DEPARTMENT OF INFORMATION SYSTEMS
FACULTY OF COMPUTING & INFORMATION SCIENCE
AIN SHAMS UNIVERSITY

Usage of Intelligent Data Mining Methodology in Cyber Security

A Thesis Submitted to Faculty of Computer and Information Sciences,
Ain Shams University, Cairo, Egypt

In Partial Fulfillment of the Requirements For
The Degree of Doctor of Philosophy in Computer and Information Sciences

By

Hanaa Mohammad Said Ibrahim

Master of Science in Information Systems 2011
Arab Academy for Science, Technology & Maritime Transport
IT General Manger
Cairo Cleaning Beautification Authority, Cairo, Egypt

Supervised By

Prof. Dr. Abdel-Badeeh M. Salem
Professor of Computer Science
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Dr. Rania Abdel Rhman Elgohary
Associated Professor, Department of Information Systems
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Dr. Mohammad Hamedy Elalamey
Lecturer, Department of Information Systems
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Cairo – December 2015

Acknowledgment

First and foremost, I humbly give my deep thanks to Allah for granting me the opportunity and the strength to accomplish this work.

I would like to thank all people who helped me to make this work possible. I would like to give my deep appreciation and thanks to my super visor, **Prof. Dr. Abdel-Badeeh M. Salem**, for the confidence that he always had in me, for his wisdom, intelligence and creative vision. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. He always gave me the precise advice at the precise moment so that I had both the freedom and support that I needed to do my research. Also, in this respect, I can't deny all made efforts done by my great and respectful above mentioned genius Doctor, his impatience and his providing me all required information, encouragement and valuable wisdoms. I announce publicly that I am much proud to present this thesis expressing the best symbol and ideal viewpoint of our science, development, advancement and modern technology which our great Doctor is considered as huge and great academy for all of these trends.

I would like also to thank Prof. Dr. Rania Elgohary and, Dr. Mohammad Hamedy Elalamey for their valuable and helpful Guidance in reviewing and directing the work presented in this research.

I would like, my love and gratitude goes to my husband who had taught me everything, and sacrificed a lot for me. I'll always be grateful for all made efforts by my son and daughters whom taught me love and compassion and also for their kind efforts that they made for economizing support, tranquility, calmness and encouragement to complete this thesis.

Finally, I would like to thank and appreciate the whole members of the great Authority for all what they supported me practically and theoretically as well as applications of all programs which I used in my thesis and all other different concerned departments that provided me all required information, descriptive statements and details.

To all of them, thank you.

Dedication

This thesis is dedicated to …………

Both my parents…..

My family, *this thesis is dedicated to my husband, my daughters and my son who have been a great source of motivation and inspiration*

My professors, *to my professors, my instructors and my supervisors with great love and respect, I dedicate my thesis for them*

Finally, *this thesis is dedicated to all those who believe in the richness of learning………*

Abstract

With the ever growing technology, its advantages and disadvantages are increasing; computer related crime is on the rise too. Technology is producing several negative impacts on society Internet hacking is worth noting. Cyber Security is the most serious issue around the world. Organizations wishing to ensure security of their systems may look towards adopting appropriate tests to protect themselves against potential security breaches. Cyberspace is known as the supposed space" it is the material space and the non- material space. It is consists of parts of all the following elements: computes; machinery device networks, and computerized information programs. It is known as the digital electronic medium for the knowing range of securing in the cyberspace with the available resources.

E-Government must ensure that; information systems are appropriately protected and individual rights are respected. The successful e-government project builds trust with any online service, Security is one of the most important issues that face the use of online services, also the Governments must be responsible custodians of the enormous amounts of personal information they hold. Security must be addressed in the phase of planning and designing of the e-government system. Management process is needed to assess security control, this management allows departments and agencies to maintain and measure the extent of data security depending on the mechanism of revealing the security weak points.

Data Mining is the process of automatically searching large volumes of data for patterns using association rules, for evaluating security threats related to the detection of cyber-attacks, moreover cybercrime, and information security. This thesis presents the analysis, studies for securing one of the minor cyber space's which the cyberspace of the authority of cleaning is and beautifying Cairo, Egypt (www.ccba.gov.eg), It is one of the important cyberspaces that provides -government services. Also we are testing Cyber space security provided by e-government systems through "suggested model (MADAM ID)" and securing the data in e-government systems. The proposed model, MADAM ID, has been used for knowing, and determining the effective and important characteristics from a citizen's point of view of solving the problems of the street sellers and overcoming its spread. The current application aims to legalize their situation in order to improve the civilized appearance and get rid of the randomness in the Egyptian streets. For solving the problems of the street sellers such as noise spread, traffic difficulties, violence, uncivilized appearance and then analyzing the data regarding the citizens' opinions about the street sellers for knowing their important characteristics as an indicator for solving this problem and overcoming the spread of this phenomenon.

