A SEARCHABLE AND VERIFIABLE DATA PROTECTION SCHEME FOR SCHOLARLY BIG DATA

[1]Busa Revathi, [2]Dr Gulab singh chauhan
[1]mtech, Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA, H.no: 20S41D5804, busarevathi674@gmail.com
[2]Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA

ABSTRACT: Scientific research achievements play a positive role in the promotion of social development. Scholarly big data include scholars' scientific research, experimental data, and their own identity information. The security of scholarly big data relates to the authors' reputation and the copyright of their works. This paper proposes a trusted third-party-aided searchable and verifiable data protection scheme that utilizes cloud computing technology. For a better description of the protocol, we first present a user-differentiated system model and a cube data storage structure. On the basis of the novel system model and data structure, the scheme helps the users review the integrity of their uploaded or downloaded data at any time and search the online scholarly data with encrypted keywords. The security analysis and performance simulation demonstrate that the novel scheme is a secure and efficient scheme for scholarly big data applications.

***Keywords** – Scholarly big data, security scheme, searchable encryption, data integrity, cloud computing*

1. INTRODUCTION

The development of society is inseparable from scientific and technological progress, both of which must rely on theoretical innovation and upgrades. Scientific research from all fields involves all aspects of people's lives. With the continuous development of research in various fields and the emergence of new fields, the achievements of various fields are becoming increasingly abundant. There are many reasons for the increase of scholarly data, including an increase in the number of scholars, the complexity of the networks of scholars, the diversification of magazines and journals, the growing readership and the continuous expansion of professional fields. People care about the development of scholarly big data because these data are related to the quality of life in the coming decades or even hundreds of years [1], [2]. The formation of scholarly big data is actually the result of the great development of scientific theory research [3]. Scholarly big data include research scholars' personal information, papers, experimental data sets, and results. These data may include information regarding the authors' privacy and social relationships, copyright and right of authorship, and experimental data related to personal privacy, such as medical data [4], [5].

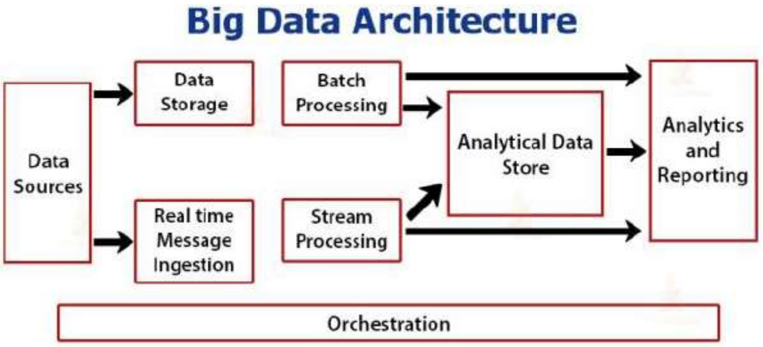


Fig.1: Example figure

These data are complex and extremely important. If the coauthor of an academic achievement is likely to be tampered with, the author's academic reputation may be affected. If malicious users are using legitimate users' identities to upload data to the system, the researchers' results may be tampered with or replaced. The keywords with which users search on academic sites may reveal the authors' latest ideas or the readers' private information. For example, a search for a disease and its associated research dynamics may reveal a reader's physical condition. Therefore, scholarly big data are not only a significant asset for researchers but also an important lifeline for data users. It is thus crucial to develop new approaches for the research, development, and conservation of scholarly big data [6]–[9].

2. LITERATURE REVIEW

Fair energy scheduling for vehicle-to-grid networks using adaptive dynamic program

