A Novel Dependability Model to Define Normal Network Behavior

Maher Salem
Network and Data Security,
Applied Computer Sciences
University of Applied Sciences Fulda
Fulda, Germany
Maher.salem@informatik.hs-fulda.de

Ulrich Buehler
Network and Data Security,
Applied Computer Sciences
University of Applied Sciences Fulda
Fulda, Germany
u.buehler@informatik.hs-fulda.de

***Abstract*—Computer networks augment in heterogeneity so that defining a normal behavior to the network becomes a severe challenge. Particularly, such a normal network behavior is essential for security issues. In addition, this behavior consolidates the intrusion detection system to significantly detect zero-day-attacks. Therefore, in this paper, we introduce a novel dependability model based on the correlation matrix of network features. Moreover, only strongly correlated features are involved in the model such that the normal connections are recognized into the online traffic in advance. The recognition is based on the distance of the incoming traffic to the linear association between the correlated features. Furthermore, the distance is compared to a threshold value to ensure correct recognition. These steps have been evaluated by the benchmark dataset NSL-KDD. The goal of this model is to build an adaptive normal network behavior that represents the intended network continuously, reduces the overhead on the classification, and supports by detecting unknown attacks respectively. The results show that the idea of dependability model in intrusion detection system promises more accuracy and preciseness in anomaly detection.**

Keywords-correlation matrix; dependability; normal network behavior; linear association.

I. INTRODUCTION

Intrusion Detection Systems (IDS) approaches can be classified into misuse detection and anomaly detection [1], misuse detection systems are using signatures to detect known attack patterns. However, they are suffering under the constantly growing number of signatures and incapability of detecting unknown attacks as well. In contrast, anomaly based intrusion detection systems are able to detect known and even unknown attacks by recognizing the deviation from the normal network behavior. Accordingly, there are two main approaches to characterize normal network behavior presented in [2], which are studying the inference of the overall network behavior through the use of network probes and the understanding of the behavior of the individual entities or nodes. Principally, IDS analyzes and studies the network traffic to establish a profile that defines a normal network behavior (NNB). Upon this profile, the IDS can detect any deviation as an anomaly and consider it most likely an attack. Presenting a significant and heuristic model that defines the normal behavior is imperative the area of networking. Therefore, we present a novel model that defines a NNB by building a dependability model from the strength of features correlations. The main idea is to capture the online traffic in the real time and match the traffic to the dependability model to investigate its normality. Moreover, for the positive strong correlations between two features a linear association is defined to the best fit of the concerned features. This is to imply that, when two feature vectors are strongly correlated there is significant of determination that ensures a linear association between these features. Thus, such relation can be exploited to determine the normality in the incoming traffic based on the linearity of correlated features. The dependability model can then be updated with the new normal traffic. This paper is structured as following; in section II, a motivation about the proposed methods in normal network behavior is discussed. On the hand, section III presents our novel methodology. Section IV describes the preparation of dataset. Section V illustrates our results and discussion. Finally, section VI concludes our work.

II. MOTIVATION

Intrusion detection system steps are summarized into feature selection, discretization, normalization, and classification. Regarding classification the IDS builds a normal profile of the network and detect the deviation from this profile.

A real time visualization platform in [3] presents a multiple visualization techniques that provide a situational understanding of real time network activity. Such platform can visualize million of records and report the network current status. However, it may not feasible in IDS research area. On the other hand, [4] proposed a modeling approach where failures and repairs of network components as well as routing and traffic handling are described by a set of stochastic activity networks (SAN). The proposed model approach serves more in the area of network routing and availability of end-to-end network components. However, it could be exploited to build a normal network behavior.

A significant work [5] explained a correlated node behavior model based on Markov process [19] for dynamic topology network. Thus, the latter classified the nodes into four categories and show that the effect from correlated failure nodes on network survivability is much severer than

 ISBN: 978-1-61208-227-1

other misbehaviors. However, this approach is investigating network nodes and builds a behavioral model accordingly. A reasonably network behavior tool in [6] exploits only the internal network traffic to monitor the internal activities on network so that a deviation from a predefined pattern model is detected as abnormal behavior. This model aims to detect anomaly indeed but it examines only the internal traffic.

