(Invited)

AN AGILE ALL-PHOTONIC NETWORK

Trevor J. Hall, Sofia A. Paredes, Gregor v. Bochmann

Photonic Network Technology Laboratory, Centre for Research in Photonics
School of Information Technology and Engineering, University of Ottawa, Canada
{thall , sparedes , bochmann}@site.uottawa.ca

ABSTRACT

This paper presents an overview of recent and current work being conducted in the "Agile All-Photonic Networks", AAPN, Research Network. An AAPN is a wavelength division multiplexed network that consists of several overlaid stars formed by edge nodes that aggregate traffic, interconnected by bufferless optical core nodes that perform fast switching in order to provide bandwidth allocation in sub-wavelength granularities. The architectures, tools and methods being developed for its operation are described, as well as the issues to be solved.

Keywords: *optical networks, WDM, TDM, OBS, traffic aggregation, bandwidth allocation, scheduling.*

1. INTRODUCTION

Most of the existing photonic networks have a mesh topology, which distributes the traffic load over many switches but warrants the use of complex routing algorithms and possibly a large number of O-E-O conversions. If the switches in the core of a photonic network have enormous capacity; an overlaid star topology is possible while maintaining robustness [1]. Complex routing is not necessary in this kind of network and there is no need to resort to technologies that are not realizable in the near future, such as optical memory and optical header recognition.

The Agile All-Photonic Networks, AAPN, research project [2] proposes such a star topology in which the photonic core space switches are all-optical and rapidly reconfigurable. This proposal is based on the observation that optical switching technologies will mature to a level where they will be able to introduce "agility" [3]; i.e., the ability to perform time domain multiplexing to dynamically allocate and share the bandwidth of each wavelength by several information flows to achieve a higher degree of data flow granularity and be able to adapt it to traffic flows as the demand varies.

AAPN envisions the use of fast switches in the core nodes to provide the agility and capacity required; as opposed to current photonic networks that are relatively static in their configuration and whose components have the luxury of relaxed requirements for switching times, settling times and clock acquisition times.

Ideally, switching would be performed on a per-packet basis to achieve the finest granularity and hence the most efficient bandwidth utilization. However, this would require switching times in the order of nanoseconds, which is not a viable technology yet. With this in mind, the constraints on switching speed are relaxed to some extent by gathering variable sized packets destined for the same output to form fixed-length slots. Switching in the core node is then performed for slots and not for individual packets.

Schemes for provision of subwavelength bandwidth granularity have been developed, for example in [4], where a time slotted WDM mesh network is described. This approach differs from AAPN in the use of ultra-fast tunable lasers at the edge nodes to avoid fast switching at the network core. In [5], a meshed WDM network is partitioned into a number of clusters. Specific nodes that serve as gateways between clusters undertake the coordination of frame switching and end-to-end routing. The core network formed by the gateways is in this case, as opposed to AAPN, a mesh.

2. ARCHITECTURE

The AAPN consists of a number of hybrid photonic/ electronic edge nodes connected together via a wavelength stack of bufferless transparent photonic switches placed at the core nodes (a set of space switches, one switch for each wavelength), of which there is a small number.

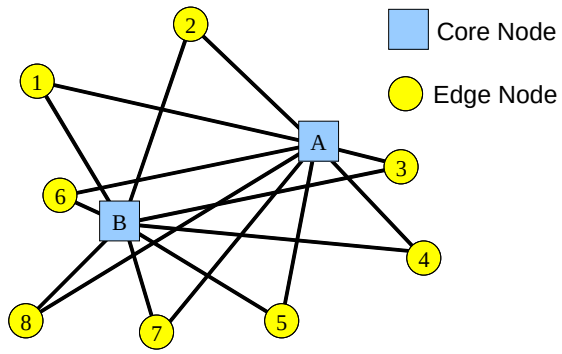

Figure 1. Overlaid star topology that characterises an Agile All-Photonic Network

Each edge node contains a separate buffer for the traffic destined to each of the other edge nodes. Traffic aggregation is performed in these buffers, where packets are collected together in slots or bursts that are then transmitted as single units across the network.

The connectivity of a core node is reconfigured in response to traffic load variations as reported by the edge nodes. The core node also coordinates the transmission

actions in the edge nodes.

Since switch fabrics with large port counts cannot provide the switching speed required at the core, port sharing is required to allow a core node to support large numbers of edge nodes. A *selector* may therefore be used between the edges and the core (Figure 2) to combine the traffic of multiple edges onto a single fibre.

