Multi-Stage Attack Detection Using Layered Hidden Markov Model Intrusion Detection System

Student: Wondimu K. Zegeye {wozeg1}@morgan.edu

Advisors: Dr. Farzad Moazzami, Dr. Richard Dean

Morgan State University, Electrical and Computer Engineering Department

ABSTRACT

Intrusion Detection Systems (IDS) based on Artificial Intelligence can be deployed to protect telemetry networks against intruders. As security solutions which encrypt radio links do not accommodate the ever evolving network attacks and vulnerabilities, new defense mechanisms using machine learning and artificial intelligence can play a significant role for telemetry networks. This paper proposes a multi-layered Hidden Markov Model (HMM) IDS that addresses multi-stage attacks. This is due to the fact that intrusions are increasingly being launched through multiple phases instead of single stage intrusion. This layered model divides the problem space into smaller manageable pieces reducing the curse of dimensionality associated with HMMs. To verify the application of this model for real network, the NSL-KDD dataset is used to train and test the model.

Key words: *Intrusion Detection System (IDS), Hidden Markov Model (HMM), Multi-stage attacks, Artificial Intelligence (AI)*

1. INTRODUCTION

As the percentage of encrypted internet traffic increases, it provided attackers better power to hide their command and control activities. However, it creates major problem for defenders as it creates challenges to properly identify normal and anomalous traffic [1].

Attackers are already applying machine learning algorithms to develop intelligent malware. In order to protect networks from these kinds of attacks, security teams need to develop Intrusion Detection Systems which use these technologies.

Anomaly based IDSs which use machine learning and artificial intelligence are at the forefront in detecting multi-stage attacks, which are executed over long spans of time. These attacks constitute several distinct events, such as network scanning, to launch the final attack.

2. BACKGROUND

Intrusion detection systems are placed at strategic points in the network or host systems, termed as network-based (NIDS) and host-based (HIDS) intrusion detection systems. The IDS basically looks for abnormal activity in the network by monitoring network traces, and alert system administrators to take corrective actions. A system administrator can block users or systems from accessing vulnerable ports, deny access to specific IP address or turn off services suspected for unusual network traces. It is a front line of defense for network administrators in the defense against adversaries.

IDS can use signatures, anomalous behaviors, and protocol behaviors to detect unusual activities in a network. Signatures are previously gathered knowledge from identified system vulnerabilities and attacks. Anomaly is used to describe divergence from a known behavior such as profiles which are collected by consistently tracking user activities and hosts in the network for a given period of time. Protocol analysis studies uncommon patterns or messages from the standardized protocol behaviors. To develop an IDS with better scalability and accuracy of malicious event detection, a hybrid of several methods are used in combination.

Several approaches are also considered to classify IDSs such as statistical analysis, pattern detection and rules. Statistical analysis approaches apply algorithms on collected data to build

models which are capable of self-study. Patterns-based ones use simple concepts and graphic depictions using user defined intrusion signatures.

Recent methodologies in the development of IDS explore machine learning and artificial intelligence. Artificial neural networks, fuzzy logic, random forest to name a few have been used in the academia and industry. Previous works in the Wireless and Network Security (WiNetS) lab focused on the Hidden Markov Model (HMM) based IDS design using single HMM and Multi-Class System (MCS) classifier designs [2-3].

3. DESIGN APPROACH: DATA PROCESSING

The design approach in this work follows Figure 1. The datasets are processed to provide the HMM IDS time series sequence of data.

