Journal Pre-proof

Fuzz-classification (*p*, *l*)-Angel: An enhanced hybrid artificial intelligence based fuzzy logic for multiple sensitive attributes against privacy breaches

Tehsin Kanwal, Hasina Attaullah, Adeel Anjum, Abid Khan, Gwanggil Jeon

PII: S2352-8648(22)00200-0

DOI: https://doi.org/10.1016/j.dcan.2022.09.025

Reference: DCAN 523

To appear in: *Digital Communications and Networks*

Received Date: 2 June 2021

Revised Date: 22 July 2022

Accepted Date: 29 September 2022

Digital Communications and Networks(DCN)

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/dcan

Fuzz-classification (p, l)-Angel: an enhanced hybrid artificial intelligence based fuzzy logic for multiple sensitive attributes against privacy breaches

Tehsin Kanwal[a], Hasina Attaullah[a], Adeel Anjum[b], Abid Khan[c], Gwanggil Jeon[*,d]

[a]Department of Computer Science, COMSATS University Islamabad, Pakistan
b Department of Information Technology, Quaid-e-Azam University Islamabad, Pakistan
c College of Science and Engineering, School of Computing and Engineering, University of Derby, DE22 1GB, UK
[d]Department of Embedded Systems Engineering, Incheon National University, South Korea

Abstract

The inability of traditional privacy-preserving models to protect multiple datasets based on sensitive attributes has prompted researchers to propose models such as SLOMS, SLAMSA, (p, k)-Angelization, and (p, l)-Angelization, but these were found to be insufficient in terms of robust privacy and performance. (p, l)-Angelization was successful against different privacy disclosures, but it was not efficient. To the best of our knowledge, no robust privacy model based on fuzzy logic has been proposed to protect the privacy of sensitive attributes with multiple records. In this paper, we suggest an improved version of (p, l)-Angelization based on a hybrid AI approach and privacy-preserving approach like Generalization. Fuzz-classification (p, l)-Angel uses artificial intelligence based fuzzy logic for classification, a high-dimensional segmentation technique for segmenting quasi-identifiers and multiple sensitive attributes. We demonstrate the feasibility of the proposed solution by modelling and analyzing privacy violations using High-Level Petri Nets. The results of the experiment demonstrate that the proposed approach produces better results in terms of efficiency and utility.

KEYWORDS: Generalization, Fuzzy-logic, MSA, Privacy disclosures, Membership function, (p, l)-Angelization, QT, HLPN

1. Introduction

Given recent advances in the Internet of Things (IoT), big data, and machine learning, which have led to a surge in requests for data resources, it is critical that data be provided in a privacy-preserving manner that does not compromise individual privacy. Companies like Apple have also encouraged the use of differential privacy in their products as an example of such practices. Furthermore, there is a wide range of enterprises and startups, such as Aircloak, whose main service is the anonymisation of data sets, providing a more widely accepted avenue for Privacy-Preserving Data Publishing (PPDP) [1]. Anonymization is used in PPDP techniques to protect an individual's sensitive information before publication [2]. The privacy of publicly available data is a critical challenge since it may include sensitive and private information about individuals, such as age, gender, and other attributes that make an individual uniquely identifiable. Additionally, sensitive information is not limited to a Single Sensitive Attribute (SSA), but can also include a person's Multiple Sensitive Attributes (MSAs). As the number of SAs in information increases, so does the threat of identification of individuals [3]. There have been many methods proposed in the literature for SSA or MSA-based datasets to anonymize sensitive data. Some of these suggested solutions, like k-anonymity [4, 5], p-sensitive k-anonymity [6], l-diversity [7], and t-closeness [8], utilized generalization to address SSA,

*Gwanggil Jeon (Corresponding author) (email:gjeon@inu.ac.kr)
[1]Tehsin Kanwal (email: tehseenkanwal@yahoo.com)
[2]Adeel Anjum (aanjum@qau.edu.pk)
[3]Hasina Attaullah (hasina.attaullah12@gmail.com)
[4]Abid Khan(a.khan3@derby.ac.uk)

while others, like anatomy [9], used bucketization to consider SSA. Several other methods have been suggested for making MSAs anonymous, including slicing [10, 11], ANGELMS [12], P-cover k-anonymity [13], p+-sensitive k-anonymity [14], the additive noise technique [15], and bucketization, used in the decomposition [16] approach. It has been demonstrated that MSA-based privacy approaches fail to protect privacy when the adversary uses MSA correlation, background, and non-membership knowledge [17] to reveal privacy. Furthermore, despite the extensive literature on single-record data sets, multi-record data sets (1:M datasets) have received little attention from the research community. As a result, in the case of 1: M data sets [18], the latest privacy work faces the possibility of severe privacy breaches. Most health-related microdata publishing entities today are more concerned with data protection and data loss, while traditional privacy protection strategies attempt to strike a balance between privacy and utility, but their effectiveness needs to be re-evaluated.

In this paper, we will employ an Artificial Intelligence (AI)-based fuzzy logic technique [19]. Fuzzy logic, a human-based reasoning system can be applied to process modelling, computer vision, deep learning, autonomous control systems, data mining, and data classification. Fuzzy logic is a rule-based technique for partitioning multidimensional data. It takes imprecise data from tables and outputs precise fuzzy sets. We may use fuzzy logic to classify Quasi Identifiers (QIs) and Sensitive Attributes (SAs) in privacy-preserving techniques. Fuzzy-based methods for privacy protection have been suggested in the literature [20, 21], but none of them provide multi-record with MSAs. The privacy preservation of multi records (1: M) with MSAs is re-investigated in this paper, and a fuzzy logic-based efficient technique is proposed for privacy protection. Fuzzy classification not only preserves privacy but also increases data utility by classifying correlated attributes using multidimensional partitioning. Fuzzy logic works for QAs and SAs, unlike techniques that suggest two separate methods for QIs and SAs, which results in minimal overhead. In this paper, an anonymization approach called Fuzz-classification (p, l)-Angel is proposed to efficiently protect the privacy of published data. Our main contributions are summarized below.

1. In (p, l)-Angelization [17], privacy disclosures based on 1: M MSA generalization was re-investigated, and an AI-based Fuzzy Logic (FL) is introduced for the design of an enhanced approach called Fuzz-classification (p, l)-Angel.
2. Formal modeling and analysis of Fuzz-classification (p, l)-Angel, is performed using High-Level Petri Nets (HLPN) [22, 23]. The formal proof shows that the proposed enhanced approach provides the same defense against the identified adversarial attacks.
3. The proposed fuzzy logic-based approach is an enhanced form of (p, l)-Angelization as it relates to privacy, efficiency, and utility. The aforesaid is also proved by performing experiments on a real-world 1: M-MSA micro data set.

The rest of the paper is organized as follows. Section 2 will go through some of the recent related work that has been done to protect the privacy of 1: M, MSAs, and a combination of 1:M and MSAs. Section 3 would include a systematic adversarial analysis of (p, l)-Angelization. The proposed Fuzz-classification (p, l)-Angel is in Section 4, and the formal verification, will be discussed in-depth in Section 5. A comparison of the proposed methodology and the (p, l)-Angelization will be used in Section 6 to highlight the experimental results. Finally, Section 7 concludes this work.

2. Related work

This section illustrates the work done so far on MSAs and 1: M datasets. The privacy-preserving approach of SSA is infeasible for MSA because the probability of re-identifying individuals in any data set is high as SAs increase [3]. The proposed MSAs-based techniques are based on generalization, decomposition, slicing, anatomization, and bucketization. The first proposed decomposition-based algorithm [16] is grounded on l-diversity principal with vertical partitioning for MSAs. Decomposition plus [24] extends the works for Decomposition [16], but it retains the noise value near to the original value.

The concept of providing privacy for MSAs using slicing was first introduced in [11] and then improved slicing models are presented in [10], which leverage the use of suppression and Mondrian slicing. In [25], a privacy-preserving technique called “SLOMS” uses the basic concept of slicing and removes the correlation between MSAs. SLASMA [26] another privacy model for MSAs is proposed, that combines anatomization [9] with slicing [11], but it does not generalize QIs, thus improving the utility. For the privacy protection of numerical MSAs, Multi-Sensitive Bucketization (MSB) based techniques have been developed [27, 28], however, these approaches ignored textual data. In [29], a rating technique for MSAs was proposed, and the algorithm generalizes the multiple sensitive attributes, leading to information loss. The author in [28] minimizes information loss in the rating algorithm by avoiding association attacks in published data. ANGELMS [12] anonymizes the MSAs data set by using anatomy with generalization and vertical partitioning. The privacy model (p, k)-angelization [30] is a weighted privacy model for MSAs. It is more important than others when it comes to information loss and privacy. But still, it has some limitations as weights are calculated based on sensitivity and dependency of SAs. The enhancement of the KC-slice

[31] model with improved utility and privacy is proposed as a novel KCi-slice [32]. Privacy preservation of multiple sensitive attributes based on the security level, with various security levels for distinct SA values, is presented in [33]. The proposed approach claims higher utility, but the execution time is also higher. In [34] fingerprint correlation attack is identified in (p, k)-angelization [30] and based on that attack an improved (c, k)-anonymization [34] algorithm is proposed. The recent work (K, L) anonymity [35], utilizes the k-anonymity model together with Laplace differential privacy to ensure privacy. The proposed approach claims to avoid a linking attack. Though we have only discussed MSAs-based techniques with a single instance of any record thus far, there may be multiple instances (1:M) of a single record in more complicated cases. The literature only makes a very minor contribution to the 1:M dataset. In [36], the preliminary research in 1:M datasets is first presented. In this paper, the authors suggested a new privacy model based on l-diversity and k-anonymity, but it excessively generalizes both sensitive and quasi attributes. It has been highlighted that their method has minimal utility and requires a lot of computing time. Additionally, it is demonstrated in [17] that it is vulnerable to the MSA correlation generalization attack. An effective privacy-preserving model for 1: M microdata, with higher utility, has been proposed in [37]. Although it was an improved work in adversarial attacks modeling and analysis it lacks MSAs consideration. The horizontal sliced permuted permutation (H-SPP) for 1: M microdata, is proposed in [38]. It makes use of slicing and anatomy to avoid identity, attribute, and membership disclosure risks. Some other privacy models for 1: M data sets are also proposed in [39] and [40].

