International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Novel Image Steganography Method for Industrial Internet of Things Security with AES

Mohammed Arkkam Achchi

Safihudheen.V.N

Mohamed Zaheer

8th semester, Department of Computer Science and Engineering
Dhaanish Ahmed College of Engineering, Chennai.

***Abstract*—The rapid development of the Industrial Internet of Things (IIoT) and artifificial intelligence (AI) brings new security threats by exposing secret and private data. Thus, information security has become a major concern in the communication environment of IIoT and AI, where security and privacy must be ensured for the messages between a sender and the intended recipient. To this end, we propose a method called HHOIWT for covert communication and secure data in the IIoT environment based on digital image steganography. The method embeds secret data in the cover images using a metaheuristic optimization algorithm called Harris hawks optimization (HHO) to effificiently select image pixels that can be used to hide bits of secret data within integer wavelet transforms. The HHO-based pixel selection operation uses an objective function evaluation depending on two phases: exploitation and exploration. The objective function is employed to determine an optimal encoding vector to transform secret data into an encoded form generated by the HHO algorithm. Several experiments are conducted to validate the performance of the proposed method with respect to visual quality, payload capacity, and security against attacks. The obtained results reveal that the HHO-IWT method achieves higher levels of security than the state-of-the-art methods and that it resists various forms of steganalysis. Thus, utilizing this approach can keep unauthorized individuals away from the transmitted information and solve some security challenges in the IIoT**

I.INTRODUCTION

With the rapid development of the Industrial Internet of Things (IIoT) and the common use of artifificial intelligence (AI) in the IIoT, several security threats have appeared in the present time [1], [2]. Today, many devices are connected to the IoT such as routers, refrigerators, we bcams, smart locks, medical devices and home automation hubs powered by AI technologies [3]. All these devices or Manuscript received September 28, 2020; revised November 21, 2020; accepted December 26, 2020. Paper no. TII-20-4521. (Corresponding author: Khan Muhammad.) M. Hassaballah is with the Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena, Egypt (Email:
m.hassaballah@svu.edu.eg) Mohamed Abdel Hameed is with the Department of Computer Science, Faculty of Computers and Information, Luxor University, Luxor, Egypt (Email: m.abdelhamid@fci.luxor.edu.eg)
Ali Ismail Awad is with the Department of Computer Science, Electrical and Space Engineering, Lule˚
a University of Technology, Lule˚a, Sweden (Email:
ali.awad@ltu.se). He is also with the Electrical Engineering Department, Faculty of Engineering, Al-Azhar University, Qena, Egypt. He is also with the Centre for Security, Communications and Network Research, University of Plymouth, United Kingdom.prohibitive cost for accessing ground-truth labels of anomalies,anomaly detection on attributed networks is predominately carried out in an unsupervised manner [13], [16]. That is to say, the algorithm has to conclude the normal pattern of data

Fig. 1. Toy examples to illustrate different types of anomalies in attributed
networks. A structural anomaly often has wrong links with other nodes but
has normal attributes. For example, in (a), an American engineer has a very low probability of being associated with a German designer and a British

manager, so the links between them are abnormal. A contextual anomaly usually has a natural neighboring structure but its attributes are corrupted.

For instance, in (b), the attribute vector of the anomaly node is disturbed by noisy information, e.g., mismatched location, employer, and occupation.

(a) Structural Anomaly. (b) Contextual Anomaly.

from the corrupted networks without supervision. Hence, a key is to fully and reasonably exploit existing information from attributed network data.