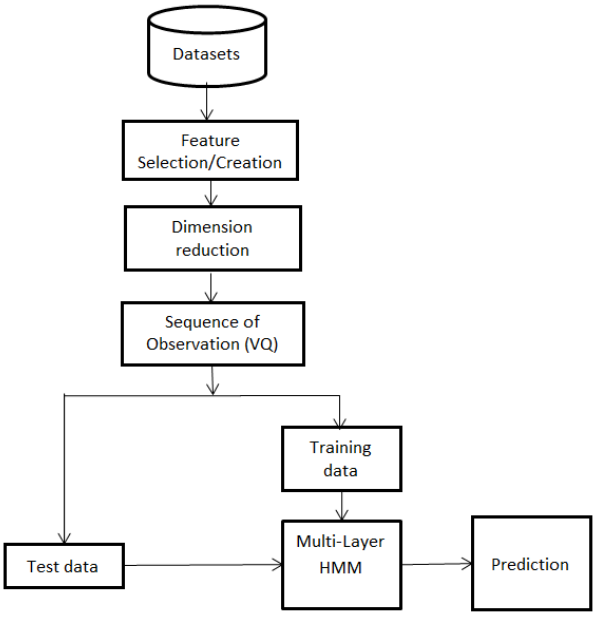

Figure 1: IDS Design Architecture

3.1 Datasets

To train and test the developed IDS, two publicly available datasets are used. The first one is NSL-KDD which is an optimized version of the original KDD Cup 1999 dataset prepared by removing duplicates in the dataset [4]. The second dataset is the CICIDS2017, which is

developed by the Canadian Institute of Cyber Security to cover what are commonly known as the eleven criteria in building a reliable benchmark dataset [5].

3.2 Feature Extraction/ Dimension Reduction

A Linear transformation function which maps a high dimensional data into a lower dimensional data is applied using a multivariate statistical function such as Principal Component Analysis (PCA) via Singular Value Decomposition (SVD) [6]. PCA is applied on the dataset after pre-processing steps so the original n features (or attributes) $A_1, A_2, \dots, A_n$ results in a set of new features $B_1, B_2, \dots, B_m$ $(m < n)$, $B_i = F_i(A_1, A_2, \dots, A_n)$ and F_i is a linear combination mapping function. It is important to note that normalization using log-transformation is applied on the datasets before applying PCA.

3.3Vector Quantization

The outputs of PCA computation can be further reduced by applying vector quantization (VQ). A basic K-Means algorithm is applied to the reduced dimension data to quantize the data into a single dimensional data encodes using the cluster labels [7]. The distinct cluster labeled data is used as an input to the HMM.

4. MULTI-LAYER HMM BASED IDS

This section discusses the architecture of the Multi-layer HMM IDS design and the results of different tests.

4.1 Multi-Layer HMM IDS Architecture

A HMM is a stochastic model that represents a dynamic process of two related random processes [8-9]. An underlying stochastic process that is not observable (hidden states) can be observed through another set of stochastic processes that produces the sequence of observed symbols. A HMM consists of a set of N distinct "hidden" states of the Markov process $Q = \{q_1, q_2, \dots, q_N\}$ and a set of M observable symbols per State $= \{v_1, v_2, \dots, v_M\}$. The overall HMM model is defined as follows with q_t and o_t denoting the state and observation symbol at time t respectively.

The HMM is specified by a set of parameters (A, B, Π):

1) The prior probability distribution $\Pi = \Pi_i$where$\Pi_i = P(q_1 = s_i)$are the probabilities of$s_i$being the first state in a state sequence.

2) The transition probability matrix $A = \{a_{ij}\}$ where $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$, are the probabilities of going from state s_i to state s_j.

3) The emission (observation) probability matrix$B = \{b_{ik}\}$ where $b_{ik} = P(o_t = v_{kj} | q_t = s_i)$ are the probabilities to observe s_k if the current state is $q_t = s_i$.

Model Structure of Two-Layered HMM (LHMM)[10]

1. At Layer 1, we have HMM_1, HMM_2,…, HMM_p with their corresponding number of hidden states S_1, S_2, …, S_p.

2. Considering the same time granularity ($t = T$) for each of the HMMs,

- The observation sequence for each of the HMMs are given as:

$O_1^T = \{O_1^1, O_1^2, \ldots, O_1^T\}$, $O_2^T = \{O_2^1, O_2^2, \ldots, O_2^T\}$, …, $O_p^T = \{O_p^1, O_p^2, \ldots, O_p^T\}$

- The probable sequence of states for each of the HMMs are given as:

$Q_1^T = \{q_1^1, q_1^2, \ldots, q_1^T\}$, $\{q_2^1, q_2^2, \ldots, q_2^T\}$, …, $\{q_p^1, q_p^2, \ldots, q_p^T\}$

3. A new feature vector is constructed from the Layer 1 HMMs probable sequence of states. This statistical feature can be considered as a new data matrix where VQ can be applied and a new sequence of observations will be created from the Layer 2 HMM.

