IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 30, 2020, accepted August 13, 2020, date of publication August 17, 2020, date of current version September 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017045

Physical Layer Detection of Malicious Relays in LTE-A Network Using Unsupervised Learning

YELIZ YENGI[1], ADNAN KAVAK[1], AND HÜSEYIN ARSLAN[2], (Fellow, IEEE)
[1]Department of Computer Engineering, Kocaeli University, 41950 İzmit, Turkey
[2]Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA

Corresponding author: Yeliz Yengi (yelizyengi@gmail.com)

ABSTRACT For Long Term Evolution Advanced (LTE-A) network, although there exist many studies that focus on improving the performance with relays, security issues are often neglected. Due to the broadcast nature of wireless channels, relay nodes in LTE-A network may act maliciously, affect communication, reduce quality, and cause delays. Recently, physical (PHY) layer security has attracted researchers to provide secure communication and data privacy. In this study, we propose using unsupervised learning approach at the destination node to detect malicious relay attacks in cooperative LTE-A network based on received source signal in the PHY layer. Outlier detection algorithms such as one class support vector machine (OCSVM), local outlier factor (LOF) and isolation forest (*i* Forest) are applied to detect various malicious relay behaviors such as garbling, regenerative, and false data injection type attacks. As input to these algorithms, feature vectors are constructed by using amplitude, phase, and relative phase information of modulated baseband symbols. The performance of the outlier detectors are evaluated with respect to precision, accuracy, and under the area curve (AUC) measures under changing signal-to-noise ratio (SNR) levels, different modulation types, allocated number of resource blocks (RBs), and varying data size. The results demonstrate the effectiveness of our proposed outlier detection approach for detecting malicious relays in the LTE-A network. Accuracy and precision of the algorithms are observed to be above 90% for 10 dB and larger SNR levels for the relay attack scenarios considered here. AUC values for all algorithms for all SNR levels is also above 0.9 for the attack detection cases, and the performance of the LOF algorithm with 0.95 and above AUC values is superior to other algorithms. The results verify the contribution of this study, which is the demonstration of the effectiveness of one class outlier detection approaches for detecting malicious relays in the LTE-A network.

INDEX TERMS Relay communication, physical layer security, outlier detection, one-class classification (OCC), unsupervised learning.

I. INTRODUCTION

LTE-A standard provides increased coverage and lower delays in cellular systems that use higher carrier frequencies at the same transmission power and higher bandwidths [1]–[4]. LTE-A supports deployment of femtocells, picocells, relays, and remote radio heads in a macrocell arrangement to provide wider coverage and higher capacity at low cost to target areas with low power nodes [5]–[7].

As cell edge performance becomes more critical, LTE-A relay technology becomes an eminent solution with low cost and high performance. LTE relays are easy to set up for their applications to improve network capacity. The main benefits of the deployment of relay nodes are to provide enhanced coverage in LTE-A targeted areas for efficient heterogeneous network and to avoid interference in surrounding areas at low cost. However, although many outstanding studies have been done on ways to improve relay performance in the LTE-A network, security issues are often neglected.

Due to the broadcast nature of the wireless channels, even though the diversity received from the relays is high quality, these relays may act maliciously by garbling the message transmitted from source, injecting false data, or regenerating new signal [8]. Therefore, cooperative systems are naturally vulnerable to these types of attacks. Recently, PHY layer security has attracted researchers to provide secure communication and data privacy. Before cooperating with a relay,

The associate editor coordinating the review of this manuscript and approving it for publication was Adnan M. Abu-Mahfouz.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

it may be possible to prevent potential attacks and possible costs by detecting whether the relay node is secure or not.

In [9]–[12], malicious node detection studies for relay network are based on statistical approaches by distinguishing the two signals with a very similar distribution using threshold values for their detectable or undetectable states. In [12], attack detection problems have been evaluated as statistical learning problems with channel state information (CSI) for different attack scenarios where measurements are observed collectively or online [12].

Studies in [13]–[15] have applied supervised machine learning technique with statistical CSI to provide security in relay network, in which artificial neural networks and decision trees techniques have been used with past data on the network.

Under the category of the PHY security with supervised and unsupervised learning in relay network, it has been seen that supervised algorithms are widely used [16]. Although supervised learning is effective method, it is a costly approach due to need for balanced data sets and it can detect only trained categories [17]–[20].

In [17], [18], relay signals have been categorized into two known classes as reliable and unreliable by using the features of I/Q constellations such as amplitude and phase in supervised learning algorithms. In [19], identity fraud has been detected through reliable channel features with a semi-supervised learning method. Anomaly detection has been studied in [20] using clustering and semi-supervised algorithms with well-known data sets, i.e., categorizing the samples on predefined data and model. Outlier detection is one of the most important approach in all data-driven scientific disciplines. Data that does not match expected behavior often has some interesting features and can help them better understand the problem at hand [21], [22].

In this study, we focus on detecting malicious relay attacks in cooperative LTE-A network in PHY layer via unsupervised approach of machine learning with a more inclusive and powerful way than the statistical and supervised machine learning algorithms. We consider outlier detection as an unsupervised learning method.

In our study, one-class support vectors (OCSVM) [23], local outlier factor (LOF) [24] and Isolation Forest (*i* Forest) [25] are used as outlier detector algorithms to detect anomalies and unreliable behaviors of relays using secure data samples from the source.

The reason for choosing these three algorithms is that they had been shown to be effective even with a small number of samples in the training phase as compared to other detection methods. Another reason is that these algorithms are representative for different categories of outlier detection methods, i.e., LOF is distance-based, *i* Forest is isolation based, and OCSVM is domain-based outlier detection methods. This allows us to show the contribution of outlier detection algorithms to the problem solution [22]. Conventional outlier detection approaches utilize the statistics of the received signals and may detect the low-density signals as an outlier. Therefore, it may not be possible to detect malicious relays as outliers, which have signal statistics that are indistinguishable as to be discussed in Section III-A.

Although supervised learning methods may achieve successful results with a balanced training dataset, it may not be feasible to always predict the attack models and find a sufficient number of examples for training the algorithms in real-time applications. Application of supervised learning approach to our problem definition is not feasible, since malicious relays must be detected as their signals are received at the destination node before further processing for diversity purposes. Therefore, in this article, we have compared performances of the aforementioned unsupervised learning approaches for detecting malicious relay attacks.

The contributions of this article are as follows:

1) We show that unsupervised learning algorithms such as OCSVM, LOF, and *i* Forest which represent broad category of outlier detection approaches are capable of detecting malicious relay attacks with high accuracy and precision in physical layer in the LTE-A network.

2) We propose feature vectors as inputs to above algorithms based on baseband modulated signal characteristics in the OFDM system and show the effectiveness of this feature selection for the problem definition in LTE-A relay network.

We explore the effects of different factors on the attack detection performance of OCSVM, LOF, and *i* Forest algorithms such as modulation type, attack type, and signal bandwidth.

The remainder of the article is as follows: We explain the system model and problem definition in Section II. In Section III, we define the feature set and explain the algorithms adapted to our problem. Section IV presents the experimental setups and performance evaluation is given in Section V. Finally, the conclusion is given in Section VI.

II. SYSTEM MODEL

In this section, we discuss the system model along with a problem definition of cooperative communication technology in LTE-A. In LTE Rel. 10, it has been specified that user equipment (UE) or mobile station can communicate to eNB via a relay node which is referred to as LTE-A donor eNB. Radio relay stations can be deployed on in different architectures in the network as being standardized in 3GPP. Potentially useful relay deployment scenarios are summarized in Table 1 as discussed in 3GPP for LTE Rel. 10 [26]. Relay technology supports one-hop for fixed position and mobile relay stations provide a feasible solution to expand the coverage area in all scenarios.

In this study, we consider deployment of Layer 1 relay nodes in downlink of LTE-A network, which is known as amplify and forward (AF) type of radio relay technology as illustrated in Figure 1. Layer 1 relay implementation is simple and has short proccessing delays that makes it relatively low-cost when compared to other layer relay implementations.

The aim of Layer 1 relay is to increase the coverage in montainous, urban and indoor areas and also to increase data

TABLE 1. Relay technology deployment scenarios.

Scenario	Deployment
Urban hot spot	Expanding coverage and increasing efficiency in urban areas with high traffic density.
Indoor hot spot	To expand the coverage of the indoor environment and to improve the efficiency.
Group mobility	To reduce handover and location-registration control signals by install the relay stations in public vehicles.
Rural area	Use to expand coverage in mountainous or sparsely populated areas.

rate in highly populated center of urban areas. However, there is a disadvantage of this implementation which is amplifying the noise as well as the desired signal. The relay communication consists of two consecutive time periods. During time period 1 (1st hop), macro base station (eNB) broadcasts the signal, and UE and relay node (RN) receive. Then, during time period 2 (2nd hop), RN amplifies the received signal from eNB and then transmits it to the UE.

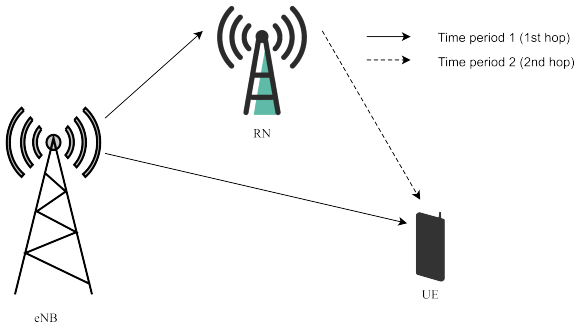


FIGURE 1. Cooperative transmission in LTE-A network with Layer 1 relay model.

