**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. 2, February 2014, pg.282 – 289***

**RESEARCH ARTICLE**

# XML Dissemination Scheme for Mobile Computing Based on Lineage Encoding

**K. Anusree[1], Mrs. D. Usha[2], C. Shiny Jennifer[3]**

[1]M. Tech. Scholar, Department of Information Technology
Hindustan University, Chennai, Tamilnadu, India
anusreeanil29@gmai.com

[2]Assistant Professor, Department of Information and Technology
Hindustan University, Chennai, Tamilnadu, India
dusha@hindustanuniv.ac.in

[3]M. Tech. Scholar, Department of Information Technology
Hindustan University, Chennai, Tamilnadu, India
shinucharles02@gmail.com

**Abstract — In wireless environments, broadcasting is an efficient and scalable method to broadcast information to a massive number of clients. We propose an energy and latency efficient XML dissemination scheme for the wireless mobile computing environments. This paper presents a novel unit structure called G-node for streaming XML data in the wireless system. It applies the benefits of the structure indexing and attributes summarization that can integrate relevant XML elements into a group. It provides a path for selective access of their attribute values and text content. The G-node structure removes structural overheads of XML documents, and enables clients to avoid downloading of unwanted data during query processing. We also introduce a lightweight and effective novel encoding technique, called Lineage Encoding, to support evaluation of predicates and twig pattern queries over the wireless stream. The Lineage Encoding technique expresses the parent-child relationships among XML elements as a sequence of bit-strings, called Lineage Code(Lineage Code(V), Lineage Code(H)), and provides basic operators and functions for efficient twig pattern query processing at clients side.**

**Keywords— Twig pattern matching, Wireless XML streaming, mobile computing, wireless data broadcasting**

## 1. INTRODUCTION

With the rise in popularity of wireless environment, wireless information scheme are currently being realized in many areas [1], [2], [3]. In these systems, information access in any place at any time is enabled by clients equipped with small, hand-held wireless mobile communication environments. Fig. 1 shows the model of our wireless XML data broadcast system [4]. The system includes an XML Data Center (the broadcast server), a broadcast program scheduler, broadcast listeners (clients) and a downlink channel (the server sends information to clients via it). Downlink channel can be shared by all clients. But clients cannot send their individual queries to the server in this model as no uplink channel is available. In wireless XML broadcasting, the broadcast server retrieves XML data to be broadcasted from the XML repository [8]. After that, it parses and generates a wireless XML data stream [10]. The XML stream is continuously broadcasted via a broadcast medium. In the mobile client side, if a query is issued by the client, the client tunes in to the broadcast channel and downloads the relevant XML stream for query processing. We want to consider energy conservation of mobile clients when broadcasting data in the wireless mobile environment, because they use mobiles with minimized to provide fast result to the users. limited battery-power. The overall query processing time

Wireless broadcasting is an attractive method of broadcasting data in wireless environments due to its beneficial characteristics such as bandwidth efficiency, energy efficiency, and scalability [1], [5]. The server broadcasts data to an unlimited number of clients (i.e. scalable) through broadcast channels (i.e. bandwidth efficient because the broadcast channels are shared by many clients), and the clients listen to the channel and collect information of interest to them without sending any requests to the server (i.e., energy efficient because the amount of energy needed for receiving a data is much smaller than that for sending one). Depending on whether or not the client has the ability to send requests, broadcast technique can be divided into two categories, *push-based broadcast* and *on-demand* broadcast [6]. For push-based broadcast, the client no need to submit any request to the server but just listens to the broadcast channels and it is the client's responsibility to find the answers to its requests. In on-demand broadcast, clients can send their requests to the server like in the ordinary on- demand access mode, while the server transfers information to clients on public broadcasting channels.

We concentrated on the access of XML documents in wireless *push-based* broadcast systems in this paper we define a novel unit structure called G-node for streaming XML data in the wireless environment [1] to facilitate the dissemination of XML documents in on-demand wireless broadcast environments. The tuning time and access time are used to check the energy-efficiency and latency-efficiency in wireless data dissemination, respectively [7], [8]. The *Tuning Time (TT)* [9], [5], [8] 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 (AT)[10]* 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. The G-node structure remove structural overheads of XML documents, and enables clients to skip downloading of irrelevant data during query processing. 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. To the best of our knowledge, our technique is the first wireless XML streaming approach that completely supports twig pattern query processing in the wireless broadcast environment. We provide algorithms for generating wireless XML stream consisting of G-nodes and query processing over the wireless XML stream.

