

Short Paper

Toward Ensuring Fair Service Among ECN and non ECN TCP Connections Over the Internet

SALAHUDDIN MUHAMMAD SALIM ZABIR[†], AHMED ASHIR^{*}, GEN KITAGATA[†],
TAKUO SUGANUMA^{†*} AND NORIO SHIRATORI^{†*}

[†]*Research Institute of Electrical Communication
Tohoku University*

Sendai, Japan

^{*}*JGN R&D Project, TAO*

Tohoku University Office

Sendai, Japan

Providing fair service among TCP connections over the Internet without hampering resource utilization has recently become a major performance concern. With the rapid expansion of the Internet along with its increasingly diversified client base, the demand for a reasonably fair service is getting even stronger. To meet these requirements, congestion sensing is essential. Explicit Congestion Notification, ECN has been proved to provide a fast indication of incipient congestion and thus better the performance of a TCP/IP network. In our previous works we have proposed a strategy for ECN called Fair In-time Marking, FIM, and observed its superiority over other ECN schemes in terms of assuring a fair service. However, all TCP connections are not ECN capable. Therefore, they use packet drops for congestion signaling. In this work, we carry out investigations on gateway or router performance in providing fairness when both FIM ECN capable and non ECN capable connections are employed. We propose a new packet dropping scheme named Fair In-time Dropping, (FID) which drops packets from a connection upon detecting an incipient indication of congestion depending on its share of buffer occupancy. We also show that the combination of FIM and FID offers the best fairness compared with combining FIM with other dropping schemes.

Keywords: fairness, congestion signaling, ECN, packet dropping, drop tail, drop front, fair in-time dropping

1. INTRODUCTION

Providing fair service among TCP connections over the Internet without hampering resource utilization has been recently becoming a major performance concern. With the rapid expansion of the Internet along with its increasingly diversified client base, the demand for a reasonably fair service is getting even stronger. To meet these requirements,

Received September 14, 2001; accepted April 15, 2002.

Communicated by Jang-Ping Sheu, Makoto Takizawa and Myongsoon Park.

congestion sensing is essential. In a typical window based network traffic management scheme like TCP, this is done implicitly by detecting packet drops due to buffer overflow at some point in the congested network. Algorithms generally employed for the purpose either use TCP time out or duplicate ACKnowledgements.

Packet drop in either case degrades network performance since it requires extra time and causes TCP retransmissions which adds to network congestion.

These disadvantages of sensing congestion implicitly pave the way for Explicit Congestion Notification (ECN). The idea is to get an indication of incipient congestion before the congestion actually occurs and thus control traffic sources so as to avoid buffer overflow. There are two approaches ECN. These are either sending an ICMP Source Quench message to the sender [1] or setting unused bits in IP and TCP headers [1, 2]. In this paper, we will deal with the later for ECN, i.e., setting bits in the packet headers.

In our previous work, we presented a new mechanism for ECN, namely Fair In-time Marking, FIM [3]. Like any ECN, FIM activates its packet marking mechanism when the queue length at the router exceeds a certain threshold. But FIM marks a packet when the corresponding connection's share of buffer occupancy is more than it deserves. This ultimately results in a reduction in packet sending rate of the corresponding connection. We have shown that FIM provides better fairness in service than other ECN strategies like mark-tail or mark-front without hampering resource utilization.

However all TCP hosts are not ECN capable. In such cases, in order to avoid a situation like buffer overflow, when the queue length at the router buffer exceeds the threshold value the router or gateway may drop packets to signal an incipient congestion implicitly [4]. Such droppings can be done from the tail or front of the router or gateway buffer queue. These are called drop-tail and drop-front strategies respectively.

In this paper, we investigate scenarios where both ECN capable, more specifically, FIM capable and non ECN capable TCP sources operate together, and observe the performance implications in terms of resource utilization and fairness. We also propose a new packet dropping scheme named Fair In-time Dropping, FID which drops packets from a connection upon detecting an incipient congestion depending on its share of buffer occupancy. If the share of a connection corresponding to the outgoing packet from queue is more than it deserves, the packet is dropped. Experiments as will be presented in the following sections show that, among the combinations of FIM ECN and different dropping schemes, FIM-FID combination provides the best fairness along with ensuring high resource utilization.

The rest of this paper is organized as follows. In section 2, we describe the problem domain we are going to address in this work. In section 3, we describe our FID approach along with other packet dropping schemes with a brief outline of FIM ECN strategy. We present evaluation of our approach in section 4. Finally we conclude the paper in section 5.

