Sudan University of Science and Technology
College of Graduate Studies

Design and Implementation of Data Mining-Based Intrusion Detection System

تصميم وتنفيذ نظام لكشف إختراق الشبكة مبنى على تقنية تنقيب البيانات

Thesis Submitted in Partial Fulfillment of Requirements for the Degree of M.Sc in Computer Science

Prepared By
Rasha Gaffer Mohammed Helali

Supervisor
Dr. Awad Mohammed Awad Al kareim

September 2008

DEDICATION

To my parents

To every person who taught me a letter. .

To every person who had a role to made this thesis see the light. .

I dedicate this work.

الايـــــــــــــــــــــه

ٹ ڑ
چ ڈ ۀ ۀ ە ۂ ۃ ۄ ھ ھ ھ ھ ے ے ۓ ۓ ڭ ڭ
ڭ ڭ ۇ ۇ ۆ ۆ ۈ ۈ ۉ ۉ ۋ ۋ ۅ ۅ ۅ ې ې ى
ﯜ ﯝ

ى ى ي چ

البقرة: ٢٥٥

ACKNOWLEDGEMENT

I am very grateful to Dr. Awad Mohammed Awad Al kareim for providing me with unwavering support. I also wish to thank Dr. Fathi almeligi and Dr. Mohammed awad for their helpful advice. Finally I wish to thank my colleagues for their assistance.

Abstract

Significant security problem for networked systems is hostile trespass by users or software. Intruder is one of the most publicized threats to security. In point of fact, most of the current systems are weak at detecting novel attacks without generating false alarms. This study proposes a solution to such limitations through data mining-based Network Intrusion Detection System NIDS. The proposed framework combines both misuse and anomaly detection techniques using data mining approaches such as decision tree (C5.0 algorithm) and distance-based clustering (Two-steps algorithm). As case study, the proposed framework is implemented to CCSIT (College of Computer Science and Information Technology) Network at Sudan University of Science and Technology (SUST), which clearly shows and reflects its applicability and effectiveness. Conclusively, the derived experimental results confirm that using of data mining approaches for both misuse and anomaly detection has a great promise in network security context.

المستخلص

إن سرية الشبكات تعتمد على نظم كشف الاختراق (Intrusion Detection System) كأحد التقنيات المهمة والاساسية لتوفير تأمين عالى للشبكة مما جعل مجموعة من الدراسات تتجه نحو تطويرها وجعلها أكثر فعالية. في الواقع ان معظم النظم الحالية ضعيفة في الكشف عن الاختراقات الغير معروفة مسبقا بل و تقوم بتوليد كم هائل من الانذارات الكاذبة.

هذه الدراسة اقترحت حل لمثل هذه المشاكل من خلال تطبيق التنقيب في البيانات.(Data Mining) تعرضت الدراسة لشرح مفصل عن تقنيات الاطار المقترح الذي جمع مابين التقنيات الخاصة باكتشاف مسيء الاستخدام والافعال الشاذة باستخدام طريقتي التنقيب في البيانات وهما شجرة القرار (خوارزمية C5.0) وطريقة تجميع البيانات بناء علي مقاييس التباعد (خوارزمية Two-steps). و طبق هذا الإطار على شبكة كلية علوم الحاسوب وتقانة المعلومات بجامعة السودان للعلوم والتكنولوجيا (SUST) والذي بين بوضوح قابليته للتطبيق و فعاليته.

ختاما , اكدت النتائج المستمدة من التجارب أن استخدام طريقة التنقيب في البيانات لكشف اختراقات مسيئي الاستخدام (misuse detection) والافعال الشاذة من قبل المستخدمين (anomaly detection) بانها طريقة واعدة في مجال امن الشبكات.

Table of Contents

List of Figures