ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Privacy Threats and Privacy Preservation Techniques for Farmer Data Collection based on Data Shuffling

Surapon Riyana[1], Nobutaka Ito[2], Tatsanee Chaiya[3], Uthaiwan Sriwichai[4], Natthawud Dussadee[5], Tanate Chaichana[6], Rittichai Assawarachan[7], Thongchai Maneechukate[8], Samerkhwan Tantikul[9] and Noppamas Riyana[10]

ABSTRACT

Aside from smart technologies, farm data collection is also important for smart farms including farm environment data collection and farmer survey data collection. With farm data collection, we observe that it is generally proposed to utilize in smart farm systems. However, it can also be released for use in the outside scope of the data collecting organization for an appropriate business reason such as improving the smart farm system, product quality, and customer service. Moreover, we can observe that the farmer survey data collection often includes sensitive data, the private data of farmers. Thus, it could lead to privacy violation issues when it is released. To address these issues in the farmer survey data collection, an anatomization model can protect the users' private data that is available in farmer survey data collection to be proposed. However, it still has disorganized issues and privacy violation issues in the sensitive table that must be addressed. To rid these vulnerabilities of anatomization models, a new privacy preservation model based on data shuffling is proposed in this work. Moreover, the proposed model is evaluated by conducting extensive experiments. The experimental results indicate that the proposed model is more efficient than the anatomization model for the farmer survey data collection. That is, the adversary can have the confidence for re-identifying every sensitive data that is available in farmer survey data collection that is after satisfied by the privacy preservation constraint of the proposed model to be at most $1/l$. Furthermore, after the farmer survey data collection satisfies the privacy preservation constraint of the proposed model, it does not have disorganized issues and privacy violation issues from considering the sensitive values.

Article information:

Keywords: Privacy Preservation, Privacy Violation Issues, Farmer Data Collection, Data Shuffling, Privacy Data Protection, Multiple Sensitive Attributes (MSA)

Article history:
Received: November 29, 2021
Revised: April 23, 2022
Accepted: June 4, 2022
Published: June 25, 2022 (Online)

DOI: 10.37936/ecti-cit.2022163.246469

1. INTRODUCTION

One of the important occupations in the world is agriculture. It is a food production source. However, only some countries can be engaged in agriculture. Moreover, some countries only have certain areas where agriculture can be possible, and some agricultural countries do not have enough efficient agricultural workers. Furthermore, we found that the quality of agricultural products is often different when they are produced from different agricultural areas. In some over-populated countries such as Ethiopia, Zambia, and Chad, two-thirds of the population faces the problem of not getting enough food. This problem is severe. It means those people do not have enough energy to lead an active life, and the children's growth and development could be harmed. In some countries in Asia (e.g., India, Bangladesh, and Pakistan), the populations also have the problem of not getting enough food. Although lives are getting better for most people in India, Bangladesh, and Pakistan, they still have the largest numbers of hungry people across the globe. To reduce the severity of these problems, agriculture experts, agriculture companies, and agriculture universities try to propose

[1,2,3,4,5,6,7,8,9,10]The authors are with Maejo University (MJU), Maejo, Sansai, Chiang Mai, Thailand 50290, E-mail: surapon_r@mju.ac.th, nobuito@mju.ac.th, tatsanee_cy@mju.ac.th, sriwichai@mju.ac.th, natthawud@mju.ac.th, tanate_c@mju.ac.th, rittichai@mju.ac.th, thongchai_m@mju.ac.th, samerkhwan@mju.ac.th and noppamas@mju.ac.th

Table 1: *An example of the farmer survey data collection.*

#	Quasi-identifier			Sensitive group 1	Sensitive group 2		
	Blood	**Gender**	**Age**	**Income**	**Chlorpyrifos**	**Grammoxone**	**Roundup**
d_1	O	F	53	$10,000.00	5	3	1
d_2	B	M	44	$15,000.00	2	3	2
d_3	O	F	50	$13,000.00	2	2	2
d_4	A	M	46	$14,000.00	1	3	3

agricultural concepts and innovations. Smart farms are outstanding agricultural concepts that are proposed. They based on agricultural information systems and intelligent agriculture technologies.

Agricultural information systems and intelligent technologies generally use the agricultural data collection that include farm environment information and farmers' information. A well-known agricultural data collection is proposed to collect the farmers' information. It is the farmer survey data collection [1], as shown in Table 1. This table is constructed from four farmer profile tuples. Every tuple consists of three quasi-identifier attributes (i.e., blood, gender, and age) and two sensitive groups (i.e., sensitive groups 1 and 2). Sensitive group 1 only includes the income attribute. Another sensitive group, sensitive group 2, consists of three sensitive attributes chlorpyrifos, grammoxone, and roundup. For example, the tuple d_1 of Table 1 represents the profile of a farmer who is a female person that is 53 years old, her blood type is O, and her income is $10,000.00. Moreover, she has the statistic of using chlorpyrifos five times, grammoxone three times, and roundup one time respectively. In Table 1, although all tuples do not have any explicit identifier of farmers such as name, surname, SSN, and Citizen ID, they still have privacy violation issues that must be addressed when they are released for use in the outside scope of the data collecting organization.

Example 1 (Privacy violation issues in the released farmer survey data collection): We suppose that Alice is the target user of the adversary. Moreover, we assume that the adversary knows that Alice is a female person who is 53 years old. The adversary highly believes one of the tuples that are available in Table 1 to be Alice's profile tuple. For this situation, the adversary can infer that the tuple d_1 must be Alice's profile tuple because only this tuple of Table 1 can match the adversary's background knowledge about Alice. With the tuple d_1, the adversary can see that Alice frequently uses chlorpyrifos and gramoxone. Thus, the adversary can further infer that Alice is a farmer who has a chance to be getting "skin cancers" and "breast cancers" because chlorpyrifos and grammoxone are produced from chlorpyrifos [2, 3] and paraquat [4, 5] respectively. Moreover, the adversary can further know that Alice's income is collected in Table 1 to be $10,000.00.

With Example 1, we can see that although the released farmer survey data collection does not include any explicit identifier of farmers, it still may have privacy violation issues that must be addressed. To address these issues, in [1], the authors propose an anatomization model for the released farmer survey data collection. It will be explained in Section 1.1.

1.1 The anatomization for releasing the farmer survey data collection

To address privacy violation issues in the released farmer survey data collection, in [1], the authors propose an anatomization model. It can ensure the privacy data in the released dataset that has multiple sensitive attributes (MSA) [11-13] to be protected. That is, let a positive integer l, where $l \geq 2$, be the privacy preservation constraint. Let $QI = \{qi_1, qi_2, \ldots, qi_n\}$ be the set of quasi-identifier attributes. Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of sensitive attributes. Let $SG = \{sg_1, sg_2, \ldots, sg_g\}$, where $sg_1, sg_2, \ldots, sg_g \subseteq S$, $sg_1 \cup sg_2 \cup \ldots \cup sg_g = S$ and $sg_1 \cap sg_2 \cap \ldots \cap sg_g = \varnothing$, be the set of sensitive attribute groups. Let $D = \{d_1, d_2, \ldots, d_x\}$ be the specified farmer survey data collection. Every $d_i \in D$ is in the form of $\{qi_1, qi_2, \ldots, qi_n, s_1, s_2, \ldots, s_m\}$. For privacy preservation, the tuples of D are partitioned such that every sensitive attribute $s_y \in S$ of each partition must consist of l different sensitive values. Furthermore, every partition of D is anatomized to be the tables $D_{QI}, D_{sg_1}, D_{sg_2}, \ldots,$ and D_{sg_g} such that each partition of these anatomized tables is related by its defined partition identifier, PID.

