A nonlinear correlation measure for Intrusion Detection

Mohammed Ambusaidi, Liang Fu Lu, Xiangjian He*, Zhiyuan Tan, Aruna Jamdagni, Priyadarsi Nanda
Centre for Innovation in IT Services and Applications (iNEXT)
University of Technology, Sydney, Australia
Contact Email: xiangjian.he@uts.edu.au

Abstract—**The popularity of using internet contains some risks of network attacks. It has attracted the attention of many researchers to overcome this problem. One of the effective ways that plays an important role to achieve higher security and protect networks against attacks is the use of intrusion detection systems. Intrusion detection systems are defined as security techniques that tend to detect individuals who are trying to sneak into a system without authorization. However, one technical challenge in intrusion detection systems is high rate of false positive alarms which affect their performance. To solve this problem, we propose an effective method, which can accurately find the correlation between network records. In this work, we compare the results using a linear measure and a nonlinear measure based on correlation coefficient and mutual information. Experiments on KDD Cup 99 data set show that our proposed method using the nonlinear correlation measure can not only reduce the rate of false alarms but also efficiently distinguish normal and abnormal behaviors, and provide higher detection and accuracy rate then using the linear correlation measure.**

Keywords-Intrusion Detection; Nonlinear correlation; Mutual Information (MI); Pearson's Correlation coefficient;

I. INTRODUCTION

Defending networks against attacks and intrusions is a delicate task and a complicated act. It requires extensive resources, time and well-trained administrators with adequate knowledge. Intrusion detection and prevention systems, firewalls and antivirus software are security mechanisms that are working together in a complementary system to provide a safe and secure communication environment for neural networks. Although such mechanisms play an important role in defending networks, they are still susceptible to some limitations that affect their performance and these need to be addressed.

The rapid growth of malicious threats and cybercrimes encourages researchers to develop and propose numerous security algorithms and prevention mechanisms. Intrusion Detection Systems, which are basically designed based on the assumption that the intruder's behavior is different from a normal flow, are necessary to achieve higher security [1]. They are security management systems that are developed to create a second level of defense together with anti-virus software.

Intrusion detection techniques can be broadly categorized into two main classes. The first class is signature-based anomaly systems or misuse detection systems, where information and feature of past intrusions are used to detect ongoing anomalies [2], [3]. This technique tends to keep records of all the different types of attacks in a database called Attack Signature Database. They can identify the types of malicious attacks based on the built signatures and even activate the response to particular intrusion. Bro system is a popular example for this technique. It is a stand-alone system for monitoring communication links in real-time traffic and detects the incoming intrusions [3]. These systems are widely used because they are simple and efficient, and; have low number of false positive alarms. However, one of the disadvantages of these systems is that the detection accuracy and efficiency which are dependent on the quality of attack signatures. Furthermore, the extraction of high quality signatures requires the involvement of knowledgeable experts in extensive study of malicious behavior, which is costly and time consuming. Moreover, intrusion's signature is required before the system can detect the corresponding attack so these IDSs cannot detect any novel attacks due to the lack of their signatures.

The second class is anomaly detection systems, which recognize anomalous behavior from the normal flow [4]. Compared with signature-based detection, anomaly detection systems offer an advantage of detecting unidentified attacks. That is because they depend on statistical methods to calculate deviations from the normal traffic behaviors. They use Network History Database instead of Attack Signature Database to collect information about the normal behavior on a network and then create a model of normal flow. After that any deviation from the normal flow will be considered as suspicious behavior [5]. The main advantages of these approaches are the ability to recognize normal and abnormal attacks, and do not require a continuously update for their attack knowledge base. However, a major weakness of these techniques is that they are susceptible to high false positive rate and low detection accuracy especially with attacks that look like normal behavior. These limitations encourage us to focus on developing an anomaly detection system that can overcome the weakness.