Research on the smart grid is being given enormous supports worldwide due to its great significance in solving environmental and energy crises. Electric vehicles (EVs), which are powered by clean energy, are adopted increasingly year by year. It is predictable that the huge charge load caused by high EV penetration will have a considerable impact on the reliability of the smart grid. Therefore, fair energy scheduling for EV charge and discharge is proposed in this paper. By using the vehicle-to-grid technology, the scheduler controls the electricity loads of EVs considering fairness in the residential distribution network. We propose contribution-based fairness, in which EVs with high contributions have high priorities to obtain charge energy. The contribution value is defined by both the charge/discharge energy and the timing of the action. EVs can achieve higher contribution values when discharging during the load peak hours. However, charging during this time will decrease the contribution values seriously. We formulate the fair energy scheduling problem as an infinite-horizon Markov decision process. The methodology of adaptive dynamic programming is employed to maximize the long-term fairness by processing online network training. The numerical results illustrate that the proposed EV energy scheduling is able to mitigate and flatten the peak load in the distribution network. Furthermore, contribution-based fairness achieves a fast recovery of EV batteries that have deeply discharged and guarantee fairness in the full charge time of all EVs.

AlgorithmSeer: A System for Extracting and Searching for Algorithms in Scholarly Big Data

Algorithms are usually published in scholarly articles, especially in the computational sciences and related disciplines. The ability to automatically find and extract these algorithms in this increasingly vast collection of scholarly digital documents would enable algorithm indexing, searching, discovery, and analysis. Recently, AlgorithmSeer , a search engine for algorithms, has been investigated as part of CiteSeer ^{X} with the intent of providing a large algorithm database. Currently, over 200,000 algorithms have been extracted from over 2 million scholarly documents. This paper proposes a novel set of scalable techniques used by AlgorithmSeer to identify and extract algorithm representations in a heterogeneous pool of scholarly documents. Specifically, hybrid machine learning approaches are proposed to discover algorithm representations. Then, techniques to extract textual metadata for each algorithm are discussed. Finally, a demonstration version of AlgorithmSeer that is built on Solr/Lucene open source indexing and search system is presented.

Compressive sensing of piezoelectric sensor response signal for phased array structural health monitoring

There are three steps for compressive sensing, such as the sparse representation of signal, the design of observation matrix and the reconstruction of signal. The existing observation matrix may lose the part information of the original signal after compressing and sampling. Then, the adaptive observation matrix is proposed to sparse samples for ultrasonic wave signals that are analysed in the phased array structural health monitoring. The matrix is generated adaptively according to the information of the sparse coefficients vector, so that the sparse signal can include all the information of the original signal after compressing and sampling. Moreover, the orthogonal matching pursuit (OMP) algorithm will be used in reconstructing the ultrasonic wave signal with high probability. Finally, experiments were carried out on the aluminium plate. The proposed method that based on the adaptive observation matrix can effectively reduce the reconstruction error and more accurately and completely reconstruct the se..

Direction density-based secure routing protocol for healthcare data in incompletely predictable networks

Healthcare data are becoming increasingly important in the life of people. By utilizing healthcare data in a proper and secure manner, the elderly may avoid some sudden diseases, whereas young people can monitor their health condition. In the hospital, for certain sizes of detection objects, an effective method of data transmission becomes very significant. In view of the movement of patients in the hospital, we introduce a type of network called incompletely predictable networks to describe such scenarios. The patients move in a certain trend or are only active in a certain limited range. To achieve high performance when transmitting healthcare data in such networks, a novel protocol called the direction density-based secure routing protocol is proposed in this paper. Both the moving direction and the influence of node group movement are considered. The novel protocol innovatively takes the density of the node moving direction into consideration, which makes full use of the relationships among the moving individuals. Moreover, the design of the secure routing with authenticated message transmission ensures secure healthcare data communication. The simulation shows that our protocol achieves a high packet delivery ratio with low overhead and end-to-end delay.

Social big data based content dissemination in internet of vehicles

By analogy with Internet of things, Internet of vehicles (IoV) that enables ubiquitous information exchange and content sharing among vehicles with little or no human intervention is a key enabler for the intelligent transportation industry. In this paper, we study how to combine both the physical and social layer information for realizing rapid content dissemination in device-to-device vehicle-to-vehicle (D2D-V2V)-based IoV networks. In the physical layer, headway distance of vehicles is modeled as a Wiener process, and the connection probability of D2D-V2V links is estimated by employing the Kolmogorov equation. In the social layer, the social relationship tightness that represents content selection similarities is obtained by Bayesian nonparametric learning based on real-world social big data, which are collected from the largest Chinese microblogging service Sina Weibo and the largest Chinese video-sharing site Youku. Then, a price-rising-based iterative matching algorithm is proposed to solve the formulated joint peer discovery, power control, and channel selection problem under various quality-of-service requirements. Finally, numerical results demonstrate the effectiveness and superiority of the proposed algorithm from the perspectives of weighted sum rate and matching satisfaction gains.

