A Low-Latency Scanning with Association Mechanism for Real-Time Communication in Mobile WiMAX

You-Lin Chen and Shiao-Li Tsao

***Abstract*—In Mobile WiMAX, a mobile station (MS) scans neighboring base stations (BSs) before conducting handover. The MS may perform the association further, i.e., initial ranging, with neighboring BSs during scanning to obtain ranging parameters and/or service availability information for a potential handover to a BS and/or the selection of target BSs. However, the scanning with association scheme may introduce additional latency for handover. This study first evaluates the conventional scanning with association process, and then proposes a mechanism to reduce the association latency. The performance models of the proposed mechanism are presented and the accuracies of analytical models are verified by simulations. Results demonstrate that the proposed mechanism reduces the association latency by 61.9%–78.0% for audio services and 41.3%–65.4% for video services.**

***Index Terms*—Mobile WiMAX, IEEE 802.16, handover, scanning with association, contention-based ranging.**

I. INTRODUCTION

THE IEEE 802.16 standard, known as WiMAX, is currently one of the most important broadband wireless access technologies. The IEEE 802.16e standard, known as Mobile WiMAX, extends WiMAX to support mobility. Mobile WiMAX allows an MS to handover its communications from the serving BS, to which the MS is currently connected, to one of its neighboring BSs, called the target BS. However, the temporary suspension of communication between the MS and the serving BS during handover may lead to service disruption. This service disruption may decrease the QoS of communications, and especially delay-sensitive communications such as voice over IP and/or video conferencing services.

According to the IEEE 802.16 specification [1], a handover process consists of four phases: network topology acquisition, scan, actual handover, and network re-entry. A number of previous studies have investigated and optimized handover procedures including handover preparation and handover execution to minimize service disruption and handover latency in Mobile WiMAX. During the handover preparation phase, an MS scans neighboring BSs before conducting handover. The association, i.e., initial ranging, with the neighboring BSs during scanning intervals is an optional procedure, the objective of which is to obtain ranging parameters and/or service availability information for a potential handover to a BS and/or the selection of target BSs. The scanning with association scheme defined in IEEE 802.16 [1] allows an MS to perform scanning and ranging with neighboring BSs while maintaining communication with the serving BS, to avoid service disruptions. A few studies have investigated the scanning with association scheme during handover. The previous studies [2] and [3] showed that applying scanning with association can minimize service disruption and reduce packet delay, but the duration of the handover process increases. This study first analyzes the scanning with association scheme. We find that sending a ranging request in later frames in a scanning interval and the pause of the backoff countdown process during interleaving intervals contribute to significant latency of the association process. Therefore, this study proposes a novel scanning with association mechanism to reduce the association latency while minimizing delay and jitter of real-time communication on MSs. The main contributions of this study are the development of the performance model and the evaluation of the conventional scanning with association scheme in Mobile WiMAX, and the proposal of a novel mechanism to reduce the association latency for real-time communication.

The rest of this paper is organized as follows: Section II presents a summary of related works; Section III details the background and problems of the scanning with association process; Section IV introduces the analysis on the scanning with association scheme and the proposed mechanism; Section V shows the analytical model of the proposed scheme; Section VI provides the simulation results; and finally, Section VII offers conclusions.

II. RELATED WORKS

Previous research [4] showed a fast handover algorithm for IEEE 802.16. By using their proposed scheme, an MS can obtain channel synchronization parameters of the target BS from the serving BS, and can receive downlink data from the serving and target BSs simultaneously. Therefore, the MS can perform scanning with association with the target BS while transmitting data with the serving BS. However, this approach assumes that an MS can maintain communications with two BSs simultaneously, which require extra hardware support on MSs. Choi *et al.* [5] proposed a new MAC message called Fast DL_MAP_IE. With the support of this new message, the target

Manuscript received September 24, 2011; revised March 3 and May 26, 2012; accepted May 30, 2012. The associate editor coordinating the review of this paper and approving it for publication was L. Libman.

The authors are with the Department of Computer Science, National Chiao Tung University, Hsin-Chu, 30010, Taiwan (e-mail: {youlin, sltsao}@cs.nctu.edu.tw).

Digital Object Identifier 10.1109/TWC.2012.081312.111765

BS can assign radio resources to an MS for downlink data transmission immediately after the MS switches to the target BS. Therefore, the MS can receive the downlink data before completing the handover procedure. Jiao *et al.* [6] further proposed a transport connection identifiers (CIDs) mapping scheme to avoid the latency of CID mapping on the target BS after handover. A cross-layer fast handover mechanism that suggests exchanging MAC messages over the network-layer protocol was proposed in [7] to reduce handover latency in Mobile WiMAX. Fattah and Alnuweiri [8] presented a service-flow aware algorithm to determine the appropriate target BS, to avoid connecting to a target BS with insufficient radio resources. Dong *et al.* [9] applied different scanning with association strategies to various types of service flows. Therefore, the traffic overhead of a handover process can be reduced while handover latency can be guaranteed. Although previous research has investigated handover latency, only a few studies had improved the initial ranging process, which may also result in service disruption.

An MS must perform initial ranging, i.e., the association, with the target BS before the MS can communicate with the target BS. Different mechanisms such as non-contention-based ranging and contention-based ranging can be used in the scanning with association process, which will be detailed in the next section. This paper focuses on contention-based ranging, which is mandated by the WiMAX Forum. The contention-based bandwidth request, which is similar to contention-based ranging, is widely used in Mobile WiMAX. A number of studies have investigated and modeled contention-based bandwidth request mechanisms [13]–[15]. Moreover, Lin *et al.* [10] proposed dynamic contention window adjustment to improve the performance of contention-based ranging. Chen and Tseng [11] suggested piggybacking the bandwidth request information in the available uplink burst; therefore, the chance of contention would be reduced. In [12], Delicado *et al.* analyzed a previous backoff window in each contention and dynamically assigned an initial backoff window to improve system performance. However, few studies have investigated the initial ranging process during scanning with association. For scanning with association based on contention-based ranging, an MS initiates a periodic scanning, which consists of a number of scanning intervals and interleaving intervals. Rouil and Golmie [2] presented an adaptive channel scanning (ACS) algorithm to allocate scanning intervals based on the jitter and delay requirements of communication services. The ACS algorithm can be used to minimize service disruptions during periodic scanning. However, applying periodic ranging prolongs overall handover latency. Tsao *et al.* [3] showed that ranging responses may be lost because of periodic scanning, and the loss of ranging response prolongs the overall handover process. They modeled the scanning with association latency and suggested suitable parameters, including lengths of scanning intervals and interleaving intervals, to minimize the latency. However, the conventional scanning with association scheme may still cause a large amount of missing ranging responses and result in a long association latency. This study investigates the scanning with association scheme and proposes a novel association mechanism to reduce the association latency.