Example 2 (Privacy preservation in the released farmer survey data collection based on anatomization constraints): Let Table 1 be the farmer survey data collection D. Let the attributes blood, gender, and age be the quasi-identifier attributes. Let the attributes income, chlorpyrifos, grammoxone, and roundup be the sensitive attributes. Moreover, the sensitive attributes are grouped into two groups, i.e., sensitive group 1 and 2. That is, sensitive group 1 only includes the income attribute. Sensitive group 2 consists of three sensitive attributes chlorpyrifos, grammoxone, and roundup. The value of l is set to be 2. Therefore, an anatomization data version of Table 1 is shown in Tables 2, 3, and 4. Every partition of these anatomization tables is related by its defined partition identifiers, PID. For this situation, if the

adversary tries to re-identify the sensitive values of the target user from these anatomized tables, he/she observes that all possibly re-identified conditions always have at least two different sensitive values to be satisfied.

Table 2: *The quasi-identifier table of Table 1.*

Blood	Gender	Age	PID
O	F	53	1
B	M	44	1
O	F	50	2
A	M	46	2

Table 3: *The table of sensitive group 1.*

Income	PID
$10,000.00	1
$15,000.00	1
$13,000.00	2
$14,000.00	2

Table 4: *The table of sensitive group 2.*

Chlorpyrifos	Grammoxone	Roundup	PID
5	3	1	1
2	3	2	1
2	2	2	2
1	3	3	2

As demonstrated in Example 2, it is so clear that the released farmer survey data collection is processed by using anatomization constraints, it can be more secure in terms of privacy preservation than its original version. However, to the best of our knowledge about the anatomization model for the released farmer survey data collection, it still has the serious vulnerabilities that should be improved such as disorganized issues and privacy violation issues. Therefore, a new privacy preservation model for the farmer survey data collection is proposed in this work.

The organization of this paper is as follows. In Section 2, the motivation is presented. The related works are reviewed in Section 3. Then, the proposed model will be presented in Section 4. In Section 5, the experimental results are discussed. Section 6 gives the conclusion of this work. Finally, Section 7, it is devoted to discussing the future work.

2. MOTIVATION

This section is devoted to identifying the significant vulnerabilities of the anatomization model that is proposed to protect the privacy data in the released farmer survey data collection [1].

2.1 Disorganized issues

Suppose that Table 1 without sensitive group 2 is the farmer survey data collection D_1. The value of l is set to be 2. For privacy preservation, the tuples of D_1 are partitioned by using the given value of l and further anatomized to be the quasi-identifier table and the sensitive table such that every partition of the anatomized tables is related by its defined identifier, PID. Thus, a version of the anatomized tables of D_1 satisfies $l = 2$, it is shown in Tables 2 and 3. With an example of transforming the farmer survey data collection to the satisfaction of anatomization constraints, we suppose that Table 1 is the farmer survey data collection D_2. The value of l is set to be 2. Thus, a version of the anatomized tables of D_2 satisfies l=2, it is shown in Tables 2, 3, and 4. With the anatomized tables of D_1 and D_2, we can conclude that if the farmer survey data collection satisfies anatomization constraints, it always has $g + 1$ anatomized tables, where g is the number of its sensitive groups. Or we can say that when the number of the sensitive groups of the farmer survey data collection is increased, it leads to disorganized issues because there are many anatomized tables that must be considered when they are utilized.

2.2 Privacy violation issues in sensitive tables

Suppose that Table 4 is the sensitive table of the farmer survey data collection such that it is after satisfied by anatomization constraints. We assume that Alice is the target user of the adversary. Moreover, the adversary knows that Alice just used “roundup” once. Furthermore, the adversary highly believes that one of the tuples is available in Table 4, it is Alice’s sensitive tuple. For this situation, the adversary can be highly confident that the first tuple of Table 4 is Alice’s sensitive tuple because only this sensitive tuple can match the adversary’s background knowledge about Alice. Thus, the adversary can know that Alice is a farmer who frequently uses chlorpyrifos and gramoxone.

In this section, it is clear that although the released farmer survey data collection satisfies anatomization constraints, it still has privacy violation issues that must be addressed. To rid these vulnerabilities, a new privacy preservation model for releasing the farmer survey data collection is proposed in this work such that it does not have disorganized issues and privacy violation issues from considering sensitive values that are available in the released farmer survey data collection. It will be presented in Section 4.

3. RELATED WORK

One of the serious issues that must be considered when datasets are released for public use, it is privacy violation issues. To address these issues, a well-known privacy preservation model, k-Anonymity

[6], is proposed. The privacy preservation idea of k-Anonymity is that before datasets are released for public use, the attributes of datasets are first grouped into three groups. The first group has the explicit identifier attributes such as SSN, citizen ID, student code, name, and surname. The second group has the quasi-identifier attributes. The quasi-identifier attributes are the set of the arbitrary attributes, such as BoD, gender, zip code, blood group, and education, which are available in datasets such that their combinations can use to identify the owner of the target sensitive value that is also available in datasets. The final group has a sensitive attribute such as the salary, lawsuit, disease, or so on. Then, all values are available in the explicit identifier attributes to be removed. Finally, all unique quasi-identifier values are distorted by using the data suppression or generalization to be k indistinguishable times. In addition, every group of k indistinguishable quasi-identifier values forms an equivalence class of datasets. Thus, after datasets are satisfied by k-Anonymity constraints, they can guarantee that all possible re-identification conditions of them always have at least k tuples to be satisfied. For this reason, they do not seem to leave any concern about privacy violation issues. Unfortunately, in [7], the authors demonstrate that even when datasets are satisfied by k-Anonymity constraints, they still have privacy violation issues that must be addressed. If the adversary can specify the equivalence class that collects the profile tuple of the target user and the sensitive values of the specified equivalence class is not diverse, the sensitive value of the target user can be violated by the adversary. To address this vulnerability of k-Anonymity, l-Diversity [7] is proposed.

l-Diversity [7] is a well-known privacy preservation model that is extended from k-Anonymity. The privacy preservation idea of l-Diversity is that aside from removing the explicit identifier values and distorting the unique quasi-identifier values, the number of distinct sensitive values is also considered in l-Diversity constraints. Every indistinguishable quasi-identifier value must relate to at least l different sensitive values. Thus, after datasets are satisfied by l-Diversity constraints, they can guarantee that all possible re-identification conditions always have at least l different sensitive values to be satisfied. For this reason, when datasets are satisfied by l-Diversity constraints, they are highly secure in terms of privacy preservation than the original dataset of them. However, in [8], the authors demonstrate that privacy preservation models based on data suppression and data generalization, they often have fake query conditions. To rid this vulnerability, a privacy preservation model, Anatomy [8], based on data anatomizations to be proposed.

