International Journal of Trend in Research and Development, Volume 3(2), ISSN: 2394-9333
www.ijtrd.com

A Survey on Secure Cloud Data in Decomposition of Encrypted Subtensor Using Homomorphic Encryption Scheme

[1]Sinduja T and [2]Dr. K. Thippeswamy,
[1]Student, [2]Professor and Head,
[1,2]Department of Computer Science & Engineering, VTU Regional Centre, Mysuru, India

Abstract: As the rapidly growing volume of data are beyond the capabilities of many computing infrastructures, to securely process them on cloud has become a preferred solution which can both utilize the powerful capabilities provided by cloud and protect data privacy. This paper presents an approach to securely decompose a tensor, a mathematical model widely used in data-intensive applications, to a core tensor multiplied with a certain number of truncated orthogonal bases. The unstructured, semi-structured, and structured data are represented as low-order sub-tensors which are then encrypted using the fully homomorphic encryption scheme. A unified high-order cipher tensor model is constructed by collecting all the cipher sub-tensors and embedding them to a base tensor space. The cipher tensor is decomposed through a proposed secure algorithm, in which the square root operations are eliminated during the Lanczos procedure. Theoretical analyses of the algorithm in terms of time complexity, memory usage, decomposition accuracy, and data security are provided. Experimental results demonstrate that the approach can securely decompose a tensor.With the advancement of fully homomorphic encryption scheme, it can be expected that the secure tensor decomposition approach has the potential to be applied on cloud for privacy-preserving data processing.

Keywords: *Tensor Decomposition, Fully Homomorphic Encryption, Lanczos Method, Cloud.*

I. INTRODUCTION

The size of data in many fields is rapidly increasing towards Terabyte level or even Petabyte level, as well as the data structures are becoming more varied. The large scale heterogeneous data have posed great challenges on current computing infrastructures, and new approaches are in urgent need to address them. Cloud Computing is a model that can enable ubiquitousand convenient network access to a shared pool of configurable computing resources such as platforms, software and services. A cloud infrastructure is the collection of hardware and software which can provide capabilities to the consumers on a pay-per-use or chargeper- use basis. It is a quite feasible approach to upload the large scale data to cloud for deeply processing and mining such as dimensionality reduction , classification , and prediction . However, carrying out such types of tasks on cloud may cause a series of security problems including loss of privacy, disclosure of business information, data tamper, etc. Therefore, thestudy of secure data mining and data analyzing on cloud is of great necessity as it is an efficient method to extract valuable information from the large scale heterogeneous data. The fully homomorphic encryption scheme, which is suggested in 1978 by Rivest, Adleman, Dertouzos , allows specific types of computations to be performed on the cyphertext to generate an encrypted result, of which the decryption is identical to the result obtainedby directly carrying out operations on the plaintext. The ideal lattice based scheme proposed by Gentry in 2009 solves the problem of limited number of operations of fully homomorphic encryption, which paves the way for trusted computing on cloud. The Learning with Errors (LWE) scheme reported in is more practical to be employed in data-intensive applications. Although the mentioned schemes provide both additive and multiplicative homomorphisms, they can cause decryption errors when be used by algorithms including non-homomorphic operations such as square root and division, which are frequently used operations during data processing. Many heterogenous data are modeled as tensors [8, 9], a type of high dimension matrix widely used in many applications. Tensor decomposition is a powerful tool to extract valuable information from large scale raw data. The decomposition is computationally expensive and is strongly suggested to be performed on cloud. Therefore, it is necessary to investigate approaches for secure tensor decomposition on cloud and address the challenges caused by non-homomorphic operations. However, little research has been devoted to such type of method.This paper presents a new computing approach which can securely decompose the tensor model generated from large scale heterogeneous data.

The major contributions are summarized as follows.

- We present a holistic framework to address the problem of secure tensor decomposition on cloud. The framework not only allows us to utilize the powerful computational capabilities of the cloud, but also ensures data security during the process of tensor decomposition.
- We introduce a Unified Cipher Tensor (UCT) model for heterogeneous data representation. The detailed procedures of how to encrypt the low-order sub-tensors constructed from heterogeneous datas cipher counterparts using the fully encryption scheme, as well as how to embed them to a base tensor space to generate a unified cipher tensor model are illustrated in this paper.
- We propose to employ the Lanczos method to decompose the generated cipher tensor model to a core tensor and a certain number of truncated orthogonal bases. A secure tensor decomposition algorithm is designed in which the nonhomomorphic square root operations are removed during the Lanczos procedure. Theoretical analyses of the algorithm in terms of time complexity, memory usage, decomposition accuracy, and data security are provided.