This thesis, presents several techniques, algorithms, approaches and different areas of data mining technique models in cyber security from different perspectives. Then the study established a

classification and comparison of various types of intrusion detection and countermeasures in E-government of this research. It reflects the important criteria of the data mining models. It summarizes various intelligent data analyses and presents an intelligent data Analysis of “Cairo Cleaning and Beautification Agency". Establishing such as classification impacts deeply guiding data mining applications towards better operations and performance. Moreover, knowing how data mining can help in the detection and prevention of these attacks.

The study uses the Mining Audit Data for Automated Models for Intrusion Detection (MADAM ID); using strategy of inferring, analyses the data, searches for them in the cyberspace by one of the technology tools (data mining). A series of the standards build on the application of data mining methods specifically represented as "Frequencies", "Logistic regression", "association rules model", "Bayesian network", "decision tress model", "Neural Networks Model", and "Hierarchical Clustering". So we analyse for making reference measurements. They form "penetration test model" to measure the extent of securing the data, and the provided services. Also this strategy is very useful to enable the decision-maker for monitoring to measure the extent of securing the cyberspace, and the provided services

In this thesis, it is found that the cyberspace needs to be improved and to enhance its sufficiency and taking the necessary arrangements to raise the efficiency of the security. The results of this study are very useful to build a strategy for measuring the extent of securing data in order to improve the management of effective government services. Any type of data to be used, any type of data was transferred in a proper way. This study could be remarkable as one of the first studies on the use of data mining tools in cyberspace. Moreover, these results could become important tools for the government and intelligence agencies in the decision-making and monitoring potential international terrorist threats in real time.

List of Contents

List of Tables

List of Figures

List of Abbreviations

Acronym	Definition
AI	Artificial Intelligence
AIGA	Annealing Immune Genetic Algorithm
ANN	Artificial Neural Networks
C2G	Citizen To Government
CBR	Case-Base Reasoning
CCBA	Cairo cleaning and beautifying Authority
CI	Computational Intelligence
CST	Chinese Soil Taxonomy
DCBRs	Distributed Case-Based Reasoning System
DLPSO	Dual layered Particle Swarm Optimization Algorithm
DM	Data Mining
DM-ID unit	Data Mining Intrusion Detection Unit
EDM	Extension Data Mining
ERSAP	Emerging Regions Support and Partnership Program
EW&PCs	Early Warning and Proactive, Control Systems
FBI’s	FBI Crisis Negotiation Unit
FKMS	Financial Knowledge Management System
Fl	Fuzzy Logic
FP	Frequent Pattern
FS	Frequencies
FST	Fuzzy Set Theory
G2B	Government to Businesses/Commerce
G2C	Government to Citizen

G2E	Government to Employees
G2G	Government to Government
G2N	Government to Business, Government to NGO
GA	Genetic Algorithms
GKFs	Group-based Knowledge Flows
GP	Gaussian Processes
GP	Gaussian processes
HTN	Hierarchical Task Network
IAS	Implicit Alternative Splicing
ICT	Information and Communications Technology
ID	Intrusion Detection
IDS	Intrusion Detection Systems
IM	Information Matrix
ISM	Industrial Scientific and Medical
KBS	Knowledge-based System
KDD	knowledge discovery in Databases
KIN	Knowledge and Information Network
K-NN	k-Nearest Neighbors Algorithm
KPI	Key Performance Indicators
LF	Likelihood Function
LG	Logistic Regressions
LLF	Log –Likelihood Function
MADAM ID	Mining Audit Data for Automated Models for Intrusion Detection
MINDS	Minnesota Intrusion Detection System
MKTPKS	Multiple Key Term Phrasal Knowledge Sequences
MMK	Multiple Media Kiosks
OLAP	Online Analytical Processing

P2P	Peer-to-Peer
PSO	Particle Swarm Optimization
PTM	Penetration Testing Model
QP	Quantitative Psychology
RFID	Radio Frequency Identification
RS	Rough Sets
RTDMM	Real-Time Data Mining Methodology
SDH	Soil Diagnostic Horizo
SDH	Synchronous Digital Hierarchy
SOM	Self-Organizing Maps
SQL	Structured Query Language
SSTs	Self-Service Technologies
SVM	Support Vector Machines
TTCN-3	Testing and Test Control Notation Version 3
VT	Virtualization Via Intel- Technology
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network