Expert Systems With Applications 145 (2020) 113105

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Robust adaptive multivariate Hotelling's T^2 control chart based on kernel density estimation for intrusion detection system

Muhammad Ahsan[a,b], Muhammad Mashuri[a,*], Muhammad Hisyam Lee[b], Heri Kuswanto[a], Dedy Dwi Prastyo[a]

[a] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[b] Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

ARTICLE INFO

Article history:
Received 19 December 2018
Revised 25 August 2019
Accepted 28 November 2019
Available online 29 November 2019

Keyword:
Fast-MCD
Kernel density estimation
Hotelling's T^2 chart
Intrusion detection
Statistical process control

ABSTRACT

The utilization of conventional multivariate control chart in network intrusion detection will deal with two main problems. First, the high false alarm occurs due to the distribution of network traffic data that is not following the theory. Second, the inability of the control chart to detect outliers caused by the masking effect. To overcome these problems, the multivariate control chart based on the fast minimum covariance determinant (MCD) algorithm and kernel density estimation (KDE) is proposed in this paper. The employment of KDE technique is expected to adaptively follow the network traffic data pattern, thereby reducing the occurrence of false alarms. Meanwhile, the usage of Fast-MCD will improve the capabilities of the proposed control chart to quickly and accurately detect the outliers. For the simulated data, the proposed chart shows a better level of accuracy when it is compared to conventional T^2 and other robust T^2 based on successive difference covariate matrix (SDSM) charts. For the data generated from some distributions, the proposed chart shows its adaptability by producing low false alarm with high detection rate. The proposed chart shows excellent performance to monitor the KDD99 dataset with 98.61% accuracy, NSL-KDD dataset with 91.71% accuracy, and UNSW-NB 15 dataset with 91.02% accuracy. The proposed method has consistent performance when monitoring the small subset of the datasets, which can minimize the computational time by more than 90% without decreasing its level of accuracy and precision. Also, the performance from the proposed chart surpasses the other benchmarks.

1. Introduction

Unlimited access to the information has brought many benefits to all parties. The fast information transfer enables the potential for security holes that result in increased crime in cyberspace. Thus, a system that can guarantee privacy, confidentiality, integrity, and authenticity in the process on a computer network is needed. Intrusion Detection System (IDS) is a system that has the ability to monitor network traffic, detect suspicious activities, and make early prevention of intrusions or harmful activities on a computer network system that has been built. The information from a network system is collected and analyzed by IDS to discover the elements which violate security policies of networks and computer. Intrusion detection is generally conducted by matching network traffic patterns with known attack patterns (misuse) or by looking for abnormal network traffic patterns or anomalies (Lee & Stolfo, 2000).

Currently, IDS based on anomalies on the network are widely studied. In general, the anomaly-based IDS in the network can be divided into three categories: knowledge-based detection approach, computational approach, and statistical procedure. This work focuses on the utilization of statistical procedure to ensure network security. For this approach, Statistical Process Control (SPC) method can be employed (Park, 2005). In contrast to other approaches, the SPC has the advantage which does not require knowledge of the attack that never happened before. Also, SDI-based SPC can guarantee the attack detection process in real time (Catania & Garino, 2012). A multivariate Control chart is one of the SPC methods that has been widely utilized in network intrusion detection. The implementation of multivariate control chart in IDS to monitor processes on a computer network system can be an effective mechanism to ensure the security and the stability of network information system (Bersimis, Sgora & Psarakis, 2016).

Along with the advantages described, there are some problems when the multivariate control chart is applied to IDS. Ahsan,

* Corresponding author.
E-mail addresses: ahsan4th@gmail.com (M. Ahsan), m_mashuri@statistika.its.ac.id (M. Mashuri), mhl@utm.my (M.H. Lee), heri_k@statistika.its.ac.id (H. Kuswanto), dedy-dp@statistika.its.ac.id (D.D. Prastyo).

https://doi.org/10.1016/j.eswa.2019.113105
0957-4174/

Mashuri, Kuswanto, and Prastyo (2018c) highlighted that most of the multivariate control charts are constructed with certain distribution assumptions, commonly the multivariate normal distribution. However, on reality, the network traffic packages are difficult to follow the multivariate normal distribution caused by the attacks on the network which results in the emergence of extreme values (Zhu, 2006). As a result, there are many false alarms found. Moreover, when it is utilizing in detecting intrusion on a network, the multivariate control chart monitors the event which is considered as an anomaly. In this term, those network anomalies can be analogous to outliers in statistic field. However, the statistic of T^2, which is developed based on the existing estimators, is easily influenced by the presence of outliers (Rousseeuw & Leroy, 2005). The ability of chart to detect the multiple outliers may reduce due to the masking effect (Alfaro & Ortega, 2009). Masking effect in monitoring a process occurs because there are undetected outliers by a control chart. Thus, if the chart is implemented in IDS, this will result in the inability of a control chart to capture the presence of an intrusion signal that occurs on a network. This situation will be hazardous for the security of the computer system because it is not only producing many false alarms but also not able to detect the attacks on the system.

To overcome the masking effect issue, the robust method may be employed to reduce the negative effect of multiple outliers by changing the existing estimators with the more robust estimators. The minimum covariance determinant (MCD) estimator is a highly robust estimator of multivariate location and scatter. Its robustness to outlying samples causes the MCD very beneficial in outlier detection problems (Hubert & Debruyne, 2010). When first introduced, this method was not so popular due to its high computational costs. This method began to attract attention when Rousseeuw and Driessen (1999) proposed the Fast-MCD algorithm which has a more efficient computing process. Thenceforth, the Fast-MCD has been used in many areas such as industry, finance, healthcare, chemistry, and image analysis.

Furthermore, to overcome the high false alarm problem, the kernel density estimation (KDE) method can be employed to estimate the more accurate control limit of the chart. Some studies have been improved the control limit of Hotelling's T^2 chart performance to monitor the non-multivariate normal or even unknown distribution by using this nonparametric techniques (Ahsan, Mashuri, Kuswanto, Prastyo & Khusna, 2018b; Chou, Mason & Young, 1999; Phaladiganon, Kim, Chen & Jiang, 2013). The adaptability of KDE techniques to follow various types of data patterns results in reducing false alarms of the multivariate chart.

Based on the aforementioned reasons, this paper proposes the integration between the Hotelling's T^2 chart, KDE, and Fast-MCD method to construct the robust and adaptive control chart for network intrusion detection. The main contribution of this work lies in the ability of the proposed method to follow any data pattern adaptively by using the KDE method. To boost the ability of the system proposed in predicting anomaly without increasing the false alarm, the Fast-MCD method is used to obtain robust estimators. The proposed method is also made to automatically update the old robust estimators by recalculating these estimators with new data to maintain its adaptability. Finally, to speed up the execution time, some subset of datasets with smaller observation will be taken. In order to prove this statement, first, by using the simulation data, the performance of the proposed chart in detecting several types of outliers and distributions is assessed. After that, its performance is also compared with the conventional or standard T^2 chart and robust T^2 chart based on successive difference covariance matrix (SDCM) with several control limits. The performance of the proposed chart to monitor anomalies in the network is evaluated using three public datasets such as KDD99, NSL-KDD, and UNSW-NB 15. The proposed chart's performance will also be compared with various control charts and several methods that have been developed for IDS to see the contribution of the proposed chart.

The rest of this paper is organized as follows: Section 2 presents the related work. In Section 3, the short review of T^2-based Fast-MCD is presented. Section 4 explains the KDE, while Section 5 describes the procedure of the proposed chart for IDS. Section 6 provides the result and discussion about the performance of the proposed chart in monitoring the outlier on the simulated data. Section 7 shows the application of the proposed chart for IDS. Finally, Section 8 is allocated for the conclusion and future development.

2. Related work

Many methods can be employed in monitoring anomalies in network and can be divided into three main categories: knowledge-based detection approach, computational approach, and statistical procedure. Knowledge-based detection method contains several rules defined by an expert in assessing the behavior of an information system. Thus, the detection system like this requires extensive experience and knowledge of a professional (Javitz & Valdes, 1994). The disadvantage of the method lies in the difficulty of making the model that will be used in the IDS. In addition, a system like this also does not allow intrusion detection in real time.

The computational approach uses data mining or machine learning methods that are applied in an IDS. The advantage of IDS based on computational approach lies in its ability to detect intrusions quickly in the detection phase and its ability to adapt when new information is obtained. The disadvantages of these two methods are the long computation process in the modeling phase and the difficulty of monitoring in the real-time (Derhab & Bouras, 2015). The data mining or machine learning techniques commonly used for intrusion detection include Bayesian Network (BN) (Devarakonda, Pamidi, Kumari & Govardhan, 2012), Naïve Bayes (Sharmai, Pande, Tiwari & Sisodiai, 2012), Decision Tree (DT) (Farid, Zhang, Rahman, Hossain & Strachan, 2014; Kim, Lee & Kim, 2014), Neural Network (NN) (Akashdeep, Manzoor & Kumar, 2017; Wang, Hao, Ma & Huang, 2010; Witten & Frank, 2005), Fuzzy Logic (Gomez & Dasgupta, 2002; Hamamoto, Carvalho, Sampaio, Abrão & Proença, 2018; Tsang, Kwong & Wang, 2007), Genetic Algorithm (GA) (Balajinath & Raghavan, 2001; Hoque, Mukit & Bikas, 2012; Kuang, Xu & Zhang, 2014a), Support Vector Machine (SVM) (Al-Yaseen, Othman & Nazri, 2017; Catania, Bromberg & Garino, 2012; Guo et al., 2014; Salo, Nassif & Essex, 2019), Extreme Learning Machine (ELM) (Singh, Kumar & Singla, 2015), Self-Organizing Map (SOM) (Karami & Guerrero-Zapata, 2014; Karami, 2018), and Skip-Gram model (Carrasco & Sicilia, 2018). Table 1 shows the detailed comparison between these procedures.

Statistical Process Control (SPC) is one of the statistical procedures commonly used for monitoring process in the industrial area. Many SPC methods, especially control chart, have been widely developed. Yang, Lin and Cheng (2011) introduced a new nonparametric EWMA Sign Control Chart for monitoring and detecting possible deviation from the process target. The proposed chart has better performance compared to the Shewhart chart. Kaya and Kahraman (2011) developed fuzzy control charts obtained from fuzzy measurements of the related quality characteristic to improve process capability analyses. The development of nonparametric Principal Component Analysis (PCA) control charts which do not need any distributional assumptions using nonparametric techniques such as kernel density estimation and bootstrap has been done by Phaladiganon, Kim, Chen, Baek and Park (2011). Huang, Tai and Lu (2014) extended the sum of squares generally weighted moving average (SS-GWMA) control chart by using the

Table 1
Summary of some references for data mining and machine learning approach.

References	Method	Dataset	Highlight
Devarakonda et al. (2012)	Bayesian Network and Hidden Markov	KDD99	Performance of the proposed model is of high order for normal and intrusion attacks classification.
Sharmai et al. (2012)	K-means clustering via naïve Bayes classification	KDD99	The proposed technique performs better in terms of Detection rate compared to naive Bayes based approach.
Farid et al. (2014)	Hybrid decision tree and naïve Bayes classifiers	NSL-KDD	The proposed techniques outperform traditional classifiers in challenging multi-class applications
Kim et al. (2014)	Hybrid C4.5 decision tree and Support Vector Machine	NSL-KDD	The proposed approach results in high detection performance with faster computational time
Akashdeep et al. (2017)	Artificial Neural Network classifier	KDD99	The proposed method outperforms other methods for attack and normal classes in term of increasing detection rate and decreasing false alarm rate
Wang et al. (2010)	Artificial Neural Networks and Fuzzy clustering	KDD99	The new proposed approach outperforms other methods in terms of detection accuracy and stability
Gomez and Dasgupta (2002)	Fuzzy Logic and Genetic Algorithms	DARPA	The proposed Algorithm can various types of attacks. However, it has a slower computational time.
Hamamoto et al. (2018)	Genetic Algorithm and Fuzzy Logic	State University of Londrina network traffic	The proposed approach achieves high accuracy and low false in real network traffic flows with higher performance compared to several other approaches.
Tsang et al. (2007)	Genetic-fuzzy rule mining	KDD99	The proposed approach outperforms other methods by providing the higher detection accuracy with intrusion attacks and lower false alarm rate for normal network traffic.
Balajinath and Raghavan (2001)	Genetic algorithms	–	Experimental evaluation of the proposed method can effectively detect the intrusion by learning user behavior with high accuracy and low false alarm rate.
Hoque et al. (2012)	Genetic algorithms	KDD99	The proposed method can efficiently detect various types of network intrusions with reasonable detection rate.
Kuang, Xu, and Zhang (2014b)	Hybrid Kernel PCA and Support Vector Machine with Genetic algorithms	KDD99	The proposed model performs higher accuracy, faster computational time with better generalization compared to other detection methods.
Al-Yaseen et al. (2017)	Multi-level hybrid Support Vector Machine and Extreme Learning Machine based on modified K-means	KDD99	Small training datasets representing the entire original training dataset can and improve the performance of the intrusion detection system with high efficiency in attack detection.
Catania et al. (2012)	Support Vector Machine	DARPA	Combination of the SVM with the autonomous labeling made by SNORT outperforms existing SVM performance. Meanwhile, under some attack distributions, SNORT also obtained improvements.
Guo et al. (2014)	Distance sum-based support vector machine (DSSVM)	KDD99, Glass, Wdbc, Cmc, Yeast, Abalone, and UNIBS 2009-traces	The proposed method did not only obtain good detection accuracy, but also achieved relatively high time efficiency in training the classifier and detecting network attacks. The proposed method also shows its effectiveness for some other pattern recognition problems.
Salo et al. (2019)	Hybrid information gain (IG) and Principal component analysis (PCA) based on support vector machine (SVM), Instance-based learning algorithms (IBK), and multilayer perceptron (MLP)	ISCX 2012, NSL-KDD, and Kyoto 2006	The hybrid method outperforms the individual approaches by achieving high accuracy and low false alarm rates.
Singh et al. (2015)	Online Sequential Extreme Learning Machine (OS-ELM)	NSL-KDD and Kyoto 2006	The proposed method outperforms other existing methods in terms of accuracy, false positive rate and faster detection time
Amin Karami (2018)	Modified Self-Organizing Map (SOM)	NSL-KDD, UNSW-NB15, AAGM, and VPN-nonVPN	The proposed method can accurately and effectively detect attacks and anomalies on the network compared to used existing approaches. It also visualizes useful information and insights about training and testing results.
Moustafa and Slay (2016)	Logistic Regression (LR), Naïve Bayes (NB), Artificial Neural Network (ANN) Decision Tree (DT), and Expectation-Maximization (EM) clustering	KDD99 and UNSW-NB 15	The evaluation results of the five techniques show that the Decision Tree (DT) technique has the best efficiency compared to others. The experimental studies point out that UNSW-NB15 is more complicated than KDD99.
Carrasco and Sicilia (2018)	Skip-gram	UNSW-NB 15	Proposed algorithm obtained high precision and accuracy with a lower false positive rate compared to the other methods. It also requires few features, easy to interpret, and produces information that can be reused by other analysis techniques.

double generally weighted moving average (DGWMA). Through a simulation study, the proposed method is superior to the other charts in all studied scenarios. Kosztyán and Katona (2016) developed a risk-based multi-dimensional T^2 chart (RBT2). The proposed method can be applied even for non-normally distributed data. Not only can be applied for the industrial field, the utilization of control chart can be found in many other areas such as public-health monitoring and outbreak detection (Frisén, Andersson & Schiöler, 2010; Hanslik, Boelle & Flahault, 2001; Schiöler & Frisén, 2012; Shmueli & Burkom, 2010; Woodall, 2006), environmental monitoring (Anderson & Thompson, 2004; George, Chen & Shaw, 2009; Morrison, 2008), financial business monitoring (Abbasi & Guillen, 2013; Chou, Chen & Chen, 2006; Frisén, 2010; Lin, Chou, Wang & Liu, 2012; Moltchanova, 2019), and network intrusion detection monitoring.

In the multivariate case, Ye, Li, Chen, Emran and Xu (2001) used the Markov Chain, Hotelling's T^2 and chi-square multivariate test for intrusion detection. To detect both counter-relations and mean-shift anomalies, Ye, Emran, Chen and Vilbert (2002) proposed a technique based on Hotelling's T^2 test. Qu, Hariri and Yousif (2005) employed Hotelling's T^2 to detect intrusion on a network. An online system which called Multivariate Analysis for Network Attack (MANA) detection algorithm has the control limits that will be updated at a specified interval. The Chi-Square Distance Monitoring (CSDM) method had been developed by Ye, Parmar and Borror (2006) to monitor uncorrelated, high correlated, autocorrelated, normally distributed, and non-normally distributed of data. Zhang, Zhu and Jin (2007))) acquainted a method to detect the anomalies on computer networks named Support Vector Clustering (SVC) based control chart. Covariance Matrix Sign (CMS) had been implemented by Tavallaee, Lu, Iqbal and Ghorbani (2008) to detect Denial of Service (DoS) attacks. The high accuracy of Hotelling's T^2 for all types of attack classes is found after comparing the performance of its control chart with Support Vector Machine (SVM) and Triangle Area-based Nearest Neighbours (TANN) methods (Sivasamy & Sundan, 2015). Hotelling's T^2 control charts based on Successive Difference Covariance Matrix (SDCM) with bootstrap control limit (M. Ahsan, Mashuri & Khusna, 2018b) and KDE control limit (Ahsan, Mashuri, Kuswanto, Prastyo & Khusna, 2018c) to monitor the anomalies in the network. Multivariate control chart using Hotelling's T^2 based on PCA has an excellent performance to detect anomaly in the network compared to conventional T^2 control chart intrusion in accuracy and computational time (Ahsan, Mashuri, Kuswanto & Prastyo, 2018a). Ahsan, Mashuri, and Khusna (2018b) proposed the combination between James-Stein and SDCM control chart using the bootstrap confidence interval in network monitoring. Table 2 reports the detailed comparison between these procedures.

3. Robust multivariate Hotelling's T^2 control chart with Fast-MCD estimator

3.1. Multivariate Hotelling's T^2 control chart

In this section, a brief review of the conventional Hotelling's T^2 control chart is presented. The Hotelling's T^2 is one of multivariate the control charts that can be used to monitor the mean of a process (Montgomery, 2009). Let $\mathbf{x}_i$, where $i = 1, 2, \ldots, n$, are identic and independently random vectors which follow multivariate normal distribution $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The data structure for this case can be written as $\mathbf{X} = [\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_n]'$. Using $\bar{\mathbf{x}} = \frac{1}{n}\sum \mathbf{x}_i$ and $\mathbf{S} = \frac{1}{n-1}\sum (\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})'$, the T^2 statistic can be calculated as follows (Hotelling, 1974):

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

By assuming the data follow the multivariate normal distribution, the control limit of Hotelling's T^2 can be obtained by 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 variables and α is the false alarm rate. The process is said to be in-control if T^2 statistic in Eq. (1) is lower than the control limit CL.

3.2. Fast-MCD algorithm

The fast-MCD algorithm uses iteration and Mahalanobis distance to calculate a robust mean vector and covariance matrix estimations. Consider a dataset $\mathbf{X} = [\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_n]'$. of p-variate observations. Let $G_1 \subset \{1,2, \cdots n\}$ with cardinality $\#G_1 = g$, and put $\mathbf{T}_1 = \frac{1}{g}\sum_{i\in G_1} \mathbf{x}_i$ and $\mathbf{S}_1 = \frac{1}{g}\sum_{i\in G_1} (\mathbf{x}_i - \mathbf{T}_1)(\mathbf{x}_i - \mathbf{T}_1)'$. If $\det(S_1) \neq 0$, then define the relative distances $d_1(i) = \sqrt{(\mathbf{x}_i - \mathbf{T}_1)'\mathbf{S}_1^{-1}(\mathbf{x}_i - \mathbf{T}_1)}$ for $i = 1, 2, \ldots, n$. Next, take G_2 such that $\{d_1(i), i \in G_2\} = \{(d_1)_{1:\,m}, \cdots, (d_1)_{g:\,m}\}$, where $(d_1)_{1:\,m} \leq (d_1)_{2:\,m} \leq \cdots \leq (d_1)_{m:\,m}$ are the ordered distances, and compute $\mathbf{T}_2$ and $\mathbf{S}_2$ based on G_2. Rousseeuw and Driessen (1999) proved that $\det(\mathbf{S}_2) \leq \det(\mathbf{S}_1)$ with equality if and only if $\mathbf{T}_2 = \mathbf{T}_1$ and $\mathbf{S}_2 = \mathbf{S}_1$. With the above theorem the core algorithm of Fast-MCD as C-step can be described as follows: For a given set of g-subset G_{old} with a pair of T_{old} and S_{old}, do the following steps:

1. Compute Mahalanobis distance $d_{old}(i)$, for $i = 1, 2, \cdots n$.
2. Sort these distances which produce a permutation π such as $d_{old}(\pi(1)) \leq d_{old}(\pi(2)) \leq \cdots \leq d_{old}(\pi(n))$.
3. Let $G = \{\pi(1), \pi(2), \cdots \pi(n)\}$.
4. Calculate $\mathbf{T}_{new} = \text{mean}(G)$ and $\mathbf{S}_{new} = \text{cov}(G)$

For n_t times of iteration, repeat step 1–4 by changing $\mathbf{S}_{n_t}$ with $\mathbf{S}_{new}$. Terminating the iteration if $\det(\mathbf{S}_{n_t}) = 0$ or $\mathbf{S}_{n_t} = \mathbf{S}_{n_t-1}$. After the iteration stopped, it can be found that the estimated robust value of the mean vector is $\mathbf{T}_{n_t}$ and estimated robust covariance matrix is $\mathbf{S}_{n_t}$. These two robust estimators will be used to construct the robust Hotelling's T^2 Control Chart.

