An Enhanced Buffer Management Scheme for Fast Handover Protocol*

Wei-Min Yao, and Yaw-Chung Chen
Department of Computer Science and Information Engineering
National Chiao Tung University
1001 Ta Hsueh Road, Hsinchu 30050, Taiwan
wmyao@csie.nctu.edu.tw, ycchen@csie.nctu.edu.tw

Abstract

The integration of IEEE 802.11 wireless LAN and Mobile IP technologies offers an affordable and high bandwidth solution for host mobility. When applying the Hierarchical Mobile IP for Fast Handover, the performance bottleneck on Mobile IP and potential disconnection during handoff period can be greatly improved. However, the Fast Handover protocol suffers several problems such as scalability and QoS support for lacking a buffer management mechanism. In this work, we propose an enhanced buffer management scheme for Fast handover. By means of the proposed scheme, we are able to improve the buffer utilization on routers as well as to support QoS services during a handoff process. Using the ns-2 simulator, we have demonstrated the significance and effectiveness of our proposed scheme.

1. Introduction

Wireless technologies have experienced immense advance in the past few years. Among them, IEEE 802.11 based wireless local area networks (WLAN) play an important role. WLAN offers convenient network connectivity and high-speed link at an affordable cost. With the cooperation of Mobile IP [1] technology, users are able to roaming around different WLAN domains without disrupting its current connections. However, once a mobile host enters another domain, a registration process between the mobile host and its home agent is required. The process may take several seconds and results in serious packet delay or loss. The handoff delay in mobile IP is unacceptable for real-time services; it also degrades TCP data transfer efficiency. Several mechanisms have been proposed and standardized to reduce the delay. Hierarchical architecture [2] aims to reduce the registration time between mobile hosts and home agents. Fast handover [3] mechanism focuses on reducing the lengthy address resolution time when entering a foreign domain. With the combination of these two mechanisms, it can minimize the handoff delay caused by network layer handoff.

The combination of Hierarchical architecture and Fast handover greatly reduces the handoff latency, but there are several potential problems. The Fast handover mechanism neither supports quality of service (QoS) nor provides buffer management mechanism. In order for mobile hosts to achieve seamless connectivity during their movement, an efficient buffer management mechanism should be included in the Fast handover protocol.

To accommodate the host mobility, Internet Engineering Task Force (IETF) proposed an extension to the Internet Protocol, Mobile Internet Protocol (Mobile IP). The Hierarchical Mobile IPv6 is an enhancement of Mobile IPv6. It tends to reduce the amount of signaling required and to improve handover speed for mobile connections. Hierarchical Mobile IPv6 reduces the registration delay of mobility by handling local movements locally and hiding them from home agents. A new Mobile IPv6 node called *Mobility Anchor Point* (MAP) is introduced to maintain the binding between itself and the mobile host. MAP replaces Mobile IPv4's foreign agent, and works as a local home agent for mobile hosts. A mobile node is assigned with two care-of-addresses, called *Regional Care-of-Address* (RCoA) and *On-Link Care-of-Address* (LCoA). RCoA is an address on the MAP's sub-network, it is used by the mobile host as the care-of-address during registration. LCoA is same as the care-of-address in the Mobile IPv6. While moving between subnets inside the MAP's domain, mobile host only changes its LCoA. This hides the movements from its home agent.

Fast handovers protocol is proposed by IETF as a way to minimize the movement detection delay during a handoff process. The key operation of Fast handover is to pre-configure a temporary address before breaking the mobile host's connection with its *previous Access Router* (PAR). Then, when the mobile host is attached to the *new Access Router* (NAR), it can resume its communications with the new care-of-address. Moreover the Fast handover protocol sets up a bi-directional tunnel between the PAR

* This research is sponsored by the Ministry of Education, the PAEU program under grant number. 89-E-FA04-1-4

and NAR. This allows a mobile host to send packets before it finishes the Mobile IP registration process.

There are three phases in the Fast handover protocol operation: *handover initiation*, *tunnel establishment*, and *packet forwarding*. The overall handoff protocol is illustrated in Figure 1. Further details can be referred in [3].

