Detecting Anomalies, Sabotage, and Malicious Acts in a Cyber-physical System Using Fractal Dimension Based on Higuchi's Algorithm

Marwan Albahar
mabahar@uqu.edu.sa
Department of Science, Umm Al Qura University, P.O. Box 715
Mecca, Saudi Arabia

Summary
With the global rise of digital data, the uncontrolled quantity of data is susceptible to cyber warfare or cyber attacks. Therefore, it is necessary to improve cyber security systems. This research studies the behavior of malicious acts and uses Higuchi Fractal Dimension (HFD), which is a non-linear mathematical method to examine the intricacy of the behavior of these malicious acts and anomalies within the cyber physical system. The HFD algorithm was tested successfully using synthetic time series network data and validated on real-time network data, producing accurate results. It was found that the highest fractal dimension value was computed from the DoS attack time series data. Furthermore, the difference in the HFD values between the DoS attack data and the normal traffic data was the highest. The malicious network data and the non-malicious network data were successfully classified using the Receiver Operating Characteristics (ROC) method in conjunction with a scaling stationary index that helps to boost the ROC technique in classifying normal and malicious traffic. Hence, the suggested methodology may be utilized to rapidly detect the existence of abnormalities in traffic with the aim of further using other methods of cyber-attack detection.

Keywords:
Cyberattacks, malicious attacks, nonlinear mathematical method, Fractal Dimension

1. Introduction

Network control systems are being deployed in increasing numbers to manage, monitor, and respond to physical infrastructure. SCADA systems compatible with Internet of Things devices are gradually displacing legacy infrastructure, enabling remote control and monitoring of isolated systems, as well as data collection and analysis. This upgrade benefits industries by increasing the flexibility as well as reliability of their deployed systems. Nevertheless, allowing external internet connectivity, may expose industries' potentially critical infrastructure to an increased risk of computer viruses [1]. Five distinct causes contribute to the vulnerability of critical infrastructure protection (CIP) infrastructure. The first is a deficiency in the availability of open protocol standards. Programmable logic controllers (PLCs) are usually controlled by proprietary network protocols that have not been thoroughly inspected by security experts. As a result, PLCs operate as black boxes, significantly increasing the attack surface [2]. The second reason is the lack of network infrastructure segmentation [3]. Due to the widespread use of PLCs that are connected to control networks, higher management can analyze data in real time. This may result in higher profit margins and shorter reaction times to market changes. However, the lack of segmentation makes transversal network attacks a very real threat The third reason is due to the network's increased reliance on commercial off-the-shelf (COTS) equipment [4]. The attack surface is increased by utilizing commercially available hardware, which opened additional potentially weak links. The fourth reason is due to a lack of training and an inability to differentiate between cyber-attacks and incidents [5,6]. The fifth is the result of organized cybercrime's exponential growth, as well as state-sponsored cyberwarfare aimed at destabilizing nations [7,8].

A Cyber-Physical System (CPS) is a complex, multidimensional system that combines control technologies, communication, and computing. Due to their critical function within the system, CPSs require a high level of security as well as robustness to ensure their continued operation is reliable. Due to its critical role in ensuring overall anomaly detection and system security, CPS is likely to remain a critical component. Additionally, due to the nature of CPS systems, CPS data is more likely to exhibit implicit correlative relationships between data points, which would be critical to exploit in more complex data environments for CPS security provisions [9]. Fractal analysis was developed as a technique for studying mathematical sets of various types based on the ideas of fractal geometry. Fractal analysis techniques have been widely applied in a variety of fields, including biology, chemistry, and physics. The most significant achievement of fractal theory has been the development of simple mathematical descriptions of extremely complex but ubiquitous natural phenomena and objects. Just like differential trigonometry, harmonic analysis, and equations, fractal analysis is a fundamental mathematical tool for modeling and explaining physical reality [10]. In view of this observation, the work demonstrates the breadth of possible applications of Higuchi Fractal Dimension (HFD) to examine the complexity of malicious traffic and anomalous behavior within a cyber-physical system.

2. MOTIVATION

Fractal properties on a large scale of network traffic have the property of self-similarity, which means that they

Manuscript received April 5, 2023
Manuscript revised April 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.4.9

appear qualitatively identical on a sufficiently large scale of the time axis and exhibit long-term dependence [11]. As computer network traffic data is stochastic and complex in nature, linear analysis techniques are limited in detecting malicious attacks on data networks. Therefore, this research employs a powerful nonlinear analysis technique that is well-reputed in the field of medical research and is called fractal dimension analysis. Moreover, the Higuchi algorithm is used to compute the fractal dimension of the data network time series data. The computed fractal dimension represents a nonlinear metric that will be eventually used to classify malicious data and normal data. A nonlinear metric is a measure of the complexity of the data under test, as most real-life problems are non-linear in nature as they change stochastically.

The contribution of this paper is as follows:

1. We propose a new approach to investigate the complexity of malicious traffic and detect anomalous behavior within a cyber-physical system based on the analysis of the fractal properties of traffic.

2. Several experiments are performed to verify the performance of the proposed approach using four datasets. In addition, Receiver Operating Characteristics (ROC) were used to assess the classification performance of the computed fractal dimension nonlinear metrics data.

3. THREAT MODEL & SOLUTION

In this section, we first introduce our threat model, then discuss the impact of malicious objects on network systems, provide a real-life scenario, and propose a solution. The behavior of malicious objects is so varied, and the types of cyber-attacks can range from passive to active to advanced types of attacks. Therefore, it is of utmost importance to understand the behavior of these various types of attacks in order to create a defense network system against such attacks. One approach to resolving this problem is to use a non-linear mathematical model to investigate the characteristics of malicious attacks.