Recent researches have widely used machine learning techniques on building IDSs. Machine learning techniques have the ability to improve and enhance the performance of detection algorithms. They can be used to increase the detection rate and decrease the false alarm rate in IDS [6]. Numerous machine learning techniques used in anomaly-based detection systems include Bayesian network [7], [8], nearest-neighbor methods [9], feature selection methods [10] and Markov models [11]. However, machine learning techniques still suffer from some challenges such as high false positive rate and low detection accuracy rate. Therefore, we propose an effective method, which can accurately find the correlation between network records. We name the correlation a nonlinear correlation coefficient-based Mutual Information (MI). The proposed method in this paper is sensitive to any relationship, not just the linear dependence [12]. It is more rational than the current Pearson's correlation coefficient (PCC) method in theory and can help intrusion detection techniques to improve their performance. In addition, we investigate our findings by proposing both a linear measure and a nonlinear measure based on correlation coefficient and mutual information for describing the relationship of random variables.

The rest of the paper is organized as follows. Section 2 provides current related works to our research. We describe the linear and nonlinear correlations based on Pearson's correlation and MI, in Section 3. Experiments and results are stated in Section 4. Finally, conclusions and future works are presented in Section 5.

II. RELATED WORKS

Correlation in the security point of view can be used to improve the threat identification and the assessment process by looking not only at individual events, but also at their groups. Nonlinear correlation coefficient (NCC) have been widely applied in many fields including fuzzy correlation method [13] and mechanical engineering [14]. In this paper, we will apply NCC based Mutual Information in the security field, which is sensitive to any relation, to measure the relationship between random traffic records.

Different IDSs have been developed using correlation technique to differentiate between the normal and abnormal flow. Ning et al. [15] proposed an intrusion detection technique based on correlation coefficient matrix. This anomaly detection method mainly focuses on calculating the quantitative correlation between TCP packets. Based on the assumption that intruders have different behavior than a normal user, Jin et al. in [16] developed a model utilizing the covariance matrix of sequential samples to detect multiple network attacks. They evaluated the correlation among the samples. In order to investigate the performance of their model, they applied two different statistical pattern recognition approaches: a threshold based detection approach and a traditional decision tree approach to detect anomalies. Their experimental results have showed that both approaches can distinguish multiple known attacks in the covariance feature space effectively.

Recently, there are some researches applying Mutual Information methods for feature selection to improve the performance of IDSs and identify the deviation behaviors from the normal flow. Sakar et al. in [17], modified MI and proposed a feature selection method named Predictive Mutual Information (PMI). The main objective of their method is to improve the feature detection capability especially in catching suspicious coincidences. They have applied their method on intrusion detection system and showed high rate in distinguishing between bad connections and normal connections. Amiri et al. stated that one of the challenges in IDS was the curse of high dimensional data [10]. To overcome this problem, they developed a feature selection algorithm using a modified mutual information-based feature selection (MMIFS). MMIFS is a feature selection method with maximum relevancy and minimum redundancy. They compared the performance of their method with two other feature selection methods: linear correlation–based feature selection (LCFS) and forward feature selection (FFSA). By applying the three methods in IDS, they have shown that MMIFS has higher classification accuracy and detection rate than LCFS and FFSA.

After several literature studies, we propose a nonlinear correlation method-based on Mutual information. The target of our method is high detection rate with reduced false rate.

III. LINEAR AND NONLINEAR CORRELATION ANALYSIS

The correlation of two variables is a statistical technique that can indicate the magnitude relationship between the two variables. It also shows how the two variables interact with each other. In this section, we present two correlation measures: Pearson's correlation and a nonlinear correlation based on MI measure.

A. Linear correlation coefficients

Pearson's correlation coefficient (PCC) [18] is one of the basic linear correlation methods used to measure dependence between two variables. It is defined as:

Assume *X* and *Y* are two random variables, x_i and y_i are their sample values for *i=1, 2,…….., n*. Their PCC is the covariance ratio of the variables divided by the product of their standard deviations.

$$Corr_{XY} = r = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ indicate the mean for *X* and *Y* respectively.

The value of PCC ranges in the closed interval [-1, 1], which indicates the linear correlation degree of the two variables. When the PCC value is close to 1 or -1, it denotes a strong relationship. If the PCC is close to 0, it means a weak relationship between the two variables. A positive or negative correlation coefficient denotes that both variables are in the same way or in the opposite way.

