Comparison of the Gigabit Ethernet Full-Duplex Repeater, CSMA/CD, and 1000/100-Mbps Switched Ethernet

Kenneth J. Christensen
Computer Science and Eng
University of South Florida
Tampa, FL 33620
Email: *christen@csee.usf.edu*

Mart Molle
Computer Science
University of California, Riverside
Riverside, CA 92521
Email: *mart@cs.ucr.edu*

Sifang Li
Computer Science and Eng
University of South Florida
Tampa, FL 33620
Email: *li@csee.usf.edu*

Abstract

Recently, the Full-Duplex Repeater (FDR) has been proposed as an alternative to half-duplex operation using CSMA/CD for controlling shared access to Gigabit Ethernet. In this paper, the basic FDR architecture is described and two extensions for traffic control are introduced. Using simulation methods, the performance of the Gigabit FDR is studied under different topologies and population sizes for a range of offered load. It is shown that the FDR provides a dramatic performance improvement over CSMA/CD (using both BEB and BLAM arbitration) at high load. The Gigabit FDR is also compared with switched Ethernet in the context of medical image retrieval. It is shown that for medical image retrieval, the performance of the Gigabit FDR is much better than 100/100 or 1000/100-Mbps switched Ethernet, and equivalent to 1000/1000-Mbps switched Ethernet for low levels of non-image background traffic.

1. Introduction

Ethernet with a 1-Gbps data rate was standardized by the IEEE 802.3 working group in June 1998 (see [8]). The Gigabit Ethernet standard supports both full-duplex operation with IEEE 802.3x PAUSE flow control (see [7]), and half-duplex operation using Carrier Sense Multiple Access / Collision Detection (CSMA/CD). In order to support 100-meter link lengths, CSMA/CD has been modified to include carrier extension and frame bursting. Under carrier extension, the minimum duration of a successful packet transmission was effectively increased from 64 bytes to 512 bytes by requiring the transmitter to append extended carrier symbols to the end of a small packet until the minimum length is met (see [6]). To improve the efficiency of carrier extension, frame bursting was added to the Gigabit Ethernet standard (see [12]). In frame bursting, a transmitting host can concatenate additional packets into the same burst of carrier until a burst length timer expires. Frame bursting allows shared Gigabit Ethernet using CSMA/CD to attain surprisingly high throughput, although the delays tend to grow more quickly with network load than in 100-Mbps Ethernet (see [12]).

An alternative approach to providing shared access to a Gigabit Ethernet channel is the Full-Duplex Repeater (FDR), as proposed in [4]. The FDR is a packet-forwarding device that broadcasts incoming packets to all ports, similar to a conventional half-duplex repeater. However, each host is connected to a port on the FDR using a full-duplex link, similar to a conventional packet switch (see [11]). Thus, the FDR represents a middle-point between shared-bandwidth, distributed arbitration (i.e., CSMA/CD) and parallel-bandwidth, centralized arbitration (i.e., switched Ethernet). The FDR implements shared-bandwidth with centralized arbitration, allowing multiple attached hosts to share 1-Gbps of bandwidth without any collisions. For the first time, in this paper the performance of 1-Gbps shared-bandwidth Ethernet using an FDR is evaluated and compared directly with the performance of switched Ethernet. Performance of the two methods is studied for transferring large image files within a workgroup of less than 130 hosts.

The remainder of this paper is organized as follows. Section 2 describes the Gigabit FDR as proposed to the IEEE 802.3 working group, along with the IEEE 802.3x PAUSE flow control. Section 2 also proposes several additions to the base Gigabit FDR architecture and briefly describes Ethernet switch architectures. In Section 3 the class of “Gigabit applications” is discussed and an empirical traffic model for medical image transfers is described. Section 4 describes a simple throughput model, the simulation model, and performance evaluation experiments. The results from the simulation study are presented in Section 5 followed by a summary.

2. The FDR and Ethernet Switching

A traditional 10BASE-T or 100BASE-T Ethernet repeater is used to forward bits from any port to all other ports. The repeater does not contain any Media Access Control (MAC) function or packet buffering. Transmission arbitration is done in the hosts via the distributed, half-duplex CSMA/CD and Binary Exponential Backoff (BEB) algorithms. In a Gigabit FDR, shown in Figure 1, full-duplex MAC function, packet buffering, and centralized transmission arbitration is added to the repeater. Connection between host and repeater is via a full-duplex link. All hosts share access to a common 1-Gbps bus in the repeater. An arbitration, or scheduling, scheme such as round robin or weighted fair queueing (see [5]) is used to select the next packet to be forwarded from the repeater input buffers. A selected packet is then forwarded to all output ports with MAC-layer filtering of packets occurring at the receiving hosts. In the case of multiple transmitting hosts, IEEE 802.3x PAUSE flow control is used to throttle the sending hosts to a maximum 1-Gbps transmission rate.

