Performance Evaluation of Opportunistic Vertical Handover Considering On-Off Characteristics of VoIP Traffic

Hyun-Ho Choi, Ohyoung Song, Yeon-Kyung Park, and Jung-Ryun Lee, *Member, IEEE*

***Abstract*—Vertical handover is one of the key technologies to support seamless connectivity across multiple wireless communication systems and guarantee the quality of service (QoS) for the applications therein. This paper proposes an opportunistic vertical handover scheme for a voice of IP (VoIP) connection, which makes use of the on-off characteristics of voice traffic by aligning the mutual silence period of the two parties engaged in conversation with the service disruption time that occurs during the vertical handover procedure. From the six-state Brady model, we derive a simplified two-state Markov model in which the characteristics of both the talk-spurt period and the mutual silence period are obtained. We then analyze the performance of the proposed scheme with respect to the handover execution time and the packet loss time. The analysis and simulation results show that the proposed scheme significantly decreases the VoIP packet loss time, while the total time required for the vertical handover procedure increases but remains within a tolerable time limit.**

***Index Terms*—Vertical handover, seamless mobility, voice over IP (VoIP), quality of service (QoS).**

I. Introduction

FUTURE wireless communication systems for Beyond 3rd Generation (B3G) or 4th Generation (4G) aim to provide users with ubiquitous information access capabilities, various wireless multimedia applications, and quality of service (QoS) guarantees. To realize the goal of B3G, a generally accepted approach is to integrate currently existing various wireless networks such as IEEE 802.11 wireless local area networks (WLANs), IEEE 802.16 wireless metropolitan area networks (WMANs), General Packet Radio Service (GPRS), and Universal Mobile Telecommunications Systems (UMTS) because no single wireless network technology can simultaneously provide high bandwidth, low latency, low access cost, and wide area service to a large number of mobile users. In this context, the 3rd Generation Partnership Project (3GPP) and 3GPP2 have standardized the interworking requirements between 3G wireless cellular systems and WLAN systems to support the mobility of users roaming between both systems [1], [2].

One of the key technologies to provide seamless mobility in multiple access environments is the *vertical handover*. Vertical handover occurs when a service connection is changed between different types of access networks (AN). It is different from conventional horizontal handover in which the mobile node (MN) changes only its base station within the same AN. In the case of horizontal handover, the MN may perform only layer 2 (L2) handover procedures such as L2 detachment and L2 attachment. However, in vertical handover, the MN may perform not only L2 handover procedures, but also connection setup procedures such as authentication and L3 (IP) registration with a new AN.[1] Since the authentication and IP registration processes consume much more time than the L2 detachment and attachment processes, vertical handover results in longer handover latency and service disruption time. The service disruption may lead to packet loss of the ongoing service and deteriorate the QoS. Therefore, it is very important to provide seamless services during vertical handover [3], [4].

To make the vertical handover seamless, conventional approaches have focused mainly on the design of efficient signaling procedures together with packet buffering and forwarding functions. The Internet Engineering Task Force (IETF) has proposed fast handover mechanisms based on the Mobile IP protocol in both IPv4 and IPv6 networks [5], [6]. The basic idea of the fast handover mechanism is to use a pre-registration or post-registration technique, which reduces the handover latency as it separates L2 and L3 handovers and performs the L3 handover that requires significant time before or after the L2 handover. In addition, the *Context Transfer Protocol* (CXTP) and a smooth handover procedure supporting the pre-authentication process have been suggested to minimize the signaling procedures required after the L2 handover [7], [8]. With efficient signaling procedures, the fast handover mechanisms adopt buffering and forwarding functions to reduce packet loss during a handover [9]-[11]. That is, a network agent buffers packets destined for the MN during the service disruption time and forwards the buffered packets to the MN after the new connection path is set up.

Although existing fast handover mechanisms can improve the performance of vertical handover in terms of handover latency, they are still hindered by several problems including

Hyun-Ho Choi is with Communication Lab., Samsung Advanced Institute of Technology, Republic of Korea (e-mail: hyun.ho.choi@samsung.com).

Jung-Ryun Lee, Yeon-Kyung Park and Ohyoung Song are with the School of the Electrical Engineering, Chung-Ang University, Republic of Korea (e-mail: {jrlee, yeonkyung, song}@cau.ac.kr).

This work was partly supported by the IT R&D program of MKE/KEIT [KI002084, A Study on Mobile Communication System for Next-Generation Vehicle with Internal Antenna Array] and a grant (CR070019) from Seoul R&BD Program funded by the Seoul Development Institute of Korean government.

[1]In this paper, we assume a loosely coupled system as an early stage interworking solution.

QoS support and scalability. Since it is difficult to apply a soft handover technique for the vertical handover, service disruption is inevitable when the MN changes the physical air interface.[2] Moreover, the pre/post-operation and related tunneling process introduce additional signaling overhead to the network. The buffering function is limited by the storage space in the network agent. Therefore, the buffered packets may be dropped by buffer overflow. Also, the forwarding function is only effective for a small number of hops because multiple hops on a wired network cause a considerable propagation delay. In particular, these problems are more critical to QoS for real time services because of their strict end-to-end delay constraint.