B. Nonlinear correlation information entropy

Although, linear correlation can detect the relationship between two dependent variables, in real world the correlations can also be nonlinear. Mutual Information, which is a quantity that measures a relationship between two discrete random variables, plays an important role in the computation of NCC. It can be thought of as a generalized correlation analogous to the linear correlation coefficient, but sensitive to any relationship including the linear correlation [19], [20]. The Mutual Information $I(X;Y)$ between two discrete random variables X and Y is given by

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2}$$

where $H(X)$ is the information entropy of the variable X, which is defined as

$$H(X) = -\sum_{i=1}^{N} p_i \ln p_i, \tag{3}$$

and $H(X,Y)$ is the joint entropy of X and Y defined by

$$H(X,Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{ij} \ln p_{ij}. \tag{4}$$

p_i denotes the probability distribution that random variable X will be in state x_i, and p_{ij} denotes the joint probability distribution that X is in state x_i and Y is in state y_j.

The disadvantage of Mutual Information is that it does not range in a definite closed interval [0, 1] as the correlation coefficient does [21]. Therefore, Wang et al. [22] developed a revised version of the MI named nonlinear correlation coefficient. The revised joint entropy of the two variables X and Y is given by

$$H^r(X,Y) = -\sum_{i=1}^{b}\sum_{j=1}^{b} \frac{n_{ij}}{N} \log_b \frac{n_{ij}}{N} \tag{5}$$

in which $b \times b$ rank grids are used to place the sample pairs $\{(x_i, y_i)\}_{1 \le i \le N}$, to the rank sequences of X and Y, and n_{ij} is the number of samples distributed in the *ij-th* rank grid. The nonlinear correlation coefficient is defined as

$$NCC(X;Y) = H^r(X) + H^r(Y) - H^r(X,Y), \tag{6}$$

where $H^r(X)$ is the revised entropy of the variable X, which is defined as

$$H^r(X) = -\sum_{i=1}^{b} \frac{n_{ij}}{N} \log_b \frac{n_{ij}}{N} \tag{7}$$

Therefore, the nonlinear correlation coefficient in (6) can be rewritten as

$$NCC(X,Y) = 2 + \sum_{i=1}^{b}\sum_{j=1}^{b} \frac{n_{ij}}{N} \log_b \frac{n_{ij}}{N} \tag{8}$$

The NCC in (8) can describe the relationships between two discrete random variables within a definite closed interval [0, 1]; with 0 and 1 indicates the weakest and the strongest relationship, respectively. In the following section, we apply (8) to measure the linear or nonlinear relationship between network records because of its sensitivity to any relation.

C. *Intrusion detection algorithm based on the correlation coefficient.*

In this section, we propose a method, which is mainly based on nonlinear correlation coefficient of different network traffic, to detect the intrusions among them. By analyzing the different correlation coefficients among the correlation matrix of the network traffic, we utilize a threshold value σ to judge whether the two different records are strong correlation or not, where σ belongs to the interval $[0,1]$.

Algorithm:

Step 1: Extract the data and generate the original matrix A.

Step 2: Compute the NCC of each two columns of A using formulas listed in the section B, which forms the correlation matrix of network traffic S_{ncc}.

Step 3: Detect the abnormal traffic from the whole data set.

3.1 Check the columns and find out the ones whose elements are lower than σ. From these values occurred in the columns, it can be considered that the records corresponding to the columns are abnormal.

3.2 After we get the abnormal records from the Step 3.1, we can easily find out elements corresponding to the records who are related to the abnormal ones in S_{ncc} that are larger than σ. It is natural that they can be considered as the abnormal records.

It should be noticed that the matrix we get in Step 2 is symmetric because of the same correlation coefficient between the *ij-th* and *ji-th* records. From the experience, we always set the threshold value σ to be 0.5 manually.

IV. EXPERIMENTS AND RESULTS

A. *The data set*

The data set used in these experiments is a 10 percent KDD Cup 99 data (kdd.ics.uci.edu//databases/kddcup99/kddcup99.html), which is a well-known dataset for intrusion evaluation. It consists of about five million connection records. Each record is defined as either normal or attack. The data set contains 24 attack types that have been categorized into four classes: Probe, Denial of Service (DOS), User to Root (U2R) and Remote to User (R2U) [23]. In order to measure the effectiveness of our method, we have chosen 1000 records randomly with six types of attacks including Smurf, Neptune, Land, Teardrop, Back and Pod attack. In fact, the method can work well for all types of attacks and the purpose of choosing these particular types of attacks is to test the effectiveness of the NCC. However, as a future work, we are planning to extend our experiments to more attacks. Each record in the data set, including normal and attack records, has 41 security parameters. We calculate the correlation between these records using both PCC and

NCC methods to identify the normal and abnormal records. Record that has strong relationship, close to 1, with other records that are considered to be normal traffic, can be judged as the normal one. On the other hand, the record that has weak relation, close to 0, can be flagged as suspicious behavior.