A. THREAT MODEL

In this research, the threat model concerns malicious objects attacking a cyber-physical subsystem. The actual network data signals that are captured comprise cyber attacks as well as anomalies, sabotages, and system breakdowns.

B. Real World Example (Petroleum management)

It is evident that there are cases of cyber-attacks on industrial production and distribution systems, which include oil, gas, and electricity. The security of a particular infrastructure relies not only on its internal vulnerabilities but also on the vulnerabilities of other infrastructures that it depends on. When coupled with that, by being aware of a system weakness, this puts a given infrastructure at risk. For example, malicious actors can use vulnerabilities to launch aggression against this weak infrastructure [13]. The severity of cyber-attacks and their extent to achieving deadly goals in real life can be measured by studying different instances. The presence of malicious objects or malicious software revealed various industrial security incidents, such as in [14,15,16,17]. Likewise, the heinous highlighted attacks against the Iranian Petroleum Plants (2016) and the petrochemical company with a plant in Saudi Arabia (2017) revealed that cyber-physical systems are subjected to attacks on their respective physical infrastructure, communication, and data management layers. These attacks demonstrated that when attackers target a CIP, they frequently conducted extensive research on the system in order to carry out the attack as stealthily as possible, tailoring their strategies to the specific system.

C. SOLUTION

It is clear that malicious objects have disastrous effects on the cyber defence mechanism. To address these issues, this research employs a state-of-the-art nonlinear mathematical analysis to identify threats and malicious attacks. This nonlinear analysis method is used to identify and distinguish malicious attacks from other non-malicious attacks that often happen in a cyber-physical system. The nonlinear mathematical method considers the dynamics and behaviour of malicious objects in a cyber-physical system because the behaviour of malicious objects is unpredictable. Therefore, the nonlinear method is important.

4. RELATED WORK

A wide range of IDSs have been developed and tested against publicly available datasets. IDS design for various applications and the machine learning techniques used to construct IDS have been the subject of numerous reviews and comparative studies. However, the dataset challenges remain unmentioned. As a result, the majority of these studies focus exclusively on one aspect of IDS evaluation rather than on the entire system. Hodo et al. [19] broadened the subject of machine learning techniques by concentrating on the importance of feature selection in the entire training and evaluation of machine learning approaches. Hamed et al. [23] categorized IDS components as follows: (a) pre-processing/feature extraction, (b) pattern analyser, which involves knowledge representation and learning procedures, and (c) decision-making. They briefly discussed the advantages of each instructional method. Features and their impact on the design and accuracy of IDS were thoroughly investigated by Varma et al. [20]. Buczak and Guven [18] described several machine learning and deep learning methods that were employed to build IDS. They explained various algorithms, including their time complexity, and provided a list of significant articles that employed each

technique to address the IDS problem. The characteristics of IDS have been discussed in detail by Debar et al. [21], Amer and Hamilton [22]. There are also several other aspects of intrusion detection systems (IDS) that are discussed. Amit et al. [24] discussed various difficulties and challenges inherent in developing IDSs that use machine learning. A single network architecture is the focus of several other perspectives, some of which have been included in recent studies. Ismail et al. discussed Wireless Sensor Networks (WSN) and their applicability to IDs. Zhou et al. [26] discussed IDs in industrial process automation, and Ghaffarian and Shahriari [27] investigated machine learning (ML) and data mining (DM) techniques for detecting software vulnerabilities. These surveys revealed the design and accuracy of different methods, but there is no comprehensive overview of the dataset's shortcomings or information on the tools that were used to conduct attacks.

5. METHODOLOGY

5.1 Data Collection from a Cyber-Physical system

In our study, we analysed a publicly available dataset [28]. The dataset includes several common scenarios (normal scenario, anomaly scenario, breakdown scenario, sabotage scenario, and cyber-attack scenario), each of which corresponds to a different real-world situation. The authors in [28] collected the dataset in the manner depicted in Fig. 1. As illustrated in Figure 1, there were two containers of varying sizes. To collect the data, one ultrasound depth measurement sensor and two pumps were used in addition to the four discrete sensors. The system was controlled by a computer via a PLC connector connected to a monitoring network system. The ultrasound sensor indicates when the main container is filled with liquid and when it is emptied. Pump 2 begins filling the second volume container, while Pump 1 begins filling the main tank. Tank 2 displays the states of the four discrete sensors (sensor 0, sensor 1, sensor 2, and sensor 3).

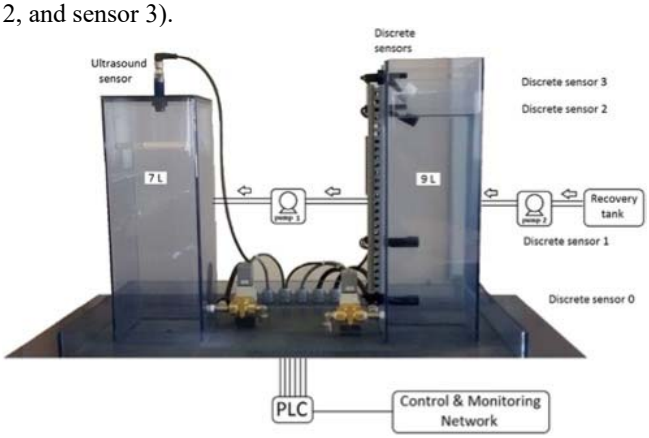

Fig 1: The platform of the cyber-physical system which was used for data collection [28]

5.2 Nonlinear mathematical method for nonlinear systems

A nonlinear system in the fields of science and mathematics is regarded as a system where input change is not proportional to output change. As most systems found in nature are inherently nonlinear, this system appears uncertain or chaotic in relation to a simple linear system. In a general perspective, the mathematical expression of nonlinear system behaviour is expressed with the help of a nonlinear system of equations with the help of variables constituting a polynomial of degree which is greater than 1. As it is difficult to solve the nonlinear dynamical equations, the nonlinear systems are normally computed and approximated through linearization [29]. The fractal dimension, however, is an important nonlinear mathematical method used to determine the behaviour and complexity of signals. The fractal analysis method is well grounded in the field of medicine to depict any changes in biosignals, and previous research demonstrated that the Higuchi's method produces a more accurate estimate of the fractal dimension (FD) of a signal when it is being tested on synthetic data, but it is noise sensitive [29,30].