2. PROBLEM DOMAIN

Internet applications are diversified in nature. Packets sharing the same buffer space in a gateway or router may belong to connections with a wide range of Round Trip Time (RTT), packet size, transmission rate and so on. As the client base of the Internet is in-

creasing at a massive rate and the demand for a reasonably fair service is strong, we have mainly addressed the impact of heterogeneity among connections in terms of different round trip times upon combinations of FIM capable connections with non-ECN capable connections employing different packet dropping schemes.

ECN is an addition to normal TCP congestion control scheme like slow start and congestion avoidance [5]. If the ACKnowledgement is not marked, TCP source sends data and increases the congestion window. When the ACKnowledgement is marked, the source halves the congestion window and reduces the slow start threshold. Similarly for an early congestion indication using packet drops, packets are dropped instead of being marked upon sensing an incipient congestion. If a packet drop occurs, no matter whether it is due to early congestion indication or buffer overflow, the normal TCP algorithms to reduce window size and retransmit the dropped packet are employed.

As stated previously, ECN *congestion experienced* bits are set upon detection of an incipient congestion. Similarly, with early congestion indication using packet drops, packets are dropped in such a situation. There may be many approaches to the detection of a probable congestion. In order to avoid sending congestion signals caused by transient traffic, and to avoid global synchronization, [2, 4] suggest that early congestion detection mechanisms should be used along with RED gateway and *average queue length*. In this paper, as in [6], we present the scheme with *actual queue length*. That is, when the *actual queue length* is smaller than the threshold, the incoming packet will not be marked; when the *actual queue length* grows larger than the threshold, the incoming packet will be marked. This corresponds to a special case of RED with an *average queue weight* value, $w_q = 1$ and $th_{min} = th_{max}$ [4] for any probability p_a of dropping.

For practical purposes, we employ the following assumptions without loss of generality. These are, (1) Data traffic is unidirectional. (2) Only ACKnowledgements are sent in the opposite direction. (3) Receiver windows are large enough so that the bottleneck is in the network. (4) Senders always have data to send in terms of as many packets as are necessary. (5) The queue builds up at only one bottleneck link. (6) Each packet received is followed by a corresponding uncompressed ACKnowledgement. (7) The queue length is measured in bytes, and packets may have different sizes. (8) Each connection can be identified by the gateway or the router capable of delivering Explicit Congestion Notification [7].

3. DROPPING STRATEGIES AND FAIR IN-TIME DROPPING

In this section we present existing common packet dropping schemes. We also present our proposed approach, Fair In-time Dropping, FID. These are employed in combination with our FIM ECN strategy. We consider TCP Reno, the most widely used TCP implementation in the following discussions.

Drop-Tail: Most of the literature dealing with early packet drop suggest that once the queue length becomes such that it indicates an incipient congestion, packets arriving at the buffer are dropped. This scheme is termed drop-tail. Indication of packet drop through duplicate ACKnowledgements in the drop-tail scheme requires that the packets following it to pass through the buffer and reach the receiver. In case there are not enough packets in the sending window so that the receiver can send three duplicate ACKnowledgements in response, the drop is detected through time out.

Drop-Front: In [8], a scheme named drop-front for marking packets has been proposed. Here, when the queue length becomes such that it indicates an incipient congestion, packets leaving the buffer are dropped. This scheme provides a faster indication of packet drop than drop-tail using duplicate ACKnowledgements if there is already enough packets in the buffer.

Fair In-time Dropping (FID): In our proposed scheme of packet dropping, we drop a packet considering a situation of incipient congestion along with the share of buffer of the particular connection corresponding to that packet.

Suppose at time t , we have N active connections sending packets to a router connected with a bottlenecked link. The queue size at the buffer is $Q(t)$, and the threshold beyond which a marking action is to take place is T . Source i contributes $n_i(t)$ packets each with size S_i . Then we may calculate $Q(t)$ as

$$Q(t) = \sum_{j=1}^{j=N} n_j(t) \times S_j \quad (1)$$

Although we have assumed constant packet sizes for each connection, the size may be varying sometime.

