Forecasting Intrusion in Critical Power Systems Infrastructure Using Advanced Autoregressive Moving Average (AARMA) Based Intrusion Detection for Efficacious Alert System

Neeraj Kumar Singh, Mahshooq Abdul Majeed and Vasundhara Mahajan*
Electrical Engineering Department, Sardar Vallabhbhai National Institute of Technology,
Surat, Gujarat, India-395007

Corresponding Author Email: vasu.daygood@gmail.com

Abstract: Cyber intrusions into critical infrastructure inflict economic and physical damage. Extensive research is needed to identify and mitigate intrusions in power grid infrastructure. The modern solution is to use a data science time-series approach to identify the intrusion based on the electric grid data collected from the sensors. This paper addresses the new vision of the data science time-series modelling approach to integrate it with the existing power system security system. In this paper, the Advanced Autoregressive Moving Average (AARMA) model is designed to detect the possible intrusion of the given data set. An attack forecast is a model to predict possible cyber intrusions using real-time data input from sensors. By investigating the statistical properties of the sensors' data set, intrusion detection is possible with a high accuracy of about 90%. Using AARMA, the operators have the benefit of an effective alert system to adjust their configuration and other resource allocation to tackle intrusions with low impact. MATLAB software is used to monitor the IEEE 9-bus and IEEE 33-bus test systems against possible cyber-attacks using the proposed AARMA model.

Keywords: Critical Infrastructure (CI), Cyber intrusion, Advanced Autoregressive Moving Average (AARMA), Statistical properties, Critical Power Systems Infrastructure (CPSI)

I. INTRODUCTION

Critical power system infrastructure plays a crucial role in providing an uninterrupted power supply to different loads [1, 2]. With new technology such as electric vehicles, renewables, and increased digitalization [3, 4]. The prevalent nature of cyber intrusion is to get maximum access and to inflict destructive economical and physical damage to the system and customers. Cyber-attacks may appear to be natural occurrences. As a result, distinguishing harmful from non-malicious data in the communication system is tough and difficult [5]. Through cyber-attack invaders tries to disrupt, network jam, deny and block system operations of Critical Infrastructure (CI) [6]. To protect critical power system infrastructure a predictive model is required which includes data science tools for data mining of real-time and historical data to analyze the statistical properties of cyber intrusion. The predictive analytics of data science technique to forecast or the capability to predict cyber intrusions ahead of a time helps the control center to prepare a defence strategy against the attack.

In contrast, nearly all prior work on cybersecurity of power systems CI lacked a statistical model for future behaviour and forecasting. For power system security, the two main approaches commonly used are state estimation-based security and analysis-based security. State estimation provides accurate, secure, and fast computation of the states of equipment, which helps to execute complex decisions for safe power system operations. But in the case of forecasting cyber intrusion and attack, the research void to be filled is quite large. Analysis-based security is based on the signature-based module to compare the data packet with a predefined signature. The system will pass the data to the next stage if the signature match is found. To make the system more accurate, an anomaly-based module is incorporated. Analysis-based security and estimation-based security lack the ability to forecast intrusions. The size of the feature set in intrusion detection applications can have a significant impact on the detection process speed and accuracy. More features need more memory, more processing time, and maybe a higher noise-to-signal ratio. Therefore, having more features does not always imply better performance. [7, 8].

Data science methods in cybersecurity focus on the statistical aspect of cyber intrusion, and can potentially benefit in the prediction of cyber threats. Figure *1* shows the different benefits that can be obtained through data science. Of these, prediction analytics and forecasting are the most crucial benefits. These two are the key features that are used for predictive modeling for forecasting intrusion in power system CI. Using these two features, researchers in [9], proposed a cyber-attack detector for CI using gray-box prediction. The proposed model shows an accuracy of 63%-73% approx. for long-range dependence systems.

Most of the research work is based on a detection scheme, only limited research has been done against predictive analytics and forecasting of cyber-attack [10, 11]. It is due to anyone of the reason stated below:

- Lack of historical and real-time data of cyber intrusion.
- In many cases, it is considered as the cyber intrusion is unpredictable.
- Non-availability of a predictable model for cyber-attack.
- Lack of operator knowledge against cyber intrusion.

In some research work [12], the authors focus to find different patterns associated with the different cyber intrusion. The model uses historical data to identify the type of attack but the system was incapable of forecasting cyber-attack. Artificial neural networks (ANNs) are commonly used as classifiers in machine learning because of their simplicity and effectiveness. It has also been used in power system intrusion detection models. The ANN's training is still a difficult problem. Traditional training algorithms have a hard time dealing with slow convergence and local optima [13, 14]. Meanwhile, because the graph method does not require model training and the impact of node changes is restricted, it outperforms the other algorithms in terms of scalability. However, the attack graph analysis method requires more work in terms of integration, integration with large data technologies, and the capacity to account for uncertainty [15, 16]. The use of game theory intrusion detection approaches to tackle security resource allocation challenges in large-scale heterogeneous networks is examined. It is worth noting that in most game theory-based models, the defender scan is supposed to always correctly identify the attackers malicious activities without making any mistakes, which may not be the case in some circumstances [17, 18]. To overcome the drawbacks of ANN, attack graph, and game theory, a new concept was introduced based on time series. However, a major flaw in the time series-based EWMA was discovered: if the attack lasts for a longer length of time without oscillations in the quantity of packages, the EWMA misidentifies it as routine traffic [19, 20]. Also, time series analysis has a number of flaws, including difficulties generalising results from a single study, getting proper metrics, and effectively finding the best model to represent the data [19, 21, 22].

Data science time-series forecasting has been used for many predictive cyber intrusion models [23, 24]. Only a few models [25] work as an early warning system with an accuracy of up to 75%. In reference [26], the researcher suggested three categories of models using time-series which exploit the statistical properties of data. The main disadvantage of this model is the non-identification of the outperform model during the intrusion detection process. In [27], the authors proposed a proactive security system to forecasting intrusion based on Distributed Denial of Service (DDoS) cyber-attack. The main finding of this research was the time of detection, attack target, and forecasting having low accuracy compared to AARMA.

In reference [28], researchers proposed forecasting using Bayesian network inference, in which the system calculates the probability of the next cyber intrusion. This study was based on previous attacks observation. In [29], the article highlights the use of time-series analysis to forecast the system variation concerning time. In a similar manner [30], the authors proposed Markov Hidden Model for predicting attacks. Also, in [31] AARMA model is used instead of Markov Model. Some more research details which can be used for Critical Power Systems Infrastructure are discussed in Table *1*. The most common model is the graph-based attack technique [32], which is used to model the system using graph nodes. The system is used to achieve an accuracy of 85%. The main drawback of the system was that the prediction was very limited to a few seconds. Then the second most popular model used to forecast cyber intrusions is the Bayesian-based attack graph. It is similar to the graph-based system, but in this model, one node acts as an intruder. But it faces problems when implementing a large network. The Markov chain and game theory models are used to give very low accuracy, even in simple networks. Neural networks and evolutionary computing are something new to this category of intrusion detection systems. It can be used to detect similarity in a system. The detailed comparison is shown in the table 1 below. The evaluation is done either on a testbed created in the lab or simulation or on real-time data sets. These methods have been used for the past few years to predict or forecast the intrusion behaviour of the system.

The literature survey highlights the drawback of previous research work. The maximum accuracy for cyber intrusion prediction is around 85% using either the graph-based approach or using a neural network approach.

From the above literature review, it is clear that intrusion detection systems need a methodology to predict the intrusion based on historical and present data sets. The Autoregressive Moving Average model predicts or forecasts malicious behaviour by combining statistical studies with precisely acquired historical data points. To analyse the data and create future predictions, the model employs time-series data and statistical analysis, making forecasting more accurate. Also, the other methods were based on single attacks and defence systems, but by using the Autoregressive Moving Average model, multiple attacks can be identified, making this method more suitable than the others.

The main objective of this research is to forecast possible intrusions using a time-series historical data set. This paper makes three crucial contributions to the security of critical power systems infrastructure and they are as follows:

1. First, AARMA is a model for predicting cyber intrusion incidents from the data sets provided by the system.
2. Second, time series modelling is done for short-term prediction which helps the operator to predict the intrusions ahead of time.
3. Third, the proposed model is evaluated on the standard IEEE 9-bus test system and IEEE 33-bus test system the accuracy of the model is estimated.

