**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

***IJCSMC, Vol. 4, Issue. 8, August 2015, pg.375 – 380***

**RESEARCH ARTICLE**

# Access Time Efficient XML Data Scheduling Algorithm for Wireless Broadcasting

**P.Prabhavathy**[1] (Research Scholar), **S.Bose**[2]

[1,2]Department of CSE, Anna University, India

pprabhavathy@gmail.com, sbs@cs.annauniv.edu

***Abstract— Wireless broadcasting system is of major two types namely push-based and pull-based. Extensible Markup Language (XML) is used to represent any information due to its inherent potential features such as scalable, self-describing, etc. Hence XML is chosen as standard for information dissemination in wireless broadcasting systems. Push based system periodically broadcasts XML data over the wireless broadcast channel. But in pull based system, mobile users request are queued up through the uplink channel and based on the request queries, corresponding XML data will be broadcasted in the downlink channel. In this paper, an access time efficient algorithm is proposed based on XML document structural difference. Lian's measure is used to focus on mismatch in structure among XML documents and Minimal Difference First (MDF) algorithm uses this measure to organize the XML documents in the broadcast channel efficiently.***

***Keywords— access latency, periodic XML broadcasting, wireless data dissemination, Lian's measure, MDF scheduling***

## I. INTRODUCTION

Wireless broadcasting has received a lot of attention from mobile users. It has been available as commercial products for many years. StarBand [1] and Hughes Network [2] are some commercially available wireless broadcasting systems. The announcement of the smart personal objects technology (SPOT) by Microsoft [3] highlighted the corporate interest in and feasibility of utilizing broadcast for wireless data services. In DirectBand Network, FM radio subcarrier frequencies are used, which helps SPOT-based devices such as PDAs and watches to continuously receive timely information such as stock quotes, airline schedules, local news, weather and traffic information.

Access efficiency and energy conservation are two main performance issues for the mobile clients in wireless data broadcast system. Access efficiency deals with how fast a client request gets satisfied. Energy conservation tries to reduce mobile client's energy consumption during its access for data of interest. Wireless data broadcasting systems are high scalable with flexibility to switch between time efficiency and energy efficiency when disseminating public information, which makes it an interesting research area. In data broadcasting systems, base station consecutively broadcasts a series of data items named broadcast program onto the air through Radio Frequency signals. Any number of subscribers within the valid broadcasting region could tune in to the broadcasting channel to search for and download data items with their mobile devices like smart phones.

 

**Access latency:** The time elapsed from the start of a query issue till the completion of its response.
**Tuning time:** Time duration during which mobile client is in active mode to download requested data.

These performance metrics, namely access latency and tuning time, have been used to measure access efficiency and energy conservation, respectively [4, 5, 6,17].

The rest of this paper is organized as follows: Section 2 explains the existing works in the scheduling and indexing of wireless broadcasting systems. Section 3 provides the explanation of periodic XML broadcasting system. Section 4 explains how access time is efficiently managed using Lian's measure and MDF algorithm in scheduling the wireless XML broadcast. Finally, this paper concludes with Section 5.

## II. RELATED WORKS

In this mobile era, there is an unprecedented demand for on-line data services. Many applications require timely data delivery from information producers to thousands of information consumers. Internet is evolving towards an information exchange XML standard that incorporates technologies, including wireless, mobile, and hybrid networking [7, 17]. Due to broadcast, a data item interested by many clients can be disseminated in dedicated public channel called a broadcast channel. Mobile clients can tune in to retrieve the desired items by listening to the broadcast channel. This wireless data broadcast technique significantly saves bandwidth required for delivering same data item many times to individual mobile clients through point-to-point wireless channels [8, 17].

In the paper [9] pull based wireless broadcast system is used and proposed scheduling algorithm for time critical non-XML data. Algorithms are devised with their performance comparison, but this is not suitable for XML based wireless broadcasting systems. Multiple channels are considered for broadcasting purpose. The data on air is susceptible to privacy attacks. Broadcasting data systems used in military requires data privacy during its transmission in wireless channel. Heavily loaded encryption algorithms are not suitable for wireless broadcasting systems.

