Limitations of Scaling Switching Capacity by Interconnecting Multiple Switches[1]

Ramu Ramamurthi
Hammerhead Systems, 1715 Marina Ct #B, San Mateo, CA 94403
rramamurthy@hammerheadsystems.com

Jean-François Labourdette
Verizon, 1095 Avenue of the Americas, NY, NY 10036
labourdette@ieee.org

Eric Bouillet
IBM Research, 19 Skyline Dr, 4S-D47, Hawthorne, NY 10532
ericbou@us.ibm.com

Abstract: Carriers are often faced with the need to scale the switching capacity of a central office site by installing multiple switches at that site. In this paper, we first provide a strong argument that multiple co-located switches should always be interconnected together because of the uncertainty in future traffic requests. However, this results in an inefficient use of a portion of the switching capacity as a number of ports are used to interconnect the switches together (the so-called "interconnect penalty"). We develop and apply a quantitative model that analyzes the interconnection capacity required to achieve certain performance criteria as a function of traffic uncertainty. We also derive bounds on the interconnection capacity required between multiple switches for different interconnection approaches and traffic patterns. The practical implication of this work is the realization of the need to eventually deploy an additional layer of higher-capacity switches of higher switching granularity in the context of optical networks to handle high bandwidth services and relieve the capacity needs of lower granularity switches.

OCIS codes: (060.0060) Fiber optics and optical communication; (060.4250) Networks

1 Introduction

We consider optical networks consisting of core sites with OEO switches interconnected by point-to-point WDM fiber links in a mesh configuration. Each WDM fiber link carries multiple wavelength channels. Multiple links are typically incident at a core site from adjacent sites. Optical switches enable re-configurable optical networking by rapidly provisioning end-to-end circuits called lightpaths between the client devices. Figure 1 illustrates a core optical network. The optical network is sparsely connected with the average number of links incident at each site around three. Some hub sites may have a higher degree of connectivity, while smaller sites may have connectivity of two. Because of the sparse connectivity, a typical lightpath travels several hops, and as a result, a *large portion* of the traffic at a backbone site is pass-through traffic, and a *small portion* of traffic is add-drop traffic. A typical node could have 70% of pass-through traffic, and 30% terminating traffic.

Carriers dimension the number of optical switches, and WDM channels on fiber links based on a demand forecast over several time periods. Initially all sites may contain a single optical switch expected to handle the traffic forecast for the initial period. As traffic at the site (both add-drop and pass-through) grows, the switching capacity at that site needs to be scaled accordingly by adding additional switches and interconnecting the switches together. Typically, the switching capacity at a site is scaled in the following fashion. First, the optical switch is scaled in-service up-to a preset threshold (say 70%) of switch capacity. Then, multiple such switches are installed at the site, and interconnected to yield a larger switching complex. The amount of intra-office interconnect is critical to the efficient operations and utilization of the network. If the switching complex at a core site blocks lightpath requests, then the network capacity may be underutilized, and circuits may be routed inefficiently. Further, upon failures, lightpath restoration may fail due to blocking at a site, even when there is sufficient protection capacity in the network.

[1] This work was performed while the authors were with Tellium.

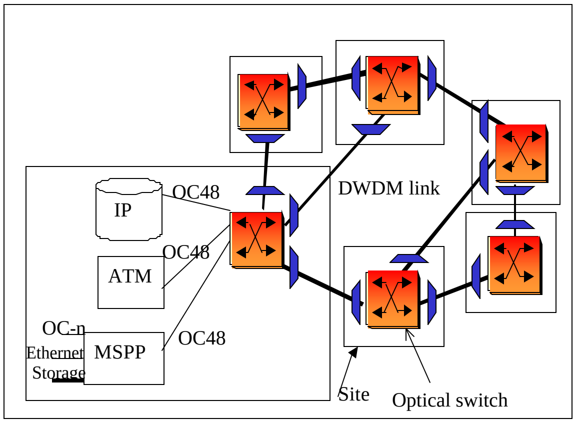

Fig. 1. Core Optical Network.

