Detecting Anomalous Network Traffic in IoT Networks

Dang Hai Hoang*, Ha Duong Nguyen**

*Posts and Telecommunication Institute of Technology, Hanoi, Vietnam

**Faculty of Information Technology, National University of Civil Engineering, Hanoi, Vietnam

hdhai.hn@gmail.com, nghaduong@gmai.com

***Abstract*—Network operators need effective tools to quickly detect anomalies in traffic data for identifying network attacks. In contrast to traditional Internet, detection of anomalous network traffic in IoT (Internet of Things) networks is becoming a challenge task due to limited network resources and performance. Comprehensive detection methods are no longer effective for IoT networks, calling for developing lightweight solutions. Principal Component Analysis (PCA) techniques can help to reduce computing complexity, thus, anomaly detection techniques based on PCA received a lot of attention in the past. However, PCA techniques could not be directly applied to IoT networks with constrained resources and limited performance. This paper investigates PCA techniques for detecting anomalous network traffic in IoT networks. We propose a novel detection scheme with two levels using PCA techniques. The first level is for quick detection with few principal components while the second level is for detailed detection with a number of principal components. We investigate the selection of parameters in a distance calculation formula using several experiments to show the feasibility of our proposed scheme.**

***Keyword*— IoT Network Traffic Anomaly, Anomaly Detection, Principal Component Analysis, Information Security, Network Security**

I. INTRODUCTION

THE world of interconnected things - the Internet of Things (IoT) opens a number of challenges regarding security. IoT can bring huge interested services nowadays, but there is still a lack of suitable security measures for mixed IoT environment [1]. IoT devices have constrained resources and limited performance and are attractive to the attackers [2,3]. Cyber attacks in IoT networks are becoming more difficult to detect. Traditional mechanisms are usually comprehensive and no longer effective for IoT networks, calling for developing new paradigm. Given the growing complexity of connected things and many constraints of IoT environment, there is a strong demand for developing new effective security solutions.

Network traffic anomaly detection (NTAD) is a promised approach for identifying network attacks since it can detect new attacks without pre-recorded signature. That's why NTAD received a lot of attention in recent years [4,5]. The definition of anomaly was formally given by Hawkins as "an observation which deviates so much from other observations as to arouse uncertainties that it was produced by an alternative mechanism" [4]. Network anomalies are unusual patterns in traffic data that do not conform to expected normal behaviour. Anomalies may be performance related (due to network failures, changes in link traffic, flash crowd, etc...) or security related (due to attacks such as denial of service attacks, network scans, etc…). NTAD is a very critical task of network operators. Effective tools are necessary for quick detection of exceptional non-conforming patterns in traffic data in order to identify abnormal traffic flows or the causes of anomalies for further handling.

Many anomaly detection techniques have been proposed in the past within diverse research areas and application domains, see e.g. [4-11]. Traditional techniques considered anomalies as outlier and typically proposed to use statistical properties of observed data to construct a normal profile based on normal traffic data. The current collected data will be compared to this profile for checking any deviation to detect anomalies [5-8, 11]. General speaking, the main principle of NTAD is to build the baseline (the normal region) using features of network traffic in normal condition and to compare online collected traffic data to this baseline to find out deviation (anomalies). However, this process is difficult in practice due to many factors such as: multivariate features of traffic data, correlation of various data features, complex dataset, required accuracy and detection speed, etc.

On the other hand, NTAD for IoT networks is facing other issues such as: heterogeneity, constrained resources and limited performance of IoT devices. In IoT networks, we can not use complex methods as in traditional Internet. Lightweight techniques with less complexity, low resource usage and lower computation requirement are necessary. The development of such techniques remains a challenging research task.

Among broad existing anomaly detection techniques [4-11] for the Internet, multivariate statistical approaches for anomaly detection received a lot of attention. Principal Component Analysis (PCA) is one of the best-known multivariate statistical analysis techniques for detecting anomalies in the context of constrained network resources like IoT networks. PCA is a dimension reduction technique, which transforms a set of correlated original variables into a set of few uncorrelated variables, called Principal Components (PC). These PCs are linear combinations of the original variables. The number of PCs is less than or equal to

Manuscript received on December 17th, 2017. This work is follow-up of the invited journal to an accepted paper of the 20th International Conference on Advance Communication Technology (ICACT2018). The work is supported by the ASEAN IVO Project "A Hybrid Security Framework for IoT Networks".

Dang Hai Hoang is with the Posts and Telecommunication Institute of Technology, Hanoi, Vietnam (Corresponding author to provide phone: +84-4- 3854-4451; fax: +84-4-3756-2036; e-mail: haihd@ptit.edu.vn).

Ha Duong Nguyen is with the National University of Civil Engineering, Hanoi, Vietnam (e-mail: nghaduong@gmail.com).

the number of original variables. Thus, PCA allows lower complexity. PCA is considered as a simple but effective method for NTAD [5, 9-19].

Using PCA for NTAD is an attractive approach presented in pioneer research works by Shyu et al. [12], Lakhina et al. [13], Brauchkoff et al. [14], Ringberg et al. [15] and Kwitt et al. [19]. A variety number of further works has been proposed based on PCA with several enhancements such as [9-11, 16-29]. However, several issues in applying PCA have still not been considered as described in [4, 5, 7, 10, 16] including sensitivity, effectiveness, dimension-depending computation complexity, feature selection. Two main issues for the goodness of PCA based anomaly detection are: how many PCs are to select and how to calculate the distance (the deviation measure).

The issue of selecting PCs has been investigated in several works including the selection of major PCs and minor PCs [11, 12, 18, 20, 23], the selection of PC subspace [14-18, 28-30]. However, heuristic is the common way of choosing PCs. It is not clear, how many PCs and which PCs are to select. On the other hand, the choice of distance formula is not clear in most of the research works. Distance is a quantitative metric for deviation of traffic pattern. Most of the proposed methods are using either T^2 distribution formula [4-6, 10-13], or Euclidean distance [7, 11, 23, 26], or Mahalanobis distance [7, 10-18, 20-22, 24-29]. It is not clear how the choice of the distance formula will affect the detection accuracy and the computation complexity. On the other hand, the high complexity of such formulas is not suitable for a quick online detection of traffic anomalies. Such problems will have impact on further development network traffic anomaly detection based on PCA.

This paper addresses the problems for detection of network traffic anomaly in the context of IoT networks with resource constraints. We investigate PCA techniques using a new general formula for deviation (distance) calculation. We show that distance formulas used in previous typical research works can be derived from our general formula. To our best knowledge, the proposed formula is the first one for interpreting the parameters in different distance formulas. Based on selecting PCs using this general formula, we propose a novel detection scheme with two levels using PCA techniques. The first level is for quick detection with few k principal components in order to reduce the complexity to $O(k)$ while retaining acceptable detection results in comparison to previous methods. The second level is for detailed detection with a number of principal components.