II. MODELING OF PROPOSED AARAM

In this section, the proposed AARAM modelling for intrusion forecasting is discussed in detail. In a time-series set of observation *O_t* from voltage and frequency of a smart grid, collected and recorded through smart devices at a specific time *t* is denoted by [O_t]. The observation O_t can be expressed as:

$$O_t = c_t + p_t + N_t \tag{1}$$

$$O_t \in \{V\} or \{F\}$$

Where t= 1, 2, 3,…t_n , c_t is the trend component of O_t, p_t represents periodic component of O_t, N_t is the stationary random noise component. It is necessary to estimate and eliminate the trend component and periodic component to make the noise component stationary in time-series. Using time-series modeling, prediction of data series becomes easy when the system consists of random components. ARMA [31] is used to model time-series data which can be used for forecasting. The standard autoregressive equation is generated by:

$$O_t - \beta O_{t-1} = U_t \tag{2}$$

Where [O_t] is time-series observation, |β|< 1 and [U_t] represent uncorrelated random variables having mean zero and variance σ^2 . With the help of backshift operator (B_s) and identity operator (I) equation 2 is modified as:

$$(I - \beta B_s)O_t = U_t \tag{3}$$

By introducing trust value of sensors [T_s; T_s>0.4 [32]], equation 3 is rewritten as:

$$(I - \beta B_s)^{T_s} O_t = U_t \tag{4}$$

Using equation 4 moving average is generated by:

$$O_t = (I - \delta B_s)U_t \;\; : \delta \to \left(|\delta| < 1\right) \tag{5}$$

To form generalised autoregressive moving average with an addition parameter T_a, the equation is written as:

$$(I - \beta B_s)^{T_s} O_t = (I - \delta B_s)^{T_a} U_t \qquad : T_a > 0 \tag{6}$$

The test is performed on CPSI where voltage and frequency parameter plays a major role in terms of security monitoring. Using the voltage and frequency data sets collected from the sensors placed near each bus of CPSI, equation 6 is modified as:

$$Voltage\ data\ set \rightarrow \{V_t\} \tag{7}$$

$$Frequency\ data\ set \rightarrow \{F_t\} \tag{8}$$

$$(I-\beta B_s)^{T_s} V_t = (I-\delta B_s)^{T_a} U_t \qquad : T_a > 0 \tag{9}$$

$$(I-\beta B_s)^{T_s} F_t = (I-\delta B_s)^{T_a} U_t \qquad : T_a > 0 \tag{10}$$

The proposed AARMA is basically a combination of the trust value of sensors and the autoregressive moving average of the data collected from the system. For forecasting purposes, model parameters estimation is required which forms the critical part of the AARMA, for which estimation algorithm is required. In section *A*, Hannan-Rissanen estimation algorithm [33] technique is described which is used to evaluate preliminary parameters for AARMA modeling.

A. Preliminary parameter evaluation using Hannan-Rissanen estimation algorithm

For modelling AARMA(*x*,*y*) for CPSI voltage data sets, Hannan-Rissanen estimation algorithm technique is used as follow:

$$V_t - \theta_1 V_{t-1} - \theta_2 V_{t-2} - ...\theta_x V_{t-x} = U_t - \vartheta_1 U_{t-1} - \vartheta_2 U_{t-2} - ...\vartheta_y U_{t-y} \tag{11}$$

In the equation 11, θ and $\boldsymbol{\vartheta}$ represents the vectors of estimated coefficients which is determined by minimizing the sum of squares:

$$S(\omega \rightarrow (\theta, \vartheta)) = \sum_{t=1+x+y}^{n} (V_t - \theta_1 V_{t-1} - \theta_2 V_{t-2} - ...\theta_x V_{t-x} + \vartheta_1 U_{t-1} + \vartheta_2 U_{t-2} + ...\vartheta_y U_{t-y}) \tag{12}$$

It can be noted that to obtain better results the parameters θ and $\boldsymbol{\vartheta}$ can be manipulated during the modelling of AARMA. For example the fitted AARMA (1,*y*) can be represented by equation:

$$V_t - \theta_1 V_{t-1} = U_t - \vartheta_1 U_{t-1} - \vartheta_2 U_{t-2} - ...\vartheta_y U_{t-y} \tag{13}$$

By using the backshift operator the AARMA (*x*,*y*) model is easily represent by:

$$\theta(B_s) V_t = \vartheta(B_s) U_t \tag{14}$$

The estimation of parameters used in equation 13 is possible by assuming:

$$T_s = \delta \tag{15}$$

It should be noted that the sensor which are working properly should have trust (T_s) more than 0.4. So δ can be estimated using equation:

$$\delta = \frac{\vartheta_1}{T_s} \tag{16}$$

The simplest model of AARMA can be expressed using equation 4 and 5:

$$V_t - \beta B_s V_{t-1} = U_t - T_s \delta U_{t-1} \tag{17}$$

In similar manner for AARMA (*x*,*y*) the corresponding variance can be expressed as:

$$\sigma^2 = S(\omega / (n - x - y)) \tag{18}$$

The above equations use voltage data set, in similar manner frequency data set is used to model the equation from 7 to 13. For any model, it is important to determine the parameters which describe the data set being used. For AARMA model Maximum Likelihood Estimation (MLE) is used for parameter estimation for the data set.

B. Maximum Likelihood Estimation (MLE) for AARMA

MLE is a statistical tool for parameter estimation and it is used for AARMA modelling. The Likelihood function for the time-series used for modelling AARMA is expressed as:

$$L(\omega, \sigma^2) = \frac{\exp(-\frac{1}{2} V^{T_p} \Gamma_n^{-1} V)}{\sqrt{(2\pi)^n \det(\Gamma_n)}} \tag{19}$$

Where Γ_n represents auto-covariance matrix. In AARMA to maximize *L* estimation of ω, σ^2 is done using equation 16:

$$LL = -\frac{n}{2} In(2\pi) - \frac{1}{2} In \det(\Gamma_n) - \frac{1}{2} V^{T_p} \Gamma_n^{-1} V \tag{20}$$

The above equation produces the maximum likelihood estimation for the proposed model.

C. Predefine Logic for intrusion detection used for AARMA

The AARMA model uses predefine logic stated in Table 2 for detecting data status as malicious or non-malicious. The AARMA logic uses the processed data collected by sensors attached to each bus of IEEE 9-bus test system. To determine the predefined logic, three priorities are considered for data status evaluation which are named as:

- Sensor Trust Value (T_s): In proposed AARMA the sensor trust is evaluated by processing the real time data and comparing with the historical data. The processed trust value is passed through threshold limit to identify the violation of equation 21. If the value is higher than 0.4 means the sensor is non-malicious otherwise vice-versa.
- Signature based detection module (S_d): Through predictive analysis of historical data sets all possible predefined pattern is stored and match with the incoming real time data pattern. Any possible mismatch alert the system with signature based detection module.
- Anomaly based detection module (A_d): Anomaly behaviour of the network like getting more frequent data request or getting sudden rise in data flow in the sensor network against the predefined behaviour triggers alert for malicious intrusion.

All three modules used by the proposed AARMA makes the system more accurate and sensitive to cyber intrusion. Figure *2* shows the priority stage for AARMA modelling. For normal operation, when the change in the data set is within an acceptable range, then the value of all the three priorities is indicated by 1. Under the abnormal condition, when data is exceeding the pre-defined acceptable range then the value of the given priorities change to 0. For example, if the data set is under acceptable range and Signature-based detection module is used then by predictive analytics, the priorities 1 otherwise 0 (in case if it matches with some past historical malicious data/attack). Similarly:

$$T_s \cong 1 \quad \text{(If sensor trust values > 0.4 otherwise 0)} \tag{21}$$

$$A_d \cong 1 \text{ (If no match found with predictive analytic and data is under acceptable range)} \tag{22}$$

So finally, the flow chart of the AARMA model is shown in Figure *3* using all the steps discussed in the above sections. In the first step, data is collected from the sensors present in the system under test. Data consists of voltage, frequency, and calculated trust value of the sensors. Then all the data set is passed through the AARMA model to predict and forecast the cyber intrusion ahead of time. To gather raw data for cyber intrusion in power systems, IEEE 9-bus test system is used which is described and discussed in reference [32]. Table *3* shows the attack severity in the AARMA model. When the predictive analytics result (***S_d, A_d***) is 0, it means the attack intensity is high. Similarly, by following figure 2 the attack intensity can be identified.

III. RESULTS AND DISCUSSION

This section presents the attack prediction/forecasting on the IEEE 9-bus test system using the proposed AARMA model. In this experiment, some assumptions had been considered which are as follow:

- Each bus consists of one sensor which monitoring voltage and frequency data.
- Maximum load is fixed.
- Attacker inject false data to break the healthy system operation.

The simulation analysis of the proposed model is conducted on the IEEE 9-bus test system using MATLAB software. For understanding purpose AARMA (1,2) and AARMA (5,5) is used for forecasting error but for overall evaluation and forecasting accuracy AARMA (5,5) model is used.

A. AARMA (1,2) Model

An auto-correlation plot indicates the observation of a single variable over a given period. It helps to understand whether the elements of a time series under study are correlated to each other. The data elements can be independent, negative correlated, or may be positively correlated. Figure 4 shows the auto-correlation for AAMRA (1, 2) which indicates at lag 2 shows a good positive correlation.

The autocorrelation function and partial correlation function for forecasting the intrusion are shown in Figure 4. Also, the data fitting errors and forecasting error for cyber intrusion prediction is shown in Figure 5 and Figure 6.

B. AARMA (5,5) Model

AARMA (5, 5) model is much better in forecasting than AARMA (1, 2). Figure 7 shows the fitting errors in data sets for forecasting the intrusion. Similarly, Figure 8 shows the forecasting errors for time of 10000 seconds. The average fitting error for AARMA (1, 2) is 0.2154, but with AARMA (5, 5) it's 0.1964. The average error for the prediction/forecasting is slightly higher for AARMA (1, 2) [0.2011] as compared with AAMRA (5, 5) model having a value of 0.1919. This is the main reason for using AARMA (5, 5) model rather than AARMA (1, 2) model.

