Optical IP Switching a Solution to Dynamic Lightpath Establishment in Disaggregated Network Architectures

Gavin Mulvihill, Marco Ruffini, Frank Smyth*, Liam Barry*, Linda Doyle and Donal O'Mahony
Centre for Telecommunications Value-chain Research, Trinity College, Dublin 2, Ireland
** Centre for Telecommunications Value-chain Research, Dublin City University, Dublin 9, Ireland*
Tel: (353) 1 608 8441, Fax: (353) 1 667 2204, e-mail: gavin.mulvihill@cs.tcd.ie

ABSTRACT
The landscape of the telecommunications environment is constantly evolving; in terms of architecture and increasing data-rate. Ensuring that routing decisions are taken at the lowest possible layer offers the possibility of greatest data throughput. We propose using wavelengths in a DWDM scheme as dedicated channels that bypass the routing lookup in a router. The future trend of telecommunications industry is, however, toward larger numbers of interlinked competing operator networks. This in turn means there is a lack of a unified control plane to allow current networks to dynamically provision optical paths. This paper will report on the concept of Optical IP Switching. This concept seeks to address optical control plane issues in disaggregated networks while providing a means to dynamically provision optical paths to cater for large data flows.
Keywords: DWDM, GMPLS, IP switching.

1. INTRODUCTION

Growth of Internet traffic is continuing, and even traditional circuit switched applications, such as voice telephony, are converging to the IP layer, in the form of VoIP. Reports have shown that VoIP access lines increased by 100 % from 2003-2004 and will increase by a further 400 % by 2008 [1]. As packet traffic increases router throughput must be scaled accordingly, with a destination lookup search and match for each packet. The largest current routers operate at line interface speeds of 40 Gbps [2]. The port capacity is scalable up to 92 Tbps, by adding line-cards.

By utilizing a circuit switched service at a physical layer level, the packet throughput could be increased to throughput speed of the physical medium. The highest optical layer modulation rate currently used in industry is 40 Gbit/s, however, modulation rates of over 160 Gbit/s have been demonstrated in research [3]. When combined with DWDM, which allows multiple wavelengths in the same fibre, throughput rates of many Tbits/s can be achieved in a single fibre.

In this paper, a means to dynamically allocate a circuit switched connection will be examined. The circuit takes the form of an optical wavelength, and is built on the ideas behind IP switching introduced by Ipsilon in 1995. Interest in optical circuit switching is heightened at present due to GMPLS, a technology which is at the moment undergoing standardization by the IETF. The focus there is on creation of optical paths, in response to user requests. Here, however, the method we propose works well in a disaggregated environment with paths that are created dynamically by the network in response to bandwidth demands.

2. OPTICAL IP SWITCHING

The concept of an optical cut-through path has its origin in the ideas behind Ipsilon's IP switching. This form of switching was proposed in the mid 1990's leading to the creation of two new protocols; Ipsilon's general switch management protocol (GSMP) [4] and Ipsilon's flow management protocol (IFMP) [5]. This technology combines the speed of link level (layer 2) switching with the flexibility of network level (layer 3) routing.

A flow is a sequence of packets with identical tuples. For IP switching this tuple is taken to be source and destination IP addresses, source and destination ports, transport protocol, TTL value and type of service [6]. In normal IP routing all packets belonging to a flow must pass through the IP routing layer. This of course adds processing overhead in the routers and introduces delay. IP switching aims to improve performance by reducing the amount of packets that have to be individually processed.

The method proposed here uses an optical wavelength instead of a VC for the cut-through path. The lightpaths are created dynamically without user interaction. The router monitors the traffic as packets are routed, if a flow occurs, which matches certain characteristics, an optical cut-through path is established by the router between its upstream and downstream neighbours. The router which has detected the flow requests the upstream node to place all packets belonging to the flow onto the new path. This newly created lightpath is optically switched optically bypassing the IP layer of the router, as the packets transparently flow from the upstream to downstream router. The flow is thus created between three routers with the middle router bypassed. By following a similar procedure the flow can be extended beyond the third node, but this decision is taken autonomously by the router depending on the traffic metrics and resources available locally.

