A DATA SHARING PROTOCOL TO MINIMIZE SECURITY AND PRIVACY RISKS OF CLOUD STORAGE IN BIG DATA ERA

Kalikota Bhavani[1], Dr. D. Srinivas Reddy[2]
[1]mtech, Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA, H.no: 20S41D5823, k.bhavani08@gmail.com
[2]Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA, srinivasreddydhava@gmail.com

ABSTRACT: A cloud-based big data sharing system utilizes a storage facility from a cloud service provider to share data with legitimate users. In contrast to traditional solutions, cloud provider stores the shared data in the large data centers outside the trust domain of the data owner, which may trigger the problem of data confidentiality. This paper proposes a secret sharing group key management protocol (SSGK) to protect the communication process and shared data from unauthorized access. Different from the prior works, a group key is used to encrypt the shared data and a secret sharing scheme is used to distribute the group key in SSGK. The extensive security and performance analyses indicate that our protocol highly minimizes the security and privacy risks of sharing data in cloud storage and saves about 12% of storage space.

Keywords *– Big data, security and privacy, cloud storage, data sharing*

1. INTRODUCTION

The emerging technologies about big data such as Cloud Computing [1], Business Intelligence [2], Data Mining [3], Industrial Information Integration Engineering (IIIE) [4] and Internet-of-Things [5] have opened a new era for future Enterprise Systems(ES) [6]. Cloud computing is a new computing model, in which all resource on Internet form a cloud resource pool and can be allocated to different applications and services dynamically. Compared with traditional distribute system, a considerable amount of investment saved and it brings exceptional elasticity, scalability and efficiency for task execution. By utilizing Cloud Computing services, the numerous enterprise investments in building and maintaining a supercomputing or grid computing environment for smart applications can be effectively reduced. Despite these advantages, security requirements dramatically rise when storing personal identifiable on cloud environment [7], [8]. This raise regulatory compliance issues since migrate the sensitive data from federate domain to distribute domain. To take the benefit enabled by big data technologies, security and privacy issues [9], [10] must be addressed firstly. Building security mechanism for cloud storage is not an easy task. Because shared data on the cloud is outside the control domain of legitimate participants, making the shared data usable upon the demand of the legitimate users should be solved. Additionally, increasing number of parties, devices and applications involved in the cloud leads to the explosive growth of numbers of access points, which makes it more difficult to take proper access control. Lastly, shared data on the cloud are vulnerable to lost or incorrectly modified by the cloud provider or network attackers. Protecting shared data from unauthorized deletion, modification and fabrication is a difficult task. Conventionally, there are two separate methods to promote the security of sharing system. One is access control, in which only authorized user recorded in the access control table has the access privilege of the shared data.

Fig.1: Example figure

Copyright @ 2022 Author

The other method is group key management, in which a group key is used to protect the shared data. Although access control makes the data only be accessed by legitimate participants, it cannot protect the attack from cloud providers. In the existing group key sharing systems, the group key is generally managed by an independent third party. Such methods assume that the third party is always honest. However, the assumption is not always real especially in the environment of cloud storage.To address the security problem of sharing data on the cloud storage, a secret sharing group key management protocol is proposed in the paper and the following means are taken by our protocol to help detect or prevent frauds. Firstly, in order to make the shared data usable upon demand by the legitimate users, symmetric encryption algorithms [17] are used to encrypt the shared data. Once one data owner wants to share data with others, the decryption key is distributed to the legitimate sharers by the data owner. Secondly, the key used to decrypt the shared data controls the access permission for shared data. Asymmetric encryption algorithms [18] are used to encrypt the interactive message and makes only legitimate participants have the ability to decrypt the key. Thirdly, in case of shared data being known by unauthorized users, this protocol uses secret sharing scheme to assign key to the legitimate participants. By adding security mechanism to conventional service-oriented clouds, we obtain a security aware cloud and guarantee the privacy of data sharing on cloud storage. Building security mechanism on cloud storage may accelerate the deployment of a cloud in mission critical business scenario.