Anatomy [8] was proposed by X.Xiao et al. in 2006. With this privacy preservation model, datasets cannot have the concern of privacy violation issues from the adversary by following the special steps. At first, the tuples of datasets are partitioned such that every partition must collect the sensitive values to be at least l different values. Then, the identifier for each partition is defined. Finally, the tuples of datasets are anatomized to be the quasi-identifier and the sensitive table such that each partition of the anatomized tables is related by its defined identifier. For this reason, anatomized tables can also guarantee that all possibly re-identified conditions always have at least l different sensitive values to be satisfied. Moreover, they do not have any fake query conditions. However, to the best of our knowledge about Anatomy, it can preserve the privacy data in datasets that only have a single sensitive attribute. Thus, Anatomy could be insufficient to address privacy violation issues in farmer survey data collection because it generally has multiple sensitive attributes. To rid this vulnerability of Anatomy in the farmer survey data collection, in [1], the authors propose a privacy preservation model that is extended from Anatomy, so-called MSA anatomization for the farmer survey data collection.

To address privacy violation issues in the farmer survey data collection, the farmer survey dataset, that based on MSA anatomization constraints, before the farmer survey dataset is released for public use, its attributes are first grouped to be two major groups as the group of quasi-identifier attributes and the group of sensitive attributes. Moreover, the attributes which are available in the group of sensitive attributes, they can further be separated to be the sub-groups. Then, the tuples of the farmer survey dataset are partitioned such that each sensitive attribute of each partition collects the sensitive values so that there are at least l different values such that each sensitive group is independently partitioned. Furthermore, the identifier of each partition is also defined by this step. Finally, the tuples of the farmer survey dataset are anatomized such that each partition of the anatomized tables is related by its defined identifier. The privacy preservation idea of MSA anatomization is more explained in Section 1.1. Also, after anatomized tables are satisfied by MSA anatomization constraints, they can guarantee that all possibly re-identified conditions always have at least l different sensitive values to be satisfied. Moreover, the anatomized tables do not have any fake query conditions. However, we discover that MSA anatomization still has both serious vulnerabilities that must be improved. It often leads to disorganized issues when the number of sensitive groups in the farmer survey dataset is increased. Moreover, we can see that if the adversary has adequate background knowledge about the target user in sensitive tables, the sensitive values of the target user can be revealed by the adversary. To address these vulnerabilities of MSA anatomization, a new privacy preservation model for the farmer survey dataset is proposed in this work, it will be pre-

Table 5: *A released data version D' of Table 1 is satisfied by $l = 2$.*

Blood	Gender	Age	Income	Chlorpyrifos	Grammoxone	Roundup	Partition
O	F	53	$15,000.00	2	3	2	1
B	M	44	$10,000.00	5	3	1	1
O	F	50	$13,000.00	1	3	2	2
A	M	46	$14,000.00	2	2	3	2

Table 6: *A released data version D' of Table 1 is also satisfied by $l = 2$.*

Blood	Gender	Age	Income	Chlorpyrifos	Grammoxone	Roundup	Partition
O	F	53	$10,000.00	5	2	2	1
O	F	50	$13,000.00	2	3	1	1
B	M	44	$14,000.00	1	3	3	2
A	M	46	$15,000.00	2	3	2	2

sented in Section 4. Aside from the above-mentioned data distortion techniques, data shuffling [9] is also a well-known data distortion technique that is often used in privacy preservation models. Moreover, data shuffling is also applied in the proposed privacy preservation model. For this reason, before the proposed privacy preservation model is presented, we would like to first present a privacy preservation model, (k, e)-Anonymous [9], that based on data shuffling. The privacy preservation idea of (k, e)-Anonymous is that before datasets are released for public use, the attributes are also grouped to be two groups as the quasi-identifier attributes and a sensitive attribute. Then, the tuples of datasets are resorted by the sensitive values in either descending or ascending order. Subsequently, the tuples of datasets are partitioned by the given value of k and the given value of e. That is, every partition of datasets must include at least k tuples, and the difference value between the lower bound and the upper bound of the sensitive values that are available in each partition of datasets must be at least e. Finally, the sensitive values or the quasi-identifier values of each partition are shuffled. For this reason, after datasets satisfy (k, e)-Anonymous constraints, they guarantee that all possibly re-identified conditions always have at least k tuples to be satisfied, and the confidence of re-identifying every sensitive value in datasets is at most the given value of e. Clearly, (k, e)-Anonymous is a privacy preservation model that can address privacy violation issues in released datasets. However, (k, e)-Anonymous is only sufficient to address privacy violation issues in datasets that only have a single sensitive attribute, and the data domain of the sensitive attribute is numerical. For this reason, we can say that (k, e)-Anonymous is also insufficient to address privacy violation issues in the farmer survey dataset.

4. THE PROPOSED MODEL

In Section 2, it was clear that although the farmer survey data collection satisfies anatomization constraints, they still have disorganized issues and privacy violation issues in the sensitive table that must be addressed. To rid the vulnerabilities of the anatomized model, a new privacy preservation model for releasing the farmer survey data collection is proposed in this section. With the proposed model, the extraction of private data in the farmer survey data collection is prevented by using data shuffling.

4.1 Privacy preservation principles

Let a positive integer l, where $l \geq 2$, be the privacy preservation constraint. Let $QI = \{qi_1, qi_2, \ldots, qi_n\}$ be the set of quasi-identifier attributes. Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of sensitive attributes. Let $D = \{d_1, d_2, \ldots, d_x\}$ be the farmer survey data collection such that every $d_i \in D$ is presented in the form of $\{qi_1, qi_2, \ldots, qi_n, s_1, s_2, \ldots, s_m\}$. Let $f_{PAR}(D) : D \rightarrow_l D'$ be the function for transforming D to become D'. That is, D' is constructed from $\{par_1, par_2, \ldots, par_p\}$, where $\cup_{q=1}^{p} par_q = D$ and $\cap_{q=1}^{p} par_q = \varnothing$. Moreover, every sensitive attribute $s_y \in S$ of each partition $par_q \in D'$, must collect the sensitive values to be at least l different values. Let $f_{SHF}(par_q[s_y]) : par_q[s_y] \rightarrow par_q[s_y]'$ be the function for shuffling the sensitive values which are available in the sensitive attribute s_y of the partition $par_q \in D'$. Therefore, the released version D' of D does not have any privacy violation issues from using the adversary's background knowledge about the target user. It is constructed from $\cup_{q=1}^{p}(par_q[QI] \cup f_{SHF}(par_q[s_1] \cup f_{SHF}(par_q[s_2] \cup \ldots \cup f_{SHF}(par_q[s_m]))$.

For example, let Table 1 be the farmer survey data collection D. Let the value of l be set at 2. Thus, a released version D' of Table 1 is shown in Table 5. With this table, we can see that every sensitive attribute of each partition always includes the sensitive values to be at least l different values. Therefore, the adversary cannot extract the sensitive value of the target user such that it is available in Table 5 because every possibly re-identified always has at least

l different sensitive values to be satisfied.

