Mobility Management of Transport Protocol Supporting Multiple Connections

Kazuya Tsukamoto[†]
kazuya@infonet.cse.kyutech.ac.jp

Yoshiaki Hori[††]
hori@csce.kyushu-u.ac.jp

Yuji Oie[†††]
oie@cse.kyutech.ac.jp

[†] [†††]Dept. of Computer Science and Electronics, Kyushu Institute of Technology
Kawazu 680–4, Iizuka, 820–8502 Japan
[††]Dept. of Computer Science and Communication Engineering, Kyushu University

ABSTRACT

A mobility management scheme for handling seamless handoffs between appropriate datalink interfaces is presented. The future mobile environment will be characterized by multimodal connectivity to the Internet with dynamic switching based on many factors. Many technologies have been proposed to support host mobility across diverse wireless access networks, and operate in various layers of the network architecture. Our major concern is the transport protocol that recovers packets lost during handoffs and controls transmission speed to achieve efficient communication. In this paper, we first examine the various latencies associated with the handoff mechanism, and discuss the limited performance of existing technologies based on a single transport connection. We then propose a new mobility management scheme that resolves the problems of limited performance by transport protocol supporting multiple connections. We compare the performance of the proposed scheme with existing technologies, and demonstrate that the proposed scheme achieves excellent goodput performance.

Categories and Subject Descriptors

C.2.2 [**Computer-Communication Networks**]: Network Protocols

General Terms

Algorithms, Design, Management, Performance

Keywords

Mobility management, Transport protocol, Seamless handoff, Network selection, Ping-pong effect

MobiWac'04, October 1, 2004, Philadelphia, Pennsylvania, USA.

1. INTRODUCTION

A wide range of datalink technologies has been introduced in recent years to provide access to the Internet backbone. In particular, these technologies have received much attention as a means of obtaining mobile access to the Internet.

In the upcoming third-generation digital mobile radio standard, the Universal Mobile Telecommunication System (UMTS) will provide packet-switched bearer services at capacities up to 2 Mb/s over a wide service area. And currently for mobile computing, the IEEE802.11a/b/g Wireless Local Area Network (WLAN) has become a widely implemented technology providing high-speed, Internet access within a limited area. The WLAN technology also currently is used to provide a public Internet access services for its customers.

WLAN and cellular networks are complementary technologies. The WLAN provides relatively inexpensive (flat rate) broadband connectivity, but offers only limited coverage. In contrast, cellular networks provide wide-area coverage, but are currently capable of only narrowband connectivity at relatively high cost (measured rate). However, while only cellular services are currently available over a wide area, WLAN hotspot services are expected to be available in many public spaces, including coffee shops and public transport waiting areas. Naturally, integrating WLAN and cellular networks will provide service at low cost to those users who need both high-speed wireless access and anytime-anywhere mobile connectivity [1].

To achieve this level of integration, mobile hosts will be designed and equipped for multiple-interface capability. It will be necessary for the mobile host to automatically select an available datalink interface based on transmission capacity, signal intensity, bit error condition and other factors. (Generally speaking, the change from one datalink interface to another is referred to as a **handoff**.)

Existing mobility technologies, such as Mobile IP [2] and Migrate TCP [3], maintain end-to-end communication even when a handoff occurs. Furthermore, Transmission Control Protocol (TCP) performance is improved with an additional function to prevent a retransmission timeout during a handoff [4]. TCP is conventionally designed for reliable data transmission service through an end-to-end connection across the Internet, and information related to a connection, such as the identity of endpoints, is held in only one Transmission Control Block (TCB) hash table per connection. When a handoff invalidates a TCB held in the transport layer due to an IP address change, existing mobile technologies based on TCP must modify the TCB state accordingly to achieve transport layer mobility, that is, it updates the single TCB of that connection. This is referred to as "vertical handoff" [5]. None of the existing vertical handoff technologies address a large change in bandwidth when the datalink interface changes. The Bandwidth Aware scheme, proposed in [4], does achieve efficient communication even for pronounced bandwidth changes at handoff, but the TCB must still be updated during the handoff resulting in large latency and a possible significant loss of packets. A particular situation that may arise is when a mobile host moves back and forth between two networks with frequent, multiple handoffs (called the **ping-pong effect**). That can cause severe performance degradation in existing vertical handoff technologies [6]. In our proposed scheme, we solve this issue by modifying the transport layer protocol to provide multiple, simultaneous TCB handling capability.

