, is that an OXC does not have buffering capability, whereas an optical packet switch does. Users thus employ scheduling prior to transmission to avoid collisions. In spite of this, the analysis of OFS in the context of this network is identical to that of OPS because the users may emulate an $M/M/1$ queueing system through coordination. The throughput-delay performance is illustrated in Figure 3, assuming poisson flow arrivals and exponentially distributed flow lengths.

C. OBS

In one OBS implementation, we assume that the access networks are passive optical broadcast networks. Usage of the channel occurs in a random-access fashion without scheduling or coordination among users. In the absence of a channel capture mechanism[3], the system may be modelled as a finite user Aloha system with variable length bursts. The throughput of such a system was derived in [15]. In this model, it is assumed that each user produces a sequence of independent transmit and idle states, which is independent of other users. Each transmit state, which corresponds to a burst being transmitted, has length drawn from density $L(t)$ with mean $\overline{L}$; and each idle state is exponentially distributed with mean λ^{-1}.

For bursts with exponentially distributed lengths of mean $\overline{L} = \mu^{-1}$, it can be shown that the throughput and mean delay are:

$$\begin{aligned} S_{obs}^{nc} &= \frac{N\rho}{(1+\rho)^N\left(1+(N-1)\rho\right)^2} \\ D_{obs}^{nc} &= \frac{(1+\rho)^{N-1}\left(\rho^{-1}+N+(N-1)\rho\right)}{\mu} - \frac{1}{\lambda}, \end{aligned}$$

[3]Channel capture is the reservation of the channel for the duration of a burst if the channel is free when the burst initially attempts transmission.

where $\rho = \lambda/\mu$. When channel capture is possible, the throughput and mean delay are given by [16]:

$$\begin{aligned} S_{obs}^{c} &= \frac{\lambda}{\mu} \\ D_{obs}^{c} &= \frac{1}{\mu-\lambda}\left(1+\frac{\lambda}{\theta}\right), \end{aligned}$$

where we assume that the time between transmission attempts of a given burst is exponentially distributed with mean θ^{-1}.

The throughput-delay performance for OBS with and without channel capture is plotted in Figure 3.

Figure 3 illustrates the benefit of scheduling, through the superior performance of OPS and OFS compared with OBS. The performance disparity is particularly pronounced when the OBS network is not capable of channel capture. When channel capture is possible for OBS, it is seen that full utilization of the wavelength channel is possible, albeit at a larger expected delay than OPS and OFS. In fact, the additional incurred delay is inversely proportional to the retransmission rate θ. It should be noted, however, that under general traffic conditions, it is not true that OBS can achieve full utilization of the wavelength channel, even when channel capture is possible. This is a consequence of the fact that a $G/G/1/1$ retrial queue — a $G/G/1/1$ queueing system with an infinitely large waiting room for rejected customers from which these customers retry for service after waiting some time — generally does not achieve full server utilization [17].

Owing to the simplicity of this example, however, some questions regarding the general relative performance of these architectures remain open. For example, in more complex networks, which of OPS and OFS performs better? How much worse will OBS perform relative to OPS and OFS in switched, multihop, core network topologies compared to the single-hop topology considered here? Figure 1, the taxonomy of the optical network architectures, provides a clue as to their relative capacity performance. While OBS is expected to have inferior performance, neither OPS nor OFS appear to be the clear winner in this respect. We investigate this question in detail in the remainder of this work.

IV. Capacity of Optical Networks

We model networks as directed graphs, where graph arcs and vertices represent directed fiber links and network nodes, respectively. Each fiber can support a maximum of t unit capacity active wavelength channels. We assume that each of these active wavelength channels carries data which is aggregated from the users associated with a particular access node. For example, an access node that has Δ_{out} outgoing fibers can support a maximum of $\Delta_{out}t$ wavelengths of traffic. Thus, $\Delta_{out}t \geq N\gamma$, where N is the number of users associated with the access node, and γ is each user's duty cycle (i.e. the fraction of time that a user has enough data to occupy a wavelength channel). We further assume that at each node there exist dedicated transmitters, receivers, and any other processing equipment (depending upon the

network architecture) that may be required to support each active wavelength channel. For the purpose of a capacity analysis, the active wavelength channels may be decoupled and considered independently, and it therefore suffices to examine only one of these channels in isolation[4]. Moreover, we assume a normalized wavelength channel capacity of unity.

In the remainder of this work, we assume that time is slotted and we neglect propagation delay. Furthermore, we only consider the case of unicast transactions. We associate a transaction type with each source-destination node pair. Let $A_{i,j}(n)$ denote the cumulative number of exogenous cell arrivals at node i destined for node j by time slot n. As in [18], we assume that the arrival process $\{A_{i,j}(n)\}_{n=1}^{\infty}$ satisfies the Strong Law of Large Numbers (SLLN) and is stationary. Let vector Λ denote the set of exogenous traffic rates $\lambda_{i,j} = \lim_{n\to\infty} \frac{A_{i,j}(n)}{n}$. For each source-destination pair (i, j), we associate a path, if one exists, from node i to node j. We define a routing as this ensemble of source-destination pairs and paths.

Let Q denote the number of queues in a network. Let X_n be a Q-dimensional vector whose k^{th} element represents the number of data transactions in the k^{th} queue at time slot n. Likewise, let the k^{th} element of the Q-dimensional vectors D_n and E_n represent the number of departures from and entrances (exogenous or endogenous) to the k^{th} queue at time slot n, respectively.

We are now ready to introduce the idea of network capacity.

Definition 1
A system of queues is rate-stable if:

$$\lim_{n\to\infty} \frac{X_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (E_i - D_i) = 0, \text{ w.p. 1.}$$

Definition 2
The capacity region of a network is the set of exogenous traffic rates for which the system of queues in the network is rate-stable for some routing under its operational constraints.

Definition 3
A set of exogenous traffic rates is admissible if a routing exists for which every link in the network is offered a rate of traffic which is strictly less than its link capacity.

