IEEE P802
Handoff ECSG

Handoff for Multi-interfaced 802 Mobile Devices

Date: May 12, 2003

Authors: **Huai-An (Paul) Lin**
Intel Corp.
2111 NE 25th Avenue, OR 97124
Phone: +1 503-264-6726
E-mail: huai-an.lin@intel.com

Abstract

This document describes the engineering requirements for IEEE 802-based access networks in support of mobile devices that can maintain simultaneous connectivity with multiple Access Points (APs) and Access Routers (ARs). Simultaneous connectivity, coupled with proper network engineering, provides a relatively simple means to overcome communication disruption during handoff. A parameter, the handoff distance, is introduced as the measure for network overlap. This document analyzes the handoff distances, at both data-link and IP layers, based on the packet delivery rates, the moving speeds of mobile devices, the upper bounds of one-way packet delay, and the handoff procedures. A brief feasibility analysis is also included in this document.

Introduction

This document examines the handoff requirements for Mobile IP mobile nodes (MNs) that can maintain simultaneous connectivity with multiple Access Points (APs) and Access Routers (ARs). Connectivity disruption due to handoff leads to packet loss, creating a major barrier for real-time applications in mobile environments. Simultaneous connectivity, coupled with proper network engineering, provides a relatively simple means to overcome the disruption problem. The focus of this document is to examine how coverage overlap should be created between adjacent APs or ARs to reduce packet loss during handoff.

Our investigation is based on the basic mobility protocols, at both IP and data-link layers. For L3, this means that only the base Mobile IP [13][14] is assumed. In L2, a generic procedure, based on the 802.11 handoff, serves as the basis for the analysis. However, we expect that multi-interfaced MNs will be able to take advantage of various enhancements, such as the micro-mobility protocols [2][3][4][5][6], context transfer [11][12] and pre-authentication [15], that are being discussed in IETF and IEEE (including this forum).

Consistent with the common wireless network architecture, we assume that an access network consists of a number of subnets, each of which is managed by an AR. A subnet also has several APs that provide the Layer 2 connectivity for MNs. An L2 handoff occurs when a MN's L2 connectivity is changed from one AP to another, and an L3 handoff changes the MN's IP layer connection to a new AR. To allow a MN to establish simultaneous L2 connectivity, the ranges of adjacent APs need to partially overlap. We call such an overlapping area the *overlapping zone* of the involved APs. Similarly, the covering area of neighboring ARs should partially overlap; we call such a geographic area the *overlapping region* of the ARs.

The size of an overlap (overlapping region or overlapping zone) is the concern of this document. If the overlap is sufficiently large, then it is possible for a multi-interfaced MN to complete a handoff before it loses the connectivity with the previous AR (or AP). On the other hand, overly large overlap increases the equipment cost. To help determine the desirable size of overlap, we will use the term *handoff distance* to represent the maximum distance that a MN can travel from the moment it makes the initial move to contact a candidate AR or AP (concerning the handoff) until it fully establishes the connectivity with a new AR or AP.

Handoff distance is determined by: (1) the packet delivery rates (how much packet loss is to be tolerated); (2) the moving speeds of MNs; (3) the upper bounds of one-way packet delay; (4) the handoff procedures[1]; and (5) the processing delays within MNs, APs, ARs and authentication servers. We will discuss how the first four variables impact the handoff distance. The last variable is highly implementation-dependent and therefore the analysis is out of the scope of this document.

Packet Delay Characteristics

Two recent studies provide helpful insight into packet delay characteristics [8][9]. One study measured the one-way packet delays between two end points over an extended period of time [9]. Packets of fixed length were dispatched periodically over sequential 300-second measurement periods, each of which created a separate measurement record. The end points were GPS synchronized to allow one-way latency to be measured with high accuracy.

Two sets of experiments were conducted. In one set (the local experiments) the end points, both in Irvine (California), communicated across a T1 Internet connection. While the two end points were physically close, the path actually traversed through 11 hops that also served other cross traffic. To complement the physically local data, the other set (the remote experiments) measured the end-to-end delays over a trans-Atlantic path. This transmission path, with one end in Irvine and the other in London (England), consisted of 22 hops.

