Optimizing Query Access Time over Broadcast Channel in a Mobile Computing Environment

Agustinus Borgy Waluyo[1], Bala Srinivasan[1], and David Taniar[2]

[1] School of Computer Science and Software Engineering, Monash University, Australia
{Agustinus.Borgy.Waluyo,Bala.Srinivasan}@infotech.monash.edu.au
[2] School of Business Systems, Monash University, Australia
David.Taniar@infotech.monash.edu.au

Abstract. Broadcast strategy is known as an effective mechanism to disseminate database information to mobile clients. However, with a very large database items, the access time of mobile client increases accordingly due to high waiting time to find their data of interest. In this paper, we propose analytical models for both query access times over broadcast channel and on-demand channel. We examine the analytical model to find optimum number of broadcast items in a channel while utilising query access time over on-demand channel as a threshold point. The optimum number indicates a point to split the broadcast cycle and allocate the data items in the new channel. Several factors are taken into account, which includes request arrival rate, service rate, number of request, size of data item, size of request, number of data item to retrieve, and bandwidth. A simulation model is developed to find out the performance of the analytical model, and it is found the model is close to accurate.

1 Introduction

Broadcast strategy refers to periodically broadcasting database items to mobile clients through one or more broadcast channels. Mobile clients filter their desired data on the fly. This strategy is known as an effective way to disseminate database information to a large set of mobile clients. In this way, mobile client enables to retrieve information without wasting power to transmit a request to the server. Other characteristics of data broadcasting includes scalability as it supports a large number of queries, query performance is not affected by the number of users in a cell as well as the request rate, and effective to a high-degree of overlap in user's request.

With the increase number of data items to be broadcast, the *access time*[1] also increase accordingly as access to data item is sequential. This situation causes some mobile clients to wait for a substantial amount of time before receiving desired data item. Consequently, the advantages of broadcast strategy will be diminished.

Alternatively, mobile client can send a request via point-to-point channel or on-demand channel to the server. The server processes the query and sends the result back to the client. However, this situation severely affects the query response time

[1] Access time is defined as elapsed time from the time a request is initiated until all data item of interest are received.

L.T. Yang et al. (Eds.): EUC 2004, LNCS 3207, pp. 439–449, 2004.

and power consumption of mobile clients, as it involves queuing and cannot scale over the capacity of the server or network. Thus, in this paper we try to *outperform* the query access time over on-demand channel in any situation by optimizing the utilization of broadcast channel.

Figure 1 illustrates two mechanisms of mobile clients to obtain desired data. Figure 1 (a) is utilizing on-demand channel, while Figure 1 (b) is through broadcast channel. These two mechanisms are also referred as pull-based and push-based approach respectively.

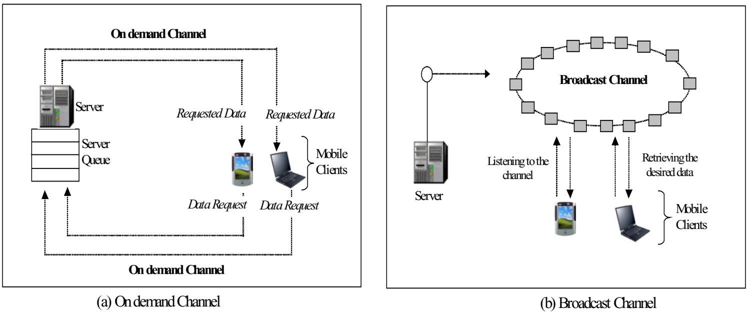


Fig. 1. The architecture of on-demand channel and broadcast channel

To meet our objective, we propose analytical models for both on-demand and broadcast channel and they are considered as independent model. Subsequently, we use simulation model to verify the results of our analytical models in finding the optimum number of broadcast items in a channel. This paper focuses on request that returns multiple data items that is an extension of our previous work, which relates to a single data item [9]. The term 'data item' corresponds to database record or tuples, and data segment contains a set of data item. A complete broadcast file is referred as a broadcast cycle [8].

The rests of the section in this paper are organized as follows. Section 2 describes the background of mobile databases; section 3 contains the related work of the proposed technique. It is then followed by the description of the optimal broadcast channel, analytical models for both broadcast and on-demand channel, and its application in section 4. Section 5 introduces parameters of concern, performance results as well as the performance analysis. Finally, section 6 concludes the paper.

