

Time Efficient Indexing Technique for Querying Location Dependent Data in Wireless Broadcast

Seema Verma
AIM & ACT, Banasthali University Banasthali
(INDIA) – 304022

Savita Kumari
Faculty of Engineering College
Zawia (LIBYA) – 16418

Rakhee
Mathematics Group
Birla Institute of Technology & Sciences
Pilani (INDIA) – 333031

Mahmoud A. Artemi
Higher Institute of Vocational Studies, Zawia
(LIBYA) – 16418

ABSTRACT

Wireless communication and mobile computing have gained a lot concentration from the computer and communication research society. Wireless dissemination has emerged as an outstanding and scalable method to broadcast information to huge number of clients equipped with hand held devices. In practice, wireless broadcast faces challenge from issues like timely delivery of data, error free quality of service, low battery capacity and limited bandwidth. Different applications need to curb out different challenges. To support location based services in wireless data broadcast system a distributed spatial index (DSI) is popular. DSI indexing technique in recent times has been becoming popular as reduce access time and energy consumption in location based queries because it has a linear yet fully distributed structure that facilitates multiple search path to be naturally mixed together by sharing links. Here, in this paper we address an issue where error free timely delivery of information is major concern. Such state of affairs generally appears in an in-house environment where recharging of battery is not a serious concern. We have proposed a Data Replicated Distributed Spatial Index (DR-DSI) which is extension of Distributed Spatial Index to get advantages of data replication to reduce access latency.

Keywords

Mobile Computing; Wireless broadcast; Distributed spatial index; Access latency; Data replication.

1. INTRODUCTION

In current years, the employ of wireless technology devices has been increasing at an exponential rate; the standard for timely access of global information is also changing. The recent advancement in mobile network, data dissemination and retrieval strategies, server and client architectures, services and applications have spurred an unprecedented emergence of various techniques that de facto gives a new generation wireless environment. The majority people are able to access information located in wired networks anywhere and anytime using portable size wireless computing devices like notebooks, tablet PCs, personal digital assistants (PDAs) and GPRS-enabled cellular phones, laptops, palmtops which are powered by small batteries. These portable computing devices communicate with a central stationary server via a wireless channel and become the integral part of the existing distributed computing environment. These mobile clients can have an access to database information systems located at the static network while they are traveling and this type of computing is known as wireless computing or mobile computing.

With the forever mounting recognition of smart mobile devices and rapid advent of wireless technology, the vision of pervasive information access has come closer to a reality. While information is important to users, it is only valuable when available at the right time, right place. The demand for location dependent data (e.g., pollution index, local traffic conditions, restaurant locations, navigation maps, etc.) is tremendous due to the broad application base. Thus, location is a very important aspect of pervasive information services.

Wireless computing provides database applications with useful aspects of wireless technology and subset of mobile computing that focuses on querying central database servers is referred as to wireless databases. Mobile service providers have established a number of information services that includes weather information or weather forecast services, news, stock indices information, foreign exchange rates, election results, tourist services, airlines, railways schedules etc. With the introduction of mobile computing, mobile client can have a higher flexibility in accessing the information from different locations with varying network connectivity.

The major shortcoming with broadcast data items in a wireless environment is that data are accessed sequentially. The growing number of broadcast items causes mobile clients to wait for larger time before receiving desired data item. Consequently, dependence of mobile devices on rechargeable batteries, which has limited capacities, is also another drawback of wireless data retrieval. The rate of increase in the chip density is much higher than the rate of increase of battery capacity. Also, due to always increasing demand for mobile information services, huge numbers of operators providing services come in to fray which cause large data broadcast rate causing deterioration in quality of services. In order to overcome these drawbacks and improve system performance, it is necessary to visualize two performance matrices viz: access time and tune time. Former is related with retrieval delay of data while later is associated with energy consumption during data retrieval. Since these parameters are at odds to each other so cannot be reduced to great extent simultaneously, but a tradeoff between two can be set for better system performance.