The rest of this paper is organized as follows. Section II presents the basic PCA techniques and related works. concept of PCA and anomaly detection. Section III describes related works. Section IV proposes a new distance formula and our PCA scheme. Section V presents the experiments. Section VI concludes the paper.

II. PCA Techniques and Related Works

A. The Basic Concept of PCA

Anomaly detection often requires a high dimension data including many features (attributes) collected from networks. Therefore, the analysis has high computation complexity, needs much time, and is not suitable for quick detection requirement. PCA is the most common technique to reduce high dimension of data [4-8]. It converts the original data into new set of axes called principal components (PC) by keeping the most essential features of the original data.

Let X be the original observed dataset with n rows and p columns $\{X_1, X_2, \ldots, X_p\}$. The dataset is a $n \times p$ matrix, each column represents an attribute (feature) of data. Each column is represented by a p-dimensional vector of p correlated variables. Let R be a pxp sample correlation matrix of $\{X_1, X_2, \ldots, X_p\}$. Let $(\lambda_1, \boldsymbol{e}_1)$, $(\lambda_2, \boldsymbol{e}_2)$, $(\lambda_3, \boldsymbol{e}_3)$,… $(\lambda_p, \boldsymbol{e}_p)$ be p eigenvalue and eigenvector pairs of the matrix R. PCA converts X into a new dataset Y using transformation matrix R as follows:

$$\boldsymbol{Y = RX} \quad (1)$$

We have the i^{th} ***PC*** as follows:

$$y_i = \boldsymbol{e}_i^T (\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad (2)$$

where $i = 1, 2, \ldots, p$

$\boldsymbol{e}_i = (e_{i1}, e_{i2}, \ldots e_{ip})^T$ is the i^{th} eigenvector.

$\boldsymbol{x}$ is the observation

$\bar{x} = (\sum_{i=1}^{n} x_i)/n$ is the sample mean of $\boldsymbol{x}$.

The eigenvalues (λ) are the roots of equation

$| R - \lambda I | = 0$.

Each eigenvalue has a corresponding non-zero eigenvector $\boldsymbol{e}$, which satisfies: $R\boldsymbol{e} = \lambda \boldsymbol{e}$

Let $\boldsymbol{z}=(z_1, z_2, \ldots z_p)'$ be the vector of standardized observations, i.e. $\boldsymbol{z} = \boldsymbol{x} - \bar{\boldsymbol{x}}$, we rewrite (2) as:

$$y_i = \boldsymbol{e}_i^T \boldsymbol{z} \quad (3)$$

PCA has the following important properties. The PCs are uncorrelated. The first PC has the highest variance; the second PC has the next highest variance, and so on. The total variance in all PCs combined is equal to the total variance of the original variables $X_1, X_2, \ldots, X_p$. The PCs are sorted in descending order of the eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p > 0$. The quotient $\lambda_i / \sum_{j=1}^{p} \lambda_j$ describes the contribution of the i^{th} PC on the variance of the data. The dataset X is transformed to PCA based on eigenvectors with the target that the produced PCs have the highest possible variances.

B. Anomaly Detection Using PCA

The common principle for anomaly detection using PCA is to calculate the statistical distance from each observed data to the normal dataset (i.e. to the centroid or statistical average of the dataset). The distance calculation is performed in the principal component space. Mathematically, distance is a quantitative degree of how far two data instances apart from each other. Observed data instances which are at far distance from the new axes represented by PCs are considered as abnormal behaviour. For the comparison, a threshold value is established. If the calculated distance of the observed data is larger than the threshold, this data instance is considered as an anomaly. The threshold value is usually determined by the statistical distribution function of the distance [4-6, 30].

The most popular distance formulas used in previous anomaly detection methods are the Euclidean distance, the Mahalanobis distance or the statistic T^2 distribution formula. We use the following notations in the formulas.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$ be the observed data, where p is the number of original attributes, and is the number of input variables.

Let $d(\boldsymbol{x}, \boldsymbol{\mu})$ is the distance between two points $\boldsymbol{x}$ and $\boldsymbol{\mu}$ in the PCA space.

Let d_N is the threshold, which is calculated using the normal data profile as presented in [5-8,11].

1) The Euclidean Distance

Intuitively, the most common distance function is the Euclidean distance. The Euclidean distance between **x** and **μ** is [7, 11]:

$$d(\boldsymbol{x}, \mu) = \sqrt{(\boldsymbol{x}-\mu)^T(\boldsymbol{x}-\mu)} \quad (4)$$

If $\boldsymbol{\mu}$ is the vector of mean values of each input variable, the Euclidean distance in PCA space is given by:

$$\Delta_{Euclid} = \sqrt{\sum_{i=1}^{p} y_i^2} \quad (5)$$

The data instance $\boldsymbol{x}$ can be determined as anomalous if $\Delta_{Euclid} \geq d_N$

In order to reduce the square root computation complexity, the square distance is also usually used:

$$d_E = \Delta_{Euclid}^2 = \sum_{i=1}^{p} y_i^2 \quad (6)$$

The Euclidean distance formula is simple, but does not take into account the variance of each variable.

2) The Mahalanobis Distance

The Mahalanobis distance between x and μ is computed as [7, 10-12, 14-18, 20-22, 25-29]:

$$d(\boldsymbol{x}, \mu) = \sqrt{(\boldsymbol{x}-\mu)^T \boldsymbol{S}^{-1}(\boldsymbol{x}-\mu)} \quad (7)$$

Where S is the covariance matrix, S^{-1} is the transposition of the sample covariance matrix between $\boldsymbol{x}$ and $\boldsymbol{\mu}$. The covariance matrix is used as weights to reduce the different variance between the variables. Thus, the covariance matrix represents the relationship between variables more efficient.

In PCA space, the Mahalanobis distance is computed as:

$$\Delta_{Ma} = \sqrt{\sum_{i=1}^{p} \frac{y_i^2}{\lambda_i}} \quad (8)$$

The data instance $\boldsymbol{x}$ can be determined as anomalous if $\Delta_{Ma} \geq d_N$

The square distance is usually used as

$$d_m = \Delta_{Ma}^2 = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} \quad (9)$$

The Mahalanobis distance formula takes into account not only the average value but also the variance and covariance of the variables. Instead of simply computing the distance from the mean value, it weights each variable by its standard deviation and covariance [25].

3) Hotelling's T2 Statistic

Another way to calculate the distance is using statistical distribution, typically the Hotelling's T^2 statistic [10-13,30]. This statistic gives the deviation (distance) from each vector **x** to the centroid **μ** of the statistical distribution of the data.

$$T^2 = (\boldsymbol{x}-\bar{\boldsymbol{x}})^T \boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}) \quad (10)$$

Where S^{-1} is the transposition of the sample covariance matrix between $\boldsymbol{x}$ and the sample mean of $\boldsymbol{x}$.