From this example, it is clear that after the farmer survey data collection satisfies the proposed privacy preservation constraint, they can be more secure in terms of privacy preservation than their original versions. Moreover, they do not have any disorganized issues when the number of sensitive groups is increased. However, each farmer survey data collection D and each value of l have variously the released versions that can be satisfied. For example, aside from Table 5, Table 6 is also a released version of Table 1 such that it also satisfies $l = 2$. Thus, only the released version of Table 1 has highly the data utility to be the desired version. For this reason, the data utility metric is a necessary importance of the proposed privacy preservation model, it will be presented in Section 4.2.

4.2 Data utility metric

With the proposed privacy preservation model, the data utility of the released farmer survey data collection can be used by defining the specified query data condition via the quasi-identifier attributes to get the target sensitive values that are collected in the specified sensitive attribute. For this reason, the size of partitions and the number of the different quasi-identifier values reasonably influence the data utility of the released farmer survey data collection. That is, the smallest partition size and fewest different quasi-identifier values lead to more data utility of the released farmer survey data collection. Therefore, the data utility of the released farmer survey data collection can be defined by Equation 2. With Equation 2, the penalty cost of the released farmer survey data collection is in the range between 0 and 1. The released data version with a penalty cost nearest to zero is the desired data version.

$$PartitionLoss(par_q) = \frac{(|par_q[qi_1]| + |par_q[qi_2]| + \ldots + |par_q[qi_n]|)}{|QI| * |par_q|} \quad (1)$$

$$DatasetLoss(par_q) = \frac{\sum_{q=1}^{p} PartitionLoss(par_q)}{p}, \text{where} |par_q| \geq 2 \quad (2)$$

where

- par_q is the specified partition of D' and $|par_q|$ is the number of tuples which are available in par_q.
- $|par_q[qi_1]|, |par_q[qi_2]|, \ldots,$ and $|par_q[qi_n]|$ is the number of the distinct quasi-identifier values which are available in the quasi-identifier attributes as $qi_1, qi_2, \ldots,$ and qi_n respectively.
- $|QI|$ is the number of quasi-identifier attributes of D'.
- p is the number of partitions which are available in D'.

For example, Table 5 has a penalty cost to be 1, i.e., (1 + 1) / 2 = 1. With Table 6 has a penalty cost to be 0.75, i.e., (0.67 + 0.83) / 2 = 0.75. For this situation, it means that Table 6 has better data utility than Table 5. Thus, Table 6 is the more desirable release data version of Table 1.

4.3 The proposed anonymization algorithms

This section is devoted to presenting a greedy privacy preservation algorithm based on the proposed privacy preservation constraint that explained in Section 4.1. The inputs of the proposed algorithm are a farmer survey data collection D and a positive integer l. Its output is an anonymized farmer survey data collection D' that does not have any privacy violation issues and which is immune to attacks using the adversary's background knowledge about the target user.

The proposed algorithm processes are separated into five parts. The first part is shown in the lines between 6 and 8. It investigates the number of sensitive values which are available in each sensitive attribute. If an arbitrary sensitive attribute of D collects the sensitive values to be at most $l - 1$ different values, the algorithm returns failure because these particular D cannot be transformed to satisfy the given value of l. The second part is shown in the lines between 9 and 12. It also investigates the number of sensitive values which are available in each sensitive attribute. If an arbitrary sensitive attribute of D just collects the sensitive values to be at most $l * 2$ different values, the tuples of D can only fit for constructing an anonymized partition of D'. The third part is shown in the lines between 13 and 26. All possible partitions of D' can be constructed by the steps as follows. At first, an arbitrary tuple d_{sel} of D is defined to be the initial tuple of the constructed anonymized partition of D'. Then, d_{sel} is removed from D. Subsequently, all possible anonymized partitions, $PAR_{(d_{sel})}$, that consist of d_{sel} and can satisfy the given value of l are generated by using $GenAllPartitions(D, d_{sel}, l)$. After that, an arbitrary partition $par_{sel} \in PAR_{(}d_{sel})$ has the penalty cost of PartitionLoss(par_{sel}) to be minimized and its size is maximized. It is chosen to be an anonymized partition of D'. Moreover, all tuples in par_{sel}, are removed from D. Finally, the algorithm investigates the remaining sensitive values which are available in each sensitive attribute. If the remaining sensitive values can still satisfy the given value of l, the algorithm runs this part again. The fourth part is shown in the lines between 27 and 36. It determines the appropriately anonymized partition of the remaining tuples of D such that these tuples cannot be assigned into any anonymized partition of

Algorithm 1: Anonymization algorithm for releasing the farmer survey data collection

Input: The original farmer dataset D and a positive integer l.
Output: An anonymized farmer dataset D' of D.
Let $|D|$ be the number of the tuples which are available in D.
Let $D[QI]$ be the quasi-identifier attributes of D.
Let $|D[s_1]|, |D[s_1]|, \ldots, |D[s_m]|$ represent the number of the distinct sensitive values that are available in the sensitive attributes as $s_1, s_2, \ldots, s_m$ respectively.
If $|D[s_1]| < l$ or $|D[s_1]| < l$ *or* ... *or* $|D[s_m]| < l$ then
    Return failure;
End if
If $|D[s_1]| < l * 2$ or $|D[s_1]| < l * 2$ *or* ... *or* $|D[s_m]| < l * 2$ then
    $D' := D[QI] \cup f_{SHF}(D[s_1]) \cup f_{SHF}(D[s_2]) \cup \ldots \cup f_{SHF}(D[s_m])$;
    Return D';
End if
While all possible tuples of D do
    $d_{sel} :=$ an arbitrary tuple of D;
    $D := D - d_{sel}$;
    $PAR_{d_{sel}} := GenAllPartitions(D, d_{sel}, l)$;
    While all possible partitions of $PAR_{d_{sel}}$ do
        $par_{sel} := PartitionLoss(par_{sel})$, where $par_{sel} \in PAR_{d_{sel}}$, is minimized and the size of par_{sel} is maximized;
        $par_{tmp} := par_{sel}[QI] \cup f_{SHF}(par_{sel}[s_1]) \cup f_{SHF}(par_{sel}[s_2]) \cup \ldots \cup f_{SHF}(par_{sel}[s_m])$;
        $D' := D' \cup par_{tmp}$;
    End while
    $D := D - par_{sel}$;
    If $|D[s_1]| < l$ or $|D[s_1]| < l$ *or* ... *or* $|D[s_m]| < l$ then
        Break;
    End if
End while
While all possible partitions of D' do
    While all possible tuples of D do
        $d_{rem} :=$ an arbitrary tuple of D;
        $D := D - d_{rem}$;
        $par_{D'} :=$ an arbitrary partition of D';
        $par_{buf} := PartitionLoss(par_{D'} \cup d_{rem})$ is minimized;
        $par_{tmp} := par_{buf}[QI] \cup f_{SHF}(par_{buf}[s_1]) \cup f_{SHF}(par_{buf}[s_2]) \cup \ldots \cup f_{SHF}(par_{buf}[s_m])$;
        $D' := (D' - par_{sel}) \cup par_{tmp}$;
    End while
End while
Return D';

D' by using the previous part processes. For determining the appropriately anonymized partition, the similar score between each existing anonymized partition $par_{D'}$ of D' and each remaining tuple d_{rem} of D is evaluated by using $PartitionLoss(par_{D'} \cup d_{rem})$. If the penalty cost of $PartitionLoss(par_{D'} \cup d_{rem})$ is minimized, $par_{D'}$ is removed from D' and $par_{D'} \cup d_{rem}$ is constructed to be the new anonymized partition of D'. In the final part, D' is returned.

