IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 3, 2019, accepted February 12, 2019, date of publication February 28, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902011

PPM: Privacy Protection Method for Outsourcing Data Entry

FENGQI LI, CHUNLI SHANG, KEMENG LIU, AIDI PANG, AND SIKAI HUANG
School of Software, Dalian University of Technology, Dalian 116600, China
Corresponding author: Fengqi Li (lifengqi@dlut.edu.cn)

ABSTRACT In China, manual approaches are still widely used in outsourcing data entry because of the poor recognition of OCR for Chinese. However, manual entries from images may cause leakage of users' privacy. In this paper, we propose a privacy protection method to reduce privacy disclosure. First, we use the EAST algorithm to segment whole images into sub-images. Second, we propose a privacy association separation algorithm to protect users' complete information which could be derived from privacy associations between attributes. Finally, we propose a sub-image allocation algorithm that varies for different relations between attributes, sub-sets, and data-entry clerks. We use AHP to model and optimize sub-contractor selection as well. The experimental results on OutsourcingData and U.S. Census Adult dataset show the advantages of our proposed method in terms of both the attribute privacy protection rate and the member privacy protection rate.

INDEX TERMS Image segmentation, outsourcing data entry, privacy protection, privacy association separation, sub-image allocation.

I. INTRODUCTION

As a type of the outsourcing service [1], Data entry, especially image data entry is commonly used in multiple industries such as finance [2], medical industry [3], publishing, etc. OCR [4] is a technology which is widely used in the image data entry. However, this approach is seldom utilized in images containing Chinese information because of its poor recognition for languages with a large number of complex glyphs like Chinese and Japanese. Thus, the OCR is not suitable for Chinese and Japanese. So manual method is still indispensable in China [5] and the process is shown in Figure 1.

Based on the process in Figure 1, we know that data-entry clerks have a high probability of obtaining complete images which may raise the risk of leaking users' privacy. Moreover, sub-images containing users' privacy associations could still contribute to privacy leakage.

Existing privacy protection technologies include encryption [6], data distortion and limited release [7]. Encryption generates a large number of public and private keys, which greatly increases the time cost and the computational cost. Data distortion cannot guarantee the exact entry of the original data. Limited release affects the accuracy and causes data missing as well. Unfortunately, as outsourcing data entry requires accurate and original data, exiting privacy protection technologies cannot be applied to the outsourcing data entry.

In cloud outsourcing, Elmehdwi *et al.* [8] focused on the query of the encrypted databases and Cheng [9] studied the security scheme for large-scale matrix computing. In addition, for the outsourcing encryption, Lee *et al.* [10] focused on the data encryption when outsourcing mobile sensitive data. And Jun [11] studied the cipher-text access control of outsourcing storage systems and the cryptography of outsourcing computing. However, they did not pay attention to the leakage of users' privacy caused by complete images and sub-images containing privacy associations.

This paper focuses on solving the two problem mentioned above. The main contributions of this paper are as follows:

- This paper proposes PPM which could increase the security of users' privacy. The algorithm consists of three steps: images segmentation, privacy association separation and sub-image allocation.
- This paper makes use of EAST [12] which utilizes a fully convolutional network (FCN) to segment images.
- This paper evaluates PPM on outsourcing dataset, OutsourcingData and the US Census Adult Dataset.

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

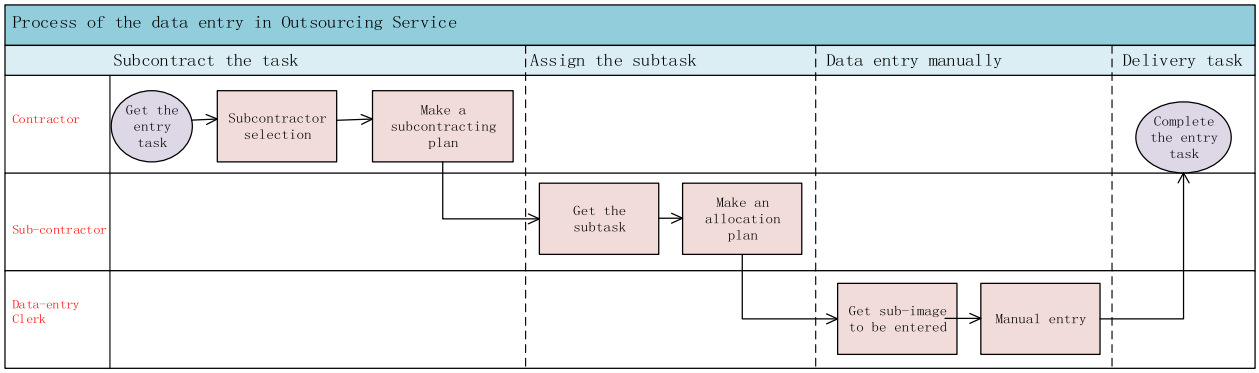


FIGURE 1. Process of the data entry in outsourcing service.