In LTE-A downlink channel, the input data stream is carried by the Orthogonal Frequency Division Multiplexing (OFDM) symbols over multiple frequency subcarriers. The block diagram of OFDM signal for LTE-A downlink data transmission channel and our attack detection approach is illustrated in Figure 2. The input data stream is converted from serial to parallel through a specific modulation scheme such as Quadrature Phase Shift Keying (QPSK), Sixteen Quadrature Amplitude Modulation (16-QAM), and 64-QAM. The modulation scheme depends on the physical channel mapped on the resource grid. The complex representation of a modulated symbol is carried by the signal $\mathbf{x_i}$ which is on i^{th} subcarrier. Consequently, the modulated symbols are mapped over the resource elements (RE) of an LTE-A RB. According to LTE-A standards, one RE consists of a subcarrier of 15 kHz in the frequency domain and 66.7 μ s of time duration in the time domain. Moreover, one RB consists of 84 REs which means 12-subcarriers and 7-symbols in frequency and time domain, respectively. Mathematically, it can be represented as,

$$x(t) = \sum_{i=1}^{P} x_i e^{j2\pi i \frac{B}{N} t} \tag{1}$$

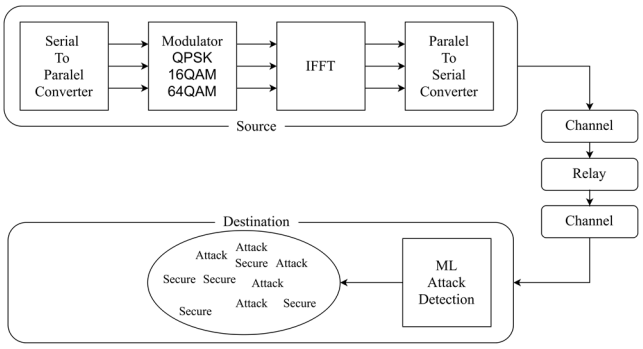


FIGURE 2. Data flow of attack detection.

where P is the number of RBs assigned to the user. The number of RBs depends on bandwidth. For instance, if we have 10 MHz bandwidth, 9 MHz is used for data transmission and 1 MHz is reserved for guard band. At the transmitter side the Inverse Fast Fourier Transform (IFFT) is applied to convert the parallel frequency-domain symbols into serial samples to generate a composite time-domain signal. The composite transmitted signal $x(t)$ is further divided into samples. To prevent the inter symbol interference (ISI), we add a guard band called as cyclic prefix (CP) which is a copy of the tail of a symbol tail.

A. PROBLEM DEFINITION

LTE-A network downlink transmission model is considered in Figure 3, where S is the eNB, R is the RN, and D is the UE when referred to Figure 1. In this system model, R is assumed to be Layer 1 relay which implements AF relaying protocol. The source node (S) communicates with the destination node (D) via direct transmission and a relay node (R).

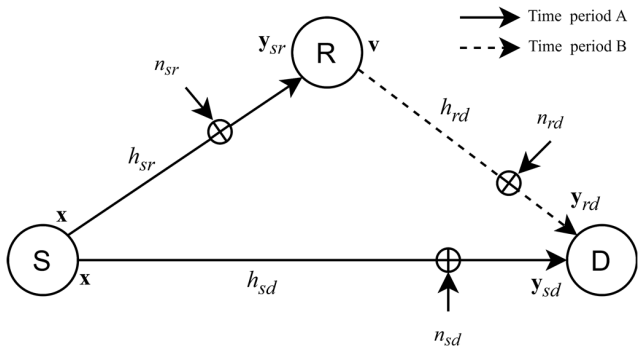


FIGURE 3. System model.

The source transmits an OFDM signal $x(t)$ within a period T1 that is equal to half subframe in LTE-A standard. There are total of N transmitted symbols from S with a half subframe, i.e. $\mathbf{x}=[x[1]\,x[2]\ldots x[N]]^T$. During the transmission time period 1, the received signal vectors at R and D are given by,

$$\mathbf{y}_{sr} = h_{sr}\mathbf{x} + \mathbf{n}_{sr}, \tag{2}$$

$$\mathbf{y}_{sd} = h_{sd}\mathbf{x} + \mathbf{n}_{sd}. \tag{3}$$

During T2, the relay node R amplifies the received signal $\mathbf{y}_{sr}$ and then retransmits it to D as given by,

$$\mathbf{v} = \alpha \mathbf{y}_{sr}, \tag{4}$$

$$\alpha = \frac{\rho}{P_{sr}\sigma_{sr}^2 + n_{sr}[i]}, i = 1, 2, \ldots, N, \tag{5}$$

$$\mathbf{y}_{rd} = h_{rd}\mathbf{v} + \mathbf{n}_{rd}. \tag{6}$$

In above equations, the vectors $\mathbf{y}$ and $\mathbf{n}$ are constructed from time samples, which can be written as,

$$\mathbf{y} = [y[1]\, y[2] \ldots y[N]]^T, \tag{7}$$

$$\mathbf{n} = [n[1]\, n[2] \ldots n[N]]^T. \tag{8}$$

where $\rho = 1$as relay power, P_{sr} is the received signal power at the relay node, and σ_{sr}^2 is the variance of S-R link channel. In Equation (2) through (6), $\mathbf{n}_{sr}$, $\mathbf{n}_{sd}$ and $\mathbf{n}_{rd}$ are i.i.d additive white Gaussian (AWGN) noise signals with variance σ^2 in the S-R, S-D and R-D links, respectively. The complex channel coefficients are denoted by h_{sr}, h_{sd} and h_{rd} in the S-R, S-D and R-D links, respectively, which include the effects of path loss and fading in the wireless channel. We assume the wireless channels over S-R, S-D and R-D links are assumed to be unchanged during the half subframe period.

In the conventional cooperative network, signals received at the source node and the relay node are combined using different diversity techniques at the receiving node to improve signal quality. However, in some cases, the relay node may manipulate the received signal from the source and intentionally send false information to the destination node. And, this relay is referred to as malicious relay. Cooperative combining techniques ignore the malicious behavior of the relay nodes, and just try to use the diversity provided by them. Given that the signal transmission from Source S to destination D is received correctly and reliably, the critical issue is to determine whether the transmission from the relay is secure before combining its signal for diversity. The aim of this study is decided to detect whether the relay is secure under various types of relay attacks. In the LTE-A relay communication, the possible attack types that a relay may create in the physical layer are discussed. Due to the nature of the communication, it is not possible to determine all the malicious behaviors that may arise from the relay, and we have considered possible attack scenarios that may occur in the physical layer. Therefore, three different malicious relay attacks are considered as described below [8].

B. GARBLING ATTACK MODEL (A1)

The malicious relay node generates a new random permutation of the modulated symbols in the constellation. The index of original symbols **y,**i.e., $\mathbf{y}_{sr}[1] \ldots \mathbf{y}_{sr}[N]$ will be repositioned by a random permutation function P,

$$\tilde{\mathbf{y}}_{sr} = \mathcal{P}(\mathbf{y}_{sr}), \tag{9}$$

$\mathcal{P}$ generates a sequence by taking a uniformly distributed random index between *1* to N in the symbol constellation.

C. REGENERATIVE ATTACK MODEL (A2)

The relay node intentionally ignores the original signal x and replaces it with newly created signal $\tilde{x}$,

$$\tilde{\mathbf{y}}_{sr} = \tilde{x}. \tag{10}$$

The new signal generated at the R has the same modulation properties of the original signal **x**.

D. FALSE DATA INJECTION ATTACK MODEL (A3)

In this attack type, the R deliberately injects a random signal vector **a**to the received signal **x** from the source before transmitting it the D as given by,

$$\tilde{\mathbf{v}} = \alpha \mathbf{y}_{sr} + \mathbf{a}, \tag{11}$$

$$\tilde{\mathbf{y}}_{rd} = h_{rd}\tilde{\mathbf{v}} + \mathbf{n}_{rd}. \tag{12}$$

III. ATTACK DETECTION USING MACHINE LEARNING METHODS

In this section, we discuss how relay attack detection can be achieved through machine learning techniques with unsupervised learning methods. Attack detection for unsupervised learning algorithms is also called as outlier detection in literature [22]. Unsupervised learning identifies outlying data by using the score of each unlabeled nominal signal sample whereas supervised learning needs labeled samples. In supervised learning, training set $Z = \{(\mathbf{z}_i, l_i)\}$ tuples where class labels l_i with secure or attack signal models and sample $\mathbf{z}_i$ can be used to construct model to predict label l_i' for the new sample $\mathbf{z}_i'$. Supervised learning methods can achieve more successful results with a balanced training dataset, however in real world applications, it may not be feasible to always predict the attack models and find a sufficient number of examples for training the algorithms. For this reason, due to nature of our problem definition to detect malicious relays, it will be faster and more independent to detect the signals received from the relay with unsupervised learning. When imbalanced class distribution for attack models is considered, this requirement is particularly costly affecting efficiency of supervised learning algorithms [27]. However, unsupervised learning does not require balanced and labeled data sets. In addition, outlier detection techniques may be more suited for the cases with difficulties in modeling attacks. Constructing the decision function with only nominal unlabeled samples significantly affect the effectiveness of the system against unknown attack models.

