TCP Performance Enhancement by Implicit Priority Forwarding (IPF) Packet Buffering Scheme for Mobile IP Based Networks

Young Sup Roh, Kyeong Hur, Doo Seop Eom, Yeonwoo Lee, and Kyun Hyon Tchah

Abstract: **The smooth handoff supported by the route optimization extension to the mobile IP standard protocol should support a packet buffering mechanism at the base station (BS), in order to reduce the degradation in TCP performance caused by packet losses within mobile network environments. The purpose of packet buffering at the BS is to recover the packets dropped during inter-subnetwork handoff by forwarding the packets buffered at the previous BS to the new BS. However, when the mobile host moves to a congested BS within a new foreign subnetwork, the buffered packets forwarded by the previous BS are likely to be dropped. This subsequently causes global synchronization to occur, resulting in the degradation of the wireless link in the congested BS, due to the increased congestion caused by the forwarded burst packets. Thus, in this paper, we propose an implicit priority forwarding (IPF) packet buffering scheme as a solution to this problem within mobile IP based networks. In the proposed IPF method, the previous BS implicitly marks the priority packets being used for inter-subnetwork handoff. Moreover, the proposed modified random early detection (M-RED) buffer at the new congested BS guarantees some degree of reliability to the priority packets. The simulation results show that the proposed IPF packet buffering scheme increases the wireless link utilization and, thus, it enhances the TCP throughput performance in the context of various inter-subnetwork handoff cases.**

Index Terms: **Mobile IP, implicit priority forwarding, packet buffering, route optimization extension, wireless link utilization.**

I. INTRODUCTION

Since 3G and beyond mobile networks started to become reality, mobile networks have been gaining momentum. Indeed, it is likely that in the future the majority of Internet users will be mobile users. As a result of the remarkable development of the world wide web (WWW) and the popularity of electronic mail (email) on the Internet; the transmission control protocol (TCP)/Internet protocol (IP) has enabled the integration of a wide range of different physical networks into a global Internet. However, in the not so distant future, it is likely that large numbers of mobile users equipped with various portable wireless IP-enabled communication devices will demand nomadic computing capabilities with seamless access to a global network of web-based mobile multimedia services. In short, these users will demand seamless mobile communication. With this in mind, the IETF (Internet engineering task force) proposed the mobile IP protocol, in order to introduce mobility to the Internet.

Mobile IP acts as an inter-subnetwork handoff protocol within an IP layer, in order to provide hosts connected to the Internet with seamless mobility without changing their IP address. The base mobile IP protocol [1] can support host mobility without any modification to existing routers or correspondent hosts (CH). The mobile IP with route optimization extension has been designed to solve the well-known triangle routing problem [2], [3] which is inherent in the base mobile IP protocol, by informing CH of the mobile host's care-of address (CoA) [4]–[6].

The major problem in mobile IP is that TCP was not specifically designed for mobile network environments. TCP interprets packet loss as a sign of network congestion, which prompts it to initiate an unnecessary congestion control function during the connection. As a result, since packet losses due to handoffs are also recognized as an occurrence of congestion [5], TCP undergoes severe performance degradation especially when a mobile host (MH) visits many subnetworks during the connection. As one of many proposed solutions to this problem, reference [6] points out that in most cases even a smooth handoff achieved through the route optimization extension of a mobile IP standard cannot prevent the degradation of TCP performance. This extension was originally intended to reduce the number of packets dropped during a handoff. Furthermore, it is also shown in [6] that by utilizing the route optimization extension, the TCP performance is occasionally worse than that achieved with the base mobile IP, when its smooth handoff fails to avoid four or more packet losses during a handoff.

It should be noted that in mobile environments, there are two causes of TCP performance degradation; that is, a high error rate on the wireless link and a handoff. In order to successfully deal with a wireless link with a high error rate, efficient retransmission over the wireless link and the avoidance of unnecessary TCP-level packet retransmissions are required. The most efficient way of accomplishing this is to have a well-tuned link-layer using forward error correction (FEC) and automatic repeat request (ARQ) techniques [7].

On the other hand, an extremely promising approach to realizing seamless handoff is to store the packets at the base station (BS) as a provision for the dropping of packets during a handoff event, i.e., the packet buffering method [6]. If the packet buffering method is successfully employed, it allows handoffs

Manuscript received February 21, 2004; approved for publication by Dong Ho Cho, Division III Editor, May 18, 2005.

Y. S. Roh is with the Department of Embedded Software at Seoul University of Venture & Information, Seoul, Korea, email: ysroh@suv.ac.kr.

K. Hur is with the Department of Computer Education at Gyeongin National University of Education, Korea, email: Khur@ginue.ac.kr.

D. S. Eom and K. H. Tchah are with the Department of Electronic Engineering at Korea University, Seoul, Korea, email: {eomds, tchah}@korea.ac.kr.

Y. Lee is with the Division of Information Engineering at Mokpo National University, Korea, email: yeonwoo.lee@gmail.com.

This research was supported by University IT Research Center Project.

to be transparent to the upper layer protocols, by enabling lost packets to be recovered during the handoff. This means that the upper layer protocols including TCP can be used on the MHs without any modification. This is very useful when considering that these protocols were designed for wired networks.

Such packet buffering methods can be implemented in either a unicast or a multicast fashion. In the former packet buffering method, only the current BS for the MH performs buffering. Immediately after the previous BS is informed of the address of the new BS, it forwards the buffered packets to the new BS to which the MH is connected after handoff. On the other hand, in the latter method of intra-subnetwork handoff, all adjacent BSs of the current BS perform packet buffering with the anticipation of the future handoff of the MH. Therefore, there is no need to forward the buffered packets to the new BS during the handoff, with the result that the handoff latency is shorter than that of the unicast-based buffering method. However, the overhead associated with buffering is increased, and more complicated multicasting routing is necessary. Such problems as this have been addressed in the split-connection protocol [8], [9] and in the hierarchical foreign agent (FA) management scheme [10]. One of our major research interests is the problem of inter-subnetwork handoff in mobile IP based networks, as described in [6]. The authors of [6] considered the possibility of making a minor modification to the route optimization extension of the mobile IP standard, in order to support the unicast-based buffering of packets at a BS without there being an associated scalability problem. They also investigated the impact of inter-subnetwork handoffs on the performance improvement of TCP.

However, both in [6] and in other works, only one TCP connection was considered. It is worth noting here that the wireless link between a CH and an MH acts as the main bottleneck link. For this reason, if other TCP connections exist in the new BS, the unicast-based packet buffering at the BS will fail to recover the packets dropped during inter-subnetwork handoff, even though the new BS employs the random early detection (RED) packet-discard scheme [11]. In such a case, the RED buffer in the new BS is likely to be fully utilized by other TCP connections already present in the new BS. Subsequently, the average queue size of the RED buffer may approach the upper threshold limit, $TH_{\max}$ [11]. Furthermore, in the case where the buffered packets are forwarded, one window of packets arrives at the congested BS within a short time when an inter-subnetwork handoff occurs. It is for this reason that, in the case of mobile IP with the route optimization extension, the packets sent by the CH, which is as yet unaware of the new care-of address, are forwarded to the new BS in addition to the buffered packets [4], [6]. Therefore, the average queue size can exceed $TH_{\max}$, due to the effect of the forwarded burst packets and, thus, the RED buffer subsequently drops all arriving packets [11]. Consequently, global synchronization may occur, resulting in even worse utilization of the wireless link.