The work debated up to this point is either in MSAs or in 1: M. There is only two privacy model proposed in the literature for 1: M together with MSAs. The earliest privacy technique for 1: M and MSAs are proposed in [17], which re-examines the findings of [36] for privacy disclosures based on 1: M and MSAs. Although the proposed method shows an effective defense against adversarial attacks, it means that it can be more effective in terms of privacy. An adversarial attacks identification in a balanced p-sensitive k-anonymity privacy model for 1:M and MSAs has been suggested in a recent study [18]. They presented the 1:M MSA-(p, l)-diversity privacy method, which is efficient, resilient, and utility aware. To the best of our knowledge, most of the work done for privacy protection and adversarial attack prevention lacks AI-based fuzzy-logic techniques. Some of the early work in privacy preservation using fuzzy logic is presented in [20, 41, 42, 43, 44, 45], but it lacks basic adversarial attack models and other relevant explanations. The recent article [46] makes use of fuzzy sets to categorize numerical and categorical attributes uniformly. Based on those categories, sensitivity levels are introduced, and (α, k)-anonymity privacy model is proposed for hierarchical data. In [47] data privacy is ensured using data perturbation. The individual's private data is perturbed using a fuzzy membership function. In article [48], the authors proposed a classification based on fuzzy logic, but it only applies to MSAs. All of the aforementioned proposed fuzzy logic methods lack the fundamental privacy adversarial models and are therefore unsuitable for MSAs and 1: M datasets.

3. A Review of privacy breaches in (p, l)-Angelization

This section revisit (p, l)-Angelization [17] working and provides a short formal overview of the privacy disclosures. MSAs correlation, adversarial background knowledge, and Non-membership knowledge are the key sources of privacy disclosures that are invalidated in (p, l)-Angelization. It also improves the 1: M generalization's high information loss. Since Angelization is a combination of bucket and batch partitioning, each bucket partitioning assures the (p, k)-anonymity principle since each bucket includes records from c groups. Each bucket contains at least k tuples, with k being the group size that minimises the linking attack. Each batch partitioning also adheres to the (p, k)-anonymity principle. Each batch and bucket must also adhere to the l-diversity principle. With MSAs, (p, l)-Angelization ensures the secure publication of a 1: M dataset. This method effectively preserves the privacy of individual publicly available data from MSAs correlation-based adversarial attacks. If a batch partitioning = $\{BA_1, BA_2, ...BA_h\}$ and a bucket partitioning = $\{BA_1, BA_2, ..., BA_K\}$, and when (p, l)-Angelization of the microdata Table T is provided, two tables are formed: a Sensitive Batch Table (SBT) and a Generalised Table (GT), where SBT is of the form: ST ,BatchID, where $ST = \{C^s_1 i, C^s_2 i, C^s_3 i, ..., C^s_n i\}$. SBT contains the row (i, ST), where i is the batch ID and ST is the set of sensitive attributes, for each batch A_i $(1 \leq i \leq g)$ and every sensitive value $s \in S$ that occurs in A_i. GT includes an additional column named Batch-ID in addition to all the QI attributes from microdata T. Each tuple $t \in T$ defines a row in GT. Each row contains a collection of the generalized QI values of t with Batch-ID. Fig. 1 displays the (p, l)-angel algorithm along with the HLPN model. Interested readers can refer to [17], for more details about algorithm steps and formal rules of (p, l)-angel algorithm. Since angelization combines bucket and batch partitioning, each bucket partitioning satisfies the (p, k)-anonymity criterion as each bucket contains records from c categories. Each bucket includes at least k tuples, where k is the minimum group size to minimize the linking attack. Each batch partitioning also adheres to the (p, k)-anonymity principle. Each batch and bucket must also meet l-diversity [7] requirements. With MSAs,

Table 1
MSAs correlation attacks description with formal rule representation

MSA correlation attacks	Attacks description	Formal representation based on (p,l)-Angelization HLPN model
Sensitive correlation attacks (Scor)	If the adversary can use MSA and background knowledge to correlate an individual's sensitive attributes, he or she can execute Scor Attacks.	R(Scor Attacks)= $\forall i18 \in x18, \forall i19 \in x19, \forall i20 \in x20, \forall i21 \in x21 \| Scor\ Dis(i18[1], i19[1]) \rightarrow (\{i18[1], i19[1]\} \cup \{i20[2]\}) \neq i2[2] \wedge i2[3](i21[2] \cup i21[3]) = \Phi$
Non-membership correlation attacks (Nmcor)	If the adversary can successfully find the absence of individual MSA in published data, he or she can launch Nmcor attacks.	R (Nm Attacks)= $\forall i22 \in x22, \forall i23 \in x23, \forall i24 \in x24, \forall i25 \in x25 \| Nm\ Dis(i22[1], i24[2]) \rightarrow i25[2] \wedge Nm\ Dis(i23[1], i24[2]) = \Phi$
Quasi-correlation attacks (Qcor)	If an adversary can map a person to a sensitive value in published data using external MSA information and quasi-identifiers like age, gender, and zipcode, he or she can execute a Qcor attack.	R (Qcor Attacks)= $\forall i26 \in x26, \forall i27 \in x27, \forall i28 \in x28, \forall i29 \in x29 \| Qcor\ Dis(i26[1], i27[1]) \rightarrow (\{i26[1], i27[1]\} \cup \{i28[2]\}) \neq i2[1] \wedge i2[3](i29[2] \cup i29[3]) = \Phi$

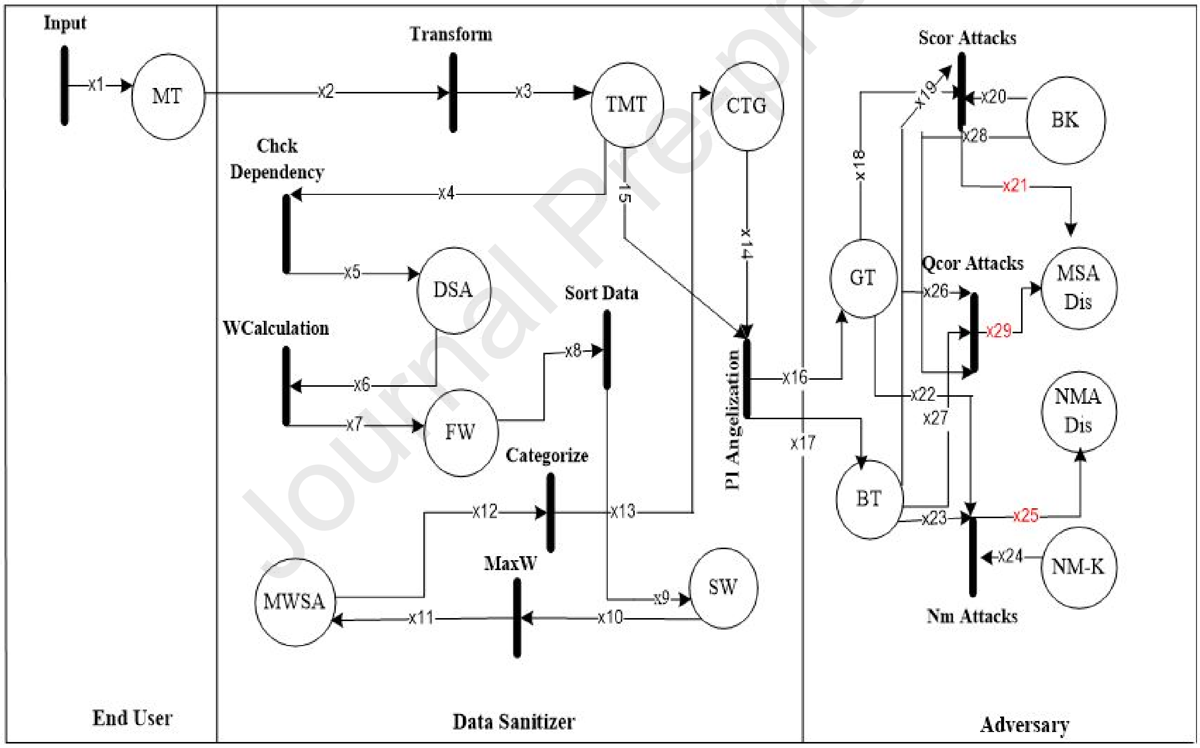

Fig. 1. HLPN model for (p, l)-Angelization

(p, l)-Angelization ensures the secure publication of a 1: M data set. This approach effectively protects the privacy of the individual published data from MSAs' adversarial disclosures as explained in Table 1. (p, l)-Angelization algorithm performs a dependency-based SA weight calculation, and the MSAs category formation depends on the maximum weight and the release of SBT and GT. In the subsequent section, we will propose a fuzzy logic-based technique that will provide privacy and utility for MSA and 1:M-based data sets.

