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. 3, Issue. 5, May 2014, pg.1061 – 1067

RESEARCH ARTICLE

Dissemination of Xml Data in Wireless Environment Supporting Twig Pattern Queries

M.S.Namose[1], R.H.Kulkarni[2]

[1]ME Computer, BSCOER, Pune, India

[2]Assistant Professor, BSCOER, Pune, India

[1] mahadevi.sggs@gmail.com; [2] rkpv2002@gmail.com

Abstract— The main aim of this paper is to improve energy and latency efficiency of XML dissemination scheme for the mobile computing, which is based on Lineage Encoding, G-node and scheduling algorithm for streaming XML data in the wireless environment. In this paper we propose a new broadcasting scheduling algorithm Frequently Access First (FAF) which effectively organize XML data on wireless channels. In this way the global query processing time must also be reduced to provide fast response to the users. The benefits of the structure indexing and attribute summarization that integrate relevant XML elements into a group and contribute a way for careful access of their quality values in a dynamic way that broadcasting can be done dynamically supporting Twig Pattern Queries. As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung "This paper presents a novel unit structure called G-node for streaming XML data in the wireless system."

Keywords— Twig pattern matching; Structure Indexing; Attribute Summarization; Lineage Encoding; Twig Pattern

I. INTRODUCTION

With the rise in popularity of wireless environment, wireless information scheme are currently being realized in many areas. Wireless technology has become deeply embedded in everyday life. At the end of 2011, there were 6 billion mobile subscriptions, estimated by the International Telecommunication Union (2011). That is equivalent to 87 percent of the world population, and is a huge increase from 5.4 billion in 2010 and 4.7 billion mobile subscriptions in 2009. According to ABI research, mobile application revenue hits 8.5 billion in 2011, and it will reach an estimated 46 billion in 2016. Obviously, the market for mobile applications will become "as big as the internet" in the near future[1]. Besides the traditional flat data information, such as records in relational databases, more and more information are described in semi-structured format over the past few years. XML has rapidly gained popularity as a standard to represent semi-structured information, and is also considered an effective format for data transmission and exchange[3].

The most Internet browsers provide support for XML in their later versions. In the near future, XML documents, like HTML Web pages, might become a part of our daily life. Consequently, XML has attracted

attentions from database community recently and there has been a large body of research work focusing on XML, such as XML filtering, querying and indexing All the major IT companies (e.g., Microsoft, Oracle, and IBM) have integrated XML into the software products which further demonstrates the popularity of XML**.** [1].
As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung "The tuning time and access time are used to check the energy-efficiency and latency-efficiency in wireless data dissemination. The *Tuning Time* is the sum of the elapsed times spent by a client to download the required data. When a client downloads the data it consumes more energy than when it waits for data. Thus, the tuning time is used as a performance measure for energy-efficiency. The *Access Time* is the time used when a client tunes in to the broadcast channel to when the relevant data is completely retrieved from the wireless stream. We define a streaming unit of a wireless XML data stream, called G-node. We propose a light weight encoding technique, called Lineage Encoding, to show the parent-child relationships among XML elements in the G-nodes. We also introduce relevant functions and operators that exploit bit wise operations on the lineage codes. This technique is the first wireless XML streaming approach that completely supports twig pattern query processing in the wireless broadcast environment."

We propose to pre-process XML data before broadcast to reduce both tuning time and access time. Taking advantage of the structured characteristics of XML data. Also we propose data broadcast scheduling algorithm **Frequently Access First(FAF)**. Broadcast schedule determines what data items to be broadcast by the server and also the order of data items on wireless channels. Access time and Tuning time are used to measure the system's performance. Data placement mainly affects access time because tuning time depends on the total content downloaded by mobile clients but not on the order of data. Hence, we use access time as our metric in this analysis[3].

II. **BACKGROUND**

A. XML Query Language

XML query language is a markup language for documents containing structured information. Originally designed to meet the demanding of large-scale computerized broadcastin.XML is also performing a highly important role in the transaction of a wide variety of data. The increasing popularity of XML is partly due to the limitations of the other two technologies are Hypertext Markup Language (HTML) and standard generalized markup languages (SGML), for represents the structured and semi-structured documents[2].

