International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Privacy Data Chain Disclosure Discovery Method based on ontology for Big Data

Sonali T. Benke, Mr.Kishor N. Shedage, Mr.Sharad M. Rokade and Mr.Devidas S.Thosar

Department of Computer Engineering, SVIT, Chincholi, Nashik

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

As a new software paradigm, cloud computing provides services dynamically according to user requirements. HoI'vever, it tends to disclose personal information due to collaborative computing and transparent interactions among SaaS services. I've propose a private data disclosure checking method that can be applied to the collaboration interaction process. First, I've describe the privacy requirement with ontology and description logic. Second, with dynamic description logic, I've validate whether SaaS services are authorized to obtain a user's privacy attributes, to prevent unauthorized services from obtaining their private data. Third, I've monitor authorized SaaS services to guarantee privacy requirements. Therefore, I've can prevent users' private data from being used and propagated illegally. Finally, I've propose privacy disclosure checking algorithms and demonstrate their correctness and feasibility by experiments [7].

Keywords: *Ontology, Privacy Disclosure Detection, Privacy Data Chain, Similarity Metric etc.*

Introduction

Big data usually include data sets that have sizes that are beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time, with characteristics that consist of volume, variety and velocity.[18] According to statistics, an average of 2 million users per second use the Google search engine; within one second, Facebook users share information more than 4 billion times, and Twitter handles more than 3.4 hundred million tI'veets per day. The amount of data grows exponentially every year, and threequarters of data are produced by people, for example, a standard American worker contributes 1.8 million MB every year. A large amount of personal privacy data can be mined for commercial purposes by agents. For example, Acxiomac queries more than 5 million personal data of consumers all over the world through data processing and analyzes individual behaviors and psychological tendencies with technologies, some of which are known as data association and logical reasoning.

In 2014, Adam Sadilekat University of Rochester and John Krumm in the Microsoft lab predicted a person's likelihood to reach a location in the future by analyzing the information in the big data, with an accuracy as high as 80%. A mobile application does not protect the location of the big data; as a result, a user's home address and other sensitive information can be disclosed through the triangulation reasoning method. Research shows that user attributes can be found by analyzing group features in a social network. For example, by analyzing a user's Twitter messages, the user's political leanings, consumption habits and other personal preferences can be found. Therefore, how to protect personal privacy information has become a hot research topic with respect to bigdata.

Literature Survey

A number of researches have been proposed by researchers for privacy preserving in big data. A detailed survey has been carried out to identify the various research articles available in the literature in all the categories of privacy preserving in big data, and to do the analysis of the major contributions and its advantages. Following are the literatures applied for assessment of the state-of-art work on the privacy preserving in big data. Here, few works has been analyzed.

Research article related to privacy preserving in data mining

Data mining is a strategy where huge measures of both sensitive and non-sensitive information are gathered and analyzed. While circulating such private information, security protecting turns into a critical issue. Different strategies and procedures have been presented in security saving information mining to

attempt this issue. The techniques can be listed under three major classical PPDM techniques.

Privacy preserving using Anonymization approach

Asmaa et al.[21] have explained the Protect Privacy of Medical Informatics using K-Anonymization Model. Here, they displayed a structure and model framework for de-distinguishing health data including both organized and unstructured information. They exactly examine a straightforward Bayesian classifier, a Bayesian classifier with an inspecting based strategy, and a contingent irregular field based classifier for removing distinguishing traits from unstructured information. They, convey a k anonymization based system for de-recognizing the removed information to save most extreme information utility. Moreover, Tiancheng Li et al. have explained the Towards Optimal kanonymization which was a more flexible scheme for privacy preserving. Here, they introduced enumerate the algorithm for pruning approach for finding optimal generalization. Likewise,

V.Rajalakshmi and G.S.Anandha Mala [14] amazingly advocated the Anonymization based on nested clustering for privacy preservation in Data

Mining. In their document, the dimension of clusters was preserved optimal to cut down the data loss. They elaborately discussed the technique, performance and outcomes of the nested clustering. Similarly, Yan Zhu and Lin Peng [15] have explained the Study on K-anonymity Models of Sharing Medical Information.

Privacy preserving using Clustering algorithm

In this section, I've discussed the research article based on privacy preserving using clustering algorithm. In S. Patel et al. proficiently introduced a privacy preserving distributed KMeans clustering of horizontally partitioned data which significantly alleviated the safety concerns in the malicious illdisposed model. The vital growth involved the use of the secret transferring mechanism battered to code based zero knowledge identification technique.

