Increasing the Trustworthiness in the Industrial IoT Networks Through a Reliable Cyberattack Detection Model

Mohammad Mehedi Hassan, *Senior Member, IEEE*, Abdu Gumaei, Shamsul Huda, and Ahmad Almogren, *Member, IEEE*

***Abstract*—The trustworthiness of an industrial Internet of Things (IIoT) network is an important stakeholder expectation. Maintaining the trustworthiness of such a network is crucial to void the loss of lives. A trustworthy IIoT system combines the security characteristics of IT trustworthiness—safety, security, privacy, reliability, and resilience. Conventional security tools and techniques are not enough to safeguard the IIoT platform due to the difference in protocols, limited upgrade opportunities, mismatch in protocols, and older versions of the operating system used in the industrial system. In this article, we propose to improve the trustworthiness of an IIoT network [i.e., supervisory control and data acquisition (SCADA) network] through a reliable and salable cyberattack detection model. In particular, an ensemble-learning model based on the combination of a random subspace (RS) learning method with random tree (RT) is proposed for detecting cyberattacks of SCADA by using the network traffics from the SCADA-based IIoT platform. The novelty of the proposed model is that it uses the industrial protocol-based network traffic and the RS to solve the sensitivity of irrelevant features and ensemble RT to reduce the overfitting problem, thereby constructs a detection engine based on industrial protocols and achieves high detection rates. The proposed model has been tested over 15 datasets of the SCADA network. Experimental results reveal that the proposed model outperforms conventional detection techniques and, thus, improves the security and related measure of the trustworthiness of the IIoT platform.**

***Index Terms*—Cyberattack, deep learning, reliability, industrial Internet of Things (IIoT), supervisory control and data acquisition (SCADA) network, trustworthiness.**

Manuscript received September 29, 2019; revised December 24, 2019; accepted January 14, 2020. Date of publication January 28, 2020; date of current version May 26, 2020. This work was supported by King Saud University, Riyadh, Saudi Arabia, under Researchers Supporting Project RSP-2019/18. Paper no. TII-19-4451. *(Corresponding author: Mohammad Mehedi Hassan.)*

M. M. Hassan, A. Gumaei, and A. Almogren are with the College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia (e-mail: mmhassan@ksu.edu.sa; abdugumaei@gmail.com; ahalmogren@ksu.edu.sa).

S. Huda is with the School of Information Technology, Deakin University, Burwood, VIC 3125, Australia (e-mail: shamsul.huda@deakin.edu.au).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2970074

I. Introduction

INDUSTRY 4.0 [1] has created the revolutionary industrial Internet [1], which connects billions of Internet of Things (IoTs) devices [1]–[4], also known as industrial IoT (IIoTs) [1]. In IIoT networks, a huge number of supervisory control and data acquisition (SCADA) based industrial control systems are connected to the corporate network over the Internet for intelligent business management [5], [6]. Usually, such a system comprises numerous field devices [7]–[10] such as sensors, actuators, intelligent electronic devices, which are connected with the enterprise network using heterogeneous communications infrastructures [8], [9] over the global internet.

This integration provides enormous flexibility [1]–[6], [8]–[10] and agility for inter/intra-industrial systems management resulting in higher productivity with resource efficiency. However, this integration exposes the SCADA-based industrial networks to severe security threats and poses a big challenge to the trustworthiness of IIoT networks and systems [9]–[14]. The trustworthiness of an IIoT system ensures how the IIoT system performs according to designed behavior under a set of security conditions such as safety, security, privacy, reliability, and resilience [15]. Fig. 1 shows the elements of trustworthiness in an IIoT network. Improving the trustworthiness of IIoT networks through the protection of data, services and identity and thereby, safeguarding of SCADA-based industrial networks and the related embedded devices in the entire industrial networks from being compromised by cybercriminals are the key objectives in the IIoT system [13], [14].

Toward the objective, a few newer versions of protocols have been proposed such as the secure distributed network protocol (DNP3.0) [16]. Secure protocols consider only the integrity and authentication that leave many loopholes for the attackers to conduct severe malicious attacks through known weaknesses such as hash collision [13]–[16]. As a complementary, in most modern automated or semiautomated industries, industrial operational technology (OT) and information technology (IT) bodies jointly develop the conventional risk management plan by following the standards such as ISO 27005:2018 [17] to identify threats and risk, prioritize the risks, and prepare mitigation strategies. However, even having a sophisticated risk management plan and with a reasonable level of preventive measures do not guarantee complete security from emerging threats and attacks.

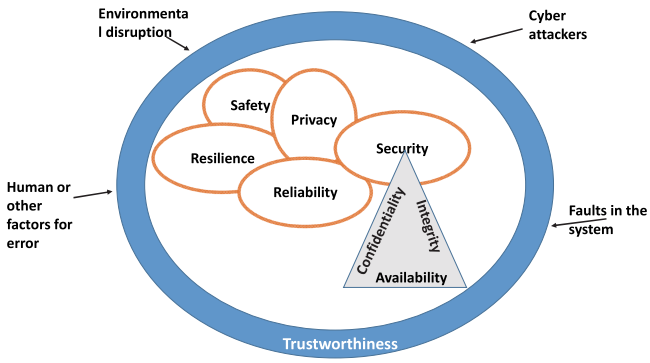


Fig. 1. Trustworthiness and security objectives (CIA triad).

This is due to the fact that industrial networks use an older version of the operating system (OS) with known vulnerabilities and limited upgrade opportunities due to the numerous installed devices with limited physical access and incompatibility with control software, expensive shutdown plan, and less availability of expertise [13]–[16]. So one of the main issues and urgent demand for OT and IT management bodies within the industrial societies is how to achieve maximum security and protection to achieve a high level of trustworthiness. This indirectly poses a challenging research problem to the cybersecurity and industrial management researchers: how to achieve the highest level of detection of threats and attacks and flag malicious activities as soon as they are detected, then isolate the affected subsystems immediately.

