Scalable Control Plane architecture for Optical Flow Switched Networks

Vincent W.S. Chan, Joan and Irwin Jacobs Professor, Fellow IEEE, OSA
Zhang Lei, Student Member IEEE
Claude E. Shannon Communication and Network Group
Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Email: chan@mit.edu, zhl@mit.edu

Keywords: *Optical networks, network management and control*

Summary

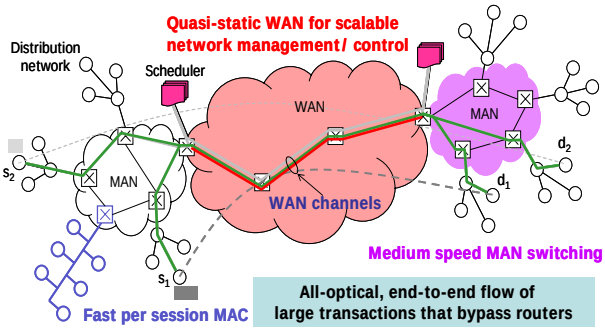

Fig. 1. OFS with transparent, end-to-end data flow between users.

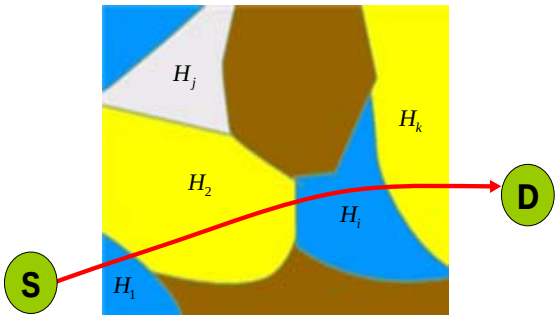

Fig. 2. Route selection based on minimized average path entropies.

In OFS, users employ an off-band signaling protocol to request lightpaths for their large unscheduled bursty transactions, and the network dynamically schedules a dedicated, end-to-end lightpath for the duration (≥100mS transaction times) of the transfer avoiding collisions due to contention, Fig.1. For OFS the scheduling and network control is dynamic and when a transaction finishes, the network resources are immediately relinquished to other users. The key to high utilization of backbone wavelength channels – a precious network resource owing to the necessary use of optical amplifiers and dispersion management – is statistical multiplexing of large flows from many users in a scheduled fashion. Efficient, dynamically assigned multi-access broadcast groups can be arranged for multiple transaction durations using the node architecture. We have shown, [1,2,3,4,6], that our approach provides significant cost savings over other transport mechanism for large transactions. OFS is technically a circuit switched service. The major difference between OFS and traditional circuit switching is the fast dynamic session setup (<100mS). This presents tremendous stresses on the network control plane. Many detractors of this technology (who mostly favors IP packet switching) question the scalability of this technology. The purpose of this paper is to address all the critical problems that prevent affordable scalability and fast dynamic setup for OFS. Major attributes in our architecture that simplify network control and potentially can realize significant simplicity and cost saving are:

1. Use of broadcast/narrowcast all-optical access network with a MAC protocol for efficient statistical multiplexing of bursty large transactions bypassing routers. This avoids fast (per session switching in the LAN and MAN) eliminating stringent hardware requirements on fast optical switching and control network signaling and decision making. In this scheme the only fast per flow hardware tuning is the transmitter/receiver wavelength of operation.

2. Use of efficient optically switched MAN mesh network topologies that minimize switching and amplifier resources to organize access to the precious WAN transport. The MAN architecture supports efficient aggregation of flow traffic without fine grain switching and only reconfigures the hardware in response to medium time scale average load changes (>100S). We use a MAN network architecture that exhibits a decreasing cost/user/data-rate as the number of users and user data-rates increase. The following

are key design features for the MAN architecture:

a. Minimize average lightpath lengths, in terms of the number of OXCs traversed.
b. Use of quasi-static MAN/LAN broadcast groups to eliminate fast per session, <100mS, network reconfigurations and only use MAC for efficient statistical multiplexing.
c. The MAN/WAN interface is quasi-static and the WAN wavelength highways are dedicated exclusively to MAN source/destination pairs essentially decoupling the numerous MAN source/destination pairs' MAC protocol and control.

