Journal of Theoretical and Applied Information Technology
31st October 2018. Vol.96. No 20
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID JAMES-STEIN AND SUCCESSIVE DIFFERENCE COVARIANCE MATRIX ESTIMATORS BASED HOTELLING'S T^2 CHART FOR NETWORK ANOMALY DETECTION USING BOOTSTRAP

[1]MUHAMMAD AHSAN, [2]MUHAMMAD MASHURI, [3]HIDAYATUL KHUSNA

[1,2,3] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

E-mail: [1]ahsan4th@gmail.com, [2]m_mashuri@statistika.its.ac.id, [3]khusna16@mhs.statistika.its.ac.id

ABSTRACT

Statistical process control (SPC) is one of the powerful statistical methods that continuously improves the manufacturing process. The advantage of using the method in network anomaly detection is the technique does not need the knowledge of an information from the previous intrusions. The Hotelling's T^2 is the mostly used control chart for network intrusion detection. However, Hotelling's T^2 chart, which uses the conventional mean and covariance matrix, is sensitive to the outlier presence. Therefore, the conventional method is not effective to be implemented in Intrusion Detection System. To overcome this problem, Successive Difference Covariance Matrix (SDCM), which is one of the robust covariance matrix estimators, can be implemented in estimating the covariance matrix. Meanwhile, the James-Stein estimator can be adopted in estimating the mean vector of the Hotelling's T^2 control chart. The utilization of the bootstrap resampling method is intended to obtain the more accurate control limit of the proposed chart. The combination of these estimators with the bootstrap resampling approach demonstrates the better performance when it is used to monitor the anomaly in the network than the other control limit approaches in training and testing dataset. In addition, the IDS based on the proposed chart has better performance than the other existing charts based on its hit rate and FN rate criteria. The proposed method also outperforms some classifier methods.

Keywords: *T^2 control chart, james-stein, successive difference covariance matrix, bootstrap, network anomaly detection*

1. INTRODUCTION

The most popular tool used in SPC is control charts. Based on the quality characteristics type used, control charts are categorized into two which are attribute and variable control chart. While based on the number of quality characteristics monitored, control charts are differentiated into univariate and multivariate control chart. Univariate control chart is a control chart only considering one characteristic such as $\overline{X}$ chart [1], Exponentially Weighted Moving Average (EWMA) chart [2] and Cumulative Sum (CUSUM) chart [3]. In addition, some control charts such as p chart [4,5], np chart [6], and multinomial chart [7,8] are developed to monitor the attribute process. On the other hand, multivariate control chart is a control chart used to control production process with several correlated or uncorrelated characteristics, and it can shortly detect the changes that occur in a manufacturing process [9]. The recent development for multivariate control chart includes Hotelling's T^2 control chart [10–13], MCUSUM control chart [14–16], and MEWMA control chart [17–19].

Not only can be practiced in industrial field, the SPC approach can also be adopted in network intrusion detection. The intrusion detection is one of the techniques which can be used in the security mechanism to monitor the events which are taking place in a computer system or network and analyze the monitoring results to find the signs of anomalies [20]. Furthermore, the intrusion detection system (IDS) has become the important piece in computer network architecture.

The benefit of using the SPC method in observing the anomalies in network is this procedures does not require the knowledge of an information from the preceding attacks. The SPC method can also be used as a powerful procedure which can confirm the system security and stability in network monitoring and intrusion detection process [22]This advantage makes the SPC based IDS can be performed well in online detection process [21]. There are many studies on SPC that have been implemented in IDS for both univariate and multivariate processes [23].

As for multivariate case, Markov Chain, Hotelling's T^2 chart and chi-square multivariate test are used in intrusion detection process [24]. To detect both counter-relations and anomalies in mean-shift, Ye et al. [25] suggested the method based on the Hotelling's T^2 control chart to detect the occurrences of intrusion in network. Qu et al. [26] has adopted the Hotelling's T^2 to invent the IDS for online monitoring system and network which called Multivariate Analysis for Network Attack (MANA) detection algorithm. The control limits of the system will be updated at certain interval to maintain the behavior changes in the network. The Chi-Square Distance Monitoring (CSDM) method has been developed by Ye et al. [27] to monitor uncorrelated, highly correlated, auto-correlated, normally distributed, and non-normally distributed of data. Zhang et al. [28] proposed the new method to detect the anomalies in computer networks by utilizing the Support Vector Clustering (SVC) in control chart. The Covariance Matrix Sign (CMS) was performed by Tavallaee et al. [29] to detect Denial of Service (DoS) type of attacks. The high accuracy of Hotelling's T^2 for all types of attack classes was discovered after comparing the performance of the control chart with Support Vector Machine (SVM) and Triangle Area based Nearest Neighbors (TANN) methods [30].

According to the previous studies in this field, it can be known that the Hotelling's T^2 chart is the most commonly used control chart for intrusion detection. The Hotelling's T^2 which employs the conventional mean and covariance matrix is sensitive to outlier. Therefore, the conventional method is not effective to use for multiple outliers case due to masking effect [10]. The masking effect in monitoring process happens due to the actual outlier which cannot be detected by the control chart. To overcome the problem arise, some robust methods must be proposed to reduce the negative effect of multiple outliers by changing the existing estimators with the more robust estimators.

The modifications in the mean vector estimator are essential to produce the better IDS. The shrinkage estimators which have lower mean squared errors than the conventional estimators can be applied in this case [31,32]. The James-Stein estimator [33], which is the improved estimator of mean vector, can be employed to get the better outcome in estimating the mean vector of the Hotelling's T^2 control chart. Wang et al. realized that the performance of the multivariate control chart with the James-Stein estimator outperforms the existing control charts [34].

Not only modify the mean vector, estimating the covariance matrix in phase I of the monitoring process become the principal feature in initiating the Hotelling's T^2 control chart. The Hotelling's T^2 control chart for individual observations had been discussed by several researchers such as Tracy et al. [19] and Lowry and Montgomery [20]. Chou et al. [37] inspected the power comparison of T^2 control chart based on different types of covariance matrix estimator in phase I monitoring process under multivariate normal distribution assumption. The necessary and sufficient requirement under those underlying multivariate normal distribution was inspected by Cambanis et al. [38]. The problem appears when the sample covariance matrix from the data is computed from the individual observation. The poor performance in detecting shift in the mean vector occurred if control chart is constructed using the conventional covariance matrix [39]. Moreover, the performance of Hotelling's T^2 control chart in detecting shift of mean vector will be increased if robust covariance matrix estimator is applied [40].

Successive Difference Covariance Matrix (SDCM) is one of the robust covariance matrix estimators. Hotelling's T^2 control chart based on SDCM is effective in detecting shift of the mean vector [39,41]. Moreover, SDCM can also be utilized for auto-correlated process such as T^2 control chart based on SDCM for multivariate process using residuals of Vector Autoregressive (VAR) model [42]. Many researchers has been proved the effectiveness of Hotelling's T^2 control chart based on SDCM. However, the exact distribution of the control chart has not been determined. Sullivan and Woodall [23] and Williams et al. [24] suggested the approximate distribution for SDCM based Hotelling's T^2 control chart.