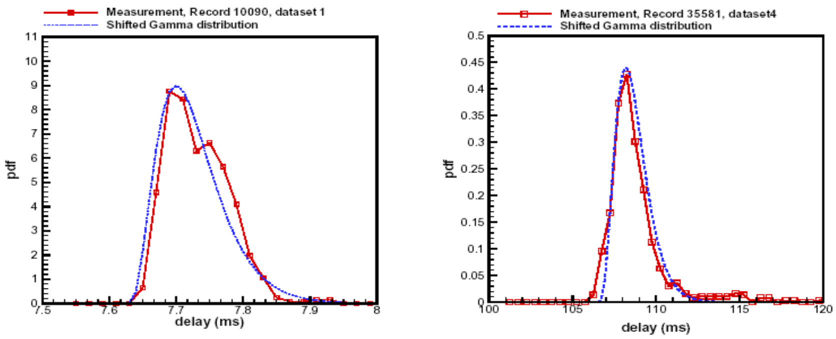
The other study performed by the RIPE (Réseaux IP Européens) Test Traffic Measurements (TTM) Project, measured packet delays over a large number of paths [8]. The TTM Project distributed about 40 test boxes over Europe (and a few in the US and New Zealand) and measured the one-way delays over 963 paths. As with the previous study, the TTM Project also synchronized the two ends of each transmission path through GPS. The large number of paths measured by the TTM Project offers a greater level of confidence of the end-to-end packet delay characteristics.

Both studies found that the vast majority of end-to-end packet delay is well approximated by a shifted gamma distribution.[2] Figure 1, reproduced from [9], shows a typical one-way latency record of the first study. The mean delays are 9.6 ms and 108.2 ms for the local and the trans-Atlantic paths, respectively. An important finding is that the delays show very sharp peaked distributions, with most delay values clustered within about 10% of both the mean and minimum values. This implies that excessive delays are in reality quite rare. Figure 2, also from [9], indicates that 99.9% of delays occur within 30 ms and 115 ms, respectively, for the said paths. Another important observation is that the trans-Atlantic path actually exhibits substantially smaller standard variation (3.083) than the local path (6.216). This is probably due to the effect of smoothing produced by a relatively large number of router hops over the international route.

---

[1] As mentioned, we will limit the discussion to the base Mobile IP and a generic L2 handoff procedure.

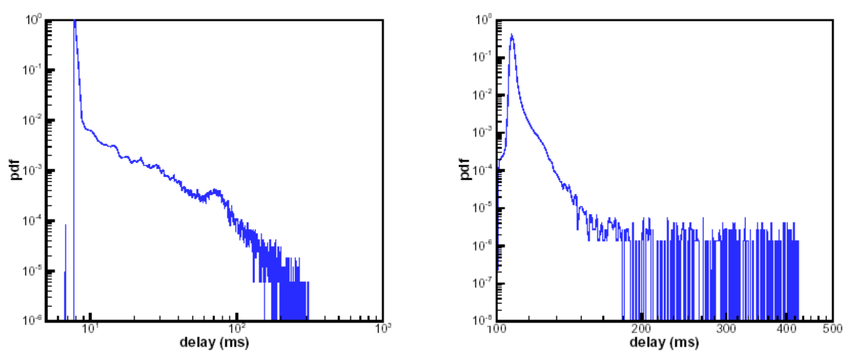
[2] An earlier analysis [10] also found that the round-trip delays follow the gamma distribution.



(a) **(b)**

Figure 1: Example of one-way latency distribution [9]. (a) The local path (both end points in Irvine, California). (b) The trans-Atlantic path (end points in Irvine and London).

The narrow distributions of delay, characterized by the sharp peaks, were also observed in the RIPE TTM Project. The vast majority of the measured paths exhibited the gamma-like shape, with the shape parameter being typically around 0.6 – 1.0 [8]. Worth mentioning is the fact that some paths in the TTM Project presented non-gamma behaviors, such as a gamma-like distribution with Gaussian lob or multiple peaks. The anomalies were cited to be likely caused by route changes during the measurements. However, such results do not negate the general pattern that the delays on a specific route follow a sharp peaked distribution, and that the frequency of excessive delay is very low. The vast majority of paths in the TTM Project had minimum delays below 100 ms, including a path that had the unusually long route-plan distance of 16154 Km. Only a very small number of paths had the minimum delays exceed 100 ms, perhaps due to some low-capacity router on the route [8].