5. EXPERIMENT

In this section, the effectiveness and efficiency of the proposed model was evaluated by both comparative models as MSA anatomization [1] and MSA l-Diversity [7].

5.1 EXPERIMENT

All experiments were proposed to evaluate the effectiveness and efficiency of the proposed model. They were conducted on Intel(R) Xeon(R) Gold 5218 @2.30 GHz CPUs with 64 GB memory and six 900 GB HDDs with RAID-5. Furthermore, they were built and executed on Microsoft Visual Studio 2019 Community Edition in conjunction with MSSQL Server 2019 and based on "Adult Dataset [10]". This dataset was constructed from about 48,843 user tuples such that every user tuple consists of six continuous attributes (i.e., age, fnlwgt, education-num, capital-gain, capital loss, and hours-per-week) and eight nominal attributes (i.e., workclass, education, marital status, occupation, relationship, race, sex, and native country). To conduct the experiments effectively, the fnlwgt, education-num, and capital gain attributes are removed. Moreover, all user tuples with the values of "0" and "?" are also removed. The experimental dataset only contains 2140 user tuples such that every user tuple consists of the attributes as education, age, sex, native country, race, relationship, marital status, capital loss, hours-per-week, occupa-

tion, and workclass. The attributes as education, age, sex, native country, and race were the quasi-identifier attributes, and other remaining attributes (relationship, marital status, capital loss, hours-per-week, occupation, and workclass) of the experimental dataset were the sensitive attributes. Furthermore, the sensitive attributes were grouped into three groups. The first group consists of both sensitive attributes as capital loss and hours-per-week. The second group contains the relationship and marital status attributes. The occupation and workclass attributes were collected in the third sensitive group. All experimental datasets were evaluated by "DatasetLoss" metric that was presented in Section 4.2.

5.2 Experimental Results and Discussion

This section evaluates and discusses the experimental results about the effectiveness and efficiency of the proposed model.

5.2.1 Effectiveness

This section is devoted to evaluating the effect of the number of quasi-identifier attributes, the number of sensitive attributes, and the given value of l and how they influence the data utility.

5.2.1.1 The effectiveness based on the number of quasi-identifier attributes

The first experiment evaluates the number of quasi-identifier attributes that influence on the data utility of the experimental datasets. For experiments, the number of quasi-identifier attributes varied from 1 to 5 attributes. The value of l was fixed to be 2. Furthermore, only Capital loss was set to be the sensitive attribute.

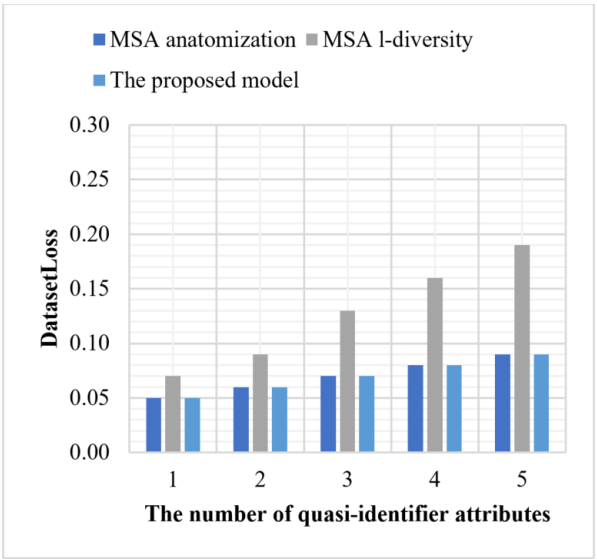

Fig.1: *The effectiveness based on the number of quasi-identifier attributes.*

From the experimental results that show in Fig.1, we conclude that the number of quasi-identifier attributes influences on the data utility of the experimental datasets which are constructed by all privacy preservation models, especially MSA l-Diversity. That is because a greater number of quasi-identifier attributes often leads to a variety of quasi-identifier values that are available in the partitions of the experimental datasets. In addition, the variety of quasi-identifier values directly influences on the data utility of the experimental datasets that are based on *DatasetLoss*. For example, let Table 7 be the experimental dataset. If we only consider Education as the quasi-identifier attribute, we can see that only Bachelors is the quasi-identifier value that is available in the experimental dataset. However, if Education and Native country are the quasi-identifier attributes, the quasi-identifier values in the experimental dataset have more variety than the experimental dataset that only has Education as the quasi-identifier attribute. Also, if Education, Native country, and Race are set to be the quasi-identifier attributes, the quasi-identifier values in the experimental dataset have more variety than the experimental datasets that have Education and Native country as the quasi-identifier attributes. With this example, it is clear that the greater number of quasi-identifier attributes influences on the data utility of experimental datasets which are constructed by all privacy preservation models. Furthermore, we observed that the experimental datasets are constructed by the proposed model and MSA anatomization, they are the same and have more data utility than MSA l-Diversity. The cause of the proposed model and MSA anatomization are the same, their partitions are the same. And the cause of lower data utility in the experimental datasets that are constructed by MSA l-Diversity, it is the effect of data generalization the are used to distort the unique quasi-identifier values. For example, let the ages 40 and 45 be the specified original quasi-identifier values from the age attribute of the experimental dataset. Their generalized values are in the range between 40 and 46. For this situation, the experimental dataset is constructed by MSA l-Diversity, it has four fake quasi-identifier values, i.e., 41, 42, 43, and 44. With this example, it is clear that the number of quasi-identifier attributes has more influence on the data utility in experimental datasets which are constructed by MSA l-Diversity than with the proposed model and MSA anatomization.

5.2.1.2 The effectiveness based on the number of sensitive attributes

The second experiment evaluates the number of sensitive attributes that influence on the data utility of the experimental datasets. For experiments, the number of sensitive attributes varied from 1 to 6 attributes. The value of l was fixed to be 2. More-

Table 7: *An example of the effect of quasi-identifier attributes.*

Education	Native country	Race
Bachelors	United States	Black
Bachelors	United States	White
Bachelors	India	Asian-Pac-Islander
Bachelors	United States	Black
Bachelors	United States	Black
Bachelors	Taiwan	Asian-Pac-Islander

over, all setting-up quasi-identifier attributes were available in the experimental datasets. From the experimental results that show in Fig.2, we observed that the number of sensitive attributes also influences on the data utility of experimental datasets which are constructed by all privacy preservation models. Moreover, the number of sensitive attributes is more influential than the number of quasi-identifier attributes because all experimental privacy preservation models are based on the number of distinct sensitive values. That is, when the number of sensitive attributes is increased, the data utility of the experimental datasets is decreased. This is because a greater number of sensitive attributes often leads to a larger size of partitions in the experimental datasets. For example, let the value of l be set at 2. Let Table 8 be the sensitive attributes that are available in the experimental dataset. If we only consider Relationship to be the sensitive attribute of the experimental datasets, the profile tuple of users in Table 8 can be partitioned into three partitions. However, if we use Relationship and Occupation as the sensitive attributes of the experimental datasets, the profile tuple of users is available in Table 8 can be partitioned into two partitions. If all attributes of Table 8 are set to be sensitive attributes of the experimental datasets, the profile tuple of users is available in Table 8 can only be constructed to be a single partition. These examples explicitly show that a greater number of sensitive attributes often leads to a larger size of partitions or less data utility in the experimental datasets of all privacy preservation models. Moreover, we observed from the experimental results that are shown in Fig.2 that MSA l-Diversity is less effective than others. The cause of the least effectiveness of MSA l-Diversity is the vulnerability of data generalization. Furthermore, we can observe that the experimental results indicate that the proposed model is less effective than MSA anatomization, but they are only a slight difference. The cause of more effectiveness in the experimental results of MSA anatomization is that every sensitive attribute group of the experimental datasets is independently considered into its appropriate partition. For this reason, the experimental datasets of MSA anatomization often lead to disorganized issues when they are utilized. However, the proposed model and MSA l-Diversity do not have any disorganized issues that must be improved when the number of sensitive groups that are available in the experimental datasets to be increased.