5.3 Fractal Dimension using Higuchi's algorithm

The fractal dimension [30] of a time series can be computed directly using Higuchi's algorithm in the signal's time domain. Higuchi's algorithm is dependent on the measurement of the length, L (k), of the curve that denotes the time series, and it uses a segment of k samples as a unit only if L (k) scales as per the following equation (See Equation 1):

$$\mathbf{L}(\boldsymbol{k}) \sim k^{-D_f} \qquad (1)$$

D_f value is always in range of 1 and 2. D_f value is equivalent to 1, which indicates a simple curve. Whereas a D_f value of 2 represents a curve that approximately fills out the whole plane. From a given time series, *X (1), X (2), X(N),* the *k* new time series of the algorithm is constructed as:

$\boldsymbol{X_{km}}$: *X(m), X(m+k), X(m+2k) ..., X* (m + int ($\frac{(N-m)}{k}$).*k*) for *m* = 1, 2, *k*

Where *m* is the initial time, *k* is the interval time, int (*r*) is integer part of a real number *r*.

The length $\boldsymbol{L_m}$ **(k)** of each curve $\boldsymbol{X_{km}}$ is determined as follows (See Equation 2):

$$L_m = \frac{1}{k}\left[\left(\sum_{i=1}^{\text{int}((N-m)/k)} |X(m+i\cdot k) - X(m+(i-1)\cdot k)|\right) \times \frac{N-1}{\text{int}\left(\frac{(N-m)}{k}\right)\cdot k}\right], \quad (2)$$

Where N is the total number of samples.
D_f (Fractal dimension) is computed through a linear least square best fitting method to determine the angular regression coefficient of the *log-log* plot of equation (1).

5.4 Pseudocode of Fractal dimension model using Higuchi's algorithm and Its Time Complexity

We computed the time complexity of Fractal dimension using Higuchi's algorithm as follows (See Equation 3):
Time complexity

$$O(1) + O(1) + O(1) + O(n).\ O(n) + \log(n) + \log(n) = O(n^2) \quad (3)$$

*In calculating complexity, we discard the lower terms

Fractal dimension using Higuchi's algorithm
Step 1: Load (dataset) *Step 2: Initialize variables* ***N*** *and* ***X*** *which represent length of data and* ***X*** *is a numeric data type* *Step 3: Initialize variable* ***k*** *= 6 (based on literature) some length or distance* *Step 4: Compute* ***Lm*** *(Length of each curve Xkm)* *Step 5: Compute* ***Df*** *using linear least square best fitting method (angular regression coefficient of log(L(k)) vs.log (k)* *End*

5.5 Classification of the FD nonlinear metrics using Receiver Operating Characteristics curve

A receiver operating characteristic curve plots the true positive rate (sensitivity) against the false positive rate for different cut-off points. The sensitivity of a test represents the proportion of the fractal dimension values that represent the malicious network data and produces a positive result. The specificity of a test (true negative rate) is the proportion of the computed fractal dimension values that do not represent any malicious network data and produce negative outcomes (See Equation4 and 5). Each point on the ROC plot represents a sensitivity/specificity pair on the ROC curve. For instance, a test that has perfect discrimination between two distributions has a ROC plot that passes through the upper left corner, which means 100% sensitivity as well as 100% specificity. As such, the closer the ROC plot is to the upper left corner of the ROC curve, the higher the overall accuracy of the ROC test.

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (4)$$

$$\text{Specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} \quad (5)$$

6. Result

The fractal analysis method using Higuchi's algorithm [30] was applied to the time series data for each type of simulated scenario, and these time series data were recorded via the PLC registers continuously. The mean of the data acquired through the various simulated scenarios was computed as well as their respective fractal dimension values. The simulated scenarios were specifically called:

- Bad Connection: simulated dataset
- *Simulated DoS Attack Dataset*
- simulated dataset hits
- A normal simulated dataset
- Simulated Plastic Bag Dataset
- Spoofing a simulated dataset
- Simulated dataset from a wet sensor

6.1 Testing of the implemented Higuchi's algorithm

The Higuchi's algorithm was tested using two artificial datasets that were both created in the Matlab environment (See Fig. 2 and 3). The first artificial dataset was generated using a *sine wave function* (Function_Sine) and the second artificial dataset was generated using a *random (rand) function* to test the reliability of the implemented Higuchi's algorithm in computing the fractal dimension of time series signals. The function_Sine = *sin* (0:0.01:100*π).

