

Optimal Channel Allocation for Data Dissemination in Mobile Computing Environments

Qinglong Hu & Dik Lun Lee
Department of Computer Science
University of Science and Technology
Clear Water Bay, Hong Kong
qinglong@cs.ust.hk dlee@cs.ust.hk

Wang-Chien Lee
GTE Laboratories Incorporated
40 Sylvan Road
Waltham, MA 02254
wlee@gte.com

Abstract

This paper discusses wireless channel allocation problem for data dissemination in mobile computing systems. Methods for accessing data through broadcast and on-demand channels are described. We provide analytical models and cost formulae for the exclusive broadcast channels and the exclusive on-demand channels and propose a dynamic channel allocation algorithm for optimizing system performance. Our performance evaluation shows that dynamic channel allocation significantly improves system performance and the channel allocation algorithm gives us the optimal solution for various system parameter settings.

1 Introduction

One of the critical issues for mobile computing systems is how to efficiently utilize the wireless communication bandwidth. This paper presents wireless channel allocation methods for data dissemination in mobile computing environments.

Wireless broadcast allows an arbitrary number of users to simultaneously receive information broadcast on the air without any performance downgrade. In a previous paper [LHL97], we have shown that broadcast technique can effectively complement traditional *on-demand channels* for data dissemination in a mobile computing environment. We also proposed a hybrid channel allocation method which mixes the broadcast and on-demand channels in order to utilize system bandwidth more efficiently. In this paper, we propose a *dynamic channel allocation method* and obtain an efficient approximation of the optimal channel allocation. We also compare the data access efficiency between exclusive on-demand and dynamic channel allocation methods.

Studies on applying broadcast technology to mobile computing have been presented in the literature [IV94, SRB97, AFZ97]. [IV94] provided an architec-

ture for wireless information services and discussed the broadcast and on-demand channels. However, their model and analysis are different from ours. Also, there are no simulation results and comparisons in that paper. [SRB97] only investigated the dynamic adjustment of the hot-spot for the broadcast program. [AFZ97] evaluated the performance for fixed channel and data allocation schemes.

The rest of the paper is organized as follows. Section 2 briefly reviews the concepts involved with data dissemination in mobile computing systems. Section 3 provides analytical models and cost formulae to evaluate the data access efficiency for on-demand and broadcast channels, respectively. Section 4 develops the cost formulae for dynamic channel allocation method and an algorithm to approximate the optimal channel allocation based on the formulae. Section 5 is the performance evaluation and comparison of the exclusive on-demand and dynamic channel allocation methods. Section 6 concludes the paper.

2 Issues in Data Dissemination

In this section, we briefly introduce the mobile computing model used in this study and the issues related to data dissemination in mobile computing systems, such as channel allocation methods, data access methods and efficiency, and broadcast scheduling strategies.

2.1 Channel Allocation

Three classes of channel allocation methods can be defined:

- **Exclusive On-demand:** All of the channels are in on-demand mode. That is, data requests and results are delivered through point-to-point connections. This method is desirable when the number of queries is small compared to the number of channels available and when energy efficiency is not an issue for the mobile computer to transmit uplink requests.

- **Exclusive Broadcast:** Contrast to the exclusive on-demand method, all of the channels are in broadcast mode. Data items are broadcast periodically in broadcast channels [HGLW87]. A complete broadcast of all of the data items in a channel is called a *broadcast cycle*. Broadcast scheduling is an important research issue. In this paper, we choose to broadcast data items in sequence because this policy guarantees a bounded data access time. The exclusive broadcast method is useful when a small number of data items is of interest to a large group of users.
- **Dynamic Allocation:** This method dynamically allocates broadcast and on-demand channels for efficient data dissemination. The main idea is to make broadcast and on-demand channels complement each other under different workloads at the server. When the load is heavy, the broadcast channels may significantly relieve the load on on-demand channels by taking care of frequent accesses to hot data items. When the load is light, on-demand channels can take over to provide instantaneous access to data.

2.2 Data Access Methods

Without prior knowledge on how the data item of interest is provided by the server, a mobile computer has to send a data access request to the server. In response, the server either directly delivers the data item if it is only available through on-demand service, or replies with the *broadcast channel access information*, such as the channel frequencies, data identifier, the decryption key, expiration time, and estimated access time. Based on the information, the mobile computer may decide if it wants to access the data through the broadcast channel. If so, it terminates the connection and monitors the broadcast channels. If the user decides it wants the data right away, it returns a confirmation request to the server. The server will deliver the data and terminate the connection.

