Intrusion Detection in Binary Process Data: Introducing the *Hamming*-distance to *Matrix Profiles*

1st Simon D Duque Anton
Intelligent Networks Research Group
German Research Center for AI
Kaiserslautern, Germany
simon.duque_anton@dfki.de

2nd Hans D Schotten
Intelligent Networks Research Group
German Research Center for AI
Kaiserslautern, Germany
hans_dieter.schotten@dfki.de

***Abstract*—The digitisation of industry provides a plethora of novel applications that increase flexibility and reduce setup and maintenance time as well as cost. Furthermore, novel use cases are created by the digitisation of industry, commonly known as *Industry 4.0* or the *Industrial Internet of Things*, applications make use of communication and computation technology that is becoming available. This enables novel business use cases, such as the digital twin, customer individual production, and data market places. However, the inter-connectivity such use cases rely on also significantly increases the attack surface of industrial enterprises. Sabotage and espionage are aimed at data, which is becoming the most crucial asset of an enterprise. Since the requirements on security solutions in industrial networks are inherently different from office networks, novel approaches for intrusion detection need to be developed. In this work, process data of a real water treatment process that contains attacks is analysed. Analysis is performed by an extension of *Matrix Profiles*, a motif discovery algorithm for time series. By extending *Matrix Profiles* with a *Hamming*-distance metric, binary and tertiary actuators can be integrated into the analysis in a meaningful fashion. This algorithm requires low training effort while providing accurate results. Furthermore, it can be employed in a real-time fashion. Selected actuators in the data set are analysed to highlight the applicability of the extended *Matrix Profiles*.**

***Index Terms*—Intrusion Detection, Industrial Networks, Time Series Analysis, Anomaly Detection, Data Mining**

I. Introduction

The introduction of *Industry 4.0*, also referred to as the Industrial Internet of Things (IIoT), enables an abundance of novel use cases [1], [2]. These use cases in turn introduce new business cases, meaning industrial value generation is changed. Customer-individual processing with minimal delay and digital twins are examples of the scenarios that can be implemented because of digitisation in industry. However, since these novel use cases rely heavily on intercommunication and computation, the network structures of industrial enterprises, the Operation Technology (OT) networks, are changing. When Supervisory Control And Data Acquisition (SCADA) systems were first introduced in the 1970's, they were meant to control industrial devices in a pre-defined, non-flexible fashion. Furthermore, the networks were physically separated from public networks and highly application specific [3]. These features limited the surface for an attacker. Commercial Off The Shelf (COTS) products for industrial application and a focus on interconnectivity drastically increase the attack surface, making a focus on intrusion detection necessary [4]. Since the requirements for industrial Intrusion Detection Systems (IDSs) are different than requirements for office Information Technology (IT) environments, novel approaches have to be developed. *Powers et al.* [5] as well as *Iturbe et al.* [6] discuss the different operational conditions of IT and OT environments. Generally, OT networks are operated for longer periods, i.e. decades, constantly with little possibility to update or change systems. IDSs must not affect the process, since availability is the highest rated requirement.

The contribution of this work consists of:

- Integration of the *Hamming*-distance [7] into the *Matrix Profile*-algorithm [8], and
- detection of attacks in a real process environment by analysing actuator information.

The remainder of this work is structured as follows. Section II presents related work to industrial intrusion detection. The data set used to evaluate the usefulness of the extended *Matrix Profiles* is discussed in Section III, the standard and extended *Matrix Profiles* are introduced in Section IV. Section V evaluates the performance. A discussion of the algorithms and results is presented in Section VI. This work is concluded in

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Foerderkennzeichen 16KIS0932, IUNO Insec). The authors alone are responsible for the content of the paper.

978-1-7281-7374-0/20/$31.00 ©2020 IEEE
DOI 10.1109/WoWMoM49955.2020.00065

Section VII.

