A Novel Approach for Detection and Location of Cyber-Attacks in Water Distribution Networks

Claudia Rodríguez Martínez[1], Marcos Quiñones-Grueiro[1], Cristina Verde[2], and Orestes Llanes-Santiago[1(✉)]

[1] Universidad Tecnológica de La Habana José Antonio Echeverría, CUJAE, 19390 La Habana, Cuba
orestes@tesla.cujae.edu.cu
[2] Instituto de Ingeniería, UNAM, Mexico City, Mexico
cverde@unam.mx

Abstract. Most scientific contributions addressing cyber-security issues in water distribution networks present proposals of detection systems and very few propose location systems. A novel methodology for detection and location of cyber-attacks in water distribution networks (WDNs) is proposed in this paper. Structural analysis and autoencoder neural networks are effectively combined with a the control chart Adaptive Exponentially Weighted Moving Average (AEWMA). In the training phase, the proposed detection and location framework only requires data from normal operating conditions and knowledge about the behavioral model of the system which represents an advantage over previous works that demand for additional data of cyber-attacks. Among other advantages of the proposed methodology are the high performance in the effective, robust and early detection and the effectiveness of the location strategy. The proposal was evaluated with the known case study BATADAL.

Keywords: Ciber-attacks · Water distribution networks · Pattern recognition · Autoencoders · Structural analysis

1 Introduction

The advances in the information technologies and the industrial computing have transformed the traditional water distribution networks (WDN) in cyber-physical systems. The combination of physical processes with cyber-physical systems improves the service of the urban WDN but expose them to the potential threats of cybernetic attacks [1,12]. In the last years, several water distribution and supply systems have received cyber-attacks [4]. This has motivated the creation of cyber-security agencies for protecting and defending WDN.

Project No. 27 of National Program of Research and Innovation ARIA of CITMA, Cuba.

Y. Hernández Heredia et al. (Eds.): IWAIPR 2021, LNCS 13055, pp. 79–90, 2021.
https://doi.org/10.1007/978-3-030-89691-1_9

In a bibliographic review about the cyber-security in WDN can be appreciated that the most algorithms address cyber-attack detection but few works are concerned with the location of the attacks being this a current open research problem [4,10,13]. At present, cyber-attack detection is based in the identification of abnormal patterns in the behaviour of the variables. Advanced methods use signal spectral analysis or the comparison between the ideal behaviour of the WDN (by using a model) with its current state [6,13].

The main objective of this paper is to propose a novel methodology for detection and location of cyber-attacks in WDN combining tools of computational intelligence and structural analysis which constitutes its main contribution. The methodology only requires the normal operating data, and it does not require a parameterized physical model of the network constituting both aspects other contributions of the paper. For the validation of the proposal, data from the BATADAL test problem were used [13].

The organization of the paper is as follows: in Sect. 2, the principal characteristics of computational tools and the structural analysis used in the paper are presented. In Sect. 3, the proposed methodology is described. In Sect. 4, this methodology is applied to the C-Town case study. An analysis and discussion of the obtained results is developed in Sect. 5. Finally, the conclusions and recommendations for future works are presented.

2 Materials and Methods

2.1 Adaptive Exponential Weighted Moving Average Chart

Exponentially weighted moving average (EWMA) is a univariate control chart used for detecting deviations in the mean of a signal [11]. The EWMA of a signal $x(t) \in \Re^p$ is defined as:

$$z(t) = \gamma \bar{x}(t) + (1-\gamma) z(t-1) \;\; z(1) = \mu \tag{1}$$

The Adaptive EWMA (AEWMA) can detect either small or large shifts or both simultaneously. In this paper, the following score function is used [2]:

$$\phi(e) = \begin{cases} e + (1-\gamma)k & if \;\; e < -k \\ \gamma e & if \;\; |e| \leq k \\ e - (1-\gamma)k & if \;\; e > k \end{cases} \quad 0 \leq \gamma \leq 1 \text{ and } k \geq 0 \text{ are constants}$$

An abnormal pattern is detected when $z(t)$ exceeds the control limits $\mu \pm h\sigma$, where μ and σ are the mean and the standard deviation parameters of a reference signal and h is chosen with the aim to achieve a desired performance. To avoid false alarms, the Average of the Run Length (ARL) of the control chart is used as performance measure of the AEWMA control chart [2]. Several methods can be used to determine the design parameters γ and k. In this paper, those parameters are calculated as in [2].

