BUILDING AN INTRUSION DETECTION SYSTEM USING A FILTER-BASED FEATURE SELECTION ALGORITHM

PITTA BINDUMADHAVI, B. SANDHYA RANI
Mtech Scholar, Assistant Professor
Department of of CSE
ISTS Women's Engineering College, Rajanagaram, Rajamahendravaram, A.P, India.

ABSTRACT:

Redundant and irrelevant features in data have caused a long-term problem in network traffic classification. These features not only slow down the process of classification but also prevent a classifier from making accurate decisions, especially when coping with big data. In this paper, we propose a mutual information based algorithm that analytically selects the optimal feature for classification. This mutual information based feature selection algorithm can handle linearly and nonlinearly dependent data features. Its effectiveness is evaluated in the cases of network intrusion detection. An Intrusion Detection System (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is built using the features selected by our proposed feature selection algorithm. The performance of LSSVM-IDS is evaluated using three intrusion detection evaluation datasets, namely KDD Cup 99, NSL-KDD and Kyoto 2006+ dataset. The evaluation results show that our feature selection algorithm contributes more critical features for LSSVM-IDS to achieve better accuracy and lower computational cost compared with the state-of-the-art methods.

Keywords: **IDS, LSSVM-IDS, KDD**

1. INTRODUCTION

Despite increasing awareness of network security, the existing solutions remain incapable of fully protecting internet applications and computer networks against the threats from ever-advancing cyber attack techniques such as DoS attack and computer malware. Developing effective and adaptive security approaches, therefore, has become more critical than ever before. The traditional security techniques, as the first line of security defence, such as user authentication, firewall and data encryption, are insufficient to fully cover the entire landscape of network security while facing challenges from ever-evolving intrusion skills and techniques [1]. Hence, another line of security defence is highly recommended, such as Intrusion Detection System (IDS). Recently, an IDS alongside with anti-virus software has become an important complement to the security infrastructure of most organizations. The combination of these two lines provides a more comprehensive defence against those threats and enhances network security.A significant amount of research has been conducted to develop intelligent intrusion detection techniques, which help achieve better network security. Bagged boosting-based on C5 decision trees [2] and Kernel Miner [3] are two of the earliest attempts to build intrusion detection schemes.\ Methods proposed in [4]

and [5] have successfully applied machine learning techniques, such as Support Vector Ma-chine (SVM), to classify network traffic patterns that do not match normal network traffic. Both systems were equipped with five distinct classifiers to detect normal traffic and four different types of attacks (i.e., DoS, probing, U2R and R2L). Experimental results show the effectiveness and robustness of using SVM in IDS. Mukkamala et al. [6] investigated the possibility of assembling various learning methods, including Artificial Neural Networks (ANN), SVMs and Multivariate Adaptive Regression Splines (MARS) to detect intrusions. They trained five different classifiers to distinguish the normal traffic from the four different types of attacks. They compared the performance of each of the learning methods with their model and found that the ensemble of ANNs, SVMs and MARS achieved the best performance in terms of classification accuracies for all the five classes. Toosi et al. [7] combined a set of neuro-fuzzy classifiers in their design of a detection system, in which a genetic algorithm was applied to optimize the structures of neuro-fuzzy systems used in the classifiers. Based on the mpre-determined fuzzy inference system (i.e., classifiers),detection decision was made on the incoming traffic.