The distribution of network traffic data does not always follow multivariate normal distribution [43]. This is caused by the attacks that occur on a network produce extreme values [44]. When the assumption does not hold, the conventional control limit may be inaccurate because a control limit determined this way can increase the rate of false alarms [45]. Some literatures have been upgraded the Hotelling's T^2 chart control limit performance by using nonparametric approaches in order to overcome the limited knowledge of Hotelling's T^2 distribution. Those studies have been conducted to improve the control limit of Hotelling's T^2 using Kernel Density Estimation (KDE) approach [37,46,47]. Using the same approach, Ahsan et al. implemented the concept of KDE into the Hotelling's T^2 based on SDCM to monitor the anomalies in the network [48]. Although can be used to approximate the exact control limit for unknown distribution data, the KDE approach is not effective while applied to a very skewed distribution of data. To overcome the problems, bootstrap, which is one of the nonparametric techniques that widely used to estimate the parameter without any distribution assumption, can be adopted [49,50]. The bootstrap resampling technique can be performed to obtain the control limits of the control chart which its statistic does not follow any distribution pattern. Phaladiganon et al. [46] developed Hotelling's T^2 control limit based on bootstrap technique and proved that bootstrap approach will be more effective in estimating the control limits when the monitoring statistics are skewed [47]. In addition, the multivariate Hotelling's T^2 based on SDCM has good performance to monitor the anomalies in network when the bootstrap resampling method is applied [51].

Based on the aforementioned reasons, this study proposes Hotelling's T^2 control chart based on Hybrid James-Stein and SDCM using bootstrap resampling approach. The Improved James-Stein estimator is used to estimate the mean vector. On the other hand, the SDCM is utilized to calculate the robust covariance matrix. The utilization of bootstrap resampling method is supposed to convey the more accurate control limit of Hotelling's T^2 based on James-Stein and SDCM. The performance of proposed method is compared with the other control limits and various control chart approaches. In addition, the performance of T^2 based on SDCM using bootstrap approach is compared with the other classification methods.

This paper is organized as follows. Section 2 describes T^2 control chart based on James-Stein and SDCM, while the control limit of Hotelling's T^2 control chart using bootstrap resampling method is explained in Section 3. Section 4 presents the dataset and methodology that used in this research. Moreover, the evaluation performance of IDS is displayed in Section 5. The performance comparisons of the proposed IDS with the other control charts and classifier methods are presented in section 6. Finally, section 7 summarizes the obtained results and presents a future research.

2. HOTELLING'S T^2 CONTROL CHART BASED ON JAMES-STEIN AND SDCM ESTIMATORS

2.1 Hotelling's T^2 Control Chart

In this section, to monitor the mean of a process, Hotelling's T^2 which is one of multivariate the control charts can be employed [52]. Let $\mathbf{x}_i, i = 1,2,\ldots,n$, where n is the number of observations, are assumed identic and independently random vectors which follow multivariate normal distribution with mean vector and covariance matrix, $\mathbf{x} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The mean vector and estimated covariance matrix of $\mathbf{x}$ can be calculated by using $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$, respectively. Thus, the T^2 statistic [53] can be calculated as follows:

$$T_i^2 = (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}). \tag{1}$$

When the data are assumed to follow multivariate normal distribution, the control limit of Hotelling's T^2 can be estimated using the following equation:

$$CL = \frac{p(n+1)(n-1)}{n^2 - np} F_{(\alpha,p,n-p)}, \tag{2}$$

where n is the number of observations, p is the number of quality characteristics and α is the false alarm rate. The process is stated to be in-control when T^2 statistic in equation (1) is not greater than the control limit CL.

2.2 Hotelling's T^2 Control Chart Based On SDCM

SDCM is another possible procedure to estimate the covariance matrix that firstly introduced by reference [54] and [55]. The statistics of T^2 based on SDCM can be computed using the following equation:

$$T_{D,i}^2 = (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{S}_D^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}), \tag{3}$$

where,

$$\mathbf{S}_D = \frac{1}{1(n-1)}\sum_{i=2}^{n}(\mathbf{x}_i - \mathbf{x}_{i-1})(\mathbf{x}_i - \mathbf{x}_{i-1})'. \tag{4}$$

Under in-control condition, the SDCM covariance matrix $\mathbf{S}_D$ is an unbiased estimator for $\boldsymbol{\Sigma}$ [39].

There are several approximations to obtain the control limit of T^2 based on SDCM. By assuming the data follow multivariate normal distribution, the control limit can be estimated using several method, such as control limit based on Sullivan and Woodall (CLSW) [39],

$$CL_{SW} = \frac{(n-1)^2}{n} BETA_{(1-\alpha),\frac{p}{2},\frac{(g-p-1)}{2}}, \tag{5}$$

control limit based on Mason and Young (CLMY) [56],

$$CL_{MY} = \frac{(f-1)^2}{f} BETA_{(1-\alpha),\frac{p}{2},\frac{(g-p-1)}{2}}, \tag{6}$$

and control limit based on chi-square distribution (CL_{χ^2}).

$$CL_{\chi^2} = \chi^2_{(1-\alpha),v}, \tag{7}$$

where $BETA_{(1-\alpha),p,g}$ is $[1-\alpha]$-th quantile of beta distribution with shape parameter p and g while $\chi^2_{(1-\alpha),v}$ is $[1-\alpha]$-th quantile of chi-square distribution with v degree of freedom and let $g = \frac{2(n-1)^2}{3n-4}$.

2.3 Hotelling's T^2 Control Chart Based On James-Stein Estimator

In this study, the James-Stein estimator is used to construct better Hotelling's T^2 control charts. The basic form of James-Stein estimator is formulated as follows:

$$\bar{\mathbf{x}}_0^{JS} = \left(1 - \frac{p-2}{n\left(\bar{\mathbf{x}}-\mathbf{v}\right)^T \boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{x}}-\mathbf{v}\right)}\right)\left(\bar{\mathbf{x}}-\mathbf{v}\right)+\mathbf{v}, \tag{8}$$

where $\mathbf{v}$ is a fixed vector that contains the target value which $\bar{\mathbf{x}}$ will be shrunk. According to Lehmann and Casella [32], $\mathbf{v}$ can be picked as any p-dimensional vector. Furthermore, the improved James-Stein estimator can be calculated as follows:

$$\bar{\mathbf{x}}^{JS} = \left(1 - \frac{p-2}{n\left(\bar{\mathbf{x}}-\mathbf{v}\right)^T \boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{x}}-\mathbf{v}\right)}\right)^{+}\left(\bar{\mathbf{x}}-\mathbf{v}\right)+\mathbf{v}, \tag{9}$$

where the notation x^+ is defined as:

$$x^+ = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

This improved James-Stein estimator is revealed to have smaller Mean Square Error (MSE) than the traditional James-Stein estimator as in (8) [34]. Consequently, in this study the improved James-Stein estimator as in (9) is adopted to develop the Hotelling's T^2 Control Chart Based on James-Stein Estimator.