In a mobile atmosphere queries can be segregate into two categories traditional queries and location-dependent queries. The queries cite in traditional wired environment are traditional queries, while when these are transmitted over a wireless communication network, called location-dependent queries. It is realized that location-dependent queries has become common and of great interest. Therefore, providing efficient and effective mobile information services that cover both traditional queries and location-dependent queries will be highly desirable. The process of queries retrieval in wireless environment can be

performed in two ways (i) pull or on demand mode and (ii) push or broadcast mode.

On-demand queries are those where the client initiates the query and sends it to the server. The server processes the query and sends the result back to the client. In broadcast-based queries, the server broadcasts the data items periodically over one or more broadcast channel. Mobile clients tune to it and select data items of interest and detain it. Through broadcast-based queries, a mobile client is capable to get information without wasting power to convey a request to the server. It supports a huge number of users at a time and the number of users in a cell or high information down loading rate does not affect query performance. It is effective in spite of a large number of user's retrieving data simultaneously.

The emergence of ubiquitous and small mobile devices along with the development of fast, reliable and accessible network has helped to advance research in pervasive computing area. The various data broadcast management techniques like indexing, scheduling and broadcast along parallel channels pivot this development. Broadcast data can be managed in various ways. The broadcast in which a data item is placed once on a single broadcast channel is called flat broadcast and a broadcast in which a single data item appear more than once is called skewed broadcast.

The behavior of broadcast system is unidirectional and cyclic hence mobile client need only tune-in to channels. For effective data broadcast tree based and signature based, approach are used. Tree based approach gives the exact time of arrival of data so that mobile client can go to doze mode and save considerable amount of battery power; while signature based approach use signature to examine data object, which result in saving access time. Further, different application scenario demands minimization of different parameters. In an in-house environment where battery can easily be recharged or devices can be connected to direct power supply, only access time minimization is sufficient and in remote area where device has to depend on previously charged battery tune time needs a control. In some applications both parameters needs a trade off.

During preceding research has deliberated a variety of system issues of wireless data broadcast [1, 2, 3]. This paper address the insist of location dependent data by proposing a novel indexing technique, called Data Replicate Distributed Spatial Index (DR-DSI), in support of location-based queries from mobile users in wireless data broadcast systems.

The design of DSI has also taken the inherently unreliable and error-prone wireless communication into consideration. We are intending to extend the existing Distributed Spatial Index to facilitate it with data replication at bucket level. The repetitions of data over broadcast channel fasten the data retrieval process. Our study has led to an innovative solution based on broadcast reorganization and resulted in very efficient query processing.

The rest of this paper organized as follows section 2 is provided Background and existing work to our study. The index structure of DR-DSI and time efficiency and energy efficient broadcast organization are presented in Section 3. Finally, we conclude this paper in section 4.

2. PRELIMINARIES

At this point the background and related work in favor of sustaining location-based wireless broadcast services are provided.

2.1. Background

Regard the same as a wireless data broadcast system to facilitate periodically broadcasts a collection of data objects to mobile clients. Every data object consists of a set of attribute values. Along with them, location and data replicated factor is mainly important because it is the focus of this paper. Access latency and power conservation are two critical issues for mobile users, about how fast a request could be satisfied and how is the energy efficient a technique.

To make possible energy protection, a smart mobile device is support two operation modes: active mode and doze mode. The device generally operates in active mode; and switches to doze mode to save power when the system becomes inactive. In the literature, two performance metrics, namely access latency and tuning time, are used to measure access efficiency and energy conservation for mobile clients in a wireless data broadcast system, respectively [5, 6]:

Access Latency: Time elapse from the moment a request is initiated until all data item of interest are received.

Tuning Time: The amount of time spent by the client to actively listen for the desired broadcast data items.

During wireless data broadcast systems, a client has to remain *active* to constantly receive and check the broadcasted data until the data of interest arrive. This method consumes plenty of energy.

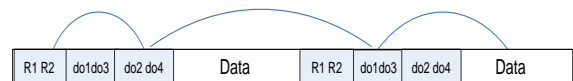
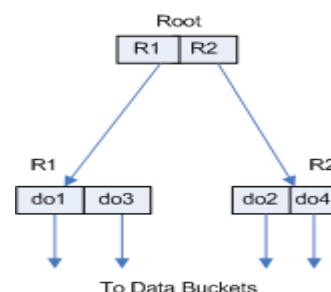


Figure 1. Linear Access on Broadcast channel

The average tuning time is a half of a broadcast cycle, a period when all the available data objects are broadcast one time.