4. Proposed enhanced Fuzz-classification (p, l)-Angel

This section describes the work of fuzzy logic-based privacy-enhancing methods.The working of proposed approach is elaborated in subsequent sections.

4.1. Proposed Fuzz-classification (p,l)-Angel

We propose a fuzzy (p, l)-Angel method for converting data attributes into fuzzy sets. Fuzzification is the process of converting data attributes into fuzzy

Table 2
Original data table

Name	Gender	Age	Zipcode	Disease	Treatment	Physician	Symptom	Diagnostic-method
P1(Johny)	M	27	14248	HIV	Antiretroviral Therapy (ART)	John	Infection	Blood Test
P2(Johny)	M	27	14248	Dyspepsia	Antibiotic	Sarah	Digestive Upset	Ultrasound
P3(Ana	F	28	14207	HIV	ART	John	Loss of Weight	ELISA Test
P4 Richard	M	26	14206	Cancer	Radiation	Alice	Loss of Weight	MRI Scan
P5 Dave	M	25	14249	Cancer	Chemotherapy	Bob	Abdominal Pain	Chest x-ray
P6 Kate	F	41	13053	Hepatitis	Drugs	Sarah	Fever	Blood Test
P7 Kate	F	41	13053	Phthisis	Antibiotic	David	Fever	Molecular Diagnostic Test
P8 Kate	F	41	13053	Flu	Medication	Suzan	Fever	RIDT tests
P9 William	M	48	13074	Phthisis	Antibiotic	David	Fever	Molecular Diagnostic Test
P10 Robert	M	45	13064	Asthma		Suzan	Difficulty in Breathing	MCCT
P11 Olivia	F	42	13062	Obesity	Nutrition Control	Steven	Eating Disorders	Body Mass Index (BMI)
P12 Emily	F	33	14248	Flu	Medication	Suzan	Fever	RIDT tests
P13 Alec	M	37	14204	Flu	Medication	Eve	Fever	RIDT tests
P14 Oliver	M	36	14205	Flu	Medication	Anas	Fever	RIDT tests
P15 James	M	35	14248	Digestive Upset	Medication	Jem	Heartburn	Chest X-Ray
P16 James	M	35	14248	Stomach Cancer	Surgery	Jem	Digestive Upset	Endoscopy
P17 Jessica	F	28	14249	Cancer	Chemotherapy	Bob	Abdominal Pain	Chest x-ray

Table 3
Transformed data table

Name	Gender	Age	Zipcode	Disease	Treatment	Physician	Symptom	Diagnostic-method
P1 {1,2} (Johny)	M	27	14248	{HIV, Dyspepsia}	{Antiretroviral therapy (ART), Antibiotic}	{John, Sarah}	{Infection, Digestive Upset}	{Blood Test, Ultrasound}
P2 (Ana	F	28	14207	HIV	ART	John	Loss of Weight	ELISA Test
P3 Richard	M	26	14206	Cancer	Radiation	Alice	Loss of Weight	MRI Scan
P4 Dave	M	25	14249	Cancer	Chemotherapy	Bob	Abdominal Pain	Chest x-ray
P5 {5,6,7} Kate	F	41	13053	{Hepatitis, phthisis,Flu}	{Drugs,Antibiotic, Medication}	{Sarah, David, Suzan}	{Fever, Fever, Fever}	{Blood test, MDM, RIDT tests}
P6 William	M	48	13074	Phthisis	Antibiotic	David	Fever	Molecular Diagnostic Test
P7 Robert	M	45	13064	Asthma	Medication	Suzan	Difficulty in Breathing	MCCT
P8 Olivia	F	42	13062	Obesity	Nutrition Control	Steven	Eating disorders	Body Mass Index (BMI)
P9 Emily	F	33	14248	Flu	Medication	Suzan	Fever	RIDT tests
P10 Alec	M	37	14204	Flu	Medication	Eve	Fever	RIDT Tests
P11 Oliver	M	36	14205	Flu	Medication	Anas	Fever	RIDT Tests
P12 {15,16} James	M	35	14248	{digestive upset, Stomach Cancer}	{Medication, Surgery}	{Jem, Jem}	{Heartburn, Digestive Upset}	{Chest X-Ray, Endoscopy}
P13 Jessica	F	28	14249	Cancer	Chemotherapy	Bob	Abdominal Pain	Chest X-Ray

sets. The first step in making fuzzy set is to categorize attributes according to their priority. This is explained in subsequent subsection.

4.1.1. Weight assignment:

The first step is to assign weights to attributes so that membership functions can be defined. Let us take the attribute physician in Table 2 as an example and convert it to a fuzzy sets. First, identify the distinct physicians in Table 4, the total number of distinct physicians (p) is ten, and they are ordered from most critical to least critical, moderately critical to less critical, and so on. The weights are computed by means of Rank Order Centroid (ROC) using rank assigned to physicians. Table 5 represents calculated ROC-based weights as given. Based on ROC based weights, fuzzy set for physicians is defined. The equation for calculating ranked based weight is given below where Ps is the number of physicians and Wk is the weight for the kth physician. We will repeat above steps for all sensitive attributes to get weights.

$$Wk = (1/Ps)\sum_{r=k}^{Ps}\frac{1}{r} \tag{1}$$

4.1.2. Fuzzy sets and rule base inference

Next Fuzzy sets are defined for each attribute and IF-THEN rules are defined based on these fuzzy sets (FSs). IF-THEN rules are evaluated to get an output of privacy classes. Let linguistic variables (qi) and (msa) represent QIs and MSAs, respectively. Let 'i' mfs be constructed for QIs, then fuzzy sets for (qi) are generated. The generalized form of fuzzy sets for two QIs X and Y are illustrated in equation 2.

$$\begin{aligned}(Fuzzy-Set)_X((qi_A)) &= \{(qi_{X1}),(qi_{X2}),.....,(qi_{Xi})\}\\ (Fuzzy-Set)_Y((qi_B)) &= \{(qi_{Y1}),(qi_{Y2}),.....,(qi_{Yi})\}\end{aligned} \tag{2}$$

Following the formation of fuzzy sets, rules are computed and privacy classes are defined using 3.

$$IF\ QI_X\ is\ (qi_{Xi})\ \cup\ QI_Y\ is\ (qi_{Yj}) action = q(PC)_{(i+j-1)} \tag{3}$$

The next step is assignment of tuples to privacy classes.

4.1.3. Assignment phase

Next step is assignment of records to Privacy Classes (PCs). PCs are created according to defined rules, and tuples in the table are assigned to each PC according to their values. Classification and assignment need to be done for every attribute of data set. Table 6 and Table 7 shows the privacy classes for QIs and MSAs respectively.

4.1.4. Anonymization phase

The classification and assignment of QIs and SAs to fuzzy privacy classes are done in last phase. The final phase is to integrate SA PCs into the QI PC table and assign PCs to tuple ids. Table 8 contains the anonymized QT and Table 9 shows anonymized SAs. Below is a line-by-line explanation of the algorithm.

