A Taxonomy of Broadcast Indexing Schemes for Multi Channel Data Dissemination in Mobile Databases

Agustinus Borgy Waluyo[1] Bala Srinivasan[1] 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

Data broadcasting strategy is known as a scalable way to disseminate information to mobile users. However, with a very large set of broadcast items, the query access time of mobile clients raise accordingly, due to high waiting time for mobile clients to find their data of interest. One possible solution is to split the database information into several broadcast channels. In this paper, we introduce taxonomy of index dissemination for multi broadcast channel based on B^+-tree structure. We consider three indexing schemes namely: (i) Non-replicated indexing scheme (NRI), (ii) Partially-replicated indexing scheme (PRI), and (iii) Fully-replicated indexing scheme (FRI). Simulation model is developed to find out the access time performance of each scheme.

1. Introduction

The emergence of wireless technology in recent years has helped community to work in a much more convenient, flexible, and efficient way. People are no longer attached to a stationary machine to do their work, with wireless application they are enabled to conduct their business anywhere and anytime using portable size wireless computer powered by battery (e.g. PDAs). These portable computers communicate with central stationary server via wireless channel. This technology is known as *mobile computing* [1,2,4]. A subset of mobile computing that focus on query to central database server refers to *mobile databases*. Mobile database face a number of limitation particularly power, storage and bandwidth capacity. Moreover, life expectancy of a battery was ever anticipated to increase only 20% for a decade [9]. Therefore, query optimization and processing in mobile database is important.

Data dissemination strategy or known as broadcast strategy is an effective way to disseminate database information to a large set of mobile clients. This strategy refers to periodically broadcast database items to mobile clients through one or more broadcast channels and they filter their desired data on the fly. With broadcast strategy, the number of mobile clients, their frequency and arrival rate of request do not affect the query access time. The use of broadcast indexing in mobile database is to minimize the client's tuning time or the time listening to the channel, which indicates power consumption. Broadcast indexing enables mobile clients to determine when the desired data arrives in the broadcast cycle. Thus, it allows them to switch into "*doze*" mode while waiting to the data and to safe power consumption. On the other hand, it is also important to minimize the index access time to retrieve the right index.

As the number of broadcast data increases over time, there is a need to disseminate the index directory via multi wireless channel. The number of broadcast channels in a wireless region that can be simultaneously utilized is up to 200 channels [7]. In our previous work [12], we introduce a Partially-replicated indexing (PRI) scheme and compared the performance with the existing scheme used in [7]. In this paper, we introduce three indexing schemes for multi broadcast channel namely: Non-replicated indexing scheme (NRI), Partially-replicated scheme (FRI), and Fully-replicated scheme (FRI). Each of these schemes will be analyzed using a simulation model. This taxonomy is an extension of our previous work in the context of parallel database systems [10]. The subsequent section of this paper is organized as

Proceedings of the 18th International Conference on Advanced Information Networking and Application (AINA'04)
0-7695-2051-0/04 $ 20.00 © 2004 IEEE

follows: section 2 describes the background of mobile database and indexing, section 3 contains the related work of the indexing technique. It is then followed by broadcast indexing schemes, and its application in section 4. The simulation model to measure the performance of each of the scheme is given in section 5. Finally, section 6 concludes the paper.

2. Background

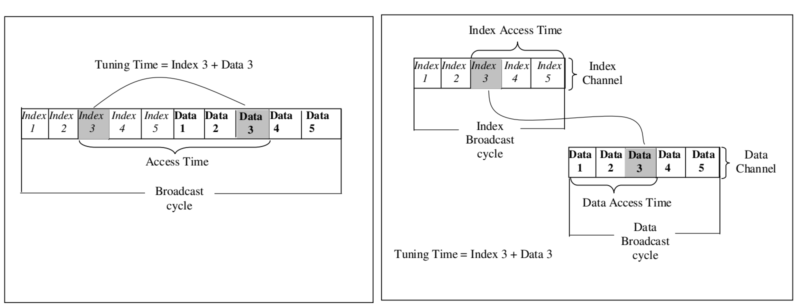
In general, each mobile user communicates with a Mobile Base Station (MBS) to carry out any activities such as transaction and information retrieval. MBS has a wireless interface to establish communication with mobile client and it serves a large number of mobile users in a specific region called *cell*. 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 asymmetry characteristic, which is uplink bandwidth (client to server) is smaller than downlink bandwidth (server to client) [5].