$$\lambda = \ln(1.2219 - 0.04697 * \ln(ARL) + 0.45985 * \sqrt{\delta_{min}} - 0.02701 * \sqrt{\delta_{max}}) \quad (2)$$

$$k = \sqrt{4.846 + 1.5852 * \ln(ARL) - 2.8679 * \sqrt{\delta_{min}} - 1.7198 * \sqrt{\delta_{max}}} \quad (3)$$

2.2 Autoencoders

Autoencoders are deep neural networks with the aim to transform input patterns into outputs with a minimum distortion. They encode the input pattern in compressed form to the feature space and after decode this information with the objective to obtain a pattern as similar as possible to the original one [3].

Autoencoders are formally defined in [3] by the following elements: 1) $\mathfrak{X} = \{x_1, ..., x_m\}$ such as $x_i \in \Re^n$ a set of m vectors representing an input pattern, 2) $A : \Re^n \rightarrow \Re^p$ such as $n > p$ a function that represents the encoder, 3) $B : \Re^p \rightarrow \Re^n$ a function that represents the decoder, and 4) Δ is a distortion function (e.g. **L2** norm) defined in $\Re^n$ which measures the distance between the output pattern and the input pattern.

The autoencoder transforms a vector $x_i \in \Re^n$ into an output vector $A \circ B(x_i) \in \Re^n$. The autoencoder problem is to find the functions A and B such that the distortion function is minimized.

The cost function used in the training process is formulated as [8]:

$$J = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{k} (x_{ij} - \hat{x}_{ij})^2 + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity}$$

where

$$\Omega_{weights} = \frac{1}{2} \sum_{l=1}^{L} \sum_{j=1}^{n} \sum_{i=1}^{k} (\omega_{ji}^{(l)})^2$$

$$\Omega_{sparsity} = \sum_{i=1}^{D^{(2)}} KL(p \| \hat{p}_i) = \sum_{i=1}^{D^{(2)}} p * \log\left(\frac{p}{\hat{p}_i}\right) + (1-p) * \log\left(\frac{1-p}{1-\hat{p}_i}\right)$$

$\Omega_{sparsity}$ is used to improve the learning in the training, $\Omega_{weights}$ is used to avoid overfitting, n represents the number of observations used in the training set, k represents the number of variables in the data training, L represents the number of hidden layers in the sparse autoencoder, ω_{ji} are the weights, $D^{(2)}$ is the number of neurons in the hidden layer, KL is the Kullback-Leibler divergence, $\hat{p}_i$ is the average activation of a single neuron obtained as $\hat{p}_i = \frac{1}{n} \sum_{j=1}^{n}, z_i^{(2)}(x_j))$, $z_i^{(2)}(x)$ is the activation of a single neuron belonging to the hidden layer, p is

a restriction imposed for $\hat{p}_i$, the weight decay parameter λ controls the relative importance of the first two terms in the cost function and the parameter β controls the weight of the sparsity penalty term.

2.3 Structural Analysis

Structural analysis is a model-based methodology that has been used for fault diagnosis in industrial processes [5]. In a similar way, it can be used to detect and locate cyber-attacks. The structural model of a system represents an abstraction of its behavioral model that permits to establish analytical redundancy relationships (ARRs) with the goal to detect and locate cyber-attacks in the system. The number of ARRs that can be defined depend on the known variables, and the structure of the system. ARRs represent a set of constraints/rules evaluated during system operation by using the variables calculated from the model and the measurements obtained from the system. If a cyber-attack occurs, one or several ARRs won't be consistent. The advantage of this methodology is the possibility to analyze the detectability and isolability of different cyber-attacks without requiring its analytical model. The behaviour model of a system can be defined by a pair $(\mathcal{C}, \mathcal{V})$ where $\mathcal{V} = v_1, v_2, ..., v_n$ represents a set of variables, and $\mathcal{C} = c_1, c_2, ..., c_m$ represents a set of constrains. The set of variables $\mathcal{V} = \mathcal{K} \cup \mathcal{X}$ where $\mathcal{K}$ represents the subset of known variables and $\mathcal{X}$ represents the subset of unknown variables.

