UNIVERSITI PUTRA MALAYSIA

FEATURES SELECTION FOR INTRUSION DETECTION SYSTEM USING HYBRIDIZE PSO-SVM

ALAA ABDULRAHMAN TABAAN

FSKTM 2017 22

FEATURES SELECTION FOR INTRUSION DETECTION SYSTEM USING HYBRIDIZE PSO-SVM

ALAA ABDULRAHMAN TABAAN

MASTER OF COMPUTER SCIENCE

UNIVERSITY PUTRA MALAYSIA

2016

FEATURES SELECTION FOR INTRUSION DETECTION SYSTEM USING HYBRIDIZE PSO-SVM

By

ALAA ABDULRAHMAN TABAAN

Thesis Submitted To the Scholl of Graduate Studies, University Putra Malaysia In Fulfillment Of the Requirement for the Degree of Master of Science

December 2016

DEDICATION

This thesis is dedicated to my parents for their endless love, support and encouragement.

Abstract of thesis presented to the Senate of University Putra Malaysia in Fulfillment of the Requirement for the Degree of Master of Computer Science

FEATURES SELECTION FOR INTRUSION DETECTION SYSTEM USING HYBRIDIZE PSO-SVM

By

ALAA ABDULRAHMAN TABAAN
December 2016

Chairman : Azizol Abdullah, PhD
Faculty : Computer Science and Information Technology

Abstract

An Intrusion Detection System is software or application which is used to detect thread, malicious activities and the unauthorized access to the computer system and warn the administrators by generating alarms. Features selection process can be considered a problem of global combinatorial optimization in machine learning. Genetic algorithm GA had been adopted to perform features selection method; however, this method could not deliver an acceptable detection rate, lower accuracy, and higher false alarm rates. Hybridize Particle Swarm Optimization (PSO) as a searching algorithm and support vector machine (SVM) as a classifier had been implemented to cope with this problem. The results reveal that the proposed hybrid algorithm is capable of achieving classification accuracy values of (95.82 % and 97.68 %), detection rates values of (95.8 % and 99.3 %) and false alarm rates values of (0.083 % and 0.045 %) on both KDD CUP 99 and NSL KDD. Electing the best set of features will help to improve the classifier predictions in

terms of the normal and abnormal pattern. The simulation will be carried on WEKA tool, which allows us to call some data mining methods under JAVA environment. The proposed model will be tested and evaluated on both NSL-KDD and KDD-CUP 99 using several performance metrics.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia Sebagai memenuhi keperluan untuk ijazah MASTER OF SCIENCE

CIRI PEMILIHAN UNTUK PENCEROBOHAN PENGESANAN SISTEM MENGGUNAKAN silang PSO-SVM

Oleh

ALAA ABDULRAHMAN TABAAN
December 2016

Pengerusi: Azizol Abdullah, PhD
Fakulti: Sains Komputer dan Teknologi Maklumat

Abstrak

Sistem Pengesanan Pencerobohan adalah perisian atau aplikasi yang digunakan untuk mengesan, aktiviti berniat jahat dan akses yang tidak dibenarkan kepada sistem komputer dan memberi amaran kepada pentadbir dengan menjana penggera. proses pemilihan ciri-ciri boleh dianggap sebagai masalah pengoptimuman kombinatorik global dalam pembelajaran mesin. algoritma genetik GA telah diterima untuk melaksanakan kaedah ciri pemilihan; Walau bagaimanapun, kaedah ini tidak dapat melepaskan kadar yang boleh diterima pengesanan, ketepatan yang lebih rendah, dan kadar penggera palsu yang lebih tinggi. Silang Particle Swarm Optimization (PSO) sebagai mesin pencarian algoritma dan sokongan vektor (SVM) sebagai pengelas yang telah dilaksanakan untuk menangani masalah ini. Keputusan menunjukkan bahawa algoritma hibrid yang dicadangkan mampu mencapai nilai pengelasan ketepatan (95,82% dan 97,68%), kadar pengesanan nilai

(95.8% dan 99.3%) dan kadar penggera palsu nilai (0,083% dan 0.045%) ke atas kedua-dua KDD CUP 99 dan NSL KDD. Memilih set yang terbaik adalah ciri-ciri yang akan membantu untuk meningkatkan ramalan pengelas dari segi corak yang normal dan tidak normal. simulasi ini akan dijalankan oleh WEKA iaitu alat, yang membolehkan kita untuk memanggil bebaerapa kaedah perlombongan data di bawah persekitaran JAVA. model yang dicadangkan akan diuji dan dinilai pada kedua-dua NSL-KDD dan KDD-CUP 99 menggunakan beberapa metrik prestasi.

