

SECURITY AND PRIVACY CONCERNS IN XML INDEXING FOR WIRELESS DATA BROADCASTING

VINAY KUMAR AHLAWAT

Research Scholar, Department of CS&E, Invertis University Bareilly, India. Email: vinahlawat@gmail.com

GAURAV AGARWAL

Professor, Department of CS&E, Invertis University Bareilly, India. Email: gaurav.a1@invertis.org

VIKAS GOEL

Professor, Department of IT, KIET Group of Institutions, Ghaziabad, India. Email: vikas.goel@kiet.edu

PRAGATI GOEL

Associate Professor, MCA Department, NCRD'S Sterling Institute of Management Studies, Nerul, Navi Mumbai. Email: goelpragati78@gmail.com

RAJU RANJAN

Associate Professor, School of CS&E, Galgotias University, Greater Noida, India.
Email: drraju.ranjan@galgotiasuniversity.edu.in

NARENDRA KUMAR

Professor, Department of CSE, Galgotias College of Engg. & Tech., Greater Noida, India.
Email: narendrakumar@galgotiacollege.edu

Abstract

There is a need to identify possible a single effective and privacy-preserving solution to all XML data indexing techniques for wireless data broadcasting. These indexing techniques are employed to secure and protect privacy in XML data. The proposed research aims to evaluate a detailed examination and comparison of these techniques to identify their gaps and possible solutions to these gaps in terms of concern parameters. The emphasis is on finding a single solution that not only enhances security and privacy but also improves accuracy in data retrieval and query performance in XML indexing techniques. Also, this proposed paper acknowledges potential challenges such as the negative impact on usability due to encryption and decryption overhead in Secure Data Transmission. In conclusion, the research suggests that an effective solution for securing XML data and protecting privacy may involve a strategic combination of techniques tailored to specific needs and compliance requirements. This highlights the importance of a comprehensive and adaptive approach to address the multifaceted challenges associated with wireless data broadcasting and privacy protection in XML data. The results and conclusions from this work can also be used to further design a novel secure dynamic indexing technique for wireless XML data broadcast.

Keywords: Partial-Tree Indexing, Encrypted Indexing, Secure Data Transmission, Access Control Mechanisms, Anonymization Methods, Regulatory Compliance, XML Indexing.

I. INTRODUCTION

A. Research Contribution:

XML, or Extensible Markup Language, is a widely used format for storing and transmitting structured data, including documents, databases, and web services. One of the key challenges in broadcasting XML data over mobile wireless networks is the limited bandwidth and varying quality of service that these networks can provide. To address this

challenge, indexing techniques are used to reduce the amount of data that needs to be transmitted and improve the speed and efficiency of data retrieval. In the process of indexing, data arrival information is established once the necessary data is available on the channel. [1]

Data integrity is a major security concern. Without protection, the indexed data is vulnerable to manipulation and corruption, which can result in inaccurate search results or even privacy violations. One method for safeguarding the indexed data is encryption. Using access control mechanisms is another strategy to regulate the kinds of data that various people can access [7]. Privacy issues emerge when the XML documents being indexed contain sensitive or personally identifiable information. This can contain financial information, private company information, or personally identifiable information (PII). Such information may be broadcast in breach of privacy laws and regulations. Sensitive data can be removed or obscured using data anonymization techniques like data masking or data perturbation to allay these worries [8].

Moreover, XML documents could include data that needs to be compliant with regulations, such as financial or medical records. Data must be safeguarded from unwanted access and handled with specific protocols to comply with standards like HIPAA or PCI DSS [9]. Using suitable security and privacy measures, such as encryption, access control, and anonymization techniques, is crucial to addressing these privacy and security concerns. In addition, it's critical to handle sensitive data by regulatory compliance standards [10][11].

Misuse of indexed data is a possible security and privacy concern as well. For instance, unapproved uses of indexed data could include profiling or targeted advertising. When private or delicate information is at stake, this can be quite alarming. Adequate privacy regulations and limits on data use must be put in place to prevent misuse of indexed data [7].

B. Research Gaps:

In the process of indexing, data arrival information is established once the necessary data is available on the channel. For wireless data broadcast using XML indexing, security, and privacy are major problems. Sensitive information from sent XML documents is extracted and stored throughout the indexing process, where it may be accessed by unauthorized individuals [1].

The potential of data breaches in XML indexing for wireless data broadcasting is a significant security and privacy issue. The security and integrity of the indexed data may be exposed if it is not securely protected from threats like hacking or interception. Since wireless networks are naturally more susceptible to interception and illegal access, this danger is particularly severe for wireless data transmission [12]. Strong authentication and encryption measures must be used to protect data transfer and storage to prevent data breaches. Also, it's critical to consistently check the system for any instances of unauthorized access or dubious activity [12].

Moreover, it is important to remember that security and privacy may also be impacted by the XML indexing approach you use. There is a higher danger of privacy violations when using indexing techniques like full-text indexing, which may extract and store more information than is necessary. Other methods, including path-based indexing, might be more effective and considerate of users' privacy [13].

Privacy and security are significant issues with XML indexing for wireless data broadcasting. There is a need to study and implement suitable security and privacy mechanisms, such as encryption, access control, and anonymization techniques. Routinely monitoring the system for any potential threats or breaches, are crucial for ensuring the confidentiality, integrity, and privacy of indexed data.

C. Research Questions

- (i) What are the gaps in the existing knowledge about XML data security and privacy-preserving XML data indexing techniques within the context of wireless data broadcasting?
- (ii) Can a one-size-fits-all security approach be identified to effectively tackle security and privacy concerns in the indexing technique for wireless XML data broadcasting?

The rest of this research paper is proposed as follows. Section 2 initiates with the background of the concern XML indexing techniques and security issues addressed in XML indexing techniques in available research articles. Section 3 studies and analyzes various secured XML indexing techniques used to generate the stream for broadcasting. All the results and findings are detailed to answer the questions in the introduction section. In the first part, various secured XML indexing techniques are compared with their key features. In the second part, various secured XML indexing techniques are compared with concerned parameters. In section 4, a detailed analysis of each secured indexing technique is described. In section 5, the conclusion of the techniques was drawn from a comparison analysis.

II. REVIEW METHODOLOGY

In the realm of wireless communication, broadcasting emerges as a highly fitting method for disseminating information, particularly appealing for handheld mobile devices with constrained resources in asymmetric communication scenarios. Incorporating the index alongside the data proves to be an energy-efficient solution. Various conventional disk-based indexing methods have been expanded to encompass techniques such as B+ trees, hashing, and signature indexing.