We emphasize that the capacity region of a network is not tied to a particular routing. Rather, it is the collection of achievable traffic rates taken over the set of all routings. Furthermore, the capacity region of a network must lie within the set of admissible rates, for otherwise, at least one of the network's queue occupancies would grow without bound.

[4]This is not true for a delay analysis, as several servers working together can achieve lower expected delay than several servers working independently under the same traffic intensity.

A. OPS

To characterize the capacity of OPS networks, we first propose a taxonomy of networks of IQ switches. Our taxonomy is based on two axes, as illustrated in Figure 4 (and Figure 1). The first axis characterizes networks according to the online/offline nature of the scheduling algorithm used. Online scheduling algorithms make use of queue occupancy information while offline algorithms do not. Thus, the capacity region for online algorithms dominates[5] that of offline algorithms. The second axis relates to the nature of the information available to switches when making scheduling decisions. In centralized scheduling each network switch is privy to network-wide information, and in distributed scheduling each switch only has access to local information at that node. The capacity region for centralized algorithms therefore dominates that of distributed algorithms. This leads us to conclude that scheduling policies in quadrant 2 of Figure 4 have the largest capacity region. We note that OPS falls into quadrants 3 and 4 of Figure 4.

Until recently, the literature on switch scheduling mostly examined the performance of a switch in isolation. An important result by McKeown [19] shows, through the use of two different scheduling policies based on maximum weight matching (MWM)[6], that stability of an IQ switch can be attained for any admissible traffic pattern with independent arrival processes. As shown by Andrews and Zhang in [20], McKeown's result does not extend to networks of IQ switches. In [18], Marsan *et al.* show, however, that there exist scheduling policies in quadrants 3 and 4 of Figure 4 which are rate-stable as long as the offered load is stationary, satisfies the SLLN, and does not overload any link in the network. While this result assumes a unique route for each transaction type, multiple transaction types may be associated with each source-destination pair. By mapping a source-destination type with a particular routing, we have the following result:

Theorem 1
The capacity region of an OPS network is the convex hull of the union (over all possible routings) of the admissible set of traffic rates. That is, OPS networks achieve the maximum possible capacity region.

B. OFS

We now address the capacity region of OFS networks, and more generally, networks with buffering at source nodes but no buffering capability in the network core. In such networks, data is scheduled to traverse the network from source to destination without being buffered at intermediate nodes.

To characterize the capacity region of this family of networks it is helpful to view a network as a large, generalized

[5]Set A is dominated by set B if $A \subseteq B$.

[6]MWM is a class of scheduling policies that employs some weighting function to assign each VOQ a weight, and then matches the switch's input ports to its output ports according to the matching which achieves the maximum weight. Examples of weighting functions are the number of cells residing in the VOQ and the age of the oldest cell in the VOQ.

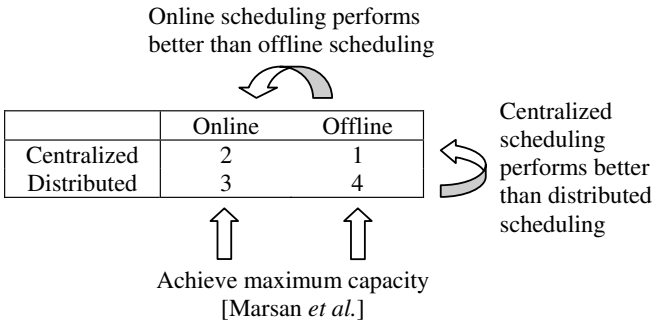


Fig. 4. Taxonomy of networks of IQ switches. This matrix is a further categorization of the core buffering/scheduling region indicated in Figure 1.

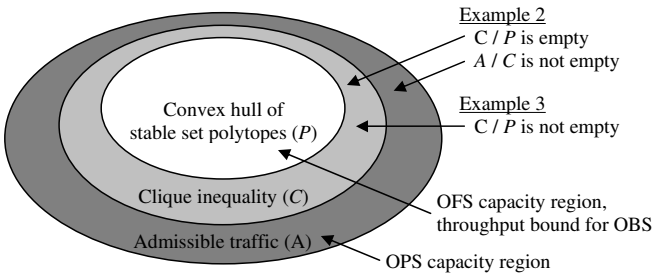


Fig. 5. Relationship among different rate regions when $w = t$.

switch. Viewed this way, input and output ports correspond to the nodes in the network, and a connection between an input and output port represents a flow between two nodes.

Building on the work in [21], [22], the capacity region of such a network is related to stable set polytopes of conflict graphs. A stable set is a set of vertices in which no two vertices have an edge connecting them, and the convex hull of the incidence vectors of stable sets is the stable set polytope. A conflict graph is an undirected graph in which vertices represent the set of flows to be served in the network. An edge exists between vertices i and j if the flows corresponding to nodes i and j cannot simultaneously exist in the network (i.e. the flows share at least one link). Note that, for a fixed routing, there is a one-to-one mapping between feasible network states and stable sets of the conflict graph. Thus, by means of time-sharing, the stable set polytope of the conflict graph is achievable. An application of [23, Proposition 1] shows that this also fully characterizes the capacity region of a network without core buffering. By further time-sharing over all the possible routings and invoking [23, Proposition 1] again, we have the following result:

Lemma 1
The capacity region P of a network without core buffering is the convex hull of the union (over all possible routings) of the stable set polytopes of the conflict graphs.

We emphasize that Lemma 1 does not characterize the capacity region of OFS networks, but provides an outer bound for its capacity region. This is because the only constraint we imposed in the lemma's derivation was the absence of buffering in the network core. In particular, we allowed for flows to be broken up into arbitrary granularities and serviced piecemeal. As discussed earlier, the assumption of being able to break up flows and service them piecemeal is contrary to the spirit of OFS. We, therefore, naturally wonder if there is an inherent sacrifice in the capacity region of OFS relative to the region P. Our main result of this section (Theorem 2) shows, somewhat surprisingly, that the answer is no.