A. STATISTICAL PROPERTIES OF THE RECEIVED SIGNAL

We first examine the statistical distributions of received signal at D. A sample illustrative probability density functions (PDFs) of received signal amplitudes $|y_{sd}|$ and $|y_{rd}|$ for different relay attack types, secure relay, and source signals are shown in Figure 4. Complementary cumulative distribution functions (CDFs) of amplitude and phase of y_{sd} and y_{rd} are also given in Figure 5. Statistical detection approaches are based on conditional PDFs $f_{y_{rd}|y_{sd}}(y_{rd}[i], y_{sd}[i])$ of the received signal samples by determining threshold value on

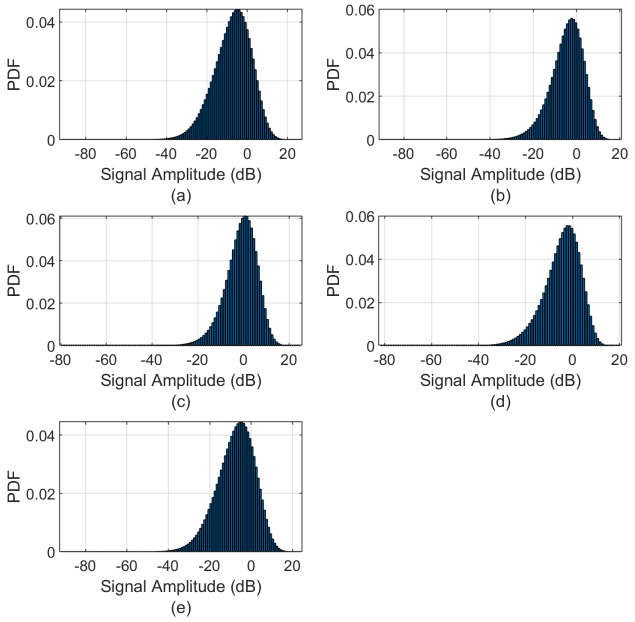


FIGURE 4. PDF's of received signal amplitudes for (a) R-D link with A1 attack, (b) R-D link with A2 attack, (c) R-D link with A3 attack, (d) S-D link with source signal, and (e) R-D link with secure relay signal.

the resultant distribution to define the normality in the data space [12].

The opposite perpective of CDF's is the complementary CDFs which show signal values above a particular level called as exceedance probability. As seen in Figure 5, CCDFs of signal amplitudes have close exceedence probabilities and are indistinguishable to be used in statistical attack detection. On the other hand, there are some differences in exceedence probabilities of signal phases.

Conventional outlier detection approaches utilize the statistics of the received signals, and may detect the low density signals as an outlier. Therefore, it is not possible to detect malicious relays as outliers, which have signal statistics that are indistinguishable as shown in Figures 4 and 5. In order to detect such relays, this study explores the feasibility of applying ML algorithms which need properly defined features of signal characteristics. For this reason, we seek for new approaches with multiple features and robust solution for attack detection problem [29]. In the next section, we define the signal featues and propose using outlier detection algorithms to provide possible solution to such a problem. Presented detection models have many benefits when we compare these models with the traditional probabilistic detection models. With our proposed detection model, we can create more complex decision functions and search hidden relation between consecutive symbols.

B. FEATURE EXTRACTION

Let $y_{n,\eta\to D}$ be the n^{th} sample of complex baseband received symbol over the $\eta \to D$ link at destination D in Figure 3, where $\eta \in \{S, R\}$. Transmitted OFDM signals from the source S are passed through the wireless channel, and therefore $y_{n,\eta\to D}$ can be treated as a modified OFDM signal. OFDM symbols at S are based upon digitally modulated symbols such as QPSK, 16-QAM, 64-QAM in LTE-A network. These symbols can be distinguished by their amplitude and phases in the signal constellation. Also, phase differences or relative phases of the symbols provide information about modulated symbol patterns.

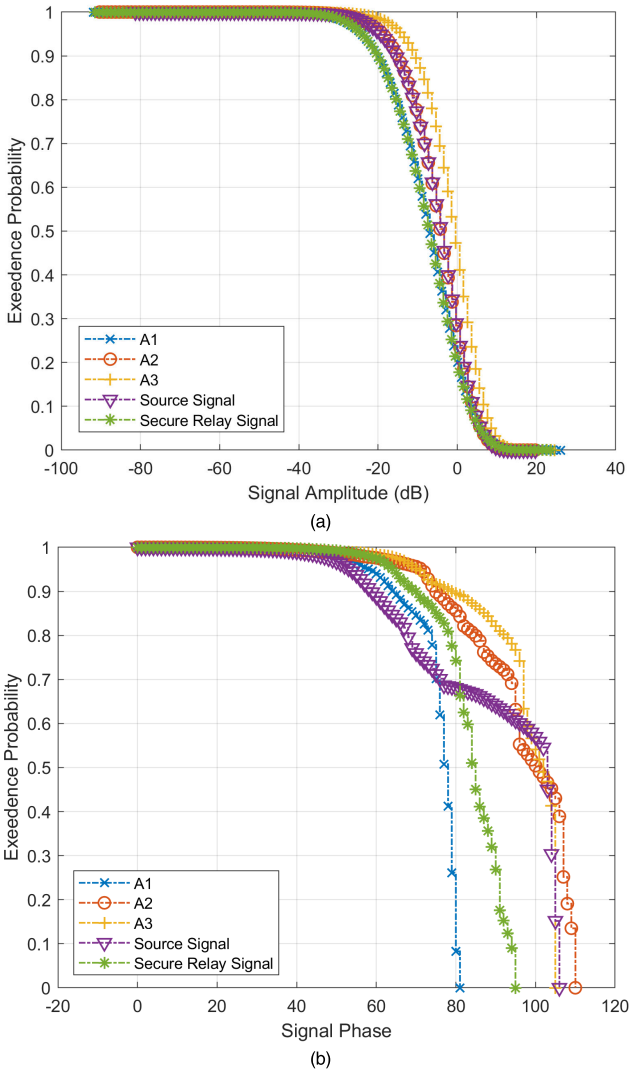


FIGURE 5. CCDF's of amplitude (a) and phase (b) feature of signal received at destination for all links respectively, R-D A1 attack, R-D A2 attack, R-D A3 attack, S-D source signal and R-D secure relay signal.

These modulated symbols can be defined as points in the constellation with different amplitudes and phases. Therefore, these amplitude and phase information or relative phases of $y_{n,\eta\to D}$ may provide some useful tool in capturing any unordinary effect on this signal. For this reason, we define following features as inputs to the outlier detection algorithms to detect relay attacks:

$$f^{(1)}_{n,\eta\to D} = |y_{n,\eta\to D}|, \tag{13}$$

$$f^{(2)}_{n,\eta\to D} = \angle y_{n,\eta\to D}, \tag{14}$$

$$f^{(3)}_{n,\eta\to D} = \angle y_{n,\eta\to D} - \angle y_{n+1,\eta\to D}, \tag{15}$$

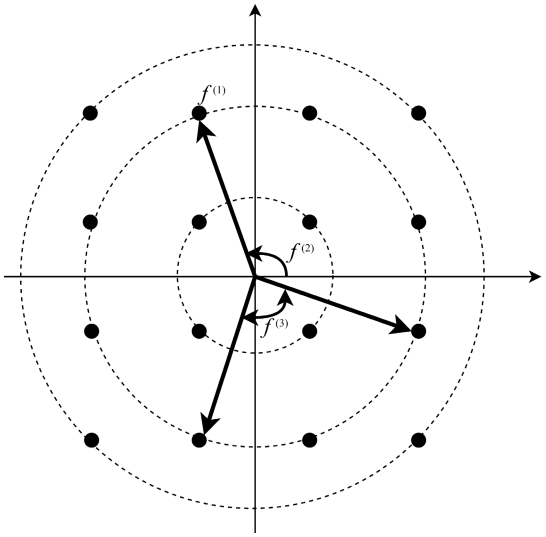


FIGURE 6. Illustration of the features for the received signal $y_{n,\eta\to D}$ with 16-QAM modulation.

where $|(.)|$ and $\angle(.)$ denote the amplitude and phase operators. The features defined in above equations are depicted in Figure 6 for 16-QAM modulation.

The feature vector of symbol $y_{n,\eta\to D}$ is then given by,

$$\mathbf{f}_n = \left[f^{(1)}_{n,\eta\to D} f^{(2)}_{n,\eta\to D} f^{(3)}_{n,\eta\to D}\right]. \tag{16}$$

For the m^{th} half subframe, i.e., slot, we have N consecutive symbols during the learning phase, and therefore feature vector can be written as,

$$\mathbf{z}_m = [\mathbf{f}_1 \mathbf{f}_2 \ldots \mathbf{f}_N]^T \, m = 1, 2, \ldots, M. \tag{17}$$

Since there are M half subframes, the dataset matrix in the learning phase is constructed by merging the feature vectors, which is given by,

$$\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \ldots \mathbf{z}_M], \tag{18}$$

where the size of $\mathbf{Z}$ is $L \times M$ and $L=3N$.

C. ONE CLASS SUPPORT VECTOR MACHINE

OCSVM introduced by Scholkopf *et al.* [23] is special version of SVM algorithm to outlier detection. The hypersphere or decision function in OCSVM is constructed for a small region wherein positive targeting samples are labeled as $+1$. The decision function or boundary is a smallest hypersphere which takes into account most of the training data which is mapped by a kernel function. However, it allows to locate data points outside of the boundary by including slack variables ξi, $i = 1, \ldots, M$.

