A light-weighted ANN architecture for the classification of cyber-threats in modern communication networks.

Eleni Ketzaki, Anastasios Drosou, Stavros Papadopoulos, Dimitrios Tzovaras
Information Technologies Institute
Centre for Research and Technology Hellas (CERTH)
Thessaloniki, Greece
{eketzaki, drosou, spap, Dimitrios.Tzovaras}@iti.gr

***Abstract*—In modern communication networks, the integrity of the security is of great importance, since the existence of cyber attacks may lead to disastrous financial and social consequences. The anomaly detection constitutes an essential part of network security. This paper proposes a two-stage procedure to provide a solution regarding the anomaly detection and threat identification. The proposed method is suitable for modern communication networks and upcoming smart networks. The first stage of the method concerns the detection of abnormal incidents and the second stage involves the identification of the type of cyber threats, in case of an attack. The method based on the development of artificial neural network models and the UNSW-NB15 dataset is used to validate the proposed methodology. The experimental results confirm that the proposed method identifies all type of threats in comparison to the already known methods that identify only the threats that appear frequently.**

***Index Terms*—5G network, Anomaly Detection, Artificial neural network model, 5G security, IoT.**

I. Introduction

The modern communication networks face new security challenges due to their network design. They consist of a mixture of different traffic types using various transport technologies [28]. The Internet of Things (IoT) and the 5G networks joined the modern communication networks.

The new enabled developments related to the IoT, the smart cities and the smart devices interact automatically with the physical world and constitute networks vulnerable to more and new threats [6]. The 5G networks challenge with similar issues since their usage expected to bring revolutionary changes in the information and communication technology field because they aim to provide very high user data rate, higher mobile data volume per geographical area, superfast bandwidth speed and connections that are more reliable [1].

In this study, we develop an efficient method of network anomaly detection on modern communication networks. We illustrate the usage of this method in the IoT and in the 5G networks. The development of our approach is of great importance because it aims to maintain the integrity of the security in the networks by detecting malicious attacks and identifying the type of attack. The existence of cyber attacks, can severe financial and social consequences, as well as the endangerment of privacy and human lives [2], [3], [4].

The main contribution of this study is the detection of all type of network attacks, in comparison with the methods that already exist on network anomaly detection [12], [13]. The proposed method based on the usage of artificial neural networks (ANNs) models. The ANNs are one of the most capable Artificial Intelligence (AI) tools for solving very complex problems in the area of AI. The ANNs are inspired by the biological neural networks, and they can develop computational models that aim to approach performance comparable to that of the brain.

This study proposes the usage of ANNs models as an anomaly-based intrusion detection system. The Intrusion Detection System (IDS) is an efficient approach for protecting wireless communications in modern networks. The IDS examines the communication traffic taking place in a network and generates reports to the management system by differentiating the malicious behaviour [14]. The network-based IDSs are grouped into five basic categories as follow: the signature-based detection (SBD), the anomaly-based detection (ABD), the specification-based detection (SPBD), the stateful protocol analysis detection (SPAD), and the hybrid intrusion detection [5].

The ABD system identifies possible differences between the target events and the predefined normal transmissions. The comparison can determine whether there is a partition between usual and unusual behaviours, considering the abnormal behaviour as an active or potential attack, which depends on the level of differences. Three main techniques are commonly used for this comparison, the statistical-based, the knowledge-based, and the machine learning techniques. The development of the machine learning techniques in an IDS creates a model which based on a training dataset that contains a collection of data instances, each instance of the dataset is described using a set of features [7]. Several machine learning-based schemes have been applied to IDS [8], [9]. The most important techniques are the Artificial neural networks (ANN), the Bayesian networks, the Markov models, the Fuzzy logic techniques, the Genetic algorithms, Clustering - outlier detection and Data mining.

The rest of this paper organized as follow; the second section contains the literature review concerning the usage of machine learning techniques for anomaly detection and multiclass classification that leads to identification of the type of threats. The third section includes the description of the proposed methodology. The validation of the proposed method obtained from the usage of the UNSW-NB15 dataset is described in the fourth section. The fifth section provides a comparative evaluation of the proposed method with the methods that already exist in literature. Finally, the last section contains the conclusion and the contribution of the proposed methodology.

II. Related work – motivation

The importance of security in modern communication networks enforces scientific research towards the anomaly detection and classification of cyber threats. Considering technologies such as Software-defined networking (SDN), IoT and 5G networks, new security aspects are expected [16].

Many researchers have highlighted the importance of security in modern communication networks and especially in the IoT. Becara et al., [25] presented a theoretical overview of the security issues and challenges for the IoT. Cyber attacks constitute an essential security issue, that should be arranged, taking into consideration the scalability and the mobility of the modern network. Zarpelao et al., [29] provide extensive literature research for the importance of security tools in IoT. They conclude that the investigation of different detection methods and the improvement of security of the alert traffic constitute main issues for future research.

The validation of the several methods that proposed by researchers to identify, detect and mitigate the attacks in the next-generation networks methods based on testbed measurements [11], [13], [16], on existing labeled datasets (e.g the KDD99 dataset, the CTU dataset) [12], [15] and on randomly generated features [10].

