Speed Adaptive MIP

Jun Tian，Abdelsalam (Sumi) Helal
Computer & Information Science and Engineering Department
University of Florida, Gainesville, FL 32611-6120, USA

Abstract

In this paper, the performance of MIP at different moving speeds is evaluated on an emulation testbed. The emulation result shows that current MIP protocol is not suitable for rapid moving environments. This paper depicts the relationship between the performance and the moving speed and breaks down the handoff latency of MIP. A speed adaptive MIP protocol extension is proposed and implemented on Hierarchical MIP. The evaluation of the speed adaptive MIP shows that it greatly improves the performance of MIP in rapid moving environments.

***Keywords*-MIP, Rapid Mobility, Speed Adaptive, Performance**

1. Introduction

While TCP/IP successfully overcomes the barriers of time and distance in a wired network, mobile IP is a promising technology to eliminate the barrier of location for the increasing wireless internet usage. Third generation (3G) services combine high speed mobile access with IP-based services. 3G networks are based on a set of radio technology standards such as CDMA2000, EDGE and WCDMA. Mobile IP (MIP) can be used as the common macro mobility management framework to merge all these technologies and allow mobile users to roam between different access networks.

Throughout history, the economic wealth of people or a nation has been closely tied to transportation efficiency. A person can drive a car on high way at speed of 80miles/h. Some high speed trains such as France TGV, Japanese bullet, German maglev can travel at speeds of over 200 miles/h. Could people surf the internet, communicate with families and enjoy an online movie while traveling at high speeds? Could the current network infrastructure support rapid mobility?

A review on recent research on MIP shows a great amount of efforts contributed to reducing MIP handoff latency. Malki [1] proposed two mobility protocols, pre- and post-registration, using L2 trigger. In pre-registration, MN may communicate with both oFA and nFA. In post-registration, data are cached in nFA before the registration is completed. Fast-handover [2] for Mobile IPv6 network combines the about two methods. But they all depend on L2. S-MIP[3], uses MN location and movement patterns to 'instruct' the MN when and how handoff should be carried out. [4] also uses MN's movement model to predict handoff. But all these efforts didn't consider the speed factor of MN, which may cause problems when the MN moving rapidly.

In this paper, section 2 introduces a rapid mobility emulator and the emulation scenario. Section 3 shows the performance of MIP and its relationship to speeds. Section 4 breaks down the handoff procedure of MIP and presents a quantitative analysis of the handoff latency. A Speed Adaptive MIP(SA-MIP) is proposed and evaluated in section 5 and 6.

2. Rapid mobility emulator and evaluation environment

MIP[5], proposed by C. Perkins in RFC3344, is designed independently from all Layer 2 technologies. In order to evaluate the performance of MIP in rapid moving environments, without losing generality, we build up a Rapid Mobile Network emulator, RAMON[6], on Wireless LAN. Fig.1 is the architecture of RAMON.

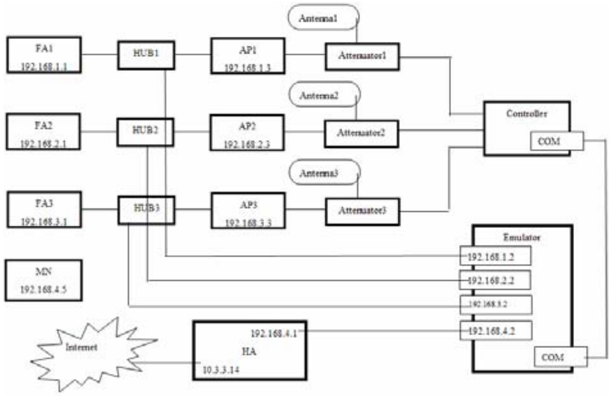

Fig. 1 Architecture of RAMON

RAMON consists of a Pentium II PC as Emulator, a circuit board as Controller, three JFW Industries Attenuators with Antennas, three Cisco 350 Access Points (AP), three Foreign Agents (FAs), a Home Agent (HA) and one or more Mobile Nodes (MNs). The FAs, HA, and MNs, which are the major entities of MIP, are running Linux kernel 2.4.20 and are installed with HUT dynamic MIP implementation version 0.8.1[7]. The Attenuators are program controllable device. The Emulator manipulates the Attenuators via the Controller to control the signal strength of the Access Points. By increasing or decreasing the signal strength of one AP, we can emulate the MN moving towards to or away from the AP. By varying the increasing or decreasing speed of the signal strength, we can emulate the speed changes of the MN.