The feature vector is constructed as follows:

$$f_i = \begin{pmatrix} q_1^i \\ \vdots \\ q_T^i \end{pmatrix}, \forall \mathrm{i} = 1, 2, \ldots, \mathrm{p} \quad (1)$$

$$F = (f_1, f_2, \ldots f_j), \forall \mathrm{j} = 1, 2, \ldots, \mathrm{p} \quad (2)$$

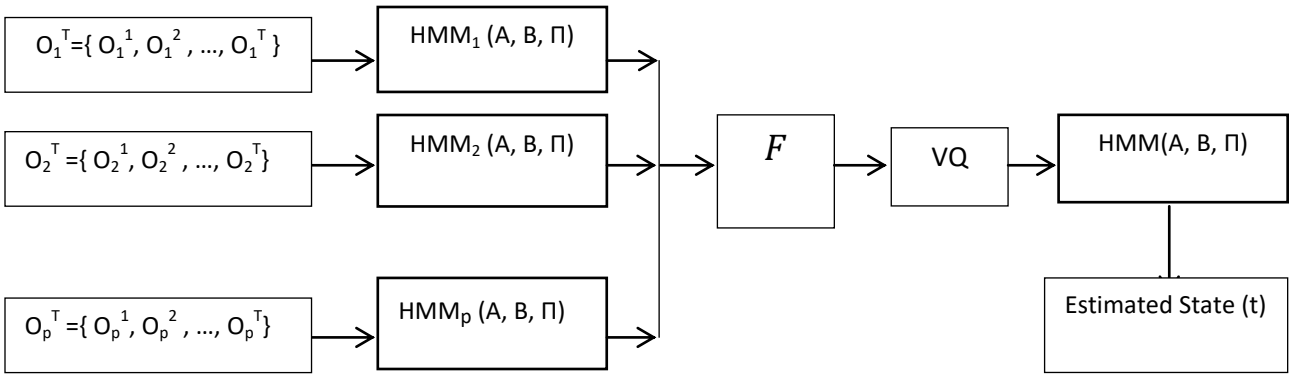

Figure 2: Multi-Layered HMM (LHMM) model structure.

4.2 Experimental Results

4.2.1. Training the LHMM

The lower layer and the upper layer HMMs are trained using their corresponding training data and the optimized model parameters (A, B, Π) are determined.

1. **HTTP HMM training:** consists 1200 time series VQ data which has normal (HTTP-BENIGN) and attack (HTTP-Web-attack-brute force) traffic.

$$A = \begin{bmatrix} 0.9827 & 0.0173 \\ 0.0140 & 0.9860 \end{bmatrix} \quad B = \begin{bmatrix} 0.3088 & 0.0973 & 0.2007 & 0.3932 \\ 0.0000 & 0.8129 & 0.0952 & 0.0919 \end{bmatrix} \quad \pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Table1: Hidden State Symbols

State Symbols	HTTP
1	HTTP-BENIGN
2	HTTP-Web-attack-brute force

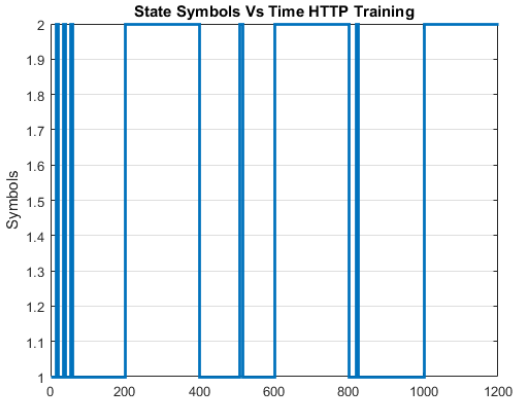

Figure 3. Time, State symbols

2. SSH Training: consists 1200 time series VQ data which has normal (SSH-BENIGN) and attack (SSH-patator) SSH traffic.

$$A = \begin{bmatrix} 0.9772 & 0.0228 \\ 0.0308 & 0.9692 \end{bmatrix} \quad B = \begin{bmatrix} 0.6135 & 0.1518 & 0.2348 \\ 0.5092 & 0.2482 & 0.2427 \end{bmatrix} \quad \pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Table 2. Hidden State Symbols