Recently, various methods have been proposed to deal with the anomaly detection task for attributed networks. The shallow methods, including AMEN [16], Radar [17], and ANOMALOUS [18], leverage shallow learning mecha nisms (e.g., ego-network analysis, residual analysis, or CUR decomposition) to detect anomalies. Unfortunately, these models cannot fully address the computational challenge on attributed networks and fail to capture the complex interactions between different information modalities due to limitations of shallow mechanisms, especially when the feature is high-dimensional [13]. With the rocketing growth of deep learning for anomaly detection [12], [19]–[21], researchers also present deep neural networks-based methods to solve the anomaly detection problem on attributed networks. DOMI NANT [13] is one of the representative methods. It constructs a graph autoencoder to reconstruct the attribute and structure information simultaneously, and the abnormality is evaluated by reconstruction error. SpecAE [22] also leverages graph autoencoder to extract low-dimensional embedding and carries out detection via density estimation. Although existing deep learning-based methods [13], [22] have achieved considerable performance for anomaly detection on graphs, they still have several shortcomings, largely attributed to the autoencoder backbone in their architectures. First, autoencoders aim to learn the latent representation by reconstructing the original data instead of detecting the anomaly itself. Although the anomaly scores can be computed according to reconstruction errors [13], these kinds of methods can only achieve suboptimal performance due to the fact that they do not target directly the anomaly detection objective. Second, autoencoder-based methods may not able to fully exploit the rich information of the attributed graph for effective graph representation learning. Specifically, autoencoders simply rebuild the original data and they do not have any refifinement for data. However, recent works [23]–[25] have shown that more useful information can be mined in an unsupervised way if we design certain pretext tasks carefully based on augmented data. Third, graph autoencoder is the bottleneck to carry out anomaly detection on large-scale networks. In general, the graph convolution operation in graph autoencoder needs to input and reconstruct the full networked data, which is unfeasible due to the explosive memory requirements when the network is large. As an alternative unsup5ervised learning technique, self-supervised contrastive learning is a promising solution to address the aforementioned limitations. By learning to contrast the elaborate instance pairs, the model can acquire informative knowledge without manual labels. Contrastive self-supervised learning has nice properties for the anomaly detection task. *First*, contrastive learning mainly studies the matching of pairs of instances, which offers helpful information for anomaly detection. For the normal instance in graphs, there is a potential matching pattern between each node and its neighbors, e.g., the homophily hypothesis. The anomalies, on the opposite, often present when there is an inconsistency/mismatch between attributes and structure, which violates the original matching pattern of networks. Moreover, different types of anomalies have different manners of mismatching: in Fig. 1, the structural anomaly has individual abnormal links with uncorrelated nodes, which is partial inconsistency; the contextual anomaly, differently, has mismatched attributes with all neighbors. Contrastive learning, naturally, is capable to learn the matching patterns and capture various mismatching patterns via its intrinsic discriminative mechanism.

Second, contrastive learning models provide a specifific predicted score to measure the agreement between the elements in each instance pair, and the score is highly related to the abnormality of instance. Since anomaly detection methods usually output a list of scores or a ranking to represent the abnormality of each node, the predicted scores of the contrastive learning model can be utilized for anomaly detection directly. In this way, we can train the model via an objective that is highly relevant to anomaly detection. In this article, we propose a novel Contrastive self-supervised Learning framework for Anomaly detection on attributed networks (CoLA for abbreviation). By sampling the well-designed instance pairs from the full network and using them to train the contrastive learning model, the information of the network is exploited better. Concretely, our framework focuses on modeling the relationship between each node and its partial neighboring substructure, which can expose the various types of anomalies within networks. Meanwhile, our CoLA framework is trained with a direct target to assist the anomaly detection task. We set the learning objective of our model to discriminate the agreement between the elements within the instance pairs, and the results can be further used to evaluate the abnormality of nodes. Besides, by splitting the network into separated lightweight instance pairs, our anomaly detection framework is compatible with large-scale networks. Specifically, our framework does not need to run graph convolution on full networks; therefore, it successfully avoids the memory explosionproblem. To summarize, the main contributions are as follows.

1) We propose a contrastive self-supervised learning frame work, CoLA, for the anomaly detection problem onThis article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LIU *et al.*: ANOMALY DETECTION ON ATTRIBUTED NETWORKS VIA CONTRASTIVE SELF-SUPERVISED LEARNING 3

attributed networks. To the best of our knowledge, this is the fifirst contrastive self-supervised learning-based method for graph anomaly detection.

2) We present a novel type of contrastive instance pair, "target node versus local subgraph," for attributed net works to adapt to the anomaly detection task, which

effifficiently captures the local information of a node and its neighboring substructure.
3) We design a contrastive learning model to learn the rep resentative information from the node-subgraph instance pairs and provide discriminative scores for abnormality ranking. The proposed learning model is friendly to large-scale networked data.
4) We conduct extensive experiments on various data sets to demonstrate the effectiveness of CoLA and its supe riority compared with a range of baseline methods.
The rest of this article is organized as follows. In Section II, we fifirst review the related works. Then, the preliminary defifinitions and notations are introduced in Section III. Section IV illustrates the overall pipeline and the components of our framework in detail. After that, we analyze the experimental results in Section V and then conclude our work in section VI.

