Bandwidth Allocation for Wireless Data Dissemination in Multi-Cell Environments*

Jianliang Xu Dik L. Lee Bo Li

Department of Computer Science
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
{xujl,dlee,bli}@cs.ust.hk

Abstract

Effective data management and resource management are vital to the success of emerging mobile data applications. This paper investigates the problem of bandwidth allocation for data dissemination in a multi-cell environment, which has received little attention in the literature. The performance objective is to minimize the overall expected access latency. Optimal allocation technique is obtained to effectively allocate bandwidth among a cell cluster. We also extend the allocation technique to the case in which there is a constraint on the bandwidth allocation granularity. Numerical results show that the proposed schemes substantially outperform the uniform allocation *and* proportional allocation *schemes.*

1 Introduction

Due to resource limitations in a mobile environment, effective data management and resource management are vital to the success of emerging mobile data applications. There are two fundamental delivery methods for wireless data dissemination [1, 4, 6]:

- **On-Demand Access:** A client sends data requests uplink to the server, and the server returns the results to the client individually.
- **Data Broadcast:** The server periodically broadcasts information to the entire client population, and the clients monitor the broadcast channel to retrieve the data of their interest.

On-demand access provides fast response for a light-load system but the performance will deteriorate rapidly as workload increases. On the contrary, data broadcast can scale up to a very large client population and fit well to an asymmetric communication environment.

A lot of work has been done on bandwidth allocation for data dissemination in order to achieve a better access performance [1, 3, 4, 6]. For data broadcast, [1, 3] proposed methods in which more bandwidth is assigned to frequently accessed items and less to infrequently accessed items such that the overall access latency is optimized. In [2, 4], a hybrid system is suggested to combine on-demand and broadcast-based data dissemination. The bandwidth allocation between on-demand access and data broadcast was analyzed in [4, 6]. Unfortunately, these studies are confined to single-cell environments.

In a multi-cell environment, such as a cellular network, bandwidth allocation becomes much more complex than that of a single-cell environment. This is because neighboring cells in a cellular network are not allowed to use the same frequencies for communication simultaneously because of signal interference [8, 13]. As such, the available bandwidth for a system is shared by a group of neighboring cells. On the other hand, if two cells are apart sufficiently, the same frequencies can be *reused* in these two cells. As a result, bandwidth allocation cannot be considered on a cell-by-cell basis. Thus, the allocation of frequencies/bandwidth to each cell so as to achieve better balance among the cells is a challenging task. Although channel allocation for multi-cell voice communications has been extensively explored (e.g., [5, 7, 8, 11, 10, 13]), these studies aim at minimizing call blocking/dropping probabilities or improving carried traffic while ensuring certain level of QoS requirement. On the other hand, the characteristics of mobile data applications, which are mostly concerned with access latency, are not considered at all.

---

*The work was supported in part by the Research Grant Council, Hong Kong SAR, China under grant numbers DAG99/00.EG09, HKUST6163/00E, and HKUST6241/00E.

This paper investigates the problem of bandwidth allocation for wireless data dissemination in a multi-cell environment. As the first step, in this paper we study this bandwidth allocation problem for a cell cluster within which frequencies are not reused. The objective is to find out the optimal bandwidth allocation scheme which minimizes the overall data access latency. Optimal allocation technique is developed to effectively allocate bandwidth among a cell cluster. We also extend the allocation technique to the case in which there is a constraint on the bandwidth allocation granularity. The proposed allocation techniques are evaluated by a set of experiments. Numerical results show that the proposed schemes substantially outperform the *uniform allocation* and *proportional allocation* schemes.

The rest of the paper is organized as follows. Section 2 provides a motivating example for bandwidth allocation among different cells. In Section 3, the bandwidth allocation problem is formulated. Section 4 describes the proposed allocation techniques. Numerical results are presented in Section 5. Finally, the paper is concluded in Section 6.

2 A Motivating Example

In this section, a simple example is presented to motivate the study on bandwidth allocation for data dissemination in a multi-cell environment. As mentioned before, the performance metric adopted in this study is *access latency*, which is defined as the time elapsed from the moment when a request is submitted to the time when the request is serviced. In this simple example, we assume that the cellular network consists of two cells A and B, each of which is associated with a database containing four data items of equal size 1 Kb. In both cells, the four data items are broadcast on air based on a *flat broadcast* program, i.e., in a round-robin manner. A mobile client monitors the broadcast in its current cell to retrieve the data of its interest. Assume that the aggregate data access rates are 1 and 4 for cell A and cell B respectively and that a total wireless bandwidth of 6 Kbps is shared by these two cells. Table 1 shows three different allocation solutions and their expected access latencies (The calculation of the expected latencies will be described in detail in Section 3).