3. METHODOLOGY

Orencik et al. present a privacy-preserving searchable scheme for encrypted data using queries with multiple keywords. Additionally, the scheme can hide the search patterns and provide an effective scoring and ranking capability. Focusing on the range query problem, Jho et al. present a novel searchable encryption protocol that providesan efficient range query by utilizing symmetric key encryption systems.

Miao et al. present a scheme that can achieve verifiable conjunctive keywords search of encrypted data without a secure channel. The scheme is proved to ensure data integrity and availability.

Ma et al. propose an efficient certificateless public key encryption scheme with multiple keywords search for industrial Internet of Things. They prove that their protocol is secure and has an acceptable communication cost.

Huang et al. present a public-key searchable encryption scheme that can solve the issue called inside keyword guessing attacks. They prove that the server in their scheme cannot encrypt a keyword itself or launch an inside keyword guessing attack successfully.

Disadvantages:

- In the existing work, the system leaks significant information for updates and it is not parallelizable.
- The existing system doesn't provide discretization of scholarly big data, a novel data structure is required to store these data for convenient encrypted searching.

A trusted third-party-aided user differentiated system model is presented. The system model presented in this paper aims to meet different application requirements from different user identities related to scholarly big data and with different technological implementation processes. Based on the different needs of the users, we divide the users into three identities: authors, editors and readers. Authors can upload their papers, procedures, experimental data and steps, and personal information. Editors can upload information about their journals, calls for papers and special issues and personal information. Valid readers have the right to search for scholarly data on the cloud by using encrypted keywords and require the system to provide an integrity verification of those data. With the help of a trusted third party, the system verifies the identities of these users and helps them search and validate the data.

A cube data storage structure is utilized. For more convenient storage and searching, we design a new cube data storage structure to store scholarly big data. This structure can not only verify the integrity of different user data separately but also effectively distinguish and store the keywords of different data
blocks.

A searchable and verifiable data protection scheme is proposed. The scheme innovatively implements data integrity verification for secure search results. In
other words, the scheme guarantees the integrity of searched data while implementing a secure search. Moreover, the scheme is scalable. Authors and editors
can also utilize the proposed scheme to obtain an integrity verification report of their own uploaded data to protect their interests. One of the most important innovations of this paper lies in the participation of keywords in the integrity verification process. This scheme ensures both the secure search function and data integrity and that the keywords are not tampered with.

Advantages:

- The system is more effective since design a suitable protection scheme to achieve better utilization of scientific research achievements.
- The system is more secured since an efficient system model must be designed to adapt to the application environment of scholarly big data

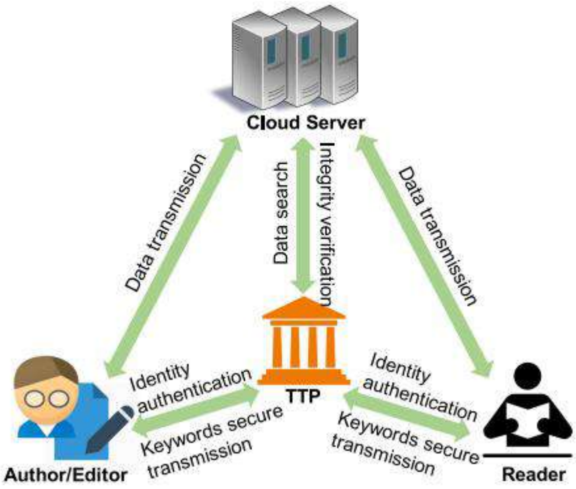


Fig.2: System architecture

4. IMPLEMENTATION

The major modules of the project are

- **Author**
  In this module, the author uploads their encrypted data in the Cloud server. For the security purpose the user encrypts the data file and then store in the server. The User can have capable of manipulating the encrypted data file and performs the following operations Upload File, Verify Block, Update Block, Delete File, View Files, View Verification.
- **Cloud Server**
  The **Cloud** server manages which is to provide data storage service for the Data Owners. Data owners encrypt their data files and store them in the Server for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the Server and then Server will decrypt them. The server will generate the aggregate key if the end user requests for file authorization to access and performs the following operations such as View Readers, View

Authors, View All Author Files, View File Requests, View Attackers, View Blocked Users, View Domains, View Time Delay Results, View Throughput Results.