Definition 1. *Structural model* [5]

The structural model of the system $(\mathcal{C}, \mathcal{V})$ *is a bi-partite graph* $(\mathcal{C}, \mathcal{V}, \mathcal{E})$ *where* $\mathcal{E} \subset \mathcal{C} \times \mathcal{V}$ *is a set of edges defined by:*

$$e_{ij} \in \mathcal{E} = \begin{cases} (c_i, v_j) & \textit{if the variable } v_j \textit{ appears in the constraint } c_i \\ 0 & \textit{in other cases} \end{cases}$$

A bi-partite graph has associated an incidence matrix [5]. In Structural Analysis, the possibility to establish ARRs implies that the graph has more constraints that unknown variables. Applying the canonical Dulmage-Mendelsohn decomposition [7] to the graph $(\mathcal{G})$, it can be divided in tree parts: the structurally over-constrained part $\mathcal{G}^+$ which has more constrains than unknown variables, the just-determined part $\mathcal{G}^0$ with the same number of constraints and unknown variables, and the structurally under-constrained part $\mathcal{G}^-$ with less constraints than unknown variables. The cyber-attacks that affect the constrains belonging $\mathcal{G}^0$ and $\mathcal{G}^-$ are not detectable.

3 Methodology for Detection and Location of Cyber-Attacks

The methodology proposed for the detection and location of cyber-attacks in this paper is shown in Fig. 1.

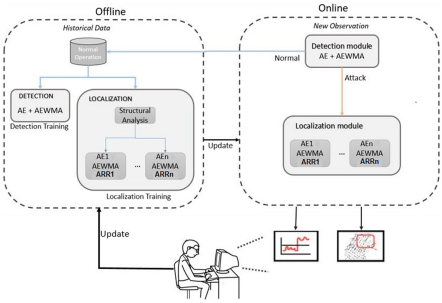


Fig. 1. Methodology proposed for detection and location of cyber-attacks

3.1 Offline and Online Stages

In the offline stage, an autoencoder is trained with data of normal operation for the detection process. In the training, a set of measurements obtained by the supervisory and data acquisition system (SCADA) are used. Together with the autoencoder, the parameters of the AEWMA control chart are tuned for detecting deviations. Also at this stage, experts will use structural analysis of the WDN to determine the ARRs that characterize each DMA into which the network is divided. For each ARR, an autoencoder is then trained with measured variables of the WDN in normal operation present in the respective ARR. Also, an AEWMA chart associated with each autoencoder is tuned to detect a deviation in the normal operation. The detection of an attack in a DMA can be characterized for the activation of one or several AEs simultaneously.

In the online stage, each new observation is analyzed by the detection module. If the observation is considered within the condition of normal operation, it can be used to update the training database in the offline stage. If the observation is not considered within the normal operation of the WDN, then it is analyzed with the combination of Autoencoder plus AEWMA established for each ARR in the location module. This observation should cause that the ARRs involved with the DMA under attack become inconsistent. This allows the location of the DMA where the cyber-attack is taking place.

4 Application of the Proposed Methodology to the C-Town Case Study

4.1 Case Study: C-Town WDN and BATADAL Datasets

C-Town is a medium-sized network based on real-world. The network consists of a single reservoir, 7 storage tanks, 429 pipes, 11 pumps distributed across 5 pumping stations (S1-S5), 388 junctions and 5 valves. Pumps, valves, and level sensors of the tanks are connected to 9 programmable logic controllers (PLCs) which form a cyber-network together with a central computer where a SCADA system coordinates the operations through the PLCs [9]. BATADAL data sets

were introduced in [13]. Three data sets were generated by using the simulation package EPANET2 including the information of 43 variables sampled with fixed hourly intervals. Of all variables, 31 are continuous and 12 are binary variables corresponding to the status of valve and pumps.

Data Set 1 and 2 can be used for training the detection algorithms. Data Set 1 was generated by simulating the operation of the C-Town WDN during 365 days without the presence of cyber-attacks. This data set allows to study the operations of the WDN under normal conditions. Data Set 2 contains information of 7 attacks produced in an interval of time of 492 h. Data Set 3 contains the information of 7 additional attacks during an interval time of 407 h, and it should be used to test the performance of the detection algorithm after training. A complete characterization about Data Sets 2 and 3 can be seen in [13].