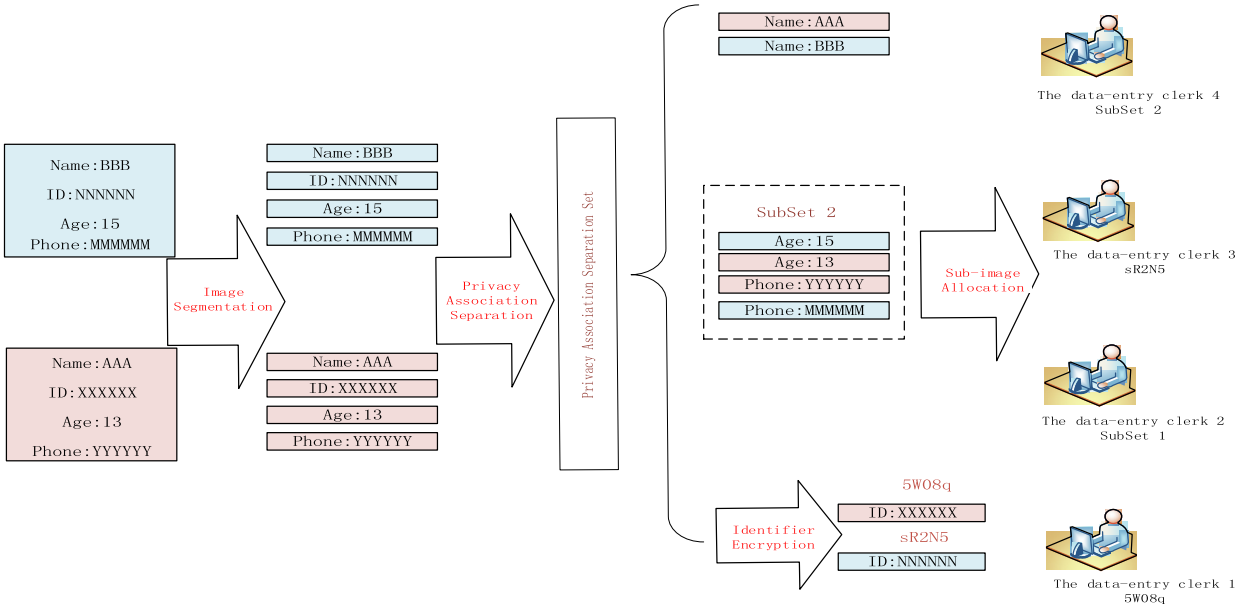


FIGURE 2. Process of privacy protection method in outsourcing data entry.

II. PRIVACY PROTECTION METHOD BASED ON OUTSOURCING DATA ENTRY

A. PROGRAMING DESIGN

The process of this method is shown in Figure 2. PPM consists of three parts: image segmentation, privacy association separation and sub-image allocation.

1) IMAGE SEGMENTATION

The image segmentation is mostly completed by the EAST algorithm. The EAST, which is based on the fully convolutional network, is used to detect the text region of complete images. EAST can allow the image to be in any directions, abbreviating the intermediate process. After the detection, the text region is segmented from the whole image to get the sub-image set.

2) PRIVACY ASSOCIATION SEPARATION

A privacy association separation algorithm should be utilized on sub-images which contain privacy associations. The algorithm has three steps. First, perform the attribute-marked to the sub-images. Second, separate the defined privacy associations by traversal to obtain the privacy association separation set. Third, after the separation, encrypt the identifier attribute that uniquely identifies the user.

3) SUB-IMAGE ALLOCATION

According to different relationships between the number of attributes, sub-sets and clerks, we choose different algorithms for sub-image allocation.

B. IMAGE SEGMENTATION ALGORITHM BASED ON THE EAST

Image segmentation is to segment the image ij ($j = \{1, 2, \ldots, k\}$) in the image set I to a set S of m sub-images according to the page character feature, wherein each sub-image is sj($j = \{1, 2, \ldots, m\}$). Then you need to satisfy $\bigcup_{j=1}^{k} ij = I$, satisfy $ip \cap iq = \phi$ for $1 \leq p \neq q \leq k$; and satisfy $\bigcup_{j=1}^{m} sj = S$, and satisfy $sp \cap sq = \phi$ for $1 \leq p \neq q \leq m$.

Image types are basic images, tables, vertical images, irregular images, and hyperspectral images. The processing of

hyperspectral images is in the literature [12], as well as the dimensionality reduction [13] and feature learning [14], [15] of hyperspectral images. But this article focuses on the first four categories.

Definition 1: Basic image: Refers to an image with a blank area separating adjacent text.

Definition 2: Table: Refers to an image with horizontal lines separating adjacent text.

Definition 3: Vertical image: Refers to an image in which the text is written in a portrait orientation.

Definition 4: Irregular image: Refers to an image with a special text distribution and multiple distinct separate text regions.

This paper cites the EAST algorithm proposed in the literature [16] for the image segmentation of the first four categories. EAST proposed End-to-End full-convolution network (FCN) [17]–[20] to solve the problem of text detection. It can generate geometric annotations in two formats: quadrilateral or rotated box according to specific applications. The article draws on the idea of inception and adopts convolution of various sizes to find the target of different sizes in the target image. Use u-net's non-pool and concat when doing the split. EAST is not only concise but also can detect the text region from the entire image intelligently.

In view of the advantages of the EAST, the main steps of the image segmentation based on the EAST are as follows:

Step 1: Feed whole images into PVANet [21] to extract features in images [22].

Step 2: The output of previous PVANet passes through three blocks containing one non-pooling layer, one $1*1$ convolution layer [23] and one $3*3$ convolution layer to generate a feature map which would be sent to output layer then.

