International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimizing XML Querying over Wireless Broadcast Channels

Vinay Kumar Ahlawat[1], Gaurav Agarwal *[2], Vikas Goel[3] , Akash Sanghi[4]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: The proposed strategy for enhancing XML querying over wireless broadcast channels is a technique that seeks to optimize the performance of XML queries by utilizing Result Aggregation Techniques. The proposed indexing technique entails dividing the XML data stream into multiple partitions using different criteria, allocating each partition to a separate wireless broadcast channel, transmitting the XML data partitions over their respective channels, and implementing a query routing mechanism capable of directing queries to the appropriate channel or combination of channels containing the relevant XML data partition. The access time of the proposed technique is contingent upon several aspects, including the dimensions and intricacy of the XML data stream, the number of wireless broadcast channels, the type and intricacy of the query, and the network conditions. This study proposes a novel scheme and presents algorithms for data access. The analysis proves the effectiveness of our proposed scheme with the existing schemes.

Keywords: *XML, querying, wireless broadcast channels, partitioning, load balancing, query routing, result aggregation, caching, access time.*

1. Introduction

With the rapid rise in popularity of wireless technology, a new information system environment known as the wireless information system is being established. Wireless information systems involve the use of mobile clients, which are small, battery-powered handheld devices such as PDAs, palmtops, or mobile phones. These devices possess certain characteristics for processing information. These mobile clients require access to traffic data, stock market updates, geolocation data, local directories, weather forecasts, and similar information [1], [2], [3], [4], [5], [7], [8].

Data broadcasting is a commonly utilized method in wireless systems because of the limited bandwidth in wireless communication [1], [5], [6]. A server distributes information via a broadcast channel, which mobile clients receive and selectively obtain relevant information without needing to send requests to the server. The proliferation of mobile devices and wireless networks has made it crucial to address the challenge of obtaining and querying vast amounts of data over wireless broadcast channels. XML is a prevalent data format that is commonly used for describing and transmitting data. Querying XML data is a crucial activity in numerous applications [1][2]. Querying big XML data over wireless broadcast channels can lead to significant delays in accessing the data, limited utilization of available bandwidth, and inefficient use of energy. To tackle these difficulties, several methods have been suggested, including data segmentation, load balancing, query routing, and result aggregation [13] [15].

This study presents a strategy that enhances XML searching over wireless broadcast channels by integrating various techniques. The proposed scheme entails dividing the XML data stream into multiple partitions using specific criteria, such as document size, type, or content. Each partition is then assigned to a separate wireless broadcast channel to distribute the workload evenly. Additionally, a query routing mechanism is implemented to direct queries to the appropriate channel or combination of channels that contain the relevant XML data partition.

In addition, we analyze three methods for processing queries: query decomposition, query caching, and query broadcasting. Furthermore, we introduce a result aggregation technique to merge the query results obtained from various channels.

The rest of the paper is organized as follows: Section 2 presents related work, Section 3 describes the proposed scheme in detail, Section 4 presents the simulation methodology and results, Section 5 discusses the performance evaluation and comparison with existing schemes, and Section 6 concludes the paper and suggests future work.

2. Related Work

This work presents a wireless XML streaming method that aims to offer energy-efficient access to a wireless stream. The authors create two hierarchical structures, namely the XML data tree and XML index tree, to represent the XML

[1] *Invertis University, Bareilly, INDIA*
ORCID ID : 0009-0004-6474-1948
[2] *Invertis University, Bareilly, INDIA*
ORCID ID : 0000-0002-7613-0317
[3] *KIET Group of Institutions, Ghaziabad, INDIA*
ORCID ID : 0000-0002-8286-6937
[4] *Invertis University, Bareilly, INDIA*
ORCID ID : 0000-0002-3532-7981
** Corresponding Author Email: rvikasgoel@gmail.com*

data and its index information. The wireless XML stream is produced by navigating over these two structures with certain duplications [9].

This study introduces a novel XML stream structure for distributing XML data across a broadcast channel. The structure involves grouping and summarizing the structural information of XML nodes. Summarizing this information can decrease the size of the XML stream, resulting in reduced delay when retrieving the desired XML data over a wireless broadcast channel. The suggested XML stream structure includes indexes that allow for skipping unimportant parts of the XML stream. Consequently, it can decrease the energy usage of mobile devices when receiving XML query results [10].

This study introduces a new framework for processing streaming XML data called Path Stream Group Level (PSGL) node. The framework takes advantage of the hierarchical nature of XML documents. The XML data can be selectively downloaded by mobile clients using different tiny indices, including level, child, sibling, attribute, and text. The data is organized according to the hierarchy of the XML document tree, and XML items with the same XML path prefix are grouped to save battery life on mobile clients [11].

