# THE USE OF FRACTAL DIMENSION (FD) ANALYSIS IN DETECTION OF ANOMALIES, SABOTAGES, AND MALICIOUS ACTS IN A CYBER-PHYSICAL SYSTEM USING HIGUCHI'S ALGORITHM

*Marwan Albahar (1)*, Mohammed Thanoon1 (1), Abdulaziz Albahr Author (2, 3)*

[1] Department of Science, Umm Al Qura University, P.O. Box 715, Mecca,
[2] College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences, Al-Ahsa 31982,
[3] King Abdullah International Medical Research Center, Al-Ahsa, Saudi Arabia

* Corresponding Author, e-mail: mabahar@uqu.edu.sa

**Abstract:** In recent years, cyber-physical systems (CPSs) have received a lot of attention because of their wide range of applications. However, because CPSs rely heavily on communication networks, they are vulnerable to deliberate cyberattacks. As a result, there are a wide variety of attack detection methods that have been proposed to protect CPSs. This research examines the behavior of malicious acts and anomalies within the cyber physical system using Higuchi Fractal Dimension (HFD). The model was constructed and evaluated using the HFD algorithm. The model categorizes various types of anomalous events, such as sabotage, cyber-attacks, and hardware failures. Furthermore, our proposed model notifies the operator by calculating the probability of an anomaly occurring, which aids in the mitigation process. A real-world dataset is used to train and test the model.

**Key words:** Cyber-attacks, malicious attacks, nonlinear mathematical method, Fractal Dimension, Receiver Operating Characteristics.

## 1. INTRODUCTION

A new class of emerging technologies known as Cyber Physical Systems (CPS) has emerged in the last couple of years. A cyber-physical system is a combination of physical and cyber systems that are segregated and interact. Simply put, a CPS is a network in which several nonphysical and physical systems are controlled by an embedded system that typically makes decisions based on input from other physical systems, most frequently a sensor network. These choices are made using a plethora of data gathered from the environment and its processes. To comprehend CPS, it is

necessary to have a thorough understanding of physical systems, communication systems, physical processes, control systems, embedded systems, software, and networking.

A CPS is a multidimensional, complex system that combines computing technologies, communication, and control. Due to their critical function within the system, CPSs require a high level of security and robustness. This ensures their continued operation is reliable. Because of its role in ensuring system-wide anomaly detection and security, CPS is likely to remain a crucial component. CPS data is more likely to show implicit correlations between data points because of the nature of CPS systems, which is necessary for leveraging CPS security provisions in more complicated data environments [1].

Examples of CPS applications include unarmed aerial vehicles (UAVs), educational systems, medical systems, smart grids, self-driving automobiles, power systems, and industrial applications [1-5]. Each of these examples is representative of a sensor-based, real-time, self-contained, communication-enabled autonomous system. CPS and the Internet of Things (IOT) share some characteristics [6]. CPS, on the other hand, is distinct from the IOT. In a CPS, cyber and physical systems are in constant communication with one another. CPS has a higher concentration of internal communication and control elements, whereas IOT is the network of numerous physical objects. IoT devices have a broader range of external communication capabilities. Cyber-security solutions are frequently unsuitable for infrastructure and fail to consider the underlying risks posed by compromised sensor data [7]. Recent years have seen a significant increase in the number of attacks on critical infrastructure. Stuxnet is a well-known SCADA attack that was discovered in a 2010 attack on an Iranian power plant [8]. Another example is the attack on the Maroochy sewer system, which led to the release of 800,000 litters of sewer waste into waterways and parks [9].

Techniques for fractal analysis are widely used in a variety of fields, including biology, chemistry, and physics. Fractal theory's crowning achievement has been the development of simple mathematical descriptions of extremely complex but ubiquitous natural phenomena and objects. Just as differential trigonometry, harmonic analysis, and equations are fundamental mathematical tools for modelling and explaining physical reality, fractal analysis is a fundamental mathematical tool for modelling and explaining physical reality [10]. Numerous researchers have investigated the possibility of their structures being self-similar (i.e., their fractality), which is a remarkable example of a non-local property [11]. Fractal theory has been used to study many artificial and natural systems that have similar patterns at different scales. Understanding fractals is critical for decoding complex systems, as it enables the behaviour of large systems to be revealed by focusing on a small subset. Fractal analysis is unique in that it identifies self-similarity, which enables us to attribute time series to a predefined model, to reveal the features of the local structure, and to identify various properties that are invisible in the normal representation, including real time. Given that self-similarity is observed over a

broad time range, the presence of long-term attacks and abnormal activity in the signal alters the self-similar nature of traffic, which can be used as an additional informative feature in machine learning.