Recently, we proposed an anomaly-based scheme for detecting DoS attacks [8]. The system has been evaluated on KDD Cup 99 and ISCX 2012 datasets and achieved promising detection accuracy of 99.95% and 90.12% respectively. However, current network traffic data, which are often huge in size, present a major challenge to IDSs [9]. These "big data" slow down the entire detection process and may lead to unsatisfactory classification accuracy due to the computational difficulties in handling such data. Classifying a huge amount of data usually causes many mathematical difficulties which then lead to higher computational complexity. As a well-known intrusion evaluation dataset, KDD Cup 99 dataset is a typical example of large-scale datasets. Samples and two million of testing samples respectively. Such a large scale dataset retards the building and testing processes of a classifier, or makes the classifier unable to perform due to system failures caused by insufficient memory. Furthermore, large-scale datasets usually contain noisy, redundant, or uninformative features which present criticalchallenges to knowledge discovery and data modeling. To address the aforementioned problems on the methods for feature selection, we have proposed a hybrid feature selection algorithm (HFSA) in [10]. HFSA consists of two phases. The upper phase conducts a preliminary search to eliminate irrelevant and redundancy features from the original data. This helps the wrapper method (the lower phase)to decrease the searching range from the entire original feature space to the pre-selected features (the output of the upper phase). In this paper, we extend our work discussed in [10]. The key contributions of this paper are listed as follows.

1. This work proposes a new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence between features and output classes. The most relevant features are retained and used to construct classifiers for respective classes. As an enhancement of Mutual Information Feature Selection (MIFS) [11] and Modified Mutual Information based Feature Selection (MMIFS) [12], the proposed feature selection method does not have any free parameter, such as _ in MIFS and MMIFS. Therefore, its performance is free from being influenced by any inappropriate assignment of value to a free parameter and can be guaranteed. Moreover, the proposed method is feasible to work in various domains, and more efficient in comparison with HFSA [10], where the computationally expensive wrapper-based feature selection mechanism is used.

2. We conduct complete experiments on two well known IDS datasets in addition to the dataset used in [10]. This is very important in evaluating the performance of IDS since KDD dataset is outdated and does not contain most novel attack patterns in it. In addition, these datasets are frequently used in the literature to evaluate the performance of IDS. Moreover, these datasets have various sample sizes and different numbers of features, so they provide a lot more challenges for comprehensively testing feature selection algorithms.

3. Different from the detection framework proposed in [10] that designs only for binary classification, we design our proposed framework to consider multiclass classification problems. This is to show the effectiveness and the feasibility of the proposed method.

2. SYSTEM STUDY

2.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

PRELIMINARY INVESTIGATION

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine ,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, ie. preliminary investigation begins. The activity has three parts:

- Request Clarification
- Feasibility Study
- Request Approval

REQUEST CLARIFICATION

After the approval of the request to the organization and project guide, with an investigation being considered, the project request must be examined to determine precisely what the system requires.

Here our project is basically meant for users within the company whose systems can be interconnected by the Local Area Network(LAN). In today's busy schedule man need everything should be provided in a readymade manner. So taking into consideration of the vastly use of the net in day to day life, the corresponding development of the portal came into existence.

FEASIBILITY ANALYSIS

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

- Operational Feasibility
- Economic Feasibility
- Technical Feasibility

Operational Feasibility

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

Economic Feasibility

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economically feasible.

Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform Independent environment. Java Server Pages, JavaScript, HTML, SQL server and WebLogic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

REQUEST APPROVAL

Not all request projects are desirable or feasible. Some organization receives so many project requests from client users that only few of them are pursued. However, those projects that are both feasible and desirable should be put into schedule. After a project request is approved, it cost, priority, completion time and personnel requirement is estimated and used to determine where to add it to any project list. Truly speaking, the approval of those above factors, development works can be launched.

3. LITERATURE SURVEY

3.1 INTEROPERABILITY OF PERSONAL HEALTH RECORDS

AUTHORS: J. L ahteenm€ aki, J. Lepp anen, and H. Kaijanranta,

The establishment of the Meaningful Use criteria has created a critical need for robust interoperability of health records. A universal definition of a personal health record (PHR) has not been agreed upon. Standardized code sets have been built for specific entities, but integration between them has not been supported. The purpose of this research study was to explore the hindrance and promotion of interoperability standards in relationship to PHRs to describe interoperability progress in this area. The study was conducted following the basic principles of a systematic review, with 61 articles used in the study. Lagging interoperability has stemmed from slow adoption by patients, creation of disparate systems due to rapid development to meet requirements for the Meaningful Use stages, and rapid early development of PHRs prior to the mandate for integration among multiple systems. Findings of this study suggest that deadlines for implementation to capture Meaningful Use incentive payments are supporting the creation of PHR data silos, thereby hindering the goal of high-level interoperability.