B. *Experimental results*

In our experiments, we set detection rate and false positive rate as standard measurements to evaluate the performance and effectiveness of our method. Firstly, to explain the details of our method, we have randomly selected six normal traffics with four different types of attacks from the chosen data set, and then calculated the corresponding correlation coefficient-based Pearson's correlation and nonlinear correlation coefficient-based Mutual Information. The following two matrices summarize the calculated results of PCC and NCC. Each element s_{ij} in the matrix behaves the linear and nonlinear coefficients between the *i-th* and *j-th records*:

$$S_{pcc} = \begin{pmatrix} 1 & 0.6785 & 0.8365 & 0.5164 & 0.9929 & 0.7926 & 0.5644 & 0.4858 & 0.6316 & -0.0405 \\ 0.6785 & 1 & 0.9592 & 0.9794 & 0.7463 & 0.9047 & 0.0533 & -0.0228 & 0.1043 & -0.0441 \\ 0.8365 & 0.9592 & 1 & 0.8877 & 0.8899 & 0.8864 & 0.2789 & 0.1021 & 0.2153 & -0.0465 \\ 0.5164 & 0.9794 & 0.8877 & 1 & 0.5969 & 0.8387 & -0.0984 & -0.0979 & -0.0355 & 0.0477 \\ 0.9929 & 0.7463 & 0.8899 & 0.5969 & 1 & 0.8184 & 0.5372 & 0.4073 & 0.5423 & -0.0266 \\ 0.7926 & 0.9047 & 0.8864 & 0.8387 & 0.8184 & 1 & 0.1059 & 0.3703 & 0.5020 & -0.0470 \\ 0.5644 & 0.0533 & 0.2789 & -0.0984 & 0.5372 & 0.1059 & 1 & 0.4045 & 0.4826 & 0.0096 \\ 0.4858 & -0.0228 & 0.1021 & -0.0979 & 0.4073 & 0.3703 & 0.4045 & 1 & 0.8173 & 0.3732 \\ 0.6316 & 0.1043 & 0.2153 & -0.0355 & 0.5423 & 0.5020 & 0.4826 & 0.8173 & 1 & -0.0423 \\ -0.0405 & -0.0441 & -0.0465 & 0.0477 & -0.0266 & -0.0470 & 0.0096 & 0.3732 & -0.0423 & 1 \end{pmatrix}$$

(a) Correlation Coefficient-based Pearson's correlation

$$S_{ncc} = \begin{pmatrix} 1 & 0.9747 & 0.9505 & 0.9262 & 0.9828 & 0.8575 & 0.0786 & -0.0226 & 0.0393 & -0.0369 \\ 0.9747 & 1 & 0.9019 & 0.9005 & 0.9303 & 0.9248 & 0.0419 & 0.0750 & 0.0068 & -0.0180 \\ 0.9505 & 0.9019 & 1 & 0.8564 & 0.9056 & 0.9019 & 0.0662 & -0.0092 & -0.0133 & 0.0124 \\ 0.9262 & 0.9005 & 0.8564 & 1 & 0.8810 & 0.8547 & -0.0984 & 0.0044 & 0.0178 & -0.0114 \\ 0.9828 & 0.9303 & 0.9056 & 0.8810 & 1 & 0.8575 & 0.1187 & 0.0419 & -0.3656 & -0.0224 \\ 0.8575 & 0.9248 & 0.9019 & 0.8547 & 0.8575 & 1 & 0.0419 & -0.0288 & 0.0068 & -0.0185 \\ 0.0786 & 0.0419 & 0.0662 & -0.0984 & 0.1187 & 0.0419 & 1 & 0.1383 & 0.0183 & -0.0021 \\ -0.0226 & 0.0750 & -0.0092 & 0.0044 & 0.0419 & -0.0288 & 0.1383 & 1 & 0.0514 & 0.0697 \\ 0.0393 & 0.0068 & -0.0133 & 0.0178 & -0.3656 & 0.0068 & 0.0183 & 0.0514 & 1 & -0.0347 \\ -0.0369 & -0.0180 & 0.0124 & -0.0114 & -0.0224 & -0.0185 & -0.0021 & 0.0697 & -0.0347 & 1 \end{pmatrix}$$

(b) Nonlinear correlation-based Mutual Information

Figure 1. Correlation Coefficient-based Pearson's correlation and Nonlinear correlation-based Mutual Information.

