Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage: www.akademiabaru.com/arca.html
ISSN: 2462-1927

Enhancing Fine Grained Technique for Maintaining Data Privacy

Open Access

Esraa A. Elwan[1,*], Mohamed Elkawkagy[1], Arabi Keshk[1]

[1] Computer Science Department, Faculty of Computers and Information, Menofyia University, Shebin El Kom, Egypt

ARTICLE INFO

Article history:
Received 18 October 2017
Received in revised form 12 December 2017
Accepted 3 March 2017
Available online 3 April 2018

Keywords:
Data privacy, fine-grained encryption, perturbation based, sensitive attributes, quasi identifiers

ABSTRACT

Privacy is a big challenge in a lot of fields such as healthcare field. As technology advances, data privacy became at risk and must to protect individuals' personal information and sensitive data. The existing privacy preserving techniques as encryption whether full or partial encryption are not sufficient because they have some defects. In this paper, we will introduce an enhanced technique whose main mission is maintaining individuals' privacy and increasing the utility of data. By combining fine-grained encryption method with the proposed perturbation technique. Where we will use fine-grained encryption method to encrypt only key and sensitive attributes. On the other side, we will use proposed perturbation technique to add a small amount of random noise to quasi-identifier in order to increase the security of data and prevent disclosure it by linking with other external data. Our evaluation proves that the introduced technique is more secure than other privacy preserving techniques and consumes less memory.

1. Introduction

Privacy is a critical aspect of protecting and securing the data itself especially sensitive and personal data in different fields such as Healthcare, Financial, etc. Privacy can only be achieved by using several approaches for data privacy preserving techniques which divided into three approaches anonymization (De-identification), perturbation and encryption.

A. In Anonymization Approach

The anonymization of the data set refers to obscured identifying attributes from the original data to maintain the privacy and secure personal and sensitive information before being released or published. Data anonymization approaches divided into three categories: k-anonymity, l-diversity, and t-closeness. There are two popular anonymization methods suppression and

* *Corresponding author.*
E-mail address: esraa.1402@gmail.com (Esraa A. Elwan)

generalization. In suppression anonymization method, attribute values or an entire tuple are replaced by an asterisk '*.' In generalization anonymization method, attribute values are replaced by a range of values as replacing value 22 in column A with (20 – 25).

B. In Perturbation Approach

Here data privacy is maintained by modifying the data record value without changing the implicit meaning of data. It classified into two categories, probability distribution that replaces the original data by the same distribution's sample data or by itself and the value distortion that changes the original values by adding some noise either additive or multiplicative. Achieving the balance between data privacy and data utility is a big challenge of data perturbation. Data perturbation techniques divided into two essential categories:

1.A.1. Noise Additive Perturbation "column-based additive randomization."

It is based on adding the amount of random noise to records' values in specified columns.

1.A.2. Multiplicative Perturbation or multi-dimensional perturbation technique

Here the rows values are multiplied with random noise to get the perturbed values. There are different ways can do this kind of perturbation, Rotation Perturbation where values of the two attributes in the matrix are rotated. Thus, the data values of multiple columns converted to a single column, but the meaning of the value is protected. Random Projection Perturbation which reduces the dimensional space of a set of data points from high to a randomly chosen lower. Geometric Data Perturbation which is an enhancement to rotation perturbation by combining additional elements to the basic form of multiplicative perturbation.

C. In encryption approach

There are two different kinds of encryption approaches. Symmetric key approach "secret key cryptosystem" which uses same keys for encrypting and decrypting such as DES, 3DES, and AES algorithms. Asymmetric key algorithms "public key cryptosystem," which uses different keys for encryption and decryption such as RSA and ElGamal algorithms. Although encryption is the best approach to maintain privacy which ensures a very high level of data privacy, it reduces system's efficiency and consumes memory. Therefore, the researchers used the fine-grained encryption method (FGEM) to encrypt only sensitive data. Data in FGEM is stored in a tree, and then sensitive attributes are determined by calculating some steps and finally using one of the encryption algorithms to encrypt them [13]. However, this solution is insufficient and is not more secure as it is easy to reach the original data by linking secured released data to other external data like voters list. In this paper, we will propose a new efficient and more secure algorithm to maintain data secure and prevent its discovery by combining the FGEM algorithm and perturbation technique. This combination will increase the efficiency of the system and reduce the processing time.