Step 3: Output the prediction information. If the predicted shape is RBOX (rotated rectangle), then output a score, four regression boxes and an angle information. If the predicted shape is QUAD (quadrilateral), output a score and eight coordinate information.

Step 4: Perform the non-maximum suppression (NMS) operation on the prediction obtained in the previous step to get the final segmentation results.

The network was implemented on Tensorflow framework. We tested performance of EAST on an image from a Japanese outsourcing company. The segmentation results are shown in Figure 3.

One can see from Figure 3 that EAST automatically generates green borders on the image to circle the Japanese text. We split the complete image into sub-images based on the green borders. In general, the text of the image is basically positioned by green borders. Results suggest that, as a part of our PPM algorithm, EAST has a good performance on text detection and image segmentation.

C. PRIVACY ASSOCIATION SEPARATION ALGORITHM

Image segmentation could decrease the risk of leaking users' privacy by separating users' complete information in image. However, clerk could still derive users' complete information from privacy associations in sub-images. For example, the combination of two attributes, age and gender, can constitute a privacy association. Therefore, it is necessary to perform a privacy association separation algorithm to protect identifier from deriving by privacy association.

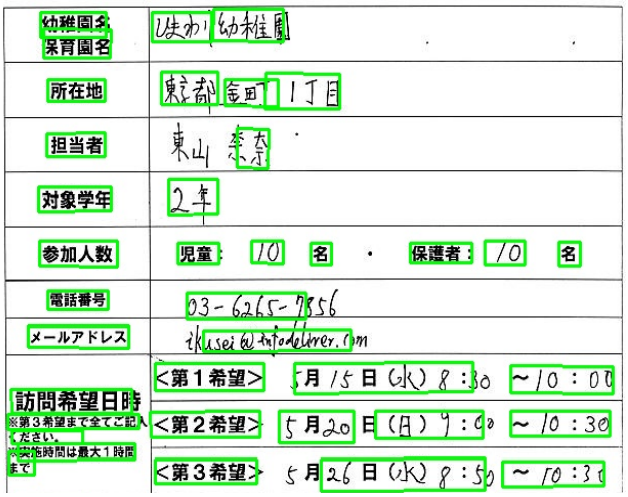


FIGURE 3. Image segmentation result based on the EAST.

The basic idea of the privacy association separation algorithm is to separate attributes in the privacy association by tagging attributes and traversing privacy associations. As a result we can get the privacy association separation set. It is necessary to ensure that attributes in each subset of the privacy association separate set do not constitute a privacy association.

In order to implement the privacy association separation algorithm, some basic concepts are defined below.

Definition 1: Identifier (*ID*). The attribute of an individual can be uniquely determined in a data set.

Definition 2: Quasi-Identifier (*QI*). The attributes can be linked to the external information in a data set, resulting in the leakage of users' information.

Definition 3: Sensitive Attribute (*SA*), the attributes that involve the user's sensitive information in a data set. Sensitive attributes are usually the user's private information and need to be protected.

Definition 4: General Attribute (*GA*), in a data set, refers to the attributes other than the identifiers, the quasi-identifiers, and the sensitive attributes. The general attributes are usually not directly or potentially linked to the user's private information.

Definition 5: Privacy Association Set (*SC*), it is a set of all privacy associations existing in the sub-image set S, and S is obtained by the image segmentation. Each subset of the *SC* may be represented by $SC_j(j = \{1, 2, \ldots, k\})$. SC_j, a set of privacy associations between the sub-images segmented from the original image ij.

Definition 6: Privacy Association Separation Set (*RE*), for the element SC_j (in which $j = \{1, 2, \ldots, k\}$) in the privacy association set *SC*, the privacy association separation algorithm is operated to get the privacy association separation

set RE. Each subset in the RE can be represented by RE_j (in which $j = \{1, 2, \ldots, k\}$). And the internal attributes of each subset in RE_j do not constitute a privacy association.

Combined definitions above, the privacy association separation algorithm in this paper is described as follows:

Step 1: Label the information of the sub-image, which contains the identifier *ID*, the quasi-identifier *QI*, the sensitive attribute *SA*, and the regular attribute *GA*.

Step 2: Combine the privacy associations between the *ID*, *QI* and *SA* in the data anonymous model. And then traverse all the privacy associations in the SC_j. For the attribute determined as the identifier, we put it into the privacy association separation set RE_j.

Step 3: First, traverse all attributes except the identifier. Second, determine the correct position of the current attribute in the subset of the RE_j. Third, remove the attributes that have been placed in RE_j from all attributes. Finally, continue the loop until all attributes in the SC_j find the correct position in the RE_j. Need to satisfy, $\bigcup_{j=1}^{k} RE_j = RE$ and for $1 \leq p \neq q \leq k$ need to satisfy that $RE_p \cap RE_q = \phi$.

The pseudo-code of privacy association separation algorithm is given in Algorithm 1.