Based on the above discussion, we demonstrate the breadth of HFD applications for analysing the complexity of malicious traffic and anomalous behaviour within a cyber-physical system. Fractal properties on a large scale of network traffic exhibit self-similarity, which means they appear qualitatively identical at sufficiently large time scales and exhibit long-term dependence [11]. Due to the stochastic and complex nature of computer network traffic data, linear analysis techniques are limited in their ability to detect malicious attacks on data networks. The method begins by calculating the fractal dimension for each segment of the signal segmented by a sliding window. The second stage detects the transient starting point by experimentally defining a threshold. The threshold value in this case is equal to the fractal dimension's mean. The Higuchi algorithm is then used to determine the fractal dimension of the time series data from the data network. The computed fractal dimension is a nonlinear metric that will be used to classify malicious and benign data in the future. Thus, the field of classifying malicious traffic and anomalous behaviour using fractal dimensions could help in terms of reducing false positives. This article makes the following contribution:

1. Based on an analysis of the fractal properties of traffic, we propose a novel approach for investigating the complexity of malicious traffic and detecting anomalous behaviour within a cyber-physical system.

2. We thoroughly evaluate our method's performance using 14 operational scenarios based on the data extracted from the dataset.

## 2. RELATED WORK

There have been many IDSs developed and validated by using publicly available data. Many reviews and comparative studies have been conducted on intrusion detection system (IDS) design for various applications, as well as on the machine learning techniques that are used to construct IDS. The challenges associated with the dataset, on the other hand, are not mentioned. Therefore, most of these studies are focused on a single aspect of IDS evaluation rather than on the system as a whole [12]. In [13-15], fractal analysis was used to solve a wide range of problems related to information security. DARPA99 network traffic data was examined for self-similarity, and it was discovered that abnormal packet traffic over a specified time period can significantly alter the self-similarity function in the traffic process [15]. The authors in [16] discussed the findings of research into the detection of multifractal dimension jumps caused by anomalous changes in the properties of telecommunication traffic on a real-time (current) time scale, which was conducted in the context of detecting anomalous changes in the properties of telecommunication traffic. A variance fractal dimension feature selection method was proposed by the authors of [17] for analysing the significant characteristics of a

cyber security attack dataset. A complexity analysis was carried out in order to determine the cognitive discriminative features of the UNSW-NB15 dataset. The authors also discovered, through a comparative analysis, that their proposed method improved the detection performance of network systems. The issue of reducing data dimensions for machine learning algorithms was addressed by the author of [18]. A smaller set of discriminative features from a cyber-attack dataset is selected using a variance fractal-based complexity analysis. In addition, the authors in [19] were able to significantly reduce the number of false positive and negative results. By decreasing the number of false negatives and positives and increasing the overall classification rate, the proposed fractal-based method outperforms the traditional Euclidean-based machine learning algorithm (k-NN).

## 3. DATASET ANALYSIS

We conducted our analysis in our study using a publicly available dataset [20]. The dataset consists of actuator and sensor readings that were taken at periodic intervals of 0.1 seconds by the PLC. Data from PLC registers 2–4 provided output data describing the system's state, which was used to conduct the data analysis. The registers that the PLC utilizes are listed in Table 1, along with a brief description of their functions. The binary state of the discrete sensors is represented by the bits in register 2. The register can be used to perform a population count in order to independently retrieve the state of each sensor. Both the pump's state and the ultrasound sensor's step value are stored in register 3 and register 4, ranging from 0 to 10,000.

*Table 1. Bit's representation of extracted registers*

| Reg. No. | Bit No. | Value |
|---|---|---|
| 2 | 4 | Discrete Sensor 3 |
| | 5 | Discrete Sensor 2 |
| | 6 | Discrete Sensor 1 |
| | 7 | Discrete Sensor 0 |
| 4 | 16-bits | Depth Sensor |

| Reg. No. | Bit No. | Value |
|---|---|---|
| 3 | 0 | Pump 2 |
| | 1 | Pump 1 |
| | 5 | Pump 2 Valve |
| | 4 | Pump 1 Valve |

Table 2 depicts the dataset's fourteen distinct scenarios. It includes a breakdown scenario, a cyber-attack scenario, an accident scenario, and six other operational scenarios indicative of a potential threat. Any parts of the system that are adversely affected by the anomaly are referred to as "affected components." Depending on the type of incident, the recorded data is split up into fifteen separate CSV files with varying durations (i.e., a DoS incident may take more time than sabotage incident).