ACKNOWLEDGMENT

First and foremost, I would like to thank my parents for their love and support throughout my life. Thank you both for giving me strength to reach for the stars and chase my dreams. My brother and little sister deserve my wholehearted thanks as well.

I would like to sincerely thank and express my deepest gratitude to Dr. Azizol Abdullah for his supervision, support and understanding during this study and in the preparation of this thesis.

To all my friends, thank you for your understanding and encouragement in my many, many moments of crisis. Your friendship makes my life a wonderful experience. I cannot list all the names here, but you are always on my mind.

I certify that a Thesis Examination Committee has met on December 2016 to conduct the final examination of (Alaa Abdulrahman Tabaan) on his thesis entitled "FEATURES SELECTION FOR INTRUSION DETECTION SYSTEM USING HYBRIDIZE PSO-SVM") in accordance with the Universities and University Colleges Act 1971 and the Constitution of the University Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science

Members of the Thesis Examination Committee were as follows:

Azizol Abdullah (Dr.)
Senior Lecturer
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Supervisor)

Zurina Mohd Hanapi (Assoc. Prof. Dr.)
Senior Lecturer
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Assessor)

Azizol Abdullah (Dr.)
Senior Lecturer
Faculty of Computer Science
and Information Technology
University Putra Malaysia
Date:

This thesis was submitted to the Senate of University Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Azizol Abdullah (Dr.)
Senior Lecturer
Faculty of Computer Science and Information Technology
University Putra Malaysia

Zurina Mohd Hanapi
(Assoc. Prof. Dr.)
Senior Lecturer
Faculty of Computer Science
and Information Technology
University Putra Malaysia
Date:

DECLARATION

Declaration by graduate student I hereby confirm that:

- This thesis is my original work;
- Quotations, illustrations and citations have been duly referenced;
- This thesis has not been submitted previously or concurrently for any other
- Degree at any other institutions;
- Intellectual property from the thesis and copyright of thesis are fully-owned by University Putra Malaysia, as according to the University Putra Malaysia.
  (Research) Rules 2012;
- Written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the University Putra Malaysia (Research) Rules 2012;
- There is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the University Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the University Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ___________________ Date: _______________

Name and Matric No.: __________________________________

List of Tables

List of Figures

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Background

With the increasing development of IT (information technology), computer networks became more important to the people for their daily uses (checking emails, news, purchase online, and so on). Securing clients' information on the network became an eccentric issue to ensure the confidentiality, integrity, and availability of the system against the insider misuse and the outside attack. Thus; developing intrusion detection systems (IDS) becomes a researchers' interest (Guolong *et al.*, 2007). Researchers have been developing some artificial intelligence (AI) and data mining algorithms such as Fuzzy Logic, K-Nearest Neighbor, Support Vector Machine, Artificial Neural Network, Genetic Algorithm, and Particle Swarms Optimization to come up with an effective and reliable intrusion detection system while it's dealing with a huge and messy data.

Computer system becomes a target for the inside misuse and outside attack because of its importance in our life. Therefore, we need to construct the best security mechanism to ensure the safety of our information. *An intrusion* can be basically defined as a set of actions or activities which might compromise the confidentiality, integrity, and availability of the system, and vulnerable the system once it happened. *An IDS* is a software or an application which is used to detect thread, malicious activities and the unauthorized access to the computer system. To add to that, IDS warn the system administrators about any event that might compromise the system security either by violating the policies or using malicious activities. IDS provide extra protection to the system despite the existing prevention technique such as firewall and the traditional security tools which can't efficiently detect the attacks because of the hidden vulnerabilities in these techniques (Aslahi-Shahri *et al.*, 2015). Despite the static protection tools like firewall and the updated software which can afford an acceptable security level, dynamic tools like IDS need to be employed as well. The IDS main purpose is to monitor the system activates by seeking for the system weaknesses, files integrity, and make an analysis based on the previous attacks. An IDS can be *categorized* based on detection type into the misuse intrusion detection system and anomaly intrusion detection system and into Network based IDS and Host based IDS based on the data source.