In PCA space, the statistic T^2 distribution is computed as:

$$T^2 = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} \quad (11)$$

In comparison to (9), the statistic T^2 is similar to the Mahalanobis distance.

T^2 is distributed as $\frac{(n-1)p}{n-p} F_{p,n-p}$, where $F_{p,n-p}$ denotes a random variable with an *F*-distribution with *p* and *n-p* degrees of freedom. A large value of T^2 indicates a deviation of the observation $\boldsymbol{x}$ from the centroid of the normal dataset and the F-statistic can be used to test for an anomaly [12, 30].

C. Related Works on NTAD using PCA

The research works on NTAD using PCA were initiated by Shyu et al. [12] and Lakhina et al. [13]. In [12], the authors proposed a principal component classifier (PCC) consisting of two functions of PC scores, one for major components (Major PCs) and one for minor components (Minor PCs). The major PCs are used to detect extreme observations with large values on some original variables. The minor PCs are used to detect observations that do not conform to the normal correlation structure. The Mahalanobis distance is used with two thresholds, one for major PCs and another for minor PCs. The detection rate is depending on the quality and the accuracy of PCC. Thus, some other authors [20, 21] proposed to improve PCC with multivariate trimming for robustness, threshold calculation.

Lakhina et al. [12] proposed to divide the network traffic data into normal subspace consisting of typical behaviour pattern and anomalous subspace accounting for uncharacteristic circumstances by mean of PCA. The Euclidean distance is employed to check the dissimilarity between data instances. Anomaly detection is using the first PCs and the mean value of the distance. However, there are several remaining issues to be solved such as the choice of PCs, the separation of normal and anomalous subspaces. Further improving approaches were indicated in [17, 26] with the concept of traffic matrix including a number of original/ destination flows (OD flows). The authors proposed to extract a list of few PCs, which contain the maximum variance of the original data. The authors in [16] proposed a control approach for pre-processing the network data to be analysed with PCA. The authors in [18] proposed to decompose the traffic variations into normal and anomalous components. A new method for identifying the anomalous flows inside the aggregated flows was introduced.

The works of Shyu et al [11] and Lakhina et al. [12] received a lot of attention in the research community. The authors in [14] presented the issue of temporal correlation in previous works of Shyu and Lakhina, and proposed a predictive filter for classical PCA in order to improve correlation effect. Ringberg et al. [15] indicated that PCA techniques are very sensitive for traffic anomaly detection, especially for the selected parameters. The authors showed that current methods for tuning PCA parameters are inadequate and presented several challenges of using PCA. However, they concluded that PCA is a promising statistical analysis technique that can be used effectively for detecting network anomalies.

In recent years, a variety number of research works has been developed based on the application of PCA for anomalous traffic detection. The authors in [9] presented the application of PCA subspace method for anomaly detection in backbone networks. PCA is used for investigate the explored data with a lower approximation. Zargar et al [22] discussed category-based selection of the features for PCA based anomaly detection. By classifying network traffic into six typical groups, the paper proposed to select most important features in order to reduce the amount of data to be

analyzed by PCA. Huang et al. [26] presented three major approaches for network traffic anomaly detection: PCA-based, sketch-based and signal-analysis-based. The authors tried to combine these approaches into a unified frame. Lee at al. [23] proposed an online over-sampling PCA algorithm for anomaly detection. Unlike other PCA based approaches, this method does not store the entire data matrix. Instead, it uses an online updating technique to update the principal direction without solving eigenvalue decomposition problems. The authors claimed that this method is favored for online applications. Camacho et al. [17] proposed a sketch-based algorithm in addition to traditional PCA for traffic anomaly detection. In this algorithm, each local monitor only maintains a series of sketches for each traffic flow in order to reduce the raw data. The anomaly detection is following the method proposed by Lakhina [12] with the novel concept of residual subspace. Horrou et al. [10] proposed an integration of PCA, Hotelling's T^2 and Q statistics to detect small or moderate anomalies in the process mean. The authors showed that T^2 statistic can result in false negatives (missed detection) and cannot detect anomalies that are orthogonal to the first PCs.

Anomalies can be considered as outliers. The authors in [6] provided a survey on outlier detection techniques and applications. The authors in [11] discussed three statistical techniques in intrusion detection: PCA-based, Chi-square distribution and Gaussian mixture distribution and discussed the comparative performance of them. The authors in [19] presented a multi-step outlier-based approach for anomaly detection in network-wide traffic. A subset of dataset is called a cluster consisting of similar data objects.

The choice of a reasonable number of PCs is important for alleviate the autocorrelation of the residual PCs. The authors [7, 10, 25-29] concluded that the goodness of the PCA model depends on a good choice of how many PCs are retained. The paper [10, 28] indicated that techniques such as Scree plot and cumulative percentage variance (CPV) can be used to determine the number of PCs for the PCA model. Nevertheless, a threshold value of cumulative variance is needed to determine the number of PCs to use. Bhuyan et al. [8] presented an overview of various facets of network anomaly detection. The paper provided a broad survey of the existing research on network anomaly detection methods in the context of network security.

Anomalies are usually detected by the help of a distance measure [19]. A good survey of distance measures used within network anomaly detection was given in [7]. The paper discussed the theoretical background in distance measures and various types of distance measures published in previous papers. The authors discussed two common distance measures including Euclidean distance [7, 11, 22, 24-26] and Mahalanobis distance [10-12, 14-21, 25-29], and several equivalent distance measures such as weighted Euclidean distance, Manhattan distance, distribution law distance T_2. The paper [7] also indicated the limitation on identifying and selecting distance measures for network anomaly detection. The paper [29] discussed the use of Mahalanobis distance for anomaly detection in time series in the context of an unsupervised learning algorithm. Fan et al. [27] proposed a modified PCA scheme which uses multiple similarity measurements to generate multiple subspaces. The similarity is measured by the Mahalanobis distance. Unlike other PCA based subspace methods that use only one measurement, MPCA uses multiple measurements. This scheme has the main disadvantage of computation complexity. It has proposed mainly for learning approaches, in particular for image processing, and it is not suitable for quick detection of anomalous network traffic. In [8], the appropriateness of proximity measures (distance measures) was discussed, but not in details. A summary of distance measures was given but there is no explanation for the choice of measures. Bayarjargal et al. [25] proposed a combination of entropy distribution and Mahalanobis distance for anomaly detection. First, entropy of selected attributes is computed for defining suspicious traffic area. After that, Mahalanobis distance is calculated to check anomalies.

D. Issues of Previous PCA Based Methods

PCA allows dimensionality reduction, thus, it is suitable to handle high dimensional data sets. It converts a multivariate high dimensional data into uncorrelated individual PCs. In principle, tests for anomaly can be applied on individual PCs. However, standard PCA based techniques are typically proportional quadratic in the number of variables regarding the distance calculation. The complexity is typically $O(kn^2)$ with k is the number of PCs and n is the number of original variables [4, 17, 23, 26]. Thus, many research works tried using a subset of PCs in order to reduce the computation complexity [6-11, 23, 26-29].