A structural model in [7] utilizes web logs to analyze user behavior based on the web-context and situation-awareness. Obviously, this model focused only on the user activities and ignores the network ones. More sufficient proposals regarding analysis of system behavior are proposed in [8],[9], and [10].

So, defining a novel model for normal network behavior is needed. Therefore, we principally focus on the network traffic to build such a model and express it as linear relations.

III. PROPOSED METHODOLOGY

In this research work, we concentrate on the definition of a normal network behavior, based on its traffic, which represents the network. In contrast to [14], we build a dependability model graph based on the correlation matrix to define a normal network behavior and to predict the normal connection in real time. Thus, the proposed idea in this work exploits network traffic statistics to build a dependability model from the feature correlations; that is, from correlation matrix. In addition, the model will be able to detect normal connections based on the linear association between the correlated features. However, selecting the most valuable features out of hundreds of network features is a provocation step. Hence we declare the proposed idea in three steps:

A. Significant Network Features

Selection of the valuable features in the area of IDS is a negotiable point in data mining research. Thus, we used the improved feature selection method proposed in [11]. It presented a novel method that abstracted the valuable features in the network based on the sequential backward search and information gain. The difference between both feature sets is that features in the most valuable feature set affect definitely the detection rate, whereas features in the most valuable and relevant feature set affect definitely the detection rate and enhance it, i.e. $MVF \subset MVRF$. Moreover, the model has been evaluated on the benchmark dataset NSL-KDD [18]. The exploited features are summarized in Table I.

TABLE I. MOST VALUABLE FEATURE SET AND MOST VALUABLE AND RELEVANT FEATURE SETS

Name of feature set	**features**
Most Valuable Features (MVF)	service, src_bytes, dst_host_serror_rate, serror_rate, dst_host_srv_diff_host_rate, protocol_type,rerror_rate,srv_rerror_rate,wrong_fragment,num_compromised,num_access_files
Most Valuable and Relevant Features (MVRF)	service, src_bytes, diff_srv_rate, same_srv_rate, dst_host_srv_count, logged_in, dst_host_serror_rate, serror_rate, srv_serror_rate, dst_host_srv_diff_host_rate, protocol_type, rerror_rate, srv_rerror_rate, hot, wrong_fragment, num_compromised, num_access_files, root_shell, num_failed_logins

In principal, we build the dependability model from the strongly correlated features of these feature sets and hence define a normal behavior for the network.

B. Correlation and Dependability Model

Network traffic has several features, which are somehow sharing an association. One of the most known methods to infer these associations is the correlation between features; that is, the correlation is used to determine the degree of association between two features [12]. Hence, let us define two network features (vectors) X and Y, which are normally distributed, such that $X=\{x_1,..., x_n\}$ and $Y=\{y_1,...,y_n\}$ where $n \in \mathrm{N}$, $x_i, y_i \in \mathfrak{R}$. Then the Pearson's correlation coefficient is

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n(\bar{x})(\bar{y})}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}} \quad (1)$$

where $\bar{x}, \bar{y}$ are the mean values of feature X and feature Y, respectively. The correlation value between two features falls between [-1,+1], so that the more positive the value, the more significant the linear association. Thus, the correlation matrix established for m network features $F_1,..,F_m$ can be built as

$$\mathbf{Corr}_{F1,..,Fm} = \begin{bmatrix} r_{11} & \dots & r_{1m} \\ \dots & \dots & \dots \\ r_{m1} & \dots & r_{mm} \end{bmatrix} \quad (2)$$

The correlation matrix is a symmetrical one, and the correlation value of the same feature is always +1. Other values of r_{ij} could be positive or negative, e.g. 0.8 means that 80% of the changes in one feature are related to the other. On the other hand, the coefficient of determination R^2 (or r_{ij}^2) of the two features F_i *and* F_j means that the percentage of variability in one feature related to variability to the other feature. In addition, R^2 gives the proportion of the variance of one feature explained by the other [13], e.g., if the value of the coefficient determination is 0.8 that indicates about 80% of the variance of one variable is explained by the other. Furthermore, it ensures about the prediction of the feature $\bar{y}$ (predicted y) from the linear association instead of the mean value. Therefore, we consider the value of coefficient of determination to assure linear association between features. Consequently, only the strongly correlated features, which reject the null hypothesis $H_0 : \rho = 0$, are considered in this research work. The Greek symbol rho is the parameter used for nonlinear correlation. However, the null hypothesis is the most common used with Pearson's correlation coefficient [20] such that the

correlation coefficient is zero and there is no linear association between the two variables. In this regard, we use the critical P-value with 0.05 of making error type 1 to check whether the correlation value between two features rejects the null hypothesis and have a linear association as well or not. Accordingly, we abstract precisely only the significant linear association between the strong correlated features.