4.2 Detection Module in the Offline Stage

In the detection module the autoencoder is trained by using 8753 observations from normal operating data, and with a maximum number of epochs of 3000. Root Mean Square Error (RMSE) function is used as performance measure. After a grid search experiment, the parameters that allow to obtain the best performance are $p = 0.99$, $\lambda = 0.0001$, $\beta = 1$. After several experiments for different dimensions of latent space and for different combinations of the activation functions sigmoid and linear, the selected latent space dimension was 24 variables and the sigmoid and linear activation functions were chosen for encoder and decoder, respectively. In the case of the AEWMA chart, the values of the parameters were chosen as $ARL = 400$, $\delta_{min} = 0.5$, $\delta_{max} = 5$ and $h = 0.669338$, considering the results presented in [2]. Furthermore, by evaluating in (2) $\gamma = 0.102013$ and $k = 2.984267$ are obtained.

4.3 Location Module in the Offline Stage

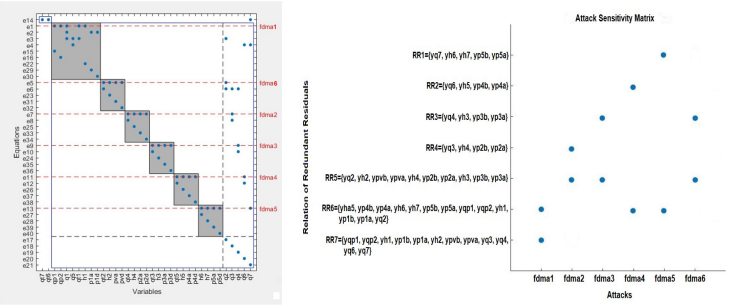
The first step in this module is to establish the possible attack areas (AAs). For C-Town WDN, there are 5 DMAs and the first idea was to identify each DMA with an AA. However, DMA1 contains two sub-areas very important: the first is the Pumping Station 1 and the second is, the valve 2 which controls the distribution of water to DMA2 and DMA3. For that reason, two AAs were established for DMA1.

In the second step, structural analysis is developed by experts to determine if each AA can be characterized by one or several ARRs which allow to locate an attack. The following variables are available: the level (h) of each tank (t), the flow rate (q), the inlet and outlet pressure (p), and the status of each pump (p), as well as the flow and status of valve (v) 2. In the development of the structural analysis three subset of variables were defined: Attacks ($fdma\#$, $\# : 1, 2, .., 6$), observable variables (always begin with $y...$), and system variables. The set of equations that describe the behavior of the system is shown in Table 1.

Table 1. Equations defined for the structural analysis

$e_1 : \{qp1, qp2, q1, qt1, h1, fdma1\}$	$e_{15} : \{yqp1, qp1\}$	$e_{29} : \{yp1a, p1a\}$
$e_2 : \{q1, p1a, p1d\}$	$e_{16} : \{yqp2, qp2\}$	$e_{30} : \{yp1d, p1d\}$
$e_3 : \{q1, qt1, q2, q5\}$	$e_{17} : \{yq2, q2\}$	$e_{31} : \{ypva, pva\}$
$e_4 : \{q5, q6, q7\}$	$e_{18} : \{yq3, q3\}$	$e_{32} : \{ypvd, pvd\}$
$e_5 : \{q2, pva, pvd, qt2, h2, fdma6\}$	$e_{19} : \{yq4, q4\}$	$e_{33} : \{yp2a, p2a\}$
$e_6 : \{q2, qt2, q3, q4\}$	$e_{20} : \{yq6, q6\}$	$e_{34} : \{yp2d, p2d\}$
$e_7 : \{q3, p2a, p2d, qt4, h4, fdma2\}$	$e_{21} : \{yq7, q7\}$	$e_{35} : \{yp3a, p3a\}$
$e_8 : \{q3, qt4\}$	$e_{22} : \{yh1, h1\}$	$e_{36} : \{yp3d, p3d\}$
$e_9 : \{q4, p3a, p3d, qt3, h3, fdma3\}$	$e_{23} : \{yh2, h2\}$	$e_{37} : \{yp4a, p4a\}$
$e_{10} : \{q4, qt3\}$	$e_{24} : \{yh3, h3\}$	$e_{38} : \{yp4d, p4d\}$
$e_{11} : \{q6, p4a, p4d, qt5, h5, fdma4\}$	$e_{25} : \{yh4, h4\}$	$e_{39} : \{yp5a, p5a\}$
$e_{12} : \{q6, qt5\}$	$e_{26} : \{yh5, h5\}$	$e_{40} : \{yp5d, p5d\}$
$e_{13} : \{q7, p5a, p5d, h6, h7, fdma5\}$	$e_{27} : \{yh6, h6\}$	
$e_{14} : \{q7, qt6, qt7\}$	$e_{28} : \{yh7, h7\}$	