II. Related Work

This section presents scientific works related to intrusion detection in industrial environments, with a focus on process data analysis. As discussed in Section I, the relevance of this topic increases. Hence, it is widely addressed in research. *Schneider and Böttinger* use *autoencoders* for detecting human-based attacks on an industrial environment [9], provided by the *iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design* [10]. The *autoencoders* are capable of detecting the attacks in the data set called *SUTD Security Showdown (S3) 2017* (*S317*) in an unsupervised fashion. Another data set provided by this institute is analysed by *Goh et al.* [11]. Recurrent Neural Networks (RNNs) are employed to detect the attacks introduced into the industrial environment. The same data set is evaluated in this work. Furthermore, this data set has been analysed by means of *Matrix Profiles* already, with a focus on the sensor data [12], [13]. A counter has been introduced that was capable of not only detecting outliers, but also detecting similar attacks that occur rarely in comparison to motifs occurring often [14]. Generally, the *Secure Water Treatment* (*SWaT*) data set has been widely regarded in research. *Inoue et al.* analyse it with Deep Neural Networks (DNNs) as well as *ocsvm* [15]. Similarly, *Kravchik and Shabtai* employ Convolutional Neural Networks (CNNs) [16]. *Li et al.* analyse the data set with Generative Adversarial Networks (GANs) [17], [18]. A method with code mutation is presented by *Chen et al.* [19]. *Lin et al.* develop a graphical model to detect the attacks [20]. Long Short-Term Memory (LSTM), a type of neural network, is applied in a multi-level approach to detect attacks in a gas pipeline data set by *Feng et al.* [21]. Means to increase the security of industrial networks are discussed by *Knapp and Langill* [22]. However, a significant disadvantage of neural networks is the immense training data and effort required to build expressive models. *Yang et al.* introduce detection methods for attacks in power system networks [23]. A publication summarising the stages and development of SCADA security systems is presented by *Larkin et al.* [24]. One task difficult to achieve is detecting attacks that are formerly unknown to the operators. Learning a model of the normal system state and detecting deviations is a strength of machine learning algorithms, such as *One Class Support Vector Machine*s (*OCSVM*s). *Maglaras and Jiang* present their application to detecting novel attacks [25]. Industrial communication protocols often do not provide means for authentication and encryption [26]. Due to their long operation times, they are still in use, resulting in the need for additional security measures. *Gao and Morris*present an approach for signature-based intrusion detection in *Modbus*-networks [27].

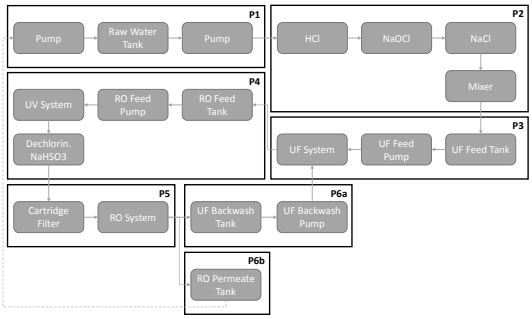

Figure 1. Relation of Sub-Processes

III. Presenting the Data Set

The data set analysed in this work is created by the *iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design*. It is named *SWaT* [28], [29]. Creating data sets for intrusion detection is a crucial, yet non-trivial task [30]. It has been previously addressed by research, e.g. by *Schneider and Böttinger* [9]. The *SWaT* data set was captured in a water processing facility consisting of real equipment. A process of water treatment was run for eleven days. During the first seven days, only normal operation occured. During the last four days, attacks were introduced. The attacker was assumed to already have broken the perimeter, thus directly impacting the Programmable Logic Controllers (PLCs). Overall, one of six PLCs was used to control a respective sub-process. The sub-processes are as follows:

- *P1*: Raw water storage
- *P2*: Pre-treatment
- *P3*: Membrane Ultra Filtration (UF)
- *P4*: Dechlorination by Ultraviolet (UV) lamps
- *P5*: Reverse Osmosis (RO)
- *P6*: Disposal

The relation of the sub-processes is shown in Figure 1. Each PLC controls a ring network, while the PLCs are controlled by a SCADA workstation. Data is collected by a data historian. The network structure is depicted in Figure 2. The raw water is contained in an initial tank and pre-processed. In a second step, filtration as well as UV light and RO treatment are applied. Then the water is stored in another tank, given the treatment has resulted in sufficiently clean water. Otherwise, the UV light and RO treatment are repeated.