C. Linear Association and Prediction

If two network features are strongly correlated, then they have a linear association that describes their correlation. The linear association between two strongly correlated features $X=\{x_1,..., x_n\}$ and $Y=\{y_1, ...,y_n\}$ can be defined as

$$y_i = \omega_0 + \omega_1 x_i + \varepsilon_i \quad (3)$$

where ω_0 is the intercept and ω_1 is the slope. The idea of least squares is exploited to find the choice of slope and intercept that give the best fit among the data points. In addition, the parameter ε_i is the normally distributed random error. In this research paper we abstract the linear line to the best fitting of the scatter data, i.e., the association is definitely not 100% linear, so that a percentage of error in the linearity and prediction is expected as well. Hence, suppose we have a pair (x_i,y_i) that is not fitting exactly on the linear line, so we can determine the distance of the point to the line as in [17], such that

$$d = \frac{|y_i - \omega_1 x_i - \omega_0|}{\sqrt{\omega_1^2 + 1}} \quad (4)$$

The distance from the linear line will be used to check if the incoming online traffic belongs to the linearity between the correlated two features or not based on a certain threshold, mainly the maximum distance d_{max}.

In brief, we select the valuable and significant network features, infer the correlation values between them, establish a correlation matrix, indicate the strong positive correlation values via rejecting the null hypothesis, find out the best fitting linear line between each two correlated features, and then detect, for the online traffic, the normal connections based on the distance from the linear line. Finally, detected normal connections will be used to rebuild the model.

For example, suppose we have three network features namely F_1, F_2, and F_3 and the values belonging to these features are shown in Table II.

TABLE II. FEATURE VALUES

F_1	F_2	F_3
0.1	0.01	0.3
0.2	0.03	0.46
0.9	0.09	0.37
0.3	0.035	0.08
0.6	0.063	0.011
0.71	0.073	0.93

According to the correlation coefficient in equation (1) the correlation coefficient matrix between these features is

$$\mathbf{Corr}_{F1,F2,F3} = \begin{bmatrix} 1 & 0.9936 & 0.2674 \\ ... & 1 & 0.2750 \\ ... & ... & 1 \end{bmatrix} \text{ P-Value} = \begin{bmatrix} 1 & 0.009 & 0.5678 \\ ... & 1 & 0.449 \\ ... & ... & 1 \end{bmatrix}$$

Both matrices are symmetric so that the lower region with (…) is the same as the upper region. Obviously, the correlation between F_1 and F_2 has a P-value with 0.009, i.e., the P-value<0.05 so that the null hypothesis is rejected at 5% significant level. Whereas, other correlation values have P-values greater than 0.05 which imply that no linear association is existed between them. Therefore, only one linear association is existed with an intercept of 0.0058 and a slope value of 0.095. Moreover, the maximum distance from the pairs to the line is 0.0053. Figure 1 shows the plot of this example.

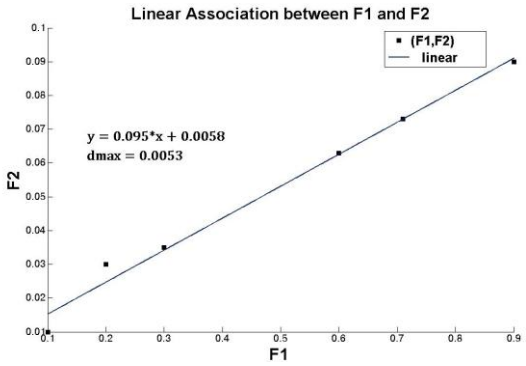

Figure 1. Linear Association between the features in the example.

Accordingly, if we receive a new feature instance in real time such that the values of F_1, F_2, and F_3 are {0.2, 0.1, 0.1} respectively then we need to calculate the distance between the pair (0.2,0.1) and the linear line then compare it to d_{max} value.