In Fig.2, a rapid moving MN travels trough 8 APs. Each AP is wired to a FA. The distance between every two adjacent APs is d= 250m, 500m or 1000m. The moving speed of MN varies from 10m/s to 80m/s. In this experiment, a large ftp file is transferred from the CN to the MN and the TCP sequence numbers are tracked when the MN is traveling.

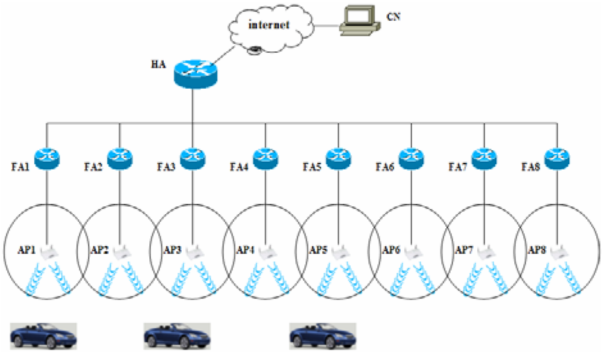

Fig. 2 Emulation scenario for MIP over WLAN

3. Performance of MIP at different speeds

The experiment results shows that the time-sequence graph and throughput graph at speed 20m/s and AP distance 1000m are similar to those graphs at speed 10m/s and AP distance 500m. Also the time-sequence graph and throughput graph at speed 80m/s and AP distance 1000m are similar to those graphs at speed 40m/s and AP distance 500m, as well as those graphs at speed 20m/s and AP distance 250m.

Fig.3 depicts the average throughput/speed relationship when d = 500m and 1000m. It shows that if we double the speed and at the same time double the AP distance, the average throughput shows no suggestive difference.

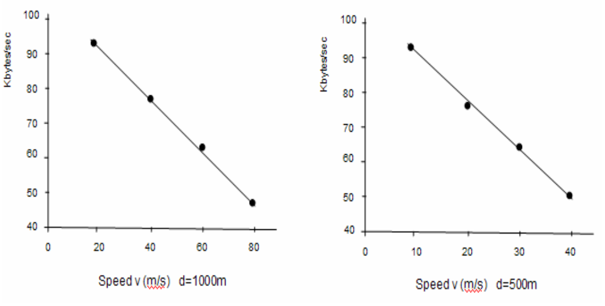

Fig.3 Average Throughputs vs Speeds

The analysis of the performance also shows:

1. The handoff time doesn't change with speed.

2. Effective time/total travel time ratio decreases when the speed increases. This is the reason why higher speed has lower throughput.

Let P_{avg} – Average throughput

P_{maxavg} – Average throughput without handoff

T_{travel} – Total travel time

$T_{effective}$ – Total effective time for ftp transmission

$T_{handoff}$ – Total handoff time while traveling

$K_{handoff}$ – The number of handoffs while traveling

$T_{handoff}$ – Average handoff time among 7 times of handoff

Then, $P_{avg} = (P_{maxavg} / T_{travel}) \times T_{effetive}$

$= P_{maxavg} (T_{travel} - T_{handoff}) / T_{travel}$

$= P_{maxavg} (1 - T_{handoff} / T_{travel})$

$= P_{maxavg}(1 - K_{handoff} \times t_{handoff} / T_{travle})$

$= P_{maxavg}(1 - (K_{handoff} / T_{travle}) \times t_{handoff}))$

Since $t_{handoff}$ doesn't change, The change of P_{avg} is caused by $K_{handoff}/T_{travel}$ ratio.

We define MN handoff rate as $r_h = v/d$, which is the ratio of the MN's speed and the cell size(AP distance). It means that how many APs or FAs the MN hands over in one second. r_h is also equal to $K_{handoff} / T_{travel}$.

3. The relationship between the performance of MIP over WLAN and the moving speed is presented in the following equation:

$$P_{avg} = P_{maxavg}(1 - r_h \, X \, t_{handoff})) \quad (1)$$

Where P_{avg} is the average throughput for the MN; P_{maxavg} is the average throughput without handoff. $t_{handoff}$ is the average handoff time (in second) for every handoff procedure.