During the four last days of operation, 41 different instances of attacks are introduced into the process. The attacks are changes in process variables with the malicious intent to disrupt the operation. Ground truth, i.e. a labelling known to be correct, is provided. Each

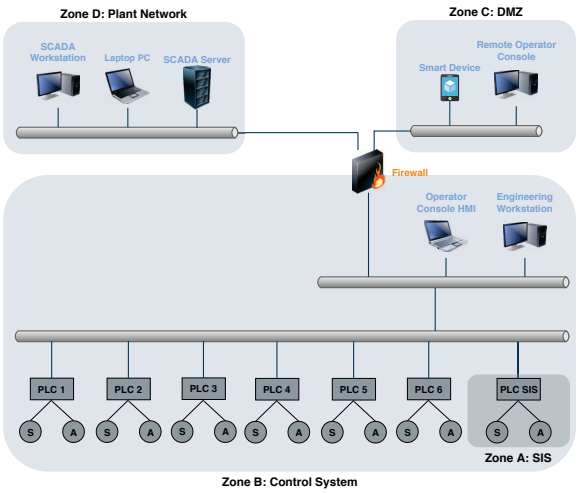

Figure 2. Schematic Overview of the Process Environment

of the 41 attacks falls into one of four categories, as introduced by *Goh et al.* [28]:

- *Single Stage Single Point (SSSP)*: Single stage attack on one point in the process, 26 instances in the data set
- *Single Stage Multi Point (SSMP)*: Single stage attack on multiple points in the process, 4 instances in the data set
- *Multi Stage Single Point (MSSP)*: Multi stage attack on one point in the process, 2 instances in the data set
- *Multi Stage Multi Point (MSMP)*: Multi stage attack on multiple points in the process, 4 instances in the data set

Some of the attacks, 18 in total, could not be observed to have an influence on the process. 51 sensors and actuators provided information about the process that could be used to detect the attacks, a complete list is provided by *Goh et al.* [28]. The top five sensors and actuators with respect to most attacks aimed at them are listed in Table I. This table includes the number of attacks and the number of attacks that did not affect the process. Each attack had a point in the process on which

Table I
SOURCES OF THE ATTACKS

Elem	Sub-P	Description	Total	No Ch
P-102	P1	Pump (backup)	3	0
P-101	P1	Pump	2	0
MV-101	P1	Motor valve	2	0
P-302	P3	UF feed pump	2	0
P-203	P2	Dosing pump	2	0

it was started and a target. Similar to the starting points, the top five sensors and actuators that were targets of an attack are listed in Table II, together with the number of attacks and the number of attacks that did not influence them.

Table II
DETECTABLE POINTS OF ATTACKS

Elem	Sub-P	Description	Total	No Ch
LIT-101	P1	Raw water tank level	7	3
P-101	P1	Pump	2	0
LIT-301	P3	UF feed tank level	5	3
MV-303	P3	Motorised valve	2	0
LIT-401	P4	RO feed tank level	3	1

IV. INTRUSION DETECTION IN INDUSTRIAL DATA

This section presents the algorithm used to detect anomalies in industrial process data. A general assumption is that in an industrial environment, deviation from expected behaviour is either an attack, a user error or malfunction and thus worth noting and inspecting. The employed algorithm is not capable of distinguishing different kinds of anomalies. In the first subsection, the general algorithm as well as the application on continuous data are explained. An extension to employ the algorithm in a meaningful fashion on binary data as well is introduced in the second subsection.