IV. DATASET PREPARATION

To evaluate our proposed method, we build a test computer network, so that only clean traffic will be aggregated. Hence, no external connections are allowed and no any USB devices will be plugged. We aggregate the traffic based on the feature set MVRF in Table I. On the other hand, a dataset with only the normal traffic from NSL-KDD is generated to establish a correlation matrix from the positively correlated features and those have linear association. Furthermore, from the correlated features a dependability model will be designed so that only features with linear association are conducted. According to the test network, we could abstract up to 15 features from the MVRF and still struggling to achieve the rest ones. Therefore, we will first test and evaluate our proposal with the normal traffic from NSL-KDD and discuss our results.

Generally, we cannot just filter a dataset out from NSL-KDD and calculate a Pearson's correlation coefficient, but we should digitize it and then normalize it. To achieve such numeric and normalized dataset we exploit the hybrid normalization method in [15] to map the nominal values into numeric and then normalize the dataset using minimum maximum normalization. Thus, a minimum maximum normalization is defined as

 ISBN: 978-1-61208-227-1

$$nv = f(v) = \frac{v - \min(v)}{\max(v) - \min(v)} \quad (5)$$

where $f: \Re \rightarrow [0,1]$ be the normalization function and $v \in \Re$ the numerical value of a feature in the feature sets, nv the normalized feature value after normalization process.

V. RESULTS AND DISCUSSION

To evaluate the novelty of the proposed method, we exploited the NSL-KDD, so that only normal traffic is abstracted and then the dataset is normalized. The selected features in our evaluation are the MVRF. Hence, a dataset with 65555 normal instances is initialized for testing and evaluation. In the following table these features are numbered to ease the explanation of our results.

TABLE III. SELECTED FEATURES IN MVRF

Feature Set	Feature number.feature name
Most Valuable and Relevant Features (MVRF)	**1**.Protocol_type, **2**. Service, **3**.scr_bytes, **4**.wrong_fragment, **5**.hot, **6**.num_failed_logins, **7**.logged_in, **8**.num_compromised, **9**.root_shell, **10**.num_access_files, **11**.serror_rate, **12**.srv_serror_rate, **13**.rerror_rate, **14**.srv_rerror_rate, **15**.same_srv_rate, **16**.diff_srv_rate, **17**.dst_host_srv_count, **18**.dst_host_srv_diff_host_rate, **19**.dst_host_serror_rate

We developed a Matlab program to calculate the correlations between features and the coefficient of determination. Due to space limitation we present a small part of the Correlation matrix and significant determination as well.

$$\mathrm{Corr}_{F1,..F19} = \begin{bmatrix} 1 & 0.5692 & ... & 0.0523 \\ ... & 1 & ... & -0.1318 \\ ... & ... & 1 & ... \\ ... & ... & ... & 1 \end{bmatrix} \quad R^2 = \begin{bmatrix} 1 & 0.3240 & ... & 0.0027 \\ ... & 1 & ... & 0.0174 \\ ... & ... & 1 & ... \\ ... & ... & ... & 1 \end{bmatrix}$$

According to the evaluated dataset MVRF, we calculate the correlation between 19 features, see Table III. Then determine the coefficient of determination, so that only the best linear association between features is considered. Table IV shows the positive correlated features from the correlation matrix, so that the correlation rejects the null hypothesis, (also, the linear equation associated between these features).

TABLE IV. STRONGLY POSITIVE CORRELATED FEATURES IN MVRF

Correlated features	r	R^2	Linear line
1↔2	0.5479	0.302	$y_i = 0.008 + 0.52x_i + \varepsilon_i$
1↔7	0.7905	0.625	$y_i = 0.23 + 1.4x_i + \varepsilon_i$
2↔7	0.5189	0.270	$y_i = 0.36 + 0.98x_i + \varepsilon_i$
11↔12	0.8748	0.765	$y_i = 0.003 + 0.8x_i + \varepsilon_i$
13↔14	0.9829	0.966	$y_i = 0.0013 + 0.98x_i + \varepsilon_i$
2↔17	0.7070	0.5	$y_i = 93 + 2.7x_i + \varepsilon_i$
15↔16	0.7620	0.58	$y_i = 0.77 - 0.77x_i + \varepsilon_i$

