**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 37, No. 3, March 2025, pp. 1845~1854
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i3.pp1845-1854

# Anomaly based detection in time series data on IoT systems using statistical models

**Mouna Boujrad[1], Mohammed Amine Kasmi[2], Noura Ouerdi[1], Yasser Lamlili El Mazoui Nadori[3]**
[1]Arithmetic, Scientific Computing and their Applications Laboratory (LACSA), Faculty of Sciences (FSO),
Mohammed First University (UMP), Oujda, Morocco
[2]Computer Sciences Research Laboratory (LARI), Faculty of Sciences (FSO), Mohammed First University (UMP), Oujda, Morocco
[3]Faculty of Sciences (FSO), Mohammed First University (UMP), Oujda, Morocco

**Article Info**

***Article history:***

Received Mar 22, 2024
Revised Oct 2, 2024
Accepted Oct 7, 2024

***Keywords:***

Anomaly detection
Integrity
IoT
Security
Statistical models
Trustworthiness

## ABSTRACT

The internet of things (IoT) has become a real revolution that represents technological innovation in all domains, it becomes more and more integrated into human activities starting from personal needs, to professional eras including industry, logistics, healthcare. Yet this technology didn’t only bring advantages to all industries, it also creates new challenges mostly security-related, due to the specifications of the IoT environment: its heterogeneity, the restricted resources, and the continuous enormous sensitive data generated and exchanged on the IoT ecosystem. In this paper, we propose a study of statistical models (autoregression, moving-average, autoregression-movingaverage, autoregression integrated movingaverage, seasonal autoregression integrated movingaverage) to build an anomaly-based detection model for times series data, to detect abnormal behavior that can be explained by a sensor failure, or a compromised sensor. The proposed anomaly-based detection relies on defining the data behavior and creating a profile on the assumed normal state, the created profile will be used as knowledge to predict future values to which real records will be compared, any deviation from the predicted data will be considered as abnormal, that indicates an anomaly has occurred on the sensor or the exchanged data. The proposed approach will help improve accuracy, reliability, trustworthiness, and data integrity.

***Corresponding Author:***

Mouna Boujrad
Arithmetic, Scientific Computing and their Applications Laboratory (LACSA), Faculty of Sciences (FSO)
Mohammed First University (UMP)
Oujda, 60000, Morocco
Email: mouna.boujrad@gmail.com

## 1. INTRODUCTION

In this article we will propose a new anomaly-based detection system in time series data on IoT systems, the novelty lies in using statistical techniques not only to predict but also to detect anomalies and improve system integrity. Our choice is justified by the specification of the IoT system [1] itself which can be defined as locally connected objects configured to work in a very specific way depending on time, for example when observing the perception layer we know that we have sensors that are recording/reporting specific data [2] on predefined time interval depending on the data send (once every second, once every minute, hourly, daily, weekly), for that time series anomaly detection can be applied for the data sent as well as the network traffic generated by the time of sending the data and also by other time, for example, if we add parameters to the sensor to send records each hour it will be weird if we detected a network traffic out of the time defined to send records [3].

There are so many techniques [4] that can be used for anomaly detection but some fits perfectly to our study which is time series analysis that relies on statistical models [5] that can be applied to detect and predict the behavior of the times series dataset. In time series each point can affect the next one periodically speaking, so that becomes a trend or even a cyclic behavior the anomaly is raised when actual data is deviating from the predicted one, by a very low or high difference. The IoT system allows us to perform time series [6] analysis for anomaly detection, it is the best environment to do so due to its specification that fits perfectly to the requirements needed to apply time series analysis or anomaly detection [7].

## 2. OBJECTIVE OF THE RESEARCH AND CONTRIBUTION

The objective of this research is to sudy statistical models, and their performance in anomaly-based detection in IoT data, the addressed data is the one collected or transfered by the IoT sensors, we took into consideration that this data is a time series dataset, the deployed sensors collect and transfer data at different times, defined based on the needs of the system, However anomalies may occure on this data, during the time of collection or transfer, whcih origine by a vulnerability or technical cause such as sesonr dysfunction. Statistical models are designed to performe on time series datasets, but in the environement of IoT the models needs to be tested. In this work we propose a literature review on the existing current work related to the research, then an overview of anomaly-based detection and statistical models, afted that we define our proposed model for anomaly-based detection in time series dataset, based on the comparative study we conduct on the choosed statistical models, the tests were validated using an IoT dataset from Kaggle plateforme.

## 3. RELATED WORK

In 2016, Jibin and Jiang [8], improved ARIMA based anomaly detection algorithm for wireless sensor networks, the improvement was applied to the structure of the ARIMA model, first using a sliding window to determine historical data for modeling, second updating the model after each time sliding window, and last making traffic prediction by short step exponential, the model was tested on real WSN network traffic using simulation on MATLAB, the results show that despite the increase of false positive rate on the proposed model, it still gives better results by decreasing the rate of true negatives and increasing the rate of true positive.