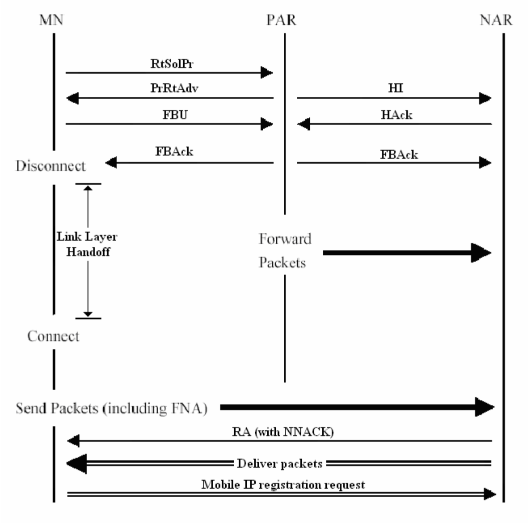

Figure 1. Fast Handover protocol messages.

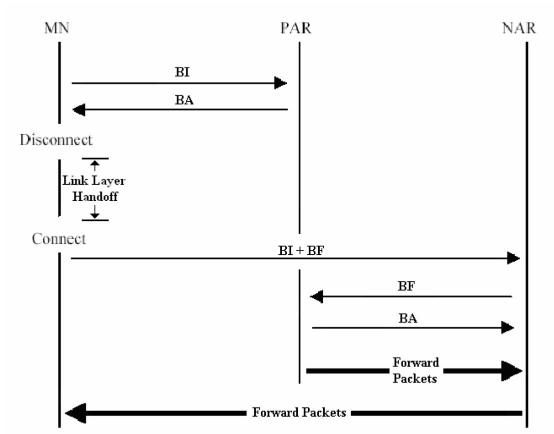

Figure 2. Buffer management for smooth handovers in IPv6 messages.

Currently available IEEE 802.11 WLAN card can only access an access point at a time. This limitation results in an inevitable link down time during handoff process. A feasible solution to avoid packet loss during the link down time is to buffer those packets. The so-called "Buffer management for smooth handovers in IPv6" [4] defines a buffering mechanism for a mobile host requesting its current access router to buffer packets as shown in Figure 2. The mechanism works as follows when a mobile host moves from one subnet into another. First, a router that enables the buffering mechanism advertises its capability by setting up a "B" flag in its router advertisements. Before the handoff process, the mobile host sends a *buffer initialization* (BI) message to requests the access router for buffering the packets. In response to the message, the PAR sends a *buffer acknowledgement* (BA) message back to the mobile host. Incoming packets to the mobile host are then buffered in the PAR (previous router). When the mobile host completes registration process to NAR and obtains a new care-of-address, it sends a *buffer forward* (BF) message to PAR. PAR forwards the buffered packets to the mobile host when receiving the message and finishes the process.

2. Proposed Approaches

With the combination of Hierarchical Mobile IPv6 and Fast Handover mechanism, the handoff latency caused by layer 3 and up can be greatly minimized [5]. However, the handoff latency resulted from link layer handover procedure is still unavoidable. With the different combination of IEEE 802.11 WLAN card and base station, the handover procedure may take from 60ms to 400ms [6]. Since the currently available IEEE 802.11 WLAN card can only link to an access point at a time, a buffer is required to prevent packet loss during handoff process. A mobile host can request its new access router for buffering packets forwarded from its previous access router. However, the simple buffering mechanism suffers several drawbacks. First, if the buffer size in access routers is not large enough, there will be scalability problem. If we simply increase the buffer size, there will be additional handoff delay when dumping packets in the buffer.

Comparing with circuit switching based telecom network, traffics on packet switching based data network are not all of a kind. Not all of them need real-time transport, examples are WWW and FTP packets. Quite a few mechanisms have been proposed to implement quality of service (QoS) on Internet, such as integrated services [6] (Intserv) and differentiated services [7] (Diffserv). Based on the similar concept, not all packets during handoff process need to be buffered, and packets with different priorities should be treated differently. In order to optimize the usage of buffers in an access router, we proposed an enhancement on buffering mechanism for Fast Handover protocol.