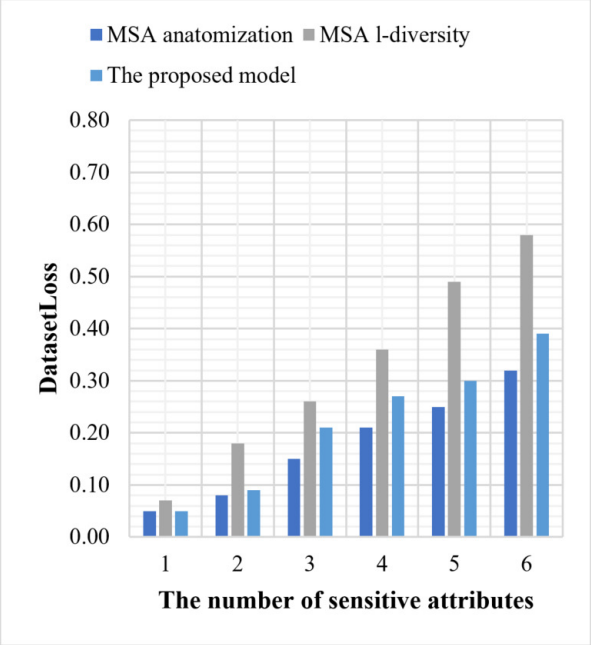

Fig.2: *The effectiveness based on the number of sensitive attributes.*

Table 8: *An example of the effect of sensitive attributes.*

Relationship	Occupation	Hours-per-week
Not-in-family	Exec-managerial	2
Own child	Exec-managerial	2
Husband	Exec-managerial	2
Unmarried	Exec-managerial	2
Wife	Prof-specialty	2
Husband	Prof-specialty	2

5.2.1.3 The effectiveness based on the value of l

The third experiment evaluates the value of l that influences on the data utility of the experimental datasets. For experiments, only Capital loss was set to be the sensitive attribute, and all setting-up quasi-identifier attributes were available in the experimental datasets. The value of l is varied in from 1 to 5.

From the experimental results that show in Fig.3, we observed that the value of l also influences on the data utility of the experimental datasets. That is when the value of l to be increased, the data utility of the experimental datasets is decreased, because a higher value of l often leads to a larger size of partitions in the experimental datasets. Furthermore,

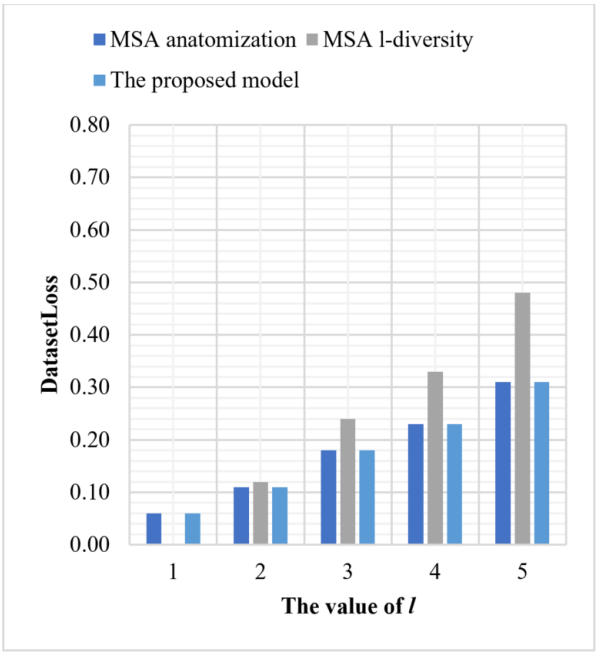

Fig.3: *The effectiveness based on the value of* l.

we observed that when the value of l is set to be 1, the experimental datasets are constructed by MSA l-Diversity, they do not lose any data utility because they always satisfy the given MSA l-Diversity constraint. However, when the experimental datasets are constructed by the proposed model and MSA anatomization, they still lose some data utility. Because some profile tuples of users are available in the experimental datasets, it has the quasi-identifier values to duplicate with other tuples, but we see its related sensitive value to be different. Furthermore, we observed that when the experimental datasets are constructed by the proposed model and MSA anatomization, they always have the same as the penalty cost of $DatasetLoss$. The experimental datasets which were constructed by the proposed model and MSA anatomization are always the same. Aside from $l = 1$, the experimental datasets were constructed by MSA l-Diversity, they have a penalty cost of $DatasetLoss$ greater than the proposed model and MSA anatomization. Also, the cause of the less data utility in the experimental datasets is constructed by MSA l-Diversity, it is the vulnerability of data generalization.

5.2.2 Efficiency

This section is devoted to evaluating the effect of the number of quasi-identifier and sensitive attributes and the given value of l such that they influence the execution time for transforming the experimental datasets to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity.

5.2.2.1 The efficiency based on the number of quasi-identifier attributes

The fourth experiment evaluates the number of quasi-identifier attributes that influence on the execution time for transforming the experimental datasets to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity. For experiments, the number of quasi-identifier attributes varied from 1 to 5 attributes. The value of l was fixed to be 2. Furthermore, only Capital loss was set to be the sensitive attribute.

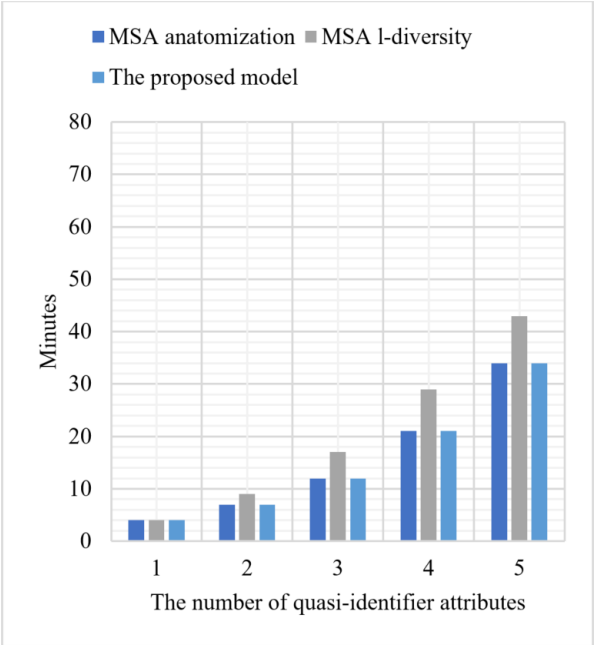

Fig.4: *The efficiency based on the number of quasi-identifier attributes.*

From the experimental results that show in Fig.4, we observe that the number of quasi-identifier attributes influences on the execution time for transforming the experimental datasets to satisfy privacy preservation constraints of all experimental privacy preservation models. That is, when the number of quasi-identifier attributes is increased, the execution time for transforming the experimental datasets is also increased. The cause of increasing the execution time is mat as the number of quasi-identifier attributes is increased, the search space about considering the quasi-identifier values for constructing the partitions of the experimental datasets also be increased. Furthermore, we observed that MSA l-Diversity often uses more execution time for transforming the experimental datasets than the proposed model and MSA anatomization. The cause of using more execution time in MSA l-Diversity is that aside from data partitioning, the experimental datasets are based on MSA l-Diversity constraints, so they have an additional data transformation cost for finding the appropriate less specific value for distorting the set of