Table 4
Rank based weight calculation

Physician	Rank	Weight (ROC)
John	1	(1+1/2+,...,+1/10)/10 = 0.31
Jem	2	(1/2+1/3+,...,+1/10)/10 =0.20
Alice	3	(1/3+1/4+,...,+1/10)/10 =0.14
Anas	4	(1/4+1/5+,...,+1/10)/10 =0.11
Bob	5	(1/5+1/6+,...,+1/10)/10 =0.08
Eve	6	(1/6+1/7+,...,+1/10)/10 =0.06
Sarah	7	(1/7+1/8+1/9+1/10)/10 =0.04
Suzan	8	(1/8+1/9+1/10)/10 =0.03
David	9	(1/9+1/10)/10 = 0.02
Steven	10	(1/10)/10=0.01

Table 5
Rank based physician weight calculation

Physician	Rank	Weight (ROC)	Fuzzy set for p	
John	1	0.31	P_{mc}	p is most critical
Jem	2	0.20	P_{lc}	p is least critical
Alice	3	0.14	P_{mrc}	p is more critical
Anas	4	0.11	P_{lsc}	p is less critical
Bob	5	0.08	P_{mdc}	p is moderate critical
Eve	6	0.06	P_{lscr}	p is lesser critical
Sarah	7	0.04	P_c	P is critical
Suzan	8	0.03	P_s	P is sensitive
David	9	0.02	P_{ls}	P is less sensitive
Steven	10	0.01	P_n	P is normal

In Algorithm 1 Line 12, 13 merge multiple records in MT into a single record representation and split the table into QIs and SAs attributes subsets (Table 3). Data attributes are called Linguistic variables (Lvn). Line 14-16: identify unique attributes and rank them (r) according to weights. Line 18-20, define linguistic variables for unique attributes using Rank order centroid (ROC). Line 21-23, define MFs for every attribute. Every attribute can have two, three, or four MFs. Line 26 makes classification rules for data-set Line 26, assigns classification rules to privacy classes. Line 33 identifies which privacy class a tuple in Table 'TMT' belongs to as shown in Table 6. Line 34-35 generate tables (QT) and (SAT) with new attribute class and attributes of subset table (Table 7, Table 7b). Every SAs subset table's QT (tuple) is checked on line 40 to see if it belongs to which privacy class (SAT). Line 41, for each subset of SAT, add a privacy class to QT. Line 46-48, publish FQT and FMSAT tables (Table 8 Table 9). In the following section, we demon-

Algorithm 1 Fuzz-classification (p,l)-Angel

procedure FUZZIFICATION
2: **Input MT:** Microdata table = {Lvn}
(MMT) : Transformed Microdata table ={Lvn}
4: Multiple Sub-sets= $\{sb1, sb2, sb3, \ldots, sbk\}$
Membership Functions MFs for Lvn
6: α = no. of attributes in one sb
Ω = no. of MFs for Lvn
8: $v = Numberoffuzzyrules(v = \Omega^{\alpha})$
Classification Rules CR [] = $CR_1, CR_2, \ldots, CR_v$
10: Privacy Classes []= $PC1, PC2, \ldots, PC_v$
Output Release Table FQT and FST
12: $(MMT) := M-Transform(Lvn)$
$sb_{(k_{(\forall k \in m)})} := Split((MMT))$
14: **for** $i = 1$ to m **do**
$RnkAtb_{i_{\forall i \in sb}} := (Rank(Distinct(sb_{k_{\forall k \in m}}))$
16: **end for**
//Step1: line 17-26 represents classification phase/
18: **for** $i = 1$ to j **do**
$Lvn_{Rb} := ROC(RnkAtb_{i_{\forall i \in sb}}), sb_{k_{\forall k \in m}}$
20: **end for**
for $k = 1$ to m **do**
22: $MF_s := Membership(Lvn_R b)$
end for
24: **for** $i = 1$ to a **do**
for $j = 1$ to ω **do**
26: $CR[i] := AND\{Lvn[1][j] \wedge Lvn[2][j] \wedge, \ldots\ldots\ldots, \wedge Lvn[\eta][j]\}$
$PC[i] := CR[i]$
28: **end for**
end for
30: /* Step2: line 30-40 represents privacy classes assignment phase */
for $i = 1$ to n **do**
32: **for** $j = 1$ to v **do**
if T(Tuple) $\in$ (PC[v]) **then**
34: Create a new table for QIs (QT) and SAT ($SAT_1, \ldots.., SAT_m$) based on the Classes.
end if
36: **end for**
end for
38: **for** $i = 1$ to n **do**
for $j = 1$ to m **do**
40: **if** QT(tuple) $\in$ (STm) **then**
Append dataset PC(STm) in QT for every SAT
42: **end if**
end for
44: **end for**
//Step 3:line 45-47 represents Fuzzy-Publication phase/
46: **for** $i = 1$ to T **do**
Publish FQT , FMSAT
48: **end for**
return FQT , FMSAT
50: **end procedure**

Table 6
Classification of QIs (Age-Zip code)

Age	Zip-code 13053-14204	14205-14249
24-27		{P1,P2,P3,P4} qPC1
28-35		{P9,P12,P13} qPC2
36-45	{P5,P6,P7,P8, P10} qPC3	{P11} qPC4

strate that the adversarial MSAs correlation attack can be successfully mitigated using the proposed approach through formal modelling and analysis.

5. Formal modeling and analysis of Fuzz-classification (p, l)-Angel with privacy attacks mitigation

Formal modelling and analysis of the proposed Fuzz-classification (p,l)-Angel-based algorithm will be demonstrated in this section. Furthermore, we will also perform the mitigation of privacy attacks through HLPN. For this purpose, we convert the proposed algorithm into the HLPN model. Descriptions of the variable types are given in Table 10. Table 11 shows the model places and its description. In formal modeling using HLPN, we identify the data types, Places (P), and mappings (Interested readers are encouraged to read [22, 23] for further details about the use of HLPN). Fig. 2 depicts the working of the HLPN model with privacy-attack invalidation. The first input transition shows the raw data table with data attributes and r number of tuples stored in Table.

$$\begin{gathered}\textbf{R(M-Merging)} = \forall i2 \in x2, i3 \in x3| \\ (i3[1], i3([2])_{i_{\forall i3[2] \in i}}) \\ := Transfrm-Rec(i2[1], (i2[2])_{m_{\forall i2[2]\in m}})\wedge \\ x3' := x3 \cup \{i3[1], i3[2]\}\end{gathered} \tag{4}$$

$$\begin{gathered}\textbf{R(D-Split)} = \forall i4 \in x4, i5 \in x5| \\ (i5[1], i5([2])_{i_{\forall i5[2]\in i}}) := DSplt(i4[2]_{m_{\forall i5[2]\in m}})\wedge \\ x5' := 5 \cup \{i5[1], i5[2]\}\end{gathered} \tag{5}$$

Algorithm starts with the transformation of multi records in Table in to single record representation by merging the same data attributes of patient records in transition M-Merging . The data attributes are divided in Transition D-Split into multiple subsets of QI and SAs. Following this linguistic variable identification, we rank each variable's value and use transition A-L-Conv to transform attributes to linguistic variables Lvn.

$$\begin{gathered}\textbf{R(Rank)} = \forall i8 \in x8, i9 \in x9| \\ (i7([1]_{i_{(\forall i7[1]\in i}}) := ((i6[1])_{m_{\forall i6[1]\in m}})\wedge \\ (i7([2])_{z_{\forall i7[2]\in z}}) := W-bsrnk(i6[2]_{m_{\forall i6[2]\in b}})\wedge \\ x7' := x7 \cup \{i7[1], i7[2]\}\end{gathered} \tag{6}$$

$$\begin{gathered}\textbf{R (A-L-Conv)} = \forall i8 \in x8, i9 \in x9| \\ (i9[1])_{n_{\vee i9[1]\in n}}) := L-Conv((i8[1])_{n_{\vee i8[1]\in n}})\wedge \\ x9' := x9 \cup \{(i9[1])\}\end{gathered} \tag{7}$$

$$\begin{gathered}\textbf{R (Classification)} = \forall i10 \in x10, i11 \in x11, \\ i14 \in x14, i15 \in x15, i17 \in x17| \\ (i9[1])s_{\forall i9[1]\in s} := Mf(i8[1])s_{\forall i8[1]\in s} \\ \wedge x9' := x9 \cup \{(i9)\}\wedge \\ (i14[1])t_{\forall i14[1]\in t} := Rules(i10[1][\rho]\wedge \\ i10[\mu][\rho])t_{\forall i10[\mu][\rho]\in t}\wedge \\ \wedge x16' := x16 \cup \{(i16)\} \\ (i17[1])_{t_{\forall i17[1]\in t}}) := (i15[1])_{t_{\forall i15[1]\in t}}) \\ \wedge x17' := x17 \cup i17[1]\end{gathered} \tag{8}$$

$$\begin{gathered}\textbf{(Assignmnt)} = \forall i18 \in x18, i19 \in x19, \\ i20 \in x20, i21 \in x21, i21 \in x21| \\ ((i18[2] \in (i20[1]) = TRUE)) \rightarrow \\ (i21[1], i21[2], i21[3]) := Fuzzytable(i18[1] \\ \parallel i19[1] \parallel i20[1])\wedge \\ x21' := x21 \cup \{i21[1], i21[2], i21[3]\}\vee \\ (i18[2] \in (i20[1]) = TRUE) \rightarrow \\ (i22[1], i22[2]) := Fuzzy-table((i18[1] \\ \parallel i19[2])_{p_{\forall i19[2]\in p}})\wedge \\ x22' := x22 \cup \{i22[1], i22[2]\}\end{gathered} \tag{9}$$

For linguistic variables x, all membership functions are defined. Now, Ω^{α} rules are created based on the combination of linguistic variable values and membership functions, and they are saved in place Rules. After this procedure, each particular rule is allocated a privacy class. The above mentioned process is represented in equations 4, 5, 6, 7 and 8.