If the mobile computer already has the broadcast channel access information regarding to the data item of interest, it does not need to query the server unless the information is expired. The mobile computer may use the information to monitor broadcast channels to receive the data item.

2.3 Data Access Efficiency

The criteria used in this paper to evaluate data access efficiency of wireless channels is *access time*, which is the period from the time a mobile computer requests for a data item until the data is received. The broadcast channels and on-demand channels introduce dif-

ferent access overheads for the mobile users. For the broadcast channels, the overhead is due to the waiting of the data item of interest to arrive. We define the duration between the time the user starts to monitor the channels and the time when the searched data item is located as *probe time*. As such, to reduce the broadcast channel access time is to reduce the probe time. On the other hand, the overhead for accessing data through the on-demand channels is the waiting time for connecting to the server, which is dependent on the service load of the system. If the load is high, the users will have to wait longer.

3 The Analytical Models

To facilitate our study on channel allocation methods, we model the on-demand and broadcast channels and derive formulae for their access time performance. We consider a cell in the mobile computing system and make the following assumptions:

- the cell consists of a server and m mobile computers;
- the server maintains n data items, D_1, D_2, \dots, D_n , with average size s ;
- the mobile computers may issue read requests, each of which has size r , to the server;
- a mobile computer does not issue a new request before the previous request is completed;
- the number of read requests generated from the m mobile computers per unit time to D_1, D_2, \dots, D_n is, respectively, $\lambda_1, \lambda_2, \dots, \lambda_n$;
- the server supports c wireless channels and each channel has bandwidth b ;
- the number of potential mobile computers is greater than the number of channels available in a cell, i.e., $m \gg c$.

3.1 On-demand Channels

To simplify our analysis, we do not consider the hand-shaking time for the client to establish connections with the server and the process time for server to retrieve the requested data items from the disk. Since there are c channels in the cell, there are at most c simultaneous communication sessions. For m mobile computers in the cell, there are at most m requests to be serviced in the system. Thus, the service rate for an on-demand channel is, $\mu = b/(r + s)$. The aggregate arrival rate of requests from the mobile computers is, $\lambda = \sum_{i=1}^n \lambda_i$. The traffic intensity, which indicates the system workload, for the queuing model is [Jai91]: $\rho = \frac{\lambda}{c\mu}$.

We use the traffic intensity of the system, ρ , as the dividing factor of the system workload for the on-demand channels. The state transitions due to the arrival of data requests and completion of data requests

can be modeled as a birth-death system, in which each state of the diagram denotes the number of mobile computers requesting a data item during that state. For lack of space, we only state the expected access time without showing the details, for which readers can refer to [LHL98].

When $\rho > 1$, the on-demand channels have a heavy system workload. Since there are c channels and at most m requests buffered for services, the system can be modeled as a $M/M/c/m$ queue.

$$E[o] \approx m/(c\mu) \quad (1)$$

At a light system workload (i.e., $\rho < 1$), since there is no buffered data requests (i.e., service rate is greater than request rate), the system can be modeled as a $M/M/c$ queue. The expected access time for a mobile computer to retrieve a data item is:

$$E[o] = 1/\mu + P_c/(c\mu(1-\rho)) \quad (2)$$

where P_c is the probability that the mobile computers request c or more data items.

3.2 Broadcast Channels

Next we develop the cost model for the access time of the broadcast channels. To simplify our analysis, we assume that the mobile computer has the valid broadcast channel access information for the data items of interest. Since there are c channels used for broadcast, the aggregate bandwidth for broadcast channels is bc . Assume that there are k data items broadcast periodically in the c broadcast channels. The size of a broadcast cycle is ks/bc . The average probe time is $ks/2bc$. Thus, the access time for retrieving data through monitoring the broadcast channels is:

$$E[b] = ks/(2bc) + s/b \quad (3)$$

4 Dynamic Channel Allocation

In order to reach optimal data access efficiency in the cell, the system should dynamically reassign channels between the on-demand and broadcast services so that data requested by a large number of users can be broadcast. In this section, we provide a method to determine the optimal allocation of on-demand and broadcast channels. We assume that the n data items, D_1, D_2, \dots, D_n , are sorted by their access rates such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Let k ($1 \leq k \leq n$) be the number of the most frequently accessed data items made available on the broadcast channels and ϵ denote the probability that a data item is accessed by mobile computers through the on-demand channels, although the data item is available on the broadcast channels.