So a light-weight methodology for privacy concern is proposed which optimally balances the performance of wireless data broadcasting is proposed in [10]. Greedy algorithm is devised in [11] for on demand or pull based XML broadcast and compared with random data placement algorithm. In on demand wireless data broadcasting systems, requests from mobile clients are gathered using low bandwidth uplink channel. The desired data for all clients will be broadcasted in high bandwidth downlink channel.

Multiple broadcast channels are considered for on demand system in [12] for access time efficiency. The wireless broadcast channel is logically divided into many channels for carrying replicated index information and clustered data based on frequency access pattern for the desired data. Cost based techniques are proposed for scheduling on-demand broadcast in [13]. Some broadcasting links will have very low bandwidth when compared with others in multi channel wireless broadcasting. Rescheduling such worst bandwidth links first is considered in [14] for cloud environments.

The queries from mobile users in pull based system may need more than one data item as response. Scheduling for such multi –item mobile user queries is proposed in [15]. Measurement of dissimilarity in XML document structure is proposed in [16] as Lian's measure. In [18] various structural similarity measures such as jaccord measure, dice measure are compared for XML document structure. A deep survey of various structural and distance based similarity algorithms are given in the paper [19].

The syntactical similarity is must for query processing of XML documents. XPath query has to match with the syntax structure of XML document to provide the query result. This is analyzed and experimentally proven in the paper [20]. Push based XML broadcasting using compact indexing and clustering by structural similarity of XML documents is proposed in [21].

## III.WIRELESS XML BROADCASTING

As Internet and mobile communication grows with rapid development, wireless data broadcasting technology has become increasingly an important research. XML is the standard used for data exchange over distributed environment, which is extensively used by researchers in a wireless environment. Wireless XML broadcasting offers great scalability, low power consumption and efficient bandwidth utilisation.

 

Existing wireless broadcast researches focus on broadcasting data items with unique key values. The requests are key-based queries and the indexing methods are also key-based for non-XML data dissemination. Traditional air indexing techniques and broadcast schedule techniques cannot be applied to XML broadcasting because XML data are semi-structured in nature unlike key-based data. So to retrieve information from the XML broadcasting, mobile users must know the XML document structure and its data organisation over the wireless channel.

Air indexing and scheduling are the two major features in wireless broadcasting whose efficiency is evaluated using tune-in time and access time performance metrics respectively. Power conserving wireless broadcasting systems focus on tune-in time metric for optimizing air indexing technique for selective tuning of information by mobile clients.

Time critical wireless broadcasting systems which disseminates stock and travel traffic information focus on access time for providing quick response to mobile clients. Since wireless broadcasting systems are prone to error in data transmission, fault tolerant schemes can be enforced along with the air indexing and scheduling techniques. In critical systems such as military purpose, wireless data has to be securely transmitted for the concern of privacy.

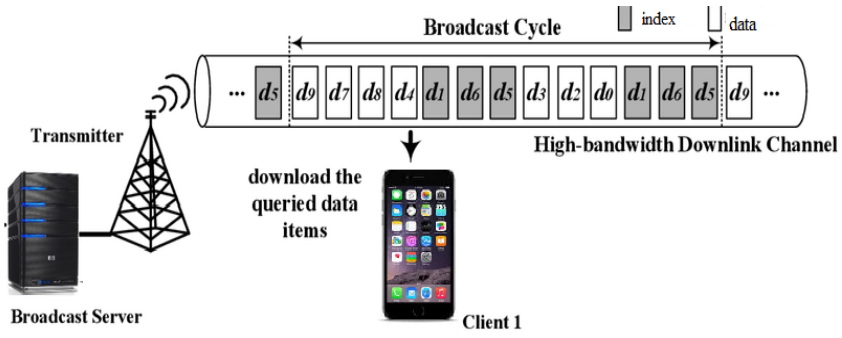


Fig. 1. Wireless data broadcasting system

The general system architecture for wireless data broadcasting system is provided in figure 1. Let us consider an example XML document and how its structure is exploited in broadcasting program and air indexing. Figure 2 provides the book XML document which is used in this paper for example purpose.

```
<?xml version="1.0"?>
<x:catalog xmlns:x="urn:books">
  <book id="bk101">
    <author>Gambardella, Matthew</author>
    <title>XML Developer's Guide</title>
    <genre>Computer</genre>
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at creating applications with
    XML.</description>
  </book>
</x:catalog>
```

Fig. 2. An example XML document

XML documents related to book information will have the structure depicted in Figure 2. Hence all XML documents having same structure can be combined with list of tag values. The document object model of XML is a tree structure which is exploited by scheduling techniques for access efficient XML broadcasting.