C. Cyber intrusion prediction and forecasting using AARMA (5,5)

To evaluate the accuracy of AARMA (5, 5) model two following cases are considered:

1. Case 1: Cyber intrusion at one bus of IEEE 9-bus test system.
2. Case 2: Cyber intrusion at multiple bus of IEEE 9-bus test system.

In both cases, five trials are done with a different number of cyber intrusions. For understanding purpose through simulation, bus number 7 forecast the single cyber intrusion at time 2500 seconds. The proposed AARMA model forecast the intrusion with almost 98% of accuracy. Figure 9 and Figure 10 show the simulation result which highlights the actual and forecast values to be almost the same with a very low error margin. Table 4 and Table 5 show two different cases as stated above. Figure 11 shows the attack severity analysis of case 1 when one bus is under cyber-attack. Similarly, Figure 12 shows the forecasting against cyber intrusion at multiple buses (Bus 5 (B5) and Bus 7 (B7)). It can be seen that the error of actual identified intrusion and it forecast is less than 2%. Figure 13 highlights severity analysis of cyber-attack on multiple bus networks.

Apart from IEEE 9-bus test system the proposed model is used to validate the model through IEEE 33-bus test system. Figure 14, shows the attack severity analysis for the IEEE 33-test bus system with five different trials. Figure 15, shows the accuracy of the model for IEEE 33-test bus system. Due to lack of data the accuracy for was found to be 87%. The proposed AARMA model is very effective and efficient to forecast the intrusion using the present and historical data sets. The three modules used to define the logic make this model non-vulnerable to new novel attack whose information are not present in the historical events. The average standard error for single bus attack and multiple bus attack is tabulated in table 6. It is used to compare how the forecasted value differs from the actual value. It can be observed that forecasting for a single bus attack is more effective when compared with a multiple bus attack, this is due to large data set handling for multiple bus intrusion.

D. Comparison of AARMA model with existing Models

For comparison of AARMA model with existing models four parameters are evaluated as follow:

- Is the system able to implement the model on critical power system infrastructure (P1).
- Can predict multiple attack on the system (P2).
- Complexity of the model (P3).
- Can identify the intensity of attack (P4).
- Accuracy of prediction/forecast (P5).

Table 7, shows the comparison of different parameters of the existing model with AARMA. The models stated in table 7 have been run on the IEEE 9-bus test system data set. Accordingly, the result is stated below.

IV. CONCLUSION

In this paper, a new modified approach is presented for predicting and forecasting cyber intrusions in critical power systems infrastructure (CPSI). The proposed methodology utilises a novel integration of the trust weight method and auto-regression moving average with the aim of predicting cyber intrusion in the IEEE 9-bus test system. The proposed AARMA model forecasts the cyber intrusion with a high accuracy of around 90% under single bus attack and multiple bus attack scenarios. Three important findings emerged from this experiment. First, the forecasting of cyber intrusion using time-series is an appropriate approach with high accuracy. Second, the proposed model helps to predict intrusion ahead of time, which helps to make an effective alert system. Third, the intensity of the attack can be determined using the attack severity analysis. The standard error for cases 1 and 2 shows that the proposed algorithm has the ability to perform efficiently. Despite this, the current results suggest that the proposed intrusion detection system can identify single and multiple assault scenarios successfully.

AARMA achieves very high accuracy with substantially shorter training and detection times, according to test results. As a result, it can be considered the most appropriate algorithm for the suggested intrusion detection system since it provides the best mix of higher accuracy and reduced processing time. In the future, research must focus on increasing the prediction or forecast accuracy above 95%. To attain high accuracy, Big Data Analytics can be an effective tool for data management and predictive modules.

REFRENCES

1. Ling, E., Lagerström, R. and Ekstedt., M. "A Systematic Literature Review of Information Sources for Threat Modeling in the Power Systems Domain". in *International Conference on Critical Information Infrastructures Security*. Springer, 2020.
2. Shojaei Berjouei, A., Moallem, M. and Manshaei, M.J.S.I. "A holistic day-ahead distributed energy management approach: Equilibrium selection for customers' game". **27**(3): p. 1437-1449, 2020.
3. Bilal, M. and Rizwan, M.J.S.I. "Intelligent Algorithm based Efficient Planning of Electric Vehicle Charging Station: A Case Study of Metropolitan City of India". 2021.
4. Baidya, S., Potdar, V., Ray, P. P., et al. "Reviewing the opportunities, challenges, and future directions for the digitalization of energy". *Energy Research & Social Science*, 81, 102243, 2021.
5. Prasad, G., Huo, Y., Lampe, L., et al. "Machine learning based physical-layer intrusion detection and location for the smart grid". In *2019 IEEE International Conference on*

Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 1-6). IEEE, 2019.

6. Sun, C. C., Cardenas, D. J. S., Hahn, A., et al. "Intrusion detection for cybersecurity of smart meters". *IEEE Transactions on Smart Grid*, 12(1), 612-622, 2021
7. Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., et al. "Machine learning and deep learning in smart manufacturing: The smart grid paradigm". *Computer Science Review*, 40, 100341, 2021.
8. Fadlullah, Z.M. and Fouda, M.M. "Combating Intrusions in Smart Grid: Practical Defense and Forecasting Approaches, in Combating Security Challenges in the Age of Big Data". Springer. p. 215-235, 2020.
9. You, J., Lv, S., Hao, Y., Feng, X., et al. "Characterizing internet-scale ics automated attacks through long-term honeypot data". In *International Conference on Information and Communications Security* (pp. 71-88). Springer, Cham, 2019.
10. Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing* 50: 159-17, 2003..
11. Xia, H., Zhang, S. S., Li, Y., et al. "An attack-resistant trust inference model for securing routing in vehicular ad hoc networks". *IEEE Transactions on Vehicular Technology,* 68(7), 7108-7120, 2019.
12. Karimipour, H., Dehghantanha, A., Parizi, et al. "A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids". *IEEE Access*, 7, 80778-80788, 2019.
13. Gamage, S., Samarabandu, J.J.J.o.N., "Deep learning methods in network intrusion detection: A survey and an objective comparison". **169**: p. 102767, 2020.
14. Zhang, Y., Jin, S., Cui, X., et al. "Network security situation prediction based on BP and RBF neural network". In *International Conference on Trustworthy Computing and Services* (pp. 659-665). Springer, Berlin, Heidelberg, 2012.
15. Zeng, J., Wu, S., Chen, Y., Zeng, et al. "Survey of attack graph analysis methods from the perspective of data and knowledge processing". *Security and Communication Networks*, 2019.
16. Cao, P., Badger, E., Kalbarczyk, Z., et al. "Preemptive intrusion detection: Theoretical framework and real-world measurements". In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security* (pp. 1-12), 2015.
17. Wang, Z., Xu, S., Xu, G., Yin, Y., et al. "Game theoretical method for anomaly-based intrusion detection". *Security and Communication Networks*, 2020.
18. Lisý, V., Píbil, R., Stiborek, J., et al. "Game-theoretic approach to adversarial plan recognition". In *ECAI* 2012 (pp. 546-551). IOS Press, 2015.
19. Babić, I., A., Čabarkapa, M., Nikolić et al., "Triple Modular Redundancy Optimization for Threshold Determination in Intrusion Detection Systems". *Symmetry* **13**(4): p. 557, 2021.
20. Silva, A., Pontes, E., Zhou, F., et al. "PRBS/EWMA based model for predicting burst attacks (Brute Froce, DoS) in computer networks". In *Ninth International Conference on Digital Information Management (ICDIM 2014)* (pp. 194-200), 2014.
21. Thakkar, A. and Lohiya, R.J.A.I.R., "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions". p. 1-111, 2021.
22. Zhan, Z., Xu, M., and S. Xu, "Predicting cyber attack rates with extreme values". *IEEE Transactions on Information Forensics and Security*. **10**(8): p. 1666-1677, 2015.
23. Werner, G., Yang, S., and K. McConky. "Time series forecasting of cyber attack intensity". In *Proceedings of the 12th Annual Conference on cyber and information security research*. 2017.
24. Goyal, P., Hossain, K. S. M., Deb, A., et al. "Discovering signals from web sources to predict cyber attacks". *arXiv preprint arXiv*:1806.03342, 2018.
25. Krakovsky, Y., Luzgin, A., and Ivanyo, Y., "Cyberattack intensity forecasting on informatization objects of critical infrastructures". *Materials Science and Engineering*. **481**(1), 2019.

26. Rege, A., Obradovic, Z., Asadi, N., et al. "Predicting adversarial cyber-intrusion stages using autoregressive neural networks". *IEEE Intelligent Systems*, 33(2), 29-39, 2018.

27. Maheshwari, V., Bhatia, A., and Kumar, K., "Faster detection and prediction of DDoS attacks using MapReduce and time series analysis". In *2018 International Conference on Information Networking (ICOIN)*. IEEE, 2018.

28. Okutan, A., Yang, S.J. , and McConky, K., "Forecasting cyber attacks with imbalanced data sets and different time granularities". *arXiv preprint arXiv*:1803.09560, 2018.