There are four design objectives in our buffer management mechanism.

- Support QoS during handoff process

Packets with different priorities should have different buffering operation during handoff process in order to fulfill their QoS requirements.

- Support real-time traffic during handoff
  For real-time packets, we should minimize their waiting time in the buffer as well as forwarding time from previous access router to the new access router.
- Maximize buffer utilization in access routers
  The buffering space in an access router is limited thus we should make good use of them. These include only buffer important packets when running out of buffer, and use both buffers in the previous access router and new access router during handoff process.
- Low signaling overhead
  The proposed buffering mechanism should minimize the signaling overhead caused by control messages.

In order to achieve the design objectives, a buffer management mechanism is integrated into the original Fast Handover protocol. There are two key ideas in our proposed scheme. First, while the original Fast Handover protocol only buffers packets in the NAR, we use buffers in both the PAR and the NAR during a handoff process. This helps improve the total buffer utilization in the network. Second, we define three types of services in the handoff process so that packets can be treated differently based on their traffic characteristics.

Our proposed mechanism can be discussed in three phases: *handover initiation*, *packet redirection*, and *buffer release*. The reference scenario for handover is shown in Figure 3 and the message flows are shown in Figure 4. The control messages used in the proposed method are marked with italic type, and messages are piggybacked on the original Fast Handover protocol if there is a "+" before it.

In the *handover initiation* phases, the handover process is triggered by specific link layer events or policy rules just as the original Fast Handover protocol. Upon receiving a trigger event, the mobile host sends a request (*buffer initialization*, BI) to the PAR for requesting the buffer spaces. During the time period while a bi-directional tunnel established between PAR and NAR, the allocation of buffer spaces for the mobile host is also negotiated via the *Buffer Request* (BR) and *Buffer Request Acknowledge* (BA).

In the *packet redirection* phases, a mobile host sends a *Fast Binding Update* (FBU) to PAR after receiving the PrRtAdv message. When PAR receives the FBU message, it starts buffering packets or forwarding them to NAR. The NAR may either buffer packets from PAR or drop them. Packets are treated differently based on their type of service.

In the *buffer release* phases, the mobile host reestablishes a connection on NAR. When connecting to NAR, the mobile host sends a *Buffer Forward* (BF) message to both NAR and PAR. Upon receiving the BF message, the NAR and PAR forward packets in their buffer to the mobile host, then end the handover process.

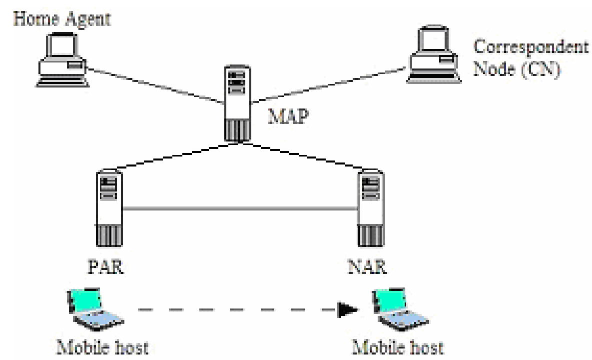

Figure 3. Reference scenario for handover.

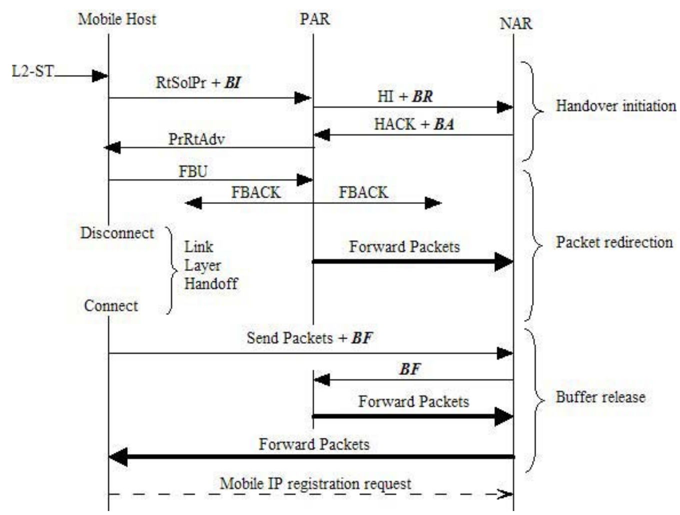

Figure 4. Enhanced buffer management mechanism for Fast Handover protocol messages.

