Performance Evaluation of Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation

Yogesh Pawar, Manar Amayri, Nizar Bouguila

Performance Evaluation of Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation

Yogesh Pawar*, Manar Amayri*†, Nizar Bouguila*
*Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC., Canada,
† Grenoble Institute of Technology, G-SCOP Lab, France
Email: yogesh.pawar@mail.concordia.ca, manar.amayri@concordia.ca, nizar.bouguila@concordia.ca

***Abstract*—Computer network technology is developing quickly, and the advancement of internet techniques is growing faster. Furthermore, people and companies have became more aware of the importance of network security. To protect the network from different attacks, it is necessary to instantly detect intrusions with significantly low False-Positive Rates (FPRs). Many Anomaly-based Detection Systems (ADS) have been proposed in the past. The performance of these systems depends on many factors such as features selection or extraction, missing or inaccurate records in data, over-fitting, under-fitting, high bias, and variance in data. Thus, it is important to take all these factors into account. Recently, a novel Geometric Area Analysis (GAA) technique based on Trapezoidal Area Estimation (TAE) has been proposed based on the Beta Mixture Model (BMM). In this work, we evaluate GAA and TAE techniques using other flexible mixture models based on inverted Beta, generalized Dirichlet, and generalized inverted Dirichlet distributions. The evaluation of this work is performed on two datasets, namely the NSL-KDD and UNSW-NB15. The results have shown the efficiency of the proposed ADS demonstrated by obtaining high accuracy and low false-positive rates in all attack types.**

***Index Terms*—Geometric area analysis, anomaly detection system, trapezoidal area estimation, large-scale network, mixture models, clustering, feature selection.**

I. Introduction

Different ADS have been proposed in the past. Unfortunately, many of them fail in terms of detection rate and time trade-off especially in the case of large-scale networks. Over the past few years, various statistical machine learning models have been proposed to mitigate this limitation. The main goal of these models is to distinguish between normal and abnormal network events represented in terms of vectors of features. The task of distinguishing between an attack profile and a normal one is very challenging [1]–[3].

Many Intrusion Detection Systems (IDS), mainly Misuse-based Systems (MDS), are build using statistical models trained on normal network data and each deviation from what is normal is considered as an outlier and then considered to be an intrusion. A recent study on ADS has shown that statistical-based mixture models are very effective to detect malicious network behaviours with significantly low FPR and high Detection Rate (DR). This work is motivated by a successful approach recently proposed to build an effective ADS statistical decision engine based on Beta mixture model and TAE for recognizing abnormal behaviours in network systems with low FPR [4]. We evaluated different algorithms for each of the commonly known attack types on the widely used UNSW-NB15 dataset. The most accurate approach is determined according to the high accuracy, DR, low FPR, and short execution time. The proposed framework consists of three modules of dimensionality reduction, training different models with typical normal data vectors, and a statistical decision engine. More precisely, we first select the most relevant features in the dataset by combining the chosen attributes using feature selection techniques widely used in the literature [5]. Next, a statistical model is learned to perform classification. The algorithms with the highest accuracy, detection rate performance, and lowest error rate for each attack type are determined automatically in our proposed system. Finally, the decision-making method for identifying anomalies is designed by specifying a minimum and maximum threshold for each typical profile and considering any deviation from it as an attack as proposed in [4]. The contributions of this paper can be summarized as follows:

- We evaluate different mixture models within the anomaly detection framework proposed by [4].
- We propose a feature selection approach based on the highest vote obtained by different techniques in the literature.
- We evaluated the GAA-TAE technique by deploying different mixture models for the widely used UNSW-NB15 and NSL-KDD data sets.

The rest of this paper is organized as follows: In Section II, we present a background and a brief introduction to different mixture models used in this study. The framework with feature selection and data preprocessing is introduced in Section III. In Section IV, the performance of the TAE technique based on different mixture models is evaluated. Finally, Section V concludes the paper.