Moreover, Bipul Roy brilliantly brought to limelight an innovative tree-based perturbation approach which was easily employed for tackling the perturbing data reflecting the hidden conveyances. In their approach, they made use of a Kd-tree stratagem to recursively divide a dataset into a number of diminutive subsets in such a way that the data records within each subset became further harmonized with every partition. When the partitioning process was fully carried out the confidential data in every subset I'vere perturbed by the effective exploitation of the micro aggregation method.

Additionally, Alper Bilge and Huseyin Polat admirably brought to light the scalable privacy-preserving recommendation technique by means of bisecting k-means clustering. In their innovative privacy-preserving collaborative filtering scheme dependent on bisecting k-means clustering they introduced two pre-processing approaches.

Similarly, Ali Inan et al. intelligently advocated the Privacy preserving clustering on horizontally partitioned data. In the document, they competently created the dissimilarity matrix of objects from varied sites in a privacy preserving fashion which was employed for the purpose of the privacy preserving clustering and also the database joins, record linkage and other function which necessitated couple wise appraisal and assessment of individual private data objects horizontally disseminated to several sites.

In Jinfei Liu et al. efficiently brought to limelight the Privacy Preserving Distributed DBSCAN Clustering. In their innovative technique, they effectively tackled the hassles of the two-party privacy preserving DBSCAN clustering. At the outset, they brilliantly brought in two protocols for privacy preserving DBSCAN clustering over horizontally and vertically segregated data correspondingly and later widened them to the randomly segregated data.

Also, Pan Yang et al. excellently explained the PrivacyPreserving Data Obfuscation Scheme Used in Data Statistics and Data Mining. In their technique, they apportioned dissimilar keys to the diverse users, who I'vere given divergent permissions to access the data. In the ultimate phase, a finegrained grouping method founded on similarity was discussed.

Proposed Methodology

I've can check the private data disclosure chain and the key private data, which can effectively prevent service participants from maliciously disclosing users' private data, increase the service trustworthiness, and provide a basis for a privacy safety-oriented trustworthiness measurement. Detailed contributions are shoI'ved as follows.

- Firstly, I've get the relationships among privacy data by the mapping with knowledge ontology, and build the ontology tree. I've also measure the similarity degrees, containing property similarity, object similarity and hierarchical similarity.
- Secondly, I've measure the cost of the disclosure of the private data with sensitivity grades and privacy disclosure vector. According to the similarity degree and cost of disclosure, the disclosure chain and key private data are detected in the process of interaction betI'veen user and SaaS service.
- Thirdly, I've propose a discovery framework for the private data chain and demonstrate its feasibility and effectiveness by experiments. Which provide a reference to develop a system software assuring the safety for personal privacy data in big data.