29. Nguyen, H. V., Naeem, M. A., Wichitaksorn, N., et al. "A smart system for short-term price prediction using time series models". *Computers & Electrical Engineering*, 76, 339-352, 2019.

30. Holgado, P., Villagrá, V.A., and Vazquez, L., "Real-time multistep attack prediction based on hidden markov models". *IEEE Transactions on Dependable and Secure Computing*, 2017.

31. Model, A., "Mixed Auto-Regressive Moving Average Model–ARMA (p, q)". 2019.

32. Singh, N.K., Gupta, P.K., and Mahajan, V., "Intrusion Detection in Wireless Network of Smart Grid Using Intelligent Trust-Weight Method". *Smart Science*. **8**(3): p. 152-162, 2020.

33. Huang, D. and Guo, L., "Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method". *The Annals of Statistics*, p. 1729-1756, 2019.

34. Husák, M., Komárková, J., Bou-Harb, E., et al. "Survey of attack projection, prediction, and forecasting in cyber security". *IEEE Communications Surveys & Tutorials*, 21(1), 640-660, 2018.

35. WANG, H., LU, S., and WANG, Y., "Intrusion Prediction Algorithm Based on Correlation Attack Graph". *Computer Engineering*, 7: p. 23, 2018.

36. Polatidis, N., Pimenidis, E., Pavlidis, M., et al. "From product recommendation to cyber-attack prediction: Generating attack graphs and predicting future attacks". *Evolving Systems,* 11(3), 479-490, 2020.

37. Osarumwense, A.S. and Osayamen, O.K., "A Distributed Denial of Service Attack with IP Information Prediction Model Based on Bayesian Belief Network".

38. Abaid, Z., Sarkar, D., Kaafar, M. A., et al. "The early bird gets the botnet: A markov chain based early warning system for botnet attacks". In *2016 IEEE 41st Conference on Local Computer Networks (LCN)* (pp. 61-68). IEEE. 2016.

39. Do, C. T., Tran, N. H., Hong, C., et al. "Game theory for cyber security and privacy". *ACM Computing Surveys* (CSUR), 50(2), 1-37, 2017.

40. Sedjelmaci, H., Senouci, S.M., and Bouali, T., "Predict and prevent from misbehaving intruders in heterogeneous vehicular networks". *Vehicular Communications*. **10**: p. 74-83, 2017.

41. Lavrova, D., Zegzhda, D., and Yarmak, A., "Using GRU neural network for cyber-attack detection in automated process control systems". In *2019 IEEE International Black Sea Conference on Communications and Networking* (BlackSeaCom). IEEE, 2019.

42. Ivanyo, Y., Krakovsky, Y., and Luzgin, A., "Interval forecasting of cyber-attacks on industrial control systems". *MS&E*. **327**(2): p. 022044, 2018.

43. Ahmed, A.A. and Mohammed, M.F., "SAIRF:A similarity approach for attack intention recognition using fuzzy min-max neural network". *Journal of Computational Science*. **25**: p. 467-473, 2018.

44. GhasemiGol, M., Ghaemi-Bafghi, A., and Takabi, H., "A comprehensive approach for network attack forecasting". *computers & security*, **58**: p. 83-105, 2016.

45. Yusof, A.R.a., Udzir, N.I., and Selamat, A., "Systematic literature review and taxonomy for DDoS attack detection and prediction". *International Journal of Digital Enterprise Technology*. **1**(3): p. 292-315, 2019.

Authors Biography:

Neeraj Kumar Singh received the master's degree in Electrical Power Systems from the Government Engineering College Aurangabad, Maharashtra, India. He is currently pursuing the Ph.D. degree in Electrical Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, Gujarat. He was working as Engineer Operations at Enercon India Ltd (Wind World India Ltd). He has also contributed to academic field as Assistant Professor at P.E.S College of Engineering, Aurangabad. His research interests include smart grid, cyber security and wireless sensor networks. Email Id: neerajksssingh90@gmail.com.
Orcid ID: 0000-0002-6005-1016

Mahshooq Majeed is pursuing M. Tech in Power Systems from Sardar Vallabhbhai National Institute of Technology Surat, completed B. Tech from National Institute of Technology Goa. Currently doing a major project on Machine Learning algorithms for Smart Grid. Email id: ckmahshooq@gmail.com
Orcid ID: 0000-0002-7147-6060

Vasundhara Mahajan, presently she is working as associate professor at department of electrical engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, Gujarat. She has obtained her doctoral (Ph. D.) and master degree (M. Tech.) form IIT Roorkee in 2014 and 2005 respectively. Graduated in Electrical Engineering from NIT Raipur (formerly GEC) in 1999. She worked as lecturer at Christian college of engineering and technology from Sept. 2000 to Oct. 2007. Then joined Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, Gujarat. She has published many research papers in International/national journals/conferences. She has organized 11 Short term training programs at SVNIT. She has delivered many expert talks in various institutes/colleges. She has guided many M.Tech. and B. Tech. projects. Presently she is guiding seven doctoral thesis. Her research area is cyber security of smart grid, power system reliability, restructuring/deregulation, congestion management, energy market, power quality improvement, active power filters, FACTS and artificial intelligence. Email Id: vasu.daygood@gmail.com.
Orcid ID: 0000-0002-2698-6096

Declarations

Funding: No funding support for this research work.

Conflicts of interest/Competing inteests: No

Availability of data and material: Authors have cited the papers.

TABLES DETAILS:

Table 1. Different research model for intrusion detection in CPSI

Model/Technique	**Year of research**[*]	**Evaluation**[^]	**Description**
Graph based attack graph [34]	2003-Till	Testbed	Attain accuracy up to 85% and suitable for a large network
Bayesian based attack graph/network [35-37]	2004-2018	Cyber-physical testbed	For large network accuracy is low
Markov chain model [30, 38]	2011-2016	Testbed	Prediction of next step in multi-stage attack
Game theory approach [39, 40]	2012-2016	Testbed for virtual attack	Very low accuracy up to 39%
Neural networks [41, 42]	2012-2017	Testbed	Intrusion prediction with accuracy up to 85%
Similarity based approach [43, 44]	2012-2017	Testbed	Reduce time of prediction with low accuracy
Evolutionary computing [45]	2014-2017	Testbed	Not so effective but alternate to similarity based approach
*Year of maximum research on that topic		^Which can be used for power system CI	

Table 2. Predefined Logic

Status	***T_s***	***S_d***	***A_d***	**Description**
Non-malicious	1	1	1	No change: all good
Malicious	1	0	1	According to figure 2, if the first two module are changed to value 0 means the data set is malicious.
	1	1	0	
	1	0	0	
	0	0	1	
	0	1	0	
	0	1	1	Data is ok but sensor is tempered

Table 3. Attack severity analysis

Attack severity	**T_s (0)**	**T_s (1)**	**S_d(0)**	**S_d(1)**	**A_d(0)**	**A_d(1)**
T_s(0)	M	NA	H	L	H	L
S_d(0)	H	L	M	NA	H	L
A_d(0)	H	L	H	L	M	NA

L= Low impact	M= Medium impact	H= High impact	NA= Not applicable

Table 4. Forecasting accuracy: For attack at single bus (Case I: Bus no.7)

Trial No.	**No. of actual attack**	**No. of attack forecast correctly**	**Accuracy (%)**
1	10	9	90
2	15	13	87
3	20	18	90
4	25	22	79
5	30	28	93
		Average accuracy	**88**

Table 5. Forecasting accuracy: For simultaneous attack on multiple bus (Case II: B5 and B7)

Trial No.	**No. of actual attack**	**No. of attack forecast correctly**	**Accuracy (%)**
1	10	9	90
2	15	14	93
3	20	18	90
4	25	24	96
5	30	27	90
		Average accuracy	**92**

Table 6. Comparison of standard Error (SE) for Case I and Case II

Trial No.	**SE for single bus attack**	**SE for multiple bus attack**
1	0.16	0.65
2	0.24	0.76
3	0.22	0.49
4	0.18	0.68
5	0.17	0.72

Table 7. Comparison of AARMA with existing models

Model/Technique	**P1**	**P2**	**P3**	**P4**	**P5**
Attack Graph [16]	✓	✓	L	✓	75%
Game theory [18]	✓	×	H	×	38.6%
Neural Networks [14]	✓	✓	M	×	85%
Time-Series (EWMA) [20]	✓	×	H	×	57.8%
Time-series (GARCHA) [22]	✓	✓	H	✓	70%-87%
Proposed AARMA	✓	✓	M	✓	90%
✓ = Yes	× = No	L= Low	M= Medium		H= High

FIGURES DEATILS:

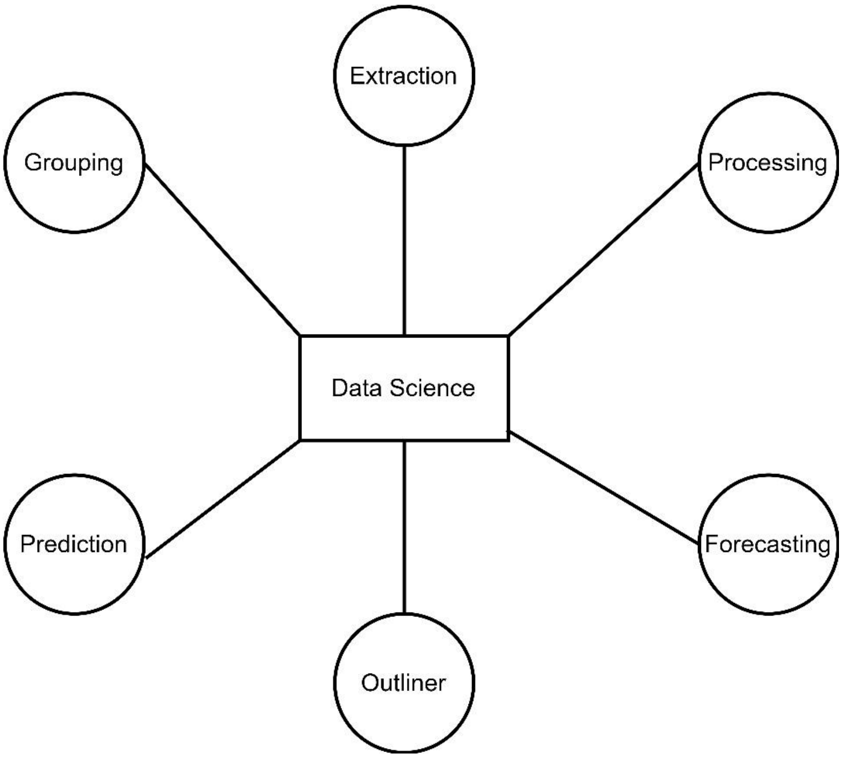

Figure 1.Features of data science

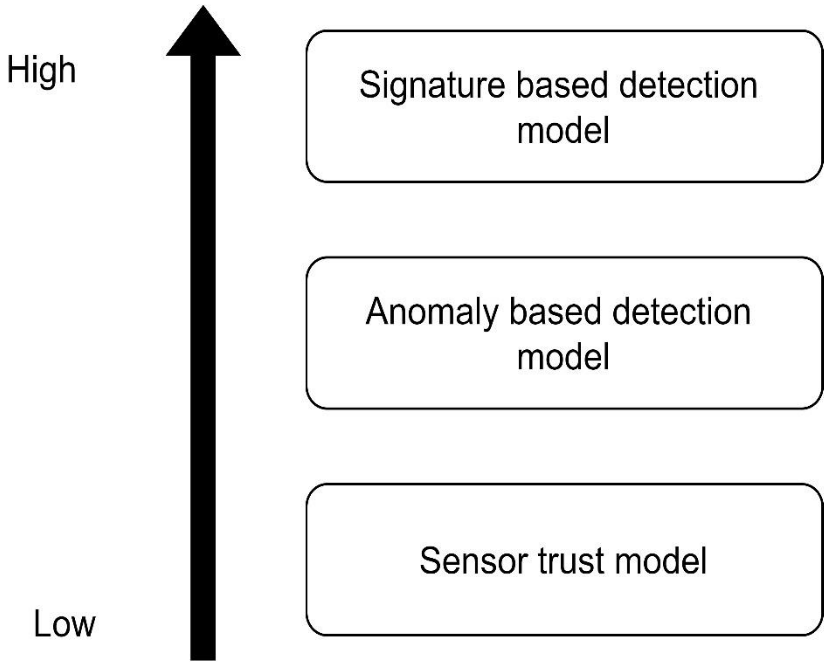
High
Signature based detection model
Anomaly based detection model
Sensor trust model
Low