3.2 APPLYING CLOUD COMPUTING MODEL IN PHR ARCHITECTURE

AUTHORS: S. Kikuchi, S. Sachdeva, and S. Bhalla,

In recent years, some practical and commercial Personal Health Records and some related services such as Google Health [1] and Microsoft HealthVault [2] have been launched. On the other hand, Cloud Computing has matured more and become the major streams to realize a more effective operational environment. However so far, there have been few studies in regards to applying Cloud architecture in the PHR explicitly despite generating volume data. In this paper, we review our trial on the general architecture design by applying the Cloud components for supporting healthcare record areas and clarify the required conditions to realize it.

3.3 HEALTH INFORMATION PRIVACY, SECURITY, AND YOUR EHR

AUTHORS: M. Bellare

If your patients lack trust in Electronic Health Records (EHRs) and Health Information Exchanges (HIEs), feeling that the confidentiality and accuracy of their electronic health information is at risk, they may not want to disclose health information to you. Withholding their health information could have life-threatening consequences. To reap the promise of digital health information to achieve better health outcomes, smarter spending, and healthier people, providers and individuals alike must trust that an individual's health information is private and secure.

Your practice, not your EHR developer, is responsible for taking the steps needed to protect the confidentiality, integrity, and availability of health information in your EHR system.

3.4 A SECURE ANTI-COLLUSION DATA SHARING SCHEME FOR DYNAMIC GROUPS IN THE CLOUD

AUTHORS: C. Ng and P. Lee. Revdedup

Benefited from cloud computing, users can achieve an effective and economical approach for data sharing among group members in the cloud with the characters of low maintenance and little management cost. Meanwhile, we must provide security guarantees for the sharing data files since they are outsourced. Unfortunately, because of the frequent change of the membership, sharing data while providing privacy-preserving is still a challenging issue, especially for an untrusted cloud due to the collusion attack. Moreover, for existing schemes, the security of key distribution is based on the secure communication channel, however, to have such channel is a strong assumption and is difficult for practice. In this paper, we propose a secure data sharing scheme for dynamic members. Firstly, we propose a secure way for key distribution without any secure communication channels, and the users can securely obtain their private keys from group manager. Secondly, our scheme can achieve fine-grained access control, any user in the group can use the source in the cloud and revoked users cannot access the cloud again after they are revoked. Thirdly, we can protect the scheme from collusion attack, which means that revoked users cannot get the original data file even if they conspire with the untrusted cloud. In our approach, by leveraging polynomial function, we can achieve a secure user revocation scheme. Finally, our scheme can

achieve fine efficiency, which means previous users need not to update their private keys for the situation either a new user joins in the group or a user is revoked from the group

3.5 ADVANCE SECURITY TO CLOUD DATA STORAGE

AUTHORS: P. Lee, and W. Lou

The proposed system is an effective and flexible distributed Scheme with explicit dynamic data support to ensure the correctness of user's data in the cloud. To fully ensure the data integrity and save the cloud users computation it is of critical importance to enable public auditing service for cloud data storage, so that users may depend on independent third party auditor to audit the outsourced data. The Third party auditor can periodically check the integrity of all the data stored in the cloud .which provides easier way for the users to ensure their storage correctness in the cloud.

4. SYSTEM DESIGN AND DEVELOPMENT

INPUT DESIGN

Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.

This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.

Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

OUTPUT DESIGN

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and

validating a new user rests with the administrator only. The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

IMPLEMENTATION

- **Source**
  In this module, the Source is responsible for register using the Biometric authentication. The Biometric authentication it's a way of logging in to project with face recognisation or login with an image. If you are entered image and already exist images are getting match or biometric login is successful then Source will get activate. Source browses the data File, and uploads their data files to the particular Receiver (Reciever1, Reciever2, Reciever3, and Reciever4).