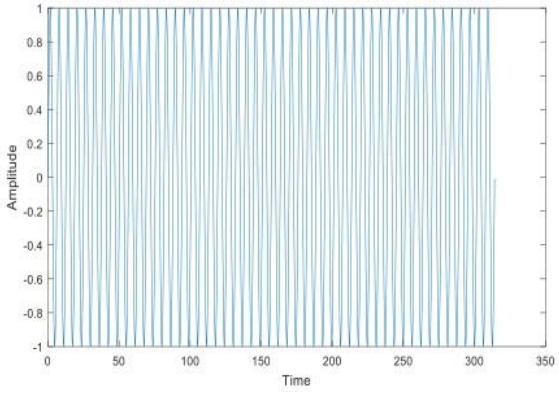

Fig 2: Sine function with amplitude 1 and computed Higuchi Fractal Dimension (HFD) is 1.0001. The accuracy of this algorithm is determined to be 99.99%.

Function_Random= *rand* (1,10000)

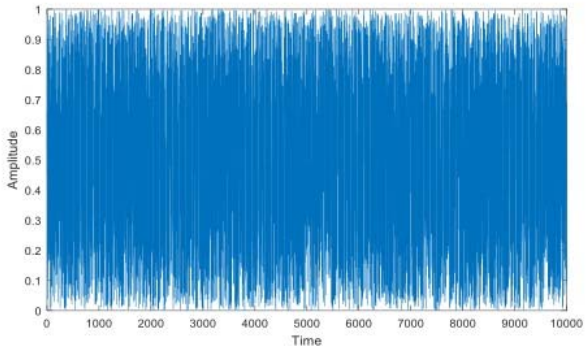

Fig 3: A random data was generated for 10000 data points and the HFD was found to be 2.0002.

6.2 **Application of Fractal Dimension using Higuchi's algorithm on the operational scenarios time series data.**

Table 2: HFD for the simulated operational scenarios

Simulated operational Scenarios	**Fractal Dimension**	**Fractal Dimension Deviation from Normal Data**
Bad Connection	1.6689	0.0747
***DoS* Attack**	**1.8440**	**0.1004**
Hits on Tank	1.7484	0.0048
Normal Data	1.7436	0
Plastic Bag	1.6632	0.0057
Spoofing Attack	1.6704	0.0804
Wet Sensor	1.7909	0.0473

∴Thus, the *DoS* Attack, $\mathbf{L(k)} \sim k^{-1.8440}$.

From Table 2, it is clearly observed that each simulated scenario produced a different fractal dimension value, and the change in fractal dimension value was included as a third column to see how a particular anomaly's characteristic behaviour changes as compared to the complexity of normal network data. In addition, we observed that the network data that is infected with malicious data (DoS attacks) produces the highest fractal dimension value. In addition, the difference between the fractal dimension value for the malicious data (1.8440) and the normal data (1.7436) is positive and the magnitude is 0.1004 (See Fig. 4a). This difference is the most significant and positive as compared to the differences between the non-malicious data and the normal data, which are much lower. Therefore, based on this key observation, a *scaling stationary index* was included, so that the ROC computation detects all those fractal values that are above a threshold value (determined by trial and error during computation) in order to achieve a very high accuracy in distinguishing between malicious and non-malicious data. By doing so, our proposed model becomes more flexible because it can be tuned to detect malicious data from normal traffic data based on the difference between the fractal dimension complexity of the malicious data and the normal data, where the fractal dimension of the time series of the network data will represent the reference point.

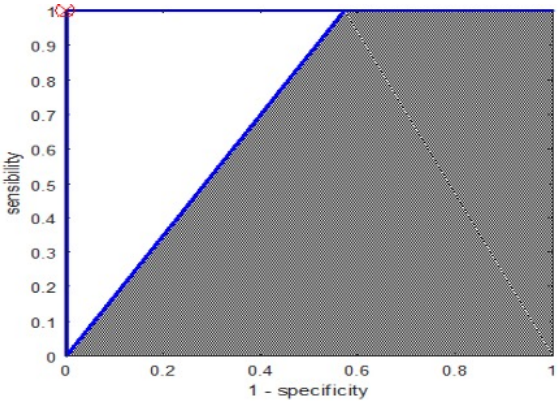

Fig 4a: Sensibility to 1-specificity curve before the scaling stationary index is added to the process.

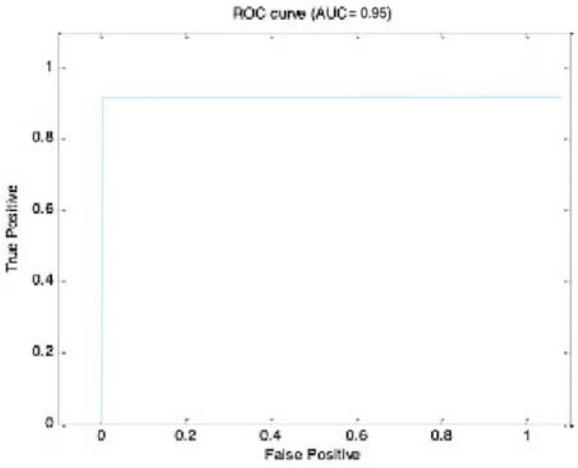

Fig 4b: ROC curve analysis of the computed fractal dimension values and the targeted outcomes after the inclusion of the scaling stationary index.

Figure 4b shows the ROC curve produced between the calculated fractal dimension of the various types of network data and the corresponding targeted output. The area under the curve was determined to be 0.95 and the accuracy was found to be 98%. The accuracy reached its maximum by trial and error while changing the threshold value in order to achieve such maximal accuracy, as well as sensitivity and

specificity. The optimum threshold value for maximum accuracy was found to be 1.69. The parameters of the ROC curve as shown in Fig 4b are summarized in the following Table 3.

Table 3: Computed ROC curve parameters

Parameter names	Computed values
AROC	0.95
Specificity	0.98
Sensitivity	1.0
Accuracy	98%
scaling stationary index	1.69

Figure 5 depicts the entire process of the various stages involved in order to classify malicious data from normal data or non-malicious data.

