Analysis of the Scalability of a Collision-Free Optical Packet Switched LAN

Xianhui Che, David K. Hunter, Ian D. Henning

University of Essex

Abstract: A local area network control protocol has been designated to implement collision-free optical packet switching. The resource scheduling takes place in advance and is based on a centralized scheme which raises scalability issues. The impact of the protocol on the network scaling parameters such as distance, nodes number and system configuration, is analyzed theoretically and via simulations to imply a stable network structure.

1. Introduction

Optical LAN technologies permit end stations to access the network via an entire optical data path, without the need for electronic switching in the central hub node. The end stations are linked to the hub by fiber, and the optical LAN protocols can be integrated into line cards in the end stations. Most existing optical LAN proposals are based on Wavelength Division Multiplexing [1], which however require sophisticated and specialized optical components, and question the associated issues of cost and complexity. Problems with those optical devices include not only expense and splitting loss, but also their inability to scale to larger LANs with many users. Often, WDM technology constrains the number of users to the number of available wavelengths.

We use a combination of electronics and optics to show how an optical packet network for computer data could be implemented using optical components that will be available in the short term. A centralized scheduling scheme [2] has been designed to avoid those limitations of WDM technology with out-of-band controlling. The successful exchange of RTS (*Request to Send*) and CTS (*Clear to Send*) messages leads to collision-free transmissions. The LAN is implemented as a star topology. A Network Allocation Vector (NAV) is maintained and updated by the arbiter, determining the switch states in the hub.

2. End-to-End Transmission

When receiving RTS messages from the end stations, the arbiter should serve all the requests as quickly as possible. The serving speed of the arbiter determines the service rate. Having updated the NAV according to the requests, the arbiter starts the process of reading the NAV columns one by one and sending CTS messages to the end stations. The service process regarding to one request starts from sending RTS and ends by the receiving CTS. To calculate the service delay, it has to include the propagation delay of the RTS and CTS messages. Assume there are N nodes and the arbiter timer is t seconds per interval and the propagation delay of the fibre link between the end station and the hub is equally d seconds. As R requests are sent at one time by each user, the NAV receives RN requests at one time. If we can calculate the overall time of serving all those requests and then divide it by RN, we will get the *mean service delay* per request. To make it easier, we partition the serving procedure into two steps. Step 1 is all the end users send requests simultaneously. The propagation delay is d. Step 2 is that NAV reads each request and gives off CTS messages to end users who will then transmit the data packets. Fig. 1 illustrates the step 2 process in the time domain. The time for step 2 is $(RN-1)t+3d$. Hence the mean service delay for each request is:

$$\text{Mean service delay} = (4d-t)/RN + t \qquad (1)$$

This equation is based on the assumption that all the links from the end users to the hub have the same length. According to the equation, theoretically the network service rate will degrade with long link propagation delay, and will yield relatively better performance with more nodes and longer length of request queue.

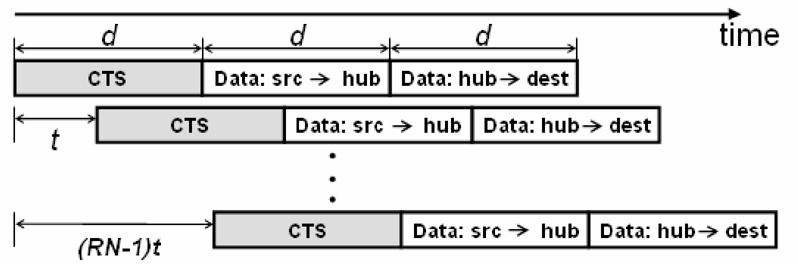

Figure 1: CTS and data transmission in Time scale

3. Scalability

3.1 Trade-off Between Latency and Service Rate

Fiber latency has a big impact on the end-to-end delay; the feasible service rate decreases markedly as the propagation delay is increased, leading to the conclusion that the end user will enjoy better network performance in a LAN of smaller physical extent. This is not only due to the transmission time of the packets themselves, but also the time required to exchange RTS and CTS packets prior to transmission. Equation (1) reflects the trade-off between the latency and the service rate. The proposed optical LAN has been studied via simulation using OPNET, with four end stations, and one hub. The traffic source has a uniformly distributed packet inter-arrival time; the traffic intensity may be varied by changing the maximum packet length. The packet size is fixed at 424 bits, the transmission rate is 10 Gb/s, and the control channel rate is 100 Mb/s. Because routing and switching of packets is scheduled in advance, there is no contention or buffering in the hub, and there is hence no delay there. Fig. 2 compares the average end-to-end delay for 100m, 10m, and 1m LAN with 128-packet buffers.