# 2. BACKGROUND

## 2.1 XML data streaming and XPath Expression

Structure information of XML document used as a stream index to transfer XML documents from the server to client via a broadcast channel. In this project, a novel XML streaming method for wireless broadcasting environments [3], [4] is introduced. An XML stream is organized to enable a selective access technique for simple Xpath [11], [12] queries, by borrowing the path summary method, which was originally devised for indexing semi-structured data. In order to utilize structure data as an XML data stream index, the structure information and text values of an XML document are separated. The proposed method demonstrates superior performance over previous approaches with regard to both access and tuning time. In this paper, we use XPath [11] as a query language. The location path selected the results of an XPath query. Location steps present in a location path. Processing each location step selects a set of nodes in the document tree. These nodes satisfy axis, node test and predicates described in the step [2].

## 2.2 Twig Pattern Query

Based on the containment labeling scheme, prior work decomposes a twig pattern into a set of binary relationships, which can be either ancestor –descendant relationships or parent–child relationships. Then, each relationship in binary is processed using structural join techniques and the final match results are obtained by "merging" individual binary join results together. The query processing of a core subset of XML query languages [11], XML twig queries. An XML twig query [2], represented as a small query tree is essentially a complex selection on the structure of an XML document.

The twig pattern query [19] is a core operation in XML query processing and popularly used as it can represent complex search conditions [12]. Fig. 2 shows two example twig pattern queries Q2 and Q3 with their tree structure. Cities located in the provinces of a country found by Q2 that has a child node "name" whose text content is "Belgium," and Q3 is to find cities in provinces located at "Middle" Matching a twig query [19] means finding all the instances of the query tree embedded in the XML data tree [17]. Finding all the occurrences of a twig pattern specified by a selection predicate on multiple elements in an XML document is a core operation for efficient evaluation of XML queries. A typical approach decompose the pattern into a set of binary structural relationships (parent- child or ancestor-descendant) between pairs of nodes then match each of the binary structural relationships against the XML database and finally stitch together the results from those basic matches.

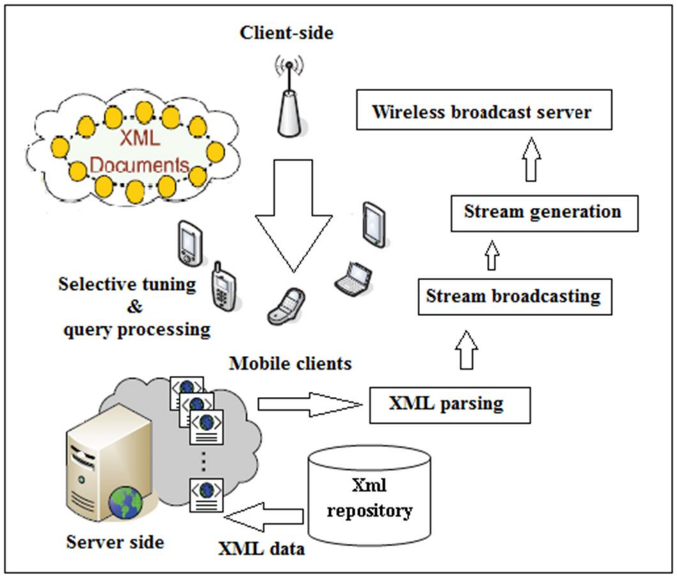


Fig.1. Wireless XML data broadcast system.

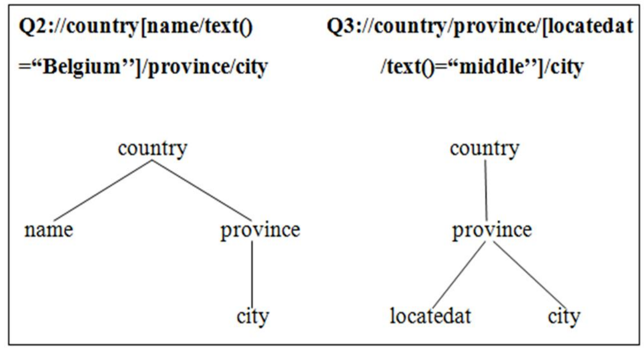


Fig.2. Example twig pattern queries and their tree representations. [12]

## 3. THE STRUCTURE INDEX

In this section, we introduce the Structure Index [13], which is document structure derived from the input XML documents. It can be used to preprocess the structure navigation part of XPath queries. At runtime, the Structure Index can be used to efficiently find queries that match a given XML document [10] by traversing its structure, and perform additional computation for predicate evaluation.

