IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Exploration Encryption Scheme Based on Hash Chain Using MD5 Algorithm

Mrs. R Surekha[1] M.Tech K Leelavathi[2] G Pushparaj[3] N Madhulatha [4] K Kishore [5]

[1]Assistant Professor, [2345]UG Student, Department of Computer Science and Engineering

[12345]Siddharth Institute Of Engineering & Technology, Tirupathi, Andhra Pradesh, India,

ABSTRACT

During this paper, for the first time, we tend to outline and solve the difficult downside of privacy preserving Hash Chain based hierarchic search over encrypted cloud data (HRSE).We establish a collection of strict privacy needs for such a secure cloud knowledge utilization system. Among numerous multi- keyword semantics, we decide the economical similarity live of "coordinate matching", i.e., as several matches as potential, to capture the connection of knowledge documents to the search question. We further use "inner product similarity" to quantitatively value such similarity live. we tend to initial propose a basic plan for the HRSE supported secure scalar product computation, then provide two considerably improved HRSE schemes to realize numerous stringent privacy needs in 2 completely different threat models. Thorough analysis work privacy and potency guarantees of planned schemes are given. Experiments on the real-world dataset additional show planned schemes so introduce low overhead on computation and communication.

Keywords: *Trapdoor, Query Processing, MRSE, Keyword, Ranking, Cloud.*

INTRODUCTION

Big-Data has turn into a trendy word which is used to illustrate the exponential increase and accessibility of data. The increasing demand for large-scale data dispensation and data analysis applications spurred the improvement of novel solutions to tackle this challenge. For about a decade, the Map Reduce framework has represented the defector regular of big data skills and has been widely utilized as a popular mechanism to harness the power of large clusters of computers. In general, the fundamental principle of the Map Reduce framework is to move analysis to the data, rather than moving the data to a system that can

analyze it. It allows programmers to think in a data-centric fashion where they can focus on applying transformations to sets of data records while the details of distributed execution and fault tolerance are transparently managed by the framework.

Cloud garage is gaining popularity these days. In organization settings, we see the upward thrust in call for records outsourcing, which assists inside the strategic control of company information. It's also used as a middle era at the back of many online offerings for private applications. Nowadays, it is simple to use without spending a dime debts for electronic mail photo album, report sharing and/or far flung get entry to, with storage length extra than 25GB (or some bucks for more than 1TB). Collectively with the cutting-edge Wi-Fi era, users can get right of entry to almost all in their documents and emails by means of a cellular telephone in any corner of the sector. Considering information privateness, a traditional manner to make certain its miles to depend on the server to put into effect the get entry to manage after authentication; this means that any unexpected privilege escalation will divulge all information. In a shared-tenancy cloud computing surroundings, matters emerge as even worse. Records from special customers can be hosted on separate digital machines (VMs) but live on a single bodily machine. Information in a goal VM might be stolen by means of instantiating another VM co-resident with the target one.

CLOUD Computing has been estimated as the subsequent-technology facts technology (IT) architecture for firms, because of its long list of extraordinary blessings inside the IT history: on-demand self-carrier, ubiquitous community get right of entry to, place unbiased resource pooling, fast useful resource elasticity, usage-based totally pricing and transference of threat. As a disruptive era with profound implications, Cloud Computing is reworking the very nature of ways companies use records technology. One fundamental issue of this paradigm shifting is that statistics is being centralized or outsourced to the Cloud. From users' angle, consisting of both individuals and IT organizations, storing statistics remotely to the cloud in a flexible on-call for way brings attractive advantages: comfort of the load for storage control, normal records get right of entry to with independent geographical locations, and avoidance of capital expenditure on hardware, software program, and personnel maintenances, and many others. at the same time as Cloud Computing makes those blessings greater appealing than ever, it also brings new and hard protection threats closer to customers' outsourced data.

PROBLEM STATEMENT

With cloud storage services, you can: Cost-effectively protect data in the cloud without sacrificing performance. Scale up your backup resources in minutes as data requirements change. Protect backups with a data center and network architecture built for security-sensitive organizations.