II. Background and Related Works

The main goal of the clustering algorithm is to find patterns and features consisting of the information of data that belongs to the same group. In the case of network data, it is very

978-1-7281-5628-6/20/$31.00 ©2020 IEEE

challenging to differentiate between normal and abnormal data vectors when both reflect the same pattern. To overcome this problem, different distance majors considered along with model-based clustering algorithms. The GAA technique, presented in [4] can be divided into following steps:

1) Normal profile creation: For D-dimensional positive vector $\vec{X}$ representing a network observation, the GAA technique is applied to calculate its area based on its TAE computed from the BMM parameters and distances of records. In this technique normal profile is constructed from legitimate network data. These legitimate data are first used to estimate BMM parameters and then used to calculate distances between the mean of normal records and each record. Finally for each data vector TAE is calculated using the BMM Probability Density Function (PDF) and distance between records [6]–[8] for training and testing dataset.

2) TAE estimation: The final PDF of normal record and test record is used to estimate TAE from Eq. 1

$$area(V) = \frac{b-a}{D}[f(x_1) + 2\sum_{i=1}^{D-1} f(x_i) + f(x_D)] \quad (1)$$

GAA technique is mainly based on trapezoidal rule which is one of the numerical integration families called Newton-Cotes formulas [9]. When this rule is used for multivariate data it is called a composite trapezoidal rule. The PDF of each vector is considered as the area under the curve as shown in Fig. 1. The final normal areas are sorted and divided into K_i intervals

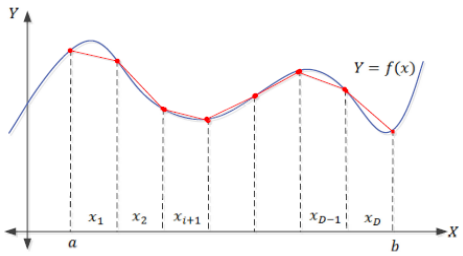


Fig. 1: Composite trapezoidal rule. [4]

where each K_i represents a minimum and maximum value ($\min_{K_i}$ and $\max_{K_i}$, respectively) and used in decision-making step.

3) Testing and Decision-making: In this step, PDF and area of test data is calculated by using same estimated parametrs from normal profile. Next, the test area is compared with normal area. If the area value for test data vector falls in between normal min and max range, it is considered a normal record, otherwise as an attack one.

A. Mixture Models

1) Beta Mixture Model: Mixture models have been widely used for data modeling. Although Gaussian mixture model, has proven to be efficient in several applications, it fails to fit the observations accurately especially when the data are clearly non-Gaussian due to its non-convex clustering properties [10]. BMM, on the other hand, has proven to be more efficient in handling many real-world applications involving one-dimensional data [11]. BMM can be more efficient on modeling the distribution of bounded data than the Gaussian mixture [12]. The PDF of a Beta distribution is given by:

$$Beta(x|v,w) = \frac{1}{beta(v,w)} x^{v-w}(1-x)^{w-1}, v, w > 0 \quad (2)$$

where $x \in [0,1]$ is the normalized feature, v and w indicate the shape parameters of the Beta distribution and $beta(v,w)$ is the Beta function given by:

$$\text{beta}(v,w) = \frac{\Gamma(v)\,\Gamma(w)}{\Gamma(v+w)} \quad (3)$$

where $\Gamma()$ is the Gamma function.
Let $\vec{X} = (x_1, \ldots, x_D)$ be a D-dimensional vector of independent normalized features supposed to follow Beta distributions described in following Eq 4.