Furthermore, a comprehensive analysis of these indexing techniques has been conducted in the context of both single and multi-level wireless channels [14]. All indexing techniques for wireless broadcast channels are depicted in Figure 1. Conventional indexing techniques designed for uniform data sets may not be suitable for XML data due to their semi-structured nature, which is identified using XPath expressions rather than

key attributes. In the case of XML data, it is streamed alongside the index and broadcasted. A survey and analysis of diverse XML-based indexing techniques have been conducted, focusing on their key factors, as outlined in [15].

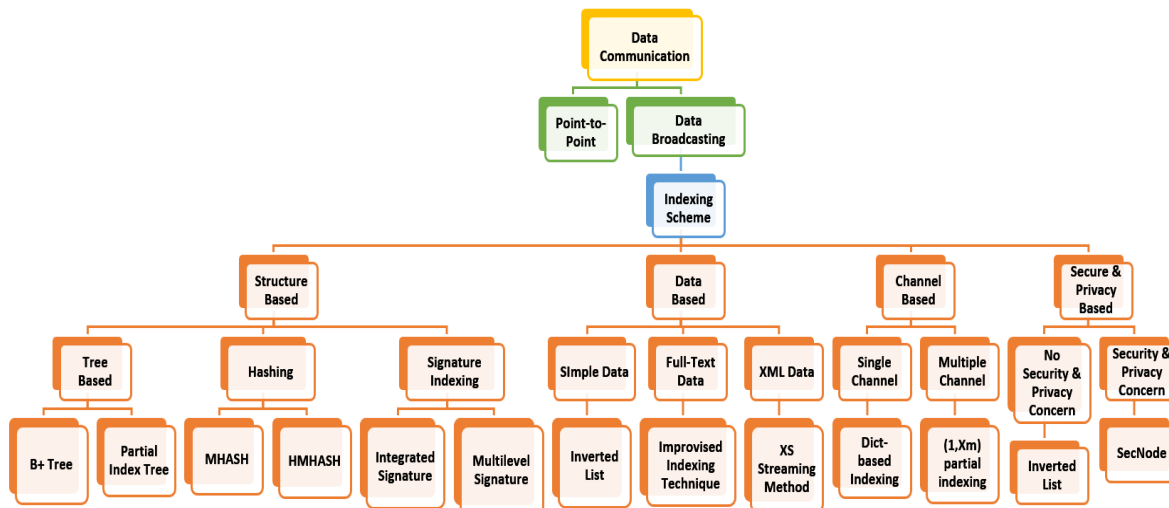


Figure 1: Indexing Techniques for Wireless Broadcast Channels

The authors of this study have concentrated on addressing the challenge of indexing methods for wireless broadcasting of XML data. They've identified that XML data is semi-structured, presenting both advantages and challenges. This semi-structured nature hinders the application of traditional indexing methods, and it poses difficulties for disc-based XML indexing methods in creating wireless streams for XML documents, mainly because the index is not static to conserve energy on the channel. In response, the authors propose an enhancement to existing indexing techniques named the (1, X_m) method. This method is specifically designed to cater to a multiple-channel environment, as outlined in [16].

The paper discusses XML, a widely used markup language on the Internet, responsible for generating substantial amounts of data in XML format. The authors introduce two search methods aimed at efficiently navigating through large volumes of data using XML indexing technology. They address an existing issue where performance tends to favor one side in keyword searches of extensive XML documents. To tackle this, they propose the Content Model, a solution applicable to both informal and formal data processing. This model facilitates the proper handling and utilization of non-structured data, aligning with the NoSQL paradigm for swift processing across different devices and enabling a responsive user interface. The authors highlight the versatility of this approach, emphasizing its applicability in various services and devices, such as N-Screen and mobile devices, without the need for specific viewers. This adaptability proves advantageous for tasks requiring efficient processing in diverse scenarios [13].

In the realm of securing sensitive data contained in XML documents within a confidential collaborative system that abstains from sharing such data, a scheme for privacy-preserving data disclosure decisions has been introduced. This scheme operates under the assumption of a trusted server. Inspired by the idea of segregating storage structure and content, the approach employs a temporary access matrix to represent structure authorization, while a vector signifies content authorization at the level of leaf nodes. The access matrix, following conversion rules, not only denotes access authorization for all nodes but also preserves the fundamental structure of the XML document. Through the amalgamation of vectors and matrices, the scheme can provide distinct access perspectives tailored to different user groups with diverse objectives. Additionally, start-end encoding is applied for encoding all nodes, simplifying the process of node and content location. The privilege matrix plays a pivotal role in resolving privacy synchronization changes for all users, thereby augmenting the overall efficacy of the privacy-preserving data disclosure decision scheme [17].

XPath is a widely utilized tool for querying XML documents across diverse applications. Nevertheless, as the XML document's node count expands, the execution complexity of queries concurrently rises. This challenge becomes particularly pronounced when handling very large XML documents, particularly in cases where there is an inadequate amount of computer memory to accommodate the storage and processing of the entire tree data. The principal objective of this research is to formulate an algorithm specifically designed for querying extensive XML trees within a distributed-memory environment [18]. This undertaking is driven by the intention to confront the computational hurdles associated with querying large XML documents by harnessing the capabilities of a distributed memory configuration.

To safeguard XML data transmitted through a mobile wireless network, mobile clients must conform to a set of access authorizations stipulated in the original XML document. Within such environments, the access of mobile clients is confined to authorized segments of an encrypted XML stream based on their specific access permissions. While several indexing methods have been proposed to enable selective access to XML data over the XML stream, these methods do not extend to encrypted XML data. This paper introduces a groundbreaking unit structure for the XML stream, termed SecNode, with the explicit purpose of supporting the data confidentiality of XML data disseminated over the wireless broadcast channel. Furthermore, two indexes, Min (NCS) and Min (NIS), are defined for the SecNode structure to adeptly process XML queries over the encrypted XML stream [19]. These innovations are geared towards fortifying the security and enhancing the query processing efficiency of XML data within wireless broadcast channels.

The "Database as a Service" (DBaaS) paradigm has garnered significant attention in recent years, leading to concerns about the security of data stored on the servers. As companies increasingly outsource their XML databases to untrusted parties, there is a growing need for secure data storage and efficient query processing. This paper delves into the realm of encrypted XML documents, exploring cryptographic index structures and

the associated query-processing methodologies. The research provides a comparative analysis of various algorithms found in the literature, offering insights into the diverse approaches taken to address the security challenges inherent in outsourcing XML databases to external entities [24].