In this paper, we focus on the handover mechanism associated with the traffic attributes of applications. Generally, the amount of packets transmitted on the air channel is time-varying due to the characteristics of traffic generation. For example, some applications such as VoIP codec do not generate any data intermittently according to its working mechanisms such as voice activity detection and silence suppression. This gives rise to a new approach with the vertical handover mechanism: executing the vertical handover when there are little or no packet(s) to be delivered. By allowing the vertical handover execution to be performed in such an opportunistic way, the effect of service disruption on the QoS of the given application is mitigated. As a consequence, this opportunistic vertical handover scheme can result in the high satisfaction of the QoS requirement of the application.

The rest of this paper is organized as follows. In Section II, the proposed vertical handover scheme is introduced. In Section III, the voice activity model of two-way conversation is derived. In Section IV, the proposed handover scheme is evaluated. Section V presents analysis and simulation results and Section VI summarizes the conclusions of this study.

II. Proposed Opportunistic Vertical Handover

VoIP traffic is highly susceptible to service disruption during handover. On the other hand, it has a high percentage of silence as enhanced voice codecs use a silence suppression scheme to prevent wasting bandwidth [12]-[14]. Thus, we adopt a VoIP telephony service as an application model for the proposed scheme. Fig. 1 illustrates the operation of the proposed opportunistic vertical handover considering the on-off characteristics of VoIP traffic in the context of the interworking scenario between 3G and WLAN systems. Since WLANs are overlapped by the coverage of a 3G system, the vertical handover is classified into two categories: *downward* and *upward* handovers [4]. The downward handover is a roaming to an AN with a smaller cell size and larger bandwidth (i.e., from 3G to WLAN), and the upward handover is the opposite (i.e., from WLAN to 3G).

[2]An MN with multi-interface can perform soft handover procedure by receiving information from multiple sources. However, this operation is generally burdens to both the MN and network because of limited battery power and radio resource. In this paper, we assume that the MN can turn on a single interface at a time, thus some service disruption always occurs during vertical handover procedures.

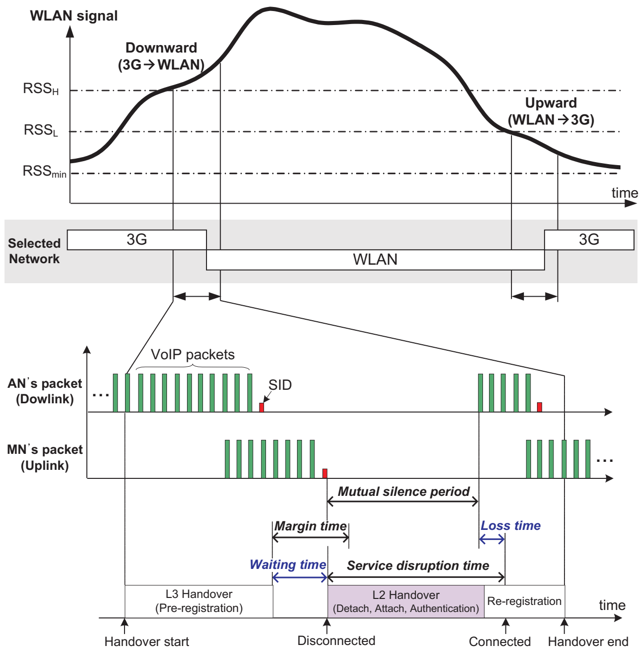

Fig. 1. Operation of proposed opportunistic vertical handover.

The vertical handover decision may depend on various parameters such as the received signal strength (RSS), minimum bandwidth, access cost, application type, delay required by the application, battery status of the MN, and the user's preference [15]. The proposed opportunistic vertical handover can be used together with any existing handover decision criteria mentioned above because the proposed algorithm is performed after the handover decision is already made. In practice, the MN prefers the WLAN to the 3G due to its low cost and high bandwidth, so we simply assume that the vertical handover decision depends on the RSS from only the WLAN channel [16]. Thus, the downward and upward vertical handovers are initiated by the threshold values of RSS_H and RSS_L, respectively. Here, RSS_H is greater than RSS_L to avoid ping-pong handovers. RSS_{min} is the minimum signal strength to maintain a connection with the WLAN.

We assume that the pre-registration technique is applied for the vertical handover execution. The MN performs the L3 handover (i.e., pre-registration), L2 handover (i.e., detachment with the serving AN, attachment with the target AN, and authentication) and L3 handover completion (i.e., re-registration) procedures in regular sequence to finish the vertical handover. Note that the MN cannot receive or transmit packets during the entire L2 handover time and some of the re-registration time until the buffered packets are forwarded to the MN, which correspond to the service disruption time [8].

After finishing the pre-registration, the legacy vertical handover initiates the L2 handover procedure without delay if the decision criterion is still satisfied. However, the proposed scheme waits for the start of a mutual silence period (i.e., both conversation parties are silent) during the predetermined *margin time*, which is defined as the maximum time duration that an MN can delay its handover progress.[3] This alignment of the service disruption time with the mutual silence period has the potential to effectively reduce packet loss because no packet is transmitted during the mutual silence period.