3.3. Multivariate Hotelling's T^2 control chart with Fast-MCD estimator

In order to construct the robust Multivariate Hotelling's T^2 Control Chart, this study uses the estimated value of mean vector and covariance matrix from Fast-MCD procedure. By changing the value of $\bar{\mathbf{x}}$ and $\mathbf{S}$ in Eq. (1) with the estimated mean vector and covariance matrix, the robust T^2 statistic can be calculated as:

$$T^2_{FMCD,i} = (\mathbf{x}_i - \mathbf{T}_{n_t})'\mathbf{S}_{n_t}^{-1}(\mathbf{x}_i - \mathbf{T}_{n_t}). \tag{3}$$

The control limit of the chart is calculated using Kernel Density Estimation to develop the adaptive control chart.

4. Kernel density estimation control limit

4.1. Kernel density estimation

The probability density function of an unknown random variable can be estimated by employing the Kernel density estimation (KDE) method. This nonparametric method was firstly introduced by Rosenblatt (1956) and Parzen (1962) so that its name is called the Rosenblatt–Parzen kernel density estimator, which is the extension of the histogram estimator.

Chou, Mason and Young (2001) proposed KDE to estimate the distribution of T^2 statistic. Let T^2 is a Hotelling's statistic which ob-

Table 2
Summary of some references for control chart approach.

References	Method	Dataset	Highlight
Ye et al. (2001)	Markov Chain, Hotelling's T^2, and chi-square multivariate test	DARPA	Hotelling's T^2 test and the chi-square multivariate test provide a rather good intrusion detection performance
Ye et al. (2002)	Hotelling's T^2 chi-square multivariate test	DARPA	Hotelling's T^2 test captures all the intrusion sessions and produces no false alarms for the normal sessions for small dataset. For a large dataset, the test detects almost all of the intrusion with no false alarm.
Qu et al. (2005)	Hotelling's T^2 control chart	–	Proposed algorithm can accurately detect well-known attacks such as Distributed Denial of Service, SQL Slammer Worm, and Email spam attacks.
Ye et al. (2006)	Chi-Square Distance Monitoring (CSDM) and Hotelling's T^2	–	The proposed procedure has an advantage in computational efficiency and scalability. Hotelling's T^2 test is superior to the proposed procedure only for multivariate process data with correlated and normally distributed data.
Zhang et al. (2007)	Support Vector Clustering (SVC) based control chart and Hotelling's T^2	KDD99	The proposed method has similar performance to T^2 control charts for the Gaussian data. On the other hand, it has a better performance for non-Gaussian distribution data. In detecting intrusion, the proposed method has a poor performance to detect the probing attack.
Tavallaee et al. (2008)	Covariance Matrix Sign (CMS)	KDD99	The experimental studies show good detection rates with low false positives rates.
Sivasamy and Sundan (2015)	Hotelling's T^2, with Support Vector Machine (SVM) and Triangle Area-based Nearest Neighbours (TANN)	KDD99	Hotelling's T^2 has higher accuracy for all types of attack classes compared to the other methods.
M. Ahsan et al. (2018b)	Hotelling's T^2 control charts based on Successive Difference Covariance Matrix (SDCM) with bootstrap control limit	NSL-KDD	Proposed IDS outperforms the other control charts and classification methods in term of accuracy detection.
M. Ahsan et al. (2018b)	Hotelling's T^2 control charts based on SDCM with KDE control limit	NSL-KDD	Proposed IDS outperforms the other approaches both in training and testing dataset.
Ahsan et al. (2018a)	Hotelling's T^2 control charts based on PCA	KDD99	The proposed method outperforms the performance of the conventional T^2 chart in term of accuracy detection and execution time.
Ahsan, Mashuri and Khusna (2018a)	A combination of James-Stein, SDCM, and bootstrap.	NSL-KDD	The proposed chart has better performance than the other existing charts based on its hit rate and FN rate criteria. Also, it outperforms some classifier methods.

tained under in-control condition. The distribution of the T^2 statistic could be calculated with the following kernel function:

$$\hat{f}_h(t) = \frac{1}{n}\sum_{i=1}^{n} K\left[\frac{\left(t - T_i^2\right)}{\hat{h}}\right], \tag{4}$$

where K and h define the kernel function and the estimated smoothing parameter, respectively. The most used kernel is Gaussian Kernel, which is also used in this analysis.

The control limit can be calculated using tables of integrals, in closed form distribution. However, the control limit might be not efficient to be calculated if the distribution is not closed form. Thus, the kernel control limit is solved using trapezoidal rule (Burden & Faires, 2011), one of the numerical integration methods to approximate the definite value of integral equation. Furthermore, the control limit of T^2 based on KDE could be estimated by taking the percentile of kernel distribution. Hence, the control limit of T^2 based on KDE equal to $[100(1-\alpha)]$-th percentile of T^2 distribution which could be calculated using as follows:

$$CL_{\text{kernel}} = \hat{F}_h(t)^{-1}(1-\alpha). \tag{5}$$

4.2. Proposed chart with KDE control limit

In this paper, the Robust multivariate T^2 control chart with KDE is proposed. The procedure of the proposed chart be defined as follows:

1. Calculate the statistic of T^2_{FMCD} as in Eq. (3).
2. Estimate the empirical density of T^2_{FMCD} statistic using KDE with the Gaussian kernel using the equation:

$$\hat{\tilde{f}}_h(\tilde{T}) = \frac{1}{n\hat{h}}\sum_{i=1}^{n} K\left(\frac{\tilde{T} - T^2_{FMCD,i}}{\hat{h}}\right), \tag{6}$$

where the Gaussian kernel is defined as:

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(u^2)\right), -\infty < u < \infty.$$

3. Calculate the distribution function of $\hat{\tilde{f}}_h(\tilde{T})$or denoted by $\hat{\tilde{F}}_h(\tilde{t})$ as:

$$\hat{\tilde{F}}_h(\tilde{t}) = \int_0^{\tilde{t}} \hat{\tilde{f}}_h(\tilde{T})d\tilde{T}, \tag{7}$$

by using a numerical approach with the trapezoid rule method, the $\hat{\tilde{F}}_h(t)$ in Eq. (7) is calculated as follows:

$$\int_{\eta_{\min}}^{\eta_{\max}} \hat{\tilde{f}}_h(T)dT \approx \frac{\eta_{\max} - \eta_{\min}}{2n}\sum_{i=1}^{n}\left(\hat{\tilde{f}}_h(T_i^2) + \hat{\tilde{f}}_h(T_{i+1}^2)\right),$$

where $\eta_{\min}$and$\eta_{\max}$ are the minimum and maximum value of $\tilde{T}$.

4. Determine $\alpha$and calculate the control limit of T^2_{FMCD} based on KDE that equal to $100(1-\alpha)^{\text{th}}$ percentile of $\hat{\tilde{f}}_h(\tilde{t})$ using the following equation:

$$CL_{KDE} = \hat{\tilde{F}}_h^{-1}(\tilde{t})(1-\alpha). \tag{8}$$

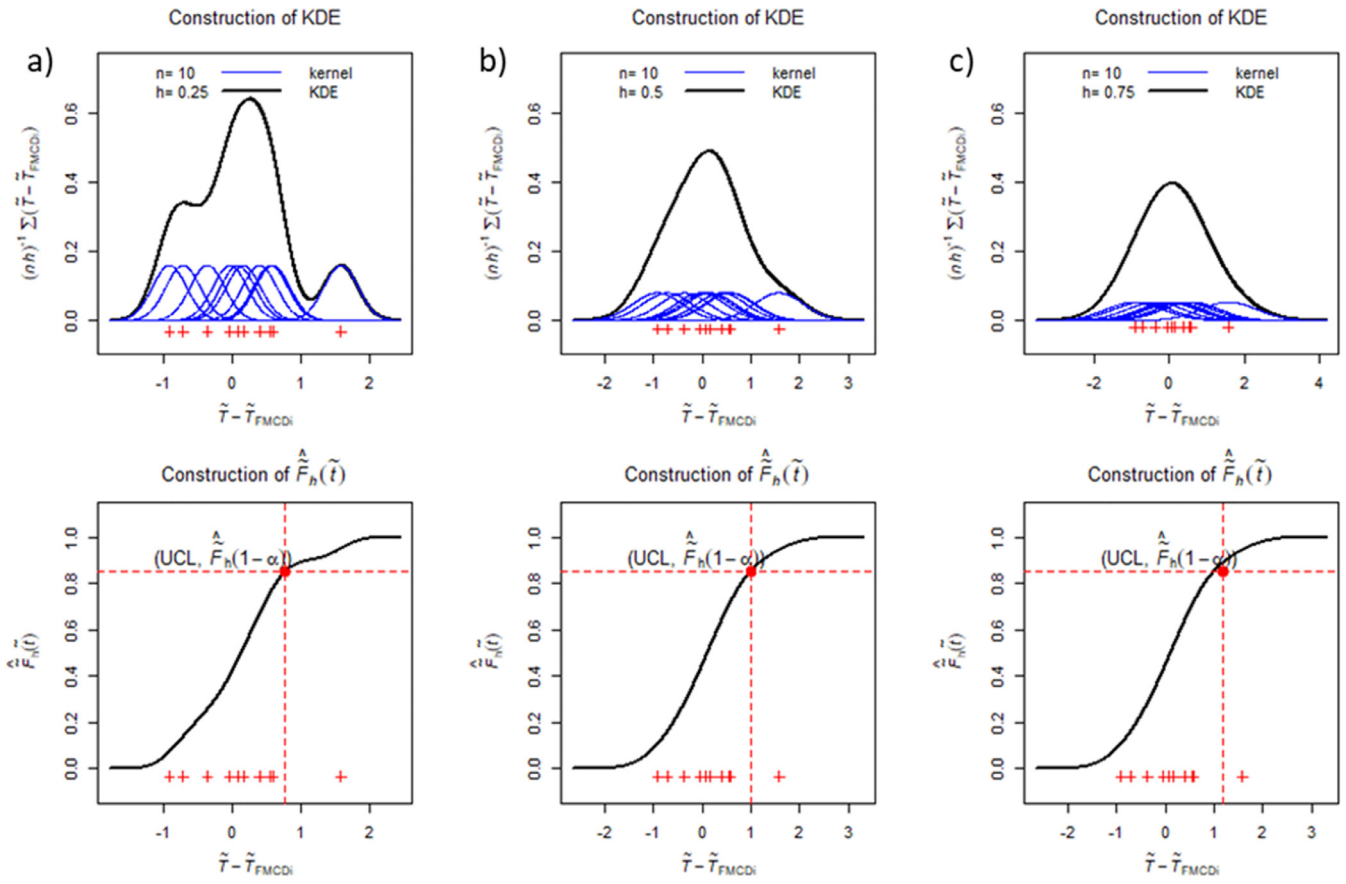

Fig. 1. Illustration of KDE construction and control limit calculation for: (a) $h = 0.25$, (b) $h = 0.5$, and (c) $h = 0.75$.

Fig. 1 illustrates the construction of KDE (first row) and the calculation of empirical distribution function $\hat{\tilde{F}}_h(\tilde{t})$as well as the corresponding control limit (second row) for the number of observation $n = 10$ and three types of bandwidth $h = 0.25$, 0.5, and 0.75. The $T^2_{FMCD,i}$statistics is represented by a red plus sign (+). It can be easily seen that the KDE is constructed by summing 10 Gaussian Kernel. It also can be seen that the inappropriate bandwidth h will result in under-smoothed (see Fig. 1(a)) or over-smoothed problem. The second row of the figure shows the empirical distribution function $\hat{\tilde{F}}_h(\tilde{t})$ calculated using the trapezoid rule method. It can be known that the under-smoothed density will result in an under-smoothed distribution function being formed. Thus, calculating the optimum bandwidth $\hat{h}$takes the vital part in this case. Finally, the CL_{KDE} is calculated by finding the value in the horizontal axis that corresponds to $\hat{\tilde{F}}_h^{-1}(1-\alpha)$.

5. The algorithm of IDS based on the proposed chart

The algorithm of the IDS based on a control chart is presented in this section. A control chart has the advantage to use in IDS, especially if there is no historical data on network traffic. By using this method, the user can create a real-time monitoring system. There are two primary procedures for constructing this monitoring system, i.e., data preparation and control chart construction.

Fig. 2 illustrates the procedure of the intrusion detection process using the control chart. The first main procedure is the preparation of data. This procedure is the most challenging part of the IDS process. This procedure is also consuming more time. In data preparation, two steps must be done, such as data sourcing and data acquisition. The data sourcing step is a 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 the control chart method.

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

The algorithm for the proposed IDS with KDE control limit can be divided into two phases as follows:

Phase I: Building Normal Profile

In this phase, the mean, the covariance matrix, and the KDE control limit are calculated from the normal profile of the dataset using the proposed method. The estimated values are then used in the next phase to monitor the new connection. The procedures of this building normal profile phase are defined as follows:

Step 1 Form matrix $\mathbf{X}_{normal}$, which is the normal connection data (clean data).

Step 2 Calculate $\mathbf{T}_{n_t,Normal}$ and $\mathbf{S}_{n_t,Normal}$, which are the robust mean vector and covariance matrix of normal connection data $\mathbf{X}_{normal}$ using Fast-MCD algorithm in section 2.2.

Step 3 Calculate statistics $T^2_{FMCD,i}$ as in Eq. (3) from normal connection data $\mathbf{X}_{normal}$ with estimated mean vector and covariance matrix in Step 2.

Step 4 Determine α and calculate the KDE control limit using CL_{KDE}as in Section 3.2.

Phase II: Detection

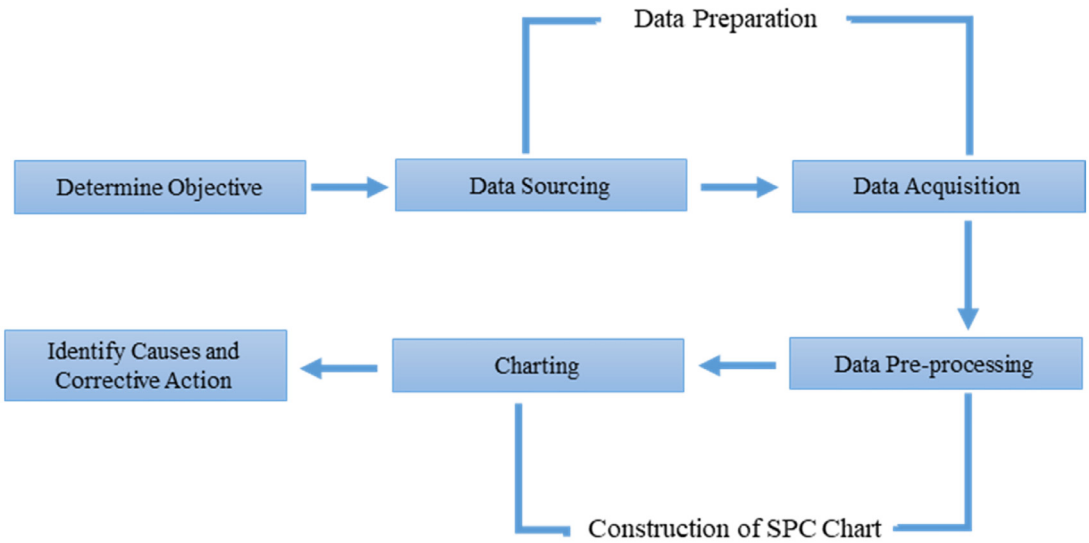

Fig. 2. Intrusion detection system using a control chart method (Park, 2005).

Table 3
Intrusion detection confusion matrix.

	Prediction	
	Intrusion	Normal
Intrusion	True Positives (TP)	False Negatives (FN)
Normal	False Positives (FP)	True Negatives (TN)

The estimated value of $\mathbf{T}_{n_t,Normal}$, $\mathbf{S}_{n_t,Normal}$, and $CL_{KDE}$from Phase I are used in this phase. The following steps explain the procedures of the detection phase:

Step 1 Form matrix $\mathbf{X}_{test}$, which is the new connection data.

Step 2 Calculate statistics $T^2_{FMCD,i}$ from new connection data $\mathbf{X}_{test}$ as follows:

$$T^2_{FMCD,i} = \left(\mathbf{x}_{test,i} - \mathbf{T}_{n_t,Normal}\right)' \mathbf{S}^{-1}_{n_t,Normal}\left(\mathbf{x}_{test,i} - \mathbf{T}_{n_t,Normal}\right),$$

where $\mathbf{T}_{n_t,Normal}$ and $\mathbf{S}_{n_t,Normal}$ are taken from normal connection data in phase I.

Step 3 If $T^2_{FMCD,i} > CL_{KDE}$ then the connection is an intrusion and if $T^2_{FMCD,i} < CL_{KDE}$ then the connection is normal, update $\mathbf{X}_{test}$, and recalculate the $\mathbf{T}_{n_t,Normal}$ and $\mathbf{S}_{n_t,Normal}$.

Moreover, the performance of IDS would be evaluated by the confusion matrix as shown in Table 3. The accuracy of a classification method could be measured by the degree of accuracy and degree of error. The accuracy in detecting intrusion can be divided into two types:

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

Errors in intrusion detection could be divided into two types:

a. False Positives (FP) is a number of normal activities that are detected as an attack.
b. False Negatives (FN) is a number of successful attacks that are detected as a normal activity.

FP causes a false alarm while FN allows an attack on the system. The level of accuracy used is the Hit Rate that can be calculated as follows:

$$\text{Hit Rate} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{9}$$

Based on the type of error, the level of error in intrusion detection can be divided into two types, namely *FP* rate and *FN* rate. The *FP* rate and *FN* Rate formula can be written as follows:

$$FP\,\text{Rate} = \frac{FP}{TN + FP}. \tag{10}$$

$$FN\,\text{Rate} = \frac{FN}{TP + FN}. \tag{11}$$

6. Simulation study

The simulation study in this section is aimed to investigate the performance of the proposed chart in detecting the labeled outliers, which are randomly mixed with clean data. There are two cases simulated in this study. First, the performance of the proposed chart is evaluated for the multivariate normal distribution generated data. Second, the non-multivariate normal distributions, i.e. multivariate gamma, multivariate exponential, and multivariate Weibull, are employed to generate the simulated data. The performance of the proposed chart is also compared to the other methods to show its superiority. In addition, the simulation studies use $\alpha = 0.00273$ in all cases.

6.1. Multivariate normal distribution

In this section, the performance of the proposed chart is evaluated for multivariate normal distribution with a different kind of outlier. Fig. 3 shows the performance of the proposed chart to detect a small amount (1%) of outlier added. It can be seen that the proposed chart can detect all of the outlier added (observations with red color). There is only a small false alarm produced by the proposed chart. It can be seen from the figure that the number of misdetected observation (observations with blue color which fall outside of the control limit) is only 3 or 4 for all cases. For the simulation study, the data are generated from a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{clean} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$=**I**, $\mathbf{X}_{clean} \sim N_p(\boldsymbol{\mu}_{clean}, \mathbf{I})$. The simulation is conducted 1000 times to calculate the Hit Rate, FN Rate, and FP Rate. The percentages of outliers ε which are mixed to the clean data are 5%, 10%, 15%, 20%, 30%, and 50% over the total observations. The simulation studies are done for six scenarios of a different number of variables, that are $p = 3$, $p = 5$, $p = 10$, $p = 15$, $p = 20$, and $p = 30$. In addition, outlier added to the clean data are generated

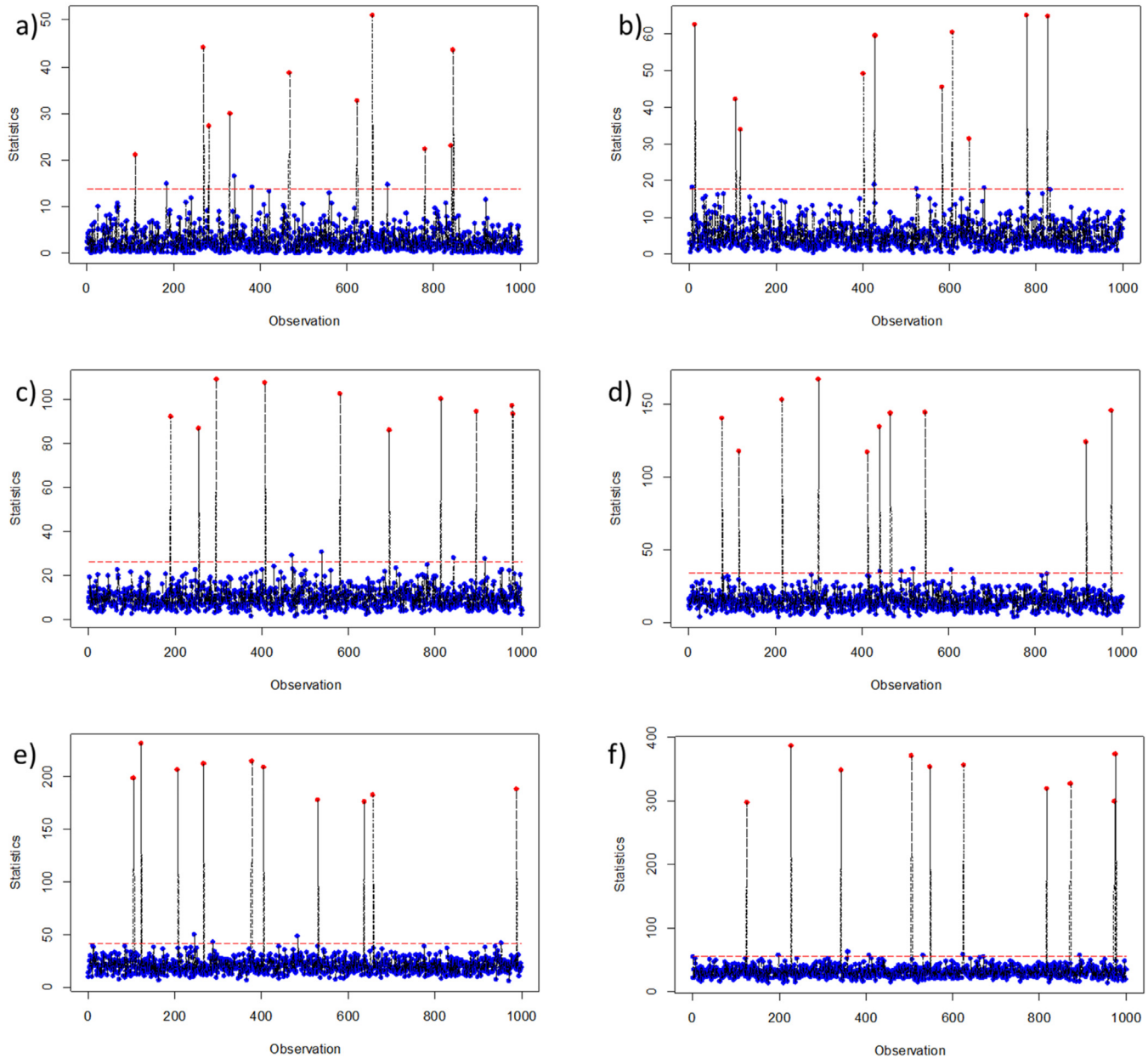

Fig. 3. The utilization of the proposed chart to monitor multivariate normal data with one percent of outlier added for: (a) $p = 3$, (b) $p = 5$, (c) $p = 10$, (d) $p = 15$, (e) $p = 20$, and (f) $p = 30$.

from Multivariate Normal Distribution with mean vector $\boldsymbol{\mu}_{cont}$ and covariance matrix $\boldsymbol{\Sigma}$=**I,** $\mathbf{X}_{cont} \sim N_p(\boldsymbol{\mu}_{cont}, \mathbf{I})$. There is three kind outliers in this study, that is small outliers for $\boldsymbol{\mu}_{cont} = \mathbf{3}$= $[3\ \ 3\ \ \ldots\ \ 3]'_{1 \times p}$, medium outliers for $\boldsymbol{\mu}_{cont} = \mathbf{6}$= $[6\ \ 6\ \ \ldots\ \ 6]'_{1 \times p}$, and large outliers for $\boldsymbol{\mu}_{cont} = \mathbf{9}$= $[9\ \ 9\ \ \ldots\ \ 9]'_{1 \times p}$.