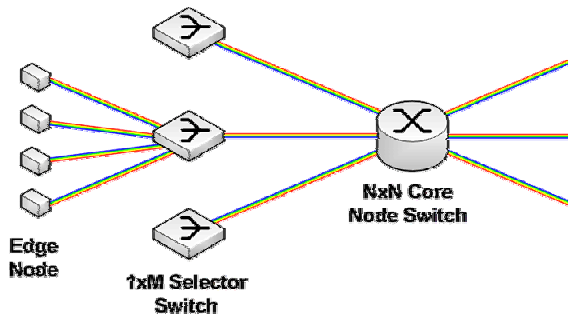

Figure 2. Use of selectors to allow for large numbers of edge nodes.

Another benefit of using a star topology is that, in the context of a single star, the routing of packets is trivial and therefore may be left to higher level protocols, e.g. MPLS; in which case the whole star may be viewed as a single MPLS switch and the same labels may be used.

2.1 Performance objectives / assumptions

The topology of AAPNs may include hundreds of edge nodes operating at 10Gbps (and even up to 100 Gbps in future upgrades). The optical switching time in the core node is aimed at 1μs. Real time traffic monitoring will be implemented for the correct adaptation of the core nodes connectivity to the most up-to-date traffic needs. For the slotted mode, the size of the slot is 10μs for links with rates of 10Gbps and the number of timeslots per frame is 100 (and up to 1000).

The photonic path lengths will be of up to thousands of kilometers; therefore all-optical 3R (as well as dynamic compensation for transmission impairments) will be required for long-haul. Wavelength conversion will be avoided in principle; although this may change as the technology matures. C and L band wavelengths with 50 to 100 GHz channel spacing will be used.

2.2 Issues

2.2.1 Switch speed and Scalability

A major concern in choosing a technology and architectural option for the switch fabric in the core is its scalability to large port counts, which is limited by two main performance degradations: insertion loss and crosstalk. The former can be compensated all optically, but preventing the accumulation of crosstalk requires signal regeneration or O-E-O conversion. It is therefore necessary to estimate the crosstalk limit that has to be achieved for the signal to remain in the optical domain.

The major performance impairment in optical switches, the in-band optical crosstalk, was analysed using a statistical crosstalk model for three architectural and technological options for the implementation of AAPN core switches [6]. The study shows that for DC switches (wavelength layered), single switch ports up to 64 can be realised with acceptable crosstalk and port counts can be scaled up using three-stage switches. For larger number of wavelengths and fibre counts, wavelength dimensional switches are more efficient.

2.2.2 Synchronization

Given the physical topology it is sufficient for the propagation delays between each edge node and the core node to be known and for each edge node to maintain a local clock locked to the core node clock. The locking of the clocks and the determination of the propagation times may be done using a suitable synchronization signaling protocol. Co-ordination of transmissions at the core node is achieved by scheduling transmissions at the edge nodes using the local clock and the known propagation delays.

2.3 Determining the Optimum Layout

The initial step for deployment of an AAPN is solving the layout design problem to determine network parameters such as optimal number, size, and placement of edge nodes, selectors/multiplexers, core nodes as well as placement of the DWDM links; the aim being to minimize network costs while satisfying performance requirements. The solution to this problem is determined by demographic and economic factors.

A mixed integer linear programming formulation is presented in [7] for core node placement and link connectivity. A number of possible solutions and their costs are discussed for a wide variety of equipment cost assumptions for both metropolitan and long-haul networks with a gravity model for traffic distribution. The model was solved also for actual population information and geographical coordinates that were obtained from a census database of 140 Canadian cites.

3. BANDWIDTH SHARING

3.1 Optical Burst Switching (OBS-AAPN)

Optical Burst Switching (OBS) is a technique where several packets with the same destination and other common attributes such as Quality of Service parameters are assembled into a “burst” (essentially a very large packet) and forwarded through a bufferless network as one entity. The header and the associated payload are sent on two different wavelength channels with the header being sent ahead in time.

3.1.1 Burst aggregation

Burst aggregation in an OBS network can be timer-based or threshold-based. In a timer-based approach, a burst is created and sent into the optical network at periodic time intervals; which produces variable length bursts and therefore might yield undesirable burst lengths at different loads. In a threshold-based approach, a limit is placed on the length of each burst; which produces

fixed-size bursts but does not give any guarantee on the delay that the packets will experience during the aggregation process.

A composite burst assembly mechanism that combines both approaches in AAPN is discussed in [8]. Results with both analytical and simulation models show that the delay experienced by the packets can be appropriately bound by choosing a time-out comparable to the maximum tolerable delay and that only the packets in the low rate flow suffer this delay. It is also shown that the hybrid mechanism reserves resources for bursts for smaller periods of time, therefore reducing blocking probabilities or reservation delays.

3.1.2 Scheduling

OBS may or may not use a two-way reservation of resources. The use of acknowledgements eliminates burst loss but increases overall delays. Simple two-way reservation approaches have been shown to work with acceptable delays for short core-edge distances.