II. RELATED WORK

Anomaly detection on attributed networks attracts considerable research interest in recent years due to the wide application of attributed networks in modeling complex systems [12]. AMEN [16] detects anomalies by leveraging ego-network information of each node on attributed networks. Radar [17] characterizes the residuals of attribute information and its coherence with network information for anomaly detection. Furthermore, ANOMALOUS [18] jointly considers CUR decomposition and residual analysis for anomaly detection on attributed networks. Zhu and Zhu [49] present a joint learning model to detect mixed anomaly by core initiating and expanding. Despite their success on low-dimensional attributed network data, these methods cannot work well when the networks have complex structures and high-dimensional attributes due to the limitation of their shallow mechanisms. With the rocketing growth of the deep learning technique [29], several deep approaches are presented to solve the anomaly detection problem for attributed networks. DOMINANT [13] constructs an autoencoder with GCN layers to reconstruct both the attribute matrix and adjacency matrix. It defifines the anomaly score of the node as the weighted sum of its reconstruction errors of attribute and structure. SpecAE [22] leverages a spectral graph autoencoder to extract the latent embedding of each node and uses the Gaussian mixture model to perform the detection. For dynamic networks, NetWalk [50] learns dynamically network representations with random walk sampling and autoencoder model and detects anomalies with a clustering-based detector. The above-mentioned deep methods achieve superior performance over the shallow methods by not only introducing deep neural networks but also have several shortcomings caused by their reconstruction mechanism with autoencoders. First, reconstruction is a naive unsupervised learning solution that fails to make full use of data. On contrary, CoLA better utilizes the attribute and structure information in a self-supervised manner. Second, their reconstructive optimization target is not associated with anomaly detection. In contrast to them, our learning objective is to discriminate the agreement between nodes and subgraphs, which can indicate the abnormality of nodes directly. Third, these methods require full adjacency and attribute matrix as model's input, which makes these algorithms unable to be run on large-sized network data due to the explosive memory requirements. In contrast, our framework learns with sampled instance pairs, rather than the full network, which makes it flflexibly adapt to large-scale networks.

III.EXISTING SYSTEM:

They proposed first to create a contrastive self-supervised learning to the anomaly detection problem of attributed networks. CoLa, is mainly consists of three components: contrastive instance pair sampling, GNN-based contrastive learning model, and multiround sampling-based anomaly score computation. Their model captures the relationship between each node and its neighbouring structure and uses an anomaly- related objective to train the contrastive learning model. We believe that the proposed framework opens a new opportunity to expand self-supervised learning and contrastive learning to increasingly graph anomaly detection applications. The multiround predicted scores by the contrastive learning model are further used to evaluate the abnormality of each node with statistical estimation. The training phase and the inference phase. In the training phase, the contrastive learning model is trained with sampled instance pairs in an unsupervised fashion. After that the anomaly score for each node is obtained in the inference phase.

IV.PROPOSED SYSTEM:

The proposed model is to build a machine learning model for anomaly detection. Anomaly detection is an important technique for recognizing fraud activities, suspicious activities, network intrusion, and other abnormal events that may have great significance but are difficult to detect. The machine learning model is built by applying proper data science techniques like variable identification that is the dependent and independent variables. Then the visualisation of the data is done to insights of the data .The model is build based on the previous dataset where the algorithm learn data and get trained different algorithms are used for better comparisons. The performance metrics are calculated and compared.

V.IMPLEMENTATIONS MODULES:

In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems

are: speech recognition, handwriting recognition, bio metric identification, document classification etc. In Supervised Learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which label should be given to new data based on pattern and associating the patterns to the unlabeled new data.

LOGISTIC REGRESSION:

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).In other words, the logistic regression model predicts P(Y=1) as a function of X. Logistic regression Assumptions:
Binary logistic regression requires the dependent variable to be binary. For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.Only the meaningful variables should be included.
The independent variables should be independent of each other. That is, the model should have little.The independent variables are linearly related to the log odds. Logistic regression requires quite large sample sizes.