In 2018, Giannoni *et al.* [9] developed an anomaly detection algorithm for IoT times series data, multiple algorithms were applied to the data to detect anomalies: running average low-high pass filter: make use of the running average computed based on the last W acquisitions (where W is a sliding window of fixed size), as a means to identify anomalies. Univariate Gaussian predictor: the model classifies each new acquisition xi based on the probability of that value in the distribution p(xi). A threshold has to be set to decide which probabilities are too low to be considered non-anomalous. Seasonal ESD algorithm: statistical-based method, applied only on time series that have their residual component symmetrically distributed.

In 2020, Braei and Wagner [10] presented in their research a survey on the state of the art on different techniques and methods used for anomaly detection in univariate time-series data, the researchers perform an experiment using five different univariate time series datasets using both statistical and classical machine learning (AR, MA, SES) and deep learning approaches (CNN, LSTM, GRU), the results show that statistical techniques perform better on univariate time series with more accurate detection on point and collective anomalies and less computation times.

In 2020 Eran *et al.* [11] reviewed in their paper anomaly detection on sensors, on different levels, comparing the techniques used to detect anomalies, the authors reveal that statistical techniques don’t work better on multivariate data, or high dimensional datasets, compared to machine learning techniques, however, the statistical technique is much easier to implement and computationally efficient.

In 2022 Oser *et al.* [12] proposed an enhanced version of SAFER’s framework using Facebook’s Prophet and ARIMA as prediction models and simple moving average (SMA) to compare the predicted results, the proposed approach applied to vulnerabilities patches of 793 IoT devices showed high prediction rate of 91 of future risks.

In 2022 Kalouptsoglou *et al.* [13] performed a time series forecasting using both statistical and deep learning techniques, applied to national vulnerabilities databased (NVD) for different software projects, the results were very close however ARIMA reached better goodness of fit level compared to other used deep learning methods.

In 2022 Awang *et al.* [14] proposed a SARIMA-based prediction model to help admins in making a decision for risk management, the results regardless of the high rate of the root mean square error (RMSE), showing that SARIMA is a good model to forecast attacks and threats in the future.

In 2023, Samaria and Thakkar [15] performed a comprehensive survey on anomaly detection algorithms, according to their research, statistical-based algorithms are considered very powerful ones in outlier detection specially when applied on certain type of data.

## 4. ANOMALY-BASED DETECTION

The anomaly-based detection is basically about defining the data behavior and creating a profile or a model on the assumed normal state [16], the profile created will be used as a knowledge to which real records will be compared any deviation from that profile will be accepted as abnormal behavior; Anomaly detection becomes the topic of the hour since it proved its performance in detecting new attacks, as long as the attack causes a deviation from the normal traffic flow or data collected. However anomalies don’t necessarily indicate an attack or threat, abnormal behavior can have several causes such as: system error, environment changes, new observations. For that reason, contradictory to rule-signature based [17] anomaly-based detection is known for its high rate of false positives and true negatives [18]. This can leads to misestimating the real danger, and the real state of the system. In spite of that, anomaly is still the best way to detect new attacks compared to rule-signature-based intrusion detection systems, that rely only on the existing signatures in their knowledge base.

To build a performant anomaly-based detection for monitoring we first need to define good data which is the main focus above all, then we can apply the different existing techniques to train the data and create the profile on which we will base anomaly detection, Figure 1 shows the different techniques used for intrusion or anomaly detection each one can be applied differently and it depends on the type of the data itself, some techniques could perform better on time series data others could give the worst results.

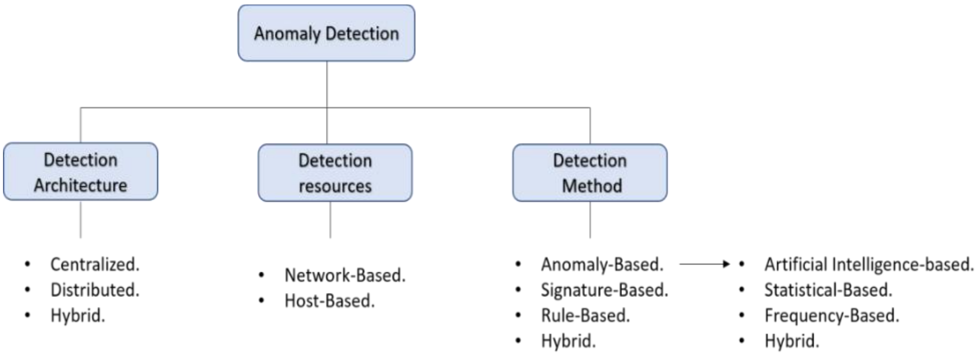


Figure 1. Classification of IDS types

### 4.1. Time series data

Time series data [19] is a very important term in monitoring different processes, it can be defined as a series of data points taken at different times, it allows tracking the changes or the movement of a very specific numerical data point over time, the time here is relevant to the data collected, it is a regular interval over a specific period such as hourly, daily, weekly or even in a given personalized period of time. The data collected in IoT systems [20] via sensors can be defined throw time series data since the sensors collect information about the environment on a given period so any records given by any sensor are related to a specific moment. A general given dataset including time series data is composed of three different patterns:

- Normal: it is the data how it is supposed to be, the correct information collected or stored.
- Anomalies: abnormalities, outliers, or deviants (a value that differs considerably from the other observations), there are many reasons for anomalies such as malicious actions, system failure, and intentional fraud.
- Novelties: it is a new observation that occurs in a non-malicious action such as changes in the environment, it is a novel pattern.