The proposed techniques that used for anomaly detection vary. However, deep learning techniques seem to gain more ground. In this respect, Maimo et al., [10] propose a 5G oriented architecture to identify cyber threats in 5G mobile networks by making use of deep learning techniques. They use an architecture that arranges the anomaly detection based on two modules. The first module classifies the behaviour either as normal or abnormal, and the second module describes a symptom sequence classification problem. Their work is not interested in the accuracy of the detection mechanism, but exclusively in the execution time. Besides, they do not make use of existing labelled datasets and the features generated randomly. The botnets draw special attention, since the considered as possible cyberthreats for the 5G networks. Maimo et al., [12] based on the CTU dataset, which constitutes of real botnet attacks and propose an architecture based machine-learning models. Santos et al., [13] present an anomaly detection solution for smart city applications based on low-power fog computing solutions. Sohal et al., [19] propose a cybersecurity framework that uses Markov model, IDS, and virtual Honeypots to identify malicious edge devices in fog computing environments. Li et al [15] propose IDS methods using the KDD99 dataset. They use random forest to select a subset of typical traffic features and classify network flows by combining K-Mean and Adaboost algorithm. Nawir et al., [20] compare three types of machine learning algorithms to determine the performance in terms of classification rate and processing time. They conclude that the average one dependence estimation (AODE) is an effective and efficient classifier for network detection of binary classification in comparison with Bayesian Network and Naive Bayes.

An important extension of anomaly detection is the multi-class classification that aims to specify which is a certain type of threat in case of an attack. In this respect, Tchakoucht et al., [21] proposed a recurrent neural network the Multilayerd echo-state machine (ML-ESM) to model intrusion detection and to develop binary and multilabel classification. They assess the proposed model on three datasets; the KDD99, the NSL-KDD and the UNSW-NB15 dataset and they measured the corresponded performance. They conclude that the performance is high if the dataset classifies the attack into a small number of categories. Moreover, it is noted that as the number of categories increases the performance of the method becomes poor, and it can not detect threats that they rarely appear. Baig et al., [22] propose a cascade structure of ANN that aims to divide the incidents into smaller sub-classes and then combine the solutions to form a classifier. They also conclude that they have successfully suggested an intrusion detection method but their proposal can not identify attacks that they rarely occur. Catac, [23] proposes a two-stage classification technique for the detection of malicious network flow. The proposed method assumes a two-stage model. In the first stage, the model decides whether the flow is malicious or not and in the second stage, it extracts the class of the network flow. The random forest, the decision tree, the neural network and the Adaboost developed as different classifiers, and the results regarding the precision and the recall presented.

In this study we examine the development of ANN models for the network anomaly detection and for the threat identification. Inspired by the work of Maimo et al., [12], that aims to detect cyberthreats in 5G networks and from the work of Tchakoucht et al., [21] and Catac, [23] that purpose methods for multilabel classification. The UNSW-NB15 is used to validate the proposed method. This dataset generates modern normal activities and attack behaviors from the network traffic. The current normal traffic proved that is different from the existing traffic described in datasets (e.g. KDD99) which were created two decades ago [18] and contains multitude threats in comparison to the CTU dataset [12].

III. Method Description

The aim of the proposed methodology is twofold; its objective is anomaly detection and the identification of the type of cyber threats. The architecture of the methodology is

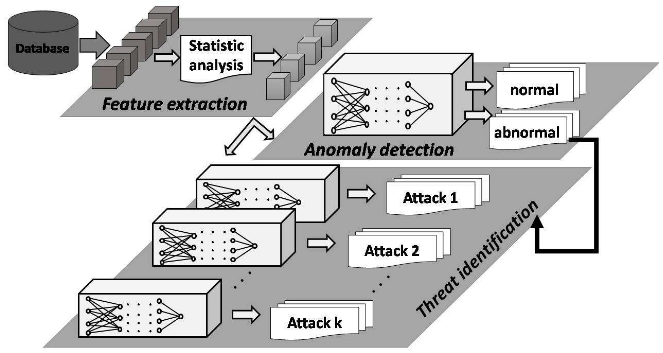

Figure 1. The Architecture of the proposed methodology that consist of three parts: the feature extraction part, the anomaly detection part that obtains from the usage of one ANN model and the Threat identification part that obtains from the usage of different ANN models per type of attack

depicted in Figure 1 and consists of three main parts: the first part is the feature extraction part, the second concerns the part of anomaly detection, and the third is the part of threat identification.

A. Feature extraction

The feature extraction part is based on statistical methods and aims to minimize the demanded number of features that set the input variables of the ANNs models, which will be used in the next parts of the procedure.

The selection of fewer features is of great importance because it accelerates the data analysis process. More specifically, the usage of fewer features leads to the reduction of large data volume that arises in modern communication networks. The feature reduction makes also quicker the training process that will set the first step to a more reliable network. Kwon et al. [26] highlight the advantages of feature selection. More specific, they claim that the feature selection part reduces the dimensionality of the feature space, speed up the learning algorithm and improve the comprehensibility of the learning results.

The proposed methodology aims to identify the correlation between two or more features. The absolute value of the coefficient correlation is to set the criterion for the feature selection. In case that two or more features are highly correlated, then the value of one can predict the value of the other. We assume that we can select only one feature from those that are highly correlated since no extra benefit arises from the use of the others to the model.

Let us suppose that q is the number of the features which are calculated for n different samples and the x_{ij} denotes the value of the j feature that belongs to the i sample, where i=1,2,..,n and j=1,2,..q. The $X_j = (x_{1j}, x_{2j}, ..., x_{nj})$ is the vector that contains the values for the j feature across n samples and the $\overline{X_j}$ denotes the mean value of the X_j.