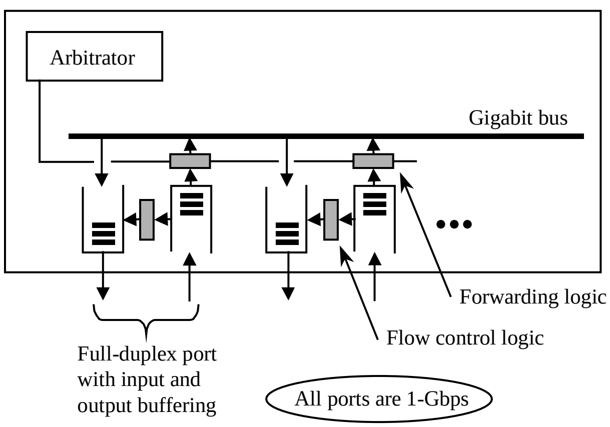

Figure 1 - Gigabit Full-Duplex Repeater (FDR)

The IEEE 802.3x task force has standardized Ethernet packet-based flow control via a specially defined PAUSE MAC frame (see [7]). The PAUSE MAC frame contains a field specifying a pause time, expressed as a multiple of a 512-bit quantum, that a receiving host MAC should withhold transmission of packets. A host Ethernet MAC receiving a PAUSE MAC frame completes any current packet transmission and then begins its forced idle pause time. If an already "pausing" host receives a PAUSE MAC frame, the host will replace its current pause time with the received pause value, possibly increasing or decreasing the forced idle time, or ending it if a pause time of zero is received. PAUSE flow control is used by the FDR to prevent overflow of the input buffers. Each input buffer has a low and high threshold defined. An arriving or departing packet that causes a threshold to be crossed can initiate a PAUSE MAC frame to be sent on the output link of the port. It is assumed that PAUSE MAC frames are priority queued, that is a PAUSE MAC frame is always the next packet sent after any current packet transmission. Figure 2 shows an algorithm for sending PAUSE MAC frames where `In_port[i]` is the input buffer for port `i` and a global variable `Over_flag` is initialized to `FALSE`. It can be seen that PAUSE flow control is used essentially as an XON/XOFF protocol to turn on and off host packet transmission.

```
if ((In_port[i] == over high threshold) &&
    (Over_flag == FALSE)) {
  send PAUSE packet with large pause value;
  Over_flag = TRUE;
}
if ((In_port[i] == under low threshold) &&
    (Over_flag == TRUE)) {
  send PAUSE packet with zero pause value;
  Over_flag = FALSE;
}
```

Figure 2 - Algorithm for sending PAUSE MAC frame

Within the Gigabit FDR, backpressure from an output to an input queue is needed to prevent multiple forwarding hosts from overflowing an output buffer. This internal backpressure is implemented via an extended poll time. A threshold is defined in the output buffer of each port and if exceeded the next poll time is extended by a time equal to the number of bits the threshold is exceeded divided by the port link data rate. This extended poll time allows the output buffer to drain below the threshold. Figure 3 shows the round robin polling algorithm with extended poll time. In the algorithm, `Num_ports` in the number of ports in the Gigabit FDR, `In_port[i]` is the input buffer for port `i`, and `Out_port[i]` is the output buffer for port `i`. If the internal backpressure causes an input buffer to fill beyond its high threshold, PAUSE flow control extends the backpressure to the sending host.

```
i = 0;
while(TRUE) {
  if ((In_port[i] has packet(s) queued) {
    remove and release a packet on the bus;
    wait for transmission on bus to complete;
  }
  Over = 0;
  for (j = 0; j < Num_ports; j++) {
    Over = bits Out_port[j] is over threshold;
    if (Over > 0)
      wait for Over bit times;
  }
  i = (i + 1) % Num_ports;
}
```

Figure 3 - Round robin polling with backpressure

2.1 Port bursting and packet spacing extensions

In this paper, two additions to the base FDR architecture are introduced: port bursting and packet spacing. Port bursting can be used at high offered loads to insure fairness for small packet traffic and to provide a per port bandwidth allocation scheme. Packet spacing can control the peak transmission rate of a host onto the Gigabit network.

The base FDR with round robin polling is fair by packet transmission count, but not necessarily by bytes transmitted. That is, if one port has small packets queued and another port large packets, then the port with large packets will receive a greater portion of the shared bandwidth when offered load is high. To minimize this unfairness and provide a provision for allocating bandwidth between ports, a port (or packet) bursting capability is proposed similar to timed token holding in FDDI and frame bursting in Gigabit CSMA/CD. In port bursting, a port may continue releasing packets from its input buffer on to the bus until a byte counter (`Byte_count` initialized to `BURST_SIZE`) that counts bytes transmitted is decremented to zero or the input buffer is empty. Ports with larger `BURST_SIZE` will be able to transmit more data than ports with smaller `BURST_SIZE` on their respective round robin polling cycle. Figure 4 shows the port burst capability added to the round robin polling algorithm of Figure 3 (the extended poll time of Figure 3 is omitted).

```
i = 0;
while(TRUE) {
  if (In_port[i] has packet(s) queued) {
    Byte_count = BURST_SIZE;
    while((Byte_count > 0) &&
          (In_port[i] has packet(s) queued)) {
      remove and release a packet on the bus;
      wait for transmission on bus to complete;
      decrement Byte_count by packet size;
    }
    i = (i + 1) % Num_ports;
  }
}
```