The objective of this paper is to analyze switching systems consisting of multiple interconnected switches. We investigate the need to interconnect the switches, and show that multiple co-located switches should always be connected together because of the uncertainty in future traffic requests. However, this results in an inefficient use of a portion of the switching capacity as a number of ports are used to interconnect the switches together (the so-called “interconnect penalty”). We analyze the interconnection capacity required to achieve certain performance criteria, and show as this inefficiency grows with the number of interconnected switches. We also derive bounds on the interconnection capacity required between multiple switches in a site for different interconnection approaches and traffic patterns.

Note that our work is different from classical work on inter-connection networks and switching fabrics used inside a network node or switch and built from smaller switching elements in typically regular interconnection patterns.

A practical implication of this analysis is a better understanding of the need to deploy higher-capacity switches as traffic demand scales over time, and current generation switches reach their capacity. Higher-capacity switches are easier to develop and manage when switching at higher granularity (e.g., STS-48 vs. STS-1), and that has traditionally resulted in a the deployment of higher-granularity switching layers over time (DS0, DS1, DS3/STS-1, STS-48, wavelength) [6-8].

The paper outline is as follows. Section 2 provides an illustration and analysis of why multiple switches at a site need to be interconnected. Section 3 presents the system model. Section 4 presents an analysis of the interconnection capacity required under uniform traffic conditions. Section 5 outlines some open problems, and directions for further work. Section 6 concludes this paper.

2 Site architecture with multiple switches

Carriers have always face the problem of exceeding the switching capacity of their equipment as traffic grows over time. They have usually addressed the problem by deploying multiple switches at a site. Figure 2 illustrates one possible site configuration where channels from each WDM fiber link could be terminated on a different but *single* switch. Traffic passing through the central office would necessarily consume interconnection capacity to be routed from its incoming WDM link onto its outgoing one. Since pass-through traffic is a significant proportion of the total traffic, such a configuration would be very inefficient in the way switch ports are used. Therefore, DWDM channels from each conduit must terminate on all (or a majority of) switches as illustrated in Fig. 3.

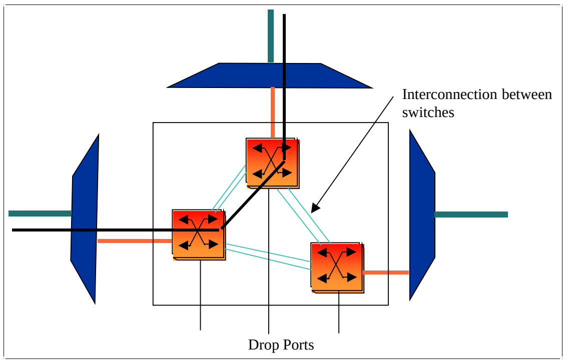

Fig. 2. Model of a site with multiple switches; channels from each WDM link are connected to a single switch.

How to terminate WDM channels in an office and to interconnect the switches together is driven by traffic forecast and network planning. Actual traffic demand is different from expected demand due to uncertainties in traffic forecast. For example, consider an unforecast originating demand from S1 routed through switch B to N1, exhausting capacity on N1.

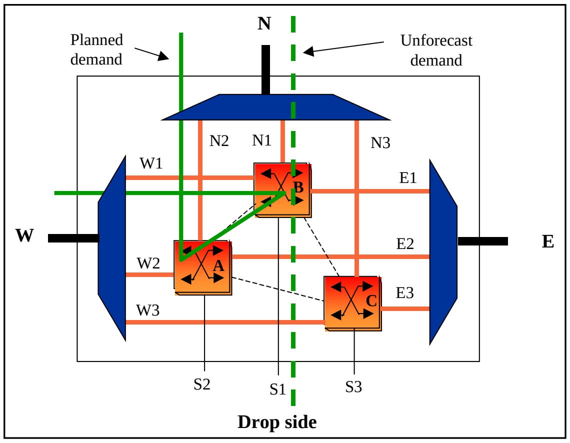

Fig. 3. Model of a site with multiple switches; channels from each WDM link are spread among switches.

Consider now that, because of similar occurrences at other sites, W2 and W3 capacity has become exhausted. With interconnect capacity between switches A and B, future planned traffic demand from **W** to **N** can still be routed from W1 to N2. The example illustrates that unless interconnections are provided between switches A and B, future planned traffic demand from **W** to **N** would block even though network capacity is available on W1 and N2.