State Symbols	SSH
1	SSH-BENIGN
2	SSH-Patator

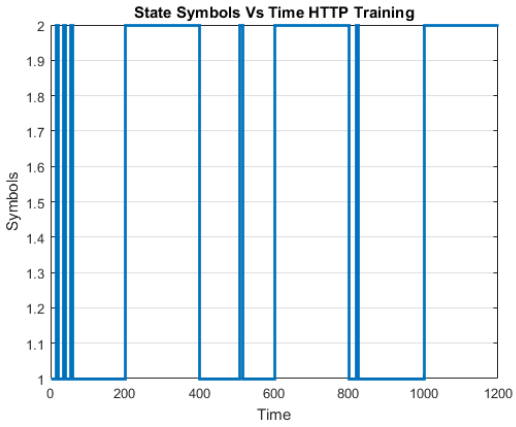

Figure 4. Time, State symbols

3. Upper Layer training: the data for the upper layer training is constructed from the posterior Viterbi paths of the HTTP and SSH training data using Equations (1) and (2).

$$A = \begin{bmatrix} 0.9860 & 0.0000 & 0.0000 & 0.0000 & 0.0140 \\ 0.0011 & 0.8962 & 0.0026 & 0.0988 & 0.0014 \\ 0.0000 & 0.2557 & 0.2870 & 0.4240 & 0.0334 \\ 0.0255 & 0.7642 & 0.0024 & 0.1401 & 0.0678 \\ 0.0568 & 0.0105 & 0.0684 & 0.0007 & 0.8636 \end{bmatrix} B = \begin{bmatrix} 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.9996 & 0.0000 & 0.0004 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 \end{bmatrix} \pi = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Table3: Hidden State Symbols

State Symbols	HTTP	SSH
1	HTTP-BENIGN	SSH-BENIGN
2	HTTP-Web-attack-bruteforce	SSH-Patator
3	HTTP-Web-attack-bruteforce	SSH-BENIGN
4	HTTP-Web-attack-bruteforce	SSH-Patator
5	HTTP-BENIGN	SSH-Patator

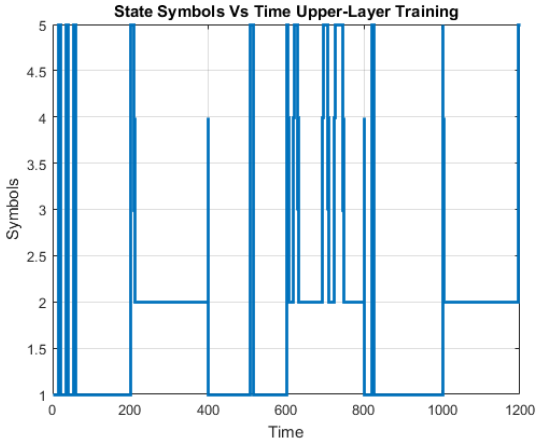

Figure 5. Time, State symbols

4.2.2. Testing the LHMM

Case I: This case shows the testing result of the HMMs where the testing data constitutes the same hidden states as in the case of the training phase. This test shows a new pattern at the upper layer which is the result of the lower layer attacks patterns. The new attack pattern emulates a multi-stage attack due to the pattern of the lower layer attacks. In addition, this test can be used as a baseline to identify when a new attack occurs as the observation matrix of the upper layer

HMM does not have a significant deviation from the observation matrix of the Upper layer HMM training phase in section 4.2.1.