Algorithm 1

Input: SC_j
Output: RE_j
(1) Attribute tagging the elements in a sub-image set
(2) **for** Traverse each subset in SC_j **do**
(3) **if** The number of attributes in current subset is 1. **then**
(4) The current attribute is *ID*, and put it into a new subset of the set RE_j.
(5) **end if**
(6) **end**
(7) Remove the attributes that have been placed in the set RE_j from SC_j
(8) **for** Traverse each attribute in SC_j (the attributes haven't been placed in the set RE_j) **do**
(9) **for** Traverse each subset in SC_j **do**
(10) **if** The current subset in the RE_j coincides with the subset in the SC_j **then**
(11) Move to the next subset in the set RE_j
(12) **break**
(13) **else** flag=flag+1
(14) **end if**
(15) **end**
(16) **if** flag=the number of the subsets in the set SC_j
(17) Put current attribute to the current subset in RE_j
(18) **end if**
(19) **end**

To increase the safety of privacy, sub-images containing identifier information need to be encrypted after the separation of privacy associations. Encryption algorithm consists of two parts. First, generate a random sequence based on the number of vertical pixels and horizontal pixels in sub-images. Second, according to the random sequence, scramble the rows and the columns of the sub-images respectively.

D. SUB-IMAGE ALLOCATION ALGORITHM

1) SUB-IMAGE ALLOCATION

Each clerk may be allocated more than one sub-image when the number of clerks is inadequate. So a reasonable allocation strategy is necessary here to hold the security of privacy. This paper proposes a sub-image allocation algorithm. This method considers the allocation strategy to ensure the rationality of the distribution and the security of the privacy.

The basic idea of the sub-image allocation algorithm is to consider the relationships among the number of attributes h and sub-sets f in the privacy association separation set, and the number of entries n. Different relationships determine different allocation strategies, such as the average allocation, the random allocation and the group allocation. The algorithm is described as follows:

Step 1: The encrypted sub-images are equally distributed to the data-entry clerks and the random sequence corresponding to the encrypted sub-image is sent to the data-entry clerks simultaneously.

Step 2: When the number of data-entry clerks n is greater than the number of attributes h in the set, each attribute can be equally distributed to n / h data-entry clerks; When the number of subsets f is less than or equal to the number of the data-entry clerks n, and n is smaller than the number of attributes h, the attributes of the subsets in the privacy association separation set may be respectively grouped. We assign the different groups to different data-entry clerks when the number of attributes of each group is less than or equal to h/n + 1;When the number of data-entry clerks n is smaller than the number of subsets f, a random allocation algorithm is selected. The process is shown in Figure 4.

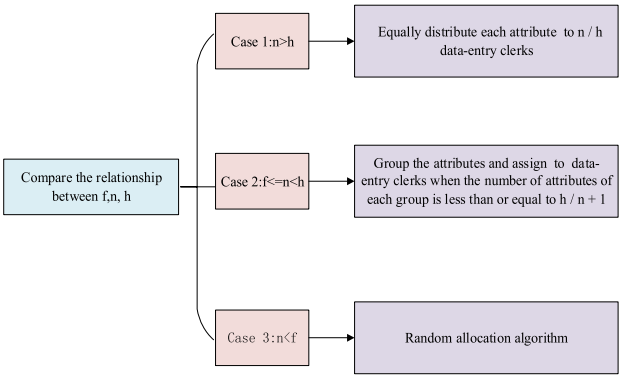


FIGURE 4. The process of step 2 in the sub-image allocation algorithm.

Step 3: Calculate the ratio between the number of sub-images to be allocated and the number of data-entry clerks to be assigned, expressed by avg. If the number of sub-images allocated to each clerk is greater than avg, the algorithm would subtract it from total amount and

recalculate the new avg. Repeat the above process until the number of sub-images is less than avg, then the loop ends.

The pseudo-code of the sub-image allocation algorithm is given in Algorithm 2.

Algorithm 2

Input: RE_j
Output: Fj (Sub-image allocation set)
(1) **for** j = 1 to **n**
(2) **do** Average allocation of encrypted sub-images **end**
(3) **if** n >= h **then**
(4) **for** k = 1 to (n - n % h) **do**
(5) Average allocation of the sub-images containing only one attribute to the n / h clerks
(6) **end**
(7) **else if** f <= n && n < h **then**
(8) $g = h / n + 1$
(9) **for** k=1 to f **do**
(10) **if** The number of attributes in current subset <=g
(11) **then** Assign attributes from current subset to a clerk and move to the next clerk
(12) **else** Assign the g attributes in current subset to a clerk and move to the next clerk
(13) **end if**
(14) **end**
(15) **else if** n < f **then** Random allocation **end if**
(16) Remove the sub-images that have been assigned from the sub-images to be assigned
(17) **while** The number of elements in the Fj >= avg **do**
(18) avg ← number of sub-images to be allocated / number of people to be assigned
(19) **if** The number of elements in the Fj > avg
(20) **then** The number of sub-images to be allocated ←f - the number of elements in the Fj
(21) Decrease one from the total number of data-entry clerks to be assigned
(22) **end if**
(23) **end**
(24) Get the data entry clerks who can continue to allocate the sub-images
(25) Image the clerk to be assigned in the previous step is at least avg.

2) SUB-CONTRACTOR SELECTION

The premise of the sub-image allocation algorithm is assigning tasks to existing data-entry clerks, but it does not involve the selection of the sub-contractors. The selection of the sub-contractors can make the allocation of the sub-images more reasonable and reduce the leakage of users' privacy. Moreover, it is conducive to maximize corporate interests.

AHP [24] is an effective method to select sub-contractors. It consists of four steps: building the hierarchical model, constructing the judgment matrix at each level, hierarchical single ranking and consistency checking, hierarchical total ranking and consistency checking.

