Research Paper Engineering

Hybrid Indexing For Optimal Data Broadcast in Wireless Environment

Dr. Savita Kumari (Sheoran)

Department of Computer Science & Applications Indira Gandhi University Meerpur, Rewari, Haryana (INDIA) - 122502

ABSTRACT

Recently the data broadcast has emerged as powerful tool for mass data dissemination in wireless environment in a scalable manner. The broadcast based system in generic form have disadvantage that mobile client have to spent more energy in search for desired data item. Various indexing schema viz. hashing, tree and table based indexing have been developed to mitigate this issue of energy consumption but no scheme have been found suitable in all instances of broadcast environment. Exponential Indexing (EI) is best known for its flexible and error resilient behavior. Further Data Replicated–Exponential Indexing (DR-EI) and Distributed Spatial–Exponential Indexing (DS-EI) have been developed for better tradeoff between performance parameters of data broadcast. Also, Signature Indexing is a hash based indexing techniques which achieve optimal energy consumption. In this paper we have developed a hybrid indexing relying on united power of Signature Indexing and DS-EI schema to design an optimal data broadcasting strategy for minimum latency and low energy consumption along with flexibility and error resilient demeanor. The simulation results attest that proposed Hybrid Indexing outperforms these two states of art indexing techniques.

KEYWORDS : wireless broadcast, air indexing, multichannel broadcast, optimal data broadcast.

Introduction

In recent years, the use of wireless technology devices has been growing at an exponential rate. The mobile devices such as smartphone, PDA and GPRS-enabled cellular phones have become common; which further enhanced due to prolifration of social networking sites such as Facebook, Twitter and LinkedIn. According to a study released by Internet And Mobile Association of India (IAMAI) is the third largest smart phone market in the world and forecasted that it will have 314 million active mobile internet users by 2017. Moreover, about 63% of adults in USA use location based services in their iphone. This enourmous use of mobile devices brough opportunity for technology savvy people while challenges for wireless research community.

In wireless environment, the data can be disseminated in two ways: (i) pull or on-demand mode and (ii) push or broadcast mode. On-demand mode is one in which 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 mode, the server broadcasts the data items periodically over one or more broadcast channel(s). Mobile clients tune to it and select data items of interest and capture [1-2]. In broadcast system, the server continously transmit data over dedicated channel based on some user statistics of past few days, week of month, irrespective of the client current use pattern. With this mechanism, the requests from the clients are not known a priori. The wireless systems are hetrogenous i.e. upstream bandwidth is always much smaller than downstream bandwidth. In addition, query uploading requires more energy in comparision to downloading. Broadcast environment is unidirectional, scalable and energy efficient in the sense that the server disseminates a set of data periodically to multiple numbers of users without any affect from simultaneous data retrieval by large number of clients and a mobile client is able to retrieve information without wasting power to transmit a request to the server. These novel features of broadcast approach makes it popular among professionals engaged in data management. Nevertheless, its major shortcoming is that data items are accessed sequentially. The increasing number of broadcast items causes mobile clients to wait for larger time before receiving desired data item [3]. 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. In addition, due to ever-increasing demand for mobile information services, huge numbers of operators providing services come into fray that 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*.

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

Tune time: The amount of time spent by the client to listen for the desired broadcast data item(s). It comprises time taken in two modes viz: active and doze mode.

Since these performance parameters are at odds to each other and can't be reduced to great extent simultaneously, but a tradeoff between two can be set for better system performance. In real time system the pure data broadcast in its generic form have low access time but to minimize energy consumption of battery we employ some techniques like indexing, partitioning, clustering of data which give rise to new version of problem of access latency. It means dilemma of high access time itself is originated from solution of problem of minimization of tune time. Imielinski et al. [1-2] have discussed this problem in one dimension, which extremely optimize one of these two performance matrices. They provide two algorithms for this as shown in figure 1 below.

The *Access_opt* algorithm provides the best access time with a very large tuning time. The best access time is obtained when no index is broadcasted along with the file. The size of the entire broadcast is minimal in this way. Clients simply tune into the broadcast channel and filter all the data until the required record found.

The *Tune_opt* algorithm provides the best tuning time with a large access time. This algorithm suggests the process of indexing the data item to save active time of MC. The MC once get the information about exact time of arrival of requisite data item and goes to doze mode and till the data item is available to him for download. In this way, it can save considerable amount of battery power. The tuning time is equal to the number of levels in the multi-leveled index tree plus one level for the final probe to download the record. This method has the worst access time because; clients have to wait until the beginning of the next broadcast cycle.