In the first solution the bandwidth allocation is uniform. In the second method, the bandwidth allocated to each cell is linearly proportional to its aggregate access rate. From Table 1, neither of these two solutions yields the best overall access performance. The best overall performance is achieved by the third solution which allocates 2 Kbps to cell A and 4 Kbps to cell B. An intuitive explanation for this phenomenon is as follows. For non-uniform traffic loads, a heavy-load cell has more impact on the overall performance than a light-load cell and hence allocating more bandwidth to the heavy-load cell B can improve the overall performance. However, if the bandwidth is over-allocated to cell B, too little bandwidth is left to the light-load cell A. As such, cell A has a very poor latency and thus leads to a worse overall performance. This observation motivates us to quantitatively analyze the impact of skewed bandwidth allocation and find a better strategy in terms of overall access performance.

3 Problem Formulation

This section formulates the bandwidth allocation problem for a cell cluster within which frequencies are not reused. The bandwidth available for the data dissemination system covers certain frequency range of the spectrum. If a certain frequency range is assigned to cell i, it cannot be used other cells to avoid interference. It is assumed that data access requests follow a Poisson process for each individual cell. For data dissemination, each cell employs the data broadcast or the on-demand access method [1, 3]. For on-demand access, the uplink cost is ignored because it is generally very small and will not affect the performance results. It is also assumed that the bandwidth allocated to a cell can be aggregated to serve data broadcast or on-demand access [4, 6]. To facilitate further discussion, the following notations are defined:

- N: number of cells in the cluster;
- N_b: number of cells that employ the broadcast for data dissemination, $0 \leq N_b \leq N$;
- B: total amount of bandwidth available for the system;
- M_i: number of data items for cell i's database;
- $item_{i,j}$: data item j for cell i's database;
- λ_i: data access rate for cell i;
- λ: total data access rate for the cluster, $\lambda = \sum_{i=1}^{N} \lambda_i$;
- $p_{i,j}$: access probability for $item_{i,j}$ in cell i, $\sum_{j=1}^{M_i} p_{i,j} = 1$;
- $\lambda_{i,j}$: mean data access rate for $item_{i,j}$, i.e., $\lambda_{i,j} = p_{i,j}\lambda_i$;
- $l_{i,j}$: length of data item $item_{i,j}$;
- l_i^s: sum of item sizes for cell i, $l_i^s = \sum_{j=1}^{M_i} l_{i,j}$;
- l_i: average item size accessed for cell i, $l_i = \sum_{j=1}^{M_i} p_{i,j} l_{i,j}$;
- $s_{i,j}$: space distance between instances of $item_{i,j}$ in cell i's broadcast program;
- b_i: the amount of bandwidth allocated to cell i, $b_i > 0$;

	Bandwidth allocation (Kbps)		Expected Access Latency (s)		Overall Expected Access Latency (s)
	Cell *A*	Cell *B*	Cell *A*	Cell *B*	
Uniform	3.0	3.0	1.0	1.0	1.0
Proportional	1.2	4.8	2.5	0.625	1.0
Best	2.0	4.0	1.5	0.75	0.9

Table 1: An Example of Bandwidth Allocation in a Two-cell System

Without loss of generality, assume that cells $1, 2, \ldots, N_b$ $(0 < N_b \leq N)$ use the broadcast method and cells $N_b + 1, \ldots, N$ $(0 \leq N_b < N)$ employ the on-demand access method. To formulate the problem, we start by deriving the formula of the data access latency for a single cell. Let's first consider cell i where the broadcast method is employed. Suppose that a total of b_i bandwidth is assigned to cell i, and instances of $item_{i,j}$ are spaced with $s_{i,j}$ in the broadcast program. Thus, $item_{i,j}$'s broadcast frequency is $\frac{b_i}{s_{i,j}}$. The expected access latency for $item_{i,j}$ is given by:

$$\bar{t}^b(i,j) = \frac{s_{i,j}}{2b_i} + \frac{l_{i,j}}{b_i}, \quad 1 \leq j \leq M_i, 1 \leq i \leq N_b. \quad (1)$$

where the first term is the average waiting time and the second is the service time.