The starting time of a mutual silence period is recognized by the arrival of a *Silence Insertion Descriptor* (SID) frame that is sent by the sender's VoIP codec at the beginning of the silent period [14]. This SID frame includes only information about the background noise that is used to generate artificial noise at the receiving side's decoder during the silent period. The SID frame is smaller in size than a VoIP data packet. Therefore, the MN and AN can identify the beginning of a mutual silence period by simply checking the size information in the received packet's header.

Once the MN recognizes the beginning of a mutual silence period during the margin time, it immediately initiates the L2 handover procedure. In the case that a mutual silence period does not occur within the margin time, the MN initiates the L2 handover procedure promptly after margin time expires so as not to excessively increase the handover execution time. The following pseudo-code summarizes the algorithm of the proposed vertical handover.

```
loop
  if RSS > RSS_H or RSS < RSS_L then
    execute pre-registration
    set timer t = 0;
    while t < margin time
      if mutual silence occurs then
        waiting time = t;
        break;
      else
        t = t + δ;      /* δ is time increment */
      end if
    end while
    execute L2 handover
    execute re-registration
  end if
end loop
```

III. Voice Activity Model

To derive the on-off characteristics of the VoIP telephony service, we consider the Brady model, which is a general six-state model that provides good statistical analysis of two-way conversation [18], [19]. Fig. 2 shows the Brady model and the values of the state transition parameters. The state transition parameter, $\alpha_{i,j}$, refers to the transition rate from state i to state j during a unit time period. Thus, Pr{Transition from state i to state j during dt}= $\alpha_{i,j} \times dt$. This figure is divided into quadrants, each of which represents a different state for parties A and B engaged in a conversation. Note that the state transitions for party A have the same characteristics as those for party B. Let π_i be the steady state probability that the Markov state stays in state i in the long run, and let P be the transition rate matrix derived from the Brady model (i.e., $P_{i,j} = \alpha_{i,j}$). π_i is obtained by the balance equation $\Pi = \Pi \cdot P$ and the normalized condition $\sum_{i=1}^{6} \pi_i = 1$, where $\Pi =$

[3] In the vertical handover, some time margin during the handover execution is allowable compared to the horizontal handover because the MN performing the vertical handover generally has a low mobility at the pedestrian level [17].

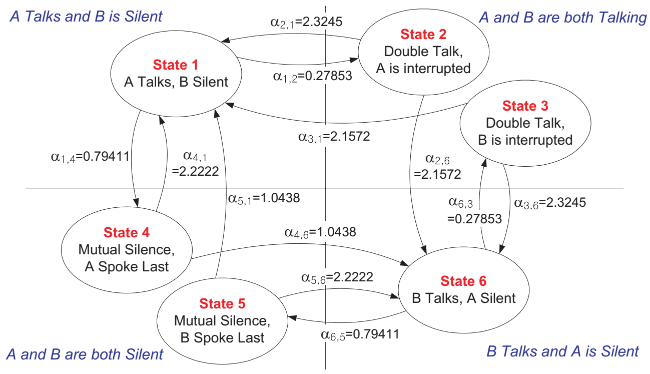

Fig. 2. Brady model for two-way conversation.

$[\pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5\ \pi_6]$ is the probability vector. The results from this computation show that the ratio of talk-spurt periods to the total conversation time is $P_T = \pi_1 + \pi_2 + \pi_3 + \pi_6 = 0.81$ and that of mutual silence periods is $P_S = \pi_4 + \pi_5 = 0.19$.

We assume that the uninterrupted length of each state i in the Brady model follows an exponential distribution with a rate of λ_i [19]. Each λ_i is given by $\lambda_1 = \alpha_{1,4} + \alpha_{1,2}$, $\lambda_2 = \alpha_{2,1} + \alpha_{2,6}$, $\lambda_4 = \alpha_{4,1} + \alpha_{4,6}$, and by the symmetry of the model, $\lambda_3 = \lambda_2$, $\lambda_5 = \lambda_4$ and $\lambda_6 = \lambda_1$. Let T_i and N_i be the average uninterrupted length of state i and the average number of visits to state i during the VoIP call duration, respectively. Then, $T_i = 1/\lambda_i$ and $N_i = D\pi_i/T_i = D\pi_i\lambda_i$ where D is the entire duration of a VoIP call.

As shown, the talk-spurt consists of states 1, 2, 3 and 6 and the mutual silence consists of states 4 and 5. Since there is no direct transition between the state 4 and state 5, the number of visits to the mutual silence state becomes $N_4 + N_5$. Therefore, the average uninterrupted length of the mutual silence state is expressed as

$$T_S = \frac{T_4 N_4 + T_5 N_5}{N_4 + N_5} = T_4 = \frac{1}{\lambda_4} = \frac{1}{\alpha_{4,1} + \alpha_{4,6}} \tag{1}$$

where $T_4 = T_5$ and $N_4 = N_5$ by the symmetry of the model. Moreover, since the talk-spurt and the mutual silence occur alternatively, the number of visits to the talk-spurt state is the same as that to the mutual silence state (i.e., $N_4 + N_5$). Thus, the average uninterrupted length of talk-spurt state is obtained by

$$\begin{aligned} T_T &= \frac{T_1 N_1 + T_2 N_2 + T_3 N_3 + T_6 N_6}{N_4 + N_5} = \frac{T_1 N_1 + T_2 N_2}{N_4} \\ &= \frac{D\pi_1 + D\pi_2}{D\pi_4 \lambda_4} = \frac{\pi_1 + \pi_2}{\pi_4(\alpha_{4,1} + \alpha_{4,6})}. \end{aligned} \tag{2}$$