Figure 4 - Round robin polling with port burst counter

Using packet spacing, the peak bandwidth usage of a port can be limited to a predetermined value. Figure 5 shows the modified round robin algorithm whereby a port is only allowed to forward packets if a calculated `Packet_gap_time[i]` has expired. `Packet_gap_time[i]` is computed as the last packet transmission time divided by the percentage of total bandwidth the port has been allocated, subtracted by the last packet transmission time. For example, for 50% allocated bandwidth the `Packet_gap_time[i]` is exactly equal to the last packet transmission time. In the case of a time greater than `Packet_gap_time[i]` before a port with packet(s) queued is subsequently polled, the next port time-out period, timed by `Port_timer[i]`, is appropriately reduced (as shown in Figure 5). Thus, the mean port time-out period will be equal to, or greater than, the computed mean `Packet_gap_time[i]`. `Port_timer[i]` is initialized to zero. Implementation of this packet spacing requires that high-resolution timers be implemented in all ports of the Gigabit FDR.

```
i = 0;
while(TRUE) {
  if ((In_port[i] has packet(s) queued) &&
      (Port_timer > 0)) {
    remove and release a packet on the bus;
    wait for transmission on bus to complete;
    compute Packet_gap_time[i];
    set Port_timer[i] to Packet_gap_time[i];
    subtract from Port_timer[i] the time since
      last packet release by In_port[i];
    start Port_timer[i] if greater than zero;
  }
  i = (i + 1) % Num_ports;
}
```

Figure 5 - Round robin polling with packet spacing

2.2 Brief overview of switched Ethernet

Packet switching of Ethernet packets is implemented via “switched Ethernet” at 10, 100, and 1000-Mbps. Switched Ethernet is an implementation of IEEE 802.1D transparent bridging where packets are filtered based on MAC address. Thus, an Ethernet switch must contain address tables, address learning, and other functions needed to support the IEEE 802.1D standard. Figure 6 shows a simplified view of a 1000/100-Mbs Ethernet switch. In such a switch, a non-blocking switch fabric (e.g., a cross-bar or high-speed bus) is used to forward packets between ports. For full-throughput operation, the bandwidth of the switch fabric must equal the total bandwidth of all connected ports. PAUSE flow control is still needed in the case of multiple hosts (input ports) simultaneously forwarding packets to a single output port and thus causing potential buffer overflows. Due to the need for MAC function, high-speed packet buffering, high speed interconnect, and address table memory and learning function, the cost of a switch is higher than that of a repeater. By providing mixed-speed switches based on expected traffic patterns, the cost can be lowered. For example, one higher-speed port can be designated for a frequently accessed server connection and multiple lower-speed ports for client connections.

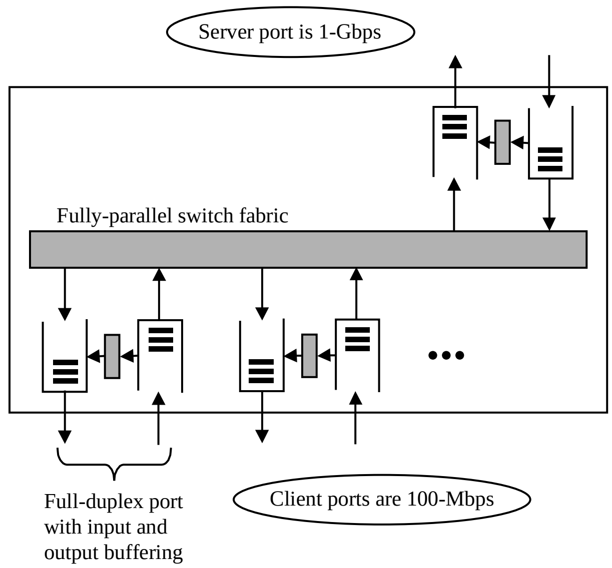

Figure 6 - 1000/100-Mbps Ethernet switch architecture

3. Model for Medical Image Transfers

Gigabit network capacity can be utilized by several classes of applications including backbones for multiple lower-speed LANs, and first-level (or workgroup) configurations for server farming, data warehousing, high-speed back-up systems, three dimensional imaging, direct transfer of print-copy from desktop to printer, and various image-rich applications including medical imaging. Developing traffic models for these types of applications is currently an area of research.

3.1 Model for medical imaging and background traffic

For this paper, we consider a medical imaging system such as Picture Archiving and Communication Systems (PACS) (see [14]) as a representative and interesting Gigabit application. A large PACS can contain more than 150 diagnostic workstations (client hosts) and the average image size can be in the range of tens or hundreds of Megabytes. A study of image sizes of a PACS at Hammersmith hospital in the United Kingdom is shown in Table 1 (from [14]) with a mean exam size of approximately 13.5 Mbytes. Doctors prefer to view images at a diagnostic workstation within 2 seconds of an image request (see [1]). In our studies we use 2 seconds as a performance bound for "satisfactory" image transfer time performance. It is expected that both the image resolution and depth will likely increase to provide better quality diagnosis (see [14]). With 3-D imaging, the network bandwidth requirements will become much larger. For example, in [3] the Zoomable Brain Database is shown to require from 720 to 11520-Mbps bandwidth to view 3-D brain images.