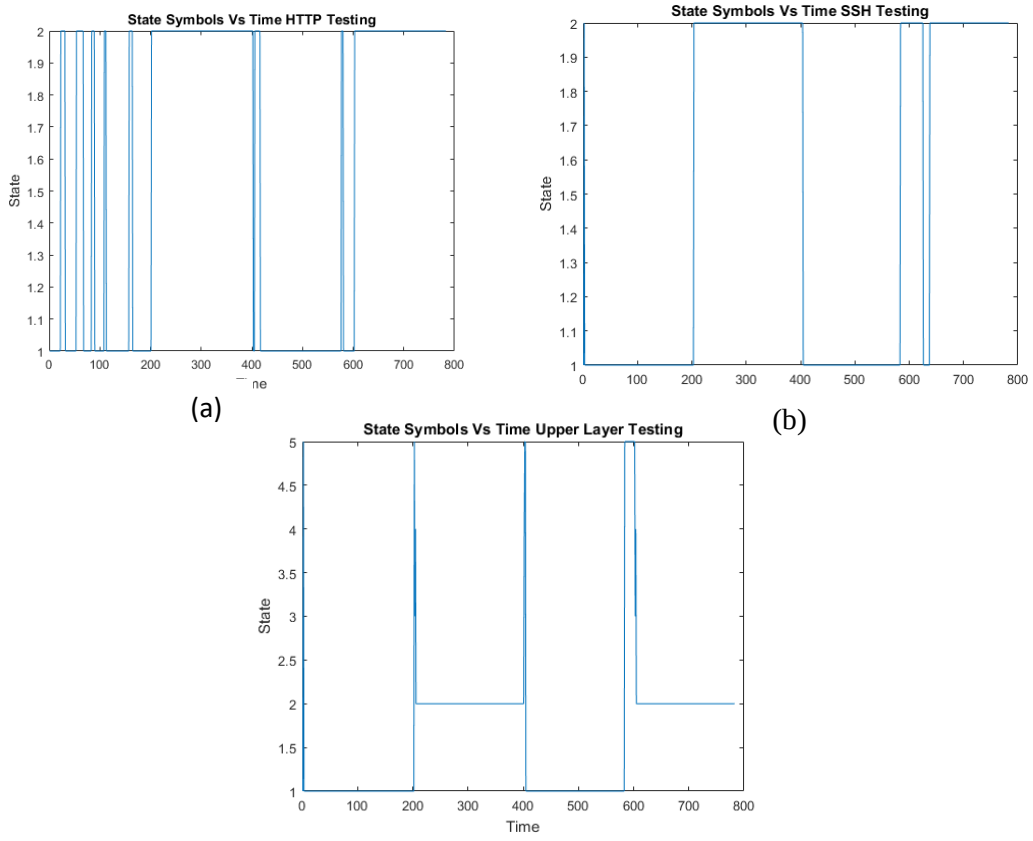

Figure 6. Time, State Symbols (a) HTTP (b) SSH (c) Upper Layer

C**ase II:** In this case the testing data of the SSH traffic consists an additional DOS traffic (τ=400 to τ=800) taken from the NSL-KDD datasets. As can be seen in Figure 7(b), the DoS data from τ=400 to τ=600 is identified as hidden state 2 instead of state 1 in case of the SSH traffic in Figure 6(b). Even if the traffic from τ=600 to τ=800 is also DoS, it is identified as hidden state 2. This change in observed new attack (DoS) traffic is reflected by the change in the observation matrix (highlighted above in RED) of the upper layer HMM as compared to the observation matrix in the training phase.