All packets are classified into three types in our method. The sender of a packet can specify a service type in the class-of-service field (in IPv6 header). The first type is *real-time* packets, and we should minimize the handoff delay for this type of packets. The second type is *high priority* packets, these packets are loss sensitive and we have to prevent them from dropping during handoff. The last one is *best effort* packets, which are low priority packets and can be discarded if there is no enough buffer space.

In the *handover initiation* phases, the PAR and the NAR negotiate the allocation of buffer spaces. The four possible results are shown in table 1, where the "Yes" and "No" fields represent the availability of buffer spaces in the NAR or PAR. For example, Case 1 means that both NAR and PAR can offer sufficient buffer spaces.

In the *packet redirection* phase, a mobile host lost its connection with the network while performing link layer handoff process. The PAR and NAR redirect packets for the mobile host during this time period. The redirection operations for each packet depend on the packet's class of service. A mobile host and its correspondent node can specify the priority of a packet on the *class-of-traffic* field in IPv6 header. In order to accommodate the characteristic of each class, we perform different buffering operations for different cases as illustrated in Table 1.

Table 1. Buffering operations

Traffic type	Buffering operation
Case 1□NAR (Yes) PAR (Yes)	
Real-time (a)	Buffer at NAR only. If buffer full, drop the first real-time packet.
High Priority (b)	Buffer at both NAR and PAR
Best effort (c)	Buffer at PAR when PAR > α
Case 2□NAR (Yes) PAR (No)	
Real-time (a)	Buffer at NAR only. If buffer full, drop the first real-time packet.
High Priority (b)	Buffer at NAR only.
Best effort (c)	Forward to NAR only. (Do not buffer)
Case 3□NAR (No) PAR (Yes)	
Real-time (a)	Forward to NAR only. (Do not buffer)
High Priority (b)	Buffer at PAR only.
Best effort (c)	Buffer at PAR when PAR > α
Case 4□NAR (No) PAR (No)	
Real-time (a)	Forward to NAR only. (Do not buffer)
High Priority (b)	Forward to NAR only. (Do not buffer)
Best effort (c)	Drop at PAR. (Do not forward to NAR)

Case 1 is the default situation and all other cases are derived from it. In case 1.a, all packets arriving at PAR will be forwarded to the NAR first, and then NAR buffers the packets. If the NAR runs out of buffer space, the first packet in the buffer will be dropped in order to buffer the new packet. In Case 1.b, packets arriving at PAR will be forwarded to NAR first and buffered there. When NAR runs out of buffer, a *Buffer Full* message will be sent to the PAR. When PAR receives the message, it will buffer the rest of the packets. In Case 1.c, packets are only buffered at the PAR when the available buffering space is greater than α. The value of α is a constant configured by the network administrator. Packets will be dropped when the buffer space is less then α. In Case 2, the PAR cannot provide buffer space. Only real-time and high priority packets are buffered in the NAR (Case 2.a, Case 2.b). The best effort packets are forwarded to NAR without buffering. In Case 3, the NAR cannot provide buffer space. Real-time packets are forwarded to NAR without buffering. High priority packets and best effort packets are buffered in PAR. In Case 4, no buffer spaces are available. The traffic load on both PAR and NAR should be heavy. We drop best effort packets in PAR directly to release the heavy loading of the network. Real-time and high priority packets are forwarded to the NAR without buffering.