Air indexing is often employing for power saving at mobile clients [6, 7, 8, 9]. The essential scheme is that the broadcast server recomputed indexing information (including indexed attributes and arrival time of data objects) and interleaves it with data objects on the broadcast channel. To provide search of data objects via air index, each data object includes an offset to the next broadcast of index information. As such, mobile clients are able to forecast the arrivals of their preferred data via examining the index information. Thus, they can stay in doze mode most of time and tune into the broadcast channel when the interested data objects arrive.



(a) R-tree Index

Various spatial index structures have been proposed for accessing spatial data, including R-tree, KD-tree etc [10, 11, 12]. Along with those, R-tree is the best received. A search algorithm based on R-tree usually expands the search space approximately the query position using a branch-and-bound approach. The direction-finding order of R-tree is dynamically determined based on the position of the query point, which results in backtracking. Thus, R-tree is better supported by random access storages, such as memory and disk.

Within a wireless broadcast system, even though data objects are broadcast based on a pre-defined sequence is called a broadcast program and hence the data object is only available when it is on the air. Thus, search algorithms designed based on arbitrary access may incur significant access latency. In Figure 1, we are depicting an example of linear access of data object broadcast. Suppose that a search algorithm visits the root node, R2 and then R1, while the server broadcasts nodes in the order of root, R1, and R2. Consequently, if a client wants to visit node R1 after it retrieves R2, it will have to wait until the next cycle because R1 has already been broadcast. This significantly extends the access time and it occurs every time a search order is different from the broadcast order. To address this issue, the navigation order of R-tree needs to follow the broadcast order of index nodes.

While nearly all location-based queries search for data objects located closely, one approach for wireless data broadcast is to schedule spatially near data objects to broadcast close to each other. In the direction of attain this, an idea is to broadcast data objects based on a space-filling curve (e.g., Hilbert Curves (HC) by C.Gotsman et al.(1996) and build an air index upon. The main thing is to keep neighbors in a high dimensional space remaining close to each others in the broadcast channel. In Figure 2 we are showing an HC of order 3. The numeric labels represent the positions of the objects in terms of HC values. In favor of instance, point (1; 1) has the HC value of 2. The order of the curve is determined by the data object allocation in the space. HC of higher order is desired for denser data object allotment because the curve has to pass through all the objects.

There is a 1-1 correspondence among the coordinate and HC value of a data object. Given the mapping function of HC, it is easy for a client to perform conversion between coordinates and HC values with in a constant time. The detailed conversion of the algorithm is available on this link. <http://www.caam.rice.edu/doug/twiddle/Hilbert>.

2.2. Related Work

The offer of broadcast disks by Acharya et. al. [1] in which hot data items are allocated more frequently than cold data items on disc from which average access time decrease; enunciate new paradigm in mobile computing. Amar and Wong describe the architecture of teletext broadcast cycles system considering data access probability [4]. Selective tuning is best possible way to reduce power consumption in single channel environment. A parameterized distributed index is proposed by Xu et. al. (2004) to reduce index load on channel by forming index on per chunk basis [7].

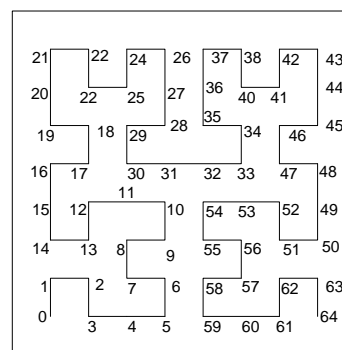
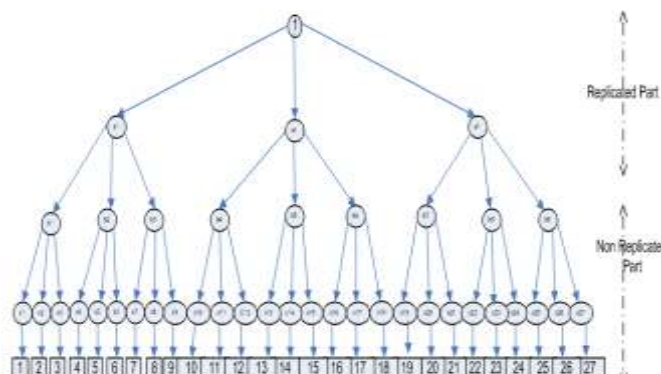