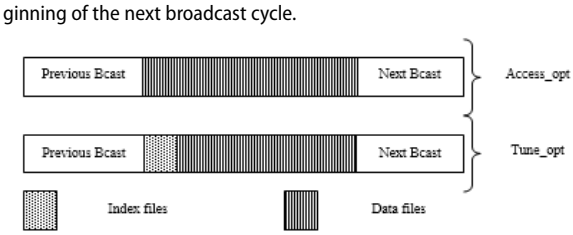

Figure 1: One Dimensional Data Organization

The performance of these two algorithms is reciprocal to each other and can't be improved without mutual adverse brunt. We need to create a balance between these two by extending the problem in two dimensions so that one parameter can be optimized with in tolerable limit of other. In this scenario role of proper data organization play an

important role.

This paper hybridize the DS-EI established by Kumari S. [4] which is an extended version of DR-EI proposed by Verma et al. [5] with Signature Indexing. DR-EI performs well in a situation where energy expense is not a major concern. But for mobile devices the battery life is very scared resource, hence need to be optimized. This amalgamation of two novel states of arts indexing techniques will inculcate minimum latency and low energy consumption along with flexibility and error resilient demeanor. The rest of this paper is organized as follow: Section II, introduces the background and gives related work on topic. In section III, an overview to DS-EI proposed by Kumari [4] is presented. Section IV, presents state of arts Signature Indexing. The first sub-section of V develop a theoretical model to evaluate the performance of our proposed strategy while second sub-section evaluates its performance experimentally through simulation results carried out using Scilab at different parameter values. Section VI, finally concludes the paper and allude to direction for further research in this field.

Background

The propose of broadcast disks by Acharya et al. [6], 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 [7] describe the architecture of teletext broadcast cycles system considering data access probability. Selective tuning is best possible way to reduce power consumption in single channel environment. A (1, m) indexing and distributed indexing method were developed to efficiently replicate and distributed the index tree in broadcast by Imielinski, Viswanathan and Badrinath [2] which is further extended for skewed broadcast by them. Various Indexing and scheduling techniques to effectively manage data have been discussed in literature. Yee et al. [8] develops efficient strategy of data allocation on multiple channels. Hsu et al. [9] considered data access frequency while allocating data and index over channels. A parameterized distributed index is proposed by Xu et al. [10] to reduce index load on channel by forming index on per chunk basis. Further, they have extended it for error resilient case to improve data retrieval in fault ridden environment. Seifert and Hung [11] have developed FlexInd to increase flexibility of access tune trade off in exponential index. Verma et al. [5] have experimented DR-EI and Kwangjin Park [12] has suggested location-based grid-index for spatial query processing, which is a distributed grid index to reduce access latency.

Hashing techniques are used in Vijayalakshmi and Kannan [13], while signature techniques are used in Lee and Lee [14]. Hu et al. [15] showed that the signature method is particularly attractive for multi-attribute indexing. Ho and Lee [16] have addressed such problem by designing the central index for all broadcast operators. To make search process easy in overloaded air spectrum environment we have proposed Unified Indexing Hub (UIH) as a new broadcast mechanism Verma et al. [17]. The performance of UIH in comparison to commonly used state of art Two Level Signature Index Model (TLSIM) proposed by Lee and Lee [14] in perspective of two key performance metrics, namely access latency and tuning time has been evaluated in Verma et al. [18].

This work is further extension of DS-EI having access latency reducing capacity of distributed spatial indexing and states of arts Signature Indexing. The proposed model is elaborated in subsequent sections.

Distributed spatial–Exponential Index (DS-EI)

This section presents an extension of DR-EI, a situation specific case parameterized and distributed exponential indexing scheme, to allow index distribution. It is an innovative data management schema, which allow, energy consumption to be reduced considerably at negligible cost of access latency. The proposed extension for data replicated exponential index called distributed spatial-exponential index (DS – EI) can serve better.

Preliminaries

In exponential Index two parameters I and r are tunable to adjust access latency and tune time. DS–EI takes data access frequency in to account 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), as defined in Verma et al. for DR-EI.

DS–EI differs from original exponential index in two aspects:

-Instead of two parameter I and r in original EI, it can be tune on three parameter I, r and BRF hence become more flexible.

-In DS-EI 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.

-In DS-EI an entry of indexkey tupple is present with in global index.