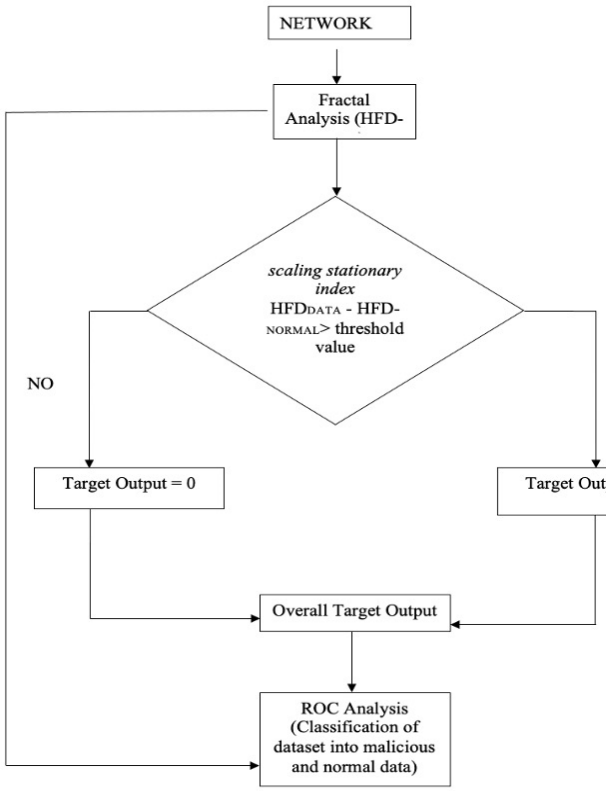

Fig 5: Stages of the entire experimental process

6.3 Validation of findings for Darpa Intrusion dataset

In order to validate the findings, the fractal dimension was applied to a real network dataset. The Darpa99 dataset [31] was used as benchmark data to see if the fractal dimension could be used to predict whether the network data encountered malicious attacks. The Darpa99 looks very stochastic and represents a real-case scenario, and any change in the network data is not easily perceptible with the naked eye. The fractal dimension was used to test some samples in the data, and it was found that the fractal dimension value of the Darpa99 dataset was 1.85, which is close to the predicted model that was formulated in the previous section. Therefore, based on the high fractal dimension values of the traffic network data, which contained *DoS* attacks for both types of data (synthetic data and real-time networked data), a *scaling stationary index* was included in the post-fractal analysis program to confirm whether a data network contained malicious. The fractal dimension value is determined from the network data as well as the slope, as shown in Figure 6 and 7.

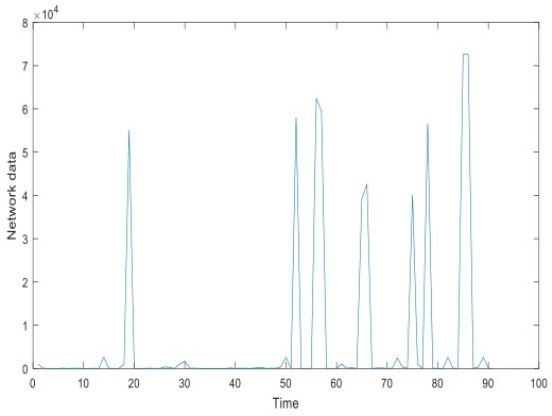

Fig 6: Darpa99 data network

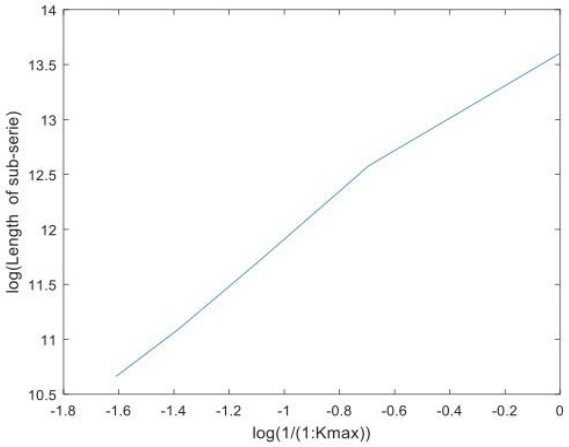

Fig 7: DARPANET network data infected with malicious objects whose fractal dimension value is 1.85 (which was determined from the slope of the graph)

The accuracy of the method was analysed using the Receiver Operating Characteristic method, and it was found to be 98%. A more detailed result based on the application of ROC on the computed fractal dimension data is provided in Table 3. ROC curves represent a very good method to see how any predictive model can distinguish between true positives and negatives.

Table 4: ROC Curve Parameters

	***Before* scaling stationary index**	***After* scaling stationary index**
AROC (AUC)	0.87	0.95
Specificity	0.81	0.90
Sensitivity	0.85	0.97
Accuracy	80%	98%

In Table 4, the variable AROC, which is also known as AUC (Area Under Curve), represents the area under the ROC curve produced from sensitivity (y-axis) vs. specificity (x-axis). The sensitivity of the ROC curve was found to be 0.95, and it is the ability of a test to identify correctly those data networks that contain malicious attacks, whereas the specificity is the ability of the test to identify correctly the network data that does not contain any malicious data. The specificity was found to be 0.90. As such, the fractal dimension analysis of the Darpa99 confirms that the complexity of the time series network data is very high in the presence of malicious data, as the fractal dimension values for malicious data networks are very high as well as positive. Based on this important observation, a *scaling stationary index* was further included in our model after the application of the fractal analysis, to facilitate the classification process by the ROC and ensure the classification is excellent.