It should be noted that without employing the buffering method, the problem mentioned above can arise when the number of packets dropped during the handoff is small and, thus, the previous FA forwards many packets to the new BS after the handoff. Therefore, addressing this particular problem, which increases in importance as the number of MHs and their handoff probability increase, is our main concern. To tackle this problem, in this paper, we propose an implicit priority forwarding (IPF) scheme, in which the previous BS marks buffered packets as priority packets during handoff, so that the RED buffer in the new BS can identify the burst packets forwarded by the previous BS. In addition, with this scheme, in order to reduce the abrupt increase in the average queue size caused by the packets forwarded by the previous BS, the RED buffer at the new congested BS does not include the priority packets in its calculation of the queue size. Therefore, it does not drop the priority packets randomly, based on the principle of the average queue size limit, during the seamless handoff. In this paper, we refer to the functionality, which performs this operation for the IPF scheme in mobile IP based networks, as the modified-RED (M-RED) buffer. We describe this M-RED buffer algorithm, as well as the BS structure of the IPF packet buffering scheme. Our proposed method is evaluated by simulation and the proposed IPF scheme is shown to increase wireless link utilization, while enhancing the TCP performance.

II. PACKET BUFFERING IN MOBILE IP BASED NETWORKS

After briefly summarizing the modification to the route optimization extension of the mobile IP standard protocol required to support unicast-based packet buffering in mobile IP based networks, we describe the implementation of the packet buffering method. In this paper, we assume that the router within each subnetwork also plays the role of the FA.

A. Route Optimization Extension of Mobile IP Standard with Packet Buffering at a BS

In the route optimization extension [4], [6], when the new FA receives the registration request message from the MH, it sends a binding update message to the old (previous) FA, in order to simultaneously deliver the new FA's care-of address and relay the registration request message. Each time the old FA receives a packet from the CH, it forwards it to the new FA. It also sends a binding warning message to the home agent (HA). As soon as the HA receives the binding warning message, it sends a binding update message to the CH, in order to inform it of the new care-of address. After receiving the binding update message, the CH can send packets to the new FA instead of the old FA. Thus, the above process, which is termed a smooth handoff, can reduce the number of packets dropped during the inter-subnetwork handoff. However, until the old FA receives the binding update message from the new FA, there is still a possibility of the in-flight packets being dropped. Therefore, in most cases, due to this packet dropping, smooth handoff cannot prevent the degradation of the TCP performance in the event of a handoff [6].

To achieve seamless inter-subnetwork handoff, in the case where only the current FA performs packet buffering, the current FA sends the buffered packets to the new FA when it receives the binding update message from the new FA. However, this leads to a scalability problem, because the FA must cover all of the TCP connections of the MHs communicating with IP hosts outside the subnetwork. Furthermore, this method inevitably leads to the new FA receiving duplicate packets due to

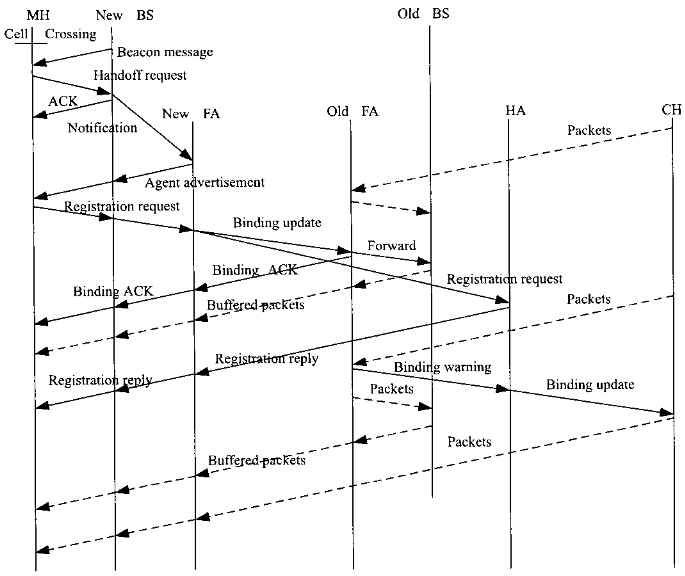


Fig. 1. The route optimization extension with packet buffering at a BS.

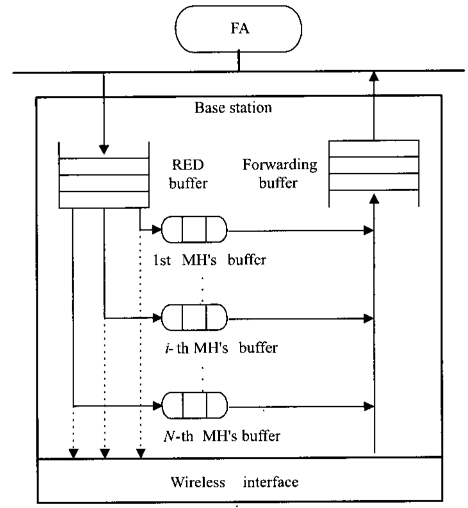


Fig. 2. Structure of packet buffering function block at BS.

buffering, which trigger an unnecessary fast retransmit [6]. To avoid this, ideally the FA should manage the buffer used for recovering packet losses due to inter-subnetwork handoff, by using TCP ACK packets. However, monitoring TCP ACK packets would induce additional overhead at the FA. On the other hand, in the case where only the current BS performs packet buffering, the scalability problem is alleviated, since the BS covers a smaller number of the TCP connections of the MHs than the FA. Moreover, if buffering is performed at the BS, the ARQ protocol of layer 2 can be used to recover the packets dropped during handoff [7], [12], where the ARQ protocol uses the retransmission buffer to perform error control on the wireless link at the data link-layer. In this case, the old BS can drop the acknowledged packets from its buffer, and thus only those packets which are not acknowledged by the MH are forwarded to the new BS after handoff. This ensures that the MH never receives duplicate packets due to buffering, when the BS performs the packet buffering.

The above reasoning formed the basis for the proposed buffering method for inter-subnetwork handoff described in [6]. This method introduced the forward message shown in Fig. 1, which serves as a link-layer control message between the FA and the BS. It contains the link-layer address of the MH, which is used to identify the buffered packets for the MH. When the old FA receives the binding update message from the new FA, it sends the forward message to the old BS, in order to request forwarding of the buffered packets. When the old FA receives these buffered packets, it forwards them to the new FA by re-encapsulating them with the address of the new FA [4], [6].

This minor modification to the route optimization extension is done to enable packet buffering to be performed at the BS in the case of inter-subnetwork handoff. On the other hand, in Fig. 1, a local handoff protocol is incorporated into the route optimization extension, in order to reduce the duration of the period during which packet loss is susceptible to occur in the case of inter-subnetwork handoff [1], [6]. Furthermore, common problems such as those associated with the buffer size and re-ordering of packets may arise when the unicast-based packet buffering method in mobile IP based networks is taken into account. These problems are presented along with their solutions in [6]. Fig. 2 shows the functional block structure of the BS required to support the packet buffering method, where the BS maintains one buffer per MH instead of maintaining one buffer per TCP connection. This resolves the possible scalability problem, while potentially enabling the network to support other transport protocols such as UDP more easily.