1.2 Problem Statement

Support vector machine (SVM) is a popular machine learning technique which successfully has been applied to construct an intrusion detection system. Features selection process can be considered a problem of global combinatorial optimization in machine learning. Genetic algorithm GA has been adopted to perform features selection method; however, these methods could not deliver an acceptable detection rate, lower accuracy and higher false alarm rates (Aslahi-Shahri *et al.*, 2015). To cope with this problem, an IDS model (hybridize Particle Swarm Optimization (PSO) and SVM) based on feature selection by PSO and SVM classifier has been implemented.

Features selection has been deployed to select the best features to show a better data representation. To add to that, feature selection helps to minimize the data dimensionality by eliminating redundant and irrelevant features from the dataset. Moreover, eliminating the less important features from the dataset before using it to train SVM will aid to increase the classification accuracy and reduce the misclassified instances. Using feature selection method will assist in creating an enhanced classifier with less number of features. Features selection has great advantages in terms of pattern recognition and machine learning.

1.3 Objectives

Our objective can be broken down as follow:

1. To propose a PSO-SVM feature selection method for IDS to reduce the data dimensionality by selecting the features that can show a better data representation.
2. To propose an enhanced classifier algorithm (SVM) and compare it against the recent IDS model in terms of Classification Accuracy, Detection Rate or Recall (DR), and False Alarm Rate (FAR).

1.4 Contribution

Our contribution will be divided as follow:

First of all, to use particle swarm optimization (PSO) as a feature selection method this will help to reduce the data dimensionality.

Secondly, to come up with an enhanced intrusion detection system compare to the recent IDS models in terms of Classification Accuracy, Detection Rate or Recall (DR), and false alarm rate (FAR).

Our IDS model consists of particle swarm optimization PSO as a feature selection method and support vector machine SVM as a classifier to evaluate the fitness of Particle Swarm Optimization (PSO).

1.5 Scope

Our scope is to simulate the proposed model of intrusion detection system (IDS) using Waikato Environment for Knowledge Analysis (WEKA), which is an open source application developed at University of Waikato and licensed under GNU General Public License. WEKA allows us to call some data mining methods under JAVA environment. It can be installed on Windows, MAC, and LINUX.

1.6 Thesis organization

The rest of this project will be distributed as following:

Chapter 2 literature reviews consists of the earlier studies and the research works which were done before in the field of intrusion detection system.

Chapter 3 the methodology describes the research methodology of our work, the proposed model of IDS and the performance of our proposed model using different performance metrics on KDD CUP 99 and NSL KDD dataset.

Chapter 4 results and discussion presents the results achieved by proposed intrusion detection's model and compare it versus GASVM using several metrics.

In chapter 5 we conclude the findings of the research and show its strength.

References

Abadeh, M. S., Habibi, J., & Lucas, C. (2007). Intrusion detection using a fuzzy genetics-based learning algorithm. *Journal of Network and Computer Applications, 30*(1), 414-428.

Aghdam, M. H., & Kabiri, P. (2016). Feature selection for intrusion detection system using ant colony optimization. *International Journal of Network Security, 18*(3), 420-432.

Ahmad, I., Hussain, M., Alghamdi, A., & Alelaiwi, A. (2014). Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components. *Neural Computing and Applications, 24*(7-8), 1671-1682.

Alazab, A., Hobbs, M., Abawajy, J., & Alazab, M. (2012). *Using feature selection for intrusion detection system.* Paper presented at the Communications and Information Technologies (ISCIT), 2012 International Symposium on.

Aslahi-Shahri, B., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M., & Ebrahimi, A. (2015). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Computing and Applications*, 1-8.

Azad, C., & Jha, V. K. (2016). Fuzzy min–max neural network and particle swarm optimization based intrusion detection system. *Microsystem Technologies*, 1-12.

Aziz, A. S. A., Azar, A. T., Salama, M. A., Hassanien, A. E., & Hanafy, S. E.-O. (2013). *Genetic algorithm with different feature selection techniques for anomaly detectors generation.* Paper presented at the Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on.