2. LITERATURE REVIEW

On minimizing energy cost in Internet-scale systems with dynamic data

With the tremendous growth of cloud computing and Internet-scale online services, massive geographically distributed infrastructures have been deployed to meet the increasing demand, resulting in significant monetary expenditure and environmental pollution caused by energy consumption. In this paper, we investigate how to minimize the long-term energy cost of dynamic Internet-scale systems by fully exploiting the energy efficiency in geographic diversity and variation over time. To this end, we formulate a stochastic optimization problem by considering the fundamental uncertainties of Internet-scale systems, such as the dynamic data. We develop a dynamic request mapping algorithm to solve the formulated problem, which balances the tradeoff between energy cost and delay performance. Our designed algorithm makes real-time decisions based on current queue backlogs and system states, and does not require any knowledge of stochastic job arrivals and service rates caused by dynamic data queries. We formally prove the optimality of our approach. Extensive trace-driven simulations verify our theoretical analysis and demonstrate that our algorithm outperforms the baseline strategies with respect to system cost, queue backlogs, and delay.

A fuzzy preference tree-based recommender system for personalized business-to-business E-services

The Web creates excellent opportunities for businesses to provide personalized online services to their customers. Recommender systems aim to automatically generate personalized suggestions of products/services to customers (businesses or individuals). Although recommender systems have been well studied, there are still two challenges in the development of a recommender system, particularly in real-world B2B e-services: (1) items or user profiles often present complicated tree structures in business applications, which cannot be handled by normal item similarity measures and (2) online users' preferences are often vague and fuzzy, and cannot be dealt with by existing recommendation methods. To handle both these challenges, this study first proposes a method for modeling fuzzy tree-structured user preferences, in which fuzzy set techniques are used to express user preferences. A recommendation approach to recommending tree-structured items is then developed. The key technique in this study is a comprehensive tree matching method, which can match two tree-structured data and identify their corresponding parts by considering all the information on tree structures, node attributes, and weights. Importantly, the proposed fuzzy preference tree-based recommendation approach is tested and validated using an Australian business dataset and the MovieLens dataset. Experimental results show that the proposed fuzzy tree-structured user preference profile reflects user preferences effectively and the recommendation approach demonstrates excellent performance for tree-structured items, especially in e-business applications. This study also applies the proposed recommendation approach to the development of a Web-based business partner recommender system.

Copyright @ 2022 Author

Data mining with big data
Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

SDN and virtualization solutions for the Internet of Things: A survey
AUTHORS: N. Bizanis and F. A. Kuipers
The imminent arrival of the Internet of Things (IoT), which consists of a vast number of devices with heterogeneous characteristics, means that future networks need a new architecture to accommodate the expected increase in data generation. Software defined networking (SDN) and network virtualization (NV) are two technologies that promise to cost-effectively provide the scale and versatility necessary for IoT services. In this paper, we survey the state of the art on the application of SDN and NV to IoT. To the best of our knowledge, we are the first to provide a comprehensive description of every possible IoT implementation aspect for the two technologies.

We start by outlining the ways of combining SDN and NV. Subsequently, we present how the two technologies can be used in the mobile and cellular context, with emphasis on forthcoming 5G networks. Afterward, we move to the study of wireless sensor networks, arguably the current foremost example of an IoT network. Finally, we review some general SDN-NV-enabled IoT architectures, along with real-life deployments and use-cases. We conclude by giving directions for future research on this topic.