II. PRELIMINARIES

In this section, the preliminaries on tensor decomposition, fully homomorphic encryption, and Lanczosmethod are reviewed.

A. Tensor Decomposition

Tensor is a type of high dimension matrix widely used in many applications such as computer vision, data mining, graph

analysis and signal processing. High- Order Singular Value Decomposition (HO-SVD) isa type of approach that can factorize the tensor to a core tensor multiplied with a number of truncated orthogonal matrices. Let T $\in R^{I1\times I2\times:::\times IN}$ denote an *N*- th order tensor model, *S* and ^ *T* refer to the core tensor and approximate tensor respectively, then the HO-SVD method is defined as

$$S=T\times_1 U_1^T\times_2 U_2^T\ldots\times_N U_N^T,$$
$$\hat{T}=S\times_1 U_1\times_2 U_2\ldots\times_N U_N \quad (1)$$

The *i*-mode product $T\times iU;\ 1\le i\le N$, of a tensor by a matrix in Eq. (1) is defined as

$$(T\times_i U)_{j_1 j_2\ldots j_{i-1} k_i j_{i+1}\ldots j_N} = \sum_{j_i=1}^{I_i} (t_{j_1 j_2\ldots j_{i-1} j_i j_{i+1}\ldots j_N}\times u_{k_i j_i}), \quad (2)$$

where$t_{j1j2\cdots ji\text{-}1jiji+1\ldots ji}$and u_{kiji}refer to the elements of tensor *T* and matrix *U*,respectively.

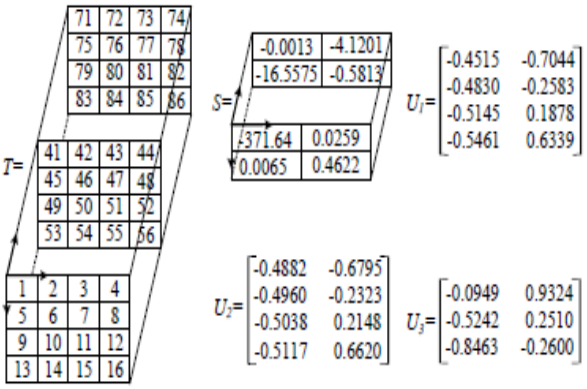


Fig. 1. Decomposing a three-order tensor to a core tensor and three truncated orthogonal bases.

For example, Fig. 1 demonstrates the generated core tensor *S* and the truncated bases *U*1, *U*2, *U*3 by decomposing the initial tensor *T*. The 4 by 4 by 3 tensor is decomposed to a 2 by 2 by 2 core tensor, two matrices of 4 by 2 and a matrix of 3 by 2. Generally, the core tensor and the truncated bases are considered as a compressed version of the initial tensor *T*. The reconstructed data in the approximate Tensor$\hat{T}$are of higher quality than the raw data as the noise, inessential and inconsistent data are removed.

B. Fully Homomorphic Encryption

Homomorphic encryption is a new type of scheme that allows specific types of operations to be performed on the cyphertext to obtain the encrypted result, of which the decryption is identical to the result directly computed by performing operations on the plaintext. Two fully homomorphic encryption schemes [6, 11] are proposed using ideal lattice and polynomial ring, respectively. A Ring Learning with Errors (RLWE) base fully homomorphic encryption scheme without bootstrapping is proposed , where a General Learning with Errors (GLWE) based scheme is reported.The encryption scheme supports the homomorphism of addition and multiplication, which can be described as follows

Enc(m1) + Enc(m2) = Enc(m1 + m2);
Enc(m1 × m2) = Enc(m1) × Enc(m2): (3)

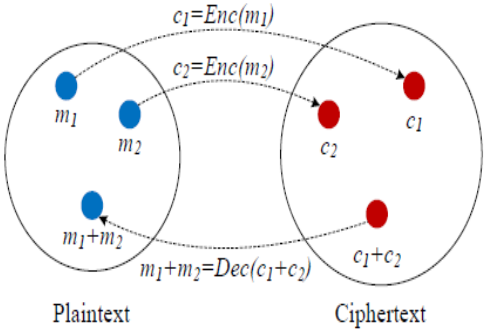


Fig. 2. Illustration of the homomorphic encryption.

Fig 2 demonstrates homomorphic encryption of an addition operation. Let m1, m2 be two elements in the plaintext, c1, c2 in the ciphertext, and c1 = Enc(m1),c2=Enc(m2), then m1+m2 = Dec(Enc(m1)+Enc(m2)).

