Proactive Transmission-based QoS Provisioning for Multicast and Broadcast Services in Mobile WiMAX Systems

Ji Hoon Lee, Ted "Taekyoung" Kwon, Yanghee Choi
School of Computer Science and Engineering
Seoul National University, Korea
jhlee@mmlab.snu.ac.kr, {tkkwon,yhchoi}@snu.ac.kr

Sangheon Pack
School of Electrical Engineering
Korea University, Korea
shpack@korea.ac.kr

ABSTRACT

Mobile WiMAX systems have gained much attention for high transmission rates in cellular environments. One of the promising applications that can leverage high bit rates is multicast and broadcast services (MBSs). Although the expectation of IP-based multimedia streaming services in mobile environments is rising, QoS provisioning for those delay sensitive services is still challenging since mobility should be supported without intolerable delay. In this paper, we propose to transmit the streaming packets of MBSs proactively and probabilistically, so that the average handover delay perceived by a user is stochastically guaranteed. To quantify the tradeoff between the perceived handover delay and the bandwidth overhead of proactive transmissions, we develop an analytical model considering user mobility, user distribution, and session popularity. Comprehensive simulation is carried out to verify the analysis.

Categories and Subject Descriptors

C.2.1 [**Computer Communication Networks**]: Network Architecture and Design—*wireless communication*; C.4 [**Performance of Systems**]: Design Studies

General Terms

Performance, Management, Design

Keywords

Multicast, Broadcast, Broadband Wireless, Multimedia Streaming, WiMAX

1. INTRODUCTION

The need for bandwidth-efficient data distribution to a large number of users in cellular network environments has led to the definition of Multicast and Broadcast Services (MBSs) in Mobile WiMAX systems [1]. With the MBS support, a WiMAX network can offer more multimedia streaming services by utilizing its bandwidth resources efficiently. However, as users move across cell boundaries, handovers should be performed, which may incur some disruption in MBS session continuity. In this paper, we focus on how to bound the degree of service degradation when WiMAX systems provide delay-sensitive multimedia streaming MBSs.

In Mobile WiMAX, a mobile station (MS) normally performs hard handovers, in which all connections to the serving base station (BS) are broken before new connections are made to the target BS. As a result, some packets sent through the serving BS during the handover may be lost. One approach to mitigate this problem is deploying an MBS zone, defined by the IEEE 802.16e standard [2]. The MBS zone is a group of (typically adjacent) BSs transmitting the same content, which allows all MSs of the same MBS session in their cells to use the same IEEE 802.16 multicast connection and security key during handovers within the same zone. In this way, the MS can perform handover with the minimum delay, which in turn minimizes the disruption (or the lost MBS packets) as long as it moves within the MBS zone. This is called an intra-MBS zone handover. However, since every BS in the same MBS zone broadcasts the packets of the MBS session regardless of the presence of a user, the scarce wireless link bandwidth can be wasted.

There have been some research efforts on the support of multimedia streaming services over WiMAX networks. To reduce the delay during hard handovers, a fast handover scheme was proposed [5] by making IEEE 802.16 connection transfer between old and new BSs faster for real-time applications. This approach is not compatible with the IEEE 802.16e standard. Furthermore, it may support unicast services (e.g., video on demand), but it may not be an effective option for MBSs. How MBS zones can improve the link conditions of IEEE 802.16 multicast connections was studied [6]. The performance of MBS packet delivery within a single MBS zone can be improved by making use of macro diversity among nearby BSs, which may help to reduce disruption for intra-MBS zone handovers. But this work lacks a study of the effect of MBS zone sizes, and the inter-MBS zone handover scenario is not considered at all. We should consider inter-MBS zone handovers, since deploying large MBS zones that may waste too much bandwidth is impractical.

MICNET'09, September 21, 2009, Beijing, China.

In this paper, we propose to transmit MBS packets proactively and probabilistically to stochastically bound the average service disruption time for an MBS user. The central idea is to broadcast MBS packets with a certain probability in a test cell[1] despite no current MBS user if there is an MBS user in the adjacent cell(s). In this way, when the MBS user in the adjacent cell hands over to the test cell, he/she can continue receiving the packets of the MBS session with minimum discontinuity. The performance of the proposed scheme is highly dependent on how to determine the probability of proactive transmissions. To provide stochastic QoS provisioning for MBSs, the probability of proactive transmissions is determined by considering user mobility, spatial user distribution, and session popularity.