Modality	Image size	Exams / yr
X-radiography	13.5 Mbytes	73.5 %
Computed tomography	18.8	5.6
Ultrasound	12.8	11.0
Fluoroscopy	7.4	3.2
Angiography	10.0	1.7
Magnetic resonance	7.5	1.1
Nuclear medicine	0.2	3.8

Table 1 - Image statistics from a PACS (from [14])

In a centralized PACS system, all medical images are stored in one high performance image server. Diagnostic workstations throughout the hospital request images from the server. In addition to the image traffic, background traffic from client-to-client interaction and other applications will also be present. For the image transfer time component of this study, two traffic models were developed:

- **Image traffic model -** The image traffic model generates large-packet (1500 byte Ethernet packet size) traffic bursts with a burst size corresponding to an empirical distribution of image sizes based on Table 1. We assume that an image server can transmit image packets at a full 1 Gigabit data rate.
- **Background traffic model -** The background traffic model generates Poisson background traffic with packet sizes corresponding to the "IEEE workstation mix" distribution (from [6]). The mean packet size of the IEEE workstation mix is 616 bytes.

Figure 7 shows the network configuration of N hosts, one image server and $N - 1$ clients. The flow of image requests is from client-to-server and images then flow from server-to-client. The number of hosts is the same as the number of ports in the Gigabit FDR or Ethernet switch. For the image traffic model, client hosts independently request images from the server with a given request rate. We assume that:

1. A client will not request a new image before it finishes receiving the previously requested image.
2. The image server serves the clients' requests concurrently if multiple unsatisfied requests exist.
3. The image server is dedicated for image requests and no background traffic goes to or comes from the server.

The second assumption is significant when a 1000/100 switch is used, since the resulting overlapping of packets destined for the multiple clients allows the higher bandwidth of the server link to be fully utilized.

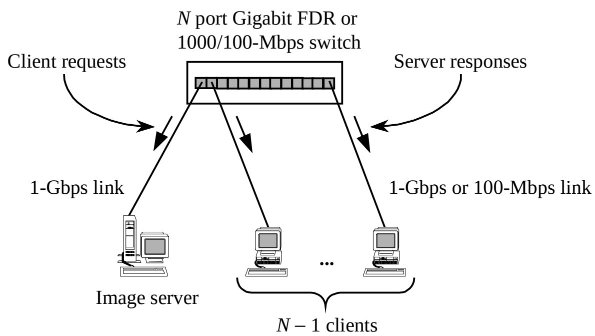

Figure 7 - Configuration for medical image retrieval

4. Performance Evaluation Experiments

Simulation models of a Gigabit FDR and an Ethernet switch were constructed using the process-oriented, station-centric method described in [2]. The CSIM18 function library (see [15]) was used to implement the models. The additions to the base FDR architecture (described in Section 2.1 of this paper) were included in the Gigabit FDR model. Settable parameters in the Gigabit FDR and switch implementations include:

- Sizes (specified in bytes) of input and output buffers for each port
- Thresholds (specified in bytes) for PAUSE flow control and internal backpressure (FDR only)
- Poll time (FDR only)
- Propagation delay from host to and through the repeater or switch
- Link speed
- Burst size and maximum bandwidth for the port bursting and packet spacing additions

Settable parameters in the host implementation include:

- Size (specified in packets) of packet queue
- Link speed

Independent to each host, a traffic source process (a model of an application) "feeds" the host packet queue with Ethernet packets. If a host queue reaches capacity, then the traffic source is blocked.

4.1 A simple throughput model for the Gigabit FDR

Using a fluid-flow approximation, we can estimate the mean image transfer time for a shared-bandwidth network configuration with a single image server and a given background traffic load. The image transfer time is,

$$T_{image} = \frac{S_{image}}{R - L_{background}} , \quad (1)$$

where T_{image} is the image transfer time in seconds for an image of S_{image} bits in size, R bits per second shared bandwidth, and $L_{background}$ bits per second background traffic load (where $0 \le L_{background} \le R$). We will compare the equation (1) model against simulation results for medical image transfer times. Offered load is expressed as a percentage of the available bandwidth (e.g., a background traffic offered load of 10% is $L_{background}$ = 100-Mbps).

4.2 Design of the experiments

Two sets of experiments are defined. The first set of experiments studies the throughput and packet delay behavior of the Gigabit FDR and CSMA/CD given simple Poisson offered traffic. The second set of experiments compares the Gigabit FDR with 100/100, 1000/100, and 1000/1000-Mbps switched Ethernet for simulated medical imaging traffic. Gigabit FDR's are available commercially (in mid-1998) in 12-ports only (see [13]). However, we study more than 12-ports to evaluate scaling of the Gigabit FDR architecture to larger host population sizes. For all experiments, packet size is measured as the number of bytes from the start of the MAC destination address field through the end of the Frame Check Sequence (FCS) field. Offered load and port throughput is measured on these fields only. The standard Ethernet 8-byte preamble and 12-byte interpacket gap are overhead. Gigabit FDR bus throughput is measured as the utilization of the shared bus.