B . XML Data Model

XML documents have a hierarchical structure and can be represented as a rooted, ordered, and labeled tree. These XML trees (twigs) are available in two forms; they are ordered and unordered XML trees. The present approach considers an ordered and labeled XML tree. XML document must contain a root element. This element is the parent of all other elements. All elements in an XML document can contain sub elements, text and attributes. The tree represented by an XML document starts at the root element and branches to the lowest level of elements. The nodes of the XML tree represent elements and the edges represent parent-child relationships among XML elements. The XML tree representation for the sample XML document is given in Fig. 2. In this paper, we use XPath as the query language. The XML query language namely XPath represent the query as ordered labeled small trees (twigs). A twig pattern query consists of two or more path expressions and it represents complex search condition. For example, twig pattern query written in XPath format and the tree representation is given in Fig. 3[4].

C. Structure Indexing

As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung "The structure indexing directly captures the structural information of XML documents and is used for XML query processing"

D. Attribute Summarization

As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung "Attribute Summarization eliminates repetitive attribute names in a set of elements when generating a stream of G-nodes."

E. Structured characteristics

Means that we can define exactly how the data is to be arranged, organized and expressed within the file. When we are given a file, we can validate that it conforms to a specific structure, prior to importing the data. As we know the structure of the file in advance, we know what it contains and how to process each item. Prior to XML, the only structure in a text file was positional – we knew the bit of text after the fourth comma should be a date of birth – and we had no way to validate whether it was a date of birth, or even a date, or whether it was in day/month/year or month/day/year order.

F. Drawbacks Of Previous Work

In this section we discuss research work related to efficiently process XML twig pattern queries. Conventional XML query processing methods cannot process XML twig pattern queries efficiently since they do not contain branching information or parent-child relationships. Previous work focus only on the problem of efficiency and scalability. None of these methods focus on the energy-efficiency issue. Example: Multipredicate merge join (MPMGJN) algorithm[14]. Several approaches have been proposed for energy and latency efficient XML query processing in the wireless mobile environment. S-node generates an XML data stream based on the unit for XML broadcasting, called a S-node[15].

```
    <Studentinfo>
    <student id =”1015511”>
    <personalinfo id=”2015511”>
    <studname id=”3015511”personalinfo=”2015511”>
    <firstname id=”4015511”> preethi </firstname>
    <lastname id=”4015512”> k</lastname>
    </studname>
    <rollno id="3015512" personalinfo=”2015511”> 11</rollno>
    <email id=”3015513” personalinfo=”2015511”> preethi@gmail.com
    </email>
    </personalinfo>
    <result id=”2015512”>
    <subject1 id=”3015514” name=”maths” student=”1015511”
    result=”2015512”>
    <theorymark id=”4015513”>80</theorymark>
    <pracmark id=”4015514”>20</pracmark>
    </subject>
    <subject2 id=”3015515” name=”science” student=”1015511”
    result=”2015512”>
    <theorymark id=”4015515”>70</theorymark>
    <pracmark id=”4015516”>20</pracmark>
    </subject>
    </result>
    </Student>
</studentsinfo>
```

Fig 1 :A Sample XML Document(Extracted from ref no-4)

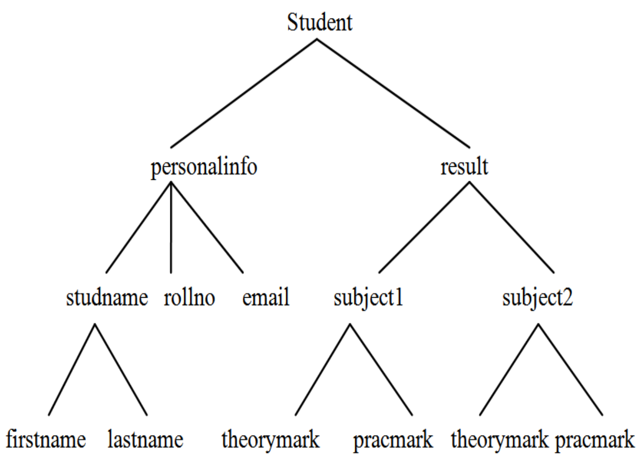

Fig 2 XML Tree Representation(Extracted from ref no-4)