$$\begin{gathered}\textbf{R (Anonymization)} = \forall i23 \in x23, i24 \in x24, \\ i25 \in x25| \\ (i23[2] \in i24[2]_{i\forall i24[2]\in i}) = TRUE)) \rightarrow \\ i25[1] := i23[1] \wedge i25[2] := i23[2]\wedge \\ i25[3] := i24[2] \wedge i25[4] := i23[3] \\ \wedge x25' := x25 \cup \{i25[1], i25[2], i25[3], i25[4]\}\end{gathered} \tag{10}$$

Table 7
Classification of sensitive attributes
(a) Classification of sensitive attributes (disease-physician-treatment)

PID	Disease	Physician	Treatment	Class
P1, P2, P3	HIV, Flu, Cancer	John, Jem, Alice	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C1
P12	Stomach Cancer, Hepatitis, Obesity	John, Jem Alice	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C2
P12	Dyspepsia, Phthisis Digestive Upset, Asthma	John, Jem Alice	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C3
P10, P11	HIV, Flu, Cancer	Anas, Bob Eve	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C4
P5, P9	HIV, Flu, Cancer	Sarah, Suzan David, Steven	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C5
P7	Dyspepsia, Phthisis Digestive Upset, Asthma	Sarah, Suzan David, Steven	Antiretroviral therapy (ART), Medication, Radiation, Surgery	C6
P4, P13	HIV, Flu, Cancer	Anas, Bob Eve	Antibiotic, Chemotherapy, Drugs, Nutrition Control	C7
P5, P8	Stomach Cancer, Hepatitis, Obesity	Sarah, Suzan David, Steven	Antibiotic, Chemotherapy, Drugs, Nutrition Control	C8
P1, P5, P6	Dyspepsia, Phthisis Digestive Upset, Asthma	Sarah, Suzan David, Steven	Antibiotic, Chemotherapy, Drugs, Nutrition Control	C9

(b) Classification of sensitive attributes (symptom-diagnostic method)

PID	Symptoms	Diagnostic-method	Class
P1,P2,P5,P9,P10.P11	Infection, Fever, Loss of Weight Digestive Upset	Blood Test, RIDT test, ELISA-Test	pC1
P1, P3	Infection, Fever, Loss of Weight Digestive Upset	Ultrasound, Chest X-Ray, MRI Scan	pC2
P4, P11, P12	Difficulty in Breathing, Abdominal Pain, Eating Disorders, Heartburn	Ultrasound, Chest X-Ray, MRI Scan,	pC3
P5, P6, P12	Infection, Fever, Loss of Weight Digestive Upset	Molecular Diagnostic Methods, MCCT, Body Mass Index (BMI), Endoscopy	pC4
P7, P8	Difficulty in Breathing, Abdominal Pain, Eating Disorders, Heartburn	Molecular Diagnostic Test, MCCT, Body Mass Index (BMI), Endoscopy	pC5

Table 8
Fuzzy Quasi Table (FQT)

PID	Age	Zip code	Class
P1, P2, P3, P4	[24-27]	[14205-14249]	qPC1
P9, P12, P13	[28-35]	[14205-14249]	qPC2
P5, P6, P7, P8, P10	[36-45]	[13053-14204]	qPC3
P11	[36-45]	[14205-14249]	qPC4

In assignment process, each record in the table is checked to see which class it belongs to, followed by construction of Q-T and S-T as depicted in 8. In this process records from tables are assigned to each privacy class according to its matched values. Next in transition anonymization the privacy classes for quasi identifiers are checked for corresponding class of sensitive attributes bucket. Each class of multiple sensitive attributes gets a quasi-based privacy class, and saved in places FQT and FST as given in equations 9, 10, and 11.

$$\begin{gathered}\textbf{R (Publication)} = \forall i26 \in x26, i27 \in x27, i28 \in x28| \\ i27[1] := i26[1] \wedge i27[2] := i26[2] \wedge i27[3] := i26[4] \\ \wedge x27' := x27 \cup \{i27[1], i27[2], i27[3]\} \\ i28[1] := i26[1] \wedge i28[2] := i26[2] \wedge i28[3] := i26[4] \\ \wedge x28' := x28 \cup \{i28[1], i28[2]\}\end{gathered} \tag{11}$$

$$\begin{gathered}\textbf{R (FScor- Attack)} = \forall i28 \in x28, \forall i29 \in x29, \\ \forall i30 \in x30, \forall i31 \in x31| \\ fcorDis((i28[2] \cup i29[2]), 30[2]) \\ \neq i31[1] \wedge i31[1] = \varphi \\ \wedge x31 := x31 \cup \{i31[1]\}\end{gathered} \tag{12}$$

Table 9
Fuzzy Multiple Sensitive Attribute Table (FMSAT))

Class	Disease-Symptom-Physician			Treatment-Diagnostic	
qPC1	{HIV, Flu, Cancer} {Dyspepsia, Phthisis Digestive Upset, Asthma}	{John Jem Alice} {Anas Bob Eve} {Sarah Suzan David Steven}	{Antiretroviral Therapy (ART), Medication, Radiation, Surgery} {Antibiotic, Chemotherapy, Drugs, Nutrition Control}	{Infection, Fever, loss of weight Digestive Upset} {Difficulty in Breathing, Abdominal Pain, Eating Disorders, Heartburn}	{Blood Test, RIDT Test, ELISA-Test} {Ultrasound Chest x-ray, MRI Scan}
qPC2	{Stomach Cancer, Hepatitis, Obesity} {Dyspepsia, Phthisis, Digestive Upset, Asthma} {HIV, Flu, Cancer}	{John Jem Alice} {Sarah Suzan David Steven} {Anas Bob Eve}	{Antiretroviral Therapy (ART), Medication, Radiation, Surgery} {Antibiotic, Chemotherapy, Drugs, Nutrition Control}	{Infection, Fever, loss of weight, digestive upset} {Difficulty in Breathing, Abdominal Pain, Eating Disorders, Heartburn}	{Blood Test, RIDT test, ELISA-Test} {Ultrasound, Chest x-ray, MRI Scan} { Molecular Diagnostic Test, MCCT, Body Mass Index (BMI), Endoscopy}
qPC3	{HIV, Flu, Cancer} {Dyspepsia, Phthisis Digestive Upset, Asthma} {Stomach Cancer, Hepatitis, Obesity}	{John Jem Alice} {Anas Bob Eve} {Sarah Suzan David Steven}	{Antiretroviral therapy (ART), Medication, Radiation, Surgery} {Antibiotic, Chemotherapy, Drugs, Nutrition Control}	{Infection, Fever, Loss of Weight, Digestive Upset} {Difficulty in Breathing, Abdominal Pain, Eating Disorders, Heartburn}	{Blood Test, RIDT test, ELISA-Test} {Molecular Diagnostic Test, MCCT, Body Mass Index (BMI) Endoscopy}
qPC4	{HIV, Flu, Cancer}	{Anas Bob Eve}	{Antiretroviral therapy (ART), Medication, Radiation, Surgery}	{Infection, Fever, Loss of Weight, Digestive Upset}	{Blood Test, RIDT Test, ELISA-Test}

$$
\begin{aligned}
\textbf{R (FQcor- Attack)} &= \forall i32 \in x32, \forall i33 \\
&\in x33, \forall i34 \in x34, \forall i35 \in x35 | \\
&fqcorDis(i32[1], i32[2], i34[1]) \\
&= i35[1]) \notin i35[2] \wedge i35[2] = \varphi \\
&\wedge x35' := x35 \cup \{i35[1], i35[2]\}
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
\textbf{R (FNm-Attack)} &= \forall i36 \in x36, \forall i37 \in x37, \\
&\forall i38 \in x38, \forall i39 \in x39 | \\
&fNmDis(i36[1], i38[2]) \neq (39[2] \\
&\wedge 37[2]) \rightarrow 39[2] = \varphi \\
&\wedge x39 := x39 \cup \{i39[1], i39[2]\}
\end{aligned} \tag{14}
$$

Last attack transitions represent the attack mitigation on proposed Fuzz-classification (p,l)-Angel approach. The adversary uses the combination of published data FQT and FST, background knowledge BGK and external available information and tries to reveal the user IDs and MSA values. In 12, 13 and 14 the value returned from the transition Attack is equal to φ. As Fuzzification-based (p, l)-Angelization use classification and ROC based approach to prevent FScor-Attack and FQcor-Attack. An attacker cannot link multiple user ID records because the target identity in the PC cannot be traced. For the FScor-Attack invalidation in our proposed approach we take a unique attribute and assign it a rank according to criticality. The ROC-based method is used for weight calculation. This process is applied to all attributes. ROC-based classifications in PCs and permutation prove to be effective for any type of MSAs correlations and external available knowledge. Likewise, in an FNm attack, the adversary cannot guess the exact appearance of the target individual in the PC, because this knowledge is not sufficient due to the aforementioned fuzzy logic classification-based methods and the permutation of the target individual presence.