In addition to those patterns, time series data can include other patterns:

- Seasonal: seasonality happens when a time series is affected by short-term variation such as seasonal factors such as the time (month) of the year or a day of the week. Seasonality produces a fixed and known frequency.
- Trend: in opposite of seasonal trend is affected when there is a long-term, it is defined as a movement in the data such as changing in a direction from increasing to decreasing or the other way.

- Cyclic: it can be defined as a novel trend it occurs due to changes in some related factors to the given measurement for example the temperature of a given place will change if the humidity changes but it will not create a frequency.

### 4.2. Type of anomaly in time series datasets

a) Point anomalies: a single abnormal observation Point that represents an irregularity or deviation that happens randomly.
b) Contextual anomaly: also known as the conditional anomaly, it is observed in specific intervals, for example, observing measurement 2 on the temperature in the winter season can be considered normal, but in summer would be abnormal, of course, we can have multiple conditions such as local because it also affects the data, for contextual anomaly a specific context must be defined.
c) Group or collective anomaly: it is a group of anomalies that can be considered as normal patterns if observed individually from the data, this type of anomaly can cause a lot of problems if the data is small.

### 4.3. Statistical models

In the following, we will detail more statistical techniques that work better than other techniques in time series anomaly detection:

a) Threshold metric: in this model, we define a threshold (Min and Max), by counting the events on a given dataset, any value that is higher than the Max defined or lower than the min should raise an alarm and be assumed as an anomaly.
b) Markov process: the Markov process fixes a specific interval of the data and track these intervals on time, any changes occurred on a given interval will be defined as an anomaly.
c) Mean and standard deviation model: the standard deviation (also spelled standard deviation) is a measure of the dispersion of the values of a statistical sample or probability distribution. It is defined as the square root of the variance or, similarly, as the square mean of the deviations from the mean. In this model, any event that falls outside the set interval defined by the standard deviation will be assumed as anomalous.
d) Multivariate model: this model takes into consideration more than one measurement to detect an anomaly, for example when talking about humidity it is affected by time and also by temperature, this model doesn't raise an alert about an abnormal measurement in humidity if it is justified by the temperature recorded, it applies correlation between two or more observations.
e) Time series model: in this model, a threshold isn't defined as in other models, since it computes the chance of having a specific measurement at a specific time taking into consideration previous measurements in time and also the time series data, its trend, seasonality. From the previous models we are going to work with time series data, for that we are going to apply different algorithms to detect anomalies by forecasting the time series data, this choice is justified by the fact that it fits the IoT architecture and also with the collected data by the sensors.

## 5. PROPOSED MODEL

In this study we are proposing a new anomaly-based Intrusion detection system using statistical models to forecast [21] and predict future values based on a small history base, the predicted value using stochastic models is close enough to the correct real values with a very small difference, once we get the real records we compare them with the predicted one if the difference was small enough it will be considered as a normal one, if not the real value will be reported as anomaly point, this technique is highly effective since it has a very small false positive rate and true negative rate compared to other techniques such as artificial intelligence.

The proposed intrusion detection system Figure 2 could be applied to any time series dataset, in our case we are proposing an approach to be applied to the collected information of sensors in an IoT system, the proposed model will help enhance the integrity and improve the accuracy, and trustworthiness of the data collected by sensors, it helps to indicate if the data was not altered by any unauthorized intruder, integrity is very important basic in security as well as in any system including IoT systems [22] and the collected information by the perception layer, in fact, sometimes what is more important in an IoT especially applied for health, industry, environment monitoring [23] is having the correct information, and the accurate values, these values that will be used later for taking actions and decisions about the whole process. The model proposed for this part is divided into four phases:

a) Collected data: the data collected by the sensors that will be compared to the predicted data.

b) Preprocess and analyze the data: this step will help to rescale the data if needed, clean the data of unnecessary values if found, and also plot the data as collected, that will help to analyze the information collected.
c) Detection engine: this is the part on which we are going to implement our model to perform the prediction using a historical dataset that is previously collected and checked, in the detection engine we implement the model, we predict values of the period given, and then we compare that values to the collected one, using other algorithms such as Mean square error, coefficient of determination (R2).
d) Report: the results obtained from the detection engine will be compared to a defined threshold, if it is higher or lower an alert is raised indicating an anomaly, otherwise everything will be reported as normal and the integrity of the data is checked.