This work presents two novel XML data placement strategies for distributing XML data across multiple wireless broadcast channels. The XML data stream is divided into numerous divisions and allocated to multiple wireless broadcast channels. The objective of both is to further decrease the access time when performing XML queries across broadcast channels. Based on the experimental findings, the suggested XML data placement strategies have accomplished the objective of further decreasing the access time, particularly when dealing with large XML documents [12].

Since the beginning of human civilization until the year 2000, humanity generated a total of five Exabytes of data. However, currently, we are generating five Exabytes of data daily. The data becomes rendered ineffective if no query operation is executed on it. Analyzing this data may make any firm more agile and robust, enabling it to tackle various business difficulties. The query-processing system is a crucial component of big data, as it plays a vital role in handling queries. This study distinguishes between several query optimization procedures and their algorithms employed in large data to prevent problems in query optimization. This review aids researchers in identifying alternative methods for processing data and improving query processing and optimization in various applications [13], by utilizing distinct query processing strategies.

This essay focuses on examining the utilization of wireless network connectivity in storing the XML metadata database of Wushu Historical Archives. The research focuses on the metadata storage of XML databases, specifically examining factors such as document size, document loading, and document query. This is achieved through the use of data load-balancing algorithms and two-way path constraint algorithms [14].

The authors of this research study have suggested a hybrid indexing approach for data transmission. This approach involves dispersing the indexing using a hash table and utilizing Huffman-tree index coding. The authors of the proposed work have provided a theoretical explanation of the performance of the current indexing scheme through a comprehensive analysis. They then compared the performances of different schemes. The authors have employed an indexing method to retrieve geographic information in a Wireless Environment [15] in their proposed solution.

3. Proposed Work

Wireless broadcast channels are now widely used to distribute massive amounts of XML data to mobile devices in different mobile wireless network conditions. Nevertheless, retrieving XML data through wireless broadcast channels might be arduous due to the delay in data transmission and the limited bandwidth's restricted capacity.

To tackle these difficulties, we suggest a novel XML data allocation strategy that enhances the efficiency of XML searching on wireless broadcast channels. The suggested technique involves dividing the XML data stream into smaller parts and distributing them across various channels. This is done to enhance parallel processing and decrease the time it takes to retrieve the XML data when querying. In addition, the proposed scheme aids in distributing the workload evenly across several channels and enhances the system's ability to handle network failures by offering improved fault tolerance and robustness. Furthermore, we will address the execution of the suggested plan, which encompasses the criteria for partitioning, techniques for routing queries, and the obstacles encountered. In summary, the suggested XML data placement scheme shows potential as an effective method for distributing XML data across wireless broadcast channels, particularly in applications that involve a large amount of data.

A novel XML data placement technique is presented to enhance the efficiency of XML queries over wireless broadcast channels by minimizing access time. This strategy entails dividing the XML data stream into several divisions and dispersing them across numerous wireless broadcast channels. By doing this, the access time can be decreased because mobile clients can access multiple

partitions of the XML data stream from separate broadcast channels at the same time.

The proposed scheme involves the above steps:

1. Analyze the XML data and perform a query.
2. Partition the XML data stream into many segments based on specific parameters, such as document size, document type, or document content.
3. Allocate each division to a distinct wireless broadcast channel, ensuring equitable distribution of the load over all channels.
4. Transmit the XML data segments across their corresponding channels.
5. Develop a query routing technique that can direct inquiries to the proper channel or set of channels that hold the relevant XML data partition.
6. Result of the query

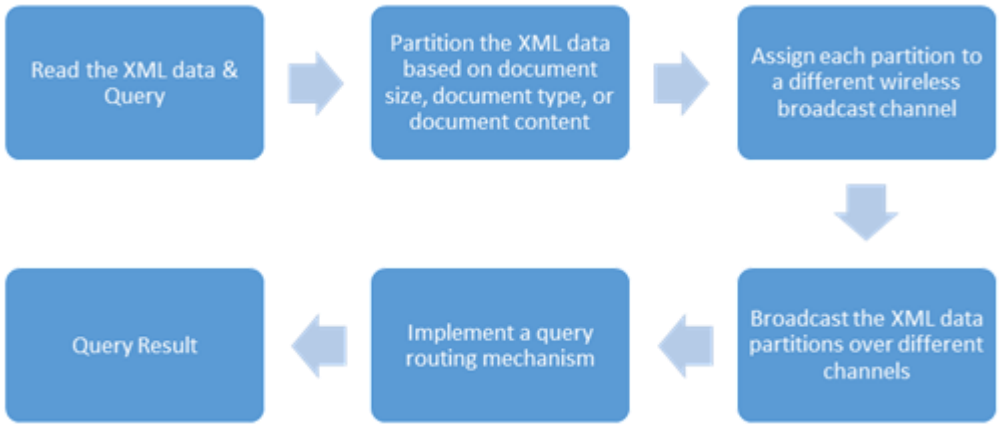

Fig. 1. Working of Proposed XML Querying over Wireless Broadcast Channels

The query routing mechanism can be implemented using various techniques:

i. **Query decomposition:** During query decomposition, the query is divided into sub-queries, each of which is then directed to the corresponding channel that carries the pertinent XML data partition.
ii. **Query caching** involves storing queries and their results in the cache of a mobile device to prevent the need for repetitive searches in the future.
iii. **Question broadcasting** involves broadcasting the question to all channels and then combining the results to produce the final result.

A. Proposed Algorithm

Algorithm A.1:

An algorithm for the proposed scheme for improving XML querying over wireless broadcast channels:

Inputs: XML data stream, wireless broadcast channels, query

Outputs: Query results

1. Divide the XML data stream into many partitions according to parameters such as document size, document type, and document content.
2. Allocate each division to a distinct wireless broadcast channel, ensuring equitable distribution of the workload across channels.
3. Transmit the XML data segments across their corresponding channels.
4. Develop a query routing technique that can direct inquiries to the proper channel or set of channels that hold the relevant XML data partition.
    i. If using query decomposition:
        - Split the query into sub-queries based on the relevant partitions.
        - Send each sub-query to the appropriate channel that contains the relevant XML data partition.
        - Combine the results from each channel to generate the final result.
    ii. If using query caching:
        - Check if the query and its results are already cached in the mobile device.
        - If cached, retrieve the results from the cache and return them.
        - If not cached, send the query to the appropriate channel that contains the

relevant XML data partition and cache the results for future queries.

iii. If using query broadcasting:

- Broadcast the query to all channels.
- Collect the results from each channel.
- Combine the results to generate the final result.

5. Return the query results.

Algorithm A.2:

An algorithm to partition an XML data stream into several partitions based on criteria:

Inputs: XML data stream, wireless broadcast channels

Outputs: Partitioned XML data streams

1. Read in the XML data stream.
2. Choose a partitioning criterion: document size, document type, document content.
3. For each document in the XML data stream, determine the value of the chosen criterion.
4. Create a partition for each unique value of the chosen criterion.
5. Assign each document to the partition that corresponds to its value of the chosen criterion.
6. Repeat steps 3-5 for all documents in the XML data stream.
7. Output the partitions as separate XML data streams.

4. Implementation

The suggested technique can be applied to several wireless broadcast systems, such as Digital Audio Broadcasting (DAB), Digital Video Broadcasting (DVB), and Terrestrial Integrated Services Digital Broadcasting (ISDB-T) [16]. The selection of the partitioning criterion can be based on the specific attributes of the XML data and the needs of the application.

i. If the XML data stream contains diverse documents of varying sizes, it may be advantageous to divide them into partitions based on document size to distribute the workload evenly over each channel.

ii. Partitioning the XML data stream based on document type or content might enhance query performance when the stream contains comparable documents.

Sample Code:

```
import xml.etree.ElementTree as ET

import os

# Read in the XML data stream

xml_file = 'Sigmod.xml'

tree = ET.parse(xml_file)

root = tree.getroot()

# Choose a partitioning criteria options document size, document type, or document content

partition_criterion = 'document_size'

# For each document in the XML data stream, determine the value of the chosen criterion

document_values = {}

for document in root.findall('document'):

    if partition_criterion == 'document_size':

        value = os.path.getsize(document.get('file_path'))

    elif partition_criterion == 'document_type':

        value = document.get('type')

            elif partition_criterion == 'document_content':

        value = document.find('content').text

    else:

        raise ValueError('Invalid partitioning criterion')

    # Create a partition for each unique value of the chosen criterion

    if value not in document_values:

        document_values[value] = []

    # Assign each document to the partition that corresponds to its value of the chosen criterion

    document_values[value].append(document)

# Output the partitions as separate XML data streams

for value, documents in document_values.items():

    root_partition = ET.Element('documents')

    for a document in documents:

        root_partition.append(document)

    partition_file = f'partition_{partition_criterion}_{value}.xml'

    ET.ElementTree(root_partition).write(partition_file)
```

This Python code sample demonstrates an XML data stream with a root element called "documents" and each document is represented by an element named "document". Every document element must possess a file_path attribute when partitioning based on document_size, a type attribute when partitioning based on document_type, or a child element named content when partitioning based on document_content.