The following paragraphs explain how to identify the normal records and abnormal records. Assume that the first six records are normal and the last four records are abnormal. We set 0.5 as the threshold value for the correlation coefficients between these records, and any values less than 0.5 will be treated as anomaly.

From the results not only in Figure 1 but also in the other numerical results such as those shown in Figure 2 and Figure 3, it can be observed that our approach using NCC achieves higher detection accuracy than PCC. For example, we can observe form Figure 1(a), that the relationship between element (1,9)-th has PCC value of 0.6316 which indicates high correlation between records 1 and 9. However, element (1,9) shown in Figure 1(b) has NCC value of 0.0393, which indicates low correlation. By comparing the correlation of records 1 and 9 with other 9 records and considering the threshold value, we can notice that record number 1 has also high PCC correlation with records number 2 to 7 and 9. Thus it can be considered as a normal record. The PCC value for record number 9 also indicates high correlation with records number 5, 6 and 8, so it can be considered as a normal record as well. However, the NCCs between record 1 and records 2 to 6 are high, and the NCC between record 1 and 9 is low. As a result, we can conclude that record 9 is anomaly.

More specifically, Figure 2 and Figure 3 show the PCC and NCC values of one normal record as first record and seven Teardrop attacks in Figure 2 and seven Back attacks in Figure 3. The NCC values (0.0786 and 0.0885 0.0172, 0.0223) showed the relationships between 1st and the 2nd records, 1st and 3rd records in Figure 2 and 1st and 5th records and 1st and 6th records in Figure 4, respectively. These values point that the anomaly records must be hidden in records 2, 3, 5 and 6. However, PCC cannot detect these anomalies.

For training purposes, we have randomly chosen 4,210 records including normal and six types of attacks. The distribution of records is listed in Table I. Next subsection shows the results of performance comparison between NCC and PCC.

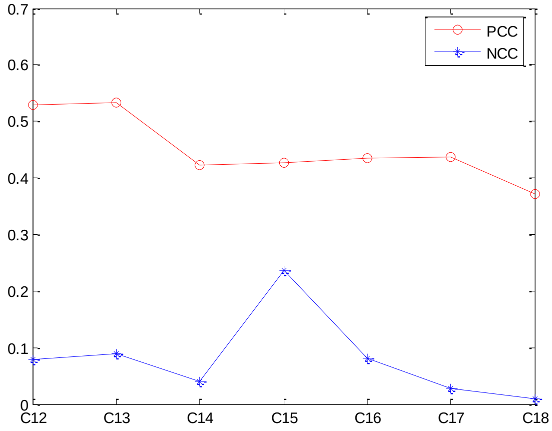

Figure 2. The comparison of the correlation coefficients calculated by PCC and NCC between one normal traffic and seven "Teardrop" attacks separately.

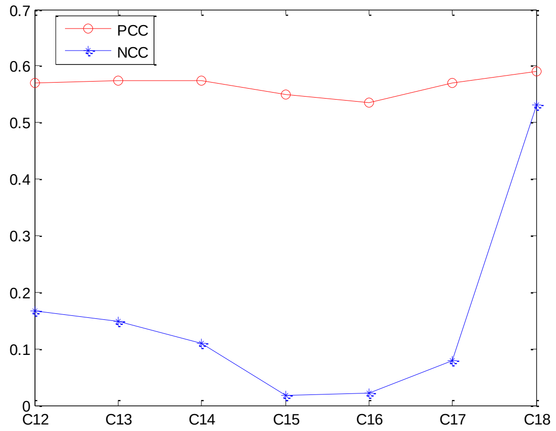

Figure 3. The comparison of the correlation coefficients calculated by PCC and NCC between one normal traffic and seven "Back" attacks separately.