LITERATURE SURVEY

What is Public-key searchable encryption?

A cryptographic system that uses two keys -- a public key known to everyone and a private or secret key known only to the recipient of the message. When John wants to send a secure message to Jane, he uses Jane's public key to encrypt the message. Jane then uses her private key to decrypt it.

An important element to the public important system is that the public and private answers are related in such a way that lone the civic key can be rummage-sale to code messages and only the agreeing private key can be used to decrypt them. Moreover, it is nearly impossible to deduce the private crucial if you know the civic key.

What is semantic security?

The simplest definition of semantic security is one-time semantic security which naturally applies to stream ciphers: Anadvers.ary sends two plaintext messages of equal length to the challenger and receives one encrypted message; semantic security means an adversary can't distinguish which plaintext message was encrypted.

What is identity-based key encapsulation mechanism?

We construct an Identity-Based Key Encapsulation Mechanism (IBKEM) in a generic "steamrolled" multi-linear map setting and prove its security under multi linear decisional Diffie-Hellmanin assumption in the selective-ID model. Then, we make our IB-KEM translated to the GGH framework, which defined an "inexact" version of a multi-linear group family from ideal lattices, and modify our proof of security to use the GGH graded algebras analogue of multi-linear maps.

What is identity based encryption?

ID-based encryption, or identity-based encryption (IBE), standsanmajorprimeval of ID-based cryptography. As such it is a type of public-key encryption in which the public basic of a user is definiteunique information about the identity of the user (e.g. a user's email address). This can use the text-value of the name or domain name as a key or the physical IP address it translates to.

RECENT WORKS

However, in recent years, with the increasing applications' requirements in the data analytics domain, various limitations of the Hadoop framework have been recognized and thus we have witnessed an unprecedented interest to tackle these challenges with new solutions which constituted a new wave of mostly domain-specific, optimized big data processing platforms .Several frameworks , have been presented to tackle the ever larger datasets on using distributed clusters of commodity machines. These frameworks significant reduce the complexity of developing big data programs and applications. However, in reality, many real-world scenarios require pipelining and integration of multiple big data jobs. There are more challenges when applying big data technology in practice. For example, consider typical online machine learning pipeline.

In the pipeline, components like feature extractor and classification trainer are normally commonly-used algorithms for many machine learning applications.

In current big data platform such as MapReduce and Spark, there is not at all proper way to share and expose a deployed and well-tuned online component to other developers. Therefore, there is massive and even unseen redundant development in big data applications.

LIMITATIONS

Real-world applications require a chain of operations or even a pipeline of data processing programs. Optimizing a complicated job is difficult and optimizing pipelined ones are even harder. Manual optimizations are time-consuming and error prone and it is almost impossible to manually optimize every program. Integration, composition and interaction with big data programs/jobs are not natively supported many practical data analytics and machine learning algorithms require combination of multiple processing components each of which is responsible for a certain analytical functionality. As a result of this, deployed jobs are not natively compassable and reusable for subsequent development and integration. Maintenance and management of evolving big data applications are complex and tedious. In a realistic data analytic process, data scientists need to explore the datasets and tune the algorithms back and force to find out a more optimal solution.

PROPOSED METHOD

We present the Hierarchically Distributed Data Matrix (HDM+: A Hierarchical Hash Chain Framework for Data Processing) along with the system implementation to support the writing and execution of compassable and integral big data applications.

HDM is a light-weight, functional and strongly-typed meta-data abstraction which contains complete information (such as data format, locations, dependencies and functions between input and output) to support parallel execution of data driven applications.

Improve the execution performance of HDM data flow.

Moreover, by drawing on the comprehensive information maintained by HDM graphs, the runtime execution engine of HDM is also able to provide provenance and history management for submitted applications.

We focus on solving the classification problem over encrypted data. In particular, we propose a secure k-NN classifier over encrypted data in the cloud. The proposed protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. To the best of our knowledge, our work is the first to develop a secure k-NN classifier over encrypted data under the semi-honest model. Also, we empirically analyze the efficiency of our proposed protocol using a real-world dataset under different parameter settings.

We proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. Specifically, we focus on the classification problem since it is one of the most common data mining tasks. Because each classification technique has their own advantage, to be concrete, this paper concentrates on executing the k nearest neighbor classification method over encrypted data in the cloud computing environment.

CONTRIBUTIONS

- Solving the classification problem over encrypted data
- Protects the confidentiality of data Secure k-NN classifier over encrypted data under the semi-honest model.
- HDM+ a lightweight, functional, strongly-typed data abstraction for developing and describing data-parallel applications.
- Based on the functional data dependency graph, optimizations include function fusion, local aggregation, operation reordering and caching are introduced to improve the performance of HDM jobs.
- A HDM system implementation provisioning that supports the execution, integration, history and dependency management of HDMs applications on distributed environments.
- Comprehensive benchmarks including: basic primitives, pipelined operations, SQL queries and iterative jobs (ML algorithms) are tested to evaluate the performance of HDM compared with the current state-of-art big data processing framework - Apache Spark.

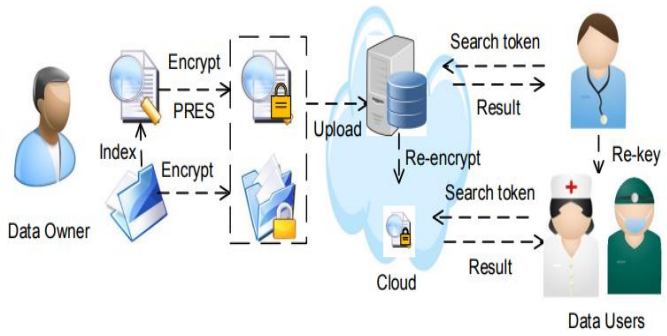


Fig 1: **Architecture Diagram**

ALGORITHM:

- *INPUT*
  - *User -> U*
  - *System -> S*
  - *Hash Tree -> HT*
  - *Data Object -> DO*
  - *Query -> Q*
  - *Classification ->C*
  - *Relationships ->Rrealtion*
- *OUTPUT*
  - *Result -> R*

HDM+ : Hierarchical Hash Chain for Data Management

- *BEGIN*
- *Step1:Usignin*
- *Step2: S encrypt element and keyword*
- *Step3: find keyword element relationships usingRrealtion*
- *Step4: classification using C*
- *Step2: U given Q -> S*
- *Step2.1: S read Q*
- *Step2.2: find weights*
- *Step2.3: retrieve relationships*
- *Step2.4: find score and sorting*
- *Step2.5: find k-nearest neighbor*
- *Step3:UgetR from S*
- *END*

MODULES DESCRIPTION

Cloud server

- A cloud server is a logical server that is built, hosted and delivered through a cloud computing platform over the Internet.
- Cloud servers possess and exhibit similar capabilities and functionality to a typical server but are accessed remotely from a cloud service provider.

- A cloud server may also be called a virtual server or virtual private sever.

Searchable encryption

- Secure Retrieval of k-Nearest Neighbors(SRKNN):
- In this stage, user initially sends his query (in encrypted form) called trapdoor to C1.
- After this, C1 securely retrieve (in encrypted form) the class labels corresponding to the k-nearest neighbors of the input query q.
- After this, C1 securely retrieve (in encrypted form) the class labels corresponding to the k-nearest neighbors of the input query q.

- Secure Computation of Majority Class (SCMCk): Following from Stage 1, C1 compute the class label with a majority voting among the k-nearest neighbors of q.

- At the end of this step, only User knows the class label corresponding to his input query record q.