Considering the limitations of existing approaches, we propose an ensemble-based detection engine for the detection of cyberattacks in a large industrial network by using the network features of industrial protocols [16]. The novelty of our proposed detection engine is that its scalability for a larger industrial network with many domains and compatibility with existing detection engine. The ensemble-based detection engine can be easily distributed over different domains of the IIoT and the decision can be combined with other existing detection engines. The proposed ensemble-based model can be easily deployed and overcome the limitations of existing works by improving both accuracy and efficiency. The novelty and contribution of this article are mentioned as follows.

1) An efficient and scalable cyber-attack detection model is proposed to improve the trustworthiness of a SCADA-based IIoT platform by using an ensemble learning.
2) An efficient probing technique is used based on SCADA network traffic to overcome the limitations of conventional security tools for the IIoT platform due to protocol mismatch.
3) A statistical analysis technique is proposed to verify the reliability, thereby the trustworthiness of our proposed detection model for IIoT networks.

The rest of this article is structured as follows. Section II presents related works. Section III explains research materials and methods including the proposed model in detail. Section IV presents the experiments and discussion. Section V concludes this article.

II. Related Works

Different approaches have been proposed in the literature for the protection of cyber–physical system, which uses supervised machine learning, semisupervised machine learning, deep learning techniques, statistical techniques, and online learning technique [9], [11], [13], [16], [18]. These approaches use a simulated cyber–physical system to collect network traffic and apply machine-learning approaches to detect the cyberattacks. The intrusion-detection methods in [9], [11], [13], [16], and [18] that use a single classifier may not be robust to the problem because random fluctuations or noises in the SCADA traffic data are learned as concepts. The existing approaches [9]–[18] consider the industrial network as an individual system. Semisupervised [11] and online learning approaches [18] can overcome the limitations of supervised approaches [9] by introducing extracted information of unknown traffic and diminish the requirement of labeling the traffic manually. However, modern industrial networks are highly interconnected and have different domains such as the smart grid. A highly interconnected network such as a smart grid incorporates different domains such as generation, distribution, consumer powered renewable-energy supplies, substation networks, and enterprise networks. Complex network events from different domains of the grid present the physical interpretation of network events as a whole, not as an individual domain event, such as coordinated cyberattacks to the ICSs. In this scenario, conventional detection engines [9], [11], [16], [18] show limitations to provide domain-wide protection due to its theoretical consideration. This directs us to develop a detection engine, which can be distributed across the domains of completely industrial Internet and can be deployed to build a coordinated engine.

III. Materials and Methods

In this section, we describe the cyberattacks in the industrial control system dataset and explain the machine learning methods used to develop the proposed model.

A. Industrial Control System Datasets

In this article, the scenario of cyberattacks in the industrial control system follows the real case scenario implemented in [19] and [20]. Through this scenario, we use the datasets of the power control system collected in [21] for industrial cyberattacks detection. The general architecture of a SCADA-based industrial control network is shown in Fig. 2. It is hierarchically constructed using a number of layers such as a physical layer, central master control and processing layer, and corporate layer. As shown in Fig. 2, the physical layer has different equipment including power generators (G1, G2), breakers (BR1–BR 4), intelligent electronic devices (R1–R4), and programmable logic controllers. These devices accumulate sensor data from the sensors implanted in the lowest physical layer and take control decisions based on the sensor data and local control logic and also receive instruction from the upper layer, i.e., "master control/process layer." The master process and control layer monitor and control the devices in the local control layer and

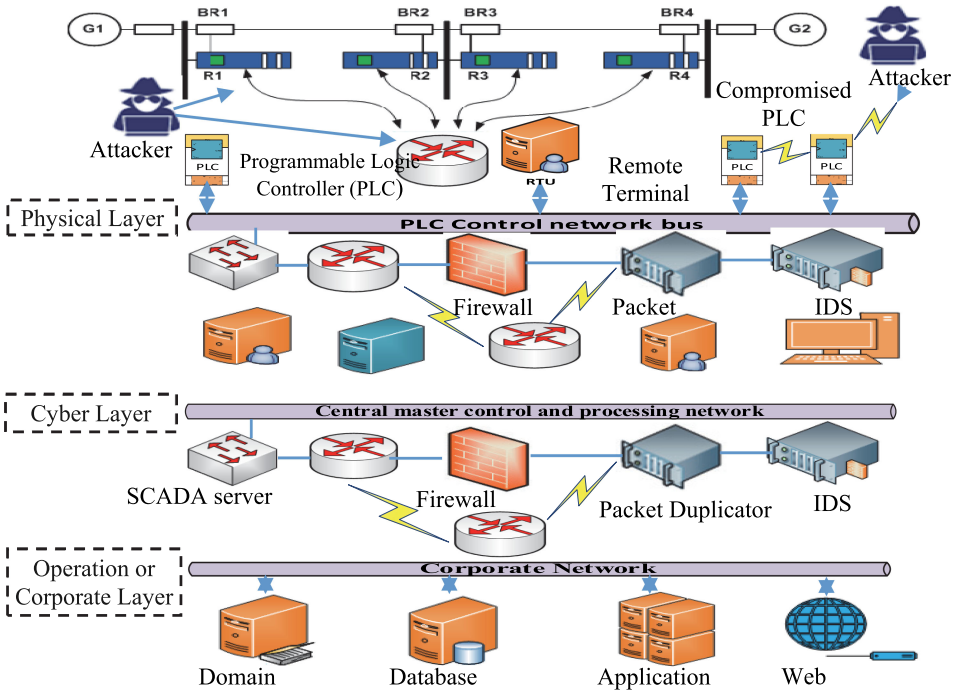


Fig. 2. General architecture of SCADA-based industrial control network.