In FID, we consider the contribution of an active connection i to the incipient congestion as its share of buffer occupancy, $O_i(t)$, in the queue

$$O_i(t) = \frac{n_i(t) \times S_i}{\sum_{j=1}^{j=N} n_j(t) \times S_j} \quad (2)$$

and our condition for dropping a packet at the front end of the queue before being transmitted to the bottleneck link is

$$\frac{n_i(t) \times S_i}{\sum_{j=1}^{j=N} n_j(t) \times S_j} > Y \times E_i \quad \& \quad Q(t) > T$$

Here Y is a tuning factor selected to be close to but less than 1 to yield optimum performance from FID; E_i is the fraction of link bandwidth deserved by connection i . We use similar conditions in FIM for marking packets to indicate an incipient congestion.

4. EXPERIMENTS AND EVALUATION

In order to compare the performance implications of employing FIM ECN capable TCP connections together with non-ECN capable ones, we carry out a set of simulations using *ns* [9] simulator. A number of sources (Src1, Src2, ..., SrcN) are connected to the router (rtr1) by 10 Mbps link. Router (rtr1) is connected to router (rtr2) by a 1.5 Mbps link. The destinations (Dst1, Dst2, ... DstN) are connected to rtr2 by 10Mbps links. The links speeds are chosen so that congestion will occur only at router (rtr1) where we conduct out experiments.

4.1 Performance Metrics

In order to compare performance obtained by employing FIM capable and non ECN capable TCP connections, we consider the following three frequently used performance metrics [10].

4.1.1 Fairness

Our target metric is fairness the index. Suppose, x_i is the throughput for connection i , and e_i is the expected throughput for the same connection. In our experiments, we have assumed all e_i s to be equal without affecting the generalizations. Then for N number of connections, the fairness index is calculated as

$$F = \frac{\left(\sum_{i=1}^{i=N} x_i / e_i \right)^2}{N \times \sum_{i=1}^{i=N} (x_i / e_i)^2} \quad (3)$$

For convenience in comparison, we consider the complement to represent it in terms of an unfairness index

$$U = 1 - F \quad (4)$$

4.1.2 Link efficiency

Link efficiency is calculated from the amount of acknowledged data excluding re-transmissions divided by the amount of data that can be transmitted during the simulated time. Because of the slow start phase and possible link idling after the window reduction, the link efficiency is always less than 1. Link efficiency should be measured using a lengthy simulation to minimize the effect of initial transient state. Suppose x_i indicates the throughput for connection i , and C denotes the maximum possible throughput for the configuration. Then link efficiency can be defined as

$$E = \frac{\sum_{i=1}^{i=N} x_i}{C} \quad (5)$$

4.1.3 Delay

It is desirable that packets queued at the congested router do not have to wait long before being transmitted. In order to be effective, along with improving other performance metrics, our proposed FID should ensure low delay when deployed together with FIM ECN strategy. Delay in the queue is therefore an important metric in such a case. We considered two metrics related to delay. These are *Average Delay* and *Variance of Delay*. A smaller value of both is desirable.

4.2 Simulation Environment

We designed the following simulation scenarios based on the basic simulation model described in Fig. 1. If not specified, all connections have an RTT of 59 ms, start at 0 second and stop at the 10th second.

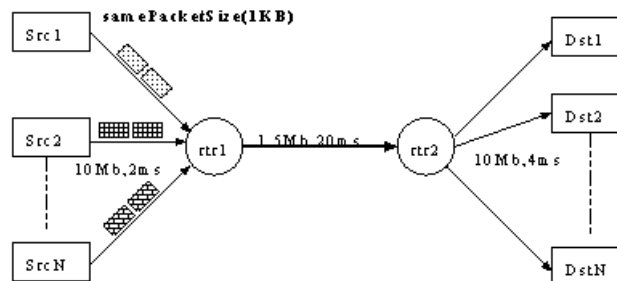


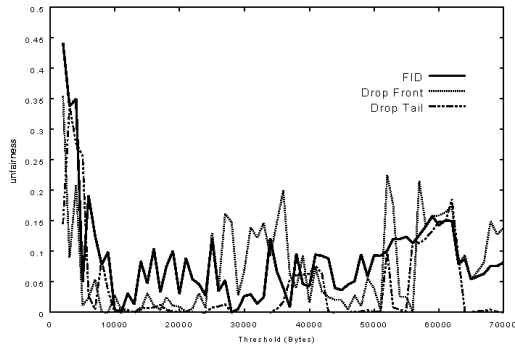
Fig. 1. Simulation model.