2 Background

In general, each mobile user communicates with a Mobile Base Station (MBS) or also known as Mobile Support Station (MSS) to carry out any activities such as transaction and information retrieval. Activities that focus on query to central database server referred as *mobile databases.* MBS has a wireless interface to establish communication with mobile client and it serves a large number of mobile users in a specific region called cell. Mobile units or mobile clients in each cell can either connect to the network via wireless radio or wireless Local Area Network (LAN).

Wireless radio bandwidth has asymmetric communication behavior in which the bandwidth for uplink communication is smaller than downlink communication [4]. In mobile environment architecture, each MBS is connected to a fixed network.

3 Related Work

Strategies to broadcast data through wireless channel have been introduced in the past few years [1, 2, 3 5, 8]. Broadcast strategy is known to be effective when involving a large number of users in the cell, which corresponds to a high frequency of requests.

Hybrid and dynamic channel allocation method, which manipulates both the on-demand channels and broadcast channels, has been presented in [2, 3]. However, in this scheme, a vigorous cost calculation introduces a high degree of complexity in the algorithm before it can decide the best channel allocation. Furthermore, the utilization of on-demand channel forces the client to consume large amount of energy especially after considering the asymmetric communication cost as indicated earlier. Thus, it is desirable to have a better strategy in utilizing broadcast channel.

Another strategy has been proposed by [5], in which the data are broadcast based on replication and partition technique through multiple channels. The data items are replicated and organized in a way that reflects the distribution of the data. Thus each channel can have different distribution pattern. This technique is made to overcome problems such as data distortion and data distribution. However, the number of broadcast channels used in that case was static. Consequently, when there are very large numbers of database items, the length of the broadcast cycle increases. This situation causes some mobile clients to wait for a substantial amount of time before receiving desired data item even with the help of different pattern of data distribution over multiple channels.

4 Optimal Broadcast Channel: Proposed Method

In general, the data items are broadcast over a single channel with underlying reason that it provides the same capacity (bit rate/bandwidth) as multiple channels, and it is used to avoid problems such data broadcast organization and allocation while having more than one channel [4]. Nevertheless, the use of single channel will show its limitation when there is a very large number of a data item to be broadcast. However, it is also not appropriate to simply utilize all available channels to broadcast a set of data items as it will cause a waste of bandwidth and resources.

The proposed strategy is used to split the length of the broadcast cycle when the number of broadcast items reaches an optimum point. This process continues until the access time is above the optimal point. To determine when and which channel the data item of interest will be broadcast, we assume to utilize Global Index, a multi channels indexing technique that we have proposed in [10].

Figure 2 illustrates a situation when the optimum number of items in a channel is known. A simple example, assume the size of the each data items are: 50 bytes. Consider a mobile client who wants to retrieve data item #8. The following two cases analyze the access time of mobile client when the data is broadcast over a single channel as compared to two channels. It is assumed that the client probes into a first data item in a channel. *Case 1*: Data items (1-10) are broadcast in a single channel. With this case, all data items are available in a single channel. Client who wants to retrieve data item #8, has to wait until the desired items are arrived. The total access time required will be (50 × 8) = 400 bytes. *Case 2:* Data items (1-10) are split into two broadcast channels: channel one and channel two with 6 and 4 data items respectively. Since data items are split into two broadcast channel, the access time will also be different. In this case, client can switch to another channel and wait for the data items of interest to arrive. The total access time for client to retrieve data item #8 would be (50 × 2) = 100 bytes. Thus, as compared to the first case, this time client can have a much more efficient data access.

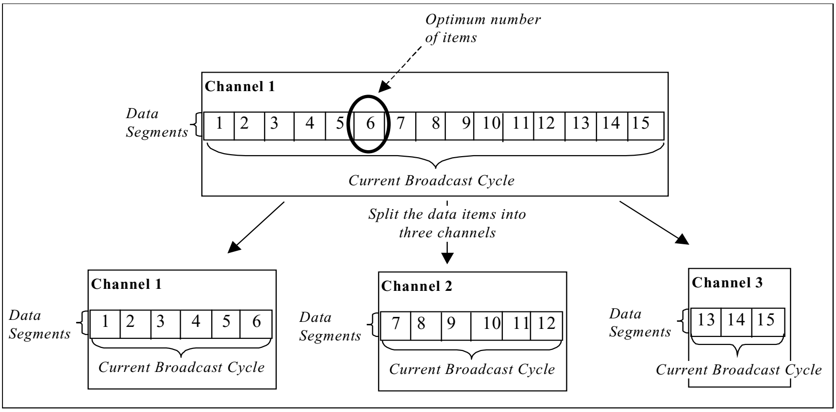


Fig. 2. Optimal Broadcast Channel

To find the optimum number of broadcast items, we develop analytical models to calculate average access time of broadcast and on-demand channel. In this context, a mobile client can only initiate a single request at a time; the next request has to wait until the first request has been completely responded.