As presented in section II.A, the PCA transformation is performed in such a way that the first PCs account for the most of the variance in the original data, and the last PCs represent linear functions of the original variables with very small variance. The first PCs represent the large cumulative proportion of the total variance of the original data. Thus, these PCs tend to be strongly related to the anomalies on one or more original variables. Intuitively, it is reasonable to detect anomalies by looking at few first PCs [18]. If most of the variance of an n-dimensional data set is accounted by $k < n$ principal components, it is possible to select only first k PCs. The dimension of the data is then reduced to k [26]. However, it is still not clear how many PCs to be selected for an effective detection.

On the other hand, last PCs are sensitive to the observation that are inconsistent with the correlation structure of the data as indicated in several works such as [12-16]. Large values on few last PCs will reflect multivariate anomalies that are not detectable using the criterion based on large values of the original variables [10]. It is useful to check the last PCs for a significant variation of an observation. That is why many works suggested using few first PCs and few last PCs.

Typically, two thresholds are used: one for major PCs (q first PCs) and one for minor PCs (last r PCs). As indicated in [12, 17, 20, 21], a data instance $\boldsymbol{x}$ is anomalous if

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} > c_1 \text{ or } \sum_{i=p-r+1}^{p} \frac{y_i^2}{\lambda_i} > c_2 \quad (12)$$

and is normal if

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} \leq c_1 \text{ and } \sum_{i=p-r+1}^{p} \frac{y_i^2}{\lambda_i} \leq c_2 \quad (13)$$

Where c_1 and c_2 are the thresholds, which are calculated with the help of the chi-square distribution. These thresholds

are very sensitive to the detection results.

Several works followed the subspace approach and proposed the decomposition of the data into normal and anomalous subspaces using projection of data on the first PCs and the last PCs respectively [13, 24, 26]. The subspace methods are also called residual analysis methods. One issue with this approach is that PCA based anomaly detection techniques are sensitive to the number of PCs in each of the two subspaces [5, 7, 16].

The choice of distance formula is one issue of most of the PCA based anomaly detection methods. As presented in the above section, most of the proposed methods are using either T^2 formula [4-6, 30], or Euclidean distance [7, 17, 24, 26] or Mahalanobis distance [10, 11, 16, 20, 21, 25]. The computation of the distance is quadratic proportional to of the variables. With high dimensional dataset, the distance calculation will increase the computational cost, and a quick (online) detection is not possible.

The application of PCA techniques for NTAD in IoT networks has been just investigated in few recent research works, e.g. [31-37]. A general security framework was recently proposed in [31], which is based on three parts of IoT systems, namely physical part, network part and application part. However, there is still not clear how to apply detection methods for network traffic anomalies. The paper just described the general scheme, which is mainly for authentication. The paper [32] provided basic three layered IoT architecture and discussed several security issues of IoT that exist in such architectures. There is a good survey on previous research works on IoT security. The authors in [33] provided a survey of several security architectures for embedded IoT systems. These architectures are mainly based on hardware-specific environment and are typically designed for specific tasks. Pajouh et al. [34] investigated the demand on detecting intrusion and malicious activities within IoT networks and proposed a model for intrusion detection based on two-layer dimension reduction and two-tier classification module. The model is using PCA and linear discriminate analysis. Sharma et al. [35] proposed a framework for coordinated processing between edge and cloud computing / processing. The paper discussed the issue of hue amount of data generated from heterogeneous wireless IoT devices. Ferrando et al. [36] investigated the issue of streaming analytical techniques for detecting events in traffic feature distribution, which can allow the classification of abnormal behaviour within an IoT network. Recently, Zhao et al. [37] proposed a model for intrusion detection based on dimension reduction algorithm using PCA and a classifier for IoT networks.

In fact, in contrast to the traditional networks, IoT opens a completely new dimension to security, where attack threats move from manipulating information to controlling actuation (in other words, moving from the digital to the physical world). In our best knowledge, there is still very few works investigating PCA based methods for NTAD in IoT networks. Moreover, there are still several issues for network traffic anomaly detection using PCA, which have been not adequately investigated for IoT environment regarding the constrained resource and limited performance. Three challenges are of most interested by applying PCA for IoT networks including: 1) selection of PCs for an effective detection regarding the IoT network constraints, 2) distance measure for lower complexity and 3) quick detection of network traffic anomalies. These issues are the topics of this paper.

III. A Novel Network Anomaly Detection Scheme using PCA for IoT Networks

A. Overview of the Proposed Concept

Figure 1 describes the network model we use for our network anomaly detection scheme. The concept of fog was introduced. A fog is a network architecture for processing data and events from IoT devices closer to the sources of data in contrast to the central data network (known as "Cloud infrastructure". In other words, fog architecture extends the "cloud" (the cyber world) into the physical world of things (the world of IoT devices). Within the network model, the network traffic anomaly detection (NTAD) can be implemented into two levels. NTAD Level 1 (NTAD-L1) is for the fog architecture (IoT network segment closed to devices) with constrained resources and performance. This detection level requires lower complexity. NTAD Level 2 is for the cloud infrastructure with powerful resources and computing performance. NTAD Level 1 uses PCA technique with few PCs for quick detection, while NTAD Level 2 can deploy more PCs for detailed detection. One NTAD Level 2 module can serve several NTAD Level 1 modules (e.g. NTAD-L1-1, NTAD-L1-2, NTAD-L1-3).

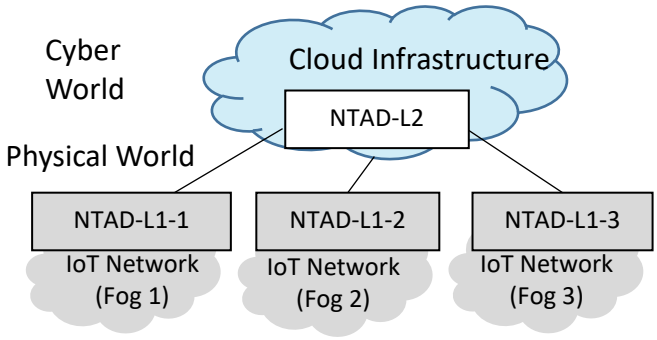


Fig. 1. The Network Model.

Figure 2 depicts the general flow of our anomaly detection scheme using two levels. Network traffic is collected from the IoT network (the fog) by a flow capture (like in other works such as [24]).