Index Structure

The exponential index naturally replicates the index and DS–EI replicates the data buckets and have provision for indexkey tupple to be present in global index. This paper has taken data transmission over single channel only in to considration because multichannel is similar to single channel with bandwidth splitting. Different data items have different access probabilities because all data are not equally demanded. According to Zipf distribution of demand probability it follow 80/20 rule i.e. 80 % client demand, 20% data and vice versa. In DS–EI all data items are sorted according to their access probabilities. There can be two types of buckets in broadcast one with index, containing index and some data according to available space and other with out index explicitly containing data. These data are filled in buckets as per available space. The buckets including replicated are adjusted on broadcast channel as shown in figure3. All buckets are divided in to data chunks. A *data chunk* is a group of data buckets. The first bucket of each chunk contains an index entry and data in remaining part. It is wisely to develop index at chunk level instead of bucket level to reduce index overhead.

The index entry consists of global index and local index. Global index indices the coming data chunk and local index indices the bucket(s) with in chunk. Global index have two fields distint (distance to end of data bucket indexed by index entry) and mkey (maximum key value). Now, each bucket is assigned a weight called BRF according to its demand. The data buckets are replicated as many times as BRF in sequence. Figure 2 show index structure of DR-Exponential Index with I=2, r =2 and BRF ranging 1-3, where 'r' is term called index base as in exponential index. In DR-EI factor 'r' have no role because it affects only the tune time and have no effect on access time but in DS-EI it is considered an important factor.

Access Protocol

The access protocol for initial data buckets is similar to DR-EI with some addition for distributed index with in global index. The access protocol to get data with attribute value 'K' is as follows:

- Tune in to the broadcast channel and get pointer to the bucket containing index.
- Tune in again to the beginning of designated data bucket. Search its global index.
- If initial_distint<K=indexkey, than tune in to initial_distint and follow step (ii) else search local index, get the pointer to data.
- If false drop, than continue to search global index of next bucket:

- If a link found there than follow the step (ii) and (iii).
- Else wait for next broadcast and follow step (ii).

Signature Indexing

This section is devoted to states of art Signature Indexing techniques better known for minimum tune in time and hence minimum energy consumption.

Preliminaries

Signature of a data frame is obtained by bit vector generated through first hashing the values in the data frame into bit strings and then superimposing them together through bitwise-OR operation (∨) to form signature. It is widely used for information retrieval in broadcast. The signature technique interleaves signatures with their associated data frames in data broadcasting. A signature is easy to generate and can be used for any type of media and have size much smaller than that of the data record. So it is considered a powerful tool to retrieve data records. Three signature indexing based access methods simple signature, integrated signature and multi-level signature have been proposed by Hu, Lee and Lee [ ] 2001). Various types of signatures are:

Simple signature: It is the simple type of signature which indexes the data frame and follows its on broadcast channel. Simple signature is shown in Figure 2(a).

Integrated Signature: An integrated signature indexes the frames group of data instead of individual data. The frames group may contain any number of data records. Integrated signature is shown in Figure 2(b).

Multilevel Signature: The multilevel scheme is a combination of the simple signature and integrated signature schemes. It consists of multiple levels of signatures. The signatures at the upper levels are integrated signatures and the lowest levels are simple signatures. Multilevel signature scheme is shown in Figure 2(c). To avoid the confusion for data retrieval different hash functions are used to generate simple and integrated signature.

The integrated and multi-level signature indexing methods are designed to handle more complex data structures.

Access Protocol

All types of signatures are together called information signature. Information signature for ith data item (S_i) is available along with data record for filtering. To process a query which contains one searched attribute, a query signature (S_q) corresponding to the query 'q' is generated based on the same hash function. Sq is compared against the signatures of examined data items using bitwise-AND operation (∧). There are two possible outcomes of the comparison:

$S_q \wedge S_i \neq S_q$, i.e. data item 'i' does not match query 'q'.

$S_q \wedge S_i \neq S_q$, i.e. a match occurred, it has two possible implications:

True match: the data item is really what the query searches for; and *False drop/match*: Signature comparison indicates a match although the data item in fact does not satisfy the search criteria.

Performance Evaluation

We theoretically analyze the performance of DS-EI to evaluate the time efficiency and compare it with exponential index for same data set. In DS-EI and DR–EI, there are three tunable parameters: r, BRF and I. In traditional exponential index, there are only two tunable parameters: r and I.

Theoretical Analysis

This sub-section is devoted to theoretical analysis of DS–EI. Let there are 'N' data items and for simplicity considering all data items of equal size. Initially there are 't' buckets available for 'N' data items. Various symbols used while synthesizing the model are shelled in table 1.