3. Use of efficient scheduling algorithms for the contention of WAN resources. Quasi-static WAN "highway" wavelength provisioning is used to slow down control plane traffic and computation loads for reconfigurations. This slowing down of the WAN control plane is important to perform near optimum global coordination of the WAN without having to perform per flow signaling and control and making the architecture unattainable.

4. Scalable ultra-fast setup using entropy function as state information reducing the complexity of network sensing, management and control.

In a typical network setting, there exists widely varying QoS requirements for data: some sessions require setup times of no longer than 100 ms, while other sessions can tolerate setup times on the order of several seconds. For the sessions requiring sub-second setup times, routing and wavelength assignment must be completed immediately. Though centralized approaches yield network configurations at least as good as those of distributed schemes, the propagation and computation times involved, in view of the stringent session setup requirements, will be barely feasible, if at all and certainly not scalable. We thus use a hybrid centralized (slow processes) / distributed (fast processes) scheme relying on up-to-date local (and slightly stale global) information to setup these sessions for a group of special users over a virtual overlay subnet of the optical network, [6,7]. We use a distributed approach in which the source node of the desired session sends out multiple pre-computed (centralized) path[1] requests (lightpath probes) to the nodes residing on these multiple paths. These path requests may contain priorities based upon path lengths and slightly stale global network information. If the network states are updated and broadcast on time scales less than the shortest session durations, then the state information is likely to be very accurate. This sets the requirement for the control network data rate and time elapsed between network state broadcasts[2]. Immediately after forwarding an ACK, a path node temporarily reserves the relevant resources in case the source/destination nodes choose to use these resources for the session[3]. In the event that multiple ACKs have arrived at the destination node, each corresponding to different lightpaths, the destination node decides which path to use. Upon making this decision, the destination node notifies the source node of the chosen path, and the source node begins data transmission immediately thereafter. Simultaneously, the destination node sends release messages along all lightpaths that have ACKed but will not be used for data transmission, so that temporarily held network resources may be released for other purposes. The centralized management system is notified so that it will refresh its lists of available resources in the next update. Lightpaths can be set up in as little time as one roundtrip time plus hardware reset times (<50 ms). The information at the beginning of the interval is accurate and yields the lowest blocking probabilities. Whereas, towards the end of the interval between updates, new traffic may have joined the network and result in higher blocking, Fig. 3. The update interval should be ~ 0.5 of the expected transaction service time. Updating the network state information on a per flow basis is still complicated and will stand in the way of a scalable and cost-efficient control plane. We propose to use a network sampled entropy function idea to further significantly simplified the control plane messaging and at the same time yield near optimum performance, Fig. 2. The network control plane globally reports to users and schedulers the degree of occupancy and time dynamics of a network region by using a single scalar quantity, the entropy. The key decision for the MAC protocol to make is the number of paths to probe. The algorithm features:

---

[1] *These paths are precomputed in an offline, centralized fashion and only have to change slowly based upon long-term traffic changes.*

[2] *Moreover, these aggressive requirements also dictate the need for active probing of unused network resources to ensure that they are functional, unless one is resigned to wasteful resource allocation for diversity routing in case of undetected failures.*

[3] *In the event that a session has to be k-protected, it is expected that tying up k-1 lightpaths for 100 ms will not substantially waste resources (at most the average session duration divided by the setup time (<100 ms).*

i. Probe K_p paths with the least entropies

ii. *Determine the number of paths to be probed assuming the **worst** case traffic distribution that is consistence with the sampled entropy.*

iii. Avoid high entropy (hot) regions: pick routes that minimize $\sum H_i$, the sum of the entropies along the routes between source and destination, as shown in Figure 2.

iv. Update intervals are determined from the entropy function and desired blocking probability

This entropy function will be path dependent and due to traffic congestion in some links may suggest congestions in nearby links, this function may also be regionally correlated. We will assess how fine grain and how accurate this entropy function must be for good network performance. Note, the algorithm does not assume any detail traffic statistical models and uses the worst possible distribution for the estimates and thus is very robust with respect to modeling errors. Fig 4. shows the number of paths required to probe for different loading, time dynamics and correlation of lightpaths due to session traversing multiple links. The use of a single scalar to report network state greatly simplifies network management and control but only sacrifice network performance slightly.