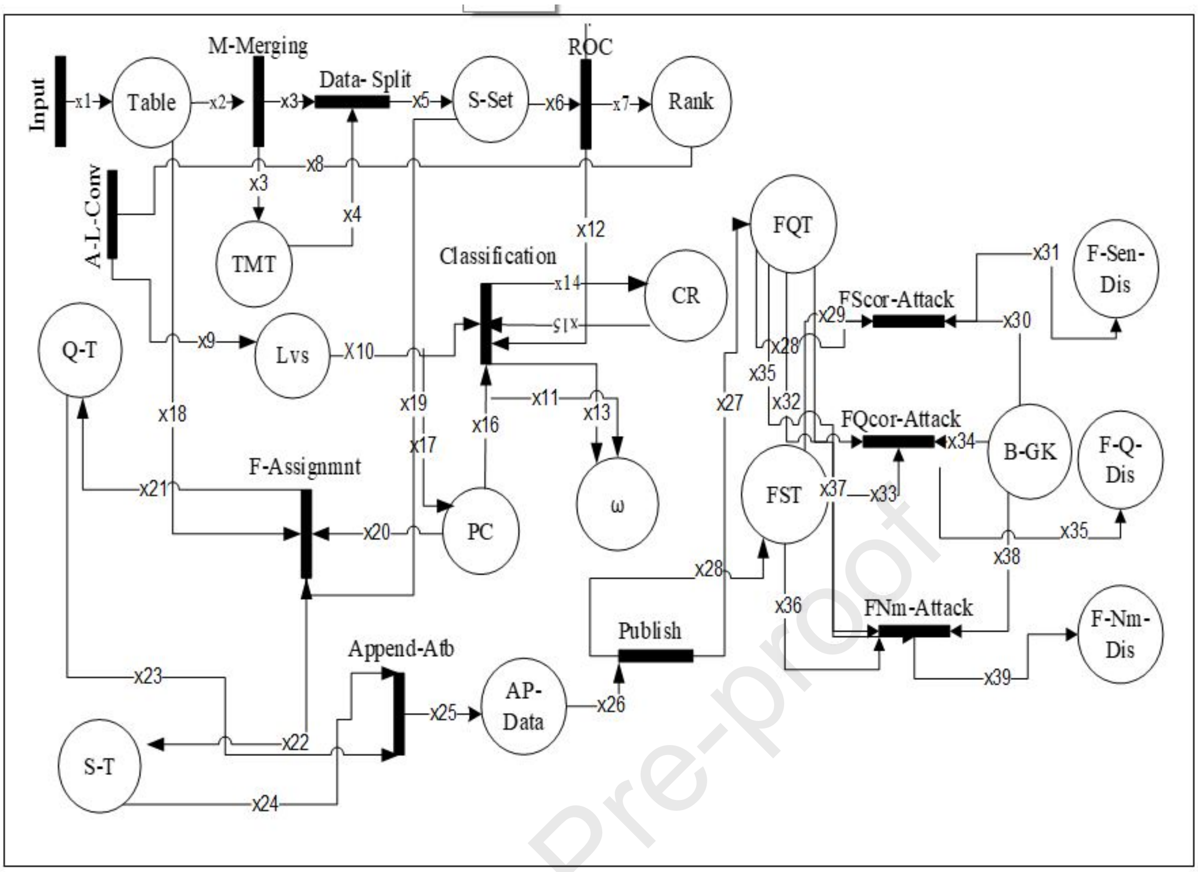

Fig. 2. HLPN for adversarial attacks mitigation on fuzz-classification (p,l)-Angel

6. Experimental results

The privacy model of the proposed Fuzz-classification (p, l)-Angel approach is validated through HLPN in the previous section. In this section, we present an experimental evaluation of the proposed approach with (p-l)-Angelization. The efficiency of the proposed privacy model is measured using computational cost and query accuracy. The proposed Fuzz- classification(p, l)-Angel algorithm, and (p, l)-Angelization is implemented in python, and experiments are carried out on a machine with operating system Windows 10, processor Intel Core i7, 500 GB hard disk, and 8 GB RAM. The data sets used are YOUTUBE and INFORMS. A total of 25000 records are used for INFORMS and YOUTUBE dataset. The total attributes in YOUTUBE dataset are 8, 6 are used as QIs uid, age, category, length, rate, and rating and 2 are SAs comments, video-ID. In INFORMS dataset 4 are QIs birth month and year, race, education year, and 2 are SAs income, code.

6.1. *Query accuracy:*

The utility of the proposed Fuzz -classification(p, l)-Angel is measured through query accuracy. The effectiveness of privacy models is obtained by comparing anonymized data sets aggregated query results [49, 50]. The aggregate query is in the form:

$$\begin{aligned}&SELECTCOUNT(*)FROMDATASET(T)\\&WHERE pred(P_1^q i) AND \dots AND\\&pred(P_q d^q i) AND pred(P^S A)\end{aligned} \tag{15}$$

In the above query, a query is executed from the original data set or anonymized data set and query predicate P comprises several QIs and SAs which we called the query dimensionality and values of each attributes called query selectivity. The Relative Query Error (RQE) is calculated using:

$$RQE = \frac{Estimated\ count\check{\ }Actual\ count}{Actual\ count} \tag{16}$$

Whereas, the actual query count is the result of the query run on the original data set T, and the estimated query count is the count returned from anonymize data set (T*). The results of the query accuracy are computed according to the number of groups and the dimensionality of the query. In Fig. 3 and Fig. 4, the number of groups and relative query error are shown for YOUTUBE and INFORMS data sets. The group size fluctuates in Fuzz-classification (p, l)-Angel, unlike (p, l)-Angelization, as a result, the relative query error in the proposed methodology is almost the same for varied group sizes. Furthermore, when compared to (p, l)-Angelization, the proposed approach uses fuzzy classification for both QIs and SAs, re-

Table 10
Data types used in HLPN for Fuzzy-PPDP

Types	Descriptions
ID	Identifier in Table
TID	Transformed identifier in Table
Tp_t	Table with t tuples
TTp_t	Transformed Table with t tuples
Sqid	Subset of quasi-identifier
$Ssen_i$	Multiple subsets of sensitive attribute values
Sa_R	Subsets of unique sensitive attribute ranking
C_{pc}	Set of privacy classes
C_q	Set of privacy classes for quasi-identifiers
C	Set of privacy classes for append table
FQI	Fuzzy quasi-identifiers
Lv_s	Linguistic variables for a attributes
R_ρ	ρ number of fuzzy rules
Q	Group of quasi identifiers
$FMSA_s$	Multiple sensitive attributes
MSAT	Fuzzy multiple sensitive attribute
FQ_bk	Fuzzy quasi-identifiers for background knowledge
$FMSA_{bk}$	Fuzzy multiple sensitive attribute of background knowledge
FQI_{cor}	Fuzzy quasi-identifiers for Qcor based adversarial disclosure
$FMSA_{cor}$	Fuzzy quasi-identifiers for Qcor based adversarial disclosure
FQI_{nm}	Multiple sensitive attributes for correlation adversarial disclosure
$FMSA_{nmm}$	Multiple sensitive attributes for non-membership disclosure.

sulting in a lower relative query error. Although better generalization is implemented for QIs in (p, l)-Angelization, query error is still greater as compared to Fuzz-classification (p, l)-Angel. In Fig. 5 and 6 relative query error is plotted against query dimensionality for YOUTUBE and INFORMS data sets. Graphs suggest that query accuracy is better in the proposed approach as compared to the renowned (p, l)-Angelization technique that uses angelization based generalization for QIs.

6.2. *Execution time:*

The execution time analysis is used to measure the computational efficiency of the proposed approach against (p, l)-Angelization. The execution time analysis of Fuzz -classification(p, l)-Angel and (p, l)-Angelization is shown in Fig. 7 and 8. In Fig. 7 execution time is plotted against the number of records in the YOUTUBE dataset. It is clear that time required to execute the proposed approach is quite short

Table 11
Mapping of data types on places

Types	Description
φ(Table)	$\mathbb{P}(ID \times Tp_t)$
φ(TMT)	$\mathbb{P}(TID \times TTp_t)$
φ(S-Set)	$\mathbb{P}(Sqid \times Ssen_i)$
φ(Rank)	$\mathbb{P}(Sqid \times Sa_R)$
φ(PCs)	$\mathbb{P}(C_{pc})$
φ(L-Variable)	$\mathbb{P}(LV_s)$
φ(MF)	$\mathbb{P}(mf_i)$
φ(CR)	$\mathbb{P}(R_\rho)$
φ(Q-T)	$\mathbb{P}(TID \times FQ \times C_q)$
φ(S-T)	$\mathbb{P}(TID \times FMSA_i)$
φ(AP-Data)	$\mathbb{P}(TID \times Q \times FMSA_s \times C)$
φ(FQT)	$\mathbb{P}(TID \times FQ \times C_q)$
φ(FST)	$\mathbb{P}(TID \times FMSAs \times C_s)$
φ(BGK)	$\mathbb{P}(FQI_{bk} \times FMSA_{bk})$
φ(F-Sen-Dis)	$\mathbb{P}(FMSA_{cor})$
φ(F-Qcor-Dis)	$\mathbb{P}(TID \times FQI_{cor})$
φ(F-Nm-Dis)	$\mathbb{P}(FQI_{nm} \times FMSA_{nmm})$