Risk and safety program performance evaluation and business process modeling
There is increasing need for agencies to coordinate their interdependent risk assessment, risk management, and risk communication activities in compliance with risk program guidelines. In particular, there is a challenge to measure risk program compliance and maturity to guidelines such as the U.S. Office of Management and Budget (OMB) memorandum “Updated Principles for Risk Analysis” among others. This paper demonstrates a systemic approach to evaluate large-scale risk program maturity with utilization of business process modeling and self-assessment methods. This approach will be helpful to agencies implementing risk guidelines such as those of the OMB, the U.S. Government Accountability Office, the U.S. Department of Homeland Security, the U.S. Department of Defense, and others. This paper will be of interest to risk managers, agencies, and risk and safety analysts engaged in the conception, implementation, and evaluation of risk and safety programs.

3. METHODOLOGY

- Rao proposed a secure sharing schemes of personal health records in cloud computing based on ciphertextpolicy attributed-based(CP-ABE) signcryption. It focus on restricting unauthorized users on access to the confidential data.
- Liu *et al.* proposed an access control policy based on CP-ABE for personal records in cloud computing as well.
- Huang *et al.* introduced a novel public key encryption with authorized equality warrants on all of its ciphertext or a specified ciphertext

Disadvantages:

- While these existing schemes use identity privacy by using attribute-based techniques which fail to protect user attribute privacy.
- Storage and communication overhead.

Copyright @ 2022 Author

To address the security problem of sharing data on the cloud storage, a secret sharing group key management protocol is proposed in the paper and the following means are taken by our protocol to help detect or prevent frauds.

- Firstly, in order to make the shared data usable upon demand by the legitimate users, symmetric encryption algorithms are used to encrypt the shared data. Once one data owner wants to share data with others, the decryption key is distributed to the legitimate sharers by the data owner.
- Secondly, the key used to decrypt the shared data controls the access permission for shared data. Asymmetric encryption algorithms are used to encrypt the interactive message and makes only legitimate participants have the ability to decrypt the key.
- Thirdly, in case of shared data being known by unauthorized users, this protocol uses secret sharing scheme to assign key to the legitimate participants.

Advantages:

- By adding security mechanism to conventional service oriented clouds, we obtain a security aware cloud and guarantee the privacy of data sharing on cloud storage.
- Building security mechanism on cloud storage may accelerate the deployment of a cloud in mission critical business scenario.

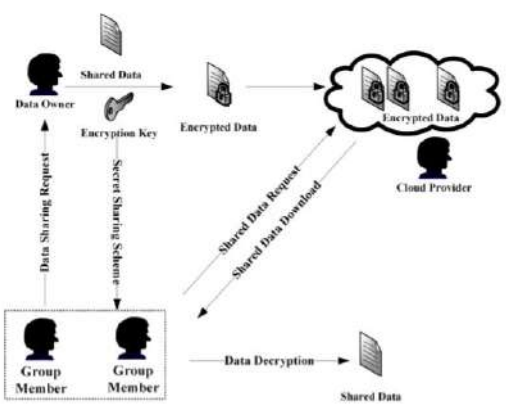

Fig.2: System architecture

MODULES:

- Data Owner
- Group Member
- Cloud Provider

MODULES DESCRIPTION:

Data Owner

Defines the access policy and encrypts its data with a symmetric encryption algorithm using a group key. The group members who satisfied the access policy constitute a sharing group. Then secret sharing scheme is used by the owner to distribute the encryption key to the sharing group

Group Member

In every group member including the data owner is assigned with a unique and a pair of keys. The group members can freely get any interested encrypted data from the public cloud. However the user can decrypt the data if and only if it gets the data decryption key from the data owner.

Cloud Provider:

Provides a public platform for data owners to store and share their encrypted data. The cloud provider doesn't conduct data access control for owners. The encrypted data can be downloaded freely by any users.

Copyright @ 2022 Author

4. IMPLEMENTATION

DATA SHARING MODEL:

Consider a cloud storage data sharing system with multiple entities and the data sharing model is shown as Figure.2. The protocol model consists of three types of entities: cloud provider, data owner and group members. The cloud provider: provides a public platform for data owners to store and share their encrypted data. The cloud provider doesn't conduct data access control for owners. The encrypted data can be download freely by any users. Data owner: defines the access policy and encrypts its data with a symmetric encryption algorithm using a group key. The group members who satisfied the access policy constitute a sharing group. Then secret sharing scheme is used by the owner to distribute the encryption key to the sharing group. Group members: every group member including the data owner is assigned with an unique and a pair of keys. The group members can freely get any interested encrypted data from the public cloud. However the user can decrypt the data if and only if it get the data decryption key from the data owner.