2. BACKGROUND

We first introduce the relevant concepts in the WiMAX MBS architecture and the motivation of this work.

MBS controller (MBSC): To provide MBSs, the WiMAX forum [3] defines a new entity, the MBSC, which is the source of MBS packets and the root of a multicast tree. An MBS session refers to a single logical multicast connection established between the MSs and the MBSC. The MBSC performs the authorization of MBS join requests, the resource reservation for the data path, the key distribution for secure MBS packet delivery and so on.

Inter-MBS zone handovers: The handover process of an MBS session from an old MBS zone to a new one consists of three parts (see [4] for details): 1) the link-level messages for the IEEE 802.16e handover, 2) the bearer signaling messages between the MS and the MBSC via the BS for the multicast connection setup/join in the new MBS zone, and 3) the MBS signaling messages to authorize the MBS join request and to reconstruct the multicast distribution tree across the WiMAX backbone network. The first delay occurs regardless of intra- and inter-MBS zone handovers. However, the second delay occurs only in inter-MBS zone handovers. The last delay occurs when an MS moves into a new MBS zone with no MBS users. Overall, reducing disruption requires that inter-MBS zone handovers should be avoided as much as possible by sizing the MBS zone appropriately. However, as the sizes of MBS zones increase, the number of BSs broadcasting MBS packets despite no MBS user in their cells may increase. Accordingly, the waste of wireless link bandwidth increases.

Location management area (LMA)-based intra-MBS zone handovers: In our previous study [4], we introduced the location management area (LMA)-based MBS scheme that partitions an MBS zone into multiple LMAs, and then selectively transmitted packets to the LMAs in which MBS users currently reside. In this way, we can increase the sizes of MBS zones to reduce the average handover delay without too much bandwidth waste. In the IEEE 802.16e context, an LMA is a paging group or a set of (typically adjacent) paging groups[2]. Like the intra-MBS zone handover, an inter-LMA handover does not require the bearer signaling messages, and hence it takes shorter delay than the inter-MBS zone handover. Instead, this scheme relies on the network capability keeping track of the location of every MS at the granularity of an LMA. Using LMAs allows large MBS zones to be used, so that the number of inter-MBS zone handovers can be reduced. Yet, the bandwidth waste is substantially reduced since MBS packets are broadcast only in LMAs with MBS users.

[1]Actually, we will consider a group of cells as a unit area of broadcasting MBS packets, to be detailed later.

[2]A paging group is a set of multiple (typically adjacent) cells, which is used to track the locations of MSs. A similar concept to a paging group is present in legacy cellular networks, which is termed a location area or a routing area.

Table 1: Notation for MBS delay analysis

μ_z	mean residence time in an MBS zone
μ_l	mean residence time in an LMA
μ_c	mean residence time in a cell
μ_s	mean MBS session duration
ρ_i	average number of users per unit area of ith most popular session
A_z	area of an MBS zone
A_l	area of an LMA
N	total number of LMAs in a network
S	total number of MBS sessions
Z_k	number of MBS zone handovers ($k = 1$: inter-MBS zone handover moving to inactive MBS zones, $k = 2$: inter-MBS zone handover moving to active MBS zones)
L_k	number of LMA handovers ($k = 1$: inter-LMA handover moving to inactive LMAs, $k = 2$: inter-LMA handover moving to active LMAs, $k = 3$: intra-LMA handover)
D_{Zk}	delay of the corresponding MBS zone handover Z_k
D_{Lk}	delay of the corresponding LMA handover L_k

Even though this LMA-based approach substantially reduces the overall disruption in an MBS session compared to the MBS zone-based broadcasting scheme, it is still difficult to configure network parameters to satisfy QoS requirements of MBSs. That is, we can analyze the average disruption time due to handovers given MBS zone and LMA sizes, but we may not be able to dynamically adjust the sizes of LMAs and MBS zones due to network deployment/operation issues. This limitation motivates us to propose how to control handover delays regardless of LMA or MBS zone sizes.