The number of data points located outside of the boundary is calculated by a penalty factor of $1/vM$. Choosing a small penalty factor will cause more data points to be located outside of the boundary. Figure 7 illustrates a non-linear mapping of original data samples with radial basis function (RBF) kernel in OCSVM [20]. Data samples are separated from the

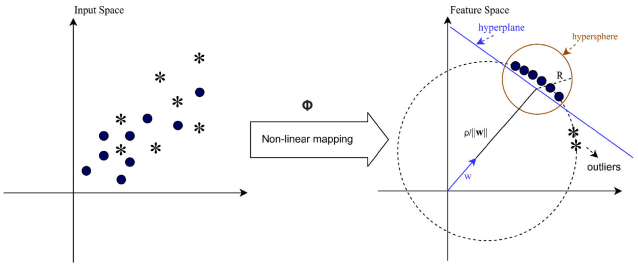


FIGURE 7. One Class SVM with RBF Kernel.

origin with a maximum margin kernel function until finding the smallest hypersphere that contains all positive samples.

OCSVM optimization model for our problem definition is formulated as follows:

$$\begin{aligned} &\min_{w,\xi_m,\rho} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{vM}\sum_{m=1}^{M}\xi_m - \rho \\ &\text{subject to } \mathbf{w}^T\phi(\mathbf{z}_m) \geq \rho - \xi_m, \quad \xi_m \geq 0, \\ &\qquad m = 1, 2, \ldots, M \end{aligned} \tag{19}$$

where ϕ is the non-linear mapping function (kernel) of the feature vectors in the training set, $\mathbf{w}$ represents the weight vector in the model, ξ_m is the slack variable for regulating weights, ρ represents the maximum deviation from the limit, and $v \in (0,1]$ specifies an upper limit for abnormal values while determining the lower limits for the support vectors. Distance and center of the data samples $\mathbf{z}_m$ are controlled by ξ_m in the optimization model.

In the test phase of the model, we try to determine whether sample vector $\mathbf{z}_j$ falls outside of the hyper plane via a decision function given by,

$$\begin{aligned} g_{OCSVM}(\mathbf{z}_j) &= sgn\left(\sum_{m=1}^{M}\mu_m K(\mathbf{z}_m, \mathbf{z}_j) - \rho\right), \\ 0 = \mu_m &< \frac{1}{vM} \end{aligned} \tag{20}$$

where μ is the Lagrange multiplier and $K(\mathbf{z}_m, \mathbf{z}_j)$ is RBF given by,

$$K(\mathbf{z}_m, \mathbf{z}_j) = e^{-\gamma\|\mathbf{z}_m - \mathbf{z}_j\|^2}, \tag{21}$$

where $||.||$ is Euclidian norm opetrator and γ is the "spread" parameter of kernel to set the length of bell-shaped curve.

D. LOCAL OUTLIER FACTOR

LOF method estimates the local density deviation of a given data sample with respect to its neighbors. LOF is a score-based method in which samples whose densities are significantly different than their neighbors are marked as outliers, i.e., LOF ≈ 1 no outlier, LOF $\gg 1$ outlier. Local density cells are formed by reaching out of the samples until capturing k neighbor samples in training data set. The decision function in LOF is given by [24],

$$g_{LOF}(\mathbf{z}_j) = \frac{k/M}{v_k \left\|\mathbf{z}_m - \mathbf{z}_j\right\|}, \tag{22}$$

where v_k is the average of local density cells, $\mathbf{z}_m \in \mathbf{Z}$, $\mathbf{z}_j \in \mathbf{Z}$ are the data vectors in (18). The LOF searches for the neighbors of a point to find its density and then compare it with the density of the other points. While a small k has a more local focus by looking at nearby points, a large k can miss local outliers. If there is too much noise in the data, inaccurate results can be obtained with small k.

E. ISOLATION FOREST

The i Forest algorithm converts the dataset into the subtree ensembles. The i Forest algorithm recursively isolates random individual instances by creating binary tree structures. The average path lengths of these random trees are normality measurement values in the decision function. For our problem, the score of the i Forest algorithm is given by [25],

$$g_{iForest}\left(\mathbf{z}_j, \mathbf{Z}\right) = 2^{-\frac{E(h(\mathbf{z}_j))}{c(\mathbf{Z})}}, \tag{23}$$

where E denotes average of $h(\mathbf{z}_j)$ which is the path length, and $c(\mathbf{Z})$ is the average path lengths of given $\mathbf{Z}$ for all data set. The data vector $\mathbf{z}_j$ with the score greater than 0.5 is marked as an outlier. The most important parameters of this algorithm are the adjustment of a subtree number. If the subtree number is too high, the model overfits and the number of false alarms will increase. However, if chosen large number of subtrees that will decrease the performance of execution therefore, there must be a compromise between performance and high execution time. Therefore, subtree parameter is chosen according to peak performance approaches.

IV. SIMULATION SETUP

Simulations are performed to evaluate the performance of algorithms. We investigate the performance of the proposed attack detection scheme by constructing the signal model in LTE-Advance system. Signal model is generated in MATLAB and used as an input to outlier detection algorithms which are run using ''Scikit Learn'' machine learning library [30].

After creating signal model in (2) through (8), we extract features from consecutive symbols as described by (13) through (18). During the training phase of the algorithms, data set is generated for 5 ms duration (half frame period) in LTE-A system model. Attack scenarios such as garbling (A1), regenerative (A2), and false data injection (A3) type attacks are created as described in Section 2. Monte Carlo simulations are performed 20 times for the same signal setup conditions (i.e., SNR, modulation type, bandwidth) and the same algorithm selected. In each run, data set is obtained from randomly generated information bits and channel conditions. Properties of generated data set and algorithm parameters are summarized in Table 2. In outlier detection algorithms, parameters have been selected to prevent overfitting problems in OCSVM, LOF, and i Forest algorithms. For OCSVM, we employ RBF kernel and the parameter v is set to 0.3, which is an upper bound on the fraction of margin errors and a lower bound of the fraction of support vectors relative to the total number of training examples. For LOF the parameter k is set to 60, which is used by neighbors' queries. The number of subtrees is chosen as 20 in i Forest setup. System simulations for LTE-A downlink signal have been performed for 1.4, 3, 5, 10, 15, and 20 MHz bandwidths which correspond to 6, 15, 25, 50, 75, 100 resource blocks, respectively.

TABLE 2. Dataset and outlier detection algorithm properties.

Bandwidth (MHz)	1.4	3	5	10	15	20
Resource Block (RB)	6	15	25	50	75	100
Training instances	6	15	25	50	75	100
Test instances secure relay	6	15	25	50	75	100
Test instances A1 attack relay	6	15	25	50	75	100
Test instances A2 attack relay	6	15	25	50	75	100
Test instances A3 attack relay	6	15	25	50	75	100
Preprocessing	none					
Data type of instances	± Double, non zero					
Training instances min, max, mean	-74.7578, 8.9066, 1.1501					
Test instances min, max, mean	-90.7578, 0.9066, 1.0420					
Data Size (number of symbols) per slot (N)	RB x 84 (Resource Block x Resource Element for 0.5 ms)					
Training data period	5 ms (Half Frame Period)					
Number of slots during training period (M)	10					
Feature vector size (L)	3N					
Training data set size	LM					
Number of OFDM subcarriers	128					
Cyclic prefix length for 1st symbol	10					
Cyclic prefix length for remaining	9					
Fading channel type	Rayleigh					
Modulation type	QPSK, 16QAM, 64QAM					
SNR values	5-30 dB					
OCSVM properties	$v = 0.3$ Kernel = Radial basis (RBF)					
LOF properties	k = 60					
*i*Forest properties	Base estimator = 20					

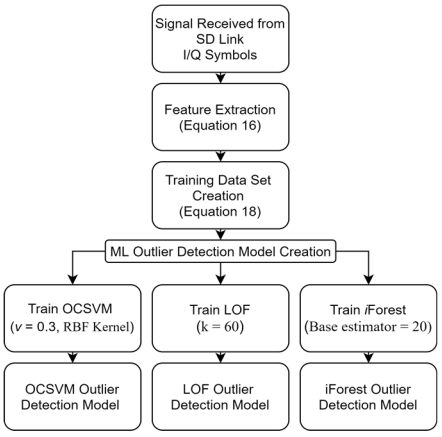


FIGURE 8. Flow chart for detection system for training phase.

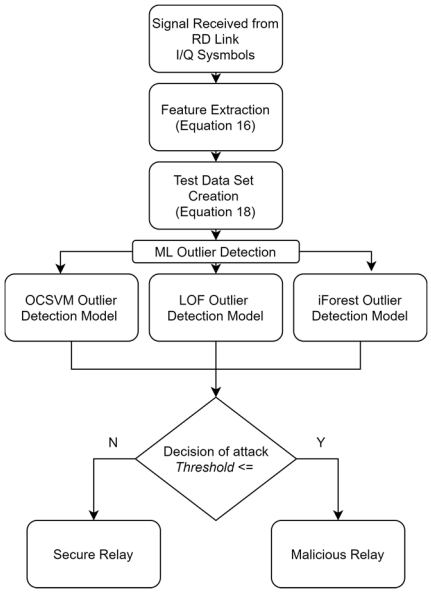


FIGURE 9. Flow chart for detection system for test phase.

TABLE 3. Confusion matrix.

		True Condition	
		Attack	Secure
Predicted Condition	Classified Attack	TP	FP
	Classified Secure	FN	TN

V. PERFORMANCE EVALUTION

For the performance evaluation of outlier detection algorithms in our problem, we do not only concern the detection of relay attacks but also try to determine the secure relay cases [31]. We consider *precision*, *accuracy*, and *AUC* as the performance evaluation metrics. Confusion matrix for attack and secure relay cases is defined in Table 3.