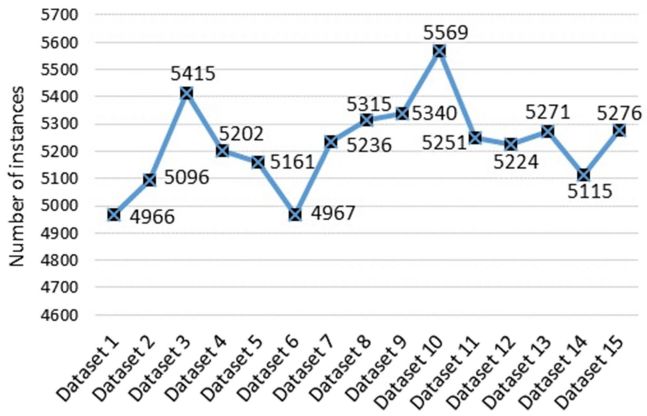


Fig. 3. Distribution of benchmark datasets.

thereby remote physical devices. They have intrusion detection systems (IDS) as well. The corporate layer sits at the top of all layers which is an IT system to support business processes and push management decisions to the master control layer. The benchmark datasets adopted in the study are 15 datasets generated from the power system of SCADA to detect the types of attacks [21]. The distribution of instances in these benchmark datasets is shown in Fig. 3. Two different classification events are used for detecting the intrusion attacks of the SCADA system. These two events are defined as follows.

1) *Classification of multiclass events:* This task of classification covers 37 scenarios of events, including normal events, natural events, and attack events with their own class labels.
2) *Classification of binary class events:* This task consists of 37 scenarios that can be grouped into nine events, which are the normal events, and 28 events that are attack events.

All 15 datasets consist of thousands of distinct attacks. To reduce the effect of a small sample size, the datasets are sampled at 1% in a random manner. Therefore, there are 294 samples of *no events*, 3711 samples of *attack events*, and 1221 samples of *natural events*. Time and date information are removed from the datasets to make the detection based only on the data.

B. Random Tree (RT) Classifier

RT is a classifier model that uses a supervised machine-learning technique to construct a tree by selecting a random feature at each node for expansion.

Consider the training examples in the dataset D are given in the form $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, where x_i is a vector of discrete or real values. These values are features of x_i that can be written in the form $\langle x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}\rangle$. In other words, the notation x_{ij} represent the jth features of vector x_i. The values of vector y_i are either discrete values representing a set of classes $\{1, \ldots, K\}$ as in the case of classification or real values as in the case of regression. Given a set D of the training dataset of E examples, the learning algorithm yields an RT classifier T. The T classifier is a hypothesis of a true function $f(x_i) = y_i$ in which new values of x_i gives a prediction of y_i values.

C. Random Subspace (RS) Method

The RS is a machine learning method, used for ensemble learning [5], [6]. The RS method constructs a set of machine learning classifiers or learners through selecting random subsets

of features. It has achieved high performance on many classification and predication applications [7]–[10]. Suppose that the RS method combines a set of different learners L on S random subspace of features $F_i \supseteq F$, denoted as $\{F_i(\cdot)\}_{j=1,\ldots,S}$. The class labels that are obtained by all individual learners can be represented as $\{y_i\}_{j=1,\ldots,S}$. For a new example x of F_i features, each learner can be classified based on its feature subspace F_i independently. The outputs of all learners can be calculated as

$$y_i = \{L_i(x, F_i)\}_{j=1,\ldots,S}. \tag{1}$$

Finally, the outputs of all learners can be ensemble by using a majority-voting rule [8] to get the final output, which is the classification label y as follows:

$$y = \arg\max_{y_i} \sum_{i=1}^{S} y_i. \tag{2}$$

D. Proposed Random Subspace-Based Random Tree (RSRT) Model

The RSRT is a fast and scalable ensemble-learning model. It combines the RS method with the RT classifier to create a reasonable set of base learners. The RSRT model is able to construct ensembles of random trees using different random subsets of features, which are selected randomly from the entire features of the training dataset. Accordingly, the ensembles of trees can help to reduce the effect of redundant features and avoid the overfitting problem, maintaining the strength of the individual trees over the split random selection.

Let D_N^F refers to the training set of F features and N samples. In addition, $D_N^{F_i}$ refers to the training set that has only F_i features of the subspace i from S RS. Similarly, let G_M^F is the testing set of F features and M samples and $G_M^{F_i}$ represents the same testing set with the same indices of the subspace features F_i. In the same context, y_i^M denotes the output class labels of M samples obtained from the ith base learners and y^M is the final output class label after majority voting of these base learners. The flowchart of the RSRT model that can be deployed in the IDS of SCADA-based IIoT is shown in Fig. 4.

The advantage of the RSRT algorithm is its efficiency in both training and testing plus its minimal requirements of memory. The model utilizes only one pass over the training dataset to construct each random tree. For the purpose of intrusion detection in SCADA IIoT, the RSRT model is suitable and scalable for its ability to reduce the irrelevant features and solve the overfitting problem.

Theorem 1: Our proposed RSRT model is trusted to detect intrusion attacks in SCADA-based IIoT by averaging the votes of its base learners.

Proof: Let S is a set of training instances of SCADA-based IIoT. The learning algorithm of an arbitrary machine-learning model M can construct a learner L. The learner L is a hypothesis about the true function f that takes a new instance x and gives its corresponding label y. The learning algorithm of the RSRT model constructs a set of learners or hypotheses and searches in a space H of these hypotheses to obtain the best hypothesis. The learning algorithm of the RSRT model can find several different