4.1 Multiple Data Item Request

This category involves a request that returns multiple data items. However, the order of the data retrieval is made arbitrary, which means first arrival of any data item of interest will be firstly retrieved. We left the data ordering retrieval for future work. Let $d_1, d_2 \ldots d_{max}$ specifies the number of relevant items for a request. We classify the number of relevant data items into uniform request that is when the occurrence rate of $d_1, d_2 \ldots d_{max}$ is equally distributed and non-uniform request, when the occurrence rate is otherwise. The non-uniform request is indicated by the corresponding frequency of occurrence *f,* for each *d,* $\sum f = f_1 + f_2 + \cdots + f_{d\max} = 1$. We define the frequency of request for each *d*, and then we calculate its average $\overline{d}$. Similarly, for server service rate μ, we determine the service rate for each request, and find the average service rate $\overline{\mu}$. The same process also applies to the size of request *r*. The formulas in table 1 are employed to find $\overline{d}$, $\overline{r}$, and $\overline{\mu}$ for uniform and non-uniform request.

Table 1. Uniform and Non-Uniform Formulas

Uniform Request	***Non Uniform Request***
$\overline{d} = \frac{\sum_{i=1}^{d\max} i}{d\max}$	$\overline{d} = \frac{\sum_{i=1}^{d\max} i \times f_i}{d\max}$
$\overline{r} = \frac{\sum_{i=1}^{d\max} \frac{r_{i\min} + r_{i\max}}{2}}{d\max}$	$\overline{r} = \frac{\sum_{i=`}^{d\max} \left(\frac{r_{i\min} + r_{i\max}}{2} \right) \times f_i}{d\max}$
$\overline{\mu} = \frac{\sum_{i=1}^{d\max} \mu_i}{d\max}$	$\overline{\mu} = \frac{\sum_{i=1}^{d\max} \mu_i \times f_i}{d\max}$

4.2 Analytical Model of Broadcast Channel

To calculate the average access time for retrieving data (T_B) using broadcast channel, we consider the following scenario:

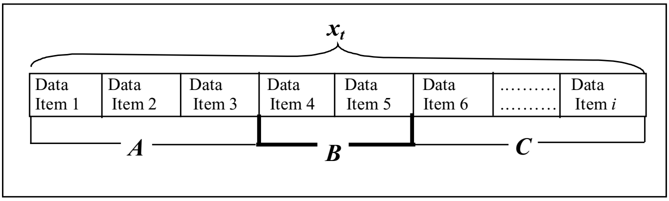


Fig. 3. Broadcast Cycle Partitioned Area

In Figure 3, the broadcast cycle is partitioned into three areas: area *A* contains data items preceding data items in area *B*, area *B* contains the number of desired data items, and area *C* includes the rest of data items in the cycle. x_t is the total number of broadcast items that makes up of *A*+*B*+*C*. There are three different scenarios when mobile client probe into one of these area. *s* corresponds to the size of data item, we consider the size is uniform, and the downlink bandwidth is denoted by b_s.

Probe A: When mobile client probes into area *A*, the total access time is given:

$$\frac{\sum_{i=0}^{A-1} (A - i + B) \times s}{b_s} \text{ or similarly } \frac{[A \times (A + 2B + 1)] \times s}{2 \times b_s} \qquad (1)$$

Probe B: When mobile client probes into area B, then the total access time is equal to the total length of broadcast cycle (A+B+C): $\frac{(A+B+C)\times B\times s}{b_s}$ (2)

Probe C: When mobile client probes into area C, the total access time can be calculated from: $\frac{\sum_{i=0}^{C-1}(C-i+A+B)\times s}{b_s}$ or equally $\frac{C\times(2A+2B+C+1)\times s}{2\times b_s}$ (3)