B. Implementation of the Packet Buffering Method in Mobile IP Based Networks.

The buffering method can be implemented at the BS by maintaining one buffer per MH, rather than maintaining one buffer per TCP connection. This resolves the possible scalability problem, while potentially allowing the network to support other transport protocols such as UDP more easily. In addition, the BSs should maintain a table containing a list of the currently attached MHs. When the old BS receives the forward message which contains the address of the MH that has moved to another wireless cell, it simply forwards the appropriate buffered packets to the MH and then deletes the entry for the MH from the table. However, the packet re-ordering problem, which causes duplicate ACK packets to be sent, also needs to be considered [6]. When the CH sends packets to the MH without knowledge of the new care-of address, the route optimization extension causes the old FA to forward these packets to the new FA. However, if these packets from the CH arrive at the old FA before it completes the forwarding of the packets buffered by the old BS, these packets could arrive at the MH in the wrong order, and this could cause the TCP performance to be worse than that observed when packet buffering is not supported. This is a direct consequence of the duplicate ACK packets caused by the re-ordering of packets [6]. To solve this problem, the old FA forwards the packets from the CH to the new FA through the old BS, as shown in Fig. 3. Namely, the old FA sends these packets to the old BS, and then the old BS places them into the buffer of the MH.

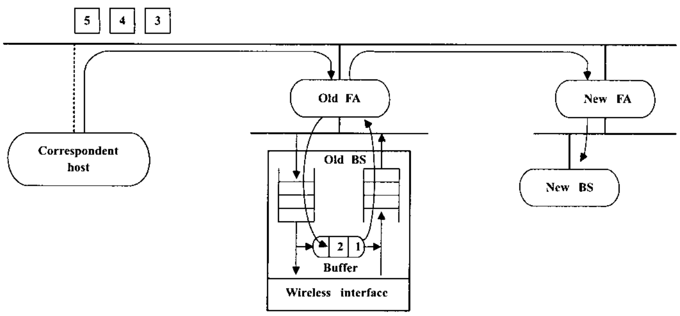


Fig. 3. Illustration of packet delivery sequence used to prevent the packets from arriving in the wrong order.

Finally, the packets in the buffer are forwarded to the new FA through the old FA. In this way, all of the packets for the MH are forwarded to the new FA in the correct order.

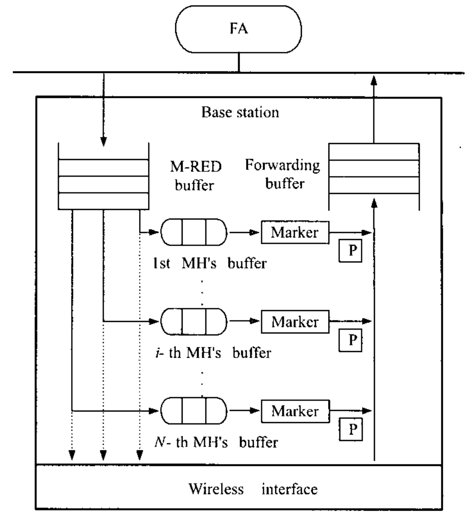


Fig. 4. Structure of the IPF packet buffering function block at the BS.

III. PROPOSED IMPLICIT PRIORITY FORWARDING (IPF) SCHEME

Although the use of the RED buffer provides quite good handoff performance, it cannot prevent the large decrease in TCP throughput caused by global synchronization, in the case where TCP connections already exist in the new BS. The RED buffer was originally proposed to prevent global synchronization from occurring at the router, where a large number of connections flow. However, the number of TCP connections in the cellular BS is much smaller than that in the router. Also, unlike in the case of the router, where a new TCP connection causes it to increase its window size, the new BS receives a maximum of one window of packets forwarded by the packet buffering method. Therefore, the RED buffer in the new BS cannot prevent global synchronization caused by an abrupt increase in the number of random packet drops, when TCP connections already exist in the new BS. To mitigate the undesirable effect of global synchronization at the RED buffer in the case of an inter-subnetwork handoff, it is necessary to prevent the average queue size from increasing abruptly, in spite of the burst arrivals of the packets forwarded by the old BS which are likely to occur.

When the new BS is heavily congested (i.e., the average queue size of the RED buffer approaches the threshold $TH_{\max}$ [11]), it is better to give up the forwarding of the buffered packets. This is because the BS employing the RED buffer drops all arriving packets when the average queue size exceeds $TH_{\max}$. Thus, it is necessary to include a new option field in the binding update message for the purpose of indicating whether the forwarding of the buffered packets is allowed by the new BS or not. However, this method would necessitate a modification to the mobile IP with route optimization extension protocol. Furthermore, it cannot prevent the performance degradation of the TCP handoff connection.

Fig. 4 shows the proposed IPF scheme, which is designed to improve the TCP performance of mobile IP based networks with packet buffering. To prevent global synchronization from occurring at the RED buffer during an inter-subnetwork handoff event, as explained above, it is necessary to prevent the average queue size from increasing abruptly, in spite of the burst arrivals of the buffered packets forwarded by the old BS which are likely to occur. In addition, for seamless handoff, it is necessary for the packets buffered at the old BS during the handoff to be delivered safely, with the least possible number of packets being dropped.

In order to satisfy these requirements, the proposed IPF scheme adopts the following new principle. The old BS marks a buffered packet as a priority packet during the inter-subnetwork handoff and forwards it, so that the M-RED buffer in the new BS can identify the burst packets forwarded by the old BS. In this way, the M-RED buffer at the new congested BS can exclude these priority packets in its calculation of the queue size, so as to reduce the abrupt increase in the average queue size, which is caused by the packets forwarded by the old BS. In addition, for the sake of achieving seamless handoff, the new BS does not drop these priority packets randomly based on such RED buffer algorithm decision criteria as the average queue size limit. This prioritized packet forwarding is performed in an implicit manner, i.e., marking and forwarding the priority packet without the extra cost that would otherwise be incurred by the use of an additional queue.

Figs. 5 and 6 show the algorithms used for the management of the RED and M-RED buffers, respectively. As shown in Fig. 5, the RED buffer algorithm performs random dropping of packets according to the value of *avg_q* (the average number of packets in the RED queue) [11], while the M-RED buffer algorithm shown in Fig. 6 drops only non-priority packets randomly according to the value of *avg_q_np* (the average number of non-priority packets in the M-RED queue). The operation of the M-RED buffer algorithm is identical to that of the RED buffer algorithm when there are no arriving priority packets generated during the inter-subnetwork handoff. The M-RED buffer in the new BS guarantees the forwarding of arriving priority packets marked by the old BS, by applying an implicit protection policy. That is to say, the M-RED buffer in the new BS does not include the priority packets in its calculation of *q_size_np* (the queue size of the non-priority packets), in order to prevent any abrupt increase in the value of *avg_q_np* caused by the forwarded burst priority packets. Thus, the value of *avg_q_np* does not increase

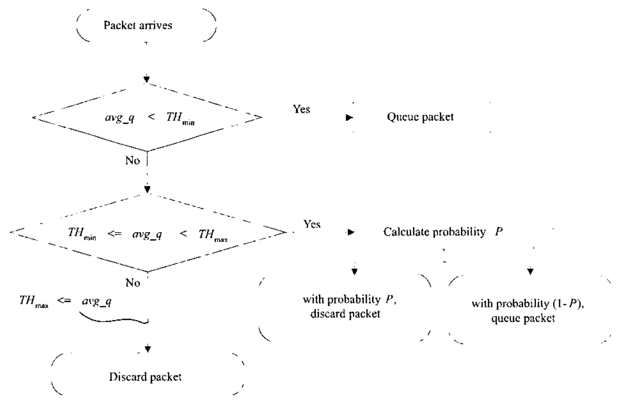


Fig. 5. RED buffer algorithm.

