Evaluation of Scan and Association Process for Real-Time Communication in Mobile WiMAX

Shiao-Li Tsao, You-Lin Chen, and Chia-Hsiang Chang

***Abstract*—One of the most important issues in offering real-time communication services in a mobile environment is support for seamless handover between base stations during a communication session. In Mobile WiMAX, a mobile station may have to perform a scan and association process before handover, but unfortunately, this scan and association process introduces service disruption. In this letter, we investigate the scan and association latency in Mobile WiMAX, and evaluate its performance through analytic models and simulations.**

***Index Terms*—Mobile WiMAX, IEEE 802.16e, handover, scan and association.**

I. Introduction

THE development of WiMAX networks based on IEEE 802.16 has grown rapidly over the past several years. Furthermore, Mobile WiMAX based on IEEE 802.16e supports mobility, thus facilitating broadband communications in a mobile environment. A critical issue pertaining to Mobile WiMAX offering voice and video services is the provision of seamless handover during a real-time communication session. According to the IEEE 802.16e specification [1], handover consists of four phases of processes: network topology acquisition, scan and association process, handover procedures, and network re-entry procedures. During these processes, a mobile station (MS) may have to leave its serving base station (BS) and switch to other channels to perform certain operations. Therefore, the communication between the MS and the serving BS is temporarily suspended, and service disruption may occur. Such a service disruption influences the quality of service (QoS) of real-time communication in Mobile WiMAX.

Many studies have investigated and proposed solutions to speed up the procedures of handover in Mobile WiMAX [2]. Most of these studies assume that the scan and association process can be done in the background and before handover without influencing the current communication. Although, with full network assistant, the scan and association process can be quickly finished, it requires support from all BSs with frame-level time synchronization. On the other hand, the scan and association process with contention-based ranging is mandated by the WiMAX Forum and should be supported by all BSs and MSs. Unfortunately, the contention-based ranging may introduce long service disruptions and packet delays if the scan and association parameters are not properly configured [3]. Rouil and Golmie [3] thus proposed a mechanism to schedule the scan and association process and maintain the QoS of a communication session. However, the loss of ranging messages was not evaluated in their study. In this letter, we consider the scan and association process with contention-based ranging. The service disruption owing to the scan and association process is modeled and the performance is evaluated.

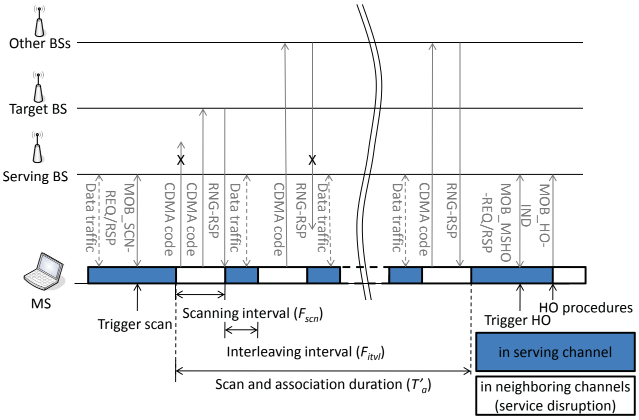

Fig. 1. Scan and association process.

II. Scan and Association Process

An MS can perform association with neighboring BSs during scan to reduce the latency introduced by initial ranging before the actual handover. Fig. 1 illustrates an example of the scan and association process. An MS which performs the scan and association has to switch to neighboring channels for ranging procedures during scanning intervals. The communication between the MS and the serving BS is temporarily suspended during these periods and service disruption may occur. To minimize the service disruption, the MS may return to the serving channel periodically and continue packet exchanges with the serving BS during interleaving intervals. Unfortunately, the neighboring BS is not aware whether the MS is in a scanning interval or not, and it may respond to the ranging request when the MS stays in the serving channel, i.e., during interval intervals. Thus, the ranging response is missing, and the MS has to perform the ranging with the neighboring BS again. Obviously, the scan and association process introduces service disruption during a communication session, and the lengths of the scanning and interleaving intervals also influence the ranging process and the overall scan and association duration. Therefore, a scan and association process that introduces less service disruptions of a real-time communication and can minimize the duration of the process is important.