3. NETWORK AND MBS MODELING

We consider a WiMAX network in which there are total N cells; we assume that the LMA-based MBS handover is supported in the network. That is, the WiMAX network consists of multiple MBS zones, each of which is divided into multiple LMAs. For the sake of exposition, we assume that cells, LMAs, and MBS zones are square-shaped, and the residence times of an MBS user in an MBS zone, an LMA, and a cell follow exponential distributions with means μ_z, μ_l and μ_c, respectively, such that $\mu_z \geq \mu_l \geq \mu_c$. The MBS session duration time follows an exponential distribution with mean μ_s.

The total number of MBS sessions on air is S, and all sessions are ranked by popularity, which indicates how many users currently receive the packets of a particular MBS session. Let β_i be the conditional probability that a request arrives for the ith most popular MBS session ($i = 1, 2, ..., S$). β_i is drawn from a cut-off Zipf-like distribution [7], and is given by

$$\beta_i = \frac{\Omega}{i^\alpha}, \qquad \text{where } \Omega = \left(\sum_{i=1}^{S} \frac{1}{i^\alpha}\right)^{-1}, \quad 0 < \alpha \le 1.$$

The spatial distribution of users in the WiMAX network follows a two-dimensional Poisson distribution with net rate ρ^*, which is defined as the average number of users per unit area: $\rho^* = \lambda^*/\mu^*$, where λ^* is the users' arrival rate into the network and μ^* is the rate at which users leave the network. Then, the average number of users of the ith most popular session per unit area is $\rho_i = \beta_i \rho^*$.

From a perspective of MSs of a particular session, there are two states of MBS zones and LMAs: active and inactive. MBS zones which currently contain users of the current MBS session are called *active* MBS zones, whereas those without such users are called *inactive* MBS zones (Similarly, there are active and inactive LMAs.). Depending on the states of target MBS zones/LMAs, the handover delay becomes different since some signaling messages in handovers to inactive ones are skipped in handovers to active ones; for details, refer to [4]. Let the delay of an inter-MBS zone handover to an inactive MBS zone be D_{Z1}, and let the delay of an inter-MBS zone handover to an active MBS zone be D_{Z2}. Since an MBS zone is partitioned into multiple LMAs, intra-MBS zone handovers are classified into three cases, which are inter-LMA handovers to an inactive LMA, inter-LMA handovers to an active LMA, and intra-LMA handovers. Let each delay be D_{L1}, D_{L2}, and D_{L3}, respectively.

Table 1 summarizes the notation to analyze the disruption of an MBS session. Note that each of the following random variables indicates how many times an MBS user will experience each kind of handover delay during his/her session. Z_1 and Z_2 are the random variables for the numbers of inter-MBS zone handovers to inactive and active MBS zones, respectively. L_1 and L_2 are the random variables for the numbers of inter-LMA handovers to inactive LMAs and to active LMAs, respectively. L_3 is the random variable for the number of intra-LMA handovers. Following the approach in [4], the service disruption time for an MBS user is defined as the sum of all the expected handover delays during the service time of an MBS session, which can be expressed as

$$\begin{aligned}\text{Disruption Time} = {} & E[Z_1] \cdot D_{Z1} + E[Z_2] \cdot D_{Z2} \\ & + E[L_1] \cdot D_{L1} + E[L_2] \cdot D_{L2} + E[L_3] \cdot D_{L3}, \quad (1)\end{aligned}$$

where

$$\begin{cases} E[Z_1] = \frac{\mu_s}{\mu_z} e^{-\rho_i A_z} \quad \text{and} \quad E[Z_2] = \frac{\mu_s}{\mu_z}(1 - e^{-\rho_i A_z}), \\ E[L_1] = \frac{\mu_s(\mu_z - \mu_l)}{\mu_l \mu_z} \cdot \Pr\{\textit{inactive LMA}\}, \\ E[L_2] = \frac{\mu_s(\mu_z - \mu_l)}{\mu_l \mu_z}(1 - \Pr\{\textit{inactive LMA}\}), \\ E[L_3] = \frac{\mu_s(\mu_l - \mu_c)}{\mu_c \mu_l}. \end{cases}$$

Note that $\Pr\{\textit{inactive LMA}\}$ denotes the probability that an LMA is inactive.