Some modifications of retransmission schemes are shown in [9] for OBS-AAPN, where a framework without any loss of bursts exploits the entire network capacity. The results indicate that for high loads the average burst transmission delay and the buffer capacity are highly dependent on the network diameter and thus confirm that these schemes are suitable for the local or metropolitan cases.

3.2 Time Division Multiplexing (TDM-AAPN)

In a time-slotted mode of operation, each star of the AAPN may be seen as a distributed three-stage Clos-like packet switch: the edge nodes can be logically split in two parts (the source and the destination modules) and the core node may be explicitely drawn with its de-multiplexers/multiplexers and its wavelength space switches in parallel. The connections between the respective source/destination edge nodes and the core node are seen now as unidirectional (Figure 3).

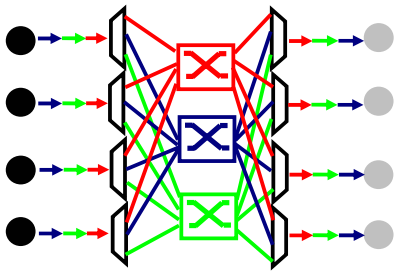

Figure 3. An AAPN star viewed as a distributed three- stage Clos switch.

This architecture is esentially the same as the one presented in [10] (and others referenced therein) and hence the same *Flexible Bandwidth Provision* scheduling algorithm may be used. Timing considerations must be made in this case: since all wavelengths and ports in the core node start a new transmission slot at the same time and the edge nodes may be located at varying distances from the core, scheduling in this modality requires synchronization between the core and the edge nodes to allow the edge nodes to know the time when a new slot should be transmitted to the core.

3.2.1 Slot aggregation

The same hybrid mechanism for burst aggregation can be extended to encapsulate variable length packets into '*envelopes*' matched to the time slots in a TDM-AAPN. Emulation results are presented in [8] for this process using real IP network traffic from a local LAN using encapsulation methods with and without packet segmentation. Bandwidth utilization measures confirm that the model with packet segmentation is more bandwidth-efficient (even if the processing delay is slightly larger) and the simplicity of the model suggests that a low cost software implementation of this process would be efficient.

3.2.2 Scheduling

Typically, TDM scheduling involves the following steps: traffic matrix estimation (bandwidth request), service matrix construction (bandwidth allocation), and decomposition into configuration matrices (connectivity for the core switches).

Traffic estimation is mainly done using queue state information gathered at the edge nodes. If the propagation delays are large and the traffic information is considered out-of-date; it is possible to use algorithms for prediction of traffic loads like the one described in [11], which uses an approach based on traffic sampling and distributed expectation-maximization for predicting the resource requirements of end-to-end flows.

The service matrix defines the bandwidth allocated to each flow in units of slots within a frame; i.e., the larger the number of slots, the larger the bandwidth granted to that flow. The construction of this matrix, can be made using many mathematical approaches or virtually any heuristic approach (or a mix of both) [12][13].

The service matrix, must be decomposed into its correspondent constituents, which will define the connectivity (schedule) of each space switch for each time slot. There are also a number of mathematical and heuristic approaches to solve this problem [10][13].

Ideally, the core switches are scheduled on a per-slot basis, but the bandwidth requests may be "out-of-date" because of signaling delays. For long-haul scenarios with large propagation delays it may therefore be a better option to take a frame-based approach and reconfigure the central nodes every frame instead of every slot.

A comparison of various OBS and TDM methods for bandwidth sharing in AAPN is reported in [14].

4. PROTOTYPE

A demonstrator is currently being built with the aim of showing that the technologies, architectures and control protocols can be combined into an operational network. The AAPN prototype will be the testbed for synchronization protocols, bandwidth allocation methods, traffic monitoring, routing protocols and fault

recovery methods. This work is currently in its first phase, in which the transmission platforms will be built with off-the-shelf components. Figure 4 shows the preliminary design of the edge node modules and their interactions. Subsequent work in this area will involve the incorporation of newly designed optical devices and components developed also by the AAPN team [2].

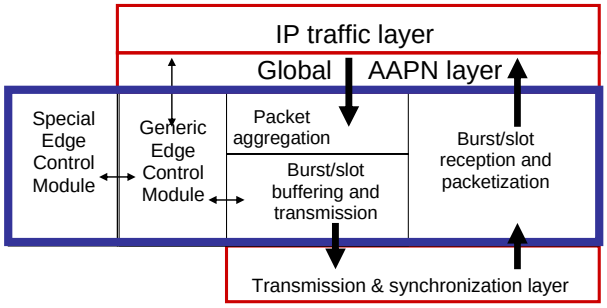

Figure 4. Edge node architecture for the prototype.