2. Related Work

Nimpal Patel *et al*., [1] proposed a technique based on Geometric data perturbation to protect the privacy of sensitive attribute. They had reduced the trade-off between mining accuracy while minimizing information loss during data perturbation. Aniket Patel *et al*., [2] proposed a Rotation

Transformation Based technique for Principal Component Analysis which transform a given data set D into another form called perturbed version D' that achieve the requirement of privacy and loss minimum information for the intended data analysis task. Bhupendra Kumar Pandya *et al*., [8] Random projection is one of Dimensionality reduction techniques which is very useful as it produces expected result where many data mining algorithms applied directly to the perturbed data. Expected Euclidean distance is preserved after perturbation. Parameshachari *et al.*, [9] introduced a new method which applies a partial encryption in images using phase manipulation block. The image is scrambled to get modified image after applying Inverse Fourier Transform then using sign encryption which encrypts the sign bits of the modified image. Yan Li *et al*., [10] proposed Horus technique which uses keyed hash trees (KHTs). This technique used to encrypt large data sets by generating different keys for each region of the data set, which cannot be used to access another. Chunguang MA *et al*., [13] invented the Fine-Grained Encryption Method (FGEM). They described the FGEM concept, mentioned its advantages and finally proposed a tree structure to achieve the FGEM. Priyanka *et al*., [14] proposed a technique which applies a partial encryption in random pixels of an image. They generate a random number to select random pixels and encrypt them using symmetric key cryptosystem. Finally, they combined the result with remaining original part of the image.

Table 1
Attributes Types

Attribute type	Property	example	Action required
Key	Can identify an individual directly	id, name	Fine-grained encryption
Quasi-identifier	It can identify the original data by linking with external information.	Age, gender	Add noise
Sensitive attribute	Data that an individual is sensitive about revealing	address, history	Fine-grained encryption
Insensitive attribute	Include those attributes which are not sensitive.	condition, suit	No action

3. Proposed Technique

In general, the fine-grained encryption method is insufficient and is not secure because it is easy to link the secured released data to other external data. In the proposed technique, we have classified data attributes into the key, sensitive attribute, quasi-identifier and insensitive attribute as shown in Table 1. In this paper, we will introduce a more secure and efficient technique to maintain the privacy of data and keep it secure. We will improve the fine grained encryption algorithm which encrypts only sensitive attributes using any encryption algorithm. As shown in Table 1, we start to determine the Key (id, name), sensitive attributes (address, history) and quasi-identifiers (age, gender) from the dataset. According to Figure 1, key and sensitive attributes are given as an input to apply fine-grained encryption algorithm using any encryption technique. Second quasi-identifiers are given as an input to proposed perturbation technique. Through the proposed perturbation technique, the quasi-identifiers are converted to strings. Then a small amount of random noise will be added to them. After that, the result will be converted to its original type (integer\ categorical) again. Finally, we will have an ambiguous and a more secure data. This will lead to preventing disclosure sensitive data when linking with external data occurs. This means that the proposed algorithm will lead to high level of privacy and high level of data utility.

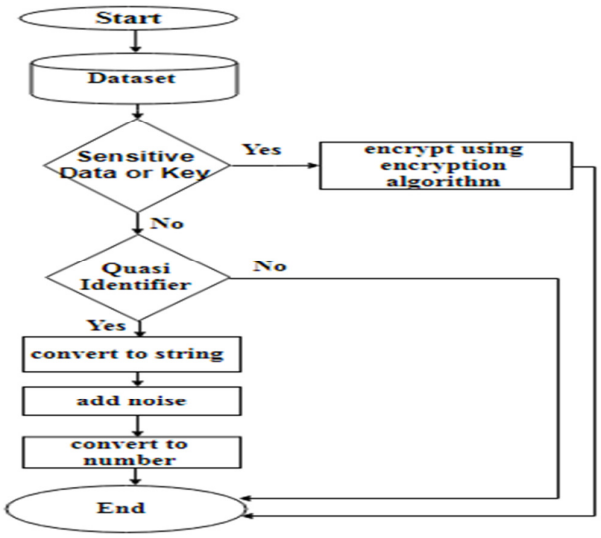

Fig. 1. Framework of FGEPT

4. Discussion and Results

Series of experiments were performed on different sizes of healthcare data set (KB) by computing memory consumption and throughput in the following steps: First, for full encryption using both RSA and DES encryption algorithms. As shown in Table 2, the throughput of DES is larger than RSA and this leads to DES consumes power less than RSA. Also, we observed that whenever data size increases, the execution time of DES becomes less than the execution time of RSA where RSA consumes execution time by 55.6% and DES consumes by 55.1%. As shown in Table 3, the RSA consumes less memory than DES where RSA consumes memory by 0.15% while DES consumes by 0.25%. Note that the throughput of Encryption algorithm equals quotient the average of total plain text in Megabytes and the average of Encryption time in seconds and average of memory consumption (Byte/Sec) = sum of memory consumption/sum of execution time.