Precision is the proportion of correctly classified attack cases out of all detected attack cases, which is given by,

$$precision = \frac{TP}{TP + FP}. \tag{24}$$

In our problem, detection of attacked relay case is considered as outlier detection. On the other hand, *accuracy* measures truely classified attacks and secure cases among all cases,

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{25}$$

Confusion matrix can also be used as evaluation metrics to measure the diagnostic ability of one class classification such as receiver operating characteristic (ROC) curve. The ROC curves can be derived from CDF of true positive rate (*TPR*) with respect to false positive rate (*FPR*). An ROC curve summarizes all of the confusion matrices that each threshold produces and shows *TPR* vs. *FPR* at different classification thresholds. By lowering the classification threshold, more items can be classified as positive, thus increasing both false positives and true positives. *TPR* and *FPR* are defined as following,

$$TPR = \frac{TP}{TP + FN}, \tag{26}$$

$$FPR = \frac{FP}{TN + FP}. \tag{27}$$

TPR is also referred to as recall. The *AUC* defines the entire two-dimensional area underneath the entire ROC curve, i.e. it provides an aggregate measure of performance across all possible classification thresholds. The statistical meaning of *AUC* is the probability that the model ranks a random positive sample more highly than a random negative sample. It also provides measure of how well predictions are ranked, rather than their absolute values. *AUC* can help us decide which classification method is better.

Below, we present the results regarding the performances of aforementioned OCSVM, LOF, and iForest algorithms in terms of *precision, accuracy* and *AUC* measures with regard to SNR level, modulation type and data size (M).

A. EFFECT OF ATTACK TYPES, MODULATION TYPES, AND ALLOCATED RESOURCE

When the performance of the algorithms is examined regarding attack types, the positive effect of increase in SNR level is observed. Even under the low SNR condition, it can be seen that all algorithms achieve at least 90% *precision* level (Fig.12, 14, and 16). The *precision* of the algorithms is more significant for the secure relay detection case when compared to the attack detection case. From Figure 10, we observe that *precision* level for *i* Forest varies between 0.4 and 0.8 while that of OCSVM and LOF varies between 0.5 and 1.

In secure relay detection with varying SNR level, both LOF and OCSVM exhibits nearly the same *precision* performance, and *precision* performance of *i* Forest is lower than these algorithms. Also, the effect of modulation types on *precision* level of the algorithms for secure relay detection is insignificant. We see that 70% *precision* level can be attained at nearly 10 dB SNR level in OCSVM and LOF, and 22 dB SNR level in *i* Forest.

Precision represents the performance of detecting only relay with attacks (true positive). On the other hand, *accuracy* provides measure about the *accuracy* of detecting both relay attacks and secure relays (true positive and true negative). Hence, in order to asses the overall detection performance, we have also obtained *accuracy* of the algorithms.

The *accuracy* in secure relay detection in Figure 11 are approximately close to the *precision* in Figure 10. However, *accuracy* of iForest algorithm is worse than its *precision*, i.e., at 20 dB SNR level its *precision* and *accuracy* values are about 70% and 60%, respectively. Overall, the effect of SNR variation on *accuracy* and *precision* performance is remarkable in secure relay detection.

The performance results of the algorithms incase of relay attack are presented in Figures 12 through 17. Comparing *precision* and *accuracy* results of Garbling attack model (A1)

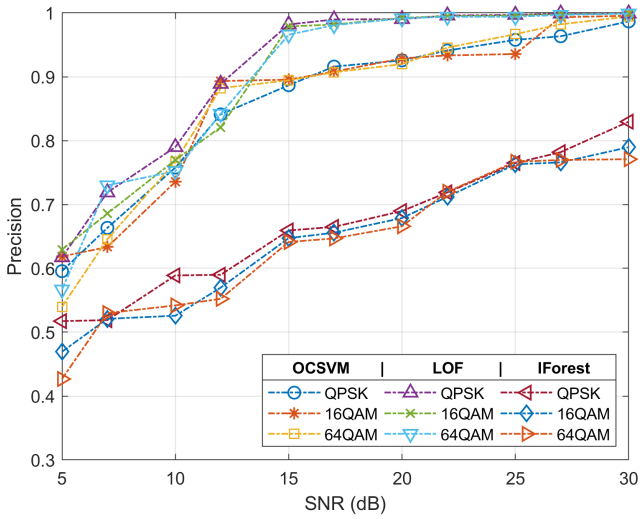


FIGURE 10. Precision vs. SNR performance of the algorithms under different modulation types when the relay has secure relay scenario.

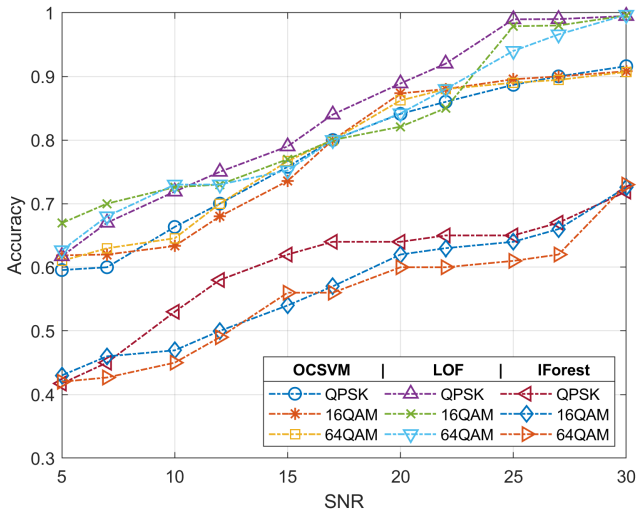


FIGURE 11. Accuracy vs. SNR performance of the algorithms under different modulation types when the relay has secure relay scenario.

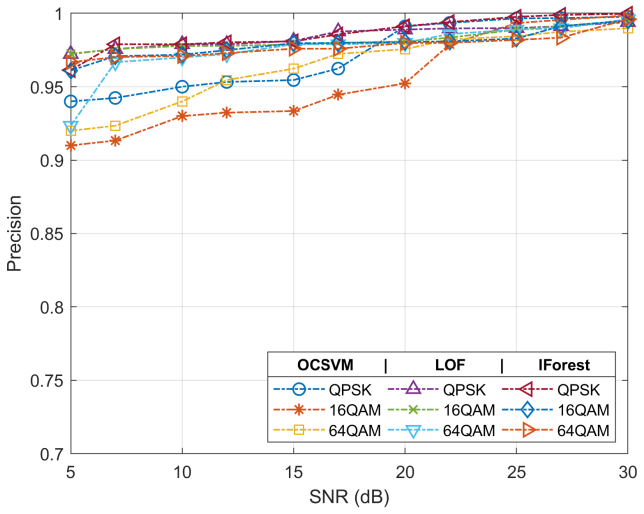


FIGURE 12. Precision vs. SNR performance of the algorithms under different modulation types when the relay has Garbling Attack Model (A1).

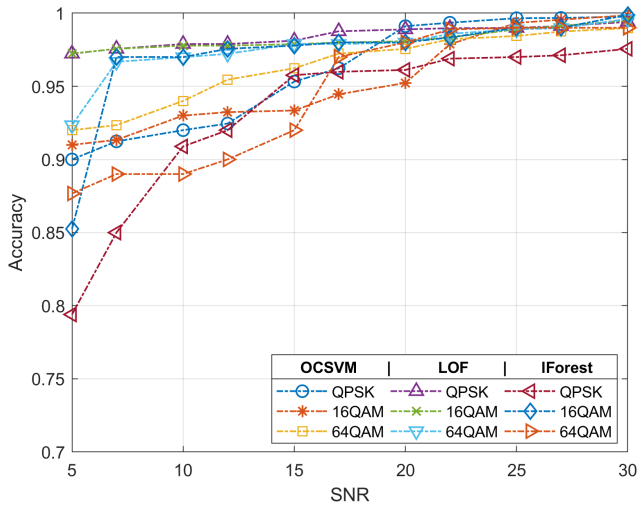


FIGURE 13. Accuracy vs. SNR performance of the algorithms under different modulation types when the relay has Garbling Attack Model (A1).

presented in Figure 12 and Figure 13, it is seen that *accuracy* is lower than its *precision* especially under low SNR conditions (10 dB and below) for all algorithms. This is because under low SNR conditions secure relay may be evaluated as false alarm and this affects detection *accuracy*. However, even with these false alarms, 80% and above *accuracy* is achieved. Considering only *precision*, over 90% performance can be achieved in all algorithms for A1 attack model.

While A1 and A2 attack type detection OCSVM more effected until 15dB SNR level than LOF and iForest with less precise performance. With regard to detection of regenerative attack (A2), as shown in Figure 14, while LOF and *i* Forest algorithms have *precision* above 95% for all SNR values, OCSVM can achieve this with SNR of approximately greater than 20 dB. When *accuracy* of the algorithms is examined in detection of regenerative (A2) attack given in Figure 15, LOF algorithm performs above 95% for all SNR values, whereas OCSVM and *i* Forest can achieve the same performance with SNR values greater than 20 dB.