We now turn our attention to Figure 5 which illustrates the relationship among P, the admissible traffic region A, and a region that, for a reason which will soon become clear, we call the clique inequality region C. We introduce the region C into our discussion as it provides intuition as to why the regions A and P differ. The concept of a clique — a fully connected subgraph — is important for the following discussion. Let F denote the number of different types of flows in a network, or equivalently, the number of vertices in a conflict graph. Let $Z = (z_1, z_2, \ldots, z_F)$ be a point in F-dimensional Euclidean space representing the flow rates. Then the following two sets of inequalities are satisfied inside the stable set polytope of the conflict graph for a particular routing [24]:

- Trivial constraints: $0 \leq z_i \leq 1$, for all i.
- Clique inequalities: $\sum_{i \in K} z_i \leq 1$, for every clique K.

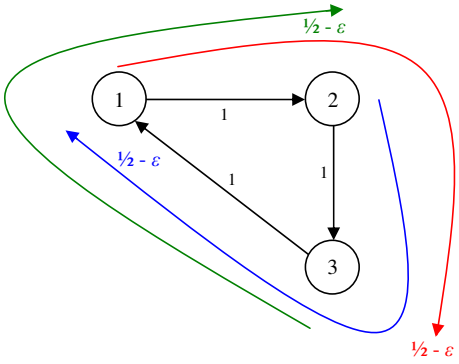
In fact, it can be shown that the stable set polytope of the conflict graph P for a particular routing is the integer hull of the polytope defined by the trivial constraints and clique inequalities. Since the problems of finding maximum-size stable sets and cliques in a graph are NP-complete, a simple inequality characterization of these regions generally does not exist [24].

We are now proceed to define the clique inequality region:

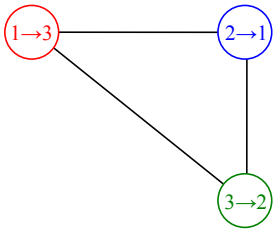
Definition 4
The clique inequality region C is the convex hull of the union (over all possible routings) of the rate regions defined by the trivial constraints and the clique inequalities.

In general, $C \subseteq A$ because the clique inequalities are stricter than the admissibility constraints, which state that no link may be oversubscribed. The clique inequalities require that any flows which form a clique in the conflict graph must have an aggregate capacity of less than unity. If all the flows in a clique share at least one common link, then the clique inequalities are equivalent to the admissibility constraints. However, as shown in Example 2, it is possible for flows to form a clique without all merging at a particular link. Therefore, we conclude that the clique inequalities are stricter than the admissibility constraints, and consequently lead to a smaller rate region. Thus, we have the result:

Lemma 2
The region A dominates the region C, which, in turn, dominates the region P. Hence, the capacity region of a network without core buffering is smaller than that of the analogous OPS network.

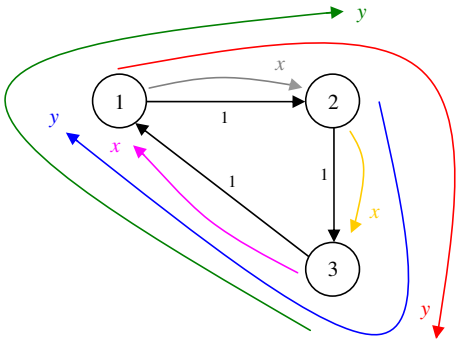


(a) Network lacking core buffering.

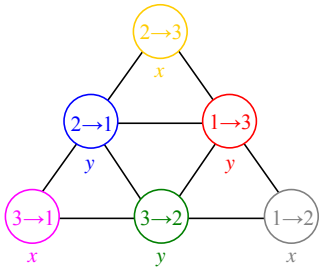


(b) Conflict graph.

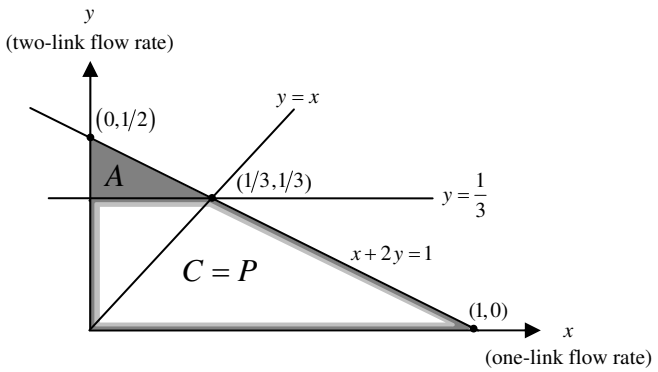
Fig. 6. Illustration of the network considered in Example 2 and its associated conflict graph when only two-link transaction types are assumed.



(a) Network lacking core buffering.



(b) Conflict graph.



(c) Rate regions

Fig. 7. Illustration of the network considered in Example 2, its associated conflict graph when an all-to-all traffic demand is assumed, and the different rate regions when one-link transaction types have rate x and two-link transaction types have rate y.

It is worth noting that, while the clique inequalities are always satisfied within the stable set polytope of a conflict graph, they exactly characterize the stable set polytope for a particular family of conflict graphs known as perfect graphs [24]. In a perfect graph, the chromatic number[7] equals the size of the largest clique for each of its induced subgraphs. For this family of graphs, the task of finding the region $C = P$ is solvable in polynomial time [24]. While it is true that for perfect graphs $P = C$, it is not necessarily true that $A = C = P$. Therefore, even for network topologies which maximize the OFS capacity region relative to the clique inequality region C, an OPS architecture atop the same network topology will generally have a larger capacity region. This is illustrated in the following example.

Example 2 (Three node ring)
Consider the network and associated transaction types drawn in Figure 6(a), where each transaction type has rate $1/2 - \varepsilon$ and the capacity of each link is unity. The corresponding conflict graph, which is perfect, is drawn in Figure 6(b). Since the clique inequalities are not satisfied, a schedule that can accommodate this traffic demand does not exist for networks which lack core buffering. However, the traffic demand is clearly admissible, which implies that an OPS network can accommodate the demand.