The Fast Handover protocol focuses on inter-domain handoffs. Since there might be multiple WLAN access points in a subnet, link layer handoff also happens when the mobile host is remaining inside a subnet. Our enhanced buffer mechanism can be applied to this condition. When a mobile host detects a handoff event, it sends an RtSolPr+BI to the current access router (PAR). If it is only a link layer handoff, the PAR allocates the buffer space and replies a PrRtAdv directly to the mobile host. The PAR starts to buffer packets when it receives an FBU message from the mobile host. After the mobile host completes the link layer handoff, it sends a BF message to PAR. The PAR then forwards the buffered packets back to the mobile host.

3. The Simulation Model and Results

We evaluate the performance of the proposed scheme using the Network Simulator (*ns-2*) [8]. Based on the standard *ns-2* distribution version *ns-allinone2.1b6*, we have added several additional features including the fast handover protocol and the proposed mechanism.

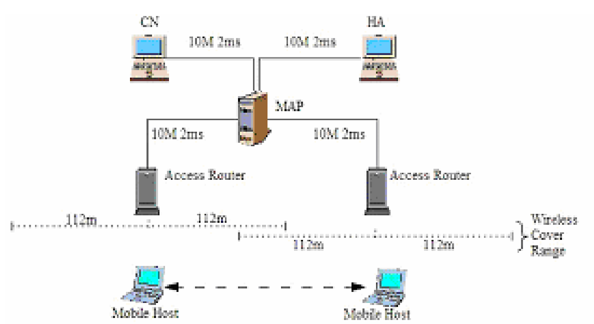

Figure 5. The network topology for simulations.

The network topology for most of the simulations is illustrated in Figure 5, which is a generic Hierarchical Mobile IPv6 network. The wireless coverage area of the access point is approximately 112 meters, and the router advertisement is once per second. We set the link layer handoff delay to 200ms in the simulation. All mobile nodes in the simulation move linearly from one access router to another at a constant speed of 10m/s (36km/Hr). Since the overlapped area between the PAR and the NAR is 12 meters, the mobile host can receive at least one router advertisement from the new access router before leaving the old one. This also assures that the mobile host has enough time to trigger Fast Handover protocols before leaving the old subnet.

All mobile hosts in the simulation as shown in Fig. 5 move along the same path simultaneously (from the PAR to the NAR), and one handoff event occurs during the movement. The CN transmits 160-byte UDP packets every 20ms (64-kb/s audio) to each mobile host. The number of mobile hosts we increased from 1 to 20, to evaluate how many handoffs the network can service at the same time.

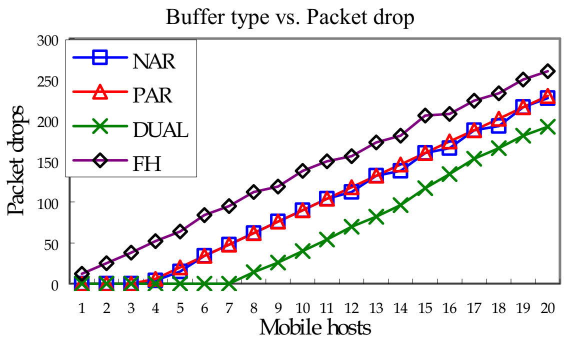

Figure 6. Buffer utilization of different handoff mechanisms.

Four types of handoffs were compared as shown in Figure 6. The NAR line represents the case that all packets are buffered in the NAR, this is also the case of original fast handover protocol. The PAR line shows the condition that all packets are buffered in the PAR. The DUAL line shows the condition where packets are buffered at both the PAR and NAR. The FH line shows the condition for fast handover protocol without buffer spaces. Overall, the DUAL line is the best case of our proposed method. We can observe that in Figure 6, the network can service twice as many simultaneously handoffs than the original fast handover protocol (the NAR line) without dropping any packets. If there is only one buffer space available (the NAR and PAR lines), the proposed method has the same performance comparing with the original fast handoff protocol. The FH line is the worst case in our method when there are no buffer spaces available. In other words, the network runs out of buffer spaces only when both the NAR and PAR are out of buffer spaces.