Broadcast indexing causes a trade-off between optimizing the client tuning time and the query access time. The consequence of minimizing one of them is the increase of the other. For instance, to minimize the access time is to reduce the length of broadcast cycles. In this case, the index can be broadcast once in each cycle but it will make the tuning time suffer since the client will have to wait for the index to arrive which happens only once in each broadcast cycle. On the other hand, when the index directory is frequently broadcast in each broadcast cycle to reduce the tuning time, the access time will be greatly affected due to the occupancy of the index in the cycle. Thus, it is necessary to find the optimal balance of these two factors. Figure 1 illustrates integrated and separated index with data channel. This paper concerns with index dissemination over index channel as depicted in Figure 1 (b).

3. Related Work

[6] proposed a signature technique, which is derived by hashing the attribute values into bit strings followed by combining them together to form a bit vector or signature. The signature is broadcast together with the data on every broadcast cycle. The signature mechanism also applies to the query that is initiated by the client. To process the query, mobile clients need to tune into the broadcast channel and verify the query signature with the data signature by performing a certain mathematical operation like 'AND' operation. If the signature is not matched, client can tune to the "doze" mode while waiting for the next signature to arrive. This technique is designed to interleave with data items in a single channel. However, in this technique mobile clients must check the every consecutive signature in order to find the right data item. Consequently, it consumes a substantial amount of power.

Another indexing technique is proposed in [7]. This technique is based on B^+-tree structure, and it incorporates a separate index channel to locate the data segment in multi data channel. The physical pointer not only indicates the time value but also the channel in which the indexed item will be broadcast. Using this mechanism, client first tunes into the index channel, find the indexed time value, switch to data channel and wait for the data item to arrive. The advantage is mobile client only needs to wait for the index broadcast cycle when it misses the desired index, which is considerably short as compared to the one that interleaves with data items. However, the number of broadcast channels used in this case is static. When there is large number of broadcast items, this situation may cause some mobile clients to wait for a substantial amount of time before receiving the desired data item even with the help of different pattern of data distribution over multiple channels



(a) Index and data in the same channel (b) Index and data in different channel

Figure 1. Integrated and Separated Index Channel

4. Broadcast Indexing Schemes over Multi Channel

There are three broadcast indexing schemes that we consider, namely

(i) Non-Replicated Index (NRI)
(ii) Partially-Replicated Index (PRI)
(iii) Fully-Replicated Index

For each scheme, it is assumed that the number of channel required to broadcast a certain amount of data items is known. 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 [5]. Nevertheless, the use of single channel will show its limitation when there is a large number of a data item to be broadcast. [11] proposed a strategy 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. Since the length of the broadcast cycle is optimal, the waiting time will be considerably short.

The index structure is designed based on B^+-tree structure. It consists of non-leaf nodes, and leaf node. Leaf node is the bottom most index that consists of up to *k* keys, where each key point to actual data items, and each node has one node pointer to a right-side neighbouring leaf node. Unlike leaf node, non-leaf node may consist of up to *k* keys and *k*+1 pointers to the nodes on the next level on the tree hierarchy (i.e. child nodes). All child nodes, which are on the left-hand side of the parent node, have the key values less than or equal to the key of their parent node. On the other hand, keys of child nodes on the right-hand side of the parent node are greater than the key of their parent node.