The expected access latency for cell i is obtained:

$$\begin{aligned} \bar{t}^b(i,b_i) &= \sum_{j=1}^{M_i} p_{i,j}\bar{t}^b(i,j) = \sum_{j=1}^{M_i} p_{i,j}(\frac{s_{i,j}}{2b_i} + \frac{l_{i,j}}{b_i}) \\ &= \frac{l_i + \frac{1}{2}\sum_{j=1}^{M_i} p_{i,j}s_{i,j}}{b_i}, \quad 1 \leq i \leq N_b. \quad (2) \end{aligned}$$

It can be seen that $\bar{t}^b(i,b_i)$ is inversely proportional to b_i, and $\bar{t}^b(i,b_i)$ can also be expressed with

$$\bar{t}^b(i,b_i) = \frac{1}{b_i}\bar{t}^b(i,1). \quad (3)$$

Now consider cell i where the on-demand access method is employed. The system is approximately modeled as an $M/M/1$ queueing model. Let l_i denote the average item size accessed for cell i and b_i the bandwidth allocated to cell i. The service rate is $\frac{b_i}{l_i}$, and we obtain the expected access latency for cell i:

$$\bar{t}^o(i,b_i) = \frac{1}{\frac{b_i}{l_i} - \lambda_i} = \frac{l_i}{b_i - \lambda_i l_i}, \quad N_b + 1 \leq i \leq N. \quad (4)$$

Thus, we have the overall expected access latency for the system as follows $(0 \leq N_b \leq N)$:[1]

$$\begin{aligned} \bar{t}(b_1, b_2, \ldots, b_N) &= \sum_{i=1}^{N} \frac{\lambda_i}{\sum_{j=1}^{N}\lambda_j}\bar{t}(i,b_i) \\ &= \frac{1}{\lambda}\Big(\sum_{i=1}^{N_b} \frac{\lambda_i \bar{t}^b(i,1)}{b_i} + \\ &\quad \sum_{i=N_b+1}^{N} \frac{\lambda_i l_i}{b_i - \lambda_i l_i}\Big). \quad (5) \end{aligned}$$

Our objective is to find out the optimal allocation of a total of B bandwidth to each cell such that the overall expected latency is minimized. The optimization problem is defined by

$$\begin{aligned} \text{(BA)} \qquad \text{Minimize} \quad & \bar{t}(b_1, b_2, \ldots, b_N), \quad (6) \\ \text{Subject to} \quad & \sum_{i=1}^{N} b_i = B, \quad (7) \end{aligned}$$

where $\bar{t}(b_1, b_2, \ldots, b_N)$ is given by (5).

4 Proposed Allocation Techniques

This section presents the proposed bandwidth allocation techniques. We first analyze the optimal allocation in Section 4.1, followed the extension to the case in which there is a constraint on the bandwidth allocation granularity.

4.1 Analysis of Optimal Bandwidth Allocation

In this section, we present the analysis for the optimal bandwidth allocation. For the *BA* problem, it is obvious that no feasible solution can be found if $B - \sum_{i=N_b+1}^{N} \lambda_i l_i \leq 0$. In this case, the system has to increase the available bandwidth to offer a good access performance. When $B - \sum_{i=N_b+1}^{N} \lambda_i l_i > 0$, the following theorem is obtained.

---

[1] Throughout this paper, it is assumed that the result is 0 for a sum function when its superscript is less than its subscript.

Theorem 1 *For the optimization problem BA, if $B - \sum_{i=N_b+1}^{N} \lambda_i l_i > 0$, the minimum overall expected access latency is achieved when the bandwidth allocated to cell i, b_i, is given by*

$$b_i = \begin{cases} \frac{\sqrt{\lambda_i \bar{t}^b(i,1)}\left(B - \sum_{i=N_b+1}^{N} \lambda_i l_i\right)}{\sum_{i=1}^{N_b} \sqrt{\lambda_i \bar{t}^b(i,1)} + \sum_{i=N_b+1}^{N} \sqrt{\lambda_i l_i}} & \text{if } 1 \le i \le N_b \\ \frac{\sqrt{\lambda_i l_i}\left(B - \sum_{i=N_b+1}^{N} \lambda_i l_i\right)}{\sum_{i=1}^{N_b} \sqrt{\lambda_i \bar{t}^b(i,1)} + \sum_{i=N_b+1}^{N} \sqrt{\lambda_i l_i}} + \lambda_i l_i & \text{if } N_b + 1 \le i \le N \end{cases} \tag{8}$$