4. PROACTIVE TRANSMISSION-BASED MULTICAST AND BROADCAST SERVICES

In this paper, we propose a proactive transmission-based (and LMA-based) MBS scheme that not only transmits the MBS packets of session i to active LMAs, but also probabilistically transmits the MBS packets to inactive LMAs adjacent to active LMAs, as shown in Fig. 1. The probability of proactively transmitting the packets of session i in inactive LMAs is denoted by p_i, and the proposed scheme is called *proactive* MBS. Intuitively, when an MS moves from an active LMA to an inactive LMA, the MS should trigger rejoining the corresponding multicast session during its handover process. This results in reconstructing the multicast distribution tree, which incurs substantial delay due to the graft latency [8]. In other words, the proactive MBS scheme removes the tree update delay when an MS hands over from an active LMA to an inactive LMA if the inactive LMA already has joined the multicast tree and proactively broadcast MBS packets. From now on, an inactive LMA where MBS packets are proactively transmitted is called a "proactive" LMA.

Determining the value of p_i relies on popularity and delay requirements of MBS session i. In general, however, as the value of p_i increases, the number of LMAs transmitting data increases. Then, the average number of inter-LMA handovers to an inactive LMA will be decreased, and thus the average handover delay is reduced.

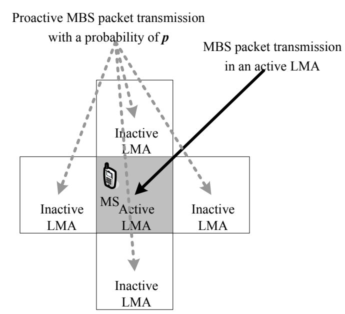


Figure 1: An illustration of the popularity-based proactive MBS.

4.1 Problem Formulation

To determine the probability p_i, we establish a constraint that the average handover delay for a session i user is kept to some specified value. We call this value *handover delay threshold*. At first, the probability that an LMA has x users of session i is $(\rho_i A_l)^x e^{-\rho_i A_l}/x!$ due to the two-dimensional Poisson process. By making x 0, the probability that a given LMA has no user is calculated. In the proposed scheme, each inactive LMA (that is adjacent to at least one active LMA of session i) will be changed to a proactive LMA with a probability of p_i. Therefore, $\Pr\{\textit{inactive LMA}\}$ is expressed as

$$\Pr\{\textit{inactive LMA}\} = (1 - p_i)e^{-\rho_i A_l}. \quad (2)$$

Then, the average handover delay can be calculated by dividing the service disruption time by the total number of handovers (μ_s/μ_c) during a session. From (1) and (2), the average handover delay for session i with proactive transmission probability p_i, $D(p_i)$, is expressed as

$$D(p_i) = \frac{\mu_c}{\mu_z}\left\{e^{-\rho_i A_z} D_{Z1} + \left(1 - e^{-\rho_i A_z}\right) D_{Z2}\right\} + \frac{\mu_c(\mu_z - \mu_l)}{\mu_l \mu_z}\left\{q_i e^{-\rho_i A_l} D_{L1} + \left(1 - q_i e^{-\rho_i A_l}\right) D_{L2}\right\} + \frac{(\mu_l - \mu_c)}{\mu_l} D_{L3}, \quad (3)$$

where $q_i = 1 - p_i$ and $0 \leq p_i \leq 1$.

We now proceed to state how the proactive MBS scheme can satisfy the QoS requirement of session i. Let the handover delay threshold of ith most popular session be γ_i. Using (3), the problem is formulated as:

Given: μ_c, μ_l, μ_z, ρ_i, A_l, A_z, D_{Z1}, D_{Z2}, D_{L1}, D_{L2}, D_{L3}, γ_i $(i = 1, 2, ..., S)$.

To find: For each session i, find the minimum p_i that satisfies $D(p_i) \leq \gamma_i$. This is shown in (4) at the top of the next page.

Increasing p_i will improve QoS for a given level of mobility, but the handover delay cutback comes at the cost of amplified traffic. Therefore, the bandwidth cost of an MBS session due to proactive transmissions should be analyzed. The bandwidth cost function for session i is denoted by $C(p_i)$, which is defined as the ratio of the number of active and proactive LMAs transmitting MBS packets to the total number of LMAs in the network. Let θ be the probability that an LMA which has no user is being considered as a proactive LMA candidate. Then, $C(p_i)$ can be expressed as

$$C(p_i) = \frac{n + p_i \theta (N - n)}{N}, \quad (5)$$

where n is the expected number of active LMAs and is given by $n = N(1 - e^{-\rho_i A_l})$. An LMA with no user can be proactive only if there should be at least one neighbor LMA with MBS user(s). Since the users of session i can be assumed to be evenly distributed with the net rate ρ_i, each user is located on any LMA with probability $1/N$. This yields

$$\theta = Pr\{\text{at least one adjacent LMA is active} \mid \text{a given LMA has no user}\} = \frac{\sum_{j=1}^{u} \binom{u}{j} [4/N]^j [(N-5)/N]^{u-j}}{[(N-1)/N]^u},$$

where u is the total number of users receiving session i in the network.