The prevailing method for data publishing heavily relies on policies, guidelines, and agreements to delineate the types of data deemed suitable for publication and the permissible uses of the published data. However, this approach is prone to either excessively distorting data or inadequately safeguarding privacy. Privacy-preserving data publishing (PPDP) emerges as a solution, providing methods and tools to disseminate valuable information while ensuring the protection of data privacy. In recent years, PPDP has gained considerable attention in research communities, resulting in the proposal of numerous approaches tailored to different data publishing scenarios. This survey systematically summarizes and evaluates diverse PPDP approaches, delving into the challenges encountered in practical data publishing. The objective of the review is to elucidate the distinctions and requirements that differentiate PPDP from other related problems. Additionally, the survey outlines potential avenues for future research in the realm of privacy-preserving data publishing [31].

This paper introduces a novel XML stream structure designed to efficiently disseminate XML data through a broadcast channel. The approach involves grouping and summarizing the structural information of XML nodes to decrease the size of the XML stream. This reduction in size translates to lower latency when retrieving specific XML data over a wireless broadcast channel. The proposed XML stream structure incorporates indexes, enabling the skipping of irrelevant parts within the XML stream. This feature contributes to minimizing the energy consumption of mobile devices during the download of XML query results. Furthermore, the suggested XML stream structure is capable of processing various types of XML queries. Experimental results demonstrate its effectiveness in enhancing the performance of XML query processing over XML data streams, surpassing existing research works in terms of access and tuning times.[39]

This study addresses the implementation of a lightweight and energy-efficient encoding technique known as lineage encoding to overcome challenges such as battery constraints, thereby facilitating the development of a sustainable large-scale Wireless Sensor Network (WSN). Our investigation centers on establishing independent communication systems for mobile data collected in casual mobile environments. The lineage encoding approach represents the parent-child relationships of XML nodes as a bit-stream arrangement, referred to as lineage code (V, H). This method incorporates an efficient twig pattern query processing for mobile clients. Through conducted experiments, we aim to ensure minimal query processing time and maximize the network's lifespan with the utilization of lineage encoding (LE). Simulation results indicate that the large-scale WSN can support an unlimited number of clients across diverse network conditions, demonstrating the effectiveness of lineage encoding in extending network range and lifetime. [40]

III. OVERVIEW

To address the security and privacy concerns in XML indexing for wireless data broadcasting, a variety of XML indexing approaches are studied and analyzed in this section. These methods concentrate on safeguarding confidential information and limiting illegal access to indexed data [17]. Some of the techniques are depicted in figure 2:

D. XML Indexing Techniques

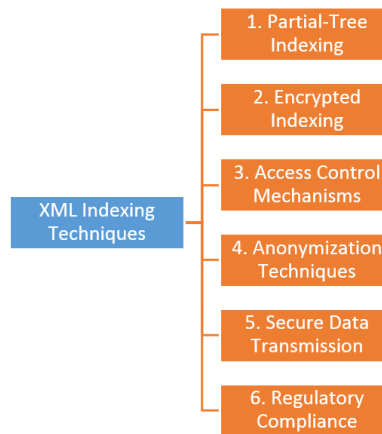


Figure 2: XML Indexing Techniques for Security & Privacy

A. Partial-Tree Indexing

As opposed to indexing the entire text of the XML document, partial-tree indexing only indexes a subset of the XML document, such as particular nodes or properties. As less sensitive data needs to be indexed and retained, this strategy is particularly helpful for addressing privacy issues [18]. Only the chosen nodes or attributes are indexed in partial tree indexing, and the index structure is constructed using the connections between these nodes or attributes. This means that the index is based on particular passages of the document that are regarded as relevant or significant rather than the entire text of the document. Compared to other indexing methods, partial-tree indexing has several benefits. As only the chosen nodes or properties must be indexed as opposed to the full document, it enables more effective indexing and searching of big XML documents. This may lead to quicker query processing times and less need for storage [3].

Moreover, by limiting the quantity of sensitive material that is indexed and retained, partial-tree indexing can assist in easing privacy issues. Sensitive data can be omitted from the index by indexing only specific nodes or properties, which can aid in preventing unwanted access to sensitive data. When only a portion of an XML document needs to be searched or when the XML document contains sensitive information that shouldn't be indexed, partial-tree indexing can be especially helpful. In situations where there are insufficient resources for indexing and querying huge XML documents, it is also a helpful technique for mobile devices or other low-power devices. It is important to keep in mind that Partial-Tree Indexing has some restrictions. For instance, it might not be appropriate

in situations requiring full-text search or when the connections between nodes or characteristics are intricate and difficult to represent in an index structure. Other XML indexing methods, including full-text indexing, may be more appropriate in these circumstances.

B. Encrypted Indexing

A method called encrypted indexing is used to safely index and search encrypted data, including XML documents. In situations where sensitive data is stored or transmitted, such as in cloud storage or wireless data broadcasting. This strategy is utilized to solve security and privacy concerns [24]. Encrypted indexing's fundamental premise is to build an index structure that enables effective searching of encrypted data without exposing the plaintext data. This is accomplished by utilizing a secure encryption technique to first encrypt the data and a secure index structure to index the encrypted data. Three different encrypted indexing schemes for wireless channel broadcasts are depicted in figure 3.

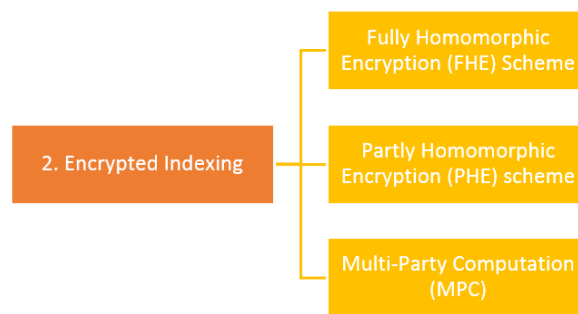


Figure 3: XML Encrypted Indexing Techniques

Using a searchable encryption method, such as a fully homomorphic encryption (FHE) or a partly homomorphic encryption (PHE) scheme, is one method for implementing encrypted indexing. Without disclosing the plaintext data, these approaches enable effective search and retrieval of encrypted data. The same encryption approach is also used to encrypt the index, enabling secure searching of the encrypted data [25].

Using a method known as secure multi-party computation is another method of using encrypted indexing (MPC). With this method, several parties work together to safely compute the index structure without disclosing any private information. The encrypted data is then searched using the index without exposing the unencrypted data [26]. For secure indexing and querying of sensitive data, encrypted indexing offers several benefits. It offers a high level of security and privacy by encrypting the data and index because the plaintext data cannot be accessed by unwanted parties. Also, it enables effective searching of encrypted data without the need for data decryption, which can be computationally expensive. Encrypted indexing does, however, have some restrictions. For instance, performing the encryption and search procedures might be computationally expensive, especially for huge datasets. The index structure might also not be as effective as conventional indexing methods, which could affect search performance. Overall,

Encrypted Indexing is a useful technique for securely indexing and searching sensitive data, such as XML documents. It allows for efficient searching of encrypted data while maintaining a high level of security and privacy. However, careful consideration must be given to the choice of encryption scheme and the impact on search performance.