Structure Index [20] is constructed by representing each element or attribute node as a node in a Structure Index called *Index Node*. The relationships between Index Nodes are the same as parent-child relationship in nodes of document structure. This index will be used (as an XML document tree) to preprocess the structural navigation part of XPath queries. It will attempt to find all possible ways that the structural parts of the queries can be matched with this XML document tree. Subsequently, each result from structure matching is stored at the Index Node that it matches. In XML indexing [20] techniques [14], XML indexes can be regarded as summary of XML source documents and thus has much smaller sizes compared to the original source.

## 4. WIRELESS XML DATA BROADCASTING

Broadcast is one of the basic ways of information access via wireless technologies. The server broadcasts information to all mobile devices in a wireless data transmission system within its transmission range via a downlink broadcast channel. Clients "listen" to the downlink channel and access information of their interest directly when related information arrives. Broadcast [3] is bandwidth efficient because all clients can share the same downlink channel and retrieve data from it simultaneously. Broadcast stream [15], [1], [3] is also energy efficient at the client ends because downloading data costs much less energy than sending data. In this work we focus on the periodic broadcast mode since it has many benefits such as saving uplink bandwidth and power at the client ends by avoiding uplink transmissions and effectively delivering information to an unlimited number of clients simultaneously.

 

### 4.1 G-Node

This paper proposes a wireless XML stream by integrating information of elements of the same path. In this project G node [9] is used for streaming XML data in the wireless environment. That is the XML data stream consists of the sequence of group nodes called G-node. A streaming unit of a wireless XML stream, called G-node. The G-node structure remove structural overheads of XML documents and enables clients to skip downloading of irrelevant data during query processing. Fig.3 illustrates the G-node structure that integrates elements of the path "/mondial/country/province". The group descriptor (GD) is a collection of indices for selective access of a wireless XML stream. Node name provide the tag name of integrated datas, and Location path of integrated elements from the root node to the element node in the XML document tree structure provided by the Xpath expression. Child Index (CI) is a set of addresses that direct to the starting positions of child G-nodes in the wireless XML stream. Attribute Index (AI) contains the Couples of attribute name and address to the starting position of the values of the attribute that are stored contiguously in attribute value list. Text Index (TI) is an address directing to the starting position of Text List. In this scheme, an address means a point in time when the relevant data is broadcast on the air. The components of the GD are used to process XML queries in the client efficiently. Generally, node name and location path are used to identify G-nodes. Indices such as CI, AI and TI are used to selectively download the next G nodes, attribute values, and text. Attribute Value List (AVL) store attribute values and Text List (TL) store text contents of the elements by the G-node. Attribute values and text contents are collected in document order of elements.

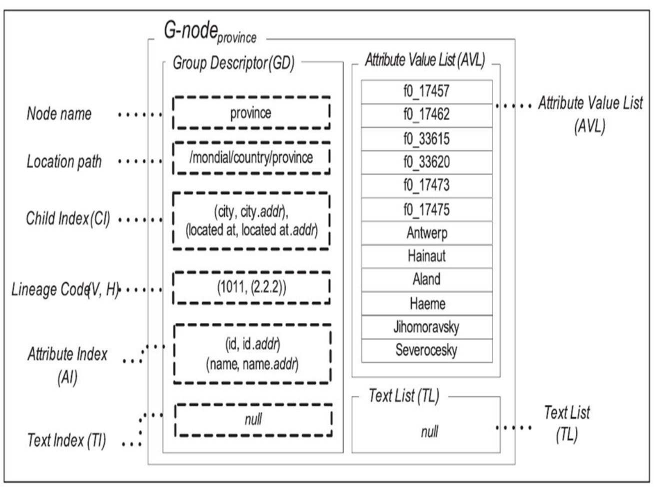


Fig.3. Structure of an example G-node. [12]