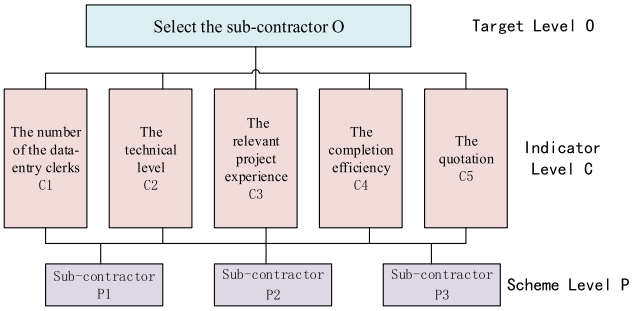


FIGURE 5. The hierarchical model for selecting the sub-contractor.

TABLE 1. The meanings of the scale of 1-9.

1-9 scale	Meaning
1	C_i and C_j are the same
3	C_i is slightly stronger than C_j
5	C_i is stronger than C_j
7	C_i is significantly stronger than C_j
9	C_i is definitely stronger than C_j
2,4,6,8	The influence ratio of C_i to C_j is between the above two adjacent levels
1,1/2,...1/9	The influence ratio of C_i to C_j is the reciprocal of the above values

(1) Building the hierarchical model

We select five indicators, the number of data-entry clerks, the technical level, the relevant project experience, the completion efficiency and the quotation, which are five key factors during choosing the sub-contractor. And the hierarchical model is established in Figure 5.

(2) Constructing the judgment matrix at each level Calculate C_i to represent the importance of the ith indicator at target level O. a_{ij} represents the ratio of C_i to C_j. And the formula of judgement matrix A is shown below.

$$A = (a_{ij})\, n \times n, \quad a_{ij} > 0, \quad aji = \frac{1}{a_{ij}} \tag{1}$$

The values in the matrix can be expressed in the scale of 1-9, and their meanings are shown in Table 1.

Obviously, in the judgment matrix A, a_{ij} and a_{ji} are reciprocal. After the C_5^2 operation, the judgment matrix A becomes

$$A = \begin{bmatrix} 1 & 4 & 4 & 3 & 3 \\ 1/4 & 1 & 1/3 & 1/2 & 1/3 \\ 1/4 & 3 & 1 & 1/2 & 1/3 \\ 1/3 & 2 & 2 & 1 & 1/2 \\ 1/3 & 3 & 3 & 2 & 1 \end{bmatrix} \tag{2}$$

Calculate the importance of each indicator at layer C in the same way. The final results are as follows.

$$C1 = \begin{bmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{bmatrix} \quad C2 = \begin{bmatrix} 1 & 1/3 & 1/2 \\ 3 & 1 & 2 \\ 2 & 1/2 & 1 \end{bmatrix}$$
$$C3 = \begin{bmatrix} 1 & 2 & 3 \\ 1/2 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{bmatrix} \quad C4 = \begin{bmatrix} 1 & 1/2 & 3 \\ 2 & 1 & 5 \\ 1/3 & 1/5 & 1 \end{bmatrix}$$
$$C5 = \begin{bmatrix} 1 & 1/3 & 1/5 \\ 3 & 1 & 1/3 \\ 5 & 3 & 1 \end{bmatrix} \tag{3}$$

(3) Hierarchical single ranking and consistency checking In hierarchical single ranking, we compute the largest eigenvalue λ_{max} and its respective eigenvector W of matrix A. We normalize the W as the relative ranking weight of the importance of the element in the level compared to the corresponding element in the previous level [25]. This method could reflect difference of relative influence on one indicator. However, it is easy to cause inconsistency between relative influences if all indicators are combined. So consistency checking is required which could refine the judgement matrix A. Furthermore, the relative ranking weight obtained by the judgment matrix A can be accepted and the inconsistency of the judgment matrix A can be limited to a certain range. Saaty defines CI as the consistency indicator and introduces the random consistency indicator RI to determine the allowable range of inconsistency of the judgment matrix A.

$$CI = \frac{\lambda - n}{n - 1} \tag{4}$$

The value of the random consistency indicator RI is related to the order n of the judgment matrix A, as shown in Table 2.

TABLE 2. The value of the random consistency indicator *RI*.

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45

The consistency ratio CR is defined as the ratio of the consistency indicator CI of judgment matrix A to the random consistency indicator RI of the same order. When the consistency ratio is less than 0.1, the judgment matrix is considered to pass the consistency checking. Feature vector is then could be used as the weight vector. Otherwise, the judgment matrix needs to be readjusted until it passes the consistency checking. In the example, $CR = 0.054$ which is less than 0.1 matches the requirement of consistency.

(4) Hierarchical total ranking and consistency checking The results of the hierarchical total ranking and consistency checking are shown in Table 3.

Results in Table 3 imply that the total order of P_1, P_2, and P_3 is 0.4254, 0.3179, and 0.2567,respectively.So the priority order of selecting the sub-contractors is: P_1, P_2, P_3. We tentatively put forward that the AHP is suitable for the sub-contractors selection. The five factors such as the number of the data-entry clerks, the technical level, the relevant project experience, the completion efficiency and the quotation may not be comprehensive enough, but it is quite useful.

TABLE 3. The results of the hierarchical total ranking and consistency checking.