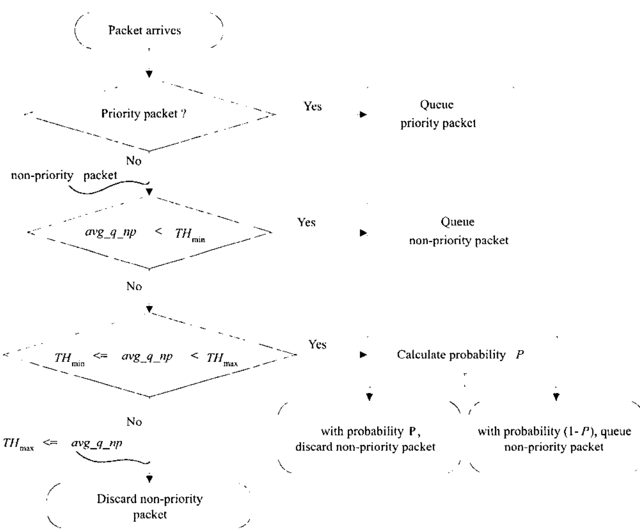


Fig. 6. M-RED buffer algorithm.

abruptly, even though the forwarded burst packets marked by the old BS arrive at the new BS. Therefore, a lower number of non-priority packets of TCP connections already present in the new BS which uses the M-RED buffer are dropped than would normally be in the case of a congested RED buffer. In addition, the packets that were buffered and forwarded by the old BS during handoff are implicitly protected for the sake of achieving seamless handoff, since the M-RED buffer does not drop these priority packets through the operation of the mechanism depending on the value of the *avg_q_np* parameter. This implies that the proposed IPF scheme can provide the marked priority packets with some form of delivery guarantee, i.e., the priority packets buffered at the old BS are implicitly guaranteed to be delivered during handoff, with minor overhead being incurred only during the inter-subnetwork handoff period.

In the proposed IPF scheme, the TCP handoff connections receive the packets that are buffered and forwarded by the old BS seamlessly during the inter-subnetwork handoff without any of them being dropped. Furthermore, the increase in the value of *avg_q_np* of the M-RED buffer is reduced in proportion to the amount of burst packets forwarded from the old BS, which in turn reduces the number of packets which are dropped at the

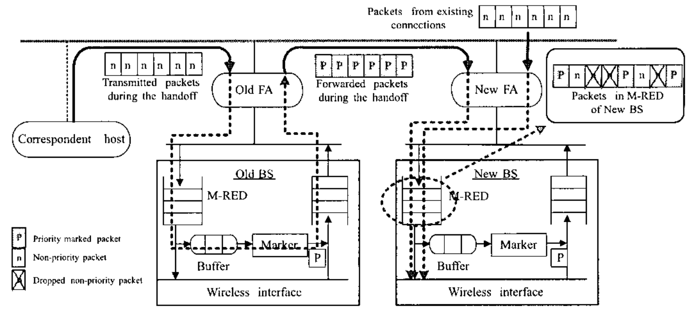


Fig. 7. Illustration of the proposed IPF packet buffering scheme with the M-RED algorithm.

M-RED buffer during the inter-subnetwork handoff. Therefore, the effect of global synchronization caused by the forwarded burst of packets during the inter-subnetwork handoff is alleviated, due to the fact that the number of packets dropped for all of the TCP connections in the new BS is reduced in comparison to that which occurs in the case where the RED buffer is used. Consequently, both the TCP handoff connections of the MH (which has moved to the new BS) and the already existing TCP connections of the new BS achieve performance improvements when the IPF scheme is used.

Fig. 7 shows an illustration of the proposed IPF packet forwarding scheme using the M-RED buffer algorithm, which demonstrates how the marked priority packets forwarded by the old BS are handled in the new BS. In the case where TCP connections already exist in the new FA, the function of the M-RED buffer with its selective dropping policy is carried out such that the priority packets forwarded from the old BS are queued with high reliability. Compared to these priority packets, non-priority packets that do not need to be forwarded in a secure manner are dropped randomly by the RED buffer algorithm, based on the average queue size of the non-priority packets (i.e., *avg_q_np*). However, the priority packets marked for seamless inter-subnetwork handoff purposes are not counted in *avg_q_np* and, consequently, these packets are not included in the random dropping action. Without any loss of generality, we can assume that the number of priority marked packets during the inter-subnetwork handoff is low enough to be handled by the M-RED buffer, with a strong guarantee of their being delivered, and that this number does not exceed its maximum queue size. In the worst case scenario, in which all of the arriving packets are priority packets, even these priority packets can be dropped due to buffer overflow at the M-RED buffer.

From the viewpoint of seamless handoff, the IPF scheme is a viable method in comparison with the conventional RED buffer scheme, unless the priority packet burst overflows the M-RED buffer of the BS. On the other hand, when considering the problem of congestion avoidance, although the IPF scheme is used during the inter-subnetwork handoff event, a maximum of one window of burst priority packets arrives at the congested new BS within a short time during the packet forwarding period. Therefore, the real queue size of the M-RED buffer could be increased abruptly. In this scenario, the IPF scheme cannot always prevent the incidence of an abrupt increase in the amount of congestion, because the *avg_q_np* parameter of the M-RED buffer reacts

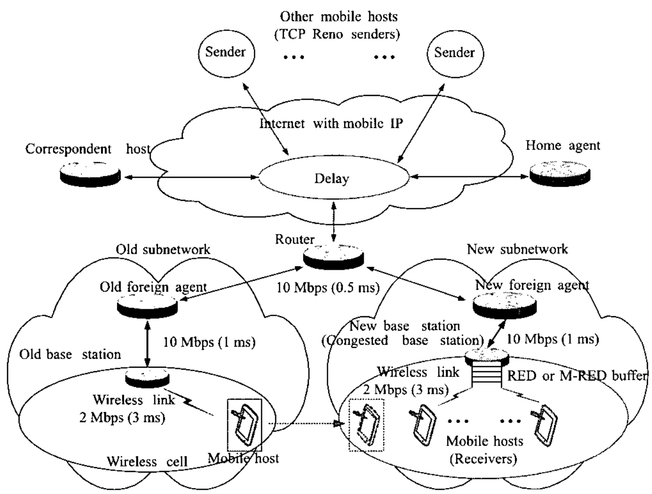


Fig. 8. Simulated network topology.