1. Two same connections, one FIM ECN capable.
2. Two connections with RTT equal to 59 and 157 ms, the latter non-ECN capable.
3. Two connections with RTT equal to 59 and 157 ms, FIM ECN capable.
4. Five connections with RTT of 59, 67, 137, 157 and 257 ms respectively, the last three connections non-ECN capable.
5. Five connections with RTT of 59, 67, 137, 157 and 257 ms respectively, the last two connections non-ECN capable.
6. Five connections with RTT of 59, 67, 137, 157 and 257 ms respectively, the last three connections FIM ECN capable.
7. Five connections with RTT of 59, 67, 137, 157 and 257 ms respectively, the last two connections FIM ECN capable.
8. Five connections with different packet sizes of 500, 1000, 1500, 2000, 4000 bytes.

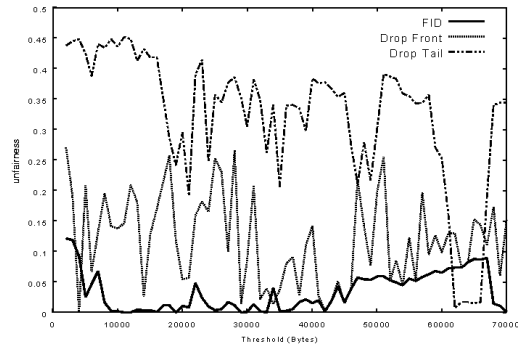
4.3 Fairness

Fig. 2 shows unfairness in FIM-FID, FIM-drop-front and FIM-drop-tail combinations in scenarios 1, 3, 4, 5, 6, 7. It is clearly visible that FIM-FID combination provides the least unfairness, i.e. the best fairness in all the cases with the minimum amount of oscillation in performance. FIM-drop-front is the next best in terms of providing fairness. As shown in Figs. 2 (b), (e), (f) that FIM-drop tail suffers the worst unfairness in cases where large RTT TCP connections are FIM capable (scenarios 3, 6, 7 in section 4.2). On the other hand, when smaller RTT connections are FIM ECN capable, the unfairness in service with FIM-drop-tail combination approaches that of FIM-drop-front combination. This indicates that in such scenarios (scenarios 4,5 in section 4.2), packet drops are less and each connection gets a fair share of resources.

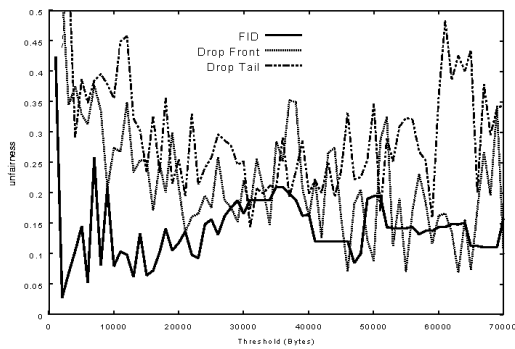
Therefore in more realistic cases when not all the connections are ECN capable at some router or gateway, we should employ FIM for ECN capable TCP connections along with our proposed FID scheme for non-ECN capable TCP connections to ensure a fair service among connections without hampering link efficiency.



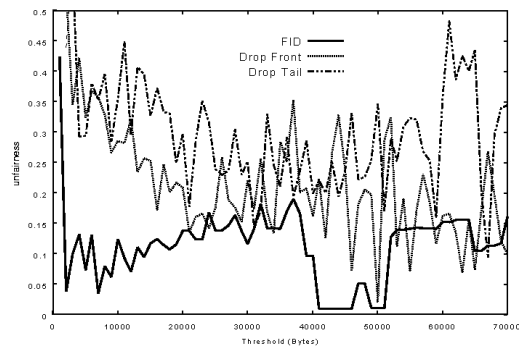
(a) Two same connections, one ECN capable (scenario 1, section 4.2).



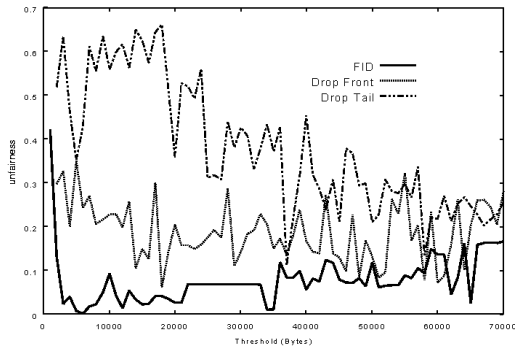
(b) Two Connections, large RTT ECN capable (scenario 3, section 4.2).