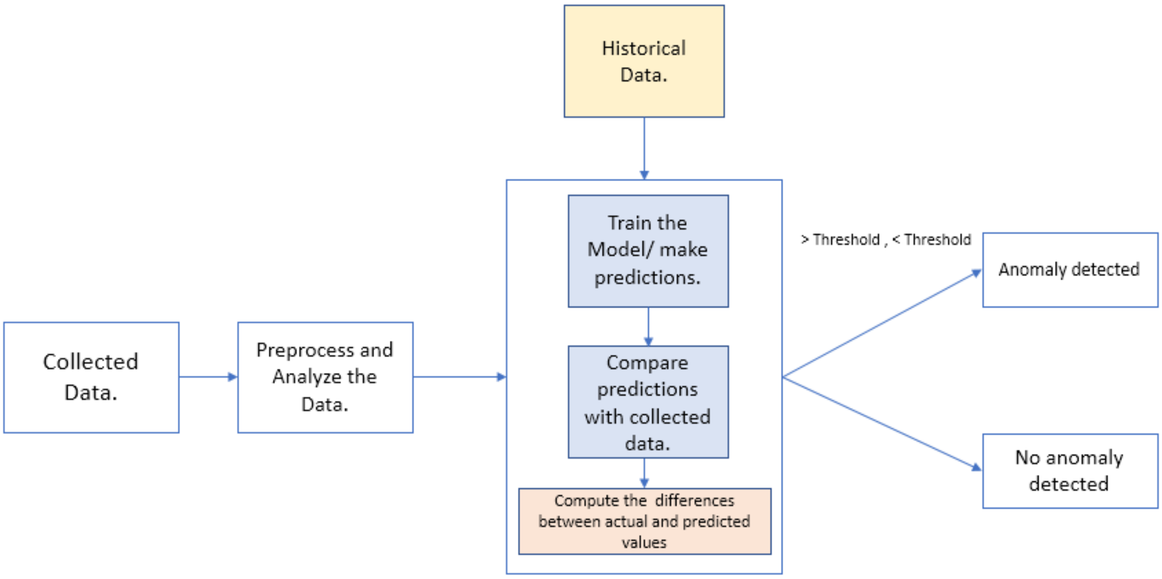

Historical Data.
Train the Model/ make predictions.
Compare predictions with collected data.
Compute the differences between actual and predicted values
Collected Data.
Preprocess and Analyze the Data.
> Threshold , < Threshold
Anomaly detected
No anomaly detected

Figure 2. Proposed model for anomaly-based intrusion detection system using forecasting models

## 6. EXPERIMENTAL TEST

### 6.1. Data analysis

Dataset: for our experiment tests we are going to use the dataset provided by Kaggle [24], the data contain 5 years of hourly measurement of different environmental values (temperature, humidity, pressure, weather direction, weather speed) the records were recorded in 36 different cities. The data fits perfectly with the needed model, it gives us exactly what we need to test on the model, we will not use the entire dataset we are going to retrieve intervals from the given datasets to perform the tests with the different models. The dataset is provided under the ODbL License. Since we are using the forecasting model for anomaly detection we don't need large history from the data, we search for a model that can forecast future values based on small history data, of course, we can use a large one but we could face the problem of seasonality, especially in environment records such as climates. We used partial data in Figure 3 going from 26-01-2017 to 31-01-201 from temperature.csv file, it is tricky for any model to predict the next value based on a small training set.

Data preprocessing: before choosing the interval on which we are going to work we analyzed the whole hourly temperature records of Portland (the choice of Portland was randomly and everything applied to Portland can be applied to other cities from the datasets provided by Kaggle).

From Figure 3 we can conclude data temperature records (or any other climates records collected by sensors) is a non-stationary one since there are so many factors that can affect the values over time, using the whole data can cause wrong predicted values, for that we split the data into the appearing seasons into the data, but it still a large dataset to use for anomaly detection, that is why we are going to plot one month from the partial dataset and then we can again define a more small part to work on in the next testing.

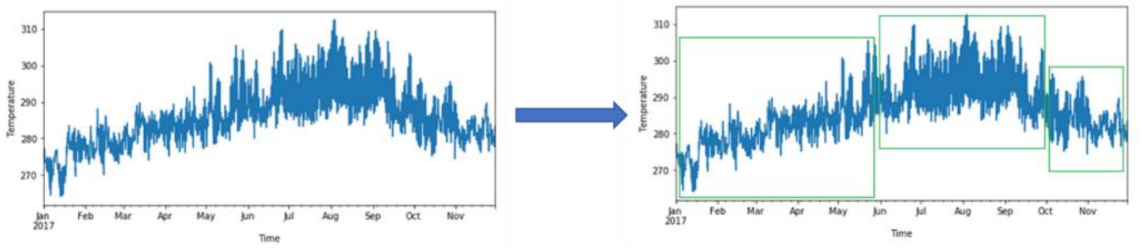


Figure 3. Hourly temperature records in 2017 in Portland

From Figure 4 we can see that even in one month of plotting there exist a lot of seasonality which is why we split once again the data into a more stationary and smaller dataset in our case we chose to work on data going from 26-01-2017 to 31-01-2017 That will be ideal interval to be applied in a real case and more precisely for anomaly detection to check the integrity of the data. The records in the data set chosen for test Figure 5 were already verified, so it doesn't include any anomalies or abnormal records Figure 5(a), for that we manually replaced two records by wrong ones to check the changes that occurred on the dataset Figure 5(b).

To study the distribution of the dataset records we plotted its histogram with and without an anomalous point Figure 6. In the original data we can see that the distribution of the values can be split into two sub gaussian distributions Figure 6(a) which is normal since in hourly temperature records in a day we always met a law temperature degree and a high one (can be justified by night and day), in the infected data we see that the distribution becomes a gaussian one Figure 6(b) since even with the low and high-temperature degrees they still close enough but the anomaly point even though they aren't massively changed they still plotted separately in the histogram.