2.4 Hotelling's T^2 Control Chart Based On James-Stein And SDCM Estimators

The Hotelling's T^2 Control Chart Based on James-Stein Estimator is constructed by replacing the $\bar{\mathbf{x}}$ in equation (1) with $\bar{\mathbf{x}}^{JS}$. The statistics of the proposed chart is formulated as follows:

$$T^2_{JSD,i} = \left(\mathbf{x}_i - \bar{\mathbf{x}}^{JS}\right)' \mathbf{S}_D^{-1}\left(\mathbf{x}_i - \bar{\mathbf{x}}^{JS}\right), \tag{11}$$

where $\bar{\mathbf{x}}^{JS}$ is calculated by changing the covariance matrix $\boldsymbol{\Sigma}$ in equation (9) with SDCM in equation (4) as follows:

$$\bar{\mathbf{x}}^{JS} = \left(1 - \frac{p-2}{n\left(\bar{\mathbf{x}}-\mathbf{v}\right)^T \mathbf{S}_D^{-1}\left(\bar{\mathbf{x}}-\mathbf{v}\right)}\right)^{+}\left(\bar{\mathbf{x}}-\mathbf{v}\right)+\mathbf{v}. \tag{12}$$

Since the distribution of the proposed chart is still unknown, its control limits are calculated using Bootstrap resampling method.

3. T^2 CONTROL LIMIT BASED ON BOOTSTRAP RESAMPLING

Bootstrap is one of the resampling methods that most widely used to estimate the parameter of random variable which has unknown distribution. The bootstrap method was firstly introduced by [49]. This method is simple to be carried out because it requires neither specification of the parameters nor a procedure for numerical integration as in KDE method [46].

Figure 1 defines an overview of the bootstrap procedure for estimating the control limit of T^2_{JSD} statistic as in [46,47,51]. The T^2_{JSD} control limit is calculated by resampling $T^2_{JSD,i}$ statistic, where $i = 1,2,\ldots,n$, for B times, where B is the large number that regularly greater than 1000. Let $T^{2(l)}_{JSD,1}, T^{2(l)}_{JSD,2}, \cdots T^{2(l)}_{JSD,N}$, $l = 1,2,\ldots B$, $N \geq n$ is a set of B bootstrap samples that randomly drawn from $T^2_{JSD,1}, T^2_{JSD,2}, \ldots, T^2_{JSD,n}$ statistics with replacement for l-th replication. For each replication of B bootstrap samples, compute the $[100(1-\alpha)]$-th percentile of T^2_{JSD} distribution, where α is false alarm rate. Furthermore, the control limit can be calculated by taking mean from the value of $[100(1-\alpha)]$-th percentile as follows:

$$CL_{boot} = \frac{1}{B}\sum_{l=1}^{B} T^{2(l)}_{JSD,(100(1-\alpha))} \tag{13}$$

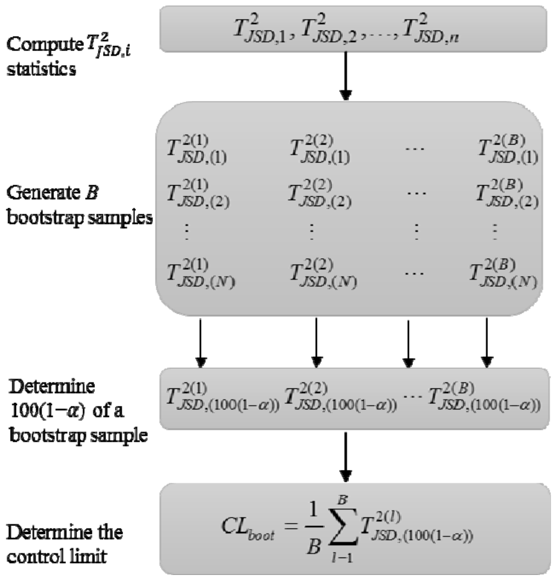
Compute $T^2_{JSD,i}$ statistics
$T^2_{JSD,1}, T^2_{JSD,2}, \ldots, T^2_{JSD,n}$
Generate B bootstrap samples
$T^{2(1)}_{JSD,(1)}$ $T^{2(2)}_{JSD,(1)}$ $\cdots$ $T^{2(B)}_{JSD,(1)}$
$T^{2(1)}_{JSD,(2)}$ $T^{2(2)}_{JSD,(2)}$ $\cdots$ $T^{2(B)}_{JSD,(2)}$
$T^{2(1)}_{JSD,(N)}$ $T^{2(2)}_{JSD,(N)}$ $\cdots$ $T^{2(B)}_{JSD,(N)}$
Determine $100(1-\alpha)$ of a bootstrap sample
$T^{2(1)}_{JSD,(100(1-\alpha))}$ $T^{2(2)}_{JSD,(100(1-\alpha))}$ $\cdots$ $T^{2(B)}_{JSD,(100(1-\alpha))}$
Determine the control limit
$CL_{boot} = \frac{1}{B}\sum_{l=1}^{B} T^{2(l)}_{JSD,(100(1-\alpha))}$

Figure 1: Bootstrap Procedure for calculating control limit of T^2_{JSD} statistics

4. METHODOLOGY

In this section, methodology of the research will be illustrated. First, the methodology of using control chart as IDS is presented. After that, the procedures of proposed T^2 James-Stein and SDCM control chart for IDS is described as well as the performance evaluation method.

4.1 Intrusion Detection System (IDS) using Control Chart

There are two main procedures for constructing this monitoring system i.e. data preparation and control chart construction. Figure 2 illustrates the algorithms of control chart based intrusion detection process. The first procedure of this approach is the preparation of data. This procedure is the most difficult part in IDS process. This procedure is also consuming more time. In data preparation, there are two step that must be done such as, data sourcing and data acquisition. The data sourcing step is process to identify the sources and select the target of the data. While, the data acquisition refers to transform the target data into the input data which can be used in control chart method.

Furthermore, the next procedure is construction of control chart. In control chart construction, the procedure is classified into two steps such as, data pre-processing and create control chart. In this step, the control limits which is previously constructed are applied to monitor the network process. The final step in this method are identifying the source of the intrusion and taking corrective actions.

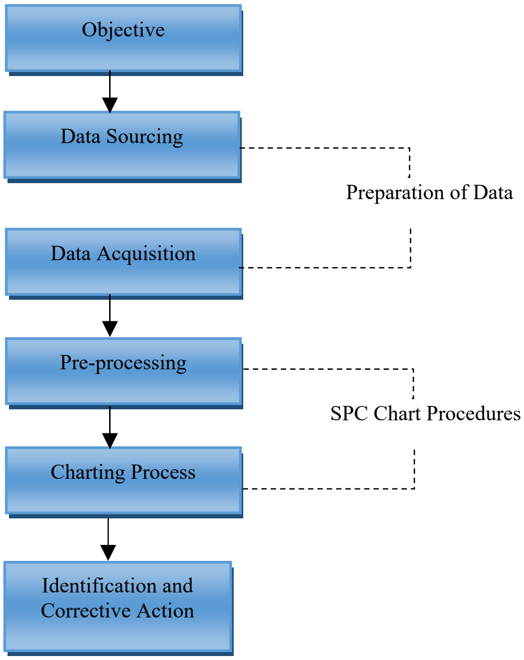

Figure 2: Control Chart based Intrusion Detection System [23]