4.2 Overall procedure

The overall procedure to determine the p_i for QoS requirements of MBS session i with the constraint on the maximum bandwidth cost is described in Algorithm 1. As the real network and MBS parameters vary over time, we may need to run this algorithm at the MBSC periodically or when the perceived QoS deviation exceeds a certain threshold.

Algorithm 1 Popularity-based Proactive MBS Scheme.

```
// C_max is the maximum cost allowed in the network.
// M_i is the set of active LMAs for session i.
Update μ_c, μ_l and μ_z.
for all i such that 1 ≤ i ≤ S do
  Update ρ_i.
  Find minimum p_i^new such that D(p_i^new) ≤ γ_i.
  if Σ_{j=1}^{S} C(p_j) − C(p_i) + C(p_i^new) ≤ C_max then
    p_i ← p_i^new
  end if
  for all m ∈ M_i do
    for all inactive neighbor LMAs of m do
      Transmit the packets of session i with a probability
      of p_i.
    end for
  end for
end for
```

5. PERFORMANCE EVALUATION

5.1 Simulation Setup

We simulate an MBS service area (or an WiMAX network) comprising of 1024 cells, which is partitioned into four MBS zones. Each cell is an 1 km × 1 km square. Three sizes of LMAs are examined; each LMA consists of 1 × 1, 2 × 2, or 4 × 4 cells. For instance, if each LMA consists of 2 × 2 cells, it means that the location of each MS is tracked at a granularity of 4 km^2. There are total 100 MBS sessions which are ranked by the popularity (i.e., session 1 is the most popular). The total number of users (across all the sessions) is 1024 × 10, which means ρ^* is 10. For each session, the number of receivers is assigned by a Zipf-like distribution with $\alpha = 0.8$. Our simulation uses a 2-D random walk mobility model of which speed range is $[0, 120]$ km/h, and the service area is actually wrapping-around to remove the boundary effect. To quantify handover delays, our simulation follows the handover message flows described in [4], setting all data transfer delays between two entities and processing delays are 1. Since we try to reduce not the IEEE 802.16 MAC-layer handover process but the multicast tree update process, the MAC-layer handover delay is ignored and no link error is assumed. The simulation duration of each run is 30 minutes.

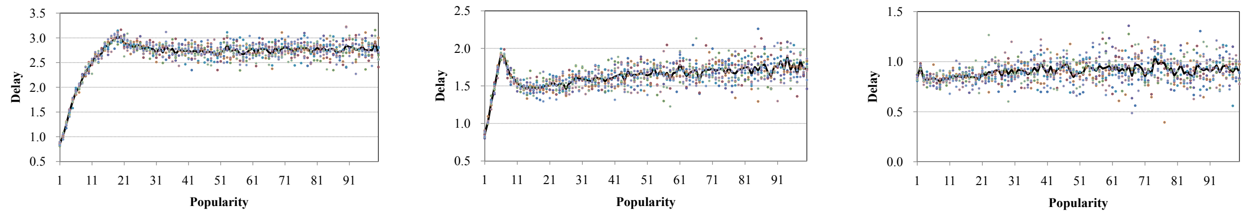
5.2 Computation of p_i

Given the above values we can get $D(p_i)$, and subsequently compute the probability p_i from (4). Fig. 2 plots the probability p_i as a function of the session rank for different handover thresholds and different LMA sizes. Notice that the less popular session has the higher values of p_i. An MS joining an unpopular session has less chance to encounter active LMAs, which results in larger average handover delay. Therefore, to reduce disruption, the number of proactive LMAs should be increased as the session popularity decreases. Also, the higher value of p_i is needed for the tighter delay requirement in Fig 2a, while Fig 2b shows that the smaller LMA size requires the higher values of p_i.