Assuming that $\rho(X_\mu, X_\lambda)$ denotes the coefficient of correlation for two features X_μ and X_λ we examine the value of the coefficient. In case that the value of the coefficient of correlation $\rho(X_\mu, X_\lambda)$ is close to one, then the features X_μ and X_λ are denoted as high correlated, the value of X_μ feature can predict the value of X_λ and only one of them will be chosen as input of the ANN model.

B. Anomaly detection

The second step of the procedure concerns the part of anomaly detection that will be developed by the usage of an ANN model. The ANNs are computational algorithms that intend to simulate the behaviour of biological systems composed by neurons. In this study, the ANN is used as a method to detect network attacks. Each ANN contains three type of layers, the input layer, the hidden layers and the output layer.

The input layer receives as input the values of the selected features which have obtained from the extraction process.

The hidden layers, apply given transformation to the values of the hidden layers via a procedure of weighted connections. The weights between the hidden layers of the ANN in this study will be the output of the Sigmoid activation function calculated from equation (1),

$$f(x) = \frac{1}{1+e^x} \tag{1}$$

The output layer receives the values that obtain from the hidden layers and returns an output value that corresponds to the prediction of the incident. In this study, the output of the model is binary and discriminates the incidents either as normal or abnormal. The binary classification set the basis to filter the abnormal incidents which are used in the third step of the procedure.

The main reason that the ANN models have been chosen in the current method is that there are plenty of disadvantages to other machine learning techniques that could affect the performance of modern communication networks. Hodo et al., [27] present an analytical comparison of the machine learning techniques. They conclude that the Bayesian network is responsible for slow classification in case that the datasets have many features. The genetic algorithm, although that can solve optimization problems during the classification process, it gets stuck in local optima. The main disadvantages of the support vector machine are the difficulty in the selection of the kernel function, the slow training and the requirement of enough memory space. They also claimed that the K-nearest neighbour is more time-consuming in training, it requires large memory space and it is computationally complex since it takes into consideration all of the training samples. Moreover, they have also noticed similar disadvantages in the machine learning techniques such as decision tree the fuzzy logic and the K-means algorithm. The above disadvantages are in contradiction to the envisions of the modern networks and IoT which are related with high user data rate, the super-fast bandwidth speed and the high mobile data volume per geographical area. Moreover, the ANNs gain a considerable interest since they can handle noisy data with high accuracy and high computational speed [27].

C. Threat identification

The third step of the procedure, aims to identify which is the type of threats for the abnormal incidents. The second part of the methodology is like a filter that distinguishes the abnormal from the normal traffic. In this part of the procedure, we choose only abnormal incidents and their corresponding features. Due to the high frequency of the normal incidents, we exclude them to avoid the creation of an imbalanced dataset. Recent surveys [23] proved that it is very difficult to detect all the types of attacks with high accuracy using only one classifier. The development of a multiclass classification model, lead to the luck of detection of attacks which rarely appear. The proposed method based on the idea to develop different models per type of attack. Let us assume that there are k-classes that represent the type of attacks denoted $\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$. Each object belongs to one of the k-classes. The "one to rest" [24] voting strategy described by expression (2), is used for developing k different samples.

$$f_i(x) = \begin{cases} 1 & \text{if } x \in \omega_i, \quad \forall \omega_i \in \Omega, \\ 0 & \text{if } x \in \omega_j, \quad \forall \omega_j \in \Omega - \{\omega_i\} \end{cases} \tag{2}$$

For each of these samples "1" is mapped if the object belongs to the corresponding class and "0" if the object belongs to the remaining k-1 classes. K separate ANN models developed as k separate classifiers, in a similar way as described in the Section III.B, and used to identify whether an incident belongs to a certain type of attack or not.

The accuracy, the precision and the recall are the metrics that will measure the efficiency of the proposed models and they are commonly used to compare the results of the usage of different methods.

IV. Experiments

The validation of the proposed methodology based on the usage of UNSW-NB15 dataset [18]. The UNSW-NB15 dataset provides the trainset and the testset that can be used for the experimental evalutation. We choose this dataset because it describes the current network traffic and it is not deteriorated in certain types of attacks, such as the CTU dataset, that contains only the botnets. The revolution in network speed and applications, demands the network traffic described by these datasets completely different compared to those of older datasets such as KDD99, proposed two decades ago. The developing and evaluation of the ANN models based on the python libraries Keras and Tensorflow. The available processor was Intel(R) Core(TM) i5-7600CPU @3.50GHz.

A. Description of the Dataset

The UNSW-NB15 dataset generates a hybrid of realistic modern normal activities and the synthetic contemporary attack behavior from the network traffic [18]. It is a new benchmark dataset for evaluating network IDS, which contaid the following types of attacks: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generics, Reconnaissance, Shellcodes and Worms. The Table I, describes the distribution of the incidents contained on the UNSW-NB15 dataset. The majority of the incidents (87.3%) have normal behavior. Generic is the type of attack that has the higher frequency (8.48%) and worms are the type of attack that have the least frequency (0.007%).