Fig. 1 depicted the data collection process; two containers of varying sizes were used. In addition to the four discrete sensors, one ultrasound depth measurement sensor and two pumps were used to collect data. The system was controlled by a computer connected to a monitoring network system via a PLC connector. When the main container is filled with liquid, the ultrasound sensor indicates when it is emptied. Pump 2 initiates the filling of the second volume container, while Pump 1

initiates the filling of the main tank. Tank 2 shows the current state of the four discrete sensors (sensor 0, sensor 1, sensor 2, and sensor 3).

*Table 2. Affected components of dataset and operational scenarios* [20]

| | Scenario | Affected Component | Operational Scenario | No. of instances |
|---|---|---|---|---|
| 1 | Normal | None | Normal | 5519 |
| 2 | Plastic bag | Ultrasound Sensor | Accident/Sabotage | 10549 |
| 3 | Blocked measure 1 | | Breakdown/Sabotage | 226 |
| 4 | Blocked measure 2 | | | 144 |
| 5 | Floating objects in main tank (2 objects) | | Accident/Sabotage | 854 |
| 6 | Floating objects in main tank (7 objects) | | | 733 |
| 7 | Humidity | | Breakdown | 157 |
| 8 | Discrete sensor failure | Discrete sensor 1 | | 1920 |
| 9 | Discrete sensor failure | Discrete sensor 2 | | 5701 |
| 10 | Denial of service attack | Network | Cyber-attack | 307 |
| 11 | Spoofing | | | 10130 |
| 12 | Wrong connection | | Breakdown/Sabotage | 6228 |
| 13 | Person hitting the tanks (low intensity) | Whole subsystem | Sabotage | 347 |
| 14 | Person hitting the tanks (medium intensity) | | | 281 |
| 15 | Person hitting the tanks (high intensity) | | | 292 |

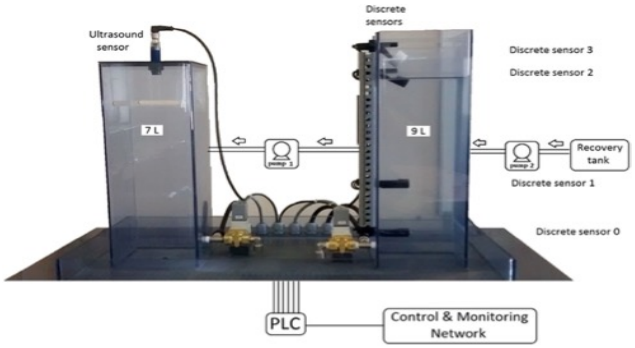


*Fig. 1. The platform of the cyber-physical system which was used for data collection* [20]

## 4. EXPERIMENTS AND RESULTS

The fractal analysis method, which was based on Higuchi's algorithm [21], was applied to the time series data for each type of simulated scenario, and the time series data was continuously recorded via the PLC registers during the experiment. Figure 2 depicts the entire process of the various stages that must be completed in order to distinguish malicious data from normal data or non-malicious data in order to classify it as such.

