International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

Predicting the Attacks in IoT Devices using DP Algorithm

P. Arul [1], N. Shanmugapriya [2]
[1]Research Supervisor, Assistant Professor, Department of Computer Science, Government Arts College (Affiliated to Bharathidasan University, Trichy-24), Tiruchirappalli-620022, Tamil Nadu, India.
[2]Assistant Professor, Department of Computer Science, Government Arts College (Affiliated to Bharathidasan University, Trichy-24), Tiruchirappalli-620022, Tamil Nadu, India.

Abstract
The fundamental goal of this study is to predict cyber-attacks before they occur and to protect the network. Most existing attack detection algorithms cannot identify zero day attacks because they lack previously known data patterns to predict the threat, which is one of the biggest issues in the existing approaches. This research work offers a novel prediction method based on Gaussian regression that identifies cyber-attacks utilizing a unique dual data pattern categorization technique with no false positives. To improve the accuracy of the prediction and to reduce the prediction time consumption, this study introduces a dual prediction technique one locally – at the fog level where non-parametric input data is dealt with two functions namely quadratic & reliability function to ease the prediction and the other universally – cloud level where result of skill mechanism is carried out. Even if the local prediction misses an attack, the universal prediction sniffs it and protects the IoT devices and the data. A detailed comparison regarding accuracy and packet drop is carried out by simulating flooding attacks using on varying numbers of dummy nodes and the proposed system found to outscore the existing methods convincingly.
Keywords: FOG computing, dual prediction, Gaussian, security, cloud.

1. INTRODUCTION

Most contemporary systems do not have the ability to forecast unknown attacks and assaults, but when trained with a past attack dataset, they can forecast the possibility of an attack. Furthermore, the accuracy level in predicting unknown assaults is quite poor, and present systems generate a lot of false positives while ignoring serious threats and attacks owing to a lack of sufficient training and data. As a result, a proactive security procedure is required to anticipate such attacks and develop counter-measures to avoid them. In addition to being attacked, IoT devices are used to create botnets. As a result, it is critical to predict an attack before the device is compromised. There are several time-series prediction mechanisms, such as the Kalman filter and ARIMA that may be used to anticipate the attack. However, all of these mechanisms are impacted by the structure of the input data; therefore we require a non-parametric model for attack prediction.

2. LIMITATIONS OF THE EXISTING METHODS

- Because the current algorithms usually depend on the input data's degree of certainty, it is exceedingly difficult for them to forecast the unpredictable data from IoT devices
- The massive amount of data generated by IoT devices is a serious worry as the scalability of data becomes challenging because the present methodologies use the complete dataset to forecast the attack.

The vast majority of IoT devices process data from end devices on the cloud where there are trade-offs between offloading computation and response latency during attack prediction

3. PROPOSED METHOD

Low-latency computations and storage are available at the edge thanks to fog computing, and fog nodes link IoT application endpoints to cloud computing resources seamlessly. By offloading data and computations from IoT devices for distributed Gaussian regression, the proposed fog computing process at the local level is utilized to foresee cyber-attacks in IoT applications using non-parametric uncertain data from the devices.

It is a known fact that every IoT application has many clusters of IoT devices, each of which is linked to a fog node which in turn is linked to the IoT application's cloud server. The cloud application servers receive traffic from IoT devices from these fog nodes. The network traffic produced by IoT endpoints is viewed as a random variable. On the basis of these random variables, the Gaussian process is described under the presumption that any finite subset of these variables forms a joint multivariate Gaussian distribution. The fog node gauges the underlying traffic characteristics and predicts the cyber-attack.

The most important advantage of the proposed dual prediction method is it utilizes supervised and unsupervised techniques, the local prediction at the fog encompasses supervised(i.e) requires training of the input data fetched from various

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

endpoints but at the global prediction at the cloud level uses unsupervised (i.e.) requires no training to predict the cyber-attack. Hence the proposed method uses dual forecasting to protect the data from cyber-attack.

4. RELATED WORKS

There has been significant advancement in network security research related to IoT devices. Despite this, there are still many hurdles like the network traffic, a lack of a uniform understanding of normality caused by network unpredictability, the absence of suitable datasets, and vulnerable environments and security flaws that allow access to attackers who actively look for and exploit security flaws. Threats are always changing for the IoT layers and new threats and attacks pops up every now and then.

To reduce assaults on IoT devices in the context of smart cities, writers in presented a variety of IDS models based on machine learning [1]. The ensemble model was created using a variety of ML algorithms and ensemble approaches, including the stacking, bagging, and boosting methods. The accuracy and recall of the suggested ensemble models were 0.999 after assessment. The IDS paradigm for IoT networks was proposed by [2] using a number of Ml algorithms. K-nearest neighbour (KNN), support vector machine (SVM), artificial neural network (ANN), and other ML methods were employed by the authors in their study.