 

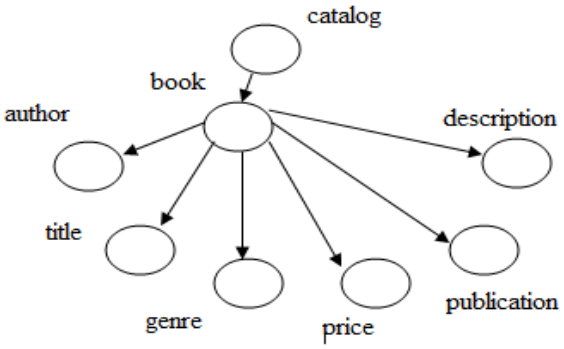


Fig. 3. DOM structure of example XML document

The structural characteristic of XML documents can be measured using various metrics. They are jaccard measure, dice measure, Lian measure and cohesion measure. These measures are used to find structural similarity among XML documents that has to be broadcasted in wireless channel. The value of these metrics always lies in the range of 0 to 1. The value 0 indicates no similarity whereas 1 indicates same structural similarity in XML documents.

Lian's measure is different from these metrics. It is used to focus on difference in the structural similarity of XML documents. More structurally different XML documents will be peculiar or outstanding in nature. Hence request for such documents will be meager from mobile clients. So this measure is utilized for scheduling rare or infrequently accessed data in the broadcast cycle for access efficiency.

Broadcast cycle is the set of data in particular sequence transmitted in the broadcast channel during unit of time. It is generated using broadcast program. Scheduling technique helps to implement broadcast program for wireless data broadcasting. The set of XML documents to be broadcasted is denoted as D, where di and dj are two XML documents and di , dj belongs to D. PS(di) and PS(dj) are the pathsets of XML documents di and dj respectively. Lian's measure for finding structural similarity among two XML documents namely di and dj is given below in the equation 1.

$$L(di , dj) = | PS(di) \cap PS(dj) | / \max\{ | PS(di) | , | PS(dj) | \} \quad (1)$$

Then a distance metric can be computed from this Lian's measure as given in equation 2.

$$dist(di , dj) = 1 – L(di , dj) \quad (2)$$

For example, pathset of book XML document is {\catalog, \catalog\book, \catalog\book\author, \catalog\book\price, \catalog\book\genre, \catalog\book\ title, \catalog\book\ publication, \catalog\book\description}

## IV. SCHEDULING THE MINIMUM DIFFERENCE FIRST (MDF)

All XML documents in the set D are compared with each other to find Lian's measure value. Distance measure got from Lian's value yields the difference among two XML documents. Documents whose distance measure is approximately same are grouped to form a cluster. Thus many clusters are generated from the set of XML documents D. These clusters are arranged in an order using MDF algorithm to achieve access efficient wireless XML broadcasting.

Wireless broadcasting system will periodically broadcast these clusters in the order of Cluster 1, 2, 3,….n. Ideally, queries from mobile clients only need XML documents within the same cluster. Consider arrival time tq of query q which needs only XML documents in Cluster 1. At time tq, there are 4 cases of on-going broadcasting:

(1) system is about to broadcast Cluster 1;

(2) system has just finished broadcasting Cluster 1;

 

(3) system is broadcasting Cluster 1;

(4) system is broadcasting other clusters.