[7]The chromatic number of a graph is the least number of colors required to color its vertices such that adjacent vertices have different colors.

Consider the same network, but under the all-to-all traffic demand illustrated in Figure 7(a). The corresponding conflict graph, which is also perfect, is drawn in Figure 7(b). We assign the single-link transaction types rate x and the two-link transaction types rate y. Figure 7(c) then illustrates the different rate regions A, C, and P as functions of x and y. We note that this capacity region, which assumes uniform traffic, is actually a two-dimensional cross-section through the unconstrained six-dimensional capacity region. We also note that, assuming uniform all-to-all traffic (i.e. all possible transaction types have equal rates), then a network without core buffering can achieve the same rates as an OPS architecture. In Figure 7(c), this

common operating point corresponds to the point $(1/3, 1/3)$.

Finally, owing to the perfectness of the conflict graph, the capacity region for networks which lack core buffering is completely characterized by the trivial and clique inequalities on the flow rates $z_{i\to j}$*:*

- $0 \leq z_{i\to j} \leq 1$, *for* $i, j = 1, 2, 3$ *and* $i \neq j$.
- $z_{1\to 2} + z_{1\to 3} + z_{3\to 2} \leq 1$.
- $z_{2\to 1} + z_{1\to 3} + z_{3\to 2} \leq 1$.
- $z_{2\to 1} + z_{3\to 1} + z_{3\to 2} \leq 1$.
- $z_{2\to 1} + z_{1\to 3} + z_{2\to 3} \leq 1$.

We now investigate online, cell-based algorithms that achieve rate-stability in the region P. In [23], [25], the authors address a family of scheduling policies known as MaxWeight scheduling (with MWM as a special case) in the context of a generalized, cell-based switch[8]. The switch model employed assumes that switch states follow a finite state, discrete time Markov chain, where, in each state, the switch has an associated finite set of scheduling choices. We may view a network lacking core buffering as such a generalized switch with a single state Markov chain in which the finite set of scheduling choices correspond to the feasible network configurations that can be used to service flows. In [23, Lemma 5], the author proves, using fluid model techniques, that MaxWeight scheduling policies achieve the maximum capacity region. This leads us to the following:

Lemma 3
The rate region P can be achieved using an online, cell-based MaxWeight scheduling algorithm (with MWM as a special case).

We highlight that Lemma 3 generalizes the optimality of MWM scheduling in two ways. First, it broadens the class of optimal scheduling policies to the MaxWeight family. Second, and more importantly for our discussion, it demonstrates the optimality of MaxWeight scheduling for generalized switches which differ from traditional nonblocking switches in that they may have additional constraints. In the context of this work, these additional constraints correspond to topology/resource constraints in the network.

The next lemma is an application of [27, Lemma 1] to our constrained switch model of networks lacking core buffering. The fluid model techniques employed in the proof in [27] are immediately applicable to our model. Specifically, in the fluid limit of a switch process, a scheduling algorithm that is "suboptimally bounded" is indistinguishable from an optimal scheduling algorithm.

Lemma 4
The capacity region of a network lacking core buffering can be achieved by an online, cell-based scheduling algorithm if the value of the weight of the matching it uses at each time slot is at most away from the maximum weight by a bounded constant.

[8]Stability results for generalized switches were originally obtained by Tassiulas and Ephremides [26].

In [27], the authors investigate the performance of scheduling algorithms for nonblocking IQ switches which are flow-based — they switch flows of variable number of cells as indivisible entities, rather than segment them into cells and switch them with cell-based schemes. The authors show that any flow-based algorithm which is work-conserving cannot be stable for all admissible traffic rates. In particular, for a scheme to be rate-stable, it is necessary to periodically "resynchronize" the switch state with the state of the queues. This requires the switch to wait for periods of time for some of the ports to become free. The authors thus propose a family of nonwork-conserving scheduling algorithms based on MWM that switch variable length flows as indivisible entities, which can be adapted to our OFS model. The previous lemma is instrumental because it implies that if we wait for bounded periods of time for the purpose of resynchronization, then the weight of our matching at any instant in time is less than the optimal weight by a bounded constant. By waiting for an arbitrarily long (but bounded) period of time before resynchronizing the switch, we can ensure that the bandwidth waste due to waiting is arbitrarily small. Thus, using the previous lemma, we have the following result:

Lemma 5
There exist online, nonwork-conserving, flow-based scheduling algorithms which are rate-stable for the capacity region of a network lacking core buffering, provided that the average flow length is bounded.

This immediately leads us to our main OFS result:

Theorem 2
The capacity region of an OFS network is the convex hull of the union (over all possible routings) of the stable set polytopes of the conflict graphs.

C. OBS

OBS networks can be viewed as incarnations of OFS networks in that they lack buffering capability in the core, and that they require bursts to be serviced as indivisible entities. However, owing to the fact that they employ random-access instead of scheduling, OBS networks are generally characterized by nonzero burst blocking probabilities. Specifically, as discussed in Section III, the fact that bursts may require retransmission can lead to instability on an individual link, even if the offered traffic is admissible [17]. Furthermore, the lack of coordination among core links implies that resources are wasted if they are consumed by bursts that are eventually discarded. This is illustrated in Example 3. For these two reasons, OBS networks are generally incapable of achieving rate-stability within the OFS capacity region. This leads to the following result:

Theorem 3
The capacity region of an OBS network is dominated by the capacity region of the analogous OFS network.

Note that the degree to which the throughput of an OBS network differs from the capacity region of the analogous OFS network depends upon the traffic statistics and the retransmission policy employed.

Obtaining analytic expressions for OBS network capacity regions is related to, and in fact more difficult than, characterizing the stability regions of retrial queues, for which analytic solutions are available only under special circumstances [16], [28]. As a result, with the exception of [29], studies of OBS network performance have usually only considered a single edge or core OBS node in isolation. These analyses, however, neglect the key property of OBS networks that resources are wasted if they are consumed by bursts that are eventually discarded.