4.2 *IDS based T^2 James-Stein and SDCM Control Chart*

In this section, algorithm of proposed IDS based on T^2 James-Stein and SDCM control chart is described. The algorithm for IDS with T^2 based SDCM using bootstrap control limit can be divided into two phase as follows:

Phase I: Building Normal Profile

1. Form matrix $\mathbf{X}_{normal}$ which is the normal connection data.

2. Calculate vector $\bar{\mathbf{x}}_{normal,l}$, $l = 1, 2, ..., p$ which is the average of each column of normal connection data $\mathbf{X}_{normal}$.
3. Calculate the matrix of $\mathbf{S}_{DN}$ as in equation (4) which is the estimated variance covariance matrix of normal connection data $\mathbf{X}_{normal}$.
4. Calculate vector $\bar{\mathbf{x}}^{JS}_{normal,l}$, $l = 1, 2, ..., p$ which is the estimated mean vector from James-Stein Estimator of normal connection data $\mathbf{X}_{normal}$.
5. Calculate statistics $T^2_{JSDN,i}$ as in equation (11) using normal connection data $\mathbf{X}_{normal}$.
6. Determine α and calculate the bootstrap control limit using CL_{boot} as in equation (13).

Phase II: Detection

1. Form matrix $\mathbf{X}_{test}$ which is the new connection data.
2. Calculate statistics $T^2_{JSDT,i}$ from new connection data $\mathbf{X}_{test}$ as follows:
$$T^2_{JSDT,i} = \left(\mathbf{x}_i - \bar{\mathbf{x}}^{JS}_{normal}\right)' \mathbf{S}^{-1}_{DN} \left(\mathbf{x}_i - \bar{\mathbf{x}}^{JS}_{normal}\right),$$
where $\bar{\mathbf{x}}^{JS}_{normal}$ and $\mathbf{S}_{DN}$ are taken from normal connection data in phase I.
3. If $T^2_{JSDT,i} > CL_{boot}$ then the connection is intrusion and $T^2_{JSDT,i} < CL_{boot}$ the connection is normal.

4.3 *NSL-KDD Dataset*

NSL-KDD dataset is used in this research. This dataset is first proposed by reference [57] as a solution for obsolete KDD-99 dataset [58] which has been obtainable for more than 15 years. NSL-KDD dataset consists of 41 variables with 34 quantitative variables and 7 qualitative variables. Similar to Ahsan et al. [48,51], this study only uses 32 quantitative variables because the value of the rest quantitative variables is equal to zero.

4.4 *Performance Evaluation*

In this study, NSL-KDD data is monitored using conventional Hotelling's T^2 based on James-Stein and SDCM using bootstrap resampling method. Under hypothesis that the proposed chart has a better performance, the performance comparison between the proposed chart and the other approaches is performed. To demonstrate the superiority of this proposed control limit over other control limits, the performance of the proposed control chart will be compared with some existing control limits. Those control limits are *F* distribution control limit according to (2), Sullivan and Woodall control limit approach based on (4), Mason and Young control limit approach according to (5), chi-square control limit based on (6). The proposed IDS is not only compared with the other control limits but also with the other control chart methods as in [51] and other classifiers that presented in [59].

In addition, the performance of each IDS approach is evaluated by confusion matrix as shown in Table 1 [51]. The performance of the IDS method can be measured by the degree of accuracy and degree of error. The accuracy in detecting intrusion can be divided into two types such as

1. True Positives (TP) is number of successful attacks that is concluded as an attack.
2. True Negatives (TN) is number of normal activities that are successfully detected as normal activity.

By contrast, the error measurement in intrusion detection can also be characterized into two kind:

1. False Positives (FP) is number of normal activities that is wrongly detected as an attack.
2. False Negatives (FN) is number of successful attacks that is wrongly detected as normal activity.

The FP happened in network causes a false alarm causing the system disorder, while FN occured in network will permit an attack on the system.

Table 1: Intrusion Detection Confusion Matrix

	Prediction	
	Intrusion	Normal
Intrusion	True Positive (TP)	False Negative (FN)
Normal	False Positive (FP)	True Negative (TN)

The level of accuracy used is the hit rate that can be calculated as follows:
$$\text{Hit Rate} = \frac{TP + TN}{TP + TN + FP + FN}.$$
The FP and FN rate formula is computed as follows:
$$FP\ \text{Rate} = \frac{FP}{TN + FP}.$$

$$FN \text{ Rate} = \frac{FN}{TP + FN}.$$

5. RESULT AND DISCUSSION

This section is aimed to provide and discuss the performance evaluation of proposed IDS compared to the other control limit approaches.

5.1 Result

The performance evaluation of IDS for NSL-KDD dataset using Hotelling's T^2 control chart based on James-Stein and SDCM is exposed. The control limit of Hotelling's T^2 based on James-Stein and SDCM is estimated using several approaches such as F distribution control limit (JS-SDCM$_F$), Sullivan and Woodall approach (JS-SDCM$_{SW}$), Mason and Young approach (JS-SDCM$_{MY}$), chi-square control limit (JS-SDCM$_{CH}$) and bootstrap control limit (JS-SDCM$_{Boot}$)

Table 2: Performance Of The Proposed IDS with various control limit For Training Dataset

IDS	Hit Rate	FN	FP	FN Rate	FP Rate
JS-SDCM$_F$	0.90758	6595	5048	0.09793	0.08610
JS-SDCM$_{SW}$	0.91748	4034	6361	0.05990	0.10849
JS-SDCM$_{MY}$	0.91073	7047	4199	0.10464	0.07162
JS-SDCM$_{CH}$	0.91074	7043	4201	0.10458	0.07165
JS-SDCM$_{Boot}$	0.91751	4115	6277	0.06111	0.10706

Table 3 : Performance Of The Proposed IDS with various control limit For Testing Dataset

IDS	Hit Rate	FN	FP	FN Rate	FP Rate
JS-SDCM$_F$	0.81396	893	3301	0.09196	0.25723
JS-SDCM$_{SW}$	0.84528	1030	2458	0.10607	0.19154
JS-SDCM$_{MY}$	0.85539	1128	2132	0.11616	0.16613
JS-SDCM$_{CH}$	0.85526	1127	2136	0.11605	0.16645
JS-SDCM$_{Boot}$	0.85535	1127	2134	0.11605	0.16629

The performance of T^2 based on James-Stein and SDCM control chart with various control limit approaches for training data is presented in Table 2. Hotelling's T^2 control chart based on James-Stein and SDCM with bootstrap control limit has hit rate 0.91751, FN rate 0.06111, and FP rate 0.10706. The T^2 based on James-Stein and SDCM with F distribution, Mason and Young, and Chi-square control limits have similar performance with hit rate about 0.907, FN rate about 0.1, and FP rate about 0.07. Moreover, the proposed chart with Sullivan and Woodall control limit has hit rate 0.91748, FN rate 0.05990, and FP rate 0.10849.