The minimum overall expected latency is

$$\bar{t}^*(b_1, b_2, \ldots, b_N) = \frac{1}{\lambda(B - \sum_{i=N_b+1}^{N} \lambda_i l_i)} \left(\sum_{i=1}^{N_b} \sqrt{\lambda_i \bar{t}^b(i,1)} + \sum_{i=N_b+1}^{N} \sqrt{\lambda_i l_i}\right)^2 \tag{9}$$

Proof: This is a constrained-minimum problem. It can be solved using the Lagrange multiplier theorem. Please see [12] for the full proof. □

In the BA problem, $N_b = 0$ refers to the case where all the cells in the cluster employ the on-demand access method. In the other extreme, $N_b = N$ refers to the case where all the cells employ the broadcast method. In the following, for $N_b = N$ we consider two kinds of broadcast schedules of particular interest to us.

Flat Broadcast: Under this model, each cell broadcasts all its data items in a round-robin manner. The space distance between instances of each $item_{i,j}$ is the sum of item sizes for cell i, i.e., $s_{i,j} = \sum_{j=1}^{M_i} l_{i,j} = l_i^s$. Thus, we have the following corollary:

Corollary 1 *When the flat broadcast schedule is employed in each cell, the minimum overall expected latency, $\frac{1}{\lambda B}\left(\sum_{i=1}^{N} \sqrt{\lambda_i(l_i + \frac{1}{2}l_i^s)}\right)^2$, is achieved when the bandwidth allocated to cell i is proportional to $\sqrt{\lambda_i(l_i + \frac{1}{2}l_i^s)}$.*

The proof of this corollary is simply by following Theorem 1 and substituting the parameters of $s_{i,j}$ and $\bar{t}^b(i,1)$. Reconsider the example presented in Section 2. According to the above corollary, the minimum latency is achieved when the bandwidth allocated to cell A and cell B follows the ratio of $\frac{\sqrt{1\times(1+4/2)}}{\sqrt{4\times(1+4/2)}} = \frac{1}{2}$, and the optimal latency is $\frac{1}{(1+4)\times 6}\left(\sqrt{1 \times (1+4/2)} + \sqrt{4 \times (1+4/2)}\right)^2 = 0.9$. The result is consistent with the example.

Optimal Broadcast: Flat broadcast is simple to implement. However, its average access performance is poor when the client access patterns are skewed. For a single cell i, the minimum expected latency is achieved when spacing $s_{i,j}$ of $item_{i,j}$ is proportional to square root of its length and inversely proportional to square root of its access probability,[2] i.e., $s_{i,j} = \sqrt{\frac{l_{i,j}}{p_{i,j}}} \sum_{j=1}^{M_i} \sqrt{p_{i,j} l_{i,j}}$. Such a schedule minimizing the expected latency is termed as the *optimal broadcast schedule*. Therefore, following Theorem 1 we have the corollary for the optimal broadcast schedule as follows:

Corollary 2 *When the optimal broadcast schedule is employed in each cell, let $q_i = \sum_{j=1}^{M_i} \sqrt{p_{i,j} l_{i,j}}$, the minimum overall expected latency, $\frac{1}{\lambda B}\left(\sum_{i=1}^{N} \sqrt{\lambda_i(l_i + \frac{1}{2}q_i{}^2)}\right)^2$, is achieved when the bandwidth allocated to cell i is proportional to $\sqrt{\lambda_i(l_i + \frac{1}{2}q_i{}^2)}$.*

4.2 Constraint of Bandwidth Allocation Granularity

So far we have quantitatively analyzed the optimal bandwidth allocation among a cell cluster. However, in some situations, there is a constraint on the bandwidth allocation granularity. Suppose that the minimum bandwidth allocation unit is B_u, then the bandwidth allocated to a cell must be $n \cdot B_u$, where n is a positive integer. In this case, the bandwidth allocation needs to approximate the optimal result obtained in the previous subsection. A heuristic is proposed for such situations.