We now analyze the performance of OBS networks under simplifying assumptions. We assume bursts to be variable length transactions that are assembled at access nodes, in contrast to our assumption in Section III that bursts are formed at end users. We further assume that bursts that lose contention at a node are dropped instead of being handled in some other way.

Our model of a link in an OBS network is a multiple-access system with a finite number of users which represent the burst types on that link. Furthermore, because access to a link is mediated by switch ports or tunable lasers, we assume that link capture is possible by a burst. Thus, if the link is free when a burst transmission is attempted on it, then the link is reserved for the duration of the burst. For analytical tractability, a burst traversing an OBS network is transmitted along a series of links that is treated as a cascade of independent multiple-access systems. Although the actual load is reduced as the network egress is approached, the load is unreduced when requests are made by the control packet preceding a burst. This is because the control packet requests resources from all nodes along the burst's intended path regardless of whether its requests at previous nodes were successful.

We assume that each burst type on a link produces, independent of other burst types, a sequence of independent transmit and idle states which represents the aggregate of fresh arrivals and retransmissions. Each transmit state, which corresponds to a burst being transmitted, has length drawn from a distribution $L(n)$ with mean $\overline{L}$; and each idle state is geometrically distributed with mean q^{-1}. While this assumption may not be entirely realistic because retransmissions corresponding to the same burst should have the same length, the derived throughput will not be affected owing to the independence of link capture and burst length. A shortcoming of our model, however, is that the delay between retransmits, and the delay between a successful (re)transmit and a subsequent initial transmit attempt are identical. Another shortcoming of our model is that the distribution of idle states is independent of whether or not there are backlogged packets attempting retransmission.

We assume that link capture requires all other burst types on that link to be idle, that the probability a burst type successfully transmits is independent of previous attempts to transmit, and that links are independent. It can be shown that the probability that a given burst of type j is successfully received at its destination d_j hops away is:

$$P_s(j) = \prod_{i=1}^{d_j} \left(\frac{q^{-1}}{q^{-1} + \overline{L}} \right)^{N_i - 1},$$

where N_i is the number of bursts traversing the i^{th} link along the burst's path.

From this probability of success, both the average delay for a burst type and the throughput of a link may be found. The average delay D experienced by a burst of type j from the time it arrives at its source node to the time it is successfully received at its destination, neglecting propagation delay, is:

$$\begin{aligned} D &= \overline{L} + (\text{avg. no. retransmissions})(1/q + \overline{L}) \\ &= \overline{L}/P_s(j) + (1/P_s(j) - 1)/q. \end{aligned}$$

The throughput of link i is:

$$S(i) = \sum_{j=1}^{F} P_s(j) \frac{\overline{L}}{\overline{L} + q^{-1}} I_i(j),$$

where F is the number of types of bursts in the network, and $I_i(j)$ is the indicator function that has the value of unity if a burst of type j traverses link i.

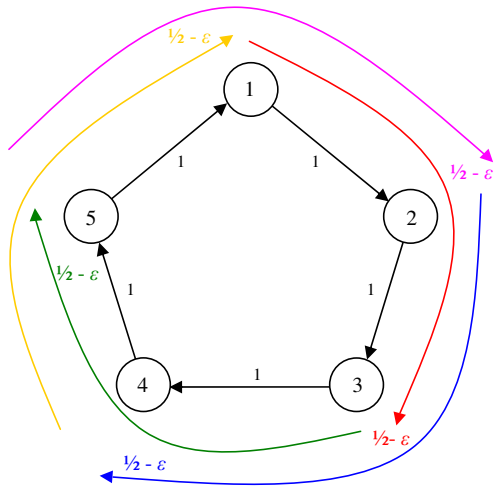
D. Five node ring example

In the following example, we illustrate through a simple, yet realistic, network topology and routing that the capacity region of OFS is smaller than that of OPS, and that the capacity region of OBS is, in turn, significantly worse than that of OFS.

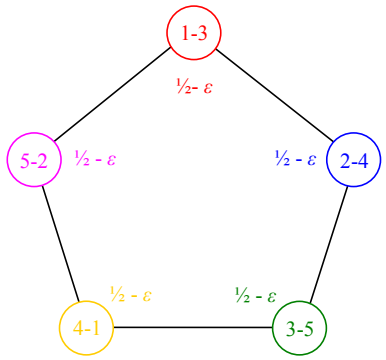
Example 3

Consider the five node ring depicted in Figure 8(a). Drawn in the figure are also the offered transaction types, each of which is of rate $1/2 - \varepsilon$. The capacity of each link is unity, implying that the traffic pattern is admissible and can therefore be serviced by an OPS architecture. In fact, if we constrain ourselves to uniform traffic, the capacity region of this OPS network contains the set of rates less than $1/2$.

By examining the conflict graph drawn in Figure 8(b), it is apparent that the clique inequalities are satisfied and that this traffic demand is therefore contained in the clique inequality region C. To determine whether the traffic pattern is contained within the OFS stability region P, we first observe that, at any instant in time, at most two of the flows may be serviced using an OFS architecture. Thus, at any instant in time, at least one link in the network is unutilized. This underutilization of at least $1/5$ of link resources implies that the traffic demand cannot be accommodated, as nearly full link utilization is required (when ε is very small). In fact, it can be shown that for uniform traffic, a rate bounded above by $2/5$ can be offered by each flow, which would achieve the $4/5$ utilization bound just mentioned. Thus, assuming uniform traffic, then the capacity region of this OFS network contains the set of flow rates less than $2/5$.



(a) Network lacking core buffering.



(b) Conflict graph.

Fig. 8. Illustration of the network considered in Example 3, and its associated conflict graph.