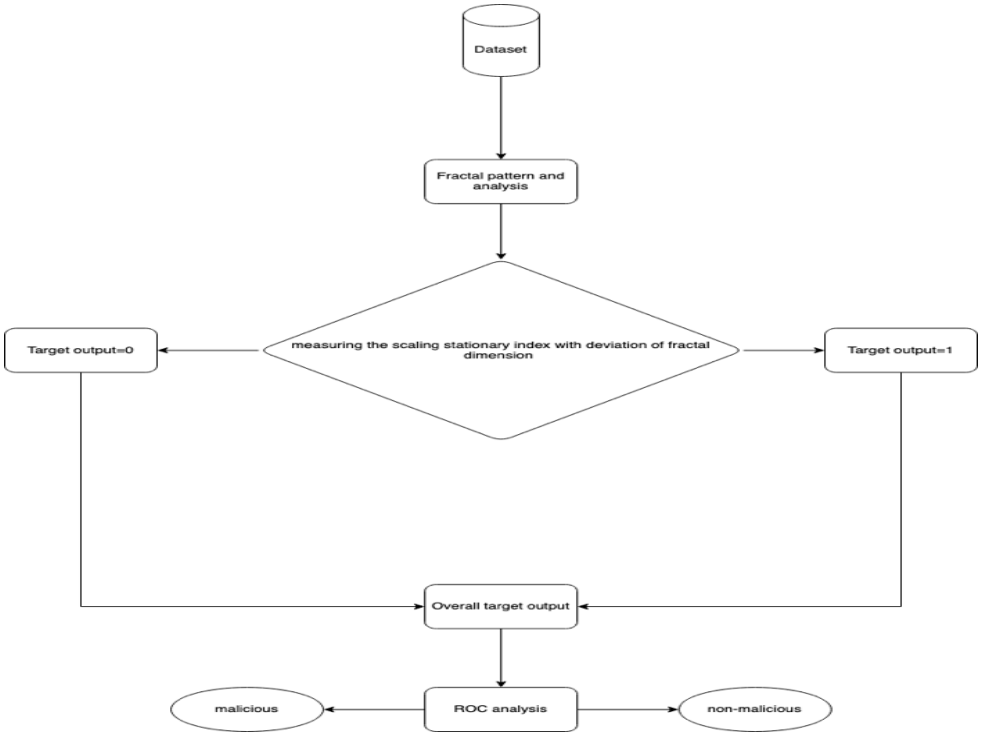


*Fig. 2. Stages of the entire experimental process*

There were two rounds of experiments carried out. In the first run, the calculation of fractal division from normal data was computed along with the fractal demission of the infected component to identify an anomaly that occurred along with the associated scenario that has been specified. The goal is to provide the fractal dimension values to the CI operator in a timely manner. As part of this process, a second run was also set up. This second run was used to check the reliability of the Higuchi algorithm to compute the fractal dimension of time series signals.

*Table 3. First experiment: the application of 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, $\mathbf{L}(\boldsymbol{k}) \sim k^{-1.8440}$.

As shown in Table 3, each simulated scenario generated a unique fractal dimension value, and the change in fractal dimension value was included as a third column to illustrate how the characteristic behaviour of a particular anomaly changes in comparison to the complexity of normal network data. Additionally, we observed that network data that has been infected with malicious data (Distributed Denial of Service attacks) has the highest fractal dimension value. Additionally, the magnitude of the difference between the malicious data's fractal dimension value (1.8440) and the normal data's fractal dimension value (1.7436) is positive (See Figure 3). This is the most significant and positive difference in comparison to the differences between malicious and non-malicious data, which are much smaller. As a result, a scaling stationary index was included to ensure that the ROC computation detects all fractal values greater than a threshold value (determined by trial and error during computation) in order to achieve a very high degree of accuracy in distinguishing malicious and benign data. In this way, our proposed model is more flexible. It can be changed to distinguish malicious data from normal traffic data based on their fractal dimension complexity differences, with the network data time series fractal dimension as a reference point for the model.

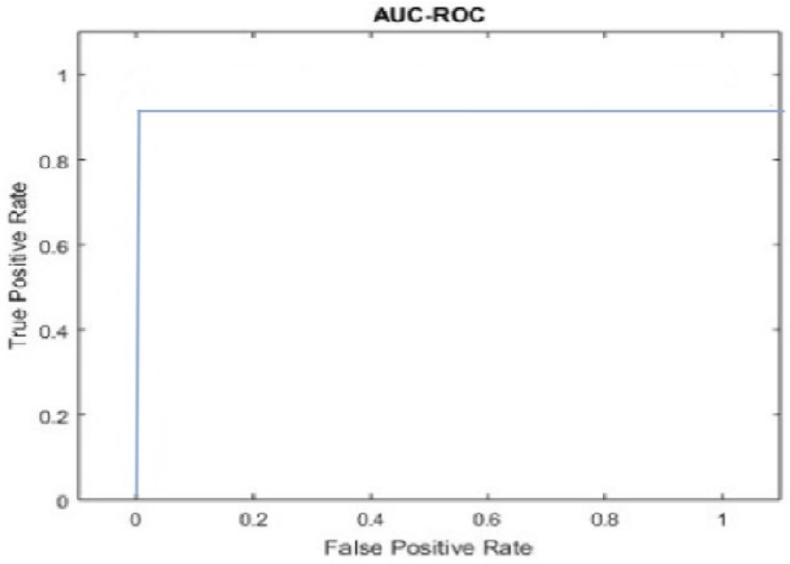


*Fig 3. ROC curve analysis of the first run of computed fractal dimension values and the targeted outcomes after the inclusion of the scaling stationary index.*

Figure 3 depicts the receiver operating characteristic (ROC) curve between the calculated fractal dimension of various types of network data and the corresponding targeted output. In this case, the area under the curve (AUC) was 0.95, and the accuracy was 96%. Accuracy was maximized through trial and error, with the threshold value being adjusted to achieve the highest possible accuracy, sensitivity, and specificity results. The optimal threshold value for maximum accuracy was

found to be 1.69, which was determined through mathematical analysis. The parameters of the ROC curve depicted in Figure 3 are summarized in the following Table 4.