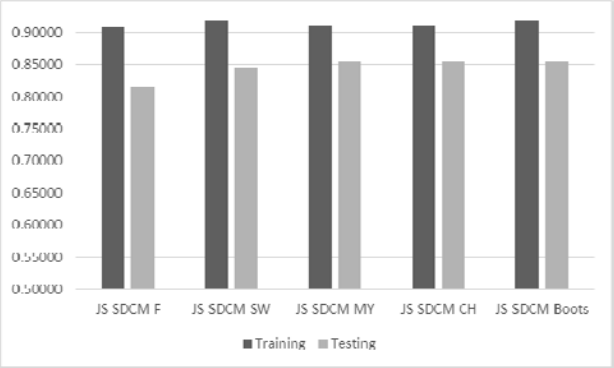

Figure 3: Hit Rate Comparison for Various Control Limit

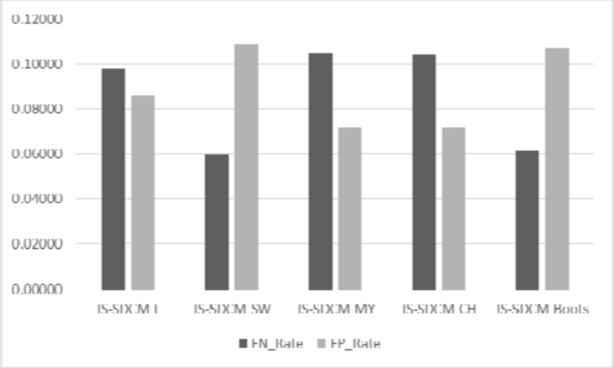

Figure 4: FN and FP Rate Comparison of Training Data

Table 3 reports the performance of Hotelling's T^2 based on James-Stein and SDCM control chart with various control limit approaches for testing dataset. The T^2 based on James-Stein and SDCM control chart with F distribution control limit has hit rate 0.81396, FN Rate 0.09196, and FP Rate 0.25723. Meanwhile, for T^2 based on James-Stein and SDCM control chart with Sullivan and Woodall control limit, the hit rate value is 0.84528, FN rate is 0.10607, and FP rate is 0.19154.

Furthermore, the similarity performance is found for T^2 based on James-Stein and SDCM control chart with Bootstrap, Mason and Young, and Chi-Square control limit. The value of hit rate, FN rate, and FP rate for those case are about 0.855, 0.116, and 0.166 respectively.

The hit rate of various control limit approaches are then visualized in single graphic in order to simplify the performance comparison for each control limit. The hit rate comparison of various control limit approaches for both training and testing dataset are depicted in Figure 3. It can be seen that for training dataset, T^2 based on James-Stein and SDCM with the bootstrap control limit has the highest hit rate. However, for testing dataset, T^2 based on James-Stein and SDCM using Bootstrap and Mason and Young control limits have similar performance. In addition, the hit rate of both T^2 based on James-Stein with *F* Distribution and Sullivan and Woodall control limits are significantly lower than that of the other control limits.

Figure 4 displays the FN rate and FP rate comparison for various control limit approaches in training dataset. The two lowest FN rate is produced by T^2 based on James-Stein and SDCM with Sullivan and Woodall as well as Bootstrap control limits. Although these two methods have highest FP rate, these methods have great performance based on FN rate criteria of training dataset. On the other hand, those methods do not have well the performance according to FP rate criteria of training dataset but superior in FN rate criteria.

The FN rate and FP rate comparison for various control limit approaches in testing dataset are presented in Figure 5. It can be seen that for testing dataset, T^2 based on James-Stein and SDCM with Bootstrap, Mason and Young, and Chi-Square control limit have similar performance for both FN rate and FP rate criteria. The T^2 based on James-Stein and SDCM with *F* Distribution control limit has the worst performance with higher FP rate than the other approaches.

.

5.2 Discussion

Based on the performance evaluation of conventional T^2 based on James-Stein and SDCM with various control limit approaches, it could be known that the T^2 based on James-Stein and SDCM with *F* Distribution control limit has the lowest hit rate for training and testing dataset. The proposed chart using the Mason and Young as well as Chi-Square control limits have same performance for training and testing dataset. For training dataset, the T^2 based on James-Stein and SDCM with Bootstrap and Sullivan and Woodall control limits have the highest performance in term of hit rate. The misdetection happens due to high value of FP rate produced by both charts. The high value of FP rate from training dataset happens due to the oversensitivity of control limit to detect an attack while attack is not actually happened in network. However, those approaches are superior due to the low value of FN rate. Consequently, IDS by those approaches would detect an attack while attack is actually happened in network but would produce high false alarm.

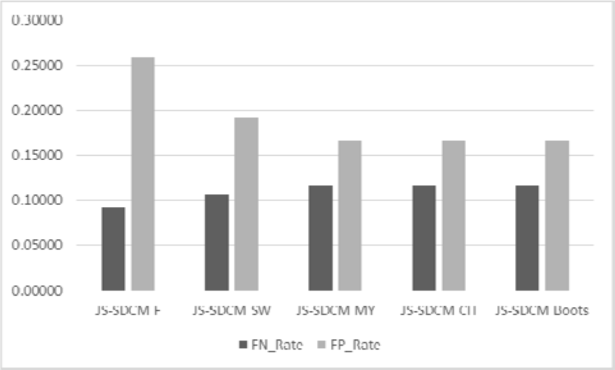

Figure 5: FN and FP Rate Comparison of Testing Dataset

For testing dataset, the highest hit rate produced by Mason and Young, Chi-Square, and Bootstrap control limits. Similar to the training dataset, the misdetection happens due to the high value of FP rate. These findings indicate that the IDS constructed by using T^2 based on James-Stein and SDCM will produce more false alarm. However, this proposed method will not let the attack occurs on the network without any warning that is shown by low level of FN rate. Thus, by considering the performance from training and testing dataset, IDS constructed by the proposed chart with bootstrap is superior to detect the attacks in network than the other approaches.

6. COMPARISON PERFORMANCE WITH OTHER METHODS

In the previous section, it is known that T^2 based on James-Stein and SDCM with Bootstrap control limit has better performance compared to the other control limits. In this section, the performance

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IDS based on the proposed chart is compared with the other control chart methods as well as the other classifiers.

6.1 Comparison with other control chart methods

The performance of the proposed chart using Bootstrap method is compared with the other control chart methods such as conventional Hotelling's T^2 control chart and Hotelling's T^2 control chart based on SDCM in [51] with various control limits. Table 4 reports the performance comparison of the proposed method and the other control charts for training dataset. The values of hit rate from the table are presented in single graphic to simplify the interpretation as shown in Figure 6. It can be seen that the proposed chart with bootstrap control limit has the highest hit rate compared to the other approaches. The proposed chart also has the better result in term of FP rate and FN rate which is depicted in Figure 7. The proposed chart with bootstrap control limit has similar value of FP rate with lower value of FN rate. Thus, the proposed chart with bootstrap control limit has better accuracy to detect anomaly in the network than the other control charts approach for training dataset.