Q1: //studentinfo/student [name/text () =”predicate”]/result

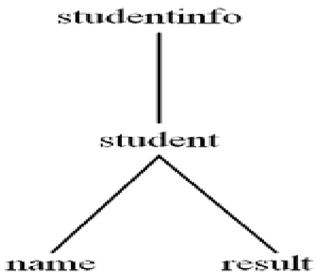

Fig 3 Example twig pattern query

III. PRE-PROCESSING XML DATA

Mobile users who have submitted queries that match some parts of this XML file all need to download the whole XML File from the wireless channel. It obviously brings down the overall access efficiency and energy efficiency. Therefore, in order to enhance the broadcast system performance, a preprocessing phase is needed in XML data broadcasts. Thus once a mobile user query matches some part of an XML file, the whole XML file will be broadcast by the server. By taking the structured characteristics of XML data into consideration, the server regards XML files as non- atomic items and all XML files are dividable. The server broadcasts every XML snippet that Satisfies any of the mobile users queries[3].

Innovative Scheduling algorithm: Frequently Access First(FAF*)*

First of all, we introduce some notation -

di : a data item (XML file) stored in the server

D: the set of data items that will be broadcast.

D = *d*1*; d*2*; : : : ; dm*

1.Initaly set Acc_No =0 for all item d from data set D.

2. select an item *d* from data item set *D* which has the *highest Acc_No*

3. place item *d* in the First position of broadcast schedule . If item d satisfies client query then update Acc_No=Acc_No+1

4. remove item *d* from *D*

5. update *Acc_No* for every item *d* in *D*

6. repeat step 2 to step 5 until *D* becomes empty

Based on this definition, when we schedule a new broadcast cycle, we first examine the Acc_No of each data item and then select out the data item which having highest Acc_No,plce that data item first position in broadcast cycle. suppose the data item set *D* contains *m* items, then step 2 and step 5 both take $O(m)$ time and step 2 and step 5 will both repeat *m* times. There- fore, the computing complexity of LEL scheduling algorithm is $O(m^2)$. First of all we pre-process given XML file then apply Lineage Encoding technique. In this way the global query processing time must also be reduced to provide fast response to the users.

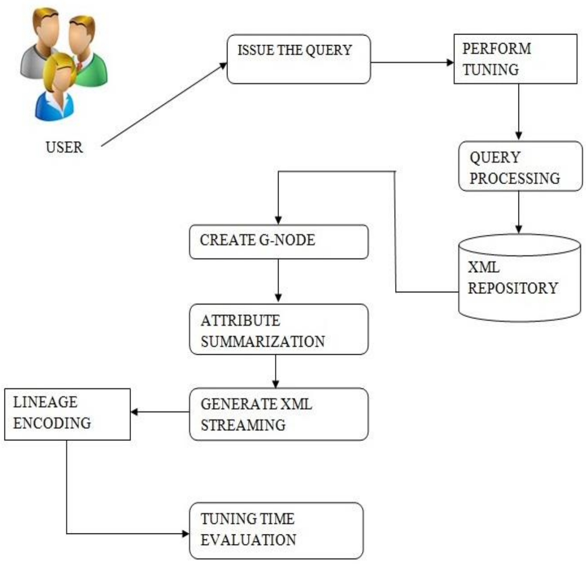

Fig .4 Architecture Diagram For Lineage Encoding

IV. LINEAGE ENCODING

A. G-Node:

As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung **"**Wireless XML stream consist of sequence of integrated G-Node. G-node integrates information of elements of the same path. G-Node is a one type of data structure, denoted by Gp=(GDp,AVLp, TLp) containing information of all the elements ep whose location path is p where GDp= is a group descriptor of Gp AV Lp= is a containing all text contents of ep."

B. Lineage Encoding

As per author Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung *"*This is the first proposed wireless xml streaming approach. Lineage Encoding is a light-weight encoding scheme. This system consist of mainly two parts 1) vertical code denoted by Lineage Code(V) 2) horizontal code denoted by Lineage Code(H). This system support twig pattern query. Twig pattern is a one type of tree patterns used in optimization process. Twig Pattern query used to represent parent-child relationships among XML elements in two G-nodes."