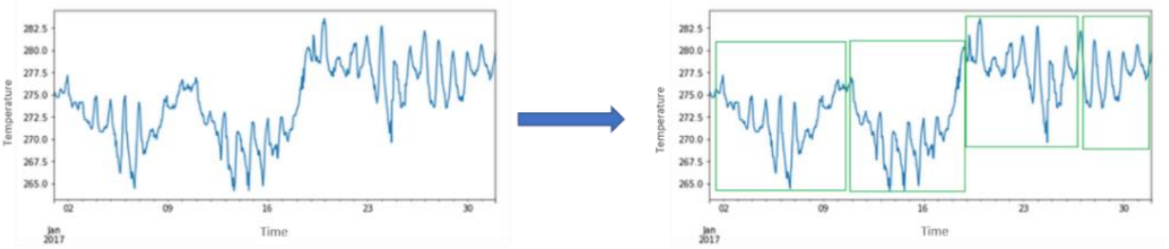
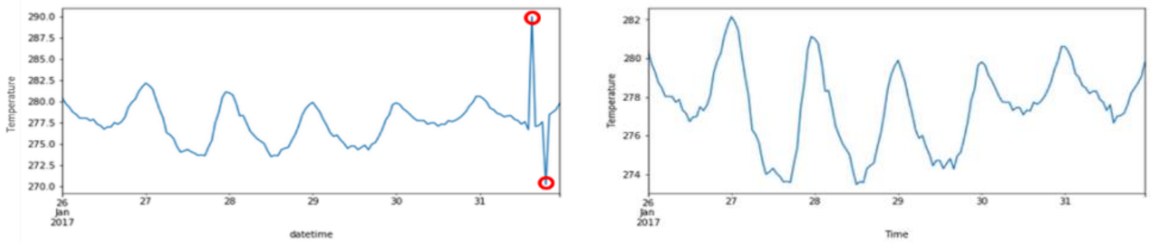


Figure 4. Hourly temperature records in January 2017 in Portland



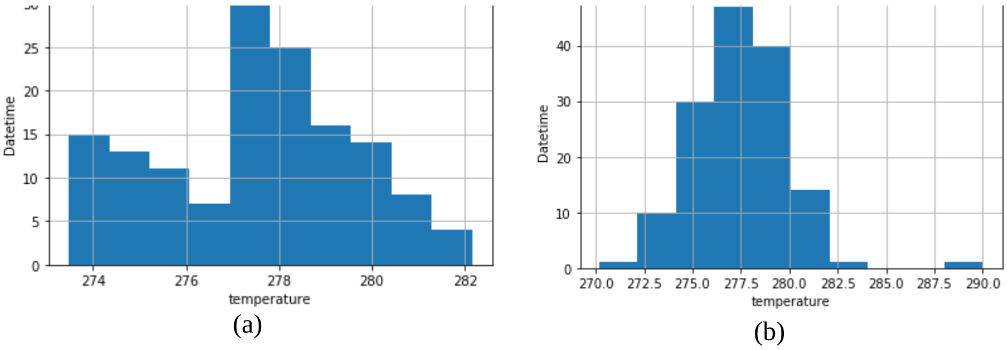
(a) (b)

Figure 5. Hourly temperature records dataset chosen for test in (a) with anomaly point added and (b) without anomalies or changes on the data

Test the stationarity of the dataset: to test if the dataset is stationary enough to apply the algorithm on it we are going to use the Dickey-Fuller Augmented test (ADF) [25]. The Dickey Fuller augmented test or ADF is a statistical test designed to determine whether a time series is stationary, i.e., whether its statistical properties (hope, variance, auto-correlation) vary over time. Using the adfuller function from statemodels.tsa library we can interpret if our dataset is stationary or not.

Null hypothesis: if not rejected, in this case, the time series has a unit root, meaning it is non-stationary. It has time-dependent structure. On the other hand, if the null hypothesis was rejected, we can assume that the time series does not have a unit root, which means it is stationary.

Applying the ADF test on our dataset (with the anomalous points) we are getting results of -4.856019 which is low than -3.479 at 1% which shows that the data can reject the null hypothesis with a significant level of less than 1% which means that our data is stationary at that point.



(a)

(b)

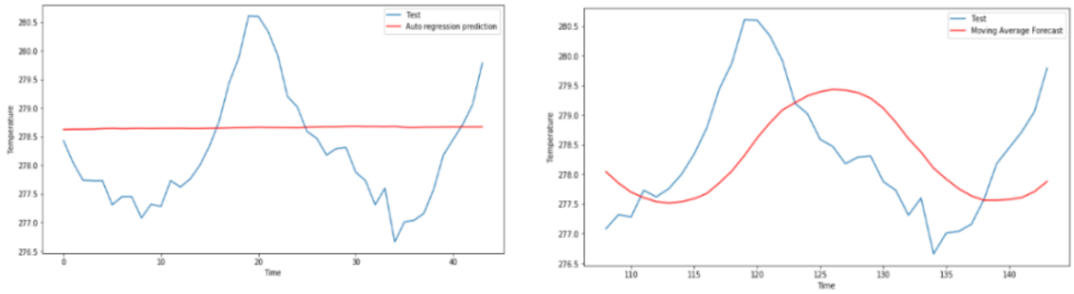
Figure 6. Histogram plot of the dataset in (a) with anomaly point and (b) without anomaly

### 6.2. Statistical models performance comparison

In this study we choose to use 5 different statistical models: i) AR: auto regression, ii) MA: moving average model, iii) ARMA: auto regression moving average, iv) ARIMA: auto regression integrated moving average [26], and v) SARIMA: seasonal auto regression moving average. To compare the models, we are using 70% of the data as training and 30% as testing, with the 70% we are going to train the models and make a prediction on the same interval as the testing part, and then we are going to compare the results obtained from the different models with the actual values of testing part [27]. For ARMA, ARIMA, and SARIMA the choice of the parameters was fulfilled using Akaike's information criterion (AIC) [28].