*Table 4. Computed ROC curve parameters*

| Parameter names | Computed values |
|---|---|
| AROC | 0.95 |
| Specificity | 0.92 |
| Sensitivity | 1.0 |
| Accuracy | 96% |
| scaling stationary index | 1.69 |

The second experiment aims to provide capital information to the operator regarding the anomaly. In this case, the alert includes the affected component. The model classifies which components are impacted by the CI's five components, as well as the "None" affected case. Figure 4 illustrates our model's classification results, with the highest accuracy reaching 80%. The result demonstrates a trade-off between the first and second runs by providing more detailed information to the operator. As demonstrated by the trade-off limits, the system's accuracy accounts for a significant number of false positives, which could confuse the operator during normal operations.

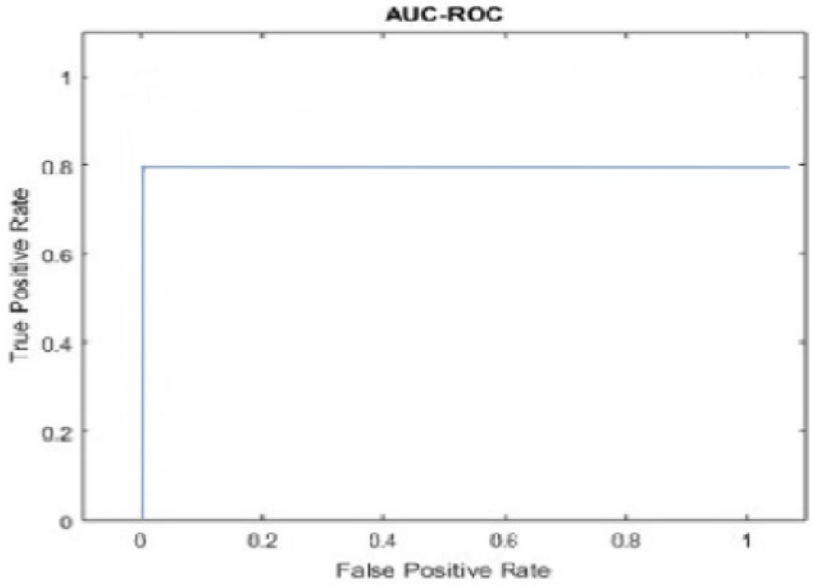


*Fig. 4. ROC curve analysis of the second run of computed fractal dimension values and the targeted outcomes after the inclusion of the scaling stationary index*

## 5. DISCUSSIONS

Fractal dimension is a nonlinear algorithm for determining the complexity of a time series of data signals that is extremely effective. By quantifying the non-stationary network dataset, it is possible to generate a unique value for each of the scenarios under consideration. This technique was employed in order to distinguish malicious data from legitimate information. In conjunction with the receiver operating characteristics, a precision of one hundred percent can be obtained when the implemented fractal dimension is used. The proposed algorithm performed admirably and produced results that were satisfactory. This experiment has yielded perfect results in every metric, including the TPR, accuracy, and AUC-ROC. According to the findings of this study, malicious data (particularly network data that has been subjected to a distributed denial of service attack) and non-malicious data were classified with the highest degree of accuracy possible. A network immune system can be monitored in a network system using the fractal dimension analysis method in conjunction with a classification technique such as ROC. It is important to note that this research does not distinguish between legitimate and malicious data, such as "Hits on Tank." This is a significant limitation of the research.

## 6. CONCLUSION

This research investigates the complexity of malicious acts and anomalies within the cyber physical subsystem using a non-linear mathematical method called Higuchi Fractal Dimension (HFD). The HFD algorithm was successfully tested and validated on synthetic time series network data and real-time network data, resulting in the production of accurate values. It was discovered that DoS-attacked time series data had the highest fractal dimension value. Additionally, the difference in HFD values between the DoS attack data and the normal traffic data was the most significant. ROC method was successfully used to classify malicious and benign network data. Using ROC method, a scaling stationary index was used to aid in the classification of both normal network data and malicious data. As a result, fractal dimension has established itself as a critical component of cyber-attack tracking.