Parallel TCP (pTCP) [7] has already been proposed as transport protocol extensions for handling multiple TCBs simultaneously. pTCP is an end-to-end transport layer approach that wraps multiple TCP connections. It provides a socket interface that aggregates multiple TCP connections on behalf of an application. Each TCP connection is managed appropriately depending on its particular characteristics. pTCP, therefore, can provide high throughput with multiple TCP connections. With pTCP, however, the number of available interfaces is constant, meaning that mobility is not addressed, and details of the handoff decision mechanism and processing are not explicitly defined. Moreover, this proposal has not been specifically compared in terms of performance with the vertical-handoff technologies.

In this paper, we propose a new mobility management scheme by transport protocol, which can achieve efficient communication even for handoffs in which multiple TCBs are handled simultaneously. We demonstrate the effectiveness of the proposed technology through a performance comparison with the vertical handoff technologies.

2. HANDOFF LATENCY

Total handoff latency is the sum of all individual latencies. The usual handoff includes the following three process latencies [8]: D_1: link-layer handoff, D_2: movement detection, and D_3: registration.

In this paper, we consider the following three additional process latencies: D_4: IP address binding by Dynamic Host Configuration Protocol (DHCP), D_5: Authentication, Authorization, Accounting (AAA), and D_6: Point-to-Point Protocol (PPP) connection establishment (UMTS only).

D_1 is the time required to establish a connection between the mobile host and a new Access Point (AP). It is dominated by the time spent finding newly available APs and then choosing one of them. APs broadcast beacon messages that include an Extended Service Set Identifier (ESSID). Mobile hosts scan all APs, detecting those with an appropriate ESSID, and then chooses the AP with the strongest signal. The time required for scanning consumes more than 90% of the total layer 2 delay. It varies from approximately 50 ms to 400 ms, depending on the hardware used [9].

D_2 is the time required to detect a handoff in layer 3 or higher. When Mobile IP is not used, mobility technology must be employed to detect movement based on the beacon message (layer 2). In this case, when the mobile host receives beacon messages from a newly detected network for x seconds, it considers the new network to be available and makes the decision to execute a handoff. On the other hand, when the host fails to receive beacon messages for y seconds from an AP currently in use, it considers the network to no longer be available and makes the decision to execute a handoff to another available network. Here, D_1 is assumed to be included in D_2, and x or y is defined as the **handoff decision latency**.

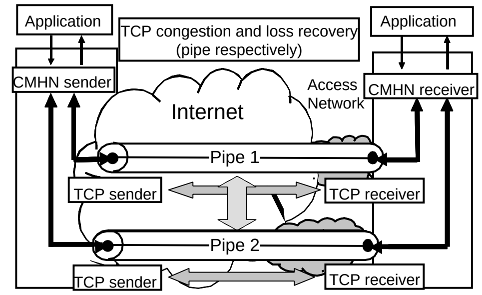

Figure 1: Architecture of the proposed scheme

After the handoff decision, the mobile host reconfigures its IP address using DHCP mechanism (D_3: 300 ms [10]), and reports the change to a location server such as DNS or HA (D_4: 50 ms). In addition, an AAA process is required to ensure the integrity, privacy and confidentiality of the user's location (D_5: 350 ms [6]). Furthermore, if the mobile host exchanges IP packets over the Internet via the UMTS access network, it is required to establish a PPP connection with a UMTS gateway (D_6: 600 ms [6]). In this paper, the time required for D_3 to D_6 is defined as the **handoff processing latency**. Vertical handoff schemes update the TCB while the handoff is being processed and thus cannot communicate with the host until handoff processing is completed.

Considering the above discussion, handoff latency therefore consists of two major elements. The first is the handoff decision latency (x or y), and the second is the handoff processing latency. The handoff processing latency of WLAN is **700 ms** (D_3-D_5) and that of UMTS is **1300 ms** (D_3-D_6).