unique quasi-identifier values. Moreover, we observe that the proposed model and MSA anatomization always use the same execution time for transforming the experimental datasets to satisfy the privacy preservation. They have the same search space about considering the quasi-identifier values for constructing the partitions of the experimental datasets.

5.2.2.2 The efficiency based on the number of sensitive attributes

The fifth experiment evaluates the number of sensitive attributes that influence on the execution time for transforming the experimental dataset to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity. For experiments, the number of sensitive attributes varied from 1 to 6 attributes. The value of l was fixed to be 2. Furthermore, only Capital loss was set to be the sensitive attribute. Moreover, all quasi-identifier attributes are available in the experimental datasets.

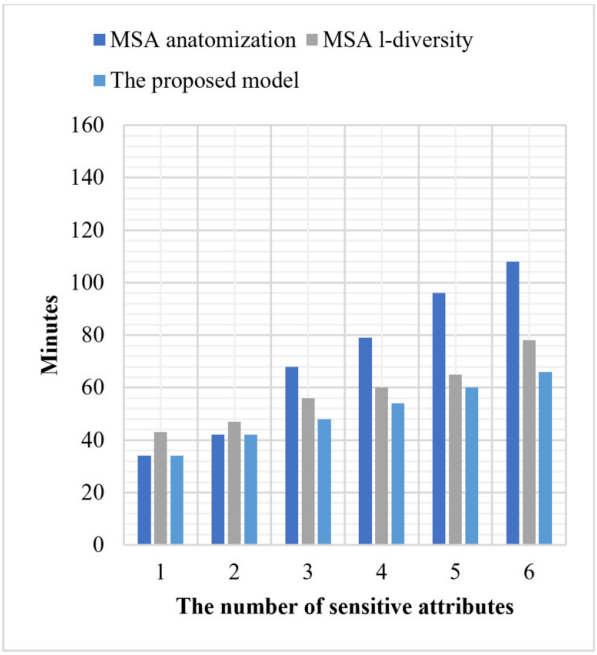

Fig.5: *The efficiency based on the number of sensitive attributes.*

From the experimental results that show in Fig.5, we observed that the number of sensitive attributes also influences on the execution time for transforming the experimental datasets to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity. In addition, we further observe that the number of sensitive attributes is more influential the execution time than the number of quasi-identifier attributes. The cause for using more execution time is that when the number of sensitive attributes is increased, the privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity are based on the number of the distinct sensitive values which are available in each sensitive attribute of the experimental datasets. Moreover, we can observe that MSA anatomization is less efficient than proposed model and MSA l-Diversity. The cause of the less efficiency in MSA anatomization is that every sensitive attribute group of the experimental datasets is independently considered into its appropriate partition. Furthermore, the experimental results show that MSA l-Diversity is less efficient than the proposed model. The cause of less efficiency of MSA l-Diversity is that MSA l-Diversity is based on data generalization.

5.2.2.3 The efficiency based on the value of l

The final experiment evaluates the given value of l that influences on the execution time which is used for transforming the experimental dataset to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity.

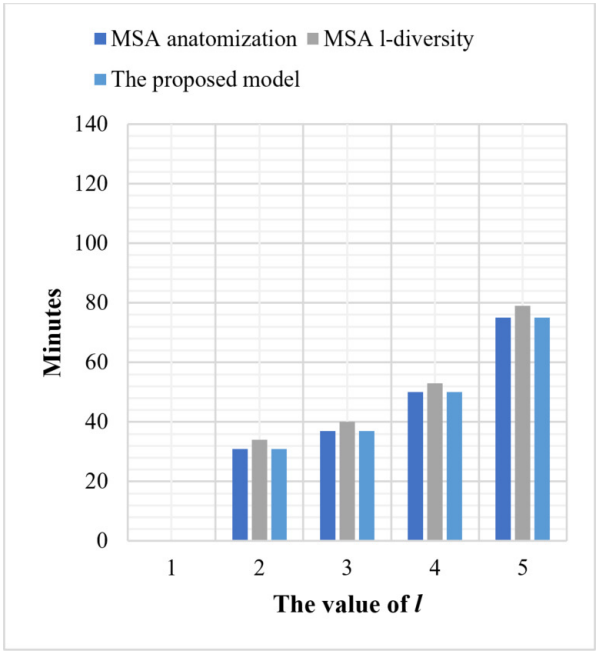

Fig.6: *The efficiency based on the value of l.*

For experiments, only Capital loss was set to be the sensitive attribute of the experiment datasets, and all quasi-identifier attributes were available in the experimental datasets. The value of l varied in the range between 1 and 5. From the experimental results that show in Fig.6, we observed that the value of l also has a greater influence on the execution time for transforming the experimental datasets to satisfy privacy preservation constraints of the proposed model, MSA anatomization, and MSA l-Diversity. When the value of l is increased, the execution time for transforming the experimental datasets also be increased. The cause of increasing the execution time is when increasing the value of l, it is that there are many sensitive values that must be considered. Moreover, MSA l-Diversity is less efficient than the proposed

model and MSA anatomization. The cause of less efficiency in MSA l-Diversity is that it is based on data generalization. Furthermore, we also observed that all experimental datasets of the proposed model and MSA anatomization always use same execution time for transforming the experimental datasets to satisfy privacy preservation constraints. That is because the partitions of the experimental datasets are constructed by the proposed model and MSA anatomization are always the same.

6. CONCLUSION

This work is devoted to resolving privacy violation issues in the farmer survey data collection to make them immune to attacks by the adversary background knowledge about the target users. To fix these issues, a privacy preservation model based on data shuffling to be proposed in this work. That is, before the farmer survey data collection is released for public use, their tuples are partitioned by the given value of l such that every sensitive attribute of each partition must include at least l different sensitive values. Furthermore, the sensitive values available in each partition are shuffled. Therefore, after the farmer survey data collection satisfy the proposed privacy preservation constraint, they are move highly secure in terms of privacy preservation than the original versions. Moreover, the proposed model is more efficient than MSA anatomization and MSA l-Diversity.

7. FUTURE WORK

Although the proposed model can address privacy violation issues in the farmer survey data collection, an adversary could discover new approaches that can be used to violate the privacy data of farmers that are collected by the farmer survey data collection in the future. Thus, improved privacy preservation models that can address the discovered privacy violation issues should be created in the future.