Table 4 shows the performance of conventional T^2 chart with an F distribution control limit in detecting the outlier for each case, and various percentages of small outlier added. In general, the Hit Rate will decrease as the percentages of outlier added to the clean data is increasing. The conventional T^2 chart has hit rate more than 0.95 only for 5% percentage of outlier added. For 10% percentage of outlier added, the conventional chart is no longer able to detect the outliers added to the clean data confirmed by the high value of FN Rate (more than 0.94). These findings indicate that the conventional T^2 chart only correctly detects less than 10% outlier added for this setting. Furthermore, the performance of the conventional chart decreases as the percentage of added outlier increases.

Tables 5–8 report the performance of T^2 chart with robust covariance matrix SDCM with several control limits such as F Distribution, Mason and Young (2002), Sullivan and Woodall (1996), as well as Chi-Square distribution control limits. In general, the performance of the robust SDCM-based T^2 chart with various control limits are better than the conventional T^2chart except for the Sullivan and Woodall control limit. From the simulation results, the superiority of this robust chart can be seen from the higher value of the Hit Rate and the lower value of the FN rate. Similar to the conventional control limit, this robust T^2 chart based on SDCM only can manage about 10% of outliers for this simulation setting. Especially for Sullivan and Woodall control limit, the chart has about zero false alarm. However, due to the high value of the control limit, the chart cannot detect the outlier added. It can be seen from the high value of FN rate

Table 9 presents the performance of the proposed chart in detecting the outlier for all scenario. In general, the Hit Rate of the

Table 4
Performance of conventional T^2 chart to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9685	0.0011	0.6088	0.9044	0.0008	0.9492	0.8518	0.0006	0.9845
$p = 5$	0.9672	0.0015	0.6280	0.9041	0.0016	0.9441	0.8515	0.0015	0.9811
$p = 10$	0.9620	0.0018	0.7256	0.9021	0.0015	0.9660	0.8502	0.0017	0.9886
$p = 15$	0.9576	0.0016	0.8172	0.9011	0.0015	0.9759	0.8500	0.0014	0.9922
$p = 20$	0.9562	0.0016	0.8450	0.9006	0.0017	0.9787	0.8498	0.0019	0.9905
$p = 30$	0.9536	0.0020	0.8898	0.8999	0.0019	0.9837	0.8497	0.0015	0.9935
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.8007	0.0015	0.9906	0.7003	0.0009	0.9969	0.5000	0.0010	0.9990
$p = 5$	0.8003	0.0012	0.9941	0.7000	0.0017	0.9960	0.4999	0.0020	0.9981
$p = 10$	0.7999	0.0016	0.9941	0.6997	0.0016	0.9971	0.4999	0.0019	0.9983
$p = 15$	0.8000	0.0013	0.9946	0.6999	0.0015	0.9968	0.5000	0.0021	0.9978
$p = 20$	0.7998	0.0015	0.9950	0.6996	0.0020	0.9966	0.5000	0.0026	0.9974
$p = 30$	0.7995	0.0019	0.9949	0.6995	0.0022	0.9967	0.5000	0.0023	0.9977

Table 5
Performance of SDCM T^2 chart with F distribution control limit ($SDCM_F$) to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9714	0.0010	0.5524	0.9112	0.0007	0.8816	0.8565	0.0015	0.9479
$p = 5$	0.9734	0.0013	0.5066	0.9108	0.0017	0.8770	0.8546	0.0013	0.9619
$p = 10$	0.9683	0.0017	0.6032	0.9061	0.0024	0.9180	0.8525	0.0021	0.9711
$p = 15$	0.9607	0.0017	0.7538	0.9025	0.0014	0.9624	0.8519	0.0026	0.9730
$p = 20$	0.9613	0.0019	0.7376	0.9028	0.0020	0.9549	0.8512	0.0018	0.9821
$p = 30$	0.9576	0.0029	0.7930	0.9010	0.0034	0.9597	0.8501	0.0023	0.9863
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.8029	0.0010	0.9818	0.7021	0.0009	0.9908	0.5037	0.0015	0.9912
$p = 5$	0.8017	0.0011	0.9869	0.7029	0.0023	0.9849	0.5021	0.0010	0.9949
$p = 10$	0.8012	0.0021	0.9858	0.7017	0.0017	0.9903	0.5021	0.0019	0.9940
$p = 15$	0.8012	0.0018	0.9868	0.7016	0.0022	0.9896	0.5026	0.0018	0.9930
$p = 20$	0.8007	0.0024	0.9869	0.7013	0.0023	0.9904	0.5021	0.0020	0.9939
$p = 30$	0.8006	0.0029	0.9854	0.7004	0.0027	0.9924	0.5018	0.0030	0.9935

Table 6
Performance of SDCM T^2 chart with Sullivan and Woodall control limit ($SDCM_{SW}$) to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9563	0.0000	0.8736	0.9010	0.0000	0.9895	0.8501	0.0000	0.9989
$p = 5$	0.9524	0.0000	0.9510	0.9004	0.0000	0.9963	0.8501	0.0000	0.9991
$p = 10$	0.9511	0.0000	0.9774	0.9002	0.0000	0.9981	0.8500	0.0000	0.9995
$p = 15$	0.9503	0.0000	0.9932	0.9001	0.0000	0.9995	0.8500	0.0000	0.9997
$p = 20$	0.9502	0.0000	0.9968	0.9000	0.0000	0.9996	0.8500	0.0000	1.0000
$p = 30$	0.9501	0.0000	0.9990	0.9000	0.0000	1.0000	0.8500	0.0000	1.0000
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.8001	0.0000	0.9995	0.7000	0.0001	0.9998	0.5001	0.0000	0.9998
$p = 5$	0.8000	0.0000	0.9998	0.7001	0.0000	0.9997	0.5001	0.0000	0.9998
$p = 10$	0.8000	0.0000	1.0000	0.7000	0.0000	0.9999	0.5000	0.0000	1.0000
$p = 15$	0.8000	0.0000	1.0000	0.7000	0.0000	1.0000	0.5000	0.0000	1.0000
$p = 20$	0.8000	0.0000	0.9999	0.7000	0.0000	0.9999	0.5000	0.0000	1.0000
$p = 30$	0.8000	0.0000	1.0000	0.7000	0.0000	1.0000	0.5000	0.0000	1.0000

proposed chart will be decreasing as the percentages of outlier added to the in-control data is increasing. Different from the other methods, the proposed chart still has Hit Rate more 0.95 when the percentage of outliers is lower than 30%. The misdetections happen due to the high value of FN. Although the value of FP rate from both charts is similar to the other charts, the proposed chart outperforms the other benchmarks based on its Hit Rate and FN rate value.

The simulation also conducted for different value of outliers such as medium and large. Fig. 4 shows the performance comparison between all charts with the different value of outlier added. Fig. 4(a) depicts the average of Hit Rate, FN rate, and FP rate for small outlier from each chart. The “average” word in this term means the average value of the metric for all cases ($p = 3$, $p = 5$, $p = 10$,$p = 15$, $p = 20$, and $p = 30$) from all charts (Tables 4–9). As explained before, for this case, the proposed chart is better than

Table 7
Performance of SDCM T^2 chart with Mason and Young control limit ($SDCM_{MY}$)to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9734	0.0016	0.5018	0.9137	0.0011	0.8533	0.8552	0.0011	0.9592
$p = 5$	0.9762	0.0019	0.4388	0.9089	0.0016	0.8959	0.8549	0.0013	0.9604
$p = 10$	0.9749	0.0025	0.4542	0.9060	0.0022	0.9202	0.8527	0.0031	0.9644
$p = 15$	0.9664	0.0033	0.6088	0.9045	0.0029	0.9290	0.8523	0.0039	0.9627
$p = 20$	0.9679	0.0037	0.5718	0.9037	0.0034	0.9317	0.8514	0.0030	0.9737
$p = 30$	0.9595	0.0059	0.6982	0.9005	0.0077	0.9260	0.8501	0.0061	0.9649
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.8025	0.0010	0.9836	0.7026	0.0012	0.9887	0.5031	0.0010	0.9927
$p = 5$	0.8038	0.0023	0.9721	0.7027	0.0012	0.9879	0.5042	0.0016	0.9900
$p = 10$	0.8021	0.0018	0.9825	0.7022	0.0019	0.9880	0.5042	0.0026	0.9891
$p = 15$	0.8021	0.0027	0.9787	0.7012	0.0032	0.9887	0.5041	0.0034	0.9883
$p = 20$	0.8009	0.0041	0.9793	0.7011	0.0048	0.9853	0.5039	0.0043	0.9880
$p = 30$	0.8002	0.0055	0.9768	0.7011	0.0062	0.9820	0.5050	0.0060	0.9840

Table 8
Performance of SDCM T^2 chart with Chi Square control limit ($SDCM_{CH}$)to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9734	0.0016	0.5018	0.9137	0.0011	0.8533	0.8552	0.0011	0.9592
$p = 5$	0.9762	0.0019	0.4388	0.9089	0.0016	0.8959	0.8549	0.0013	0.9604
$p = 10$	0.9749	0.0025	0.4542	0.9060	0.0022	0.9202	0.8527	0.0031	0.9644
$p = 15$	0.9664	0.0033	0.6088	0.9045	0.0029	0.9290	0.8523	0.0039	0.9627
$p = 20$	0.9679	0.0037	0.5718	0.9037	0.0034	0.9317	0.8514	0.0030	0.9737
$p = 30$	0.9595	0.0059	0.6982	0.9005	0.0077	0.9260	0.8501	0.0061	0.9649
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.8025	0.0010	0.9836	0.7026	0.0012	0.9887	0.5031	0.0010	0.9927
$p = 5$	0.8038	0.0023	0.9721	0.7027	0.0012	0.9879	0.5042	0.0016	0.9900
$p = 10$	0.8021	0.0018	0.9825	0.7022	0.0019	0.9880	0.5042	0.0026	0.9891
$p = 15$	0.8021	0.0027	0.9787	0.7012	0.0032	0.9887	0.5041	0.0034	0.9883
$p = 20$	0.8009	0.0041	0.9793	0.7011	0.0048	0.9853	0.5039	0.0043	0.9880
$p = 30$	0.8002	0.0055	0.9768	0.7011	0.0062	0.9820	0.5050	0.0060	0.9840

Table 9
Performance of the proposed chart to detect small outlier.

	ε=5%			ε=10%			ε=15%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9944	0.0018	0.0782	0.9893	0.0013	0.0955	0.9804	0.0002	0.1292
$p = 5$	0.9973	0.0027	0.0032	0.9976	0.0021	0.0051	0.9984	0.0012	0.0039
$p = 10$	0.9976	0.0025	0.0000	0.9986	0.0016	0.0000	0.9990	0.0012	0.0000
$p = 15$	0.9974	0.0028	0.0000	0.9979	0.0023	0.0000	0.9987	0.0015	0.0000
$p = 20$	0.9965	0.0037	0.0000	0.9977	0.0025	0.0000	0.9991	0.0011	0.0000
$p = 30$	0.9951	0.0052	0.0000	0.9963	0.0041	0.0000	0.9975	0.0029	0.0000
	ε=20%			ε=30%			ε=50%		
	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate	Hit rate	FP rate	FN rate
$p = 3$	0.9558	0.0002	0.2200	0.7001	0.0017	0.9956	0.4998	0.0021	0.9983
$p = 5$	0.9976	0.0007	0.0093	0.7000	0.0029	0.9934	0.5000	0.0036	0.9964
$p = 10$	0.9996	0.0006	0.0000	0.7000	0.0022	0.9949	0.5001	0.0023	0.9975
$p = 15$	0.9914	0.0107	0.0000	0.6941	0.0237	0.9643	0.4996	0.0239	0.9768
$p = 20$	0.9989	0.0014	0.0000	0.6994	0.0032	0.9947	0.4999	0.0037	0.9965
$p = 30$	0.9983	0.0021	0.0000	0.6991	0.0046	0.9924	0.5000	0.0058	0.9942

the other charts in term of Hit Rate and FN rate. As the value of the outlier increased (see Fig. 4(b) and (c)), the performance of the proposed chart increase confirmed by the similar value of FP rate for the medium and large outlier.

6.2. Non-multivariate normal distribution

In this section, the performance of the proposed chart is evaluated for non-multivariate normal distribution. Data is generated from three kinds of the non-multivariate normal distribution that is Multivariate Gamma, Multivariate Exponential and Multivariate Weibull. For the multivariate gamma distribution, the clean data are generated with the shape parameter $\boldsymbol{\gamma} = \mathbf{2}$ dan rate parameter $\boldsymbol{\theta} = \mathbf{1}$, $\mathbf{X}_{clean} \sim Mgamm(\mathbf{2}, \mathbf{1})$. Meanwhile, the Multivariate exponential data are generated with rate parameter $\boldsymbol{\theta} = \mathbf{2}$, $\mathbf{X}_{clean} \sim Mexp_p(\mathbf{2})$ and Multivaraite Weibull data are generated with shape parameter $\boldsymbol{\gamma} = \mathbf{2}$ and decay parameter $\boldsymbol{\delta} = \mathbf{1}$,$\mathbf{X}_{clean} \sim Mweib(\mathbf{2}, \mathbf{1})$. These

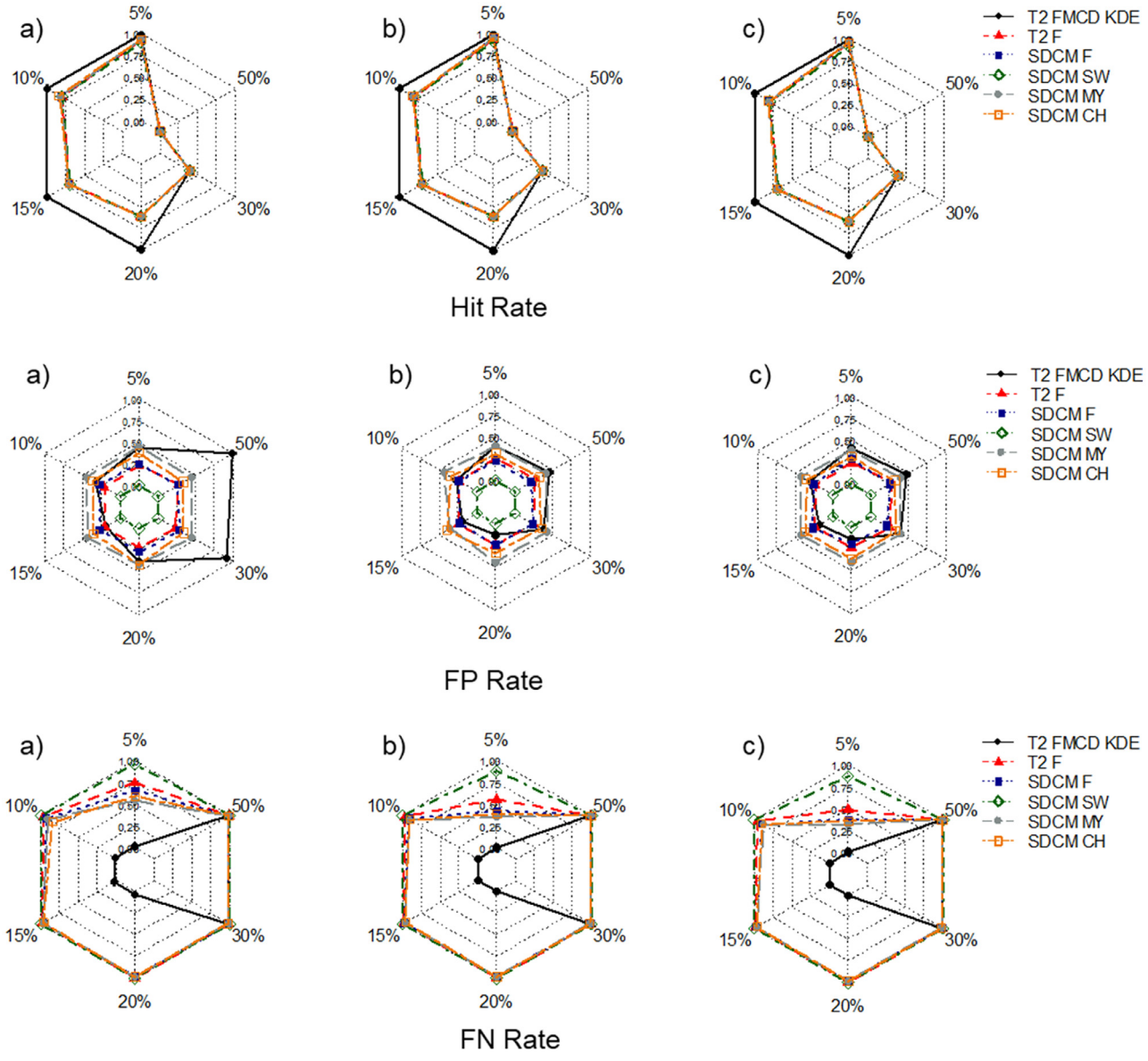

Fig. 4. Average of Hit Rate, FN Rate, and FP Rate comparison for all charts to detect: (a) small outlier, (b) medium outlier, and (c) large outlier.

distributions are chosen because they have different characteristics from symmetrical normal multivariate distributions. Thus, it can be clearly shown the adaptability of the proposed to the various types of data patterns. The percentages of outliers ε which are mixed to the clean data are 10% out of the total observations, while the number of variables $p = 10$. Fig. 5 shows the performance of the proposed chart to monitor outlier for three kinds of the non-multivariate normal distribution for $\varepsilon = 1\%$. For Multivariate Gamma and Weibull distributions, and it can be seen that the proposed chart can detect all of the outlier added. Meanwhile, for Multivariate Exponential distribution, the proposed chart has a slightly poor performance by producing more false alarm and cannot detect some of the outlier added.

Performance comparison of the proposed chart to conventional T^2 chart and the other robust charts in detecting outlier for Multivariate Gamma distribution is shown in Table 10. The T^2 based F-MCD chart with conventional F distribution control limit (Fast-MCD_F) is also included in this comparison in order to show the advantage of using the KDE control limit. It can be seen that the proposed chart can detect all outlier added while producing a small false alarm. T^2 based F-MCD with conventional control limit also detect all outliers, but it produces more false alarm. Furthermore, although the conventional T^2 chart and T^2 based on SDCM with various control limit produce smaller FP rate, these charts cannot correctly detect more outlier compared to the proposed chart.

Table 10
Performance comparison of the proposed chart to the other methods in detecting outlier for Multivariate Gamma distribution.