$$p(\vec{X}|\vec{v},\vec{w}) = \prod_{d=1}^{D} Beta(x_d; v_d, w_d) \quad (4)$$

where $\vec{v} = (v_1, \ldots, v_D)$ and $\vec{w} = (w_1, \ldots, w_D)$. In [4], the authors have considered a finite mixture model based on the distribution in Eq.4, by normalizing semi-bounded positive features, given by:

$$p(\vec{X}|\Theta) = \sum_{j=1}^{M} p(\vec{X}|\vec{v}_j, \vec{w}_j) p_j \quad (5)$$

where $\Theta = \{p_j, \vec{v}_j, \vec{w}_j\}$ refers to the entire set of parameters to be estimated, p_j are positive mixing proportions, with $\sum_{j=1}^{M} p_j = 1$, $p(\vec{X}|\vec{v}_j, \vec{w}_j)$ is the joint density function for a D-dimensional positive vector given by Eq.4. These parameters can be learned using the maximum likelihood approach proposed in [13] or the Bayesian one proposed in [12]. By investigating this mixture model, we can notice a main shortcoming which is related to supposing that the features are independent which may not be the case. The goal of this paper is to consider other mixture models to handle this shortcoming.

2) Other Mixture Models: In order to avoid supposing that the features are independent, we consider the generalized Dirichlet mixture [14]. The generalized Dirichlet distribution with parameters $\vec{\alpha} = (\alpha_1, \ldots, \alpha_D)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_D)$ is given by:

$$p(\vec{X}|\vec{\alpha},\vec{\beta}) = \prod_{d=1}^{D} \frac{\Gamma(\alpha_d + \beta_d)}{\Gamma(\alpha_d)\Gamma(\beta_d)} x_d^{\alpha_d - 1}(1 - \sum_{l=1}^{d} x_l)^{\gamma_d} \quad (6)$$

for $\sum_{d=1}^{D} x_d < 1$ and $0 < x_d < 1$ for $d = 1 \ldots D$, where $\alpha_d > 0, \beta_d > 0, \gamma_d = \beta_d - \alpha_{d+1} - \beta_{d+1}$ for $d = 1 \ldots D-1$ and $\gamma_D = \beta_D - 1$. As discussed extensively in a series of papers [14], [15] the consideration of the generalized Dirichlet allows the transformation of the data using a geometric transformation in such a way that the independence between the features becomes a fact and not an assumption. Each original vector $\vec{X}$ is geometrically transformed into a vector $\vec{Y} = (y_1, \ldots, y_D)$ as:

$y_d = x_d$ if d = 1 and $y_d = \frac{x_d}{(1-\sum_{l=1}^{D-1} x_l)}$ for $d = 2, 3, \ldots, D$. Hence, each feature y_d has a Beta distribution and the resulting vector $\vec{Y}$ follows the distribution in Eq.4 .

In the following we consider two other techniques. The first one is based on a mixture model based on Inverted Beta Distribution (IBMM) [16], [17] as a replacement to the Beta distribution considered in the original approach in [4]. The inverted Beta distribution is given by:

$$iBeta(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1+x)^{-(\alpha+\beta)} \quad (7)$$

where $x > 0$ and $\Gamma()$ is the Gamma function. The learning of the resulting mixture model could be based on the approaches proposed in [16]–[18].

However, it is clear that the previous technique supposes that the features are independent. A better alternative is the Generalized Inverted Dirichlet (GID) mixture as introduced [19]. The GID distribution is given by [19].

$$p(\vec{X}|\vec{\alpha},\vec{\beta}) = \prod_{d=1}^{D} \frac{\Gamma(\alpha_d+\beta_d)}{\Gamma(\alpha_d)\Gamma(\beta_d)} \frac{x_d^{\alpha_{d-1}}}{(1+\sum_{l=1}^{D} x_l)^{\gamma_d}} \quad (8)$$

where $\gamma_d = \beta_d + \alpha_d - \beta_{d+1}$ for $d = 1, \ldots, D$ with $\beta_{D+1} = 0$. As described in [19], we can factorize GID distribution as a product of inverted Beta distributions by using the following geometric transformation: $y_1 = x_1$ and $y_d = \frac{x_d}{(1+\sum_{d=1}^{D-1} x_d)}$ for $d = 2, 3, \ldots, D$. Thus, each feature y_d has an inverted beta distribution with parameters α_d and β_d as described in eq.7 . The learning of the parameters of a GID mixture model could be based on the approaches proposed in [19], [20]. The PDFs estimated from above mixture models will then be used to calculate trapezoid area by applying the composite trapezoidal rule and uniform-grid property (features of equal length) as described in [21] and using Eq. 1.