III. Scan and Association Strategies

A. *Scan and Association Without Interleaving (SA)*

The first scan and association strategy is called Scan and Association without Interleaving (SA). An MS requests a long scanning interval, temporarily stops exchanging packets with the serving BS, and switches to neighboring channels. The

---

Manuscript received April 30, 2009; revised October 3, 2009 and May 10, 2010; accepted August 14, 2010. The associate editor coordinating the review of this letter and approving it for publication was D. Zeghlache.

The authors are with the Department of Computer Science, National Chiao Tung University, Hsin-Chu, 30010, Taiwan (e-mail: {sltsao, youlin}@cs.nctu.edu.tw; nelsonchang1218@gmail.com).

Digital Object Identifier 10.1109/TWC.2010.091510.090619

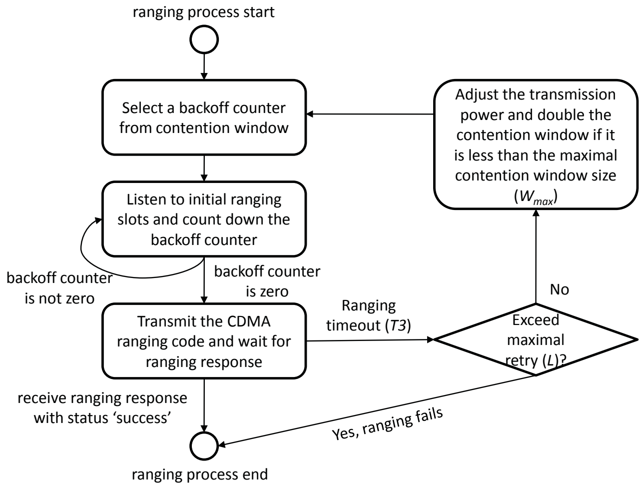

Fig. 2. Flowchart of a contention-based ranging.

MS does not switch back to the serving channel until the scan is completed; in other words, there is no interleaving interval during the scan and association operations. This approach is simple and straightforward, and can minimize the scan and association duration since no interleaving interval is involved. However, this approach introduces a long service disruption.

Fig. 2 illustrates the procedures for a contention-based ranging based on the binary exponential back-off (BEB) algorithm in IEEE 802.16e. Kwak et al. [4] analyzed the performance of the exponential back-off protocol, and Peyre et al. [5] also evaluated the contention resolution scheme in a WiMAX network. The BEB algorithm, which is used in bandwidth request and initial ranging in WiMAX networks, was studied in [6] and [7]. In this letter, we model the service disruption and the duration of the scan and association process based on these previous studies.

The parameters advertised by the serving BS for applying scan include the window size of the ranging back-off start B_s, the window size of the ranging back-off end B_e, and the ranging retry limit L for the neighboring BS. These parameters indicate the minimum and maximal contention window sizes, i.e., $W_0 = 2^{B_s}$ and $W_{max} = 2^{B_e}$, and the maximal retry of the ranging request. The initial ranging procedure is complete when the MS successfully receives the ranging response from the neighboring BS. If the MS cannot receive the ranging response within *T3* ms, which is defined in IEEE 802.16e, the ranging fails and the MS doubles its contention window size to perform the next ranging. Therefore, the contention window size of the i^{th} ranging request is denoted as:

$$W_i = \begin{cases} 2^i \times W_0, i \le m \\ 2^m \times W_0, m < i \le L \end{cases}, m = B_e - B_s \qquad (1)$$

In this letter, we assume that each frame has an average of S available slots for initial ranging and the MS picks up the back-off counter from the contention window based on a uniform random process. The average delay for waiting back-off counter at the i^{th} transmission, say d_i, is

$$d_i = \left\lceil \frac{W_i - 1}{2S} \right\rceil \times T_f, 0 \le i \le L \qquad (2)$$

where T_f denotes the frame duration in milliseconds.