(a) **(b)**

Figure 2: One-way latency distribution over all measurement records [9]. (a) The local path. (b) The trans-Atlantic path.

The implication of these two empirical one-way packet delay studies on handoff is two-fold:

1. For L3 handoff, 115 ms appears to be a reasonable upper bound for the one-way delay between a MN and its Home Agent (HA) and Corresponding Nodes (CNs). The base Mobile IP requires that when a MN changes its Care-of-Address (COA) it must informs the HA and perhaps the CNs. Based on the studies as discussed, an upper bound of 115 ms seems to be more than adequate, assuming the access network is reasonably engineered. For services provisioned only for metropolitan or regional areas, the upper bound can be significantly lower.
2. For L2 handoff, 10 ms should be a realistic upper bound. L2 handoff signals are exchanged locally; consequently, the one-way packet delay should be much lower. As shown in one of the afore-mentioned studies, a local path not particularly well-engineered (T1 connection over 11 intermediate hops) has a mean delay less than 10 ms.

L3 Handoff Distance

To take advantage of the overlapping coverage of ARs (overlapping region), a MN may immediately acquire a new COA when it moves into an overlapping region between the current AR and a Candidate AR (CAR). This prepares the MN for a handoff. The MN may also check if it has the credentials to connect to the new subnet. To avoid prematurely kicking off the handoff process, the MN should initiate the handoff only after the signal strength from the CAR becomes stronger than the current AR. This implies that the handoff is initiated roughly after the MN is half way into the overlapping region.[3] This procedure based on Mobile IP is illustrated in Figure 3.

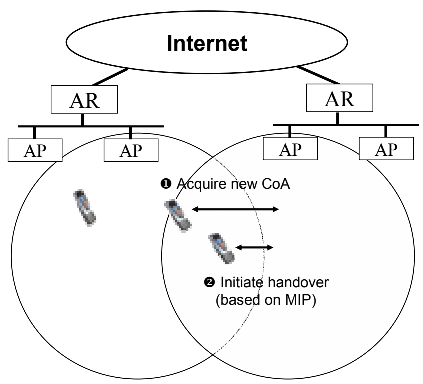


Figure 3: (MIP-based) L3 Handoff for Multi-Interfaced MN

Under this scenario, the L3 handoff distances, for a targeted 99.9% packet delivery rate, are listed in Table 1.[4] Four maximum moving speeds are considered: 10 Km/h for hot-spot deployment, 60 Km/h for normal mobility in metropolitan areas, and 120 Km/h and 180 Km/h for high-speed traveling (such as on freeways).

[3] The analysis herein serves only as an illustration. It assumes certain “normal” conditions: no use of directional antennae, and roughly equal signaling strengths emitted by the adjacent AP’s. The calculation of handoff distance should be adjusted under different conditions.

[4] Each of the values in the table is calculated to achieve 99.9% packet delivery for the longest path in the supported group, and for the fastest moving clients. The average packet delivery rate over all the supported paths and mobile clients should be much higher than 99.9%.

What follows is an example of how two of the values in this table are derived. In this example, which almost represents a worst-case scenario, we let 180 Km (~110 miles) per hour be the moving speed and 115 ms the packet delay upper bound, respectively. Because of its target for supporting very high-speed and very long end-to-end delay, one should expect a fairly sizable overlapping region.

Two Mobile IP (MIP) change of COA procedures are considered. Without route optimization, it takes one round-trip for an MIP handoff, which involves an exchange of Binding Update and Binding Acknowledgement between the MN and the HA, to take effect. At 180Km/h, a MN can traverse 5.75 meters in distance in 115 ms or 11.5 meters in a round trip. However, the handoff distance should be doubled, since it is assumed that the handoff will not be initiated until the MN is half way into the overlapping region. As a result, the handoff distance should be at least 23 meters. When route optimization is supported, completion of a handoff involves the Return Routability Procedure [13] and a Binding Update. In total five one-way delays are required, including four one-way delays for the RR Procedure[5], followed by another one-way delay for the MN to inform Binding Update to the CN. In this case, the handoff distance should be 57.5 meters.