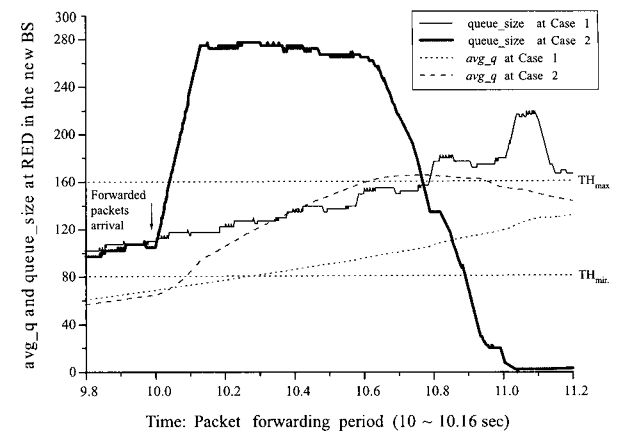


Fig. 9. Burst arrival behavior of the packets forwarded by the old BS.

slowly to this type of scenario. However, this helps to clarify the role of the IPF scheme, which is to avoid any abrupt increase in the amount of random dropping performed by the RED buffer that might occur immediately following such an event. Thus, the IPF scheme can reduce the impact of global synchronization on the RED buffer during inter-subnetwork handoff.

There exist users who desire preferential treatment. So, if it is possible for the sender to generate priority packets, some specific mechanisms are required in the router connecting the old FA to the new FA, in order to deal with this situation. These mechanisms should monitor the wrongly marked priority packets arriving from the core network, and then drop them or remark them as non-priority packets, before they arrive at the FAs.

IV. SIMULATION RESULTS AND DISCUSSIONS

A. Simulation Model

Fig. 8 shows the mobile IP based network topology considered in our simulation for the route optimization extension with packet buffering at the BS. The model includes an MH, which moves between two wireless cells within different subnetworks, while maintaining a TCP connection with a fixed CH, and the bandwidth of the wireless part in the new BS is shared among nine TCP connections. In this case, the buffer in the new BS is likely to be fully utilized by those TCP connections that are already present in the new BS, given that the bottleneck is assumed to be the wireless part. The RTTs (round trip times) of the TCP connections in the new BS are 86, 88, 90, 92, 94, 96, 98, 100, and 102 ms, respectively.

In our simulation model, the delay taken to send a packet between adjacent FAs is set to 1 ms, based on the use of 10 Mbps Ethernet. Similarly, the delay taken to send a packet between an FA and a BS is also set to 1 ms. We set the bandwidth of the wireless part in the BS to 2.0 Mbps, the delay taken to send a packet between the BS and the MH to 3 ms, and the broadcasting period for the beacon messages to 50 ms [6], [12]. We set the delay taken to send a packet between an FA and an HA to 60 ms, the delay between an HA and a CH to 60 ms, and the RTT between the CH and the MH which handoffs to the new BS is set to 88 ms. In our simulations, each CH transmits packets to the MH using TCP Reno. The packet size and maximum window size for the TCP connection are set to 512 bytes and 64 packets, respectively. We assume that there is no packet loss caused by transmission errors.

B. Effect of the Packet Buffering Method in the Simple Inter-Subnetwork Handoff Case

We implemented a RED queue with a queue length of 300 packets and parameters ($TH_{\min}$, $TH_{\max}$, $P_{\max}$, w_q) = (80, 160, 0.02, 0.002) in the new BS [11]. The burst traffic arrival behavior of the packets forwarded by the old BS for the following two cases is taken into consideration and demonstrated in Fig. 9.

Case 1 : We consider the case where the packets corresponding to a new mobile connection start to arrive at the new BS at 10 sec. That is, an MH in the new BS newly sets up a TCP connection with a CH through the new BS, and the CH transmits packets to the MH gradually using TCP Reno.

Case 2 : The packets forwarded from the old BS of a mobile handoff connection start to arrive at the new BS at 10 sec. Here, one window of packets arrives at the new BS within a short time, or in the worst case scenario, a number of packets equal to the maximum window size arrives at the new BS.

Fig. 9 shows the queue size and average queue size variations of the RED buffer for the above two cases. The RED buffer performs low pass filtering, using an exponential weighted moving average (EWMA), to calculate the average queue size, *avg_q* [12], which is used to randomly drop packets. Since packets are dropped in proportion to the average queue size, when it increases beyond the value of $TH_{\min}$, more random dropping of packets occur in case 2. We can see that due to this random dropping of packets, the queue size declines abruptly for a considerable period of time in case 2. Although the queue size of the RED buffer can be maintained at a low level by using this random dropping mechanism, the steep increase in the queue size caused by the forwarded packets results in a marked increase in the average queue size, which is much larger than that observed in case 1. Thus, global synchronization caused by an abrupt increase in the queue size is unavoidable in case 2, as shown in Fig. 9. This means that when using the packet buffering method, global synchronization occurs due to the ran-

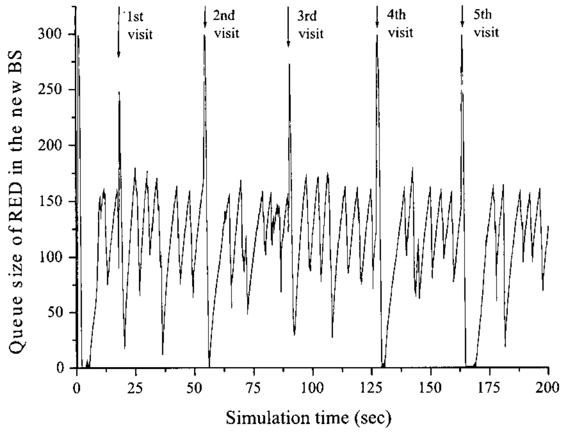


Fig. 10. Abrupt increase in RED queue size at each visit of the MHs.

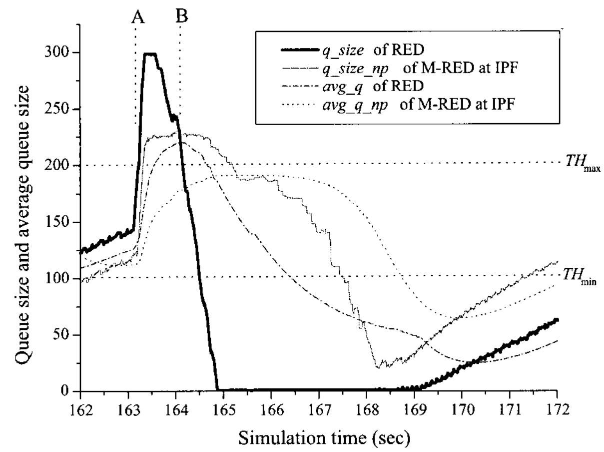


Fig. 11. Avoidance of global synchronization by using the IPF scheme at 5-th visit of the MH.

dom dropping of packets and, thus, the link utilization of the wireless link becomes low. The actual queue size around 11 sec is zero. Another important point which should be noted is that compared to routers, a much smaller number of TCP connections pass through BSs. Thus, it is likely that the large number of packets which are randomly dropped from the RED buffer at the BS have an effect on all of the TCP connections, ultimately leading to global synchronization.