For the first set of experiments, the primary measures of interest are the *throughput*, and *packet delay* mean and standard deviation. Packet delay is defined as the time of packet entry into a sending host packet queue to the packet received at its destination host. The Gigabit FDR contains 16 Kbytes of input buffering and 16 Kbytes of output buffering per port. The low and high thresholds for PAUSE flow control are 5 Kbytes and 10 Kbytes. For output port backpressure, the threshold is set at 8 Kbytes. Polling time is 16 nanoseconds per port. Unless otherwise stated, the end-to-end propagation time is 2000 bit times. The host packet queue is of size 100. For the CSMA/CD comparisons (experiment #1 only), the same propagation delay and host parameters are assumed. Both BEB and BLAM (see [10]) arbitration are used. A CSMA/CD burst size (see [12]) of 8-Kbytes is assumed. The first set of experiments consists of:

- **Experiment #1 -** general Gigabit FDR performance and comparison with CSMA/CD. Increasing offered load with 12 and 96 port configurations. Offered load is IEEE workstation mix. Destination addresses are uniformly randomly chosen for each packet.

- **Experiment #2 -** non-symmetrical topology delay fairness. Same as experiment #1 (for 12 ports only) except one link has host-to-repeater propagation delay of 10 bit times, the remaining hosts have 1000 bit times.
- **Experiment #3 -** small packet fairness and port bursting. Same as experiment #1 (for 12 ports only) except offered load is 64-byte packets from one host and 1500-byte packets from all remaining hosts. The port burst counter of Section 2.1 is evaluated for improving small packet fairness.
- **Experiment #4 -** bandwidth management with combined port bursting and packet spacing. Same as experiment #1 (for 12 ports only) except port 1 has a maximum bandwidth of 41.66-Mbps (1-Gbps divided by 24, or 50% offered load shared by all ports) set by packet spacing. In addition, all ports have a `BURST_SIZE` of 1500 bytes except port 2 has a `BURST_SIZE` of 3000 bytes. Port bursting is evaluated for increasing the bandwidth share for a port and packet spacing as a means of limiting the bandwidth of a port.

For the second set of experiments, the primary measure of interest is the *mean image transfer time* defined as the total time from client image request to the last packet of a requested image be received by the client. For all experiments, the Gigabit FDR contains 12 Kbytes of input buffering and 4 Kbytes of output buffering per port. The low and high thresholds for PAUSE flow control are 4 Kbytes and 8 Kbytes. For output port back pressure, the threshold is set at 3 Kbytes. Polling time is 16 nanoseconds per port. Unless otherwise stated, the end-to-end propagation time is 250 bit times. The Ethernet switch models are configured with the same buffer sizes as the Gigabit FDR. There is no poll time in the Ethernet switch, immediate forwarding of a received packet is assumed if the destination output port has sufficient space for the packet. The transfer time of the packet within the Ethernet switch is assumed to be the same as the port speed (i.e., either 1-Gbps or 100-Mbps). The second set of experiments consists of:

- **Experiment #5 -** increasing port count. Comparative evaluation of Gigabit FDR, 100/100, and 1000/100-Mbps Ethernet switching. Increasing port count of 12, 24, 48, 64, 80, 96, 112, and 128. Each client requests one new image with an exponentially distributed mean of 204 seconds (following the download of the previous image) and generates 4-Mbps of background traffic. The mean image request rate is based on the observation in [9] that medical doctors spend a mean of 204 seconds per image study.
- **Experiment #6 -** increasing background traffic for a fixed port count of 12. Comparative evaluation of Gigabit FDR and 1000/1000-Mbps Ethernet switching. The background traffic offered load is increased as, 10%, 20%, …, 80% of the Gigabit bandwidth. The image requests arrive as in experiment #5.

5. Performance Evaluation Results

5.1 Gigabit FDR and CSMA/CD comparison

Experiment #1 evaluates the general performance of the Gigabit FDR in terms of the mean and standard deviation of packet delay as a function of throughput for a range of offered loads and populations sizes, and compares the results with CSMA/CD. This simulation experiment was run for 1.1 million packets for all offered loads and the first 100-thousand packet completions were discarded as a "warm-up" period. Confidence intervals, measured with the CSIM18 built-in batch means method, are within 1% accuracy (for 95% confidence interval) in all cases except for a few of the very high offered load measurements. Figure 8 shows packet delay as a function of throughput for a 12 port configuration. In this case, the FDR offers a very significant performance improvement over CSMA/CD, although CSMA/CD with BLAM is somewhat better than with BEB. However, one must keep in mind that a one-millisecond delay is almost a hundred times larger than the transmission time for a maximum length packet on a Gigabit network. Thus, in reality CSMA/CD only provides good delay performance below 30% load, whereas the FDR provides good performance all the way up to 90% load. Figure 9 shows throughput as a function of offered load for the same experiment. In this case, the FDR throughput matches the offered load until complete saturation (100% offered load). BLAM offers some throughput improvement over BEB, but the FDR provides much better performance than either one.

