Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 4, Issue 1; January-March, 2017, pp. 85-88

http://www.krishisanskriti.org/Publication.html

Exploring Traditional Energy Efficient Indexing for Wireless Big Data

Savita Kumari Sheoran

Department of Computer Science & Applications Chaudhary Ranbir Singh University, Jind (India) –126102
Email: savita.sheoran@gmail.com

Abstract—*The radio network which is aged more than a century has witnessed a tremendous growth thruway wireless and cellular industry. Now we reach at a limit where the available architecture, bandwidth, broadcast and data management system are constrained. The enormous amount of data broadcasted over wireless networks reach to scale of 'Big data' and tradition data handling techniques are insufficient to disseminate query for such large data set through energy and time bounded approach. This paper carry out a unified study of resource constrained wireless network and recent large data handling technique Big data to find a solution for energy and time efficient data management technique. I have explored traditional indexing technique to address these issues and results were compared with state of art, which show that proposed model outperform in all situation.*

Keywords*: Indexing Big data, Wireless Big data, Resource constrained wireless network, Energy Efficient Indexing and Wireless Data Indexing.*

1. INTRODUCTION

The history of wireless is aged back to century age with the invention of first radio network by Heinrich Hertz in 1886. Since then it constantly grow to pervasive and gigantic network connecting desktop, palmtop and handheld devices. The communication over wireless network is asymmetric, heterogeneous and fragmented in channels and the available bandwidth is very scared i.e. only a few Giga Hertz is available worldwide. The wireless communication devices generate and guzzle data from social media, financial analysis and personnel interaction etc in the form of text, image and video etc. Until now the data set comprising these data are reckonable and traditional techniques are sufficient to manage it. But now the data has been surpassed the limit of volume, velocity and variety of traditional data to ensnare the range of 'Big data'.

The major shortcoming with broadcast data items in a wireless environment is that data are accessed sequentially. The increasing 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 ever 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 indexing the data broadcast has been proved effective, which can be visualized based upon 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.

Recently Radix Adaptive and Hash Tables based index were employed to solve this data complexity issue in wireless environment. But one of the drawbacks of these indexing is that they are not energy efficient. Since the traditional tree based indexing and signature based indexing have been used since a long in multi channel wireless environment. Here in this paper we have tested these two indexing to retrieve energy efficient timely retrieval of wireless big data. The experimental results show that these two indexing outperform the state of art Adaptive Radix based indexing. The rest of this paper is organized as follow: Section 2 gives related literature; section 3 and 4 represents traditional indexing techniques and concept of big data respectively. Section 5 represents results and finally section 6 concludes the paper.

2. LITERATURE REVIEW

Wireless computing allows mobile clients to enjoy unrestricted mobility and ubiquitous information access in a region called cell. With data broadcast, clients listen to a broadcast channel to retrieve data based on their queries and hence are responsible for query processing. Acharya et al (1995) has designed a broadcast disk for query processing in asymmetric environment, which is different from traditional environment. An item-based solution to the problem of broadcast query processing by providing air indexing was proposed by Imielinski et al. (1994). One major limitation of

such item-based schemes is their lack of semantics associated with the broadcast. Thus, it is hard for mobile clients to determine if their queries could be answered from the broadcast entirely, forcing them to contact the server for possibly additional items. To remedy this, a semantic-based broadcast approach was suggested by Lee et al. This approach attaches a semantic description to each broadcast unit, called a chunk, which is a cluster of data items. This allows clients to determine if a query can be answered based solely on the broadcast and to define precisely the remaining items in the form of a "supplementary" query. Mobile clients can move between cells while being active and the intercell movement is known as a handoff process. Query management in intercell environment is discussed by Imielinski et al. (1994) Waluyo, A.B. (2004) and Pfoser D. (2000) develop analytical model for query processing in moving objects and provides a comprehensive taxonomy of wireless query. Cai et al. (2002) have studied real time query delivery by involving the adaptability concept hierarchy. This paper develops collective model comprising query selection, classification and processing, which are essential components of wireless communication environment.