Having all data pointers stored on the leaf nodes is considered better than storing data pointers in the non-leaf nodes like the original *B* trees [3]. Moreover, by having node pointers in the leaf level, it becomes possible to trace all leaf nodes from the left most to the right most nodes producing a sorted list of keys. When being broadcast, each physical pointer to the neighbouring leaf node as well as actual data item are replaced by a time value, which indicates when the leaf node or data item will be broadcast.

Assume that in the index tree, the maximum number of node pointers from any non-leaf node is 4, and the maximum number of data pointers from any leaf node is 3. In this example, we use the ID attribute as the index-partitioning attribute, which is different from the table partitioning. The range partitioning rules that we used are that index channel 1 holds data IDs between 1 to 30, index channel 2 holds data IDs between 31 to 60, and the rest go to index channel 3. We must stress that the location of each leaf node is the not same as where the actual data is broadcast.

4.1. Non-replicated indexing (NRI) scheme

A *Non-Replicated Indexing* (NRI) scheme is where the index structure is partitioned into a number of disjoint and smaller indices. Each of these small indices is placed in a separate index channel.

After the table is partitioned, each relevant index channel builds its local index accordingly.

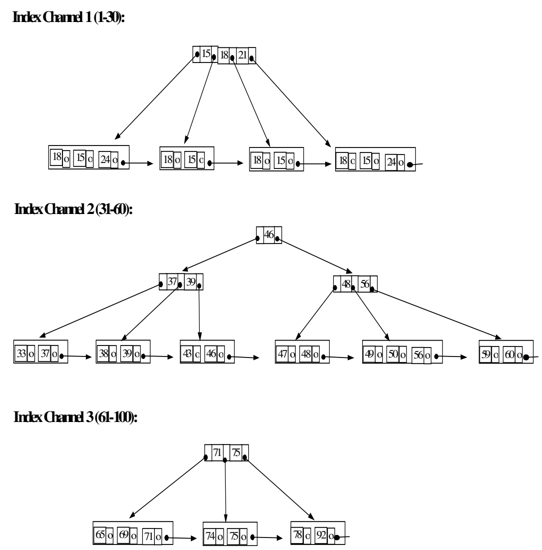


Figure 2. Non-replicated indexing scheme

Figure 2 shows the composition of each index channel with its local index. With this indexing scheme, mobile client needs to know in advance the index channel that contains the desired index key. Assuming that each mobile client is only able to listen to single channel at any time, index channel directory can be broadcast prior to each index cycle or transmitted upon request. Data retrieval mechanism in this scheme can be described as follows:

- Mobile client tunes in one of the index channel (i.e. must be the index channel that has the right index key).
- Mobile client follow the index pointer to the right index key. While waiting for the index to arrive, mobile clients can switch to doze mode.

Proceedings of the 18th International Conference on Advanced Information Networking and Application (AINA'04)
0-7695-2051-0/04 $ 20.00 © 2004 IEEE

- Mobile client tunes back on at the right index key. The index points to the data channel that contains the relevant data item. It indicates a time value of the desired data to arrive in the data channel.
- Mobile client tunes into the relevant data channel, and switch back to doze mode while waiting for the data item to come.
- Mobile client switches back to active mode just before the desired data item arrives, and retrieves the information.

4.2. Partially-replicated indexing (PRI) scheme

A Partially-replicated indexing (PRI) scheme is different from the NRI in a way that it has some degree of replication while NRI has not. Each index channel has a different part of the entire index structure, and the overall structure of the entire index is still preserved.

PRI index is exhibited in Figure 3. Notice that the fifth leaf node (28, 33, 37) is replicated to channel 1 and 2 because key 28 belongs to index channel 1, while keys 33 and 37 belong to index channel 2. Also notice that some non-leaf nodes are replicated whereas others are not. For example, the non-leaf node 15 is not replicated and located only in index channel 1, whereas non-leaf node 18 is replicated to index channel 1 and 2. It is also clear that the root node is fully replicated. As mentioned earlier, the index is broadcast separately with the data, and each index key points to the relevant data channel.