The trained NCC-based MI is evaluated on the testing dataset containing normal and attack records. By considering the chosen threshold, the confusion matrix in Table I shows the correlation results of 2,352 normal records placed at the beginning of the testing dataset with 1,857 attack records. The distribution of records is listed in Table II. Next subsection shows the results of performance comparison between NCC-based MI and PCC methods.

TABLE I. CONFUSION MATRIX FOR NCC

Predicted **Actual**	**Normal**	**Attack**	**%Correct**
Normal	2319	33	98.59
Attack	21	1836	98.86

TABLE II. SAMPLE DISTRIBUTION ON THE TETSTING DATASET

Class	**Total number of records**
Normal	2353
Nebtune	939
Land	20
Smurf	266
Teardrop	164
Back	365
Pod	103
Total	4210

C. *Experimental analysis*

The quality of the detection techniques can be identified by the probabilities of detection rate and false positive rate. We compare these two measures of our method NCC-based MI with PCC method. Assuming that the first 2,353 records of the dataset are normal records, the comparison results are shown in Table III.

Zhiyuan et al. in [24] shows more detailed description in calculating the detection rate (DR) and false positive rate (FPR). The following formulas are used to compute accuracy, DR and FAR:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (9)$$

$$DR = \frac{TP}{TP+FN} \times 100\% \qquad (10)$$

$$FAR = \frac{FP}{FP+TN} \times 100\% \qquad (11)$$

where,

- TP: the number of actual attack classified as attack
- TN: the number of actual normal classified as normal
- FP: the number of actual normal classified as attack
- FN: the number of actual attack classified as normal

TABLE III. COMPARISON DETECTION RATE AND FALSE ALARM BETWEEN NCC AND PCC

Methods	**Measures**		
	Detection rate (%)	**False positive rate (%)**	**Accuracy (%)**
NCC-MI	98.86	1.4	98.69
PCC	96.6	3.96	96.27

As shown in Table III that the detection rate of NCC-based MI is 98.86% which outperforms the detection rate of PCC (95.69). In addition, for the false positive alarm rate NCC-based MI (1.4) also performs better than PCC (3.96). Furthermore, the accuracy rate of our method is 98.69%, which is better than the accuracy rate of the PCC method.

V. CONCLUSION AND FUTURE WORK

This paper has introduced a nonlinear correlation coefficient-based Mutual Information for measuring the relationship between discrete variables. This technique applies information theory and machine learning technique for calculating the correlation coefficient between network records. We have designed our method based on the assumption that intruder's behavior is different from a normal flow. One of the ways to identify the deviation from the normal flow is to calculate the relationship between communication records. Although, linear correlation can detect the relationship between two dependent variables, in real world the correlations can also be nonlinear. The proposed method is sensitive to any relationship, including linear relationship. We have investigated our findings by comparing with Pearson's correlation coefficient method. The experimental results have shown that NCC-based MI achieves higher detection rate, lower false alarms and accuracy rate than PCC.

However, the proposed method still needs further research in some aspects, which will be our target in the future works. For example, we will consider when the attack occurs and what the type of an attack is.

REFERENCES

[1] Hssanzadeh, A., and Sadeghian, B.: 'Intrusion Detection with Data Correlation Relation Graph'. Proc. Proceedings of the 2008 Third International Conference on Availability, Reliability and Security2008 pp. 982-989

[2] Vigna, G., and Kemmerer, R.A.: 'NetSTAT: a network-based intrusion detection approach'. Proc. Proceedings of the 14th Annual Computer Security Applications Conference 1998 pp. 22-34

[3] Paxson, V.: 'Bro: A System for Detecting Network Intruders in Real-Time', Computer Networks, 1999, 31, (23-24), pp. 2435-2463

[4] Barford, P., Kline, J., Plonka, D., and Ron, A.: 'A signal analysis of network traffic anomalies'. Proc. Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment, Marseille, France2002 pp. 71-82

[5] Jamdagni, A., Tan, Z., Nanda, P., He, X., and Liu, R.: 'Intrusion Detection Using Geometrical Structure'. Proc. Proceedings of the 2009 Fourth International Conference on Frontier of Computer Science and Technology2009 pp. 327-333