Table I
Description of the frequency and the percentage (%) of normal connections and of different type of attacks lead to the distribution of the UNSW-NB15 dataset

Type of attack	Frequency	Percent
Normal	1,218,760	87.351
Fuzzers	24,246	0.955
Analysis	2,677	0.105
Backdoors	2,329	0.092
DoS	16,353	0.644
Exploits	44,525	1.753
Generic	215,481	8.483
Reconnaisance	13,987	0.551
Shellcode	1,511	0.059
Worms	174	0.007
Total	2,000,000	100

B. Feature extraction

The UNSW-NB15 dataset consists of 45 features, which belong into 5 categories: the flow features, the basic features, the content features, the time features and the additional generated features.

We assume the absolute value of the coefficient correlation is the criterion to select the most suitable among the numerical features of the dataset. Based on the correlation matrix for all the selected features the Table II, concludes the selected and the corresponded excluded features.

More specific, the feature that measures "the number of source to destination packets" is highly correlated with the features that measures "the source to destination transaction bytes" (ρ=$0,96$) and with the feature that measures "the number of source packets retransmitted or dropped" (ρ=$0,9711$) therefore the "the number of source to destination packets" is selected since can replace the other features which are extracted. Following the same procedure "the number of destination to source packets" replace the "the destination to source transaction bytes" and the "the number of destination packets retransmitted or dropped" since the values of the coefficient of correlation are (ρ =0.9719) and (ρ =0.9786) respectively. Table II describes in details all the selected and the extracted features that obtained from the correlation matrix. The features that are not included in this table either as selected or as included, are accepted as selected features.

The usage of less features lead to the reduction of large data volume and make quicker the training and detection process. In order to check how the number of features affect the training and the detection process we calculate the elapsed time for this procedure assuming a specific ANN. The elapsed time for an ANN that use 41 features as input layer in comparison to an ANN that has input 26 features is bigger, while the number of training samples increase. The figure 2 describes the difference between mean elapsed time of an ANN that has 41 features as inputs and for exactly the same ANN that uses 25 features

Table II

DESCRIPTION OF THE SELECTED AND THE CORRESPONDED EXCLUDED FEATURES OF THE UNSW-NB15 DATASET BASED ON THE VALUE OF COEFFICIENT OF CORRELATION. IF THE VALUE OF COEFFICIENTS OF CORRELATION FOR TWO FEATURES IS CLOSE TO 1, THEN THE FEATURES ARE HIGHLY CORRELATED AND THE EXCLUDED FEATURES CAN BE REPLACED BY THE SELECTED.

Selected features	**Excluded features**	ρ
Source to destination packet	Source to destination transaction bytes	0.9637
	Source packets retransmitted or dropped	0.9711
Destination to source packet	Destination to source transaction bytes	0.9719
	Destination packets retransmitted or dropped	0.9786
Source interpacket arrival time (mSec)	If source and destination IP addresses equal and port numbers are equal or not	0.9413
Source TCP window advertisement value	Destination TCP window advertisement value	0.9901
TCP connection setup time, time between the syn and the synack	TCP connection setup roundtrip time, the sum of synack and ackdat	0.9495
No of connections of the same source address and destination port	No. of connections that contain the same service and source address	0.8660
	No. of connections of the same destination address.	0.9603
	No of connections of the same destination address and the source port	0.9068
	No of connections of the same source and the destination address	0.8699
	No. of connections of the same source address	0.8974
	No. of connections that contain the same service and destination address	0.8685

as inputs. The feature reduction proves that the elapsed time is reduced as the number of training samples increases from 30000 to 80000 samples (Figure 2).

C. Anomaly detection - Threat identification

The selected features of the UNSW-NB15 dataset consist the inputs to develop an ANN model for the network anomaly detection. This model treats like a filter that discriminates the abnormal from the normal incidents. The ANN model consists of seven hidden layers 52, 26, 16, 14, 4, 2, 2 nodes respectively. The input layer has 25 nodes, the output layer that leads to the binary classification consists of one node and the model has been trained for 45 epochs. The accuracy of the ANN model is 81.9%, the precision and the recall are 83.9% and 90.78% respectively.

Based on the threat identification process, that is described in the second part of the proposed methodology, we assume only the abnormal incidents. Then nine ANN models were developed separately from each other, so that each model concerns only a specific type of attack. The selection of the number of layers, the number of nodes and the number of epochs obtained as a result of many experiments. We carried out different combinations of layers, nodes and epochs and the architecture of models that presented Table III are the ones that show the best performance in terms of accuracy and precision.

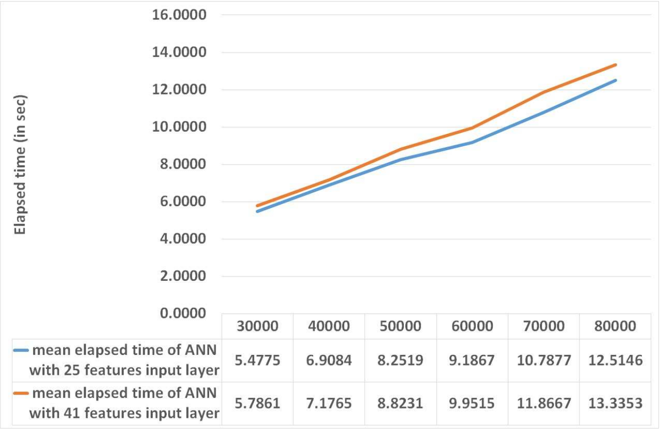

Figure 2. Comparison of the mean elapsed time (in sec) between the ANN that has 41 features as inputs and for the same ANN that uses 25 features as inputs. This figure depicts the reduction of the elapsed time while the training samples ranges from 30000 to 80000.