From the considerations mentioned above, we remodel the six-state Brady model into a simplified two-state Markov model. The probability density functions (pdf) of the talk-spurt period and mutual silence period are respectively expressed as

$$f_X(x) = \lambda_T e^{-\lambda_T x}, \quad f_Y(y) = \lambda_S e^{-\lambda_S y} \tag{3}$$

where the rate parameters calculated using (1) and (2) are $\lambda_T = 1/T_T = 0.7476$ and $\lambda_S = 1/T_S = 3.266$, respectively

TABLE I
PARAMETERS DEFINITION

Notation	Meaning
S	Random variable denoting the time when the handover occurs
S'	Random variable denoting the waiting time occurrence
X	Random variable denoting the length of talk-spurt period ($X \sim$ Exp(λ_T))
Y	Random variable denoting the length of mutual silence period ($Y \sim$ Exp(λ_S))
P_T	Ratio of talk-spurt periods ($P_T = \frac{\lambda_S}{\lambda_T+\lambda_S}$)
P_S	Ratio of mutual silence periods ($P_S = \frac{\lambda_T}{\lambda_T+\lambda_S}$)
D	Duration of VoIP call
N	Number of cycles of talk-spurt and mutual silence during D ($N = \frac{D}{(1/\lambda_T+1/\lambda_S)}$)
T_m	Margin time
T_d	Service disruption time
T_w	Waiting time
T_l	Packet loss time
T_{pre}	Time needed for pre-registration
T_{re}	Time needed for re-registration

($T_T = 1.337$ s and $T_S = 0.306$ s).[4]

IV. Performance Evaluation

We derive the waiting time and the packet loss time as the performance metrics. The waiting time is the time difference between the end of pre-registration and the start of the L2 handover procedure whose maximum value is the margin time (see Fig. 1). The loss time is defined as the time fraction of a service disruption period during which VoIP packets are transmitted (i.e., the talk-spurt period). Hence, the VoIP packets are actually lost during this loss time. The notations used in this analysis are given in Table I. Here, the handover event S has a uniform distribution because the vertical handover happens randomly throughout the VoIP call duration. By definition, $S' = S + T_{pre}$ and S' also has a uniform distribution assuming the pre-registration time, T_{pre}, is fixed.

A. Waiting Time

To analyze the waiting time, we divide the talk-spurt periods into three time zones, as shown in Fig. 3. Here, we do not consider that the event S' occurs in a mutual silence period because its waiting time is zero. There are three possible cases with different waiting times, as follows:

- Case 1: The event S' occurs in a talk-spurt period and the mutual silence does not occur within the margin time (i.e., S' happens in zone 1).
- Case 2: The event S' occurs in a talk-spurt period whose length is longer than T_m and the mutual silence occurs within the margin time (i.e., S' happens in zone 2).
- Case 3: The event S' occurs in a talk-spurt period whose length is shorter than T_m. (i.e., S' happens in zone 3).

Let n_i denote the total number of occurrences of zone i during D where $i \in \{1,2,3\}$ and $d_{i,j}$ denote the length of j-th zone i where $1 \leq j \leq n_i$. Also let T_w^i be the average waiting time in the case i and P_w^i be the probability that the case i occurs during D.

[4]The validation of the two-state Markov model was achieved by comparison with the simulation experiment based on the original six-state Brady model, but not presented in this paper because of the page limit.

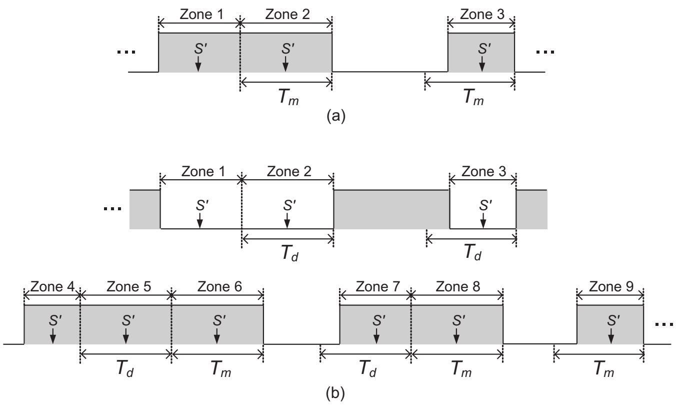

Fig. 3. Analyses of (a) waiting time and (b) loss time.