C. Access Control Mechanisms

To implement security and privacy policies and limit access to the indexed data based on a set of rules or policies, Access Control Mechanisms can be employed in XML indexing for wireless data broadcasting as depicted in figure 4 [27].

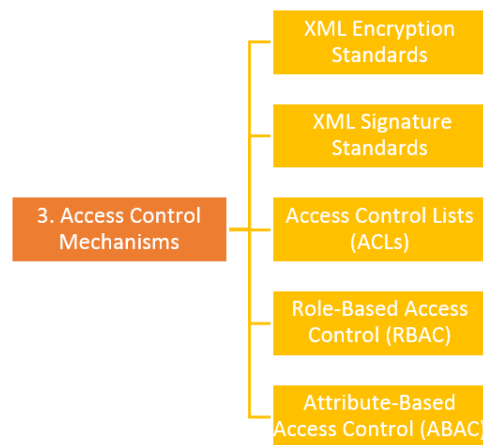


Figure 4: XML Access Control Mechanism Indexing Techniques

Using XML Encryption and XML Signature standards to encrypt and sign the indexed data is one method of access control in XML indexing. This can be used to guarantee the integrity and validity of the data and limit access to the indexed data to only authorized parties [28]. Another strategy is to limit access to the indexed data based on user roles or permissions using access control lists (ACLs) or role-based access control (RBAC). For instance, certain users might be given read-only access while others might be given read-write access to the indexed data [29]. In the context of Access Control Mechanisms, fine-grained access models like Attribute-Based Access Control (ABAC) complement traditional access control methods by allowing more granular control over access permissions. ABAC is an access control model that considers various attributes of entities involved in a system to make access decisions. Entities include subjects (users or processes), objects (resources), and the environment. XACML is a standard language for expressing ABAC policies. It defines a flexible framework for specifying access control policies based on attributes. Another method that can be utilized for Access Control in XML indexing is the XML Security Gateway. In this method, a security gateway is installed between the clients accessing the data and the XML data source, serving as a proxy. The gateway can include monitoring and auditing capabilities to track user access to the indexed data, as well as the ability to impose security policies including access control, authentication, and permission [2]. The security and privacy of indexed data in wireless data transmission are often ensured by access control mechanisms. Depending on the

requirements and policies of the system, the choice of access control mechanisms may combine encryption, signature, access control lists, role-based access control, and security gateways.

D. Anonymization Techniques

XML data can have sensitive information removed or obscured while still being valuable for analysis or indexing thanks to anonymization techniques. This is done to safeguard people's privacy and to abide by privacy laws like the GDPR or HIPAA [30]. Some of the Anonymization Techniques used for indexing XML data include in figure 5:

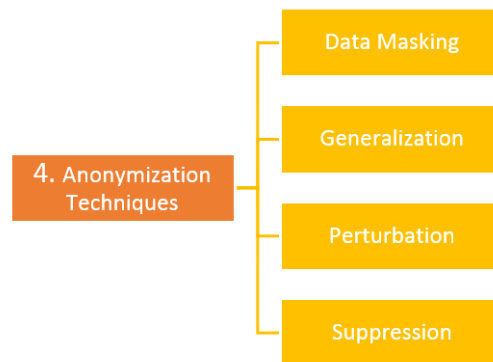


Figure 5: XML Anonymization Indexing Techniques

- (i) *Data Masking*: This technique involves replacing sensitive data, such as names or social security numbers, with a masking character, such as an asterisk or a hash. This technique preserves the structure of the XML data but obscures the sensitive information [30].
- (ii) *Generalization*: This technique involves replacing sensitive data with a more general value, such as replacing the exact age of a person with an age range (e.g. 20-30). This technique preserves the meaning of the data but obscures the specific details [31].
- (iii) *Perturbation*: This technique involves adding random noise to the sensitive data, such as adding a random value to a person's income. This technique preserves the statistical properties of the data but obscures the exact values [32].
- (iv) *Suppression*: This technique involves removing sensitive data entirely from the XML data, such as removing a person's address or phone number. This technique preserves the structure of the data but removes the sensitive information entirely [30].

The selection of an anonymization strategy may comprise a combination of techniques and will depend on the particular requirements and policies of the system. To ensure that the data is appropriately safeguarded while being valuable for indexing and analysis, it is crucial to carefully assess the effectiveness of the anonymization technique used.

E. Secure Data Transmission

Throughout the indexing process, Secure Data Transmission is employed to make sure that XML data is securely transported from the source to the destination. This prevents illegal access, interception, and manipulation of the data [33]. There are several techniques used for Secure Data Transmission of XML data, including in figure 6:

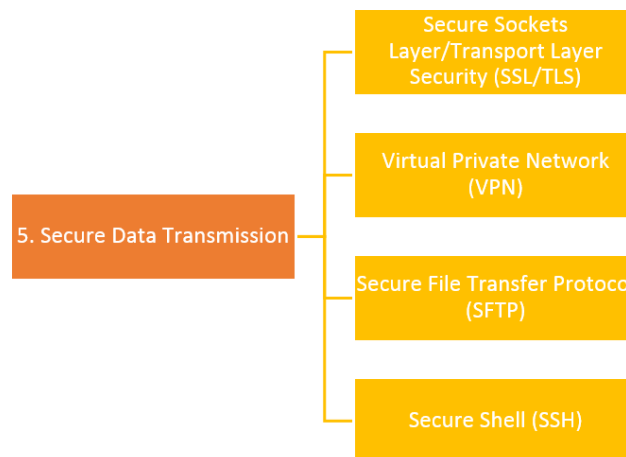


Figure 6: XML Secure Data Transmission Indexing Techniques

- (i) *Secure Sockets Layer/Transport Layer Security (SSL/TLS)*: This technique involves encrypting the XML data during transmission using SSL/TLS protocols. SSL/TLS protocols use a combination of public-key and symmetric-key encryption to ensure the confidentiality, integrity, and authenticity of the data [34].
- (ii) *Virtual Private Network (VPN)*: This technique involves creating a secure tunnel between the source and destination networks to transmit the XML data. VPNs use encryption and authentication to ensure the confidentiality and integrity of the data during transmission [35].
- (iii) *Secure File Transfer Protocol (SFTP)*: This technique involves encrypting the XML data using SSH protocols during transmission. SFTP provides strong authentication and encryption to ensure the security of the data [36].
- (iv) *Secure Shell (SSH)*: This technique involves encrypting the XML data during transmission using SSH protocols. SSH provides strong authentication and encryption to ensure the confidentiality and integrity of the data [37].