4.1 Actual Service Rate of On-demand

In order to calculate the average data access time of the dynamic channel allocation method, we have to

derive the actual service rate of the on-demand channels in a dynamic channel allocation environment.

Assume that the size of the message for broadcast channel access information is q . The average service rate for each on-demand channel is:

$$\mu' = b/(\omega_o(r+s) + \omega_b(r+q) + \omega_{b|o}(2r+q+s)) \quad (4)$$

where ω_o, ω_b and $\omega_{b|o}$, respectively, are the percentages of communication sessions established between the mobile computers and the MSS for requests of data which are only available through on-demand channels, requests of broadcast channel access information, and requests to deliver through on-demand channels data which are otherwise available on broadcast channels.

4.2 Channel Allocation Algorithms

Let us assume the total number of the communication channels in a cell is c and each channel has the same bandwidth b . These channels consist of two pools: the broadcast pool and the on-demand pool. Let c_b and c_o denote the number of channels in the broadcast pool and the number of channels in the on-demand pool, respectively. Thus,

$$c = c_b + c_o \quad (5)$$

To make the discussion simple, we define the arrival rate of the requests for hot data items available on the broadcast channels as, $\lambda_{hot} = \sum_{i=1}^k \lambda_i$, and the arrival rate of the requests for cold data items that can only be accessed via on-demand channels as, $\lambda_{cold} = \sum_{i=k+1}^n \lambda_i$.

Since ϵ percents of the hot data items are retrieved from on-demand channels, the actual arrival rate of requests for data items answered via on-demand channels is $\epsilon * \lambda_{hot} + \lambda_{cold}$. Thus, the overall expected access time for any data item is:

$$E[d] = \frac{1}{\lambda} * (\lambda_{hot} * E[b] * (1-\epsilon) + \lambda_{hot} * \epsilon * E[o] + \lambda_{cold} * E[o]) \quad (6)$$

where $E[o]$ denotes the expected access time for a mobile computer to retrieve a data item through an on-demand channel and $E[b]$ denotes the expected access time for retrieving data through monitoring broadcast channels.

Heavy load

In the following, we adopt approximation methods for a faster estimation of the optimal access time and channel allocation. First, we explore the situation when the system is under a heavy system workload ($\rho > 1$). Based on Equations (1), (3) and (5), we substitute $E[o]$, $E[b]$ and c_b in Equation (6) to obtain:

$$E[d] = \frac{1}{\lambda} (\lambda_{hot} * (1-\epsilon) * (\frac{ks}{2b(c-c_o)} + \frac{s}{b}) + (\lambda_{hot} * \epsilon + \lambda_{cold}) * (\frac{m}{c_o \mu'})) \quad (7)$$

where μ' is the actual service rate of on-demand channels in the dynamic channel allocation environment (see equation (4)).

To simplify further development of the formula, we define the following constants:

$$\begin{aligned} A &= \frac{1}{\lambda} * \lambda_{hot} * (1 - \epsilon) * \frac{ks}{2b} \\ B &= \frac{1}{\lambda} * (\lambda_{hot} * \epsilon + \lambda_{cold}) \\ C &= \frac{1}{\lambda} * \lambda_{hot} * (1 - \epsilon) * \frac{s}{b} \end{aligned}$$

Thus, Equation (7) can be rewritten as:

$$E[d] = \frac{A}{c - c_o} + \frac{Bm}{c_o\mu'} + C \quad (8)$$

We differentiate the function $E[d]$ on c_o and make it equal to zero (i.e., $\partial E[d]/\partial c_o = 0$) in order to evaluate the optimal point for c_o . As a result,

$$\frac{\partial E[d]}{\partial c_o} = \frac{(A\mu' - Bm)(c_o - c1)(c_o - c2)}{\mu'c_o^2(c - c_o)^2} \quad (9)$$

where,

$$\begin{aligned} c1 &= c\sqrt{mB}/(\sqrt{mB} + \sqrt{\mu'A}) \\ c2 &= c\sqrt{mB}/(\sqrt{mB} - \sqrt{\mu'A}) \end{aligned}$$

Since the number of channels working in on-demand mode c_o should satisfy $0 \leq c_o \leq c$, it is obvious that $0 \leq c1 \leq c$. On the other hand, if $mB > \mu'A$, then $c2 > c$; if $mB < \mu'A$, then $c2 < 0$. Therefore, $c1$ is an effective differential point, while $c2$ is not.