III. Data preprocessing and Framework

In this section, we describe the ADS framework to train the high dimensional vectors to create the normal profile using estimated parameters, distances between the means of the records, and normal area using the GAA-TAE method as described in [4]. This module can be divided into several sub-modules namely data preprocessing including dimensionality reduction, area estimation, and a decision engine to distinguish between normal and abnormal data instances. The block diagram of the system is shown in Fig.2.

A. Data preprocessing

Data preprocessing involves different steps such as data cleaning, instance selection, normalization, transformation, feature selection, and dimensionality reduction. Raw data collected from network traffic is often in an undependable format, incomplete, and/or deprived of certain trends or behaviors and likely to contain many missing values. Machine learning models are generally built using structured data with numeric values. Unstructured data such as text or categorical data, images, video, and audio need to be converted into numeric representation using appropriate methods. After removing missing, inconsistent, and duplicated values, data are aggregated into appropriate numerical format and structure. At this point, each feature in the dataset needs to be scaled into specific range using normalization, in this study we used min-max normalization in the equation 9:

$$X_i^{normalized} = (X_i - min(X))/(max(X) - min(X)) \quad (9)$$

In many research works, it is found that normalization has a great impact to improve overall accuracy and efficiency while building a statistical model.

To build effective ADS, it is very crucial to identify important relevant features in a dataset. In a machine learning model, considering irrelevant features might decrease the performance of the model and overall time efficiency. Thus, feature selection plays an important role while building any statistical model. In this study, we applied feature selection based on the voting method for each dataset. In this method, we selected features by considering multiple data aspects like dimensionality, mean, length, and variance of the dataset. To fulfill the above-listed aspects of data we combined different feature selection techniques such as LightGBM, Random forest, embedded approach, RFE wrapper, data variance, feature correlation, and Chi-2. Based on the highest vote of different feature selection methods, we selected optimal features.

In addition to feature selection, to increase the computational and modeling efficiency we applied Principal Component Analysis (PCA) as a dimensionality reduction technique as described in [4] on selected optimal features. Using the above feature selection approach, the most relevant features obtained for NSL-KDD dataset are: $dst\_host\_srv\_count$, $dst\_host\_same\_srv\_rate$, dst_host_count, $same\_srv\_rate$, $protocol\_type$, $logged\_in$, $flag$, $dst\_host\_srv\_diff\_host\_rate$, $dst\_host\_serror\_rate$, $dst\_host\_same\_src\_port\_rate$, $count$. The optimal features selected for UNSW-NB15 dataset using the proposed voting approach are: $synack$, $sttl$, $sinpkt$, $dttl$, $dload$, ct_srv_dst, $swin$, $smean$, $sload$, $sbytes$, $rate$, dur, $dmean$, $dbytes$, ct_state_ttl, ct_srv_src, $ct\_dst\_src\_ltm$, $ackdat$. After applying the PCA on selected features, we reduced the dimensionality to 10 for NSL-KDD and 3 for the UNSW-NB15 dataset.

B. Area estimation and standard profile creation

To build the ADS, we divided both datasets into training and testing sets. Only normal data vectors were selected to create a standard profile. Besides, we further divided the UNSW-NB15 dataset according to attack types, including Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, and Worms to evaluate detection of each category using different models. In the training phase, for every given normal data vector we estimated the parameters of the tested mixture models using Maximum Likelihood (ML) and Expectation-Maximization (EM) algorithms. The normal profile includes the estimated parameters, PDFs for normal