SECURITY MODEL:

In SSGK, we have the following assumptions: The data owner is totally trusted and will never be corrupted by any adversaries. Cloud provider is semi-trusted, it correctly executes the task assigned to them for profits, but they would try to find out as much secret information as possible based on the data owners uploaded data. We now describe the security model of SSGK by listing possible attacks. The group key is distributed by running the secret sharing scheme. Parts of the group members can gather their subsecret shares to reconstruct the group key. Moreover, the communication channel of our protocol is defined as: Every pair of participants have a point-to-point channel to send messages. Additionally, all the participants access to a broadcast channel: when a participant puts a message m on this channel, all the other participants receive m. The group key is distributed on the public channel and the key may be tempered by adversaries.

5. EXPERIMENTAL RESULTS

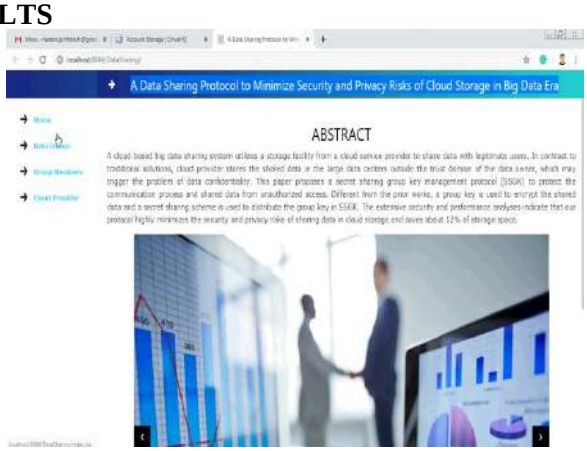

Fig.3: Home screen

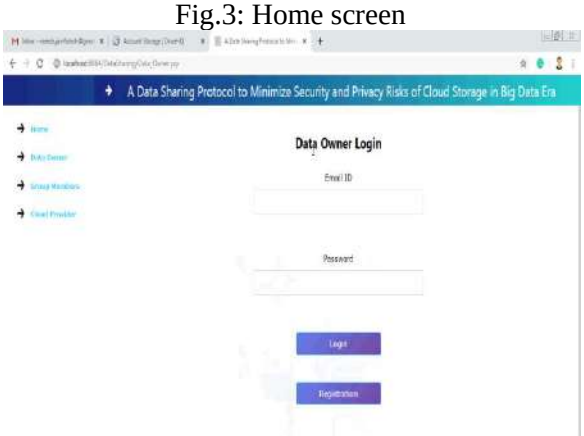

Fig.4: Data owner login

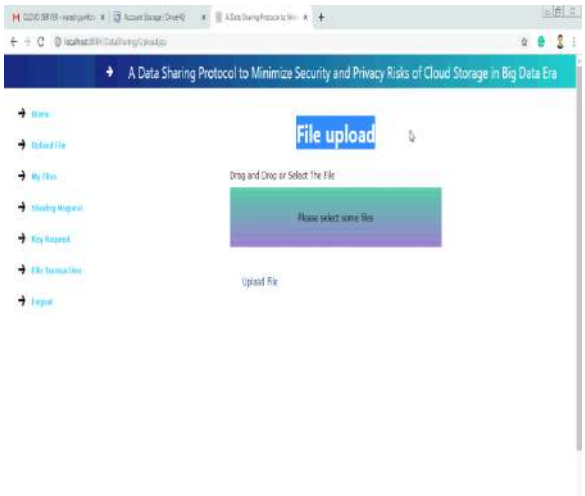

Fig.5: Upload file

Fig.6: My files

Fig.7: Sharing request

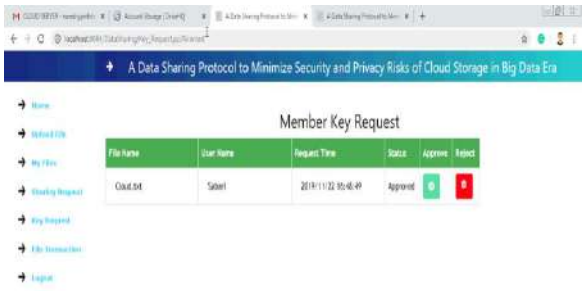

Fig.8: Key request

Fig.9: Download history

6. CONCLUSION

In this paper, we propose a novel group key management protocol for the data sharing in the cloud storage. In SSGK, we uses RSA and verified secret sharing to make the data owner achieve _ne-grained control over the outsourced data without relying on any third party. In addition, we give detailed analysis of possible attacks and corresponding defenses, which demonstrates that GKMP is secure under weaker assumptions. Moreover we demonstrate that our protocol exhibits less storage and computing complexity. Security mechanism in our scheme guarantees the privacy of grids data in cloud storage. Encryption secures the transmission on the public channel; verified security scheme make the grids data only accessed by authorized parties. The better performance in terms of storage and computation make our scheme more practical. The problem of forward and backward security in group key management may require some additions to our protocol. An efficient dynamic mechanism of group members remains as future work.