We need to calculate the average access time from equation (1), (2), and (3) as follows:

$$\frac{[(A\times(A+2B+1))+(2B\times(A+B+C))+(C\times(2A+2B+C+1))]\times s}{2\times x_t\times b_s} \quad (4)$$

If $\bar{d}$ substitutes for B and x_t for A+B+C, then equation (4) can be rewritten as below:

$$T_B \approx \frac{[(x_t-\bar{d})^2+(2\times\bar{d}\times x_t)+(2\times\bar{d}\times(x_t-\bar{d}))+(x_t-\bar{d})]\times s}{2\times x_t\times b_s} \quad (5)$$

4.3 Analytical Model of On-Demand Channel

As mentioned earlier, mobile clients can send their request to be processed via point-to-point or on-demand channel. Thus, the average access time of on-demand channel (T_D) is made for comparison. A comparison is made to locate the optimal point where the average access time of broadcast channel is equal or less than on-demand channel. To calculate the average access time of on-demand channel, we classify the analytical model into the following categories:

a. Pre-determined Number of Request

In this scenario, the total number of request has been known. The average access time can be calculated using the following analytical model:

$$T_D \approx \frac{\bar{r}}{b_r}+\frac{\sum_{i=1}^{n} i\times\frac{1}{\bar{\mu}}}{n}+\frac{\bar{d}\times s}{b_s} \quad (6)$$

The analytical model to calculate the access time of on-demand channel in equation (6) is based on the architecture that is depicted in Figure 1(a). The access time of on-demand channel comprises of the time needed to transmit a request to the server, process the request that involves the time spent in the server queue and retrieve the relevant data, and send the result back to the client. The transmission of request to the server is affected by uplink bandwidth, b_r. n denotes the total number of request, and i reads the request number. However, in real situation the number of requests is hardly known. Thus, we consider arrival rate of request, which is described in the subsequent strategy.

b. Request Arrival Rate
We replace the determined number of request with the arrival rate of request ( ). Thus, the average access time of on demand channel (T_D) is now calculated using the following analytical model:

If the server is at light load $\rho = \left(\frac{\lambda}{\mu} \le 1\right)$, which means there is no waiting time then:

$$T_D \approx \left(\frac{\bar{r}}{b_r} + \frac{1}{\bar{\mu}} + \frac{\bar{d} \times s}{b_s}\right) \tag{7}$$

Else if the server is at heavy loads $\rho = \left(\frac{\lambda}{\mu} > 1\right)$, then:

$$T_D \approx \left(\frac{\bar{r}}{b_r}\right) + \frac{\sum_{i=1}^{q\max} \left(i \times \frac{1}{\bar{\mu}}\right) - \left[(i-1) \times \frac{1}{\lambda}\right]}{q\max} + \left(\frac{\bar{d} \times s}{b_s}\right) \tag{8}$$

The average time in server queue and average processing time by the server are dynamically changed depending on the arrival rate of request ( ) and service rate. The maximum number of request in the server queue is defined by $q\max$.

5 Performance Evaluation

In this section, we introduce three cases, and employ the analytical model to determine the optimum number of broadcast data items. We develop some simulation models to verify the results of our analytical model. The simulation is carried out using a simulation package *Planimate, a*nimated planning platforms [7].

5.1 Simulation Setup

The simulation environment is set to apply exponential distribution for both arrival rate and server service time with given average value. In the simulation model we incorporate three facilities as multi client environment, we run the simulation process for five iterations and derive the average result accordingly. In each case, we consider request that return one or more number of items. Subsequently, we classify the request into two scenarios, one is uniform request and the other is non-uniform request. Uniform request is when the frequency of occurrence for all type of request is equal. Whilst, non-uniform request is when the frequency of occurrence varies. Table 2 and 3 contains a set of parameter used in the simulation. The bandwidth values are determined according to common transmission rates for packet-switching in GSM cellular network [6].