D. Validation finding for and UNSW-NB15 and NSL KDD Datasets

Apart from the Darpa99 Intrusion dataset, we also tested the method on two other well-known benchmark datasets known as the NSL KDD and UNSW-NB15 datasets [32,33]. Fractal dimensions were calculated for both datasets and it was found that the fractal dimension value for the NSL KDD data set for DoS attacks is 1.9945, while for the UNSW dataset, the fractal dimension value for DoS attacks is 1.9696. Upon testing with the model given in Figure 8,9,10,11, and 12, we found the accuracy for NSL dataset and UNSW dataset were 93% and 99% respectively. In addition, the AUC computed was 0.95 for NSL dataset and 0.94 for UNSW dataset. However, the algorithm yields a specificity of 0.90 for the NSL dataset and for the UNSW dataset it is 0.98. Similarly, sensitivity turned out to be in the same range as specificity for both datasets.

Table 5: For each category in NSL dataset has a different fractal dimension value and the change in fractal dimension value was included as a third column.

Categories	*Fractal Dimension*	*Difference with Normal value*
Normal	**2.0006**	**0**
DoS	1.9945	0.0061
Probe	**2.0459**	**0.0453**
R2L	1.9531	0.0475
U2R	1.9928	0.0078

Table 6: For each category in UNSW dataset has a different fractal dimension value and the change in fractal dimension value was included as a third column.

Categories	*Fractal Dimension*	*Deviation from Normal*
Normal	**2.0032**	**0**
DoS	1.9696	0.0336
Analysis	1.9717	0.0315
Backdoor	1.9406	0.0626
Exploits	1.9766	0.0266
Fuzzers	1.9553	0.0479
Generic	1.9902	0.0130
Reconnaissance	1.9948	0.0084
Shellcode	1.9758	0.0274
Worms	**2.0285**	**0.0253**

Table 7: Results obtained for different measures for NSL and UNSW datasets

	NSL Dataset	UNSW Dataset
AROC (AUC)	0.95	0.94
Specificity	0.90	0.98
Sensitivity	0.91	0.99
Accuracy	93%	99%

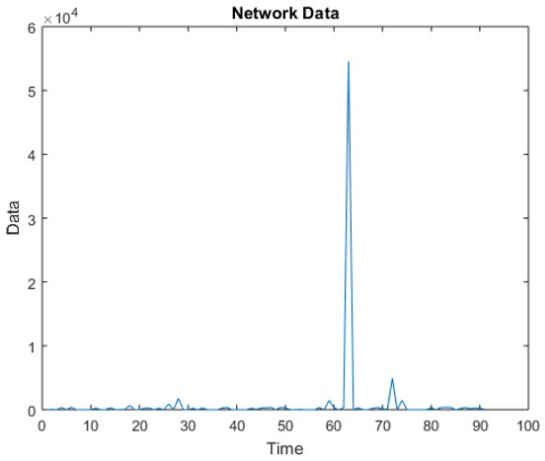

Fig 8: NSL KDD data network

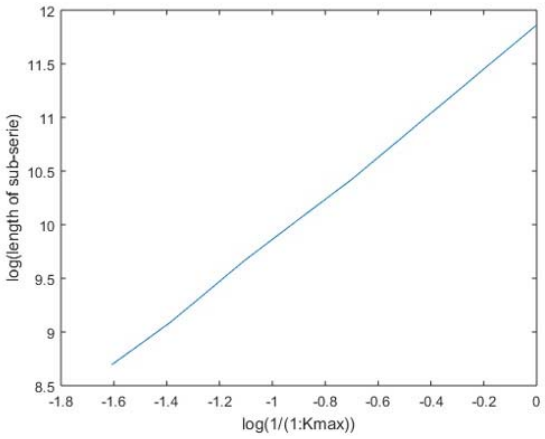

Fig 9: NSL network data infected with malicious objects whose fractal dimension value is 1.9945

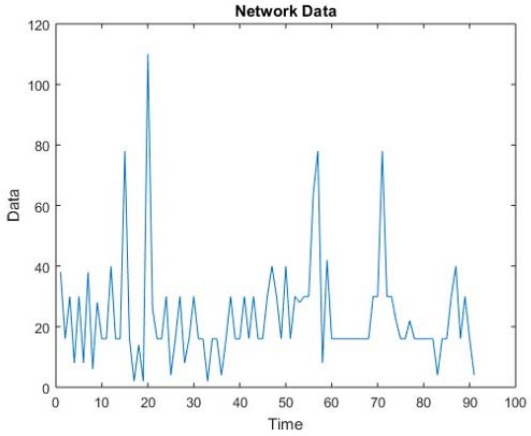

Figure 10: UNSW-NB15 data network

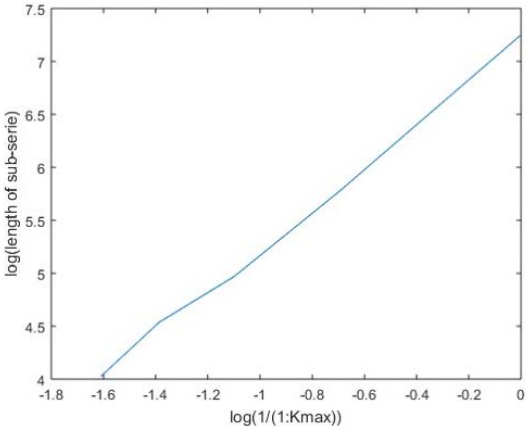

Fig 11: UNSW-NB15 network data infected with malicious objects whose fractal dimension value is 1.9696