In an OBS implementation of the network, each link is shared by two contending bursts and each burst traverses two links. Thus, using the above approximate OBS analysis, the throughput of each link (and by symmetry, the network) is:

$$S = \frac{2\overline{L}q^{-2}}{(\overline{L}+q^{-1})^3}.$$

Assuming that $\overline{L} = q^{-1}$, which is equivalent to letting each burst type's aggregate (fresh arrivals plus retransmissions) rate be $1/2$, then the throughput of the system is $1/4$. If we assume that $3\overline{L} = 2q^{-1}$, which is equivalent to letting each burst type's aggregate rate be $2/5$, then the throughput of the system is $36/125 = 0.288$. It can be shown that throughput is maximized when $2\overline{L} = q^{-1}$, or when bursts are offered at aggregate rates of $1/3$. This yields a maximum throughput of $8/27 \approx 0.296$. Thus, under the above traffic assumptions, the capacity of this OBS network is limited to burst rates of less than approximately $4/27$.

The above example illustrates that OPS networks, owing to their ability to buffer in the core, have a larger capacity region than OFS networks which lack core buffering. A significant performance disparity is also observed between OFS and OBS which are physically similar architectures. This performance difference can be attributed to the benefit of scheduling over random-access. Example 3, and its generalization to larger rings, thus illustrate that the capacity inequivalence of OPS, OFS and OBS exists not just for contrived networks, but also for realistic network topologies such as rings, and for realistic routings such as shortest-path routing.

V. Dependence on number of switch ports

In Section IV, we investigated the capacity performance of OPS, OFS and OBS assuming equal switch port counts in optical packet switches and OXCs at core nodes. Specifically, we assumed that for a core node with Δ_{in} incoming fibers and Δ_{out} outgoing fibers, the optical switching device residing at the node possessed $\Delta_{in}t$ input ports and $\Delta_{out}t$ output ports. This assumption led us to the result that the capacity region of OPS dominates that of OFS, and that the capacity region of OFS dominates that of OBS.

In this section, we investigate the capacity regions of the optical network architectures as a function of the number of switch port counts available at switching devices. This is motivated by the incommensurate complexity/cost of commensurate OPS, OFS and OBS architectures. For example, the present cost of an all-optical logic gate, a building block of OPS networks, is several orders of magnitude more than the cost of an electronic logic gate, the analogous building block for OFS and OBS networks.

As before, we assume that each node generates (terminates) a maximum of t unit capacity wavelengths of traffic per outgoing (incoming) fiber. However, depending upon the architecture, we may permit a larger number of wavelengths to be carried and switched on a fiber.

A. OPS

As in Section IV-A, we assume that optical packet switches may switch up to t wavelengths of traffic per fiber. Thus, each fiber may carry a maximum of t unit capacity active wavelength channels. The capacity region of the network is therefore given by Theorem 1, with each link assumed to have capacity of t.

B. OFS

In OFS, we allow the OXCs at core nodes to switch $w \geq t$ wavelength channels per fiber. We allow a larger number of switch ports in the OFS architecture because OFS core nodes are much simpler, and thus cheaper, to build than OPS core nodes. As we see in the following example, this relaxation in the number of OXC switch ports allows for certain traffic rates to be carried on OFS networks but not on OPS networks.

Example 4 (Bottleneck)
Consider the network drawn in Figure 9, where nodes $s_1,\ldots,s_t$ each send data at rate $1-\varepsilon$ to node d_1 and nodes $s_{t+1},\ldots,s_{2t}$ each send data at rate $1-\varepsilon$ to node d_2 via the intermediate nodes i_1 and i_2. Since the offered load to the link between i_1 and i_2 is close to $2t$ when ε is very small, this traffic pattern

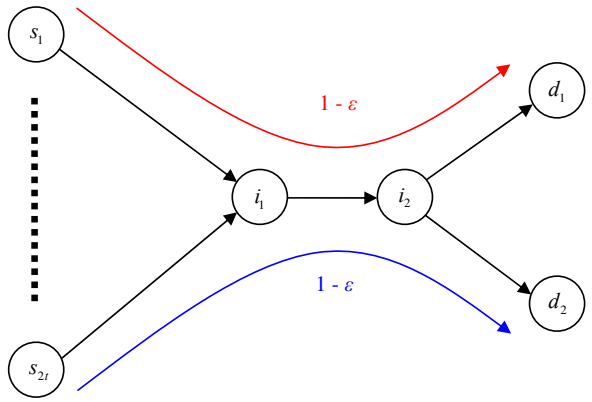


Fig. 9. Illustration of the network considered in Example 4.

cannot be serviced with an OPS architecture. However, an OFS architecture can accommodate this traffic provided that $w \geq 2t$.

In characterizing the capacity region of OFS under this relaxed assumption, we first note that a feasible network configuration corresponds to an ensemble of w-stable sets (one for each wavelength), subject to the constraint that no more than t flows per fiber may originate or terminate at a network node. We define a w-stable set of a graph as the sum of the incidence vectors of w-stable sets (some of which are possibly identical) of the graph. Then, by similar reasoning as in Section IV-B, we have the following:

Lemma 6

The capacity region of an OFS network is the convex hull of the union (over all possible routings) of the w*-stable set polytopes of the conflict graphs, subject to the constraint that the* w*-stable sets correspond to no more than* t *flows per fiber originating or terminating at a network node.*

C. OBS

By similar reasoning as in Section IV-C, Theorem 3 still holds.