The system's requirements and rules will determine the best Secure Data Transmission method, which may comprise a combination of methods. To make sure that the data is adequately safeguarded during transmission, it is crucial to thoroughly assess the efficiency of the Secure Data Transmission technique utilized.

F. Regulatory Compliance

The process of verifying that the transmission of XML data through wireless networks complies with regulatory requirements and standards is known as regulatory compliance.

The security and privacy of the data, as well as avoiding the financial and legal repercussions of non-compliance, depend on adherence to regulatory requirements [38]. Some regulatory guidelines that may be relevant for indexing XML data for broadcasting over wireless channels are included in figure 7:

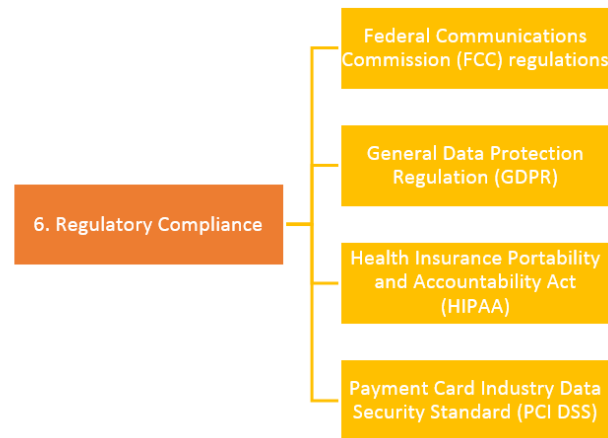


Figure 7: XML Regulatory Compliance Indexing Techniques

- (i) *Federal Communications Commission (FCC) regulations*: These regulations specify the requirements for wireless communication devices, including the use of licensed and unlicensed spectrum, power levels, and other technical requirements.
- (ii) *General Data Protection Regulation (GDPR)*: This regulation applies to the protection of personal data of European Union citizens, and requires that the data be processed and transmitted in a secure and compliant manner.
- (iii) *Health Insurance Portability and Accountability Act (HIPAA)*: This regulation applies to the protection of medical data, and requires that the data be transmitted in a secure and compliant manner.
- (iv) *Payment Card Industry Data Security Standard (PCI DSS)*: This standard applies to the protection of credit card data, and requires that the data be transmitted in a secure and compliant manner.

Using suitable security measures, such as encryption, access control, and authentication, to safeguard the data during transmission through wireless channels is crucial for ensuring compliance with these legal requirements. Regular audits and assessments should be performed to verify continued adherence to the pertinent regulatory requirements.

IV. RESULTS AND FINDINGS

E. Comparison of Secured XML indexing techniques based on their key features

Here is a comparison of the different secured XML indexing techniques based on their key features, advantages, and disadvantages:

Table 1: Comparison of Secured Xml Indexing Techniques Based On Key Features, Advantages, and Disadvantages

Indexing Scheme\Parameters	Key Features	Advantages	Gaps
Partial-Tree Indexing [18]	<ul style="list-style-type: none"> Indexes only a portion of the XML tree to improve efficiency and reduce storage requirements. 	<ul style="list-style-type: none"> Improve query performance Reduce indexing overhead, particularly for large XML documents. 	<ul style="list-style-type: none"> not be suitable for all types of XML data structures Result is incomplete or inaccurate if the indexed portion of the tree does not include all relevant information.
Encrypted Indexing [19][24]	<ul style="list-style-type: none"> Indexes encrypted XML data to protect it from unauthorized access 	<ul style="list-style-type: none"> Provides an additional layer of security for XML data, particularly sensitive information 	<ul style="list-style-type: none"> Add overhead and complexity to the indexing and querying process Require additional hardware or software to support encryption
Access Control Mechanisms [2][20][27]	<ul style="list-style-type: none"> Implements security controls to restrict access to XML data based on user roles and permissions 	<ul style="list-style-type: none"> Allows administrators to control access to XML data, Prevents unauthorized access and data breaches 	<ul style="list-style-type: none"> Complex to implement and maintain, particularly for large and complex systems
Anonymization Techniques [21][30]	<ul style="list-style-type: none"> Removes or obfuscates personally identifiable information (PII) from XML data to protect user privacy 	<ul style="list-style-type: none"> Reduces the risk of data breaches Protects user privacy, particularly for sensitive information 	<ul style="list-style-type: none"> Result is incomplete or inaccurate data if too much Difficult to implement for complex XML data structures
Secure Data Transmission [22][33]	<ul style="list-style-type: none"> Implements encryption and other security measures to protect XML data during transmission over wireless channels. 	<ul style="list-style-type: none"> Provides a secure method for transmitting XML data over wireless channels, reducing the risk of interception or unauthorized access. 	<ul style="list-style-type: none"> Add overhead and complexity to the transmission process Require additional hardware or software to support encryption.
Regulatory Compliance [23][38]	<ul style="list-style-type: none"> Ensures compliance with relevant regulatory guidelines and standards for the storage and transmission of XML data.[1] 	<ul style="list-style-type: none"> Ensures that XML data is stored and transmitted in a secure and compliant manner, reducing the risk of legal and financial penalties.[1] 	<ul style="list-style-type: none"> Add overhead and complexity to the indexing and querying process Require additional hardware or software to support compliance[1]

The research gaps between XML indexing techniques concerned about the security and privacy of XML data are described here to answer the first question i.e. “What are the gaps in the existing knowledge about XML data security and privacy-preserving XML data indexing techniques within the context of wireless data broadcasting?”. In conclusion, depending on the particular system requirements, each XML indexing technique has pros and cons. For an XML data indexing and broadcasting system to be secure and compliant, a mix of these methods might be necessary. To determine which indexing method is best for a specific system, it is critical to assess and contrast different methods based on significant criteria such as security, privacy, efficiency, scalability, complexity, flexibility, accuracy, usability, cost, and regulatory compliance.