Multilevel interrelationship

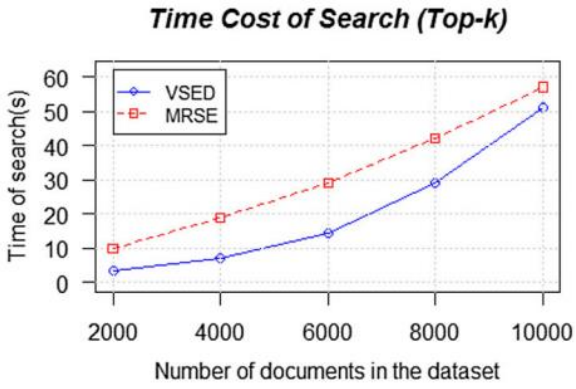
- The generating encrypted keyword-mapping, multilevel inter-relationship.
- The inter-relationships between elements at different levels keyword is mentioned in some entity descriptions at the element level.
- Entities at the element level are associated with a set-level element via type.
- A set-level element is contained in a source.
- There is an edge between two keywords if two elements at the element level *mentioning these keywords are connected via a path.
- We propose a ranking scheme that deals with relevance at many levels.

AES cryptosystem

- AES, the U.S. government announced that AES could be used to protect classified information.
- It soon became the default encryption algorithm for protecting classified information as well as the first publicly accessible and open cipher approved by the NSA for top-secret information.
- The NSA chose AES as one of the cryptographic algorithms to be used by its Information Assurance Directorate to protect national security systems.
- Its successful use by the U.S. government led to widespread use in the private sector, leading AES to become the most popular algorithm used in symmetric key cryptography
- The transparent selection process helped create a high level of confidence in AES among security and cryptography experts.

- AES is more secure than its predecessors DES and 3DES as the algorithm is stronger and uses longer key lengths.
- It also enables faster encryption than DES and 3DES, making it ideal for software application

RESULT ANALYSIS



First we conduct an ablation study with threshold-based similarity search to evaluate our proposed techniques in the filter and verification phases, respectively. To begin with, we look at techniques in the filter phase proposed in Section 4 to reduce filter cost. We propose 3 methods: Dimension is the dimension-aware method that builds triplets by dimensions to fast compute the lower bounds (Algorithm 2); Index uses the proposed index to share the computation among different objects (Algorithm 3); Batch uses the progressive filtering technique to prune dissimilar objects in batch base on proposed index (Algorithm 4). Because we only test the performance of proposed techniques, we directly apply the above methods to our datasets without other lower bounds nor verification phase. Figure 11 shows the performance of such methods. We see that Index significantly outperforms Dimension as it can utilize the relationships between objects. And Batch has better performance than Index as it can further avoid unnecessary.

CONCLUSION

To address the problem of semantic retrieval, we propose effective schemes based on concept hierarchy. Our solutions use two cloud servers for encrypted retrieval and make contributions both on search accuracy and efficiency. To improve accuracy, we extend the concept hierarchy to expand the search conditions. In addition, a tree-based index structure is constructed to organize all the document index vectors, which are built based on the concept hierarchy for the aspect of search efficiency. The security analysis shows that the proposed scheme is secure in the threat models. Experiments on real world dataset illustrate that our scheme is efficient.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.

[2] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE TPDS, vol. 23, no. 8, pp. 1467–1479, 2012.

[3] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.

[4] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," In Proc. of ACM CCS, 2006, pp. 79–88.

[5] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L.Varna, S. He, M. Wu, and D. W. Oard, "Confidentiality-preserving rank-ordered search," in Proc. of the 2007 ACM Workshop on Storage Security and Survivability, 2007, pp. 7–12.

[6] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Topk retrieval from a confidential index," in Proc. of EDBT, 2009, pp. 439–449.

[7] N. Cao, C. Wang, and M. Li, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," Parallel and Distributed Systems,IEEE Transactions on, vol.25,no.1, pp.222-233,2014.

[8] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, YT. Hou, and H.L, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. of ACM SIGSAC symposium on Information, computer and communications security, 2013, pp. 71–82.

[9] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data," in Proc. of the 31st ICDCSW, 2011, pp. 273–281.

[10] Ayad Ibrahim, Hai Jin, Ali A. Yassin, and Deqing Zou, "Secure Rank-ordered Search of Multi-keyword Trapdoor over Encrypted Cloud