Helwan University
Faculty of Computers and Information
Information Systems Department

"Securing Big Data using Negative Databases"

Thesis Submitted For Partial Fulfillment of the Requirements for the Master Degree in Computers and Information, Information Systems Department, Faculty of Computers and Information, Helwan University

Submitted by

Azza Ahmed Mohammed Ahmed

Under Supervision of

Prof. Sayed Abdel Gaber
Information Systems Department,
Faculty of Computers and Information,
Helwan University

Dr. Hanan Fahmy
Information Systems Department,
Faculty of Computers and Information
Helwan University

Dr. Mohammed Hassan
Faculty of Computers and Information,
Fayoum University

2018

Acknowledgement

I can not express my thanks to ALLAH for every thing ALLAH gives to me beginning from being a Muslim person.

I am grateful to a number of people, which supported me to carry out this work. I would like to thank my thesis co-advisers; Prof. Dr. Sayed Abdel Gaber, Dr. Hanan Fahmy for her guidance, patience, and encouragement, Dr. Mohammed Hassan and Eng. Hussien Shahata they have given me during the work in this thesis.

Finally, a heart- felt “Thanks” to my wonderful Family for their pray, unconditional support and understanding, and for always being there for me, my dearest friend Asmaa for always being there for me.

Abstract

Security related issues have many threats to the original data in any organization. There are users who always try to get into the internal information system and the data systems. There are organizations such as credit card companies, government agencies and security agencies that need their data secured to the highest extent. Hence, such organizations want their applications to provide high security. There are numerous ways to keep data secure but there is no way to secure it 100%. Anywhere, because of the different kinds of attacks that could be vulnerable at any time and the consequences that may be worse than expected. Therefore, no one knows the types and the severity of such attacks.

This thesis presents a new framework that addresses the problem of big data security. The proposed framework consists of two main phases, which are: NDG (Negative Data Generator) and NQC (Negative Query Converter). The first phase explains how to generate a Negative Data; this phase consists of big dataset, mapper, mapper result**,** NDG, and Negative Data. The second phase shows how to get a positive query on a Negative Data and retrieve the positive result of the query as the Positive Data; this phase consists of authority model and NQC.

The experiments that were conducted on the proposed framework are shown that the generator is capable of generating automatic data named ND from the original data (big data) in a low time and high processing and the malicious users can't get any useful information from ND. The increasing in the volume of big data leads to a decreasing in the volume of ND. Therefore, applying the concept of negative databases to the distributed

environment rather than just use it in the traditional environment to overcome the shortcomings of the traditional environment. This thesis consider the first research that presents a NQC that helps in dealing with negative data such as positive data, the experiments were provided the correctness of the NQC , and the importance of NQC that helps the user to extract data efficiently and easily from negative database.

List of Abbreviations

3Vs	Volume, ,Velocity, Varity
4Vs	Volume, ,Velocity, Varity, Value
ABAC	Attribute Based Access Control
ACLs	Access Control Lists
AM	Authority Model
ANN	Artificial Neural Network
CNF-SAT	Conjunctive Normal Form- Satifiability Problem
CRM	Customer Relationship Management
DB	Data Base
DPLL	Davis-Putnam-Logemann-Loveland
ERM	Environmental Resources Management
ETL	Extract, Transform, and Load
FHE	Fully Homomorphic Encryption
GFS	Google File System
HDFS	Hadoop Distributed File System

HE	Homomorphic Encryption
IDC	International Data Corporation
IDS	Intrusion Detection Systems
IoT	Internet of things
IPS	intrusion protection systems
MPC	Multi-Party Computation
MR	MapReduce
ND	Negative Data
NDBs	Negative Databases
NDG	Negative Data Generator
NoSQL	Not Only Structure Query Language
NQC	Negative Query Converter
OLAP	Online Analytical Processing
PD	Positive Data
RDBMS	Relational Database Management Systems
RNDB	Randomize Negative Database
SAT	Satisfiability Problem
SNA	Social Network Analysis
SQL	Structure Query Language

SRP	Secure Remote Password Protocol
U	Universal
USB	Universal Serial Bus
VC	Verifiable Computation

Table of Contents

List of Figures

List of Tables