To quantify this phenomenon, we define Q to be the fraction of the total traffic that has ingress and egress ports (including drop side ports) on the same switch, and therefore does not require any interconnection hops at a site. When $Q=1$, no interconnection is theoretically required. The assumption that all traffic will have ingress and egress on the same switch at a site is however unrealistic because:

- As illustrated above, due to the dynamism of traffic patterns, actual traffic may deviate from planned traffic patterns requiring the use of interconnection channels
- During restoration upon failures, routes for failed traffic may be determined dynamically, and such routes may require the use of interconnection channels.

As a result, the fraction of total traffic, Q, which has ingress and egress ports on the same switch is less than one in practice, creating the need for interconnections.

When a site with multiple switches blocks traffic, the network capacity is used inefficiently. This is illustrated by the following argument. Consider an H-hop path between site A and site Z, with a capacity of C channels on each intermediate link along the path. Assume that the network is dimensioned to carry S lightpaths between A and Z. Further assume that each site blocks cross-connection requests with a probability B. Each site is assumed to block independently. The probability that a lightpath request is blocked is $P = 1 - (1\text{-}B)^{H+1}$. The effective utilization of the network capacity on the *H* hop path is thus *1-P*. Figure 4 plots the network utilization against the site blocking probability for a path with hop-distance varying from one to three.

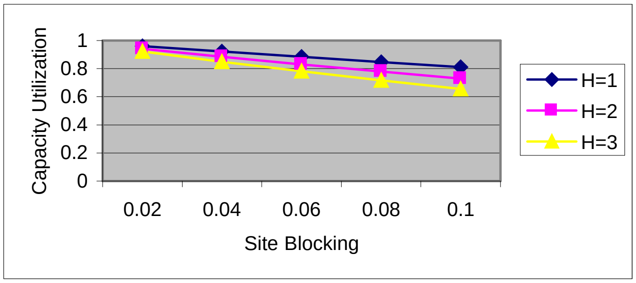

Fig. 4. Capacity utilization versus site blocking probability for lightpaths of different hop distances.

Figure 4 illustrates that for networks with lightpaths on average 3-hop long, and with site blocking probability of 0.1, the network capacity utilization is limited to 65%. It is therefore important to interconnect switches at a site so that the site does not cause, or minimizes, blocking. In the next section, we will model and analyze the intra-office interconnection of multiple switches within a site, and its relationship with traffic blocking.

3 System Model and Analysis

Each site has conduits to neighboring sites in the network. A conduit contains multiple fibers, and a fiber carries multiple WDM channels. Each WDM channel terminates at a port on a switch at the node. Sites also have client devices that connect to switches using drop ports at the switch and terminate traffic. Each switch is strictly non-blocking of size M x M, and has M unidirectional ports. As traffic grows, switches are incrementally added as illustrated in Fig. 5, and inter-connected with previously installed switches using intra-office links. Initially, we assume a simplified model of interconnection between switches, which we call the "uniform" case. There are N switches interconnected using a full-mesh interconnection topology. Each interconnection link between switches has C unidirectional channels using C unidirectional ports at each of the end-switches. The useful ports on a switch are those ports that are not used for interconnections between switches. Such useful ports can be used to connect to the DWDM systems or to client devices. In subsequent work [9], we have relaxed the assumptions of a full-mesh interconnection topology, of equal distribution of interconnection capacity on interconnection links, and of even distribution of useful ports among switches.

A traffic request is a unidirectional connection request from an ingress port to an egress port. A connection request is setup as follows: if the ingress and egress ports belong to the same switch then the connection is established within the switch. If the ports belong to different switches, the connection is established with a path between the ports, with cross-connections made at each switch on the path.

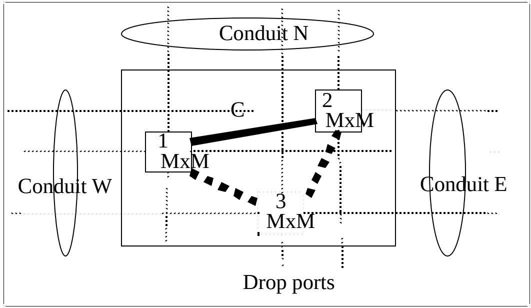

Fig. 5. Model of a site. Multiple switches are interconnected in a full-mesh interconnection topology. Switch 3 is incrementally added by interconnecting to switches 1 and 2, and by connecting its ports to the DWDM systems on the line side, and client devices on the drop side.