C. Twig Pattern Query Processing

As per author K. Preethi, S. Ganesh Kumar *"*Twig pattern query processing consists of three phases namely tree traversal phase, sub path traversal phase, and main path traversal phase. While processing a twig pattern query with predicates, we should select subset of elements satisfying the given predicates. Then, for the selected elements, we should find their parent elements or child elements. For example, to process the query in Fig. 3,
Q1://studentinfo/student[name/text()="predicate"]/result,we should find a subset of "name" elements satisfying the given predicate condition, select their parent "student" elements, and then identify "result" elements which are children of those "student" elements. A subset of the elements selected in a G-node can be represented by a bit string, called a selection bit string (SB) for the G-node where 1-value bits identify the selected elements. First, a function to obtain a selection bit string identifying a subset of elements in a particular child G-node is defined as follows:

Function 1
V_p=Shrink&Mask(V,SB_m)
SB_n=Unpack(V_p,H)

A selection bit string SBn for child G-node N can be computed based on the Lineage Code of N (V, H) using Shrink&Mask and Unpack operators in order. Shrink&Mask (V,SBm), where V denotes LC (V) of child G-node and elements in m with one or more child elements in N are selected by SBm. Shrink&Mask operator computesVp by shrinking 0's in LC (V) and then it shrinks SBm by eliminating the bits in same position as those removed in V. Unpack(Vp, H), where Vp is the shrunken bit string computed by Shrink&Mask operator and H denotes LC (H) of child G-node. Unpack operator extends Vp based on H to obtain the result selection bit string for the G-node N. Thus a subset of elements in a particular child G-node is selected. Second, a function to identify the parent elements of a subset of elements selected in a G-node is defined as follows:

Function 2:
V_p=Pack(SB_n,H)
SB_m=Expand&Mask(V,V_p)

To identify the parent elements of a subset of elements selected in a G-node, Pack and Expand&Mask operators are used. Pack(**SB_n**,**H**) operator computes **V_p** by shrinking the bit string **SB_n** based on H. V_p denotes the elements in the parent G-node of N which are parents of the elements in N selected by **SB_n** . Expand&Mask (V, V_p) operator expands V_p and masks V_p with it to obtain the result selection bit string SB_m for parent G-node of N. Thus selection bitstring to identify a subset of elements in a particular child G-node and to identify the parent elements of a subset of elements selected in a G-node is found. Finally, we define a function GetSelectionBitString (J) to select elements in a Gnode contained in the query tree of a given twig pattern query, which satisfy all the branching paths and predicate conditions in the sub-tree. The selection bit string SBj for J can be computed by performing bitwise AND operations over all the selection bit strings SBm obtained from the child nodes of J where J is a G-node in the query tree T[4]."

D. Sub path and Main path Traversal Phase

As per author K. Preethi, S. Ganesh Kumar **"**The main path denotes a path from the root node to a leaf node which represents the target element of the query and the sub path denotes branch paths excluding the main path in the query tree. The mobile client enters query and decryption key into the application. The query is then modeled into a query tree. In the tree traversal phase, the query tree is traversed in a depth-first manner; it selectively downloads components of the relevant G-nodes into the nodes in the query tree. Attribute values and texts involved in the given predicates are decrypted using the decryption key and downloaded into the relevant nodes. In the Subpaths traversal phase, the mobile client performs a postorder depth-first traversal starting from the highest branching node in the query tree using the GetSelectionBitString () function. In the sub path traversal phase each sub path is explored from the leaf node. Thus, the selection bit string for the branching node is calculated from all the sub paths in a bottom-up manner using Pack and Expand&Mask operators. Finally, the Main path traversal phase propagates the selection bit string on the branching node along the main path using Shrink&Mask and Unpack operators. Finally, the mobile client retrieves the required data which satisfies the given twig pattern query and with the help of encryption technique [4]".Fig 4 is the Architecture Diagram For Lineage Encoding, Fig 5 display the XML tree, this is the first step query processing. Fig 6 is the output screen of Lineage encoding technique, in which we show the number of nodes traverse.