Let's evaluate the differential result of equation (9). For $c_o \in (0, \dots, c1)$, we have $\partial E[d]/\partial c_o < 0$, (8) is monotonous descendent; for $c_o \in (c1, \dots, c)$, we have $\partial E[d]/\partial c_o > 0$, (8) is monotonous ascendent. We can see that (8) is a concave function and that $c_o = c1$ is a minimum point. Thus, the optimal point can be reached when $\hat{c}_o = c1$ channels are allocated for on-demand and $c - \hat{c}_o$ channels are allocated for broadcast. That is, $\hat{c}_o = c\sqrt{mB}/(\sqrt{mB} + \sqrt{\mu'A})$.

Notice that sometimes the value we obtain via the above formula will make $\rho = (\lambda_{cold} + \lambda_{hot} * \epsilon)/(\hat{c}_o\mu') < 1$. That means if we allocate \hat{c}_o for on-demand channels, the on-demand channels will work under light load. Thus, (7) does not apply. We shall find out the optimal point of c_o , where the on-demand channels are working under heavy load. Then we can use the following light load formula to handle the light load case and compare which one is better, e.g. on-demand channels work in heavy load or light load. Since (7) is a concave function, the optimal point for on-demand channels working in heavy mode can be obtained as:

$$\begin{aligned} \bar{c}_o &= \text{Max}\{c_o | (\lambda_{cold} + \lambda_{hot} * \epsilon)/(c_o\mu') > 1 \\ &\text{and } 0 \leq c_o \leq \hat{c}_o\}. \end{aligned} \quad (10)$$

Light load

Next, we look at the situation when the system is under light workload ($\rho < 1$). Based on Equations (2), (3) and (5), we substitute $E[o]$, $E[b]$ and c_b in Equation (6) to obtain:

$$\begin{aligned} E[d] &= \frac{1}{\lambda} \left(\lambda_{hot}(1 - \epsilon) \left(\frac{ks}{2b(c - c_o)} + \frac{s}{b} \right) + \right. \\ &\quad \left. (\lambda_{hot}\epsilon + \lambda_{cold}) \left(\frac{1}{\mu'} + \frac{P_{c_o}}{c_o\mu'(1 - \rho)} \right) \right) \end{aligned} \quad (11)$$

where P_{c_o} is the probability that the mobile computers are requesting c_o or more data items from the on-demand channels.

We continue to use the constants A, B and C defined above and further define one more constant as follows.

$$D = \frac{1}{\lambda} * (\lambda_{hot} * \epsilon + \lambda_{cold}) * \frac{1}{\mu'}$$

Thus, Equation (11) becomes:

$$E[d] = \frac{A}{c - c_o} + \frac{BP_{c_o}}{c_o\mu'(1 - \rho)} + C + D$$

Without losing the upper bound of the access time $E[d]$, we assume P_{c_o} to be 1 in order to simplify the complexity of differential derivation in the next step. Thus, we obtain

$$E[d] \leq \frac{A}{c - c_o} + \frac{B}{c_o\mu'(1 - \rho)} + C + D$$

By differentiating the above upper bound to evaluate the optimal point for c_o , we have:

$$\frac{\partial E[d]}{\partial c_o} = \frac{\mu'(A\mu' - B)(c_o - c1)(c_o - c2)}{(c - c_o)^2(\mu'c_o - \lambda_{cold} - \lambda_{hot} * \epsilon)^2} \quad (12)$$

where,

$$\begin{aligned} c1 &= c \cdot (\rho\sqrt{\mu'A} + \sqrt{B})/(\sqrt{\mu'A} + \sqrt{B}) \\ c2 &= c \cdot (\rho\sqrt{\mu'A} - \sqrt{B})/(\sqrt{\mu'A} - \sqrt{B}) \\ \rho &= (\lambda_{cold} + \lambda_{hot} * \epsilon)/(\mu'c) \end{aligned}$$

As in the heavy-load case, $c_o = c1$ is a minimum point of the system access time. The optimal point is $\hat{c}_o = c(\rho\sqrt{\mu'A} + \sqrt{B})/(\sqrt{\mu'A} + \sqrt{B})$.