The data structure for our PRI index employs single node pointers model. In the single node pointer model, each node pointer has only one outgoing node pointer.

If a child node exists locally, the node pointer points to this local node only, even when this child node also replicated to other index channels. For example, from node 37 at index channel 1, there is only one node pointer to the local node 18. The child node 18 at index channel 2 will not receive an incoming node pointer from the root node 37 at index channel 1; instead it will receive one node pointer from the local root node 37 only.

If a child node does not exist locally, the node pointer will choose one node pointer pointing to the nearest child node (in case if multiple child nodes exist somewhere else). For example, from the root node 37 at index channel 1, there is only one outgoing right node pointer to child node (48,60) at Index channel 2. In this case, we assume that index channel 2 is the nearest neighbour of index channel 3.The child node (48,60), which also exists at index channel 3, will not receive a node pointer from root node 37 at index channel 1.Using this single node pointer model, it is always possible to trace a node from any parent node. For example, it is possible to trace to node (71,75) from the root node 37 at index channel 1, although there is no direct link from root node 37 at index channel 1 to its direct child node (48,60) at index channel 3. Tracing to node (71,75) can still be done through node (48,60) at index channel 2.

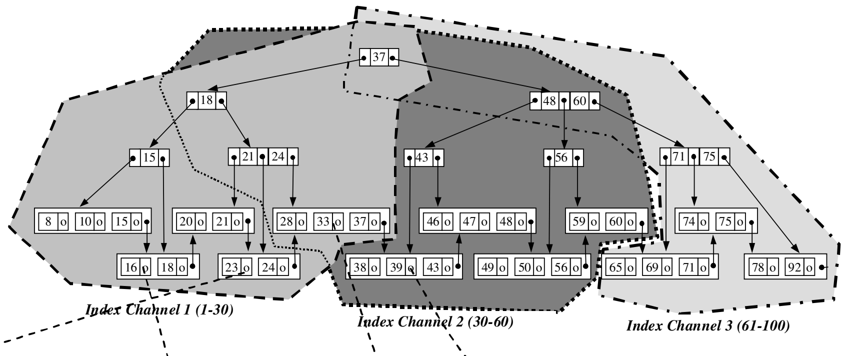
A more formal definition for the single node pointer model is as follows. First, given a parent node is replicated when its child nodes are scattered at multiple locations, there is always a direct link from whichever copy of this parent node to any of its child nodes. Second, using the same methodology as the first statement above, given a replicated grandparent node, there is always a direct link from whichever copy of this grandparent node to any of the parent nodes. Considering the first and the second statements above, we can conclude that there is always a direct link from whichever copy of the grandparent node to any of its child nodes. In this scheme, mobile client can tune in any of the index channel without having to understand the index channel directory. The index pointer may lead to another index channel that contains the right index key. The rest of the mechanism is similar to NRI

4.3. Fully-replicated indexing (FRI) scheme

A *Fully-Replicated Indexing* (FRI) scheme is where the global index structure is fully replicated to all index channels. The index in each channel can be broadcast in different interval. The data retrieval mechanism in this scheme is similar to PRI. However, this scheme does not have a pointer to the other index channel. FRI can also be considered as the solution of problems like data distortion, noise, and signal distortion. Mobile client can switch to another index channel to avoid the problems.

5. Performance Evaluation

In this section, we introduce two cases to analyze the performance of the three broadcast indexing schemes. The simulation is carried out using a simulation package *Planimate,* animated planning platforms [8].