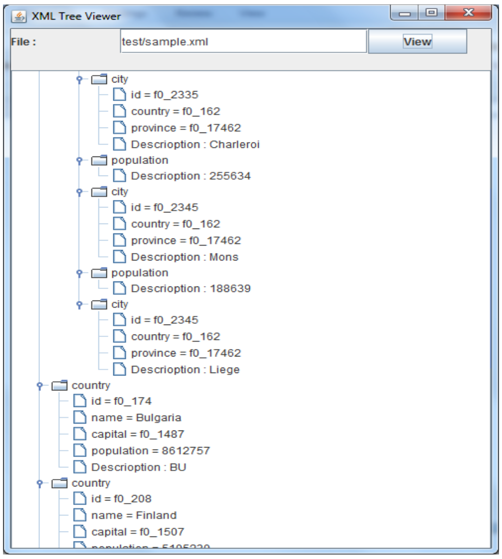

Fig 5. XML Tree

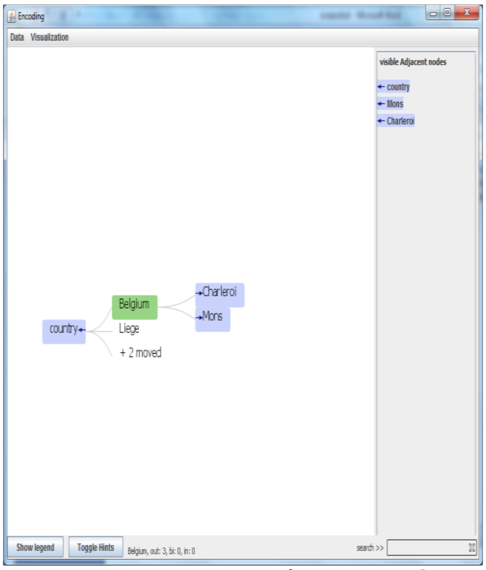

Fig 6. Output Screen Of Lineage Encoding

V.CONCLUSION

Twig pattern queries containing complex conditions are popular and critical in XML query processing. In this paper, we propose an efficient wireless XML streaming method supporting twig pattern queries. We defined Lineage Encoding scheme and relevant operators to efficiently process twig pattern queries and for selective access of XML elements as well as their attribute values and text. We also propose a one broadcasting scheduling algorithm Frequently Access First(FAF) which effectively organize XML data on wireless channels. In this way the global query processing time must also be reduced to provide fast response to the .Thus the mobile client can retrieve the required data more efficiently. First of all we pre-process given XML file then apply Lineage Encoding technique. In this way the global query processing time must also be reduced to provide fast response to the users.

REFERENCES

[1] Weiwei Sun, Yongrui Qin, Jingjing Wu, Baihua Zheng, Zhuoyao Zhang, Ping Yu, Peng Liu, Jian Zhang" Air Indexing for On-Demand XML Data Broadcast" Parallel and Distributed Systems, IEEE Transactions on (Volume:25 , Issue: 6)

[2] U. Ravindran, J. Sribhavani, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 1, January 2014* "AN EFFICIENT XML DATA STREAMING METHOD IN MOBILE WIRELESS NETWORK "

[3] Effective Scheduling Algorithm for On-Demand XML Data Broadcasts in Wireless Environments Yongrui Qin1 Hua Wang1 Jitian Xiao2

[4] Secured Wireless XML Streaming Supporting Twig Pattern Queries K. Preethi1, S. Ganesh Kumar2 IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2112-211

[5] Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung "Lineage Encoding: An Efficient Wireless XML Streaming Supporting Twig Pattern Queries **"** Knowledge and Data Engineering, IEEE Transactions on Volume: 25 , Issue: 7 Publication Year: 2013 , Page(s): 1559 - 157.

[6] Marouane Hachicha and Jerome Darmont "A Survey of XML Tree Patterns " , IEEE transactions on knowledge and data engineering, vol. 25, no. 1, January **2013**

[7] Yongrui Qin1 Hua Wang1 Jitian Xiao" Effective Scheduling Algorithm for On-Demand XML Data Broadcasts in Wireless Environments" Australian Computer Society, Inc.22nd Australasian Database Conference(ADC 2011), Perth, Australia, January2011
[8] S. Vidya, R. Uma "Dynamic XML Dissemination Supporting Twig Pattern Queries" International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682 Vol.17, No.17, September 2013.
[9] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik, "Broadcast Disks: Data Management for Asymmetric Communication Environments," Proc. ACM SIGMOD Int'l Conf. Management of Data Conf., pp. 199-210, Mar. 1995