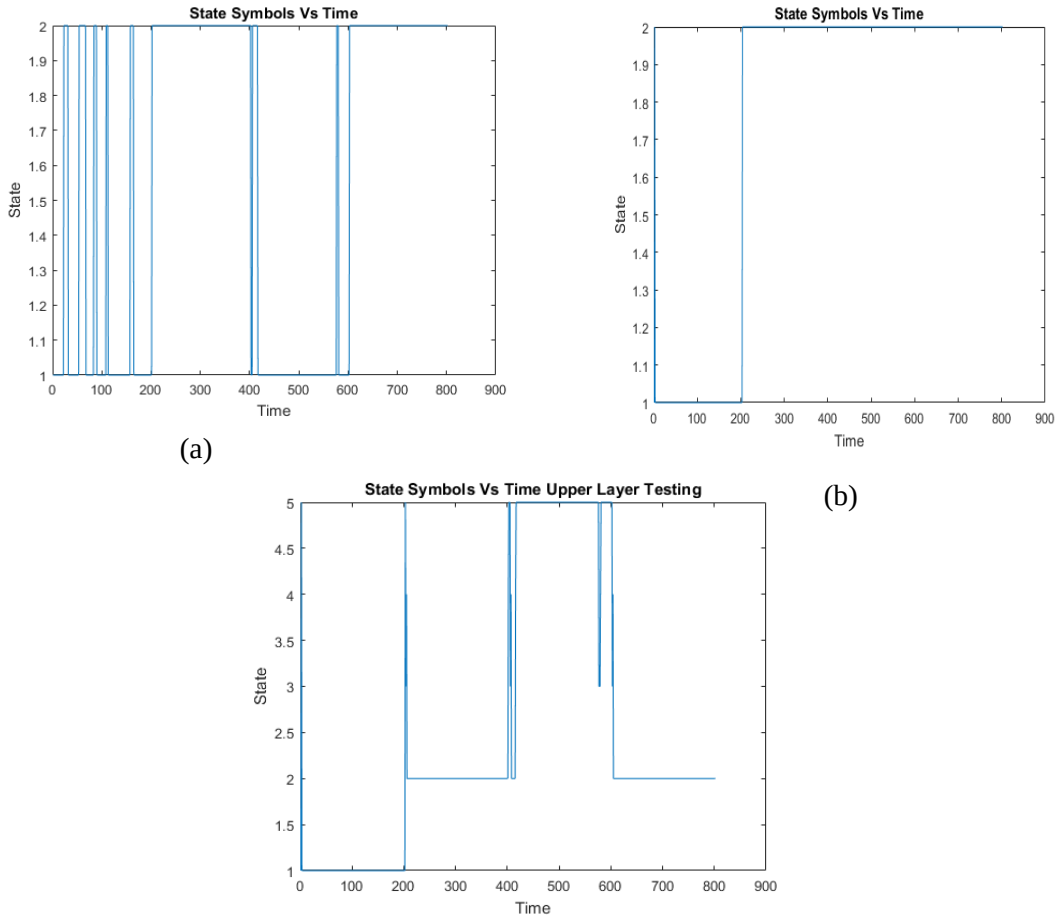

Figure 7. Time, State Symbols (a) HTTP (b) SSH with new attack (c) Upper Layer with new attack

$$B = \begin{bmatrix} 0.0000 & 0.7882 & 0.0049 & 0.2069 \\ 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.9996 & 0.0000 & 0.0004 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 \end{bmatrix}$$

From the change in the observation matrix of the upper layer HMM in case II can also be seen in Figure 7(c) as hidden state 5 as compared to hidden state 1 in Figure 6 (c) for τ=400 to τ=600. For τ=600 to τ=800, the hidden state stays the same in both cases as the traffic is virtually recognized similar in the lower layer SSH HMM.

The change in the observation matrix at the upper layer can be used as a proof of existence a new attack which plays an important role as it detects the attack in zero-day. The capability

of the model based on machine learning algorithm to detect attacks which are referred as unknown-unknown, attacks which are previously unseen, makes it very useful for zero-day attack detection.

5. CONCLUSIONS

This work highlights the potential for use of multi-layer HMM-based IDS for telemetry applications. This approach can factor a problem of huge dimensions into small and manageable pieces as lower layers. The upper layer can be used to identify multi-phase events which can stem from the lower layer events that can be observed over longer duration.

REFERENCES

[1]Cisco 2018 Annual Cybersecurity Report. Available online: https://www.cisco.com/c/en/us/products/security/security-reports.html (accessed on 17 July 2019).

[2]Nadim Maharjan, , Paria Moazzemi, "Telemetry Network Intrusion Detection System",*International Telemetering Conference (ITC)*, Las Vegas, USA, Oct 24-26, 2012.

[3] Zegeye, W.K.; Moazzami, F.; Richard Dean, R.A. Design of Intrusion Detection System (IDS) Using Hidden Markov Model (HMM). In Proceedings of the International Telemetering Conference, Glendale, AZ, USA, 5–8 November 2018.

[4]KDD Cup 1999 Data, retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[5]Sharafaldin, I.; Gharib, A.; Lashkari, A.; Ghorbani, A.A. Towards a reliable intrusion detection benchmark dataset. Softw. Netw. 2017, 177–200, doi:10.13052/jsn2445-9739.2017.009.

[6]Lindsay I Smith, "A tutorial on Principal Components Analysis", February 26, 2002

[7] Stuart P. Lloyd.,"Least squares quantization in pcm", IEEE Transactions on Information Theory, 28:129-137, 1982.

[8] L. Rabiner and B.Juang, "An Introduction to hidden Markov Models", IEEE ASSP Mag., vol. 3, no. 1, pp. 4H16, Jan. 1986.

[9] L.Rabiner, "A tutorial on hidden Markov Models and selected applications in speech recognition," Proc. IEEE, Vol. 77, no.2, pp.257-286, Feb. 1989.[11]L.Baum and T.Petrie, "Statistical inference for probabilistic functions of finite state Markov Chains," Ann.Math.Stat.,Vol. 37,no. 6, pp. 1554-1563, Dec. 1966.

[10] Zegeye, W., Dean, R., Moazzami,F.;Multi-layer Hidden Markov Model based Intrusion Detection System. Mach. Learn. Knowl.Extr. (1), 265-286(2019).