In fact, case (1) and case (2) rarely happens because they are actually special cases of case (4). So consider only case (3) and case (4). In case (3), the access latency is possibly worse due to query q missing XML documents in Cluster 1, which needs to wait a whole broadcast cycle to get it. As result, access latency can be considered the same as length of broadcast cycle, denoted as LenBC. In case (4), expected access latency can be

$$(\text{LenBC} - \text{Length of Cluster1})/\ 2 + \text{Length of Cluster1} \qquad (3)$$

From the equation 3, it is clear that clustering helps to reduce access latency which in turn yields better access efficiency in wireless broadcasting.

**MDF Scheduling Algorithm:**
Input: n clusters $\{C1, C2, \ldots, Cn\}$
Output: Broadcast program of XML documents on air
1) Find cluster with least Lian's measure valued XML documents, denoted as $C_L$ which will be the first cluster to be broadcasted.
2) Determine next cluster $C_{next}$ which has maximum distance with $C_L$, calculated by $dist(C_{next}, C_L)$. $C_{next}$ will be broadcasted immediately after $C_L$.
3) Assign $C_L$ with $C_{next}$
4) Repeat the above steps until all clusters have been processed.
4) Record the order of all n clusters and broadcast them one by one periodically.

Fig. 4. MDF scheduling algorithm for access efficient broadcasting

*A. Experimental setup*

Windows 7 Operating system is used and implementation of algorithm is done using java programming. All experiments are run using synthetic data set with almost 250 XML documents defined by News Industry Text Format (NITF) DTD. Average depth of all documents is between 5 and 9 while maximum depth could be up to 12. The probability of XPath queries having wildcards (*) and // axes are varied from 9 to 27% for experimentation purpose. Our MDF scheduling technique is compared with random scheduling to emphasize its performance enhancement in access efficiency during wireless XML broadcasting.

*B. Performance analysis*

Random method is compared with our algorithm for the above mentioned experimental setup. Random placement means to broadcast XML documents in a random order. Figure 5 shows the performance of our MDF scheduling algorithm when adjusting probability of wildcards and axes from 9% to 27%.

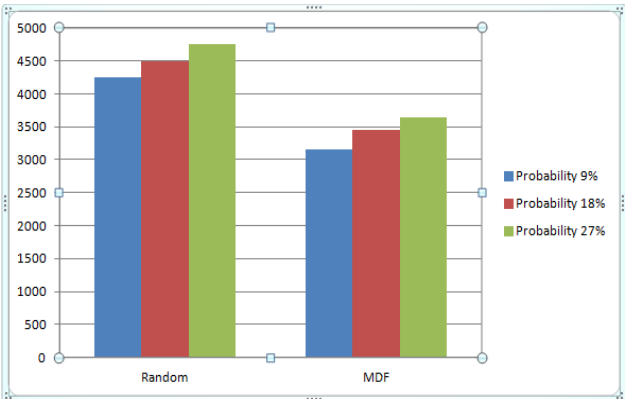


Fig. 5. Adjusting probability of * and // in queries for analysing performance

 

From this figure 5, it is clear that access latency goes up for these methods when probability increases. But our MDF scheduling method outperforms random scheduling by efficiently accessing the wireless XML broadcasting channel.

## V. CONCLUSIONS

A novel Minimum Difference First (MDF) scheduling algorithm is devised to enhance the access efficiency of periodic wireless XML data broadcasts. It takes advantage of structured characteristics of XML data and no prior knowledge from mobile clients is required before starting XML data schedule. Pathset is defined by finding all full paths and sub paths of all XML documents.

We apply Lian's measure on these pathsets to measure the similarity between XML documents. Using the similarity measure and the distance metric analysis, we devised a novel MDF scheduling algorithm to effectively organize XML data on wireless channels for periodic broadcasts.

Our experiments demonstrate that the proposed MDF scheduling algorithm achieves better access efficiency when compared with random scheduling. Fault tolerance feature can be coupled with scheduling to have reliable and access efficient wireless XML broadcasting system in future.

## REFERENCES


[1] StarBand. Available at http://www.starband.com/.

[2] Hughes Network Systems. Available at http://www.direcway.com/.

[3] DirectBand Network. Microsoft Smart Personal Objects Technology (SPOT). Available: http://www.microsoft.com/resources/spot/.