5.3 Simulation Results

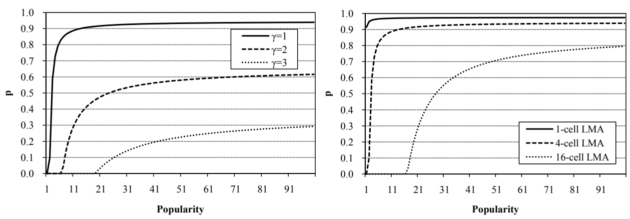
We simulate handover delays of MSs using the value of p_i obtained from the analytic modeling (Fig. 2). For each session rank, we carry out 10 simulation runs (plotted as dots) and the average handover delay of MSs is drawn as bold lines. Figs. 3a, 3b, and 3c show the average handover delays for different handover thresholds with the 4-cell LMA size. In most cases, the handover delays are below the threshold. When $\gamma = 3$ in Fig. 2a, our model suggests that p_i is zero for the session index $i \leq 18$, since the handover delay is expected to be less than 3 without proactive transmissions due to popularity of sessions. In Fig. 3a, the delays are rapidly

$$\text{Minimum} \quad p_i = 1 - \left[\frac{\left(\frac{\mu_l \mu_z}{\mu_z - \mu_l}\right) \left\{ \frac{\gamma_i}{\mu_c} - \frac{D_{Z1}}{\mu_z} e^{-\rho_i A_z} - \frac{D_{Z2}}{\mu_z} \left(1 - e^{-\rho_i A_z}\right) - \frac{D_{L3}(\mu_l - \mu_c)}{\mu_c \mu_l} \right\} - D_{L2}}{e^{-\rho_i A_l} \left(D_{L1} - D_{L2}\right)} \right]. \tag{4}$$



(a) $\gamma = 3$ (b) $\gamma = 2$ (c) $\gamma = 1$

Figure 3: Handover delays for different γ values are plotted with 4-cell LMA configuration; we conduct 10 simulation runs for every session. A dot is plotted for the delay of each simulation run, and the solid line represents their average values.



(a) 4-cell LMA (b) $\gamma = 1$

Figure 2: Computation of p_i: (a) For different handover thresholds; (b) For different LMA sizes.

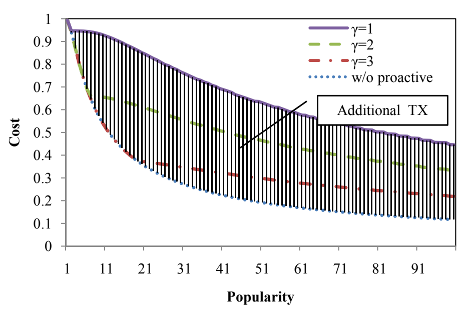


Figure 5: Bandwidth cost.

increased up to 3 as i goes to 18. As i exceeds 18, p_i is increased, so that the delays are reduced by proactive transmissions. Figs. 3b and 3c also confirm the results obtained from Fig. 2a.

Figs. 4a, 4b, and 4c show the average handover delays for different LMA sizes with $\gamma = 1$. As the LMA size increases, the number of crossing LMAs during a session is reduced; thus, p_i becomes smaller as shown in Fig. 2b. By adjusting p_i depending on LMA sizes, the delays are shown to be readily controlled. Therefore, our proposed scheme can adjust the probability of proactive transmissions to satisfy QoS requirements, irrespective of the LMA sizes.

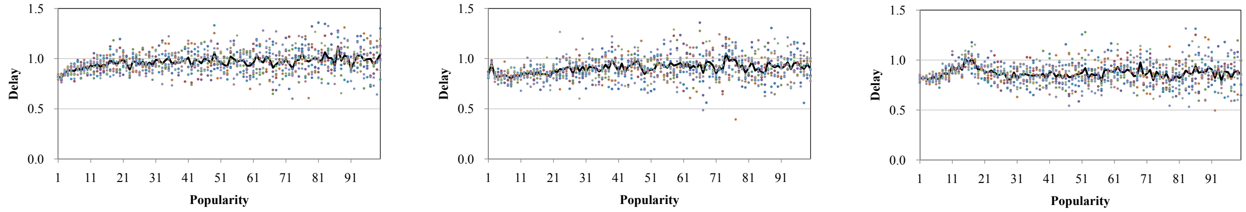
Fig. 5 depicts the bandwidth cost from Eq. (5) with 4-cell LMA sizes. The dotted curve (which is the minimum) represents the cost for transmitting MBS packets only to active LMAs, and therefore the shaded area between the dotted curve and the cost curve for $\gamma = 1$ shows how much cost is raised by the proposed scheme. The bandwidth cost significantly increases as the delay requirement becomes tighter. Figs. 6a, 6b, and 6c show the bandwidth costs measured by the simulations, compared to the dashed lines computed by (5). The dots are the bandwidth cost of each simulation run; the average cost of 10 runs is drawn as the solid line.