If $\hat{\rho}_o = (\lambda_{cold} + \lambda_{hot} * \epsilon)/(\mu'\hat{c}_o) > 1$, the minimum point where on-demand channels are working in light load mode can be recalculated as:

$$\begin{aligned} \bar{c}_o &= \text{Min}\{c_o | c \geq c_o \geq \hat{c}_o \\ &\text{and } (\lambda_{cold} + \lambda_{hot} * \epsilon)/(\mu'c_o) < 1\}. \end{aligned} \quad (13)$$

This gives us an approximate method for allocating broadcast channels and on-demand channels in a way that the access time is no more than $A/(c - \bar{c}_o) + B/(\bar{c}_o\mu' - \lambda_{cold} - \lambda_{hot} * \epsilon) + C + D$ for any data item.

Based on the formulae we derived above, we present the following algorithm to approximate the optimal channel allocation for optimizing the system performance.

Algorithm 1 Approx-Optimal Channel Allocation.

Given n data items, D_1, D_2, \dots, D_n , which have been sorted in descending order by data access frequency, and c wireless channels in a cell.

Decide the (approximate) optimal allocation of channels for broadcast and on-demand services and the set of data items to be disseminated on the broadcast channels.

begin

1. Initialize the allocation to be exclusive on-demand (i.e., $C_O = c, C_B = 0$, and the set of broadcast data (i.e., $I = 0$).
2. If $\rho > 1$, initialize the approximate lowest access time T using Equation (1) else initialize the approximate lowest access time T using Equation (2).
3. For $i = 1$ to n do
begin
 - Identify D_1, \dots, D_i as the data items to be broadcast.
 - Compute c_o^h using Equation (10) and c_o^l using Equation (13).
 - Compute the approximate lowest access time t^h using Equation (7) and t^l using Equation (11) based on c_o^h and c_o^l , respectively.
 - Let t be $\text{MIN}(t^h, t^l)$ and c_o be the corresponding assignment of on-demand channels.
 - Compute c_b using Equation (5).
 - Keep track of the optimal access time T , the corresponding allocation, C_O and C_B , and the corresponding number of broadcast data items, I .

end;

4. Return the corresponding I, C_O , and C_B values with the optimal access time T .

end.

5 Performance Evaluation

Our model consists of a single server, a set of clients and a fixed number of channels. The server can either continuously broadcast the most frequently accessed data items (we call them hot items) on broadcast channels or reply client requests on on-demand channels. We evaluate the performance of on-demand and broadcast channels using the formulae developed

in Sections 3.1 and 3.2, respectively. Moreover, we model the activity on the on-demand channels by a M/M/m queuing system. The simulation is implemented using C and CSIM [Sch92].

We make the following assumptions in the comparisons. The number of accesses to the data items is exponentially distributed. About 25% of the data access requests is on the top 50 most frequently accessed data items and 50% of the requests is on the top 174 most frequently accessed data items.

Unless otherwise specified, the parameter values used in the simulations and the analytical models are provided in Table 1. The primary performance criteria employed in this study is the data access time.

| | |
|----------------------|------------|
| No. of Clients | 500 |
| Data Item Size | 1000 bytes |
| No. of Data Items | 2000 |
| Access Info. Size | 10 bytes |
| No. of Channels | 50 |
| Channel Bandwidth | 1000 bps |
| Request Arrival Rate | 500/sec. |

Table 1: System Parameter Settings

5.1 Exclusive versus Dynamic Allocation

In this section, we compare the data access time of the exclusive on-demand and dynamic channel allocation under different system workloads. Under *heavy-loaded*, the number of clients is 1000 and the request arrival rate is 500/sec, while under *light-loaded* the number of clients is 100 and the request arrival rate is 45 per sec.

In the experiments, we move n hottest data items from the on-demand channels to the broadcast channels in order to illustrate the processing of our algorithm. Both simulations and the analytical formulae specified in the approx-optimal channel allocation algorithm are used to find the best channel assignment for on-demand and broadcast services.

Figures 1, and 2 illustrate the best access time corresponding to the number of channels assigned to on-demand service¹ and the number of hot data items on broadcast channels under heavy-load and light-load, respectively. Table 2 summarizes the access time and optimal channel allocations obtained in our experiments.