Method	ε=10%		
	Hit rate	FP rate	FN rate
T^2	0.9822	0.0024	0.1643
$SDCM_F$	0.9814	0.0027	0.1722
$SDCM_{SW}$	0.9832	0.0022	0.1522
$SDCM_{MY}$	0.9616	0.0001	0.3855
$SDCM_{CH}$	0.9827	0.0022	0.1612
Fast-MCD_F	0.9592	0.0424	0.0000
Proposed chart	0.9958	0.0056	0.0000

Table 11 shows the performance comparison of the proposed chart to the other methods in detecting outlier for Multivariate Exponential distribution. For this distribution, all charts have poor

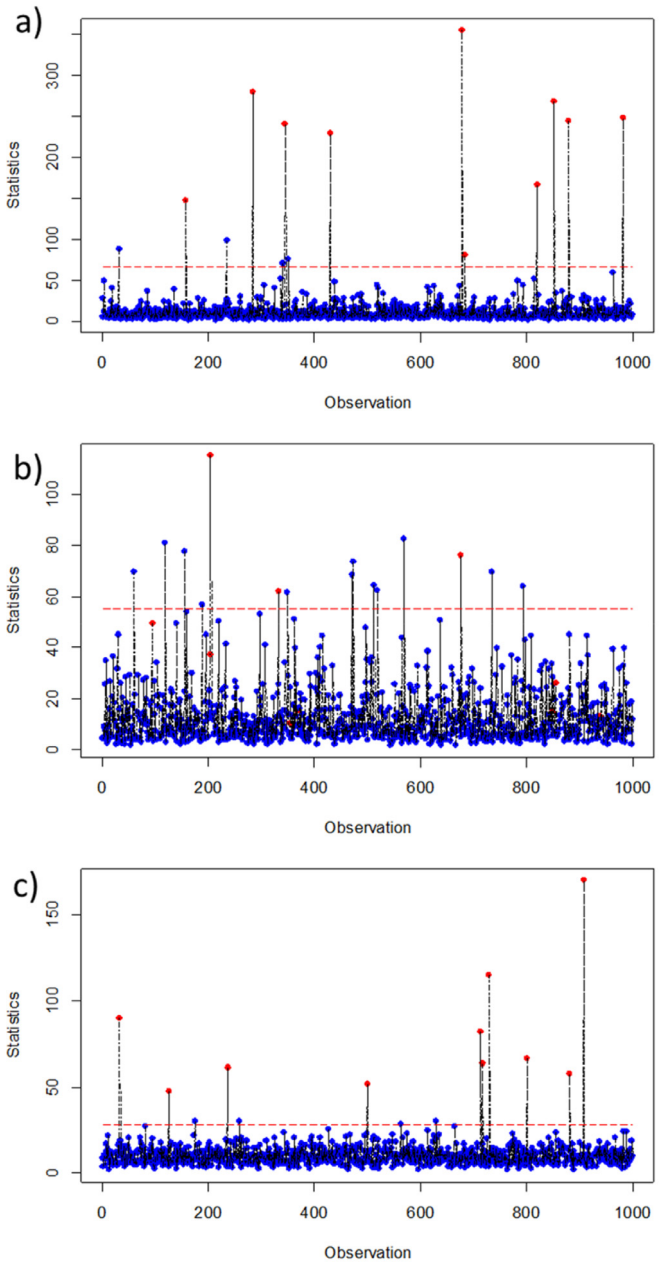

Fig. 5. The utilization of the proposed chart to one percent of outlier added for: (a) multivariate gamma distribution, (b) multivariate exponential distribution, (c) multivariate Weibull distribution.

Table 11
Performance comparison of the proposed chart to the other methods in detecting outlier for Multivariate Exponential distribution.

Method	ε=10%		
	Hit rate	FP rate	FN rate
T^2	0.9386	0.0151	0.5012
$SDCM_F$	0.9313	0.0156	0.5422
$SDCM_{SW}$	0.9312	0.0156	0.5378
$SDCM_{MY}$	0.9254	0.0033	0.7134
$SDCM_{CH}$	0.9323	0.0156	0.5376
Fast-MCD_F	0.9042	0.0811	0.2376
Proposed chart	0.9518	0.0052	0.4256

Table 12
Performance comparison of the proposed chart to the other methods in detecting outlier for Multivariate Weibull distribution.

Method	ε=10%		
	Hit rate	FP rate	FN rate
T^2	0.9855	0.0000	0.1501
$SDCM_F$	0.9867	0.0000	0.1345
$SDCM_{SW}$	0.9873	0.0000	0.1328
$SDCM_{MY}$	0.9623	0.0000	0.3776
$SDCM_{CH}$	0.9854	0.0000	0.1487
Fast-MCD_F	0.9935	0.0044	0.0293
Proposed chart	0.9956	0.0011	0.0367

performance by producing more FN. However, the proposed chart has better performance due to a smaller false alarm produced. The performance comparison for Multivariate Weibull distribution is tabulated in Table 12. An impressive result can be seen for the conventional and SDCM based T^2 charts by producing zero false alarm. However, the proposed chart obtain a higher hit rate by detecting more outlier correctly with a small false alarm. In addition, for this case, the T^2 based F-MCD chart with conventional F distribution control limit has similar performance with the proposed chart.

7. Illustration for intrusion detection system (IDS)

7.1. KDD 99 dataset

The KDD99 dataset is the most widely used and accepted benchmark dataset for network IDS. Defense Advanced Research Project Agency (DARPA) is a primary dataset derived from the MIT Lincoln Laboratory, while (Knowledge Discovery and Data Mining) KDD99 is a feature extraction feature of DARPA. KDD99 has included the types of attacks that occur, so many researchers use them to test the merits of the proposed new method. In this work, the 10% subset KDD99 dataset (494,021 records) is used because the original one is a large dataset (about five million package records). This study only uses 32 out of 34 quantitative variables because the two other quantitative variables have same values (entirely zero) (Ahsan et al., 2018b; Ahsan et al., 2018c).

Table 13 shows the performance of the proposed chart to monitor the attacks on the KDD99 dataset. For all records analyzed (n = 494,021), the proposed chart has the Hit Rate of 0.9861, FP rate of 0.0169, and FN rate of 0.0132. These results are obtained by using optimum α = 0.017 based on the highest hit rate criteria as illustrated in Fig. 6. For α = 0.017the obtained KDE control limit is 214.0065 (see Appendix 2). The full analyzed records are then used as the benchmark to see the performance of the proposed chart in monitoring the subset of the dataset. This analysis is done to see the consistency of the proposed chart in monitoring the network. The results can be used as a basis in determining how large a subset n should be considered in monitoring intrusion in a dataset without having to monitor the entire records in the dataset. For 1000 replications, the sample was chosen randomly from the full dataset for the specified number of subset n. After that, the average of the Hit Rate, FN rate, and FP rate from 1000 replications are calculated.

According to Table 13 (column 2–4), it can be seen that the proposed chart has similar performance for all subsets randomly taken from the dataset. However, when the difference average (over 1000 replications) of Hit Rate, FP Rate, and FN Rate from each replication to the benchmark value is calculated, it can be seen that there is a trend found (Table 13 column 5–7). The average of the difference will decrease monotonically when the number of samples taken is getting larger. As a consequence, there is a negative relationship between the number of samples and the average difference to the

Table 13
Performance of the proposed chart to detect intrusion in KDD99 dataset for 1000 replications using a various number of sample.

n	Hit Rate	FP Rate	FN Rate	Hit Rate Difference	FP Rate Difference	FN Rate Difference
All	0.9861	0.0169	0.0132	–	–	–
400,000	0.9861	0.0169	0.0132	0.0001	0.0002	0.0001
300,000	0.9861	0.0169	0.0132	0.0001	0.0003	0.0001
200,000	0.9861	0.0169	0.0132	0.0002	0.0004	0.0002
100,000	0.9861	0.0169	0.0131	0.0003	0.0006	0.0003
75,000	0.9861	0.0169	0.0132	0.0003	0.0008	0.0003
50,000	0.9861	0.0168	0.0132	0.0004	0.0010	0.0004
20,000	0.9861	0.0169	0.0132	0.0007	0.0017	0.0007
10,000	0.9861	0.0170	0.0131	0.0009	0.0022	0.0010
5000	0.9861	0.0171	0.0132	0.0013	0.0033	0.0014
4000	0.9861	0.0168	0.0131	0.0015	0.0036	0.0016
3000	0.9861	0.0169	0.0131	0.0017	0.0043	0.0019
2000	0.9862	0.0170	0.0130	0.0020	0.0050	0.0021
1000	0.9861	0.0170	0.0131	0.0029	0.0075	0.0032
500	0.9863	0.0157	0.0132	0.0042	0.0102	0.0047
300	0.9862	0.0174	0.0129	0.0053	0.0126	0.0060
200	0.9858	0.0171	0.0135	0.0068	0.0178	0.0071

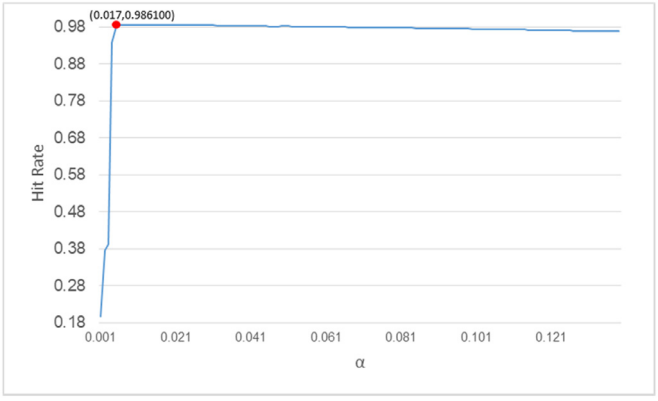

Fig. 6. Selection of α for KDD99 dataset.

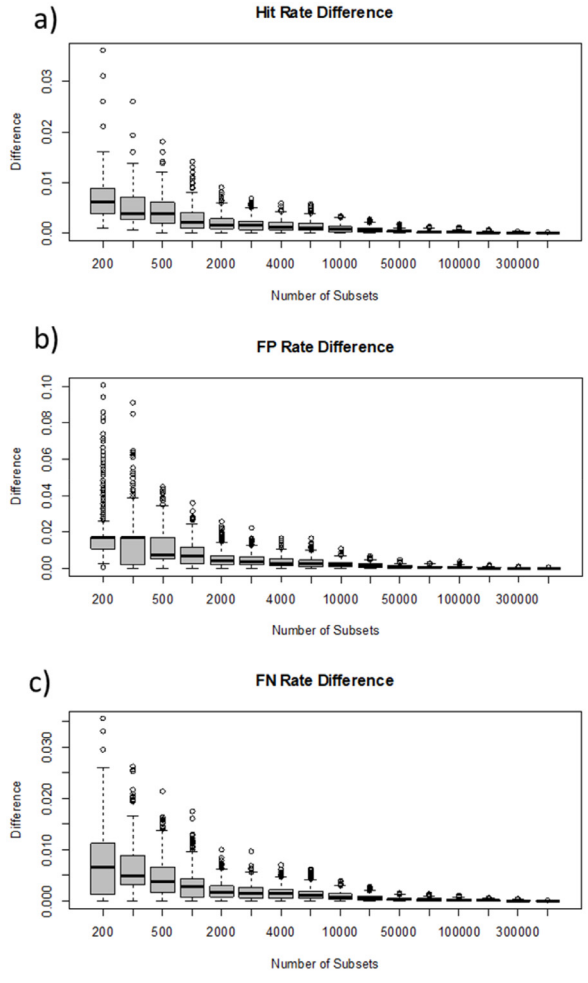

Fig. 7. Boxplot from 1000 replications for KDD99 dataset of: (a) Hit Rate Difference, (b) FP Rate Difference, and (c) FN Rate Difference.

full records. The variability of the difference of Hit Rate, FP rate, and FN rate which will also decrease when the number of samples taken is increasing as shown in Fig. 7. Therefore, it can be concluded that the smaller the number of samples taken randomly will result in more inconsistency of the proposed chart for detecting intrusion.

Furthermore, the Least Significance Different (LSD) test is employed to determine the subset which has a significant difference to the others. This test is done by comparing the average value of differences in Hit Rate, FP Rate, and FN Rate from larger samples (n_1) to smaller samples (n_2) in sequence as reported in Table 14. Under the null hypothesis that the value of the average difference of each subset is the same. The average value of the differences is still considered not statistically different value until 50,000 samples for the significance level 0.05. Thus, under the assumption that the largest samples were taken (400,000) produce the same result as produced from the complete dataset. The proposed method still at its best performance for the subset with 10.12% samples randomly taken from the total dataset records of KDD99 dataset. Using this small subset will reduce 90.02% detecting time as reported in Table 15.

7.2. NSL-KDD dataset

The NSL-KDD dataset used in this research to enrich the empirical evidence about the performance of the proposed chart. This dataset is firstly introduced by Tavallaee, Bagheri, Lu and Ghorbani (2009) as a solution for obsolete KDD-99 dataset (Stolfo, 1999) which has been attainable for more than 15 years. The NSL-KDD dataset, which has 125,973 connection records, consists of 41 variables with 34 continuous variables and seven categorical variables. Similar to the previous dataset, this study only uses 32 continuous variables because the value of the rest continuous variables is equal to zero.

Table 14
Comparison of Hit Rate, FP Rate, and FN Rate average from each subset for KDD99 dataset.

Number of Samples		Hit Rate		FP Rate		FN Rate	
n_1	n_2	Difference	p-value	Difference	p-value	Difference	p-value
400,000	300,000	0.0001	0.6510	0.0001	0.6297	0.0001	0.6354
300,000	200,000	0.0001	0.5504	0.0001	0.5226	0.0001	0.5035
200,000	100,000	0.0001	0.2266	0.0002	0.3025	0.0001	0.2595
100,000	75,000	0.0001	0.6159	0.0001	0.5199	0.0001	0.5960
75,000	**50,000**	**0.0001**	**0.4813**	**0.0002**	**0.3492**	**0.0001**	**0.4563**
50,000	20,000	0.0003	0.0013	0.0007	0.0028	0.0003	0.0021
20,000	10,000	0.0003	0.0048	0.0006	0.0096	0.0003	0.0024
10,000	5000	0.0004	0.0001	0.0011	0.0001	0.0004	0.0001
5000	4000	0.0001	0.1274	0.0003	0.1577	0.0002	0.0794
4000	3000	0.0003	0.0037	0.0006	0.0047	0.0003	0.0017
3000	2000	0.0003	0.0018	0.0007	0.0011	0.0002	0.0121
2000	1000	0.0009	0.0001	0.0025	0.0001	0.0011	0.0001
1000	500	0.0013	0.0001	0.0026	0.0001	0.0015	0.0001
500	300	0.0011	0.0001	0.0024	0.0001	0.0013	0.0001
300	200	0.0014	0.0001	0.0052	0.0001	0.0011	0.0001

Table 15
Comparison of execution time from each subset for KDD99 dataset.

n	Building Normal Profile (sec)	Detection (sec)	Time efficiency (%)
All	1.3689	1.6339	–
400,000	1.3718	1.3271	18.7772
300,000	1.3706	0.9685	40.7246
200,000	1.3704	0.6471	60.3954
100,000	1.3681	0.3216	80.3170
75,000	1.3868	0.2473	84.8644
50,000	**1.3688**	**0.1631**	**90.0177**
20,000	1.3828	0.0678	95.8504
10,000	1.3775	0.0555	96.6032
5000	1.3653	0.0170	98.9595
4000	1.3724	0.0132	99.1921
3000	1.3861	0.0098	99.4002
2000	1.3737	0.0068	99.5838
1000	1.3895	0.0035	99.7858
500	1.3886	0.0019	99.8837
300	1.3776	0.0013	99.9204
200	1.3856	0.0009	99.9449

Table 16
Performance of the proposed chart to detect intrusion in NSL-KDD dataset for 1000 replications using a various number of samples.

n	Hit Rate	FP Rate	FN Rate	Hit Rate Difference	FP Rate Difference	FN Rate Difference
All	0.9171	0.0624	0.1064	–	–	–
100,000	0.9171	0.0625	0.1064	0.0003	0.0004	0.0005
50,000	0.9171	0.0624	0.1064	0.0008	0.0009	0.0012
30,000	0.9172	0.0623	0.1064	0.0011	0.0014	0.0018
20,000	0.9172	0.0623	0.1064	0.0015	0.0018	0.0023
10,000	0.9171	0.0625	0.1065	0.0022	0.0026	0.0035
5000	0.9171	0.0624	0.1063	0.0030	0.0038	0.0051
3000	0.9171	0.0625	0.1063	0.0038	0.0047	0.0065
1000	0.9171	0.0625	0.1063	0.0070	0.0084	0.0113
500	0.9171	0.0628	0.1061	0.0098	0.0116	0.0158
300	0.9171	0.0621	0.1067	0.0129	0.0153	0.0203
100	0.9171	0.0612	0.1079	0.0221	0.0273	0.0362
50	0.9171	0.0604	0.1089	0.0296	0.0371	0.0514

The performance of the proposed chart to detect the anomalies in the NSL-KDD dataset for a various number of samples is reported in Table 16. The optimum $\alpha = 0.062$ is selected based on the highest hit rate criteria as shown in Fig. 8. This optimum α yields KDE control limit of 80.3441 (see Appendix 3). According to Table 16, the proposed chart has Hit Rate 0.9171, FP rate 0.0624, and FN rate 0.1064 for a full dataset of 125,973 connection records. Similar to the previous dataset, the consistency of the proposed chart to detect an intrusion is decreasing as the number of samples randomly taken is getting smaller as depicted in Fig. 9. By applying the same procedure applied on the previous dataset, the proposed is chart starting to show a significant performance difference for the subset with 5000 samples taken as tabulated in Table 17. This empirical result shows that the proposed chart is still reliable when used to monitor 7.94% of all records in NSL-KDD dataset with 90.53% time efficiency, as reported in Table 18.

7.3. UNSW-NB 15

The UNSW-NB 15 raw dataset version was generated by the IXIA Perfect Storm device for capturing a hybrid of real modern normal activities and synthetic contemporary attack behaviors. The tool is employed to capture the raw traffic with the size of 100 GB with nine types of attacks, such as Fuzzers, Analysis, Backdoors,

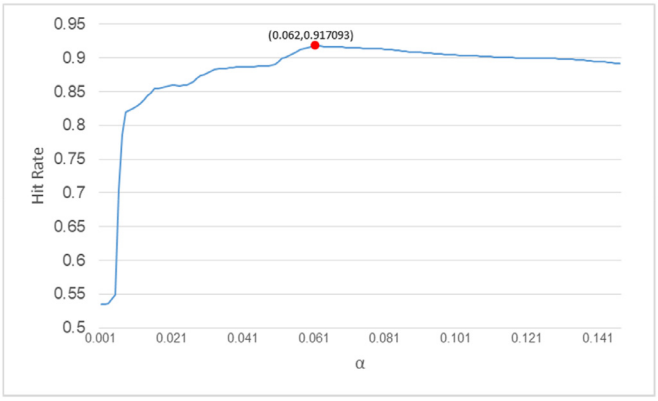

Fig. 8. Selection of α for NSL dataset.

DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms are presented. The 49 features with the class label creation were created by utilizing Argus, Bro-IDS tools, and twelve algorithms. In this paper, the training set of the dataset with 175,341 records is used to evaluate the performance of the proposed chart.

By choosing the highest hit rate criterion, the optimum $\alpha = 0.274$ with KDE control limit of 5.30588 (see Appendix 4) is selected for this dataset as depicted in Fig. 10. The proposed chart has the Hit Rate 0.9101, FP rate 0.2748, and FN rate 0.0031 as shown in Table 19. For this dataset, using the same steps of analysis on the previous dataset, the performance of the proposed chart has similar monotonically increasing trend (in term of the average of difference and consistency) to the other datasets when the samples size is getting larger as presented in Fig. 11 and Table 19 (column 5–7). According to Table 20, by using the LSD test, the proposed chart still has similar performance for the subset with 20,000 samples (11.41% of the total records). According to Table 21, this can reduce the execution time by 94.16% in the detection phase.

Therefore, based on the information from the three datasets, it can be concluded that the proposed method still has the same level of accuracy and precision to the full dataset when the number of samples n randomly taken is at the interval of 8%−12% of the total dataset records. This way will also increase time efficiency by more than 90%. For the smaller proportion of samples taken, there is no statistical guarantee that the proposed chart will be performed better or similar to the result of the complete dataset.

7.4. The comparison to the other approaches of control charts

In this subsection, the performance of the proposed chart is compared with conventional T^2 control charts and robust T^2 chart based on SDCM with various control limits. The control limits

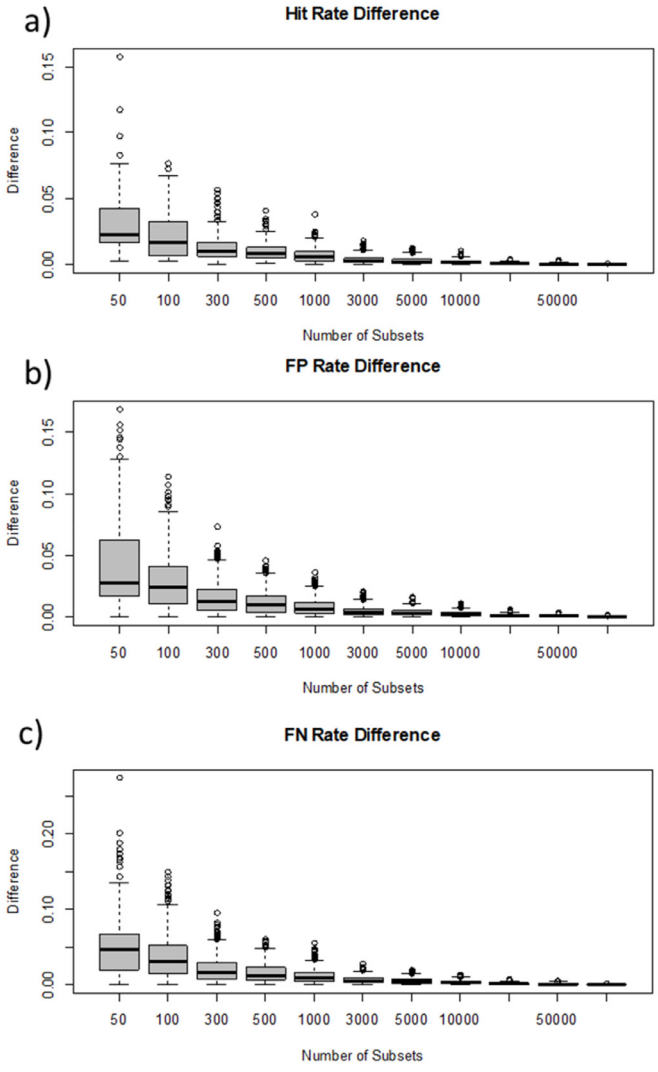

Fig. 9. Boxplot from 1000 replications for NSL-KDD dataset of: (a) Hit Rate Difference, (b) FP Rate Difference, and (c) FN Rate Difference.

of the T^2 chart based on SDCM are estimated using F distribution control limit ($SDCM_F$), Sullivan and Woodall control limit ($SDCM_{SW}$), Mason and Young control limit ($SDCM_{MY}$), and Chi-Square control limit ($SDCM_{CH}$). The performance comparison between the proposed chart with the other control chart methods for three datasets is shown in Table 22.