- **Intrusion Classifier**
  The classifier is responsible to scan their contents a Biometric Scan and Vulnerable word scan/ Span message scan.

  Biometric scan
  Authenticate the user with Biometric / image and then activate the Source Otherwise your image and existed image are not matched or Biometric authentication fails in pattern classifiers then related message and Image will be stores in pattern manager.

  Span message scan
  The classifiers check if uploaded file contains any vulnerable or bad words then pattern classifier removes those words and these words stores in pattern classifier Manager.

- **Intrusion Classifier manager**
  The Pattern classifier manager is responsible for capturing the whole transaction of the authentication and spam messages. You can check all the details regarding biometric authentication with their tags (Image-name, Date and time and status). The IDM can view the scanning report of spam message with their tags Filename, invalid words, Receivers and Date and time, and also can filter the fake injected data and captures in the attacker table with their tags File name, Injected data, Date and Time.

  Reciever
  In this module, the Receivers (Reciever1, Reciever2, Reciever3, and Reciever4) can receive the file sent from the Sourcevia Pattern classifier.

  Threat model
  In this model, Attacker adds fake data when Source wants to send a file to receivers, in the middle of Source and pattern classifier. The Attacker may have chance to attack on file or he can inject a false data. And these attacked details are recognized by pattern classifiers. If injected data found then all these

details are sent to Intrusion Detection Manager (IDM), after removing this injected data, file will be sent safely to respective receiver(Reciever1, Reciever2, Reciever3, and Reciever4).

6. SYSTEM ANALYSIS

EXISTING SYSTEM

- A significant amount of research has been conducted to develop intelligent intrusion detection techniques, which help achieve better network security. Bagged boosting-based on C5 decision trees and Kernel Miner are two of the earliest attempts to build intrusion detection schemes.

- Mukkamala et al. investigated the possibility of assembling various learning methods, including Artificial Neural Networks (ANN), SVMs and Multivariate Adaptive Regression Splines (MARS) to detect intrusions.

DISADVANTAGES OF EXISTING SYSTEM:

- Existing solutions remain incapable of fully protecting internet applications and computer networks against the threats from ever-advancing cyber attack techniques such as DoS attack and computer malware.
- Current network traffic data, which are often huge in size, present a major challenge to IDSs. These **"**big data**"** slow down the entire detection process and may lead to unsatisfactory classification accuracy due to the computational difficulties in handling such data.
- Classifying a huge amount of data usually causes many mathematical difficulties which then lead to higher computational complexity.
- Large-scale datasets usually contain noisy, redundant, or uninformative features which present critical challenges to knowledge discovery and data modeling.

PROPOSED SYSTEM:

- We have proposed a hybrid feature selection algorithm (HFSA). HFSA consists of two phases.
- The upper phase conducts a preliminary search to eliminate irrelevant and redundancy features from the original data. This helps the wrapper method (the lower phase) to decrease the searching range from the entire original feature space to the pre-selected features (the output of the upper phase). The key contributions of this paper are listed as follows.
- This work proposes a new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence between features and output classes.
- The most relevant features are retained and used to construct classifiers for respective classes. As an enhancement of Mutual Information Feature Selection (MIFS) and Modified Mutual Information based Feature Selection (MMIFS), the proposed feature selection method does not have any free parameter, such as in MIFS and MMIFS. Therefore, its performance is free from being influenced by any inappropriate assignment of value to a free parameter and can be guaranteed. Moreover, the proposed

method is feasible to work in various domains, and more efficient in comparison with HFSA, where the computationally expensive wrapper-based feature selection mechanism is used.

- We conduct complete experiments on two well known IDS datasets in addition to the dataset used. This is very important in evaluating the performance of IDS since KDD dataset is outdated and does not contain most novel attack patterns in it. In addition, these datasets are frequently used in the literature to evaluate the performance of IDS. Moreover, these datasets have various sample sizes and different numbers of features, so they provide a lot more challenges for comprehensively testing feature selection algorithms.
- Different from the detection framework proposed that designs only for binary classification, we design our proposed framework to consider multiclass classification problems. This is to show the effectiveness and the feasibility of the proposed method.