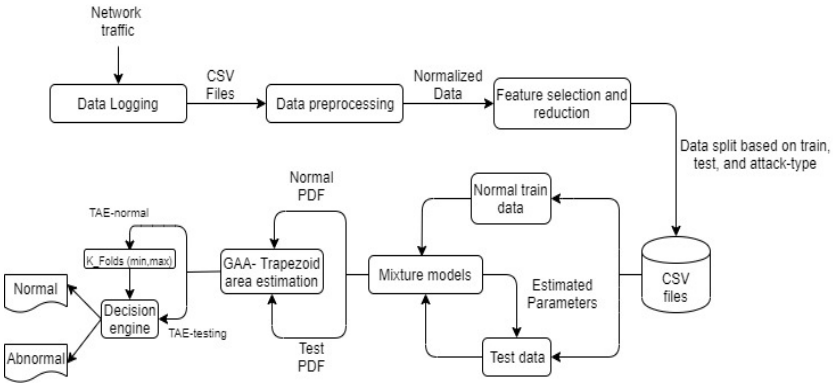


Fig. 2: Framework for anomaly detection system.

data vector (PDF^{normal}), absolute distance (calculated using mean of all the normal records (μ) and mean of each normal record (μ_n) using following equations 10 to 12, and mixing weights.

$$\mu = 1/N \sum_{i=1}^{N} v_i/(v_i + w_i) \tag{10}$$

$$\mu_n = 1/D \sum_{d=1}^{D} v_{nd}/(v_{nd} + w_{nd}) \tag{11}$$

$$abs_{distance} = |\mu - \mu_n| \tag{12}$$

In testing phase, we use same estimated parameters from normal profile to calculate PDFs for testing set ($PDF^{testing}$) and use mean of normal profile (μ_n) to calculate distance measure for testing records. After calculating (PDF^{normal}) and ($PDF^{testing}$), we use the PDFs to calculate TAE area using eq. 1 to get ($area^{normal}$) and ($area^{testing}$). Further, we divided ($area^{normal}$) into (K^{normal}) folds, where each fold contains minimum and maximum values of normal area for each normal vector. (K^{normal}) folds can be calculated using following equation [4]:

$$K_{folds} = [N/2], [(N-1)/2], [(N-2)/2], ..., [4/2] \tag{13}$$

Finally, we use this (K^{normal})) fold in the decision engine to classify normal and abnormal data instances proposed in [4]. Any observation falling in the range of ($area^{testing} >= min_{K_i}) and (area^{testing} <= max_{K_i}$) is considered as normal otherwise as abnormal data vector.

IV. Experimental results

In this section, we evaluate the performance of the GAA technique using TAE. We deployed two widely used datasets, namely, UNSW-NB15[1] and NSL-KDD[2] for the following models: Beta Mixture Model (BMM), Inverted Beta Mixture Model (IBeta), Generalized Dirichlet Mixture Model (GDir), and Generalized Inverted Dirichlet (GID) mixture model. The parameters of these models have been estimated using the maximum likelihood approach within expectation-maximization framework as detailed in [13], [17], [14], and [19], respectively.

The first set of experiments evaluates the effectiveness of absolute distance with the TAE method to detect malicious attack vectors in a network with low FPR. For BMM, Fig. 3a and Fig. 3b represent the areas estimated for normal and abnormal data vectors without using any distance measure with minimum and maximum range for the normal profile from 0.2 and 0.89, respectively. TAE estimate for abnormal data instance is 0.87, which is under normal range and produces FPR.

Fig. 3c and Fig. 3d, represent IBeta-TAE for normal and abnormal data vectors. It can be observed that some abnormal data vectors fit the same as the normal vectors and thus generate high FPR. We can observe that by adding distance measure, in Fig. 3c and Fig. 3d, the model can distinguish between normal range (0.383 to 0.841) and abnormal data vectors. In this case, the overall IBeta-TAE value (0.836) is close to the normal range, but still, some abnormal data are considered as normal.

Fig. 3e and Fig. 3f shows the effectiveness of the GID model using TAE technique and distance measure to detect the same normal and abnormal data vectors very effectively with a normal profile range from 0.27 to 0.82 and significant change in respective TAE values for abnormal data vector value as 0.895. Thus, the GID-TAE gives good detection capability with low FPR as we prove it also in the next set of experiments.