	C_1 0.440	C_2 0.070	C_3 0.108	C_4 0.150	C_5 0.233	TOTAL ORDER	CR
P_1	0.648	0.163	0.540	0.309	0.105	0.425	
P_2	0.230	0.540	0.297	0.582	0.254	0.318	0.01
P_3	0.122	0.297	0.163	0.110	0.637	0.257	

III. PRIVACY PROTECTION EXPERIMENT

The experiment was performed on data from an outsourcing dataset, OutsourcingData and a 1994 US Census Adult Dataset [26]. This dataset covers a wider range of attributes which could make our results more convincing. Moreover, data from this dataset contains many sensitive attributes like salary which could test our algorithm's performance of privacy protection.

There are three types of the privacy leakage during the process of data entry in the outsourcing service, attribute leakage, the member leakage, and the identity leakage respectively [27]. We execute specific tests on the attribute privacy protection rate and the member privacy protection rate.

A. ATTRIBUTE PRIVACY PROTECTION RATE

The attribute privacy protection rate refers to the ratio of the number of sensitive attributes that cannot identify an individual to the total number of sensitive attributes.

The experimental results about the attribute privacy protection rate on OutsourcingData and Adult are shown in Figure 6(a) and Figure 6(b). It can be seen that the attribute privacy protection rate is 100%, no matter with the dataset and the number of entries. This is because the attribute privacy protection rate reflects the impact of sensitive attributes on users' privacy. However, PPM has no sensitive attributes after image segmentation and privacy association separation.

In addition, in Figure 6, the method proposed in this paper (referred to as PPM) is compared with the random allocation method and the data blocking method [28] on OutsourcingData and Adult. It can be seen that as the increase of the data-entry clerks, the attribute privacy protection rate of the random allocation method and the data blocking method improves as well. The reason is that the more clerks, the harder it is for individuals to get sensitive information. However, PPM is obviously superior to the other two methods. Even with a small number of the data-entry clerks,

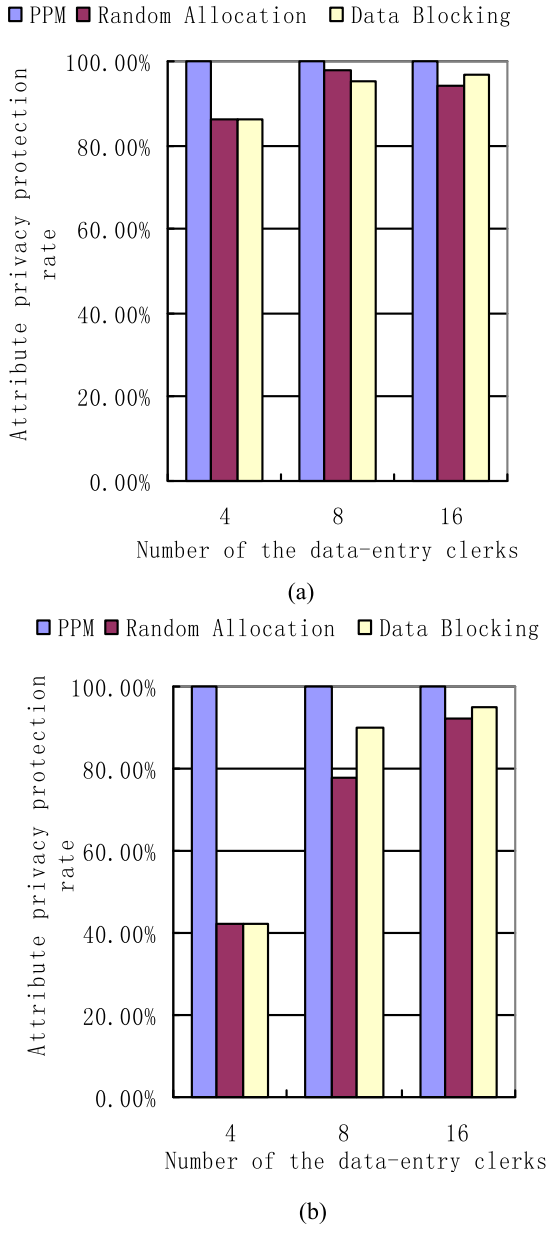


FIGURE 6. (a) Attribute privacy protection rate of three methods on outsourcingData. (b) Attribute privacy protection rate of three methods on adult.

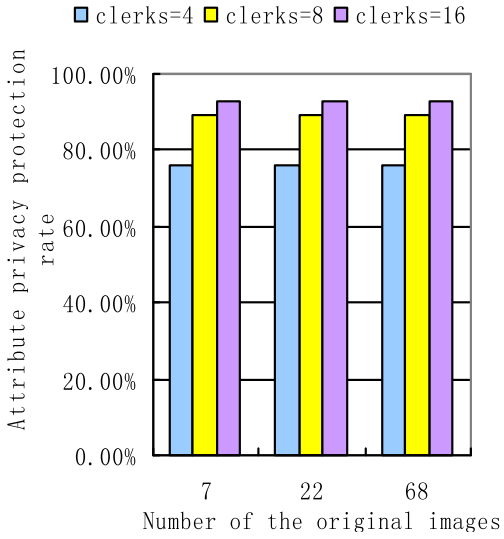


FIGURE 7. Change of the member privacy protection rate.