Bostani, H., & Sheikhan, M. (2015). Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems. *Soft Computing*, 1-18.

Bukhtoyarov, V., & Zhukov, V. (2014). *Ensemble-distributed approach in classification problem solution for intrusion detection systems.* Paper presented at the International Conference on Intelligent Data Engineering and Automated Learning.

Chen, R.-C., Cheng, K.-F., Chen, Y.-H., & Hsieh, C.-F. (2009). *Using rough set and support vector machine for network intrusion detection system.* Paper presented at the Intelligent Information and Database Systems, 2009. ACIIDS 2009. First Asian Conference on.

Chen, Y., Abraham, A., & Yang, J. (2005). *Feature selection and intrusion detection using hybrid flexible neural tree.* Paper presented at the International Symposium on Neural Networks.

Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing, 12*(9), 3014-3022.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Crosbie, M., & Spafford, G. (1995). *Applying genetic programming to intrusion detection.* Paper presented at the Working Notes for the AAAI Symposium on Genetic Programming.

Deng, N., Tian, Y., & Zhang, C. (2012). *Support vector machines: optimization based theory, algorithms, and extensions*: CRC press.

Eberhart, R. C., & Kennedy, J. (1995). *A new optimizer using particle swarm theory.* Paper presented at the Proceedings of the sixth international symposium on micro machine and human science.

Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Systems with Applications, 42*(5), 2670-2679.

Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. *Expert Systems with Applications, 42*(1), 193-202.

Eunhye, K., & Kim, S. (2014). A Novel Anomaly Detection System Based on HFR-MLR Method *Mobile, Ubiquitous, and Intelligent Computing* (pp. 279-286): Springer.

Fan, W., Bouguila, N., & Ziou, D. (2011). *Unsupervised anomaly intrusion detection via localized bayesian feature selection.* Paper presented at the 2011 IEEE 11th International Conference on Data Mining.

Gao, H.-H., Yang, H.-H., & Wang, X.-Y. (2005). *Ant colony optimization based network intrusion feature selection and detection.* Paper presented at the 2005 International Conference on Machine Learning and Cybernetics.

Ghali, N. I. (2009). Feature selection for effective anomaly-based intrusion detection. *International Journal of Computer Science and Network Security, 9*(3), 285-289.

Golmah, V. (2014). An efficient hybrid intrusion detection system based on C5. 0 and SVM. *International Journal of Database Theory and Application, 7*(2), 59-70.

Guolong, C., Qingliang, C., & Wenzhong, G. (2007). A PSO-based approach to rule learning in network intrusion detection *Fuzzy Information and Engineering* (pp. 666-673): Springer.

Hall, M. A., & Smith, L. A. (1999). *Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper.* Paper presented at the FLAIRS conference.

Ibrahim, L. M., Basheer, D. T., & Mahmod, M. S. (2013). A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network. *Journal of Engineering Science and Technology, 8*(1), 107-119.

John, G. H., Kohavi, R., & Pfleger, K. (1994). *Irrelevant features and the subset selection problem.* Paper presented at the Machine learning: proceedings of the eleventh international conference.

Kim, D. S., Nguyen, H.-N., Ohn, S.-Y., & Park, J. S. (2005). *Fusions of GA and SVM for anomaly detection in intrusion detection system.* Paper presented at the International Symposium on Neural Networks.

Koc, L., Mazzuchi, T. A., & Sarkani, S. (2012). A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Expert Systems with Applications, 39*(18), 13492-13500.

Kosamkar, V. (2013). *Improved Intrusion Detection System using C4. 5 Decision Tree and Support Vector Machine.* Mumbai University.

Kumar, V., Chauhan, H., & Panwar, D. (2013). K-means clustering approach to analyze NSL-KDD intrusion detection dataset. *International Journal of Soft*.

Laamari, M. A., & Kamel, N. (2014). A hybrid bat based feature selection approach for intrusion detection *Bio-Inspired Computing-Theories and Applications* (pp. 230-238): Springer.

Li, W. (2004). Using genetic algorithm for network intrusion detection. *Proceedings of the United States Department of Energy Cyber Security Group, 1*, 1-8.

