

Contents

List of Figures	v
List of Tables	vii
Glossary	ix
1 Introduction	1
1.1 Motivation	2
1.2 Contribution	4
1.3 Report Organization	4
2 Background	5
2.1 General Principles	5
2.2 Privacy Models	6
2.2.1 k -anonymity	6
2.2.2 ℓ -diversity	7
2.3 From SSA towards the MSA case	7
2.4 Discussion	9
2.5 ℓ -diversity in the MSA case	9
3 Decomposition	11
3.1 Partitioning	11
3.2 Partitioning in Decomposition	12
3.3 Addition of noise	13
3.4 Discussion	16

CONTENTS

4	Decomposition+	17
4.1	Extending Decomposition to the continuous release scenario	17
4.1.1	Multiple Publishing: A practical necessity of anonymization techniques	17
4.1.2	Continuous Release in Decomposition+	18
4.2	Improving the noise selection procedure	20
4.3	Algorithm for Decomposition+	21
4.4	Discussion	23
5	Experiments	25
5.1	Occurrence of residual tuples	26
5.1.1	Discussion	26
5.2	Performance	27
5.2.1	Discussion	27
6	Conclusion and Further Work	29
7	Appendix	31
	Bibliography	37

List of Figures

1.1	A typical Anonymization Model	3
3.1	The Decomposition Data Model	12
5.1	ℓ_{pri} versus performance (in CPU seconds)	28

LIST OF FIGURES

List of Tables

2.1	General Scheme of a data to be anonymized	6
3.1	The Microdata table	14
3.2	QIDs and non-sensitive attributes of Table 3.1 after Decomposition	14
3.3	Sensitive attributes of Table 3.1 after Decomposition	15
3.4	Sensitive attributes in their original form provided as Marginals	15
4.1	Sensitive attributes of Table 3.1 without addition of noise	21
5.1	Description of Attributes	26
5.2	Residual tuples in each group for different values of ℓ_{pri}	27

GLOSSARY

//////// //////////////

Glossary

MSA Multiple Sensitive Attribute the scenario where a dataset contains two or more sensitive attribute. See SSA

SSA Single Sensitive Attribute the scenario where a dataset contains one sensitive attribute. See MSA

PPDM Privacy Preserving Data Mining the study of data mining methods which take into account privacy related issues

PPDP Privacy Preserving Data Publishing the study of methods which publish data sets while reducing its privacy disclosure risk

Microdata Data at the level of individual records representing each entity. Term comes from the census community

PSA Primary Sensitive Attribute the attribute on which decomposition is done