Figure 2: Hilbert Curve of order 3

Imielinski et al. has extended B+-tree to maintain the access of broadcast data. Two strategies are proposed to interrupt the index and data in a broadcast channel, namely (1, m) and distributed index [9]. The previous treats the entire index as a fragment and replicates the index segment m times during one broadcast cycle. Therefore, the clients suffer longer access latency because the m duplicated index fragment extends the overall broadcast cycle. The distributed index replicates only the top part of an index tree illustration in Figure 3 for. It has been revealed that the distributed index scheme is more efficient than (1, m) in terms of access latency. Ideas of indexing the attribute ranges of exponentially increasing number of data objects were discussed in the literature, e.g., flexible index [5] and exponential index [7].



However, the focus of these works is totally different from our study. Chord aims at providing peer-to-peer lookup based on a hashed search key, while flexible index and exponential index investigate optimal tuning of the access latency and tuning time in support of simple search of broadcast data based on single attribute. None of them considered complex location-based queries with hot data and cold data replication as we do in this paper.

A few air indexing techniques have been in recent times proposed to support broadcast of location dependent data [13, 14, 15]. Amongst them, the D-tree is a paged binary search tree to index a given solution space in support of planar point queries [13], as the Grid-partition index is specialized for the (one) nearest neighbor problem [15]. Hilbert Curve Index (HCI) is designed to support both of window queries and nearest neighbor queries at the same time. It adopts a B+-tree to index data objects broadcast according to the Hilbert Curve order [15].

3. DATA REPLICATE DISTRIBUTED SPATIAL INDEX (DR-DSI)

The entire spatial air index techniques evaluated earlier are based on tree structures. Therefore, a search always set up by the root, which results in some deficiencies. First, the patrons have to wait for the coming of the root node to begin the search. Besides, the search has to be stopped once an index node alongside the search path will be lost. The client must wait for the next root or blindly scan all the following nodes in order to resume the search. On other hand, some information that is broadcasted is accessed very less that also increase the access latency of the more required data. Therefore, those indexes only perform well under an ideal situation where no packet loss happens. To address these inherited deficiencies of tree indexes on air, the data replicate fully distributed spatial index structure, namely DR-DSI, to allow a client to start query processing as soon as possible in order to minimize the access latency while still conserving tuning time.

DR-DSI distributes the index information over the whole broadcast cycle and equips a client, every time it tunes into the channel, with adequate information to conduct the location-based search as well as with broadcasting frequency factor. The data replicated full distributed index if decide according the hot data (the data have more demand) and cold data (the data that have the less demand) then it will recover the access time as well as compare to DSI and will not increase the tuning time because index will be less to comparison

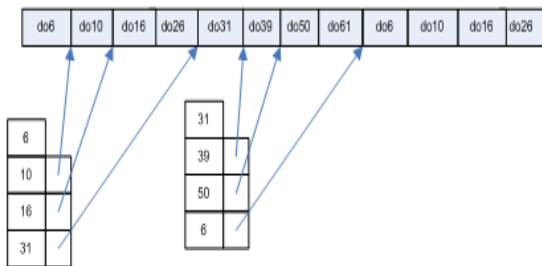


Figure 4: DR-DRI for the Processing Example

the DSI. To follow this, we introduce the index structure of DR-DSI, time efficient and power efficient forwarding, using the search algorithms for window query and k-Nearest-Neighbor (kNN) query search is discussed by Lee, W.C et.al. (2005).

3.1. The Index Structure

The index structure of the DR-DSI is taking into description the in sequence access property of wireless broadcast, Hilbert Curve (HC) is adopted in the preliminary design of DR-DSI to detect broadcast order of data objects. By default, data objects are broadcast in the ascending order of their HC and BRF (broadcast frequency factor) values.