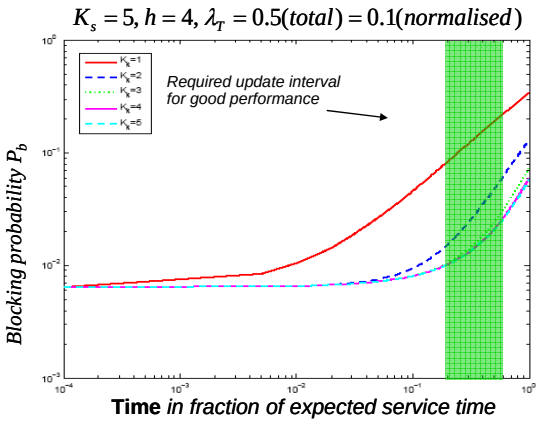

Fig. 3. Required network state update interval for very fast service.

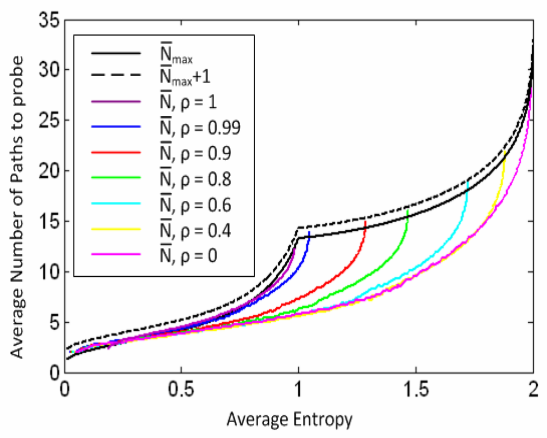

Fig. 4. Number of paths to probe vs entropy of underlying network regions and link correlation ρ.

In our presentation, we will describe in detail an OFS architecture that exploits the strengths of optics to serve large transactions. It will enable orders of magnitude of cost and power reductions. The shift towards OFS requires some architectural elements of the network – from the physical layer to the higher network layers, as well as network management and control, be substantially redesigned at the fundamental level for the network architecture to be scalable and implementable.

References

[1] Guy Weichenberg, Vincent Chan and Muriel Medard, "Performance Analysis of Optical Flow Switching", IEEE/ICC Dresden Germany June 2009.

[2] Guy Weichenberg and Vincent Chan, "Design and Analysis of Optically Flow Switched Networks," IEEE/OSA Journal on Optical Communications and Networking, August, 2009.

[3] Kyle Guan and Vincent Chan, "Cost-Efficient Fiber Connection Topology Design for Metropolitan Area WDM Networks," IEEE/OSA Journal on Optical Communications and Networking, June 2009.

[4] V.W.S. Chan, "Editorial: Optical Network Architecture from the Point of View of the End User and Cost," IEEE Journal on Selected Areas in Communications, Optical Communications and Networking, Volume 24, Issue 12, pp. 1-2, December 2006.

[5] Anurupa Ganguly, G. Weichenberg, and Vincent Chan, "Optical Flow Switching with Time Deadlines for High-Performance Applications," IEEE Globecom 2009, Honolulu Hawaii, Dec, 2009.

[6] G. Weichenberg and V.W.S. Chan, "Access Network Design for Optical Flow Switching," Proceedings of IEEE Global Telecommunications Conference (Globecom 2007), Washington, D.C., Nov. 2007.

[7] Y.G. Wen, Vincent Chan and L. Z. Zheng, "Efficient Fault Diagnosis Algorithms for All-Optical WDM Networks with Probabilistic Link Failures (invited paper)," IEEE/OSA Journal of Lightwave Technology, October 2005

[8] Lei Zhang and Vincent Chan, "Fast Scheduling for Optical Flow Switching," Globecom 2010, Miami Dec 2010.