## 7. RESULTS AND DISCUSSION

In our study, the different statistical models were trained using 70% of the chosen dataset, error measurement of the predicted values in the models. To compare the models results we performed many tests to calculate the precision of the detected values, as well as the difference between the actual and estimated ones. In this research we tried to propose an approach for anomaly detection using statistical techniques, The results from plotting autoregression and moving average models Figure 7, clearly show that the autoregression model tends to be stationary over time with the same value Figure 7(a), this result is the same when using a much larger dataset (we used the entire month and still have the same stationary action from AR model) that means that this model is very weak for prediction and it cannot be used in anomaly detection, this model will work fine in a stationary environment when the data tend to be fixed but not on data with seasonality even small ones, this result was justified by the scores obtained by the different tests, mean square error (MSE) output a value of 1.156 for the AR, also the variance was very low to be good at 0.004 Table 1, proves there is almost no relation between the predicted values and the actual ones. Moving average plotting didn't quite reflect the data since it smooths the data in a way that it was plotted without showing the differences between the hourly records Figure 7(b), the MSE spotted this problem by giving a high score of 1.204, the mean absolute error (MAE) score was also high with a value of 0.943 Table 1, this model although it gives the shape of the data and how the whole data act, it doesn't give accurate and close predicted values that can be used in anomaly detection to check data integrity. In Figure 8 shows the plotted results of ARMA Figure 8(a) and ARIMA Figure 8(b) predictions, they were very close in the plotting, although ARIMA plotting was smoother and closer to the actual one, even if the error tests show that ARMA and ARIMA perform similarly with a small difference, that ARIMA tests were relatively better, the mean error of ARMA and ARIMA was of 0.882 and 0.707 respectively Table 1, they still high scores, but in fact, the predicted values were close enough to be considered as good models to use. However, the coefficient of determination (R2) was significantly low for both models, 0.143 for ARMA and 0.287 for ARIMA, the problem with these models is the seasonality, ARMA doesn't handle seasonality, and ARIMA removes the differences in the training data. SARIMA was the best model to use, it shows a great plot Figure 9, and also very good error test results, an MSE of 0.099 Table 1, close to 0 which means that the values estimated were very close to the real ones in the dataset.

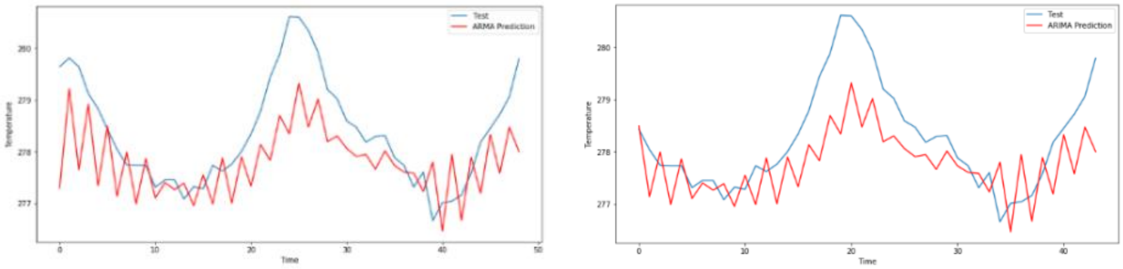


(a) (b)

Figure 7. Obtained prediction results using different statistical models in (a) by AR model and (b) by MA model

Table 1. Measurement error of different models using different tests

| Model | SE | Variance | R1 | MAE |
|---|---|---|---|---|
| AR | 1.156 | 0.004 | -0.166 | 0.931 |
| ME | 1.204 | -0.071 | -0.071 | 0.943 |
| ARMA | 0.882 | 0.458 | 0.143 | 0.728 |
| ARIMA | 0.707 | 0.523 | 0.287 | 0.660 |
| SARIMA | 0.099 | 0.903 | 0.903 | 0.249 |



(a) (b)

Figure 8. Obtained prediction results using different statistical models in (a) by ARMA and (b) by ARIMA model

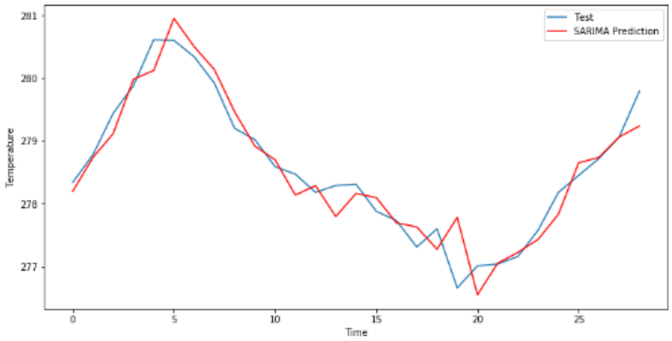


Figure 9. Prediction results by SARIMA model

The obtained results can be used in anomaly detection since any anomalous point will be spotted even if it was very small, and we can check the accuracy of the information. However statistical models may not be sufficient for anomaly detection when it comes to non-seasonal or big data, which needs a more complex algorithm such as in machine learning techniques to detect anomalous points.