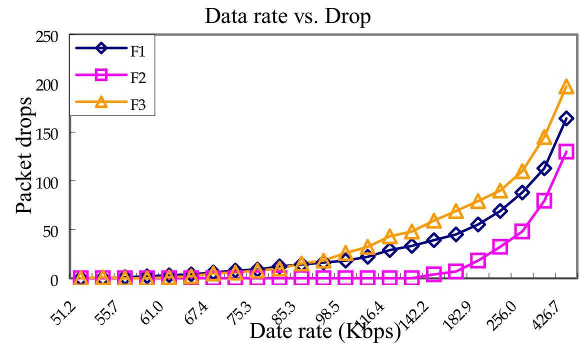

Figure 7. Packet loss for different data rates in the proposed method.

There is only one mobile host moving back and forth between the two access routers as shown in Fig. 5. The CN transmits three UDP flows with different priorities to the mobile node. We use F1, F2, and F3 to represent each of the flows in the following description. We define F1 as the real-time traffic, F2 as the high priority traffic, and F3 as the best effort traffic. In Figure 7, with the increasing of data rate, packet dropping rate from the high priority flow (F2) is always the lowest. When running out of buffering spaces, we successfully saved most of the high priority packets at the cost of dropping best effort and real-time packets.

Real-time packets should be transferred to the mobile host as soon as possible after the handoff process. In the proposed method, real-time packets are forwarded to the NAR and buffered there during link layer handoff process. This saves the transfer delay from the PAR to the NAR when forwarding these buffered packets to the mobile host.

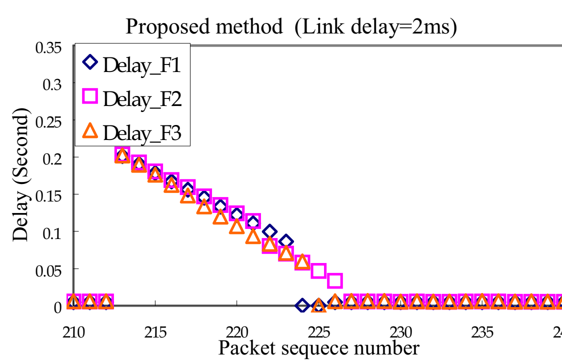

Figure 8. End-to-end delay in fast handover protocol. (Link delay=2ms)

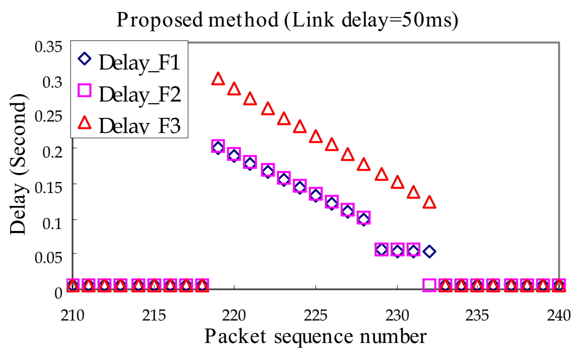

Figure 9. End-to-end delay in fast handover protocol. (Link delay=50ms)

The buffer size is configured large enough so as to accommodate all packets during the handoff process. Since the interval of UDP traffic (F1~F3) is 20ms, the handoff delay for the i^{th} packet coming during the handoff period is "Link layer handoff time (200ms)" – *20 * i + forward time*. While the forward time represents the time for buffered packets being transmitted to the mobile host after a link layer handoff. In Figure 8, the link delay between the NAR and the PAR is set to 2ms, while we set it to 50ms in Figure 9. The link delay may be caused by either heavy traffic in an access router or the transmission latency, thus there is no guarantee for this delay. However, with our proposed method, real-time packets (F1) are forwarded and buffered at the NAR during the link layer handoff period. We can avoid unnecessary transmission delay when forwarding the buffered real-time packets. Besides, since the best effort packets are buffered in the PAR, the queuing delay for real-time packets also decreases. Thus the proposed method assures that real-time packets can be transferred to the mobile host without any unnecessary delay caused by the handoff process.