In addition, these equations represent the network traffic when the network is in its normal behavior. Of course, one of the drawbacks of NSL-KDD is that there is no 100% linearity between the correlated features. Therefore, we expect an error (false positive) when detecting the incoming online traffic based on these equations. Figure 2 shows a linear relation between two features. Based on this figure, most of the data fit on the linear line and other are on the area around, so we can consider the points on the line or nearby are only the related ones to this equation. To do so, a maximum distance must be determined and an error value must be defined such as $d_{max} \leq 0.5$ and $\varepsilon_i \approx 0.0005$ where d_{max} is determined from $\{(Point_{max} - Point_{min})/2\}$ and because the dataset is normalized then $Point_{max} = 1$ *and* $Point_{min} = 0$. We found after testing several cases that the maximum distance is the average between the minimum point and maximum point from the normalized dataset. On the other hand, the error is selected manually to a small value to avoid incorrect distance calculation.

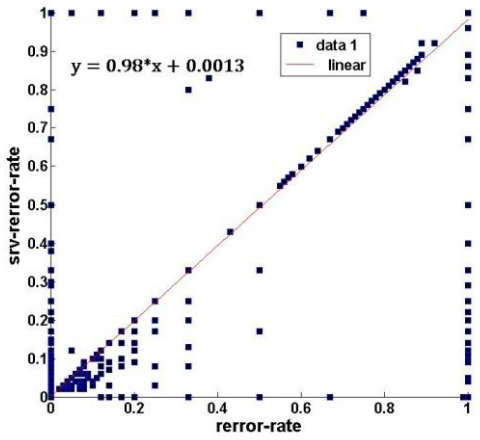

Figure 2. Linear Association between rerror_rate and srv_rerror_rate.

In contrast, Figure 3 shows more stable linear association between the feature *protocol_type* and *logged_in* so that a better detection is expected.

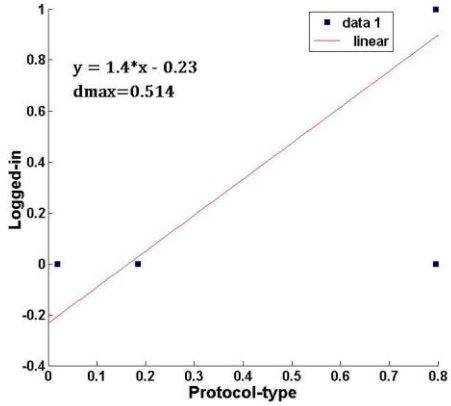

Figure 3. Linear Association between protocol_type and logged_in.

The figure shows few points because the value of the feature *logged_in* is mostly 1 or 0, so that several points are overwritten. Maximum distance is calculated from the longest distance to the line.

Moreover, although some correlated features have a high value of coefficient of determination, they could have no adequate linear association. Therefore, we prune the association with a small value of error and ignore the point on the border to achieve better linearity. So, we have derived various linear equations from the correlated features. Therefore, we present a dependability model that shows the correlated features and hence their dependencies (correlation coefficient values). Intuitively, the concerned pairs from online traffic will be matched to the related linear line. Figure 4 depicts a general dependability model of the MVRF that represents the normal network behavior based on the offline dataset NSL-KDD.

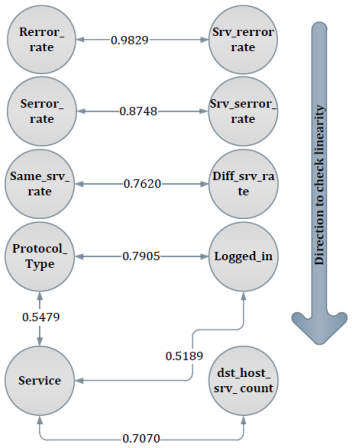

Figure 4. Dependability model of MVRF.

This model implies that, these features are strongly correlated when the network traffic is normal. That means, based on the benchmark dataset NSL-KDD and the selected feature set MVRF, this model can be used to analyze the online traffic directly and detect the normal connections or the abnormal ones. Generally, the online traffic is prepared so that firstly the distance from the pair (*rerror_rate*, *srv_rerror_rate*) and the linear line $y_i = 0.0013 + 0.98x_i + \varepsilon_i$ is calculated and compared to the value of d_{max}. Consequently, the distance from the pair (*serror_rate*, *srv_serror_rate*) and the linear line $y_i = 0.003 + 0.8x_i + \varepsilon_i$ is calculated and compared to the value of d_{max}. In the last step, distances must be evaluated so that all must fulfill the condition $d \leq d_{max}$. Finally, if the online traffic is detected as normal it will be considered in the dataset to adjust the linearity accordingly, such that the dependability model stays adaptive.