6. CONCLUSIONS

One crucial issue for multimedia streaming services in the WiMAX multicast and broadcast service (MBS) framework is how to provide QoS for recipients despite their mobility. We leverage our earlier work, the location management area (LMA)-based MBS scheme, which partitions an MBS zone into multiple LMAs to balance the average handover delay and the bandwidth usage overhead. In this paper, assuming the same LMA-based MBS framework, we seek to bound the disruption of an MBS session stochastically. To this end, the proposed scheme transmits MBS packets not only to the LMAs with MBS users, but also to their neighbor LMAs without MBS users probabilistically and proactively. By considering session popularity, user distribution, and user mobility, the probability of proactive transmissions in the neighbor LMAs without users is determined for each MBS session. Through extensive simulations, the analytic model to determine the probability of proactive transmissions is verified to satisfy the QoS requirements.

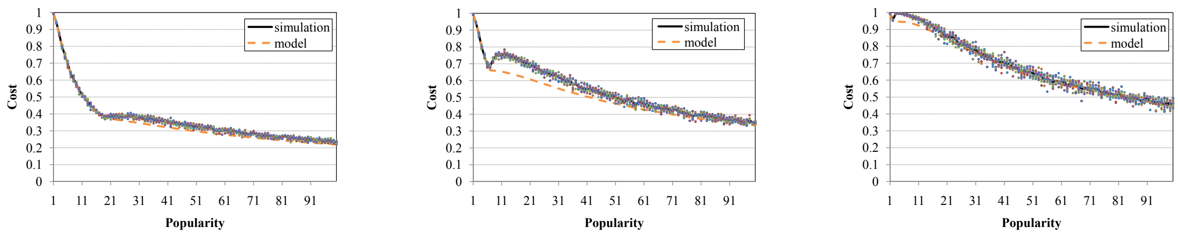
7. ACKNOWLEDGMENTS

This work has been supported in part by the IT R&D program of MKE/IITA [2007-F-038-03, Fundamental Technologies for the Future Internet] and in part by MEST, Korea under WCU (World Class University) Project [R33-2008-000-10044-0]. The ICT at Seoul National University provides research facilities for this study.



(a) 1-cell LMA (b) 4-cell LMA (c) 16-cell LMA

Figure 4: Handover delays for different LMA sizes are plotted with $\gamma = 1$.



(a) $\gamma = 3$ (b) $\gamma = 2$ (c) $\gamma = 1$

Figure 6: Bandwidth cost simulations with 4-cell LMA configuration: The dots and solid line represent the cost values for 10 simulation runs and their average, respectively; the dashed line represents the cost values obtained from (5).

8. REFERENCES

[1] B. Li, Y. Qin, C.-P. Low, and C.-L. Gwee. A Survey on Mobile WiMAX. *IEEE Communications Magazine*, 45(12):70–75, December 2007.

[2] IEEE Standard for Local and Metropolitan Area Networks–Part 16: Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1. IEEE Std. 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005.

[3] http://www.wimaxforum.org/.

[4] J. H. Lee, S. Pack, T. Kwon, and Y. Choi. Location Management Area (LMA)-based MBS Handover in Mobile WiMAX Systems. In *Proc. IEEE COMSWARE*, January 2008.

[5] W. Jiao, P. Jiang and Y. Ma. Fast Handover Scheme for Real-Time Applications in Mobile WiMAX. In *Proc. IEEE International Conference on Communications*, 2007.

[6] J. Wang, M. Venkatachalam, and Y. Fang, System Architecture and Cross-Layer Optimization of Video Broadcast over WiMAX. *IEEE Journals on Selected Areas in Communications*, 25(4):712–721, May 2007.

[7] L. Berslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proc. INFOCOM*, March 1999.

[8] A. Acharya, A. Bakre, and B. R. Badrinath. IP Multicast Extentions for Mobile Internetworking. In *Proc. INFOCOM*, March 1996.