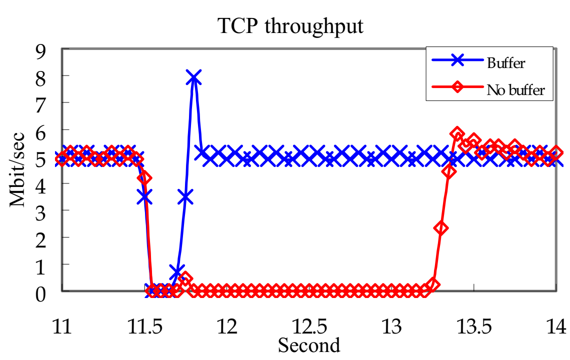

Figure 10. TCP throughput during link layer handoff.

The original fast handover protocol does not support buffering mechanism during a pure link layer handoff. This means that an access router is unable to buffer packets for a mobile host which is moving between different WLAN access points within the same subnet. This temporary disconnection results in packet loss and degrades the throughput of TCP connections. However, the proposed buffering mechanism supports buffering packets under any handoff conditions.

In Figure 10, there is an FTP connection between the MH and the CN using TCP Reno. The link layer handoff occurs at 11.7s and results a 200ms disconnection for the MH. When there is no buffering mechanism, the link layer handoff results in long TCP connection timeout, and the throughput drops significantly. The link layer handoff process starts at 11.47 second and finishes 200ms after. All packets sent to the mobile host during this period are lost. The TCP retransmission starts at 11.7 second. However, not all the lost packets can be recovered since the congestion window is full. The TCP connection now must wait until the timeout occurs. In most TCP implementations, the minimum TCP retransmission timeout is 1 second. Considering the TCP tick interval (500ms), the connection takes 1 to 1.5 second to resume the transmission. However, when applying our method, packets are buffered and forwarded to the MH after 200ms. Thus the CN starts transmission right after the handoff process.

4. Conclusions

An enhanced buffer management mechanism for Fast handover protocol has been proposed. There are several advantages of the proposed buffer management mechanism. First, the cooperation of both buffering spaces in the PAR and the NAR assures the full utilization of buffer spaces. The network will be able to serve more handoffs simultaneously. Second, the proposed method supports QoS mechanism. The high priority packets are protected from being dropped and the real-time packets are protected by minimizing the delay during a handoff process. Third, with the buffering mechanism integrated into the fast handover protocol, the proposed method also supports buffering operations during a link layer handoff. This helps improve the performance of TCP connections when a mobile host handoffs. Finally, the proposed method piggybacks most of the control messages on the original fast handover protocol thus it does not generate additional signaling overhead. We have evaluated the performance of the proposed method using the ns-2 simulator, and it is proven that our method really improves the quality of communications during a handoff process.

Reference

[1] C. Perkins, “IP mobility support,” RFC 2002, IETF October 1996
[2] H. Soliman, C.Castelluccia, K. Malki and L.Bellier, “Hierarchical MIPv6 mobility management,” Internet Draft, IETF, October 2002. Work in progress.
[3] G. Dommety et al., “Fast handovers for mobile IPv6,” Internet Draft, IETF, March2002. Work in progress.
[4] Govind Krishnamurthi, Robert C. Chalmers, Charles E. Perkins, “Buffer Management for Smooth Handovers in IPv6,” Internet Draft, IETF, March 2001
[5] Robert Hsieh, Aruna Seneviratne, Hesham Soliman, Karim El-Malki, “Performance analysis oh Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP,” Proceedings of GLOBECOM, Taipei, Taiwan 2002.
[6] Arunesh Mishra, Minho Shin, Willian Arbaugh, “An empirical analysis of the IEEE 802.11 MAC layer handoff process”, University of Maryland Technical Report, UMIACS-TR-2002-75, 2002
[6] Braden, R., Clark, D. and Shenker, S., "Integrated Services in the Internet Architecture:an Overview", Internet RFC 1633, June 1994
[7] Blake, S., et al., "An Architecture for Differentiated Services." RFC 2475, 1998.
[8] “The Network Simulator – ns (version 2) Website,” http://www.isi.edu/nsnam.