1) Case 1: It is obvious that the waiting time is equal to the margin time in this case. So simply $T_w^1 = T_m$. On the other hand, P_w^1 is expressed as

$$P_w^1 = \frac{\sum_{j=1}^{n_1} d_{1,j}}{D} = \frac{n_1 \overline{d}_1}{D} \tag{4}$$

where $\overline{d}_1$ is the average of $\{d_{1,j}\}_{\forall j}$. Here, n_1 and $\overline{d}_1$ are respectively calculated as

$$n_1 = N \cdot P[X \geq T_m] = N e^{-\lambda_T T_m} \tag{5}$$

$$\overline{d}_1 = E[X | X \geq T_m] - T_m = \frac{\int_{T_m}^{\infty} x f_X(x) dx}{\int_{T_m}^{\infty} f_X(x) dx} - T_m = \frac{1}{\lambda_T}. \tag{6}$$

Combining (5), (6) and the value of N in Table I, P_w^1 is finally obtained by

$$P_w^1 = \frac{N e^{-\lambda_T T_m} \cdot \frac{1}{\lambda_T}}{D} = \frac{\lambda_S}{\lambda_T + \lambda_S} e^{-\lambda_T T_m} = P_T e^{-\lambda_T T_m}. \tag{7}$$

2) Case 2: In this case, the event S' occurs uniformly in zone 2 whose length is always T_m, thus $T_w^2 = \frac{T_m}{2}$. Similar to the case 1, P_w^2 is obtained by

$$\begin{aligned} P_w^2 &= \frac{\sum_{j=1}^{n_2} d_{2,j}}{D} = \frac{n_2 \overline{d}_2}{D} = \frac{N e^{-\lambda_T T_m} \cdot T_m}{D} \\ &= \frac{\lambda_T \lambda_S}{\lambda_T + \lambda_S} T_m e^{-\lambda_T T_m} = P_T \lambda_T T_m e^{-\lambda_T T_m} \end{aligned} \tag{8}$$

where $n_2 = n_1 = N e^{-\lambda_T T_m}$ and $\overline{d}_2 = T_m$.

3) Case 3: We define $p_{3,j}$ as the conditional probability that the event S' happens in the j-th zone 3 given that the event S' occurs during zone 3's. Then, T_w^3 and P_w^3 are respectively calculated as

$$\begin{aligned} T_w^3 &= E\left[\frac{d_{3,j}}{2}\right] = \sum_{j=1}^{n_3} p_{3,j} \frac{d_{3,j}}{2} = \sum_{j=1}^{n_3} \left(\frac{d_{3,j}}{\sum_{j=1}^{n_3} d_{3,j}} \cdot \frac{d_{3,j}}{2} \right) \\ &= \frac{1}{2} \frac{\sum_{j=1}^{n_3} d_{3,j}^2}{\sum_{j=1}^{n_3} d_{3,j}} = \frac{1}{2} \frac{n_3 \overline{d}_3^2}{n_3 \overline{d}_3} = \frac{1}{2} \frac{E[X^2 | X \leq T_m]}{E[X | X \leq T_m]} \\ &= \frac{1}{2} \frac{\int_0^{T_m} x^2 f_X(x) dx}{\int_0^{T_m} x f_X(x) dx} = \frac{1}{\lambda_T} + \frac{1}{2} \cdot \frac{T_m^2}{T_m + \frac{1}{\lambda_T}(1 - e^{\lambda_T T_m})} \end{aligned} \tag{9}$$

$$\begin{aligned} P_w^3 &= P_T - P_w^1 - P_w^2 \\ &= P_T \left(1 - e^{-\lambda_T T_m} - \lambda_T T_m e^{-\lambda_T T_m}\right). \end{aligned} \tag{10}$$

Finally, the average waiting time considering all cases is obtained by $\overline{T}_w = \sum_{i=1}^{3} P_w^i \cdot T_w^i$.

B. Loss Time

To analyze the packet loss time, we divide the talk-spurt and mutual silence periods into nine time zones, as shown in Fig. 3. There are seven possible cases with different loss times, as follows:

- Case 1: The event S' occurs in a mutual silence period and the talk-spurt does not occur within the service disruption time (i.e., S' happens in zone 1).
- Case 2: The event S' occurs in a mutual silence period whose length is longer than T_d and the talk-spurt occurs within the service disruption time (i.e., S' happens in zone 2).
- Case 3: The event S' occurs in a mutual silence period whose length is shorter than T_d (i.e., S' happens in zone 3).
- Case 4: The event S' occurs in a talk-spurt period and the mutual silence does not occur within the margin time plus the service disruption time (i.e., S' happens in zone 4).
- Case 5: The event S' occurs in a talk-spurt period whose length is longer than $T_m + T_d$ and the mutual silence occurs within the service disruption time after the margin time expires (i.e., S' happens in zone 5).
- Case 6: The event S' occurs in a talk-spurt period whose length is shorter than $T_m + T_d$ and the mutual silence occurs after the margin time expires (i.e., S' happens in zone 7).
- Case 7: The event S' occurs in a talk-spurt period and the mutual silence occurs within the margin time (i.e., S' happens in zones 6, 8 or 9).

Let T_l^i be the average loss time in the case i and P_l^i be the probability that the case i occurs during D. For simplicity of analysis, we consider only one state transition of VoIP traffic during the service disruption time.