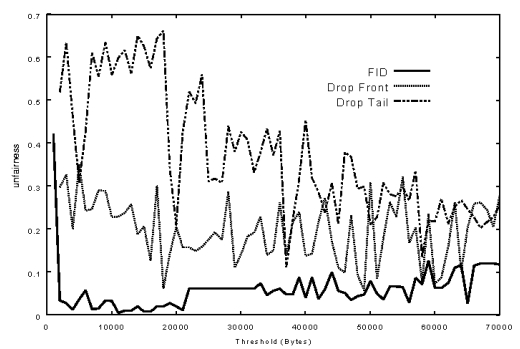
(c) Five connections, 2 small RTT ECN capable (scenario 4, section 4.2).



(d) Five connections, 3 small RTT ECN capable (scenario 5, section 4.2).



(e) Five connections, 3 large RTT ECN capable (scenario 6, section 4.2).



(f) Five connections, 2 large RTT ECN capable (scenario 7, section 4.2).

Fig. 2. Unfairness in various scenarios.

4.4 Link Efficiency

Fig. 3 shows link efficiency vs. congestion detection thresholds in several simulation scenarios representing several hundreds of simulation runs with different combinations. We observe that too small a threshold causes early backoff of the TCP sources and results in bottleneck link idling. So at small threshold region, link efficiency is low. At moderate thresholds, the efficiency increases.