Moreover, we assume that the loss of ranging requests comes only from collisions, not channel error. Then, the probability of ranging failure, P_c, can be derived from the number of MSs performing ranging, the size of the contention window, the ranging retry limit, and number of initial ranging CDMA codes, based on the studies in [4] and [5]. Also, we assume that the response delay is a uniform distribution and the neighboring BS responds to the MS in T_w ms if it receives the ranging request. This implies that the MS may receive a ranging response after one frame (T_f ms) to T_w ms. However, since the MS does not know whether the BS receives the ranging or not, the MS has to wait for the responses for *T3* ms, according to the specification. Therefore, the average time of a successful ranging, say T_a, for performing the scan and association with a neighboring BS can be calculated as:

$$T_a = \sum_{k=0}^{L-1} P_c^k \cdot (1-P_c) \cdot (D_k + k \cdot T3 + \frac{T_w}{2}), D_k = \sum_{i=0}^{k} d_i \quad (3)$$

The above equation only models the scan and association time for one neighboring BS. The equation can be also applied to multiple neighboring BSs, which may have different ranging parameters. Then, the summation of T_a for different neighboring BSs becomes the average scan and association duration. The total average scan and association duration can be also considered as the service disruption time for the SA approach, because the communication session is temporarily paused during the scan and association process.

B. Scan and Association With Interleaving (SAI)

The SA strategy performs consecutive scan and association processes without interleaving and may introduce a long service disruption. For the Scan and Association with Interleaving (SAI) strategy, an MS stays in the neighboring channel for F_{scn} frames for initial ranging and then returns to the serving channel for F_{itvl} frames for communication. The total duration of the scan and association process increases, since extra time is introduced for the interleaving intervals. Moreover, the ranging response may be missing if the MS leaves the neighboring channel just when the neighboring BS responds to the ranging request. P_c is the probability that the ranging request is lost due to collision, and P_m is the probability that the ranging response is missing due to the MS leaving the neighboring channel. Thus, we can model the average scan and association duration in frames as:

$$T_a{}' = \sum_{k=0}^{L-1} \Bigg((P_c + (1-P_c) \cdot P_m)^k \cdot (1-P_c) \cdot (1-P_m) \cdot (D_k{}' + k \cdot (T3 + \frac{F_{itvl} \cdot T_f}{2}) + \frac{T_w}{2}) \Bigg) \qquad (4)$$

where $D_k{}' = \sum_{i=0}^{k} d_i{}'$ and $d_i{}'$ is the average waiting time for the i^{th} transmission of the ranging request. Due to the involvement of interleaving intervals, the average delay for the back-off window increases to

$$d_i{}' = \left\lceil \frac{W_i - 1}{2S} \right\rceil \times T_f \times (1 + \frac{F_{itvl}}{F_{scn}}) \qquad (5)$$

In the above equation, we also consider the case that the timeout of the ranging response occurs when the MS is in

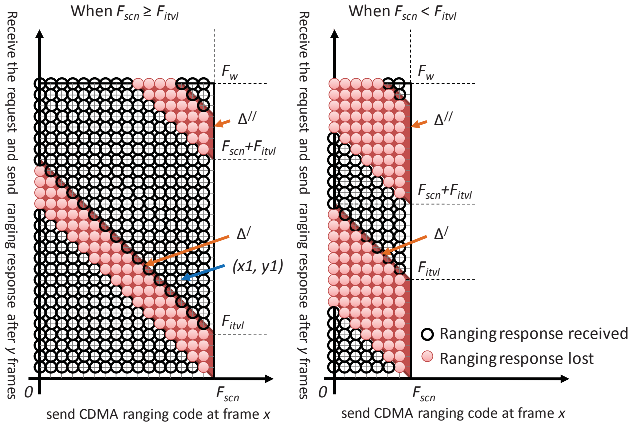

Fig. 3. Possible situations for ranging requests and responses.

an interleaving interval. The MS cannot immediately initiate the next ranging request when it is still in an interleaving interval. Therefore, the average time for waiting a ranging response timeout and initiating the next ranging request should be modified as $T3 + \frac{F_{itvl} \times T_f}{2}$.