[6] Patcha, A., and Park, J.-M.: 'An overview of anomaly detection techniques: Existing solutions and latest technological trends', Computer Networks, 2007, 51, (12), pp. 3448-3470

[7] Kruegel, C., Mutz, D., Robertson, W., and Valeur, F.: 'Bayesian event classification for intrusion detection'. Proc. Proceedings of the 19th Annual Computer Security Applications Conference 2003 pp. 14-23

[8] Valdes, A., and Skinner, K.: 'Adaptive, Model-Based Monitoring for Cyber Attack Detection Recent Advances in Intrusion Detection', in Debar, H., Mé, L., and Wu, S. (Eds.) (Springer Berlin / Heidelberg, 2000), pp. 80-93

[9] Tsai, C.-F., and Lin, C.-Y.: 'A triangle area based nearest neighbors approach to intrusion detection', Pattern Recognition, 2010, 43, (1), pp. 222-229

[10] Amiri, F., Yousefi, M.R., Lucas, C., Shakery, A., and Yazdani, N.: 'Mutual information-based feature selection for intrusion detection systems', J. Netw. Comput. Appl., 2011, 34, (4), pp. 1184-1199

[11] Nong, Y., Yebin, Z., and Borror, C.M.: 'Robustness of the Markov-chain model for cyber-attack detection', Reliability, IEEE Transactions on, 2004, 53, (1), pp. 116-123

[12] Zhiyuan, S., Qiang, W., and Yi, S.: 'Effects of statistical distribution on nonlinear correlation coefficient'. Proc. Proceedings of the 2011 Instrumentation and Measurement Technology Conference (12 MTC), 2011 IEEE, pp. 1-4

[13] Yu, D., Hu, Q., and Wu, C.: 'Uncertainty measures for fuzzy relations and their applications', Applied Soft Computing, 2007, 7, (3), pp. 1135-1143

[14] Jardine, A.K.S., Lin, D., and Banjevic, D.: 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', Mechanical Systems and Signal Processing, 2006, 20, (7), pp. 1483-1510

[15] Chen, N., Chen, X.-S., Xiong, B., and Lu, H.-W.: 'An Anomaly Detection and Analysis Method for Network Traffic Based on Correlation Coefficient Matrix'. Proc. Proceedings of the 2009 International Conference on Scalable Computing and Communications; English International Conference on Embedded Computing 2009, pp. 238-244

[16] Jin, S., Yeung, D.S., and Wang, X.: 'Network intrusion detection in covariance feature space', Pattern Recognition, 2007, 40, (8), pp. 2185-2197

[17] Sakar, C.O., and Kursun, O.: 'A Hybrid Method for Feature Selection Based on Mutual Information and Canonical Correlation Analysis'. Proc. Proceedings of the 2010 20th International Conference on Pattern Recognition 2010, pp. 4360-4363

[18] Rodgers, J.L., and Nicewander, A.: 'Thirteen Ways to Look at the Correlation Coefficient', The American Statistician, 1988, 42 (1), pp. 59-66

[19] Roulston, M.S.: 'Significance testing of information theoretic functionals', Physica D: Nonlinear Phenomena, 1997, 110, (1–2), pp. 62-66

[20] Roulston, M.S.: 'Estimating the errors on measured entropy and mutual information', Physica D: Nonlinear Phenomena, 1999, 125, (3–4), pp. 285-294

[21] Shen, Z., Wang, Q., and Shen, Y.: 'A new non-liner correlation measure'. Proc Proceedings of the 2009 IEEE Youth Conference on Information, Computin and Telecommunication 2009, pp. 11-14

[22] Wang, Q., Shen, Y., and Zhang, J.Q.: 'A nonlinear correlation measure for multivariable data set', Physica D: Nonlinear Phenomena, 2005, 200, (3–4), pp. 287-295

[23] Mukkamala, S., Sung, A.H., and Abraham, A.: 'Intrusion detection using an ensemble of intelligent paradigms', Journal of Network and Computer Applications, 2005, 28, (2), pp. 167-182

[24] Zhiyuan, T., Jamdagni, A., Xiangjian, H., and Nanda, P.: 'Network Intrusion Detection based on LDA for payload feature selection'[M], Network Intrusion Detection based on LDA for payload feature selection (2010), pp. 1545-1549