

Contents

1	Introduction	5
1.1	State of The Art	6
1.2	Motivation	6
1.3	Aim of the Work	7
1.3.1	Objectives	7
1.4	Contributions	8
1.5	Report Organization	8
2	Basics	10
2.1	Privacy Preserving Data Mining (PPDM)	10
2.1.1	Data Availability	11
2.2	Privacy Preserving Data Publishing(PPDP)	12
2.3	Cellular Automata(CA)	12
2.4	Secret Sharing	14
2.5	Discussion	15
3	Related works	16
3.1	Works in PPDM	16
3.1.1	Methods of PPDM for Centralized Data	16
3.1.2	Methods of PPDM for Distributed Data	19
3.2	Works in PPDP	20
3.2.1	Taxonomy of PPDP	21
3.2.2	Works in Randomization	22
3.2.3	Works in Anonymization	26
3.2.4	Works for distributed environment	30
3.3	Works in Secret Sharing	31

3.4	Discussion	34
4	Proposed Method	36
4.1	Protocol	37
4.1.1	Data collection and Share generation	37
4.1.2	Original data regeneration	38
5	Security Analysis	41
5.1	Attacks while being transferred	41
5.2	Stealing original data from T	42
5.3	Attacks on published data	42
5.3.1	Homogeneity Attack	42
5.3.2	Background knowledge Attack	42
5.3.3	Similarity attack	43
6	Conclusion and future work	44

List of Figures

2.1	Data Availability	11
3.1	Taxonomy for PPDP Methods	24
4.1	Data Collection	38
4.2	Secret Share Generation	39
4.3	Original Data Regeneration	39

List of Tables

3.1	Prior works in clustering of distributed data	21
3.2	Prior works in Association Rule Mining of distributed data	22
3.3	Prior works in Classification of distributed data	23
3.4	Original data ready to be published	27
3.5	A 3-Anonymous Data	27
3.6	A 3-Diverse Data	28
3.7	Sliced data ready for anonymization	30
3.8	Comparison of our survey with other surveys	33
4.1	Symbols used and their meaning	37