Next, we model P_m, as shown in Fig. 3. The *x*-axis is the time that an MS sends the ranging request to the neighboring BS, and the *y*-axis presents the delay time of the BS response to the ranging request if the BS receives the request. Only situations where y is less than $F_w = \left\lfloor \frac{T_w}{T_f} \right\rfloor$ are considered, since we assume that the BS can respond to the ranging request in T_w ms if it receives the request. As can be seen from the figure, (x_1, y_1) means that the MS sends the ranging request at frame x_1, and the BS responds to the request after y_1 frame. When the value of $(x_1 + y_1)$ modulo $(F_{scn} + F_{itvl})$ is between F_{scn} and $F_{scn} + F_{itvl} - 1$, the neighboring BS responds to the ranging request when the MS returns to the serving channel. The ranging response is missing. Otherwise, the MS can receive the ranging response, since the MS is in a scanning interval. Therefore, the area in which the *x*-axis is less than F_{scn} and the *y*-axis is less than F_w shows all possible situations for ranging requests and responses. It is important to note that x and y must be integers. For the shadow area within the $F_{scn} \times F_w$ rectangle, the ranging response is missing. Thus, we can derive $P_m = \frac{\Delta}{F_{scn} \times F_w}$, where Δ is the size of the shadow area. If $F_w > F_{scn} + F_{itvl}$, we can rewrite $F_w = r \times (F_{scn} + F_{itvl}) + F_w{}'$ where r is an integer and $0 \leq F_w{}' < F_{scn} + F_{itvl}$. There are $r + 1$ separated shadow areas, i.e., $\Delta = r \times \Delta' + \Delta''$. $\Delta' = F_{scn} \times F_{itvl}$, and for Δ'', three possible situations should be considered.

$$\Delta'' = \frac{F_w{}' \times (F_w{}' - 1)}{2}, \text{if } 0 \leq F_w{}' < \min(F_{scn}, F_{itvl}) \tag{6a}$$

$$\begin{aligned} \Delta'' &= F_w{}' \times \min(F_{scn}, F_{itvl}) \\ &\quad - \min(F_{scn}, F_{itvl}) \times (\min(F_{scn}, F_{itvl}) - 1)/2, \\ &\quad \text{if } \min(F_{scn}, F_{itvl}) \leq F_w{}' < \max(F_{scn}, F_{itvl}) \end{aligned} \tag{6b}$$

$$\begin{aligned} \Delta'' &= F_{scn} \times F_{itvl} \\ &\quad - (F_{scn} + F_{itvl} - F_w{}') \times (F_{scn} + F_{itvl} - F_w{}' - 1)/2, \\ &\quad \text{if } \max(F_{scn}, F_{itvl}) \leq F_w{}' < F_{scn} + F_{itvl} \end{aligned} \tag{6c}$$

The service disruption time during the scan and association process is reduced to $F_{scn} \times T_f$ because the MS returns to the serving channel and resumes communication during interleaving intervals. However, the packets which are buffered during scanning intervals may suffer from additional delay. The average additional packet delay introduced by applying the SAI can be denoted as $\frac{F_{scn}}{2} \times T_f \times \frac{F_{scn}}{F_{scn}+F_{itvl}} = \frac{F_{scn}^2 \cdot T_f}{2(F_{scn}+F_{itvl})}$, which is no longer than a half of a scanning interval.

From the above equations and discussion, we can learn that, for a long scanning interval and short interleaving interval, the duration of the scan and association process can be reduced. However, the QoS of communication is scarified because the packet delay and service disruption time during the scan and association process increase for a long scanning interval.