Fig. 3g and Fig. 3h, represent GDir-TAE for normal and abnormal data vectors. Although It can be observed that some abnormal data vectors look like normal ones and thus reduce the overall accuracy, the model can distinguish between normal range (0.12 to 0.77) and abnormal data vectors. In this case, the overall GDir-TAE value is (0.86) and generates less FPR.

In the second set of experiments, the performance evaluation of mixture models was conducted with selected features, and the principal components using the PCA technique for

[1]https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/

[2]https://www.unb.ca/cic/datasets/nsl.html

TABLE I: Overall accuracy for NSL-KDD.

Model	ACC (%)	DR (%)	FPR (%)
BMM	92.12	99.16	0.29
IBeta	90.50	97.00	0.29
GID	91.12	94.33	0.18
GDir	87.28	90.60	0.21

TABLE II: Overall accuracy for UNSW-NB15.

Model	ACC (%)	DR (%)	FPR (%)
BMM	95.86	93.50	0.023
IBeta	96.40	96.00	0.014
GID	96.44	97.50	0.010
GDir	98.85	96.00	0.010

both datasets. The overall accuracy of the different models is measured by Detection Rate and False Positive Rate. Tables I and II summarize the obtained results for NSL-KDD and UNSW-NB15 data sets, respectively.

We evaluated the performance of the proposed model with the TAE technique in detecting each type of attack in the UNSWNB15 dataset in the final set of experiments. Table III shows the comparison of the performance test results for accuracy for each attack category in the UNSW-NB15 dataset. The performance of GID with TAE techniques gives us higher accuracy in each attack type as compared to other models except the shellcode attack type.

V. Conclusion

The main goal of this work is to evaluate the Geometric Area Analysis technique based on Trapezoidal Area Estimation with different mixture models. We evaluated the proposed hybrid ADS through extensive experiments involving two well-known datasets, namely, NSL-KDD and UNSW-NB15. We have shown that the GID and GDir using the TAE technique provide promising results by accurately detecting abnormal network behavior. Moreover, results indicate that the false-positive rate in the GID is much less as compared to other mixture models. Thus, it can be considered to build ADS for a high-speed network to detect malicious activity. By selecting an appropriate number of folds and selecting the optimal number of features using the voting approach with the PCA technique for dimensionality reduction, we achieved a significant improvement in the modeling accuracy. Future works could be devoted to deploying infinite mixture models instead of the finite ones as developed in [22]–[25].

Acknowledgement

The completion of this research was made possible thanks to Natural Sciences and Engineering Research Council of Canada (NSERC) and the INVOLVED ANR-14-CE22-0020-01 project of the French National Research Agency.

References

[1] S. Fu and N. Bouguila, "A bayesian intrusion detection framework," in *2018 International Conference on Cyber Security and Protection of Digital Services, Cyber Security*, 2018, pp. 1–8.

[2] W. Fan, N. Bouguila, and D. Ziou, "Unsupervised anomaly intrusion detection via localized bayesian feature selection," in *11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011*, 2011, pp. 1032–1037.

[3] N. Bouguila and T. Elguebaly, "A fully bayesian model based on reversible jump MCMC and finite beta mixtures for clustering," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5946–5959, 2012.

[4] N. Moustafa, J. Slay, and G. Creech, "Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks," *IEEE Transactions on Big Data*, vol. 5, no. 4, pp. 481–494, 2019.

[5] Ü. Çavuşoğlu, "A new hybrid approach for intrusion detection using machine learning methods," *Applied Intelligence*, vol. 49, no. 7, pp. 2735–2761, 2019.

[6] C. C. Aggarwal, "Outlier analysis," in *Data mining*. Springer, 2015, pp. 237–263.

[7] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*. Springer, 2005.

[8] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *Int. J. Math. Model. Meth. Appl. Sci.*, vol. 1, pp. 300–307, 2007.

[9] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*. Springer Science & Business Media, 2013, vol. 12.

[10] W. Pan, X. Shen, and B. Liu, "Cluster analysis: Unsupervised learning via supervised learning with a non-convex penalty," *Journal of machine learning research : JMLR*, vol. 14, p. 1865, 07 2013.