A. Comparison of Secured XML Indexing Techniques Based on Concern Parameters

Several parameters can be used to compare various XML indexing techniques that address security and privacy issues in wireless data broadcasting. Some of these parameters include:

- (i) *Security*: The effectiveness of the technique in providing secure storage and transmission of XML data.
- (ii) *Privacy*: The ability of the technique to protect the privacy of XML data by preventing unauthorized access, interception, or disclosure.
- (iii) *Efficiency*: The performance of the technique in terms of indexing and querying speed, storage requirements, and processing overhead.
- (iv) *Scalability*: The ability of the technique to handle large volumes of XML data and support a growing number of users.
- (v) *Complexity*: The complexity of the technique in terms of implementation, configuration, and maintenance.
- (vi) *Flexibility*: The ability of the technique to support different types of XML data structures and query types.
- (vii) *Accuracy*: The accuracy of the technique in terms of indexing and querying results.
- (viii) *Usability*: The ease of use of the technique for developers and end-users.
- (ix) *Cost*: The cost of implementing and maintaining the technique, including hardware, software, and personnel costs.
- (x) *Regulatory compliance*: The extent to which the technique complies with relevant regulatory guidelines and standards.

It is feasible to assess and compare different XML indexing strategies for how well they solve security and privacy concerns in wireless data broadcasting by taking these factors into account. The technique selected will be determined by the particular guidelines and rules of the system as well as the relative weight of each parameter.

Table 2: Comparison of Various Secured Xml Indexing Techniques Based On Various Parameters

XML Techniques	Security	Privacy	Efficiency	Scalability	Complexity	Flexibility	Accuracy	Usability	Cost
Partial-Tree Indexing [18]	Provides weak security not as strong as encrypted indexing or access control mechanisms.	Does not provide any privacy protection.	Improves query performance and reduces indexing overhead	Improve scalability for large XML documents.	Simple to implement and maintain.	Not suitable for all types of XML data structures.	Incomplete or inaccurate results if only a subset of the XML document is indexed.	More user-friendly as they require minimal user interaction.	O(log h) where h is the height of the tree. Divided into non-partial and partial tree.
Encrypted Indexing [19][24]	Provides strong security by encrypting XML data, ensuring that it can only be accessed by authorized users.	Provides privacy protection by encrypting XML data, ensuring that only authorized users can access it.	Add overhead and complexity to the indexing and querying process.	Require additional hardware or software to support encryption and may affect scalability	Complex to implement and maintain, particularly for large and complex systems.	Used with any type of xml data structure.	Does not impact the accuracy of the xml data	More user-friendly as they require minimal user interaction.	Traditional dsse achieves optimal bandwidth overhead of o(r) , where r is the number of data corresponding with the search/update query [41].
Access Control Mechanisms [2][20][27]	Provides strong security by controlling access to XML data based on user roles and permissions.	Provides some level of privacy protection by controlling access to XML data based on user roles and permissions.	Complex to implement and maintain, particularly for large and complex systems.	Challenging to scale for large and complex systems.	Complex to implement and maintain, particularly for large and complex systems.	Customized to different user roles and permissions.	Improve accuracy by ensuring that users only see xml data that they are authorized to access.	require users to provide authentication information	The proposed labeling scheme is superior in terms of the number of nodes used to search for the proper location of a node, and the location searching time. [42]

<p>Anonymization Techniques[21][30]</p>	<p>Provides weak security by removing personally identifiable information from XML data.</p>	<p>Provides privacy protection by removing personally identifiable information from XML data.</p>	<p>Result in incomplete or inaccurate data if too much information is removed and may be difficult to implement for complex XML data structures.</p>	<p>Be difficult to scale for complex XML data structures</p>	<p>Complex to implement and maintain, particularly for complex XML data structures</p>	<p>Not suitable for all types of XML data structures</p>	<p>Impact accuracy if too much information is removed particularly for complex XML data structures.</p>	<p>require additional processing to remove sensitive information</p>	<p>The d-dependency privacy scheme provides additional protection beyond current privacy-preserving properties. [44][45]</p>
<p>Secure Data Transmission[22][33]</p>	<p>Provides security during transmission of XML data over wireless channels by implementing encryption and other security measures.</p>	<p>Provides privacy protection during transmission of XML data over wireless channels by implementing encryption and other security measures.</p>	<p>Adds overhead and complexity to the transmission process.</p>	<p>Affect scalability for large XML documents.</p>	<p>Complex to implement and maintain, particularly for large and complex systems.</p>	<p>Used with any type of XML data structure.</p>	<p>Does not impact the accuracy of the XML data.</p>	<p>Impact usability due to the overhead of encryption and decryption.</p>	<p>More expensive to implement and maintain due to the additional hardware or software required to support secure transmission.</p>
<p>Regulatory Compliance[23][38]</p>	<p>Ensures security by storing and transmitting XML data in a secure and compliant manner.[1]</p>	<p>Ensures security by storing and transmitting XML data in compliance with relevant privacy regulations.[1]</p>	<p>Add overhead and complexity to the indexing and querying process.</p>	<p>Affect scalability for large and complex systems.</p>	<p>Add complexity to the implementation and maintenance of xml indexing systems.</p>	<p>Customized to comply with different regulatory guidelines and standards</p>	<p>Ensures that xml data is stored and transmitted accurately in compliance with relevant regulations and standards.[1]</p>	<p>XML enables data interoperability. [46]</p>	<p>The LRSL can be used to shape the conditionality of regulatory coverage. [46]</p>

V. ANALYSIS

Overall, the analysis shows that each XML indexing method has pros and cons in terms of security, privacy, efficiency, scalability, complexity, flexibility, accuracy, usability, cost, and compliance with legal requirements. For big XML documents, partial-tree indexing is effective and scalable, but it offers no security and no privacy protection.

Even though encrypted indexing offers high protection and privacy, it may complicate and burden the indexing process. Strong security and some privacy protection are provided by access control mechanisms, but their implementation and upkeep can be challenging. Although some amount of security and privacy is provided by anonymization techniques, doing so for intricate XML data structures can be challenging.

While ensuring confidentiality and privacy during transmission, secure data transfer may also increase complexity and overhead. Regulatory Compliance measures ensure compliance with industry security standards but may add complexity and overhead to the implementation process.

All indexing methods are capable of delivering great accuracy in terms of data retrieval and query performance. Nevertheless, partial-tree indexing and encrypted indexing approaches can provide greater accuracy than other methods because they offer improved indexing strategies that make it possible to search for and retrieve particular data pieces quickly.

Access Control Mechanisms and Anonymization Methods can influence usability because they may make it more difficult for users to obtain data. Because they need less user input, Partial-Tree Indexing and Encrypted Indexing might be more user-friendly. The overhead of encryption and decryption may have an impact on usability during secure data transmission.

Cost varies according to the implementation's specifics and required levels of security and privacy, with Anonymization Methods and Encrypted Indexing perhaps being more expensive due to higher processing demands.

Due to the requirement for additional hardware or software to handle access, access control mechanisms may also result in cost increases. Measures for Regulatory Compliance, Secure Data Transfer, and Partial-Tree Indexing are typically less expensive to adopt. Last but not least, Regulatory Compliance is essential for guaranteeing adherence to industry security standards, and it is frequently achieved through the use of Access Control Mechanisms, Anonymization Methods, and Encrypted Indexing.