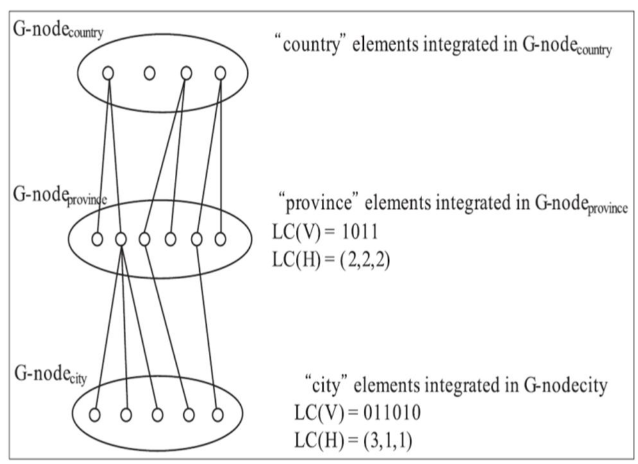


Fig.4. Lineage codes of G-nodecountry, G-node province and G-nodecity

### 4.2 Attribute Summarization

Attribute Summarization [16] technique is used to reduce the size of a wireless XML stream and that eliminates repetitive attribute names in a set of elements when generating a stream of G-nodes. In this project we exploit the benefits of the Attribute Summarization technique [11] which is used to reduce the size of a wireless XML element stream. An element in XML document may have multiple attributes, each attributes consists of a name and value pair. There is a structural characteristic that elements with the same tag name and location path often contain the attributes of the same name. During the generation of G-nodes stream Attribute Summarization eliminates repetitive attribute names in a set of elements.

## 5. LINEAGE ENCODING

This work introduces a novel encoding technique, called Lineage Encoding, to permit queries involving predicates and twig pattern matching. In the proposed system, two kinds of lineage codes, i.e., vertical code denoted by Lineage Code (V) and horizontal code denoted by Lineage Code (H) are used to represent parent-child relationships among XML elements in two G-nodes.

### 5.1 Lineage Code

This paper provide a light-weight encoding scheme, called Lineage Encoding [9] to represent parent-child relationships among XML elements in the G-nodes. We propose applicable functions and operators that apply bit-wise operations on the Lineage codes. Our scheme is the first wireless XML streaming approach that completely supports twig pattern query processing in the wireless broadcast environment. The Lineage Code scheme encodes parent-child relationships between two sets of elements in two G-nodes based on light-weight and efficient bit string representation. Fig. 4 demonstrates an example of Lineage Codes in G node country, G-node province, and G-node city. Note that Lineage Code (V) of G-node province is defined by 1,011 since the elements collected in G-node province are mapped to only the first, third, and fourth elements in G-node country. Lineage Code (H) of G-node province is (2, 2, 2) where each value represent the number of child elements in G-node province mapped to the same parent element in G-node country in document order.

## 6. WIRELESS XML STREAM GENERATION

In wireless XML stream generation [14] we explain how to produce wireless XML element stream [21]. A server gets an XML document to be broadcasted from the XML repository and it generates wireless XML steam by using SAX (Simple API for XML) [17], which is an event-driven API. During the parsing of an XML document SAX invokes content handlers. Then streaming of XML data streamed XML data are disseminated via a broadcast channel.

## 7. CONCLUSIONS AND FUTURE WORK

In wireless environments, data broadcasting is widely used for information delivery services due to its beneficial natures like bandwidth efficiency, energy efficiency, and scalability. In this paper, we proposed Lineage Encoding to support queries involving predicates and twig pattern matching and also defined the relevant operators and functions to efficiently process twig pattern matching. The client can retrieve the required data satisfying the given twig pattern by performing bit-wise operations on the Lineage Codes in the efficient G-nodes. Thus, our mechanism can support twig pattern query processing while providing both latency efficiencies and energy.

In the future, we plan to examine the areas that were not fully discussed in this project. First, depth-first traversal of XML elements builds up the access time for particular queries. Second, in wireless broadcasting system communication is not stable; the indexing performance should consider tail drops and packet losses network failures.

**ACKNOWLEDGEMENT**

It's a matter of pride and privilege for me to express my deep gratitude to the management of HITS for providing me the necessary facilities and support.