To test this model and the proposed idea of detecting the normal traffic in real time, we selected arbitrary instances from the NSL-KDD dataset so that two instances are normal and the third is anomaly. We calculated the distances as defined before. Moreover, the error value is fixed to 0.0005 for all linear lines just to ease the calculation of all distances. The instances are shown in Table V, instances values are sorted the same way as in Table III. Moreover, Table VI depicts the detection result for the test instances. It proved that the proposed idea could detect all instances significantly based on the distance from the linear line.

TABLE V. TEST INSTANCE DATA FROM NSL-KDD

Instance number	Instance values (normalized)
Instance 1	0.7957,0.5646,0,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0.08,0
Instance 2	0.1848,0.1343,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0.9882,0,0
Instance 3	0.8152,0.0545,0.0037216,0,0.051948,0,1,0,0,0,0.9,0.9,0,0,1,0,0.15294,0.13,0

The first two instances are normal and the third one is anomaly. Table VI shows the results, so that linearity means the linear line between the correlated features; they are sorted according to the strong linearity in descending order. Moreover, d_{max} is the maximum distance for each linear line between features, **d** means in the table the distance of this instance from the linear line, as mentioned before, the concerned pair from each instance is selected for the suitable linear line. The following table summarizes the testing results.

TABLE VI. TEST RESULT OF SELECTED INSTANCES

Linearity	d_{max}	Instance 1 (d)	Instance 2 (d)	Instance3 (d)
13↔14	0.02	0.00092	..	0.0009
11↔12	0.02	0.001	..	**0.139**
1↔7	0.514	0.199	..	0.22
15↔16	0.2	0	..	0
2↔17	0.65	0.6	..	0.08
1↔2	0.4	0.336	..	**0.4**
2↔7	0.64	0.064	..	0.42

Obviously, the distances of instance1 are all minimum than the maximum distance of each linear association, so it is a normal traffic. In contrast, instance3 has two distances (in bold) greater than the maximum distance in the linear line for the intended correlated features, which is lead to predict this traffic as anomaly. Therefore, the new detected normal instances will be added to the normal dataset so that a new linear line and dependability model with a roughly modified correlation values will be enhanced.

Principally, we focus in this paper on the idea of dependability model and how it represents the normal network behavior. Hence, to declare this idea we have introduced a test and evaluation example from an offline dataset. But we have explained how to use this idea to predict online normal traffic using the distance threshold.

Another main discussion point is the linearity between features. We notice that when the dataset increases the features are not more correlated and they lose the linearity. The association between these features becomes nonlinear. Therefore, in the incoming research work we will exploit the idea in [16], so that a nonlinear association between features

 ISBN: 978-1-61208-227-1

will be established by exploiting the idea of Maximum Information Coefficient (MIC).

VI. CONCLUSION AND FUTURE WORK

Defining a normal network behavior is a necessary step in intrusion detection system. However, it is a challenge under research in the data mining area. In this research work, we present a novel dependability model from the positive correlations between network features. In addition, we abstract the linear associations between these correlated features and exploit them to predict the normal connection from the online traffic in the real time. The prediction is examined so that each features pair from the online traffic instance is exploited to calculate their distance from the linear line related to these pair exclusively. Furthermore, all distances of all feature pairs in the online traffic must be greater than the threshold distance (d_{max}) to consider it a linear connection. Our test results show that the model could detect the normal connection and anomaly as well from test dataset NSL-KDD. In addition, we looked to enhance the model by examine the nonlinear association in large dataset. Finally, we proved that the dependability model can represents the normal network behavior and can support the IDS to detect the attacks in online traffic. Therefore, it promises more accuracy, less overhead in classification, and enhancement in network performance.

ACKNOWLEDGMENT

This research project “SecMonet” is supported by the German Federal Ministry of Education and Research (BMBF) under the funding line “FHprofUnt”.