Table 1: Symbols for various parameters

Symbol	Parameter
N	Number of data items to be broadcasted
B	Capacity of data bucket with out index
B'	Capacity of data bucket with index
I	Number of buckets in data chunk
T	Number of data buckets containing data without replication
R_i	BRF for ith bucket
C	Number of data chunks in broadcast cycle
S_o	Index overhead in first bucket of chunk

The bucket with index contains less data than bucket with out index by an amount equal to index load.

Therefore, $B' = B - S_o$

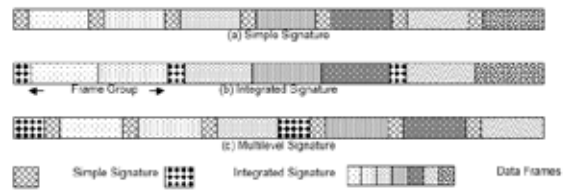

Figure 2: Signature with Data Frames on Broadcast Channel

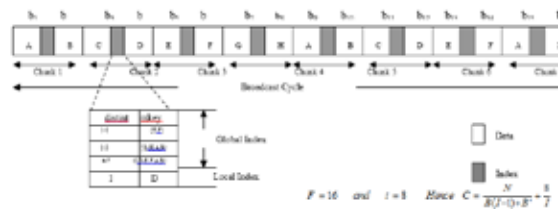

Figure3. : Distributed Spatial – Exponential Index (I=2, r =2, BRF = 1-3)

Let $b=\{b_1, b_2, b_3, \ldots\ldots b_t\}$ be set of data buckets containing data without replication with $P=\{p_1, p_2, p_3, \ldots\ldots p_t\}$ as set of

their access probabilities and the set R of BRF's for all buckets $R=\{R_1,R_2,R_3,R_4,R_5,R_6,R_7,R_8\}$ with R_i as BRF for ith item. A data chunk consists of one bucket with index entries and remaining $(I-1)$ buckets with out index entry, it can hold $[B(I-1)+B']$ data items. Therefore, number of chunks formed from non-replicated buckets is

$C'=\frac{N}{B(I-1)+B'}$ The total number of buckets in broadcast cycle is.

$F = \sum_{i=1}^{i=t} R_t$

Out of these buckets $F-t$ are replicated. Hence, number of chunks formed from replicated buckets is

$C'' = \frac{F-t}{I}$

Total number of chunks in broadcast cycle

$C = C' + C'' = \frac{N}{B(I-1)+B'} + \frac{(F-t)}{I}$

The average access latency consists of initial probe in chunk and half of the length of broadcast cycle. In DS – EI some buckets appears more than once on broadcast channel, such buckets can be accessed with shorter latency. The ith bucket can be accessed with IC/R_i times. The effective broadcast length is

$$L = \frac{1}{t}\left(\frac{IC}{R_1} + \frac{IC}{R_2} + \frac{IC}{R_3} + \ldots\ldots\ldots + \frac{IC}{R_t}\right) + 2^{-r}$$

$$AverageAccessTime = \frac{I}{2} + \frac{L}{2} = \frac{I}{2} + \frac{IC}{2t}\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots\ldots\ldots + \frac{1}{R_t}\right) + 2^{-r}$$

$$F = 16 \quad and \quad t = 8 \quad Hence \quad C = \frac{N}{B(I-1)+B'} + \frac{8}{I}$$

The average access latency calculated here can be used as general formula for 16 bucket broadcast system with out serious consequences for any number of data item and chunk size.

$$AverageTuneTime = \frac{I}{2} + \frac{L}{2.r} = \frac{I}{2} + \frac{IC}{2.r.t}\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots\ldots\ldots + \frac{1}{R_t}\right) + \frac{1}{2}$$

Results

This section presents simulation results for DS-EI presented in section V (A) over Scilab. The performance of DS-EI is compared with DR-EI and EI for access latency and tune time. For calculation purpose, different parameters are set as mentioned in table 2.