The material is based on work supported by Science Foundation Ireland under Grant No. 03/CE3/1405.

The disaggregated network is one where many telecommunication operator networks are encountered across the path from source to destination. Deregulation and other factors are leading to an increased number of operators, rather than the small number of large networks of incumbent telecommunication companies. In our concept the paths are created quickly and automatically in response to dynamic network conditions. No delay is encountered in waiting for a path across the complete network prior to transmission and no signaling is needed between the different operators as the cut-through path decision is handled locally at each router.

One wavelength is permanently allocated to allow the IP layer to route packets and exchange routing information, using existing protocols. This channel is also used for signaling messages to dynamically allocate the new cut-through paths. The upstream node has the task of injecting packets onto the flow wavelength. It also controls the flow by sending periodic refresh messages along the default link to ensure that the path is reserved, a reset message is sent when the flow is expired.

At present much work is ongoing in developing the architecture for GMPLS, addressing link management, topology discovery, connection provisioning, and the protection and restoration issues. The GMPLS standards, however, are focusing on reserving end to end lightpaths, using one of two competing protocols RSVP-TE [7] or CR-LDP [8]. Both protocols require the user to forward a request to reserve dedicated bandwidth to the network operator, using a user-network interface (UNI). The network control plane then configures an optical path with the required characteristics. If the path spans different providers, things are complicated by the need for inter-domain routing. Furthermore, interaction by the user with the management control plane, adds management and administrative overheads as well as contributing to delay. The scenario envisioned here allocates the lightpaths dynamically at a much higher speed, according to bandwidth usage. A provision may be also made to increase the quality of service (QOS), by providing priority access to resources for flows which have requested this QOS.

The ideal types of flows which would be suitable for this type of optical cut-through paths are the ones termed "elephants". These are the small number of flows which carry most of the network traffic. If there are not many "elephants" a number of smaller flows can be aggregated together to fill most of the bandwidth of the optical carrier. In the next section a study of some of the characteristics of flows from real networks is carried out.

3. THE CASE FOR OPTICAL IP SWITCHING

In order to test the potential for optical cut-through paths we examined a number of recent real Internet traffic traces taken from a trans-pacific link, on the WIDE backbone MAWI working group traffic archive [9]. The traces were chosen because they were the most recent traces available, additionally as the trace is from a trans-pacific (US – Japan) link the potential numbers of IP source and destinations is huge. Only one 15 minute trace per day was available taken at 2 pm. We examined the trace to find a base line for the number of flows present and their particular metrics such as number of packets, number of bytes, time of the first and last packet, port numbers and protocols. Other traces, from CAIDA [10], not shown here, were also examined which displayed similar characteristics to those shown in Fig. °1.

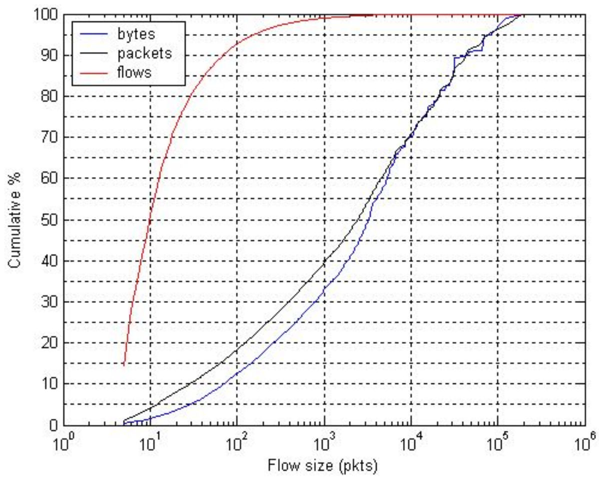

Figure 1. Cumulative data for flows of greater than 4 packets.

The flow tuple here is taken to be the source and destination IP addresses. The source and destination ports, protocol and TTL are all recorded but are not used to strictly define a flow. This is because we are just interested in the flow not flow content. The fields are important, however, in predicting the length of a particular flow. A flow is instantiated when two or more packets which have the same tuple arrive. Detection of a flow does not mean, however, that resources should be allocated to cut-through this flow. The means of choosing the flows which will be allocated resources will be examined elsewhere.