3. OUR PROPOSED SCHEME

3.1 Architecture of the proposed scheme

The proposed scheme uses an end-to-end transport layer approach with dynamic handoff mechanism, acting as a wrapper of multiple TCP connections as shown in Figure 1. It provides a socket interface that aggregates multiple TCP connections on behalf of an application. In the scheme, a Connection Manager for Heterogeneous Network (CMHN) is responsible for detecting the link-up or link-down state and selecting the optimum datalink interface. It stripes the data packets across multiple pipes, with each pipe managed by individual TCP congestion control. TCP requests new data packets from the CMHN until the TCP congestion window (cwnd) is filled with packets, in a manner similar to pTCP.

3.2 Available datalink detection

Figure 2 shows the detection procedure for changing the datalink interface. The mobile host, after making a decision to execute a handoff as described in Section 2, conveys infor-

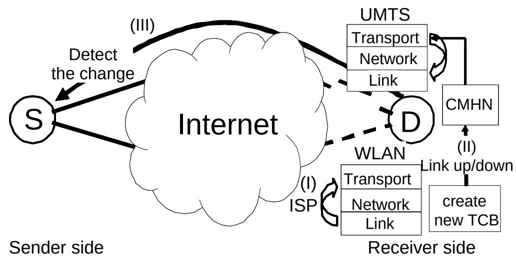

Figure 2: Detection procedure (WLAN link)

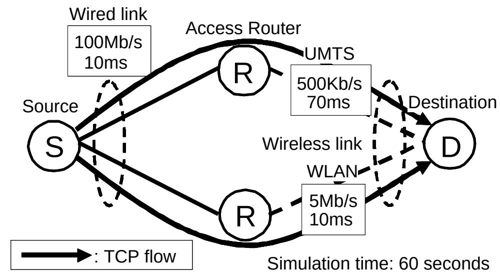

Figure 3: Simulation model

mation about the datalink interface state (such as link-up or link-down) from the link layer to the transport layer through a cross-layer mechanism such as an interlayer signaling pipe (ISP) [11] (I). After that, the transport layer actually processes the handoff by creating a new TCB based on the conveyed information such as the new IP address. Note that the proposed scheme can maintain multiple TCBs and retain the TCP connection on the inactive interface for a fixed period of time. Because it does not release the TCB, a new TCB is not required if the new network is only intermittently available for some reason. Such a case is examined later in the simulation results.

After creating the TCB, the TCP receiver informs the CMHN receiver of any change in the link state, that is, link-up or link-down (II). The CMHN receiver then sends a special packet with a link-up/link-down flag to the CMHN sender on an available and stable connection (typically UMTS link) (III). In this way, any change in the link state on the receiver side is known immediately by the CMHN sender.

3.3 Optimized datalink selection

3.3.1 Link up

If CMHN sender detects a change to the link-up state on some datalink interface, it starts striping data across multiple interfaces, including the newly available interface, as parallel transmissions. The TCP sender that established connection on the new datalink interface calculates the bandwidth delay product (BDP) during the start-up period and sets the BDP according to the method proposed by [12]. The CMHN determines if a network interface is appropriate by comparing the BDP value of the newly available network with the BDP values of any links currently in use for data transmissions. If the BDP value of the new datalink interface is greater than that of a current interface, the CMHN changes from the current to the new interface and begins immediately transmitting packets. Otherwise, the CMHN continues data transmission on the current interface.

When the CMHN decides to execute a handoff, it halts TCP communication over the current link. At this time, the TCP sender informs the CMHN of any unacknowledged packets already sent to the TCP receiver but not confirmed by received ACK packets. Upon switching to the new datalink interface, the CMHN first retransmits the unacknowledged packets then continues transmitting new packets.

3.3.2 Link down

The CMHN sender detects a change to the link-down state on the current datalink interface using a procedure similar to that for the link-up case. The CMHN receiver informs the CMHN sender of the state change over one of the other available datalinks (*backup* links). By using the *backup* link, the receiver side can reliably inform the sender side of the change even if the current interface goes down suddenly. In that case, the CMHN sender will immediately restart data transmission on the *backup* datalink interface.

After the handoff decision, the CMHN stops TCP communication over the current link. The TCP sender informs the CMHN sender of any unacknowledged packets, and the CMHN retransmits them afterward, as in the link-up case.