In the figures, 'exclusive on-demand' denotes the data access time of exclusive on-demand method and the 'analytical dynamic' and 'simulated dynamic' denote the data access time of dynamic channel allocation method obtained by the analytical formulae and

¹The rest of the channels are assigned to broadcast service.

| | Analytical Optimal | | Simulated Optimal | |
|--|--------------------|------------|-------------------|------------|
| | Heavy Load | Light Load | Heavy Load | Light Load |
| Optimal Allocation (on-demand/broadcast) | 11/39 | 43/7 | 11/39 | 41/9 |
| Number of Hot Data Items | 671 | 3 | 671 | 3 |
| Optimal Access Time | 9.489 | 1.049 | 9.470 | 1.066 |
| Access time for Exclusive On-Demand | 20.20 | 1.100 | 20.14 | 1.082 |

Table 2: Optimal Channels and Data Item Allocation

by simulation, respectively. The optimal channel allocation and the number of data items which should be disseminated through broadcast channels correspond to the point in the figures where the access time is lowest. Also, from the figures, we can easily observe that the access time obtained analytically is consistently close to the simulated result, meaning that it is feasible to utilize approx-optimal channel allocation algorithm to determine the best assignment of on-demand/broadcast channels and the number of data item to broadcast.

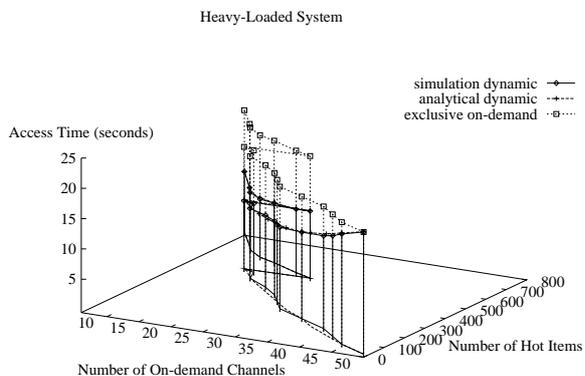


Figure 1: Optimal Access Time under Heavy Load

As demonstrated in Figure 1, the system access time of a heavy-loaded system may be significantly improved by moving some of the hot data items to the broadcast channels. The difference between the access time for ‘exclusive on-demand’ and ‘dynamic’ represents the improvement we can get when certain number of the hottest items are broadcast.

We can also observe from the experiments that the access time for dynamic channel allocation method decreases as the number of hot data items moving to the broadcast channels increases and the corresponding number of channels assigned to on-demand channels decreases. However, this trend changes when the number of hot items reaches 525. The number of channels assigned to on-demand increases again when the number of hot items increases to 525, while the access time continues to drop. From then on, the number of on-demand channels increases again to deal with the increased broadcast data items. This turning behavior

of channel assignments can be explained as follows. As the system moves more and more hot data items to the broadcast channels, the workload for on-demand channels will gradually become light-loaded. In a light-loaded system, the waiting time for on-demand channels drops greatly, the broadcast channels don’t have as significant an effect on alleviating on-demand workload as they do in a heavy-loaded system. Thus, the dynamic channel allocation method reassigns certain number of channels back to the on-demand service in order to improve the overall system performance. In our experiments, when around 525 data items are disseminated via broadcast channels, the workload of on-demand services become light.

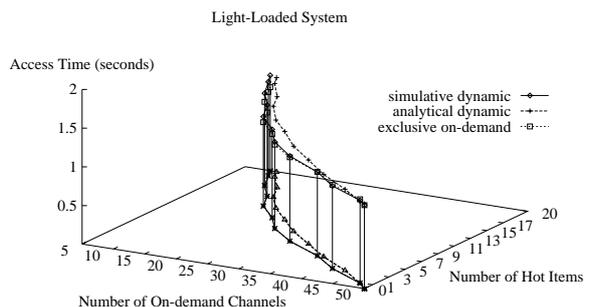


Figure 2: Optimal Access Time under Light Load

For the light-loaded system (Figure 2), the exclusive on-demand channel outperforms the most of the dynamic allocations which provides broadcast services. Even the optimal dynamic allocation has only 1.51% of improvement over the exclusive on-demand method. General speaking, compared with exclusive on-demand, the dynamic channel allocation method doesn’t improve data access time much for light-loaded systems and the improvement only happens when the probe time of the broadcast channels is shorter than the transmission time of the data request message for on-demand services. In our analysis and experiments, we have neglected the setup time for on-demand services. In a realistic environment, the dynamic channel allocation method may still significantly improve the performance of light-loaded systems.