An analysis of the structural isolability properties shown the possibility to isolate all defined attacks in a unique way. Figure 2a) shows the corresponding Dulmage-Mendelsohn decomposition where it is appreciated the isolability properties because attacks appear in different classes. For designing the residual generators, the structurally over-constrained minimal equation sets (SOMs) are determined. With them, and based on the Fault Sensitivity Matrix, the set of constrains that satisfy the specifications of isolability are determined. In this case, 28 SOMs were obtained and seven analytical redundant residuals (ARRs) were established which are shown in Fig. 2b).



(a) (b)

Fig. 2. (a) Dulmage-Mendelsohn decomposition (b) Attack sensitivity matrix

For each ARR an autoencoder was trained. In the training process, 8753 samples of normal operation of the WDN were used but only with the variables present in each ARR. Furthermore, the parameters of the AEWMA control charts associated with each autoencoder were established and they are shown in Table 2a) where DLS represents the dimension of the latent space. The same activation functions were used for the seven autoenconders: linear for the encoders and sigmoid for the decoders. After several experiments and considering the results presented in [2] the parameters for the seven AEWMA control charts of the location module in the offline stage were chosen. They are shown in Table 2b). A set of 289 observations of the data set 3 of the C-Town case study were used in the analysis of the performance of the AEWMA control charts because it contains data of attacks made to all AAs except to the AA4.

Table 2. a) Parameters of autoencoders and b) Parameters of AEWMA control charts in the location module

	Parameters of autoencoders			
ARRs	λ	p	β	DLS
ARR1	0.0001	0.1	1	6
ARR2	0.01	0.1	1	5
ARR3	0.00001	0.1	1	3
ARR4	0.00001	0.1	1	3
ARR5	0.0001	0.1	1	9
ARR6	0.001	0.1	1	12
ARR7	0.01	0.1	1	12

Parameters of AEWMA control charts					
ARL	δ_{min}	δ_{max}	h	γ	k
100	1.5	4	1.416963	0.416893	2.205357
600	0.25	4	0.664883	0.095301	3.178406
300	0.5	5	0.956107	0.192733	2.817520
500	0.25	6	0.660161	0.095073	3.035565
400	0.25	6	0.666165	0.100697	2.969459
500	0.25	6	0.660161	0.095073	3.035565
600	0.25	6	0.660161	0.095073	3.035565

4.4 Performance Assessment

There are two important characteristics to be satisfied by a system for detection of cyber-attacks: 1) the early and reliable detection of an attack and 2) its correct classification. To evaluate the early detection of an attack, and the performance in the classification process the indexes S_{TTD} and S_{CLF} are defined respectively as:

$$S_{TTD} = 1 - \frac{1}{n_a} \sum_{i=1}^{n_a} \frac{TTD_i}{\Delta t_i}, \; S_{CLF} = \frac{TPR + TNR}{2} \tag{4}$$

where n_a is the number of attacks contained in the data set, TTD_i is the time to detection of the $i-th$ attack and its corresponding duration time is Δt_i. TPR and TNR represent the True Positive Rate and True Negative Rate respectively, and they are calculated as:

$$TPR = \frac{TP}{TP + FN}, \; TNR = \frac{TN}{TN + FP} \tag{5}$$

where TP is the true positive alarms, FN is false negative alarms, TN is true negative alarms and FP is false positive alarms.