The Table III, provides the number of layers and the corresponded nodes and the number of epochs for each model.

Table III

DESCRIPTION OF THE ARCHITECTURE AND THE VALUES THAT MEASURES THE EFFICIENCY OF THE ANN MODELS WHICH DEVELOPED SEPARATELY FOR EACH TYPE OF ATTACK. THE DESCRIPTION OF THE ARCHITECTURE CONTAINS THE NUMBER OF LAYERS AND THE CORRESPONDED NUMBER OF NODES AND THE NUMBER OF EPOCHS.

Type of attack	Number of layers	Number of nodes per Layer	Epochs
Analysis	6	(25-300-100-50-25-1)	81
Backdoors	8	(25-75-50-25-15-10-5-1)	500
DoS	3	(25-1000-1)	1000
Exploits	4	(25-600-300-1)	30
Fuzzers	9	(25-175-125-100-50-20-10-5-1)	26
Generic	7	(25-50-25-20-10-2-1)	150
Recon.	5	(25-1000-550-550-1)	125
Shellcode	6	(25-500-500-500-400-1)	250
Worms	5	(25-1000-500-500-1)	200

The calculated precision and accuracy diverse among the different ANN models due to the existence of imbalance among the classes of the dataset. The ANN model associated to the DoS attack has the highest precision 100% and on the other hand the ANN model associated to the Worm attack has the lower precision 25% due to lack of a large number of Worm incidents. The The ANN model associated to the DoS attack has also the highest accuracy 90.90% and on the other hand the ANN model associated to the Shellcode attack has the lower accuracy 51.25% due to lack of a large number of Worm incidents. The results of the accuracy and precision for each ANN model depicted in Table IV.

The Roc curves described in Figure 3 illustrates the ability of the detection for each model. The Area Under Curve (AUC) for each type of attack is also a performance measurement for classification problem and denotes the capability of the models

Table IV
DESCRIPTION OF THE ACCURACY AND THE PRECISION FOR THE ANN MODELS WHICH DEVELOPED SEPARATELY FOR EACH TYPE OF ATTACK.

Attack	Accuracy	Precision
Analysis	63.12%	58.33%
Backdoors	64.57%	60.88%
DoS	90.90%	100.0%
Exploits	71.79%	50.04%
Fuzzers	81.37%	89.16%
Generic	73.11%	61.14%
Recon.	59.27%	41.03%
Shellcode	51.25%	51.20%
Worms	87.83%	25.00%

to distinguish the incidents between classes. A value of the AUC close to 1 denotes that the performance of the specific model regarding the classification procedure is very good. Calculating the AUC for each type of threats the Analysis type of attack have 0.653, the Backdoors 0.639, the DoS 0.553, the Exploits 0.698, the Fuzzers 0.861, the Generic 0.762, the Reconnaissance 0.511, the Shellcode 0.536 and the AUC for Worms is 0.797.

V. COMPARATIVE EVALUATION

This section contains the comparative evaluation of the experimental results. The comparison concerns the results of the proposed method the results of the methods that are already available in the literature. It takes into account mainly the methods that based on the ANN models for the multiclassification procedure but it is also exceeded to other classifiers such Decision Tree (DT), Random Forest (RF) and AdaBoost (AB) classifiers. We assume for this procedure the studies that validate their methodology with the UNSW-NB15 dataset as well.

Comparing the methods that use ANN-based models we conclude that the proposed method is superior because it can detect all the types of attack. Even if there are some attacks that are detected with low precision it set the basis to indentify all the type of attacks in modern communication networks. The multiclassification methods that have been used so far divided into those methods that use ANN-based models and to the methods that use other classifiers such DT, the RF and the AB classifiers. Catac [23], Tchakouht et al., [21] and Baig et al., [22] proposed multiclassification methods that use ANN-based models. The results of their methodsprove that is very difficult to detect all the type of attacks especially those that they rarely appear such as Analysis, Backdoors, Shellcodes and Worms due to lack of information that associated to them. In the table V are presented the comparison of the precision and recall between the experimental results and the results that arisen from the literature [23], [21], [22]. Since the precision and recall have been used as the measures of efficiency from the state of the art methods we calculate the same measures to obtain a sufficient comparison. The highest values of the precision or recall is denoted as bold in the table V for each type of attack.

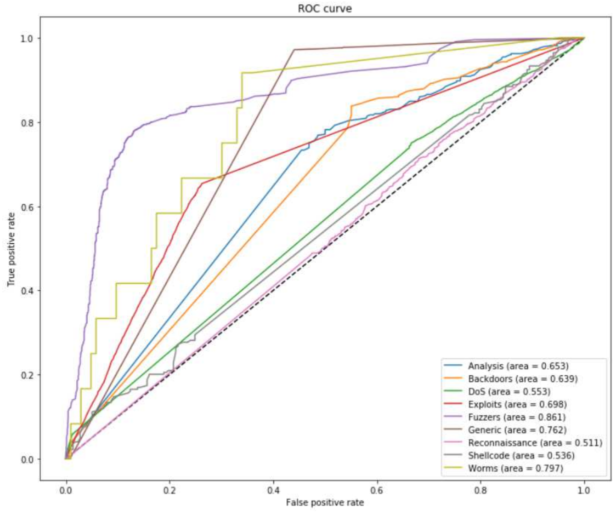

Figure 3. The Roc curve per each type of attack illustrates the ability of the detection for each ANN model. The Area Under Curve denotes the capability of the each model to distinguish the incidents between the classes of the model