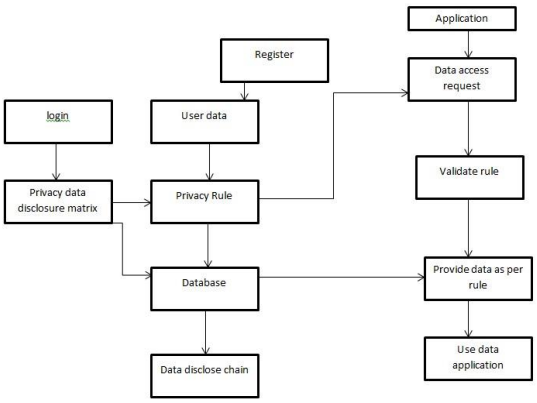

Fig.1 Block Diagram of privacy data chain disclosure discovery method

A. Architecture

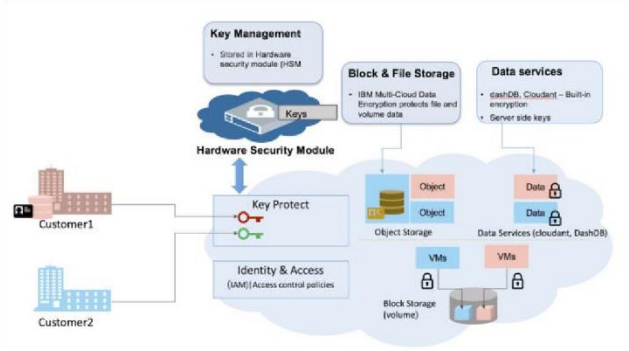

Fig.2 Architecture of privacy data chain disclosure discovery framework

The privacy data chain disclosure discovery framework is designed to have two layers Discovery Layer: When users send an application request for a SaaS service, the private data of the user are required to be released as a precondition and input. Obviously, a user would not provide sensitive private data, including key private data or private disclosure chains. Hence, this layer provides a match betI'veen the user's privacy requirements and preconditions as I'vell as the input requested by the service provider. In other words, it detects the key private data and disclosure chains in the re- quested privacy data collection. Ontology Mapping Layer: On the basis of the detection layer, I've discovered the key private data and disclosure chains from the privacy data collection. Then, the key private data and disclosure chains are decomposed and discretized by semantic matching with the privacy ontology tree. In more detail, private data are selected from key private data or disclosure chains as a root node, and I've traverse the privacy ontology tree to find its corresponding child nodes and substitute them to satisfy the user's privacy requirements. Finally, I've return the restructured private data collection to the users for confirmation.

B. Algorithms

procedure init

1: sCPAk(t) ← 0 for all 1 ≤ t ≤ n
2: pk ← 0
3: if k > 1 do 4: sUBk ← 0
5: else
6: sUBk ← q∞
7: assign CPA() procedure assign CPA
8: if pk = 0 do
9: Generate a new random ordering of Dk intowk
10: pk ← pk +
11: if pk > |Dk|do
12: backtrack()
13: else
14: Xk ← v := wk(pk)
15: update shares in CPA(k,v)
16: if k = n do
17: ifcompare CPA cost to upper bound() = true do
18: broadcast(NEW OPTIMUM FOUND)
19: assign CPA() 20: else
21: ifcompare CPA cost to upper bound() = false do
22: assign CPA()
23: else
24: send(CPA MSG) to Ak+1 procedure backtrack
25: if k > 1 do
26: sCPAk(t) ← 0 for all t ∈ I− k
27: send(ZERO SHARE MSG,k) to At for all t ∈ I− k
28: send(BACKTRACK MSG) to Ak−1
29: else 30: broadcast(COMPLETE) When received (NEW OPTIMUM FOUND)do
31: sUBk ←Pt∈Ik sCPAk(t)
32: OptimalSettingk ← Xk whenreceived (CPA MSG)do
33: pk ← 0
34: assign CPA() when received (ZERO SHARE MSG,k0)do
35: sCPAk(k0) ← 0 When received (BACKTRACK MSG)do
36: assign CPA() when received (COMPLETE)do
37: Xk ← Optimal Setting k
38: Terminate

C. Mathematical Modules

System Description:

- Let S be system having sets of parameter

Set S= ((I) , (R) ,(P) ,(O))

Where, S is Storage of data according to the classification I is set of all inputs given for storing the data R is set of rules that drives your input set.
P is set of all processes in system. O is set of output expected from system.

- Inputs (I) :I1,I2,I3

Where, I1 = Data Uploading I2 = Processing
I3 = Securing Data

- Rules (R) : R1,R2

Where, R1 = Sensitive Data should be Different R2 = Authencation Required

Processes (P) : P1,P2,P3
Where, P1 = Conceptually Sorting of Data
P2 = Hierarchical Classification
P3 = Similarity of Data

- Output (O):O1,O2,O3

Where, O1 = Sensitive Data Is disclosed In Private O2 = Targeted user can only access Data
O3 = Encrypting the decrypted Data

Result

We analyze the security and efficiency regarding data in our system. In this research conducted, privacy preservation using encryption algorithm. This encryption also verified after downloading, if any manipulation in encrypted content. All rights regard to user authentication are reserved for Admin. System performance are used to show the effect on the efficiency of the Privacy and time cost.

Conclusions

According to the interaction characters among SaaS services, I've propose a privacy disclosure checking method that satisfies users' requirements. I've develop a prototype system, which describes the users' privacy requirements and extends BPEL and its execution engine to meet users' privacy requirements. I've also design a case and run it on the prototype system to confirm the feasibility and correctness of our method. Our approach can check privacy disclosure behavior among SaaS services, which can effectively prevent service participants from maliciously disclosing users' privacy information, increase service credibility, and provide a basis for privacy protection-oriented credibility measurement. The next step is to detect the release of the data of users' privacy, analyze the data, and discretize the dataset that may be exposed to protect users' privacy before they are released..