The pseudo code of the heuristic is described in Algorithm 1. It is a two-round bandwidth allocation scheme. In the first-round, the algorithm assigns bandwidth to a cell according to the optimal solution but truncates its fractional bandwidth. If the bandwidth is 0 after truncation, we set it to B_u to prevent the expected latency becoming infinite. In the second-round, if the remaining bandwidth is larger than 0, a bandwidth of B_u is repeatedly assigned to a cell which achieves the greatest incremental decrease[3] in the objective function $\bar{t}(1, 2, \ldots, N)$ until the remaining bandwidth is exhausted. On the other hand, if the

---

[2] This result can be easily obtained by extending the study performed in [3, 9], where waiting time was employed as the performance metric.

[3] That is, $\lambda_i \bar{t}(i, b_i) - \lambda_i \bar{t}(i, b_i + B_u)$, where $\bar{t}(i, b_i)$ is given in (2) and (4) for broadcast and on-demand access respectively.

remaining bandwidth is less than 0,[4] B_u bandwidth is repeatedly extracted from a cell which has more than B_u bandwidth and introduces the least incremental increase (i.e., $\lambda_i\bar{t}(i, b_i - B_u) - \lambda_i\bar{t}(i, b_i)$) in the objective function. The performance of this heuristic is investigated in Section 5.3.

Algorithm 1 Bandwidth Allocation with Allocation Granularity Constraint

```
(1) 1st-round allocation
remain_band := $B$;
sum_para := $\sum_{i=1}^{N_b} \sqrt{\lambda_i \bar{t}^b(i,1)} + \sum_{i=N_b+1}^{N} \sqrt{\lambda_i l_i}$;
$L := \sum_{i=N_b+1}^{N} \lambda_i l_i$;
for $i := 1$ to $N$ do
    if $i <= N_b$ then
        $b_i := \lfloor \frac{\sqrt{\lambda_i \bar{t}^b(i,1)}}{sum\_para} \cdot \frac{B-L}{B_u} \rfloor \cdot B_u$;
    else
        $b_i := \lfloor (\frac{\sqrt{\lambda_i l_i}}{sum\_para}(B - L) + \lambda_i l_i) \frac{1}{B_u} \rfloor \cdot B_u$;
    if $b_i == 0$ then $b_i := B_u$;
    remain_band := remain_band $- b_i$;
end for
(2) 2nd-round allocation
while remain_band $> 0$ do
    select $i$ for which $\lambda_i\bar{t}(i, b_i) - \lambda_i\bar{t}(i, b_i + B_u)$
    is the largest;
    $b_i := b_i + B_u$;
    remain_band := remain_band $- B_u$;
end while
while remain_band $< 0$ do
    select $i$ for which $b_i > B_u$ and
    $\lambda_i\bar{t}(i, b_i - B_u) - \lambda_i\bar{t}(i, b_i)$ is the least;
    $b_i := b_i - B_u$;
    remain_band := remain_band $+ B_u$;
end while
```

5 Performance Evaluation

This section presents the performance evaluation of the proposed bandwidth allocation techniques. We first describe the experiment configuration in Section 5.1. Section 5.2 evaluates the performance under different levels of access skewness. The effect of the bandwidth allocation granularity is examined in Section 5.3.

5.1 Basic Experiment Configuration

The basic experiment configuration is as follows. It is assumed that there is a total of 672 Kbps bandwidth available for the data dissemination system. The database associated with each cell consists of 1,000 data items. Data item sizes follow an exponential distribution with a mean of 10 Kb. Data access rates for the cells are set as follows. Each cell i is assigned a relative access rate λ_i^r. If cell i uses data broadcast, we set cell i's access rate to λ_i^r since the cell's access performance depends on access probabilities over items and is independent to its overall data access rate. If cell i uses on-demand access, we introduce a factor of utilization rate ρ, unless explicitly specified, cell i's access rate is set to $\frac{B \cdot \rho}{\overline{\lambda^r} \cdot \bar{l} \cdot N}\lambda_i^r$, where B is the total available bandwidth, $\overline{\lambda^r}$ is the average of λ_i^r's for the system, $\bar{l}$ is the average item length and N is the size of an interference cluster. In the experiments, ρ is set to 0.5 by default. In each cell i, data requests over the items follow the *Zipf* distribution [14] with a skewness parameter of θ_i, and θ_i is randomly chosen between 0 and 1. The experiment results are obtained after the system has reached the stable state, i.e., after 5,000 queries have been made in each cell. A total of 1,000,000 queries are evaluated.