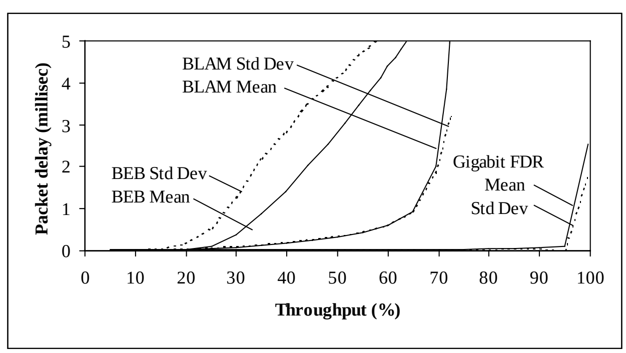

Figure 8 - Gigabit FDR and CSMA/CD packet delay

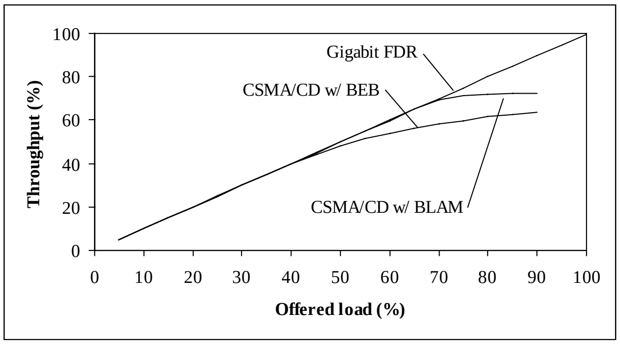

Figure 9 - Gigabit FDR and CSMA/CD throughput

Figure 10 shows the analogous delay performance for a 96 port configuration. Increasing the total number of hosts had no effect on the Gigabit FDR throughput. In fact, no packet losses were recorded in any of the simulation trials. Similarly, the delays are also quite similar to the 12 port configuration until the network approaches saturation, at which point the delays are significantly higher than for the 12 port configuration. This is expected because of Little's Law: since we have eight times as many hosts, the total number of waiting packets will be much higher when all hosts are saturated. The results for CSMA/CD show more of a change compared to the 12 port configuration. In this case, CSMA/CD using both BEB and BLAM only provide good delay performance up to an offered load of about 20%. Moreover, BLAM provides significantly worse mean delay (but better standard deviation of delay) than BEB. This performance degradation of BLAM is caused by the large slot time, which adds significant overhead to the BLAM arbitration mechanism in large population systems.

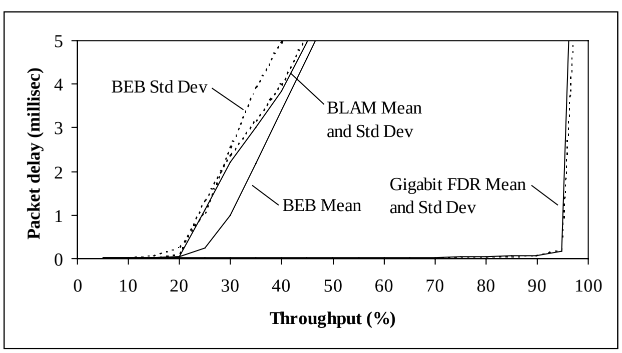

Figure 10 - Gigabit FDR and CSMA/CD packet delay

5.2 Results for a non-symmetrical topology

Experiment #2 evaluates the general performance of the Gigabit FDR given a non-symmetrical physical topology. Figure 11 shows the throughput in Mbps, measured at port 1 and average of ports 2 through 12 (the series are directly on top of each other!). This experiment shows that the Gigabit FDR does not favor "near" or "far" hosts.

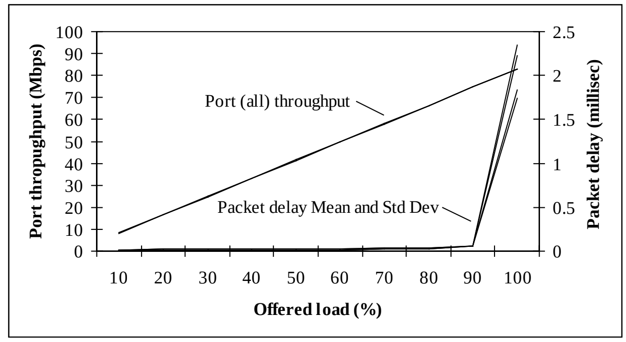

Figure 11 - Port throughput and packet delay results

5.3 Results for small packet throughput fairness

Experiment #3 evaluates the throughput performance of the Gigabit FDR given offered load consisting of small packets from one port and large packets from the remaining ports. This simulation experiment was run identically to experiment #1. Figure 12 shows the throughput at port 1 and average of ports 2 to 12 for two cases, 1) without burst counter enabled, and 2) with burst counter enabled and `BURST_SIZE` of 1500 bytes. It can be seen that without the burst counter at high offered loads ("Infin" is fully saturated queueing), the port with small packets is starved. With the addition of the port burst counter, all ports receive a fair share of the bandwidth.