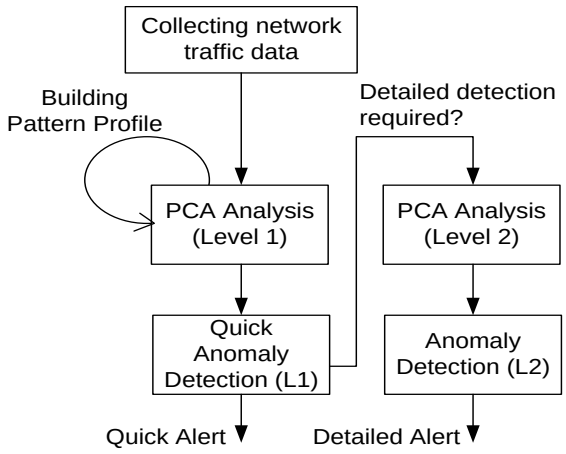


Fig. 2. The Detection Flow.

The captured data is standardized and transformed into PCA. Based on traffic data in normal condition (no attacks), the model builds a pattern profile (the baseline) for comparison later. The Level 1 detection module uses few PCs (e.g. 3 PCs) for quick anomaly detection. The number of suitable PCs will be given in the next section. If the network operator wants to have more detailed detection alerts, he can proceed the Level 2 detection module, which uses more PCs (e.g. major PCs and minor PCs as presented in [12, 17-22, 24-26]) for providing detailed detection.

In the following section, we present our detection scheme in details.

B. A Novel Distance Formula for the Detection Scheme

As presented in Section II, popular NTAD methods use Euclidean distance or Mahalanobis distance or statistic T^2 distribution with high computing complexity. Thus, such methods are not suitable for online anomaly detection, especially for IoT networks (the fogs).

For quick online anomaly detection, we suggest to use only few PCs and a simple distance formula as possible. Intuitively, the price to be paid is the accuracy. However, the detection can be implemented into two steps as described above. At the first step, network operators just want to know whether there are abnormal traffic flows or not. A quick detection with acceptable accuracy is desirable. At the next step, network operators can apply a detailed analysis for higher accuracy.

For implementing our proposed detection scheme with two levels, we want to develop a general distance calculation formula, which can use few PCs as well as a number of PCs. Our idea is to use the well-known Minkowski formula [38] for developing the new formula. The derivation of the general formula is as follows.

As indicated in [38], the Minkowski distance between two observed data $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$ is as follows:

$$\Delta_{Minkowski} = \left(\sum_{i=1}^{p} |x_i - \mu_i|^c \right)^{1/c}, c \geq 1; \qquad (14)$$

$$d = \Delta^c_{Minkowski} = \sum_{i=1}^{p} |x_i - \mu_i|^c, c \geq 1; \qquad (15)$$

By transforming into PCA domain with y_i principal components, we derive the Minkowski distance from each observation to the centre (the origin of PC axes) as follows:

$$d = \sum_{i=1}^{p} |y_i|^c, c \geq 1; \qquad (16)$$

Since each variable can have different variance, we can follow the approach of weighted Euclidean distance as illustrated in [16, 30] to give a weight to each PC in the equation (16) as follows:

$$d = \sum_{i=1}^{p} w_i |y_i|^c, \quad c \geq 1; \qquad (17)$$

Where w_i is the weight for the i^{th} PC ($w_i \neq 0$).

The formula (17) is our new proposed distance formula, which has the following properties.

Lemma 1: *The formula (17) is a general form of the Euclidean distance.*

Proof: Since c is arbitrary variable, we can choose c=2. All PCs can be considered equally, i.e. no variance of each input variable, the weight w_i can be set to 1. Thus, from (17) we can have:

$$d = \sum_{i=1}^{p} y_i^2 = y_i^2 + y_1^2 + \cdots + y_p^2 \qquad (18)$$

The formula (18) is the square Euclidean distance. In case we have consider variance of the input variables, i.e. PCs have different impacts on the distance calculation, we get:

$$d = \sum_{i=1}^{p} w_i y_i^2 \qquad (19)$$

The formula (19) is the weighted Euclidean distance. □

Lemma 2: *The formula (17) is a general form of the Mahalanobis distance.*

Proof: Since c is arbitrary variable, we can choose c=2. The impact of PCs is inverse proportional to their eigenvalues [38]. PCs with lower eigenvalues have large contribution on distance deviation, and PCs with large eigenvalues have less contribution on distance deviation, respectively. Thus, from (17) we can have:

$$d = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \cdots + \frac{y_p^2}{\lambda_p} \qquad (20)$$

The formula (20) is the Mahalanobis distance. □

Lemma 3: *The formula (17) is a general form of the Mahattan distance.*

Proof: Since c is arbitrary variable, we can choose c=1. Thus, from (17) we can have:

$$d = \sum_{i=1}^{p} w_i |y_i| \qquad (21)$$

All PCs can be considered equally, i.e. no variance of each input variable, the weight w_i can be set to 1. Thus, from (17) we have:

$$d = \sum_{i=1}^{p} |y_i| = |y_1| + |y_2| + \cdots + |y_p| \qquad (22)$$

The formula (22) is the Mahalanobis distance. □

Conclusion:

We can use the general formula (17) for distance calculation in our proposed NTAD scheme based on PCA. For quick detection (Level 1 detection), we can use formula (21) or formula (22) with few PCs. In our previous works [28, 39], we have indicated that 3 or 4 PCs are enough for quick detection with an acceptable detection true rate of 92%. In this case, the computation complexity is O(k) with k=3 or 4.

For detailed detection (Level 2 detection), we can use more PCs using the formula (21) or (22), or even the formula (19) or (20), depending on the performance of the detection module NTAD-L2 in the cloud infrastructure.

C. Description of the Novel NTAD Scheme

In this section, we describe how to use the distance formula for our novel NTAD scheme.

At the initial stage, we have to collect the normal network

traffic (in case of no attacks) in order to build the pattern profile. The captured data is standardized and transformed into PCA dataset. By transforming into PCA space, we classify PCs into *m* major PCs and *p-m* minor PCs similar to the approach in [12, 17-22, 24- 27]. As suggested in [12, 26], the major PCs include the PCs representing 50% of the variance of total eigenvalues. The minor PCs have eigenvalues smaller than or equal to 0.2.

From the pattern profile, we can use the empirical cumulative distribution function (ECDF) to determine the thresholds d_{1N} for major PCs and d_{2N} for minor PCs, respectively. The ECDF function is defined as follows:

$$F_{ECDF} = P(d \le d_N) \quad (23)$$

That means the probability for a distance *d* smaller or equal to the threshold d_N. With the assumption of statistical distribution and an estimation rate α for false estimation, we can look at the table for F_{ECDF} - distribution function and seek for the values of d_N corresponding to (1-α) of the ECDF. For instance, if α = *5%,* the thresholds d_N can be determined corresponding to *95%* of ECDF.

At the detection stage, online traffic data will be gathered into datasets. Similar to the previous stage, the collected dataset will be standardized and transformed into PCA domain using two subspaces, one for *m* major PCs and *p-m* minor PCs (see e.g. [12].

For quick detection (NTAD-L1), we apply the formula (22) with 3 or 4 PCs as follows:

$$d_{L1} = \sum_{i=1}^{k} |y_i| = |y_1| + |y_2| + \cdots + |y_k| \quad (24)$$

Where *k=3* or *k=4*.