C. Lanczos Method

The Lanczosmethod is efficient for computingthe eigenvalues and eigenvectors of a sparse symmetric matrix. It transforms the matrix *M* with an orthogonal matrix *W*, where *W* = [*w*1,. ,.*wk*] and $W^TW = I$, to a tridiagonal matrix as follows

$$L=\begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ & & \beta_k & \alpha_k \end{bmatrix}. \quad (4)$$

Equating columns in the expression *MW* = *WL*, the tridiagonal matrix *L* can be generated by carrying out the iteration procedures

$$\alpha j=w_i^T M w j,$$
$$rj=Mwj-\alpha jwj-\beta jwj-1,$$
$$\beta j+1=\|rj\|_2, wj+1=rj/\beta j+1. \quad (5)$$

The components of α,β,rcan be progressively calculated. Let the eigenvalue decomposition of matrix *L* be defined as *L* = *Q*^*Q*T, then the eigenvalues and eigenvectors of matrix *M* are ^ and *WQ*, respectively. In the matrix-vector product is the frequently called
linear transformation during the Lanczos procedure.

III. PROBLEM DEFINITION AND SOLUTIONFRAMEWORK

This section formalizes the problem of secure tensor decomposition on the bases of the fully homomorphic encryption scheme, and provides an overview of the proposed solution framework.

A. Problem Definition

Heterogeneous data consist of unstructured data Du, semi structured data Dsemi, and structured data Ds. Let core denote the core data including the core tensor S and the truncated orthogonal bases U1, U2,….,UN, then the secure tensor decomposition problem can be formalized as

fr:{Enc(Du),Enc(Dsemi),Enc(Ds)} →Enc(T),
fd:Enc(T) →{Enc(S),Enc(U_1), …. , Enc(U_N)}: (6)

In Eq. (6)the data representation function *fr* integrates all encrypted data as a unified cipher tensor model (*UCT*), on which the decomposition function *fd* is performed to generate the encrypted core tensor as well as the encrypted truncated orthogonal bases.

As the decomposition operations are carried on the encrypted data, the user's privacy are protected. In order to guarantee the correctness of the decomposition resultEq. (6) satisfies $S = T\times_1 U_1^T \times_2 U_2^T \ldots \times_N U_N^T$. Accordingto the fully homomorphic encryption scheme, the securedecomposition process satisfies the following equation

$$Dec(sk; Eva(pk; Cfd; Enc(T))) = Cfd(T); (7)$$

where *Eva*, *Enc*, *Dec* refer to the evaluation, encryption,and decryption function, *pk*and *sk*denote the public keyand private key, *Cfd*refers to the boolean circuits of thetensor decomposition function *fd*defined in Eq. (6).

The homomorphism can be guaranteed by performing addition, subtraction, and multiplication operations on the cipher data during the tensor decomposition process. However, new challenges arise when the nonhomomorphic operations such as square root and division are adopted in some types of decomposition methods, for example, Lanczos-based algorithm. A secure tensor decomposition algorithm is proposed in this paper to address these challenges.

For convenience, in the following sections this paper adopts the symbol Ψ^E to denote the cipher data in gtothepla in dataΨ, namely $\Psi^E = Enc(\Psi)$.Therefore, the encrypted tensor $Enc(T)$ is denoted as T^E.

B. Overview of the Solution Framework

To address the problem defined above, this paper proposes a secure tensor decomposition approach based on the fully homomorphic encryption scheme. Fig.3 provides an overview of the framework where the unstructured, semi-structured, and structured data a reencrypted and represented as a unified tensor model, which is then securely decomposed to a core tensor multiplied with a certain number of truncated orthogonal bases. The four representative step so the solution framework are summarized as follows.

1. Data Representation, Encryption and Submission:

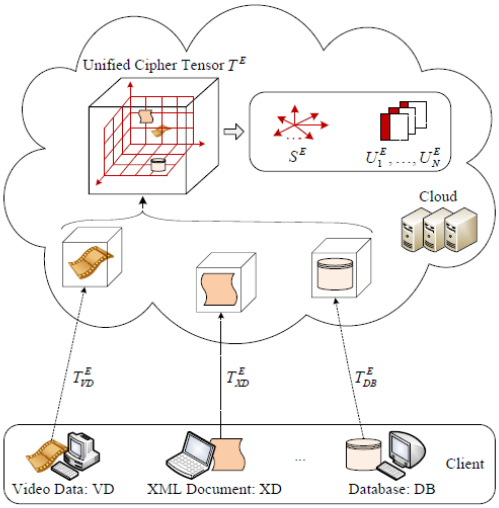


Fig. 3. Framework overview of the secure tensor decomposition approach.