[11] Y. Ji, C. Wu, P. Liu, J. Wang, and K. R. Coombes, "Applications of beta-mixture models in bioinformatics," *Bioinformatics*, vol. 21, no. 9, pp. 2118–2122, 02 2005.

[12] N. Bouguila, D. Ziou, and E. Monga, "Practical bayesian estimation of a finite beta mixture through gibbs sampling and its applications," *Statistics and Computing*, vol. 16, pp. 215–225, 06 2006.

[13] N. Bouguila and D. Ziou, "Using unsupervised learning of a finite dirichlet mixture model to improve pattern recognition applications," *Pattern Recognit. Lett.*, vol. 26, no. 12, pp. 1916–1925, 2005.

[14] N. Bouguila and D. Ziou, "A powreful finite mixture model based on the generalized dirichlet distribution: Unsupervised learning and applications," in *17th ICPR 2004, Cambridge, UK, August 23-26, 2004*, 2004, pp. 280–283.

[15] N. Bouguila and D. Ziou, "A countably infinite mixture model for clustering and feature selection," *Knowl. Inf. Syst.*, vol. 33, no. 2, 2012.

[16] T. Bdiri and N. Bouguila, "Learning inverted dirichlet mixtures for positive data clustering," in *Proc. of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, ser. RSFDGrC'11. Springer-Verlag, 2011, pp. 265–272.

[17] T. Bdiri and N. Bouguila, "Positive vectors clustering using inverted dirichlet finite mixture models," *Expert Syst. Appl.*, vol. 39, 02 2012.

[18] T. Bdiri and N. Bouguila, "Bayesian learning of inverted dirichlet mixtures for SVM kernels generation," *Neural Computing and Applications*, vol. 23, no. 5, pp. 1443–1458, 2013.

[19] M. A. Mashrgy, T. Bdiri, and N. Bouguila, "Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted dirichlet mixture models," *Knowl. Based Syst.*, vol. 59, pp. 182–195, 2014.

[20] S. Bourouis, M. A. Mashrgy, and N. Bouguila, "Bayesian learning of finite generalized inverted dirichlet mixtures: Application to object classification and forgery detection," *Expert Syst. Appl.*, vol. 41, no. 5, 2014.

[21] T. Asano, T. Asano, and H. Imai, "Partitioning a polygonal region into trapezoids," *Journal of the ACM*, vol. 33, pp. 290–312, 1986.

[22] N. Bouguila and D. Ziou, "A dirichlet process mixture of dirichlet distributions for classification and prediction," in *2008 IEEE Workshop on Machine Learning for Signal Processing*, 2008, pp. 297–302.

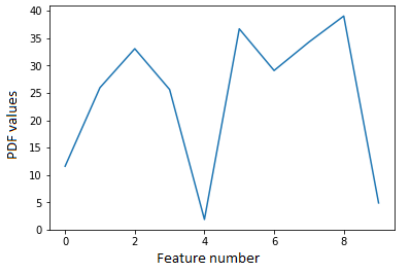
[23] W. Fan and N. Bouguila, "Variational learning of a dirichlet process of generalized dirichlet distributions for simultaneous clustering and feature selection," *Pattern Recognit.*, vol. 46, no. 10, 2013.

[24] T. Bdiri and N. Bouguila, "An infinite mixture of inverted dirichlet distributions," in *Neural Information Processing - 18th International Conference, ICONIP , Proceedings, Part II*, ser. Lecture Notes in Computer Science, vol. 7063. Springer, 2011, pp. 71–78.

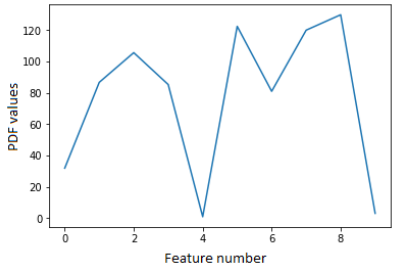
[25] T. Bdiri, N. Bouguila, and D. Ziou, "Variational bayesian inference for infinite generalized inverted dirichlet mixtures with feature selection and its application to clustering," *Appl. Intell.*, vol. 44, no. 3, pp. 507–525, 2016.