Table 1: L3 Handoff Distance for 99.9% Packet Delivery

L3 Handoff Distance (in meters)		Max MN Moving Speed w/o route optimization				Max MN Moving Speed with route optimization			
		10 Km/h	60 Km/h	120 Km/h	180 Km/h	10 Km/h	60 Km/h	120 Km/h	180 Km/h
Max. one-way packet delay	50 (ms)	0.6 (m)	3.3	6.7	10.0	1.4	8.3	16.7	25.0
	75	0.8	5.0	10.0	15.0	2.1	12.5	25.0	37.5
	115	1.3	7.7	15.3	23	3.2	19.2	38.4	57.5

L2 Handoff Distance

L2 handoff typically consists of a period of discovery, followed by authentication and re-association. During the discovery phase the MN detects the candidate APs to associate by either passively listening to potential APs (in range) or actively probing the nearby APs. In the authentication phase the MN attempts to authenticate itself to one or more candidate APs. The re-association phase involves transferring the credentials and other state information from the old-AP to the new AP(s).

Figure 4 shows the typical sequence of messages.[6] The process starts with the first probe request message and ends with a re-association response message from an AP.

Although not formally specified, it has been suggested that the probe phase can be initiated even before the MN moves into the overlapping zone (between the current AP and a new AP). The MN can either cache or deduce AP information without having to actually perform a complete active scan. Alternatively, the current AP can simply provide the candidate AP information to the MN. In an empirical study it is noted that the probe phase accounts for more than 90% of the overall handoff delay [1]. As such, a scheme that allows the client to obtain the information of candidate APs ahead of time will clearly stand to benefit from the dominating cost of the scan process. In

[5] We make a gross simplification by assuming one maximum round-trip between the MN and the HA, and another maximum one between the HA and the CN. Obviously, it is highly unlikely for the RR Procedure to require two 230ms round-trip delays.

[6] The process is typical in 802.11. It is quite generic and will be used as the generic L2 handoff procedure in the following analysis.

the following discussion we assume that the probe phase is conducted before the MN has moved into the overlapping zone and will not be considered in the handoff distance calculations.

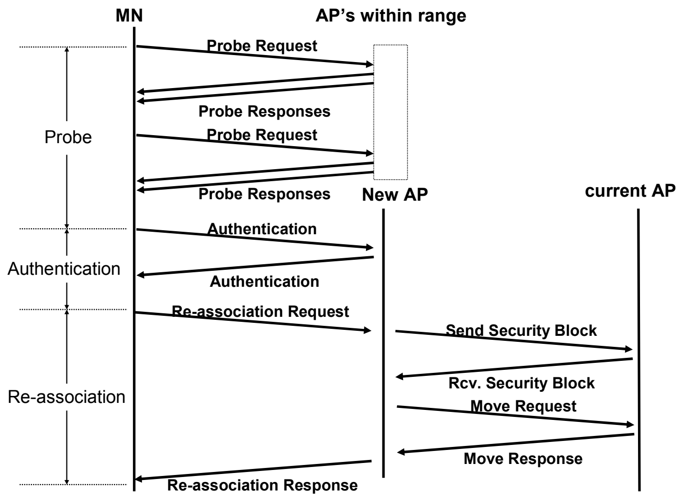


Figure 4: Generic L2 Handoff Procedure

We note that the authentication phase may also be initiated outside the overlapping zone. This can be done if the MN entrusts the current AP to relay authentication messages. However, pre-authentication handoff is still under discussion and we will assume that authentication and the subsequent re-association phase will not be initiated until the MN is in the overlapping zone.

Table 2 lists the L2 handoff distances, again for four maximum moving speeds and for a targeted 99.9% packet delivery rate. However, only 10 ms upper limit on one-way packet delay is assumed. As previously mentioned, during a L2 handoff signals are exchanged locally and 10 ms is a realistic upper bound. Following the same method as in the L3 analysis, one can get the L2 handoff distances as shown in Table 2.