The heterogeneous data collected in the clients are represented a slow-order sub-tensors using the method proposed in previous work then the sub- tensors are encrypted using the fully homomorphic encryption scheme and the generated cipher results are submitted to the cloud forunification and decomposition. In Fig.3, the unstructured video data VD, semi-structured XML document XD, and structured database DB are transformed to cipher low – order sub-tensors $T^E_{VD}, T^E_{XD}, T^E_{DB}$respectively

2. Construction of Cipher Tensor:

The generated sub-tensors $T^E_{VD}, T^E_{XD}, T^E_{DB}$ are then embedded to a base tensor model $T_{base} \in R^{Itim\times Ispa\times Iclt}$ to generate a unified cipher tensor model T^E using the tensor extension operation T^E Tbase→×T^EV D→×T^EXD×T^EDB the three orders I_{tim}; I_{spa}; I_{clt} of the base tensor model denote the time, space and client characteristics.

3. Secure Tensor Decomposition:

After unfolding the unified cipher tensor $T^E$to matrices $T^E(1),\ldots, T^E(N)$,where N is the number of orders of tensor T^E, the symmetrization transformation is performed on each tensor unfolding to generate the symmetric matrix $sym(T^E(i)) = T^E(i)(T^E(i))T$; $1 \le i \le N$. The eigen vectors of the symmetric matrix $sym(T^E(i))$are corresponding to the left singular vectors of matrix $T^E(i)$. The Lanczos method is employed to perform the eigen value decomposition, namely, $sym(T^E(i)) = U^E i \wedge^E U^E i)T$. The cipher core tensor S^E can be computed by applying Eq. (1) to the truncated bases $U^E1,\ldots, U^EN$ and the unified cipher tensor T^E.

4. Obtain the Plain Core Tensor and Bases:

By decrypting the cipher core tensor and cipher truncated bases generated in Step 3, the plain core tensor Sand plain truncated orthogonal bases $U_1,\ldots., U_N$can be computed. As the homomorphism are supported during the secure tensor decomposition, the generated results are correct and are identical to that directly computed using the plain data. This paper focuses on Step 2 and Step 3, which correspond to the secure representation function frand secure tensor decomposition function fd.

IV. CONSTRUCTION ON CIPHER TENSOR VIA FULLY HOMOMORPHICENCRYPTION SCHEME ON CLOUD

This section illustrates the process of representing the heterogeneous data as a unified cipher tensor model via the fully homomorphic encryption scheme. New concepts and operations closely related to the cipher tensor model are introduced.

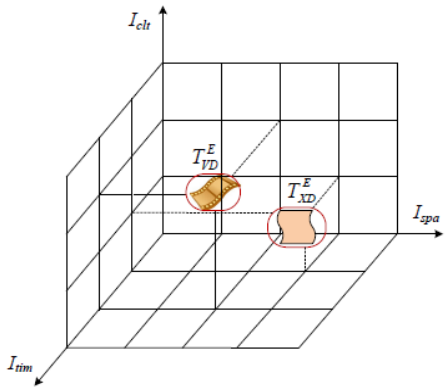
A. Cipher Tensor and Nil Element

In order to clearly describe the process of representing the unstructured, semi structured, and structured data as a unified cipher tensor model.

B. Constructing a Unified Cipher Tensor Model on Cloud

In this paper, the heterogenous data are first represented and encrypted as cipher low-order sub-tensors on the clients, then they are submitted to the cloud for unification. To integrate all the cipher sub-tensors, a base tensor model is proposed, which is defined as Tbase∈RItim×Ispa×Iclt, where Itim; Ispa; Icltrefer to the time, space and client characteristics. The three orders serve as a basis to which various types of encrypted

subtensors can be appended to generate a unified cipher tensor model.



Embedding two encrypted sub-tensors to the base model on cloud.

C. Tensor Unfolding and Memory Storage Scheme

When the unified cipher tensor is generated, the next critical step is to obtain the tensor unfolding, which are then transformed to symmetric matrices. For sparse tensor, the Compressed Row Storage (CRS) method is employed to store the unfolded matrices. The CRS scheme is efficient for matrix-vector product and can reduce memory usage during tensor decomposition. Additionally, in order to decrease execution time of the secure tensor decomposition algorithm, the data-intensive application can employ $T^E(i)((T^E(i))T$ v) toperform the matrix-vector operation on the symmetric matrix of the i-mode tensor unfolding.