Suzhi B et al (2015) has introduced indexing wireless big data for boudless and free query dissemination over wireless environment. Jianguo M (2016) have addressed the design challenge and opportunity in the regime. Yang K et al (2015) and Schuh S et al (2015) have addressed the energy efficient big data retrieval in mobile environment.

3. BIG DATA

In 1997, NASA scientist used the term big data because they found difficulty with data set that is so large that they even do not get fit inside memory disk etc. They called that big data problem. It is an interesting challenge to manage such data. Then some American scientist in 2008 popularized the term big data because they predict that big data analysis can enable unlocking of valuable knowledge and helps in decision making in varieties of fields like medical, science, agriculture, industries etc. Big data cannot be considered in isolation .it is combination of various data management technologies that progresses over the time. Big data is all about large volume, diversified data that is captured from various data sources and analyzed at high velocity to provide valuable insights. In the past years big data size was in terabytes but today it is in petabytes and soon it will be in Exabyte (millions of terabytes) So It is necessary to understand big data from all aspects and dimensions.

Owing to the scientific decision making with great accuracy the Big data is gaining popularity as decision making tool in business. Fig, 1, depicts the popularity of data driven decision making process. Today's every activity is controlled by data. Every year we are producing new data, which is roughly equal to the data generated since dawn of civilization to 2003. The data storage is expanding exponential and we have entered in Zettabyte Data Age in 2010.

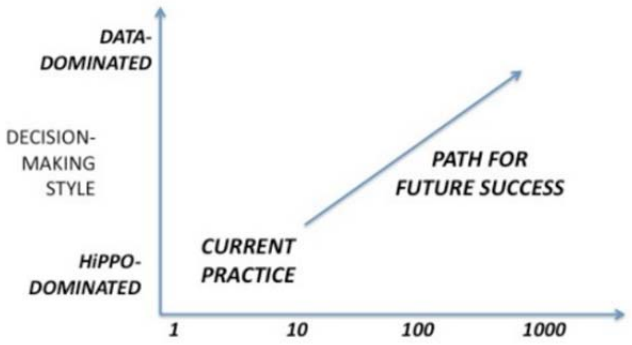


Fig. 1: Popularity of Data Driven Decision Making

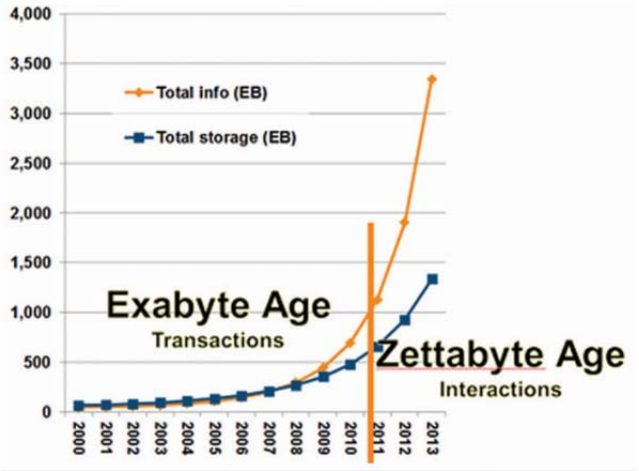


Fig. 2: Data Age Break-through

Since the Zettabyte is too large data size to be handelled with ordinary data management techniques. Different data size scale is depicted in Fig, 3. Owing to advance analytical cpacity and avaialble data storage capacity, Big data reseaches have increased leap and bound in last few years. Fig, 4, envisage the number of researches papers publiched on Big data. This increased reseach shows the testimony for capacity of Big data.

4. TRADITIONAL WIRELESS INDEXING

Traditionally we use two index structures for multichannel wireless broadcast environment viz: Index Tree Based Indexing and Signature Based Indexing. The subsequent subsections of this section will represent them in detail.