Both index are integrated in an index of global performance

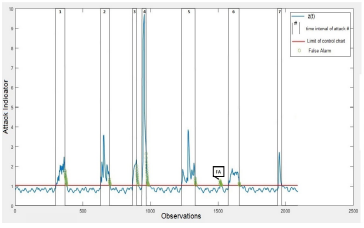
$$S_{GP} = \varsigma * S_{TTD} + (1 - \varsigma) * S_{CLF} \tag{6}$$

where ς determines the relative importance of the indexes S_{TTD} and S_{CLF} in the general index S_{GP}. In this paper $\varsigma = 0.5$ to give the same weight to early detection and correct classification. The False Alarm Rate (FAR) index is calculated as:

$$FAR = 1 - TNR \tag{7}$$

5 Results and Discussion

The performance of the detection system was evaluated by using 2089 observations from the data set 3 of the C-Town case study which contains the information about 7 cyber-attacks. The results of the detection system, the time to detection (TTD), the duration time (DT) of each cyber-attack expressed in number of observations, and the elements and attack area affected in each attack are shown in Fig. 3. In this figure, the blue line indicates the behaviour of the $z(t)$ variable corresponding to the AEWMA control chart of the detection system. The red horizontal line represents the limit of the AEWMA control chart for the normal operation of the WDN. With vertical lines, the start and the end time of each cyber-attack has been indicated. As can be observed, the seven attacks are detected.



Cyber-Attack Number (i)	1	2	3	4	5	6	7
TTD_i	4	1	1	0	0	1	8
DT	70	65	31	31	100	80	30

Attack number	Elements affected	AA number
1	Level of tank 3, Flow of pumps 4 and 5	3
2	Level of Tank 2, Flow of valve 2	6
3	Flow of pump 3	1
4	Flow of pump 3	1
5	Level of tank 2, Flow of valve 2, Pressure in valve 2	6
6	Level of tank 7, Flow in pumps 10 an 11	5
7	Level of tank 6	2

Fig. 3. Detection, time to detection, duration time, and elements and attack area affected in each cyber-attack for the data set 3,

Using the expression (4) the index $S_{TTD} = 0.9451$ was calculated. The observations of the data set were classified as: TP: 350, TN: 1612, FP: 70 and FN: 17. With these values and by using the expressions (5) and (7), the indicators $TNR = 0.9584$, $FAR = 0.0416$ and $TPR = 0.9582$ as well as the indexes $S_{CLF} = 0.9583$ and $S_{GP} = 0.9517$ were obtained. These results indicate the effectiveness of the detection system proposed.

To locate an attack satisfactorily, the observations during the attack should make non-consistent the ARRs that characterize the attacked area. However,

four different alternatives can occur when an observation detected as attack is analyzed: A.1) Only the actual AA is activated, A.2) Several AAs are activated and the actual AA is included in them, A.3) One or several AAs are activated and the actual AA is not included and A.4) No AA es activated.

The first result to be pointed out is that the 70 observations classified as FP did not make non-consistent any ARR. Table 3 shows the other results obtained by the cyber-attack location system. The first column (**A**) indicates the attack number. The second column (**B**) indicates the number of observations during the presence of each attack. The third column (**C**) indicates the number of observations that active at least one AA. The next four columns (**D**, **E**, **F**, **G**) indicate the number of observations that satisfy the alternatives A.1, A.2, A.3 and A.4 respectively. The column **H** indicates the location rate (LR, percent of observations that activate at least one AA with respect the total number of observations during the presence of the attack $\frac{C}{B} * 100$). The column **I** shows the exact location rate (ExLR, percent of observations that activate only the actual AA with respect the number of observations that activate at least one AA during an attack $\frac{D}{C} * 100$) The last column (**J**) indicates the effective location rate (EfLR, percent of observations that activate the real AA with respect to the number of observations that activate at least one AA $\frac{D+E}{C} * 100$)

Table 3. Results of the cyber-attack location process

A	B	C	D	E	F	G	**H** = $\frac{C}{B} * 100$	**I** = $\frac{D}{C} * 100$	**J** = $\frac{D+E}{C}$
1	70	44	41	3	0	26	62.85	93.18	100
2	65	48	41	7	0	17	73.84	85.41	100
3	31	29	29	0	0	2	93.54	100	100
4	31	31	29	2	0	0	100	93.54	100
5	100	74	51	23	0	26	74	68.91	100
6	80	23	19	4	0	57	28.75	82.60	100
7	30	16	16	0	0	14	53.33	100	100
Total	407	265	226	39	0	142	65.11	85.28	100

The most relevant aspects of the analysis considering the results in Table 3 are the following:

- The LR index was 65.11%. Note that the cyber-attack 6 is the most affected.
- The ExLR index was 85.25% which represents a satisfactory result. In this case, the most negative incidence is in the attack 5 where 23 observations activate several AAs simultaneously.
- The EfLR index was 100% which is relevant because this indicates that the real AA is always activated for all observations that activate at least one AA. This result is very important for the operators because if the location process indicates several AAs, they can be sure that one of them is the real AA. If operators simultaneously analyze which AA has been activated for the greatest number of observations during the period of time under attack, they can identify the real AA.