Table 4 : Performance Of Various IDS For Training Data

IDS	Hit Rate	FN	FP	FN Rate	FP Rate
T^2	0.91330	5428	5494	0.0806	0.0937
SDCM$_{F}$	0.91338	5417	5495	0.0804	0.0937
SDCM$_{SW}$	0.91705	4280	6170	0.0636	0.1052
SDCM$_{MY}$	0.91331	5429	5492	0.0806	0.0937
SDCM$_{CH}$	0.91332	5427	5492	0.0806	0.0937
SDCM$_{KDE}$	0.91710	4124	6319	0.0612	0.1078
SDCM$_{Boot}$	0.91706	4238	6210	0.0629	0.1059
JS-SDCM$_{Boot}$	0.91751	4115	6277	0.0611	0.1071

Table 5 exhibits the performance comparison between the proposed chart and the other control chart methods for testing dataset. Similar to the previous result the performance of the proposed chart with Bootstrap control limit is better than the other approaches based on the hit rate criteria as shown in Figure 8. Although it has similar performance with the T^2 control chart based on SDCM with KDE and Bootstrap method, the proposed chart with bootstrap control limit also outperforms the other methods based on the FN rate criteria as depicted in Figure 9.

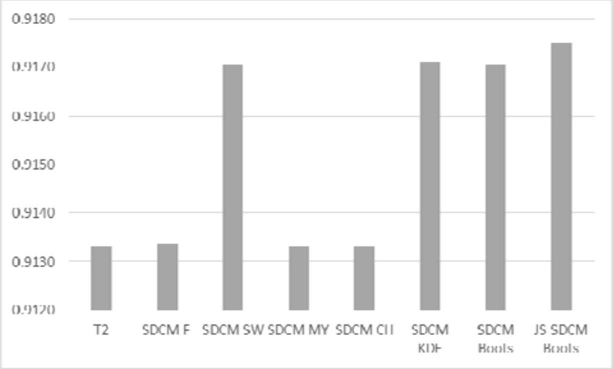

Figure 6: Hit Rate Comparison of Various Control Charts for Training Dataset

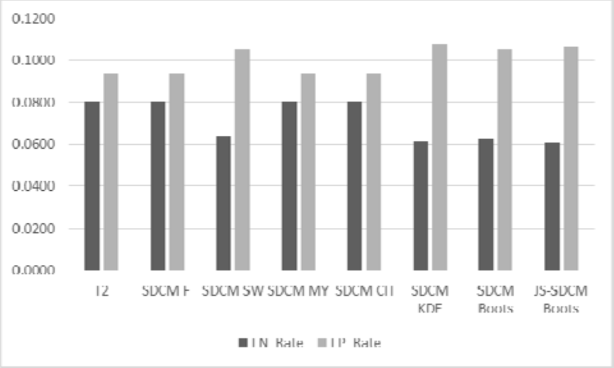

Figure 7: FN and FP Rate Comparison of Various Control Charts for Training Dataset

Table 5 : Performance Of Various IDS For Testing Data

IDS	Hit Rate	FN	FP	FN Rate	FP Rate
T^2	0.8049	814	3584	0.0838	0.2793
SDCM$_{F}$	0.8049	814	3585	0.0838	0.2794
SDCM$_{SW}$	0.7911	731	3978	0.0753	0.3100
SDCM$_{MY}$	0.8049	814	3584	0.0838	0.2793
SDCM$_{CH}$	0.8049	814	3584	0.0838	0.2793
SDCM$_{KDE}$	0.8558	1236	2014	0.1273	0.1569
SDCM$_{Boot}$	0.8562	1221	2020	0.1257	0.1574
JS-SDCM$_{Boot}$	0.8554	1127	2134	0.1160	0.1663

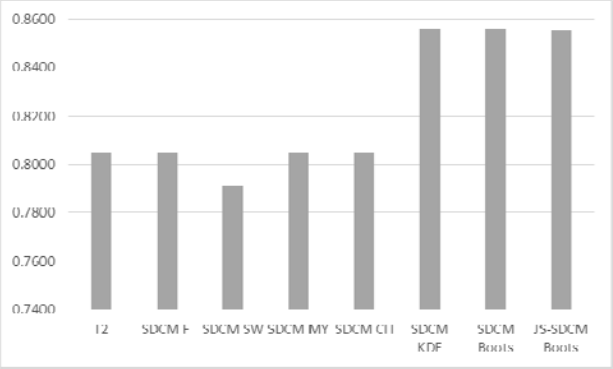

Figure 8: Hit Rate Comparison of Various Control Chart for Testing Dataset

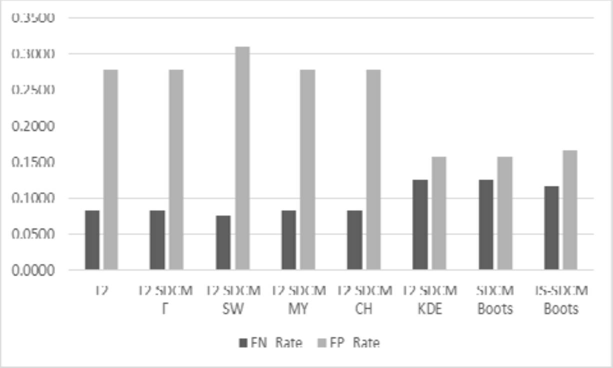

Figure 9: FN and FP Rate Comparison of Various Control Chart for Testing Dataset

6.2 Comparison with other classifiers

Several studies have also proposed IDS by using some classification methods for NSL KDD dataset. Hence, it is important to compare the performance of proposed IDS based on T^2 James-Stein and SDCM using bootstrap control limit not only with the other control limits and control chart methods but also with the other classifiers. In this section, the performance of proposed chart is compared with the other classifiers.

The proposed Hotelling's T^2 chart based on James-Stein and SDCM with bootstrap control limit is then compared with the other classification methods such as Random Forest Classification (RFC), Logistic Regression (LR), Gaussian Naive Bayes (GNB), and Support Vector Machine (SVM) methods as explored in Belavagi and Muniyal [59].

Figure 6 depicts the hit rate comparison between the proposed IDS using T^2 based on James-Stein and SDCM with bootstrap control limit and the other classification methods. It can be seen that the hit rate of proposed IDS based on James-Stein and SDCM with bootstrap approach is 0.9171. The hit rate for Logistic Regression (LR) method is 0.840 and the Gaussian Naïve Bayes (GNB) method has hit rate of 0.79. Meanwhile, the Support Vector Machine (SVM) and Random Forest Classification (RFC) have hit rate of 0.750 and 0.990, respectively.

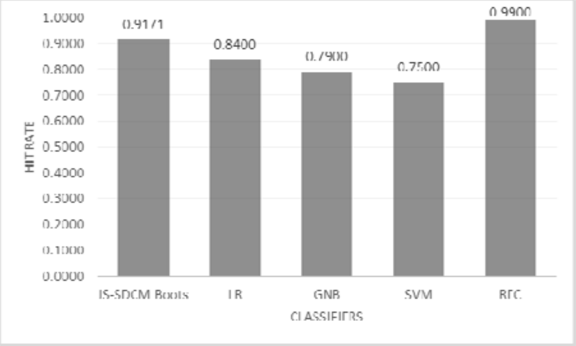

Figure 10: Hit Rate Comparison of Proposed IDS with the other classification methods in Belavagi and Muniyal [59]

The hit rate of proposed T^2 based on James-Stein and SDCM with bootstrap resampling method is higher than that of the other classification methods, except for Random Forest Classification (RFC). This fact reveals that the proposed T^2 based on James-Stein and SDCM with bootstrap is more effective to be used as IDS than the other classification methods inspected by Belavagi and Muniyal except for the Random Forest Classification (RFC).

