

DIT IS DE TITEL VAN MIJN AFSTUDEERVERSLAG

by

Tjibbe van der Ende

in partial fulfillment of the requirements for the degree of

Master of Science
in Software Engineering

at the Open University, faculty of Management, Science and Technology

Master Software Engineering

to be defended publicly on Day Month DD, YYYY at HH:00 PM.

Student number: student number

Course code: IMA0002

Thesis committee: titles and name of the chairman (chairman), Open University
titles and name of the supervisor (supervisor), Open University

CONTENTS

1	Introduction	1
1.1	Research questions	1
2	Theoretical Framework	2
2.1	Differential privacy	2
2.1.1	Laplace algorithm.	2
2.1.2	2D-Laplace	2
2.1.3	3D-Laplace	2
2.2	Clustering.	2
2.2.1	Methods	2
2.2.2	Evaluation	2
3	Methodology	3
3.1	Datasets	3
3.2	Environmental setup	3
3.3	Experiment setup	4
3.3.1	Research question 1	4
3.3.2	Research question 2	4
3.3.3	Research question 3	4
3.4	Results	4
3.4.1	Research question 1	4
3.4.2	Research question 2	4
3.4.3	Research question 3	4
	Bibliography	i

1

INTRODUCTION

This is just to show how to include a tex file for a chapter, with a reference [Dijkstra, 1968].

1.1. RESEARCH QUESTIONS

Main question:

How can the n D-Laplace algorithm be applied in training privacy-preserving clustering algorithms on distributed n -dimensional data?

1. RQ1: How can 2D-Laplace be used to protect the data privacy of 2-dimensional data which is employed for training clustering algorithms?
2. RQ2: How can 3D-Laplace be extended to protect the data privacy of n -dimensional data which is employed for training clustering algorithms?
3. RQ3: What is the impact of different privacy budgets, dataset properties, and other clustering algorithms on the research conducted for research question 2?

2

THEORETICAL FRAMEWORK

2.1. DIFFERENTIAL PRIVACY

2.1.1. LAPLACE ALGORITHM

2.1.2. 2D-LAPLACE

2.1.3. 3D-LAPLACE

2.2. CLUSTERING

2.2.1. METHODS

2.2.2. EVALUATION

3

METHODOLOGY

To gain insights into the proposed methods for researching the appliance of (ND)-Laplace for cluster algorithms we conducted experiments. The experiment results are used to evaluate our method against other literature. In this chapter we explain:

1. Datasets
2. Environmental setup.
3. For each research question: Description of the different experiments.
4. For each research question: Results.

3.1. DATASETS

For this research, we will use a synthetic dataset for all three research questions.

Records	Centers	Dimensions	Standard deviation	Research
200	4	2	0.60	RQ 1

Research question 3 uses a "real-world" dataset to properly assess the different dataset properties that are the subject of this research question.

Describe datasets

3.2. ENVIRONMENTAL SETUP

Describe the exact environment details

3.3. EXPERIMENT SETUP

3.3.1. RESEARCH QUESTION 1

3.3.2. RESEARCH QUESTION 2

3.3.3. RESEARCH QUESTION 3

3.4. RESULTS

3.4.1. RESEARCH QUESTION 1

3.4.2. RESEARCH QUESTION 2

3.4.3. RESEARCH QUESTION 3

BIBLIOGRAPHY

Edsger W Dijkstra. Go to statement considered harmful. *Communications of the ACM*, 11 (3):147–148, 1968. 1