4. SIMULATION MODEL

We compare our proposed scheme with existing vertical handoff technology using the Virtual Inter-Network Testbed (VINT) network simulator (NS) version 2. In the simulations, the end-to-end path from multi-homed mobile host (S) to destination (D) is emulated using the bandwidth and propagation delay. The bandwidth and propagation delay of a link between the access router (AR) and the destination are varied dynamically according to channel conditions, as shown in Figure 3.

A topology with two active links between the source and destination is used. The TCP variant employed is NewReno, and the TCP traffic is greedy file transfer, such as ftp data, using TCP packets of 1500 bytes. Two datalink types are used in the simulations: (i) UMTS link and (ii) WLAN link.

We assume that the UMTS link is always available throughout the simulation and that the WLAN link is available during a limited period due to movement of the mobile host and variations in the signal intensity. The experiments therefore compare performance of the proposed scheme with that of vertical handoff technology when the number of available datalink interfaces changes dynamically. The Bandwidth Aware [4] vertical handoff technology is used. Through extensive simulations, we examine in detail the characteristics of pTCP goodput to clarify the impact of the proposed mobility management scheme when the handoff decision time (x or y) is varied.

5. RESULTS AND DISCUSSION

5.1 Case 1: WLAN becomes active (link up)

In this simulation, the mobile host begins receiving beacon messages from the WLAN AP at 20 s. Typically, the mobile host continues to receive beacon messages for x seconds, decides to execute a handoff at $20 + x$ s, then starts handoff processing. The handoff processing time considered

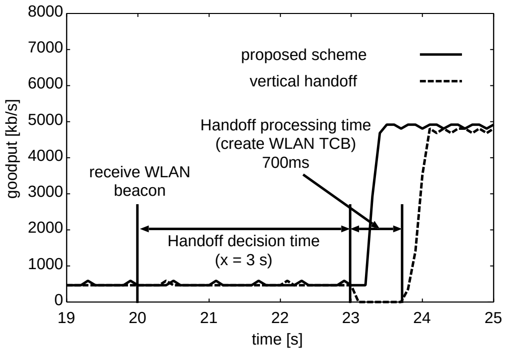

Figure 4: Goodput (link-up, $x = 3$)

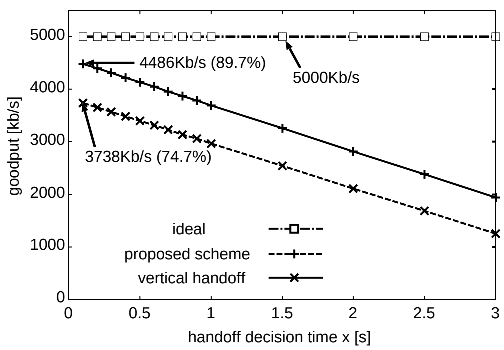

Figure 5: Average goodput (20-25 s) vs. Handoff decision time (x) :link-up

here is 700 ms (WLAN). We assume the mobile host has connected with the WLAN and then reconnects while its TCB is still available. Based on this assumption, however, the proposed scheme need not create the TCB since it can reuse the previous TCB.

Figure 4 shows the handoff timing and change in goodput when a new WLAN link comes up, with x set to 3 s, indicating that the mobile host decides the handoff carefully. In both the proposed and vertical handoff schemes, packets are transmitted through the UMTS link until 23 s (20 s + 3 s) even though the mobile host does in fact begin receiving beacon messages at 20 s. The proposed scheme can use the WLAN shortly after 23 s, whereas communication ceases with the vertical handoff scheme from 23 to 23.7 s due to handoff processing. This is because in the vertical handoff case, transmitted packets are dropped while the TCB of the WLAN is being updated.

Figure 5 shows the average goodput for the period from 20 s to 25 s, with a handoff decision time x of 0.1 to 3 s. The ideal goodput occurs at the bottleneck bandwidth (5000 Kb/s in this case) after handoff. Goodput achieved by the proposed scheme is greater than that of vertical handoff in all cases. With a decrease in x, the goodput performance improves. When x is 0.1 s, the vertical handoff scheme achieves a goodput of 3738 Kb/s, a 75% utilization of the ideal value. The proposed scheme as well achieves a goodput of 4486 Kb/s, as much as 90% utilization. These results demonstrate that aggressive handoff decisions, achieved by