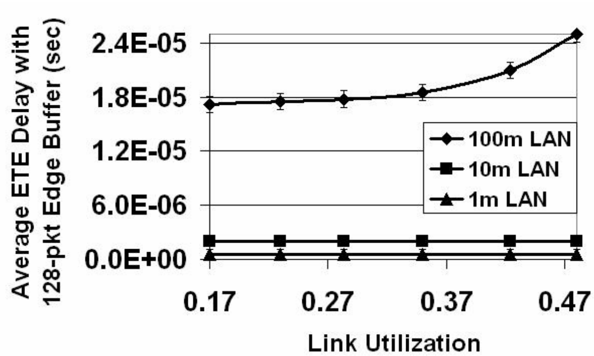

Figure 2: Average ETE Delay for 100m, 10m, and 1m LANs (seconds)

There is another factor restricts this relationship. Each switch in the hub maintains a switching table which is updated by the arbiter signalling based on the NAV contents. The system creates two pointers referring to the index of the switching table. One is for the signals from the arbiter, which increments by one every time the switch receives a configuration signal from the arbiter; in such a way that the arbiter is able to update the switching table line by line in the right sequence. The other pointer is for the signals of packet arrivals which moves forward every time the switch receives an incoming packet; the pointer helps to locate the proper line of the switching table so that the switch can be configured to the correct state.

If we use the current popular CPU of 3.5GHz, it gives the arbiter timer interval of 2.8×10^{-10} seconds. And the propagation delay in the fibre should be $m\times5\times10^{-9}$ seconds (i.e. m: integral; light speed = 2×10^{8} m/s). So presumably the link delay is normally longer than the arbiter timer interval which means the pointer for the packet arrivals always go behind the pointer for the arbiter signals on their way of scanning the switching table. In order to avoid the table information being overwritten by the arbiter signal before the packet arrival pointer reads it, the following formula must be met:

$$(RN)t + 2d \geq 2d(RN) \quad (2)$$

Hence we can get:

$$d \leq RNt/2(RN - 1) \quad (3)$$

The cause of this requirement is the limited fixed-length of the switching table. One solution for this trade-off may be the application of *Link List* [3] to make up switching table instead of array. With link list, the table can insert a new element whenever in need; also the existing elements can be removed anytime as well. In this case, the arbiter will be able to keep signalling to update the switching table without worry to overwrite any existing information. The packet arrival pointer will read the switch table asynchronously on the other hand.

3.2 Variation of the Number of Nodes

To evaluate the scalability in terms of number of nodes, we built a network model with 16 nodes which runs the same protocol. Same traffic source are employed in order to make explicit contrast. The simulation results give an optimistic view of the network performance when it is extended to more nodes.

Fig. 3 shows the average end-to-end delay of the 100m LAN with 4 and 16 nodes. By comparison 16 nodes network performance remains relatively stable with small amount of traffic; the end-to-end delay accelerates quickly when there is congestion in the network. According to equation(1), the actual service delay should go slightly smaller with more node number; however the end-to-end delay increases due to the long queuing delay in the edge buffer. The congestion possibility becomes bigger with the added nodes. Loads of packets have to wait in the edge buffer before getting transmitted, which accumulates a long delay.