The offered traffic *L* represents the set of connection requests, which can be a subset of any permutation of the useful ports of the system. The system may block at most a fraction B of the offered traffic. The carried traffic, S, is the set of connection requests that are set up $S = L(1-B)$. A fraction Q of the offered traffic has ingress and egress ports on the same switch. The amount of interconnection capacity used, *X*, is the total number of interconnect channels (on inter-switch links) used to carry *S(1-Q)* connections. Given such a model of the switching system, we are interested in obtaining its switching capacity, i.e., fraction of ports that can carry traffic, and the interconnection capacity that is necessary and sufficient to carry the traffic demand. The set of parameters used to model a switching system is summarized in Table 1.

M	Number of (unidirectional) ports per switch
N	Number of switches at site
B	Blocking probability of site
L	Offered traffic: size of the set of unidirectional connection requests that are offered to a site and could be carried through the site without internal blocking at the site
S	Carried traffic: size of the set of connection requests that are carried through the site; *S = L(1-B)*
Q	Fraction of traffic for which ingress & egress ports are on the same switch (measure of forecast accuracy)
T	Multi-hop carried traffic: traffic that is carried in one or more hops over the interconnect capacity between multiple switches; *T = S(1-Q) = L(1-B)(1-Q)*
C	Number of available unidirectional channels (in each direction) on each interconnection link between switches
P	Useful ports: number of (unidirectional) ports connected to DWDM systems and client equipments [note: $P \geq 2L$]
X	Interconnection capacity required, equal to the number of unidirectional interconnection channels used to carry traffic; $X \leq C * N(N-1)$

Table 1. Parameter set used to model a switching system.

The sum of useful ports and of interconnection ports needed to carry the offered traffic is at most equal to the total number of ports across all switches: $2L + 2X \leq P + 2X \leq NM$. One can then express the number N of switches of size M required to carry an amount of offered traffic L to achieve a particular level of blocking B as follows:

$$N \geq \frac{2L}{M}\left(1 + \frac{X}{T}(1-B)(1-Q)\right) \quad (1)$$

X/T can be obtained from the appropriate equations derived later for different interconnect topology and traffic patterns. Note that *X/T* can depend on N as shown later, in which case Eq. (1) is a fixed-point equation.

4 Analysis of Interconnection Capacity – Uniform Case

In this section we assume that the switch configuration pattern is uniform, and that the traffic demand can be arbitrary in its pattern but uniform at each node.

Lemma 1: *Given N non-blocking switches interconnected in a uniform pattern, the interconnection capacity, X, needed to carry an arbitrary but uniform set of connection requests satisfies:*

$$X \geq 2T\left(\frac{N-1}{N}\right) \quad (2)$$

Proof: See appendix.

Combining Equations (1) and (2), and replacing T in Eq. (1) using *T = L(1-B)(1-Q)*, we can derive the ratio of interconnection ports to total number of ports *2X/NM* as:

$$\frac{2X}{NM} \geq \frac{2(1-B)(1-Q)\left(\frac{N-1}{N}\right)}{1 + 2(1-B)(1-Q)\left(\frac{N-1}{N}\right)} \quad (3)$$

Figure 6 illustrates the number of interconnection ports required (as a fraction of the total ports at the site) as a function of Q, for different values of N. Here we assume that B=0.

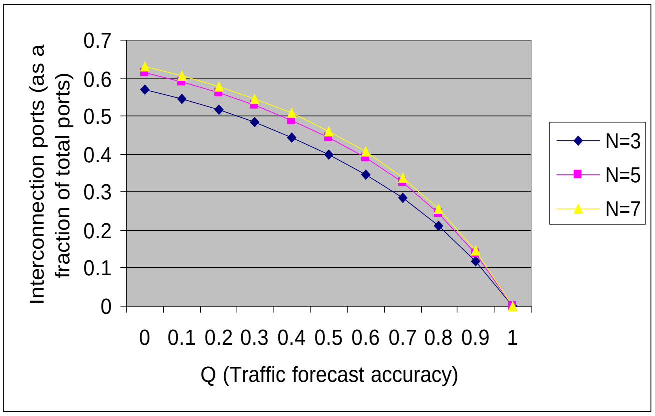

Fig. 6. Interconnection ports (as a fraction of the total ports) as a function of Q for different values of the number of switches N.