By recording the TCP/IP packets in the networks, a number of open-source network monitoring technologies are now used to provide network security. To identify malicious attacks, the authors Suricata and Snort rely on pre-established rules [3]. One of these systems' biggest flaws is that any departure from the established restrictions will trigger a false alarm. Once more, it necessitates that a security professional to investigate both current assaults and innovative network deviations under predetermined circumstances that establish the database's signatures. Attackers manipulate the events protocol by taking advantage of the vulnerabilities [4] that are frequently found in IoT networks.

5. METHODOLODY

The main goal of this research work is to propose a novel technique to foresee the zero day attacks present in the present day IoT space and detect the attacks efficiently without error and false positives. The Gaussian process uses random variables to create random function to predict the attacks without any training and prior data patterns [5]. The Gaussian function is shown in the following section,

$$F(z) = Gaussian\ (m(z), Co\text{-}Var\ (z,z^1))$$

Where $m(z) = E\ [\ F(z)\ ]$

The m(z) is assumed to be zero initially and $Co\text{-}var\ (z,z^1)$ is the kernel that is being used in the Gaussian process. This is computed using the following formula

$$Co\text{-}var(z,z^1) = E\ [\ (f(z) - m(z))\ (F(z^1) - m(z^1)\ ]$$

The Gaussian process is used for the regression and let us assume that the given input data $D_i=\{a_i,b_i)$ and the regression will foresee the output b for the input a where b= F(z) + Gaussian noise

The noise is added and the co-variance function becomes

$$Co\text{-}Var(z,z^1) = kernel(z,z1) + \sigma^2{}_n delta(z,z^1)$$

Using this covariance function the relationship between the training data are computed and the covariance matrix is calculated

$$\text{Matrix K} = \begin{vmatrix} Co\text{-}Var(z1,z1^1) & - - - - & Co\text{-}Var(z1,zn^1) \\ - & & \\ - & & \\ Co\text{-}Var(zn,z1^1) & - - - - & Co\text{-}Var(zn,zn^1) \end{vmatrix}$$

The proposed method uses three major attributes named variation and reliability. The heterogeneous nature of the IoT devices has lot of traffic variance since the devices are either synchronized or unsynchronized [6] [7]. The formula to compute the variation using the quadratic kernel Q is

$$Variation = K_Q\ (r_l, l_l, \alpha) = \sigma\ (1 + r_l\ /\ 2\alpha)^{-\alpha}$$

where α is considered as the shape , l is considered as the length and σ is the actual variance of the IoT.

The reliability of the devices is computed using the exponential kernel and the formula is

Reliance $K_s = \sigma\ (-\ r\ /\ l) + \sigma 1\ (-\ r1\ /\ l1)$

Where *l* and *l1* and length

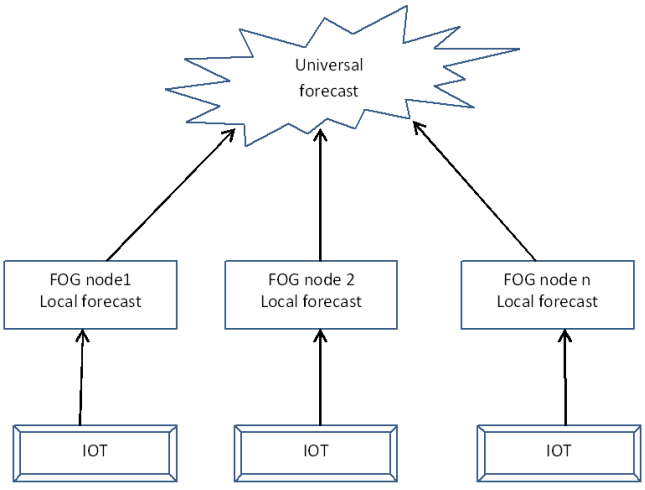

FIGURE 1: PROPOSED PREDICT ATTACK MODEL

The IOT devices are connected to the fog nodes where the local prediction takes place and then finally the cloud performs the universal prediction to ensure that the attack will be mitigated at the initial stage itself. The proposed algorithm Dual prediction algorithm is shown in the following figure 2.

LOCAL PREDICTION AT FOG LEVEL

Let the non-parametric training data be $Ds = \{A_i, B_i\}$ *where* $i = 1, 2, 3, \ldots n$, *A* is the IoT parameter value, *B* is the output value after n observations.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

Here $B_i = f(Ai) + \Xi_i$

The secret function f is the learning or training the IoT input parameter and Ξ is the independent noise.

The full Gaussian process is carried out using the following formula,

f(Ai) = G(mean(Ai), cv(Ai,A'i)), Where cv is the co-variance

The covariance of the output Bi is,

cv(Bi) = Kernal(Ai,A'i)

Since the input data is non-parametric, two types of functions are used at the local level to attain higher accuracy

1. *Quadratic function = Qf→(R1, L1, A)*

Where R is quadratic co-variance, L is the length of data, and A is shape parameter.