Table 2. Parameters of Concern

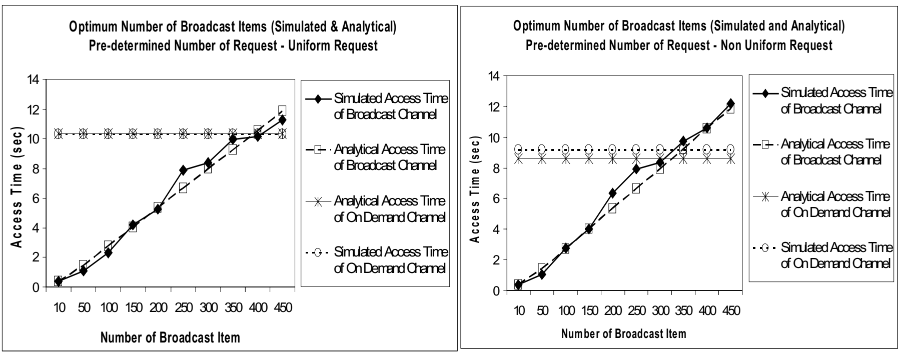
Parameter	Description	Initial Value
Broadcast channel		
x_t	Total number of Broadcast Items	450
On demand channel		
n	Number of request (pre-determined number)	50
b_r	Uplink Bandwidth	9600 bytes
Broadcast and On demand channel		
b_s	Downlink Bandwidth	19200 bytes
s	Size of each data item	1000 bytes

Table 3. Set of Parameters for Uniform and Non-Uniform Request

Number of Returned Data items (*d*)	1	2	3	4
Non-Uniform Request				
Frequency of Occurrence (*f*)	0.4	0.3	0.2	0.1
Uniform Request				
Frequency of Occurrence (*f*)	0.25	0.25	0.25	0.25
Non-Uniform and Uniform Request				
Size of Request (r)	50-100 bytes	101-150 bytes	151-200 bytes	201-250 bytes
Service Rate (μ)	4 request per sec	3 request per sec	2 request per sec	1 request per sec

5.2 Performance Results

Case 1: To find the optimum number of broadcast items with pre-determined number of request (uniform and non-uniform request).

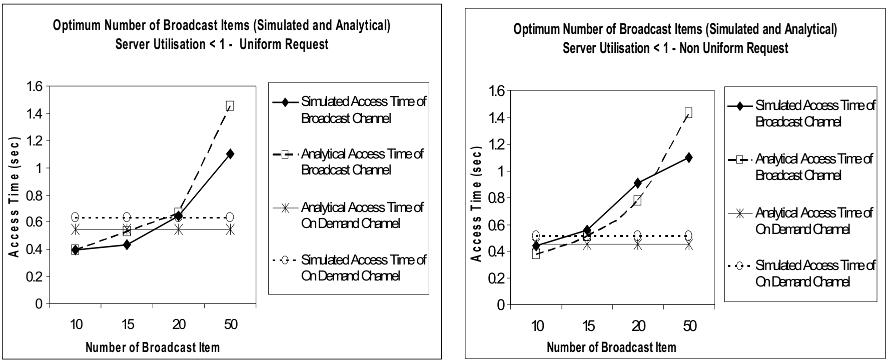


(a) Pre-determined Number of Request with Uniform Request (b) Pre-determined Number of Request with Non-Uniform Request

Fig. 4. Optimum Number of Broadcast Items with Pre-determined Number of Request (Uniform and Non-Uniform request)

This case is categorized into pre-determined number of requests with uniform and non-uniform request. As shown in Figure 4, we try to find the cross line of on-demand channel and broadcast channel. The average access time of on-demand channel is the threshold point, which appears in a straight line since the number of broadcast items does not affect its value. The intersection point indicates the optimum number of broadcast items.

Case 2: Introducing a new parameter, arrival rate of request (λ). We specify λ = 1 per two seconds, to obtain server utilization $\rho < 1$. As shown in Figure 5, the on-demand channel performs well when the traffic request is low. The broadcast channel seems to have to allocate its data items in every 10-15 data items to a new channel to keep up with the performance of on demand channel. However, this is not a good sign since the number of channels may be excessive and the bandwidth may be well wasted. Thus, on-demand channel in this case is a better way. Figure 6 denotes optimum number of items when the server is at heavy load.

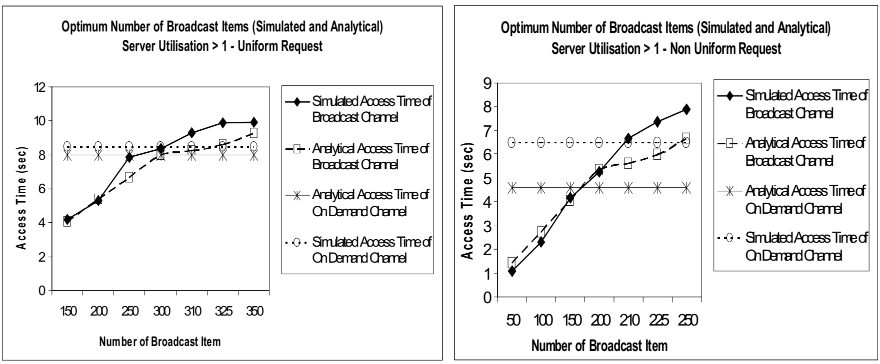


(a) Server Utilisation < 1 with Uniform Request

(b) Server Utilisation < 1 with Non Uniform Request

Fig. 5. Optimum Number of Broadcast Items with Server Utilisation $\rho < 1$ (Uniform and Non-Uniform Request)