The development of an approximate throughput analysis of OBS under the assumption that OBS core nodes have $w \geq t$ ports per fiber resembles that of Section IV-C. Let $B(N, i, p)$ denote the binomial probability of i successes from N trials with individual trial success probability p. Let $p = \frac{\overline{L}}{q^{-1}+\overline{L}}$ denote the probability that a burst of a particular type is attempting transmission. We denote the number of burst types passing through a link by N_p, and the number of burst types originating at a link by N_s. At a source node, it can be shown that the probability that a particular burst may begin transmission on a link is:

$$P_{s,b}(N_s, N_p) = \sum_{r=0}^{N_s+N_p-1} \left(\frac{w-1}{w}\right)^r \sum_{l=0}^{\min(t-1,r,N_s-1)} B\left(N_s - 1, l, p\right) B\left(N_p, r-l, p\right).$$

At a downstream link along the burst's path, the probability that the burst is carried is:

$$P_{s,c}(N_s, N_p) = \sum_{r=0}^{N_s+N_p-1} \left(\frac{w-1}{w}\right)^r \sum_{l=0}^{\min(t,r,N_s)} B\left(N_s, l, p\right) B\left(N_p - 1, r-l, p\right).$$

In the above two expressions, the summation index r represents the number of burst types which are transmitting on the link at that instant in time, and the summation index l represents the number of those burst types that originate at the link.

Owing to the independence of OBS links, the probability that a given burst of type j is successfully received at its destination d_j hops away is:

$$P_s(j) = P_{s,b}(N_{s,1}, N_{p,1}) \prod_{i=2}^{d_j} P_{s,c}(N_{s,i}, N_{p,i}),$$

where $N_{s,i}$ and $N_{p,i}$ are the number of burst types originating and passing through the i^{th} link of the burst's path, respectively. Using this probability of success, the average delay of a burst type, denoted D, and the throughput of the link i, denoted $S(i)$, are the same as in Section IV-C:

$$\begin{aligned} D &= \overline{L}/P_s(j) + (1/P_s(j) - 1)/q \\ S(i) &= \sum_{j=1}^{F} pP_s(j)I_i(j), \end{aligned}$$

where F is the number of types of bursts in the network, and $I_i(j)$ is the indicator function that has a value of unity if a burst of type j traverses link i.

In a manner akin to Example 3, such a throughput analysis can be used to obtain an approximate OBS capacity region.

D. On the impact of varying w with respect to t

Our last result relates the capacity regions of OPS, OFS, and OBS:

Lemma 7

1) *If* $w = t$*, then the capacity region of OPS dominates that of OFS, and the capacity region of OFS dominates that of OBS.*
2) *If* $w > t$*, then it is possible for the capacity regions of OFS and OBS to contain traffic patterns not in the capacity region of OPS. Likewise, it is possible for the capacity region of OPS to contain traffic patterns not in the capacity regions of OFS and OBS.*
3) *If* $w - t$ *is sufficiently large, then the capacity region of OFS dominates the capacity region of OPS. If there are dedicated wavelength receivers in each fiber (to avoid receiver collisions), then the capacity region of OBS is identical to that of OFS.*

Proof. (1) follows from Lemma 2 and Theorem 3.

To show (2), consider the five node ring illustrated in Figure 8 with $t = 10$ and $w = 11$. For an example of a traffic pattern that can be accommodated by OPS but not by OFS or OBS, let the five flows illustrated in Figure 8 each have rate $5 - \varepsilon$, where ε is very small. In order for this traffic pattern to be accommodated, links must carry an average of $10 - 2\varepsilon$ wavelengths of traffic. However, even with 11 wavelengths available on each fiber, this is impossible to support with an OFS architecture, and hence, an OBS architecture. On the other hand, the traffic pattern is clearly admissible, and thus serviceable with an OPS architecture. For an example of a traffic pattern that can be accommodated by OFS and OBS but not OPS, define flows of rate $5 - \varepsilon$ and $6 - \varepsilon$ between nodes 1 and 5, and nodes 2 and 4, respectively. Since $11 - 2\varepsilon$ units of traffic pass through node 3, this traffic pattern can be supported by an OFS and an OBS architecture but not by an OPS architecture (when ε is very small).

To prove (3), we show that any admissible OPS traffic pattern can be accommodated by an OFS network with a large enough number of wavelengths. When there are at least as many wavelengths as there are flow types, then the OFS network may be viewed as a large nonblocking switch, in which the set of all tunable transmitters in the network correspond to the switch's input ports and the set of all tunable receivers represent the switch's output ports. By the results of [27], there exist flow-based scheduling algorithms that are rate-stable for the set of admissible traffic rates.

If we assume that there are sufficiently many receivers per fiber at destination nodes to avoid receiver collisions, then the capacity region of OBS is identical to that of OFS. This is because, provided that we can assign each burst type its own wavelength, once a burst enters the network it is guaranteed to reach its destination without collision. If, however, we assume that receiver collisions are possible, then it is no longer true that the OBS capacity region is equivalent to the OFS capacity region. □

Lemma 7 is important as it provides an indication of the relative performance of the OPS and OFS architectures if the costs of the architectures are made comparable. When switch ports have commensurate costs in OPS and OFS core nodes, then the capacity performance of OPS dominates that of OFS. However, when switch ports in OPS core nodes are far more expensive than in OFS core nodes, then the converse is true: OFS outperforms OPS. Finally, when switch port cost in OPS core nodes is only moderately more expensive than in OPS core nodes, then it is possible that neither architecture dominates.

VI. Discussion and conclusion

In this work, we put forth a framework for comparing optical transport network architectures which is based on a notion of core network capacity as the set of exogenous traffic rates that can be stably supported under operational constraints. Using this framework, we characterized the capacity regions of OPS, OFS, and OBS. We showed that, under the assumption of an equal number of switch ports per fiber at core nodes, the capacity region of OPS dominates that of OFS, and that the capacity region of OFS dominates that of OBS. These differences in capacity performance arose because of the benefits of core buffering and scheduling.

Motivated by the incommensurate complexity/cost of commensurate transport architectures, we investigated the dependence of relative capacity performance on the number of switch ports per fiber at core nodes. When this number is significantly larger in OFS than in OPS, we found that OFS outperforms OPS. This can be attributed to the fact that OFS exploits its higher capacity network core in spite of its lack of core buffering. This is a useful result because core routers are more expensive than OXCs with the same number of ports operating at the same line rates. Finally, we showed that when the number of switch ports per fiber in core nodes is only moderately larger in OFS than in OPS, then it is possible that neither OPS nor OFS dominates.