For the detection of false data injection attack (A3), *precision* and *accuracy* of the algorithms are presented in Figure 16 and Figure 17, respectively. All algorithms have *precision* and *accuracy* performance above 95% and 90%, respectively. In general, all algorithms in A1, A2, and A3 attack detection show more than 90% *precision* performance. In terms of *accuracy* performance, it is seen that *i* Forest algorithm is affected by low SNR level, i.e., 10 dB and below, especially for detecting A1 and A2 attacks.

The performance of attack detection for LOF is able to predict a very accurate density than OCSVM and *i* Forest with sufficiently high neighbor numbers to detect high number of outliers. A high accurate is also observed for OCSVM by mapping high-dimensional hypersphere. However, *i* Forest has good average performance even in high dimensionality

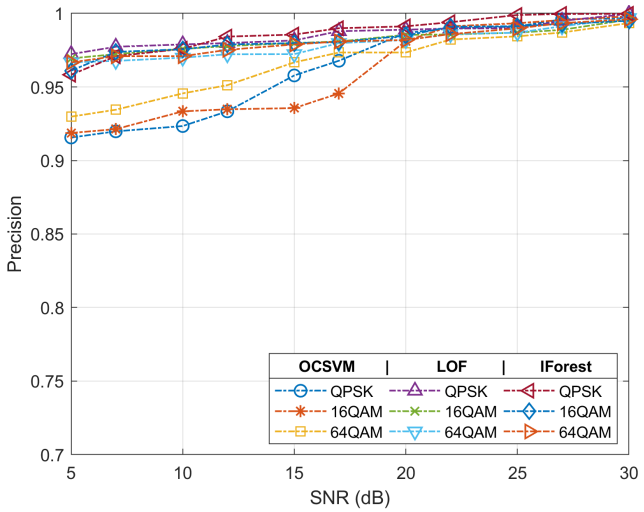


FIGURE 14. Precision vs. SNR performance of the algorithms under different modulation types when the relay has Regenerative Attack Model (A2).

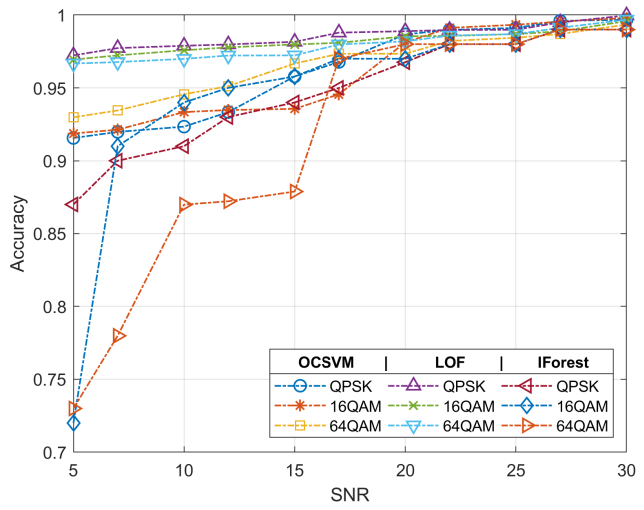


FIGURE 15. Accuracy vs. SNR performance of the algorithms under different modulation types when the relay has Regenerative Attack Model (A2).

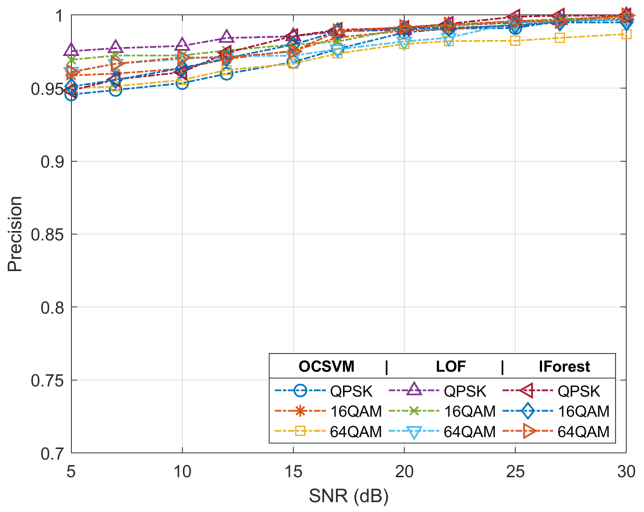


FIGURE 16. Precision vs. SNR performance of the algorithms under different modulation types when the relay has False Data Injection Attack Model (A3).

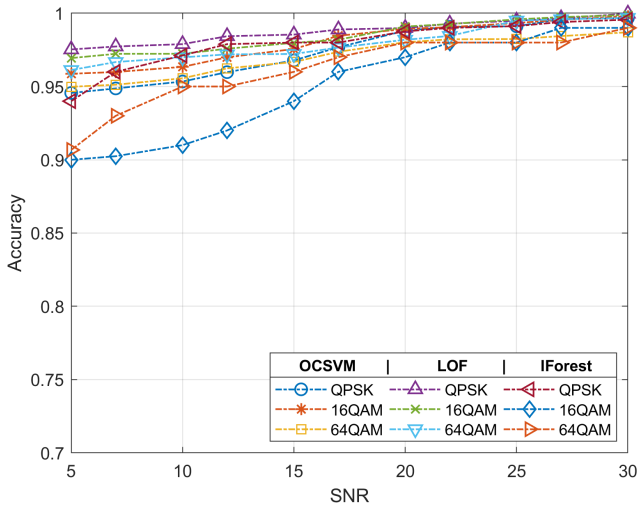


FIGURE 17. Accuracy vs. SNR performance of the algorithms under different modulation types when the relay has False Data Injection Attack Model (A3).

than LOF and OCSVM but less accurate due to high false alarm rate because of overfitting problem. As result of secure relay detection, OCSVM is a good candidate with small amount of data performance without requiring significant tuning in our benchmark.

When the effect of modulation types (QPSK, 16QAM, 64QAM) on the performance of algorithms are examined, it is seen that *accuracy* is not affected considerably for the same attack type as indicated in Table 5. This situation can be evaluated as an indicator of the robustness of the attack detection model we proposed in this study.

In the presented results, the parameters of the algorithms specified in Table 2 were used. However large values of the parameters v, k, and base estimator causes overfitting problem and therefore increases the false alarm rate for all three algorithms and decreases attack detection accuracy. It is of great importance that the parameters used in one-class outlier detection algorithms are to be selected at the point where they result in the most accurate results.

Another parameter used to evaluate the performance of the system is the data size M. The results for the effect of data size is summarized in Table 6. No remarkable effect of data size on attack detection is observed. The data size is especially important in secure relay detection case. As M is increased from 6 to 100, *accuracy* of the algorithms increases especially for secure relay detection case.

AUC performance of the algorithms is presented in Figure 16. AUC of all algorithms is above 0.9 for the attack detection cases. For all attack types, the performance of the LOF algorithm with 0.95 and above AUC values is superior to other algorithms.

Regarding secure relay detection, when AUC values for 15 dB SNR level are examined, AUC values of LOF, OCSVM, and *i* Forest algorithms are 0.81, 0.78 and 0.58,

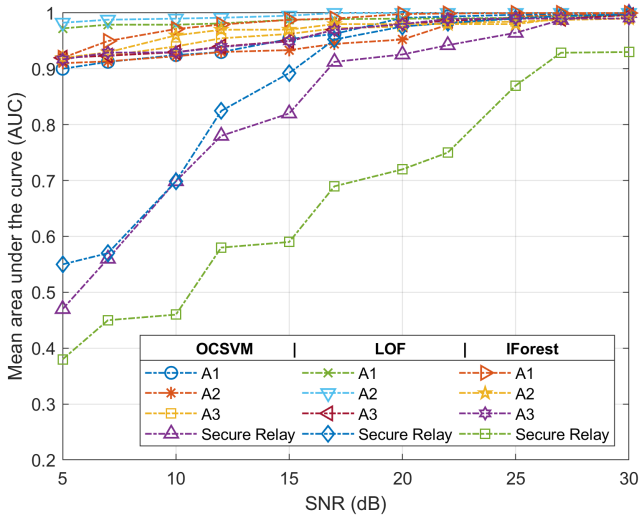


FIGURE 18. AUC performance comparison of the algorithms with different attack types (modulation = QPSK, Bandwidth = 1.4 MHz, M=6).

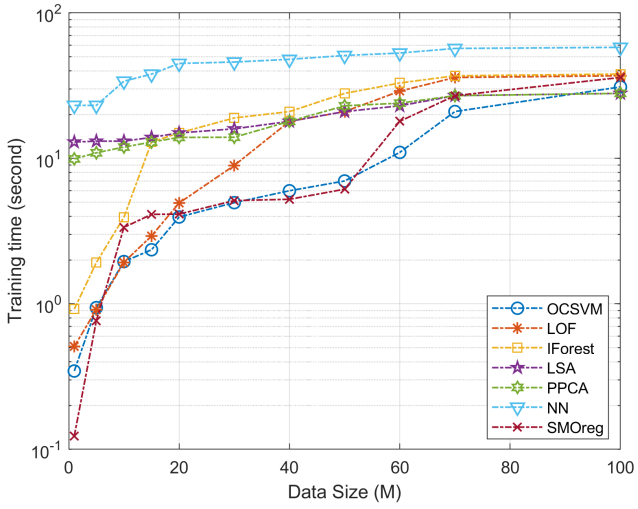


FIGURE 19. Training time of the algorithms (modulation = QPSK, SNR = 15, Bandwidth = 1.4 MHz).