REFERENCES

[1] R. Karthick, V. P. Hattiwale, and B. Ravindran, “Adaptive Network Intrusion Detection System using a Hybrid Approach,” IEEE Fourth International Conference in COMSNETS, pp. 1-7, Januray, 2012.

[2] M. Thottan and C. Ji, “Anomaly Detection in IP Networks,” IEEE Transactions On Signal Processing, vol. 51, NO. 8, pp. 2191-2204, August, 2003, doi: 10.1109/TSP.2003.814797.

[3] D.M. Best, S. Bohn, D. Love, A. Wynne, and W. Pike, “Real-Time Visualization of Network Behaviors for Situational Awareness,” ACM Proceedings of the Seventh International Symposium on Visualization for Cyber Security, pp. 79-90, 2010, doi: 10.1145/1850795.1850805.

[4] Q. Gan and B. E. Helvik, “Dependability Modelling and Analysis of Networks as Taking Routing and Traffic into Account,” IEEE Conference in Next Generation Internet Design and Engineering, 2006, doi: 10.1109/NGI.2006.1678219.

[5] A.H. Aznin, R. Ahmad, Z. Muhamad, A. Basari, and B. Hussin, “Correlated Node Behavior Model based on Semi Markov Process for MANETS,” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January, 2012, ISSN:1694-0814.

[6] S. Kakuru, “Behavior Based Network Traffic Analysis Tool,” IEEE third conference in communication software and networks (ICCSN), pp. 649–652, 2011, doi: 10.1109/ICCSN.2011.6014810.

[7] W. Liu, D. Huang, and L. zhang, “Analysis of Network User Behavior” IEEE youth conference on information computing and telecommunications, pp. 126-129, November, 2010, doi: 10.1109/YCICT.2010.5713061.

[8] M. Burgess, H. Haugerud, S. Straumsnes, and T. Reitan, „Measuring System Normality“ ACM Transactions on Computer Systems, Vol. 20, No. 2, pp. 125-160, May, 2002, doi: 10.1145/507052.507054.

[9] J. M. Estevez-Tapiador, P. Gracia-Teodoro, and J. E. Diaz-Verdejo, “Measuring normality in HTTP traffic for anomaly-based intrusion detection” Elsevier Computer Networks, pp. 175-193, 2004.

[10] M. V. Mahoney and P. K. Chan, “Learning nonstationary models of normal network traffic for detecting novel attacks” ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 376-385, 2002, doi: 10.1145/775047.775102.

[11] M. Salem, U. Buehler, and S. Reissmann, “Improved Feature Selection Method using SBS-IG-Plus”, ISSE Proceeding on Securing Electronic Business Processes, pp. 352-361, 2011.

[12] A.G. Asuero, A. Sayago, and A.G. Gonzalez, “The Correlation Coefficient: An Overview” Critical Review in Analytical Chemistry, pp. 41-59, 2006, doi: 10.1080/10408340500526766.

[13] J. Freeman and T. Young, “Correlation Coefficient: Association Between Two Continuous Variables” Scope, pp. 31-33, June, 2009.

[14] N. Chen, X. Chen, B. Xiong, and H. Lu, “An Anomaly Detection and Anaylsis Method for Network Traffic Based on Correlation Coefficient Matrix,” IEEE conference on Embedded Computing, pp. 238-244, 2009, doi: 10.1109/EmbeddedCom-ScalCom.2009.50.

[15] M. Salem, U. Buehler, “ Hybrid Normalization Method in Data Mining Toward Improving the Network Intrusion Detection System” submitted to the IEEE conference on Data Mining, ICDM 2012.

[16] D. Reshef, et al. “ Detecting Novel Association in Large Data Sets”, 2011, doi: 10.1126/science.1205438.

[17] http://math.ucsd.edu/~wgarner/math4c/derivations/distance/distptline.htm.

[18] Nsl-kdd dataset: http://nsl.cs.unb.ca/NSL-KDD/, March, 2009.

[19] Stroock, D. “An Introduction to Markov Processes”. Graduate Text Series #230, Springer–Verlag, Heidelberg, 2005.

[20] Edwards, A. L. "The Correlation Coefficient." Ch. 4 in An Introduction to Linear Regression and Correlation. San Francisco, CA: W. H. Freeman, pp. 33-46, 1976.

 ISBN: 978-1-61208-227-1