Furthermore, crucial to compliance are partial-tree indexing, secure data transmission, and regulatory compliance methods, the latter of which are created particularly to satisfy regulatory criteria. And may be the most effective solution for ensuring compliance.

VI. CONCLUSION

Two research questions have been raised and answered in this paper. The first question is “What are the gaps in the existing knowledge about XML data security and privacy-preserving XML data indexing techniques within the context of wireless data broadcasting?”. To answer this question, a detailed study has been done and highlights all the gaps between XML indexing techniques. The second question is “Can a one-size-fits-all security approach be identified to effectively tackle security and privacy concerns in the indexing technique for wireless XML data broadcasting?” To answer this question, the choice of a suitable XML indexing method depends on the organization's unique needs. It includes those for security, privacy, efficiency, scalability, complexity, flexibility, accuracy, usability, cost, and regulatory compliance.

Before deploying an indexing solution, organizations must thoroughly assess their unique demands. Each technique has advantages and disadvantages. It's also crucial to remember that combining various indexing strategies can result in a perfect answer that satisfies the demands of both security and legal compliance. Finally, while preserving the confidentiality and privacy of sensitive data, a well-designed and executed indexing system can considerably increase the efficiency and efficacy of XML data retrieval and query performance.

Statements and Declarations

Conflict of Interest: The authors declare that they have no conflict of interest.

References

- 1) Sen, J., “Security and Privacy Issues in Cloud Computing Computing”. *Innovation Labs, Tata Consultancy Services Ltd.*, Kolkata, INDIA. <https://doi.org/10.4018/978-1-4666-6539-2.ch074>
- 2) Crampton J. “Applying hierarchical and role-based access control to XML documents”. In Proceedings of the 2004 workshop on Secure web service 2004 Oct 29 (pp. 37-46). <https://doi.org/10.1145/1111348.1111353>
- 3) Runapongsa, K., Patel, J.M., Jagadish, H.V., Chen, Y. and Al-Khalifa, S., 2006. “The Michigan benchmark: towards XML query performance diagnostics”. *Information Systems*. 2006; 31(2): 73-97. <https://dl.acm.org/doi/abs/10.5555/1126965.1711914>
- 4) Catania B, Maddalena A, Vakali A. “XML document indexes: a classification”. *IEEE Internet Computing*. 2005 Sep 19;9(5):64-71. <https://doi.org/10.1109/MIC.2005.115>
- 5) Abdullah F, Peng L, Tak B. “A Survey of IoT Stream Query Execution Latency Optimization within Edge and Cloud”. *Wireless Communications and Mobile Computing*. 2021 Nov 16; 2021:1-6. <https://doi.org/10.1155/2021/4811018>
- 6) He Q, Otto P, Anton AI, Jones L. “Ensuring compliance between policies, requirements, and software design: A case study”. In Fourth IEEE International Workshop on Information Assurance (IWIA'06) 2006 Apr 13 (pp. 14-pp). IEEE. <https://doi.org/10.1109/IWIA.2006.7>
- 7) Jang-Jaccard J, Nepal S. “A survey of emerging threats in cybersecurity”. *Journal of computer and system sciences*. 2014 Aug 1; 80(5):973-93. <https://doi.org/10.1016/j.jcss.2014.02.005>
- 8) Abdulsalam YS, Hedabou M. “Security and privacy in cloud computing: a technical review”. *Future Internet*. 2021 Dec 27; 14(1):11. <https://doi.org/10.3390/fi14010011>

- 9) Yimam D, Fernandez EB. "A survey of compliance issues in cloud computing". *Journal of Internet Services and Applications*. 2016 Dec; 7:1-2. <https://doi.org/10.1186/s13174-016-0046-8>
- 10) Abouelmehdi K, Beni-Hessane A, Khaloufi H. "Big healthcare data: preserving security and privacy". *Journal of big data*. 2018 Dec; 5(1):1-8. <https://doi.org/10.1186/s40537-017-0110-7>
- 11) Jain P, Gyanchandani M, Khare N. "Big data privacy: a technological perspective and review". *Journal of Big Data*. 2016 Dec; 3:1-25. <https://doi.org/10.1186/s40537-016-0059-y>
- 12) Wheelus C, Zhu X. "IoT network security: Threats, risks, and a data-driven defense framework". *IoT*. 2020 Oct 19; 1(2):259-85. <https://doi.org/10.3390/iot1020016>
- 13) Pyo CK, Yun SJ, Ryu GS. "XML Indexing Techniques for Handling Large Amounts of Data". *Indian Journal of Science and Technology*. 2016 Oct; 9:40. <https://doi.org/10.17485/ijst/2016/v9i40/103278>.
- 14) Goel V, Panwar G, Ahlawat AK. "Energy efficient air indexing schemes for single and multi-level wireless channels". In 2013 3rd IEEE International Advance Computing Conference (IACC) 2013 Feb 22 (pp. 525-530). IEEE. <https://doi.org/10.1109/IAdCC.2013.6514281>
- 15) Gautam D, Goel V. "XML-based streaming strategies for indexing the wireless broadcast data". In 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE) 2016 Sep 22 (pp. 629-634). IEEE. <https://doi.org/10.1109/ICMETE.2016.54>.
- 16) Goel V, Gautam D, Gupta A, Kumar S. "An improvised indexing technique for XML data over multiple channels in the wireless environment: (1, Xm) method". *International Journal of Communication Systems*. 2019 Nov 10; 32(16):e4122. <https://doi.org/10.1002/dac.4122>
- 17) Guo L, Wu H. "An XML Privacy-Preserving Data Disclosure Decision Scheme". *Security and Communication Networks*. 2022 Feb 24; 2022. <https://doi.org/10.1155/2022/9099722>.
- 18) Hao W, Matsuzaki K. "A partial-tree-based approach for XPath query on large XML trees". *Journal of Information Processing*. 2016; 24(2):425-38. <https://doi.org/10.2197/ipsjip.24.425>
- 19) Fathi L, Ibrahim H, Mirabi M. "An air indexing method for encrypted XML data broadcast in a mobile wireless network". In Proceedings of the 12th International ACM Workshop on Data Engineering for Wireless and Mobile Access 2013 Jun 23 (pp. 28-35). <https://doi.org/10.1145/2486084.2486090>
- 20) Sandhu RS, Samarati P. "Access control: principle and practice". *IEEE Communications magazine*. 1994 Sep; 32(9):40-8. <http://dx.doi.org/10.1109/35.312842>
- 21) Murthy S, Bakar AA, Rahim FA, Ramli R. "A comparative study of data anonymization techniques". In 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) 2019 May 27 (pp. 306-309). IEEE. <https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2019.00063>
- 22) Seddigh N, Nandy B, Makkar R, Beaumont JF. "Security advances and challenges in 4G wireless networks". In 2010 Eighth International Conference on Privacy, Security and Trust 2010 Aug 17 (pp. 62-71). IEEE. <https://doi.org/10.1109/PST.2010.5593244>
- 23) Sattarova Feruza Y, Kim TH. "IT security review: Privacy, protection, access control, assurance, and system security". *International journal of multimedia and ubiquitous engineering*. 2007 Apr; 2(2):17-32. <http://dx.doi.org/10.14257/ijmue.2007.2.2.02>
- 24) Ünay O, Gündem Tİ. "A survey on querying encrypted XML documents for databases as a service". *ACM SIGMOD Record*. 2008 Mar 1; 37(1):12-20. <https://doi.org/10.1145/1374780.1374783>
- 25) Saha TK, Rathee M, Koshiya T. "Efficient private database queries using ring-LWE somewhat homomorphic encryption". *Journal of Information Security and Applications*. 2019 Dec 1; 49:102406. <https://doi.org/10.1016/j.jisa.2019.102406>