To take a close look at the effects of non-uniform workloads, the size of the cell cluster, N, is set to 3. For the three cells, the relative access rates are set to $base\_rate^i$ ($i = 0, 1, 2$), and the access skewness parameters θ_i's ($i = 0, 1, 2$) are randomly set to 0.20, 0.59, 0.38, respectively. In order to evaluate the performance under different data dissemination methods, all the cells employ either the broadcast or the on-demand access method. For data broadcast, two kinds of schedules, flat broadcast and optimal broadcast [3], are considered.

In the experiments, besides the proposed allocation algorithms, two additional allocation schemes are evaluated:

- **Uniform Allocation**: The bandwidth is uniformly allocated to each cell regardless of their workloads.
- **Proportional Allocation**: This scheme allocates bandwidth to the cells skewly in an intuitive way. The bandwidth allocated to a cell is linearly proportional to its data access rate.

5.2 Effect of Access Skewness

In this subsection, the bandwidth allocation schemes are evaluated under various levels of workload skewness for the cluster. In the experiments, $base\_rate$

---

[4] This occurs when there are too many cells that have bandwidth of 0 after truncation and are assigned B_u bandwidth.

is varied from 1 to 6. The larger the *base_rate* value, the more skewed the workload of the system. The results are presented in Figures 1(a), 1(b), and 1(c) for flat broadcast, optimal broadcast, and on-demand access, respectively. In Figures 1(a) and 1(b), theoretical lower bounds are calculated according to (9). In Figure 1(c), theoretical lower bounds are not available because they require solving the $M/G/1$ model which is infeasible analytically. In the figures, the optimal allocation denotes the scheme that allocates bandwidth for the cluster according to (8).

From Figure 1, two observations are made. First, among the three allocation schemes, the proposed optimal allocation scheme performs the best. For data broadcast, in all cases, the performance of the optimal allocation approaches the theoretical lower bounds very closely. When the traffic loads are uniform in the cluster (i.e., *base_rate* = 1), all the schemes have the same performance. With increasing the workload skewness, the performance improvement of the optimal allocation scheme over the uniform and proportional allocation schemes becomes much greater. This is particularly true for on-demand access. For example, for on-demand access, when *base_rate* is larger than 3, the uniform allocation scheme makes the system exhausted because the heavy-load cells have too little bandwidth, whereas the optimal allocation scheme can adjust bandwidth allocation among the cells and achieve a very good performance. Second, although more bandwidth is allocated to the heavy-load cells in the proportional allocation scheme, its performance is not good, on average 24% and 23% worse than the optimal allocation scheme for broadcast and on-demand access, respectively. It is observed in the experiments that, in terms of variance (see Figure 2), the optimal allocation scheme performs similarly to the uniform allocation scheme but the proportional allocation scheme has a very worse performance. This verifies our intuition that the need of allocating more bandwidth to the heavy-load cells is over-estimated in the proportional allocation scheme.

5.3 Effect of the Bandwidth Allocation Granularity

This subsection investigates the performance with a constraint on the bandwidth allocation granularity. The performance of the proposed scheme is compared to the theoretical lower bounds, which are calculated without the constraint. Since on-demand access has no available lower bounds, only data broadcast is evaluated. The uniform and proportional allocation schemes are also included for comparison. For the uniform and proportional allocation, if the number of allocated bandwidth units for a cell is not an integer, the aggregate fractional bandwidth is re-assigned unit by unit in a way such that the overall performance is the best. The minimum bandwidth allocation unit is varied from 224 Kbps to ∞^{-1}, where ∞^{-1} denotes the case where there is no constraint on the bandwidth allocation granularity. In other words, the number of bandwidth allocation units is varied from 3 to ∞. In the figures, the optimal allocation denotes the scheme that allocates bandwidth for the cluster using Algorithm 1.

Figure 3(a) and Figure 3(b) show the results for flat broadcast and optimal broadcast, respectively. As can be seen, the optimal allocation scheme performs pretty well. When the number of allocation units is larger than 15, it has a very close performance to the theoretical lower bounds. When the number of allocation units is smaller than 15, its performance is only a little worse. For example, when there are only 3 allocation units available, the optimal allocation scheme is 19% worse than the lower bound for both flat broadcast and optimal broadcast. However, since in this case the only reasonable solution is to allocate one unit for one cell, in fact this is the best result that a practical scheme could obtain. In all cases, the optimal allocation scheme performs no worse than the uniform and proportional allocation schemes.