Table V
COMPARISON OF THE EXPERIMENTAL RESULTS IN TERMS OF PRECISION (AND RECALL) THAT OBTAIN FROM THE PROPOSED METHOD WITH THE EXPERIMENTAL RESULTS THAT OBTAIN FROM THE STATE OF THE ART METHODS AND BASED ON ARTIFICIAL NEURAL NETWORKS (ANN), DECISION TREE (DT), RANDOM FOREST (RF) AND ADABOOST (AB) CLASSIFIERS

Classification per attack	Proposed method (%)	ANN [23] (%)	ANN [21] (%)	ANN [22] (%)	DT [23] (%)	RF [23] (%)	AB [23] (%)
Analysis	**58.33** (**75.47**)	NA (NA)	0.21 (0.14)	0.0 (0.0)	NA (NA)	NA (NA)	NA (NA)
Backdoors	**60.88** (**83.76**)	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆
DoS	**100.0** (0.15)	27 (6.0)	30 (**49**)	6.25 (0.02)	**100** (0.0)⋆	**100** (0.0)⋆	32 (**49**)
Exploits	50.04 (61.40)	62 (84)	**65** (69)	31.58 (56.93)	54 (**94**)	54 (**94**)	59 (62)
Fuzzers	**89.16** (76.04)	5.0 (63)	33 (75)	0.0⋆ (0.0)⋆	77 **(79)**	77 **(79)**	78 (52)
Generic	61.14 (97.15)	49 (39)	99 (96)	90.81 **(97.81)**	**100** (97)	**100** (97)	98 (97)
Recon.	41.03 **(97.42)**	7.0 (0.0)⋆	20 (0.028)	40.45 (33.74)	**91** (60)	**91** (60)	65 (74)
Shellcode	**51.20** (**85.27**)	0.0⋆ (0.0)⋆	5 (16)	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	26 (5.0)
Worms	25 **(8.33)**	0.0⋆ (0.0)⋆	**50** (0.02)	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆	0.0⋆ (0.0)⋆
Classification binary	83.9 (90.78)	75 (2.0)	NA (NA)	86.74 (93.3)	92 **(99)**	**98** (98)	95 (96)

⋆ based on the state of the art [21], [22], [23].

Based on the state of the art results the values for the precision and recall per type of attack for the ANN model that is proposed by Catac, [23] are 27% and 6% for the DoS attacks, 62% and 84% for the Exploits, 5% and 63% for the Fuzzers, 49% and 39% for the Generic and 7% and 0% for the Reconnaissance respectively. The Analysis, the Backdoors, the Shellcodes and the Worms type of attacks are not detected from the model proposed by Catac.

Similar are also the results that obtain from the model of Tchakoucht et al., [21]. The values of the precision and recall per type of attack of the ANN model are 0.2% and 0.14% for the Analysis attacks 30% and 49% for the DoS, 65% and 69% for the Exploits, 33% and 75% for the Fuzzers, 99% and 96% for the Generic, 20% and 0.02% for the Reconnaissance, 5% and 16% for the Shellcodes and 50% and 0.02% for the Worms respectively. The Backdoors are not detected by the aforementioned model.

Finally, the outcomes obtain from the model proposed by Baig et al., [22]. The values of the precision and recall per type of attack of the ANN model are 6.25% and 0.02% for the DoS attacks, 31.58% and 56.93% for the Exploits, 90.1% and 97.81% for the Generic, 33.74% and 40.45% for the Reconnaissance respectively. The Analysis, the Backdoors, the Fuzzers, the Shellcodes and the Worms are not detected from the ANN model proposed by Baig et al.

The comparison of the experimental results proves that the proposed method is superior in terms of the classification to all type of attacks (Figure 4.). More specifically, the ANN-based model proposed by Catak, [23] can not detect the Analysis, the Backdoors, the Reconnaissaince, the Shellcode and the Worms type of attacks. The ANN-basd model proposed by Tchakoucht et al.,[21] can not detect the Backdoors type of attack and the ANN-based model proposed by Baig et al., [22] can not detect the Analysis, the Backdoor, the Fuzzers, the Shellcode and the Worms type of attack.The values of the precision for the Exploits, Generic and Reconnaisance are lower in the proposed method in comparison with other methods. In current research the primary aim is to provide a method that can detect all type of attacks, that is the reason that more measures like F1-score that provide information regarding the balance between the precision and recall are omitted.

The existence of imbalanced classes is an issue that affects not only the usage of ANN models but the usage of other classifiers as well. The DT, the RF and the AB are not also detect attacks that they rarely appear [23].