A. Continuous Data

The *Matrix Profile* algorithm has previously been employed successfully to detect attacks in the data set presented in Section III [12]–[14]. However, the algorithm as presented by *Yeh et al.* relies on continuous data. *Matrix Profiles* were presented by *Yeh et al.* in 2016 as an algorithm for motif discovery [8]. It worked by analysing the sequences of length m starting at each point t_n of the time series and comparing it to each other sequence of length m. The sequences are analysed in a sliding window-fashion. A distance of the sequences is calculated, for example the z-normalised distance as described in (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{m} (\hat{x}_i - \hat{y}_i)^2} \qquad \hat{x}_i = \frac{x_i - \mu_x}{\sigma_x}, \quad \hat{y}_i = \frac{y_i - \mu_y}{\sigma_y} \tag{1}$$

After applying *Pearson's Correlation Coefficient* [31] (2)

$$corr(x,y) = \frac{E((x-\mu_x)(y-\mu_y))}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{m} x_i y_i - m\mu_x\mu_y}{m\sigma_x\sigma_y}, \tag{2}$$

where

$$\mu_x = \frac{\sum_{i=1}^{m} x_i}{m}, \quad \mu_y = \frac{\sum_{i=1}^{m} y_i}{m} \tag{3}$$

and

$$\sigma_x^2 = \frac{\sum_{i=1}^{m} x_i^2}{m} - \mu_x^2, \quad \sigma_y^2 = \frac{\sum_{i=1}^{m} y_i^2}{m} - \mu_y^2. \quad (4)$$

The Euclidean distance is derived as indicated in (5) [32],

$$d(x, y) = \sqrt{2m(1 - corr(x, y))} \quad (5)$$

the resulting distance metric is shown in (6).

$$d(x, y) = \sqrt{2m\left(1 - \frac{\sum_{i=1}^{m} x_i y_i - m\mu_x\mu_y}{m\sigma_x\sigma_y}\right)} \quad (6)$$

x and y are two distinct time series, μ is the respective mean and σ the respective standard deviation. The minimal distances are calculated and stored in a matrix fashion, which leads to the name *Matrix Profiles*. If the minimal distance is high in comparison to the minimal distances of other sequences, the corresponding sequence is an outlier or anomaly in the time series, since there is no similar sequence contained. On the other hand, small minimal distances indicate the presence of at least one similar sequence.

B. Binary Data

In contrast to sensors in industrial environments that typically produce continuous values, e.g. the temperature, pressure or flow volume, actuators often return a binary value. This value indicates the operation state of the actuators, commonly either active or inactive. The event space is binary, i.e. either on or off, which can be represented by boolean values of 0 and 1. If these values are considered as a time series and the *Matrix Profiles* are computed with the distance metrics provided by the authors, the calculation often breaks in praxis. If there are constant values, or means of 0, divisions by zero occur. This characteristic makes the current distance metrics unsuited for binary data. To bridge this gap, the *Hamming* distance [7] is proposed as an additional distance metric for calculation of the *Matrix Profiles*. The *Hamming* distance $D(x, y)$ is defined "as the number of coordinates for which x and y are different" [7]. Applied to the task of evaluating the distance of binary sensors, that means a sequence of length m is compared to any other sequence in the time series of the same length m in a sliding window fashion. The sequences are compared, the distance is derived by calculating the number of bits that are different between the two sequences for a given position. This approach results in a distance which can then be used to compute the *Matrix Profiles* as introduced in the previous subsection.

V. Evaluation

This section presents the application of the extended *Matrix Profiles* as presented in Section IV to the data set introduced in Section III. The application of *Matrix Profiles* to continuous sensor data obtained from the same data set has been successfully evaluated in related works [12]–[14].

In this work, a total of three time intervals was selected from the data set and evaluated for attacks. These intervals contain selected attacks on the top four sources for attacks, listed in Table I. The sequence length m was set to 2 000 or 500. Those values were used as preliminary evaluations provided promising results for these values. In contrast to machine learning-based intrusion detection, the *Matrix Profiles* algorithm does not require a training phase and training data as such. Instead, each sequence is compared to every other sequence. A general assumption about batch processing is the periodicity of behaviour, i.e. the process variables repeat themselves over and over again. In case of water treatment, a batch of water is introduced to the treatment process, treated in each stage, and removed from the environment. This process is then repeated with the next batch of water, which creates highly similar values in terms of process control. This leads to the assumption that events occurring repeatedly are intended, while events only happening once are malicious or non-intentional. Furthermore, one period of normal operation is sufficient to detect normal and anomalous behaviour. In the course of this work, several periods of normal operation were employed, since there might be small deviations.