Data Channel 1			
23	Adelaide	Mostly Sunny	24°C
65	Albany	Local Shower	15°C
37	Ballarat	A few showers	12°C
60	Barton	Mostly Fine	19°C
46	Beechworth	Shower or Two	14°C
92	Belmont	Partly Cloudy	10°C
48	Brisbane	Fine	17°C
71	Broome	Shower or two	15°C
56	Bunbury	Mostly sunny	20°C
59	Busselton	Shower	21°C

Data Channel 2			
18	Cairns	Shower	16°C
21	Canberra	Partly Cloudy	23°C
10	Cervantes	Fine	26°C
74	Darwin	A few shower	21°C
78	Gosford	Mostly fine	23°C
15	Grafton	Thunderstorm	21°C
16	Kingscote	Possible Hail	13°C
20	Kingston	Fine	15°C
24	Lismore	Partly cloudy	17°C
28	Maitland	Shower	15°C

Data Channel 3			
39	Melbourne	Mostly fine	25°C
43	Miranda	Thunderstorms and Hail	12°C
47	Newcastle	A few shower or two	15°C
50	Northbridge	Fine	25°C
69	Philip	Cloudy	26°C
75	Queensland	Fine	30°C
8	Queenstown	Mostly sunny	28°C
49	Surfers Paradise	Shower	22°C
33	Sydney	Cloudy	21°C
38	Torguay	Fine	23°C

Figure 3. Partially-replicated indexing scheme

In this simulation, we use the weather scenario as illustrated in Figure 3, and the same set of data items. We run the simulation model for fifty iteration times, and calculate the average access time for given number of request: 5, 10, 15, and 20 requests. Table 1 presents the parameters of concern

Table 1. Parameters of Concern

Parameters	Value
Partially-replicated indexing (PRI)	
Index page in Channel 1	17
Index page in Channel 2	22
Index page in Channel 3	12
Fully-replicated indexing (FRI)	
Index page in Channel 1, 2, and 3	41
Non-replicated indexing (NRI)	
Index page in Channel 1	13
Index page in Channel 2	18
Index page in Channel 3	9
PRI - NRI - FRI	
Node Pointer Size	5 bytes
Data Pointer Size	5 bytes
Indexed Attribute Size	4 bytes
Bandwidth	19200 bytes
Index Arrival Rate	2 index page per sec

.

Case 1: To compare the performance of NRI, PRI and FRI in the context of single index retrieval.

As shown in Figure 4, we can see that NRI outperforms both PRI and FRI respectively. NRI provides approximately three and a half times lower average access time as compared to FRI, and about one and a half times better than PRI.

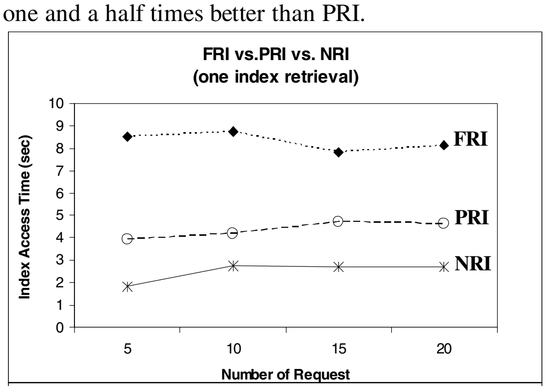


Figure 4. NRI vs. PRI vs. FRI – access time performance (Single Index Retrieval)

Case 2: To compare the performance of NRI, PRI, and FRI in the context of two indexes retrieval. In this case the two indexes are retrieved in unordered fashion, which means the first desired index that arrives in the channel is obtained first. The distance between the first index and the second index is set to random interval. Figure 5 shows that NRI still provides a better access time than the other two. This case follows the trend in case 1 with overall larger access time. The performance of PRI is getting close to NRI with less than one and a half times lower access time

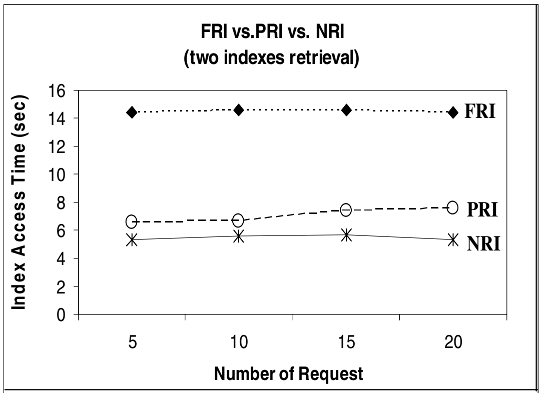


Figure 5. FRI vs. PRI vs. NRI – access time performance (Two Indexes Retrieval)