In the extreme case when *Q=0*, and *B=0*, i.e., for completely arbitrary but uniform traffic patterns, and no blocking, the above result indicates that approximately 2/3 of the ports at a switch must be dedicated for interconnections with other switches. When *Q=0.7*, i.e., 70% of traffic does not need interconnection hops, about 30% of the ports need to be interconnection ports. The number of useful ports is at most a third of the total ports at the site. Figure 7 illustrates the number of interconnection ports required as a function of Q when N=5, and the blocking rate of the site varies from 0 to 0.5. At a blocking rate of 20%, and when Q = 70%, about 27% of the ports are interconnection ports.

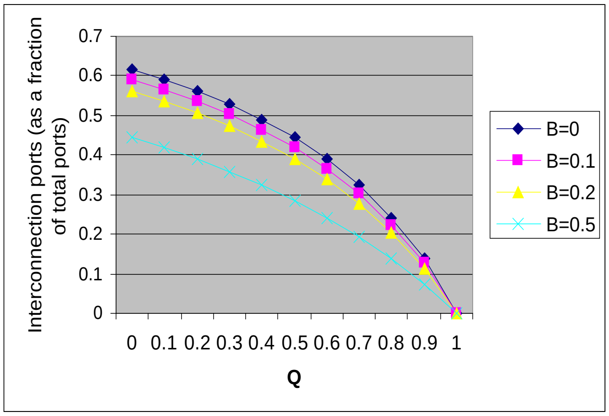

Fig. 7. Interconnection ports (as a fraction of the total ports) as a function of Q for different values of blocking B.

Figure 8 plots the switching capacity (number of ports that can carry traffic) as a function of the number N of switches at a site, when Q=0, and for different levels of blocking B. With no blocking, the switching capacity (i.e., the number of ports that can carry traffic) of three interconnected switches is 1.28 times that of a single switch. With a blocking of 0.1 the switching capacity of three interconnected switches is 1.36 times that of a single switch.

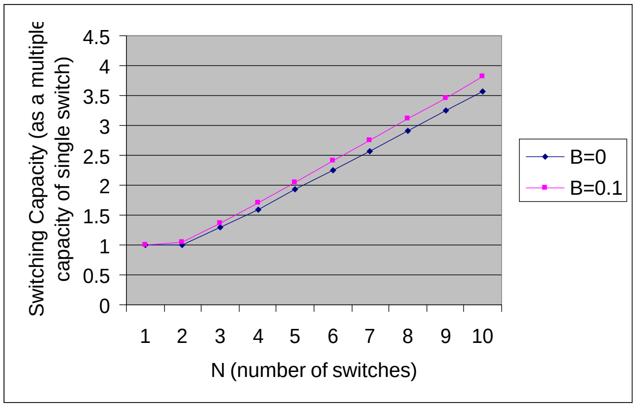

Fig. 8. Switching capacity of site (expressed as a multiple of switching capacity of single switch) as a function of the number of interconnected switches for different levels of blocking B.

5 Open Problems and Future Work

The problems considered in this paper and the results can be extended in several directions. Algorithms to decide when and how to incrementally add switches at a site deserves further study. A study of sufficient conditions on interconnection capacity (including algorithms to route connections) for the different switching systems for both unicast and multicast connections are a fertile area of research. The results in this paper have implications on how to architect and build networks. In particular, they indicate that multi-layered networks can avoid the interconnection penalty of a single-layer network when switch size is limited [6]. This will be the subject of future research.

6 Conclusion

We have investigated the problem of whether to interconnect multiple switches at the same site, and if so, how many interconnection ports are required. We find that the need to interconnect switches at the same site arises from the need to accommodate uncertain traffic patterns without blocking. A blocking site reduces the capacity utilization of the network with the reduction in capacity utilization a function of the length in hops of lightpaths. Assuming that the switches are interconnected in a full-mesh interconnection topology, we determine a relationship between the interconnection ports required as a function of the number of switches, and the blocking. We find that, when 70% of traffic have ingress and egress on the same switch, about 30% of ports need to be used for interconnecting switches. For arbitrary traffic, we have showed [9] that the interconnection capacity must be at least (about) twice the carried traffic, and this interconnection capacity is also sufficient for re-arrangeable switching. We also showed [9] that the full-mesh interconnection topology is the best possible among all regular interconnection topologies.