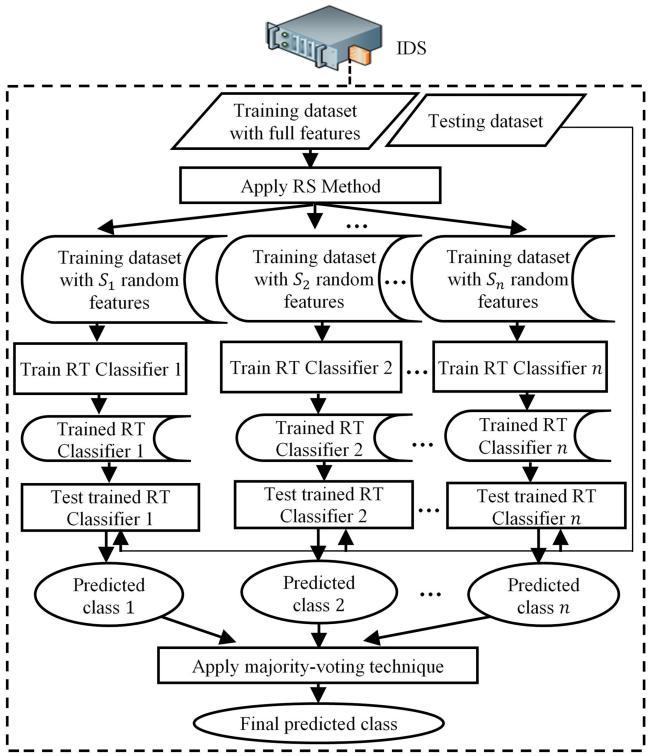


Fig. 4. Flowchart of RSRT model deployed in IDS of SCADA-based IIoT architecture.

TABLE I
PARAMETERS SETTINGS OF RSRT MODEL

Parameter	Value
Batch Size	100
number of learners (iterations)	5, 15, and 20
Subspace Size Ratio	0.1

hypotheses in H, which may give the same or different accuracy results on the training instances of different random feature sets. By constructing a set of accurate learners, the RSRT model can average their votes, reducing the risk of choosing wrong learners, which are not able to detect intrusion attacks in SCADA-based IIoT. Statistically, averaging the votes of accurate hypotheses can observe a good approximation to the function f. Therefore, the RSRT model is trusted to detect intrusion attacks in SCADA-based IIoT by averaging the votes of its base learners.

IV. EXPERIMENTS AND DISCUSSION

A large set of experiments are conducted on the benchmark datasets described in Section II. Through these experiments, the parameters of the RSRT model are initialized with a number of values, as given in Table I.

The different values for the number of learners are considered to test their effect on the result of detection accuracy. In general, the specific values of the model's parameters can be determined by the application owner based on the nature of data and requirements. A tenfold cross-validation strategy is also chosen for training and testing the proposed model. This strategy divides

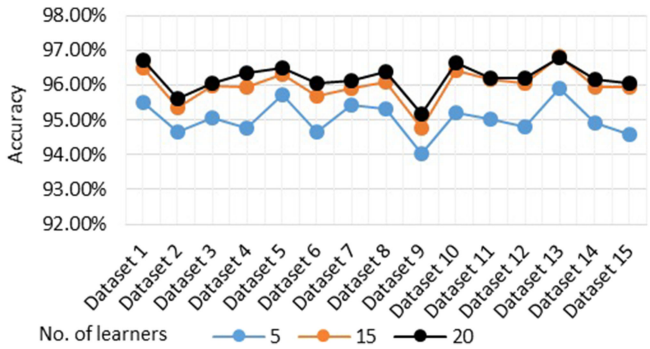


Fig. 5. Accuracy results of binary classification for the proposed RSRT model.

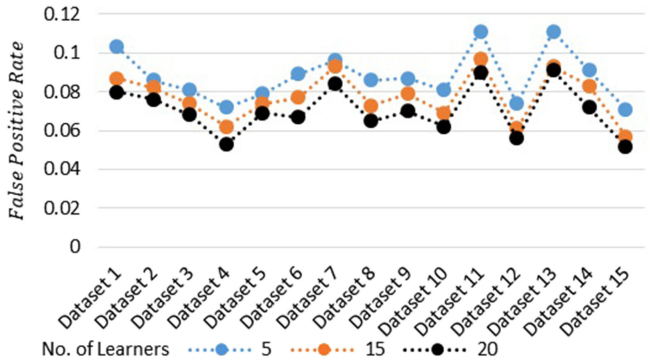


Fig. 6. False-positive rates of binary classification for the proposed RSRT model.

randomly the dataset into ten parts and chooses one part for testing and the other nine parts for training. This division will be repeated ten times, and the testing result is the average of testing results for the ten times. The results are measured using detection time and the following evaluation metrics:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{4}$$

where TP, FP, TN, and FN are the true positive, false positive, true negative, and false negative, respectively.

A. Results

The experimental results for the proposed model are shown in Figs. 5–8. Figs. 5 and 6 illustrate the accuracy results and false-positive rate for detecting the normal and abnormal events. In addition, Figs. 7 and 8 demonstrate the accuracy results and false-positive rates to classify the normal and different types of attacks in traffic events.

B. Comparison With Recent Related Works

In order to show the superiority of the proposed model against the latest work in the field, we compare its accuracy results and computational time costs to the accuracy results and computational time costs of the RSKNN model proposed in [17]. In this

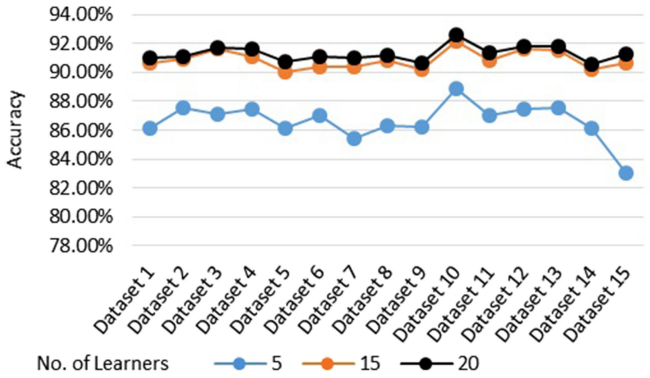


Fig. 7. Accuracy results of multiclassification for the proposed RSRT model.

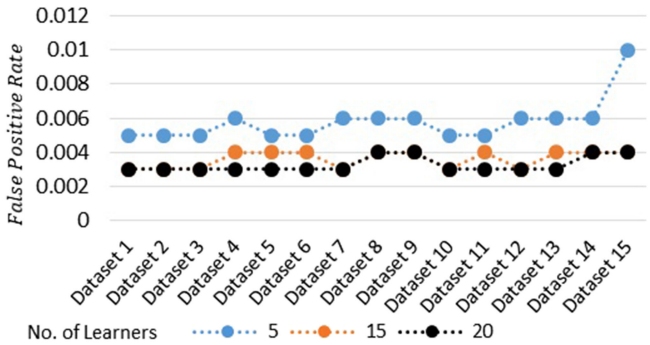


Fig. 8. False-positive rates of multiclassification for the proposed RSRT model.