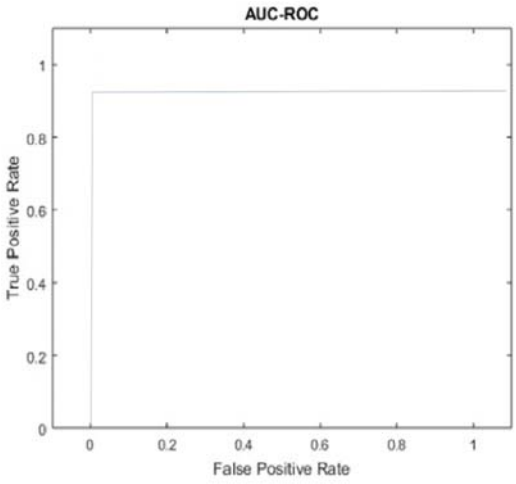

Fig 12: AUC-ROC Curve for both NSL and UNSW datasets

7. DISCUSSIONS

It is clearly observed that Fractal Dimension is a powerful nonlinear algorithm that can be used in computing the complexity of a particular time series of data signals. Fractal dimension is a more powerful technique in terms of robustness to noisy stochastic data than linear analysis techniques, which are often used for malicious object detection, such as frequency analysis and wavelet analysis, because the latter exploit the visual properties (time and frequency) of the network data signals. The fractal dimension analysis produces a unique value for each assessed scenario by quantifying the non-stationary network dataset. Moreover, the fractal dimension can be easily implemented owing to its low algorithm complexity. This method was used in order to dissociate malicious data from normal datasets. The accuracy of the implemented fractal dimension together with the receiver operating characteristics yields an accuracy of 100%. It was found that the HFD for the network data that contained *DoS* attacks was the highest, and the difference between the HFD values for the *DoS* attack traffic data and the normal traffic data was the highest. However, even though the HFD value for the spoofing attack was low, the difference between the HFD value for the spoofing attack and that of the normal dataset ranked below the *DoS* Attack. Furthermore, it is also tested on well-known benchmark datasets such as NSL-KDD and UNSW-NB15 datasets. The NSL dataset has an HFD value of 1.9945 for DoS attacks, and the UNSW dataset has a value of 1.9696. For the NSL dataset, the Probe attacks have the highest HFD value of 2.0459, but when the difference from the normal value is analysed, the highest deviation from the normal value is 0.0475 for the R2L attacks. Similarly, for the UNSW dataset, the worm attacks have the highest HFD value of 2.0285 but the backdoor attacks have more deviation from the normal data. Instead of such an abnormal

deviation, the proposed algorithm performed well and has produced satisfactory results. In terms of numbers, it can be seen from table 7. For each measure, such as TPR, accuracy, and AUC-ROC, it has given perfect results instead of specificity for the UNSW dataset, as it is a complex and recent dataset in this domain and contains advanced attacks. However, overall, the fractal dimension analysis performed well, as evident from the results. Therefore, the fractal dimension analysis method can be implemented in a network system together with a classification method such as the Receiver Operating Characteristics in order to monitor the network immune system as well as to detect any malicious activities that occur in the network. This research successfully classified malicious data (especially DoS-attacked network data) and non-malicious data with maximum accuracy. One key limitation of this research is that it does not discriminate between normal and non-malicious data, such as "Hits on Tank" as an external factor.

8. CONCLUSION

A non-linear mathematical method called Higuchi Fractal Dimension (HFD) is used in this research to investigate the intricacy of the behaviour of malicious acts and anomalies within the cyber physical subsystem. The HFD algorithm was successfully tested on synthetic time series network data and validated on real-time network data, resulting in accurate values being produced. It was discovered that the time series data that had been subjected to DoS attacks had the highest fractal dimension value. Furthermore, the difference in HFD values between the data from the DoS attack and the data from normal traffic was the most significant. The Receiver Operating Characteristics (ROC) method was successfully applied in the classification of both malicious network data and non-malicious network data. A *scaling stationary index* was used to aid in the classification of both normal network data and malicious data using the Receiver Operating Characteristics (ROC) method. As a result, Fractal Dimension has proven to be an important component in the tracking of cyber attacks.

References

[1] C.-W. Ten, G. Manimaran, and C.-C. Liu, "Cybersecurity for Critical Infrastructures: Attack and Defense Modeling," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 40, no. 4. Institute of Electrical and Electronics Engineers (IEEE), pp. 853–865, Jul-2010.

[2] A. Calderón Godoy and I. González Pérez, "Integration of Sensor and Actuator Networks and the SCADA System to Promote the Migration of the Legacy Flexible Manufacturing System towards the Industry 4.0 Concept," Journal of Sensor and Actuator Networks, vol. 7, no. 2. MDPI AG, p. 23, 21-May-2018.

[3] N. Jiang, H. Lin, Z. Yin, and C. Xi, "Research of paired industrial firewalls in defense-in-depth architecture of integrated manufacturing or production system," 2017 IEEE International Conference on Information and Automation (ICIA). IEEE, Jul-2017.

[4] A. Bujari, M. Furini, F. Mandreoli, R. Martoglia, M. Montangero, and D. Ronzani, "Standards, Security and Business Models: Key Challenges for the IoT Scenario," Mobile Networks and Applications, vol. 23, no. 1. Springer Science and Business Media LLC, pp. 147–154, 20-Feb-2017.

[5] X. Bellekens, R. Atkinson, A. Seeam, C. Tachtatzis, I. Andonovic, and K. Nieradzinska, "Cyber-Physical-Security Model for Safety-Critical IoT Infrastructures," figshare, 2016.