B.1 Effect of Packet Buffering Scheme on Queue Size Variation

In order to show the efficiency of the proposed IPF scheme and the improvement that it provides, our simulation assumes that there are seven TCP connections already present in the new BS, and that there are five MHs, which visit the new BS a total of five times during a period of 200 sec, with all of the TCP connections being maintained during this period. The RED queue with a queue length of 300 packets has $TH_{\min}$ set to 100 and $TH_{\max}$ set to 200 in this case. Fig. 10, which shows the variation in the queue size of the RED buffer in the new BS, demonstrates that the burst arrivals of the forwarded packets of the handoff TCP connections increase the queue size abruptly at each visit. Fig. 11 shows a comparison of the variation in the queue size of the IPF buffering scheme with the M-RED buffer and the conventional scheme with the RED buffer. In Fig. 11, it can be seen that, unlike in the case of the RED buffer, the proposed IPF scheme prevents the average queue size from increasing abruptly due to the burst arrivals of the buffered packets forwarded by the old BS. Therefore, when using the IPF scheme, the average queue size does not exceed $TH_{\max}$, with the result that this scheme prevents global synchronization from occurring in the RED buffer during the inter-subnetwork handoff. This is the main feature of the proposed IPF packet buffering scheme.

As explained above, the value of *avg_q_np* for the M-RED buffer is calculated based on the corresponding value of *q_size_np*. In Fig. 11, it can be seen that in comparison with the *q_size* and *avg_q* of the RED buffer case, the *q_size_np* and *avg_q_np* parameters of the M-RED buffer in the IPF scheme are less sensitive to the abrupt increase in the number of burst packet arrivals, since the M-RED buffer does not include the priority packets in its calculation of *q_size_np* and *avg_q_np*. However, it is noticed that the actual queue size of the M-RED buffer after the MH's visit is similar to that of the RED buffer from point A to point B in Fig. 11, since the actual queue of the M-RED buffer contains priority packets as well as non-priority packets. However, we should also notice that, in the proposed IPF scheme, the priority packets are not included in the random dropping mechanism of the RED algorithm, with only non-priority packets being randomly dropped when the average queue size is above $TH_{\min}$. In the case of the RED buffer, the sharp increase in the queue size caused by the forwarded packets results in a remarkable increase in the average queue size, which in turn causes large random droppings of packets. As previously mentioned, compared to routers, a much smaller number of TCP connections pass through BSs. Thus, it is probable that the large random dropping of packets from the RED buffer at the BS affects all of the TCP connections passing through, ultimately resulting in global synchronization. On the other hand, after point B, the proposed IPF scheme does not cause global synchronization, maintaining a stable average queue size. Even though the average queue size of the M-RED buffer (*avg_q_np*) does not reflect the existing priority packets, after point B the priority packets have already left the queue.

B.2 Wireless Link Utilization Performance of the Proposed IPF Packet Buffering Scheme

Figs. 12 and 13 show a comparison of the wireless link utilization for the above two schemes (the IPF buffering scheme with the M-RED buffer and the conventional scheme with the RED buffer). In Fig. 12, the wireless link utilization for the two schemes is simulated for 200 sec for various numbers of visits to the congested new BS, viz., 3, 4, 5, 6, 7, and 8, in the situation where the five MHs handoff to the new BS. In Fig. 13, on the other hand, the wireless link utilization is simulated with a different number of MHs handing off to the new BS, viz., 3, 4, 5, 6, and 7, where each MH visits the new BS five times during a total duration of 200 sec. It can be observed from the simulation results in Figs. 12 and 13 that the proposed IPF scheme with M-RED buffering increases the wireless link utilization, irrespective of the number of MHs and their handoff probability, as compared to when the RED buffer is used. Based on these simu-

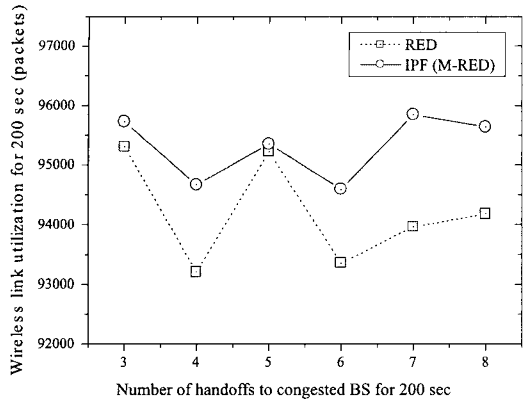


Fig. 12. Wireless link utilization with different numbers of handoffs to the congested BS.

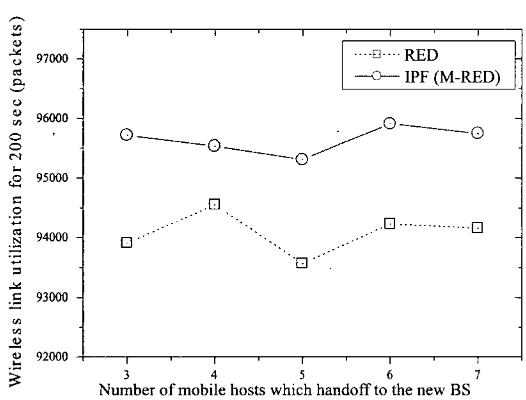


Fig. 13. Wireless link utilization with different numbers of handoff MHs.

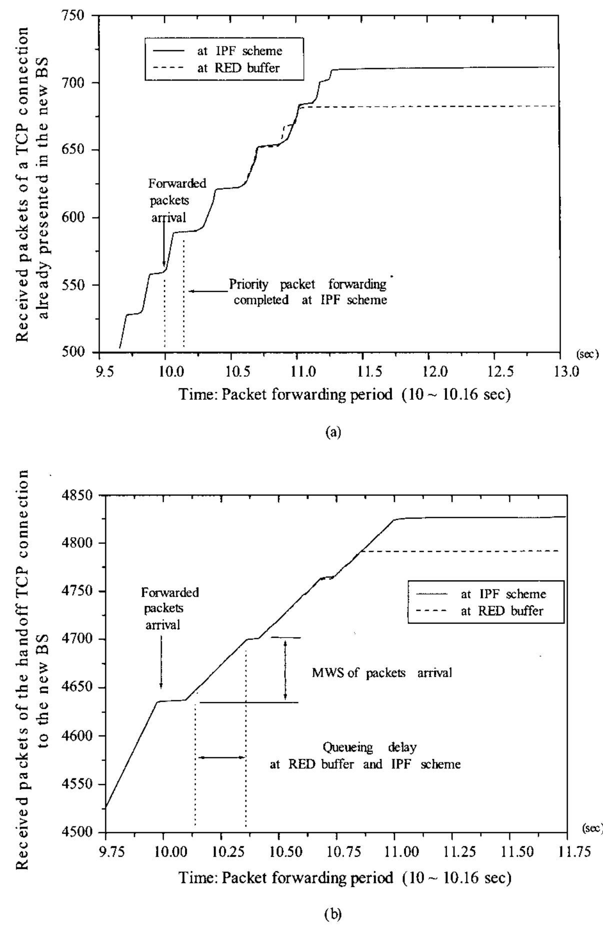


Fig. 14. Comparison of TCP performance in the new BS for the two schemes in the case of an inter-subnetwork handoff: (a) TCP performance of a TCP connection already present in the new BS, (b) TPC performance of the handoff TCP connection to the new BS.

lation results, we believe that the proposed IPF packet buffering scheme can guarantee the controllability of packet forwarding and buffering at a congested BS, offering enhanced TCP performance, the avoidance of global synchronization and high wireless link utilization.