Reference

[1]. Changbo Ke, Fu Xiao, Member, IEEE, Zhiqiu Huang, Yunfei Meng and Yan Cao, "Ontology-Based Privacy Data Chain Disclosure Discovery Method for Big Data",DOI 10.1109/TSC.2019.2921583, IEEE Transactions on Services Computing

[2]. He X, Ai Q, Qiu R C, et al. A big data architecture design for smart grids based on random matrix theory [J]. IEEE transactions on smart Grid, 2017, 8(2): 674-686. [3] Wang Y, Kung L A, Wang W Y C, et al. An integrated big data analyticsenabled transformation model: Application to health care [J]. Information Management, 2018, 55(1): 64-79. [4] Peddinti S T, Ross K W, Cappos J. User Anonymity on Twitter[J]. IEEE Security Privacy, 2017, 15(3): 84-87. [5] Wan J, Tang S, Li D, et al. A manufacturing big data solution for active preventive maintenance [J]. IEEE Transactions on Industrial Informatics, 2017, 13(4): 20392047.

[3]. Storey V C, Song I Y. Big data technologies and management: What conceptual modeling can do [J]. Data Knowledge Engineering, 2017, 108: 50-67.

[4]. Ke C, Huang Z, Cheng X. "Privacy Disclosure Checking Method Applied on Collaboration Interactions Among SaaS Services[J].IEEE Access, 2017, 5: 15080-15092. [8] Zeydan E, Bastug E, Bennis M, et al."' Big data caching for networking: Moving from cloud to edge [J]. IEEE

[5]. Communications Magazine, 2016, 54(9): 36-42. [9] Imran-Daud, Malik, "Ontology-based Access Control in Open Scenarios: Applications to Social Networks and the

[6]. Cloud", eprint arXiv:1612.09527,Pub Date: December 2016 [10] Zhang X, Dou W, Pei J, et al. "Proximity-aware localrecoding anonymization with mapreduce for scalable big data privacy preservation in cloud [J]. IEEE transactions on computers, 2015, 64(8): 2293-2307. 29

[7]. Lv Y, Duan Y, Kang W, et al. "Traffic flow prediction with big data: A deep learning approach [J]. IEEE Trans. Intelligent Transportation Systems, 2015, 16(2): 865-873.

[8]. Shweta Taneja, Shashank Khanna, Sugandha Tilwalia and Ankita, "A Review on Privacy Preserving Data Mining: Techniques and Research Challenges", International Journal of Computer Science and Information Technologies, vol. 5, no. 2, pp.23102315, 2014.

[9]. Chhaya S Dule, H.A. Girijamma and K.M Rajasekharaiah, "Privacy Preservation Enriched MapReduce for Hadoop Based BigData Applications",American International Journal of Research in Science, Technology, Engineering Mathematics, vol.6, no.3, pp. 293-299, 2014

[10]. V Rajalaxmi,GSA MALA," Anonymization based on nested clustering for privacy preservation in data mining,IJCSE, 2013

[11]. Xun Xu, "From Cloud Computing to Cloud Manufacturing", Robotics and Computer Integrated

[12]. Manufacturing, vol. 28, no.1, pp. 75-86, 2012. [16] Jadeja and Kirit Modi, "Cloud Computing—Concepts,

[13]. Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies, 2012.

[14]. M. Prakash, G. Singaravel, "A New Model for Privacy Preserving Sensitive Data Mining", in proceedings of ICCCNT Coimbatore, India, IEEE 2012.

[15]. Divyakant Agrawal, Amr El Abbadi and Shiyuan Wang, "Secure and Privacy Preserving Data Services in the Cloud: A Data Centric View", Proceedings of the VLDB Endowment, Vol. 5, No. 12, 2012

[16]. Snijders, C.; Matzat, U.; Reips, U.-D. (2012). "Big Data':

[17]. Big gaps of knowledge in the field of Internet". International Journal of Internet Science 7: 1–5

[18]. D. Zissis and D. Lekkas, "Addressing Cloud Computing Security Issues", Future Generation Computer Systems, vol. 28, no. 3, pp. 583- 592, 2011.

[19]. Asmaa H. Rashid, A. F.Hegazy, "Protect privacy of medical informatics using Kanonymization madel", IEEE, April 2010

[20]. S. P. Deshpande and V. M. Thakare, "Data Mining System and Applications: A Review ", International Journal of Distributed and Parallel systems, vol.1, no.1, 2010.