Based on the measurements elsewhere [6, 11] a good balance of flow timeout is found to be 60 seconds. A packet cache is used to hold packet headers, which have not been assigned to a flow. When a new packet arrives the cache is searched for a match and a flow is assigned if a match is found. As the flow timeout is given as 60 seconds, a packet's header must also be held for 60 seconds before it is timed out. Therefore a time out of 60 seconds for each flow, and a cache size of 600,000 packets, for packets which were not assigned to a flow, was chosen. The packet cache size was chosen because there are on average 10,000 packet arrivals per second, resulting in 600,000 packets in a 60 second time period. Running the program on this basis will provide the baseline data of the amount of flows in a particular dataset.

Analysis of our selected dataset showed that 86% of the packets were TCP, in agreement with previous analysis of IP traffic [12]. The remainder of the packets comprised of UDP 9.5%, ICMP 1.5% and the final 3% was from other protocols. The data also revealed that 96% of the packets were part of a set with at least one other packet with the same IP source and destination addresses, corresponding to 314,913 flows. To further refine the flows into larger flows we deemed a flow to be a full flow if more than 4 packets were detected within a 60 second time window. This is analogous to the X/Y classifier used in previous work [13]. The implication here is that if some packets are sent it is likely that more will be sent. Using this method 82,195 flows over the 15 minute period were detected. This method does not attempt to predict how long the flow will last for, and the performance is not very good. From Fig. 1 it can be seen that 93 % of the flows have fewer than 100 packets. The larger 7 % of the flows carry 80 % of the packets and 88 % of the bytes. The fact that so many packets and bytes are carried by so little flows provides enormous potential for the possibility of optical cut-through paths.

4. TESTBED

The optical testbed setup to explore the concept of optical IP switching is shown in Fig. °2. The network is composed of 6 nodes, and stretches across Dublin from Trinity College to Dublin City University, a distance of 16 km. The three routers are connected to each other and one end user machine is connected to each router.

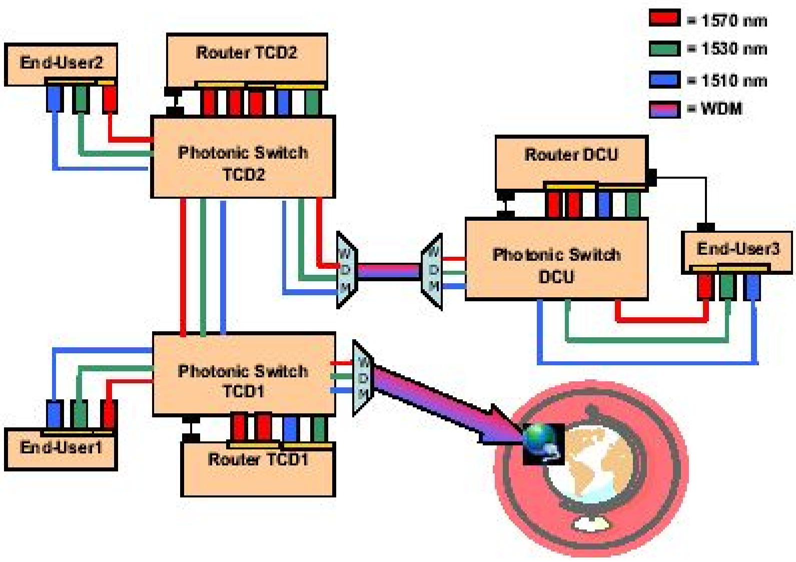

Figure 2. Optical router testbed setup.

The optical connections are based on Optical Gigabit Ethernet, and the connection between Trinity and DCU, a three wavelength dedicated bi-directional link, is provided by HEAnet, Ireland's National Education and Research Network. The photonics switches used are 16x16 port Glimmerglass MEMS switches. At present the routers used are blade servers running our proprietary protocol stack, on a Windows OS. Using the blade servers in this way allows for testing and development of the optical IP concepts with more control and flexibility. We wish to study and develop solutions for issues related both to the signaling mechanisms to create and extend flows, and to the flow analysis algorithm to trigger the path creation.