7 References

[1] T.E. Stern and K. Bala, “Multi-wavelength Optical Networks: A Layered Approach”, Reading, MA: Addison Wesley, 1999.
[2] J. Y. Hui, “Switching and Traffic Theory for Integrated Broadband Networks”, Kluwer Academic Publishers, 1990.
[3] G. Chartrand and L. Lesniak, “Graphs and Digraphs”, Third Edition, Chapman and Hall, 1996.
[4] J-F. Labourdette et al., “Routing Strategies for Capacity-Efficient and Fast-Restorable Mesh Optical Networks”, special issue of Photonic Network Communications on "Routing, Protection, and Restoration Strategies and Algorithms for WDM Optical networks"., 4:3/4, 2002.
[5] G. Ellinas et al., "Routing and Restoration Architectures in Mesh Optical Networks", Optical Networks Magazine, Jan/Feb 2003.
[6] C. Olszewski et al,. “Two-Tier Network Economics”, NFOEC 2002, Sept. 2002, Dallas, TX.
[7] S. French et al., “Efficient Network Switching Hierarchy“, NFOEC 2002, Sept. 2002, Dallas, TX.
[8] J-F. Labourdette et al., “Layered Architecture for Scalability in Core Mesh Optical Networks“, NFOEC 2003, Orlando, FL, Sept 2003.
[9] R. Ramamurthy, J.-F. Labourdette, Eric Bouillet, Technical Report on Scaling Switching Capacity for Non-Uniform Interconnection.

8 Appendix

Proof of Lemma 1: It is easy to show that the interconnection capacity X must be at least as much as the multi-hop carried traffic T, $X \geq T$. This results from the fact that there is a multi-hop carried traffic demand for which each traffic request must take at least one hop across the interconnection network. For example, consider a permutation of the switches such as (1->2, 2->3,…, N->1). For each element of the permutation, e.g., i->j, and for each useful port at i, select a useful port at j, for a total of T/N traffic units from i to j. Each traffic request has ingress and egress ports on different nodes, and must ride at least one hop over the interconnection capacity between the multiple switches, thus $X \geq T$. We can show a stricter lower bound on X as follows: Consider the same permutation traffic pattern as above. Consider the traffic from node A to node B as shown in Fig. 9.

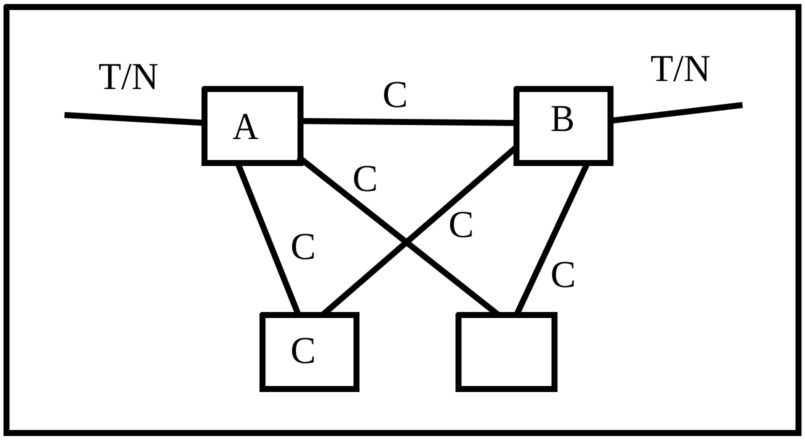

Fig. 9. Traffic from A to B. At most C traffic requests go over the direct one-hop path, the rest go over at least 2 hops.

At most C traffic requests can be carried on the direct one-hop link between A and B, and the rest of the traffic requests must be carried on paths that have two or more hops. Therefore, the interconnection capacity consumed by the traffic from A to B is at least $C + 2(T / N - C)$, assuming $T / N \geq C$. Since there are N elements in the permutation traffic pattern, $X \geq N\big(C + 2\big(T / N - C\big)\big)$. By definition, X can be at most $X = C * N(N - 1)$, and therefore $X \geq 2T(N-1)/N$. **QED**.