III. THE HANDOVER PROCESS

Mobility management schemes in Mobile WiMAX handle link and network layer handover, and they have been jointly developed by the IEEE 802.16 Working Group and the WiMAX Forum. Readers may refer to [16] for more detailed information and descriptions of Mobile WiMAX and all types of handover in Mobile WiMAX. In Mobile WiMAX, hard handover is mandatory, whereas soft handover mechanisms such as Fast BS Switching (FBSS) and Macro Diversity Handover (MDHO) are optional. This study considers hard handover. Fig. 1 shows the link-layer handover procedures in Mobile WiMAX. The first phase of handover is the acquisition of network topology. A BS periodically broadcasts a control message, MOB_NBR-ADV, which includes information of neighboring BSs, such as channel configuration and handover parameters. When the signal strength of the serving BS is lower than a predefined threshold, the MS triggers the second-phase procedures, i.e., the scan phase, to seek potential BSs for handover. During the scan phase, the MS can perform scanning only or scanning with association. For scanning only, the MS first negotiates the parameters of a scan, e.g., scanning intervals, interleaving intervals, and a number of scan iterations, with the serving BS, and then scans neighboring BSs in a round-robin basis. The MS switches to the neighboring BS for scanning intervals and the serving BS for interleaving intervals. In a scanning interval, the MS measures the neighboring BSs' channel qualities. During scanning intervals, packets sent to the MS are buffered on the serving BS. After a scanning interval, the MS returns to the serving BS for an interleaving interval, resumes communication with the serving BS, and retrieves the buffered packets. The scan lasts for a number of scanning and interleaving iterations, according to the negotiated agreement between the MS and the serving BS. With extra hardware support, an MS can perform autonomous neighbor cell scanning without switching to the neighboring BSs, and can maintain the signal quality database for neighbor cells through preamble detection in the same carrier frequency [1]. For scanning with association, the MS requests a scan and performs the association, i.e., initial ranging, procedures during scanning to reduce the service disruption. After the scan is complete, the MS reports the measurement results to the serving BS, so that the serving BS can suggest a list of potential BSs to the MS for handover. Once the signal strengths of the serving BS and/or neighboring BSs meet the handover criteria, the third phase, i.e., the actual handover phase, is triggered. The MS sends the handover messages, disconnects from the serving BS, and reconnects to the target BS. Finally, in the fourth phase, the MS performs network re-entry procedures and completes handover.

An MS may perform the association, i.e., initial ranging, with the neighboring BSs during scanning intervals. The IEEE 802.16 standard specifies three association levels: Association Levels 0, 1, and 2, i.e., association without coordination, association with coordination, and network-assisted association reporting, respectively. In Association Level 0, an MS performs contention-based ranging with its neighboring BSs. In Association Level 1, the serving BS can negotiate with a neighboring BS to reserve the dedicated resources on the

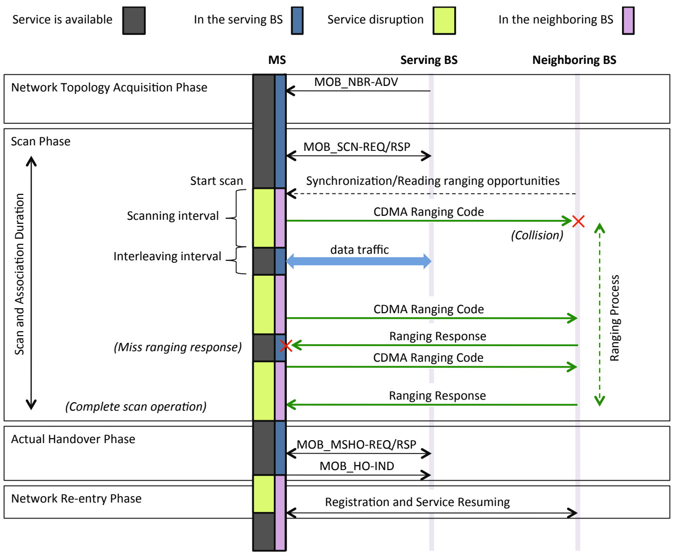

Fig. 1. Mobile WiMAX handover process.

neighboring BS for a ranging procedure. This implies that the MS can perform non-contention-based ranging with the neighboring BS in Mobile WiMAX. Association Level 2 is similar to Association Level 1; the difference is that the ranging response messages are not delivered to the MS directly from the neighboring BS over the air, and they are forwarded to the serving BS over the backbone network first and then sent to the MS. Thus, the MS is not required to stay in the neighboring channel to wait for a ranging response. Although Association Levels 1 and 2 achieve superior performance for latency to Association Level 0, they require all BSs and network supports. Moreover, the frame-level time synchronization of all BSs is also required for Association Levels 1 and 2. This study considers Association Level 0, which relies on contention-based ranging and is mandated by the WiMAX Forum certification, and focuses on reducing the association latency.

IV. ENHANCED SCAN AND ASSOCIATION MECHANISM

A. Analysis of the conventional approach

Previous research [3] presented a model of the association latency while an MS adopts the scanning with association scheme. This model shows that higher collision probability of a ranging request, defined as P_c, and the missing probability of a ranging response, defined as P_m, prolong the association latency. Although P_m was evaluated in [3], P_m can represent only the average missing probability of a ranging response for a neighboring BS. However, the average missing probability of a ranging response is insufficient for evaluating the exact missing probability of a ranging response when an MS sends a ranging request at a specific frame. Therefore, this study defines $P_m(x)$ for a precise evaluation on the missing probability of a ranging response for one neighboring BS when a ranging request is sent at a specific frame x. A ranging response can be received if it arrives on a scanning interval, but is missed if it falls on an interleaving interval. Therefore, the missing probability of a ranging response is the percentage of ranging responses falling on interleaving intervals, whereas

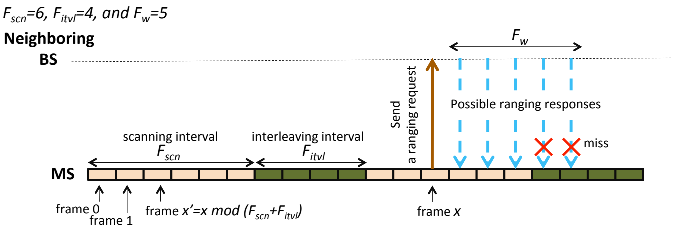

Fig. 2. An example of missing ranging responses.

we assume that ranging responses from a BS are uniformly distributed over the next F_w frames after the BS receives a ranging request. For example, in Fig. 2, the neighboring BS replies with a ranging response in next five frames after receiving a ranging request. Assume that two of the five frames are in an interleaving interval, and thus the missing probability of a ranging response is $\frac{2}{5} = 40\%$. In other words, the missing probability of a ranging response varies when an MS sends a ranging request at different frames.

The missing probability of a ranging response depends on the frame in a scanning interval where an MS sends a ranging request, and it is irrelevant to which scanning interval that the ranging request is sent. Therefore, we use x' which is an auxiliary notation to indicate the relative position of frame x in a scanning cycle. A scanning cycle is defined as a scanning interval and an interleaving interval, i.e., $F_{scn}+F_{itvl}$. In other words, $x' = x \bmod (F_{scn} + F_{itvl})$ and $P_m(x) := P_m(x')$. To compute $P_m(x)$, x' is first derived in (1) and (2), and then calculate $P_m(x')$, i.e., $P_m(x)$.

$$t = \left\lfloor \frac{x}{F_{scn} + F_{itvl}} \right\rfloor \tag{1}$$

$$x' = x - (F_{scn} + F_{itvl}) \times t \tag{2}$$

$$P_m(x) := P_m(x') \tag{3}$$

After an MS sends a ranging request without a collision and the ranging request is received by the neighboring BS, the ranging response is replied by the neighboring BS within F_w frames. Assume that the neighboring BS replies the ranging response after R_{fw} frames, where R_{fw} is a random variable of delay. The ranging response is missed if it is sent to the MS when the MS is in an interleaving interval, i.e., $F_{scn} \leq X_{rsp} < F_{scn} + F_{itvl}$ and X_{rsp} is defined in (4). It indicates $P_m(x) = P\{F_{scn} \leq X_{rsp} < F_{scn} + F_{itvl}\}$.

$$X_{rsp} = x + R_{fw} - (F_{scn} + F_{itvl}) \times \left\lfloor \frac{x + R_{fw}}{F_{scn} + F_{itvl}} \right\rfloor \tag{4}$$