D. Cipher Tensor Representation Algorithm on Cloud

Based on the above mentioned methods, this paper proposes Algorithm 1 to represent the heterogeneous data as a unified cipher tensor (UCT) model on cloud.

Algorithm 1 Cipher Tensor Representation.TE =fr (Du; Dsemi; Ds)

Input:
The unstructured data Du, semi-structured dataDsemi, and structured data Ds.

Output:
The unified cipher tensor model TE.

1. Represent the local heterogeneous data as low-order sub-tensors, and encrypt them to cipher low-order sub-tensors on clients.
2. Upload the generated cipher sub-tensors to cloud.
3. Embed all the cipher sub-tensors to the base tensor model $T_{base} \in R_{Itim \times Ispa \times Iclt}$, and obtain the unified cipher tensor model TE.
4. Unfold the cipher tensor to matrices and generate the symmetric matrices for decomposition.

In Line 1 of the proposed Algorithm 1, the unstructured, semi-structured, and structured data are transformed to low-order sub-tensors, which are then encrypted using the fully homomorphic encryption scheme on clients. All the cipher sub-tensors are uploaded to cloud for unified representation. In this paper, the zero elements of the plain data are removed during the encryption procedure. The cloud embeds all the cipher sub-tensors to the base tensor model in Line 3 to obtain the unified cipher tensor model T^E. Line 4 generates the symmetric matrices of each cipher tensor unfolding for secure tensor decomposition.

CONCLUSION

Aiming to propose an efficient approach that can securely process large scale heterogeneous data, this paper formalizes the secure tensor decomposition problem, and proposes a holistic solution framework to address it. A unified cipher tensor model is presented to integrate all the encrypted low-order sub-tensors as a unified model. Concise examples are provided for illustrating the process of cipher tensor construction and unfolding. A Lanczos-based secure tensor decomposition algorithm is introduced, in which the non-homomorphic square root operations in Lanczos procedure are removed. Theoretical analyses in terms of time complexity, memory usage, decomposition accuracy, and data security are provided. Some very preliminary experiments are carried out to evaluate the performance of the presented methods. The results support that the proposed approach is feasible and can pave a way for secure data processing on cloud.

References

[1] P. Mell and T. Grance, "The NIST Definition ofCloud Computing (Draft)," NIST Special Publication,vol. 800, no. 145, p. 7, 2011.
[2] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality Reduction: A Comparative Review," J. Machine Learning Research, vol. 10, no. 1-41, pp. 66–71, 2009.
[3] J. Han and M. Kamber, Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann,2006.
[4] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons, 2011.
[5] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On Data Banks and Privacy Homomorphisms," Foundations of Secure Computation, vol. 4, no. 11, pp. 169–180, 1978.
[6] C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices," in Proc. 41st Ann. ACM Symp. Theoryof Computing, vol. 9, 2009, pp. 169–178.
[7] K. Lauter, M. Naehrig, and V. Vaikuntanathan, "Can Homomorphic Encryption be Practical?" Cryptology ePrint Archive, Report 405, 2011, http://eprint.iacr.org/2011/
[8] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," SIAM Review, vol. 51, no. 3, pp. 455–500, 2009.
[9] L. Kuang, F. Hao, L. T. Yang, M. Lin, C. Luo, and G. Min, "A Tensor-Based Approach for Big Data Representation and Dimensionality Reduction," IEEE Trans. Emerging Topics in Computing, vol. 2, no. 3, pp. 280–291, 2014.
[10] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," SIAM J.Matrix Analysis and Applications, vol. 21, no. 4, pp. 1253–1278, 2000.
[11] M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption Over the Integers," in Advances inCryptology–EUROCRYPT 2010. Springer, 2010, pp.24–43.
[12] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "Fully Homomorphic Encryption Without Bootstrapping," in Proc. 3rd Innovations in TheoreticalComputer Science Conference.ACM, 2012, pp. 309– 325.
[13] J. K. Cullum and R. A. Willoughby, Lanczos Algorithms for Large Symmetric Eigenvalue Computations: Vol. 1: Theory. SIAM, 2002, vol. 41.

[14] M. Gr¨ uning, A. Marini, and X. Gonze, "Implementation and Testing of Lanczos-based Algorithms for Random-Phase Approximation Eigenproblems," Computational Materials Science, vol. 50, no. 7, pp. 2148–2156, 2011.

[15] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, Templates for the Solution of AlgebraicEigenvalue Problems: A Practical Guide. SIAM,