Table 2
Throughput of RSA and DES (full encryption, fine-grained encryption), perturbation and proposed technique using both RSA and DES

Input Size (KB)	RSA	DES	Fine Grained Encryption (RSA)	Fine Grained Encryption (DES)	Perturbation	Proposed Technique using (RSA)	Proposed Technique using (DES)
50	2.26	2.57	1.12	1.33	2.11	1.67	1.84
100	7.19	7.55	3.73	3.85	7.38	5.65	5.80
200	27.16	27.36	13.92	13.76	29.60	22.47	20.55
300	57.71	57.67	28.74	29.16	57.60	43.43	43.29
400	110.57	102.43	52.04	51.53	107.48	84.89	77.30
500	161.77	157.33	83.09	78.48	156.02	116.16	113.08
600	237.67	231.81	143.23	141.52	244.12	203.01	198.32
800	487.61	485.86	261.64	260.87	427.84	373.85	366.59
1000	850.52	863.94	448.39	412.51	796.2	647.37	611.49
2000	3619.61	3575.37	1977.28	1869.74	3317.48	2725.85	2421.14
Throughput (MB/Sec.)	5.95/5562.07 =0.001070	5.95/5511.89 =0.001079	5.95/3013.18 =0.001975	5.95/2862.75 =0.002078	5.95/5145.83 =0.001156	5.95/4224.35 =0.001409	5.95/3859.40 =0.001542

Second for Fine Grained Encryption algorithm using RSA and DES to encrypt Key and sensitive attributes. Tables 2 and 3 show that the DES is slightly faster than RSA when data size is small and whenever data size increases the execution time of FGEM with DES become the smallest execution time of other techniques. Where FGEM using RSA consumes execution time by 30.13%, while FGEM using DES consumes by 28.63%. Contrariwise FGEM with DES consumes more memory than FGEM with RSA where FGEM using RSA consumes by 0.21%, and FGEM using DES consumes by 0.41%. Third using proposed perturbation technique to perturb all attributes. The experimental in Tables 2 and 3 show that the throughput value of proposed perturbation technique lies between throughput values of fine-grained encryption technique and full encryption technique. As it consumes execution time by 51.46%. The proposed perturbation technique has the lowest memory consumption of other techniques as it consumes by 0.07%. Finally, using our proposed technique with both RSA and DES to encrypt key and sensitive data and perturb quasi-identifiers. Tables 2 and 3 show that whenever data size increases the throughput and the execution time of proposed technique using RSA is larger than the throughput of proposed technique using DES. As proposed technique using RSA consumes execution time by 42.24% and proposed technique using DES consumes by 38.59%. The proposed technique using DES consumes more memory than proposed technique using RSA. As proposed technique using RSA consumes memory 0.17% and proposed technique using DES consumes by 0.32%. Thus, proposed technique using DES consumes less power, execution time and more memory than proposed technique using RSA. The simulation results of Tables 2 and 3, proposed perturbation technique has the lowest memory consumption of other techniques (consumes 0.07%) and FGEM technique using DES consumes memory more than other techniques (consumes by 0.41%). As a result, FGEPT using DES consumed a convenient memory (consumes 0.32%) that is smaller than FGEM and larger than proposed perturbation technique. And also proposed a technique using DES consumes execution time (consumes 38.59%) more than FGEM using DES (consumes 28.63%) and less than other techniques. Thus, the proposed technique using DES is more secure and a small information loss, fast and consumes less power than proposed technique using RSA.

Table 3
Memory consumption of RSA and DES (full encryption, fine-grained encryption), perturbation and proposed technique using both RSA and DES