7. CONCLUSION

In this paper, the multivariate Hotelling's T^2 control chart is improved by James-Stein and SDCM estimators while its control limit is calculated using bootstrap resampling method before it is applied into IDS. The performance of proposed IDS is evaluated and compared with the other control limits by hit rate, FN rate, and FP rate criteria. Furthermore, the performance of proposed IDS is also compared with some existing control charts and classifiers.

The performance evaluation by confusion matrix reveals that the proposed IDS using T^2 based on James-Stein and SDCM with bootstrap control limit surpasses the other control limit approaches in training and testing dataset. Its performance also better than the other existing control chart methods applied to the same dataset. In addition, the proposed T^2 based on James-Stein and SDCM with bootstrap control limit outperforms the other existing classification methods except for random forest classification.

Thus, the proposed method is effective to be utilized in IDS based on its ability to detect anomaly in the network confirmed by the high value of hit rate and low value of FN rate. The multiclass detection for each type of attack with incremental algorithm with KDE and mixed chart [60] based IDS can be considered as a future research.

DISCLOSURE STATEMENT

The authors declare that there are no potential conflict of interest

ACKNOWLEDGMENT

This work was supported by The Ministry of Research, Technology, and Higher Education, The Republic of Indonesia through PMDSU scheme year 2018

REFERENCES

[1] W.A. Shewhart, Some Applications of Statistical Methods to the Analysis of Physical and Engineering Data, Bell Labs Technical Journal. 3 (1924) 43–87.

[2] S.W. Roberts, Control Chart Tests Based on Geometric Moving Averages, Technometrics. 1 (1959) 239–250. doi:10.1080/00401706.1959.10489860.

[3] E.S. Page, Cumulative Sum Charts, Technometrics. 3 (1961) 1–9. doi:10.1080/00401706.1961.10489922.

[4] D.B. Laney, Improved control charts for attributes, Quality Engineering. 14 (2002) 531–537.

[5] M. Ahsan, M. Mashuri, H. Khusna, Evaluation of Laney p' Chart Performance, International Journal of Applied Engineering Research. 12 (2017) 14208–14217.

[6] W.H. Woodall, Control charts based on attribute data: Bibliography and review, Journal of Quality Technology. 29 (1997) 172. http://proquest.umi.com/pqdweb?did=11613494&Fmt=7&clientId=43036&RQT=309&VName=PQD.

[7] Wibawati, M. Mashuri, Purhadi, Irhamah, Fuzzy multinomial control chart and its application, in: AIP Conference Proceedings, 2016. doi:10.1063/1.4943351.

[8] Wibawati, M. Mashuri, Purhadi, Irhamah, M. Ahsan, Perfomance Fuzzy Multinomial Control Chart, Journal of Physics: Conference Series. 1028 (2018) 12120. http://stacks.iop.org/1742-6596/1028/i=1/a=012120.

[9] J.-N. Pan, S.-C. Chen, New robust estimators for detecting non-random patterns in multivariate control charts: a simulation approach, Journal of Statistical Computation and Simulation. 81 (2011) 289–300.

[10] J.L. Alfaro, J.F. Ortega, A comparison of robust alternatives to Hotelling's T2 control chart, Journal of Applied Statistics. 36 (2009) 1385–1396. doi:10.1080/02664760902810813.

[11] H. Ali, S.S. Syed Yahaya, Z. Omar, Robust hotelling T2 control chart with consistent minimum vector variance, Mathematical Problems in Engineering. 2013 (2013). doi:10.1155/2013/401350.

[12] M.O.A. Abu-Shawiesh, G. Kibria, F. George, A Robust Bivariate Control Chart Alternative to the Hotelling's T2 Control Chart, Quality and Reliability Engineering International. 30 (2014) 25–35.

[13] Alkindi, M. Mashuri, D.D. Prastyo, T2 hotelling fuzzy and W2 control chart with application to wheat flour production process, in: AIP Conference Proceedings, 2016. doi:10.1063/1.4953977.

[14] R. Noorossana, S.J.M. Vaghefi, Effect of autocorrelation on performance of the MCUSUM control chart, Quality and Reliability Engineering International. 22 (2006) 191–197. doi:10.1002/qre.695.

[15] J. Arkat, S.T.A. Niaki, B. Abbasi, Artificial neural networks in applying MCUSUM residuals charts for AR(1) processes, Applied Mathematics and Computation. 189 (2007) 1889–1901. doi:10.1016/j.amc.2006.12.081.

[16] B.K. Issam, L. Mohamed, Support vector regression based residual MCUSUM control chart for autocorrelated process, Applied Mathematics and Computation. 201 (2008) 565–574. doi:10.1016/j.amc.2007.12.059.

[17] G. Chen, S.W. CHENG, H. Xie, A new multivariate control chart for monitoring both location and dispersion,

Communications in Statistics—Simulation and Computation®. 34 (2005) 203–217.

[18] M. Pirhooshyaran, S.T.A. Niaki, A double-max MEWMA scheme for simultaneous monitoring and fault isolation of multivariate multistage auto-correlated processes based on novel reduced-dimension statistics, Journal of Process Control. 29 (2015) 11–22. doi:10.1016/j.jprocont.2015.03.008.

[19] H. Khusna, M. Mashuri, Suhartono, D.D. Prastyo, M. Ahsan, Multioutput Least Square SVR Based Multivariate EWMA Control Chart, Journal of Physics: Conference Series. 1028 (2018) 12221. http://stacks.iop.org/1742-6596/1028/i=1/a=012221.

[20] R. Bace, P. Mell, NIST special publication on intrusion detection systems, 2001. doi:10.1016/S1361-3723(01)00614-5.

[21] C.A. Catania, C.G. Garino, Automatic network intrusion detection: Current techniques and open issues, Computers &amp; Electrical Engineering. 38 (2012) 1062–1072. doi:10.1016/j.compeleceng.2012.05.013.

[22] S. Bersimis, A. Sgora, S. Psarakis, The application of multivariate statistical process monitoring in non-industrial processes, Quality Technology and Quantitative Management. 3703 (2016) 1–24. doi:10.1080/16843703.2016.1226711.

[23] Y. Park, A Statistical Process Control Approach for Network Intrusion Detection, Georgia Insitute of Technology, 2005.

[24] N. Ye, X. Li, Q. Chen, S.M. Emran, M. Xu, Probabilistic techniques for intrusion detection based on computer audit data, IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans. 31 (2001) 266–274. doi:10.1109/3468.935043.

[25] N. Ye, S.M. Emran, Q. Chen, S. Vilbert, Multivariate statistical analysis of audit trails for host-based intrusion detection, IEEE Transactions on Computers. 51 (2002) 810–820. doi:10.1109/TC.2002.1017701.