1) Cases 1, 2 & 3: These cases have a reciprocity with the three cases of waiting time analysis. In the results of T_w^i and P_w^i where $i = \{1, 2, 3\}$, by replacing λ_T, P_T and T_m with λ_S, P_S and T_d, respectively, we can simply obtain T_l^i and P_l^i where $i = \{1, 2, 3\}$ as follows:

$$T_l^1 = 0, \quad P_l^1 = P_S e^{-\lambda_S T_d} \tag{11}$$

$$T_l^2 = \frac{T_d}{2}, \quad P_l^2 = P_S \lambda_S T_d e^{-\lambda_S T_d} \tag{12}$$

$$T_l^3 = T_d - \frac{1}{\lambda_S} - \frac{1}{2} \cdot \frac{T_d^2}{T_d + \frac{1}{\lambda_S}\left(1 - e^{\lambda_S T_d}\right)}, \tag{13}$$

$$P_l^3 = P_S\left(1 - e^{-\lambda_S T_d} - \lambda_S T_d e^{-\lambda_S T_d}\right). \tag{14}$$

2) Cases 4 & 5: The cases 4 and 5 have analogy to the cases 1 and 2 of the waiting time analysis, respectively. By using the same derivation process, we can obtain following:

$$T_l^4 = T_d, \quad P_l^4 = P_T e^{-\lambda_T (T_m + T_d)} \tag{15}$$

$$T_l^5 = \frac{T_d}{2}, \quad P_l^5 = P_T \lambda_T T_d e^{-\lambda_T (T_m + T_d)}. \tag{16}$$

3) Case 6: This case is similar to the case 3 of the waiting time analysis. By using the same derivation, we can obtain T_l^6 and P_l^6 as follows:

$$\begin{aligned} T_l^6 &= E\left[\frac{d_{7,j} - T_m}{2}\right] = \sum_{j=1}^{n_7}\left(\frac{d_{7,j}}{\sum_{j=1}^{n_7} d_{7,j}} \cdot \frac{d_{7,j} - T_m}{2}\right) \\ &= \frac{1}{2}\frac{\sum_{j=1}^{n_7} d_{7,j}^2}{\sum_{j=1}^{n_7} d_{7,j}} - \frac{T_m}{2} = \frac{1}{2}\frac{\int_{T_m}^{T_m+T_d} x^2 f_X(x)dx}{\int_{T_m}^{T_m+T_d} x f_X(x)dx} - \frac{T_m}{2} \\ &= \frac{1}{2} \cdot \frac{\frac{T_m}{\lambda_T} + \frac{2}{\lambda_T^2} - e^{-\lambda_T T_d}\left(T_d^2 + T_m T_d + \frac{T_m}{\lambda_T} + \frac{2}{\lambda_T} T_d + \frac{2}{\lambda_T^2}\right)}{T_m + \frac{1}{\lambda_T} - e^{-\lambda_T T_d}\left(T_m + T_d + \frac{1}{\lambda_T}\right)} \end{aligned} \tag{17}$$

$$\begin{aligned} P_l^6 &= \frac{\sum_{j=1}^{n_7} d_{7,j}}{D} = \frac{n_7 \overline{d}_7}{D} \\ &= \frac{\lambda_T \lambda_S}{\lambda_T + \lambda_S}\left\{\frac{1}{\lambda_T} e^{-\lambda_T T_m} - \left(T_d + \frac{1}{\lambda_T}\right) e^{-\lambda_T (T_m + T_d)}\right\} \\ &= P_T\left\{e^{-\lambda_T T_m} - (1 + \lambda_T T_d)\, e^{-\lambda_T (T_m + T_d)}\right\} \end{aligned} \tag{18}$$

where the total number of occurrences of zone 7 and the average of $\{d_{7,j}\}_{\forall j}$ are respectively given by

$$n_7 = \left(-e^{-\lambda_T (T_m + T_d)} + e^{-\lambda_T T_m}\right) N \tag{19}$$

$$\begin{aligned} \overline{d}_7 &= E[X | T_m \le X \le T_m + T_d] = \frac{\int_{T_m}^{T_m+T_d} x f_X(x)dx}{\int_{T_m}^{T_m+T_d} f_X(x)dx} \\ &= \frac{-e^{-\lambda_T (T_m + T_d)}\left(T_m + T_d + \frac{1}{\lambda_T}\right) + e^{-\lambda_T T_m}\left(T_m + \frac{1}{\lambda_T}\right)}{-e^{-\lambda_T (T_m + T_d)} + e^{-\lambda_T T_m}}. \end{aligned} \tag{20}$$

4) Case 7: In this case, the start of the disruption time is always aligned with the start of a mutual silence period. Therefore, the loss time becomes the difference between the disruption time and the mutual silence period, i.e., $T_l^7 = T_d - Y$. The average loss time and the probability of this case are obtained by

$$\begin{aligned} T_l^7 &= \int_0^{T_d} (T_d - y) f_Y(y)dy = \int_0^{T_d} (T_d - y)\lambda_S e^{-\lambda_S y} dy \\ &= T_d - \frac{1}{\lambda_S}(1 - e^{-\lambda_S T_d}) \end{aligned} \tag{21}$$

$$P_l^7 = P_T - (P_l^4 + P_l^5 + P_l^6) = P_T\left(1 - e^{-\lambda_T T_m}\right). \tag{22}$$

Finally, the average packet loss time considering all cases is expressed as $\overline{T}_l = \sum_{i=1}^{7} P_l^i \cdot T_l^i$. In the proposed scheme, the handover execution time is affected by the additional waiting time, so the total handover execution time is defined as $T_{HO} = T_{pre} + \overline{T}_w + T_d + T_{re}$.