## REFERENCES

[1] Krishna C., I. Koren. Thermal-aware management techniques for cyber-physical systems. *Sustainable Computing: Informatics and Systems*, 2017.

[2] Yetis, H., M. Baygin, M. Karakose. An investigation for benefits of cyber-physical systems in higher education courses. *15th IEEE International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2016, pp. 1–5.

[3] Pawlick J., Q. Zhu. Strategic trust in cloud-enabled cyber-physical systems with an application to glucose control. *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, 2017, pp. 2906–2919.

[4] Trappey, A.J.C., C.V. Trappey, U.H. Govindarajan, J.J. Sun, A.C. Chuang. A review of technology standards and patent portfolios for enabling cyber-physical systems in advanced manufacturing. *IEEE Access*, vol. 4, 2016, pp. 7356–7382.

[5] J¨arvinen, T., G.S. Lorite, A.-R. Rautio, K.L. Juh´asz,´A.Kukovecz, Z.K´onya, K.Kordas, G. Toth. Portable cyber-physical system for indoor and outdoor gas sensing. *Sensors and Actuators B: Chemical*, 2017.

[6] Wan, J., H. Yan, H. Suo, F. Li. Advances in cyber-physical systems research. *TIIS*, vol. 5, no. 11, 2011, pp. 1891–1908.

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

[8] Langner, R., Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy Magazine*, vol. 9, no. 3, May 2011, pp. 49–51.

[9] Brenner, J.F. Eyes wide shut: The growing threat of cyber-attacks on industrial control systems. *Bulletin of the Atomic Scientists*, vol. 69, no. 5. 2013, pp. 15–20.

[10] Kotenko, I., I. Saenko, O. Lauta, 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. Sep 2020, p. 5031.

[11] Leland, W.E., M. S. Taqqu, W. Willinger, D. V. Wilson. On the self-similar nature of Ethernet traffic/. *ACM SIGCOMM Computer Communication Review*, vol. 23, no. 4, Oct 1993, pp. 183–193.

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

[13] Lippmann P., K. Cunningham. Improving intrusion detection performance using keyword selection and neural networks. *Computer Networks*, vol. 34, no. 4, 2000, pp. 597-603.

[14] Sheluhin, O., I. Lukin. Network traffic anomalies detection using fixing method of jumps of multifractal dimension in the real-time mode. *Automation Control Computer Science*, vol. 52, no. 5, 2018, pp. 421–430. https://doi.org/10.3103/S0146411618050115.

[15] Sheng, Z., Z. Qifei, P. Xuezeng, Z. Xuhui. Detection of low-rate DDoS attack based on self-similarity. *2010 Second International Workshop on Education Technology and Computer Science*, vol. 1, 2010, pp. 333–336.

[16] Sheluhin, O. *Multifractals. Information Applications*. Moscow: Hotline–Telecom, 2011.

[17] Kaiser, S., K. Ferens. Variance fractal dimension feature selection for detection of cyber security attacks. *Transactions on Computational Science and Computational Intelligence*. Springer International Publishing, 2021, pp. 1029-1045.

[18] Kaiser, S. *Cognitive Discriminative Feature Selection Using Variance Fractal Dimension for the Detection of Cyber Attacks*. University of Manitoba, 2020.

[19] Siddiqui, S., M. S. Khan, K. Ferens, W. Kinsner. Detecting advanced persistent threats using fractal dimension based machine learning classification. *Proceedings of the 2016 ACM on International Workshop on Security and Privacy Analytics*, 11th March 2016.

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

[21] Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, Jun 1998, pp. 277–283.

### *Information about the authors:*

**Dr. Marwan Ali 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 chief information security officer, 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 4+. His main areas of research Computer Networks & Security, Cybersecurity, Artificial intelligence.

**Dr. Mohammed Thanoon** has taught several subjects in computer science and computer engineering at Umm Al-Qura University and Tennessee State University. His academic research interests involve the areas of human and machine teaming, data fusion, decision-making, intelligent control systems, artificial intelligence, machine learning, deep learning, federated learning, edge computing, medical image processing, computer vision, robotics, and IoT. He is certified as a "MathWorks Certified MATLAB Associate."

**Dr. Abdulaziz Albahr** is currently an assistant professor at King Saud Bin Abdulaziz University for Health Science. He earned his Ph.D. in computer science from Southern Illinois University Carbondale in 2019. His research interests include natural language processing (NLP), security, and artificial intelligence in medical imaging.

**Manuscript received on 28 February 2022**