For KDD99 dataset, the proposed chart competes closely with the standard T^2 chart. However, the proposed chart has a bet-

Table 17
Comparison of Hit Rate, FP Rate, and FN Rate average from each subset for NSL-KDD dataset.

Number of Samples		Hit Rate		FP Rate		FN Rate	
n_1	n_2	Difference	p-value	Difference	p-value	Difference	p-value
100,000	50,000	0.0004	0.2819	0.0005	0.2898	0.0007	0.2884
50,000	30,000	0.0004	0.3418	0.0005	0.3101	0.0006	0.3810
30,000	20,000	0.0003	0.4334	0.0004	0.4354	0.0005	0.4388
20,000	**10,000**	**0.0007**	**0.0681**	**0.0008**	**0.0921**	**0.0011**	**0.0954**
10,000	5000	0.0008	0.0414	0.0012	0.0127	0.0017	0.0136
5000	3000	0.0008	0.0451	0.0009	0.0501	0.0014	0.0344
3000	1000	0.0032	0.0001	0.0037	0.0001	0.0048	0.0001
1000	500	0.0028	0.0001	0.0032	0.0001	0.0045	0.0001
500	300	0.0032	0.0001	0.0037	0.0001	0.0045	0.0001
300	100	0.0092	0.0001	0.0121	0.0001	0.0159	0.0001
100	50	0.0075	0.0001	0.0098	0.0001	0.0152	0.0001

Table 18
Comparison of execution time from each subset for NSL-KDD dataset.

N	Building Normal Profile (sec)	Detection (sec)	Time efficiency (%)
All	0.6406	0.4383	–
100,000	0.6587	0.3637	17.0203
50,000	0.6247	0.1794	59.0691
30,000	0.6288	0.1097	74.9715
20,000	0.6460	0.0772	82.3865
10,000	**0.6369**	**0.0415**	**90.5316**
5000	0.6378	0.0375	91.4442
3000	0.6212	0.0160	96.3495
1000	0.6332	0.0086	98.0379
500	0.6280	0.0067	98.4714
300	0.6137	0.0055	98.7452
100	0.6443	0.0004	99.9087
50	0.6197	0.0009	99.7947

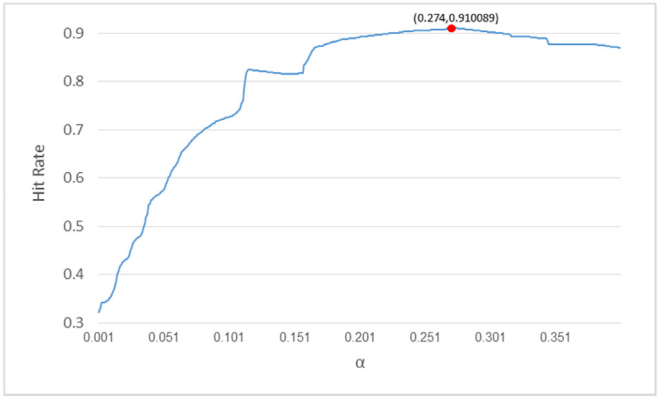

Fig. 10. Selection of α for UNSW-NB 15 dataset.

ter performance than the standard chart in the term of Hit Rate and FP rates. For this dataset, SDCM based T^2 charts show poor performance as can be seen from the high FP rate resulted. The proposed chart also shows the excellent performance for the NSL-KDD dataset. Despite having the results that are not much different from other methods, especially the T^2 chart based on SDCM with Mason and Young control limit, the proposed chart shows better results than other charts in terms of Hit Rate and FP rate. Moreover, for the UNSW-NB 15 dataset, the proposed chart has a better performance than the other methods in all aspects measured.

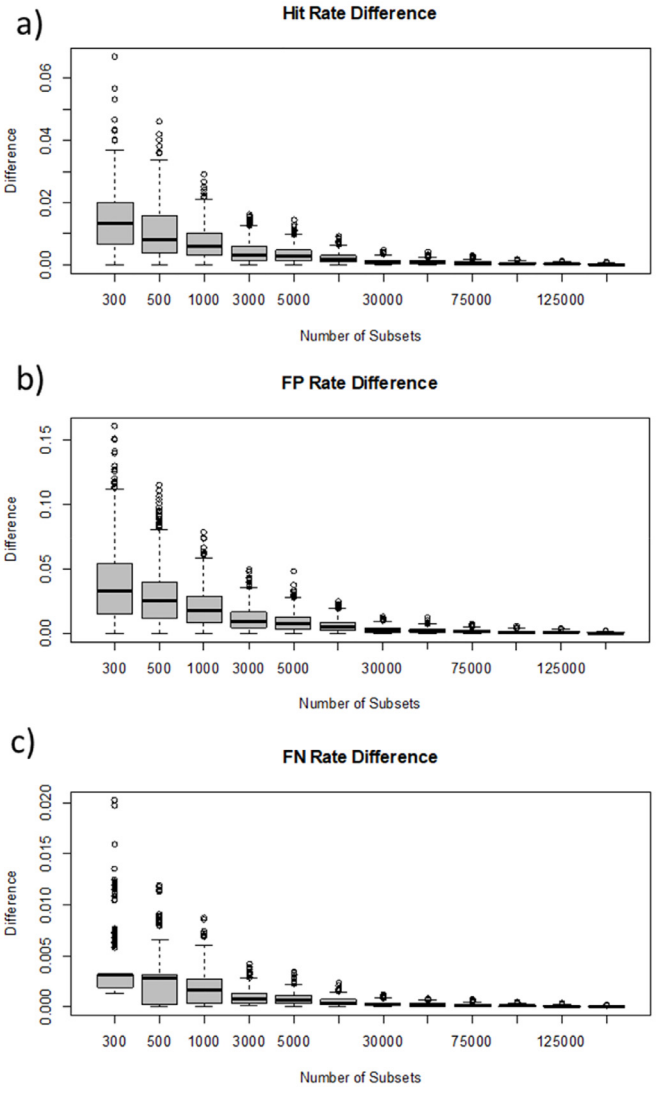

Fig. 11. Boxplot from 1000 replications for UNSW-NB 15 dataset of: (a) Hit Rate Difference, (b) FP Rate Difference, and (c) FN Rate Difference.

7.5. Comparison to the other methods

In this subsection, the proposed chart is then compared to the other techniques from some literature to show the contri-

Table 19
Performance of the proposed chart to detect intrusion in UNSW-NB 15 dataset for 1000 replications using a various number of sample.

n	Hit Rate	FP Rate	FN Rate	Hit Rate Difference	FP Rate Difference	FN Rate Difference
All	0.9101	0.2748	0.0031	–	–	–
150,000	0.9101	0.2748	0.0031	0.00023	0.0006	0.0001
125,000	0.9101	0.2748	0.0031	0.00034	0.0010	0.0001
100,000	0.9101	0.2748	0.0032	0.00047	0.0013	0.0001
75,000	0.9101	0.2748	0.0031	0.00062	0.0017	0.0002
50,000	0.9101	0.2747	0.0031	0.00088	0.0024	0.0002
30,000	0.9101	0.2748	0.0031	0.00120	0.0033	0.0003
20,000	0.9101	0.2750	0.0031	0.00151	0.0015	0.0004
10,000	0.9101	0.2748	0.0031	0.00220	0.0061	0.0005
5000	0.9102	0.2748	0.0032	0.00324	0.0090	0.0007
3000	0.9102	0.2747	0.0032	0.00410	0.0117	0.0010
1000	0.9101	0.2752	0.0032	0.00722	0.0198	0.0017
500	0.9101	0.2746	0.0031	0.01038	0.0286	0.0023
300	0.9101	0.2741	0.0032	0.01376	0.0378	0.0033

Table 20
Comparison of Hit Rate, FP Rate, and FN Rate average from each subset for UNSW-NB 15 dataset.

Number of Samples		Hit Rate		FP Rate		FN Rate	
n_1	n_2	Difference	p-value	Difference	p-value	Difference	p-value
150,000	125,000	0.0001	0.5267	0.0003	0.4904	0.0000	0.5052
125,000	100,000	0.0001	0.4557	0.0003	0.5126	0.0000	0.5425
100,000	75,000	0.0001	0.4126	0.0004	0.4287	0.0000	0.2442
75,000	50,000	0.0003	0.1531	0.0007	0.1767	0.0001	0.2245
50,000	30,000	0.0003	0.0797	0.0009	0.0614	0.0001	0.0431
30,000	**20,000**	**0.0003**	**0.0800**	**0.0010**	**0.0588**	**0.0001**	**0.1177**
20,000	10,000	0.0007	0.0001	0.0018	0.0003	0.0002	0.0004
10,000	5000	0.0010	0.0000	0.0029	0.0000	0.0002	0.0000
5000	3000	0.0009	0.0000	0.0026	0.0000	0.0003	0.0000
3000	1000	0.0031	0.0000	0.0081	0.0000	0.0007	0.0000
1000	500	0.0032	0.0000	0.0088	0.0000	0.0006	0.0000
500	300	0.0034	0.0000	0.0092	0.0000	0.0010	0.0000

Table 21
Comparison of execution time from each subset for UNSW-NB 15 dataset.

n	Building Normal Profile (sec)	Detection (sec)	Time efficiency (%)
All	0.6367	0.5533	13.6635
150,000	0.6248	0.4777	30.3452
125,000	0.6129	0.3854	43.0146
100,000	0.6019	0.3153	57.4191
75,000	0.6427	0.2356	71.8778
50,000	0.6192	0.1556	82.9387
30,000	0.6249	0.0944	88.5957
20,000	**0.6389**	**0.0631**	**94.1623**
10,000	0.6361	0.0323	94.2888
5000	0.6182	0.0316	98.1927
3000	0.6236	0.0100	98.4818
1000	0.6023	0.0084	99.2590
500	0.6383	0.0041	99.5301
300	0.6070	0.0026	13.6635

Table 22
Performance comparison of the proposed chart to the other control charts in monitoring intrusion.

Control Chart	Hit Rate	FP	FN	FP Rate	FN Rate
KDD99 dataset					
T^2	0.9799	6542	3384	0.0673	0.0085
DCM_F	0.9108	42,789	1284	0.4399	0.0032
$SDCM_{SW}$	0.9484	23,674	1808	0.2434	0.0046
$SDCM_{MY}$	0.9107	42,811	1283	0.4401	0.0032
$SDCM_{CH}$	0.9108	42,803	1283	0.4400	0.0032
Proposed chart	0.9861	1645	5222	0.0169	0.0132
NSL-KDD dataset					
T^2	0.9133	5494	5428	0.0937	0.0806
$SDCM_F$	0.9134	5495	5417	0.0937	0.0804
$SDCM_{SW}$	0.9171	6170	4280	0.1052	0.0636
$SDCM_{MY}$	0.9133	5492	5429	0.0937	0.0806
$SDCM_{CH}$	0.9133	5492	5427	0.0937	0.0806
Proposed chart	0.9171	4204	6240	0.0624	0.1064
UNSW-NB 15 dataset					
T^2	0.7023	8022	44,170	0.1433	0.3701
$SDCM_F$	0.8803	19,730	1258	0.3523	0.0105
$SDCM_{SW}$	0.8854	16,809	3281	0.3002	0.0275
$SDCM_{MY}$	0.8803	19,733	1256	0.3524	0.0105
$SDCM_{CH}$	0.8803	19,732	1256	0.3524	0.0105
Proposed chart	0.9101	15,390	375	0.2748	0.0031

bution and superiority of the proposed method. For the KDD99 dataset, the proposed chart is compared with Logistic Regression (LR), Naïve Bayes (NB), Artificial Neural Network (ANN) Decision Tree (DT), and Expectation-Maximization (EM) clustering as provided in Moustafa and Slay (2016), Long Short Term Memory Recurrent Neural Network Classifier (LSTM-RNN) by Kim, Kim, Thu and Kim (2016), as well as Multi-level hybrid SVM-extreme learning machine (ELM) based on modified K-means by Al-Yaseen et al. (2017). For this dataset, the proposed chart demonstrated a high performance, which can be seen from the high hit rate and the low FP rate, as shown in Table 23.

Table 23
Performance comparison of the proposed chart to the other methods in monitoring intrusion.

Methods	Hit Rate	FP Rate
KDD99 dataset		
LR (Witten & Frank, 2005)	0.9275	NA
NB (Shyu et al., 2005)	0.9500	0.0500
ANN (Witten & Frank, 2005)	0.9704	0.0148
EM Clustering (Salem & Buehler, 2012)	0.7806	0.1037
GA (Hoque et al., 2012)	0.9000	0.3046
DT (Bro-IDS Tool, 2014)	0.9230	0.1171
LSTM-RNN (Kim et al., 2016)	0.9693	0.1004
Multi-level hybrid SVM-ELM K-means (Al-Yaseen et al., 2017)	0.9575	0.0187
Proposed Method	0.9861	0.0169
NSL –KDD dataset		
Hybrid DT (Farid et al., 2014)	0.8192	0.1740
Hybrid NB (Farid et al., 2014)	0.8239	0.1640
NB (Singh et al., 2015)	0.8729	0.1735
LR (Belavagi & Muniyal, 2016)	0.8400	0.1700
SVM (Belavagi & Muniyal, 2016)	0.7500	0.2400
Proposed Method	0.9171	0.0624
UNSW-NB 15 dataset		
DT (Bro-IDS Tool, 2014)	0.8556	0.1578
LR (Witten & Frank, 2005)	0.8315	0.1848
NB (Shyu et al., 2005)	0.8207	0.1856
ANN (Witten & Frank, 2005)	0.8134	0.2113
EM (Salem & Buehler, 2012)	0.7847	0.2379
Fuzzy SOM (A Karami & Guerrero-Zapata, 2014)	0.9061	0.0942
Skip-Gram (Carrasco & Sicilia, 2018)	0.9102	0.0061
Proposed Method	0.9101	0.2748

For the NSL-KDD dataset, the proposed chart is compared with Hybrid DT, Hybrid NB, NB, LR, and SVM methods. According to Table 23, the proposed chart has a better performance than the other techniques compared in this dataset. Furthermore, LR, DT, ANN, EM clustering, and DT as found in Moustafa and Slay, (2016), Fuzzy Self Organizing Map (SOM), and Skip-Gram techniques are used as the benchmarks of the proposed chart for UNSW-NB 15 dataset. For this dataset, there are three techniques which have similar performance to the proposed chart. Compared to the Skip-Gram method, the proposed chart has the similar Hit Rate to the compared method. However, the proposed chart has slightly higher FP rate for the similar Hit Rate for this case. As a consequence, the value FN rate of the proposed chart is much lower than Skip-Gram. Thus, it can be concluded that the proposed chart is able to detect more attacks than the method confirmed by the low value of the FN rate.

8. Conclusion

This paper proposes a robust and adaptive T^2 Hotelling's chart for network anomalies detection. The Fast-MCD algorithm is em-

ployed to ensure the robust mean vector and covariance matrix of the proposed chart. Meanwhile, the KDE is used to estimate the control limit of the proposed chart adaptively. Through the simulation studies, for the non-multivariate normal distribution, the proposed chart shows its adaptability to follow any pattern of data by producing low false alarm. However, the proposed chart has a difficulty to monitor the outliers for the data generated from Multivariate Exponential distribution. The proposed chart has a better performance in detecting outlier than the conventional T^2 and Robust T^2 chart based on SDCM with various control limit. The same results was also found when the proposed chart is utilized to monitor intrusion in KDD99, NSL-KDD, and UNSW-NB 15 datasets. When the proposed chart is applied to monitor the subsets of the dataset, the smaller subset produce the larger average of the performance difference when applied on the full dataset. Surprisingly, by using a small subset (8%−12% samples), the proposed chart performs as good as when it is applied on the complete dataset. This evidence shows that the proposed chart is still reliable to monitor a small portion of the network traffic packages which can reduce the computational time at about 90%. Compared to the other techniques, the performance of the proposed chart demonstrated high performance by surpassing all of the benchmarks.

In this work, it can be seen that the proposed chart can achieve at about 98% accuracy for KDD99 dataset, while for the other datasets, the proposed chart can only obtain accuracy at about 91%. This different performance can be caused by the nonlinear relationship which cannot be accommodated by the current work. Thus, combining methods that can handle nonlinearity problems such as Support Vector Machine (SVM) or Neural Networks (NN) can improve the performance of the current work. The Fast-MCD algorithm is the old method. The utilization of newest robust method which has faster computational time such as Det-MCD (Hubert, Rousseeuw & Verdonck, 2012) and Index Set Equality (ISE) (Lim & Midi, 2016) can strengthen the performance of current work. Furthermore, PCA Mix method can also be considered as the future work to monitor not only the continuous features but also the categorical features. Finally, by considering the robustness, adaptability, and rapid calculation of current work, it is possible to extend the current work in monitoring the intrusion in the Big Data Systems, Cloud, and Internet of Things (IoT).

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Muhammad Ahsan: Formal analysis, Data curation, Conceptualization, Resources, Writing - original draft. **Muhammad Mashuri:** Formal analysis, Resources, Resources. **Muhammad Hisyam Lee:** Data curation, Formal analysis, Resources. **Heri Kuswanto:** Data curation, Formal analysis. **Dedy Dwi Prastyo:** Data curation, Formal analysis, Writing - original draft.

Acknowledgement

The authors are thankful to the anonymous reviewers for their constructive comments help us to improved this manuscript. The authors also appreciate the Research, Technology, and Higher Education Ministry, Republic of Indonesia for providing PDUPT program under grant 5/E1/KP.PTNBH/2019.

Appendix 1. List of Symbol and Notation

Symbor or Notation	Meaning or definition
$\mathbf{x}_i$	Vector random for i observation
$\bar{\mathbf{x}}$	Sample mean vector
$\mathbf{S}$	Sample covariance matrix
T_i^2	Statistic of T^2 Hotelling for i observation
n	Number of observations
p	Number of variables
CL	Conventional control limit
$\mathbf{S}_{n_t}$	Robust estimator for covariance matrix
$\mathbf{T}_{n_t}$	Robust estimator for mean vector
d	Mahalanobis distance
$\#G$	The number of elements of set G
$\subset$	Subset
$\{1, 2, \cdots n\}$	Set with element 1,2,..n
$T^2_{FMCD,i}$	Statistic of T^2 Hotelling with fast MCD estimator for i observation
$K(u)$	Kernel function
h	Smoothing parameter
$\hat{h}$	Optimum Smoothing parameter
$\hat{f}_h(t)$	Probability distribution function of T^2 statistic calculated using kernel function
CL_{kernel}	KDE control limit for T^2 statistic
$\hat{\tilde{f}}_h(\tilde{T})$	Empirical density of T^2_{FMCD} statistic using KDE
$\hat{\tilde{F}}_h(\tilde{t})$	Distribution function of $\hat{\tilde{f}}_h(\tilde{T})$
$\eta_{\min}$	Minimum value of $\tilde{T}$
$\eta_{\max}$	Maximum value of $\tilde{T}$
CL_{KDE}	KDE control limit for T^2_{FMCD} statistic
$\mathbf{X}_{normal}$	Normal connection data
$\mathbf{X}_{test}$	New connection data
$\mathbf{T}_{n_t.Normal}$	Robust estimator for mean vector of normal connection data
$\mathbf{S}_{n_t.Normal}$	Robust estimator for covariance matrix of normal connection data
$\mathbf{X}_{clean}$	Clean data
$\mathbf{X}_{cont}$	Contaminated data
$\boldsymbol{\mu}$	Mean vector
$\boldsymbol{\Sigma}$	Covariance matrix
$N_p(\boldsymbol{\mu},\mathbf{I})$	Multivariate Distribution with mean vector $\boldsymbol{\mu}$ dan covariance matrix $\mathbf{I}$
$\boldsymbol{\mu}_{clean}$	Mean vector of clean data
$\boldsymbol{\mu}_{cont}$	Mean vector of contaminated data
ε	Percentages of outliers
α	Significance level
Fast-MCD_F	T^2 based F-MCD chart with conventional F distribution control limit Fast-MCD_F
$SDCM_F$	Robust T^2 chart based on SDCM using F distribution control limit
$SDCM_{SW}$	Robust T^2 chart based on SDCM using Sullivan and Woodall control limit
$SDCM_{MY}$	Robust T^2 chart based on SDCM using Mason and Young control limit
$SDCM_{CH}$	Robust T^2 chart based on SDCM using Chi-Square distribution control limit