Table2: Values of different parameters for results

Parameter	Value	Parameter	Value
B	20-110	S_o	14
I	1-10	N	300-1200
R	2-4	F	16

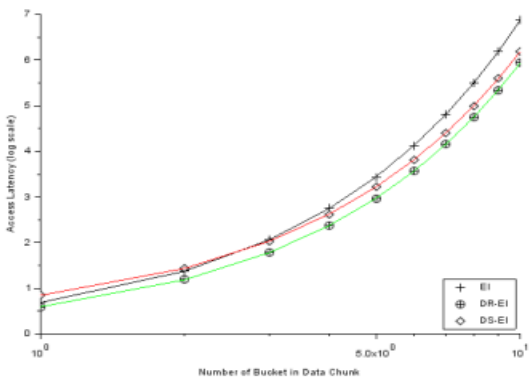

Figure 4: Effect of chunk size on access latency

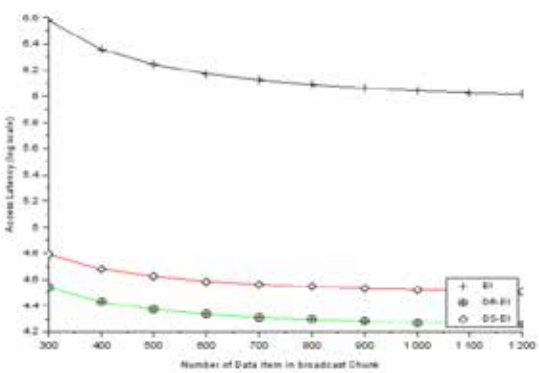

Figure 5: Effect of number of data items on access latency (I=5)

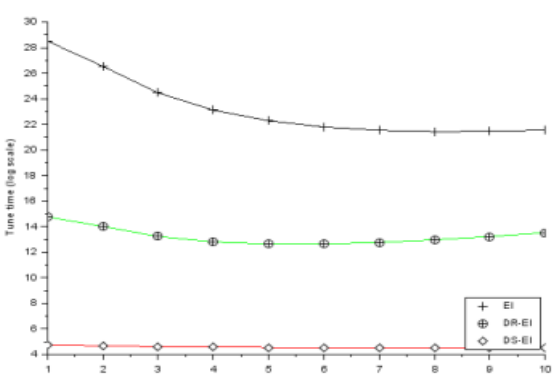

Figure 6: Effect of number of chunk size on tune time

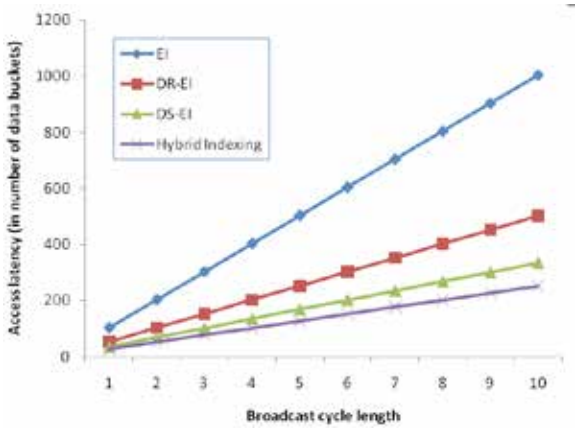

Figure 7: Variation in Access latency with broadcast cycle length

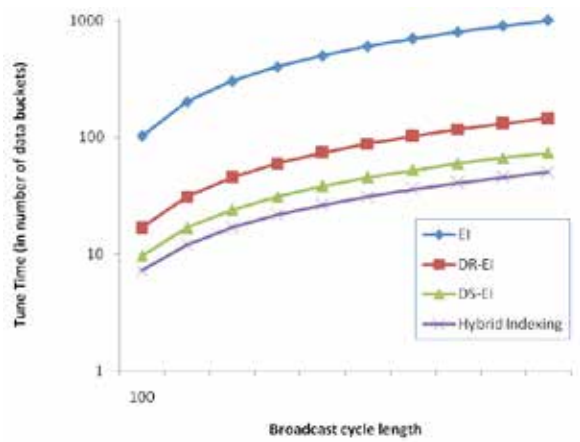

Figure 8: Variation in tune time with broadcast cycle length

Number of data items and chunk size are two important factors affecting the access time and energy consumption in broadcast system. The number of data items in broadcast cycle is important for comparing normalized performance of broadcast designs. The variation of access latency with chunk size for Distributed Spatial-Exponential Index, Data Replicated-Exponential Index (DR-EI) and Exponential Index (EI) for N=300 data items is presented in figure 3. Simulation results depicts that DS-EI has slight higher access latency than DR-EI but still it is lesser than EI. The variation of access latency (I=5) for these schema is presented in figure 5. The figure show similar trends. It simply means that DS-RI has reduced access latency for all size of chunk or broadcast cycle but it is still much lesser than EI.