Since R_{fw} is in a uniform distribution, from 1 to F_w, the missing probability of a ranging response can be simplified as the number of frames in interleaving intervals divided by F_w. Moreover, since F_w may be longer than $F_{scn} + F_{itvl}$ frames, we thus divided F_w into q segments, each containing $F_{scn} + F_{itvl}$ frames, and the remaining $F_w{}'$ frames, which is smaller than a segment, as (6) shows. Each segment contains $F_{scn} + F_{itvl}$ frames, and $F_{itvl} \times q$ frames of these segments are in interleaving intervals, i.e., (8b). These $F_w{}'$ frames may also overlap with some frames in an interleaving interval, and the number of these frames is F_x, defined in (7). It implies that no frame in an interleaving interval overlaps with the ${F_w}'$ frames if $F_x < 0$. Conversely, the ${F_w}'$ frames cover the whole interleaving interval if $F_x \geq F_{itvl}$, i.e., (8d). Therefore, the total number of frames in interleaving intervals is $F_x + F_{itvl} \times q$ as (8c) shows.

$$q = \left\lfloor \frac{F_w}{F_{scn} + F_{itvl}} \right\rfloor \tag{5}$$

$${F_w}' = F_w - (F_{scn} + F_{itvl}) \times q \tag{6}$$

$$F_x = x' + {F_w}' - F_{scn} + 1 \tag{7}$$

Equation (8) shows that the missing probability of a ranging response, increases monotonically by x' in a scanning interval. Thus, an MS should send a ranging request as early as possible when it switches to the neighboring BS for a scanning interval to avoid missing a ranging response. Although a long scanning interval and short interleaving interval can reduce the missing probability of a ranging response, these two parameters also affect the handover latency. A long scanning interval introduces more packet delay and requires additional radio resources in the next interleaving interval to transfer the packets buffered on the serving BS. A short interleaving interval implies that the serving BS must allocate sufficient enough radio resources to transmit buffered packets to the MS within a short interleaving interval. Therefore, F_{scn} and F_{itvl} should be carefully decided for each MS to fit its delay and jitter requirements under the serving BS's loading.

B. Scanning with Self-backoff (SSB)

Two occasions exist for performing initial ranging during handover. In the first case, an MS does not perform the association during scanning. The MS performs initial ranging with the target BS after disconnecting from the serving BS. We call it "scanning without association." In this scheme, communication between the MS and the BS during the initial ranging period must be suspended, and the service disruption may violate the maximal delay constraint of real-time communication. The other occasion is to perform initial ranging during scanning, i.e., the scanning with association scheme. In this scheme, the MS performs initial ranging with the neighboring BSs during scanning intervals to reduce service disruptions. However, the scanning with association scheme may suffer from the miss of ranging responses and result in a long association latency. Hence, this study proposes a novel scanning with association mechanism, called scanning with self-backoff (SSB), inspired by the above observations and analyses.

In WiMAX, an MS performs backoff based on a variable contention window for contention resolution. A contention-based ranging is based on the binary exponential backoff (BEB) algorithm. The initial ranging procedure is complete when the MS successfully receives a ranging response from the neighboring BS. If the MS cannot receive the ranging response within *T3* ms, the ranging fails and the MS doubles its contention window size to perform next ranging attempt. The proposed SSB scheme maintains a backoff window W, much like the conventional WiMAX backoff, and a self-backoff window $W_{ssb} = \left\lceil \frac{W}{S_{ro}} \right\rceil$. S_{ro} is defined as the average number of ranging opportunities per frame for a neighboring BS. There are several ways to decide S_{ro}. For example, a neighboring BS may determine the future resource allocation of ranging opportunities, and periodically broadcasts its S_{ro}. An alternative approach is that a neighboring BS may calculate the average number of ranging opportunities per frame during the past period of time, and provides MSs the average value for estimating future ranging opportunities. Fig. 3 gives an example and illustrates the differences between the conventional scanning with association scheme and the proposed SSB scheme. We assume that an MS has to wait for five ranging opportunities before it can perform the ranging. As can be seen from the upper part of Fig. 3, the conventional scheme performs backoff only in scanning intervals. The MS finds three ranging opportunities in the first scanning interval, stops tracking ranging opportunities during the interleaving interval, and detects another two ranging opportunities in the second scanning interval. Finally, the MS sends a ranging request when it detects the sixth ranging opportunity. On the other hand, an MS stays connected with the serving BS and performs self-backoff simultaneously when the SSB scheme is employed. The MS waits for ten frames, which are expected to have five ranging opportunities, switches to a scanning interval, and sends a ranging request as soon as it detects the first ranging opportunity. Therefore, the MS can reduce the time in tracking ranging opportunities during scanning intervals, and can send a ranging request in early frames of a scanning interval to avoid missing of a ranging response.

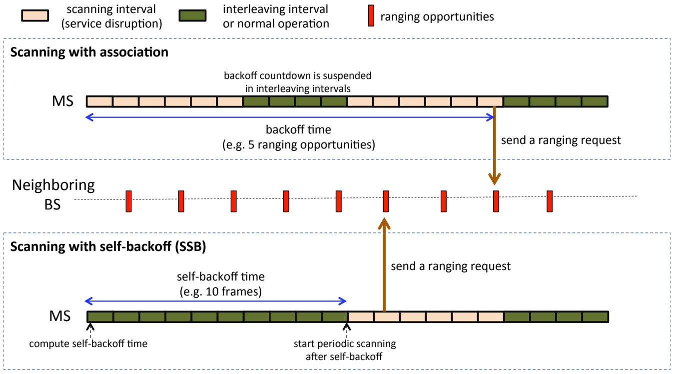

Fig. 3. Comparison between the scanning with association and the SSB scheme.

The MS randomly picks up a self-backoff counter, says R_w, from 1 to W_{ssb}. The term R_w represents the number of frames that the MS must wait for self-backoff countdown before sending a ranging request. When the MS waits for the backoff time, the MS remains connected with the serving BS. After R_w frames, the MS starts a periodic scanning with the neighboring BSs. The parameters of a periodic scanning such as the lengths of a scanning interval and an interleaving interval are chosen by the serving BS to prevent violation of the delay constraints of real-time communication on the MS. To avoid service disruption, a scanning interval is usually short, and an interleaving interval should suffice in duration to transmit buffered packets during the previous scanning interval.

In the periodic scanning, the MS sends a ranging request when finding a frame that offers the resources for ranging. If

$$P_m(x') = \begin{cases} 0 & , F_{scn} \le x' < F_{scn} + F_{itvl} & \text{(8a)} \\ \dfrac{F_{itvl} \times q}{F_w} & , 0 \le x' < F_{scn} \text{ and } F_x < 0 & \text{(8b)} \\ \dfrac{F_x + F_{itvl} \times q}{F_w} & , 0 \le x' < F_{scn} \text{ and } 0 \le F_x \le F_{itvl} & \text{(8c)} \\ \dfrac{F_{itvl} \times (q+1)}{F_w} & , 0 \le x' < F_{scn} \text{ and } F_{itvl} < F_x & \text{(8d)} \end{cases}$$