V. Results and Discussions

For the result, the constants T_{pre} and T_{re} are set to 1 s and 0.1 s, respectively [11]. The service disruption time, T_d, varies within the range of a few hundred milliseconds, which is reasonable for vertical handover [20]. To validate the numerical analysis, Monte Carlo simulations for 10,000 VoIP calls are performed by using C++. The VoIP call duration is set to 200 s and one vertical handover event happens randomly throughout the time duration of each VoIP call.

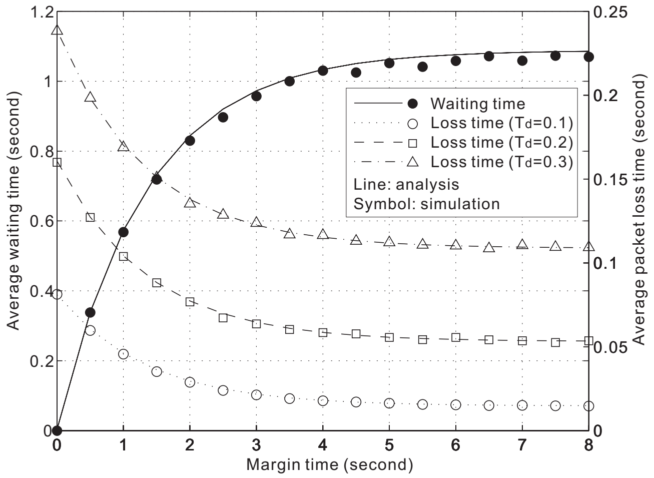

Fig. 4. Average waiting time and packet loss time vs. margin time.

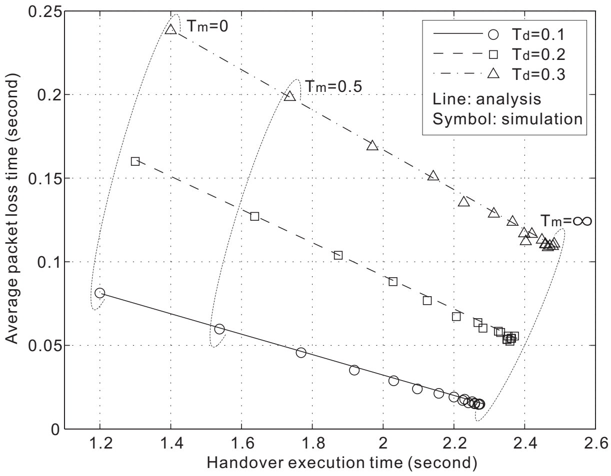

Fig. 5. Average packet loss time vs. handover execution time.

For the generation of VoIP traffic, the G.729 codec with the packet generation period of 20 ms is adopted and the on-off characteristic of traffic conforms to the Brady model [13].

Fig. 4 shows the average waiting time and packet loss time versus the margin time. As the margin time increases, the average waiting time increases before eventually becoming saturated. From (7)-(10), the waiting time converges to $\frac{P_T}{\lambda_T}$ (=1.088 s) as the margin time goes to infinity. Note that when the margin time goes to infinity, the waiting time is entirely dependent to the traffic characteristics (i.e., the distribution of talk-spurt period) and independent to the value of the margin time. On the other hand, the average loss time decreases as the margin time increases. This is due to the possibility that the mutual silence happens within the margin time increases as the margin time increases. Similar to the waiting time, the loss time is saturated as the margin time is increased because it is less affected by the margin time and eventually depends on only the on-off distribution of VoIP traffic. In addition, the loss time necessarily increases as the service disruption time increases.

Fig. 5 shows the trade-off relationship between the average packet loss time and the handover execution time. As the margin time increases, the handover execution time increases, but the packet loss time decreases, and vice versa. It is seen that both the packet loss time and the handover execution time converge as the margin time goes to infinity. Compared to the legacy handover case where the margin time is not applied ($T_m = 0$), for just $T_m = 0.5$, the packet loss time significantly decreases by 20∼30% while the handover execution time increases by the waiting time of 0.34 s.

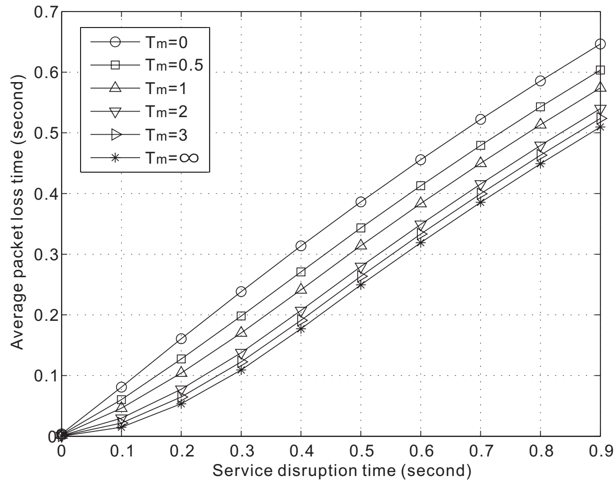

Fig. 6. Average packet loss time vs. service disruption time.