Appendix 2. KDE control limit calculation for KDD99 dataset

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.001	49,434.19	0.196910658	2	396,741	2.06E-05	0.999995
0.002	1946.421	0.375633829	168	308,282	0.001727009	0.777032
0.003	1169.655	0.391667561	264	300,265	0.002713872	0.756825
0.004	781.4577	0.93973738	361	29,410	0.003711014	0.074129
0.005	621.5025	0.981075703	483	8866	0.004965151	0.022347
0.006	503.8557	0.984970275	573	6852	0.005890335	0.017271
0.007	436.6819	0.98530022	658	6604	0.006764119	0.016646
0.008	392.8891	0.985433818	725	6471	0.007452867	0.01631
0.009	358.3744	0.985156501	926	6407	0.00951911	0.016149
0.01	341.6737	0.985124114	966	6383	0.009930303	0.016089
0.011	312.7259	0.985201034	1038	6273	0.01067045	0.015811
0.012	288.9739	0.985028977	1166	6230	0.011986266	0.015703
0.013	272.6443	0.984974323	1231	6192	0.012654454	0.015607
0.014	255.5726	0.98475166	1387	6146	0.014258106	0.015491
0.015	242.5832	0.984745588	1482	6054	0.015234688	0.015259
0.016	228.1093	0.985235445	1555	5739	0.015985115	0.014465
0.017	**214.0065**	**0.986100**	**1645**	**5222**	**0.016910298**	**0.013162**
0.018	203.2438	0.985994523	1728	5191	0.017763523	0.013084
0.019	194.3368	0.985826513	1852	5150	0.01903822	0.012981
0.02	189.1411	0.98567065	1942	5137	0.019963404	0.012948
0.021	185.8009	0.985514786	2030	5126	0.020868028	0.01292
0.022	183.9453	0.985033025	2277	5117	0.023407142	0.012898
0.023	180.234	0.98498242	2313	5106	0.023777216	0.01287
0.024	178.0073	0.984972299	2330	5094	0.023951973	0.01284
0.025	172.4404	0.985061364	2439	4941	0.025072473	0.012454
0.026	166.1313	0.984836677	2587	4904	0.026593886	0.012361
0.027	159.451	0.98479012	2639	4875	0.027128436	0.012288
0.028	151.2862	0.984723321	2728	4819	0.02804334	0.012146
0.029	143.8637	0.98484275	2813	4675	0.028917124	0.011783
0.03	137.9257	0.984737491	2917	4623	0.029986225	0.011652
0.031	132.7299	0.984579603	3049	4569	0.031343161	0.011516
0.032	128.2764	0.98447232	3129	4542	0.032165546	0.011448
0.033	124.5652	0.984385279	3193	4521	0.032823454	0.011395
0.034	120.4828	0.984152496	3329	4500	0.034221509	0.011342
0.035	117.5138	0.983988535	3432	4478	0.035280331	0.011287
0.036	114.9159	0.983877204	3506	4459	0.036041037	0.011239
0.037	112.6892	0.983709195	3604	4444	0.037048459	0.011201
0.038	110.4624	0.98342783	3753	4434	0.038580152	0.011176
0.039	108.9779	0.9832497	3850	4425	0.039577294	0.011153
0.04	107.1223	0.983065497	3952	4414	0.040625835	0.011126
0.041	105.6378	0.982848907	4067	4406	0.041808014	0.011105
0.042	104.1533	0.982709237	4149	4393	0.042650959	0.011073
0.043	102.6688	0.982551349	4235	4385	0.043535023	0.011052
0.044	101.1843	0.982330711	4349	4380	0.044706922	0.01104
0.045	99.69975	0.982209258	4421	4368	0.045447069	0.01101
0.046	97.84412	0.982008862	4527	4361	0.04653673	0.010992
0.047	95.24624	0.981895506	4606	4338	0.047348835	0.010934
0.048	91.90611	0.981830732	4695	4281	0.048263739	0.01079
0.049	88.56598	0.98246026	4774	3891	0.049075844	0.009807
0.05	85.22585	0.982371195	4875	3834	0.050114106	0.009664
0.051	81.88572	0.982193065	4997	3800	0.051368244	0.009578
0.052	79.65896	0.982126266	5064	3766	0.052056991	0.009492
0.053	76.68996	0.981935991	5190	3734	0.053352248	0.009412
0.054	74.4632	0.981792272	5293	3702	0.054411069	0.009331
0.055	72.9787	0.981642481	5385	3684	0.055356812	0.009286
0.056	71.4942	0.981385407	5533	3663	0.056878225	0.009233
0.057	70.38082	0.98123764	5617	3652	0.05774173	0.009205
0.058	69.63857	0.981120236	5687	3640	0.058461317	0.009175
0.059	68.15406	0.98086721	5829	3623	0.059921051	0.009132
0.06	67.41181	0.980774097	5887	3611	0.06051728	0.009102
0.061	66.29844	0.980648596	6001	3559	0.061689179	0.008971
0.062	65.55618	0.980506902	6115	3515	0.062861079	0.00886
0.063	64.81393	0.9804482	6206	3453	0.063796542	0.008703
0.064	64.07168	0.980336868	6305	3409	0.064814244	0.008592
0.065	63.32943	0.98014052	6420	3391	0.065996423	0.008547
0.066	62.58718	0.979853083	6581	3372	0.067651473	0.008499

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.067	61.84493	0.979804502	6642	3335	0.068278542	0.008406
0.068	61.10268	0.97975997	6723	3276	0.069111207	0.008257
0.069	60.36042	0.979618275	6812	3257	0.070026111	0.008209
0.07	59.61817	0.979454315	6909	3241	0.071023253	0.008169
0.071	58.87592	0.979434073	6980	3180	0.07175312	0.008015
0.072	57.76254	0.979280233	7078	3158	0.072760542	0.00796
0.073	57.02029	0.979150684	7153	3147	0.073531528	0.007932
0.074	55.53579	0.978936118	7293	3113	0.074970703	0.007846
0.075	54.79354	0.978800496	7381	3092	0.075875326	0.007793
0.076	53.68016	0.97862439	7488	3072	0.076975267	0.007743
0.077	52.93791	0.978464478	7581	3058	0.07793129	0.007708
0.078	51.82453	0.97830659	7675	3042	0.078897592	0.007667
0.079	51.08228	0.978280276	7748	2982	0.079648019	0.007516
0.08	49.9689	0.978053565	7889	2953	0.081097473	0.007443
0.081	49.22665	0.977946282	7971	2924	0.081940418	0.00737
0.082	48.4844	0.977756006	8078	2911	0.083040359	0.007337
0.083	47.74215	0.97761836	8177	2880	0.08405806	0.007259
0.084	46.9999	0.977470593	8317	2813	0.085497235	0.00709
0.085	46.62877	0.977432133	8385	2764	0.086196262	0.006967
0.086	45.88652	0.977312705	8504	2704	0.08741956	0.006815
0.087	45.51539	0.977235786	8555	2691	0.087943831	0.006783
0.088	44.77314	0.977073849	8674	2652	0.089167129	0.006684
0.089	44.03089	0.976897743	8783	2630	0.090287629	0.006629
0.09	43.28864	0.976745928	8898	2590	0.091469808	0.006528
0.091	42.91751	0.976618403	8978	2573	0.092292194	0.006485
0.092	42.17526	0.976359305	9136	2543	0.093916405	0.00641
0.093	41.80414	0.976201417	9221	2536	0.094790189	0.006392
0.094	41.43301	0.976049601	9309	2523	0.095694813	0.006359
0.095	40.69076	0.975741922	9493	2491	0.097586299	0.006279
0.096	40.69076	0.975741922	9493	2491	0.097586299	0.006279
0.097	39.94851	0.975446388	9668	2462	0.099385267	0.006206
0.098	39.94851	0.975446388	9668	2462	0.099385267	0.006206
0.099	39.20626	0.975227774	9819	2419	0.100937519	0.006097
0.1	38.83513	0.97505774	9921	2401	0.101986061	0.006052
0.101	38.46401	0.974802691	10,069	2379	0.103507473	0.005996
0.102	38.46401	0.974802691	10,069	2379	0.103507473	0.005996
0.103	37.72175	0.974341172	10,333	2343	0.106221345	0.005906
0.104	37.72175	0.974341172	10,333	2343	0.106221345	0.005906
0.105	37.72175	0.196910658	10,333	2343	0.106221345	0.005906
0.106	36.9795	0.973851314	10,601	2317	0.108976336	0.00584
0.107	36.9795	0.973851314	10,601	2317	0.108976336	0.00584
0.108	36.23725	0.973272391	10,917	2287	0.112224758	0.005764
0.109	36.23725	0.973272391	10,917	2287	0.112224758	0.005764
0.11	36.23725	0.973272391	10,917	2287	0.112224758	0.005764
0.111	36.23725	0.973272391	10,917	2287	0.112224758	0.005764
0.112	35.495	0.972774437	11,189	2261	0.115020868	0.005699
0.113	35.495	0.972774437	11,189	2261	0.115020868	0.005699
0.114	35.495	0.972774437	11,189	2261	0.115020868	0.005699
0.115	34.75275	0.972383765	11,412	2231	0.117313267	0.005623
0.116	34.75275	0.972383765	11,412	2231	0.117313267	0.005623
0.117	34.0105	0.972003214	11,629	2202	0.119543987	0.00555
0.118	34.0105	0.972003214	11,629	2202	0.119543987	0.00555
0.119	33.26824	0.971527526	11,900	2166	0.122329818	0.005459
0.12	33.26824	0.971527526	11,900	2166	0.122329818	0.005459
0.121	32.52599	0.971049814	12,167	2135	0.125074529	0.005381
0.122	32.52599	0.971049814	12,167	2135	0.125074529	0.005381
0.123	32.52599	0.971049814	12,167	2135	0.125074529	0.005381
0.124	32.52599	0.971049814	12,167	2135	0.125074529	0.005381
0.125	31.78374	0.970564004	12,439	2103	0.127870639	0.005301
0.126	31.78374	0.970564004	12,439	2103	0.127870639	0.005301
0.127	31.04149	0.970029614	12,773	2033	0.131304098	0.005124
0.128	31.04149	0.970029614	12,773	2033	0.131304098	0.005124
0.129	30.29924	0.969466885	13,082	2002	0.134480561	0.005046
0.13	30.29924	0.969466885	13,082	2002	0.134480561	0.005046
0.131	30.29924	0.969466885	13,082	2002	0.134480561	0.005046
0.132	30.29924	0.969466885	13,082	2002	0.134480561	0.005046
0.133	30.29924	0.969466885	13,082	2002	0.134480561	0.005046
0.134	29.55699	0.968865696	13,416	1965	0.13791402	0.004953
0.135	29.55699	0.968865696	13,416	1965	0.13791402	0.004953
0.136	29.55699	0.968865696	13,416	1965	0.13791402	0.004953
0.137	28.81474	0.968215926	13,770	1932	0.141553075	0.00487

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.138	28.81474	0.968215926	13,770	1932	0.141553075	0.00487
0.139	28.81474	0.968215926	13,770	1932	0.141553075	0.00487
0.14	28.81474	0.968215926	13,770	1932	0.141553075	0.00487
0.141	28.07248	0.967554011	14,135	1894	0.145305208	0.004774
0.142	28.07248	0.967554011	14,135	1894	0.145305208	0.004774
0.143	28.07248	0.967554011	14,135	1894	0.145305208	0.004774
0.144	28.07248	0.967554011	14,135	1894	0.145305208	0.004774
0.145	27.33023	0.966703845	14,593	1856	0.150013364	0.004678
0.146	27.33023	0.966703845	14,593	1856	0.150013364	0.004678
0.147	27.33023	0.966703845	14,593	1856	0.150013364	0.004678
0.148	27.33023	0.966703845	14,593	1856	0.150013364	0.004678
0.149	27.33023	0.966703845	14,593	1856	0.150013364	0.004678
0.15	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.151	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.152	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.153	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.154	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.155	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.156	26.58798	0.965248441	15,354	1814	0.157836304	0.004572
0.157	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.158	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.159	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.16	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.161	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.162	25.84573	0.96307242	16,465	1778	0.16925718	0.004481
0.163	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.164	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.165	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.166	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.167	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.168	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.169	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.17	25.10348	0.961673694	17,182	1752	0.176627809	0.004416
0.171	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.172	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.173	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.174	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.175	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.176	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.177	24.36123	0.960248653	17,910	1728	0.184111515	0.004355
0.178	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.179	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.18	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.181	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.182	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.183	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.184	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.185	23.61898	0.95922643	18,442	1701	0.189580378	0.004287
0.186	22.87672	0.958030124	19,076	1658	0.196097782	0.004179
0.187	22.87672	0.958030124	19,076	1658	0.196097782	0.004179
0.188	22.87672	0.958030124	19,076	1658	0.196097782	0.004179
0.189	22.87672	0.958030124	19,076	1658	0.196097782	0.004179
0.19	22.87672	0.958030124	19,076	1658	0.196097782	0.004179
0.191	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.192	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.193	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.194	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.195	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.196	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.197	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.198	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.199	22.13447	0.956677955	19,791	1611	0.20344785	0.004061
0.2	22.13447	0.956677955	19,791	1611	0.20344785	0.004061

Appendix 3. KDE control limit calculation for NSL-KDD dataset

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.001	31,871.86628	0.534638	2	58,621	2.97E-05	0.999846
0.002	1501.735659	0.53521	118	58,433	0.001752	0.99664
0.003	1173.94148	0.536282	167	58,249	0.00248	0.993502
0.004	796.9781751	0.542783	283	57,314	0.004202	0.977554
0.005	692.7016213	0.549777	306	56,410	0.004544	0.962135
0.006	522.8662318	0.704238	405	36,853	0.006014	0.628569
0.007	470.8467209	0.786327	450	26,467	0.006682	0.451424
0.008	396.9742647	0.820136	555	22,103	0.008241	0.376991
0.009	371.5583393	0.82193	578	21,854	0.008583	0.372744
0.01	325.0020647	0.825463	675	21,312	0.010023	0.3635
0.011	301.2488634	0.828416	734	20,881	0.010899	0.356149
0.012	281.7712383	0.83329	795	20,206	0.011805	0.344636
0.013	265.6190614	0.837846	879	19,548	0.013053	0.333413
0.014	247.0915643	0.844784	961	18,592	0.01427	0.317107
0.015	235.9275597	0.848182	983	18,142	0.014597	0.309432
0.016	217.4000627	0.854143	1082	17,292	0.016067	0.294934
0.017	207.8987822	0.855191	1141	17,101	0.016943	0.291677
0.018	196.4972455	0.855961	1234	16,911	0.018324	0.288436
0.019	187.946093	0.857596	1274	16,665	0.018918	0.28424
0.02	179.1574085	0.858557	1339	16,479	0.019883	0.281068
0.021	174.1692363	0.859335	1390	16,330	0.020641	0.278526
0.022	170.8437881	0.859335	1487	16,233	0.022081	0.276872
0.023	169.89366	0.859049	1560	16,196	0.023165	0.276241
0.024	166.5682118	0.859081	1669	16,083	0.024784	0.274313
0.025	163.2427637	0.859859	1688	15,966	0.025066	0.272318
0.026	157.5419953	0.862629	1753	15,552	0.026031	0.265257
0.027	152.5538231	0.865479	1812	15,134	0.026907	0.258127
0.028	148.5157788	0.870425	1880	14,443	0.027917	0.246341
0.029	145.1903306	0.873449	1967	13,975	0.029209	0.238359
0.03	142.5774785	0.87545	2049	13,641	0.030426	0.232662
0.031	139.0144983	0.878109	2102	13,253	0.031213	0.226045
0.032	135.6890501	0.880546	2158	12,890	0.032045	0.219853
0.033	132.1260699	0.882467	2217	12,589	0.032921	0.214719
0.034	129.0381537	0.883602	2297	12,366	0.034109	0.210916
0.035	126.6628336	0.884197	2367	12,221	0.035148	0.208443
0.036	124.2875135	0.884713	2442	12,081	0.036262	0.206055
0.037	122.1497254	0.884999	2506	11,981	0.037212	0.204349
0.038	119.7744052	0.886087	2574	11,776	0.038222	0.200853
0.039	117.8741491	0.886388	2634	11,678	0.039113	0.199181
0.04	115.736361	0.886809	2707	11,552	0.040197	0.197032
0.041	114.0736369	0.886896	2777	11,471	0.041237	0.195651
0.042	112.4109128	0.88715	2833	11,383	0.042068	0.19415
0.043	110.9857207	0.887254	2906	11,297	0.043152	0.192683
0.044	109.5605287	0.887325	2985	11,209	0.044325	0.191182
0.045	108.1353366	0.887595	3052	11,108	0.04532	0.189459
0.046	106.7101445	0.888087	3113	10,985	0.046226	0.187361
0.047	105.7600164	0.888325	3141	10,927	0.046642	0.186372
0.048	104.3348244	0.888278	3292	10,782	0.048884	0.183899
0.049	102.9096323	0.889056	3326	10,650	0.049389	0.181648
0.05	101.0093762	0.890905	3371	10,372	0.050057	0.176906
0.051	98.39652403	0.89435	3442	9867	0.051111	0.168293
0.052	96.0212039	0.899852	3514	9102	0.052181	0.155245
0.053	94.35847981	0.901558	3565	8836	0.052938	0.150708
0.054	92.4582237	0.90394	3678	8423	0.054616	0.143664
0.055	91.03303162	0.906528	3735	8040	0.055462	0.137131
0.056	89.60783954	0.909123	3780	7668	0.056131	0.130786
0.057	87.94511545	0.912299	3859	7189	0.057304	0.122616
0.058	86.51992337	0.91387	3928	6922	0.058328	0.118062
0.059	84.85719927	0.915672	4003	6620	0.059442	0.112911
0.06	83.43200719	0.916879	4056	6415	0.060229	0.109415
0.061	82.00681511	0.916871	4131	6341	0.061343	0.108153
0.062	**80.34409102**	**0.917093**	**4204**	**6240**	**0.062427**	**0.10643**
0.063	79.39396297	0.917077	4259	6187	0.063243	0.105526
0.064	77.73123888	0.916927	4340	6125	0.064446	0.104469
0.065	76.78111082	0.916641	4399	6102	0.065322	0.104076
0.066	75.35591874	0.916395	4473	6059	0.066421	0.103343
0.067	74.40579069	0.916339	4531	6008	0.067282	0.102473

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.068	73.21813062	0.916292	4600	5945	0.068307	0.101399
0.069	72.26800257	0.915974	4688	5897	0.069614	0.10058
0.07	71.31787452	0.915736	4750	5865	0.070534	0.100034
0.071	70.36774646	0.915291	4835	5836	0.071797	0.099539
0.072	69.65515042	0.915029	4888	5816	0.072584	0.099198
0.073	68.70502237	0.914728	4955	5787	0.073579	0.098704
0.074	67.99242633	0.91449	5019	5753	0.074529	0.098124
0.075	67.04229828	0.914124	5102	5716	0.075761	0.097493
0.076	66.09217023	0.913894	5165	5682	0.076697	0.096913
0.077	65.14204217	0.913704	5222	5649	0.077543	0.09635
0.078	64.42944613	0.913458	5289	5613	0.078538	0.095736
0.079	63.71685009	0.913466	5366	5535	0.079682	0.094406
0.08	63.00425405	0.913188	5429	5507	0.080617	0.093928
0.081	62.29165801	0.912862	5492	5485	0.081553	0.093553
0.082	61.81659399	0.912521	5554	5466	0.082473	0.093229
0.083	61.10399795	0.911918	5660	5436	0.084047	0.092717
0.084	60.86646593	0.911727	5689	5431	0.084478	0.092632
0.085	60.39140191	0.911124	5775	5421	0.085755	0.092461
0.086	59.91633788	0.910266	5900	5404	0.087611	0.092171
0.087	59.67880587	0.909862	5959	5396	0.088487	0.092035
0.088	59.20374184	0.909044	6075	5383	0.09021	0.091813
0.089	59.20374184	0.909044	6075	5383	0.09021	0.091813
0.09	58.72867781	0.908496	6159	5368	0.091457	0.091557
0.091	58.4911458	0.908234	6198	5362	0.092036	0.091455
0.092	58.01608177	0.90779	6266	5350	0.093046	0.09125
0.093	57.54101775	0.907425	6325	5337	0.093922	0.091028
0.094	57.06595372	0.906964	6403	5317	0.09508	0.090687
0.095	56.59088969	0.906416	6489	5300	0.096357	0.090397
0.096	56.11582567	0.905948	6567	5281	0.097516	0.090073
0.097	55.87829365	0.905742	6606	5268	0.098095	0.089852
0.098	55.40322963	0.905202	6688	5254	0.099312	0.089613
0.099	55.16569761	0.905123	6711	5241	0.099654	0.089391
0.1	54.45310157	0.904741	6784	5216	0.100738	0.088965
0.101	53.97803755	0.904495	6835	5196	0.101495	0.088624
0.102	53.02790949	0.904138	6928	5148	0.102876	0.087805
0.103	52.55284547	0.90394	6974	5127	0.103559	0.087447
0.104	51.84024943	0.903583	7064	5082	0.104896	0.086679
0.105	51.3651854	0.903424	7114	5052	0.105638	0.086167
0.106	50.89012138	0.90313	7197	5006	0.106871	0.085383
0.107	50.41505735	0.9029	7275	4957	0.108029	0.084547
0.108	49.93999332	0.902614	7352	4916	0.109172	0.083848
0.109	49.4649293	0.902304	7416	4891	0.110123	0.083421
0.11	48.98986527	0.902058	7481	4857	0.111088	0.082842
0.111	48.51480124	0.901749	7560	4817	0.112261	0.082159
0.112	48.03973722	0.901439	7632	4784	0.11333	0.081596
0.113	47.56467319	0.901328	7689	4741	0.114177	0.080863
0.114	47.08960916	0.901392	7746	4676	0.115023	0.079754
0.115	46.61454514	0.901249	7817	4623	0.116077	0.07885
0.116	46.13948111	0.900939	7895	4584	0.117236	0.078185
0.117	45.66441708	0.900439	7994	4548	0.118706	0.077571
0.118	45.42688507	0.900272	8037	4526	0.119344	0.077196
0.119	44.95182104	0.90005	8125	4466	0.120651	0.076173
0.12	44.71428903	0.899828	8178	4441	0.121438	0.075746
0.121	44.239225	0.899915	8246	4362	0.122448	0.074399
0.122	43.76416098	0.899931	8322	4284	0.123576	0.073068
0.123	43.52662896	0.899828	8365	4254	0.124215	0.072557
0.124	43.05156494	0.899725	8442	4190	0.125358	0.071465
0.125	42.81403292	0.899764	8464	4163	0.125685	0.071005
0.126	42.3389689	0.899859	8538	4077	0.126784	0.069538
0.127	41.86390487	0.899828	8621	3998	0.128016	0.06819
0.128	41.38884084	0.899518	8719	3939	0.129472	0.067184
0.129	41.15130883	0.899256	8774	3917	0.130288	0.066809