4.1. Index Tree Based Indexing

To reduce access time while maintaining similar tune in time the indexing techniques are using B-tree, B+-tree, R-tree, alphabetic huffman tree structure. The B+-tree has advantage over traditional B-tree and other trees because it contains data pointer stored only in leaf nodes. Hence it is easier for mobile client to differentiate between index and data buckets

contained in nodes. It organizes index and data on broadcast channel to accelerate the search processes. The use of B+-tree indexing in wireless environments is very similar to that of traditional disk based environments. However, there are some differences that introduce new challenges to wireless environments. For example, in disk based environments, offset value is the location of the data item on disk, whereas in wireless environments, offset value is the arrival time of the bucket containing the data item. The B+-tree index was first introduced in (1, m) indexing, which is fully replicated index (FRI), and distributed indexing, which is partially replicated index (PRI). We only discuss distributed indexing in this section, because all other are derived from it and have very similar data structure.

The index structure of B+-tree contains leaf nodes comprising bottom most indexes containing of up to 'k' keys all pointing to actual data while non-leaf nodes comprise upto 'k' keys and 'k +1' pointers to the leaf nodes on the next level of on the tree hierarchy. The scanning process requires frequent shifts between index and data nodes. The index nodes have two types on index pointers viz: local and control (global) while each data bucket has local pointer only. Each index bucket contains pointers that point to the buckets containing its child nodes placed on same channel. These pointers are referred to as local index. Since users, initially, may not tune to the index of desired data. In this case, more information is needed to direct users to other index segment containing the required information on any channel. These pointers are referred to as control index. The control index consists of pointers that point at the next occurrence of the buckets containing the parent nodes in its index path. The access method for retrieving data buckets with index tree technique consists of following three steps:

Initial probe: The client tunes into the broadcast channel and determines when the next index tree is broadcast.

Search: The client follows a list of pointers to find out the arrival time of the desired data frames. The number of pointers retrieved is equal to the height of the index tree.

Retrieve: The client tunes into the channel and downloads all the required data frames.

4.2. Index Tree Based Indexing

A signature of a data frame is basically a bit vector generated by first hashing the values in the data frame into bit strings and then superimposing them together through bitwise-OR operation (v) to form signature. It shares some similarities to the Bloom Filter commonly used in networking, hence widely used for information retrieval in broadcast. Access methods making use of signatures of data records are called signature indexing. 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 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 can be defined as:

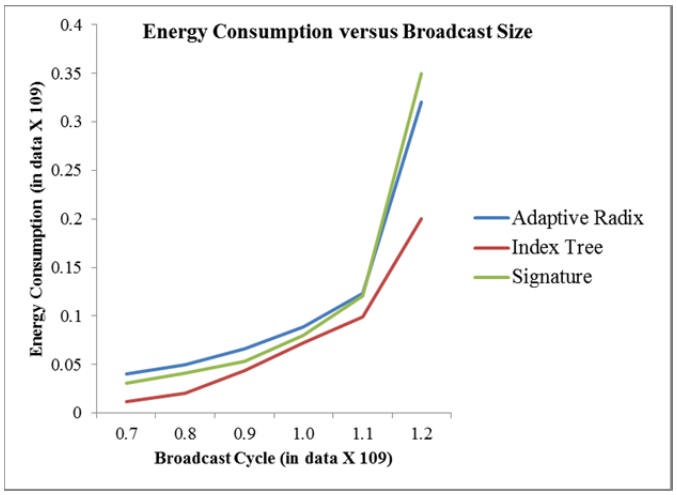
Simple signature: It is the simple type of signature which indexes the data frame which follows it on broadcast channel.

Integrated Signature: An integrated signature indexes a frame group of data instead of individual data. A frame group may contain any number of data records.

Multilevel Signature: The multilevel scheme is a combination of the simple signature and integrated signature schemes. It consists of multiple levels of signatures. Signatures at the upper levels are integrated signatures and those at the lowest level are simple signatures.