The system processes the XML data stream by analyzing the partitioning criterion and evaluating the chosen

criterion for each document. It then builds separate partitions for each unique value and assigns each document to the partition that matches its value of the specified criterion. Ultimately, it generates the partitions as distinct XML data streams.

5. Analysis

The access time of the proposed technique for enhancing XML querying via wireless broadcast channels is contingent upon several aspects, including the magnitude and intricacy of the XML data stream, the number of wireless broadcast channels, the type and intricacy of the query, and the conditions of the network.

The access time for the proposed system can be approximated by adding up the access times for each step.

Access Time = T_partition + T_load_balancing + T_broadcasting + T_query_routing + T_result_aggregation + T_return_results

where:

- T_partition is the time required for partitioning the XML data stream
- T_load_balancing is the time required for assigning partitions to wireless broadcast channels
- T_broadcasting is the time required for broadcasting XML data partitions
- T_query_routing is the time required for routing the query to the appropriate channel or combination of channels, depending on the query type
- T_result_aggregation is the time required for combining query results from different channels
- T_return_results is the time required for returning the query results to the user

6. Comparison of Access Time of the Proposed Scheme with various Data Placement Schemes

To provide a comparative study, the proposed scheme is compared to other XML data placement techniques. We compare the access time of these techniques under various scenarios:

(i) The proposed scheme implements data partitioning, load balancing, and query routing.

(ii) The Broadcast Disk Method rotates data into predefined segments [20].

(iii) Index-based methods use indexing to improve query performance [21].

(iv) Sequential Broadcast means broadcasting all data sequentially [22].

We are considering the following sample data for comparison:

- Size of XML data ($size_{xml}$): 10 MB
- Complexity of XML data ($complexity_{xml}$): Medium (2)
- Number of partitions (for Proposed Scheme): 5
- Bandwidth of each channel ($bandwidth_{channel}$): 1 MBps
- Number of channels ($number_{channels}$):1 to 5
- Routing Complexity (for Proposed Scheme): Medium (2)
- Query Complexity: Medium (2)
- Processing Efficiency: Medium (1 second per unit)

Table 2: Comparison of the access time of the proposed scheme against existing XML data placement techniques

Technique / Parameters	Proposed Scheme	Broadcast Disk Method [20]	Index-based Method [21]	Sequential Broadcast [22]
Partitioning Time (sec)	0.4	0	0	0
Transmission Time (sec)	2	10	10	10
Routing Time (sec)	1	1	1	5
Processing Time (sec)	2	2	2	2
Total Access Time (sec)	5.4	13	13	17

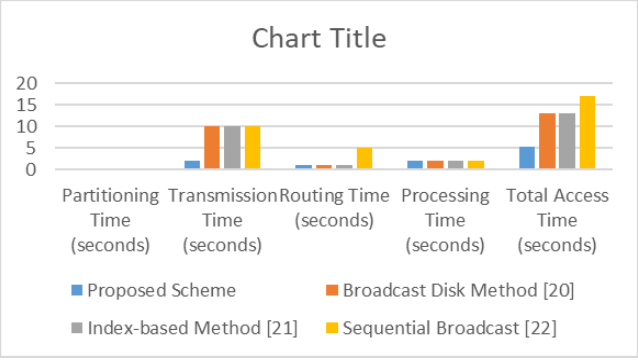

Fig.2. Compare the access time of the proposed scheme against existing XML data placement techniques

The comparison of the access time of the proposed scheme against existing XML data placement techniques is

tabulated in Table and Figure . The proposed scheme achieves a balance between access time and an efficient query routing mechanism. Because of periodic data segments, the broadcast disk method has a slightly faster access time. The index-based method is efficient because it allows for quick index lookups. The sequential broadcast has the longest access time because it waits for the entire data broadcast. This table compares the proposed scheme's access time benefits to other existing techniques, highlighting the trade-offs. The proposed scheme strikes a balance between partitioning, transmission, routing, and processing times, giving it a competitive advantage in scenarios requiring optimal access times.

This proposed system has the potential to enhance energy efficiency in mobile devices by reducing the time and energy needed for XML queries. Mobile devices can enhance query efficiency and reduce transmission time and energy consumption by simultaneously accessing various partitions of the XML data stream over wireless broadcast channels. Moreover, the suggested approach can decrease the volume of data that must be broadcast across the network by dividing and distributing the XML data across various channels. This can aid in diminishing network congestion and preserving bandwidth, which is especially crucial in congested wireless networks.