Figure 2. Priority for Predefined Logic

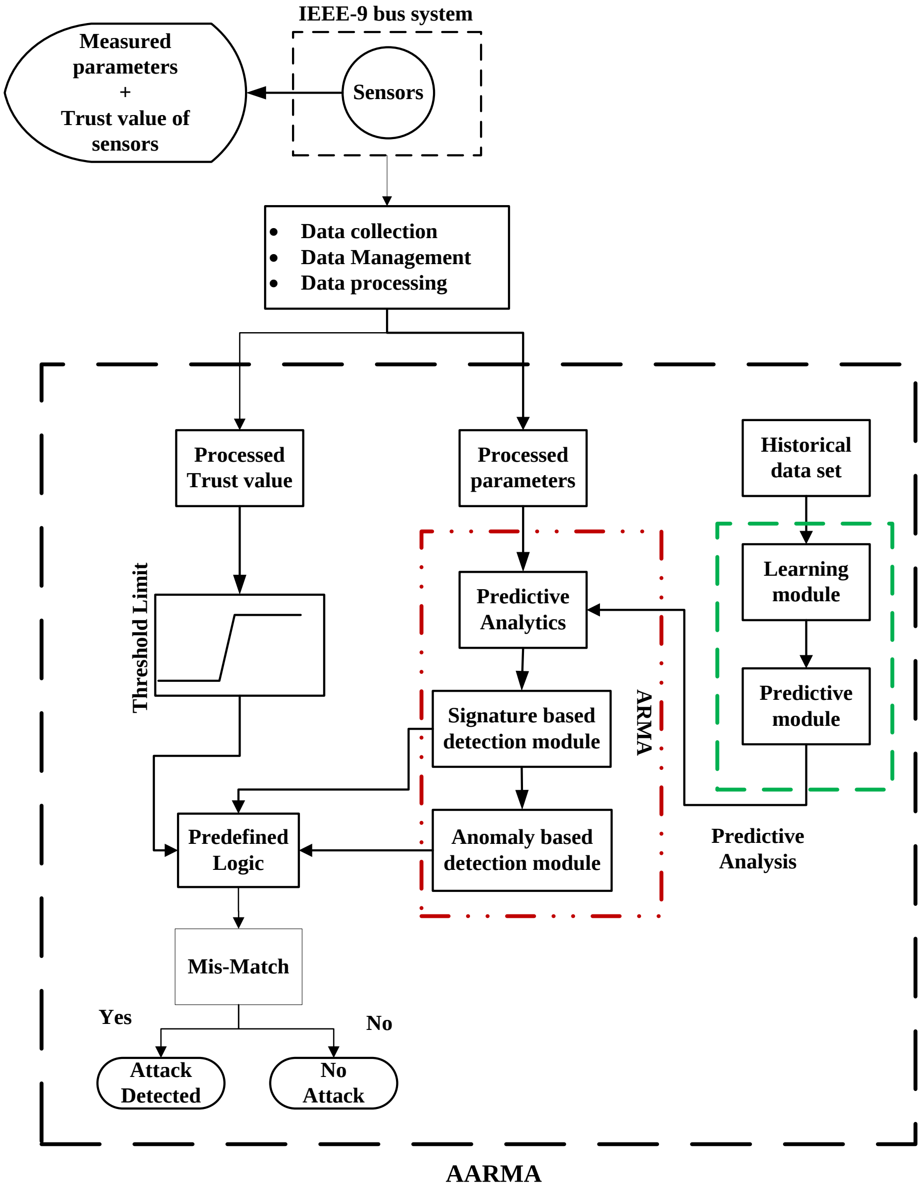

Figure 3. Proposed AARMA model

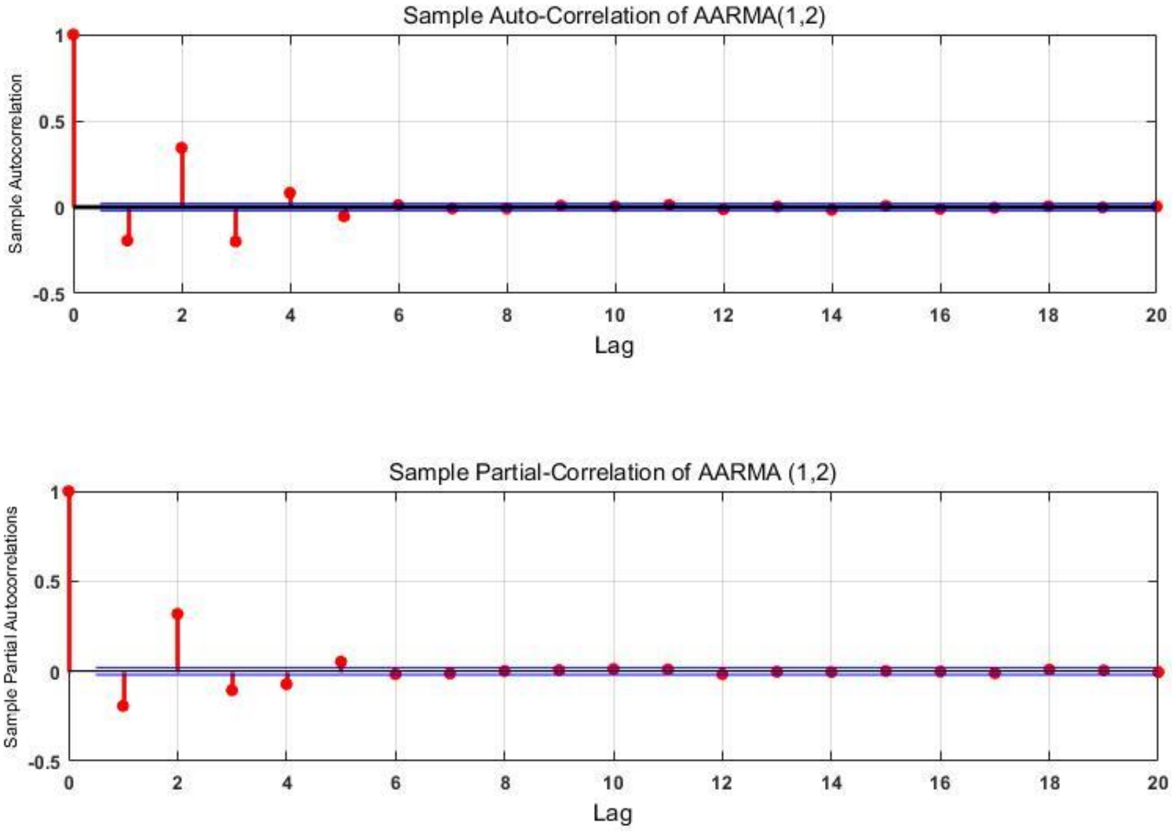

Figure 4. Auto-correlation for the time series used in AARMA

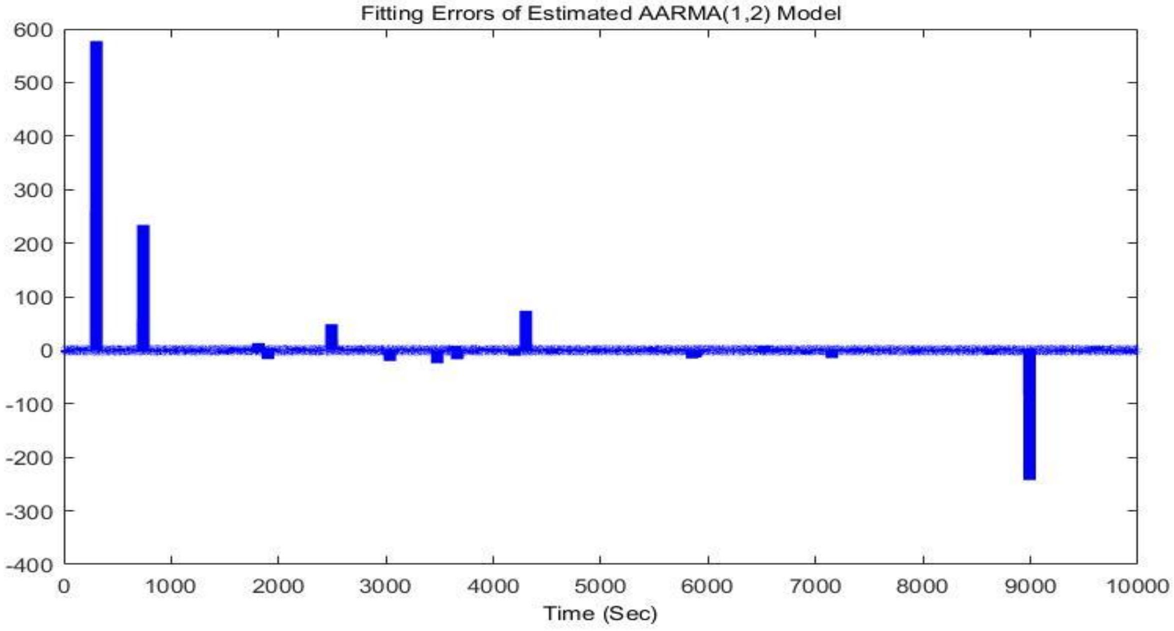

Figure 5. Fitting errors of estimation using AARMA (1,2)

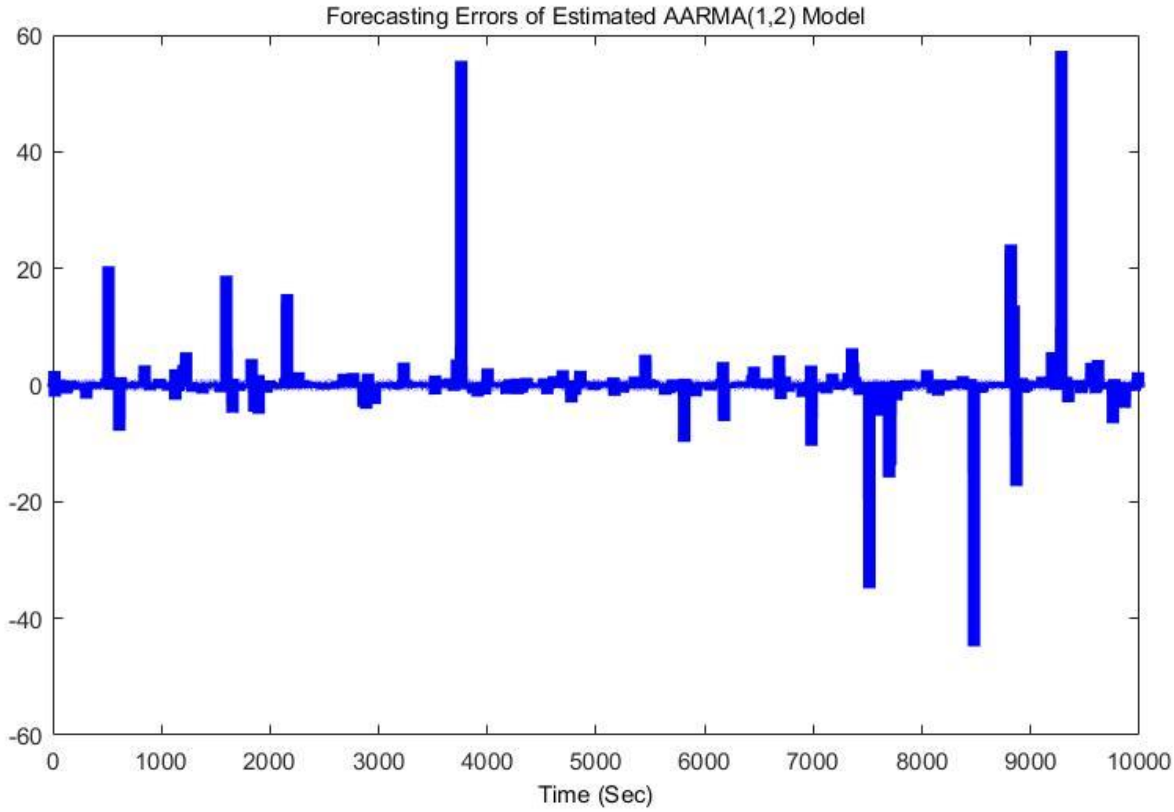

Figure 6. Forecasting errors of estimation using AARMA (1, 2)

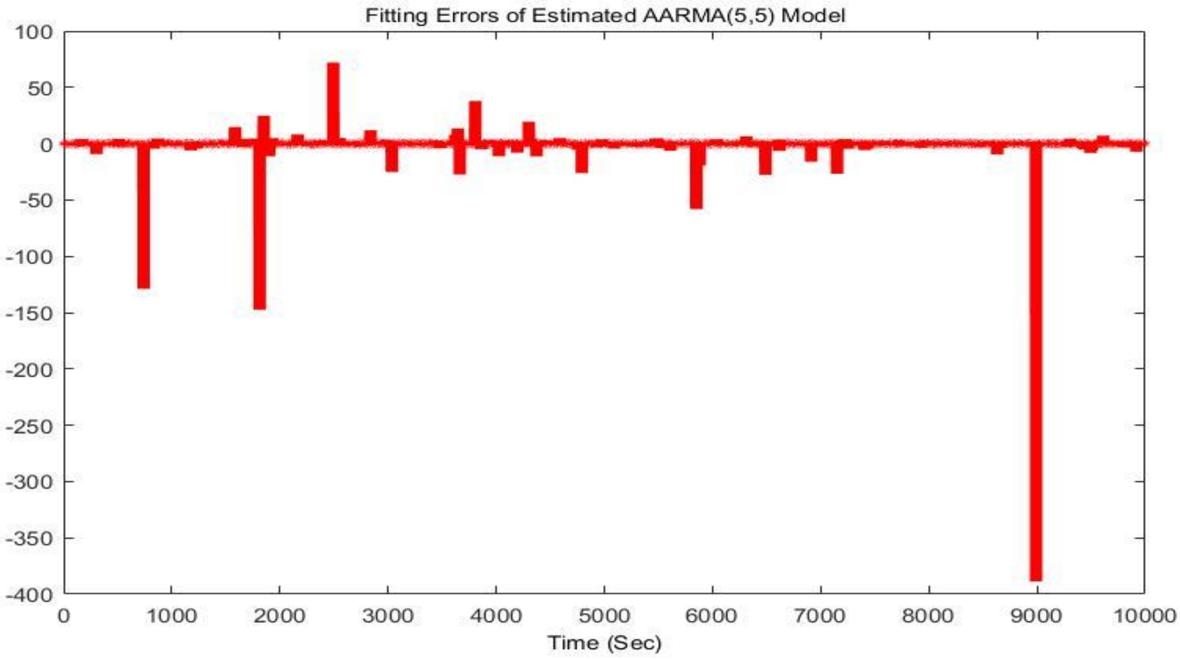

Figure 7. Fitting errors of estimation using AARMA (5, 5)