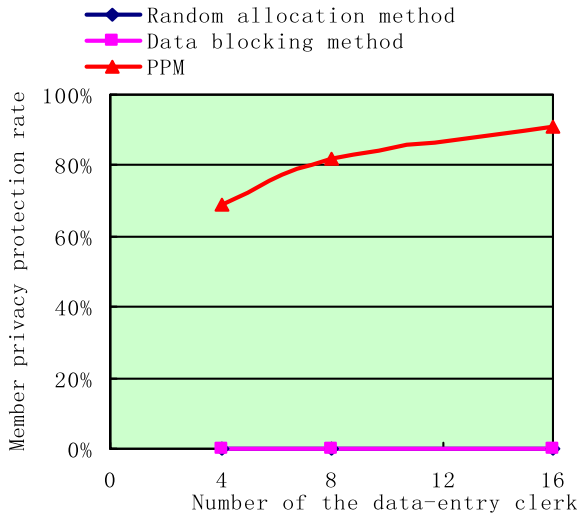


FIGURE 8. Changes in the member privacy protection rate of the three methods.

this method can also ensure the attribute privacy protection rate reaches 100%.

B. MEMBER PRIVACY PROTECTION RATE

The member privacy protection rate refers to the ratio of the number of individuals that cannot be identified to the total number of individuals. The member privacy protection rate on Outsourcing and Adult is same and shown in Figure 7. It can be seen that the member privacy protection rate has no change as the number of original images but increases gradually as the increase of the number of data-entry clerks. This is because we perform image segmentation on the original image to get a sub-images set. The number of sub-images is unpredictable and is generated intelligently by the EAST. Therefore, the number of original images has no effect on the member privacy protection rate.

We compare PPM with the random allocation method and the data blocking method. The changes of the member privacy protection rate in different methods are shown in Figure 8.

PPM has a better performance on privacy protection. Furthermore, it can increase the member privacy protection rate to more than 70%. This is because member privacy protection rate is concerned with the possibility of identifying individuals. PPM performs identifier encryption and privacy association separation to ensure the member privacy protection rate. However, both the random allocation method and the data blocking method do not take into account the individual identification. Therefore, there is no member privacy protection for these two methods.

The experimental results show that the proposed method has a good performance on the protection of attribute privacy and member privacy, which can reduce the probability of the leakage of users' privacy effectively. PPM has a big advantage over the random allocation method and the data blocking method even with inadequate clerks. Therefore, the method proposed in this paper is safe and practical.

IV. CONCLUSION

The purpose of this paper is to solve the problem of the leakage of user's privacy during the process of manual image entry when privacy associations and images with complete user's information are accessible to clerks. This method embeds the image segmentation, the privacy association separation and the sub-image allocation into the process of the image entry. We consider each essential during the process of image entry and propose methods to prevent privacy leakage accordingly. Results indicate that the method is practical and safe. It does increase the security and the availability of image data entry. However, we have to point out that the privacy association separation is not intelligent and efficient. Thus, in the following work, we are going to utilize deep leaning to improve the privacy association separation.

REFERENCES

[1] K. W. Hamlen and B. Thuraisingham, "Data security services, solutions and standards for outsourcing," *Comput. Standards Interfaces*, vol. 35, no. 1, pp. 1–5, 2013.

[2] J. C. Corena and T. Ohtsuki, "Secure and fast aggregation of financial data in cloud-based expense tracking applications," *J. Netw. Syst. Manage.*, vol. 20, no. 4, pp. 534–560, 2012.

[3] J. Zhou, Z. Cao, X. Dong, and X. Lin, "TR-MABE: White-box traceable and revocable multi-authority attribute-based encryption and its applications to multi-level privacy-preserving e-healthcare cloud computing systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, Apr./May 2015, pp. 2398–2406. [Online]. Available: http://ieeexplore.ieee.org/stamp/stmp.jsp?tp=&arnumber=7218628&isnumber=7218353. doi: 10.1109/INFOCOM.2015.7218628.

[4] A. S. Sawant and D. G. Chougule, "Notice of retraction script independent text pre-processing and segmentation for OCR," in *Proc. Int. Conf. Elect., Electron., Signals, Commun. Optim. (EESCO)*, Visakhapatnam, India, Jan. 2015, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/stamp/stmp.jsp?tp=&arnumber=7253643&isnumber=7253613. doi: 10.1109/EESCO.2015.7253643.

[5] M. Geethanjali and S. Slochanal, "Optimal coordination of directional over current relays using evolutionary programming," *CAAI Trans. Intell. Syst.*, vol. 4, no. 6, pp. 549–560, 2009. doi: 10.3969/j.issn.1673-4785.2009.06.014.

[6] M. Sharma and R. B. Garg, "DES: The oldest symmetric block key encryption algorithm," in *Proc. Int. Conf. System Modeling Adv. Res. Trends (SMART)*, Moradabad, India, Nov. 2016, pp. 53–58. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7894489&isnumber=7894471. doi: 10.1109/SYSMART.2016.7894489.

[7] Z. Gengshui *et al.*, "Privacy preservation in database applications: A survey," *Chin. J. Comput.*, vol. 32, no. 5, pp. 847–861, 2009.