The quick alert for traffic anomaly is determined if:

$$d_{L1} = \sum_{i=1}^{k} |y_i| > d_{NL1} \quad (25)$$

The threshold value d_{NL1} is determined using ECDF function as described above.

For detailed detection (NTAD-L2), we apply the formula (17) for *q* first PCs (*q*<*p*) including *m* major PCs and *q-r* minor PCs with *1*<*m*<*q-r*<*p*. An observed dataset is considered as anomaly if:

$$d_1 = \sum_{i=1}^{m} w_i |y_i|^c > d_{N1} \quad \text{OR} \quad (26)$$

$$d_2 = \sum_{i=r}^{q} w_i |y_i|^c > d_{N2} \quad (27)$$

Where d_1 and d_2 are the distances according to major PCs and minor PCs respectively, d_{1N} and d_{2N} are the corresponding thresholds of d_1 and d_2, OR and AND are the logic operations.

An observed dataset is considered as normal if:

$$d_1 = \sum_{i=1}^{m} w_i |y_i|^c \le d_{N1} \quad \text{AND} \quad (28)$$

$$d_2 = \sum_{i=r}^{q} w_i |y_i|^c \le d_{N2} \quad (29)$$

In case of *c=1* and w_i =1, the formulas for d_1 and d_2 will be:

$$d_1 = \sum_{i=1}^{m} |y_i| = |y_1| + |y_2| + \cdots + |y_m| \quad (30)$$

$$d_2 = \sum_{i=r}^{q} |y_i| = |y_r| + |y_{r+1}| + \cdots + |y_q| \quad (31)$$

The determination of the thresholds d_{1N} and d_{2N} is using the empirical cumulative distribution function (ECDF) as described above.

IV. Experiments

A. Dataset and Metrics

The dataset Kyoto Honeypot [40] is a real dataset collected from networks using a Honeypot at the Kyoto University. This dataset was used in many works for anomaly detection. The dataset includes most of the anomalies originated by the Internet. Thus, this dataset reflects objectively the anomalous events of collected network traffic.

We use in our experiments 14 essential features of the Kyoto Honeypot dataset including the most important features (attributes) of the network layer and transport layer traffic data. These features (attributes) are selected similar to the previous works [9, 12-21, 23-27] for comparison purpose.

For the experiments, we select three random datasets of the Kyoto Honeypot datasets collected from networks as indicated in the Table I.

TABLE I
THREE RANDOM DATASETS FOR EXPERIMENTS

Dataset	Number of flows	Number of anomalous flows	Number of normal flows
Dataset 1	125643	84476	41167
Dataset 2	114148	38790	75358
Dataset 3	120857	57337	63520

Following parameters are used for checking the accuracy:

$$TPR = TP / (TP + FN) \quad (32)$$

$$FPR = FP / (TN + FP) \quad (33)$$

Where *TPR* is True Positive Rate, *FPR* is False Positive Rate, *TP* is the number of true positive alerts, *FN* is the number of false negative alerts, *FP* is the number of false positive alerts, and *TN* is the number of true negative alerts. The total number of anomalous events is *TP+FN*; the total number of normal events is *TN+FP*.

B. Experiment Results

To build the sample dataset (the profile), we use traffic data from 5000 connections (similar to previous works in [12,13] for the consequent comparison). We transform this dataset into PCA space; calculate the centroid of the data using the distance formula. We determine the threshold using the empirical cumulative distribution function as described in the previous section. The thresholds are 95% of the ECDF function.

In the detection phase, the scheme uses the online collected datasets from Kyoto Honeypot datasets [40] and carries out the following steps: data standardization, PCA transformation, distance calculation, threshold comparison.

Table II shows the experiment result with dataset 1 using the threshold 95% for different combination of the

parameters k (number of PCs), *c* and w_i (three cases). As indicated in the grey row, our scheme provides acceptable TPR (94.3%) and FRP (4.8%) while keeping a low complexity with *k*=3 (only 3 PCs), c=1 and w_i =1.

TABLE II
RESULTS WITH DATASET 1, THRESHOLD 95%

Case	*k*	*c*	w_i	TPR (%)	FPR (%)
1	3	2	1	92.4	4.7
1	5	2	1	91.8	5.2
1	14	2	1	94.4	5.3
2	3	2	$1/\lambda_i$	94.6	4.9
2	5	2	$1/\lambda_i$	91.9	5.6
2	14	2	$1/\lambda_i$	93.8	5.0
3	3	1	1	94.3	4.8
3	5	1	1	91.3	5.2
3	14	1	1	92.1	5.4

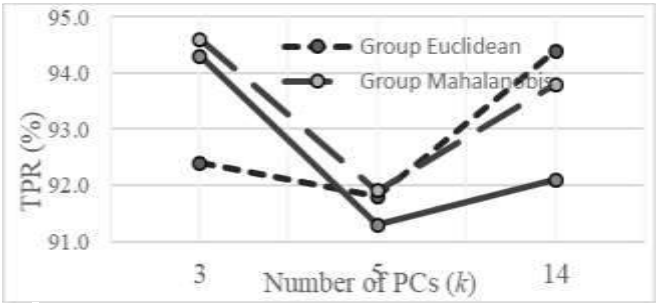
Table III shows the number of anomalous flows that are true detected in dataset 1 for case 1, 2 and 3, respectively. In fact, the total number of anomalous flows in dataset 1 is 84476.

TABLE III
NUMBER OF TRUE DETECTED ANOMALY FLOWS (TP), DATASET 1

Case	Parameters	TP			
		k=3	*k=4*	*k=5*	*k=14*
1	*c=2, w_i=1*	78106	77430	77611	80082
2	*c=2, w_i= $1/\lambda_i$*	79951	78626	77633	79302
3	*c=1, w_i=1*	79783	77678	78465	78431

Figure 3 shows the number of true detected anomalous flows in dataset 1 for three cases regarding different *k* (number of PCs)..

Fig. 3. TP for Anomalous Flows in Dataset 1.



.

Fig. 4. TPR Comparison using Dataset 1.

Figure 4 shows the comparison of TPR between our scheme and typical previous works according to the number of PCs used. As presented in the Section II, the Group Euclidean represents the TPR using Euclidean distance formula, which is used in previous works such as [7, 11, 17, 24-26]. The Group Mahalanobis represents the TPR using Mahalanobis distance formula, which is used in previous works such as [10, 12, 14-16, 18, 20-22].

As showed in figure 4, our scheme provides acceptable TPR (94.3%) in comparison to other works by *k*=3. The Group Mahalanobis gives better TPR (94.6%), but the complexity of our scheme is O(3), while the complexity of Group Mahalanobis is $O(w3^2)$. Thus, our scheme can be used for quick detection using *k=3*.