## 8. CONCLUSION

Data integrity is a critical issue in security, we cannot focus on securing the network systems without taking into consideration verifying the accuracy of the data that is being transferred into the network. Anomalies detected in the data can indicate a possible malicious action, the main idea was to propose a model that can detect correct future values that could be collected by the sensors and then compare these predictions with the actual values, taking into consideration that after training the model with historical dataset, it will be able to provide accurate predictions, for that we proposed statistical models that fit perfectly with this idea, since IoT data collected by sensors are known as time series ones, in this work we tested several statistical models in order to compare theire performance, the results show that SARIMA outperforms other models (AR, MA, ARMA, and ARIMA), the obtained results were validated using a temperature dataset from Kaggle. The choice of usign statistical models come from the litrature review we did, from which we concluded that statictical models have a high-performance on time series datasets. For our future work we are planning to apply the best model on different datasets,to validate the results, and also compare it later with machine learning and deep learning techniques especially sequential algorithms such as RNN, this work will en with the creation of robust and well designed anomaly-based model for IoT datasets.

## REFERENCES

[1] M. Boujrad, S. Lazaar, and M. Hassine, “Performance assessment of open source IDS for improving IoT architecture security implemented on WBANs,” in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, New York, NY, USA: ACM, Mar. 2020, pp. 1–4, doi: 10.1145/3386723.3387892.

[2] V. Chaudhary, A. Kaushik, H. Furukawa, and A. Khosla, “Review—towards 5th generation AI and IoT driven sustainable intelligent sensors based on 2D MXenes and borophene,” *ECS Sensors Plus*, vol. 1, no. 1, p. 013601, Mar. 2022, doi: 10.1149/2754-2726/ac5ac6.

[3] M. Nooribakhsh and M. Mollamotalebi, “A review on statistical approaches for anomaly detection in DDoS attacks,” *Information Security Journal*, vol. 29, no. 3, pp. 118–133, 2020, doi: 10.1080/19393555.2020.1717019.

[4] A. Wahab, O. Ahmad, M. Muhammad, and M. Ali, “A comprehensive analysis on the security threats and their countermeasures of IoT,” *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 7, 2017, doi: 10.14569/ijacsa.2017.080768.

[5] H. Zare Moayedi and M. A. Masnadi-Shirazi, “Arima model for network traffic prediction and anomaly detection,” in *2008 International Symposium on Information Technology*, IEEE, 2008, pp. 1–6, doi: 10.1109/ITSIM.2008.4631947.

[6] A. H. Yaacob, I. K. T. Tan, S. F. Chien, and H. K. Tan, “ARIMA based network anomaly detection,” in *2010 Second International Conference on Communication Software and Networks*, IEEE, 2010, pp. 205–209, doi: 10.1109/ICCSN.2010.55.

[7] E. H. M. Pena, M. V. O. de Assis, and M. L. Proenca, “Anomaly detection using forecasting methods ARIMA and HWDS,” in *2013 32nd International Conference of the Chilean Computer Science Society (SCCC)*, IEEE, Nov. 2013, pp. 63–66, doi: 10.1109/SCCC.2013.18.

[8] Q. Yu, L. Jibin, and L. Jiang, “An improved ARIMA-based traffic anomaly detection algorithm for wireless sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, p. 9653230, Jan. 2016, doi: 10.1155/2016/9653230.

[9] F. Giannoni, M. Mancini, and F. Marinelli, “Anomaly detection models for IoT time series data,” *ArXiv (Prepint),* 2018.

[10] M. Braei and S. Wagner, “Anomaly detection in univariate time-series: a survey on the state-of-the-art,” *ArXiv (Prepint),* 2020, [Online]. Available: http://arxiv.org/abs/2004.00433

[11] L. Erhan *et al.*, “Smart anomaly detection in sensor systems: a multi-perspective review,” *Information Fusion*, vol. 67, pp. 64–79, 2021, doi: 10.1016/j.inffus.2020.10.001.

[12] P. Oser, F. Engelmann, S. Lüders, and F. Kargl, “Evaluating the future device security risk indicator for hundreds of IoT devices,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13867 LNCS, 2023, pp. 52–70,. doi: 10.1007/978-3-031-29504-1_3.

[13] I. Kalouptsoglou, D. Tsoukalas, M. Siavvas, D. Kehagias, A. Chatzigeorgiou, and A. Ampatzoglou, “Time series forecasting of software vulnerabilities using statistical and deep learning models,” *Electronics,* vol. 11, no. 18, 2022, doi: 10.3390/electronics11182820.