TABLE III: Accuracy with TAE technique for all the attack types in UNSW-NB15 dataset

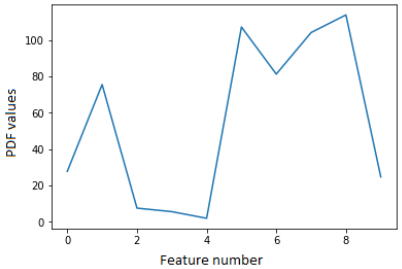
Dataset	Analysis (%)	Backdoor (%)	DoS (%)	Exploits (%)	Fuzzers (%)	Generic (%)	Reconnaissance (%)	Shellcode (%)	Worms (%)
BMM	92.93	95.06	94.53	90.80	90.53	91.33	93.46	87.06	89.83
IBMM	88.53	96.80	94.13	92.93	90.00	96.66	93.73	83.20	92.66
GID	91.28	99.57	96.71	100.00	90.85	99.85	97.14	85.71	94.28
GDir	97.57	93.57	96.71	83.85	90.28	99.85	90.10	88.42	96.82



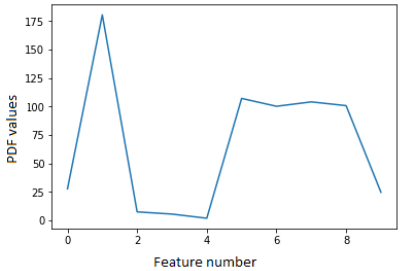
(a) Normal vector PDF (Beta without Dist. TAE [0.2-0.89])



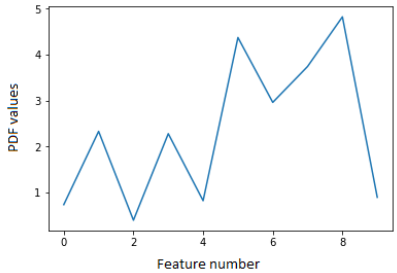
(b) Abnormal vector PDF (Beta with Dist. TAE-0.87)



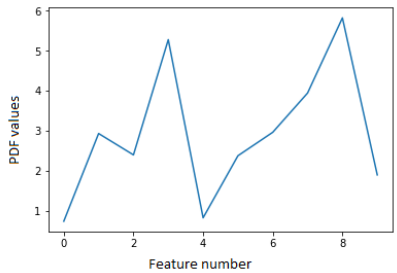
(c) Normal vector PDF (IBeta without Dist. TAE [0.38-0.84])



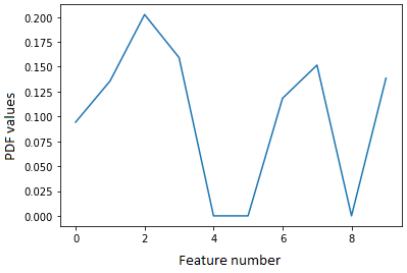
(d) Abnormal vector PDF (IBeta with Dist. TAE-0.83)



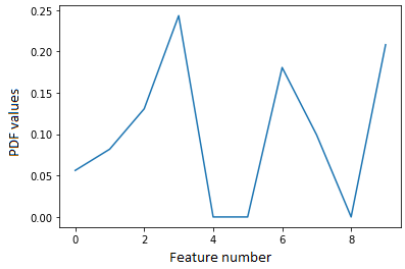
(e) Normal vector PDF (GID without Dist. TAE [0.27-0.82])



(f) Abnormal vector PDF (GID with Dist. TAE-0.89)



(g) Normal vector PDF (GDir without Dist. TAE [0.12-0.77])



(h) Abnormal vector PDF (GDir with Dist. TAE-0.86)

Fig. 3: Normal profile range with TAE values Beta, IBeta, and GID model to detect abnormal data vector