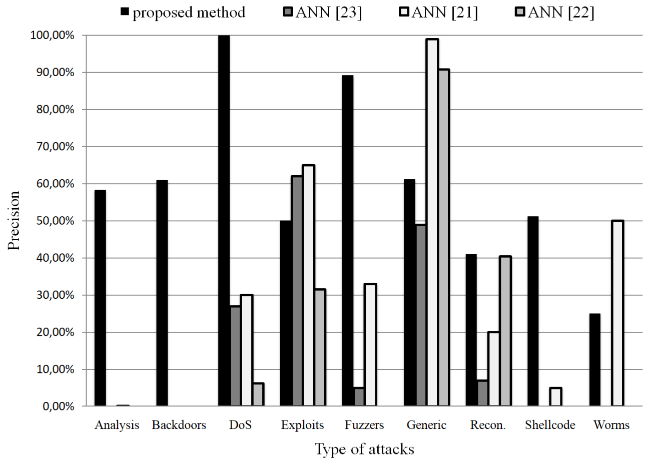

Figure 4. Comparison of the precision between the proposed and the State of the Art methods that use ANN-based models and the UNSW-NB15 dataset

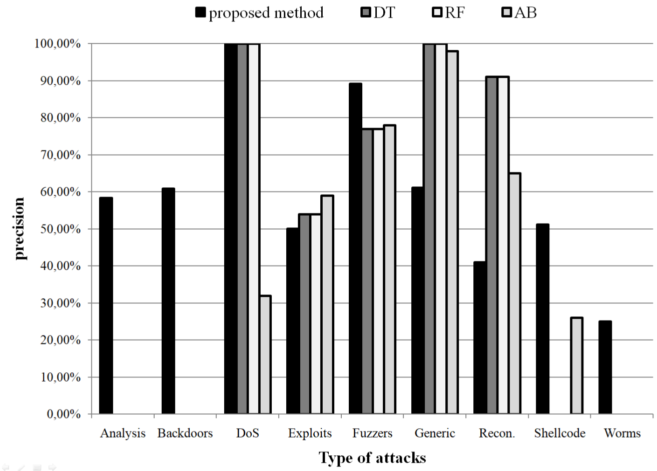

Figure 5. Comparison of the precision between the proposed method and the State of the Art methods that based on the Decision Tree, the Random forest and the Ada boost classifier

Comparing the experimental results of the proposed method with the experimental results that use the same dataset in the state of the art and based on the DT, the RF and the AB classifier the proposed method is superior in terms of classification per attack as well (Figure 5.). The DT and the RF classifiers are not detect the Analysis, the Backdoors, the Shellcodes and the Worms type of attacks. The AB classifier can not detect the Analysis, the Backdoors and the Worms type of attack. Finally in terms of of binary classification, the RF classifier detects with higher precision and recall the abnormal incidents from the normal. The advantage of the RF classifier is not exceed in case of the classification per attack since it can not detect all type of threats in case of abnormal incidents.

More specific, the values of the precision and recall per type of attack of using a DT classifier as proposed by Catac, [23] are 100% and 0% for the DoS attacks, 54% and 94% for the Exploits, 77% and 79% for the Fuzzers, 100% and 97% for the Generic and 91% and 60% for the Reconnaissance respectively. The Analysis, the Backdoors, the Shellcodes and the Worms are not detected from the model that obtained from the DT classifier. The same exactly are the results that obtain from the usage of the RF classifier the values of the precision and recall per type of attack are 100% and 0% for the DoS attacks, 54% and 94% for the Exploits, 77% and 79% for the Fuzzers, 100% and 97% for the Generic and 91% and 60% for the Reconnaissance and 26% for the Shellcodes. The Analysis, the Backdoors and the Worms are not detected from the model that obtained from the RF classifier [23]. The AB classifier leads to similar results as well since the Analysis, Backdoors and Worms type of attacks are not detected from the model that obtained from the AB classifier [23].

VI. Conclusions

In this study, we developed a method for anomaly detection and for the threat classification in modern communication networks. The proposed method concerns a two stages procedure that uses ANN-based models. In the first part of the method we proposed an effective process that assumes the coefficient of correlation as the measure to reduce the number of features that demanded in the training procedure of the ANN model. Comparing the mean elapsed time before and after the feature extraction process, we noticed that there is significant reduction. The reduction of the mean elapsed time validates the effectiveness of the feature extraction process. The feature extraction part of the methodology sets the first step for the analysis of the large volume of data that expected in modern communication networks such as IoT and 5G networks. An important contribution of the proposed method cited in the second part of the methodology, where we proposed the usage of ANN models for the network anomaly detection and to identify the type of threats for the corresponded attacks. The main contribution of this part is that it set the basis to identify all the type of attacks. The proposed method deals effectively with the problem of the large volume of data and uses the ANN model that provide high computational speed to modern communication networks. We have also verified the above results based on the UNSW-NB15 dataset that contains modern type of attacks. In the future research, our aim is to improve the performance of the proposed models in terms of the precision and the recall and to develop with high accuracy detection that would have the inability to identify unknown type of attacks as well.

VII. Acknowledgments

This work has been partially supported by the European Commission through project 5G-MoNArch funded by the European Union Horizon 2020 program under Grant Agreement n^o 761445 and through project H2020-FORTIKA "cybersecurity Accelerator for trusted SMEs IT Ecosystem" under Grant Agreement n^o 740690.

The opinions expressed in this paper are those of authors and do not necessarily reflect the views of the European Commission.

References

[1] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila and A. Gurtov, "5G security: Analysis of threats and solutions", In Standards for Communications and Networking (CSCN), September 2017, IEEE Conference, pp.193–199.

[2] A. Gupta, R. K. Jha and S. Jain, "Attack modeling and intrusion detection system for 5G wireless communication network", International Journal of Communication Systems, 2017, 30(10), e3237.