- 26) Sepehri M, Cimato S, Damiani E. "Privacy-preserving query processing by multi-party computation". *The Computer Journal*. 2015 Oct 1; 58(10):2195-212. <https://doi.org/10.1093/comjnl/bxu093>
- 27) Nehme RV, Rundensteiner EA, Bertino E. "A security punctuation framework for enforcing access control on streaming data". In 2008 IEEE 24th International Conference on Data Engineering 2008 Apr 7 (pp. 406-415). IEEE. <https://doi.org/10.1109/ICDE.2008.4497449>
- 28) Liu B. XML Security in XML Data Integrity, Authentication, and Confidentiality (Doctoral dissertation, University of Huddersfield). <https://eprints.hud.ac.uk/id/eprint/9671/>
- 29) Samarati P, de Vimercati SC. "Access control: Policies, models, and mechanisms". In International school on foundations of security analysis and design 2000 Sep 18 (pp. 137-196). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45608-2_3.
- 30) El Ouazzani Z, El Bakkali H. "A classification of non-cryptographic anonymization techniques ensuring privacy in big data". *International Journal of Communication Networks and Information Security*. 2020 Apr 1; 12(1):142-52. <https://doi.org/10.17762/ijcnis.v12i1.4401>
- 31) Kiran P, Kavya NP. "A survey on methods, attacks and metric for privacy-preserving data publishing". *International Journal of Computer Applications*. 2012 Jan 1; 53(18). <https://doi.org/10.1109/ACCESS.2017.2706947>
- 32) Al-Rubaie M, Chang JM. "Privacy-preserving machine learning: Threats and solutions". *IEEE Security & Privacy*. 2019 Mar 29; 17(2):49-58. <https://doi.org/10.1109/MSEC.2018.2888775>
- 33) Lv T, Yan P. "A web security solution based on XML technology". In 2006 International Conference on Communication Technology 2006 Nov 27 (pp. 1-4). IEEE. <https://doi.org/10.1109/ICCT.2006.341975>
- 34) Eteng IE. "A Multilayer Secured Messaging Protocol for REST-based Services". *Journal of International Technology and Information Management*. 2019; 28(3):43-66. <https://doi.org/10.58729/1941-6679.1378>
- 35) Conti M, Hasani A, Crispo B. "Virtual private social networks". In Proceedings of the first ACM conference on Data and application security and privacy 2011 Feb 21 (pp. 39-50). <https://doi.org/10.1145/1943513.1943521>
- 36) Charest G, Rogers M, Planning S, Leader EA, Sevier R, Thomas M, Clevenger J, Tikofsky B. Data Exchange Methods and Considerations. Technical report, Version: 1.0, Last Revised February 07, 2020, Cambridge, MA: Harvard University; 2020. https://enterprisearchitecture.harvard.edu/files/enterprise/files/data_exchange_advisory_v1_final.pdf
- 37) Wasserman M. Using the netconf protocol over secure shell (ssh). 2011 Jun. <https://datatracker.ietf.org/doc/rfc6242/>
- 38) Ramgovind S, Eloff MM, Smith E. "The management of security in cloud computing". In 2010 Information Security for South Africa 2010 Aug 2 (pp. 1-7). IEEE. <https://doi.org/10.1109/ISSA.2010.5588290>
- 39) Shekarriz, M., Babamir, S.M. and Mirabi, M. "Query processing optimization in broadcasting XML data in mobile communications". *The Journal of Supercomputing*. 2021:77, 5354-5380. <https://doi.org/10.1007/s11227-020-03479-5>
- 40) Jaya Lakshmi, A. and Joe Prathap, P.M., 2023. "An Energy-Efficient Approach to Transfer Data from WSN to Mobile Devices". In Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022 (pp. 127-137). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5443-6_11

- 41) A. A. Yavuz and J. Guajardo. "Dynamic searchable symmetric encryption with minimal leakage and efficient updates on commodity hardware". In Selected Areas in Cryptography – SAC Aug 2015:9566. (pp.20-45), Lecture Notes in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-31301-6_15
- 42) Shammar, E.A., Zahary, A.T. and Al-Shargabi, A.A., 2021. "A survey of IoT and blockchain integration: Security perspective". IEEE Access, 9, pp.156114-156150.<https://doi.org/10.1109/ACCESS.2021.3129697>.
- 43) Ko, H.K., Kim, M.J. and Lee, S. "On the efficiency of secure XML broadcasting". *Information Sciences*, 2007: 177(24), pp.5505-5521, <https://doi.org/10.1016/j.ins.2007.05.020>.
- 44) Landberg, A.H., Nguyen, K., Pardede, E. and Rahayu, J.W. "δ-dependency for privacy-preserving XML data publishing". *Journal of Biomedical Informatics*, 2014:50, pp.77-94, <https://doi.org/10.1016/j.jbi.2014.01.013>.
- 45) Cunha, M., Mendes, R. and Vilela, J.P. "A survey of privacy-preserving mechanisms for heterogeneous data types". *Computer science review*, 2021:41, p.100403, <https://doi.org/10.1016/j.cosrev.2021.100403>.
- 46) Breaux, T.D. and Gordon, D.G. "Regulatory requirements traceability and analysis using semi-formal specifications". In Requirements Engineering: Foundation for Software Quality: 19th International Working Conference, REFSQ 2013, Essen, Germany, April 8-11, 2013. Proceedings 19 (pp. 141-157). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-37422-7_11