IV. Analytic and Simulation Results

An IEEE 802.16e MAC-layer simulator written in C++ was developed to evaluate the performance of scan and association process. The simulator used a 5 ms frame length ($T_f = 5$), 50 ms timeout ($T3 = 50$), and a BPSK $1/2$ modulation and coding scheme. The low modulation and coding scheme—i.e., BPSK $1/2$—was assumed because the scan and association process was usually triggered when the signal strength of the serving BS became poor. It was assumed that ten MSs were performing ranging for handover to the neighboring BS. Moreover, the configuration of the neighboring BS was: $B_s = 5$, $B_e = 10$, $L = 16$, $T_w = 50$, two CDMA codes for ranging, and two slots per ranging region every 50 ms on average.

The first simulation considers the ideal scenario, where the MS performs the scan and association with only one neighboring BS and subscribes no real-time services. Fig. 4 shows both analytical and simulation results of the scan and association duration under different scanning and interleaving intervals. When comparing the simulation and analytical results, the differences are less than 5%. These results demonstrate the accuracy of the proposed models.

Without an interleaving interval ($F_{itvl} = 0$), the result presents the scan and association duration by applying the SA strategy. Fig. 4 shows that the scan and association duration when employing the SA is about 122 frames, i.e., 610 ms, which is the service disruption time during the scan and association process. As can be seen in Fig. 4, short scanning intervals and long interleaving intervals result in long scan and association durations. To reduce the duration of the scan and association process, long scanning intervals and short interleaving intervals are preferred. However, a long scanning interval introduces more communication service disruptions.

In the second simulation, we assume that an MS establishes a G.711 voice communication, which is 64 kbps and 20ms frame length, and the serving BS suffers from different workloads. When the serving BS has a high workload, the MS may have to stay in the serving channel for more frames—i.e., a longer interleaving interval— to maintain service. Two lines in Fig. 5 indicate feasible configurations when the serving BS workloads are 10% and 70%. A feasible configuration means that the configuration of scanning and interleaving intervals satisfies the bandwidth requirements for the real-time communication under the workload of the specific serving BS. A minimal interleaving interval can be determined for each given scanning interval to generate a feasible configuration.

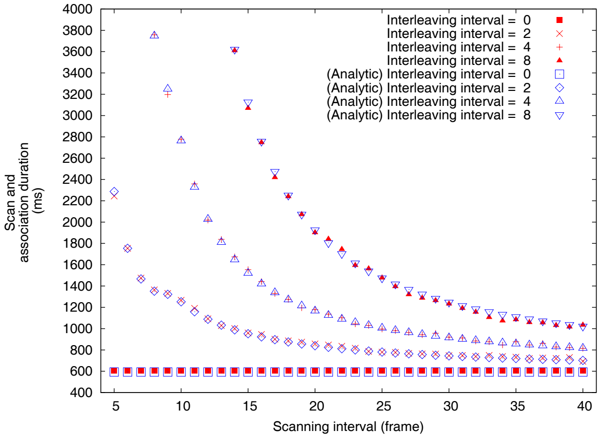

Fig. 4. Scan and association duration under different scanning and interleaving intervals.

If we consider the delay constraint, some configurations in the figure become infeasible. For example, the SA strategy introduces a 610 ms service disruption time, which may be too long for a real-time communication. If we set the service disruption time as 50 ms, configurations where the scanning interval is less than 10 frames are feasible for the SAI strategy. If real-time commutation can tolerate a maximum of 200 ms service disruption time, the configurations can accommodate scanning intervals with 40 frames. The above examples demonstrate that the scan and association strategy and the scanning and interleaving interval should be carefully chosen to minimize the service disruption time and the scan and association duration. The proposed analytic model helps the MS and/or the BS to easily determine the parameters of the scan and association process.

The next simulation considers an H.261 video communication, which is 256 kbps and 25 video frames per second. Video communication requires more radio resources, which implies a longer interleaving interval, when the scanning interval increases. In these cases, increasing the scanning interval may often increase the scan and association duration, especially when the serving BS suffers from a heavy workload. Fig. 6 demonstrates the scan and association duration when an MS establishes a video communication. In Fig. 5 and Fig. 6, each peak of the feasible configurations implies a change of the length of the interleaving interval owing to the bandwidth constraint of the real-time communication.