TABLE II
ACCURACY RESULTS OF BINARY CLASSIFICATION FOR RSRT MODEL COMPARED WITH RSKNN MODEL

Dataset \ Model	RSKNN [17]	Proposed RSRT
Dataset 1	95.8719 %	96.7177 %
Dataset 2	95.1272 %	95.6007 %
Dataset 3	95.9187 %	96.0665 %
Dataset 4	95.1749 %	96.3476 %
Dataset 5	96.5511 %	96.4929 %
Dataset 6	95.7318 %	96.054 %
Dataset 7	95.2636 %	96.123 %
Dataset 8	95.5974 %	96.3876 %
Dataset 9	95.1685 %	95.1873 %
Dataset 10	96.1393 %	96.6601 %
Dataset 11	95.7722 %	96.1912 %
Dataset 12	95.9418 %	96.1907 %
Dataset 13	96.7369 %	96.7748 %
Dataset 14	95.4839 %	96.1681 %
Dataset 15	95.1099 %	96.0576 %

comparison, the accuracy results for both models are computed for the 15 datasets and the computational time costs are calculated for the dataset 9. Furthermore, the statistical differences for accuracy results will be analyzed using a statistical analysis test.

1) Comparison of Accuracy Results: The experiments are conducted on 15 datasets using each model separately. Tables II provides the accuracy results of our RSRT model as compared to the RSKNN model for detecting normal and abnormal events.

TABLE III
ACCURACY RESULTS OF MULTICLASSIFICATION FOR RSRT MODEL COMPARED WITH RSKNN MODEL

Model / Dataset	RSKNN [17]	Proposed RSRT
Dataset 1	89.9919 %	91.0189 %
Dataset 2	89.1300 %	91.1028 %
Dataset 3	90.7295 %	91.6713 %
Dataset 4	89.7924 %	91.6378 %
Dataset 5	90.2151 %	90.7770 %
Dataset 6	90.7389 %	91.1013 %
Dataset 7	89.8778 %	91.0237 %
Dataset 8	90.6115 %	91.1947 %
Dataset 9	88.8015 %	90.6929 %
Dataset 10	90.5189 %	92.6199 %
Dataset 11	90.3828 %	91.354 %
Dataset 12	91.0605 %	91.8453 %
Dataset 13	90.4003 %	91.8422 %
Dataset 14	89.4233 %	90.5376 %
Dataset 15	90.0872 %	91.2434 %

Similarly, Table III summarizes the accuracy results for both models to classify the multiclass attacks. As shown in Tables II and III, the better results of accuracy are achieved by the RSRT model with an important difference over all datasets.

2) Statistical Analysis of Accuracy Results: To test the significance of the accuracy results for RSKNN and RSRT models, we used the nonparametric Mann–Whitney T test drawing the statistical analysis. The nonparametric Mann–Whitney T test is more suitable for small sample size, robust for outliers, and does not depend on distributional assumptions [22]. The Mann–Whitney (WMW) T test takes all the observations of the two groups and ranks them in order of size. The ranks of each group are then summed, and the test statistic is computed as

$$T = R_1 - \frac{n_1(n_1+1)}{2} + R_2 - \frac{n_2(n_2+1)}{2} \tag{5}$$

where n_1, n_2, R_1, and R_2 are the size of sample 1, the size of sample 2, the sum of the ranks in sample 1, and the sum of the ranks in sample 2, respectively, by using the mean rank and sum of ranks for each group on the data. In this case, the best group has the rank one and the second best has the rank two. The testing question of this statistical analysis can be formulated to be "*Are there statistically significant differences in accuracy results achieved by RSKNN and RSRT models*?" We start the test by stating the hypothesis and categorize the claim as follows

Null Hypothesis: There are no statistical differences between the accuracy results of two models to classify the normal and abnormal events.

Alternate Hypothesis: There are statistical differences between the accuracy results of two models to classify the normal and abnormal events.

In Fig. 9, we illustrate the standard error of standard deviation for the accuracy results of normal and abnormal classification obtained by the two models groups. The SPSS statistical tool is used to perform the test. Tables IV–VI tabulate the descriptive statistics, ranks, and test statistics of the two models regarding the accuracy results of binary classification.

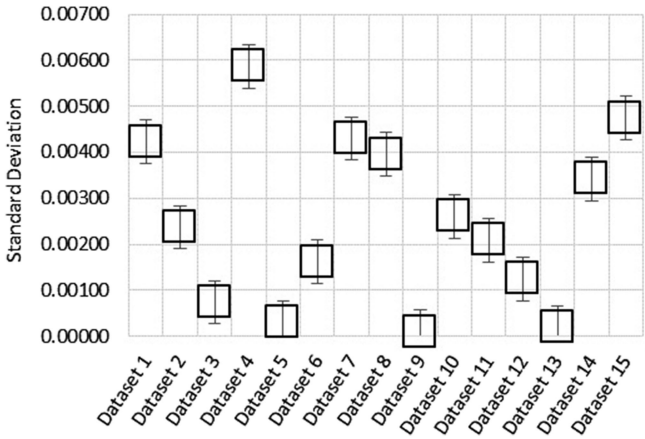


Fig. 9. Standard deviation for the accuracy results of normal and abnormal classification.