Figure 6, shows the variation of tune time with chunk size. It reveals that for N=800 data items, DS-EI is good bidder than both DR-EI and EI. The tradeoff between access latency and tune time for proposed Hybrid Indexing is shown in figure 7 and figure 8. These figures show that proposed Hybrid Indexing outperform DS-EI, DR-EI and EI irrespective of the size of chunk or broadcast cycle length.

Conclusion

This research paper presents a model, which amalgamate DS-EI, which is an extension of Data Replicated-Exponential Index to distribute index with in global index with hash based Signature Indexing. The proposed model is examined through simulation study for different broadcast size and chunk size. The results of analysis model calculated for values of parameter settings contained in table 2 are presented in section V(B). The results show that the DR-EI has much lesser tune time than DR-EI and EI but its access latency is slighly greater than DR-EI but less than EI. Also Hybrid Indexing is

better option to put up a tradeoff between access latency and tune time. Hence proposed model reduce tune time and hence energy consumption considerably without putting any cost to access latency. From this discussion, it is clear that Hybrid Indexing is preferable index strategy, which gets benefit from bucket replication as well as distribution along with flexibility.

For further research, we are planning to examine the proposed model for multichannel with error prone broadcast environment.

References

1. T. Imielinski, S. Viswanathan and B.R. Badrinath, "Energy Efficient Indexing on Air," Proceedings of the ACM SIGMOD Conference, pp25-36, 1994.
2. 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.
3. S. Verma, Rakhee and S. Kumari, "Data Broadcast Management in Wireless Communication: An Emerging Research Area", Applied Signal and Image Processing: Multidisciplinary Advancements, IGI Publications USA chapter 4, pp 61-75, 2011.
4. S. Kumari, "DS–Exponential Index for Energy Efficient Broadcast Data", In the proceeding of National Conference on Advanced Computing Research (NCACR 2015), pp 214-216, 2015.
5. S. Verma, Rakhee and S. Kumari, "Data Replicated–Exponential Index to Reduce Access Latency", International Journal of Advances in Communication Engineering, 2(2), 2009.
6. S. Acharya, R. Alonso, M. Franklin and S. Zdonik, "Broadcast Disks: Data Management for Asymmetric Communication Environments," In Proceedings of ACM Sigmod, pp.199-210, 1995.
7. M. H. Ammar and J. Wong, "The Design of Teletext Broadcast Cycles," Performance Evaluation 5(4), pp 235–242, 1985.
8. W.G. Yee, S.B. Navathe, E.Omiencinski, and C. Jermaine, "Efficient Data Allocation over Multiple Channels at Broadcast Servers," IEEE Transaction on Computers 51(10), pp 1231-1236, 2002.
9. C. H. Hsu, G. Lee, and A. L. P. Chen, "Index and data allocation on multiple broadcast channels considering data access frequencies," in Proceedings of 3rd International Conference on Mobile Data Management (MDM 2002), pp 87-93, 2002.
10. 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.
11. A. Seifert, J. Hung, "FlexInd: A Flexible and Parameterizable Air-Indexing Scheme for Data Broadcast Systems," EDBT 2006, LNCS 3896, pp 902–920, 2006.
12. K. Park "Location-based grid-index for spatial query processing" Expert Systems with Applications 41(1) pp 1294–1300, 2014.
13. M. Vijayalakshmi and A. Kannan, "A Hashing Scheme for Multichannel Wireless Broadcast", Journal of Computing and Information Technology, 3, pp 197–207, 2008.
14. W. C. Lee and D. L. Lee, "Using Signature Techniques for Information Filtering in Wireless and Mobile Environments", DPDB 4(3), pp 205–227, 1996.
15. Q. L.Hu, W. C. Lee, and D. L.Lee, "A Hybrid Index Technique for Power Efficient Data Broadcast", DPDB-2001, 9(2), pp 151–177, 2001.
16. A.Y. Ho and D.L. Lee, "Data Indexing for Heterogeneous Multiple Broadcast Channel", In the proceeding of IEEE International Conference on Mobile Data Management (MDM 04), pp 274-283, 2004.
17. S. Verma, Rakhee and S. Kumari, "Dynamic Broadcast Scheduling At Using Unified Index Hub (UIH)", International Journal of Intelligent Information Processing, No. 2(2), pp 199-206, 2008.
18. S. Verma, Rakhee and S. Kumari, "Two Level Signature Model for Multiple Broadcast Channel Using Unified Index Hub (UIH)", 2nd International Conference on Soft Computing (ICSC-2008) , pp 422-431, 2008.