DECISION TREE:

It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Assumptions of Decision tree: At the beginning, we consider the whole training set as the root.Attributes are assumed to be categorical for information gain, attributes are assumed to be continuous.On the basis of attribute values records are distributed recursively.We use statistical methods for ordering attributes as root or internal node.
Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into
smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set
which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed. This process is continued on the training set until meeting a termination condition. It is constructed in a top-down recursive divide-and-conquer manner. All the attributes should be categorical. Otherwise, they should be discretized in
advance. Attributes in the top of the tree have more impact towards in the classification and they are identified using the information gain concept. A decision tree can be easily over-fitted generating too many branches and may reflect
anomalies due to noise or outliers.

RANDOM FOREST:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join
different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

SUPPORT VECTOR MACHINE:

A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines are perhaps one of the most popular and talked about machine learning algorithms. They were extremely popular around the time they were developed in the 1990s and continue to be the go-to method for a high-performing algorithm with little tuning.How to disentangle the many names used to refer to support vector machines.The representation used by SVM when the model is actually stored on disk. How a learned SVM model representation can be used to make predictions for new data.How to learn an SVM model from training data.How to best prepare your data for the SVM algorithm.Where you might look to get more information on SVM.

GRAPHICAL USER INTERFACE:

GUI means Graphical User Interface. It is the common user Interface that includes Graphical representation like buttons and icons, and communication can be performed by interacting with these icons rather than the usual text-based or command-based communication. A common example of a GUI is Microsoft operating systems.

The graphical user interface (GUI) is a form of user interface that allows users to interact with electronic devices through graphical icons and audio

indicator such as primary notation, instead of text-based user interfaces, typed command labels or text navigation. GUIs were introduced in reaction to the perceived steep learning curve of command-line interfaces (CLIs) which require commands to be typed on a computer keyboard.

TKINTER

Tkinter is Python's de-facto standard GUI (Graphical User Interface) package. It is a thin object-oriented layer on top of Tcl/Tk. Tkinter is not the only GUI Programming toolkit for Python. It is however the most commonly used one. ... Graphical User Interfaces with Tk, a chapter from the Python Documentation.

The tkinter package ("Tk interface") is the standard Python interface to the Tcl/Tk GUI toolkit. Both Tk and tkinter are available on most Unix platforms, including macOS, as well as on Windows systems.

Running python –m tkinter from the command line should open a window demonstrating a simple Tk interface, letting you know that tkinter is properly installed on your system, and also showing what version of Tcl/Tk is installed, so you can read the Tcl/Tk documentation specific to that version.

Tkinter supports a range of Tcl/Tk versions, built either with or without thread support. The official Python binary release bundles Tcl/Tk 8.6 threaded. See the source code for the _tkinter module for more information about supported versions.

ADVANTAGES OF PROPOSED SYSTEM

1. The anomaly detection can be automated process using the machine learning.
2. Performance metric are compared in order to get better model.

VI.CONCLUSION:

The analytical process starting from data cleaning and processing , missing values, exploratory analysis and finally model building and evaluation.

- The best accuracy on public test set is higher accuracy score will be found out by comparing each algorithm with the type of all network attacks for future prediction results by finding best connections.
- This brings some of the following insights about the diagnose of the network attack of each new connection.
- To present a prediction model, with the aid of artificial intelligence to improve over human accuracy and provide with the scope of early detection.

VII.REFERENCE :

[1] Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song, "Heterogeneous graph neural networks for malicious account detection," in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2018, pp. 2077–2085.

[2] L. Tang and H. Liu, "Relational learning via latent social dimensions," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2009, pp. 817–826.

[3] Y. Zhang et al., "Your style your identity: Leveraging writing and photography styles for drug trafficker identification in darknet markets over attributed heterogeneous information network," in Proc. World Wide Web Conf. (WWW), 2019, pp. 3448–3454.

[4] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for Web-scale rec ommender systems," in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp. 974–983.

[5] W. Fan et al., "Graph neural networks for social recommendation," in Proc. World Wide Web Conf. (WWW), 2019, pp. 417–426.

[6] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Represent., 2017, pp. 1–14.