Input Size (KB)	RSA	DES	FGEM (RSA)	FGEM (DES)	Perturbation	Proposed Technique using (RSA)	Proposed Technique using (DES)
50	0.3756	1.0606	0.4693	1.0363	0.3284	0.2788	1.4200
100	0.2838	0.6897	-0.0088	0.7079	-0.3476	0.0805	0.7387
200	0.0720	0.8075	-0.1792	0.4640	-0.0741	0.0518	0.7435
300	0.2489	0.8692	0.1483	0.6780	-0.0588	0.0663	0.6162
400	0.3082	0.8681	0.2006	0.7817	0.0451	0.2251	0.7542
500	0.5543	1.0214	0.4791	0.8998	0.1050	0.3279	0.8427
600	0.7984	1.4552	0.5309	1.1174	0.8077	0.5455	1.0730
800	1.0825	1.5368	0.9125	1.2655	0.4902	0.9430	1.6095
1000	1.0807	1.9154	1.2759	1.8983	0.8261	1.4528	1.6761
2000	2.9526	3.4072	2.4768	2.7928	1.4535	3.2279	2.9820
Average of memory consumption (B/S)	8.3977/5562.07=0.001499	13.6241/5511.89=0.002472	6.30543013.18=0.002093	11.5997/2862.75=0.004052	3.5755/5145.83=0.000695	7.1996/4224.35=0.001704	12.4359/3659.40=0.003227

5. Conclusion

There are many techniques to maintain and protect data privacy. The most important one is encryption techniques, but the full encryption has many defects as memory consumption and the

slow speed of the system. The partial encryption is not sufficiently secure and does not preserve on the data protection fully. So the proposed technique add a small amount of random noise to fine-grained encryption (partial encryption) to increase the privacy of data and make it secure. Our evaluation proves that the proposed technique not only can preserve data privacy but also can consume less power, memory than other privacy preserving techniques.

References

[1] Patel, Nimpal, and Shreya Patel. "Geometric Data Perturbation Techniques in Privacy Preserving On Data Stream Mining." (2016).

[2] Aniket Patel and Rahul Shrimali," Rotation Transformation Based Multiplicative Data Perturbation for Privacy Preserving Data Mining,"1st International Conference on "Advances in Engineering at India, Volume - 5 | Issue - 1 Jan Special Issue - 2015 | ISSN - 2249-555X.

[3] Nimpal Patel and Shreya Patel, "A Study on Data Perturbation Techniques in Privacy Preserving Data Mining," International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 09, Dec-2015.

[4] Ankleshwaria, Twinkle, and J. S. Dhobi. "A SURVEY OF PERTURBATION TECHNIQUE FOR PRIVACY–PRESERVING OF DATA." *International Journal* 2, no. 1 (2014).

[5] Patel, Aniket, and H. S. Patel. "A Study of Data Perturbation Techniques For Privacy Preserving Data Mining." *Keyvanpour, M., & Moradi, SS (2011). Classification and evaluation the privacy preserving data mining techniques by using a data modification-based framework. arXiv preprint* (2014).

[6] Patel, Samir, Gargi Shah, and Aniket Patel. "Techniques of Data Perturbation for privacy Preserving Data Mining"." *International Journal of Advent Research in Computer & Electronics (IJARCE) Vol* 1: 5-10.

[7] Srivastava, Avruti. "Comparative Study of Privacy Preservation Techniques in Data Mining."

[8] Pandya, B., Umesh Kumar Singh, and Keerti Dixit. "A Study of Projection Based Multiplicative Data Perturbation for Privacy Preserving Data Mining." *International Journal of Application or Innovation in Engineering and Management* 3, no. 11 (2014): 180-182.

[9] Parameshachari, B. D., KM Sunjiv Soyjaudah, and Sumitra Devi KA. "Secure transmission of an image using partial encryption based algorithm." *International Journal of Computer Applications* 63, no. 16 (2013).

[10] Li, Yan, Nakul Sanjay Dhotre, Yasuhiro Ohara, Thomas M. Kroeger, Ethan L. Miller, and Darrell DE Long. "Horus: fine-grained encryption-based security for large-scale storage." In *FAST*, pp. 147-160. 2013.

[11] Ilavarasi, A. K., B. Sathiyabhama, and S. Poorani. "A survey on privacy preserving data mining techniques." *International Journal of Computer Science and Business Informatics* 7, no. 1 (2013): 209-221.

[12] Patel, Aniket, Keyur Dodiya, and Samir Pate. "A Survey On Geometric Data Perturbation In Multiplicative Data Perturbation." *International Journal* 1, no. 5 (2013).

[13] Ma, Chunguang, Changli Zhou, Jiuru WANG, and Xiaorui ZHONG. "A Research of Fine-grained Encryption Method for IoT." *Journal of Computational Information Systems* 8, no. 24 (2012): 10213-10222.

[14] Agrawal, Priyanka, and Manisha Rajpoot. "Partial Encryption Algorithm for Secure Transmission of Multimedia Messages 1." (2012).

[15] Krishna Priya .J and Geetha Mary. A " Perturbation of String Values," International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1257-1259.