REFERENCES

[1] P. Zhao, W. Yu, S. Yang, X. Yang, and J. Lin, ``On minimizing energycost in Internet-scale systems with dynamic data," *IEEE Access*, vol. 5,pp. 20068_20082, 2017.

[2] D.Wu, G. Zhang, and J. Lu, ``A fuzzy preference tree-based recommendersystem for personalized business-to-business E-services," *IEEE Trans.Fuzzy Syst.*, vol. 23, no. 1, pp. 29_43, Feb. 2015.

[3] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, ``Data mining with big data,"*IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97_107, Jan. 2014.

[4] X. Shi, L. X. Li, L. Yang, Z. Li, and J. Y. Choi, ``Information flow inreverse logistics: An industrial information integration study," *Inf. Technol.Manage.*, vol. 13, no. 4, pp. 217_232, Dec. 2012.

[5] N. Bizanis and F. A. Kuipers, ``SDN and virtualization solutions forthe Internet of Things: A survey," *IEEE Access*, vol. 4, pp. 5591_5606,May 2016.

[6] S. Li, L. Xu, X. Wang, and J. Wang, ``Integration of hybrid wireless networksin cloud services oriented enterprise information systems," *Enterprise Inf. Syst.*, vol. 6, no. 2, pp. 165_187, Nov. 2012.

Copyright @ 2022 Author

[7] K.-Y. Teng, S. A. Thekdi, and J. H. Lambert, ``Risk and safety programperformance evaluation and business process modeling," *IEEE Trans.Syst., Man, Cybern. A, Syst. Humans*, vol. 42, no. 6, pp. 1504_1513,Nov. 2012.

[8] Z. Fu, X. Sun, S. Ji, and G. Xie, ``Towards efficient content-aware searchover encrypted outsourced data in cloud," in *Proc. 35th Annu. IEEE Int.Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1_9.

[9] J. Han, W. Susio, Y. Mu, and J. Hou, ``Improving privacy and security indecentralized ciphertext-policy attribute-based encryption," *IEEE Trans.Inf. Forensics Security*, vol. 10, no. 3, pp. 665_678, Mar. 2015.

[10] D. Zou, Y. Xiang, and G. Min, ``Privacy preserving in cloud computingenvironment," *Secur. Commun. Netw.*, vol. 9, no. 15, pp. 2752_2753Oct. 2016.

Copyright @ 2022 Author