[6] I. Ahmed, V. Roussev, W. Johnson, S. Senthivel, and S. Sudhakaran, "A SCADA System Testbed for Cybersecurity and Forensic Research and Pedagogy," Proceedings of the 2nd Annual Industrial Control System Security Workshop on - ICSS '16. ACM Press, 2016.

[7] E.T. Jensen, "Computer attacks on critical national infrastructure: A use of force invoking the right of self-defense," Stan. J. Int'l L. 38, 207, 2002.

[8] E.E. Tan, "Cyber deterrence in Singapore: Framework & recommendations," 2018.

[9] B. B. Gupta, D. P. Agrawal, S. Yamaguchi, N. A. G. Arachchilage, and S. Veluru, "Editorial security, privacy, and forensics in the critical infrastructure: advances and future directions," Annals of Telecommunications, vol. 72, no. 9–10. Springer Science and Business Media LLC, pp. 513–515, 14-Sep-2017.

[10] I. Kotenko, I. Saenko, O. Lauta, and A. Kribel, "An Approach to Detecting Cyber Attacks against Smart Power Grids Based on the Analysis of Network Traffic Self-Similarity," Energies, vol. 13, no. 19. MDPI AG, p. 5031, 24-Sep-2020.

[11] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," ACM SIGCOMM Computer Communication Review, vol. 23, no. 4. Association for Computing Machinery (ACM), pp. 183–193, Oct-1993.

[12] B. Vamanu and M. Masera, "Understanding Malicious Attacks Against Infrastructures - Overview on the Assessment and Management of Threats and Attacks to Industrial Control Systems". EUR 23681 EN. Luxembourg (Luxembourg): OPOCE; 2008.

[13] J. Slay and M. Miller, "Lessons Learned from the Maroochy Water Breach," IFIP International Federation for Information Processing. Springer US, pp. 73–82.

[14] J. P. Conti, "The day the samba stopped [power blackouts]," Engineering & Technology, vol. 5, no. 4. Institution of Engineering and Technology (IET), pp. 46–47, 06-Mar-2010.

[15] S. Kuvshinkova, "SQL Slammer worm lessons learned for consideration by the electricity sector," North American Electric Reliability Council, 2003.

[16] J. P. Farwell and R. Rohozinski, "Stuxnet and the Future of Cyber War," Survival, vol. 53, no. 1. Informa UK Limited, pp. 23–40, 2011.

[17] G. Richards, "Hackers vs slackers," Engineering & technology, vol. 3, no. 19, pp. 40–43, 2008

[18] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2. Institute of Electrical and Electronics Engineers (IEEE), pp. 1153–1176, 2016.

[19] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson, "Shallow and deep networks intrusion detection system: A taxonomy and survey." arXiv preprint arXiv:1701.02145, pp. 1–43, 2017.

[20] P. Ravi Kiran Varma, V. Valli Kumari, and S. Srinivas Kumar, "A Survey of Feature Selection Techniques in Intrusion Detection System: A Soft Computing Perspective," Advances in Intelligent Systems and Computing. Springer Singapore, pp. 785–793, 2018.

[21] H. Debar, M. Dacier, and A. Wespi, "Towards a taxonomy of intrusion detection systems." pp. 805–822, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128698000176

[22] S. H. Amer and J. Hamilton, "Intrusion detection systems (IDS) taxonomy - a short review." Defense Cyber Security, vol. 13, no. 2, pp. 23–30, 2010.

[23] T. Hamed, J. B. Ernst, and S. C. Kremer, "A Survey and Taxonomy of Classifiers of Intrusion Detection Systems," Computer and

Network Security Essentials. Springer International Publishing, pp. 21–39, 13-Aug-2017.

[24] I. Amit, J. Matherly, W. Hewlett, Z. Xu, Y. Meshi, and Y. Weinberger, "Machine learning in cyber-security - problems, challenges and data sets." arXiv preprint arXiv:1812.07858, pp. 1–8, 2018.

[25] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks." IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 266–282, 2014.

[26] C. Zhou, S. Huang, N. Xiong, S.-H. Yang, H. Li, Y. Qin, and X. Li, "Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation." IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 10, pp. 1345–1360, 2015.

[27] S. M. Ghaffarian and H. R. Shahriari, "Software vulnerability analysis and discovery using machine-learning and data-mining techniques: A survey." ACM Computing Surveys (CSUR), vol. 50, no. 4, p. 56, 2017.

[28] P. M. Laso, D. Brosset, and J. Puentes, "Dataset of anomalies and malicious acts in a cyber-physical subsystem," Data in Brief, vol. 14. Elsevier BV, pp. 186–191, Oct-2017.

[29] D. K. Saini, "A mathematical model for the effect of malicious object on computer network immune system," Applied Mathematical Modelling, vol. 35, no. 8. Elsevier BV, pp. 3777–3787, Aug-2011.

[30] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," Physica D: Nonlinear Phenomena, vol. 31, no. 2. Elsevier BV, pp. 277–283, Jun-1988.

[31] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham, and M. A. Zissman, "Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation," Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00. IEEE Comput. Soc.

[32] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[33] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS). IEEE, Nov-2015.

Marwan Albahar received his B.S. in computer science from King Faisal University in 2011, Saudi Arabia, and his M.Sc. in computer science with honor in 2015 from Frostburg State University, USA. Dr. Albahar received 2018 his Ph.D. from the University of Eastern Finland. Dr. Albahar a senior Information Security, Privacy, and Risk Management Professional with a solid technical background and a highly analytical mind. He has been involved within the information security field for the last 3+. His main areas of research Computer Networks & Security, Cybersecurity, Artificial intelligence.