With this in mind, each interval analysed in this work consists of a period without attacks, followed by two instances of attacks. For each interval, one or two sensors are analysed. In the first and third interval, two actuators are evaluated. In the second interval, one actuator was analysed. The first interval starts with the last 10 000 events of the normal data set, i.e. the seven day period during which no attacks were introduced. Appended are the first 7 203 events of the malicious data set, i.e. the four day period during which all attacks occurred. That means 10 000 events can be considered as training data, the next 7 203 events contain the attacks, but they contain normal operation as well. An overview of the interval is shown in Figure 3. The actuator output for the motorised valve *MV-101* controlling water flow to the raw water tank as well as for the backup pump *P-102* pumping water from the raw water tank to the next stage are shown in the top line with the dotted line indicating *MV-101* and the dashed line *P-102*. For some reason that is not explained by the authors of the data set, *MV-101* provides tertiary output, i.e. zero, one, and two. In the second line, the minimal distances are shown, again with the dotted line indicating *MV-101* and the dashed line *P-102*. Attacks are indicated in the bottom line. Overall, four attacks occur during the interval, however, the last two attacks are affecting neither *MV-101* nor *P-102* and are not expected to be discoverable by observing these

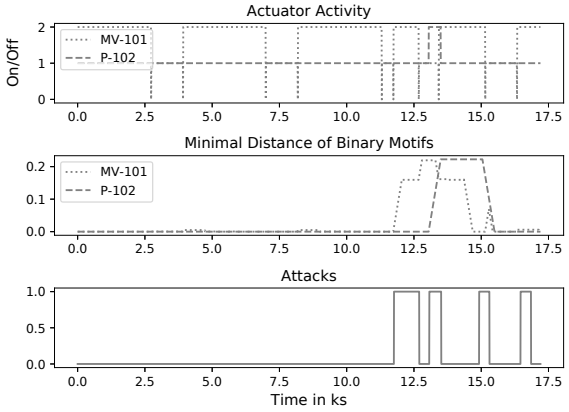

Figure 3. *Matrix Profile* of Interval 1 for *MV-101* and *P-102*

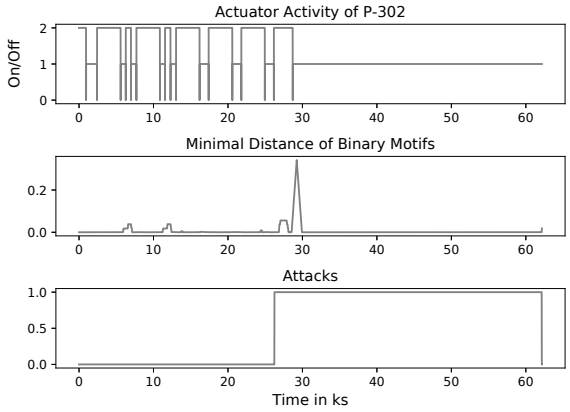

Figure 4. *Matrix Profile* of Interval 2 for *P-302*

actuators. Still, the third attack aims at the raw water tank level so that it implicitly also affects the valve *MV-101*, leading to a smaller raise in minimal distance around second 15 000. Due to the inter-dependability of components, attacks can be detected by looking at devices that are not target of the given attack. The first attack unexpectedly opens the motorised valve *MV-101*, leading to an uncontrolled water flow into the raw water tank and potential overflow. This unexpected behaviour is clearly distinguishable in Figure 3, meaning this attack can be detected. The second attack is turning on the backup pump *P-102* increasing the pressure in the pipe from raw water tank to initial stage of treatment. This could lead to a pipe burst. Similar to the first attack, this attack can be detected by the change of minimal *Hamming*-distance of the *Matrix Profile* as the backup pump is expected to remain inactive unless a malfunction requires it to become active. Figure 3 shows that the minimal distance is continuously increasing with the start of the attack, as for each sequence after the attack, more data points are different from known sequences. Hence, defining a sensible threshold value allows for early discovery of the attack. In this case, a threshold around 0.1 would easily detect all attacks with no false positives.