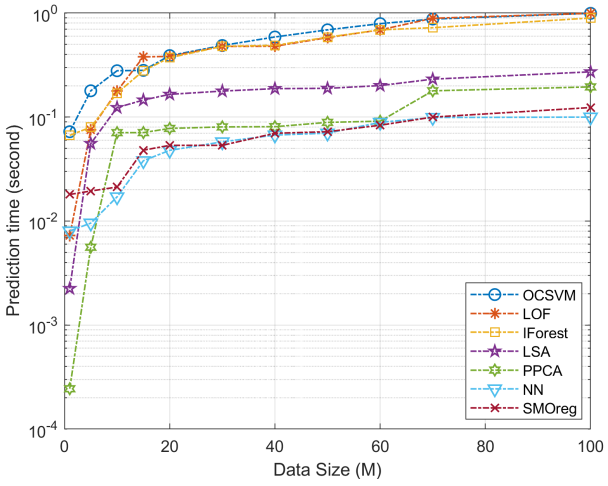


FIGURE 20. Prediction time of the algorithms (modulation = QPSK, SNR = 15, Bandwidth = 1.4 MHz).

respectively. The *i* Forest algorithm is not so successful in detecting secure relays. The LOF algorithm has the largest AUC performance and obtaining feasible results with the OCSVM algorithm indicates that these two algorithms can be used to detect malicious and secure relays even under low SNR levels around 10 dB.

In Table 4, performance of outlier detection methods are compared with conventional and supervised learning methods. Conventional methods that are Least-squares Anomaly detection (LSA) [32] and Probabilistic Principal Component Analysis (PPCA) [33] have failed to distinguish secure and attack samples with high false alarm rate. The accuracy values of these methods are 0.5 because they detect all secure relay cases as attack cases. As statistical properties are described before, the data set is not distinguishable with such conventional approaches. For malicious relay detection, accuracy performance of supervised learning methods that are Neural Networks (NN) [34] and Support Vector Machine for Regression (SMOreg) [35] with labeled and predefined dataset remains below that of unsupervised methods.

Based on the above presented results, two conclusive remarks can be made about malicious relay detection problem in LTE-A network. First observation is about the comparison of unsupervised learning or outlier detection methods among themselves. High accuracy is observed for OCSVM and LOF. However, due to overfitting problem that causes high false alarm rate, *i* Forest has average accuracy performance with high-dimensionality. LOF has higher accuracy and precision when compared to OCSVM and *i* Forest in detecting relay attacks but its performance depends on selecting sufficiently large number of neighbors. Thus, OCSVM is a good candidate for the malicious relay detection problem with small amount of data performance without requiring significant tuning in our benchmark. Second observation is about the performance comparison of studied unsupervised learning methods with supervised learning and conventional methods. When average detection accuracy of these methods are considered, LOF and OCSVM outperforms supervised learning (NN, SMOreg) and conventional methods. The reason for that is data sparsity in training phase which is an important challenge for the performance comparison of the proposed approach.

B. COMPLEXITY COMPARISON

Another performance assessment in the applications of algorithms is the complexity analysis. In this study, time and hardware needs are analyzed as complexity analysis of machine learning algorithms applied to our attack detection. The experiments are performed on Intel (R) Core (TM) i7-7500U, CPU 2.90 GHz and 8GB RAM hardware computer.

Training time against to changing data size (M) for all three algorithms applied in this study is obtained and plotted in Figure 19. As expected, larger data size causes larger training time, and OCSVM is observed as the fastest algorithm in terms of training time.

TABLE 4. Accuracy of the algorithms (SNR = 15, bandwidth = 1.4 MHZ, M = 6).

	OCSVM	LOF	*i*Forest	LSA	PPCA	NN	SMOreg
Secure Relay	0.75	0.78	0.62	0.5	0.5	0.62	0.72
Garbling Attack (A1)	0.95	0.98	0.96	0.5	0.5	0.73	0.82
Regenerative Attack (A2)	0.99	0.98	0.91	0.5	0.5	0.75	0.82
False Data Injection Attack (A3)	0.99	0.98	0.98	0.5	0.5	0.74	0.84

TABLE 5. Accuracy of the algorithms based on modulation type (SNR = 15, BANDWIDTH = 1.4 MHZ, M = 6).

	QPSK			16QAM			64QAM		
	OCSVM	LOF	*i*Forest	OCSVM	LOF	*i*Forest	OCSVM	LOF	*i*Forest
Secure Relay	0.75	0.78	0.62	0.82	0.89	0.57	0.84	0.88	0.56
Garbling Attack (A1)	0.95	0.98	0.96	0.93	0.98	0.97	0.97	0.98	0.92
Regenerative Attack (A2)	0.99	0.98	0.91	0.97	0.98	0.95	0.98	0.97	0.87
False Data Injection Attack (A3)	0.99	0.98	0.98	0.98	0.98	0.94	0.98	0.98	0.97

TABLE 6. Accuracy of the algorithms based on allocated resource blocks (SNR = 15, MODULATION = QPSK, M = [6, 25, 100]).

	6 (1.4 MHz)			25 (5MHz)			100 (20MHz)		
	OCSVM	LOF	*i*Forest	OCSVM	LOF	*i*Forest	OCSVM	LOF	*i*Forest
Secure Relay	0.75	0.78	0.62	0.88	0.92	0.68	0.91	0.99	0.73
GarblingAttack (A1)	0.95	0.98	0.96	0.95	0.98	0.96	0.96	0.99	0.96
Regenerative Attack (A2)	0.99	0.98	0.91	0.99	0.99	0.91	0.99	0.99	0.92
False Data Injection Attack (A3)	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.99	0.98

Time complexity of the algorithms is shown in Figure 18 during prediction phase in which the decision is made whether the test data is attack or secure. The prediction time cost is observed to be approximately the same for all three algorithms. The memory usage of the algorithms during the training phase is also analyzed with respect to varying data size and it is given in Figure 21. As the data size increases, the amount of memory used by the algorithms increases, but in the worst case, it does not exceed 1MB of memory space. OCSVM algorithm consumes the largest memory space. Since the memory usage of algorithms in the prediction phase is too small to be ignored, they are not shown in the plots.

Considering the memory consumption cost of the algorithms it is observed that applying these algorithms even in small capacity devices may be feasible. In the context of the practicability of the relay attack detection system, the memory and time requirements of the detection algorithms imply their applicability for small-capacity user equipment. As illustrated in Figures 19 and 21, training data size for 6 RB with 1.4 MHz bandwidth requires small memory size and time to train the model.

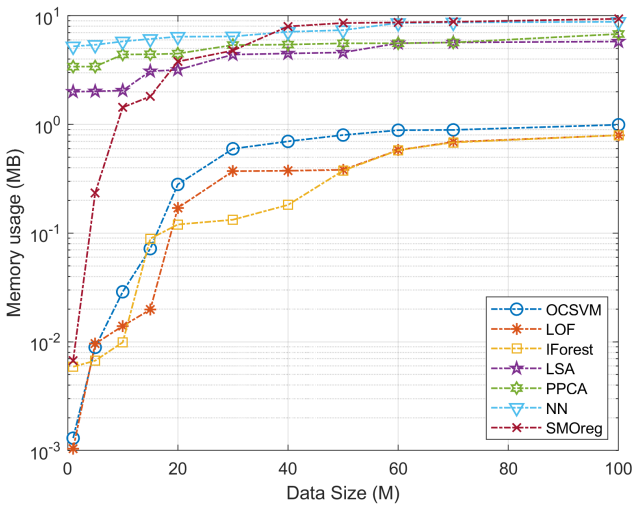


FIGURE 21. Memory usage of the algorithms in training phase under different data amount (modulation = QPSK, SNR = 15, Bandwidth = 1.4 MHz).

Complexity of the algorithms in terms of training times and prediction times are given in Figure 19 and 20. In terms of training time which is a dominant factor in time complexity of the algorithms and relatively larger than prediction times, OCSVM outperforms other methods. Other unsupervised methods, i.e., LOF and *i* Forest have training times which are smaller as compared supervised methods, i.e., NN. Unsupervised methods have larger prediction times when compared to supervised and conventional methods. Memory usage performance of supervised and conventional methods is noticeably worse than unsupervised methods as seen from Figure 21. Memory usage requirements of outlier detection methods is much lower (nearly 10 times) than that of conventional and supervised learning methods. For the algorithms applied in this study, training complexities of NN, SMOreg, LSA, PPCA, LOF, OCSVM, and *i* Forest are O(ML^2), O(L^3M), O(L^2M), O($L^2M + M^3$), O(Mk+ML), O(M^2L+M^3), and, O($LMlog(M)$), respectively. *i* Forest algorithm has exponential complexity, which is greater than complexity of LOF algorithm but less than OCSVM algorithm. PPCA method has the largest complexity. LSA and NN has similar complexity but it is still higher than that of outlier detection methods.

Table 7 summarizes the training time, prediction time, and memory usage dependencies of the algorithms to data size M and feature size L. When we increase the number of features L, its effect on training time, prediction time and memory

TABLE 7. Performance of the algorithms on dimension.

	Training time		Prediction time		Memory Usage	
Algorithm	$\nearrow M$	$\nearrow L$	$\nearrow M$	$\nearrow L$	$\nearrow M$	$\nearrow L$
LOF	Medium	Low	Medium	Low	Low	Low
OCSVM	High	Low	High	Low	High	Low
*i*Forest	Low	Low	Low	Low	Medium	Low
LSA	Low	Low	Medium	Low	Medium	Low
PPCA	Low	High	Low	Low	Low	High
NN	Medium	Low	Medium	Low	Medium	High
SMOreg	Medium	High	Low	Low	Medium	High

usage for all three algorithms remains low. On the other hand, the effect of data size is varying depending on the algorithm.