V. CONCLUSIONS

In this letter, we investigated the scan and association process and proposed analytical models of the scan and association duration for real-time communication in Mobile WiMAX. The accuracies of the proposed models were verified by simulations. With the proposed models, an MS or a serving BS can easily determine the parameters for performing the scan and association process without affecting the QoS of communication. The results demonstrate that with proper selection of the scanning and association interval, the service disruption time of real-time communication and the duration of the scan and association process could be both reduced.

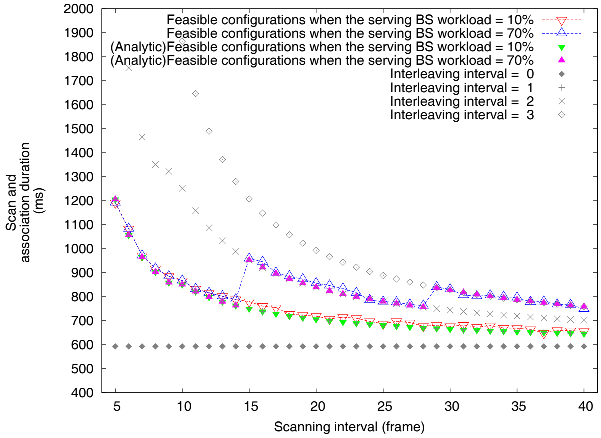

Fig. 5. Scan and association duration under different scanning and interleaving intervals and serving BS's workload for a voice communication.

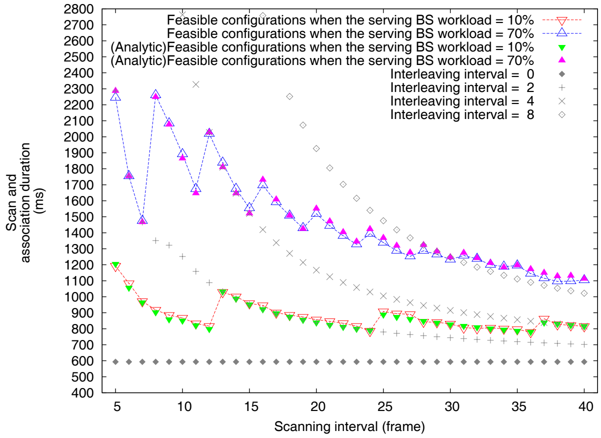

Fig. 6. Scan and association duration under different scanning and interleaving intervals and serving BS's workload for a video communication.

ACKNOWLEDGMENT

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 98-2220-E-009-013-, NSC 99-2220-E-009-045-, and NSC 98-2219-E-009-019-.

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] J. Chen, C.-C. Wang, and J.-D. Lee, "Pre-coordination mechanism for fast handover in WiMAX networks," in *Proc. 2nd International Conf. Wireless Broadband Ultra Wideband Commun.*, Aug. 2007.

[3] R. Rouil and N. Golmie, "Adaptive channel scanning for IEEE 802.16e," in *Proc. IEEE Military Commun. Conf.*, Oct. 2006, pp. 1–6.

[4] B. J. Kwak, N. O. Song, and L. E. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Trans. Networking*, vol. 13, no. 2, pp. 343–355, Apr. 2005.

[5] T. Peyre and R. ElAzouzi, "Performance analysis of single cell IEEE 802.16e wireless man," in *Proc. 32nd IEEE Conf. Local Computer Networks*, Oct. 2007, pp. 262–263.

[6] Y. P. Fallah, F. Agharebparast, M. R. Minhas, H. M. Alnuweiri, and V. C. M. Leung, "Analytical modeling of contention-based bandwidth request mechanism in IEEE 802.16 wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3094–3107, Sep. 2008.

[7] 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. IEEE Wireless Commun. Networking Conf.*, Mar. 2007, pp. 1877–1882.