Li, Y., Wang, J.-L., Tian, Z.-H., Lu, T.-B., & Young, C. (2009). Building lightweight intrusion detection system using wrapper-based feature selection mechanisms. *Computers & Security, 28*(6), 466-475.

Li, Y., Xia, J., Zhang, S., Yan, J., Ai, X., & Dai, K. (2012). An efficient intrusion detection system based on support vector machines and gradually feature removal method. *Expert Systems with Applications, 39*(1), 424-430.

Om, H., & Kundu, A. (2012). *A hybrid system for reducing the false alarm rate of anomaly intrusion detection system.* Paper presented at the Recent Advances in Information Technology (RAIT), 2012 1st International Conference on.

Park, J. S., Shazzad, K. M., & Kim, D. S. (2005). *Toward modeling lightweight intrusion detection system through correlation-based hybrid feature selection.* Paper presented at the International Conference on Information Security and Cryptology.

Patra, M. P. a. M. R. (2009). Evaluating machine learning algorithms for detecting network intrusions. *Int. J. of Recent Trends in Engineering and Technology, 1*(1).

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.

Ravale, U., Marathe, N., & Padiya, P. (2015). Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function. *Procedia Computer Science, 45*, 428-435.

Revathi, S., & Malathi, A. (2013). A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection. *International Journal of Engineering Research and Technology. ESRSA Publications*.

Sahu, S. K., Sarangi, S., & Jena, S. K. (2014). *A detail analysis on intrusion detection datasets.* Paper presented at the Advance Computing Conference (IACC), 2014 IEEE International.

Sathya, S. S., Ramani, R. G., & Sivaselvi, K. (2011). Discriminant analysis based feature selection in kdd intrusion dataset. *International Journal of Computer Applications, 31*(11), 1-7.

Shu, G., Fu, G., Li, P., & Geng, H. (2014). Violent Behavior Detection Based on SVM in the Elevator. *System, 8*(5).

Sinclair, C., Pierce, L., & Matzner, S. (1999). *An application of machine learning to network intrusion detection.* Paper presented at the Computer Security Applications Conference, 1999.(ACSAC'99) Proceedings. 15th Annual.

Singh, A., Banafar, H., & Pippal, R. S. (2015). Intrusion Detection on KDD99cup Dataset using K-means, PSO and GA: A Review. *International Journal of Electrical, Electronics and Computer Engineering, 4*(1), 40.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*(4857), 1285-1293.

Tan, Z., Jamdagni, A., He, X., & Nanda, P. (2010). *Network Intrusion Detection based on LDA for payload feature selection.* Paper presented at the 2010 IEEE Globecom Workshops.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A.-A. (2009). *A detailed analysis of the KDD CUP 99 data set.* Paper presented at the Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009.

Tran, B., Xue, B., & Zhang, M. (2014). *Improved PSO for feature selection on high-dimensional datasets.* Paper presented at the Asia-Pacific Conference on Simulated Evolution and Learning.

Tsang, C.-H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition, 40*(9), 2373-2391.

Wang, Y., Wong, J., & Miner, A. (2004). *Anomaly intrusion detection using one class SVM.* Paper presented at the Information Assurance Workshop, 2004. Proceedings from the Fifth Annual IEEE SMC.

Xiao, L., Chen, Y., & Chang, C. K. (2014). *Bayesian model averaging of bayesian network classifiers for intrusion detection.* Paper presented at the Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International.

Xue, B., Zhang, M., & Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing, 18*, 261-276.

Yao, J., Zhao, S., & Fan, L. (2006). *An enhanced support vector machine model for intrusion detection.* Paper presented at the International Conference on Rough Sets and Knowledge Technology.

Yin, C., Ma, L., & Feng, L. (2015). Towards accurate intrusion detection based on improved clonal selection algorithm. *Multimedia Tools and Applications*, 1-14.

Zainal, A., Maarof, M. A., & Shamsuddin, S. M. (2006). *Feature selection using rough set in intrusion detection.* Paper presented at the TENCON 2006-2006 IEEE Region 10 Conference.

Zainal, A., Maarof, M. A., & Shamsuddin, S. M. (2007). *Feature selection using rough-DPSO in anomaly intrusion detection.* Paper presented at the International Conference on Computational Science and Its Applications.