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.13	40.6762448	0.898851	8882	3860	0.131892	0.065837
0.131	40.43871279	0.898756	8931	3823	0.13262	0.065206
0.132	40.20118078	0.898613	8989	3783	0.133481	0.064523
0.133	39.72611675	0.898081	9096	3743	0.13507	0.063841
0.134	39.72611675	0.898081	9096	3743	0.13507	0.063841
0.135	39.25105272	0.897391	9202	3724	0.136644	0.063517
0.136	39.01352071	0.89701	9257	3717	0.13746	0.063398
0.137	38.7759887	0.896645	9317	3703	0.138351	0.063159
0.138	38.30092467	0.895882	9435	3681	0.140104	0.062784
0.139	38.30092467	0.895882	9435	3681	0.140104	0.062784
0.14	37.82586064	0.89485	9586	3660	0.142346	0.062425
0.141	37.82586064	0.89485	9586	3660	0.142346	0.062425
0.142	37.35079662	0.893946	9721	3639	0.144351	0.062067
0.143	37.35079662	0.893946	9721	3639	0.144351	0.062067
0.144	36.87573259	0.892914	9867	3623	0.146519	0.061794
0.145	36.87573259	0.892914	9867	3623	0.146519	0.061794
0.146	36.40066856	0.891778	10,042	3591	0.149117	0.061249
0.147	36.40066856	0.891778	10,042	3591	0.149117	0.061249
0.148	NaN	0.534583	0	58,630	0	1
0.149	35.92560454	0.890532	10,241	3549	0.152072	0.060532
0.15	35.92560454	0.890532	10,241	3549	0.152072	0.060532
0.151	NaN	0.534583	0	58,630	0	1
0.152	35.45054051	0.889794	10,386	3497	0.154225	0.059645
0.153	35.45054051	0.889794	10,386	3497	0.154225	0.059645
0.154	34.97547649	0.889222	10,504	3451	0.155978	0.058861
0.155	34.73794447	0.888889	10,559	3438	0.156794	0.058639
0.156	34.50041246	0.888643	10,616	3412	0.157641	0.058195
0.157	34.02534843	0.888008	10,741	3367	0.159497	0.057428
0.158	34.02534843	0.888008	10,741	3367	0.159497	0.057428
0.159	33.55028441	0.887468	10,853	3323	0.16116	0.056677
0.16	33.55028441	0.887468	10,853	3323	0.16116	0.056677
0.161	33.07522038	0.886468	11,023	3279	0.163684	0.055927
0.162	33.07522038	0.886468	11,023	3279	0.163684	0.055927
0.163	32.60015635	0.885618	11,178	3231	0.165986	0.055108
0.164	32.60015635	0.885618	11,178	3231	0.165986	0.055108
0.165	32.12509233	0.884975	11,331	3159	0.168258	0.05388
0.166	32.12509233	0.884975	11,331	3159	0.168258	0.05388
0.167	31.6500283	0.884396	11,474	3089	0.170381	0.052686
0.168	31.6500283	0.884396	11,474	3089	0.170381	0.052686
0.169	31.6500283	0.884396	11,474	3089	0.170381	0.052686
0.17	31.17496427	0.883594	11,620	3044	0.172549	0.051919
0.171	31.17496427	0.883594	11,620	3044	0.172549	0.051919
0.172	30.69990025	0.882761	11,794	2975	0.175133	0.050742
0.173	30.69990025	0.882761	11,794	2975	0.175133	0.050742
0.174	30.69990025	0.882761	11,794	2975	0.175133	0.050742
0.175	30.22483622	0.881848	11,967	2917	0.177702	0.049753
0.176	30.22483622	0.881848	11,967	2917	0.177702	0.049753
0.177	29.74977219	0.881046	12,129	2856	0.180108	0.048712
0.178	29.74977219	0.881046	12,129	2856	0.180108	0.048712
0.179	29.74977219	0.881046	12,129	2856	0.180108	0.048712
0.18	29.27470817	0.880077	12,332	2775	0.183122	0.047331
0.181	29.27470817	0.880077	12,332	2775	0.183122	0.047331
0.182	28.79964414	0.878887	12,540	2717	0.186211	0.046341
0.183	28.79964414	0.878887	12,540	2717	0.186211	0.046341
0.184	28.79964414	0.878887	12,540	2717	0.186211	0.046341
0.185	28.32458011	0.876632	12,884	2657	0.191319	0.045318
0.186	28.32458011	0.876632	12,884	2657	0.191319	0.045318
0.187	28.32458011	0.876632	12,884	2657	0.191319	0.045318
0.188	27.84951609	0.874457	13,227	2588	0.196412	0.044141
0.189	27.84951609	0.874457	13,227	2588	0.196412	0.044141
0.19	27.84951609	0.874457	13,227	2588	0.196412	0.044141
0.191	27.84951609	0.874457	13,227	2588	0.196412	0.044141
0.192	27.84951609	0.874457	13,227	2588	0.196412	0.044141
0.193	27.37445206	0.873306	13,464	2496	0.199932	0.042572
0.194	27.37445206	0.873306	13,464	2496	0.199932	0.042572
0.195	27.37445206	0.873306	13,464	2496	0.199932	0.042572
0.196	27.37445206	0.873306	13,464	2496	0.199932	0.042572
0.197	27.37445206	0.873306	13,464	2496	0.199932	0.042572
0.198	26.89938803	0.872639	13,649	2395	0.202679	0.040849
0.199	26.89938803	0.872639	13,649	2395	0.202679	0.040849
0.2	26.89938803	0.872639	13,649	2395	0.202679	0.040849

Appendix 4. KDE control limit calculation for NSL-KDD dataset

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.001	822.8187	0.321869956	18	118,886	0.000321	0.996187
0.002	265.142	0.332814345	99	116,886	0.001768	0.979429
0.003	228.572	0.341625746	186	115,254	0.003321	0.965754
0.004	199.1843	0.342121922	213	115,140	0.003804	0.964798
0.005	159.0019	0.342469816	248	115,044	0.004429	0.963994
0.006	123.3775	0.344026782	337	114,682	0.006018	0.960961
0.007	103.1907	0.345583748	395	114,351	0.007054	0.958187
0.008	94.18099	0.3472947	401	114,045	0.007161	0.955623
0.009	79.51898	0.350631056	516	113,345	0.009214	0.949757
0.01	74.36604	0.353705066	594	112,728	0.010607	0.944587
0.011	68.39499	0.359396832	599	111,725	0.010696	0.936183
0.012	62.08395	0.367255804	653	110,293	0.011661	0.924184
0.013	56.49539	0.37321562	724	109,177	0.012929	0.914832
0.014	51.91617	0.385945101	801	106,868	0.014304	0.895484
0.015	48.12318	0.399792405	839	104,402	0.014982	0.874821
0.016	43.6927	0.407856691	897	102,930	0.016018	0.862486
0.017	40.98342	0.414820265	933	101,673	0.016661	0.851954
0.018	38.10414	0.421584227	1012	100,408	0.018071	0.841354
0.019	37.2648	0.42444722	1048	99,870	0.018714	0.836846
0.02	36.38295	0.428439441	1087	99,131	0.019411	0.830653
0.021	36.22358	0.429779686	1189	98,794	0.021232	0.827829
0.022	36.11734	0.431171261	1261	98,478	0.022518	0.825182
0.023	35.57548	0.433720579	1329	97,963	0.023732	0.820866
0.024	34.65114	0.439452267	1351	96,936	0.024125	0.812261
0.025	33.20619	0.448143902	1405	95,358	0.025089	0.799038
0.026	31.83561	0.457063665	1446	93,753	0.025821	0.785589
0.027	30.42253	0.464791463	1514	92,330	0.027036	0.773665
0.028	29.65756	0.468903451	1557	91,566	0.027804	0.767264
0.029	28.9882	0.472228401	1620	90,920	0.028929	0.76185
0.03	28.63759	0.474937408	1671	90,394	0.029839	0.757443
0.031	28.2976	0.476471561	1760	90,036	0.031429	0.754443
0.032	27.98949	0.478587438	1802	89,623	0.032179	0.750982
0.033	27.54325	0.481324961	1864	89,081	0.033286	0.746441
0.034	26.9164	0.487575638	1927	87,922	0.034411	0.736729
0.035	26.34267	0.49437382	1953	86,704	0.034875	0.726523
0.036	25.67332	0.505586258	2014	84,677	0.035964	0.709538
0.037	24.62148	0.517545811	2096	82,498	0.037429	0.69128
0.038	24.05837	0.524731808	2121	81,213	0.037875	0.680512
0.039	22.88966	0.542742428	2190	77,986	0.039107	0.653472
0.04	22.4328	0.547892393	2211	77,062	0.039482	0.645729
0.041	21.77408	0.553601268	2312	75,960	0.041286	0.636495
0.042	21.49783	0.555494722	2344	75,596	0.041857	0.633445
0.043	21.07285	0.558791156	2421	74,941	0.043232	0.627957
0.044	20.80723	0.560696015	2455	74,573	0.043839	0.624873
0.045	20.53099	0.56278908	2519	74,142	0.044982	0.621262
0.046	20.28663	0.565064645	2581	73,681	0.046089	0.617399
0.047	20.10601	0.566963802	2629	73,300	0.046946	0.614206
0.048	19.87226	0.569610074	2724	72,741	0.048643	0.609522
0.049	19.64915	0.571412277	2759	72,390	0.049268	0.606581
0.05	19.36228	0.574817071	2811	71,741	0.050196	0.601143
0.051	19.08604	0.579122966	2848	70,949	0.050857	0.594506
0.052	18.7673	0.58750663	2922	69,405	0.052179	0.581569
0.053	18.51231	0.594213561	2968	68,183	0.053	0.571329
0.054	18.22545	0.600327362	3029	67,050	0.054089	0.561835
0.055	17.98108	0.605260606	3082	66,132	0.055036	0.554143
0.056	17.76859	0.611345892	3131	65,016	0.055911	0.544792
0.057	17.54547	0.617790477	3200	63,817	0.057143	0.534745
0.058	17.39672	0.620408233	3244	63,314	0.057929	0.53053
0.059	17.21611	0.624525924	3315	62,521	0.059196	0.523885
0.06	17.05674	0.62832994	3369	61,800	0.060161	0.517844

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.061	16.89737	0.633497014	3418	60,845	0.061036	0.509842
0.062	16.72737	0.64132747	3485	59,405	0.062232	0.497775
0.063	16.57863	0.648661751	3537	58,067	0.063161	0.486564
0.064	16.44051	0.652602643	3586	57,327	0.064036	0.480363
0.065	16.31301	0.656110094	3649	56,649	0.065161	0.474682
0.066	16.21739	0.659206917	3708	56,047	0.066214	0.469637
0.067	16.13239	0.661921627	3759	55,520	0.067125	0.465222
0.068	16.03677	0.665012747	3829	54,908	0.068375	0.460093
0.069	15.95177	0.668126679	3881	54,310	0.069304	0.455082
0.07	15.85615	0.671645536	3922	53,652	0.070036	0.449569
0.071	15.77115	0.675227129	3991	52,955	0.071268	0.443728
0.072	15.67553	0.679225053	4054	52,191	0.072393	0.437327
0.073	15.61178	0.680930302	4082	51,864	0.072893	0.434587
0.074	15.51616	0.684203923	4172	51,200	0.0745	0.429023
0.075	15.45242	0.686120189	4208	50,828	0.075143	0.425906
0.076	15.36742	0.688960369	4274	50,264	0.076321	0.42118
0.077	15.3143	0.690540147	4326	49,935	0.07725	0.418423
0.078	15.23992	0.693232045	4392	49,397	0.078429	0.413915
0.079	15.1868	0.694532368	4440	49,121	0.079286	0.411602
0.08	15.13368	0.696956217	4502	48,634	0.080393	0.407521
0.081	15.08055	0.700264057	4572	47,984	0.081643	0.402075
0.082	15.02743	0.70228298	4619	47,583	0.082482	0.398715
0.083	14.98493	0.703486349	4664	47,327	0.083286	0.396569
0.084	14.94243	0.704301903	4726	47,122	0.084393	0.394852
0.085	14.88931	0.706919659	4796	46,593	0.085643	0.390419
0.086	14.85744	0.708054591	4839	46,351	0.086411	0.388391
0.087	14.80431	0.710392892	4906	45,874	0.087607	0.384394
0.088	14.77244	0.711875716	4946	45,574	0.088321	0.38188
0.089	14.72994	0.713501121	5018	45,217	0.089607	0.378889
0.09	14.68744	0.715508637	5068	44,815	0.0905	0.375521
0.091	14.64494	0.71790397	5112	44,351	0.091286	0.371633
0.092	14.60244	0.718440068	5208	44,161	0.093	0.37004
0.093	14.55995	0.719660547	5248	43,907	0.093714	0.367912
0.094	14.52807	0.720692821	5278	43,696	0.09425	0.366144
0.095	14.4962	0.72216424	5335	43,381	0.095268	0.363505
0.096	14.46432	0.722557759	5436	43,211	0.097071	0.36208
0.097	14.43245	0.723812457	5485	42,942	0.097946	0.359826
0.098	14.4112	0.724667933	5518	42,759	0.098536	0.358293
0.099	14.37933	0.725249656	5605	42,570	0.100089	0.356709
0.1	14.35808	0.726002475	5618	42,425	0.100321	0.355494
0.101	14.33683	0.726607011	5737	42,200	0.102446	0.353609
0.102	14.30495	0.727639286	5761	41,995	0.102875	0.351891
0.103	14.27308	0.729612583	5811	41,599	0.103768	0.348573
0.104	14.24121	0.731009861	5852	41,313	0.1045	0.346176
0.105	14.20933	0.733296833	5929	40,835	0.105875	0.342171
0.106	14.18808	0.734180825	5972	40,637	0.106643	0.340512
0.107	14.15621	0.737157881	6016	40,071	0.107429	0.335769
0.108	14.13496	0.739376415	6104	39,594	0.109	0.331772
0.109	14.11371	0.744406613	6173	38,643	0.110232	0.323803
0.11	14.06059	0.75374841	6211	36,967	0.110911	0.309759
0.111	14.03934	0.759999087	6222	35,860	0.111107	0.300483
0.112	13.98622	0.781363172	6323	32,013	0.112911	0.268248
0.113	13.55061	0.809987396	6373	26,944	0.113804	0.225773
0.114	13.13624	0.81873036	6386	25,398	0.114036	0.212819
0.115	12.48814	0.823555244	6421	24,517	0.114661	0.205437
0.116	11.05382	0.823937356	6517	24,354	0.116375	0.204071
0.117	10.7457	0.823988685	6521	24,341	0.116446	0.203962
0.118	9.693861	0.823749152	6584	24,320	0.117571	0.203786
0.119	9.608864	0.823167428	6693	24,313	0.119518	0.203727
0.12	9.57699	0.822819535	6758	24,309	0.120679	0.203694
0.121	9.545116	0.822414609	6829	24,309	0.121946	0.203694
0.122	9.523867	0.822112341	6882	24,309	0.122893	0.203694
0.123	9.491993	0.821758744	6947	24,306	0.124054	0.203668
0.124	9.470744	0.821530617	6989	24,304	0.124804	0.203652
0.125	9.43887	0.821188427	7052	24,301	0.125929	0.203627
0.126	9.417621	0.820908972	7103	24,299	0.126839	0.20361
0.127	9.385747	0.820595297	7159	24,298	0.127839	0.203601
0.128	9.364498	0.82025881	7220	24,296	0.128929	0.203585
0.129	9.343248	0.819922323	7279	24,296	0.129982	0.203585

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.13	9.321999	0.819602945	7338	24,293	0.131036	0.20356
0.131	9.30075	0.819255052	7403	24,289	0.132196	0.203526
0.132	9.279501	0.818833017	7479	24,287	0.133554	0.203509
0.133	9.268876	0.818673328	7509	24,285	0.134089	0.203493
0.134	9.247627	0.818194261	7594	24,284	0.135607	0.203484
0.135	9.237002	0.818045979	7621	24,283	0.136089	0.203476
0.136	9.215753	0.817629647	7695	24,282	0.137411	0.203467
0.137	9.205128	0.817390114	7737	24,282	0.138161	0.203467
0.138	9.183879	0.816905344	7823	24,281	0.139696	0.203459
0.139	9.173254	0.81659167	7879	24,280	0.140696	0.203451
0.14	9.16263	0.81635784	7921	24,279	0.141446	0.203442
0.141	9.14138	0.816044165	7982	24,273	0.142536	0.203392
0.142	9.130756	0.81589018	8012	24,270	0.143071	0.203367
0.143	9.109506	0.815593615	8067	24,267	0.144054	0.203342
0.144	9.088257	0.815285643	8123	24,265	0.145054	0.203325
0.145	9.067008	0.815108845	8160	24,259	0.145714	0.203275
0.146	8.992635	0.81481228	8220	24,251	0.146786	0.203208
0.147	8.822641	0.814766655	8255	24,224	0.147411	0.202981
0.148	8.5464	0.81482939	8299	24,169	0.148196	0.202521
0.149	8.217037	0.814835093	8341	24,126	0.148946	0.20216
0.15	7.728303	0.814789467	8419	24,056	0.150339	0.201574
0.151	7.409564	0.81517158	8441	23,967	0.150732	0.200828
0.152	7.027077	0.815553693	8506	23,835	0.151893	0.199722
0.153	6.814584	0.81612401	8562	23,679	0.152893	0.198415
0.154	6.655215	0.81634073	8633	23,570	0.154161	0.197501
0.155	6.548968	0.816489013	8697	23,480	0.155304	0.196747
0.156	6.432097	0.816985189	8756	23,334	0.156357	0.195524
0.157	6.389599	0.817053627	8774	23,304	0.156679	0.195272
0.158	6.315226	0.833136574	8954	20,304	0.159893	0.170134
0.159	6.304602	0.837071763	8974	19,594	0.16025	0.164185
0.16	6.283353	0.841069687	9008	18,859	0.160857	0.158026
0.161	6.251479	0.845615116	9053	18,017	0.161661	0.150971
0.162	6.219605	0.851586337	9107	16,916	0.162625	0.141745
0.163	6.187731	0.857295213	9158	15,864	0.163536	0.13293
0.164	6.145232	0.861891971	9226	14,990	0.16475	0.125606
0.165	6.113358	0.865821456	9269	14,258	0.165518	0.119473
0.166	6.081484	0.869089374	9330	13,624	0.166607	0.11416
0.167	6.060235	0.870902983	9498	13,138	0.169607	0.110088
0.168	6.060235	0.870902983	9498	13,138	0.169607	0.110088
0.169	6.038986	0.872956125	9698	12,578	0.173179	0.105395
0.17	6.038986	0.872956125	9698	12,578	0.173179	0.105395
0.171	6.038986	0.872956125	9698	12,578	0.173179	0.105395
0.172	6.038986	0.872956125	9698	12,578	0.173179	0.105395
0.173	6.017737	0.874986455	9776	12,144	0.174571	0.101759
0.174	6.007112	0.875762086	9841	11,943	0.175732	0.100075
0.175	5.996487	0.876800064	9889	11,713	0.176589	0.098147
0.176	5.975238	0.878208748	9985	11,370	0.178304	0.095273
0.177	5.975238	0.878208748	9985	11,370	0.178304	0.095273
0.178	5.953989	0.879959622	10,061	10,987	0.179661	0.092064
0.179	5.943364	0.880410172	10,102	10,867	0.180393	0.091058
0.18	5.932739	0.881174397	10,155	10,680	0.181339	0.089491
0.181	5.91149	0.88274277	10,279	10,281	0.183554	0.086148
0.182	5.91149	0.88274277	10,279	10,281	0.183554	0.086148
0.183	5.890241	0.884179969	10,364	9944	0.185071	0.083324
0.184	5.890241	0.884179969	10,364	9944	0.185071	0.083324
0.185	5.868992	0.885503105	10,437	9639	0.186375	0.080769
0.186	5.858367	0.886113345	10,477	9492	0.187089	0.079537
0.187	5.837118	0.887179838	10,641	9141	0.190018	0.076596
0.188	5.826493	0.887407965	10,778	8964	0.192464	0.075112
0.189	5.805244	0.888862274	10,922	8565	0.195036	0.071769
0.19	5.805244	0.888862274	10,922	8565	0.195036	0.071769
0.191	5.805244	0.888862274	10,922	8565	0.195036	0.071769
0.192	5.805244	0.888862274	10,922	8565	0.195036	0.071769
0.193	5.805244	0.888862274	10,922	8565	0.195036	0.071769
0.194	5.783995	0.889917361	11,098	8204	0.198179	0.068744
0.195	5.783995	0.889917361	11,098	8204	0.198179	0.068744
0.196	5.783995	0.889917361	11,098	8204	0.198179	0.068744
0.197	5.762745	0.891046589	11,260	7844	0.201071	0.065728
0.198	5.762745	0.891046589	11,260	7844	0.201071	0.065728