5. RESULT AND DISCUSSION

The simulation results for synthetic data to calculate Energy Consumption are presented in Fig, 3. For simplicity both the broadcast cycle length and energy consumption are represented in number of data size of billions data. However, it is too big amount of data but near real time we have to face in congested environment.



Fig, 3: Plot of Energy Consumption and Broadcast Cycle Length in term of Data items

As envisioned from plot in Fig, 3, the Index Tree based Indexing outperform both Adaptive Redix and Signature Indexing in all situation but Signature Indexing outperform for small broadcast length, rather prove weak for large broadcast length.

6. CONCLUSION

This research study explores two real time concepts of Indexing Wireless Data Broadcast and Big Data to harness their common potential for synthetic data. Since The Wireless data in real time start reaching to the level of Big Data in term of volume, velocity and variety. The simulation results are obtained and presented. The result show that traditional Tree

based index outperform the state of in all situation while Signature based indexing may prove good for small broadcast length. Since this type of work exploring wireless big data is at primary stage. The research study to look at these techniques in real time environment have much scopes and can revolute the field of Mobile and Wireless Computing.

REFERENCES

[1] Acharya S., Alonso R., Franklin M. and Zdonik S., "Broadcast Disks: Data Management for Asymmetric Communication Environments", In Proceedings of ACM Sigmod, pp.199-210, 1995.

[2] Barbara D., "Mobile Computing and Databases- A Survey", IEEE Transactions on Knowledge and Data Engineering, 11(1): 108-117, 1999.

[3] Urson A., and Jiao Y., "Data Broadcasting in a Mobile Environment, Wireless Information Highways: Chapter 3", IRM Press, pp. 96- 154, 2005.

[4] Lee W. C. and Lee D. L., "Using Signature Techniques for Information Filtering in Wireless and Mobile Environments", DPDB 4(3), pp. 205–227, 1996.

[5] Imielinski T., Viswanathan S. and Badrinath B. R., "Energy Efficient Indexing on Air", In the proceeding of the ACM Sigmod Conference, pp.25-36, 1994.

[6] Verma S., Rakhee and Kumari Savita, "Signature Model for Heterogeneous Multiple Broadcast Channel Using Unified Index Hub (UIH)", In the proceeding of NCRAET at Govt. Engg. College, Ajmer (Rajasthan) – India, pp. 45-46, 2008.

[7] Verma S., Rakhee and Kumari Savita, "Data Replicated – Exponential Index to Reduce Access Latency", International Journal of Advances in Communication Engineering, 2(2), 2009.

[8] Hu Q. L., Lee W. C., and Lee D. L., "A Hybrid Index Technique for Power Efficient Data Broadcast", DPDB-2001, 9(2), pp. 151–177, 2001.

[9] Hu Q.L., Lee, W.C., and Lee D.L., "Power Conservative Multi-Attribute Queries on Data Broadcast", In the proceeding of 16th International Conference Data Engineering (ICDE '00), pp. 157-166, 2000.

[10] Hsu C.H., Lee G., and Chen A. L. P., "Index and Data Allocation on Multiple Broadcast Channels Considering Data Access Frequencies", MDM 2002, pp. 87–93, 2002.

[11] Suzhi Bi ; Rui Zhang ; Zhi Ding and Shuguang Cui, , "Wireless communications in the era of big data," IEEE Communications Magazine, 53(10), 2015 pp: 190 – 199.

[12] Jianguo Ma and Haipeng Fu, "Wireless big-data: Opportunity and the design challenging," IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO) 27-29 July 2016.

[13] Kun Yang ; Qin Yu ; Supeng Leng ; Bo Fan and Fan Wu, "Data and Energy Integrated Communication Networks for Wireless Big Data" IEEE Special Section Big Data for Green Communications and Computing, Vol 4, 2015, pp: 713 – 723.

[14] Stefan Schuh and Jens Dittrich, "AIR: Adaptive Index Replacement in Hadoop," 31st IEEE International Conference on Data Engineering Workshops, ICDE Workshops 2015, Seoul, South Korea, April 13-17, 2015, pp: 22–29.