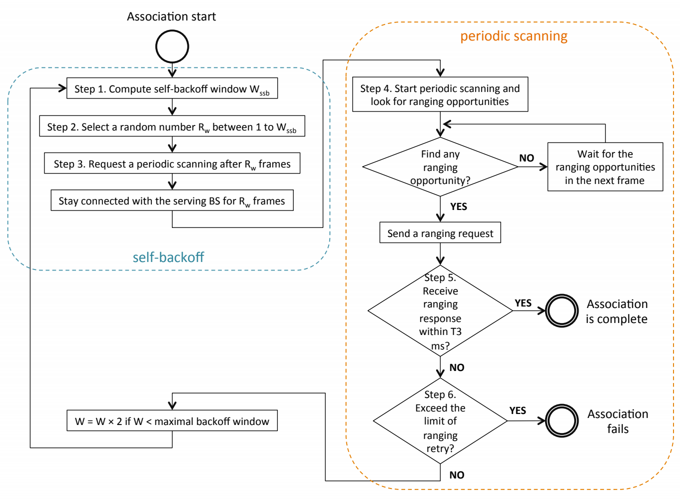

Fig. 4. Flowchart of the proposed scanning with self-backoff (SSB) scheme.

no ranging response is received within *T3* ms after sending a ranging request, the MS regards the request as a ranging failure and doubles the backoff window, W, for the next ranging attempt. The SSB procedures are illustrated in Fig. 4 and presented as follows:

- Step 1: The MS computes W_{ssb}.
- Step 2: The MS randomly selects the value R_w between 1 and W_{ssb}.
- Step 3: The MS remains connected with the serving BS and requests a periodic scanning after R_w frames.
- Step 4: The MS starts periodic scanning and searches for the first frame that contains ranging opportunities. The MS randomly selects a ranging opportunity to send a ranging request.
- Step 5: The association is complete if the MS receives the ranging response within *T3* ms. Otherwise, the ranging attempt fails, and the MS proceeds to Step 6.
- Step 6: The association fails if the number of ranging retries exceeds the retry limit. Otherwise, the MS doubles the backoff window W if W is smaller than the maximal backoff window size and returns to Step 1.

V. ANALYTICAL MODEL OF THE PROPOSED MECHANISM

A. Probability of finding the first frame containing ranging opportunities

The SSB scheme requires an MS to estimate the number of frames for backoff and to stay in the serving BS during backoff time. W_{ssb} denotes the self-backoff time in terms of the number of frames. After the self-backoff time reaches zero, the MS starts a periodic scanning and switches between the serving and neighboring BSs. The MS sends a ranging request immediately after detecting the first ranging opportunity in a scanning interval. Therefore, it is necessary to predict where the MS can find the first frame containing ranging opportunities after it switches to the neighboring BS, and the success probability if the MS sends the ranging request. These two probabilities make it possible to derive the success probability of a ranging attempt, which in turn can help derive the association latency of an MS.

To derive the probability that an MS can find the ranging opportunity to send a ranging request on each frame, we first define a frame that offers radio resources for a contention-based ranging as an RO-frame and a frame containing no ranging opportunity as an NRO-frame. If “0” denotes an NRO-frame and “1” denotes an RO-frame, then the frames on the neighboring BS can be represented as a sequence of “0” and “1”. For an MS employing the SSB scheme, seeking the first RO-frame for ranging can be regarded as the process of finding the first “1” in a binary sequence. Since the MS does not know the previous frames before it switches to the neighboring BS, we use “ϵ” to represent the previous frames before the MS switches to the neighboring BS as the unknown “0”s or “1”s in the binary sequence. If the MS switches to a neighboring BS and immediately finds an RO-frame for sending a ranging request, the binary sequence representing the frames in the neighboring BS can be written as “$\epsilon 1$”. The probability of the MS finding the first RO-frame at frame x is then denoted as $\mathit{DRQ}(x)$, as Eq. (9) shows. “$0^x 1$” in (9) indicates a binary sequence with first x successive “0”s followed by one “1.” The definition of x is the same as that defined in Section III.

$$\mathit{DRQ}(x) = P\{\epsilon 0^x 1\}, \forall x \ge 0 \tag{9}$$

The scenarios “$\epsilon 0^x 1$” are mutually exclusive for different x because the MS sends a ranging request as soon as it finds the first RO-frame. After the MS sends the ranging request, it waits for a ranging response. A pattern “$\epsilon 0^x 1$” appears only if the patterns “$\epsilon 0^i 1$”, $\forall i < x$, do not appear. Therefore, (9) can be rewritten as (10). The probability $P\{\epsilon 0^x 1 | \epsilon 0^x\}$ defined in (10) represents the conditional probability that the next frame is “1” if x successive “0”s have been detected. Noted that $P\{\epsilon\} = 1$ in Eq. (10) because no frame has been tracked yet.

$$\begin{aligned} \mathit{DRQ}(x) &= P\{\epsilon 0^x 1 | \epsilon 0^x\} \times P\{\epsilon 0^x\}, \forall x \ge 0 \\ P\{\epsilon 0^x\} &= 1 - \sum_{i<x} \mathit{DRQ}(i), \forall x \ge 0 \end{aligned} \tag{10}$$

The variable RO_P indicates the number of successive NRO-frames between two RO-frames on the neighboring BS. For example, $RO_P = 0$ implies that the next frame offers radio resources for contention-based ranging, while $RO_P = 1$ means the next frame does not have a ranging opportunity but the next second frame has a ranging opportunity. $RO_P = i$ can also be denoted as pattern "0^i1". Equation (11) represents the probabilities of the patterns "0^i1" on the neighboring BS. For example, $P\{RO_P = 0\} = 1$ and $P\{RO_P \neq 0\} = 0$ if a neighboring BS allocates resources for ranging on each frame. $P\{\epsilon 0^x 1|\epsilon 0^x\}$ can thus be derived by (12). $P\{\epsilon 0^x 1\} \cap P\{\epsilon 0^x\} = P\{\epsilon 0^x 1\}$ denotes the probability that we can find x successive "0"s and then a "1" in the binary sequence. $P\{\epsilon 0^x 1\}$ can be evaluated by accumulating the probabilities of "0^i1" that $i \geq x$, i.e., at least x successive "0"s and then a "1". On the other hand, $P\{\epsilon 0^x\}$ is the probability that we can find x successive "0"s. x successive "0"s are obtained from "0^i1" where $i \geq x$; however, there are multiple possible matches with x successive "0"s among "0^i1" if $i > x$. For example, there are two matches of "$\epsilon 00$" in "0001" if $x = 2$ and $i = 3$. The first match is "$\underline{00}01$", and the second match is "$0\underline{00}1$". In other words, there are $i - x + 1$ possible matches when we find "$\epsilon 0^x$" in "0^i1". Equation (9) through (12) thus can derive the probability of finding the first frame containing ranging opportunity at frame x, i.e., *DRQ*(x).

$$P\{0^i1\} = P\{RO_P = i\}$$
$$\sum_{i \geq 0} P\{RO_P = i\} = 1 \tag{11}$$

$$P\{\epsilon 0^x 1|\epsilon 0^x\} = \frac{P\{\epsilon 0^x 1\} \cap P\{\epsilon 0^x\}}{P\{\epsilon 0^x\}} = \frac{\sum_{i \geq x} P\{0^i1\}}{\sum_{i \geq x} P\{0^i1\} \times (i - x + 1)} \tag{12}$$

B. Model DRQ(x) considering interleaving intervals

In the section above, *DRQ(x)* $= P\{\epsilon 0^x 1\}$ only considers cases without an interleaving interval. The binary sequence pattern for the case with interleaving intervals can be represented as "$\epsilon(0^{F_{scn}} X^{F_{itvl}})^t 0^{x'} 1$" where x' and t have been defined in (3) in Section III. "$(0^{F_{scn}} X^{F_{itvl}})^t$" indicates t duplicated segments and each segment comprises a binary sequence with F_{scn} successive "0"s and F_{itvl} successive "X"s. "X" denotes an unknown frame which could be either an RO-frame or an NRO-frame. Therefore, the probability of finding the first frame containing ranging opportunity while considering interleaving intervals, defined as *DRQ*$'(x)$, can be written as (13). The term *DRQ*$'(x) = 0$ in (13) means that frame x is in an interleaving interval and the MS cannot send ranging request during interleaving intervals.