[3] V. Hadjioannou, C. X. Mavromoustakis, G. Mastorakis, J. M. Batalla, I. Kopanakis, E. Perakakis and S. Panagiotakis, "Security in smart grids and smart spaces for smooth IoT deployment in 5G", Internet of Things (IoT) in 5G Mobile Technologies, Springer, 2016, pp.371–397.

[4] R. T. Tiburski, L. A. Amaral, and F. Hessel,"Security Challenges in 5G-Based IoT Middleware Systems", Internet of Things (IoT) in 5G Mobile Technologies, Springer, 2016, pp.399–418.

[5] K. Gai, M. Qiu, L. Tao and Y. Zhu, "Intrusion detection techniques for mobile cloud computing in heterogeneous 5G", Security and Communication Networks, 2016, 9(16), pp.3049–3058.

[6] K. Tai-hoon, C. Ramos and S. Mohammed, "Smart city and IoT", Future generation computer systems ,Elsevier, 2017, pp.159–162.

[7] S. K. Wagh, V. K. Pachghare, and S. R. Kolhe, "Survey on intrusion detection system using machine learning techniques", International Journal of Computer Applications, 2013, 78(16).

[8] S. Omar, A. Ngadi, and H. H. Jebur, "Machine learning techniques for anomaly detection: an overview", International Journal of Computer Applications, 2013, 79(2).

[9] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen and L. Hanzo, " Machine learning paradigms for next-generation wireless networks", IEEE Wireless Communications, 2017, 24(2), pp.98–105.

[10] L. F. Maimó, F. J. G. Clemente, M. G. Pérez, and G. M. Pérez, "On the Performance of a Deep Learning-Based Anomaly Detection System for 5G Networks", 2017 IEEE 3rd Smart World Congress, August 2017, pp. 303–310.

[11] J. Suomalainen, K. Ahola, M. Majanen, O. Mämmelä and P. Ruuska, "Security Awareness in Software-Defined Multi-Domain 5G Networks", Future Internet, 2018, 10(3), 27.

[12] L. F. Maimó, A. L. P. Gómez, F. J. G. Clemente, and M. G. Pérez, "A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks",IEEE Access, 2018, 6, 7700-7712.

[13] J. Santos, P. Leroux, T. Wauters, B. Volckaert, and F. De Turck, "Anomaly detection for Smart City applications over 5G low power wide area networks", In NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium, April 2018, pp. 1–9.

[14] L. F. Maimó, A. H. Celdrán, M. G. Pérez, F. J. G. Clemente, and G. M. Pérez, "Dynamic management of a deep learning-based anomaly detection system for 5G networks", Journal of Ambient Intelligence and Humanized Computing, 2018, pp. 1–15.

[15] J. Li, Z. Zhao, and R. Li. "A Machine Learning Based Intrusion Detection System for Software Defined 5G Network", IET Networks, 2017, 7, no. 2,pp. 53–60.

[16] D. Fang, Y. Qian, and R. Q. Hu. "Security for 5G Mobile Wireless Networks", IEEE Access, 2018, (6), pp. 4850–4874.

[17] P. Schneider, and G. Horn. "Towards 5G security", In Trustcom/BigDataSE/ISPA, August 2015, IEEE, pp. 1165–1170.

[18] N. Moustafa, and J. Slay. "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set", Information Security Journal: A Global Perspective, 2016, (25), pp. 18–31.

[19] A. S. Sohal, R. Sandhu, S. K. Sood, and V. Chang, "A cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments." Computers and Security, 2018, 74, pp. 340–354.

[20] M. Nawir, A. Amir, O. B Lynn, N. Yaakob, and R. B. Ahmad. "Performances of Machine Learning Algorithms for Binary Classification of Network Anomaly Detection System", In Journal of Physics: Conference Series, 2018, 1018 No(1). IOP Publishing.

[21] T. A. Tchakoucht and M. Ezziyyani. "Multilayered Echo-State Machine: A Novel architecture for efficient intrusion detection", IEEE Access, 2018 6, 72458-72468.

[22] M. M. Baig, M. M. Awais, and E. S. M El-Alfy. "A multiclass cascade of artificial neural network for network intrusion detection", Journal of Intelligent and Fuzzy Systems, 2017, 32(4), pp, 2875-2883.

[23] F. Catak, "Two-layer malicious network flow detection system with sparse linear model based feature selection", Journal of the National Science Foundation of Sri Lanka, 2018, 46(4).

[24] D. M. Tax, and R. P. Duin, "Using two-class classifiers for multiclass classification", Object recognition supported by user interaction for service robots, IEEE, 2002, Vol. 2, pp. 124–127.

[25] C. Bekara, "Security issues and challenges for the IoT-based smart grid." Procedia Computer Science, 34, (2014).pp 532-537.

[26] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K.J. Kim. "A survey of deep learning-based network anomaly detection." Cluster Computing, 2019.

[27] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson. "Shallow and deep networks intrusion detection system: A taxonomy and survey." arXiv preprint arXiv:1701.02145. (2017).

[28] C. Larsson. "Design of modern communication networks: methods and applications." Academic Press.(2014).

[29] B. B. Zarpelao, R. S. Miani, C. T. Kawakani and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things." Journal of Network and Computer Applications, 84, (2017) pp. 25–37.