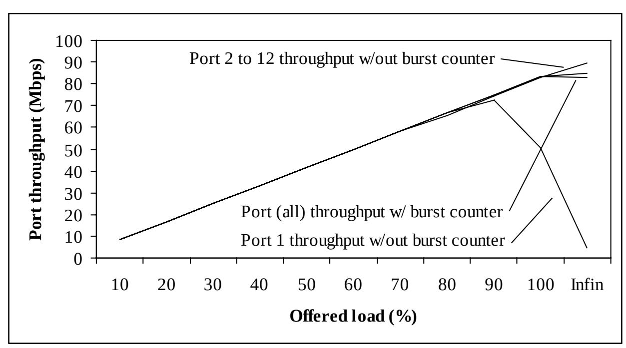

Figure 12 - Port throughput results

5.4 Results for simple bandwidth management

Experiment #4 demonstrates simple bandwidth management in a Gigabit FDR using the port bursting and packet spacing additions of Section 2.1. Figure 13 shows the experiment #4 results for port throughput for port 1 (with packet spacing for 41.66-Mbps maximum bandwidth), port 2 (with BURST_SIZE double that of all other ports), and the remaining ports. It can be seen that up to 50% offered load all ports have the same throughput, above 50%, port 1 is throttled to its designated bandwidth, and then at saturation port 2 has twice the throughput of ports 3 through 12.

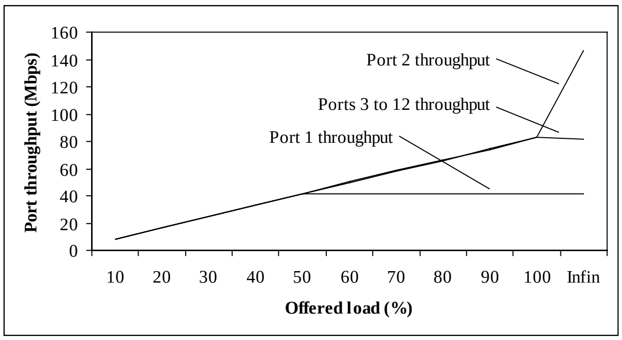

Figure 13 - Port throughput results

5.5 Results for medical image retrieval

Figure 14 shows the mean image transfer time results for experiment #5. It can be seen that both the Gigabit FDR and 1000/100-Mbps Ethernet switch are within the performance bound for up to, at least, 128 hosts. The 100/100-Mbps Ethernet switch does not achieve acceptable performance for greater than about 80 hosts.

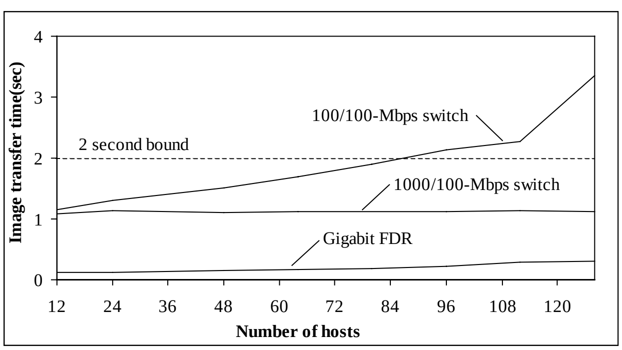

Figure 14 - Number of hosts and image transfer time

The histogram of image transfer times was plotted for each of the network configurations for a population of 80 hosts. For the Gigabit FDR and 1000/100-Mbps switched Ethernet, all image transfer times were less than 2 seconds even for a population of 128 hosts. However, for the 100/100-Mbps switched Ethernet approximately 34% of the image transfer times were greater than 2 seconds for 80 hosts. This is shown in Figure 15.

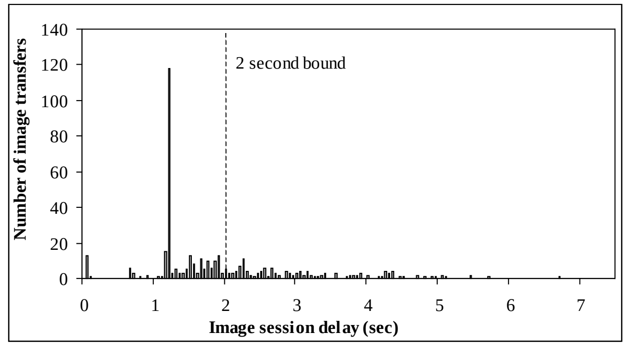

Figure 15 - Image transfer times for 80 hosts

Figure 16 shows the mean image transfer time results for experiment #6 illustrating the impact of background traffic on image transfer time. For the 1000/1000-Mbps Ethernet switch the background traffic utilizes the parallel bandwidth of the switch and does affect image transfer time. In a Gigabit FDR, however, the background traffic shares the same bandwidth as the image traffic and image transfer times are affected. The "ideal Gigabit FDR" is based on equation (1). Due to overheads the actual performance is slightly worse than the ideal case.