Table 2: L2 Handoff Distance for 99.9% Packet Delivery

Max MN Moving Speed	10 Km/h	60 Km/h	120 Km/h	180 Km/h
L2 Handoff Distance (in meters)	2.4 (m)	13.3	26.7	40.0

Feasibility Considerations

As pointed out previously, the size of an overlap is the concern of this document. For a MN to complete a handoff while it still has the connectivity with the current AR or AP requires the overlap (overlapping region or overlapping zone) to be sufficiently large. However, it is also important to keep overlap small to control equipment cost. Based on the preceding analyses, one may ask: Can the access network be practically engineered to create the overlaps?

The coverage area of an 802.11 AP is typically 150-2000 feet without special antenna. Primarily used in hot-spots for low-speed roaming users, the moving speeds of the MNs are expected to be under 10 Km per hour. The handoff distances, even when doubled, tripled or quadrupled to accommodate the processing delays, seem to be small enough to be reasonable. For medium- or high-speed roaming users in metropolitan areas, 802.16e/20, with much larger coverage, seems to be a more feasible solution.

Another issue that requires further study is the “blind spot.” A blind spot is an area within the overlapping region (or overlapping zone), normally at the corners, where the diameter is less than the handoff distance. We believe that the end-to-end Internet packet delay characteristics, nonetheless, are still helpful, as they provide a strong incentive for developing a packet-forwarding scheme. Given that access routers (or access points) involved in a handoff should normally be fairly close to each other, we expect that adding the forwarding function should have a small impact on the overall packet latency. However, this issue also needs further analysis.

References

[1] A. Mishra, M. Shin and W. Arbaugh, “An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process,” Submitted to ACM CCR, 2002.

[2] A. Campbell, et al., “Comparison of IP Mobility Protocols,” *IEEE Wireless Communications*, Feb. 2002.

[3] A. Campbell, et al., “Design, Implementation, and Evaluation of Cellular IP,” *IEEE Personal Communications*, Aug. 2000.

[4] R. Ramjee, et al., “HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks”, IEEE/ACM Transactions on Networking, vol. 10, no. 3, June 2002.

[5] H. Soliman, et al., “Hierarchical Mobile IPv6 mobility management (HMIPv6),” Internet Draft, Internet Engineering Task Force, draft-ietf-mobileip-hmipv6-07.txt, Work in Progress.

[6] A. Lopez, et al, “BRAIN architecture specifications and models, BRAIN functionality and protocol specification,” IST-1999-10050 BRAIN D2.2, March 2001

[7] P. Roberts and J. Loughney, “Local Subnet Mobility Problem Statement,” Internet Draft, Internet Research Task Force, draft-proberts-local-subnet-mobility-problem-02.txt, Work in Progress.

[8] C. Bovy, et al., “Analysis of End-to-End Delay Measurements in the Internet,” RIPE41, January 2002, Amsterdam.A. Corlett, et al., “Statistics of One-Way Internet Packet Delays,” 53rd IETF, Minneapolis, March 2002.

[10] Mukherjee, A., “On the dynamics and significance of low frequency components of internet load,” Journal of Internetworking: Research and Experience, vol. 5, no. 4, pp. 163-205, Dec. 1994.

[11] J. Loughney (ed.), “Context Transfer Protocol,” Internet Draft, Internet Engineering Task Force, draft-ietf-seamoby-ctp-00.txt, Work in Progress.

[12] M. Liebsch & A. Singh (ed.), “Candidate Access Router Discovery,” Internet Draft, Internet Engineering Task Force, draft-ietf-seamoby-card-protocol-00.txt, Work in Progress.

[13] D. Johnson, C. Perkins and J. Arkko, “Mobility Support in IPv6,” Internet Draft, Internet Engineering Task Force, draft-ietf-mobileip-ipv6-20.txt, Work in Progress.

[14] C. Perkins (ed.), “IP Mobility Support for IPv4,” RFC 3344, August 2002.

[15] B. Aboba, “IEEE 802.1X Pre-Authentication,” June 2002.