With respect to performance, the most salient limitation of our work is the absence of a treatment of delay. A characterization of the throughput-delay trade-off in a network is important, as a network may operate significantly below its capacity in order to ensure reasonable delay. This work also neglects propagation delay. Propagation delays complicate scheduling because simultaneous data transmissions at sources do not necessarily imply simultaneous arrivals at destinations. Furthermore, nonsimultaneous data transmissions at different sources may arrive at the same destination simultaneously. A more complete performance study could lend support to the case for heterogeneous optical networks comprising more than one of the architectures considered in this work.

Finally, as mentioned in the introductory section, the goal of the network architect should be to determine which network architecture meets end user requirements with the minimum complexity/cost. While we were motivated in Section V by the importance of complexity/cost in assessing a network, our work omits a detailed consideration of the optical transport network architectures in this respect. Indeed, such a consideration would likely place OFS and OBS in a more positive light than a strict performance comparison.

Acknowledgments

The authors would like to thank Andrew Brzezinski, Devavrat Shah, and Jay Kumar Sundararajan for insightful discussions.

References

[1] G. Birkhoff, "Tres observaciones sobre el algebra lineal," *Universidad Nacional de Tucuman Revista, Serie A*, vol. 5, pp. 147–151, 1946.

[2] J. von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem," *Contributions to the Theory of Games*, vol. 2, pp. 5–12, 1953.

[3] C. S. Chang, W. J. Chen, and H. Y. Huang, "On service guarantees for input buffered crossbar switches: A capacity decomposition approach by Birkhoff and von Neumann," *Proceedings of IEEE International Workshop on Quality of Service*, pp. 79–86, 1999.

[4] I. Widjaja, I. Saniee, R. Giles, and D. Mitra, "Light core and intelligent edge for a flexible, thin-layered and cost-effective optical transport network," *IEEE Communications Magazine*, vol. 41, pp. S30–S36, 2003.

[5] A. Brzezinski and E. Modiano, "Dynamic reconfiguration and routing algorithms for IP-over-WDM networks with stochastic traffic," *Proceedings of IEEE Infocom*, 2005.

[6] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 11, pp. 1204–1216, 2000.

[7] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 782–795, 2003.

[8] B. Ganguly and V. W. S. Chan, "A scheduled approach to optical flow switching in the ONRAMP optical access network testbed," *Proceedings of the Optical Fiber Communication Conference*, pp. 215–216, 2002.

[9] I. Widjaja, "Performance analysis of burst admission-control protocols," *IEE Proceedings of Communications*, vol. 142, pp. 7–14, 1995.

[10] J. Turner, "Terabit burst switching," *Journal of High Speed Networks*, vol. 8, pp. 3–16, 1999.

[11] M. Yoo, M. Jeong, and C. Qiao, "A high speed protocol for bursty traffic in optical networks," *SPIE All-Optical Communication Systems*, vol. 3230, pp. 79–90, 1997.

[12] C. Qiao and M. Yoo, "Optical burst switching OBS – a new paradigm for an optical internet," *Journal of High Speed Networks*, vol. 8, pp. 69–84, 1999.

[13] S. Verma, H. Chaskar, and R. Ravikanth, "Optical burst switching: A viable solution for terabit IP backbone," *IEEE Network*, vol. 14, pp. 48–53, 2000.

[14] T. Battestilli and H. Perros, "An introduction to optical burst switching," *IEEE Communications Magazine*, vol. 41, pp. S10–S15, 2003.

[15] M. J. Ferguson, "A study of unslotted Aloha with arbitrary message lengths," *Proceedings of the Symposium on Data Communications*, pp. 5.20–5.25, 1975.

[16] J. H. Dshalalow, Ed., *Frontiers in Queueing*. Boca Raton, Florida: CRC Press, 1996.

[17] E. Altman and A. A. Borovkov, "On the stability of retrial queues," *Queueing Systems*, vol. 26, pp. 343–363, 1997.

[18] M. Marsan, P. Giaccone, E. Leonardi, and F. Neri, "On the stability of local scheduling policies in networks of packet switches with input queues," *Journal on Selected Areas of Communications*, vol. 21, 2003.

[19] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch (extended version)," *IEEE Transactions on Communications*, vol. 47, 1999.

[20] M. Andrews and L. Zhang, "Achieving stability in networks of input-queued switches," *IEEE/ACM Transactions on Networking*, pp. 848–857, 2003.

[21] J. K. Sundararajan, S. Deb, and M. Médard, "Extending the Birkhoff–von Neumann switching strategy to multicast switches," *Proceedings of IFIP*, 2005.

[22] C. Caramanis, M. Rosenblum, M. X. Goemans, and V. Tarokh, "Scheduling algorithms for providing flexible, rate-based, quality of service guarantees for packet-switching in Banyan networks," *Proceedings of the Conference on Information Sciences and Systems*, pp. 160–166, 2004.

[23] A. Stolyar, "Maxweight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic," *Annals of Applied Probability*, vol. 14, pp. 1–53, 2004.

[24] A. Schrijver, *Combinatorial Optimization*. Germany: Springer-Verlag, 2003.

[25] M. Andrews *et al.*, "Scheduling in a queueing system with asynchronously varying service rates," *Probability in the Engineering and Informational Sciences*, vol. 18, pp. 191–217, 2004.

[26] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, pp. 1936–1948, 1992.

[27] Y. Ganjali, A. Keshavarzian, and D. Shah, "Input queued switches: Cell switching vs. packet switching," *Proceedings of IEEE Infocom*, 2003.

[28] G. I. Falin and J. G. C. Templeton, *Retrial Queues*. Great Britain: Chapman & Hall, 1997.

[29] Z. Rosberg, H. L. Vu, M. Zukerman, and J. White, "Performance analyses of optical burst-switching networks," *Journal on Selected Areas of Communications*, vol. 21, 2003.