5. FURTHER ON

The results presented in [7] correspond to the deployment of new infrastructure for the realisation of the AAPN. This is, however, a costly alternative and therefore it is of great interest to study how an AAPN can be built by using and adapting the already deployed WDM technology. Migration strategies is therefore an important ongoing research topic.

In the context of a TDM-AAPN, if the load is perfectly balanced, then each space switch in the core node would have the same schedule. In this case, a single wavelength-independent crossbar could be used in the core node to switch the wavelength multiplex as a whole, thus simplifying the scheduling problem significantly. This is called *Photonic Slot Routing* [15] and its application to AAPN, is currently under investigation.

6. SUMMARY AND REMARKS

The architectures, tools and methods so far designed for the correct and efficient operation of an Agile All-Photonic Network have been described. Topology dimensioning, traffic aggregation, bandwidth allocation, scheduling methods and prototyping have been discussed. The proposed overlaid star architecture provides high bandwidth connectivity in sub-wavelength granularities by dynamically reconfiguring the photonic network according to the traffic demands.

It is important to note that a large amount of work in the development of enabling technologies (optical switches, transmission and amplification) is being carried out by several investigators of the AAPN Research Network but these topics are out of the scope of this paper and have not been discussed. The authors refer the reader to the AAPN website [2] for information.

7. ACKNOWLEDGEMENTS

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners, through the Agile All-Photonic Networks (AAPN) Research Network. Dr. Trevor Hall holds a Canada Research Chair in Photonic Network Technology at the University of Ottawa. He and Sofia Paredes are grateful to the Canada Research Chairs Programme for their support of this work.

8. REFERENCES

[1] R. Vickers and M. Beshai, “PetaWeb architecture,” 9th Int Telecom. Netw. Planning Symp., Toronto, Canada, 2000.

[2] The Agile All-Photonic Networks (AAPN) Research Network, http://www.aapn.mcgill.ca/, 2003-2005.

[3] G. v. Bochmann, et al. “The Agile All-Photonic Network: An architectural outline”, in Proc. 22nd Bien. Symp. on Comm., Kingston, Canada, 2004, pp. 217-218.

[4] I. Saniee and I. Widjaja, “A New Optical Network Architecture that Exploits Joint Time and Wavelength Interleaving,” in Proc. OFC, Los Angeles, 2004.

[5] A. Stavdas, H. Leligou, K. Kanonakis, C. Linardakis, and J.D. Angelopoulos, "Scheme for performing statistical multiplexing in the optical layer," J. Opt. Netw., 4(5), 237-247, 2005.

[6] R. Shankar and T.J. Hall, “Core Switch Architectures for the Agile All-Photonic Network: Performance Evaluation with Crosstalk”, Indicon 2005, Chennai, India, 2005.

[7] L.G. Mason, A. Vinokurov, N. Zhao and D. Plant, “Topological design and dimensioning of Agile All-Photonic Networks”, J. of Comp. Comm., Special issue on Optical Networking, 2005 (accepted).

[8] S. Parveen, R. Radziwilovicz, S.A. Paredes, T.J. Hall. "Evaluation of burst aggregation methods in an optical burst switched agile all-photonic network". SPIE Photonics North 2005, Toronto, 2005.

[9] A. Agustí-Torra, G. v. Bochmann and C. Cervelló-Pastor, “Retransmission schemes for Optical Burst Switching over star networks”, IFIP Int. Conf. Wireless Opt. Commun. Netw., 2004.

[10] S.A. Paredes and T.J. Hall, “Flexible bandwidth provision and scheduling in a packet switch with an optical core”, J. of Opt. Netw., 4(5), 260-270, 2005.

[11] T. Ahmed, N. Saberi, M.J. Coates, “Time-slot reservation in all-photonic networks based on flow prediction”, 2nd Workshop on Optim. of Opt. Netw., Montreal, 2005.

[12] N. Saberi and M.J. Coates, “Bandwidth Reservation in Optical WDM/TDM Star Networks”, in Proc. 22nd Bienn. Symp. on Comm, Kingston, 2004, pp. 219-221.

[13] C. Peng, S.A. Paredes, G. v. Bochmann and T.J. Hall, “Bandwidth Provisioning in the Core of an Agile All-Photonic Network”, submitted to Elsevier’s Optical Switching and Networking, 2005.

[14] X. Liu, A. Vinokurov and L.G. Mason, “Performance Comparison of OTDM and OBS Scheduling for Agile All-Photonic Networks”, Proc. IFIP MAN Conference HCNC, Vietnam, 2005.

[15] H. Zang, J.P. Jue and B. Mukherjee, “Capacity Allocation and Contention Resolution in a Photonic Slot Routing All-Optical WDM Mesh Network,” J. of Lightw. Tech., 18(12), 1728-1741, 2000.