Case 3: In this case, we specify request arrival rate, λ = 4 per second to obtain server utilization $\rho > 1$.



(a) Server Utilisation >1 with Uniform Request

(b) Server Utilisation >1 with Non Uniform Request

Fig. 6. Optimum Number of Broadcast Items with Server Utilisation $\rho > 1$ (Uniform and Non-Uniform Request)

5.3 Performance Analysis

Table 4 shows the optimum number of broadcast items derived from our analytical and simulation results on the three cases. Our analytical calculation differs about 5 percent in average from the simulation result. Thus, the analytical models are considered very close to accurate. Having known the optimum number of items in a broadcast channel, we can decide how many channels are required to broadcast a certain amount of data items. Consequently, the average access time of broadcast channel is minimized.

Table 4. Simulated and Analytical Performance

Optimum Number of Broadcast Items						
		Uniform/Non Uniform	**Analytical**	**Simulated**	**Difference**	**Error Rate**
Case 1	Pre-determined Number of Request	Uniform	391	400	9	2.3%
		Non Uniform	325	311	14	4.5%
Case 2	Server Utilization <1	Uniform	15	16	1	6.2%
		Non Uniform	12	12	0	0%
Case 3	Server Utilization > 1	Uniform	300	305	5	1.6%
		Non Uniform	170	205	35	17.1%
Average Error Rate: 5.28%						

6 Conclusions and Future Work

We have introduced analytical models for both broadcast and on-demand channel to find the average query access time over the channel. To locate the optimum value of broadcast items, we utilize the access time of on-demand channel as a threshold point. We conduct a simulation model to see the performance of our analytical model in determining the optimum broadcast items. The optimum number is used as an indication of when the broadcast cycle needs to be split into different channel, forming multiple channels. It is found that our analytical model is considerably very close to accurate as compared to the simulation results. For future work, we will vary the size of broadcast data items, and ordered the items in a way to meet majority requests.

References

1. Acharya S., Alonso R., Franklin M. and Zdonik S., "Broadcast Disks: Data Management for Asymmetric Communication Environments", In *Proc. of ACM Sigmod,* pp. 199-210, 1995.
2. Hu Q., Lee D. L. and Lee W.C., "Optimal Channel Allocation for Data Dissemination in Mobile Computing Environments", In *Proc. of 18th ICDCS*, pp. 480-487, 1998.
3. Hu Q., Lee W.C. and Lee D. L., "Indexing Techniques for Wireless Data Broadcast under Data Clustering and Scheduling", In *Proc. of the 8th CIKM***,** pp. 351-358, 1999.

4. Imielinski T., Viswanathan S. and Badrinath B. R., "Data on Air: Organisation and Access", *IEEE TKDE,* **9**(3):353-371, 1997.
5. Leong H. V. and Si A., "Data Broadcasting Strategies over Multiple Unreliable Wireless Channels", In *Proc.of the 4th CIKM***,** pp. 96-104, 1995.
6. Pitoura E., and Samaras G., *Data Management for Mobile Computing*, Kluwer Academic Publishers, London, 1998.
7. Seeley D. et al, *Planimate™-Animated Planning Platforms*, InterDynamics Pty Ltd, 1997.
8. Si A. and Leong H. V., "Query Optimization for Broadcast Database", *DKE,* **29**(3):.351-380, 1999.
9. Waluyo, A.B., Srinivasan B., Taniar D., "Optimal Broadcast Channel for Data Dissemination in Mobile Database Environment", In *Proc. of the 5th APPT*, pp. 655-664, 2003a.
10. Waluyo, A.B., Srinivasan B., Taniar D., "Global Index for Multi Channels Data Dissemination in Mobile Databases", In *Proc. of the ISCIS,* pp. 210-217, 2003b.