5.1. Remarks

Based on the two cases, we can say that Non-replicated indexing (NRI) scheme offers a lower index access time as compared to the other two schemes. However, this scheme (NRI) requires mobile client to figure out the right index channel that contains the index of interest. This can be easily solved if client has the ability to listen to more than one channel at any time but it also consumes a substantial amount of energy.

In the simulation, we assume where NRI is utilized, each mobile client has got the knowledge of the index channel directory. Thus, the draw-back of this scheme is not carried out in the simulation, but it certainly needs to be considered. Based on this reason, Partially-replicated indexing scheme (PRI) is considered a better scheme for index dissemination over multi broadcast channel. Fully-replicated indexing scheme (FRI) might be the best option when the channel is vulnerable to noise and signal distortion, as client can move to the other channel and retrieve the desired data.

6. Conclusions and Future Work

Data dissemination or known as data broadcasting is an effective way to keep up with number of clients in a cell and their frequency of requests. We have introduced taxonomy of index dissemination for multi broadcast channel. Three broadcast indexing schemes are considered namely: (*i*) Non-replicated indexing (NRI), (*ii*) Partially-replicated indexing, and (*iii*) Fully-replicated indexing scheme (FRI). These schemes incorporate index channels, and data channels. We have developed a simulation model and analyzed the performance of the three schemes, which includes single index and two indexes retrieval. For future work, we will consider to modify the index structure to accommodate multi index attributes. Consequently, it will be able to support mobile clients to find the data item efficiently using alternative search key or combination thereof.

References

1. Barbara D., "Mobile Computing and Databases-A Survey", *IEEE TKDE*, **11**(1): 108-117, 1999.
2. Badrinath B. R. and Phatak S. H., "An Architecture for Mobile Databases", *Technical Report DCS-TR-351*, Department of Computer Science, Rutgers University, New Jersey.
3. Elmasri R. and Navathe S. B., *Fundamentals of Database Systems,* Third Edition, Addison Wesley, U.S.A., 2000.
4. Imielinski T. and Viswanathan S., "Adaptive Wireless Information Systems", *Proc. of SIGDBS Conf.*, 1994.
5. Imielinski T., Viswanathan S. and Badrinath B. R., "Data on Air: Organisation and Access", *IEEE TKDE,* **9**(3): 353-371, 1997.
6. Lee W.C. and Lee D. L., "Using Signature techniques for Information Filtering in Wireless and Mobile Environments", *Journal on Distributed and Parallel Databases,* **4**(3):205-227, 1996.
7. Leong H. V. and Si A., "Data Broadcasting Strategies Over Multiple Unreliable Wireless Channels", *Proc. of the 4th CIKM*, 96-104, 1995.
8. Seeley D. et al, *Planimate tm-Animated Planning Platforms*, InterDynamics Pty Ltd, 1997.
9. Sheng S., Chandrasekaran A. and Broderson R. W., "A Portable Multimedia Terminal for Personal Communication", *IEEE Communications magazine*, 64-75, Dec, 1992.
10. Taniar, D., Rahayu J.W., "A Taxonomy of Indexing Schemes for Parallel Database Systems", *Distributed and Parallel Databases*, **12**: 73-106, 2002.
11. Waluyo, A.B., Srinivasan B., Taniar D., "Optimal Broadcast Channel for Data Dissemination in Mobile Database Environment", *Proc.of the 5th APPT*, LNCS, **2834**: 655-664, 2003a.
12. Waluyo, A.B., Srinivasan B., Taniar D., "Global Index for Multi Channel Data Dissemination in Mobile Databases", *Proc. of the 18th ISCIS'03,* LNCS, **2869**:210-217, 2003b.