[8] Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k-nearest neighbor query over encrypted data in outsourced environments," in *Proc. IEEE 30th Int. Conf. Data Eng.*, Chicago, IL, USA, Mar./Apr. 2014, pp. 664–675. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6816690&isnumber=6816620. doi: 10.1109/ICDE.2014.6816690.

[9] Q. Cheng, "Research on security outsourcing of large scale matrix computing for cloud platform," M.S. thesis, Dept. Comput. Sci. Technol., Nanjing Univ Aeronaut. Astronaut., Nanjing, China, 2016.

[10] H. Lee, H.-J. Kim, J.-W. Chang, and H.-I. Kim, "Bitmap-based distributed index structure and encrypted query processing schemes for outsourcing mobile sensitive data," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, Feb. 2017, pp. 288–295.

[11] Z. Jun, "Research on key issues of security and privacy of outsourcing system," Ph.D. dissertation, Dept. Comput. Sci. Technol., Shanghai Jiao Tong Univ., Shanghai, China, 2015.

[12] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3071–3083, Jun. 2017. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7885075&isnumber=7932278. doi: 10.1109/TGRS.2017.2650938.

[13] F. Luo, H. Huang, Y. Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, p. 790, 2017.

[14] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7564440&isnumber=8207702. doi: 10.1109/TCYB.2016.2605044.

[15] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8351966&isnumber=6352949. doi: 10.1109/TCYB.2018.2810806.

[16] X. Zhou *et al.*, "EAST: An efficient and accurate scene text detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2642–2651. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8099766&isnumber=8099483. doi: 10.1109/CVPR.2017.283.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7485869&isnumber=7919342. doi: 10.1109/TPAMI.2016.2577031.

[18] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1480–1500, Jul. 2015. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6945320&isnumber=7116666. doi: 10.1109/TPAMI.2014.2366765.

[19] S. Uchida, "Text localization and recognition in images and video," in *Handbook of Document Image Processing and Recognition*. London, U.K.: Springer, 2014, pp. 843–883. doi: 10.1007/978-0-85729-859-1_28.

[20] Y. Zhu, C. Yao, and X. Bai, "Scene text detection and recognition: Recent advances and future trends," *Frontiers Comput. Sci.*, vol. 10, no. 1, pp. 19–36, 2016.

[21] K.-H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park. (2016). "PVANET: Deep but lightweight neural networks for real-time object detection." [Online]. Available: https://arxiv.org/abs/1608.08021

[22] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, 2015.

[23] K. He, X. Zhang, S. Ren, and J. Sun. (2015). "Deep residual learning for image recognition." [Online]. Available: https://arxiv.org/abs/1512.03385

[24] T. L. Saaty, "Modeling unstructured decision problems—The theory of analytical hierarchies," *Math. Comput. Simul.*, vol. 20, no. 3, pp. 147–158, 1978.

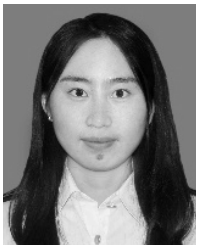
[25] Z. Yongzheng, Z. Tangsen, and F. Chenghong, *Mathematical Modeling*. Shanghai, China: Tongji Univ. Press, 2010.

[26] T. Li, N. Li, J. Zhang, and I. Molloy, "Slicing: A new approach for privacy preserving data publishing," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 561–574, Mar. 2012. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5645625&isnumber=6138960. doi: 10.1109/TKDE.2010.236.

[27] Z. Yu, "Transaction-based data privacy protection method based on slicing technology," M.S. thesis, Dept. Electron., Dalian Univ. Technol., Dalian, China, 2015.

[28] W. Yingwang, "Research on privacy protection mechanism based on relationship hiding," M.S. thesis, Dept. Electron., Beijing Inst. Technol., Beijing, China, 2015.

FENGQI LI received the Ph.D. degree from the Dalian University of Technology, in 2014. He is currently a Professorate Senior Engineer. He is also the Vice-President of the School of Software, Dalian University of Technology, and the Vice-President of the Dalian Software Industry Association. His main interests include blockchain, 3-D printing, big data, and enterprise information. More than 30 of his academic papers have been included in SCI or EI.



CHUNLI SHANG received the B.S. degree in information security from the Civil Aviation University of China, in 2017. She is currently pursuing the M.S. degree in software engineering with the School of Dalian University of Technology. Her main interests include block chain and deep learning.

KEMENG LIU received the B.S. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2018, respectively, where he is currently pursuing the M.S. degree in software engineering. He is currently involved in improving the PBFT algorithm to solve the centralization problem in the alliance chain. His main interest is blockchain technology, especially the consensus algorithm in the alliance chain.

AIDI PANG received the B.S. degree in medical information engineering from the Liaoning University of Traditional Chinese Medicine, in 2016, and the master's degree in software engineering from the Dalian University of Technology, in 2018. Her main interest includes intelligent medical treatment.

SIKAI HUANG received the B.S. degree in software engineering from the Dalian University of Technology, China, in 2014, and the M.Sci. degree in computer science from Columbia University, USA, in 2017. He is currently pursuing the Ph.D. degree with the National University of Singapore, Singapore.

From 2017 to 2018, he was an Assistant Research Scientist with the Research Foundation for Mental Hygiene. From 2018 to 2019, he was a Researcher with iFlytek's AI Research. His research interest includes machine learning, specifically probabilistic graphical model, Gaussian processes, deep learning for computer vision, and medical imaging.