The probability of the patterns, $P\{\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1\}$, is equivalent to the summation of probabilities of all possible patterns. There are $t \times F_{itvl}$ unknown frames, which are either "0" or "1", during interleaving intervals. Therefore, there are $2^{t \times F_{itvl}}$ cases for the patterns. Define $str_k, 1 \leq k \leq 2^{t \times F_{itvl}}$ for the $2^{t \times F_{itvl}}$ patterns, and then we can rewrite $P\{\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1\}$ as (14). Note that the patterns str_k in (14) include one or more "1"s. The probability for a pattern that includes multiple "1"s, such as "$\epsilon 0^{q_0} 1 0^{q_1} 1 \cdots 0^{q_z} 1$", can be calculated by (15).

$$\begin{aligned} & P\{\epsilon 0^{q_0} 1 0^{q_1} 1 \cdots 0^{q_z} 1\} \\ = & P\{\epsilon 0^{q_0} 1\} \times P\{0^{q_1} 1\} \times \cdots P\{0^{q_z} 1\} \\ = & P\{\epsilon 0^{q_0} 1\} \times \prod_{i=1}^{z} P\{0^{q_i} 1\} \end{aligned} \tag{15}$$

C. Collision probability of a ranging request

As mentioned above, an MS may suffer from different probabilities of a ranging failure if it sends the ranging request at different frames. Therefore, the probability of a ranging failure for a ranging attempt, says P_f, can be represented as (16). The terms $P_f(x)$, $P_c(x)$, and $P_m(x)$ denote the probability of a ranging failure, collision probability of a ranging request, and missing probability of a ranging response, respectively when an MS sends a ranging request at frame x. In (16), P_f is the summation of the probability of a ranging failure when an MS sends a ranging request at frame x and the probability that the MS finds the first RO-frame at frame x. In (17), ranging failure occurs when either the ranging request has a collision or the ranging request has no collision but the ranging response is lost.

$$P_f = \sum_{x \geq 0} DRQ'(x) \times P_f(x) \tag{16}$$

$$P_f(x) = P_c(x) + (1 - P_c(x)) \times P_m(x), \forall x \in \mathbb{N}^0 \tag{17}$$

$P_m(x)$ has been modeled as (8). To estimate $P_c(x)$, we first define the average rate of a ranging attempt, say α, which denotes the average number of ranging attempts per frame from one MS. When applying the SSB scheme, MSs send ranging requests at the first RO-frame they detect after backoff time. The rate of a ranging attempt at RO-frames therefore becomes higher than α because ranging requests cannot be transmitted during NRO-frames. Pending ranging attempts from MSs during the NRO-frames accumulate and may be sent at the same RO-frame. Therefore, it is possible to calculate the accumulated rate for ranging attempts, defined as $A(x)$, from one MS for a specific RO-frame, say frame x, when the MS sends a ranging request at frame x. $A(x)$ can also be represented as $A(\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1)$ and split into $2^{t \times F_{itvl}}$ patterns of $A(str_k)$, such as in (14). Subsequently, (18) calculates $A(x)$ by accumulating rates of ranging attempts for each $A(str_k)$. $A(str_k)$ is determined by counting the number of successive NRO-frames in addition to one in (19). While str_k contains exactly one "1", we can derive $A(str_k)$ based on the average number of successive NRO-frames. Otherwise, we consider the last segment "$0^{q_z} 1$" where the

$$\boldsymbol{DRQ}'(x) = \begin{cases} 0 & , F_{scn} \le x' < F_{scn} + F_{itvl} \\ P\{\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1\} & , 0 \le x' < F_{scn} \end{cases} \tag{13}$$

$$\begin{aligned} P\{\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1\} &= \sum_{k=1}^{2^{t \times F_{itvl}}} P\{str_k\} \\ &= \sum_{\substack{Q_{i,j} \in \{0,1\}, \\ 1 \le i \le t, \\ 1 \le j \le F_{itvl}}} P\{0^{F_{scn}} Q_{1,1} Q_{1,2} \cdots Q_{1,F_{itvl}} \cdots 0^{F_{scn}} Q_{t,1} \cdots Q_{t,F_{itvl}} 0^{x'} 1\} \end{aligned} \tag{14}$$

frame x is located, and calculate $A(str_k)$.

$$\begin{aligned} A(x) &= A(\epsilon \left(0^{F_{scn}} X^{F_{itvl}}\right)^t 0^{x'} 1) \\ &= \frac{\sum_{k=1}^{2^{t \times F_{itvl}}} P\{str_k\} \times A(str_k)}{\sum_{k=1}^{2^{t \times F_{itvl}}} P\{str_k\}} \\ &= \frac{\sum_{k=1}^{2^{t \times F_{itvl}}} P\{str_k\} \times A(str_k)}{\boldsymbol{DRQ}'(x)} \end{aligned} \tag{18}$$

If the number of MSs that are contending for ranging opportunities is N_{ms}, the number of CDMA ranging codes is N_{cdma}, and the number of ranging opportunities in an RO-frame is N_{ro}. Then, $P_c(x)$ can be modeled as (20) based on the assumption that a collision occurs if two or more MSs send the same ranging code in the same ranging opportunity.

$$\begin{aligned} & P_c(x) \\ =& 1 - \sum_{i=0}^{N_{ms}-1} \left\{ \binom{N_{ms}-1}{i} A(x)^i (1-A(x))^{N_{ms}-i-1} H(i) \right\} \\ & H(i) = \sum_{j=0}^{i} \binom{i}{j} \left(\frac{N_{cdma}-1}{N_{ro} \times N_{cdma}}\right)^j \left(1 - \frac{1}{N_{ro}}\right)^{i-j} \end{aligned} \tag{20}$$

Equations (8), (13), (16), and (20) make it possible to derive the probability of a ranging failure, P_f. Meanwhile, the association latency for the proposed SSB scheme, say T_{a_ssb}, can also be calculated as (21) where L denotes the maximal retry limit for a ranging process. T_{a_ssb} denotes the time required to complete the scanning with association process by employing the SSB scheme regardless of whether the association is successful. The failure of the scanning with association process means that the MS cannot receive a ranging response in L retries. Equation (21) shows that T_{a_ssb} comprises of three parts: $T_{(backoff,j)}$, $T_{(wait,j)}$, and $T_{(rspdelay,j)}$ where j indicates the j-th ranging attempt. $T_{(backoff,j)}$ gives the time that the MS spends in backoff. $T_{(backoff,j)}$ depends on the contention window size for each ranging attempt. $T_{(wait,j)}$ indicates the time that the MS must wait for an RO-frame after backoff, and $T_{(rspdelay,j)}$ indicates the time spent in waiting for a ranging response regardless of the association process success or failure. In $T_{(wait,j)}$, $\boldsymbol{DRQ}'(x)$ implies the MS requires x frames to detect the first RO-frame, introducing a delay of $x+1$ frames.