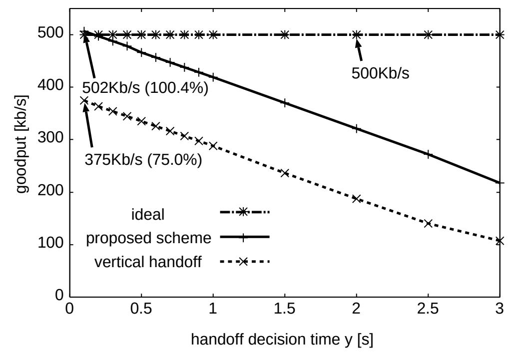

Figure 6: Average goodput (30-35 s) vs. Handoff decision time (y) :link-down

shortening x in both schemes, have a significant impact on goodput when the WLAN link goes up.

5.2 Case 2: WLAN becomes inactive (link down)

In this case, we examine goodput performance when a WLAN becomes inactive; i.e., the mobile host does not receive WLAN beacon messages for 30 s. After the interruption continues for y seconds, the mobile host decides to execute handoff at 30 + y s and proceeds with handoff processing as in the link-up case. The handoff processing time considered here is 1300ms (UMTS).

Figure 6 shows the average goodput for the period from 30 s to 35 s, with a handoff decision time y of 0.1 s to 3 s. The ideal goodput occurs at the bottleneck bandwidth (500Kb/s in this case) after the handoff. As in the case for link-up, the average goodput achieved by the proposed scheme is greater than that of the vertical handoff in all cases. In both schemes, decreasing y improves the goodput performance as in the case of link-up. In particular, when y is 0.1s, vertical handoff achieves a goodput of 375 Kb/s, a 75% utilization of the ideal value. The proposed scheme as well achieves a goodput of 504 Kb/s, a 100.4% utilization. This goodput is greater than the ideal value due to the retransmission of unacknowledged packets. These results demonstrate that aggressive handoff decisions, achieved by shortening y in both schemes, have a significant impact on goodput when the WLAN link goes down.

5.3 Case 3: multiple frequent handoffs

Future mobile environments will be characterized by diverse wireless networks. In this environment, mobile hosts will often move into the boundary areas of different networks types. This will result in frequent handoffs at short intervals due to dynamic changes in wireless signal strength. Simulation is performed to examine the effect of multiple frequent handoffs between UMTS and a one WLAN at 1-s intervals. The handoff decision time (x and y) is set to a small value, i.e., 0.1 s, because a smaller handoff decision time can attain higher goodput, as demonstrated by the previous results.

Figure 7 shows how goodput changes when the wireless signal strength changes five times at 1-s intervals. Communication is rarely successful for the vertical handoff scheme, because it ceases communication for the entire handoff processing time (700 ms and 1300 ms) after every handoff de-

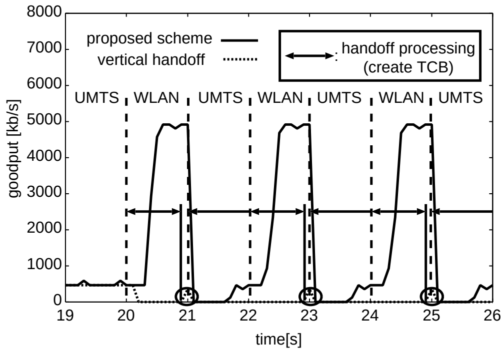

Figure 7: Goodput (multiple handoffs, 1 s interval)

cision, even though the new datalink interface is available shortly after the handoff decision (0.1 s). This results in severe degradation in goodput performance (95 Kb/s), which is quite inferior to the case in which there is no handoff mechanism (denoted by "no handoff") (500 Kb/s). These results demonstrate that aggressive handoff decisions degrade performance in the case of vertical handoff.

The proposed scheme, on the other hand, starts handoff processing quickly after making aggressive handoff decisions, and restarts communication while reusing the TCBs of both UMTS and the one WLAN immediately after the decisions are made. This leads to an effective utilization of the available interfaces. The proposed scheme achieves a goodput of 1562 Kb/s, about three times that of "no handoff". These results clearly demonstrate that the proposed scheme greatly improves goodput performance even when the wireless signal strength is changed consecutively at 1-s intervals.