The first version of the flow detection algorithm has been implemented in the testbed. This permits us to automatically allocate cut-through paths based on the flow rate of the incoming traffic. The current implementation uses more dynamic protocols, such as those used in ad-hoc networks for example DSR and OLSR. Nevertheless, the only information used to create the paths is the source and destination IP address, making the fully system compatible with the IP layer. Work is underway to fully implement the IP layer, by making use of the CLICK modular router system [14]. It is hoped also that the testbed maybe linked to other testbeds worldwide that are used to explore similar issues.

5. CONCLUSIONS

There is enormous potential for the use of optical cut-through paths in a modern IP network. The large amount of traffic carried by relatively few flows: 10% of the flows carrying 90% of the traffic, means that a huge proportion of the traffic can be diverted on to a small number of wavelengths. This increases the router bandwidth by switching some capacity to the optical lightpath, thus, freeing capacity on the packet processor of the router, which can then be used to route more packets. Furthermore, increased usefulness of cut-through paths is likely with the adoption of bandwidth heavy applications, such as IPTV, which would increase the traffic carried by single flows. It is imperative, however, that the large flows are accurately predicted to increase resource allocation efficiency. Work is underway on algorithms which will choose the larger flows automatically. In the trace analysed from a total of 82,195 flows we wish to choose the small amount of large flows which carry most of the traffic: 7% of flows carry over 88% of the bytes and 80% of the packets. It must be pointed out, however, that the flows examined here represent a worst case scenario. In a network environment aggregation of flows will take place, so many small flows may be placed on an optical wavelength leading to increased routing efficiency. The physical barriers to adoption of optical cut-through paths are minimal. Advances in DWDM technology means that there is now access to many wavelengths on a single fibre. Future work will focus on developing algorithms to choose the largest flows accurately, in addition to resolving issues such as aggregation and extension of flows and the resultant optical control plane concerns.

REFERENCES

[1] M.D. Gallagher, "VoIP: The catalyst for universal, affordable broadband access by 2007," *presented at Voice on the Net*, San Jose, CA, 2005.

[2] Cisco, "Cisco carrier routing system data sheet," www.cisco.com/en/us/products/ps5763, 2004.

[3] M.S.E. Lach, M. Witte, K. Schuh, H. Buelow, F. Buchali,A.T. Clausen, L.M. Oxenløwe, B. Sartorius, C. Kazmierski, D. Rouvillain, P. Pecci, S. Vorbeck, R. Leppla, E. Le-Rouzic, B. Cuenot, A. Richter, F. Ramos, R. Llorente, "DWDM Transmission at Ultra-High Channel Bitrates: European TOPRATE Project View," *presented at ECOC*, Stockholm, 2004.

[4] P. Newman and W. Edwards, "Ipsilon's General Switch Management Protocol Specification," *RFC 1987*, 1996.

[5] P. Newman and W. Edwards, "Ipsilon Flow Management Protocol Specification for IPv4 version 1.0," *RFC 1953*, 1996.

[6] P. Newman, G. Minshall, and T. L. Lyon, "IP Switching - ATM Under IP," *IEEE/ACM Transactions on Networking*, vol. 6, pp. 117-129, 1998.

[7] "Generalised MPLS - RSVP - TE Extensions IETF RFC 3472," 2003.

[8] "Generalised MPLS - CR - LDP Extensions. IETF RFC 3472," 2003.

[9] MAWI, "Working group traffic archive." http://tracer.csl.sony.co.jp/mawi/, 2005.

[10] CAIDA, "Anonymized OC-48 traces." www.caida.org, 2003.

[11] N. Brownlee, "Some Observations of Internet Stream Lifetimes," 2004.

[12] N. Duffield, C. Lund, and M. Thorup, "Estimating Flow Distributions from Sampled Flow Statistics," *presented at SIGCOMM,* Karlsruhe, Germany, 2003.

[13] S. Li and N. McKeown, "A Simulation Study of IP Switching," *presented at SIGCOMM*, Cannes, France, 1997.

[14] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek, "The click modular router," *ACM Trans. on Computer Systems*, vol. 18, pp. 263-297, 2000.