$$\begin{aligned} T_{a\_ssb} &= (P_f)^L \left(T_{(backoff,L-1)} + T_{(wait,L-1)} + L \times \mathit{T3}\right) \\ &+ \sum_{j=0}^{L-1} (P_f)^j (1-P_f)(T_{(backoff,j)} + T_{(wait,j)} + T_{(rspdelay,j)}) \end{aligned} \tag{21}$$

$$T_{(backoff,j)} = \sum_{i=0}^{j} d_i \text{ , } d_i \text{ is the backoff time for the } i\text{-th ranging attempt} \tag{22}$$

$$T_{(wait,j)} = (j+1) \times T_f \times \sum_{x \ge 0} (x+1) \times \boldsymbol{DRQ}'(x) \tag{23}$$

$$T_{(rspdelay,j)} = \mathit{T3} \times j + \frac{F_w \times T_f}{2} \tag{24}$$

Consider a saturation environment in which MSs continue performing scanning with association, and all MSs are adopting the proposed SSB scheme. In this case, the rate of ranging attempts, i.e., α in (25), can be obtained by the total number of ranging requests and overall time spent in the scanning with association process. The previous section uses the rate of ranging attempts α to evaluate variables, such as $P_c(x)$, P_f, and T_{a_ssb}. Therefore, we can find a feasible solution to compute these variables by numerical analysis.

$$\alpha = \left(1 + \sum_{i=1}^{L-1} P_f^i\right) \times \frac{T_f}{T_{a\_ssb}} \tag{25}$$

VI. SIMULATION RESULTS

To verify the accuracy of the proposed model and evaluate the performance improvement of the proposed SSB scheme, simulations are conducted. The simulator was written in C++ language[1]. Results were obtained by running 1,000,000 successful ranging processes from an MS in a saturated environment, where the saturated environment means that we assume a different number of MSs, i.e., $N_{ms} \in \{1, 5, \cdots, 50\}$, in the same neighboring BS, and every MS continually performs the initial ranging. Fig. 5 shows the network topology for the simulation. The simulation uses the MAC parameters of the IEEE 802.16 OFDMA TDD system suggested by the

---

[1] http://brass.cs.nctu.edu.tw/ssb/

$$A(str_k) = \begin{cases} A(\epsilon 0^x 1) = \dfrac{\displaystyle\sum_{i \ge x} P\{0^i 1\} \times (i+1)}{\displaystyle\sum_{i \ge x} P\{0^i 1\}} \times \alpha & , str_k \text{ contains exactly one '1'} \\ A(\epsilon 0^{q_0} 1 0^{q_1} \cdots 0^{q_z} 1) = A(0^{q_z} 1) = (q_z + 1) \times \alpha & , str_k \text{ contains two or more '1's} \end{cases} \tag{19}$$

TABLE I
DESCRIPTION OF NOTATION

Notation	Description
W	Backoff window size for ranging
W_{ssb}	Self-backoff window size used in the SSB scheme
R_w	Backoff counter before starting a periodic scanning used in the SSB scheme
S_{ro}	Average ranging opportunities per frame on the neighboring BS
P_c	Collision probability of a ranging request
P_m	Missing probability of a ranging response
T_f	802.16 MAC frame duration (in ms)
F_{scn}	Number of frames in a scanning interval
F_{itvl}	Number of frames in an interleaving interval
F_w	Maximal delay for a BS to reply a ranging response (in number of frames)
RO_P	A variable denotes the period between two RO-frames (in number of frames)
T3	Timeout for waiting a ranging response (in ms)
N_{ms}	Number of MSs which are performing ranging
N_{cdma}	Number of CDMA codes for ranging
N_{ro}	Number of ranging opportunities on an RO-frame
P_f	Probability of a ranging failure
α	Average rate of ranging attempts per frame from an MS
T_{a_ssb}	Average association latency for the SSB scheme

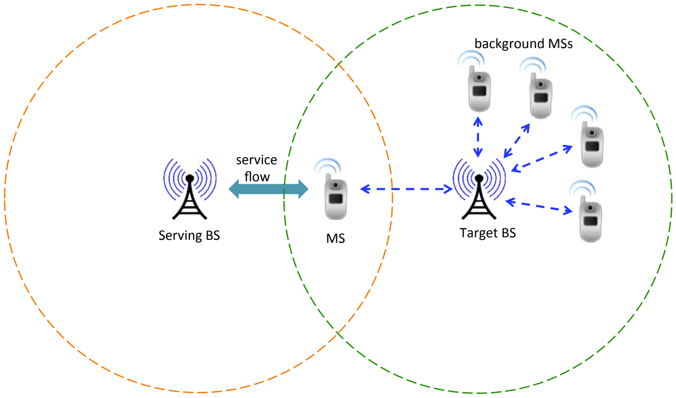

Fig. 5. Network topology for the simulation.

IEEE 802.16 specification [1] and the WiMAX Forum [17], and the scan parameters suggested by the references [2]–[5]. The simulations in this study assume that the neighboring BS allocates an RO-frame every two to six frames and each RO-frame has one ranging opportunity, i.e., one ranging opportunity every four frames on average. The serving BS or the neighboring BS advertises $S_{ro} = 0.25$, and thus an MS can estimate the self-backoff window. For example, if an MS's backoff window is four ranging opportunities, the self-backoff window of the MS is 16 frames ($W_{ssb} = W/S_{ro}$, $16 = 4/0.25$). An IEEE 802.16 MAC frame was set to five ms, the number of CDMA code for the contention-based ranging was eight, the initial backoff window size for IEEE 802.16e was 16, and the maximal backoff window size was 1024. After the BS detected a ranging request, it replied a ranging response within five frames. The delay for replying ranging response was uniformly distributed within five frames. The timeout for waiting the ranging response for an MS was 200 ms, i.e., *T3* = 200, and the retry limit of ranging was 16. This study focuses on the examination of procedures, protocols, and state machines of the association process. To evaluate the performance of the conventional and proposed association mechanisms, we do not take the physical layer characteristics such as channel model and MSs' mobility model into consideration at this stage, and we assume that no channel error is involved. The duration of a scanning interval for each session was decided by the delay and jitter constraints of real-time communication [2]. For example, the delay is less than 50 ms (10 frames) for an audio session and 100 ms (20 frames) service for a video session. The simulations consider two real-time communication sessions: an audio and a video session. The 10-frame scanning interval and 10-frame interleaving interval, and the 20-frame scanning interval and 20-frame interleaving interval are applied to an audio and a video session, respectively.

In this paper, we assume that an MS may pause the communication with the serving BS and scan the neighboring BSs before conducting handover. The packets are buffered temporarily on the serving BS during scanning intervals. During interleaving intervals, extra radio resources are allocated to transfer the buffered packets to the MS. Therefore, ertPS and rtPS service classes, which are able to request extra resources, have been used for delivering real-time communication services. Because our main focus is the examination of the contention of ranging opportunities and the ranging procedures during handover, we simplify the radio resource scheduling and management for uplink and downlink data bursts. We assume that the radio resource schedulers on BSs always assign a higher priority to the packet transmissions of real-time communication sessions, i.e., rtPS/ertPS connections, than nrtPS/BE connections. Therefore, we simulate the contention of ranging opportunities from all camped MSs and the contention of uplink/downlink data bursts from all rtPS/ertPS connections in the simulation. Moreover, we assumed that a BS does not deny MSs that request to camp to the BS, and that camped MSs are able to access initial ranging opportunities without an access control.