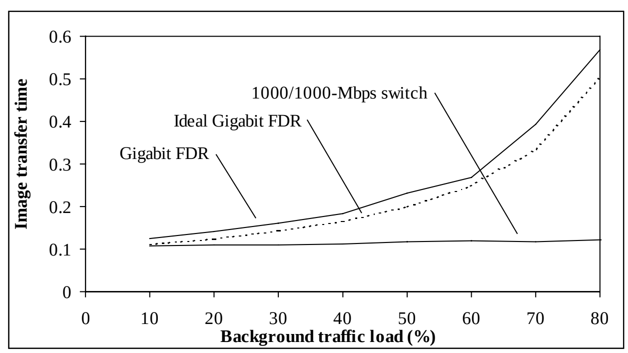

Figure 16 - Background traffic and image transfer time

6. Summary

The Gigabit FDR has been shown to be an efficient and fair architecture for sharing 1-Gbps bandwidth, which offers significantly better performance than CSMA/CD. The Gigabit FDR can offer full-throughput performance and is stable across a range of offered loads and host populations (port counts). The improvement in both throughput and packet delay for the Gigabit FDR, compared to CSMA/CD operation, is remarkable.

The FDR architecture fairly allocates bandwidth for small and large population networks and is not affected by "near" and "far" hosts. To improve small packet fairness at high offered loads, a port burst counter was introduced. The port burst counter and a packet spacing capability can be used to manage bandwidth in a Gigabit FDR. Port bursting allows long-term bandwidth allocations per port, and packet-spacing allows instantaneous transmission rates to be set. A likely and representative application for Gigabit Ethernet is medical PACS systems. A comparison of Gigabit FDR and 100/100, 1000/100, and 1000/1000-Mbps switched Ethernet demonstrated the suitability of Gigabit FDR for this type of application. Future work will focus on improving scheduling disciplines for Gigabit Ethernet for better bandwidth management and Quality of Service support

Acknowledgements

The authors would like to thank Boaz Yeger, of the University of California, Riverside, for his assistance in the execution of the CSMA/CD simulation experiments.

References

[1] W. Chimiak, "The Radiology Environment," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 7, pp. 1133 - 1144, September 1992.

[2] K. Christensen, M. Molle, and B. Yeger, "The Design of a Station-Centric Network Model for Evaluating Changes to the IEEE 802.3 Ethernet Standard," under review for *Simulation*.

[3] M. Claypool, J. Riedl, J. Carlis, G. Wilcox, R. Elde, E. Retzel, A. Georgopoulos, J. Pardo, K. Ugurbil, B. Miller and C. Honda, "Network Requirement for 3-D Flying in a Zoomable Brain Database," *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 5, pp. 816 - 827, June 1995.

[4] B. Daines, "Gigabit Buffered Distributor Proposal," *IEEE 802.3z Plenary Meeting*, Vancouver, Canada, November 1996. URL: h*ttp://grouper.ieee.org/groups/802/3/z/public/presentations/nov1996/BDbufdis.pdf*.

[5] A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm," *Computer Communication Review*, Vol. 19, No. 4, pp. 1 - 12, September 1989.

[6] H. Frazier, Jr., "Review and Update of Carrier Extension Proposal," *IEEE 802.3z Plenary Meeting*, Vancouver, Canada, November 1996. URL: *http://grouper.ieee.org/groups/802/3/z/public/presentations/nov1996/HFcarext.pdf*.

[7] IEEE P802.3x (Specification for 802.3x Full Duplex Operation), P802.3x/D3.1, *Institute of Electrical and Electronic Engineers*, New York, December 16, 1996.

[8] IEEE P802.3z (Specification for 802.3z Gigabit Operation), P802.3z/D5, *Institute of Electrical and Electronic Engineers*, New York, June 29, 1998.

[9] H. Kalyoncu and B. Sankur, "Design and Performance Evaluation of Picture Archival and Communication Systems," *Simulation*, Vol. 64, No. 2, pp. 77 - 90, February 1995.

[10] M. Molle, "A New Binary Logarithmic Arbitration Method for Ethernet," *Technical Report CSRI-298*, Computer Systems Research Institute, University of Toronto, 1994.

[11] M. Molle and G. Watson, "100Base-T / IEEE 802.12 / Packet Switching," *IEEE Communications Magazine*, Vol. 34, No. 8, pp. 64 - 73, August 1996.

[12] M. Molle, M. Kalkunte, and J. Kadambi, "Scaling CSMA/CD to 1 Gb/s with Frame Bursting," *Proceedings of the 22nd IEEE Conference on Local Computer Networks*, pp. 211 - 219, November, 1997.

[13] Packet Engines, Inc. FDR(tm) Gigabit Ethernet Hub, 1998. URL: *http://www.packetengines.com/f_productsandsolutions.htm*.

[14] R. Reynolds, "Digital Radiology and PACS: Pictorial Archiving and Communication Systems," Grainger and Allison's Diagnostic Radiology : A Textbook of Medical Imaging, Churchill Livingstone, New York, 1997.

[15] H. Schwetman, "Introduction to Process-Oriented Simulation and CSIM," *Proceedings of the 1990 Winter Simulation Conference*, pp. 154 - 157, December 1990.