[26] G. Qu, S. Hariri, M. Yousif, Multivariate Statistical Analysis for Network Attacks Detection, The 3rd ACS/IEEE Interenational Conference on Computer Systems and Applications. (2005) 9–14. doi:10.1109/AICCSA.2005.1387011.

[27] N. Ye, D. Parmar, C.M. Borror, A Hybrid SPC Method with the Chi-Square Distance Monitoring Procedure for Large-scale, Complex Process Data, Quality and Reliability Engineering International. 22 (2006) 393–402. doi:10.1002/qre.717.

[28] Z. Zhang, X. Zhu, J. Jin, SVC-Based Multivariate Control Charts for Automatic Anomaly Detection in Computer Networks, in: IEEE, 2007: p. 56. doi:10.1109/CONIELECOMP.2007.99.

[29] M. Tavallaee, W. Lu, S.A. Iqbal, A. Ghorbani, A Novel Covariance Matrix based Approach for Detecting Network Anomalies, in: Sixth Annual Conference on Communication Networks and Services Research, 2008.

[30] A. Avalappampatty Sivasamy, B. Sundan, A Dynamic Intrusion Detection System Based on Multivariate Hotelling's T ^{2} Statistics Approach for Network Environments, The Scientific World Journal. 2015 (2015) 1–9. doi:10.1155/2015/850153.

[31] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, STANFORD UNIVERSITY STANFORD United States, 1956.

[32] E.L. Lehmann, G. Casella, Theory of point estimation, Springer Science & Business Media, 2006.

[33] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1961: pp. 361–379.

[34] H. Wang, L. Huwang, J.H. Yu, Multivariate control charts based on the James–Stein estimator, European Journal of Operational Research. 246 (2015) 119–127.

[35] N.D. Tracy, J.C. Young, R.L. Mason, Multivariate Control Charts for Individual Observations, Journal of Quality Technology. 24 (1992) 88. https://ezproxy.bibl.ulaval.ca/login?url=http://search.proquest.com/docview/214483701?accountid=12008%5Cnhttp://sfx.bibl.ulaval.ca:9003/sfx_local??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Aabiglobal&atitle=.

[36] C.A. LOWRY, D.C. MONTGOMERY, A review of multivariate control charts, IIE

Transactions. 27 (1995) 800–810. doi:10.1080/07408179508936797.

[37] Y. Chou, R.L. Mason, J.C. Young, Power comparisons for a hotelling's t 2 STATISTIC, Communications in Statistics - Simulation and Computation. 28 (1999) 1031–1050. doi:10.1080/03610919908813591.

[38] S. Cambanis, S. Huang, G. Simons, On the theory of elliptically contoured distributions, Journal of Multivariate Analysis. 11 (1981) 368–385. doi:10.1016/0047-259X(81)90082-8.

[39] J.H. Sullivan, W.H. Woodall, A Comparison of Multivariate Control Charts for Individual Observations, Journal of Quality Technology. 28 (1996) 398–408.

[40] J.D. Williams, W.H. Woodall, J.B. Birch, J.O.E.H. Sullivan, On the Distribution of Hotelling ' s T 2 Statistic Based on the Successive Differences Covariance Matrix Estimator, Journal of Quality Technology. 38 (2006) 217–229.

[41] N.J. Vargas, Robust estimation in multivariate control charts for individual observations, Journal of Quality Technology. 35 (2003) 367–376.

[42] J.K. Wororomi, M. Mashuri, Irhamah, A.Z. Arifin, On monitoring shift in the mean processes with vector autoregressive residual control charts of individual observation, Applied Mathematical Sciences. 8 (2014) 3491–3499. doi:10.12988/ams.2014.44298.

[43] M. Ahsan, M. Mashuri, H. Kuswanto, D.D. Prastyo, Intrusion Detection System using Multivariate Control Chart Hotelling's T2 based on PCA, International Journal on Advanced Science, Engineering and Information Technology. 8 (2018).

[44] X. Zhu, Anomaly Detection Through Statistics-Based Machine Learning For Computer Networks, The University of Arizona, 2006.

[45] Y.-M. Chou, R. Mason, J. Young, the Control Chart for Individual Observations From a Multivariate Non-Normal Distribution, Communications in Statistics: Theory & Methods. 30 (2001) 1937. doi:10.1081/STA-100105706.

[46] P. Phaladiganon, S.B. Kim, V.C.P. Chen, J.-G. Baek, S.-K. Park, Bootstrap-Based T 2 Multivariate Control Charts, Communications in Statistics - Simulation and Computation. 40 (2011) 645–662. doi:10.1080/03610918.2010.549989.

[47] P. Phaladiganon, S.B. Kim, V.C.P. Chen, W. Jiang, Principal component analysis-based control charts for multivariate nonnormal distributions, Expert Systems with Applications. 40 (2013) 3044–3054. doi:10.1016/j.eswa.2012.12.020.

[48] M. Ahsan, M. Mashuri, H. Kuswanto, D.D. Prastyo, H. Khusna, T2 Control Chart based on Successive Difference Covariance Matrix for Intrusion Detection System, in: Journal of Physics: Conference Series, IOP Publishing, 2018: p. 12220.

[49] B. Efron, Bootstrap Methods: Another Look at the Jacknife, The Annals of Statistics. 7 (1979) 1–26. doi:10.1214/aoms/1177692541.

[50] B. Efron, R.J. Tibshirani, An introduction to the bootstrap, CRC press, 1994.

[51] M. Ahsan, M. Mashuri, H. Khusna, Intrusion Detection System Using Bootstrap Resampling Approach Of T2 Control Chart Based On Successive Difference Covariance Matrix, Journal of Theoretical and Applied Information Technology. 96 (2018) 2128–2138.

[52] D. Montgomery, Introduction to statistical quality control, 2009. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

[53] H. Hotelling, Multivariate quality control, in: Techniques of Statistical Analysis, McGraw-Hill., New York, 1974.

[54] D.M. Hawkins, D.F. Merriam, Zonation of multivariate sequences of digitized geologic data, Journal of the International Association for Mathematical Geology. 6 (1974) 263–269. doi:10.1007/BF02082892.

[55] D.S. Holmes, A.E. Mergen, Improving the performance of the T2 control chart, Quality Engineering. 5 (1993) 619–625. doi:10.1080/08982119308919004.

[56] R.L. Mason, J.C. Young, Multivariate Statistical Process Control with Industrial Applications, Society for Industrial and Applied Mathematics, 2002. http://epubs.siam.org/doi/book/10.1137/1.9780898718461.

[57] M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD

CUP 99 data set, in: IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009, 2009. doi:10.1109/CISDA.2009.5356528.

[58] S.J. Stolfo, KDD cup 1999 dataset, UCI KDD Repository. Http://Kdd.Ics.Uci.Edu. (1999) 0.

[59] M.C. Belavagi, B. Muniyal, Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection, in: Procedia Computer Science, 2016: pp. 117–123. doi:10.1016/j.procs.2016.06.016.

[60] M. Ahsan, M. Mashuri, H. Kuswanto, D.D. Prastyo, H. Khusna, Multivariate Control Chart based on PCA Mix for Variable and Attribute Quality Characteristics, Production & Manufacturing Research. 6 (2018). doi:10.1080/21693277.2018.1517055.