5.2 Sensitivity Analysis

Several factors affect the data access performance of a mobile computing system, i.e., the sizes and the number of data items and messages, the frequency of data requests, the number and the bandwidth of channels, and the number of users. In this section, we present the results of a sensitivity analysis on the data access time using some of the basic factors.

We assume the size of the database is 500. This assumption is to facilitate the delivery of exclusive broadcast service, which is not feasible for serving the hardly accessed data items in the database.

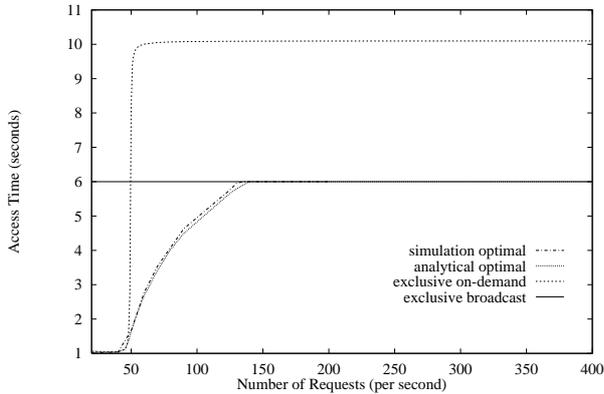


Figure 3: Access Time vs Request Arrival Rate

Figure 3 illustrates that, for exclusive on-demand method, when the request arrival rate is over 50 data items per second, the system is overloaded. When the request arrival rate is less than 50 data items per second, the access time for on-demand channels is close to the transmission time for a data item. Since we assume that the mobile computers don't attempt to issue new requests before their previous requests are completed, the access time for an overloaded system is constantly close to 10.01 seconds. On the other hand, the broadcast channels have the advantage of not being affected by the frequency of data access. No matter how frequent the broadcast data items are requested, the access time for broadcast channels remains constant. The dynamic channel allocation method wisely chooses exclusive on-demand when the system workload is light (i.e., less than 50 requests per second) and switch to exclusive broadcast method when the system workload is really heavy (i.e., greater than 150 requests per second). When the request arrival rate is between 50 and 150 per second, its performance is better than the exclusive on-demand and exclusive broadcast methods. Note that if we vary the number of mobile computers, we obtain similar results, since the number of mobile computers has a direct impact

on the request arrival rate².

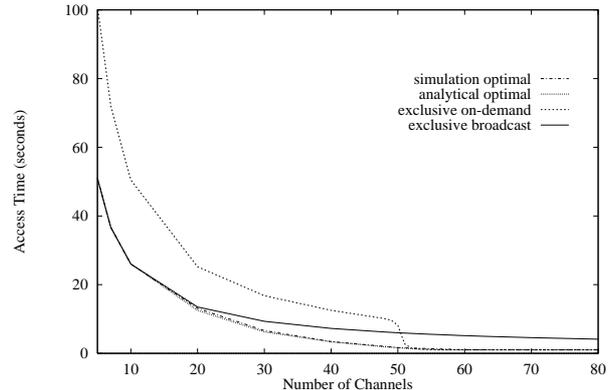


Figure 4: Access Time vs Number of Channels

Next, we vary the number of channels available in the system to observe the change in access time for the channel allocation methods. Figure 4 demonstrates that the access time of broadcast channels drops rapidly when the number of broadcast channels increases from 1 to 20. From then on, however, increasing the number of broadcast channels has diminishing impact on the broadcast channel access time. This observation suggests not to invest too many broadcast channels in order to lower the access time on broadcast channels. For the on-demand services, the number of channels used has a dramatical impact on the access time performance. In other words, with a small increase of the channels at the dividing point between a heavy-loaded system and a light-loaded system, the system access time may improve significantly. As expected, the dynamic channel allocation method has the best access time.