Fig. 6 illustrates the association latency while various association strategies are used. The scanning with association scheme suffers from the highest association latency while an MS has an audio session. This is because an MS with an audio session requests a short scanning interval to meet the delay constraint of the audio session. The short scanning interval causes a high missing probability of a ranging response. The SSB and the scanning without association scheme both can reduce the association latency significantly. Fig. 7 shows the missing probability of a ranging response while different schemes are applied. The simulation results reveal that the missing probability of a ranging response is unaffected by the number of MSs competing for ranging opportunities, and it is determined by the lengths of a scanning interval, an interleaving interval, and the delay of the neighboring BS responses to the ranging request. As shown in Fig. 7, the missing probability of a ranging response is almost zero when the scanning without association and the SSB scheme are employed. A ranging response is never missed in the scanning without association scheme because no interleaving interval is involved. For the SSB scheme, we can reduce missing of ranging response messages by sending a ranging request in early frames of a scanning interval. Fig. 8 illustrates the collision probability of a ranging request, and Fig. 9 shows the probability of a ranging failure which presents the jointed effects from misses and collisions of ranging requests. The scanning without association scheme and the SSB scheme have a higher collision probability of a ranging request than the scanning with association scheme. This is because both schemes can reduce the duration of the association process, and more ranging attempts occur in a shorter association period. The scanning without association scheme and the SSB scheme can reduce overall ranging failure and the association latency. Average packet delay and packet jitter are shown in Figs. 10 and 11, respectively. The scanning without association scheme introduces 100–250 ms packet delay for both audio and video sessions. On the other hand, the scanning with association scheme and the SSB scheme can reduce the average packet delay by interleaving association procedures during scanning intervals. While the scanning with association scheme or the SSB scheme is applied, the packet jitter can be under 50 ms and 100 ms for an audio session and a video session, respectively. The packet jitter would be smaller with applying the SSB scheme.

To examine the reasons of improvement by applying the proposed SSB scheme, we further compare the performance of the conventional and SSB approach using different figures. When only one MS with an audio session is contending initial ranging resources, the probability of a ranging failure is contributed only by the missing probability of a ranging response. As shown in Fig. 7, the missing probability of a ranging response reaches 28.9% when the scanning with association scheme is applied. The proposed SSB scheme can reduce 25% of missing probability of a ranging response, i.e., save 700 ms association latency (an improvement of 350%), compared with the conventional approach. Avoiding loss of ranging responses reduces the association latency and packet delays significantly. Conversely, as shown in Fig. 9, the SSB scheme suffers from a higher probability of a ranging failure than the scanning with association scheme when more than 35 MSs are contending initial ranging resources. In this case, the proposed SSB scheme can still reduce 46% (approximately 800 ms) association latency, compared with the conventional approach. The improvement is derived mainly from the property of the proposed SSB scheme performing the backoff countdown process while staying connected with the serving BS.

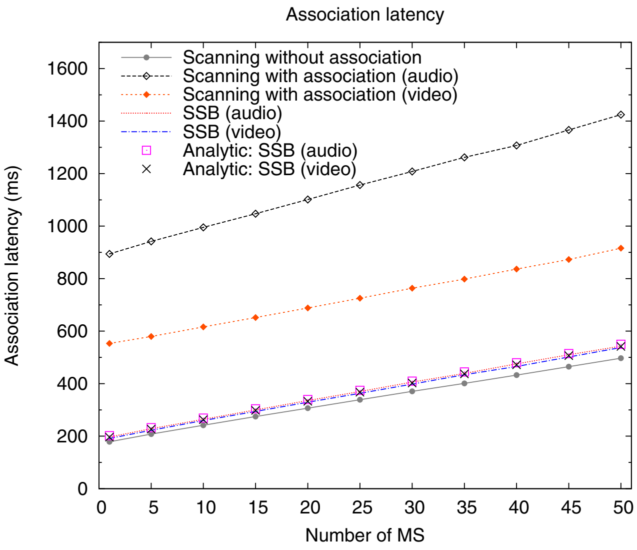

Fig. 6. Association latency while various scanning schemes are used.

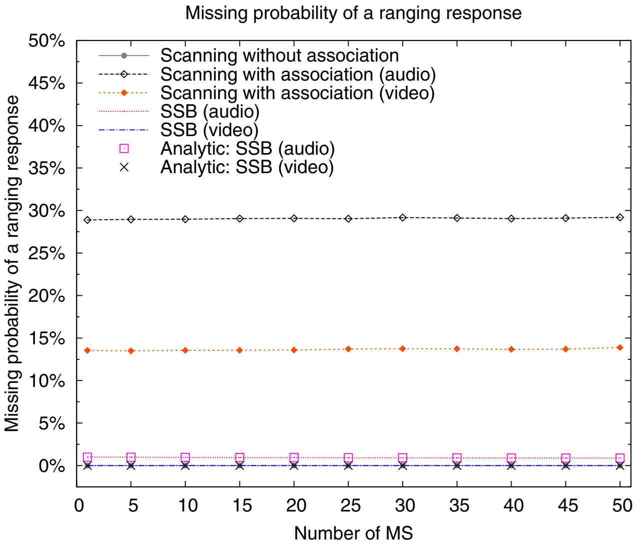

Fig. 7. Missing probability of a ranging response while various scanning schemes are used.

VII. Conclusions

This study proposed a low-latency scanning with association mechanism, the SSB scheme, for smooth handover in a Mobile WiMAX network. The SSB scheme does not require an MS to track all ranging opportunities, and can reduce power consumption during the scanning with association. SSB can be also utilized in other broadband wireless systems, such as 3GPP LTE for contention-based accessing. The performance

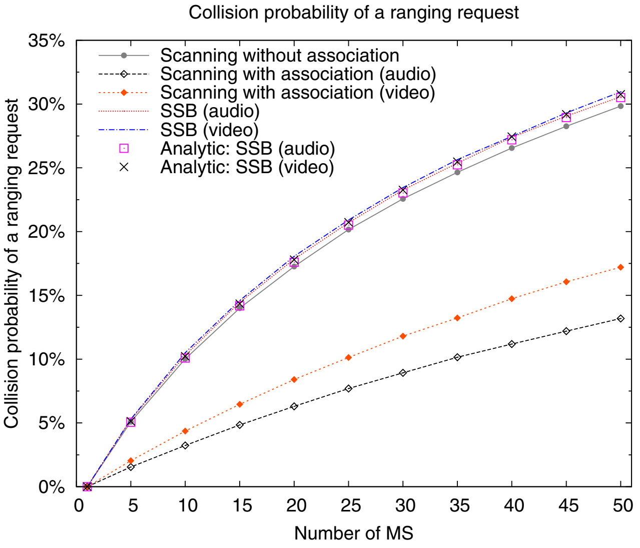

Fig. 8. Collision probability of a ranging request while various scanning schemes are used.

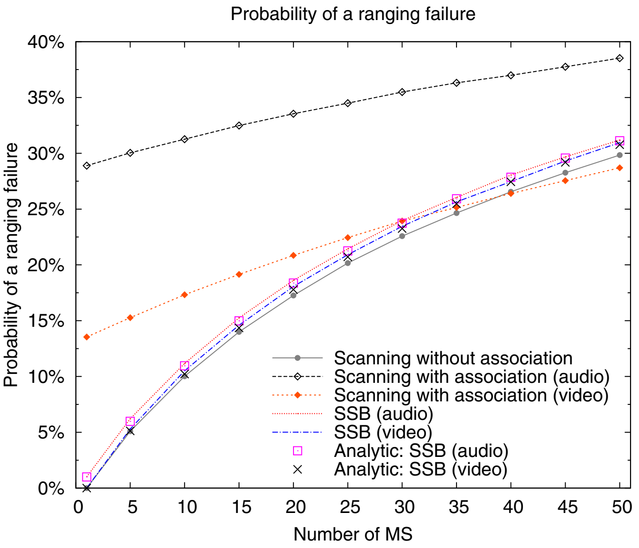

Fig. 9. Probability of a ranging failure while various scanning schemes are used.

of the SSB scheme is formulated and verified by simulations. These simulations show that the SSB scheme reduces the association latency by 61.9%–78.0% for audio services and 41.3%–65.4% for video services compared with the conventional scanning with association, and can also minimize packet delay and jitter during handover.