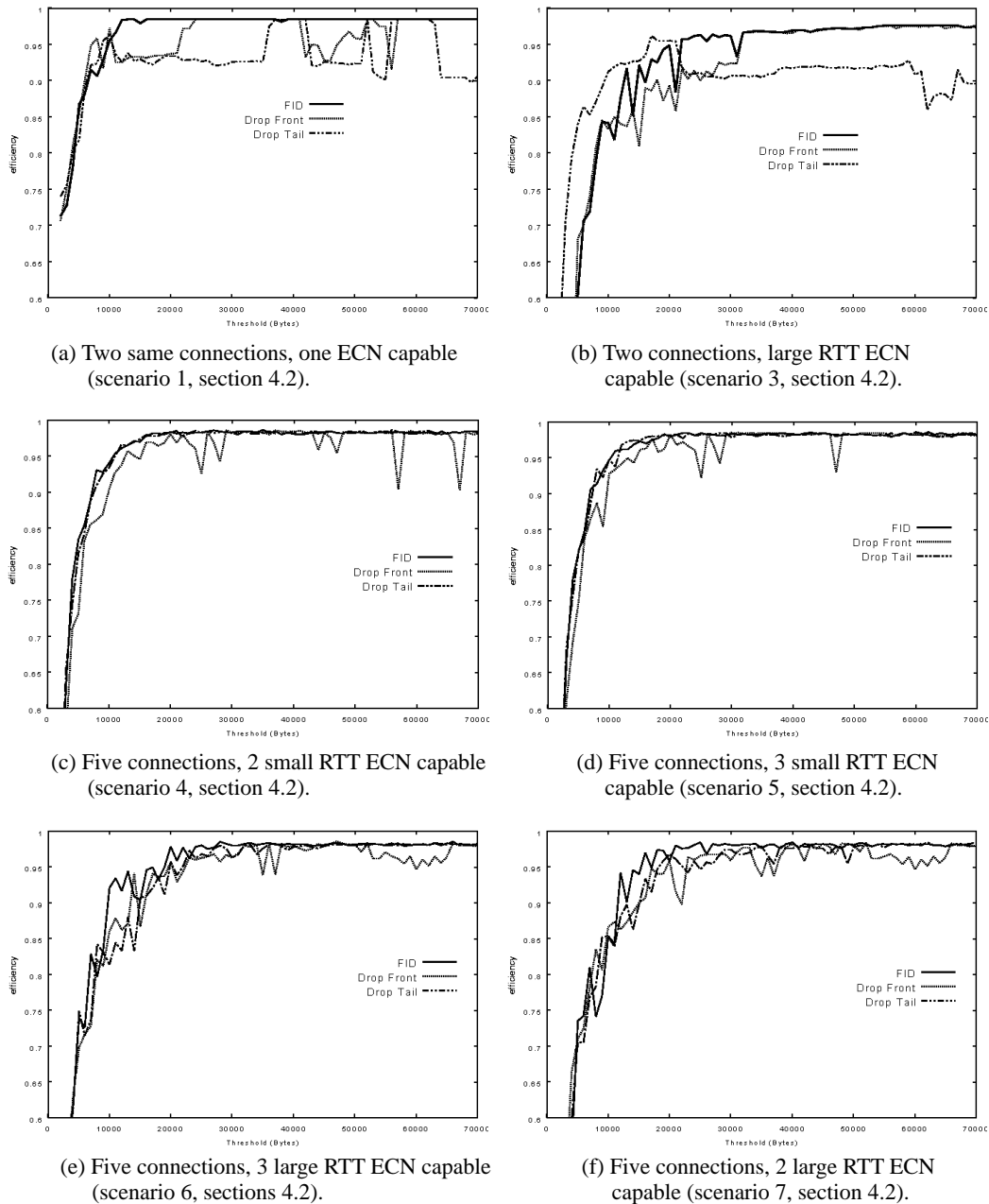


Fig. 3. Link efficiency in various scenarios.

The efficiency obtained from the combination of FIM and our proposed new dropping scheme FID is higher than FIM-drop-front or FIM-drop-tail combinations. It is interesting to note that efficiency decreases in case of FIM-drop-tail combination when larger RTT flows are FIM ECN capable while the smaller RTT flows are not (Figs. 3 (b), (e), (f), corresponding to scenarios 3, 6, 7 in section 4.2). On the other hand, when large RTT TCP connections are non ECN capable, FIM-drop-tail combination enjoys comparatively higher link efficiency. FIM-drop-tail combination is found to be non responsive to such conditions.

When large RTT connections are FIM ECN capable, FIM ensures fair allocation of buffer space for them. In such a scenario, drop-tail scheme drops packets of small RTT connections at the entry of the queue. But when FIM ensures fair allocation of buffer space for small RTT connections, packets from large RTT drop-tail connections suffer relatively less number of losses. A look at Fig. 2 establishes this fact further. This explains the above mentioned differences in link efficiency.

4.5 Delay

Fig. 4 shows average delay and variance of delay in scenario 7, section 4.2. This is typical a pattern for other scenarios as well. From the figure we observe that average delay is almost similar for all the cases while the variance of delay for FID is the lowest most of the time. A low variance indicates a stable value of the average. Therefore, for FID, average delay is representative of the delay to be expected by an incoming packet. On the other hand, with drop-tail, variance of delay is higher. But, since average delay remains close to the other two schemes, there are considerably higher and lower values of delay in the queue. That is, even though some packets will experience a small delay, the maximum delay experienced by some packets in the queue will be rather high. Such unusually high delay adversely affects TCP performance and is quite undesirable. Therefore, it is evident that when applied with FIM ECN strategy, FID provides good delay behavior, making it suitable for practical deployment.

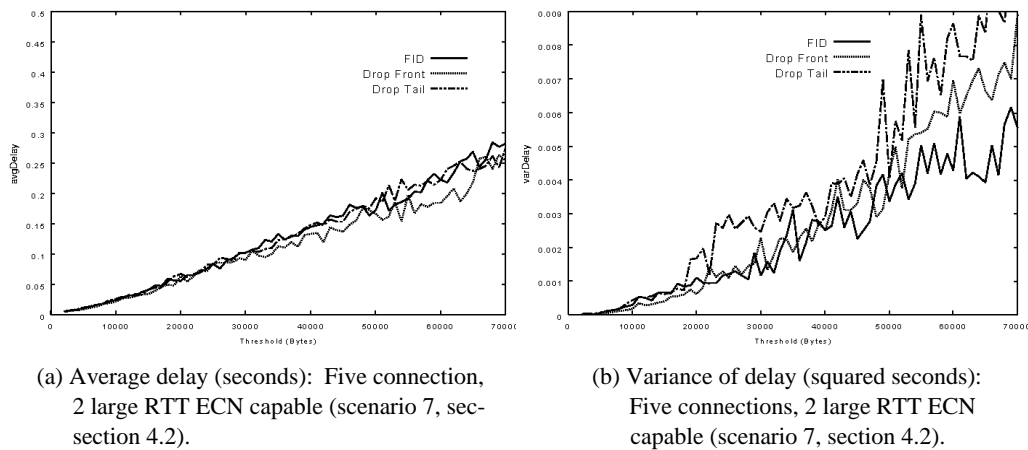


Fig. 4. Average delay and variance of delay.