6 Conclusions

In this paper, a methodology for detection and location of cyber-attacks in a water distribution network is presented. The proposal is based on the combined use of structural analysis with two computational intelligence tools: autoencoders and AEWMA control chart. The proposal permits to locate an attack which represents an advantage with respect to the most proposals present in the literature. Another advantage of the proposed methodology is that it only needs data corresponding to the normal operation of the WDN for training. To evaluate the detection process the indexes of early detection S_{TTD}, performance in classification S_{CLF} and global performance S_{GP} were defined. The obtained results demonstrates the robustness and attack detection capability of the first part of the proposed methodology. With respect to the location process, very satisfactory results were obtained in the indexes exact location rate, and the effective location rate which ensure the possibility of always determining the real attack area.

References

1. Adepu, S., Palleti, V.R., Mishra, G., Mathur, A., et al.: Investigation of cyber attacks on a water distribution system. In: Zhou, J. (ed.) ACNS 2020. LNCS, vol. 12418, pp. 274–291. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-61638-0_16
2. Aly, A., Hamed, R., Mahmoud, M.: Optimal design of the adaptive exponentially weighted moving average control chart over a range of mean shifts. Commun. Stat.-Simul. Comput. **46**(2), 890–902 (2015). https://doi.org/10.1080/03610918.2014.983650
3. Baldi, P.: Autoencoders, unsupervised learning, and deep architectures. In: Guyon, I., Dror, G., Lemaire, V., Taylor, G., Silver, D. (eds.) Proceedings of ICML Workshop on Unsupervised and Transfer Learning. Proceedings of Machine Learning Research, vol. 27, pp. 37–49. JMLR Workshop and Conference Proceedings, Bellevue, 02 July 2012
4. Berglund, E., Pesantez, J., Rasekh, A., Shafiee, M., Sela, L., Haxton, T.: Review of modeling methodologies for managing water distribution security. J. Water Resour. Plan. Manag. **146**(8), 1–23 (2020). https://doi.org/10.1061/(ASCE)WR.1943-5452.0001265
5. Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M.: Diagnosis and Fault-Tolerant Control. Springer, Heidelberg (2006). https://doi.org/10.1007/978-3-540-35653-0
6. Quiñones Grueiro, M., Llanes-Santiago, O., Prieto Moreno, A., Verde, C.: Decision support system for cyber attack diagnosis in smart water networks. IFAC-PapersOnLine **51**(34), 329–334 (2019). https://doi.org/10.1016/j.ifacol.2019.01.024
7. Krysander, M., Frisk, E.: Sensor placement for fault diagnosis. IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum. **38**(6), 1398–1410 (2008)
8. Ng, A.: Sparse autoencoder. Technical report, University of Stanford (2010). https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf

9. Ostfeld, A., et al.: Battle of the water calibration networks. J. Water Resour. Plann. Manag. **138**(5), 523–532 (2012). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000191
10. Ramotsoela, D.T., Hancke, G.P., Abu-Mahfouz, A.M.: Attack detection in water distribution systems using machine learning. HCIS **9**(1), 1–22 (2019). https://doi.org/10.1186/s13673-019-0175-8
11. Roberts, S.: Control chart tests based on geometric moving averages. Technometrics **1**, 239–250 (1959)
12. Taormina, R., Galelii, S., Tippenhauer, N., Salomons, E., Ostfeld, A.: Characterizing cyber-physical attacks on water distribution systems. J. Water Resour. Plann. Manag. **143**(5), 04017009 (2017). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000749
13. Taormina, R., et al.: The battle of the attack detection algorithms disclosing cyber attacks on water distribution networks. J. Water Resour. Plann. Manag. **144**(8), 04018048 (2018). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000749