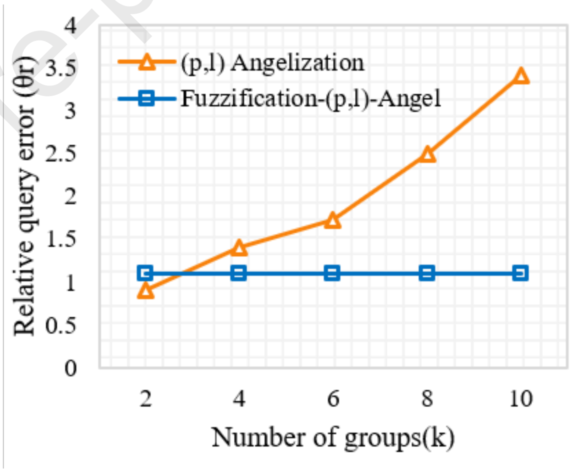

Fig. 3. Relative query error for varying k groups on YOUTUBE data set

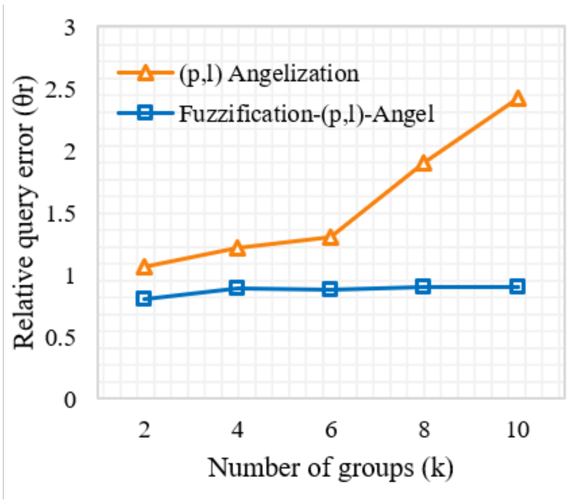

Fig. 4. Relative query error for varying k groups on INFORMS data set

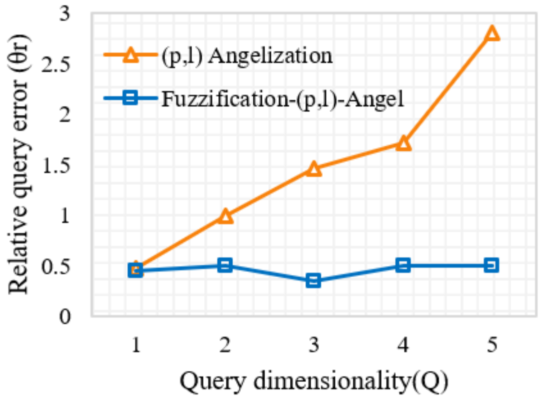

Fig. 5. Relative query error for different query dimensions on YOUTUBE data set

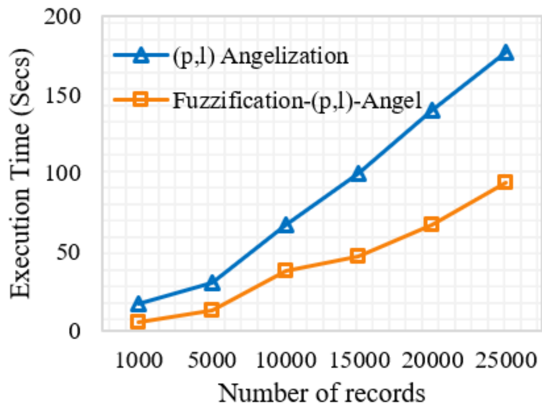

Fig. 7. Execution time analysis for different number of records on YOUTUBE data set

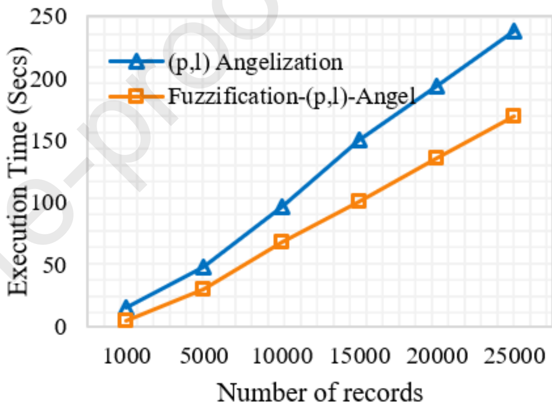

Fig. 8. Execution time analysis for different number of records on INFORMS data set

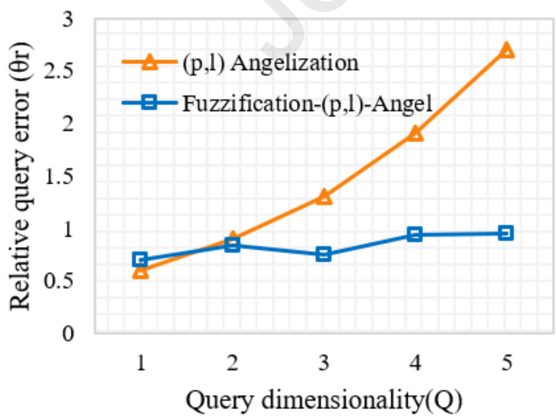

Fig. 6. Relative query error for different query dimensions on INFORMS data set

as compared to (p, l)-Angelization. The proposed method is AI-based and uses fuzzy logic to classify QI and SA, so it is computationally faster than (p, l)-Angelization. In Fig. 8 execution time of INFORMS dataset with varying records is shown. The results show that the execution time increases with the number of records, but the proposed method with different records has very short execution time compared to (p, l)-Angelization. The reason for the longer execution time in (p, l)-Angelization is due to generalization using QI anonymization, using weight assignment and balancing steps to discover MSA dependencies.

7. Conclusion

Privacy and utility are factors that are mutually dependent on privacy-preserving data publishing, it is crucial to design effective privacy techniques for data publication that strike a balance between the two. The main objective of this research is to maximize the utility of health care data while protecting the privacy of multiple sensitive attributes and multiple records data sets. The work in this area is very limited and is still not well explored. State-of-the-art privacy work for multiple sensitive attributes requires reliable techniques like AI-based fuzzy logic. In this article, we propose an enhanced version of our previously proposed technique (p, l)-Angelization. The proposed Fuzz-classification (p, l)-Angel uses permutation and fuzzy logic to classify multiple sensitive attributes and quasi-identifiers. Privacy Classes are mapped to MSAs data sets using specific fuzzy rules. The suggested technique is an improved version of (p, l)-Angelization in terms of privacy, as shown through modelling and analysis of privacy disclosures using HLPN. The experiments' results show that the suggested approach outperforms its counterpart in terms of utility and performance. In the future, we plan to develop secure smart homes and transportation systems based on fuzzy logic, and we will look into their privacy and reliability. There is also a need for privacy-aware federated learning-based mechanisms re-investigation and their use for secure sensitive health data collection and transmission.

References

[1] A. Zigomitros, F. Casino, A. Solanas, C. Patsakis, A survey on privacy properties for data publishing of relational data, IEEE Access 8 (2020) 51071–51099.

[2] B. C. Fung, K. Wang, R. Chen, P. S. Yu, Privacy-preserving data publishing: A survey of recent developments, ACM Computing Surveys (Csur) 42 (4) (2010) 1–53.

[3] L. Rocher, J. M. Hendrickx, Y.-A. De Montjoye, Estimating the success of re-identifications in incomplete datasets using generative models, Nature communications 10 (1) (2019) 1–9.

[4] L. Sweeney, k-anonymity: A model for protecting privacy, International journal of uncertainty, fuzziness and knowledge-based systems 10 (05) (2002) 557–570.

[5] X. Huang, J. Liu, Z. Han, J. Yang, A new anonymity model for privacy-preserving data publishing, China Communications 11 (9) (2014) 47–59.

[6] T. M. Truta, B. Vinay, Privacy protection: p-sensitive k-anonymity property, in: 22nd International Conference on Data Engineering Workshops (ICDEW'06), IEEE, 2006, pp. 94–94.

[7] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, l-diversity: Privacy beyond k-anonymity, ACM Transactions on Knowledge Discovery from Data (TKDD) 1 (1) (2007) 3–es.

[8] N. Li, T. Li, S. Venkatasubramanian, t-closeness: Privacy beyond k-anonymity and l-diversity, in: 2007 IEEE 23rd international conference on data engineering, IEEE, 2006, pp. 106–115.

[9] X. Xiao, Y. Tao, Anatomy: Simple and effective privacy preservation, in: Proceedings of the 32nd international conference on Very large data bases, 2006, pp. 139–150.

[10] S. Kiruthika, M. M. Raseen, Enhanced slicing models for preserving privacy in data publication, in: 2013 International Conference on Current Trends in Engineering and Technology (ICCTET), IEEE, 2013, pp. 406–409.

[11] T. Li, N. Li, J. Zhang, I. Molloy, Slicing: A new approach for privacy preserving data publishing, IEEE transactions on knowledge and data engineering 24 (3) (2010) 561–574.