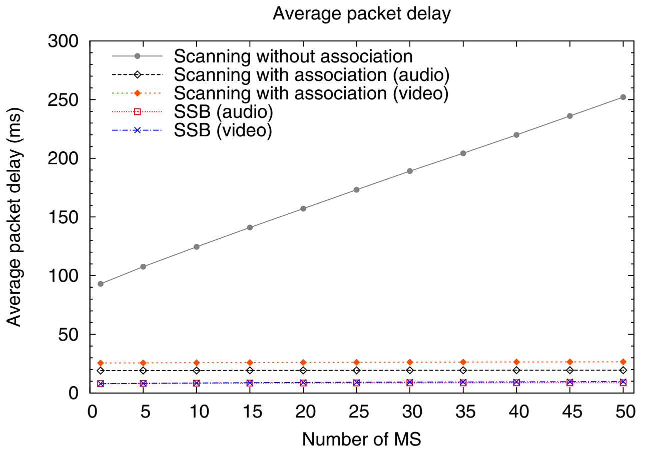

Fig. 10. Average packet delay during a handover period while various scanning schemes are used.

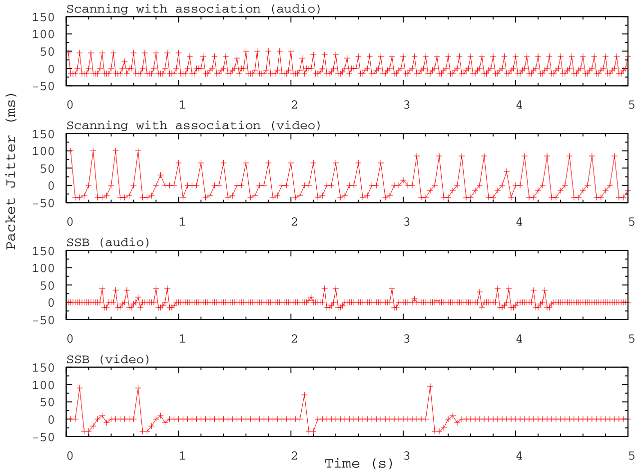

Fig. 11. Packet jitter during a handover period while various scanning with association schemes are employed.

ACKNOWLEDGMENT

The authors would like to thank MediaTek Inc. and National Science Council of the Republic of China for financially supporting this research under Contract NSC101-2219-E-009-010-, NSC101-2220-E-009-036-, and NSC 101-2219-E-009-001-.

REFERENCES

[1] IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems, IEEE Std. 802.16-2009, May 2009.

[2] R. Rouil and N. Golmie, "Adaptive channel scanning for IEEE 802.16e," in *Proc. 2006 IEEE Military Communications Conference*, pp. 1–6.

[3] S.-L. Tsao, Y.-L. Chen, and C.-H. Chang, "Evaluation of scan and association process for real-time communication in mobile WiMAX," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3320–3323, Oct. 2010.

[4] D. H. Lee, K. Kyamakya, and J. P. Umondi, "Fast handover algorithm for IEEE 802.16e broadband wireless access system," in *Proc. 2006 International Symposium on Wireless Pervasive Computing*, pp. 1–6.

[5] S. Choi, G.-H. Hwang, T. Kwon, A.-R. Lim, and D.-H. Cho, "Fast handover scheme for real-time downlink services in IEEE 802.16e BWA system," in *Proc. 2005 IEEE Vehicular Technology Conference*, vol. 3, pp. 2028–2032.

[6] W. Jiao, P. Jiang, and Y. Ma, "Fast handover scheme for real-time applications in mobile WiMAX," in *Proc. 2007 IEEE International Conference on Communications*, pp. 6038–6042.

[7] L. Chen, X. Cai, R. Sofia, and Z. Huang, "A cross-layer fast handover scheme for mobile WiMAX," in *Proc. 2007 IEEE Vehicular Technology Conference*, pp. 1578–1582.

[8] H. Fattah and H. Alnuweiri, "A new handover mechanism for IEEE 802.16e wireless networks," in *Proc. 2008 Wireless Communications and Mobile Computing Conference*, pp. 661–665.

[9] G. Dong and J. Dai, "An improved handover algorithm for scheduling

services in IEEE8O2.16e," in *Proc. 2007 IEEE Mobile WiMAX Symposium*, pp. 38–42.

[10] L. Lin, W. Jia, B. Han, and L. Zhang, "Performance improvement using dynamic contention window adjustment for initial ranging in IEEE 802.16 P2MP networks," in *Proc. 2007 IEEE Wireless Communications and Networking Conference*, pp. 1877–1882.

[11] L.-W. Chen and Y.-C. Tseng, "Design and analysis of contention-based request schemes for best-effort traffics in IEEE 802.16 networks," *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 602–604, Aug. 2008.

[12] J. Delicado, Q. Ni, F. Delicado, and L. Orozco-Barbosa, "New contention resolution schemes for WiMAX," in *Proc. 2009 IEEE Wireless Communications and Networking Conference*, pp. 1–6.

[13] J. He, K. Guild, K. Yang, and H.-H. Chen, "Modeling contention based bandwidth request scheme for IEEE 802.16 networks," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 689–700, Aug. 2007.

[14] H. Vu, S. Chan, and L. Andrew, "Performance analysis of best-effort service in saturated IEEE 802.16 networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 460–472, Jan. 2010.

[15] Q. Ni, L. Hu, A. Vinel, Y. Xiao, and M. Hadjinicolaou, "Performance analysis of contention based bandwidth request mechanisms in WiMAX networks," *IEEE Systems J.*, vol. 4, no. 4, pp. 477–486, Dec. 2010.

[16] S.-L. Tsao and Y.-L. Chen, "Mobility management in Mobile WiMAX," in *Mobile WiMAX: Toward Broadband Wireless Metropolitan Area Networks*. Auerbach Publications, pp. 219–232.

[17] WiMAX System Evaluation Methodology v2.0, WiMAX Forum, Dec. 2007.

You-Lin Chen received his B.S. degree in Dept. of Computer Information Science and M.S. degree in Dept. of Computer Science and Information Engineering from National Chiao Tung University in 2004 and 2006, respectively. His research interests include mobile computing and broadband wireless networks. He is currently a PhD candidate in Dept. of Computer Science, National Chiao Tung University.

Shiao-Li Tsao earned his PhD degree in engineering science from National Cheng Kung University in 1999. His research interests include embedded software and system, and mobile communication and wireless network. He was a visiting scholar at Bell Labs, Lucent technologies, USA, in the summer of 1998, a visiting professor at Dept. of Electrical and Computer Engineering, University of Waterloo, Canada, in the summer of 2007, and Dept. of Computer Science, ETH Zurich, Switzerland, in the summer of 2010 and 2011. From 1999 to 2003, Dr. Tsao joined Computers and Communications Re-search Labs (CCL) of Industrial Technology Research Institute (ITRI) as a researcher and a section manager. Dr. Tsao is currently an associate professor of Dept. of Computer Science of National Chiao Tung University. Prof. Tsao has published more than 85 international journal and conference papers, and has held or applied 18 US patents. Prof. Tsao received the Research Achievement Awards of ITRI in 2000 and 2004, Highly Cited Patent Award of ITRI in 2007, Outstanding Project Award of Ministry of Economic Affairs (MOEA) in 2003, and Advanced Technologies Award of MOEA in 2003. He also received the Young Engineer Award from the Chinese Institute of Electrical Engineering in 2007, Outstanding Teaching Award of National Chiao Tung University, and K. T. Li Outstanding Young Scholar Award from ACM Taipei/Taiwan chapter in 2008. He is a member of IEEE.