6 Conclusion and Future Work

Studies on bandwidth allocation for wireless data dissemination in a multi-cell mobile network have been limited in the literature. With the booming of mobile data applications, it is believed that this bandwidth allocation problem is becoming more and more important.

This paper has formulated the bandwidth allocation problem for a cell cluster, with the objective of minimizing the overall expected data access latency. Optimal allocation technique has been analyzed. We have also proposed a heuristic for the case in which there is a constraint on the bandwidth allocation granularity. Numerical results demonstrated that the proposed optimal allocation techniques show a substantially better performance than the uniform and proportional allocation schemes for non-uniform workloads.

This paper represents the first step into the multi-cell bandwidth allocation for wireless data dissemination. We are extending the work into a cellular mobile network where frequency reuse is allowed. It is assumed in this paper that each cell employs either broadcast or on-demand access for data dissemination. A third model is hybrid data dissemination, in which one needs to consider how much bandwidth is assigned to a cell and how to allocate the assigned bandwidth between broadcast and on-demand access. Thus, this hierarchical bandwidth allocation problem becomes more complicated, which deserves further work.

References

[1] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik. Broadcast disks: Data management for asymmetric communications environments. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 199–210, San Jose, CA, USA, May 1995.

[2] S. Acharya, M. Franklin, and S. Zdonik. Balancing push and pull for data broadcast. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 183–194, Tucson, AZ, USA, May 1997.

[3] S. Hameed and N. H. Vaidya. Efficient algorithms for scheduling data broadcast. *ACM/Baltzer Journal of Wireless Networks (WINET)*, 5(3):183–193, 1999.

[4] T. Imielinski and S. Viswanathan. Adaptive wireless information systems. In *Proceedings of the Special Interest Group on DataBase Systems (SIGDBS) Conference*, pages 19–41, Tokyo, Japan, October 1994.

[5] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey. *IEEE Personal Communications*, 3(3):10–31, June 1996.

[6] W.-C. Lee, Q. L. Hu, and D. L. Lee. A study of channel allocation methods for data dissemination in mobile computing environments. *ACM/Baltzer Journal of Mobile Networking and Applications (MONET)*, 4(2):117–129, 1999.

[7] B. Li, L. Yin, K. Y. Michael Wong, and S. Wu. An efficient and adaptive bandwidth allocation scheme for mobile wireless networks based on on-line local parameter estimations. *ACM/Kluwer Journal of Wireless Networks (WINET)*, 7(2):107–116, March 2001.

[8] S.-H. Oh and D.-W. Tcha. Prioritized channel assignment in a cellular radio network. *IEEE Transactions on Communications*, 40(7):1259–1269, July 1992.

[9] N. H. Vaidya and S. Hameed. Scheduling data broadcast in asymmetric communication environments. *ACM/Baltzer Journal of Wireless Networks (WINET)*, 5(3):171–182, 1999.

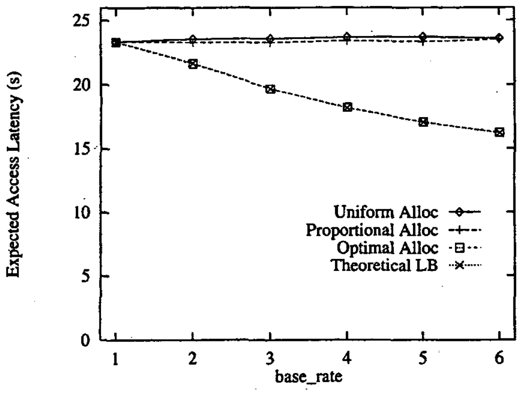
[10] A. L. Wijesinha, S. P. Kumar, and D. P. Sidhu. Handover and new call blocking performance with dynamic single-channel assignment in linear cellular arrays. *ACM/Baltzer Journal of Wireless Networks (WINET)*, 6(2):121–129, 2000.

[11] S. Wu, K. Y. Michael Wong, and B. Li. A dynamic call admission policy with precision qos guarantee using stochastic control for mobile wireless networks. *IEEE Transaction on Networking, to appear*, 2002.