TABLE IV
DESCRIPTIVE STATISTICS OF ACCURACY RESULTS FOR BINARY CLASSIFICATION

N	Mean	Std. Deviation	Minimum	Maximum
30	95.9536	.5206386	95.1099	96.7748

TABLE V
RANKS OF RSKNN MODEL AND OUR RSRT MODEL FOR BINARY CLASSIFICATION COMPARISON

Model	N	Mean Rank	Sum of Ranks
RSKNN [17]	15	11.00	165.00
Our RSRT	15	20.00	300.00
Total	30		

TABLE VI
TEST STATISTICS OF ACCURACY RESULTS FOR BINARY CLASSIFICATION

Statistic	Value
Mann-Whitney U	45.000
Wilcoxon W	165.000
Z	-2.800
Asymp. Sig. (2-tailed p-value)	.005
Exact Sig. [2*(1-tailed Sig.)]	.004

From the output in Table VI, we can confirm that the two-tailed p-value is less than 0.05; then, we reject the null hypothesis and accept the alternate hypothesis at a 95% confidence interval. Hence, we can conclude that there are statistical differences between the accuracy results of the two models to classify the normal and abnormal events. From the results of the mean rank and sum of ranks in Table V, we can also conclude that these differences are for the RSRT model, proving its superiority against the RSKNN model.

Similarly, we formulate the hypothesis for the two models to classify the normal and different types of attacks in traffic events as follows.

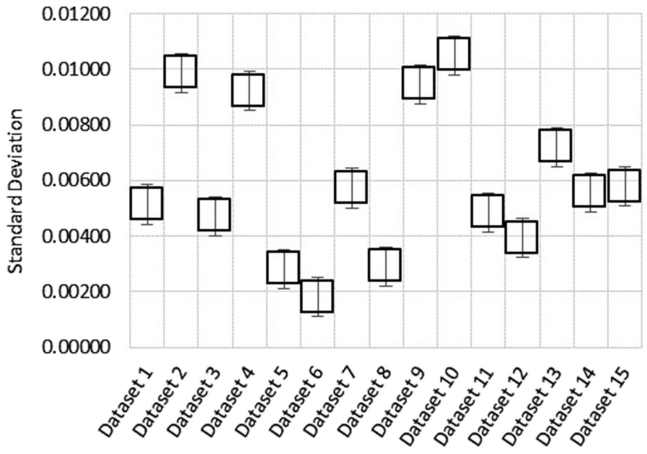


Fig. 10. Standard deviation for the multiclass classification accuracy results.

TABLE VII
DESCRIPTIVE STATISTICS OF ACCURACY RESULTS FOR MULTICLASSIFICATION

N	Mean	Std. Deviation	Minimum	Maximum
30	90.7141	.8371435	88.8015	92.6199

TABLE VIII
RANKS OF RSKNN MODEL AND OUR RSRT MODEL FOR MULTICLASSIFICATION COMPARISON

Model	N	Mean Rank	Sum of Ranks
RSKNN [17]	15	8.67	130.00
Our RSRT	15	22.33	335.00
Total	30		

TABLE IX
TEST STATISTICS OF ACCURACY RESULTS FOR MULTICLASSIFICATION

Statistic	Value
Mann-Whitney U	10.000
Wilcoxon W	130.000
Z	-4.252
Asymp. Sig. (2-tailed p-value)	.000
Exact Sig. [2*(1-tailed Sig.)]	.000

Null Hypothesis: There are no statistical differences between the accuracy results of two models to classify the normal and different types of attacks in traffic events.

Alternate Hypothesis: There are statistical differences between the accuracy results of two models to classify the normal and different types of attacks in traffic events.

Fig. 10 shows the standard error of standard deviation for the multiclass classification accuracy results of the two models. The outputs of the Mann–Whitney T test for accuracy results of multiclassification are stated in Tables VII–IX. Since the two-tailed p-value in Table IX is less than 0.05, we reject the null hypothesis and accept the alternate hypothesis at a 95% confidence interval. Therefore, we can conclude that there are statistical differences between the accuracy results of the two models to classify the normal and different types of attacks.

TABLE X
AVERAGE TIME COST OF TRAINING AND TESTING IN SECONDS FOR BINARY AND MULTICLASS ATTACKS DETECTION

Model	Training time on 3738 instances	Testing Time on 1602 instances	Task
RSKNN [17]	0.05	3.64	Binary class classification
	0.05	15.04	Multiclass classification
Proposed RSRT	0.22	0.01	Binary class classification
	0.36	0.02	Multiclass classification

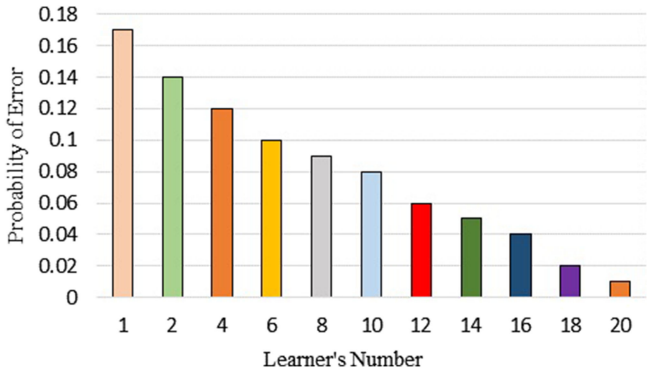


Fig. 11. Probability of error for 20 individual learners.

Moreover, the mean rank and sum of ranks of the RSRT model outperform clearly the RSKNN model.

3) Comparison of Computational Time Costs: To compare the efficiency of the RSRT and RSKNN models, we calculated the time cost of training and testing on the dataset 9, which contains 5340 instances. This dataset was divided into 3738 instances for training and 1602 for testing. The time cost was measured for binary and multiclass. The experiment is accomplished on a laptop Intel(R) Core(TM) i7-4510U with 2.0 GHz, 8 GB RAM, and Windows 10 OS. Table X demonstrates the average time cost of training and testing on the selected dataset.

We can see that the training time of RSRT is slightly higher than the training time of the RSKNN model. This is because the RSKNN model does not build any trained model. However, the testing time of the RSRT model for detection is significantly less than the testing time of the RSKNN model. This makes the proposed RSRT model is efficient for real-time intrusion detection in IIoT.