5. CONCLUSIONS

In this paper we have presented results from investigation of the effects on performance of employing ECN capable and non-ECN capable congestion signaling TCP connections. We introduced a new non-ECN capable congestion signaling scheme called Fair In-time Dropping. Experimentations show that combination of FIM and FID offers a high link efficiency, ensures fair service among connections and maintains a stable average value of delay in the queue.

Therefore we conclude that for routers or gateways handling both ECN capable and non-ECN capable TCP connections, our proposed method can yield high link efficiency and fair service among connections by employing a FIM-FID combination.

REFERENCES

1. S. Floyd, "TCP and explicit congestion notification," *ACM Computer Communication Review*, Vol. 24, 1994, pp. 10-23.
2. K. Ramakrishnan and S. Floyd, "A proposal to add explicit congestion notification (ECN) to IP," RFC 2481, Internet Engineering Task Force, 1999.
3. S. M. S. Zabir, A. Ashir, and N. Shiratori, "Providing fair service over the Internet: an approach based on packet marking," in *Proceedings of International Conference on Internet Computing, IC'2001*, 2001, pp. 609-615.
4. S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, Vol. 1, 1993, pp. 397-413.
5. W. Stevens, "TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms," RFC 2001, Internet Engineering Task Force, 1997.
6. C. Liu and R. Jain, "Improving explicit congestion notification with the mark-front strategy," *Computer Networks*, Vol. 35, 2001, pp. 285-201.
7. G. Hasegawa, T. Matsuo, M. Murata, and H. Miyahara, "Comparisons of packet scheduling algorithms for fair service among connections on the Internet," in *Proceedings of IEEE INFOCOM'2000*, 2000, pp. 272-281.
8. T. V. Lakshman, A. Neidhardt, and T. J. Ott, "The drop from front strategy in TCP and in TCP over ATM," in *Proceedings of IEEE INFOCOM'1996*, 1996, pp. 1414-1424.
9. "UCB/LBNL/VINT network simulator – ns (version 2)," <http://www.isi.edu/nsnam/ns/>.
10. R. Jain, *The Art of Computer System Performance Analysis*, John Wiley and Sons Inc., 1991.
11. V. Jacobson, "Congestion avoidance and control," in *Proceedings of ACM SIGCOMM'88*, 1988, pp. 314-329.
12. J. H. Salim and U. Ahmed, "Performance evaluation of explicit congestion notification (ECN) in IP networks," RFC 2884, Internet Engineering Task Force, 2000.

Salahuddin Muhammad Salim Zabir joined the Department of Computer Science and Engineering of Bangladesh University of Engineering and Technology as a Lecturer

in 1995 and later got promoted as Assistant Professor in 1997. At present, he is with RIEC, Tohoku University, Japan. He received a Best Paper Award in SCI, 2001. He is a member of IEEE, BCS and BAAS.

Ahmed Ashir after receiving his PhD in 1999 from Tohoku University, Japan, worked as a JSPS Research Associate in RIEC of the same university. At present, he is with the Japan Gigabit Network (JGN) Project of Telecommunication Advancement Organization (TAO), Tohoku University Office. He received a Best Paper Award in SCI, 2001. He is a member of IEEE.

Gen Kitagata is a research associate of Research Institute of Electrical Communication of Tohoku University. He received a Dr. Eng. degree in information engineering from Tohoku University, Japan in 2002. His research interests include agent-based computing and network middleware design. He is a member of IEICE.

Takuo Suganuma is a research associate of Research Institute of Electrical Communication of Tohoku University. He received a Dr. Eng. degree from Chiba Institute of Technology in 1997. His research interests include agent-based computing and design methodology for distributed systems. He is a member of IPSJ, IEICE and IEEE.

Norio Shiratori after receiving his doctoral degree at Tohoku University, joined the Research Institute of Electrical Communication (RIEC) where he is now a professor. He has been engaged in research on distributed processing systems and flexible intelligent networks. He received IPSJ Memorial Prize Winning Paper Award in 1985, Telecommunications Advancement Foundation Incorporation Award in 1991, the Best Paper Award of ICOIN-9 in 1994, the IPSJ Best Paper Award in 1997, etc. In recognition of his outstanding contributions to the field of computer communication networks, he has been named a Fellow of the IEEE. He is also a member of IPSJ and IEICE.