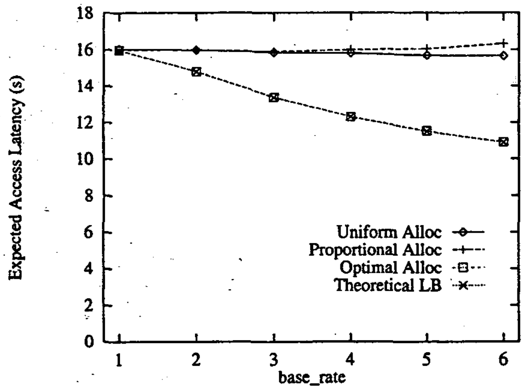
[12] J. Xu, D. L. Lee, and B. Li. A study on bandwidth allocation for data dissemination in cellular mobile networks. Technical Report HKUST-CS01-05, Department of Computer Science, Hong Kong Univ. of Science and Technology, February 2001.

[13] M. Zhang and T.P. Yum. The nonuniform compact pattern allocation algorithm for cellular mobile system. *IEEE Transactions on Vehicular Technology*, 40(2):387–391, May 1991.

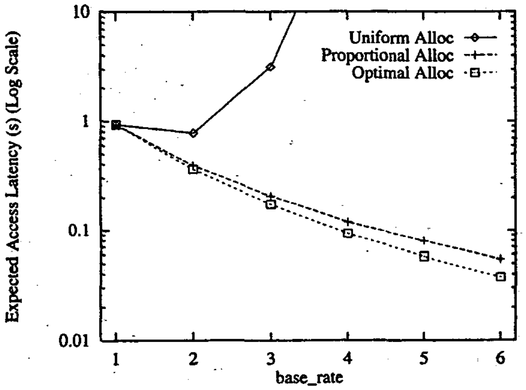
[14] G. K. Zipf. *Human Behaviour and the Principle of Least Effort.* Addison-Wesley, MA, USA, 1949.



(a) Flat Broadcast



(b) Optimal Broadcast



(c) On-Demand Access

Figure 1: Access Latency Performance under Skewed Workloads

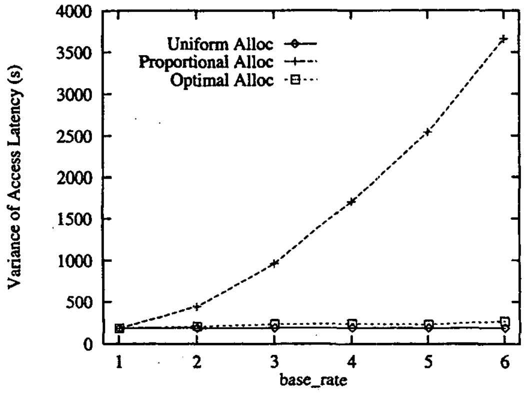
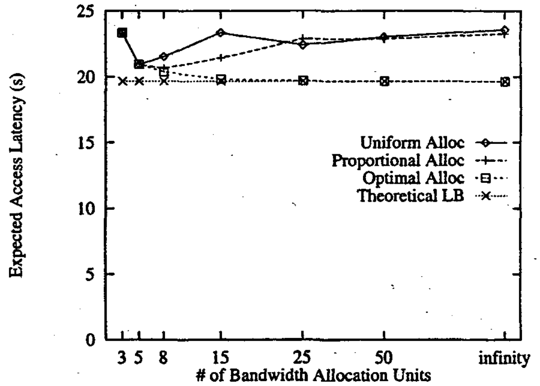
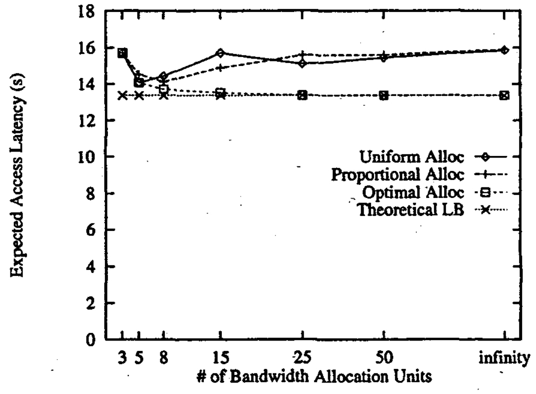


Figure 2: Variance Performance of Different Allocation Schemes (Flat Broadcast)



(a) Flat Broadcast



(b) Optimal Broadcast

Figure 3: Effect of Different Bandwidth Allocation Granularity