Fig.4 shows the relationship between average throughput and handoff rate in equation 1. At handoff rate 0.02 FAs/s, the average throughput is 92.72 kB/s. When the handoff rate goes up to 0.08 FAs/s, the average throughput drops to 49.97 kB/s.

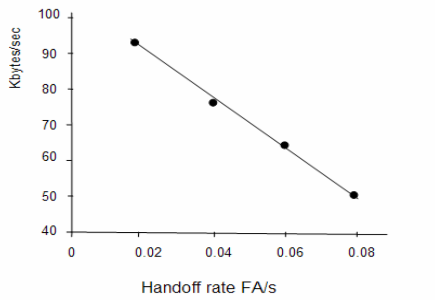

Fig. 4 Average Throughput vs Handoff Rate

This section shows that the performance of MIP depends on the MN handoff rate. In section 5, we propose an approach to improve the performance of MIP in rapid moving environment using this throughput/handoff-rate relationship. In the following section, we will take a deep view of the handoff latency by breaking down the handoff procedure of MIP over WLAN.

4. Quantitative analysis of the handoff latency

Fig. 5 is a global view of the handoff procedure of MIP over two WLAN. When a MN moves from WLAN1 to WLAN2, it performs a layer2 802.11b handoff between Access Point 1 (AP1) and Access Point 2(AP2). After the layer2 handoff, the MN begins a layer3 handoff, which is MIP handoff. Suppose there is a communication, e.g. a TCP stream, between MN and CN before the handoff. After the layer2 and layer3 handoff, it will require a significant time interval to recover the communication. This time internal is called layer4 handoff delay, which is also a part of the whole handoff cost. Equation 2 gives the life-cycle of MIP over WLAN handoff procedure:

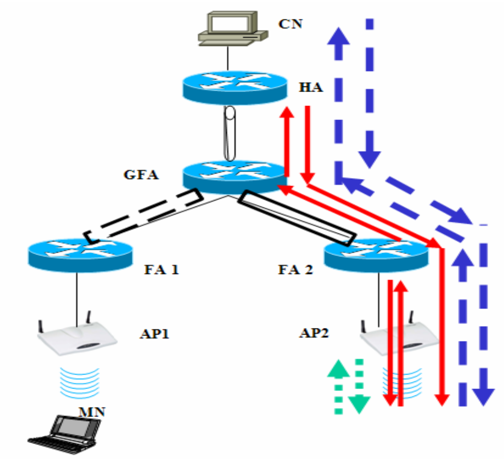

Fig.5 MIP/WLAN handoff procedure

$$t_{handoff} = t_{L2handoff} + t_{L3handoff} + t_{L4handoff} \qquad (2)$$

Where $t_{handoff}$ is the total handoff delay of MIP over WLAN, $t_{L2handoff}$, $t_{L3handoff}$, and $t_{L4handoff}$ are the handoff cost of Layer2, Layer3, and Layer4 separately.

Layer2 handoff time can be broken into three parts: movement detection, AP searching and reassociation[8], which are shown in equation 3. The detail analysis of three phases of Layer 2 handoff is not covered in this paper.

$$t_{L2handoff} = t_{L2detection} + t_{L2seraching} + t_{L2reassociation} \qquad (3)$$

Where $t_{L2detection}$, $t_{L2seraching}$ and $t_{L2reassociation}$ are the time costs for Layer2 movement detection, Layer2 AP searching and Layer2 reassociation.

Only after the layer 2 link has been established could the Layer 3 handoff start, because the MN can only communicate with the FA on the same link[9]. The Layer 3 handoff involves 2 phases: agent discovery and registration.

$$t_{L3handoff} = t_{mipagentdicovery} + t_{mipregistration} \qquad (4)$$

TCP is a connection-oriented, end-to-end reliable protocol designed to support error recovery and flow control. Reliability is ensured by a sliding-window acknowledgement and re-transmission mechanism. All data sent by TCP must be acknowledged by the receiver. TCP maintains a variable-sized window of data that is unacknowledged for a given time. If the window is full, no data will be sent until an acknowledgement is received. TCP maintains a Retransmission Time Out (RTO) timer. If no ACK has been received when the RTO timer expires, TCP assumes that the data has been lost and retransmits all of the data in the window. The re-transmission follows the exponential back-off algorithm. According to this algorithm TCP doubles the timeout value on unsuccessful successive retransmissions [10]. In this case, during the Layer2 and layer3 handoff, the TCP doubles the retransmission timeout value several times. Therefore, even after the layer2 and layer3 handoff is over, TCP still has to wait for RTO to timeout to recover the retransmission. This latency is caused by TCP exponential back-off algorithm. We call it TCP back-off delay $t_{tcp\text{-}back\text{-}off}$.