VI. CONCLUSION

In this study, the problem of detecting malicious relays in cooperative LTE-A has been studied. The detection of the attacks was performed without applying demodulation and channel estimation steps on the received signal in the physical layer. Detecting an attack in the first step for cooperative relay systems, i.e., before further processing the received relay signal at the destination, provides a significant gain in wireless communication systems. The physical layer detection approach by means of unsupervised machine learning techniques has been applied for this problem. More specifically, we have considered outlier detection as an unsupervised learning method. We have proposed using one class approach with OCSVM, LOF and iForest algorithms which can distinguish between malicious and secure relays. The effects of SNR, modulation type, data size on the precision, accuracy, AUC performances of the algorithms have been investigated for detecting relay attacks such as garbling, false data injection, and regenerative attacks and secure relay cases. This study has shown that malicious relay attacks can be detected by using the one class method with high precision and accuracy performances even under low SNR conditions. AUC values for all algorithms were above 0.9 for the attack detection cases. For all attack types, the performance of the LOF algorithm has been found to be superior to OCSVM and iForest algorithms. Accuracy performance of the algorithms for detection of malicious relays have not been affected considerably by changing the modulation type (QPSK, 16-QAM, 64-QAM) in LTE-A network. Also, the effect of changing data size M on the attack detection performance was observed to be insignificant. This study has showed that the minimum resources in the training phase, i.e., the lowest amount of RB that can be carried in a frame with the lowest bandwidth in LTE-A systems, can be used detect malicious relay attacks effectively. The results also implied that applying these algorithms even in small capacity devices was feasible.

The major contribution of this work is the demonstration of feasibility of applying unsupervised learning with outlier detection approach for effective detection of malicious relay nodes in the LTE-A network. We have also shown that baseband signal characteristics of malicious and secure relays in physical layer, which is quite difficult to detect using conventional statistical approaches, can be distinguished by proper selection of features in ML based approaches.

REFERENCES

[1] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. New York, NY, USA: Academic, 2011.

[2] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*. Hoboken, NJ, USA: Wiley, 2012.

[3] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. Hoboken, NJ, USA: Wiley, 2011.

[4] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: Next-generation wireless broadband technology," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10–22, Jun. 2010.

[5] *Long Term Evolution (LTE) Physical Layer*, document TS 36.201, General Description, 3GPP, Release 8, V8.3.0, Mar. 2009.

[6] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 100–105, Oct. 2009, doi: 10.1109/MCOM.2009.5273815.

[7] M. Ramarakula, "An efficient uplink power control algorithm for LTE-advanced relay networks to improve coverage area," *Wireless Pers. Commun.*, vol. 110, no. 3, pp. 1075–1087, Feb. 2020.

[8] Y. Yengi, A. Kavak, H. Arslan, K. Kucuk, and H. Yigit, "Malicious relay node detection with unsupervised learning in amplify-forward cooperative networks," in *Proc. Int. Conf. Innov. Intell. for Informat., Comput., Technol. (3ICT)*, Sakhier, Bahrain, Sep. 2019, pp. 1–5, doi: 10.1109/3ICT.2019.8910328.

[9] Y. Mao and M. Wu, "Tracing malicious relays in cooperative wireless communications," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 2, pp. 198–212, Jun. 2007, doi: 10.1109/TIFS.2007.897242.

[10] L.-C. Lo and W.-J. Huang, "Misbehavior detection without channel information in cooperative networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, San Francisco, CA, UA, Sep. 2011, pp. 1–5, doi: 10.1109/VETECF.2011.6092907.

[11] W. Hou, X. Wang, and A. Refaey, "Misbehavior detection in amplify-and-forward cooperative OFDM systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 5345–5349, doi: 10.1109/ICC.2013.6655437.

[12] T. Lv, Y. Yin, Y. Lu, S. Yang, E. Liu, and G. Clapworthy, "Physical detection of misbehavior in relay systems with unreliable channel state information," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1517–1530, Jul. 2018, doi: 10.1109/JSAC.2018.2824728.

[13] J. Xing, T. Lv, and X. Zhang, "Cooperative relay based on machine learning for enhancing physical layer security," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Istanbul, Turkey, Sep. 2019, pp. 1–6, doi: 10.1109/PIMRC.2019.8904319.

[14] Z. Deng, Q. Sang, Y. Gao, and C. Cai, "Optimal relay selection for wireless relay channel with external eavesdropper: A NN-based approach," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Beijing, China, Aug. 2018, pp. 515–519, doi: 10.1109/ICCChina.2018.8641223.

[15] X. Wang and F. Liu, "Data-driven relay selection for physical-layer security: A decision tree approach," *IEEE Access*, vol. 8, pp. 12105–12116, 2020, doi: 10.1109/ACCESS.2020.2965963.

[16] A. Carreño, I. Inza, and J. A. Lozano, "Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework," *Artif. Intell. Rev.*, vol. 53, pp. 3575–3594, Oct. 2019.

[17] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep learning convolutional neural networks for radio identification," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 146–152, Sep. 2018, doi: 10.1109/MCOM.2018.1800153.

[18] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018, doi: 10.1109/ACCESS.2018.2818794.

[19] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016, doi: 10.1109/TVT.2016.2524258.

[20] E. R. de Faria, A. C. P. de Leon Ferreira Carvalho, and J. Gama, "MINAS: Multiclass learning algorithm for novelty detection in data streams," *Data Mining Knowl. Discovery*, vol. 30, no. 3, pp. 640–680, May 2016.

[21] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Comput. Sci.*, vol. 60, pp. 708–713, Sep. 2015.

[22] R. Domingues, M. Filippone, P. Michiardi, and J. Zouaoui, "A comparative evaluation of outlier detection algorithms: Experiments and analyses," *Pattern Recognit.*, vol. 74, pp. 406–421, Feb. 2018.
[23] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 582–588.
[24] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, Jun. 2000.
[25] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discovery Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012.
[26] M. Iwamura, H. Takahashi, and S. Nagata, "Relay technology in LTE-advanced," *NTT DoCoMo Tech. J.*, vol. 12, no. 2, pp. 29–36, 2010.
[27] H. Ali, M. N. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: A review," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 14, no. 10, pp. 1560–1571, 2019.
[28] Y. J. Feng and J. A. Fan Suykens, "A statistical learning approach to modal regression," *J. Mach. Learn. Res.*, vol. 21, no. 2, pp. 1–35, 2020.
[29] J. K. Dutta and B. Banerjee, "Improved outlier detection using sparse coding-based methods," *Pattern Recognit. Lett.*, vol. 122, pp. 99–105, May 2019.
[30] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, Chicago, IL, USA, 2009, pp. 21–32.
[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, pp. 2825–2830, Nov. 2011.
[32] S. Zhang, S. Bahrampour, N. Ramakrishnan, L. Schott, and M. Shah, "Deep learning on symbolic representations for large-scale heterogeneous time-series event prediction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5970–5974.
[33] B. Zhao, X. Xiao, W. Zhang, B. Zhang, G. Gan, and S. Xia, "Self-paced probabilistic principal component analysis for data with outliers," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 3737–3741.
[34] R. Xin, J. Zhang, and Y. Shao, "Complex network classification with convolutional neural network," *Tsinghua Sci. Technol.*, vol. 25, no. 4, pp. 447–457, Aug. 2020.
[35] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans. Neural Netw.*, vol. 11, no. 5, pp. 1188–1193, Sep. 2000.



YELIZ YENGI received the B.S. degree from the Computer Engineering Department, Sakarya University, and the M.S. degree from the Computer Engineering Department, Kocaeli University, Turkey, where she is currently pursuing the Ph.D. degree. Her research interests include machine learning for security in LTE-A networks, natural language processing, and recommendation systems in machine learning.



ADNAN KAVAK received the M.S. and Ph.D. degrees from the Electrical-Computer Engineering Department, The University of Texas at Austin. He was a Senior Researcher with Samsung Telecommunication America, Dallas, TX, USA, from 2000 to 2001. From 2008 to 2009, he was a Visiting Professor with the Electrical-Computer Engineering Department, The University of Texas at Austin. He is currently a Professor with the Department of Computer Engineering, Kocaeli University, Kocaeli, Turkey. His current research interests include 5G and beyond wireless communication, cooperative networks, physical layer security in next generation networks, smart antennas and MIMO Systems, synchronization in cloud computing, E-Health systems, and machine learning applications in communication systems.



HÜSEYIN ARSLAN (Fellow, IEEE) received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1992, and the M.S. and Ph.D. degrees from Southern Methodist University, Dallas, TX, USA, in 1994 and 1998, respectively. From 1998 to 2002, he was with the Research Group, Ericsson, Inc., Raleigh, NC, USA, where he was involved with several projects related to 2G and 3G wireless communication systems. Since 2002, he has been with the Electrical Engineering Department, University of South Florida, Tampa, FL, USA. He has also been the Dean of the College of Engineering and Natural Sciences, İstanbul Medipol University, since 2014. He was a part-time consultant for various companies and institutions, including Anritsu Company, Morgan Hill, CA, USA, and The Scientific and Technological Research Council of Turkey (TÜBİTAK). His research interests include physical layer security, mmWave communications, small cells, multicarrier wireless technologies, co-existence issues on heterogeneous networks, aeronautical (high-altitude platform) communications, *in vivo* channel modeling, and system design.