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.199	5.762745	0.891046589	11,260	7844	0.201071	0.065728
0.2	5.741496	0.892210036	11,403	7497	0.203625	0.06282
0.201	5.741496	0.892210036	11,403	7497	0.203625	0.06282
0.202	5.741496	0.892210036	11,403	7497	0.203625	0.06282
0.203	5.720247	0.892871604	11,639	7145	0.207839	0.05987
0.204	5.720247	0.892871604	11,639	7145	0.207839	0.05987
0.205	5.698997	0.894000833	11,762	6824	0.210036	0.057181
0.206	5.698997	0.894000833	11,762	6824	0.210036	0.057181
0.207	5.698997	0.894000833	11,762	6824	0.210036	0.057181
0.208	5.698997	0.894000833	11,762	6824	0.210036	0.057181
0.209	5.677748	0.894901934	11,886	6542	0.21225	0.054818
0.21	5.677748	0.894901934	11,886	6542	0.21225	0.054818
0.211	5.677748	0.894901934	11,886	6542	0.21225	0.054818
0.212	5.656499	0.895899989	12,093	6160	0.215946	0.051617
0.213	5.656499	0.895899989	12,093	6160	0.215946	0.051617
0.214	5.656499	0.895899989	12,093	6160	0.215946	0.051617
0.215	5.63525	0.897137578	12,260	5776	0.218929	0.048399
0.216	5.63525	0.897137578	12,260	5776	0.218929	0.048399
0.217	5.614	0.898101414	12,494	5373	0.223107	0.045022
0.218	5.614	0.898101414	12,494	5373	0.223107	0.045022
0.219	5.614	0.898101414	12,494	5373	0.223107	0.045022
0.22	5.592751	0.899219236	12,744	4927	0.227571	0.041285
0.221	5.592751	0.899219236	12,744	4927	0.227571	0.041285
0.222	5.592751	0.899219236	12,744	4927	0.227571	0.041285
0.223	5.592751	0.899219236	12,744	4927	0.227571	0.041285
0.224	5.592751	0.899219236	12,744	4927	0.227571	0.041285
0.225	5.571502	0.900456824	12,977	4477	0.231732	0.037514
0.226	5.571502	0.900456824	12,977	4477	0.231732	0.037514
0.227	5.571502	0.900456824	12,977	4477	0.231732	0.037514
0.228	5.571502	0.900456824	12,977	4477	0.231732	0.037514
0.229	5.550252	0.90179707	13,204	4015	0.235786	0.033643
0.23	5.550252	0.90179707	13,204	4015	0.235786	0.033643
0.231	5.550252	0.90179707	13,204	4015	0.235786	0.033643
0.232	5.550252	0.90179707	13,204	4015	0.235786	0.033643
0.233	5.529003	0.902880673	13,479	3550	0.240696	0.029747
0.234	5.529003	0.902880673	13,479	3550	0.240696	0.029747
0.235	5.529003	0.902880673	13,479	3550	0.240696	0.029747
0.236	5.529003	0.902880673	13,479	3550	0.240696	0.029747
0.237	5.507754	0.904192402	13,835	2964	0.247054	0.024836
0.238	5.507754	0.904192402	13,835	2964	0.247054	0.024836
0.239	5.507754	0.904192402	13,835	2964	0.247054	0.024836
0.24	5.507754	0.904192402	13,835	2964	0.247054	0.024836
0.241	5.507754	0.904192402	13,835	2964	0.247054	0.024836
0.242	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.243	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.244	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.245	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.246	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.247	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.248	5.486505	0.905201864	14,315	2307	0.255625	0.019331
0.249	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.25	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.251	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.252	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.253	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.254	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.255	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.256	5.465255	0.906513594	14,741	1651	0.263232	0.013834
0.257	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.258	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.259	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.26	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.261	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.262	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.263	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.264	5.444006	0.90792798	14,956	1188	0.267071	0.009955
0.265	5.422757	0.908578142	15,081	949	0.269304	0.007952
0.266	5.422757	0.908578142	15,081	949	0.269304	0.007952
0.267	5.422757	0.908578142	15,081	949	0.269304	0.007952
0.268	5.422757	0.908578142	15,081	949	0.269304	0.007952

(*continued on next page*)

Alpha	CL	Hit_Rate	False_Positive	False_Negative	FP_Rate	FN_Rate
0.269	5.401508	0.908937442	15,177	790	0.271018	0.00662
0.27	5.401508	0.908937442	15,177	790	0.271018	0.00662
0.271	5.380258	0.909319554	15,240	660	0.272143	0.00553
0.272	5.369634	0.909530572	15,265	598	0.272589	0.005011
0.273	5.348384	0.909821434	15,337	475	0.273875	0.00398
0.274	**5.305886**	**0.910089483**	**15,390**	**375**	**0.274821**	**0.003142**
0.275	5.210264	0.910049561	15,420	352	0.275357	0.00295
0.276	4.997771	0.909752996	15,472	352	0.276286	0.00295
0.277	4.817152	0.909587604	15,501	352	0.276804	0.00295
0.278	4.530287	0.909125647	15,584	350	0.278286	0.002933
0.279	4.339044	0.908903223	15,626	347	0.279036	0.002908
0.28	4.190299	0.90866369	15,668	347	0.279786	0.002908
0.281	4.052178	0.908167514	15,757	345	0.281375	0.002891
0.282	3.956557	0.907910871	15,802	345	0.282179	0.002891
0.283	3.860935	0.907580087	15,861	344	0.283232	0.002882
0.284	3.744064	0.907237896	15,923	342	0.284339	0.002866
0.285	3.637818	0.907032582	15,965	336	0.285089	0.002815
0.286	3.510322	0.906690392	16,027	334	0.286196	0.002799
0.287	3.329703	0.906388124	16,085	329	0.287232	0.002757
0.288	3.191583	0.906182809	16,122	328	0.287893	0.002748
0.289	3.010964	0.905891948	16,184	317	0.289	0.002656
0.29	2.883468	0.905544054	16,248	314	0.290143	0.002631
0.291	2.745348	0.905218973	16,308	311	0.291214	0.002606
0.292	2.649726	0.905042175	16,341	309	0.291804	0.002589
0.293	2.522231	0.904603031	16,419	308	0.293196	0.002581
0.294	2.447858	0.904249434	16,481	308	0.294304	0.002581
0.295	2.38411	0.903958572	16,532	308	0.295214	0.002581
0.296	2.320362	0.903553647	16,604	307	0.2965	0.002572
0.297	2.288488	0.90332552	16,645	306	0.297232	0.002564
0.298	2.235365	0.902920595	16,716	306	0.2985	0.002564
0.299	2.203491	0.902589811	16,774	306	0.299536	0.002564
0.3	2.171617	0.902333168	16,819	306	0.300339	0.002564

References

Abbasi, B., & Guillen, M. (2013). Bootstrap control charts in monitoring value at risk in insurance. *Expert Systems with Applications, 40*(15), 6125–6135. doi: 10.1016/j.eswa.2013.05.028.

Ahsan, M., Mashuri, M., & Khusna, H. (2018a). Hybrid James-Stein and successive difference covariance matrix estimators based Hotelling's T^2 chart for network anomaly detection using bootstrap. *Journal of Theoretical and Applied Information Technology, 96*(20), 6828–6841.

Ahsan, M., Mashuri, M., & Khusna, H. (2018b). Intrusion detection system using bootstrap resampling approach of T^2 control chart based on successive difference covariance matrix. *Journal of Theoretical and Applied Information Technology, 96*(8), 2128–2138.

Ahsan, M., Mashuri, M., Kuswanto, H., & Prastyo, D. D. (2018c). Intrusion detection system using multivariate control chart Hotelling's T^2 based on PCA. *International Journal on Advanced Science, Engineering and Information Technology, 8*(5), 1905–1911. doi: 10.18517/ijaseit.8.5.3421.

Ahsan, M., Mashuri, M., Kuswanto, H., Prastyo, D. D., & Khusna, H. (2018). Multivariate control chart based on PCA mix for variable and attribute quality characteristics. *Production & Manufacturing Research, 6*(1), 364–384. doi: 10.1080/21693277.2018.1517055.

Ahsan, M., Mashuri, M., Kuswanto, H., Prastyo, D. D., & Khusna, H. (2018c). T^2 control chart based on successive difference covariance matrix for intrusion detection system. In *Proceedings of the Journal of Physics: Conference Series: 1028* (p. 12220). IOP Publishing.

Akashdeep, Manzoor, I., & Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2017.07.005.

Alfaro, J. L., & Ortega, J. F. (2009). A comparison of robust alternatives to Hotelling's T2 control chart. *Journal of Applied Statistics, 36*(12), 1385–1396. doi: 10.1080/02664760902810813.

Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A. (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Systems with Applications, 67*, 296–303. doi: 10.1016/j.eswa.2016.09.041.

Anderson, M. J., & Thompson, A. A. (2004). Multivariate control charts for ecological and environmental monitoring. *Ecological Applications, 14*(6), 1921–1935. doi: 10.1890/03-5379.

Balajinath, B., & Raghavan, S. V. (2001). Intrusion detection through learning behavior model. *Computer Communications, 24*(12), 1202–1212. doi: 10.1016/S0140-3664(00)00364-9

Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. In *Procedia computer science: 89* (pp. 117–123). doi: 10.1016/j.procs.2016.06.016.

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

Burden, R. L., & Faires, J. D. (2011). *Numerical Analysis. Cengage Learning*. doi: 10.1017/CBO9781107415324.004.

Carrasco, R. S. M., & Sicilia, M.-A. (2018). Unsupervised intrusion detection through skip-gram models of network behavior. *Computers & Security, 78*, 187–197.

Catania, C. A., Bromberg, F., & Garino, C. G. (2012). An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection. *Expert Systems with Applications, 39*(2), 1822–1829. doi: 10.1016/j.eswa.2011.08.068.

Catania, C. A., & Garino, C. G. (2012). Automatic network intrusion detection: Current techniques and open issues. *Computers & Electrical Engineering, 38*(5), 1062–1072. doi: 10.1016/j.compeleceng.2012.05.013.

Chou, C.-Y., Chen, C.-H., & Chen, C.-H. (2006). Economic design of variable sampling intervals T^2 control charts using genetic algorithms. *Expert Systems with Applications, 30*(2), 233–242. doi: 10.1016/j.eswa.2005.07.010.

Chou, Y., Mason, R. L., & Young, J. C. (1999). Power comparisons for a Hotelling's T^2 statistic. *Communications in Statistics-Simulation and Computation, 28*(4), 1031–1050. doi: 10.1080/03610919908813591.

Chou, Y.-M., Mason, R., & Young, J. (2001). The control chart for individual observations from a multivariate non-normal distribution. *Communications in Statistics: Theory & Methods, 30*(8/9), 1937. doi: 10.1081/STA-100105706.

Derhab, A., & Bouras, A. (2015). Multivariate correlation analysis and geometric linear similarity for real-time intrusion detection systems. *Security and Communication Networks, 8*(7), 1193–1212. doi: 10.1002/sec.1074.

Devarakonda, N., Pamidi, S., Kumari, V. V., & Govardhan, A. (2012). Intrusion detection system using Bayesian network and hidden Markov model. *Procedia Technology, 4*, 506–514. doi: 10.1016/j.protcy.2012.05.081.

Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications, 41*(4, Part 2), 1937–1946. doi: 10.1016/j.eswa.2013.08.089.

Frisén, M. (2010). Principles for multivariate surveillance. In *Frontiers in statistical quality control 9* (pp. 133–144). Springer.

Frisén, M., Andersson, E., & Schiöler, L. (2010). Evaluation of multivariate surveillance. *Journal of Applied Statistics, 37*(12), 2089–2100. doi: 10.1080/02664760903222208.

George, J. P., Chen, Z., & Shaw, P. (2009). Fault detection of drinking water treatment process using PCA and Hotelling's T^2 chart. *World Academy of Science, Engineering and Technology, 50*, 970–975.

Gomez, J., & Dasgupta, D. (2002). Evolving fuzzy classifiers for intrusion detection. In *Proceedings of the IEEE workshop on information assurance* (pp. 1–5). (June 2001) doi:citeulike-article-id:9927895.

Guo, C., Zhou, Y., Ping, Y., Zhang, Z., Liu, G., & Yang, Y. (2014). A distance sum-based hybrid method for intrusion detection. *Applied Intelligence, 40*(1), 178–188. doi: 10.1007/s10489-013-0452-6.

Hamamoto, A. H., Carvalho, L. F., Sampaio, L. D. H., Abrão, T., & Proença, M. L. (2018). Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Systems with Applications, 92*, 390–402. doi: 10.1016/j.eswa.2017.09.013.

Hanslik, T., Boelle, P.-Y., & Flahault, A. (2001). The control chart: An epidemiological tool for public health monitoring. *Public health, 115*(4), 277–281.

Hoque, M. S., Mukit, M. A., & Bikas, M. A. N. (2012). An implementation of intrusion detection system using genetic algorithm. *International Journal of Network Security & Its Applications (IJNSA), 4*(2), 109–120.

Hotelling, H. (1974). Multivariate quality control. *Techniques of statistical analysis*. New York: McGraw-Hill.

Huang, C.-J., Tai, S.-H., & Lu, S.-L. (2014). Measuring the performance improvement of a double generally weighted moving average control chart. *Expert Systems with Applications, 41*(7), 3313–3322. doi: 10.1016/j.eswa.2013.11.023.

Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(1), 36–43.

Hubert, M., Rousseeuw, P. J., & Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics, 21*(3), 618–637.

Javitz, H. S., & Valdes, A. (1994). *The Nides statistical component: Description and justification* citeulike-article-id:7899275.

Karami, A. (2018). An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities. *Expert Systems with Applications, 108*, 36–60. doi: 10.1016/j.eswa.2018.04.038.

Karami, A., & Guerrero-Zapata, M. (2014). Mining and visualizing uncertain data objects and named data networking traffics by fuzzy self-organizing map. In *Proceedings of the CEUR Workshop: 1315* (pp. 156–163). Retrieved from. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919674294&partnerID=40&md5=83d4ec6676415195a72a31eaebbffdf2 .

Kaya, İ., & Kahraman, C. (2011). Process capability analyses based on fuzzy measurements and fuzzy control charts. *Expert Systems with Applications, 38*(4), 3172–3184. doi: 10.1016/j.eswa.2010.09.004.

Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications, 41*(4, Part 2), 1690–1700. doi: 10.1016/j.eswa.2013.08.066.

Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long short term memory recurrent neural network classifier for intrusion detection. In *Proceedings of the International Conference on Platform Technology and Service (PlatCon)* (pp. 1–5). IEEE.

Kosztyán, Z. T., & Katona, A. I. (2016). Risk-based multivariate control chart. *Expert Systems with Applications, 62*, 250–262. doi: 10.1016/j.eswa.2016.06.019.

Kuang, F., Xu, W., & Zhang, S. (2014a). A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing Journal, 18*, 178–184. doi: 10.1016/j.asoc.2014.01.028.

Kuang, F., Xu, W., & Zhang, S. (2014b). A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing Journal*. doi: 10.1016/j.asoc.2014.01.028.

Lee, W., & Stolfo, S. J. (2000). *A framework for constructing features and models for intrusion detection systems*: 3. ACM Transactions on Information and System Security. doi: 10.1145/382912.382914.

Lim, H. A., & Midi, H. (2016). Diagnostic robust generalized potential based on index set equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics, 31*(3), 859–877.

Lin, S.-N., Chou, C.-Y., Wang, S.-L., & Liu, H.-R. (2012). Economic design of autoregressive moving average control chart using genetic algorithms. *Expert Systems with Applications, 39*(2), 1793–1798. doi: 10.1016/j.eswa.2011.08.073.

Mason, R. L., & Young, J. C. (2002). *Multivariate statistical process control with industrial applications* Society for Industrial and Applied Mathematics. Retrieved from http://epubs.siam.org/doi/book/10.1137/1.9780898718461 .

Moltchanova, E. (2019). Real options economic control chart for binomial and normal processes. *Quality and Reliability Engineering International, 35*(1), 385–391.

Montgomery, D. (2009). *Introduction to statistical quality control*. New York: John Wiley & Sons Inc 10.1002/1521-3773(20010316)40:6<9823::AID−ANIE9823>3.3.CO;2-C.

Morrison, L. W. (2008). The use of control charts to interpret environmental monitoring data. *Natural Areas Journal, 28*(1), 66–74.

Moustafa, N., & Slay, J. (2016). The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective, 25*(1–3), 18–31. doi: 10.1080/19393555.2015.1125974.

Park, Y. (2005). *A statistical process control approach for network intrusion detection*. Georgia Insitute of Technology.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*(3), 1065–1076. doi: 10.1214/aoms/1177704472.

Phaladiganon, P., Kim, S. B., Chen, V. C. P., Baek, J.-G., & Park, S.-K. (2011). Bootstrap-based T^2 multivariate control charts. *Communications in Statistics-Simulation and Computation, 40*(5), 645–662. doi: 10.1080/03610918.2010.549989.

Phaladiganon, P., Kim, S. B., Chen, V. C. P., & Jiang, W. (2013). Principal component analysis-based control charts for multivariate nonnormal distributions. *Expert Systems with Applications, 40*(8), 3044–3054. doi: 10.1016/j.eswa.2012.12.020.

Qu, G., Hariri, S., & Yousif, M. (2005). Multivariate statistical analysis for network attacks detection. In *Proceedings of the 3rd ACS/IEEE international conference on computer systems and applications* (pp. 9–14). doi: 10.1109/AICCSA.2005.1387011.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*, 832–837. doi: 10.1214/aoms/1177728190.

Rousseeuw, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection*: 589. John Wiley & sons.

Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences, 41*(3), 212–223.

Salem, M., & Buehler, U. (2012). Mining techniques in network security to enhance intrusion detection systems. *International Journal of Network Security & Its Applications, 4*(6), 51–66. doi: 10.5121/ijnsa.2012.4604.

Salo, F., Nassif, A. B., & Essex, A. (2019). Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection. *Computer Networks, 148*, 164–175. doi: 10.1016/j.comnet.2018.11.010.

Schiöler, L., & Frisén, M. (2012). Multivariate outbreak detection. *Journal of Applied Statistics, 39*(2), 223–242. doi: 10.1080/02664763.2011.584522.

Sharmai, S. K., Pande, P., Tiwari, S. K., & Sisodiai, M. S. (2012). An improved network intrusion detection technique based on k-Means clustering via Naive Bayes classification, *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012), Nagapattinam, Tamil Nadu* (pp. 417–422).

Shmueli, G., & Burkom, H. (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics : A Journal of Statistics for the Physical, Chemical, and Engineering Sciences, 52*(1), 39–51. doi: 10.1198/TECH.2010.06134.

Shyu, M.-L., Sarinnapakorn, K., Kuruppu-Appuhamilage, I., Chen, S.-C., Chang, L., & Goldring, T. (2005). Handling nominal features in anomaly intrusion detection problems. In *Research issues in data engineering: Stream data mining and applications, 2005* (pp. 55–62). IEEE. *RIDE-SDMA 2005. 15th International Workshop on*.

Singh, R., Kumar, H., & Singla, R. K. (2015). An intrusion detection system using network traffic profiling and online sequential extreme learning machine. *Expert Systems with Applications, 42*(22), 8609–8624. doi: 10.1016/j.eswa.2015.07.015.

Sivasamy, A., & Sundan, B. (2015). A dynamic intrusion detection system based on multivariate Hotelling's T^2 statistics approach for network environments. *The Scientific World Journal, 2015*, 1–9. doi: 10.1155/2015/850153.

Stolfo, S. J. (1999). *KDD cup 1999 dataset*. UCI KDD Repository 0.

Sullivan, J. H., & Woodall, W. H. (1996). A comparison of multivariate control charts for individual observations. *Journal of Quality Technology, 28*(4), 398–408.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*. doi: 10.1109/CISDA.2009.5356528.

Tavallaee, M., Lu, W., Iqbal, S. A., & Ghorbani, A. (2008). A novel covariance matrix based approach for detecting network anomalies. *Sixth Annual Conference on Communication Networks and Services Research*.

Tool, B.-I. (2014). Bro-IDS tool. Retrieved from https://www.bro.org/

Tsang, C.-H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition, 40*(9), 2373–2391. doi: 10.1016/j.patcog.2006.12.009.

Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert Systems with Applications, 37*(9), 6225–6232. doi: 10.1016/j.eswa.2010.02.102.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques (The Morgan Kaufmann Series in Data Management Systems)*. San Francisco: Elsevier.

Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology, 38*(2), 89–104.

Yang, S.-F., Lin, J.-S., & Cheng, S. W. (2011). A new nonparametric Ewma Sign Control Chart. *Expert Systems with Applications, 38*(5), 6239–6243. doi: 10.1016/j.eswa.2010.11.044.

Ye, N., Emran, S. M., Chen, Q., & Vilbert, S. (2002). Multivariate statistical analysis of audit trails for host-based intrusion detection. *IEEE Transactions on Computers, 51*(7), 810–820. doi: 10.1109/TC.2002.1017701.

Ye, N., Li, X., Chen, Q., Emran, S. M., & Xu, M. (2001). Probabilistic techniques for intrusion detection based on computer audit data. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 31*(4), 266–274. doi: 10.1109/3468.935043.

Ye, N., Parmar, D., & Borror, C. M. (2006). A hybrid SPC method with the chi-square distance monitoring procedure for large-scale, complex process data. *Quality and Reliability Engineering International, 22*(4), 393–402. doi: 10.1002/qre.717.

Zhang, Z., Zhu, X., & Jin, J. (2007). *SVC-Based multivariate control charts for automatic anomaly detection in computer networks* (p. 56). IEEE. doi: 10.1109/CONIELECOMP.2007.99.

Zhu, X. (2006). *Anomaly detection through statistics-based machine learning for computer networks*. The University of Arizona.