Interval 2 is the period on December 31st from 18:00:00 to 23:16:00, covering 62 160 events during which two attacks occur. The actuator under observation is the pump *P-302* that pumps water from the UF feed tank into the RO feed tank. As a first attack, *P-302* is kept on despite of the RO feed tank having reached capacity. As a second attack, *P-302* is turned off to stop the flow of water. This behaviour is shown in Figure 4. Since the two attacks in this interval are occurring back to back, there is only one attack-peak visible. The second attack lasts for several hours, thus the length. The first attack leads to a slight peak in the minimal distance, as it consists of leaving the pump open. Since the first attack only affects the system briefly, its characteristic is hardly different than normal operation. The second attack, however, leads to a notable peak in the minimal distance, clearly marking an anomaly. Both attacks can be detected automatically, given an appropriate threshold is selected. Additionally, the width of the peak can be used as an indicator of an anomaly. The window size m was set to *2000* for analysis of this interval. A value of *500* led to constant minimal distance of 0, presumably since a window size smaller than a period of operation does not contain relevant information, i.e. the length and structure of one period.

Interval 3 is the period on January 1st from 14:30:00 to 20:00:00, covering 19 801 events during which two attacks occur. The window size m is set to 500, as it produces better results than window sizes of 1 000 and 2 000. This contrasts the results of other works employing *Matrix Profiles* for process data evaluation stating that the window size m should not be smaller than the first peak of the autocorrelation function, but might as well be larger [13]. In interval 3, the pump *P-101* transporting water from the raw water tank into the second stage is observed. For both attacks, *P-101* is turned off to stop the water flow. However, during the first attack, backup pump *P-102* is turned on so that the effect on the water level is not visible for the operator. This behaviour of pumps *P-101* and *P-102* is shown in Figure 5. The minimal distances of both actuators peak notably during the attacks. Due to the length of the window size and the closeness of the attacks, the minimal distance only contains one peak, which starts at the beginning of the first attack. The behaviour of the water level sensor *LIT-301* during that period is shown in Figure 6. The minimal distance for this sensor is calculated with the Scalable Time series Anytime Matrix Profile (STAMP) algorithm [8]. Due to the backup pump *P-102*, the first attack cannot be

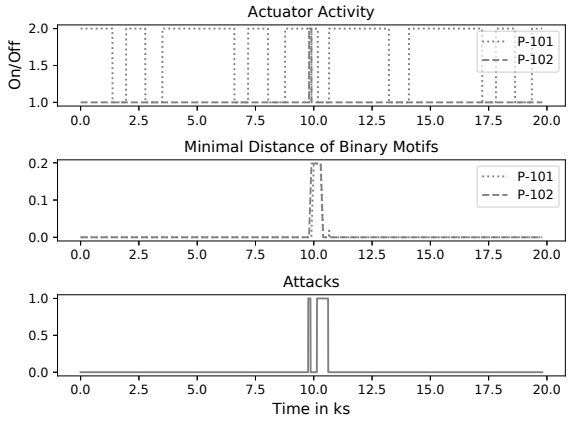

Figure 5. *Matrix Profile* of Interval 3 for *P-101* and *P-102*

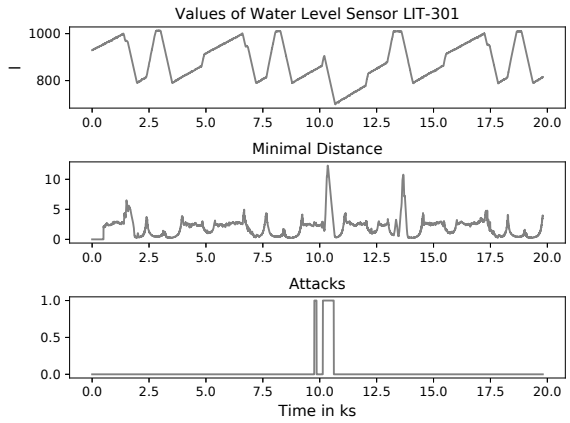

Figure 6. *Matrix Profile* of Interval 3 for *LIT-301*

detected by analysing the water level, as there is no effect on the process, except for an anomalous pump use. In this case, evaluating actuators can detect attacks that could not be detected otherwise.

