ENSEMBLE METHODS IN INTRUSION DETECTION

KEKERE TEMITOPE JOSIAH

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2015

This dissertation is dedicated to my family for their endless support and encouragement.

ACKNOWLEDGEMENT

First, I want to thank God for the gift of life without which I wouldn’t be able to complete my study let alone this thesis. I want to thank my parents, Pastor & Mrs Kekere for their unending love, sacrificial support especially for sponsoring me to Malaysia to acquire advanced expertise. I sincerely thank my supervisor, **Dr Shukor Razak**, for his mentorship, guidance, support and confidence in me throughout my research in the faculty of computing, UTM.

I want to appreciate the efforts of my brothers and sister; Dr Victor Kekere, Engr Daniel Kekere and Mrs Ruth Umahi. Worthy of mention is the effort of Dr Clement Folorunsho for his response squad during my health crisis. I am not ungrateful to Dr Philip Achimugu for his enlightenment, time and support during my study in UTM. Special thanks to beloved people in The Redeemed Christian Church of God, Jalan Mutiara Emas, Johor Bahru, (a.k.a. PowerPalace) Pst Ruthie Gunasundari A/P, Pst Alabi Michael, Pst Olowoyo Austin, Pst Goke Oladokun, Pst Chidi Okpechi, Pst Charles Uti, Oluwagbemi Oluwatolani (Mummy Success), Augusta Nduka, Lilian Oladokun (Mummy Church), Bayo-Philip Patrick, Ayo Sanwoolu, Esther Tazamu, Bosede Edwards, Kemi Adedokun, Isaac Katuka, Kavitha, Carol, Previn, Daniel, Stanley, Paul, Pa and Mummy Esther. I also want to thank Tobi Akanbi, Marwa Alhazmi, Aminu Muhammed, Salisu Borodo, Nansukusa Hidaya Ngaya, Rabiu Idris, my friends in the faculty of computing, UTM. I also acknowledge the following Dr Olukayode Obasan, Dr Femi Folorunsho Ayinde, Dr Adebanjo Adekiigbe, Adeyemi Ajao.

ABSTRACT

As services are being deployed on the internet, there is the need to secure the infrastructure from malicious attacks. Intrusion detection serves as a second line of defense apart from firewall and cryptography. There are many techniques employed in intrusion detection which include signature detection, anomaly and specification based detection system. These techniques often trade off accuracy with false positive rate. In this study, anomaly detection using ensembles is used to automatically classify and detect attack patterns. It has been proven that ensembles of classifier outperform their base classifiers. Several multiples of classifiers have been combined to improve the performance of intrusion detection system. Commonly used classifiers include Support Vector Machines, Decision Trees, Genetic Algorithms, Fuzzy, Principal Component Analysis. The study employed KStar clustering and Instance Based classification algorithms to detect intrusions in NSL-KDD dataset. The results show that the ensemble we designed has a *1-error rate* of 99.67% and false positive 0.33%. The response time of the anomaly is 0.18seconds. The chosen ensemble outperformed the rest of the ensembles (rPART & SMO and J48) and the base classifiers. The performance of the combiners has showed that the study has built a model with high detection, and reduced error.

ABSTRAK

Sebagai perkhidmatan sedang diperluaskan di internet, terdapat keperluan untuk menjamin infrastruktur daripada serangan jahat. Pengesanan pencerobohan berfungsi sebagai pertahanan peringkat kedua selain dari "firewall" dan kriptografi. Terdapat pelbagai teknik yang digunakan dalam pengesanan pencerobohan iaitu pengesanan tandatangan, anomali dan spesifikasi berasaskan sistem pengesanan. Teknik tersebut mempertimbangkan ketepatan berdasarkan kadar kesalahan positif. Dalam kajian ini, pengesanan anomali berasaskan pengumpulan digunakan untuk mengkelaskan dan mengesan corak serangan secara automatik. Ia terbukti dapat mengumpul pengelasan yang melebihi pengelasannya. Beberapa pengelas digabungkan untuk meningkatkan prestasi sistem pengesanan pencerobohan. Pengelas yang selalu digunakan adalah Sokongan Mesin Vektor, Pokok Keputusan, Algoritma Genetik, Kabur, Analisis Komponen Utama. Kajian ini menggunakan pergkelasan KStar algoritma pengkelasan segera untuk mengeson pencerobohan dalam set data NSL-KDD. Kajian menunjukkan bahawa pengumpulan yang dibangunkan mempunyai kadar 1-kesilapan sebanyak 99.67% dan kesalahan positif 0.33%. Masa tindak balas daripada anomali adalah 0.18saat. Pengumpul yang dipilih telah mengatasi (rPART & SMO dan J48) dan Pengelas asas. Prestasi daripada penggambungan ini telah menunjukkan bahawa kajian telah membina sebuah model dengan pengesanan tinggi, dan kesilapan dikurangkan.