Table IV and Table V show the experiment results with dataset 2 and dataset 3, respectively. Threshold is *95%*. The results are showed for different combination of the parameters k (number of PCs), *c* and w_i. The grey row in the table indicates that our scheme (*k=3, c=1,* w_i =1) provides acceptable TPR and FPR with lower complexity in comparison to other previous works.

TABLE IV
NUMBER OF TRUE DETECTED ANOMALY FLOWS (TP), DATASET 2

Case	*k*	*c*	w_i	*TPR (%)*	*FPR (%)*
1	3	2	1	95.9	5.4
1	5	2	1	99.9	5.1
1	14	2	1	93.1	4.6
2	3	2	$1/\lambda_i$	93.8	4.7
2	5	2	$1/\lambda_i$	96.9	4.8
2	14	2	$1/\lambda_i$	99.0	4.9
3	3	1	1	95.9	5.0
3	5	1	1	92.3	4.7
3	14	1	1	93.8	5.1

TABLE V
NUMBER OF TRUE DETECTED ANOMALY FLOWS (TP), DATASET 3

Case	*k*	*c*	w_i	TPR (%)	FPR (%)
1	3	2	1	99.4	5.6
1	5	2	1	99.9	5.3
1	14	2	1	99.9	5.4
2	3	2	$1/\lambda_i$	99.7	4.9
2	5	2	$1/\lambda_i$	99.8	5.0
2	14	2	$1/\lambda_i$	99.9	5.3
3	3	1	1	98.6	4.9
3	5	1	1	97.6	5.3
3	14	1	1	100	4.9

Table VI shows the number of normal traffic flows that are true detected in dataset 2 for case 1, case 2 and case 3, respectively. The total number of actual normal flows in dataset 2 is 75358.

TABLE VI
TRUE DETECTED ANOMALY FLOWS (TP) IN DIFFERENCE CASES

Case	Parameters	TN			
		k=3	*k=4*	*k=5*	*k=14*
1	*c=2, w_i=1*	71275	71385	71500	71886
2	*c=2, w_i= $1/\lambda_i$*	71820	71417	71668	71675
3	*c=1, w_i=1*	71586	71699	71767	71504

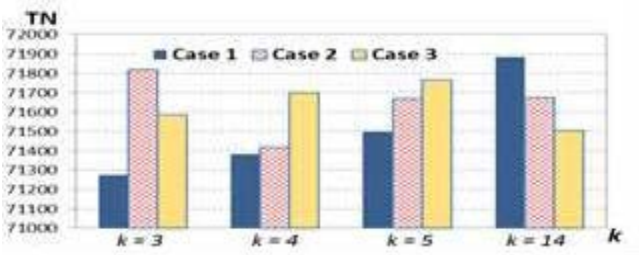


Fig. 5. TN for Normal Flows in Dataset 2.

Figure 5 shows the number of true detected normal flows in dataset 2 for three cases according to the number of PCs. The reason for TN results is that the last PCs are responsible for the variance of the anomalies.

The experiment results showed that our scheme can provide comparable accuracy with lower complexity (O(kn) instead of O(kn^2) in other schemes) in comparison to typical works using Euclidean distance formula (Group Euclidean) or using Mahalanobis distance formula (Group Mahalanobis). For quick network anomaly detection, we can use the general distance formula with few PCs instead of using all PCs or a large number of PCs as presented in previous works.

V. Conclusion

Network traffic anomaly detection is one of the challenging tasks of network operators, especially in IoT network environment due to many constraints such as limited resource and performance. The detection scheme should be simple (lower complexity) and quick as possible. Although many detection methods have been proposed in the past, methods based on Principal Component Analysis (PCA) received a lot of attention. PCA based methods promise suitable solutions for IoT networks, but the remaining issues need to be investigated including selection of principal components, distance formula, complexity reduction. There is still very few works investigating PCA based methods for network anomaly detection in IoT networks

The paper investigated problems of PCA based methods for IoT networks including: 1) selection of PCs for an effective detection regarding the IoT network constraints, 2) distance measure for lower complexity and 3) quick detection of network traffic anomalies. Based on selecting few PCs using our developed general distance formula, we proposed a novel detection scheme with two levels. The first level is for quick detection with few k principal components in order to reduce the complexity to O(k) while keeping an acceptable detection rate in comparison to previous methods. The second level is for detailed detection with a number of principal components. The paper presented various experiments using three random datasets from Kyoto Honeypot to show the feasibility of our proposed scheme.

Acknowledgment

This work is supported by the ASEAN IVO project “A Hybrid Security Framework for IoT Networks”.

The authors thank the National Institute of Information and Communications (NICT, Japan) and NES (NEC, Japan) for the supports.

References

[1] Q.Jing, et al., "Security of the IoT: Perspectives and Challenges". Wireless Networks, Vol 20, Issue 8, Nov. 2014, pp.2481-2501

[2] Y.M. Pa pa, S. Suzuki, K. Yoshioka, T. Matsumoto, T. Kasama, C. Rossow. "IoTPOT: A Novel Hoeypot for Revealing Current IoT. Threats". Journal of Information Processing, Vol 24, No.3, May 2016, pp.522-533.

[3] A.V. Vijayalakshmi, L. Arockiam, "A Study on Security Issues and Challenges in IoT". Engineering Sciences & Management Research. Vol 3, No 11, Nov. 2016, pp. 34-43

[4] V. Chandola, A. Banerjee, V. Kumar, “Anomaly detection: a survey”, ACM Computing Surveys, 2009, 41, (3), pp. 1-58.

[5] M. Ahmed, A. Mahmood, J. Hu, “A survey of network anomaly detection techniques”, Network & Computer Applications, 2016, 60 (2016), pp. 19-31.

[6] R. Bansal, et al., “Outlier detection: applications and techniques in data mining”, Cloud Syst./Big Data Eng. 2016, pp. 373-377.

[7] D. Weller-Fahy, B. Borgehtti, A. Sodemann, “A survey of distance and similarity measures used within network intrusion anomaly detection”, IEEE Communication Survey & Tutorials, 2015, 17, (1), pp. 70-91.

[8] M. Bhuyan, D. Bhattacharyya, J. Kalita, “Network anomaly detection: methods, systems and tools”, IEEE Com. 2014, 16, 303-336.

[9] M. Molina, et al., “Operational experiences with anomaly detection in backbone networks”, Computer & Security, 2012, 31, (3), pp. 273-285.

[10] F. Harrou, et al.,“Amalgamation of anomaly detection indices for enhanced process monitoring”, Loss Prevention in the Process Industries, 2016, 40, (3), pp. 365-377.

[11] H. Om, T. Hazra, “Statistical techniques in anomaly intrusion detection system”, Advanced in Eng. & Technology, 2012, 5, (1), pp.387-398.

[12] M. Shyu, S. Chen, K. Sarinnapakorn, L. Chang, “A novel anomaly detection scheme based on principle component classifier”, Proc. IEEE foundation and New Directions of Data Mining Workshop, Florida, USA, Nov. 2003, pp. 172-179.