2. *Reliability function = Rf(R2, L2) + Rf(R3,L3)*

Where L2 and L3 are length of the parameters

Mf = *Quadratic function + Reliability function*

= Qf→(R1, L1, A) + Rf(R2, L2) + Rf(R3,L3)

Algorithm Dual Prediction

For each Fog Nodes FN do begin

1. Identify the covariance function
2. Apply the covariance function locally
3. Train the input data
4. Perform GP← Gaussian Process
5. Fetch the predicted values ←Pval
6. Check the Pval with threshold value Tv
7. IF [Pval varies with the Tv]
8. Declare the Fog node FN as malicious
9. The FN sends the locally predicted Pval to Cloud

Collect Pval from FN

Compute the universal prediction

Confirm the node attack and alert the device

End For

End Algorithm

FIGURE 2: PSEUDO CODE OF DUAL PREDICTION ALGORITHM

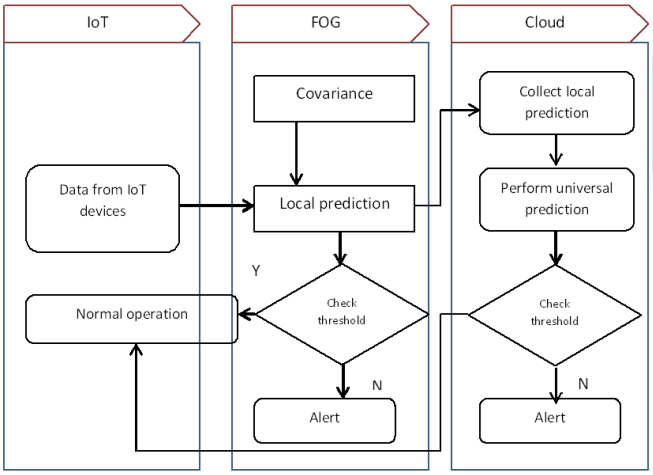

FIGURE 3: OVERALL WORK FLOW OF THE PROPOSED ALGORITHM

6. EXPERIMENTAL EVALUATION AND DISCUSSION

The experimental evaluation is conducted using WS2 simulator tool and Azure cloud with Matlab code with a set of parameters as shown in the table 1 and the attack parameters are shown in the table 2.

TABLE 1: EVALUATION PARAMETER USED

PARAMETER	VALUE
Number of Nodes	150
Packet Size	512 Bytes
Transmission pattern	Constant baud
Bandwidth	12 MBPS

TABLE 2: ATTACK PARAMETER USED IN SIMULATION

PARAMETER	VALUE
Rate of Attack	40 packets per sec
Number of Malicious nodes	3 nodes per Fog cluster
Number of Dummy	10 per Fog cluster
Attack mode	Route Request packet

The simulation is carried out by generating the denial of service attack [5] in the wireless sensor. When a node sends a packet to the destination and if the node does not have any route, the node sends the packet to its nearest neighbor node where the route id collected from its cache and then the packet is redirected via that route. The node sends a request message RREQ to the neighbor and if the neighbor node has a route then it sends the reply RREP to the source node. To generate the attack simulation, dummy nodes are used in the cluster and in each dummy 50 requests are flooded with RREQ message. The entire simulation is carried out with varying number of dummy nodes and the packet delivery rate PDR is noted as shown in the following graph figure 4.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

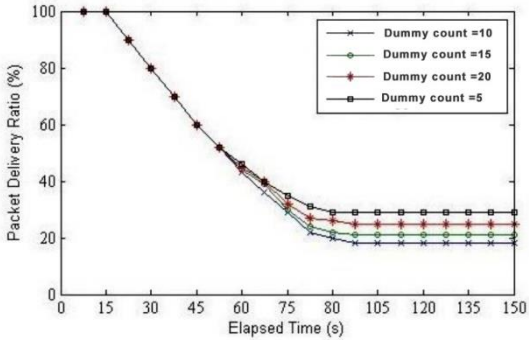

FIGURE 4: PDR COMPARISON WITH RESPECT TO VARYING DUMMY NODES

From the graph it is quite obvious that the packet drops threshold is computed and found to be 40% as the attack is predicted at this level. The training time and the prediction time are calculated for the training data set and shown in the table 3.

TABLE 3: TRAINING TIME AND PREDICTION TIME

Fog node level local forecast		**Cloud level universal forecast**	
Training time (sec)	**Predicted time (sec)**	**Training time (sec)**	**Predicted time (sec)**
41	59	-	59.7

To gauge the performance of the proposed algorithm, a similar assault with 10 dummy nodes considered for prediction with Full Gaussian and Gaussian based on subnet of data from the cloud and the MSE mean squared error is computed for these three approaches and noted as shown in the table4.