VI. Discussion

The evaluation of the actuators provides an accurate detection of attacks on the process. In case of interval 3, the analysis of actuator behaviour is capable of detecting attacks that do not influence the system because of a backup pump *P-102*. Knowledge about an attempted attack that has failed is valuable for threat intelligence as it indicates the presence of an attacker. Due to the uniform behaviour of actuators, attacks are clearly distinguishable from normal operation by looking at the distance metrics. It is typical for processes in batch processing to repeat for a large amount of time, providing a sound normal behaviour that is used as the basis for intrusion detection. An advantage of *Matrix Profiles* is the unsupervised fashion to deploy it. In contrast to many machine learning-based approaches, neither training nor labelled data are necessary to create sound models. Furthermore, it requires one hyper-parameter that is relatively robust. Still, choosing a window size m is the most difficult task in employing *Matrix Profiles* Values that are too small lead to an increase in false negatives, while window sizes that are too large lead to false positives. Similar to previous works, window sizes around the period length seem to be an optimal choice.

VII. Conclusion

This work showed that the extended *Matrix Profiles* employing the *Hamming*-distance are capable of detecting all evaluated attacks without any false positives. Furthermore, training was performed automatically. For future work, extending the algorithm to extract motifs could prove beneficial in terms of computation time. Since many sequences are expected to be identical, a dictionary with a motif and its number of occurrences reduces the amount of comparisons required, while providing the same amount of information. Furthermore, the number of occurrences can be used as an additional metric for an outlier, as this would detect an attack that was deployed twice. The second time would not be detected by regular *Matrix Profiles* as the exact same motif was already present. However, any motif with a comparably low number of occurrences could be considered suspicious, similar to the work of *Duque Anton et al.* [14].

Acknowledgement

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Foerderkennzeichen 16KIS0932, IUNO Insec). The data set used in this work has been provided by *iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design*. The authors alone are responsible for the content of the paper.

References

[1] "Study on communication for automation in vertical domains," 2017, 3GPP TR 22.804, V1.0.0.

[2] S. Haller, S. Karnouskos, and C. Schroth, "The internet of things in an enterprise context," in *Future Internet Symposium*. Springer-Verlag, Berlin, Heidelberg, 2008, pp. 14–28.

[3] V. M. Igure, S. A. Laughter, and R. D. Williams, "Security issues in SCADA networks," *Computers & Security*, vol. 25, pp. 498–506, 2006.

[4] S. Duque Anton, D. Fraunholz, C. Lipps, F. Pohl, M. Zimmermann, and H. D. Schotten, "Two decades of SCADA exploitation: A brief history," in *2017 IEEE Conference on Application, Information and Network Security (AINS)*, November 2017, pp. 98–104.

[5] E. Powers, S. Peasley, R. Waslo, B. Fletcher, and D. Dinh, "Examining the industrial control system cyber risk gap," https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-aers-ics-white-paper.pdf, Deloitte, Tech. Rep., 2015.

[6] M. Iturbe, I. Garitano, U. Zurutuza, and R. Uribeetxeberria, "Towards large-scale, heterogeneous anomaly detection systems in industrial networks: A survey of current trends," *Security and Communication Networks*, 2017.

[7] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, 1950.

[8] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, "Matrix Profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, December 2016, pp. 1317–1322.

[9] P. Schneider and K. Böttinger, "High-performance unsupervised anomaly detection for cyber-physical system networks," in *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy*, ser. CPS-SPC '18. New York, NY, USA: ACM, 2018, pp. 1–12. [Online]. Available: http://doi.acm.org/10.1145/3264888.3264890

[10] iTrust Center for Research in Cyber Security, "S317 dataset," https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/\#s317.