We define $t_{L4handoff} = t_{tcp\text{-}back\text{-}off}$ $\qquad (5)$

We examine the data captured in the scenario shown in Fig.2. Table I gives 20 instances of experimental data. Each row is one expriment. Each column is the time latency for that handoff phase. The data in the last column are the total handoff latencies for every experiment. The number in the bottom right cell is the overall average handoff latency.

TABLE I
HANDOFF LATENCIES OF MIP OVER WLAN

Delay exp #	L2 movement detection	L2 AP searching	L2 reassociation	MIP agent discovery	MIP registration	TCP backoff	Handoff delay
1	1.033	0.061	0.005	2.996	0.073	5.058	9.226
2	1.064	0.044	0.009	1.945	0.042	6.01	9.511
3	1.133	0.063	0.006	3.023	0.052	5.345	9.622
4	1.032	0.100	0.008	2.563	0.050	5.323	9.076
5	1.044	0.065	0.003	2.756	0.052	5.125	9.045
6	1.131	0.057	0.004	2.578	0.043	5.004	8.817
7	1.009	0.056	0.010	2.436	0.060	5.625	9.196
8	1.120	0.060	0.006	3.001	0.704	5.002	9.893
9	1.023	0.059	0.026	2.213	0.054	4.998	8.373
10	1.039	0.076	0.005	3.008	0.053	5.006	9.187
11	1.100	0.045	0.030	2.770	0.041	5.728	9.714
12	1.013	0.049	0.010	2.545	0.042	4.768	8.427

13	1.021	0.051	0.009	3.001	0.065	5.202	8.896
14	1.006	0.043	0.017	2.600	0.046	5.312	9.024
15	1.104	0.069	0.006	2.598	0.047	4.544	8.368
16	1.003	0.064	0.013	2.674	0.062	4.806	8.622
17	1.110	0.054	0.010	2.783	0.054	5.705	9.716
18	1.100	0.064	0.006	3.012	0.057	5.602	9.841
19	1.302	0.056	0.009	2.349	0.070	5.71	9.496
20	1.098	0.044	0.004	2.404	0.062	5.172	8.784
Avg	1.074	0.059	0.010	2.660	0.086	5.253	9.142

5. Speed adaptive MIP

The above quantitative analysis of handoff latencies shows the largest part is the TCP back-off delay $t_{tcp\text{-}back\text{-}off}$. Because of TCP exponential back-off algorithm, if we reduce the L2 and L3 delay, $t_{tcp\text{-}back\text{-}off}$ will be reduced exponentially. The next largest part is L3 latency. In this paper, we start with L3 latency, and L2 and L4 latencyies will be discussed later. In section3, we define MN handoff rate as $r_h = v / d$, which means how many APs or FAs the MN moves through per second. Equation 1 shows that the performance of MIP over WLAN depends on the MN handoff rate. r_h is also equal to the ratio of $K_{handoff}/T_{travel}$, where $K_{handoff}$ is the number of handoffs occurred during the MN traveling. T_{travel} is MN's total travel time. To reduce r_h without changing total travel time, we can reduce the number of handoffs. The optimal is $K_{handoff} = 0$

Let N be the total FA numbers on the way MN travels. Let's assume somehow M is the number of FAs with whom the MN can communicate without L3 handoff delay. The optimal is let M = N. But this costs too many resources, especially when the number of active MNs is large. Also we don't know how long the MN will travel at the beginning.

We call M the size of the FA Set with whom the MN can communicate without L3 handoff delay. From IP level of view, M is the number of FAs that MN has registered to and can communicate with at that moment.