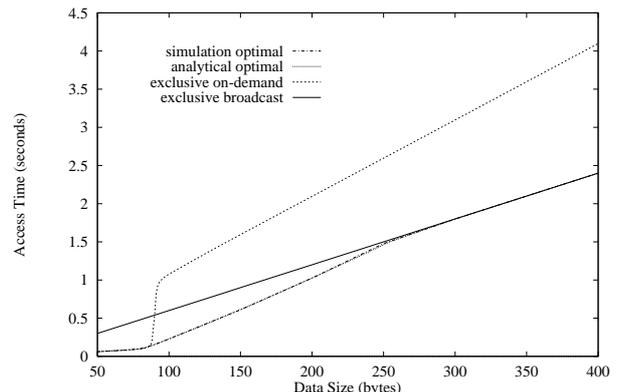


Figure 5: Access Time vs Data Size

In Figure 5, we increase the size of data items from 50 bytes to 400 bytes, while fixing the other parameters. The access time of on-demand service shows that the system is rapidly loaded after the data size

² Assuming the request rate for each computer is fixed.

is greater than 90 bytes. The leap on the access time of on-demand service is due to the increase of system workload caused by the increase of waiting time for mobile computers to connect with the MSS. On the other hand, the access time of broadcast services is proportional to the size of the data items broadcast. With 50 channels, the access time of broadcast method is worse than that of the on-demand method when the data access load is light, i.e., when data item size is less than 90 bytes, and is better than that of the on-demand method when the load is heavy. Dynamic channels allocation method is realized by broadcasting some of the hottest data items on broadcast channels. Thus, the broadcast channels relief a significant part of the system workload off on-demand channels and improve overall system performance.

In all of the above experiments, the performance of on-demand channels has certain turning point due to the change of system workload. The turning point corresponds to the system traffic intensity, $\rho = 1$. The factors varied in the experiments have an impact on the traffic intensity ρ and thus causing a dramatic change of the data access time. On the other hand, the performance of broadcast channels has no such abrupt change.

In the studies, we have deliberately set the database size to be small, i.e., 500 data items, so that it is feasible to disseminate the entirely database on broadcast channels. In a realistic environment, the database usually is quite large, exclusive broadcast mode may never be adopted. Broadcast is useful in the dynamic channel allocation method as a means to offload frequently accessed data from the on-demand channels.

6 Conclusion

Wireless broadcast is an important technique for disseminating data to mobile users, because it scales up to an arbitrary number of users. In this paper, we study systems where broadcast and on-demand channels are used to complement each other to maximize system performance. We first developed analytical models for broadcast and on-demand channels. Cost formulae for the access time of the exclusive broadcast, exclusive on-demand, and dynamic channel allocation methods are developed based on the analytical models. Based on the models, we develop a method to approximate the optimal channel allocation. Simulation studies are performed to validate the analytical models. The result shows that the dynamic channel allocation method improves the system performance significantly when the system is heavy-loaded, thus showing the importance of dynamic channel allocation in the efficient utilization of wireless bandwidth.

As for future research, we will study the global channels management issues for neighboring cells, in which channel borrowing is allowed. Moreover, use of the broadcast channels as shared cache for the mobile users is an interesting idea. A coherent approach for the caching, invalidation and maintenance of data items through a combination of the broadcast and on-demand channels needs further research.

References

- [AFZ97] S. Acharya, M. Franklin, and S. Zdonik. Balancing push and pull for data broadcast. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 183–194, Tuscon, Arizona, May 1997.
- [HGLW87] G. Hermen, G. Gopal, K. Lee, and A. Weinrib. Datacycle architecture for very high throughput database systems. In *Proceedings of the 1987 ACM SIGMOD*, pages 97–103, May 1987.
- [IV94] T. Imielinski and S. Viswanathan. Adaptive wireless information systems. In *proceedings of SIGDBS (Special Interest Group in DataBase Systems) Conference*, Tokyo - Japan, October 1994.
- [Jai91] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, New York, 1991.
- [LHL97] W.-C. Lee, Q.L. Hu, and D. L. Lee. Channel allocation methods for data dissemination in mobile computing environments. In *Proceedings of the 6th IEEE International Symposium on High Performance Distributed Computing*, pages 274–281, August 1997.
- [LHL98] W.-C. Lee, Q.L. Hu, and D. L. Lee. A study of channel allocation methods for data dissemination in mobile computing environments. *To appear in MONET: Special Issue on Resource Management in Wireless Networks*, 1998.
- [Sch92] H. Schwetman. *Csim user's guide (version 17)*. MCC Corporation, 1992.
- [SRB97] K. Stathatos, N. Roussopoulos, and J. S. Baras. Adaptive data broadcast in hybrid networks. In *Proceedings of the 23rd VLDB Conference*, pages 326–335, Athens, Greece, 1997.