**REFERENCES**

[1] C.-S. Park, C.S. Kim, and Y.D. Chung,"Efficient Stream Organization for Wireless Broadcasting of Xml Data," Proc. Int'l Conf. Asian Computing Science Conf., pp. 223-235, 2005.
[2] J.P. Park, C.-S.Park, and Y.D. Chung, "Energy and Latency Efficient Access of Wireless XML Stream," J. Database Management, vol. 21, no. 1, pp. 58-79, 2010.
[3] Y.D. Chung, "An Indexing Scheme for Energy-Efficient Processing of Content-Based Retrieval Queries on a Wireless Data Stream," Information Sciences, vol. 177,no.2,pp.525-542, 2007.
[4] J.P.Park, C.-S.Park, and Y.D. Chung, "Energy and Latency Efficient Access of Wireless XML Stream," J. Database Management, vol. 21, no. 1, pp. 58-79, 2010. 2003.
[5] T. Imielinski and B.Badrinath, "Energy Efficient Indexing on Air," 1994 Proc. ACM SIGMOD Int'l Conf. Management of Data Conf., pp.25-36.
[6] Sun, Weiwei; Qin, Rongrui; Wu, Jinjin; and Zheng, Baihua. (2013). Air Indexing for On- Demand XML Data Broadcast.
[7] Y.D. Chung, S. Yoo, and M.H. Kim, "Energy- and Latency-Efficient Processing of Full- Text Searches on a Wireless Broadcast Stream," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 2,pp. 207-218, Feb. 2010.
[8] XML Data Repository, Research/xmldata sets, 2012. http://www.cs.washington.edu
[9] Jun Pyo Park, Chang-Sup Park, and Yon Dohn Chung, "Lineage Encoding: An Efficient Wireless XML Streaming Supporting Twig Pattern Queries". IEEE transactions on knowledge and data engineering, vol. 25, no. 7, july 2013.
[10] M. Altinel and M. Franklin, "Efficient Filtering of XML Documents for Selective Dissemination of Information," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 53-64, 2000.
[11] A. Gupta and D. Suciu, "Stream Processing of XPath Queries with Predicates," Proc. ACM SIGMOD Int'l Management of Data Conf., pp. 419-430, 2003.
[12] S. Al-Khalifa, H.V. Jagadish, N. Koudas, J.M. Patel, D. Srivastava, and Y. Wu, "Structural Joins: A Primitive for Efficient XML Query Pattern Matching," Proc. Int'l Conf.Data Eng. (ICDE), pp. 141-152, Feb. 2002.
[13] R.Kaushik,R. Krishnamurthy, J.F.Naughton, and R.Ramakrishnan,"On the Integration of Structure Indexes and Inverted Lists," Proc.ACM SIGMOD Int'l Management of Data Conf, June 2004.
[14] H. He and J. Yang, Multiresolution indexing of XML for frequent queries, In Proceedings of the 20nd International Conference on Data Engineering, 2004.
[15] M.S. Chen, K.-L. Wu, and P.S. Yu, "Optimizing Index Allocation for Sequential Data Broadcasting in Wireless Mobile Computing," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 1, pp. 161-173, Jan./Feb.
[16] J.P.Park, C.-S. Park, and Y.D. Chung, "Attribute Summarization: A Technique for Wireless XML Streaming," Proc. Int'l Conf. Interaction Sciences, pp. 492-496, Dec. 2009.

 

[17] SAX (Simple API for XML), org, 2004. http://www.saxproject.

[18] S. Chen, H. Li, J. Tatemura, W. Hsiung, D. Agrawal, and K.S. Candan, "Scalable Filtering of Multiple Generalizaed-Tree-Pattern Queries over XML Streams," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 12, pp. 1627-1640, Dec. 2008.

[19] H. Jiang and W. Wang, "Efficient Processing of XML Twig Queries with OR-Predicates," Proc. ACM SIGMOD Int'l Management of Data Conf,. pp. 59-70, June 2004.

[20] R.Kaushik,R.Krishnamurthy,J.F. Naughton, and R. Ramakrishnan,"On the Integration of Structure Indexes and Inverted Lists," Proc. ACM SIGMOD Int'l Management of Data Conf., June 2004.

[21] S.H. Park, J.H. Choi, and S. Lee, "An Effective, Efficient XML Data Broadcasting Method in Mobile Wireless Network," Proc. 17th Int'l Conf. Database and Expert Systems Applications (DEXA), pp. 358-367, 2006.

 