B.3 TCP Performance of the Proposed IPF Packet Buffering Scheme

Fig. 14 compares the TCP performance in the new BS using the two schemes in the case of an inter-subnetwork handoff. Fig. 14(a) shows the variation in the throughput of a TCP connection already present in the new BS for the two schemes. Fig. 14(b) shows the variation in the throughput of the handoff TCP connection to the new BS for the two schemes. In Fig. 11, the period of global synchronization during inter-subnetwork handoff is made shorter by using the IPF scheme. As shown in Fig. 14(a), in the case of the IPF scheme, the already existing TCP connection of the new BS receives more packets than it would in the case where the RED buffer is employed, due to the reduction in the amount of packet dropping afforded by the IPF scheme using the M-RED buffer. Furthermore, as shown in Fig. 14(b), the new BS also receives packets via the handoff TCP connection at a much higher rate, in the case of the IPF scheme. This is because, when using the IPF scheme to accomplish seamless handoff, it receives the buffered and forwarded packets from the old BS during the handoff without dropping any of them. This improvement in the TCP performance afforded by the IPF scheme could have been inferred from the result shown in Fig. 11.

C. TCP Performance of the Proposed IPF Scheme for Various Inter-Subnetwork Handoffs

In this subsection, we investigate the influence of packet buffering on the TCP performance in the new BS when multiple TCP connections perform inter-subnetwork handoffs with various inter-handoff-arrival intervals, as shown in Fig. 15. For the cases shown in Fig. 15, we observe the TCP performance in the new BS for a period of 300 sec. In this simulation, unlike in the simulations described in Subsection IV-B, there are four MHs that visit the new BS three times each, while each MH maintains one TCP connection. The RTTs of the four MH TCP connections are 88, 90, 92, and 94 ms, respectively.

In each case, the three visits of the first MH occur at 42.9 sec, 128.6 sec, and 214 sec, respectively, and those of the other three MHs follow with a specified inter-handoff arrival time. For example, in case 1, the three visits of the second MH occur at 62.9 sec, 148.6 sec, and 234 sec, respectively. That is, case 1 represents a uniform inter-subnetwork handoff-arrival distribution over a period of 300 sec at the new BS. On the other

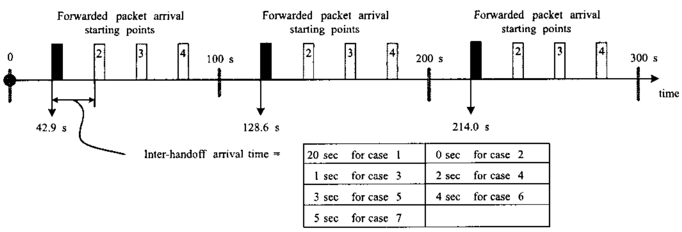


Fig. 15. Handoff arrival distributions considered in the TCP performance simulation (3 forwarded-packet-arrival starting points and 7 different inter-handoff arrival times).

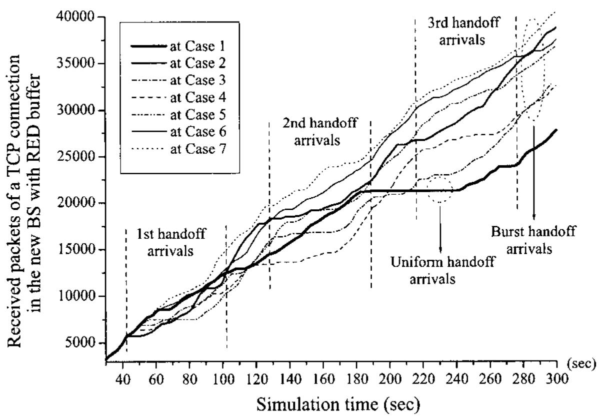


Fig. 16. TCP throughput performances of a TCP connection already present in the new BS with the RED buffer algorithm for the multiple inter-subnetwork handoff cases.

hand, the other cases shown in Fig. 15 represent the burst inter-subnetwork handoff-arrival distributions over the period of 300 sec at the new BS. Each case has a different inter-handoff-arrival time ranging from 0 to 5 sec.

For the above cases, Fig. 16 shows the TCP throughput performance of a TCP connection, which is already present in the new BS. We can see that the inter-subnetwork handoffs of each visit of the four MHs reduce the TCP throughput. In the case of the RED buffer shown in Fig. 16, we can see that the TCP throughput in the case of uniform handoff arrivals is lower than that of the other cases. This result implies that the burst packets are dropped more frequently in the case of uniform handoff arrivals than in the case of burst handoff arrivals.

In Fig. 16, we can see that for the general multiple inter-subnetwork handoff case, the burst arrival characteristic of the forwarded packets seriously impacts the TCP performance at the RED buffer. Although the RED buffer provides moderate TCP throughput performance depending on the characteristic of the inter-subnetwork handoff arrivals, i.e., uniform or burst arrival, the variation in the TCP performance dictates the need for improvement. This means that the RED buffer should be employed within the BS, where a smaller number of TCP connections flow through the mobile IP based networks compared to the router. However, as shown in Fig. 11, the RED buffer cannot prevent the reduction in the TCP throughput that occurs when an inter-subnetwork handoff TCP connection is added at the BS, which is larger than that which occurs in the case where a new TCP connection is added at the BS.

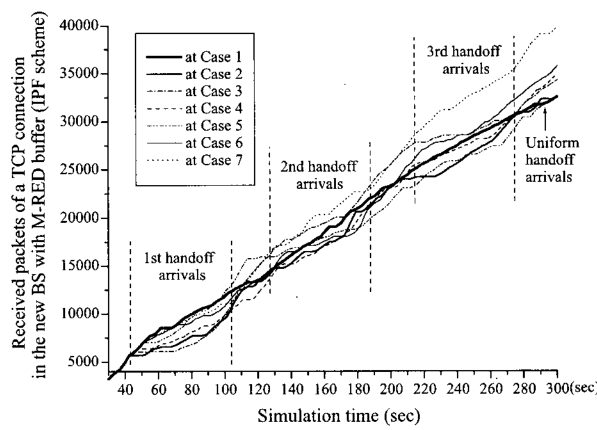


Fig. 17. TCP throughput performances of a TCP connection already present in the new BS with the M-RED buffer algorithm (IPF scheme) for the multiple inter-subnetwork handoff cases.

For the multiple inter-subnetwork handoff cases shown in Fig. 15, we show the improvement in the TCP performance obtained when using the IPF scheme. In Fig. 17, it can be seen how the throughput of a TCP connection already present in the new BS varies in each multiple handoff case when using the IPF scheme. As explained above, the IPF scheme slows the increase in the value of *avg_q_np* in proportion to the amount of burst priority packets forwarded to the TCP handoff connections in the new BS. This reduces the number of packets which are dropped according to the *avg_q_np* parameter of the M-RED buffer. For this reason, as shown in Fig. 17, it is noticed that the IPF scheme reduces the decrease in the TCP throughput as compared to the RED buffer for the multiple inter-subnetwork handoff cases. Furthermore, the difference in the TCP throughput between the cases of uniform and burst handoff arrivals is much smaller than that observed for the RED buffer. Hence, regardless of the characteristic of the inter-subnetwork handoff arrivals, the IPF scheme provides stable, fair, and better TCP throughput performance than the RED buffer.

When considering the average values and standard deviations of the TCP throughputs shown in each figure, we can see that the IPF scheme considerably reduces the standard deviation value for the handoff cases, as compared to those observed with the M-RED buffer, (i.e., IPF with 37 kbps and RED with 60 kbps). This result indicates that the IPF scheme guarantees an improvement in the TCP throughput, regardless of the characteristic of the inter-subnetwork handoff arrivals.