4) Analyzing Reliability and Trustworthiness of the Proposed Model: To analyze the reliability of the proposed model, imagine that the RSRT model has an ensemble of 20 learners. Due to the diversity property of ensemble learning, the errors made by these learners are uncorrelated. Therefore, if less than half of learners are wrong, the others may be correct, so that the majority vote will correctly classify the intrusion attack occurred in the SCADA-based IIoT. Fig. 11 demonstrates a simulated probability of error for 20 individual learners, each of them has an error less than or equal 0.17.

By using a majority vote, we can see that 12 learners have a probability of error less than 0.09 to classify the intrusion attack

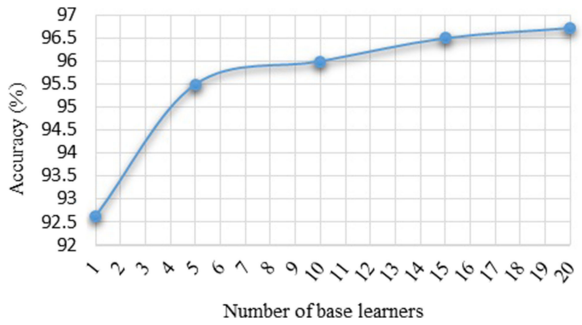


Fig. 12. Accuracy results of the proposed RSRT model at the different number of base learners.

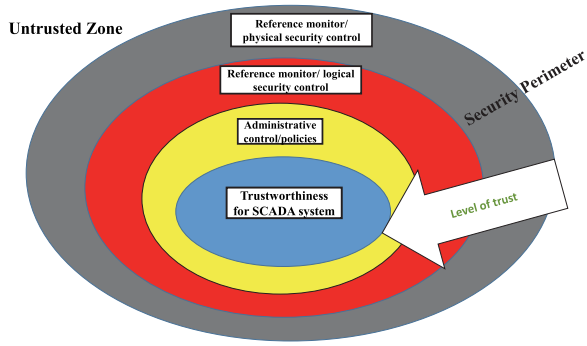


Fig. 13. Defense-in-depth strategy in the TCB security model to provide the assurance of trustworthiness.

that makes the model is reliable to detect the attacks in SCADA-based IIoT. For validating the trustworthiness property of the proposed model, we conducted an experiment on the dataset 1 to classify the normal and abnormal traffic using a different number of base learners. Fig. 12 presents the accuracy results of the proposed model using an ensemble of 20 base learners against a single base learner. From Fig. 12, we can see how the accuracy is improved based on the majority voting of different base learners that achieves the trustworthiness property of the proposed RSRT model.

In addition, the trustworthiness of the proposed SCADA system can also be explained by presenting the proposed SCADA model using a mapping of trusted computing base (TCB) model to defense-in-depth model and how the confidentiality-integrity-availability (CIA) are preserved through this mapping.

The proposed SCADA model embeds the concept of TCB security model (see Fig. 13). Inside security perimeter, hardware and software, security control, and policies constitute the components of the trusted zone which are combined to ensure the preservation of CIA triad and overall security system contributes toward trustworthiness. The reference monitor/physical security control of the TCB/SCADA model deters and blocks different unauthorized access and activities to the resources inside the perimeter of the trusted zone. Generally automated physical access control systems (PACS) such as motion detector or CCTV cameras or mantraps are part of this layer. However, SCADA systems and related subsystems are mostly located in remote places where the implementation of PACS is difficult. Therefore, a defense-in-depth strategy for this situation needs to be compensated through additional measures such as implementing IDSs or antimalware solutions or access control list based control in the logical control, which is the second layer in Fig. 13.

However, such conventional preventive or detective security controls fail to prevent unauthorized access as they are based on a set of protocols or application program interfaces (APIs) that are different from the SCADA system. Therefore, for ensuring defense-in-depth strategy and thereby improving the trustworthiness of the SCADA system, reliable and scalable security control based on SCADA network protocols and network traffic analysis need to be developed. In our proposed model, we addressed these issues and have developed a reliable and salable cyber attack detection model. We have also validated our model based on a large SCADA network traffic with different types of attacks that exploit different vulnerabilities of SCADA components and the overall systems.

V. Conclusion

Safeguarding of SCADA-based industrial networks from cyberattacks improves the trustworthiness of IIoT networks. However, existing security methods and software products are still not efficient and accurate for safeguarding SCADA-based IIoT networks. In this article, an improved ensemble-learning model for detecting the cyberattacks of SCADA network traffic was proposed. In particular, a scalable and reliable ensemble detection model based on the combination of the RS learning method and RT was developed. The proposed model had been tested over 15 datasets of the SCADA network and significant improvement in classification accuracy was achieved. The results showed a good tradeoff between the model complexity, classification accuracy, and reliability, thereby ensuring better performance compared to previous state-of-the-art methods. However, our RSRT model had a limitation for selecting optimal random feature subsets when there is a very small number of features. Another limitation was related to the execution time which is slightly more than the execution time of a single RT classifier. In the future, we will enhance our model to address these issues.

References

[1] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial Internet of Things (IIoT): An analysis framework," *Comput. Ind.*, vol. 101, pp. 1–12, 2018.

[2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[3] R. Casadei, G. Fortino, D. Pianini, W. Russo, C. Savaglio, and M. Viroli, "Modelling and simulation of opportunistic IoT services with aggregate computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 252–262, 2019.

[4] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.

[5] Y. Luo, Y. Duan, W. Li, P. Pace, and G. Fortino, "A novel mobile and hierarchical data transmission architecture for smart factories," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3534–3546, Aug. 2018.

[6] C. Sanin, Z. Haoxi, I. Shafiq, M. M. Waris, C. S. de Oliveira, and E. Szczerbicki, "Experience based knowledge representation for Internet of Things and cyber physical systems with case studies," *Future Gener. Comput. Syst.*, vol. 92, pp. 604–616, 2019.