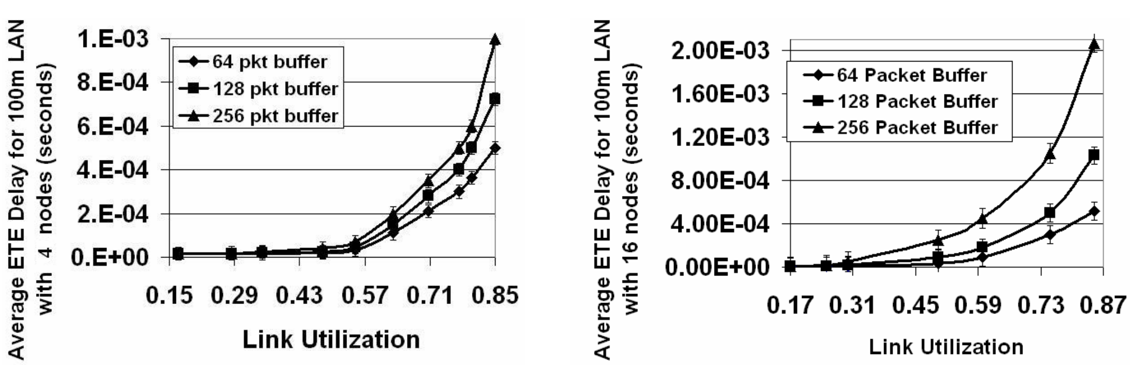

Figure 3: Average ETE Delay of 100m LAN with 4 nodes and 16 nodes (seconds)

3.3 Variation of the Request Queue Length

The exchange time of RTS and CTS messages is twice as much as the link propagation delay. This kind of delay takes a big percentage of the overall service delay. If the request queue can carry more requests at one time, the arbiter will be able to serve more requests at one turn with the same affection of the latency. According to (1), the service rate will be enhanced if the request queue length is long enough. On the contrary, if the request queue is very short, the network will less efficient. An extreme case is that R=1 (*R* is the request queue length) where each end node just sends one request at a time. Fig. 4 shows the average end-to-end delay of a four nodes LAN with R=1 and R=5 respectively. In the first case, we see that the network is able to handle small traffic well, but works less efficiently with the increased load.

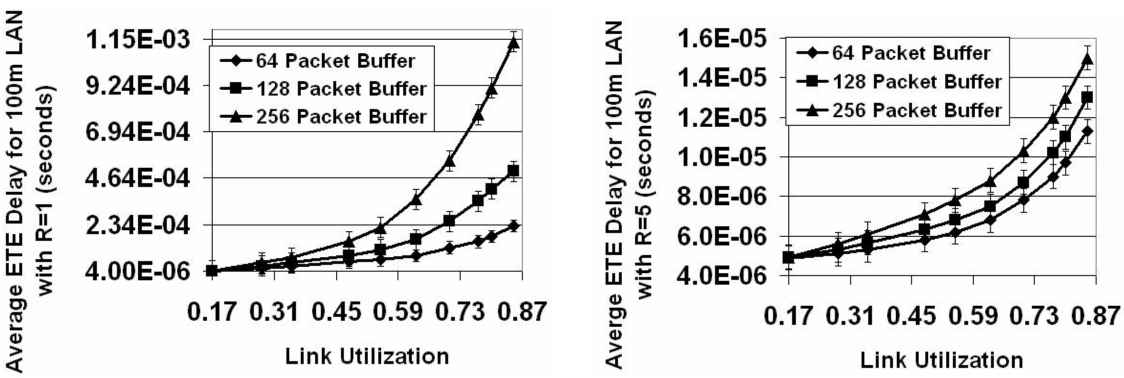

Figure 4: Average ETE Delay of 100m LAN with R = 1 and R = 5 (seconds)

4. Conclusion

The network scalability of the proposed optical LAN protocol has been analyzed theoretically. Simulations results agree with the equation brought up in chapter 2. There are both advantages and disadvantages. As expected, the fibre link between the end user and the hub influences the network service rate, so dimensional extension is relatively difficult. With longer fiber links (such as1000m), the network does not perform well with respect to either delay or packet loss, as the increase latency is detrimental to the service rate. So in practice, this LAN, as currently proposed, is restricted to fiber lengths under 100m. On the other hand, the network remains steady with more number of nodes. With small amount of traffic, the network delay of 16 nodes model is in the similar level as that of 4 nodes model. When there is more traffic, the average end-to-end delay of 16 nodes network model goes up by comparison with the one of 4 nodes model, but still remains in acceptable situation. Moreover, mean network delay can be reduced by increasing the request queue length. When each node sends more requests at one time, with the same impact of fibre link latency the network efficiency gets enhanced.

References

[1] Dan Stevenson, Ilia Baldine, George N. Rouskas, Helios, "Regional testbed optical access network for ip multicast & differentiated services", DARPA NGI PI Meeting, Jan 5, 2002.

[2] X Che, David K. Hunter, Ian D. Henning, "A MAC Protocol for a Switched Optical LAN", PREP Proceedings pp.15-16, 2005.

[3] Nick Parlant, "Link list basics", Stanford CS Education Library, http://cslibrary.stanford.edu/.