[4] Q. L. Hu, W.-C. Lee, and D. L. Lee, Power conservative multi-attribute queries on data broadcast. In Proceedings of 16th International Conference on Data Engineering, pp: 157-166, San Diego, CA, February 2000.

[5] T. Imielinski, S. Viswanathan, and B. R. Badrinath. Power effcient filtering of data on air. In Proceedings of 4th International Conference on Extending Database Technology, pp: 245-258, Cambridge, UK, March 1994.

[6] T. Imielinski, S. Viswanathan, and B. R. Badrinath. Data on air: Organization and access. IEEE Transactions on Knowledge and Data Engineering (TKDE),Vol 9(3),pp:353-372, May/June 1997.

[7] Acharya S, Franklin MJ, Zdonik SB. Balancing push and pull for data broadcast. In Proceedings of 1997 ACM SIGMOD International Conference on Management of Data, Tucson, AZ,USA, May 13-15, 1997, pp. 183–194.

[8] Acharya S, Alonso R, Franklin MJ, Zdonik SB. Broadcast disks: Data management for asymmetric communications environments. In Proceedings of 1995 ACM SIGMOD International Conference on Management of Data, San Jose, CA, USA, May 22–25, 1995, pp. 199–210.

[9] J. Xu, X. Tang, and W.-C. Lee, "On Scheduling Time-Critical On-Demand Broadcast", IEEE Transaction on Distributed and Parallel Systems, Volume 17, Issue 1, Jan. 2006, pp. 3-14.

[10] Q. Tan, W.-C. Lee, B. Zheng, P. Liu, and D. Lee, "Balancing Performance and Confidentiality in Air Index", ACM Fourteenth Conference on Information and Knowledge Management (CIKM'05), Breman, Germany, October 2005.

[11] Yongrui Qin, Hua Wang, Jitian Xiao, Effective Scheduling Algorithm for On-Demand XML Data Broadcasts in Wireless Environments, In 22nd Australasian Database Conference (ADC 2011), Perth, Australia, January 2011.

[12] Ta-Chih Su, Chuan-Ming Liu, On-demand data broadcasting for data items with time constraints on multiple broadcast channels, in Proceedings of 15th international conference on Database systems for advanced applications, pp 458-469, Springer-Verlag Berlin, Heidelberg ,2010

[13] W. Sun, W. Shi, B. Shi, and Y. Yu, "A cost-efficient scheduling algorithm of on-demand broadcasts," Wireless Networks, vol. 9, no. 3, pp. 239–247, 2003.

[14] Y. Huang, Y. Zhang, and J. He, "Optimizing broadcast completion time by rescheduling worst-bandwidths-links first in clouds environments," Journal of Computer and System Science, 2010

[15] W. Sun, Z. Zhang, P. Yu, and Y. Qin, "Efficient data scheduling for multi-item queries in on-demand broadcast," in EUC (1), 2008, pp. 499–505.

[16] Wang Lian, David W. Cheung, Nikos Mamoulis, and Siu-Ming Yiu,An Efficient and Scalable Algorithm for Clustering XML Documents by Structure, IEEE transactions on knowledge and data engineering, vol. 16, no. 1, january 2004,pp.1-14

[17] J. Xu, D.-L. Lee, Q. Hu, and W.-C. Lee, Handbook of wireless networks and mobile computing. John Wiley & Sons, Inc., 2002, pp. 243–265

[18] S. Helmer, "Measuring the structural similarity of semistructured documents using entropy," in VLDB, 2007, pp. 1022– 1032.

[19] D. Buttler. A short survey of document structure similarity algorithms. In 5th Int. Conf. on Internet Computing, Las Vegas, Nevada, 2004.

[20] D. Rafiei, D. L. Moise, and D. Sun, "Finding syntactic similarities between xml documents," in DEXA Workshops, 2006, pp. 512–516.

[21] Yongrui Qin, Quan Z. Sheng, Muntazir Mehdi, Hua Wang, Dong Xie, Effectively Delivering XML Information in Periodic Broadcast Environments, 24th International Conference, DEXA 2013, Prague, Czech Republic, August 26-29, 2013.pp 165-179


 