Now the question is:

1. How to decide FA set size M
2. How to guarantee MN can communicate with a FA set almost like to do with a single FA.

For question 1, equation 6 gives the FA set size.

$$M = \left| t_{handoff} \times r_h \right| + 1 \qquad (6)$$

In this equation, $t_{handoff}$ is the handoff time for every handoff procedure, and r_h is the handoff rate. Here, we use the experimental average handoff time 9.142s for $t_{handoff}$. r_h is dynamic.

For question 2, our solution is to let MN pre-register M potential FAs along the way MN travels, at the same time let IP packets be multicasted to those M FAs in this FA set. So MN will not experience any handoff delay from the IP level of view.

In SA-MIP, the set of FAs that MN can talk to without L3 latency is extended from one point at low moving speed to a line at high moving speed. The length of the line dynamically changes with the MN handoff rate. The behavior of SA-MIP will automatically adapt to the handoff rate of the MN so that the performance of SA-MIP won't decline dramatically in a rapid moving environment. At the same time, SA-MIP only cost reasonable amount of resource that is appropriate for seamless handoff.

Speed detection and location tracking is an interesting topic on mobile computing. Bahl [11] and Youssef[12] both utilize signal strength information to locate and track wireless users. Wijngaert and Blondia [13] use GPS to inform mobile users about the prospective future location and to improve performance of the ad hoc routing. In this paper, we assume the MN has GPS system to detect its location. When the MN moves at speed v, if $v \leq 30$m/s (108km/h), it performs a normal registration. If $30\text{m/s} < v \leq 40\text{m/s}$ (144km/h), it initializes registration after receiving two successive agent advertisements. If $v > 40$m/s, it is assumed that the MN does not change its direction significantly in a short distance. It initializes registration once it gets a new agent advertisement. MN's registration message is extended by speed extension. According to Mobile IP Vendor/ Organization-Specific-Extensions[14], two types of extensions are allowed for MIP, Critical (CVSE) and Normal (NVSE) Vendor/Organization Specific Extensions. The basic difference is when the CVSE is encountered but not recognized, the message containing the extension must be silently discarded, whereas when a NVSE is encountered but not recognized, the extension should be ignored, but the rest of the Extensions and message data must still be processed. We use the NVSE extension to extend MIP with handoff rate information.

Whenever the MN needs to handoff to a new FA set, after it receives certain times of agent advertisements according to the MN moving speed(step 1 in Fig. 6), it sends a registration request with up-to-date handoff rate information to the very first FA in a new FA set(step 2). The first FA relays the registration request to upper FA or HA(step 3). Meanwhile, it decapsulates the speed extension, refills the MIP header and authentication extension and then forwards it to other FAs(M-1 FAs) in this FA set(step 4). These other FAs relay the registration request to upper FA or HA, in the same way the request comes from the MN (step 5). When the GFA or HA receives these registration requests, it builds up tunnels downwards to each FA and responses with registration reply (step 6 and 7). When the FA receives the registration reply, it builds up tunnel upwards to the GFA or HA. Whenever the MN sets up the Link-layer contact with the FA, the later forwards the registration reply to the former (step8, 9 or 10). The MN gets the care-of-address from agent advertisement message (step

10 or 9) or registration reply message (step 9 or 10), then begins data communication. At the same time, it sends registration requests to the new FA with up-to-date speed information (step 11). This new FA decapsulates the message, sets up a new FA set, forwards the request (12,13) and repeats the above process. In Fig.6, the FA set size M changes from 2 to 3 when the MN handoff rate changes from 0.08 to 0.12.

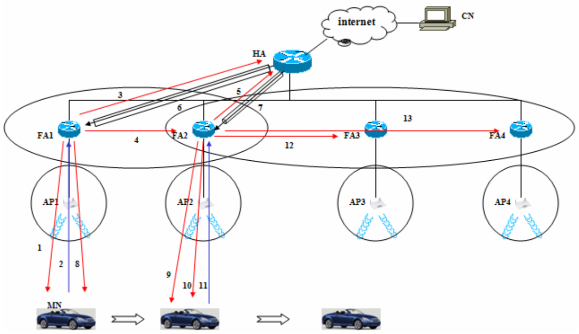

Fig. 6 Speed adaptive handoff procedure

6. Perfomance evaluation of SA-MIP

We evaluate the performance of speed-adaptive MIP over WLAN under the same scenario as in Fig.2 except the SA-MIP is installed. Fig.7 is the average throughput vs. handoff rate before and after the SA-MIP is installed. At handoff rate 0.02 FA/s, the average throughput is improved by (100.12 – 92.72)/ 92.72 = 7.98%. At handoff rate 0.04, 0.06 and 0.08 FA/s, the average throughput is improved by 12.99%, 16.81% and 26.45% respectively.