The basic idea is to divide the whole set of data objects into nF frames as well as BRF and associate with each frame an index table. The index table maintains information regarding to the HC and BRF values of data objects to be broadcast with specific waiting interval from the current frame. Various symbols used while synthesizing the model are shelled in table 1.

DR-DSI index table consist of a number of table entries, Ti in the form of (HC_i', Pi, BRF_i) where 0 ≤ i ≤ (logr(nF)-1), r is selected exponential base is called index base, and nF is the number of the frame with in broadcast cycle and DR-DSI takes data access

frequency in to description and allows the data to replicate at bucket level (logically basic unit of data transfer in broadcast) according to access frequency. The bucket containing hot data (most demanded data) appears more times than bucket containing cold data (least demanded data) on broadcast channel. A critical bucket access frequency is decided by broadcast scheduler based on bucket access statistics gathered by it during last day, week or months time. Each data is assigned a number called bucket replication factor (BRF). BRF may be defined as below:

Definition BRF: - If P_c is critical bucket access frequency and

P_i is demand frequency for i^{th} bucket at present than

$$BRF_i = \lfloor \frac{P_i}{P_c} \rfloor \quad \forall i \in t$$

, where t is the numbers of data bucket without replication, is called BRF for i^{th} data bucket. It gives number of times data bucket appear on broadcast cycle.

DR-DSI differs from original Distributed Spatial Index in three aspects:

1. Index table consist of a number of table entries, Ti in the form of (HC_i', P_i, BRF_i) where 0 ≤ i ≤ (logr(nF)-1),
2. Instead of that it will have another factor BRF for more flexibility.
3. In later data buckets access pattern is taken in to consideration and buckets are allowed to replicate on broadcast channel hence client get access gain from data repetition as well as missed data can be accessed without waiting for next broadcast resulting in easy data access with less error.

Table 1: Symbols for various parameters

Symbol	Parameter
n _F	Number of frames within one broadcasted cycle
R	Exponential base (called <i>index base</i>)
HC ₀ '	HC _i ' is the smallest HC value of the objects within the frame pointed by Pi.
P _i	Point to the next i^{th} frame
t _i	Table Entries
R _i	BRF for i^{th} bucket
F	Set of frames
n ₀	Data Object factor
do	Data object

Logically, the set of frames beginning from any arbitrary frame F until the frame before reappearance of F forms a broadcast cycle. Therefore, the index table associated with a frame F is designed to cover the next $(nF - 1)$ frames following F , i.e., providing index information on HC value of all the frames within a broadcast cycle. P_i points to the next i th frame. HC_i is the smallest HC value of the objects within the frame pointed by P_i and BRF_i .

Thus, the i th entry of the index table provides range information of HC and BRF values amongst data objects in the i th to $(i+1 - 1)$ th frames. In other words, the number of frames covered is exponentially increased with the order of index table entries. The number of data objects within a frame n_0 can be decided based on various parameters such as the size of data packets, etc. data object factor is correlated to the overall index size and the average waiting time to reach the next index table and the bucket replicate factor.

Additionally, the index base r and BRF can be chosen to control the overhead of index table, i.e. the number of entries within one index table. For simplicity, we assume the object factor to be 1 and the index base to be 2 and BRF to be 1 to 3 in our discussion.

Figure 4 shows the broadcast of data objects based on the example in Figure 2 and the index tables. In this example, $nF = 8$ and thus each index table has 3 entries. The DR-DSI table corresponding to frames of data objects do_6 and do_{32} are also shown in the figure 4. Take the index table for frame do_6 as an example: T_0 contains a pointer to the next upcoming frame whose HC value is 11, T_1 contains a pointer to the second frame with HC value 17, and the last entry T_2 points to the fourth frame with HC value 32.

3.2. Access Protocol

One critical process for DR-DSI is to proficiently achieve a frame containing the data object of a given location. This operation will become point query if they reached frame is scanned to find the queried data object. The access protocol based on DR-DSI is informally given as follows.

1. Specified a target point p , a client first calculate the HC and BRF value of point p , denoted as HC_p , and BRF_p and then tunes into the broadcast channel.
2. After the initial probe, it retrieves the index table linked with the first frame encountered. By comparing HC_p with HC_i 's is maintained in the index table, the client chases the pointer P_i , where HC_p belongs $[HC_i; HC_{i+1}]$.
3. The client goes into the doze mode and wakes up when the frame specified by P_i arrives. This process continues until the frame containing data object located at p is reached.
4. As shown earlier, DR-DSI organizes the index information in a way such that the ranges of HC values amongst exponentially increasing segments of frames are maintained in the index tables.