ADVANTAGES OF PROPOSED SYSTEM:

- FMIFS is an improvement over MIFS and MMIFS.
- FMIFS suggests a modification to Battiti's algorithm to reduce the redundancy among features.
- FMIFS eliminates the redundancy parameter required in MIFS and MMIFS.

CONCLUSION

Recent studies have shown that two main components are essential to build an IDS. They are a robust classification method and an efficient feature selection algorithm. In this paper, a supervised filter-based feature selection algorithm has been proposed, namely Flexible Mutual Information Feature Selection (FMIFS). FMIFS is an improvement over MIFS and MMIFS. FMIFS suggests a modification to Battiti's algorithm to reduce the redundancy among features. FMIFS eliminates the redundancy parameter _ required in MIFS and MMIFS. This is desirable in practice since there is no specific procedure or guideline to select the best value for this parameter. FMIFS is then combined with the LSSVM method to build an IDS. LSSVM is a least square version of SVM that works with equality constraints instead of inequality constraints in the formulation designed to solve a set of linear equations for classification problems rather than a quadratic programming problem. The proposed LSSVMIDS + FMIFS has been evaluated using three well known intrusion detection datasets: KDD Cup 99, NSL-KDD and Kyoto 2006+ datasets. The performance of LSSVM-IDS + FMIFS on KDD Cup test data, KDDTest+ and the data, collected on 1, 2 and 3 November 2007, from Kyoto dataset has exhibited better classification performance in terms of classification accuracy, detection rate, false positive rate and F-measure than some of the existing detection approaches. In addition, the proposed LSSVM-IDS + FMIFS has shown comparable results with other state-of the-art approaches when using the Corrected Labels sub-date set of the KDD Cup 99 dataset and tested on Normal, DoS, and Probe classes; it outperforms other detection models when tested on U2R and R2L classes. Furthermore, for the experiments on the KDDTest 21 dataset, LSSVM-IDS + FMIFS produces the best classification accuracy compared with other detection systems tested on the same dataset. Finally, based on the experimental results achieved on all datasets, it achieved promising performance in detecting intrusions over computer networks. Overall, LSSVM-IDS + FMIFS has performed the best when compared with the other state-ofthe- art models. Although the proposed feature selection algorithm FMIFS has shown encouraging performance, it could be further

enhanced by optimizing the search strategy. In addition, the impact of the unbalanced sample distribution on an IDS needs to be given a careful consideration in our future studies

REFERENCES

[1] S. Pontarelli, G. Bianchi, S. Teofili, Traffic-aware design of a high speed fpga network intrusion detection system, Computers, IEEE Transactions on 62 (11) (2013) 2322–2334.

[2] B. Pfahringer, Winning the kdd99 classification cup: Bagged boosting, SIGKDD Explorations 1 (2) (2000) 65–66.

[3] I. Levin, Kdd-99 classifier learning contest: Llsoft's results overview, SIGKDD explorations 1 (2) (2000) 67–75.

[4] D. S. Kim, J. S. Park, Network-based intrusion detection with support vector machines, in: Information Networking, Vol. 2662, Springer, 2003, pp. 747–756.

[5] A. Chandrasekhar, K. Raghuveer, An effective technique for intrusion detection using neuro- fuzzy and radial svm classifier, in: Computer Networks & Communications (NetCom), Vol. 131, Springer, 2013, pp. 499–507.

[6] S. Mukkamala, A. H. Sung, A. Abraham, Intrusion detection using an ensemble of intelligent paradigms, Journal of network and computer applications 28 (2) (2005) 167–182.

[7] A. N. Toosi, M. Kahani, A new approach to intrusion detection based on an evolutionary soft computing model using neurofuzzy classifiers, Computer communications 30 (10) (2007) 2201–2212.