Reader

In this module, the reader can only access the data file with the secret key. The user can search the file for a specified keyword and end user and can do the following operations like Register and Login, Search File, Request File, View File Response.

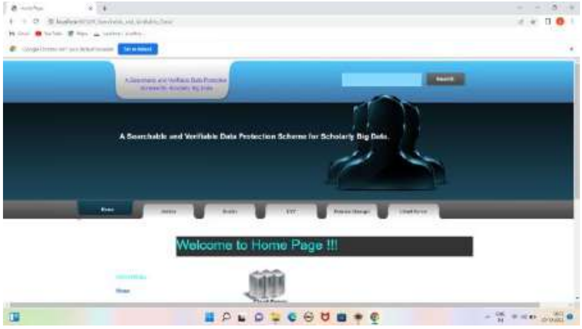
- **TTP**

  In this module, the ttp performs the following operations View Meta Data
  ,Verify All Blocks, Send Block Verification

- **Domain Manager**

In this module, the domain manager performs the following operations Add Domains, View All Domains, View verified Blocks.

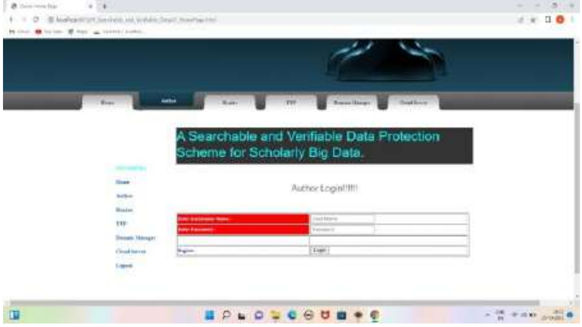
5. EXPERIMENTAL RESULTS



A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

Welcome to Home Page !!!

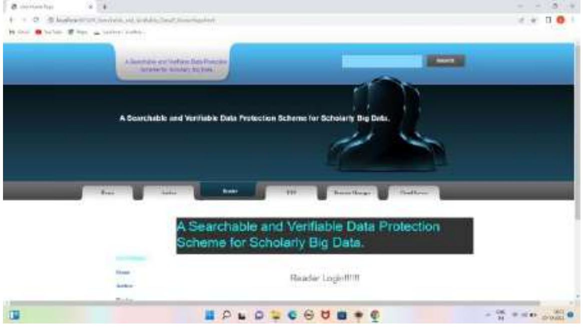
Fig.3: Home screen



A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

Author Login!!!!

Fig.4: Author login

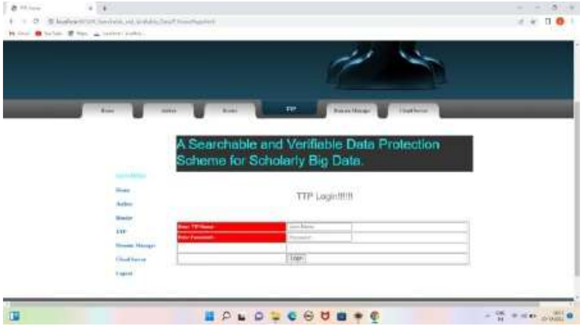


A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

Reader Login!!!!