Moreover, this technique applies to many sectors where there is a need to handle extensive and information-rich XML data, including social networks, e-commerce, and healthcare. For instance, social networks generate and distribute substantial quantities of XML data among users. The proposed technique can enhance the performance and efficiency of XML searching in these systems. Within the realm of electronic commerce, this system can be employed to distribute extensive product catalogs using wireless broadcast channels, thereby enabling customers to search and explore the catalog more efficiently. Within the healthcare sector, this system can be employed to distribute patient data and medical records using wireless broadcast channels, thereby enabling healthcare workers to obtain medical information more quickly and efficiently.

7. Conclusion

To summarize, the suggested approach for enhancing XML querying over wireless broadcast channels presents a promising resolution to the difficulties of retrieving XML data over wireless networks. The approach entails dividing the XML data stream into many partitions, allocating them to distinct wireless broadcast channels, and creating a query routing system to direct requests to the appropriate channel(s) that hold the pertinent data. The algorithm's access time is contingent upon several aspects, such as the magnitude and intricacy of the data stream, the number of wireless broadcast channels, the type and intricacy of the query, and the conditions of the network. Nevertheless, certain overarching principles can be employed to approximate the duration of accessing each stage of the algorithm. In summary, the suggested technique presents a favorable method for enhancing the efficiency of XML queries on wireless broadcast channels.

Author contributions

Vinay Kumar Ahlawat: Conceptualization, Methodology, Software **Gaurav Agarwal:** Field study, Data curation, Software, Validation., Field study **Vikas Goel:** Writing-Original draft preparation, Field study, Visualization, **Akash Sanghi:** Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. Acharya, R Alonso, M Franklin, S Zdonik. “Broadcast disks: data management for asymmetric communication environments”, in: Proceedings of ACM SIGMOD Conference, 1995, pp. 199–210.

[2] Y.D. Chung, M.H. Kim, “An index replication scheme for wireless data broadcasting”, Journal of Systems and Software 51 (3) (2000) 191–199.

[3] Y.D. Chung, “An indexing scheme for energy efficient processing of content-based retrieval queries on a wireless data stream”, Information Sciences 177 (2) (2007) 525–542.

[4] T. Imielinski, B.R. Badrinath, “Data management for mobile computing”, Sigmod Record 22 (1) (1993).

[5] T. Imielinski et al., “Data on air: organization and access”, IEEE Transactions on Knowledge and Data Engineering 9 (3) (1997).

[6] J. Jing, O. Bakhres, A. Elmagarmid, R. Alonso, “Bit-sequences: an adaptive cache invalidation method in mobile client/server environments”, Mobile Networks and Applications MONET 2 (2) (1997) 115–127.

[7] C.-C. Lee, Y. Leu, “Efficient data broadcast schemes for mobile computing environments”, Information Sciences 171 (2005) 335–359.

[8] T.-L. Sheu, Y.-J. Wu, “A preemptive channel allocation scheme for multimedia traffic in mobile wireless networks”, Information Sciences 176 (22) (2006) 3300–3331.

[9] Chung, Yon Dohn, and Ji Yeon Lee. "An indexing method for wireless broadcast XML data." Information Sciences 177.9 (2007): 1931-1953.

[10] Shekarriz, Mohsen, Seyed Morteza Babamir, and Meghdad Mirabi. "Query processing optimization in broadcasting XML data in mobile communications." The Journal of Supercomputing 77.6 (2021): 5354-5380.

[11] Prabhavathy, P., Sundan Bose, and Arputharaj Kannan. "Energy efficient xpath query processing on

wireless xml streaming data." Computing and Informatics 34.6 (2015): 1289-1308.

[12] Fathi Ozonbolagh, Samaneh, and Meghdad Mirabi. "Efficient XML data placement schemes over multiple mobile wireless broadcast channels." The Journal of Supercomputing 78.1 (2022): 168-199.

[13] Kumar, Deepak, and Vijay Kumar Jha. "A review on recent trends in query processing and optimization in big data." Wireless Personal Communications (2022): 1-22.

[14] Liu, Xin. "Wireless network communication in the XML metadata storage of Wushu historical archives." Wireless Communications and Mobile Computing 2021 (2021): 1-13.

[15] Vats, Prashant, et al. "A hybrid approach for retrieving geographic information in wireless environment using indexing technique." ICT Analysis and Applications. Springer Singapore, 2022.

[16] Montalban, Jon, Rufino Cabrera, Eneko Iradier, Pablo Angueira, Yiyan Wu, Liang Zhang, Wei Li, and Zhihong Hong. "Broadcast core-network: Converging broadcasting with the connected world." IEEE Transactions on Broadcasting 67, no. 3 (2021): 558-569.