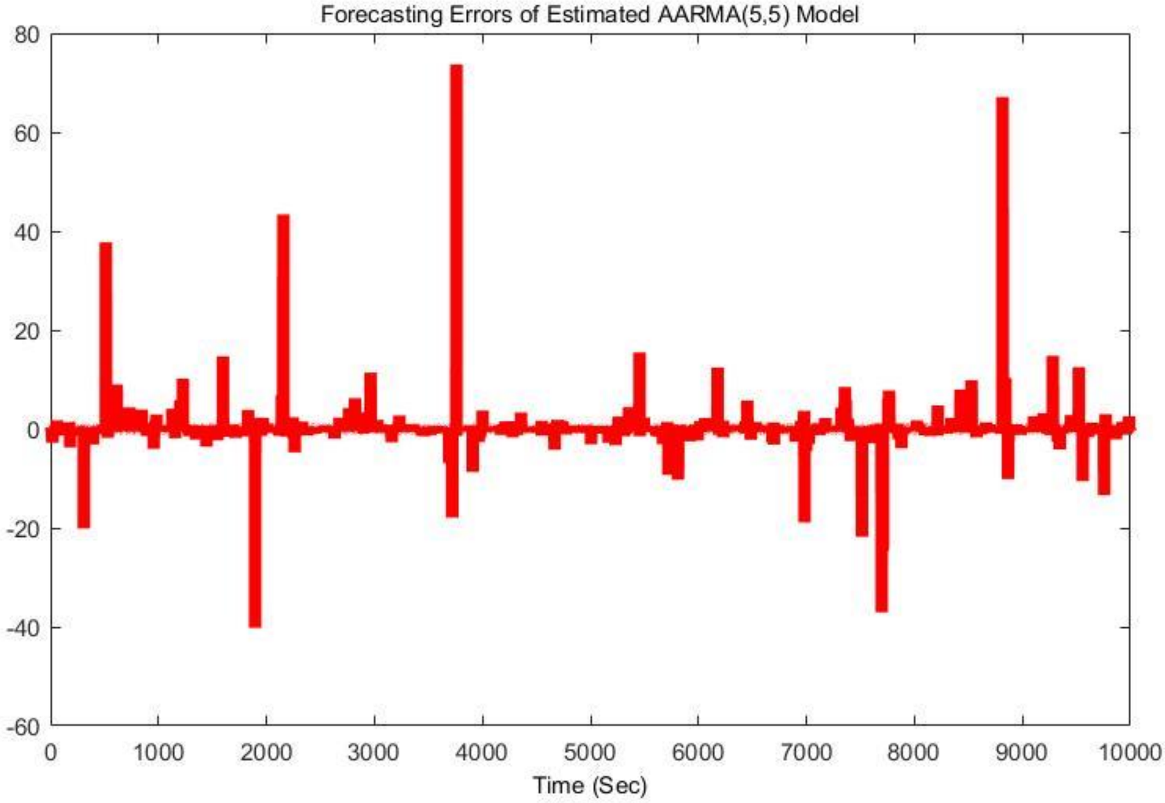

Figure 8. Forecasting errors of estimation using AARMA (5, 5)

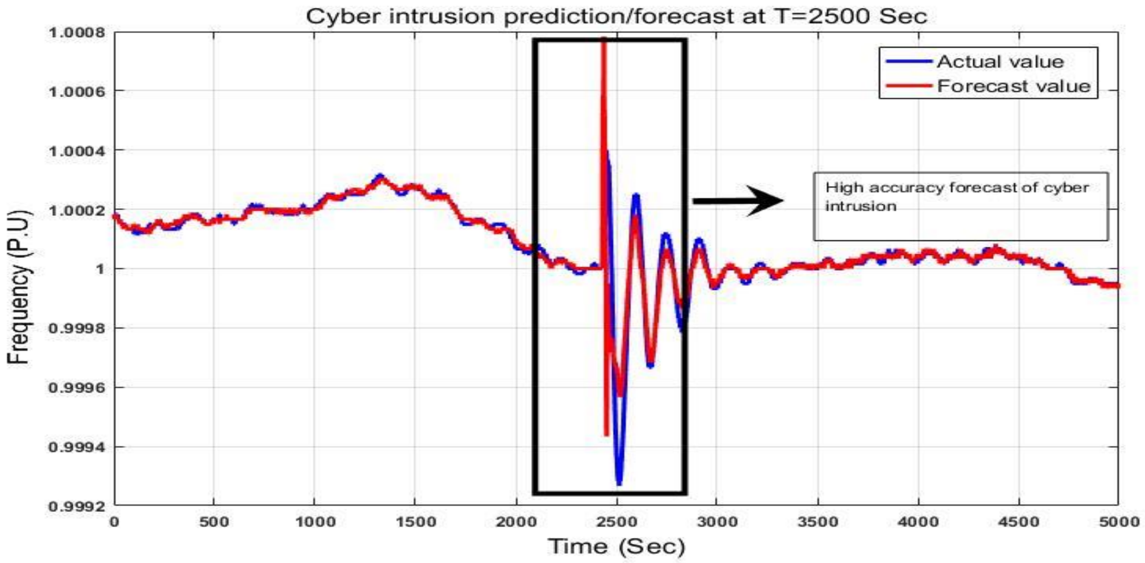

Figure 9. Cyber intrusion forecasting at Bus no. 7 in terms of frequency (Case I)

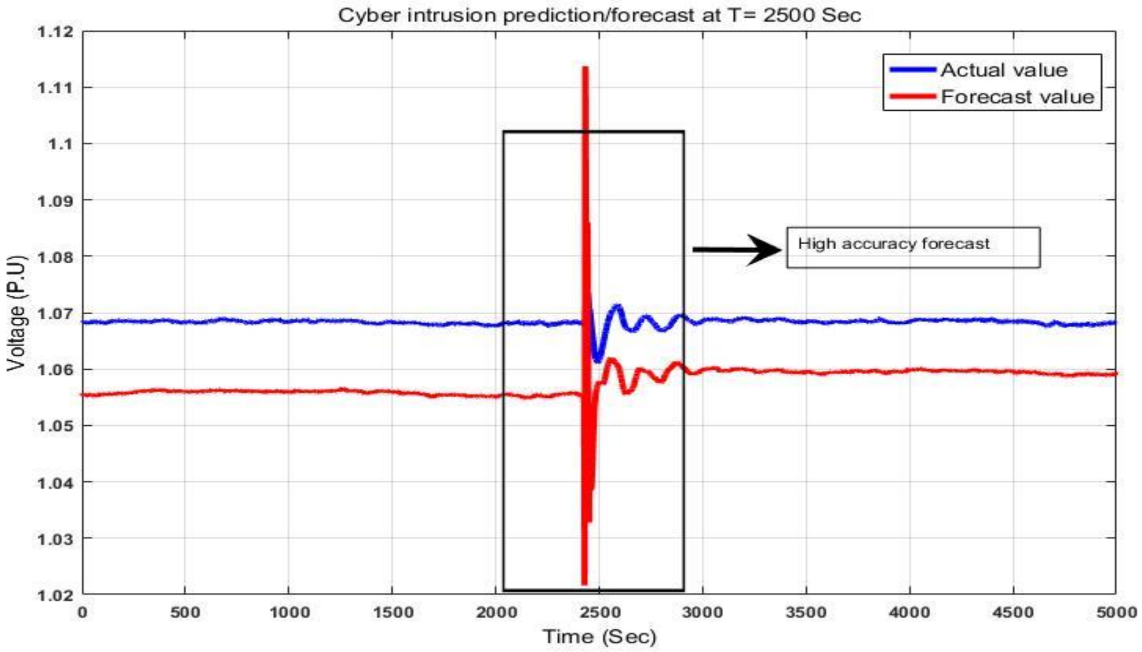

Figure 10. Cyber intrusion forecasting at Bus no. 7 in terms of voltage (Case I)

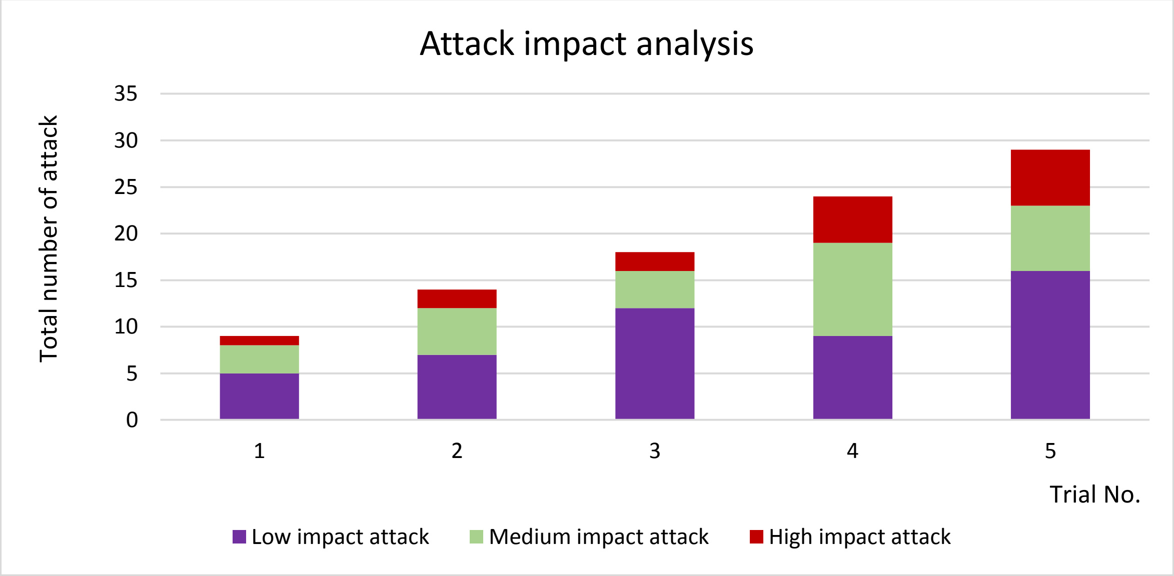

Figure 11. Attack severity analysis for single bus attack (Case I)