[8] Z. Tan, A. Jamdagni, X. He, P. Nanda, L. R. Ping Ren, J. Hu, Detection of denial-of-service attacks based on computer vision techniques, IEEE Transactions on Computers 64 (9) (2015) 2519– 2533.

[9] A. M. Ambusaidi, X. He, P. Nanda, Unsupervised feature selection method for intrusion detection system, in: International Conference on Trust, Security and Privacy in Computing and Communications, IEEE, 2015.

[10] A. M. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, T. U. Nagar, A novel feature selection approach for intrusion detection data classification, in: International Conference on Trust, Security and Privacy in Computing and Communications, IEEE, 2014, pp. 82–89.

[11] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks 5 (4) (1994) 537–550.

[12] . Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, N. Yazdani, Mutual information-based feature selection for intrusion detection systems, Journal of Network and Computer Applications 34 (4) (2011) 1184–1199.

[13] A. Abraham, R. Jain, J. Thomas, S. Y. Han, D-scids: Distributed soft computing intrusion detection system, Journal of Network and Computer Applications 30 (1) (2007) 81–98.

[14] S. Mukkamala, A. H. Sung, Significant feature selection using computational intelligent techniques for intrusion detection, in: Advanced Methods for Knowledge Discovery from Complex Data, Springer, 2005, pp. 285–306.

[15] S. Chebrolu, A. Abraham, J. P. Thomas, Feature deduction and ensemble design of intrusion detection systems, Computers & Security 24 (4) (2005) 295–307.

[16] Y. Chen, A. Abraham, B. Yang, Feature selection and classification flexible neural tree, Neurocomputing 70 (1) (2006) 305–313.

[17] S.-J. Horng, M.-Y. Su, Y.-H. Chen, T.-W. Kao, R.-J. Chen, J.-L. Lai, C. D. Perkasa, A novel intrusion detection system based on hierarchical clustering and support vector machines, Expert systems with Applications 38 (1) (2011) 306–313.
[18] G. Kim, S. Lee, S. Kim, A novel hybrid intrusion detection method integrating anomaly detection with misuse detection, Expert Systems with Applications 41 (4) (2014) 1690–1700.
[19] P. Gogoi, M. H. Bhuyan, D. Bhattacharyya, J. K. Kalita, Packet and flow based network intrusion dataset, in: Contemporary Computing, Vol. 306, Springer, 2012, pp. 322–334.
[20] R. Chitrakar, C. Huang, Selection of candidate support vectors in incremental svm for network intrusion detection, Computers & Security 45 (2014) 231–241.
[21] H. F. Eid, M. A. Salama, A. E. Hassanien, T.-h. Kim, Bi-layer behavioral-based feature selection approach for network intrusion classification, in: Security Technology, Vol. 259, Springer, 2011, pp. 195–203.E. de la Hoz, A. Ortiz, J. Ortega, E. de la Hoz, Network anomaly classification by support vector classifiers ensemble and non-linear projection techniques, in: Hybrid Artificial Intelligent Systems, Vol. 8073, Springer, 2013, pp. 103–111. M. M. Abd-Eldayem, A proposed http service based ids, Egyptian Informatics Journal 15 (2014) 13–24.
[22] M. Tavallaee, E. Bagheri, W. Lu, A.-A. Ghorbani, A detailed analysis of the kdd cup 99 data set, in: Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications, 2009, pp. 1–6.
[23] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, K. Nakao, Statistical analysis of honeypot data and building of kyoto 2006+ dataset for nids evaluation, in: Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, ACM, 2011, pp. 29–36.
[24] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, 2012.
[25] M. S. Roulston, Estimating the errors on measured entropy and mutual information, Physica D: Nonlinear Phenomena 125 (3) (1999) 285–294.
[26] K. Fukunaga, Introduction to statistical pattern recognition, Academic press, 2013.
[27] Y.-I. Moon, B. Rajagopalan, U. Lall, Estimation of mutual information using kernel density estimators, Physical Review E 52 (3) (1995) 2318–2321.
[28] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min redundancy.