[12] F. Luo, J. Han, J. Lu, H. Peng, Angelms: a privacy preserving data publishing framework for microdata with multiple sensitive attributes, in: 2013 IEEE Third International Conference on Information Science and Technology (ICIST), IEEE, 2013, pp. 393–398.

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

[14] X. Sun, L. Sun, H. Wang, Extended k-anonymity models against sensitive attribute disclosure, Computer Communications 34 (4) (2011) 526–535.

[15] H. Zhu, S. Tian, M. Xie, M. Yang, Preserving privacy for sensitive values of individuals in data publishing based on a new additive noise approach, in: 2014 23rd International Conference on Computer Communication and Networks (ICCCN), IEEE, 2014, pp. 1–6.

[16] Y. Ye, Y. Liu, C. Wang, D. Lv, J. Feng, Decomposition: privacy preservation for multiple sensitive attributes, in: International Conference on Database Systems for Advanced Applications, Springer, 2009, pp. 486–490.

[17] T. Kanwal, S. A. A. Shaukat, A. Anjum, K.-K. R. Choo, A. Khan, N. Ahmad, M. Ahmad, S. U. Khan, et al., Privacy-preserving model and generalization correlation attacks for 1: M data with multiple sensitive attributes, Information Sciences 488 (2019) 238–256.

[18] T. Kanwal, A. Anjum, S. U. Malik, H. Sajjad, A. Khan, U. Manzoor, A. Asheralieva, A robust privacy preserving approach for electronic health records using multiple dataset with multiple sensitive attributes, Computers & Security 105 (2021) 102224.

[19] G. Klir, B. Yuan, Fuzzy sets and fuzzy logic, Vol. 4, Prentice hall New Jersey, 1995.

[20] V. V. Kumari, S. S. Rao, K. Raju, K. Ramana, B. Avadhani, Fuzzy based approach for privacy preserving publication of data, International Journal of Computer Science and Network Security 8 (1) (2008) 115–121.

[21] P. Kumar, K. I. Varma, A. Sureka, et al., Fuzzy based clustering algorithm for privacy preserving data mining, International Journal of Business Information Systems 7 (1) (2011) 27.

[22] S. U. Malik, S. U. Khan, S. K. Srinivasan, Modeling and analysis of state-of-the-art vm-based cloud management platforms, IEEE Transactions on Cloud Computing 1 (1) (2013) 1–1.

[23] S. U. Malik, K. Bilal, S. U. Khan, B. Veeravalli, K. Li, A. Y. Zomaya, Modeling and analysis of the thermal properties exhibited by cyberphysical data centers, IEEE Systems Journal 11 (1) (2015) 163–172.

[24] D. Das, D. K. Bhattacharyya, Decomposition+: improving l-diversity for multiple sensitive attributes, in: International Conference on Computer Science and Information Technology, Springer, 2012, pp. 403–412.

[25] J. Han, F. Luo, J. Lu, H. Peng, Sloms: A privacy preserving data publishing method for multiple sensitive attributes microdata., J. Softw. 8 (12) (2013) 3096–3104.

[26] V. S. Susan, T. Christopher, Anatomisation with slicing: a new privacy preservation approach for multiple sensitive attributes, SpringerPlus 5 (1) (2016) 1–21.

[27] Q. Liu, H. Shen, Y. Sang, Privacy-preserving data publishing for multiple numerical sensitive attributes, Tsinghua Science and Technology 20 (3) (2015) 246–254.

[28] T. Yi, M. Shi, Privacy protection method for multiple sensitive attributes based on strong rule, Mathematical Problems in Engineering 2015.

[29] J. Liu, J. Luo, J. Z. Huang, Rating: privacy preservation for multiple attributes with different sensitivity requirements, in: 2011 IEEE 11th International conference on data mining workshops, IEEE, 2011, pp. 666–673.

[30] A. Anjum, N. Ahmad, S. U. Malik, S. Zubair, B. Shahzad, An efficient approach for publishing microdata for multiple sensitive attributes, The Journal of Supercomputing 74 (10) (2018) 5127–5155.

[31] S. Onashoga, B. Bamiro, A. Akinwale, J. Oguntuase, Kc-slice: A dynamic privacy-preserving data publishing technique for multisensitive attributes, Information Security Journal: A Global Perspective 26 (3) (2017) 121–135.

[32] N. L. Raju, M. Seetaramanath, P. S. Rao, A novel dynamic kci-slice publishing prototype for retaining privacy and utility of multiple sensitive attributes, International Journal of Informa-

tion Technology and Computer Science 11 (4) (2019) 18–32.
[33] Y. Xiao, H. Li, Privacy preserving data publishing for multiple sensitive attributes based on security level, Information 11 (3) (2020) 166.
[34] R. Khan, X. Tao, A. Anjum, H. Sajjad, A. Khan, F. Amiri, et al., Privacy preserving for multiple sensitive attributes against fingerprint correlation attack satisfying c-diversity, Wireless Communications and Mobile Computing 2020.
[35] J. Andrew, J. Karthikeyan, Privacy-preserving big data publication:(k, l) anonymity, in: Intelligence in Big Data Technologies—Beyond the Hype, Springer, 2021, pp. 77–88.
[36] Q. Gong, J. Luo, M. Yang, W. Ni, X.-B. Li, Anonymizing 1: M microdata with high utility, Knowledge-based systems 115 (2017) 15–26.
[37] A. Anjum, N. Farooq, S. U. R. Malik, A. Khan, M. Ahmed, M. Gohar, An effective privacy preserving mechanism for 1: M microdata with high utility, Sustainable cities and society 45 (2019) 213.
[38] T. Kanwal, A. Anjum, A. Khan, A. Asheralieva, G. Jeon, A formal adversarial perspective: Secure and efficient electronic health records collection scheme for multi-records datasets, Transactions on Emerging Telecommunications Technologies 32 (8) (2021) e4180.
[39] X. Li, Z. Zhou, A generalization model for multi-record privacy preservation, Journal of Ambient Intelligence and Humanized Computing 11 (7) (2020) 2899–2912.
[40] K. Albulayhi, P. T. Tošić, F. T. Sheldon, G-model: a novel approach to privacy-preserving 1: M microdata publication, in: 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), IEEE, 2020, pp. 88–99.
[41] R. Mukkamala, V. G. Ashok, Fuzzy-based methods for privacy-preserving data mining, in: 2011 Eighth International Conference on Information Technology: New Generations, IEEE, 2011, pp. 348–353.
[42] T. Jahan, G. Narasimha, C. Rao, A comparative study of data perturbation using fuzzy logic to preserve privacy, in: Networks and Communications (NetCom2013), Springer, 2014, pp. 161–170.
[43] M. Sridhar, B. R. Babu, A fuzzy approach for privacy preserving in data mining, International Journal of Computer Applications 57 (18).
[44] G.-r. Zhang, Privacy data preserving method based on fuzzy discretization, in: 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Vol. 2, IEEE, 2011, pp. 1201–1205.
[45] S. Ahmed, S. Haque, S. F. Tauhid, A fuzzy based approach for privacy preserving clustering, Int. J. Sci. Eng. Res 5 (2) (2014) 1067–1071.
[46] J. Wang, G. Cai, C. Liu, J. Wu, X. Li, A multi-level privacy-preserving approach to hierarchical data based on fuzzy set theory, Symmetry 10 (8) (2018) 333.
[47] T. Jahan, K. Pavani, G. Narsimha, A. Rao CG., Data perturbation method to preserve privacy using fuzzy rules, in: Proceedings of the Second International Conference on Computational Intelligence and Informatics. , Singapore, 2018, pp. 9–16.
[48] H. Attaullah, A. Anjum, T. Kanwal, S. U. R. Malik, A. Asheralieva, H. Malik, A. Zoha, K. Arshad, M. A. Imran, F-classify: Fuzzy rule based classification method for privacy preservation of multiple sensitive attributes, Sensors 21 (14) (2021) 4933.
[49] X. Xiao, Y. Tao, Dynamic anonymization: accurate statistical analysis with privacy preservation, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, 107–120, 2008.
[50] R. C.-W. Wong, Y. Liu, J. Yin, Z. Huang, A. W.-C. Fu, J. Pei, (α, k)-anonymity based privacy preservation by lossy join, in: Advances in Data and Web Management, Springer, 2007, pp. 733–744.

Declaration of Interest Statement

Title:

Fuzz-Classification (p, l)-Angel: An enhanced Hybrid Artificial Intelligence based Fuzzy logic for Multiple Sensitive Attributes against Privacy Breaches

Authors:

Tehsin Kanwal[1], Hasina Attaullah[1], Adeel Anjum[1], Abid Khan[2,] Gwanggil Jeon[3]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan,
[2]Department of Computer Science, Aberystwyth University, SY23 3DB, Aberystwyth, United Kingdom.
[3]Department of Embedded Systems Engineering, Incheon National University, 119 Academy-ro, Yeonsugu, Incheon, 22012, South Korea.

Conflict of Interest

None Declared.