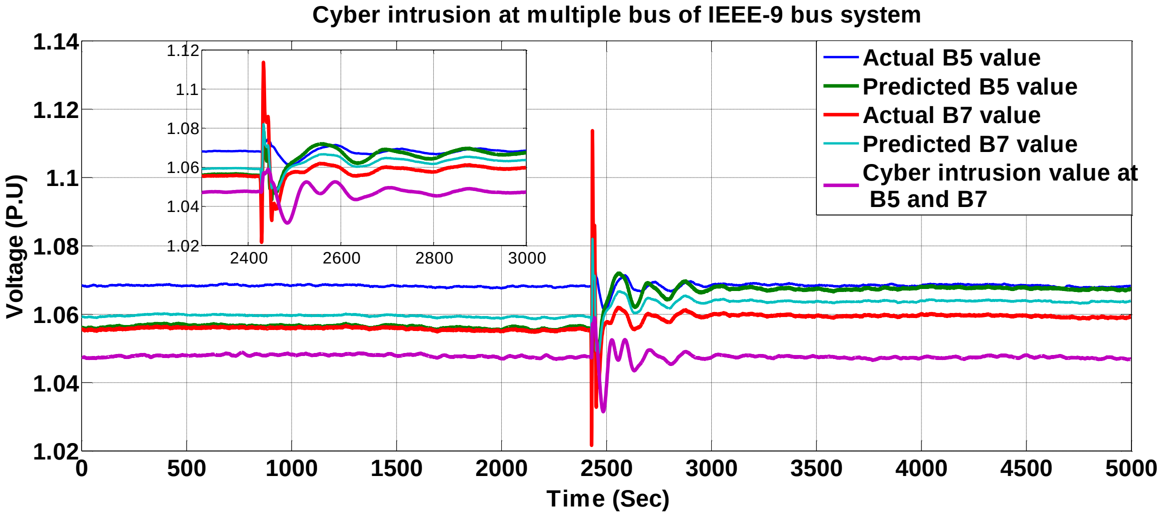

Figure 12. Cyber intrusion forecasting at multiple bus (B5 and B7) in terms of voltage (Case II)

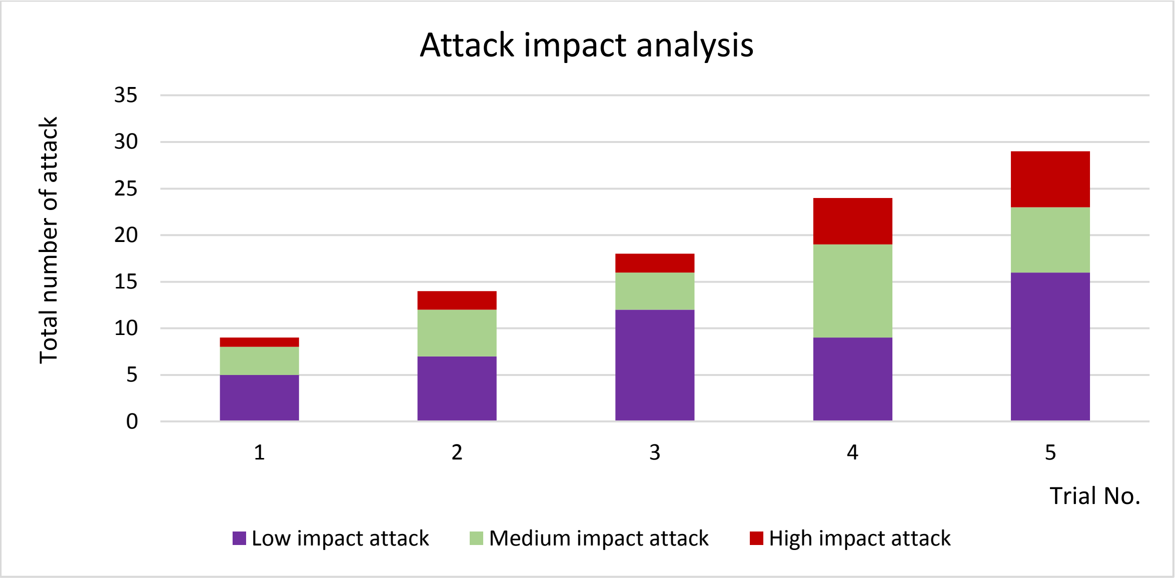

Figure 13. Attack severity analysis for multiple bus attack (Case II)

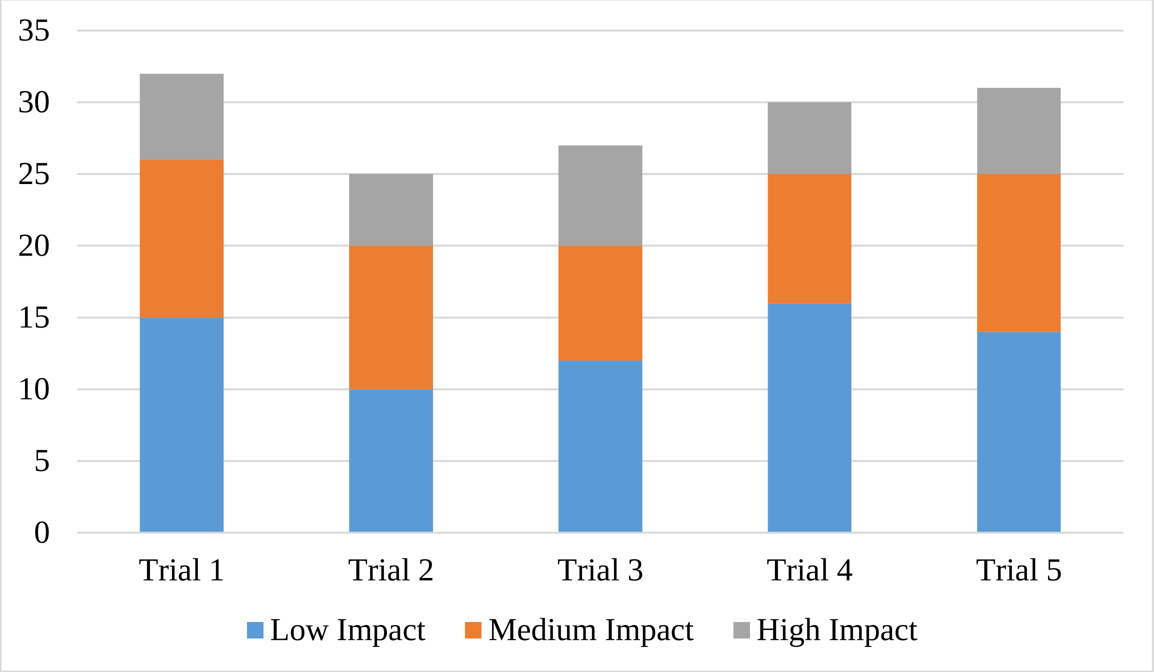

Figure 14. Attack severity analysis for multiple bus attack for IEEE 33-bus test system

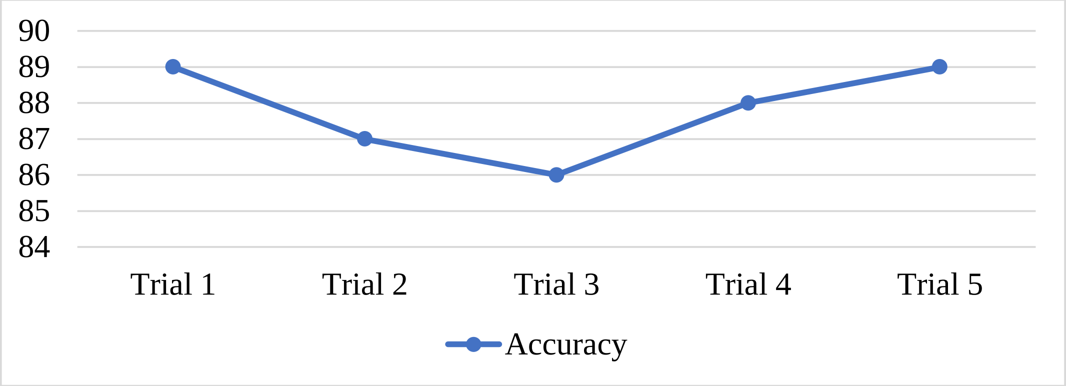

Figure 15. Accuracy for the IEEE 33-bus test system