V. CONCLUSION

In this study, we showed that when an inter-subnetwork handoff of a TCP connection occurs at the congested new BS in a mobile IP based network with packet buffering at the BS, the burst arrivals of the packets forwarded from the old BS sharply increase the average queue size of the RED buffer. Global synchronization subsequently occurs and, thus, the utilization of the wireless link deteriorates. In this paper, we proposed the implicit priority forwarding (IPF) packet buffering scheme as a solution to this problem. The proposed IPF packet buffering scheme assures the implicit protection of the priority marked forwarded packets (i.e., the packets buffered at the old BS during the handoff) in order to guarantee seamless handoff, by using the marking process and the M-RED buffer algorithm, with only minor overhead being incurred, and this only during inter-subnetwork handoff. In addition, the simulation results show that the IPF scheme prevents situations from arising, wherein the average queue size increases abruptly due to the burst arrivals of the buffered packets forwarded from the old BS. Thus, the IPF scheme relieves the effect of global synchronization occurring in the RED buffer during inter-subnetwork handoff, in such a way that the IPF scheme increases wireless link utilization, enhances TCP throughput performance, and increases the fairness of all of the connections with low standard deviation.

REFERENCES

[1] C. E. Perkins, "IP mobility support," *IETF RFC 2002*, Oct. 1996.

[2] C. E. Perkins, "Mobile IP," *Int. J. Commun. Syst.*, pp. 3–20, 1998.

[3] C. E. Perkins, "Mobile IP," *IEEE Commun. Mag.*, pp. 84–99, 1997.

[4] C. E. Perkins and D. E. Johnson, "Route optimization in mobile-IP," draft-ietf-mobileip-optim-10.txt, 2000.

[5] R. Caceres and L. Iftode, "Improving the performance of reliable transport protocols in mobile computing environments," *IEEE J. Select. Areas Commun.*, vol. 13, no. 5, pp. 100–109, 1995.

[6] D. S. Eom, H. S. Lee, M. Sugano, M. Murata, and H. Miyahara, "Improving TCP handoff performance in mobile IP based networks," *Computer Commun.*, vol. 25, pp. 635–646, 2002.

[7] H. Balakrishanan, V. N. Padmanabhan, S. Sehan, and R. H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 756–769, 1997.

[8] A. Bakre and B. R. Badrinath, "I-TCP: Indirect TCP for mobile hosts," in *Proc. Int. Conf. Dist. Computing Syst.*, 1995, pp. 136–143.

[9] R. Yavaatkar and N. Bhagwat, "Improving end-to-end performance of TCP over mobile internetworks," in *Proc. Workshop Mobile Computing Appl.*, 1994, pp. 146–152.

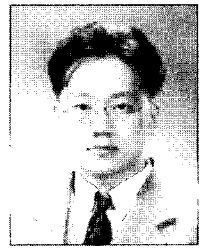
[10] C. E. Perkins and K. Y. Wang, "Optimized smooth handoffs in mobile IP," in *Proc. IEEE Int. Symp. Computers Commun.*, 1999, pp. 340–346.

[11] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Networking*, vol. 1, no. 4, pp. 397–413, Aug. 1993.

[12] R. Caceres and V. Padmanabhan, "Fast and scalable handoffs for wireless internetworks," in *Proc. ACM Mobicom'96*, 1996, pp. 56–66.

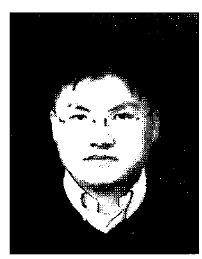
[13] D. Clark and W. Fang, "Explicit allocation of best effort packet delivery service," *IEEE/ACM Trans. Networking*, vol. 6, no. 4, pp. 362–373, Aug. 1998.

Young Sup Roh received the B.S. degree in Electronics Engineering from Inha University, Inchon, Korea in 1988 and M.S. degree from Korea Advanced Institute of Science Technology (KAIST) in 1996. In 2005, he received the Ph.D. degree in Department of Electronics Engineering from Korea University, Seoul, Korea. From November 1987 to March 1998, he was a Senior Researcher at the PIC Gr. of LG Electronics Co. Ltd., and from March 1998 to February 2001, he was an Associate Professor of Chung Kang College of Cultural Industries, Korea. From February 2001 to February 2005, he was a CTO at the Smart phone research center of Cyberbank Co. Ltd., Seoul, Korea. Since March 2005, he has been a professor in the Department of Embedded Software at Seoul University of Venture & Information. His research interests include; mobile computing, signal processing, ubiquitous networking, and future mobile IP networks.

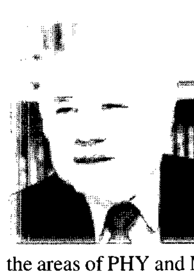


Kyeong Hur is currently a Professor in the Department of Computer Education at Gyeongin National University of Education, Korea. He was senior researcher with Samsung Advanced Institute of Technology (SAIT), Korea from September 2004 to August 2005. He received a M.S. and Ph.D. in Department of Electronics and Computer Engineering from Korea University, Seoul, Korea, in 2000 and 2004, respectively. His research interests include; computer network designs, next generation Internet, Internet QoS, and future All-IP networks.

Doo Seop Eom received the B.S. and M.S. degrees in Electronics Engineering from Korea University, Seoul, Korea in 1987 and 1989, respectively. In 1999, he received the Ph.D. degree in Information and Computer Sciences from Osaka University, Osaka, Japan. He joined the Communication Systems Division, Electronics and Telecommunications Research Institute (ETRI), Korea, in 1989. From September 1999 to August 2000, he was an Associate Professor of Wonkwang University, Korea. Since September 2000, he has been an Associate professor in the Department of Electronics Engineering at Korea University. His research interests include; communication network design, Bluetooth, ubiquitous networking, and Internet QoS.



Yeonwoo Lee is currently a Professor in the Division of Information Engineering at Mokpo National University, Korea. He was senior researcher with Samsung Advanced Institute of Technology (SAIT), Korea from January 2004 to August 2005. He was a research fellow with the School of Electronics and Engineering at the University of Edinburgh, UK, from Oct. 2000 to Dec. 2003. He joined the core 2 team at the Mobile VCE (Virtual Centre of Excellence in Mobile and Personal Communications Ltd.) programmer in UK, from Oct. 2000 to Dec. 2002. He received a M.S. and Ph.D. in Department of Electronics Engineering from Korea University, Seoul, Korea, in 1994 and 2000, respectively. His research interests include; wireless multimedia mobile telecommunication systems, radio resource management, radio resource metric estimations, with specific interest in their applicable issues to 3G and beyond mobile communication systems.



Kyun Hyon Tchah received the B.S. and Ph.D. degrees in Electrical Engineering from Seoul National University, Seoul, Korea in 1965 and 1976, and M.S. degree in Electrical Engineering from Illinois University. Since 1978, he has been a Professor in the Department of Electronics Engineering at Korea University. He was the President of the Korean Institute of Communication Sciences (KICS) from January 1998 to December 1998, and the chairman of IEEE Seoul Section from 2001 to 2002. He has been the Fellow in IEE since 2000. His current research interests are in the areas of PHY and MAC layer solutions for wireless communication systems, wireless LAN/PAN, and ubiquitous networking.