[14] N. Awang, G. A. N. Samy, N. H. Hassan, N. Maarop, and S. Perumal, “Implementation of SARIMA algorithm in understanding cybersecurity threats in university network,” *Turkish Online Journal of Qualitative Inquiry (TOJQI),* vol. 6, no. 3, pp. 8442–8451, 2022, [Online]. Available: https://www.journalppw.com/index.php/jpsp/article/view/5110

[15] D. Samariya and A. Thakkar, “A comprehensive survey of anomaly detection algorithms,” *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, Nov. 2021, doi: 10.1007/s40745-021-00362-9.

[16] F. Hafeez, U. U. Sheikh, A. Khidrani, M. A. Bhayo, S. M. Abdallah Altbawi, and T. A. Jumani, “Distant temperature and humidity monitoring: Prediction and measurement,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, pp. 1405–1413, 2021, doi: 10.11591/ijeecs.v24.i3.pp1405-1413.

[17] N. S. Nordin *et al.*, “A comparative analysis of metaheuristic algorithms in fuzzy modelling for phishing attack detection,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 2, pp. 1146–1158, 2021, doi: 10.11591/ijeecs.v23.i2.pp1146-1158.

[18] R. Van Den Goorbergh, M. Van Smeden, D. Timmerman, and Ben Van Calster, “The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression,” *Journal of the American Medical Informatics Association*, vol. 29, no. 9, pp. 1525–1534, 2022, doi: 10.1093/jamia/ocac093.

[19] W. A. Fuller, *Introduction to statistical time series,* vol. 20, no. 2. in Wiley Series in Probability and Statistics, vol. 20,Wiley, 1995. doi: 10.1002/9780470316917.

[20] H. G. Hamid and Z. T. Alisa, “A survey on IoT application layer protocols,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, p. 1663, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1663-1672.

[21] J. Scott Armstrong, “Research needs in forecasting,” *International Journal of Forecasting*, vol. 4, no. 3, pp. 449–465, Jan. 1988, doi: 10.1016/0169-2070(88)90111-2.

[22] A. A. Abbood, Q. M. Shallal, and M. A. Fadhel, “Internet of things (IoT): A technology review, security issues, threats, and open challenges,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1685–1692, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1685-1692.

[23] S. C. Poh, Y. F. Tan, S. N. Cheong, C. P. Ooi, and W. H. Tan, "Anomaly detection for home activity based on sequence pattern," *International Journal of Technology*, vol. 10, no. 7, pp. 1276–1285, Nov. 2019, doi: 10.14716/ijtech.v10i7.3230.

[24] D. Beniaguev, "Historical Hourly Weather Data 2012-2017." [Online]. Available: https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data

[25] J. Wolters and U. Hassler, "Unit root testing," *Allgemeines Statistisches Archiv*, vol. 90, no. 1, pp. 43–58, Mar. 2006, doi: 10.1007/s10182-006-0220-6.

[26] S. F. M. Hussein *et al.*, "Black box modelling and simulating the dynamic indoor air temperature of a laboratory using (ARMA) model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 791–800, Feb. 2020, doi: 10.11591/ijeecs.v21.i2.pp791-800.

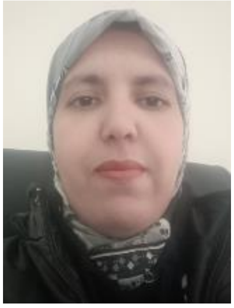
[27] K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria Engineering Journal*, vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.

[28] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, Sep. 1987, doi: 10.1007/BF02294361.

## BIOGRAPHIES OF AUTHORS

**Mouna Boujrad** received her Master's degree in Cyber-criminality and cybersecurity from the National School of Applied science in Tangier, Morocco. She is currently a Ph.D. student at the faculty of sciences, Mohammed First University in Oujda, Morocco. Her research area is on Artificial Intelligence for anomaly detection on IoT systems. Her research interests include security, internet of things, soft computing and artificial intelligence. She can be contacted at email: mouna.boujrad@gmail.com.

**Mohammed Amine Kasmi** received his Ph.D. in computer science from Mohammed First University in Oujda, Morocco. He is currently a senior lecturer at the faculty of sciences, Mohammed First University in Oujda, Morocco. His research interests include computer science security, internet of things and artificial intelligence. He can be contacted at email: a.kasmi@ump.ac.ma.



**Noura Ouerdi** received her Ph.D. in computer science from Mohammed First University in Oujda, Morocco. She is currently a senior lecturer at the faculty of sciences, Mohammed First University in Oujda, Morocco. Her research interests include computer science security, internet of things and artificial intelligence. She can be contacted at email: n.ouerdi@ump.ac.ma.

**Yasser Lamlili El Mazoui Nadori** is a Dr. at Mohammed First University in the Faculty of Sciences. He got a degree in engineering in Computer Sciences from the National School of Applied Sciences at Oujda. He received his M.Sc. degree in New Information and Communication Technologies from the faculty of sciences and Techniques at Sidi Mohamed Ben Abdellah University. His research activities at the MATSI Laboratory (Applied Mathematics, Signal Processing, and Computer Science) have focused on web marketing in social networks using the MDA (Model Driven Architecture) approach. He can be contacted at email: lamliliyasser@gmail.com.