These results clearly demonstrate that performance degradation of the proposed scheme depends on the handoff decision time only, and the proposed scheme greatly improves goodput performance through aggressive handoff decisions irrespective of the interval of change in wireless signal strength. It should be noted that a smaller handoff decision time leads to higher goodput performance for a single handoff, but a too-small handoff decision time can adversely affect goodput performance in the case of vertical handoff.

6. CONCLUSIONS

In future mobile network environments, the optimal wireless network available to a mobile host will change due to signal strength and/or available bandwidth variations due to movement of the mobile host. Thus, mobile hosts will continuously change the wireless networks to which they are connected; i.e., handoffs will occur frequently.

The mobility management scheme we proposed uses a new transport protocol that can handle multiple connections simultaneously and can transmit packets on multiple separate datalink interfaces, i.e., multiple TCBs can be maintained. Our scheme achieves seamless and efficient communication during handoffs using multiple TCBs. In this paper, we focused on handoff time in terms of (1) handoff decision and (2) handoff processing. We evaluated the proposed scheme through extensive simulations and compared it with vertical handoff technology which can use only one connection at a time. The simulation results demonstrate that the proposed scheme attains excellent performance in all cases. The performance is improved by using smaller handoff decision time for a single handoff. On the other hand, goodput performance in the vertical handoff scheme is degraded for frequent multiple handoffs if a too-small handoff decision time is used, while the proposed scheme for the same case attains excellent performance due to its use of multiple TCBs.

From these results we can see that the handoff decision time at different handoff intervals must be carefully adjusted in the case of vertical handoff. And, we can conclude that the proposed scheme enables mobile hosts to use the available bandwidth efficiently at any handoff interval with aggressive handoff decisions and multiple TCBs.

7. ACKNOWLEDGMENTS

This work was supported in part by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (A)(15200005) and in part by the Ministry of Public Management, Home Affairs, Posts and Telecommunications, Japan.

8. REFERENCES

[1] H. Luo et. al, "Integrating Wireless LAN and Cellular Data for the Enterprise," *IEEE Internet Computing Magazine*, pp. 25-33, Mar/Apr. 2003.

[2] C. Perkins, "IP Mobility Support," *Networking Group*, RFC *2002*, Oct. 1996.

[3] A. C. Snoeren et. al, "An End-to-End Approach to Host Mobility," *ACM/IEEE MOBICOM'00*, no. 6, pp. 155-166, Boston, Aug. 2000.

[4] K. Tsukamoto et. al, "New flow control schemes of TCP for multimodal mobile hosts," *Proc. of the IEEE VTC2003-Spring*, Vol. 3, pp. 1720-1724, Apr. 22-25, 2003.

[5] M. Stemm et. al, "Vertical handoffs in wirelessoverlay networks," *Mobile Networks and Applications*, 3(4): pp. 335-350, 1998.

[6] H. Parikh et. al, "Seamless Handoff of Mobile Terminal from WLAN to cdma2000 Network," *Proc. of IEEE 3G Wireless2003*, San Francisco, CA, USA, 2003.

[7] H. Y. Hsieh et. al, "A Transport Layer Approach for Achieving Aggregate Bandwidths on Multihomed Mobile Hosts," *Proc. of ACM MOBICOM2002*, pp. 83-94, Atlanta, Sep. 2002.

[8] N. A. Fikouras et. al, "Accelerating Mobile IP hand-offs through Link-layer Information, an Experimental Investigation with 802.11b and Internet Audio," *Proc. of MMB2001*, Germany, Sep. 2001.

[9] A. Mishra et. al, "An Empirical Analysis of the IEEE 802.11 Mac Layer Handoff Process," *ACM SIGCOMM Computer Communication Review*, Vol. 33, Issue 2, Apr. 2003.

[10] A. Dutta et. al, "Application Layer Mobility Management Scheme for Wireless Internet," *Proc. of IEEE 3G Wireless2001*, San francisco, 2001.

[11] G. Wu et. al, "Interactions between TCP and RLP in Wireless Internet," *Proc. of IEEE GlobeCom'99*, pp. 661-666, Dec. 1999.

[12] J. C. Hoe, "Improving the Start-up Behavior of a Congestion Control Scheme for TCP," *Proc. ACM SIGCOMM'96*, pp. 270-280, Aug. 1996.