References

[1] S. Riyana, N. Riyana and W. Sujinda, "An Anatomization Model for Farmer Data Collections," *SN COMPUT. SCI*, no.353, 2021.

[2] S. Thongprakaisang, A. Thiantanawat, N. Rangkadilok, T. Suriyo and J. Satayavivad, "Glyphosate induces human breast cancer cells growth via estrogen receptors," *Food and Chemical Toxicology*, vol. 59, pp.129-136, 2013.

[3] C. Ventura, MRR. Nieto, N. Bourguignon, V. Lux-Lantos, H. Rodriguez, G. Cao, A. Randi, C. Cocca and M. Núñez, "Pesticide chlorpyrifos acts as an endocrine disruptor in adult rats causing changes in mammary gland and hormonal balance," *The Journal of Steroid Biochemistry and Molecular Biology*, vol. 156, pp. 1-9, 2016.

[4] S.H. Jee, H.W. Kuo, W. Su, C. Chang, C. Sun and J.D. Wang, "Photodamage and skin cancer among paraquat workers," *International Journal of Dermatology*, vol.34, no. 7, pp.466-469, 1995.

[5] J.D. Wang, W.E. Li, F.C. Hu and K.H. Hu, "Occupational risk and development of premalignant skin lesions among paraquat manufacturers," *British Journal of Industrial Medicine*, vol. 44, no. 3, pp.196-200, 1987.

[6] L. Sweeney, "k-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol.10, no. 5, pp.557-570, 2002.

[7] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," *22nd International Conference on Data Engineering (ICDE'06)*, pp.24-24, 2006.

[8] X. Xiao and Y. Tao, "Anatomy: simple and effective privacy preservation," in *Proceedings of the 32nd international conference on Very large data bases (VLDB'06)*, 2006.

[9] Q. Zhang, N. Koudas, D. Srivastava and T. Yu, "Aggregate Query Answering on Anonymized Tables," *2007 IEEE 23rd International Conference on Data Engineering*, pp. 116-125, 2007.

[10] R. Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid," *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 202-207, 1996.

[11] T. Kanwal, S. A. A. Shaukat, A. Anjum, S. R. Malik, K.-K. R. Choo, A. Khan, N. Ahmad, M. Ahmad and S. U. Khan, "Privacy-preserving model and generalization correlation attacks for 1: M data with multiple sensitive attributes," *Information Sciences*, vol. 488, pp. 238-256, 2019.

[12] A. Anjum,N. Ahmad, S. U. R. Malik, S. Zubair and B. Shahzad, "An efficient approach for publishing microdata for multiple sensitive attributes," *The Journal of Supercomputing*, vol. 74, pp. 5127-5155, 2018.

[13] Y. Wu, X. Ruan, S. Liao and X. Wang, "P-cover k-anonymity model for protecting multiple sensitive attributes," *2010 5th International Conference on Computer Science & Education*, pp. 179-183, 2010.

Surapon Riyana received a B.S. degree in computer science from Payap University (PYU), Chiangmai, Thailand, in 2005. Moreover, He further received a M.S. degree and a Ph.D. degree in computer engineering from Chiangmai University (CMU), Thailand, in 2012 and 2019 respectively. Currently, he is a lecturer in Smart Farming and Agricultural innovation Engineering (Continuing Program), School of Renewable Energy, Maejo University (MJU), Thailand. His research interests include data mining, databases, data models, privacy preservation, data security, and the internet of things.

Nobutaka Ito received a B.S in agricultural machinery from Mie University, Japan, in 1965. Moreover, he further received M.S and Ph.D. in agricultural engineering from Kyoto University, Japan, in 1967 and 1975 respectively. Currently, he is a visiting professor in Smart Farming and Agricultural innovation Engineering (Continuing Program), School of Renewable Energy, Maejo University (MJU), Thailand. His research interests include smart agricultures (precision agriculture, robotics, plant (green) factory, the application of micro nano bubble to agriculture, rice mechanization, food processing, community - based Asian agriculture; future farmer of Asia growing project, Asian agriculture: policy or strategy proposal, haze free issue in Thailand and Asia, and terramechanics: soil & machine system.

Tatsanee Chaiya received a B.S. degree in Computer Engineering from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, in 2013. She received the M.S. in computer engineering from Florida Institute of Technology (FIT), the United State of America, in 2016. She worked as an iOS and Android mobile application developer at Click Connect Co., Ltd, Thailand, from 2013 to 2014. Currently, she is a lecturer in Smart Farming and Agricultural innovation Engineering (Continuing Program), Maejo University (MJU), Chiangmai, Thailand. Her research interests include data mining, machine learning, data science, and artificial intelligence.

Uthaiwan Sriwichai received a B.S. degree and a M.S. degree in computer science from Chiangmai University (CMU), Thailand. Currently, she is a lecturer in Smart Farming and Agricultural innovation Engineering (Continuing Program), Maejo University (MJU), Chiangmai, Thailand. Her research interests include data mining, machine learning, data science, and artificial intelligence.

Natthawud Dussadee received a B.S. degree in physics from Srinakharin wirot University, Thailand, in 1989. He received a M.S. degree in energy technology from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, in 1991, and he further received a M.S. degree in mechanical engineering from Chiangmai University (CMU), Thailand, in 1997. Moreover, He received a Ph.D. degree in energy technology from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, in 2003. His research interests include energy technologies, renewable technologies, and agriculture technologies.

Tanate Chaichana (Asst.Prof) received a B.Sc (Physics) from Prince of Songkla University, Thailand, in 2001. He received a M.Eng and D.Eng in the field of Energy Engineering from Chiangmai University Thailand, in 2004 and 2010 respectively. His research interests include renewable energy technologies and management for agriculture and community.

Rittichai Assawarachan received a B.S. degree in food engineering from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, in 2000. He received a M.S. degree in food engineering from King Mongkut's University of Technology Thonburi (KMUTT), Thailand, in 2000. Moreover, he received a Ph.D. in food engineering and bioprocess technology from Asian Institute of Technology (AIT), Phahonyothin Rd, Khlong Nueng, Khlong Luang District, Pathum Thani, Thailand. His research interests include the design and development of food processing machinery.

Thongchai Maneechukate received a B.S. degree in physics from Srinakharinwirot University, Thailand. Moreover, he received a M.S. degree and a Ph.D. degree in electrical engineering from King Mongkut's University of Technology Thonburi (KMUTT), Thailand. His research interests include renewable energies.

Samerkhwan Tantikul received a B.S. degree in mechanical engineering from Rajamangala University of Technology, Thailand. Moreover, he received a M.S. degree in educational psychology and a M.S. degree in agricultural machinery engineering from Mahasarakham University, Thailand and Khon Kaen University, Thailand respectively. His research interests include agricultural machinery engineering.

Noppamas Riyana received a B.S. degree in computer science from Payap University (PYU), Thailand, in 2005. Moreover, she received a M.S. degree in business administration from Payap University (PYU), Thailand, in 2012. Currently, she is a computer technical officer at Maejo University (MJU), Thailand. Her research interests include data mining, databases, data models, and privacy preservation.