TABLE 4: PERFORMANCE COMPARISON

Approach	**Mean squared error**	**Train time(sec)**
Dual prediction algorithm	0.12	29
Gaussian regression	0.13	43
Gaussian with subset	0.35	30

From the table it is clear that the proposed dual prediction algorithm performed the best when compared with the other two approaches, the Gaussian regression performed quite well but the training time taken was bit longer, in the Gaussian with subset, the training time was lower but the error rate was on the higher side since it fetched the data from the cloud randomly. The prediction of attacks and the training time are used to find the efficiency of the overall approach and the proposed DP algorithm fared quite well on both without any tradeoff but if you note the figure 5, the Gaussian compromised with the time and the Gaussian with subset compromised with the error rate.

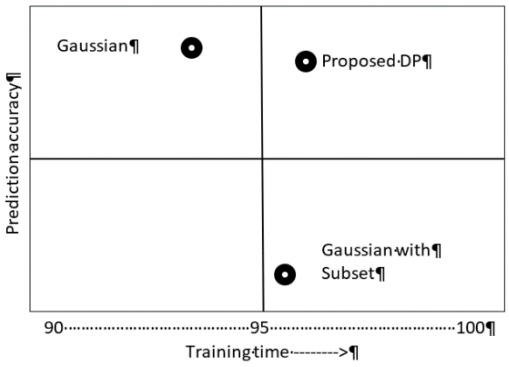

FIGURE 5: PREDICTION ACCURACY COMPARISON

The proposed algorithm along with the other two methods are evaluated after generating flooding attacks on dummy nodes and the prediction accuracy is computed for these. The number of dummy nodes used in the experiment varies between 5 and 20 and from the experimental result it is found that the proposed DP algorithm predicted with less false positives as shown in the figure 6.

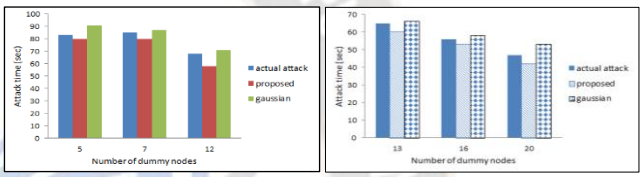

FIGURE 6: ATTACK PREDICTION TIME

7. CONCLUSION

Two co-variance functions are used in the proposed method to greatly increase the prediction accuracy utilizing the data obtained from the different devices, both synchronized and unsynchronized, data with low statistical volatility, short & long packets acquired from the endpoints. The key advantage of the proposed DP method is that fog nodes require less training time to locally anticipate low-rate attacks. To demonstrate this, low-rate flooding assaults with training times of 50s were constructed utilizing only 5 and 12 dummy nodes, and the attack prediction was carried out using the proposed method. Because the Gaussian strategy needed much longer training time and had lower accuracy for subsets, the new DP approach must be examined with cloud random data to see how it works in future.

REFERENCES

[1] Rashid, M.M.; Kamruzzaman, J.; Hassan, M.M.; Imam, T.; Gordon, S. Cyberattacks detection in iot-based smart city applications using machine learning techniques. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9347. https://doi,org/10.3390/ijerph17249347.

[2] Churcher, A.; Ullah, R.; Ahmad, J.; Ur Rehman, S.; Masood, F.; Gogate, M.; Alqahtani, F.; Nour, B.; Buchanan, W.J. An experimental analysis of attack classification using machine learning in IoT networks. *Sensors* **2021**, *21*, 446. https://doi.org/10.3390/s21020446.

[3] Murphy, B.R. Comparing the Performance of Intrusion Detection Systems: Snort and Suricata. Ph.D. Thesis, Colorado Technical University, Colorado Springs, CO, USA, 2019.

[4] Jingyi Su, Shan He, Yan Wu, Features selection and prediction for IoT attacks, High-Confidence Computing, Volume 2, Issue 2, 2022, https://doi.org/10.1016/j.hcc.2021.100047

[5] Avcı, İ.; Koca, M. Predicting DDoS Attacks Using Machine Learning Algorithms in Building Management Systems. *Electronics* **2023**, *12*, 4142.

[6] Mazhar, T.; Talpur, D.B.; Shloul, T.A.; Ghadi, Y.Y.; Haq, I.; Ullah, I.; Ouahada, K.; Hamam, H. Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. *Brain Sci.* **2023**, *13*, 683. https://doi.org/10.3390/brainsci13040683

[7] Subrato Bharati, Prajoy Podder. (2022). Machine and Deep Learning for IoT Security and Privacy: Applications, Challenges, and Future Directions. Security and Communication Networks. 2022. 8951961. 1-41. https://doi.org/10.1155/2022/8951961