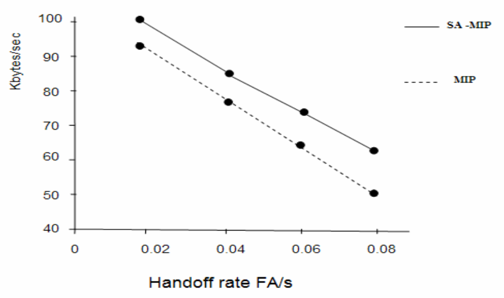

Fig. 7 Performance of Speed-Adaptive MIP

7. Conclusion

Compared to the mechanisms of Malki[1] and Koodli's mechnism[2], SA-MIP combines the pre- and post-registration methods, but keeps indenpendency from L2 infrastructure. Compared to Hsieh[3] and Wijngaert's mechnism[4], SA-MIP not only predicts its next move but also involves next M number of FAs according to MN's moving speed. The emulation shows that the SA-MIP can improve the performance from 8% to 27% when the handoff rate changes from 0.02 FA/s to 0.08 FA/s. In this paper, SA-MIP only deal with L3 handoff latency. But there is still physical link break from the Layer 2 handoff. In addition, we notice that even in SA-MIP, the biggest part of handoff delay is still the layer4 TCP back-off-latency. In future works, we plan to apply the speed adaptive scheme to layer 2 and layer 4 handoff latencies.

REFERENCES:

[1]. K. El Malki et al., "Low latency handoffs in mobile IPv4",IETF draft-ietf-monileip-low latency - handoffs -v4-04.txt (2002).

[2]. Koodli, R. (eds.). "Fast Handovers for Mobile IPv6. Internet Draft", IETF, Sep 2002.

[3]. Hsieh, R., Zhou, Z.-G., and Seneviratne, A. S-MIP: A Seamless Handoff Architecture for Mobile IP. In Proceedings of INFOCOM, San Francisco, 2003.

[4]. Van den Wijngaert, N. and Blondia, C.: "An Urban Mobility Model and Predictive Handover Scheme for Mobile IP", Proceedings of OPNETWORK 2004, Washington D.C., (2004).

[5]. C.Perkins,RFC3344,"IP Mobility Support for IPv4"

[6]. E. Hernandez and Sumi Helal, "RAMON: Rapid Mobility Network Emulator," Proceedings of the 27th Annual IEEE Conference on Local Computer Networks (LCN), November 2002, Tampa, Florida

[7]. http://dynamics.sourceforge.net/?page=main

[8]. Héctor Velayos and Gunnar Karlsson "Techniques to Reduce IEEE 802.11b Handoff Time" IEEE ICC 2004, Paris, France, June 2004.

[9]. N. A. Fikouras, A. J. Könsgen. "Accelerating Mobile IP Hand-offs through Link-layer Information". In Proceedings of the International Multiconference on Measurement, Modelling, and Evaluation of Computer- Communication Systems (MMB), Aachen, Germany, September 2001.

[10]. P. Hsieh and A. Seneviratne. "A Comparison of Mechanisms for Improving Mobile IP Handoff Latency for End-to-End TCP".MobiCom'03 San Diego,CA,Sep.2003.

[11]. Paramvir Bahl and Venkata N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," IEEE Infocom 2000, volume 2, pages 775-784, March 2000.

[12]. Moustafa Youssef, Ashok Agrawala, "Small-Scale Compensation for WLAN Location Determination Systems," IEEE Wireless Communications and Networking Conference (WCNC) 2003 New Orleans, Louisiana, March 16-20, 2003.

[13]. Ergen M., Coleri S., Dundar B., Puri A., and Varaiya P., "Fast Handoff with GPS routing for Mobile IP", IPCN 2002, April 2002, Paris, France.

[14]. RFC3115:MIPVendor/Organization-SpecificExtensions