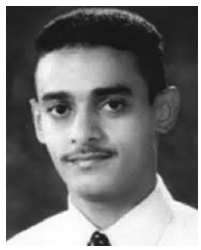
[7] H. Sándor, B. Genge, Z. Szántó, L. Márton, and P. Haller, "Cyber attack detection and mitigation: Software defined survivable industrial control systems," *Int. J. Crit. Infrastruct. Protection*, vol. 25, pp. 152–168, 2019.

[8] M. S. Mahmoud, M. M. Hamdan, and U. A. Baroudi, "Modeling and control of cyber–physical systems subject to cyber attacks: A survey of recent advances and challenges," *Neurocomputing*, vol. 338, pp. 101–115, 2019.
[9] D. Wang, X. Wang, Y. Zhang, and L. Jin, "Detection of power grid disturbances and cyber-attacks based on machine learning," *J. Inf. Secur. Appl.*, vol. 46, pp. 42–52, 2019.
[10] C. Gavriluta, C. Boudinet, F. Kupzog, A. Gomez-Exposito, and R. Caire, "cyber–physical framework for emulating distributed control systems in smart grids," *Int. J. Elect. Power Energy Syst.*, vol. 114, 2020, Art. no. 105375.
[11] S. Huda *et al.*, "Defending unknown attacks on cyber–physical systems by semi-supervised approach and available unlabeled data," *Inf. Sci.*, vol. 379, pp. 211–228, 2017.
[12] S. Rahimi and M. Zargham, "Analysis of the security of VPN configurations in industrial control environments," *Int. J. Crit. Infrastruct. Protection*, vol. 5, no. 1, pp. 3–13, 2012.
[13] S. Huda, J. Abawajy, B. Al-Rubaie, L. Pan, and M. M. Hassan, "Automatic extraction and integration of behavioural indicators of malware for protection of cyber–physical networks," *Future Gener. Comput. Syst.*, vol. 101, pp. 1247–1258, 2019.
[14] S. Huda, J. Yearwood, M. Hassan, and A. Almogren, "Securing the operations in SCADA-IoT platform based industrial control system using ensemble of deep belief networks," *Appl. Soft Comput.*, vol. 71, pp. 66–77, Oct. 2018.
[15] *Industrial Internet of Things, Volume G4: Security Framework*. Needham, MA, USA: Ind. Internet Consortium, 2016.
[16] R. A. Abouhogail and M. S. Gadelrab, "A new secure and privacy preserved protocol for IEEE802. 11s networks," *Comput. Secur.*, vol. 77, pp. 745–755, 2018.
[17] *Information Technology-Security Techniques-Information Security Risk Management*, ISO/IEC 27005:2018. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:27005:en, Accessed on: Sep. 15, 2019.
[18] G. Li, Y. Shen, P. Zhao, X. Lu, J. Liu, Y. Liu, and S. C. Hoi, "Detecting cyberattacks in industrial control systems using online learning algorithms," *Neurocomputing*, vol. 364, pp. 338–348, 2019.
[19] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *Proc. 7th Int. Symp. Resilient Control Syst.*, Denver, CO, USA, Aug. 2014, pp. 1–8.
[20] A. Derhab *et al.*, "Blockchain and random subspace learning-based IDS for SDN-enabled industrial IoT security," *Sensors*, vol. 19, no. 14, 2019, Art. no. 3119.
[21] U. Adhikari, S. Pan, T. Morris, R. Borges, and J. Beave, "Industrial control system (ICS) cyber attack datasets." [Online]. Available: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets Accessed on: Aug. 15, 2019.
[22] G. W. Zeoli and T. S. Fong, "Performance of a two-sample Mann–Whitney nonparametric detector in a radar application," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-7, no. 5, pp. 951–959, Sep. 1971.

Mohammad Mehedi Hassan (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Kyung Hee University, Seoul, South Korea, in February 2011.

He is currently an Associate Professor with the Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He has authored and coauthored around 180+ publications including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include cloud computing, edge computing, Internet of Things, body sensor network, big data, deep learning, mobile cloud, smart computing, wireless sensor network, 5G network, and social network.

Dr. Hassan is a recipient of a number of awards including the Best Journal Paper Award from the IEEE SYSTEMS JOURNAL in 2018, Best Paper Award from CloudComp in 2014 conference, and the Excellence in Research Award from King Saud University (two times in a row, 2015 and 2016).



Abdu Gumaei received the Ph.D. degree in computer science from King Saud University, Riyadh, Saudi Arabia, in 2019.

He is currently an Assistant Professor with the College of Computer and Information Sciences, King Saud University. He worked as a Lecturer and taught many courses such as programming languages at the Computer Science Department, Taiz University, Yemen. He has authored and coauthored more than 30 journal and conference papers in well-reputed international journals. He has received a patent from the United States Patent and Trademark Office (USPTO), in 2013. His main areas of interest are software engineering, image processing, computer vision, machine learning, networks, and Internet of Things (IoT).

Shamsul Huda received the Ph.D. degree in computer science from the Center for Informatics and Applied Optimization (CIAO), Federation University, Australia in 2010.

He is currently a Lecturer with the School of Information Technology, Deakin University, Melbourne, VIC, Australia. Previously, he also worked in Federation University as a Research Fellow. He worked as an Assistant Professor with the Computer Science Department, Khulna University of Engineering and Technology, Bangladesh. He has authored or coauthored more than 50 journal and conference papers in well-reputed journals including IEEE transactions. His main research interests include information security, cyber–physical systems, computational intelligence, and machine learning.

Ahmad Almogren (Member, IEEE) received the Ph.D. degree in computer science from Southern Methodist University, Dallas, TX, USA, in 2002.

He is currently a Professor with the Computer Science Department, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. He is also the Director of the Cyber Security Chair at CCIS, KSU. He was the Vice Dean for the Development and Quality at CCIS. He also served as the Dean of the College of Computer and Information Sciences and the Head of Academic Accreditation Council at Al Yamamah University. He served as the General Chair for the IEEE Smart World Symposium and a Technical Program Committee member in numerous international conferences/workshops such as IEEE CCNC, ACM BodyNets, IEEE HPCC. His research interests include mobile-pervasive computing and cybersecurity.