[13] A. Lakhina, M. Crowella, C. Diot, “Diagnosing network-wide traffic anomalies”, Proc. ACM SIGCOMM '04, Portland, Oregon, USA, Aug. 2004, pp. 219-230.

[14] D. Brauckhoff, K. Salamatian, M. May, “Applying PCA for traffic anomaly detection: problems and solutions”, Proc. INFOCOM’09, Rio de Janeiro, Brazil, Apr. 2009, pp. 2866-2870.

[15] H. Ringberg, A. Soule, J. Rexford, C. Diot, “Sensitivity of PCA for traffic anomaly detection”, Proc. ACM SIGMETRICS ’07, San Diego, USA, Jun.2007, pp. 109–120.

[16] J. Camacho, A. Perez-Villegas, P. Garcia-Teodoro, G.Macia-Fernandez, “PCA-based multivariate statistical network monitoring for anomaly detection”, Computer & Security, 2016, 59 (2016), pp. 118-137.

[17] Y. Liu, L. Zhang, Y. Guan, “Sketch-based streaming PCA algorithm for network-wide traffic anomaly detection”, Proc. IEEE 30th Int. Conf. on Dist. Computing Systems, June 2010, Genoa, Italy, pp. 807-816.

[18] C. Callegari, L. Gazzarrini, S. Giordano, M. Pagano, T. Pepe, “A novel PCA-based network anomaly detection”, Prof. IEEE Int. Conf. on Communications, Kyoto, Japan, Jun. 2011, pp. 1-5.

[19] M. Bhuyan, D. Bhattacharyya, J. Kalita, “A multi-step outlier-based anomaly detection approach to network-wide traffic”, Information Sciences 2016, 348 (2016), pp. 243-271.

[20] U. Kwitt, Hofmann, “Robust methods for unsupervised PCA-based anomaly detection”, Proc. IEEE/IST Work-shop on Monitoring, Attack Detection and Mitigation, Tubingen, Germany, 2006, pp. 1-3.

[21] A. Das, S. Misra, S. Joshi, J. Zambreno, G. Memik, A. Choudhary, “An efficient FPGA implementation of principle component analysis based network intrusion detection system”, Proc. Design, Automation and Test in Europe, Munic, Germany, Mar. 2008, pp. 1160-1165.

[22] G. Zargar, T. Baghaie, “Category-based intrusion detection using PCA”, Journal of Information Security, 2012, (3), pp. 259-271.

[23] Y. Lee, Y. Yeh, Y. Wang, “Anomaly detection via online oversampling principal component analysis”, IEEE Trans. On Knowledge & Data Engineering, 2013, 25, (7), pp. 1460-1470.

[24] M. Elrawy, T. Abdelhamid, A. Mohamed, “IDS in telecommunication network using PCA”, International Journal of Computer Networks & Communications, 2013, 5, (4), pp.147-157.

[25] D. Bayarjargal, G. Cho, “Detecting an anomalous traffic attack area based on entropy distribution and Mahalanobis distance”, International Journal of Security and Its Applications, 2014, 8, (2), pp. 87-94.

[26] H. Huang, H. Al-Azzawi, H. Brani, “Network Traffic Anomaly Detection”, arXiv:1402.0856 preprint, 2014.

[27] Z. Fan, Y. Xu, W. Zuo, J. Yang, J. Tang, Z. Lai, D. Zhang, “Modified principal component analysis: an integration of multiple similarity subspace models”, IEEE Trans. On Neural Networks & Learning Systems, 2014, 25, (8), pp. 1538-1552.

[28] H.D. Nguyen, D.H. Hoang, “A model for network traffic anomaly detection”, ICACT Trans. on Advanced Communications Technology, 2015, 4, (4), pp.644-650.

[29] E.G. Nascimento, O. Tavares, A.F. Souza, “A cluster-based algorithm for anomaly detection in time series using Mahalanobis distance”, Proc. Int. Conf. Artificial Intelligence, July 2015, Nevada, USA, pp. 622-628.

[30] S. Franklin, M. Brodeur, “A practical application of a robust multivariate outlier detection method”, Proc. of Survey research methods section, American Statistical Association, 1997, pp. 186-191, http://www.amstat.org/ sections/srms/proceedings.

[31] M. Pasha, S.M. Waqas Shah, U. Pasha, "Security Framework for IoT Systems" International Journal of Computer Science and Information Security (IJCSIS), Vol 14, No 11, Nov. 2016, pp.99-104.
[32] A.V. Vijayalakshmi, L. Arockiam. "A Study on Security Issues and Challenges in IoT", Intl. Journal of Engineering Sciences & Management Research. Vol 3, No 11, Nov. 2016, pp. 34-43.
[33] N.C. Winget, Aa.R. Sadeghi, Y. Jin. "INVITED: Can IoT be Secured: Emerging Challenges in Connecting the Unconnected". DAC' 2016, June 05-09, 2016, Austin, TX, USA.
[34] H.H. Pajough, R. Javidan, R. Khayami, D. Ali, K.K. Raymond Choo, "A Two-layer Dimention Reduction and Two-tier Classification Model for Anomaly-based Intrusion Detection in IoT Backbone Networks", IEEE Trans. On Emerging Topics in Computing, 29 Nov. 2016.
[35] S.K. Sharma, X. Wang, "Live Data Analytics with Collaborative Edge and Cloud Processing in Wireless IoT Networks", IEEE Access, Vol 5, March 2017, pp. 4621-4635.
[36] R. Ferrando, P. Stacey, "Classification of Device Behaviour in Internet of Things Infrastructures", Proc. of International Conference on Internet of Things and Machine Learning, Oct. 2017.
[37] S. Zhao, W.Li, T. Zia, A.Y. Zomaya, "A Dimension Reduction Model and Classifier for Anomaly-Based Intrusion Detection in Internet of Things", IEEE 15th International Conference on Pervasive Intelligence and Computing (PICom 2017), Nov. 2017.
[38] J.P. Geer, "Some aspects of Minkowski distance", research report, University of Leiden (1995), pp. 1-31.
[39] H.D. Nguyen, D.H. Hoang. "Network traffic anomaly detection in the condition of noisy training dataset". Journal of Science & Technology on Information and Communications, Vol 1, CS01, 2016, pp.5-18.
[40] Traffic data from Kyoto University's Honeypots, http://www.takakuara.com/Kyoto_data/

Dang Hai Hoang, A/Prof. Dr. DSc., PhD (1999), DSc (2002) at TU Ilmenau, Germany. Current institution: Posts and Telecommunication Institute of Technology. Research interests: Communication network, IoT networks, information security, network security.

Ha Duong Nguyen, BSc (2001), MSc (2003) at TU Hanoi, PhD (2017) at PTIT, Vietnam. Current institution: Faculty of Information Technology. Research interests: Communication network, telecommunication networks, information security, network security.