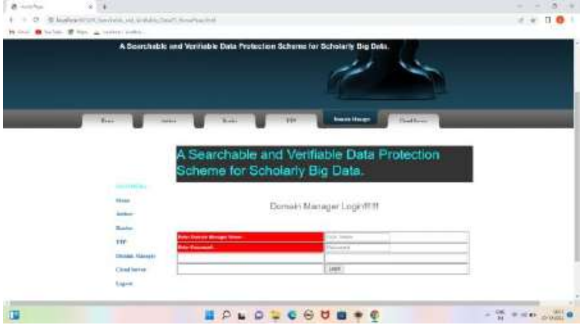
Fig.5: reader login


A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

TTP Login!!!!!!

Enter TTP Name:	
Enter Password:	
	Login

Fig.6: TTP login

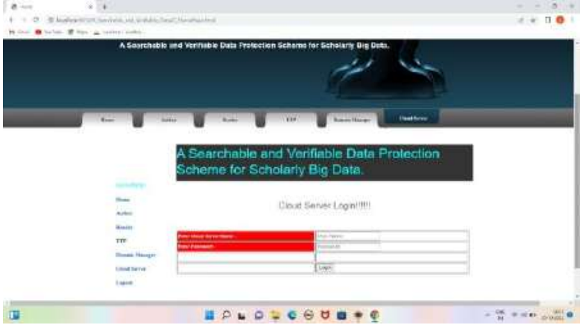

A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

Domain Manager Login!!!!!!

Enter Domain Manager Name:	
Enter Password:	
	Login

Fig.7: Domain manager login


A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

A Searchable and Verifiable Data Protection Scheme for Scholarly Big Data.

Cloud Server Login!!!!!!

Enter Cloud Server Name:	
Enter Password:	
	Login

Fig.8: Cloud server login

6. CONCLUSION

In this paper, we construct a system model that can distinguish the users according to their roles and special requirements of scholarly big data. Moreover, an innovative cube data storage structure is proposed. On the basis of the novel system and data structure, we present a novel searchable and verifiable data protection scheme for scholarly big data. The security and performance analyses show that our scheme is efficient for scholarly big data.

In the future, we will design a secure data sharing scheme for scholarly big data to supplement our current scheme.

REFERENCES

[1] S. Xie, W. Zhong, K. Xie, R. Yu, and Y. Zhang, “Fair energy scheduling for vehicle-to-grid networks using adaptive dynamic programming,” IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 8, pp. 1697–1707, Aug. 2016.

[2] Y. Zhang, R. Yu, S. Xie, and W. Yao, “Home M2M networks: Architectures, standards, and QoS improvement,” IEEE Commun. Mag., vol. 49, no. 4, pp. 44–52, Apr. 2011.

[3] S. Tuarob, S. Bhatia, P. Mitra, and C. L. Giles, “AlgorithmSeer: A system for extracting and searching for algorithms in scholarly big data,” IEEE Trans. Big Data, vol. 2, no. 1, pp. 3–17, Mar. 2016.

[4] Y. Sun and F. Gu, “Compressive sensing of piezoelectric sensor response signal for phased array structural health monitoring,” Int. J. Sensor Netw., vol. 23, no. 4, pp. 258–264, 2017.

[5] J. Shen, C. Wang, C.-F. Lai, A. Wang, and H.-C. Chao, “Direction densitybased secure routing protocol for healthcare data in incompletely predictable networks,” IEEE Access, vol. 4, pp. 9163–9173, Dec. 2016.

[6] F. Xia, W. Wang, T. M. Bekele, and H. Liu, “Big scholarly data: A survey,” IEEE Trans. Big Data, vol. 3, no. 1, pp. 18–35, Mar. 2017.

[7] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, “Social big data based content dissemination in internet of vehicles,” IEEE Trans. Ind. Informat., vol. 14, no. 2, pp. 768–777, Feb. 2018.

[8] W.-L. Chin, W. Li, and H.-H. Chen, “Energy big data security threats in IoT-based smart grid communications,” IEEE Commun. Mag., vol. 55, no. 10, pp. 70–75, Oct. 2017.

[9] Y. Xiang, W. Zhou, and M. Guo, “Flexible deterministic packet marking: An IP traceback system to find the real source of attacks,” IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 4, pp. 567–580, Apr. 2009.

[10] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, “New algorithms for secure outsourcing of modular exponentiations,” IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 9, pp. 2386–2396, Sep. 2014.