[11] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*, January 2017, pp. 140–145.

[12] S. Duque Anton, D. Fraunholz, and H. D. Schotten, "Using temporal and topological features for intrusion detection in operational networks," in *ARES '19: Proceedings of the 13th International Conference on Availability, Reliability and Security*, ACM. ACM, August 2019.

[13] S. Duque Anton, L. Ahrens, D. Fraunholz, and H. D. Schotten, "Time is of the essence: Machine learning-based intrusion detection in industrial time series data," in *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, November 2018.

[14] S. Duque Anton, A. P. Lohfink, C. Garth, and H. D. Schotten, "Security in process: Detecting attacks in industrial process data," in *Proceedings of the 3rd Central European Cybersecurity Conference (CECC)*, ACM. ACM, November 2019.

[15] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2017, pp. 1058–1065.

[16] M. Kravchik and A. Shabtai, "Detecting cyber attacks in industrial control systems using convolutional neural networks," in *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy*, 2018, pp. 72–83.

[17] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "Madgan: Multivariate anomaly detection for time series data with generative adversarial networks," in *International Conference on Artificial Neural Networks*. Springer, 2019, pp. 703–716.

[18] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," *arXiv preprint arXiv:1809.04758*, 2018.

[19] Y. Chen, C. M. Poskitt, and J. Sun, "Learning from mutants: Using code mutation to learn and monitor invariants of a cyber-physical system," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 648–660.

[20] Q. Lin, S. Adepu, S. Verwer, and A. Mathur, "Tabor: A graphical model-based approach for anomaly detection in industrial control systems," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, 2018, pp. 525–536.

[21] C. Feng, T. Li, and D. Chana, "Multi-level anomaly detection in industrial control systems via package signatures and lstm networks," in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, June 2017, pp. 261–272.

[22] E. D. Knapp and J. T. Langill, *Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems*. Syngress, 2014.

[23] Y. Yang, K. McLaughlin, S. Sezer, T. Littler, E. G. Im, B. Pranggono, and H. F. Wang, "Multiattribute SCADA-specific intrusion detection system for power networks," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1092–1102, June 2014.

[24] R. D. Larkin, J. Lopez Jr., J. W. Butts, and M. R. Grimaila, "Evaluation of security solutions in the SCADA environment," *SIGMIS Database*, vol. 45, no. 1, pp. 38–53, Mar. 2014.

[25] L. A. Maglaras and J. Jiang, "Intrusion detection in SCADA systems using machine learning techniques," in *2014 Science and Information Conference*. IEEE, 2014, pp. 626–631.

[26] M. Herrero Collantes and A. López Padilla, "Protocols and network security in ICS infrastructures," https://www.incibe.es/extfrontinteco/img/File/intecocert/ManualesGuias/incibe_protocol_net_security_ics.pdf, Spanish National Cybersecurity Institute, Tech. Rep., 2015.

[27] W. Gao and T. H. Morris, "On cyber attacks and signature based intrusion detection for modbus based industrial control systems," *Journal of Digital Forensics, Security and Law*, vol. 9, no. 1, 2014.

[28] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *Proceedings of the 11th International Conference on Critical Information Infrastructures Security*, October 2016.

[29] iTrust Centre for Research in Cyber Security, "Secure water treatment (SWaT) testbed," Singapore University of Technology and Design, Tech. Rep. 4.2, October 2018.

[30] S. Duque Anton, M. Gundall, D. Fraunholz, and H. D. Schotten, "Implementing scada scenarios and introducing attacks to obtain training data for intrusion detection methods," in *ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS 2019*. Academic Conferences and publishing limited, 2019, p. 56.

[31] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, vol. 2. Springer, Berlin, Heidelberg, 2009, pp. 1–4.

[32] A. Mueen, S. Nath, and J. Liu, "Fast approximate correlation for massive time-series data," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 171–182.