Fig. 6 shows the packet loss time versus the service disruption time. A longer disruption time naturally results in an increased packet loss time. Compared to the legacy handover ($T_m = 0$), the proposed handover using the margin time shows better performance regardless of the disruption time. It is shown that the performance improvement is maximized after the disruption time is 0.3 s, which is similar to the average length of the mutual silence period ($T_S = 0.306$ s).

VI. Conclusions and Further Work

We proposed an opportunistic vertical handover scheme based on the on-off characteristics of VoIP traffic. The proposed vertical handover scheme introduces a margin time before the L2 handover procedure and therefore aligns the service disruption time with the mutual silence period as much as possible. The results showed that there is a trade-off between the packet loss time and the handover execution time as the margin time varies. Considering that the allowable handover execution time of vertical handover is longer than that of horizontal handover, the efficiency of the proposed scheme can be maximized in the context of the vertical handover environment.

Obviously, this opportunistic handover scheme may be used for any traffic having an on-off property if a certain off-time duration is guaranteed. Moreover, the basic concept of the proposed algorithm can be extended to traffic whose transmission rate is not fixed but varies by aligning the service disruption with the time when the lowest transmission rate starts. This operation will mitigate the detrimental effect of service disruption on the QoS of the application. In future work, we will investigate adaptation of the proposed scheme to other traffic.

REFERENCES

[1] 3GPP, *Requirement on 3GPP System to Wireless Local Area Network (WLAN) Interworking*, TS 22.234 (v8.0.0), Mar. 2007.

[2] 3GPP2, *3GPP2-WLAN Interworking*, S.R0087-A(v1.0), Feb. 2006.

[3] N. Nasser, A. Hasswa, and H. Hassanein, ”Handoffs in fourth generation heterogeneous networks,” *IEEE Communications Magazine*, vol. 44, no. 10, pp. 96-103, Oct. 2006.

[4] J. McNair and Fang Zhu, ”Vertical handoffs in fourth-generation multi-network environments,”, *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 8-15, June 2004.

[5] K. El Malki *et al.*, *Low Latency Handoff in Mobile IPv4*, IETF, draft-ietf-mobileip-lowlatency-handoffs-v4-01, May 2001.

[6] G. Koodli, *Fast Handovers for Mobile IPv6*, IETF RFC 4068, Jul. 2005.

[7] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli, *Context Transfer Protocol (CXTP)*, IETF RFC 4067, Jul. 2005.

[8] H. H. Choi, O. Song, and D. H. Cho, ”Seamless Handoff Scheme Based on Pre-registration and Pre-authentication for UMTS-WLAN Interworking,” *Wireless Personal Communications*, vol. 41 , no. 3, pp. 345-364, May 2007.

[9] C. E. Perkins, ”Mobile IP Perkins,” *IEEE Communications Magazine*, vol. 35, no. 5, pp. 84-99, May 1997.

[10] C. E. Perkins *et al.*, ”Optimized Smooth Handoffs in Mobile IP,” in Proc. *IEEE Int. Symp. Comput. Commun.*, pp. 340-346, Jul. 1999.

[11] Shou-Chih Lo, Guanling Lee, Wen-Tsuen Chen, and Jen-Chi Liu ”Architecture for mobility and QoS support in all-IP wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 4, pp. 691-705, May 2004.

[12] ITU-T Recommendation G.723.1 Annex A, ”Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, ANNEX A: Silence Comppression Scheme,” Approved in Nov. 1996.

[13] ITU-T Recommendation G.729 Annex B, ”A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70,” Approved in Oct. 1996.

[14] A. Estepa, R. Estepa, and J. Vozmediano, ”A New Approach for VoIP Traffic Characterization,” *IEEE Commun. Lett.*, vol. 8, no. 10, pp. 644-646, Oct. 2004.

[15] E. Stevens-Navarro, Yuxia Lin, V. W. S. Wong, ”An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks,” *IEEE Trans. on Veh. Tech.*, vol. 57, no. 2, pp. 1243-1254, Mar. 2008.

[16] Jae-Young Pyun, ”Context-Aware Streaming Video System for Vertical Handover over Wireless Overlay Network,” *IEEE Trans. on Consumer Electronics*, vol. 54, no. 1, pp. 71-79, Feb. 2008.

[17] T. Melia, A. de la Oliva, I. Soto, C. J. Bernardos, and A. Vidal, ”Analysis of the Effect of Mobile Terminal Speed on WLAN/3G Vertical Handovers,” *IEEE GLOBECOM*, pp. 1-6, Nov. 2006.

[18] P. T. Brady, ”A model for generating on-off speech patterns in two-way conversation,” Bell Syst. Tech. J., pp. 2445-2472, Sept. 1969.

[19] H. P. Stern, S. A. Mahmoud, and K.-K. Wong, ”A Model for Generating On-Off Patterns in Conversational Speech, Including Short Silence Gaps and the Effects of Interaction Between Parties,” *IEEE Trans. on Veh. Tech.*, vol. 43, no. 4, pp. 1094-1100, Nov. 1994.

[20] H. Fathi, S. S. Chakraborty, and R. Prasad, ”Optimization of Mobile IPv6-Based Handovers to Support VoIP Services in Wireless Heterogeneous Networks,” *IEEE Trans. on Veh. Tech.*, vol. 56, no. 1, pp. 260-270, Jan. 2007.