

**ANALYSIS OF RANDOM EQUIVALENCE CLASS
SELECTION ALGORITHMS FOR K-ANONYMITY AND
ANOTHER $O(K)$ -APPROXIMATION OF OPTIMAL
K-ANONYMITY BY CELL SUPPRESSION**

By

David M. Cerna

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: **COMPUTER SCIENCE**

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Approved:

Professor Bülent Yener, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

July 2010
(For Graduation August 2010)

In this thesis, we statistically analyze random equivalence class selection algorithms under different constraints and information loss metrics. This is done in an attempt to heuristically improve the algorithms found in [1,2] by adding a mixing constraint found in each algorithm or designed by us. We test the algorithms on six different distributions and also take 16 different statistics on the resulting k-anonymous databases. As well as analyzing the algorithms, we test our information loss metric based on information entropy. We also present another $O(k)$ -approximation algorithm for cell suppression K-anonymity based on the work done in [3].