As a result, TE-EEF is logically like a binary search (when the index base is 2 and BRF is 1 to 3) and the distances among visited frames and the final target frame decrease rapidly.

4. CONCLUSION

In this paper attractive air indexing technique which exploit properties of wireless broadcast, facilitate access and time and energy efficient query dealing out, and be resilient in error-prone wireless environments.

Here a data replicated distributed spatial index, named DR-DSI, is proposed to address these requirements. DR-DSI naturally mixes multiple search paths into a linear index structure which is fully distributed into the whole broadcast cycle with the hot data more replication. Therefore, it consented to a search to initiate immediately and make possible direct recovery of interrupted query processing when a packet is corrupted or lost.

In this paper we are currently looking into prototyping of a location-based wireless broadcast system. Meanwhile, we continue to exploit the potential impacts and use of broadcast reorganization. For the future work the research issues in unreliable wireless communication environments and implements with the result.

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] A. Datta, A. Celik, J. Kim, D. VanderMeer, and V. Kumar. Adaptive broadcast protocols to support power conservation retrieval by mobile users. In *Proceedings of IEEE International Conference Data engineering*, pages 124.133, Birmingham, UK, April 1997.
- [3] A. Datta, D. E. VanderMeer, A. Celik, and V. Kumar. Broadcast protocols to support efficient retrieval from databases by mobile users. *ACM Transactions on Database Systems (TODS)*, 24(1):1.79, March 1999.
- [4] M. H. Ammar and J. Wong, "The Design of Teletext Broadcast Cycles," *Performance Evaluation* 5(4):235–242, 1985.
- [5] T. Imielinski, S. Viswanathan and B.R. Badrinath, "Energy Efficient Indexing on Air," *Proceedings of the ACM SIGMOD Conference*, pp.25-36, 1994.
- [6] T. Imielinski, S. Viswanathan and B.R. Badrinath, "Data on Air: Organization and Access," *IEEE Transaction on knowledge and Data Engineering*, 9(3) May/June 1997.
- [7] J. Xu, W.C. Lee, and X. Tang, "Exponential Index: A Parameterized Distributed Indexing Scheme for Data on Air," *MobiSys 2004*, pp. 153–164, 2004.
- [8] Q. L. Hu, W.-C. Lee, and D. L. Lee. Power conservative multi-attribute queries on data broadcast. In *Proceedings of the 16th International Conference on Data Engineering (ICDE'2000)*, pages 157.166, San Diego, CA, USA, February 2000.
- [9] W.-C. Lee and D. L. Lee. Using signature techniques for information filtering in wireless and mobile environments. *Journal of Distributed and Parallel Databases (DPDB)*, 4(3):205.227, July 1996.
- [10] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 47.54, 1984.

- [11] J. Robinson. The kdb tree: A search structure for large multidimensional dynamic indexes. In Proceedings of the 1981 ACM SIGMOD International Conference on Management of Data, Ann Arbor, Michigan, 1981.
- [12] H. Samet. The quad tree and related hierarchical data structures. *ACM Computing Surveys*, 16(2), 1984.
- [13] J. Xu, B. Zheng, W.-C. Lee, and D. L. Lee. Energy efficient index for querying location-dependent data in mobile broadcast environments. In Proceedings of the 19th IEEE International Conference on Data Engineering (ICDE'03), Bangalore, India, March 2003.
- [14] B. Zheng, J. Xu, W. C. Lee, and D. L. Lee. Energyconserving air indexes for nearest neighbor search. In Proceedings of the 9th International Conference on Extending Database Technology (EDBT'04), Heraklion - Crete, Greece, March 2004.
- [15] B. Zheng, W. C. Lee, and D. L. Lee. Spatial index on air. In Proceedings of the first IEEE International Conference on Pervasive Computing and Communications PerCom'03), Dallas-Fort Worth, Texas, USA, March 2003.