The Detection of Sensor Signal Attacks in Industrial Control Systems

Dušan Nedeljković
Teaching Assistant
University of Belgrade
Faculty of Mechanical Engineering

Živana Jakovljević
Associate Professor
University of Belgrade
Faculty of Mechanical Engineering

Zoran Miljković
Full Professor
University of Belgrade
Faculty of Mechanical Engineering

To improve productivity and efficiency in industrial manufacturing, the fourth industrial revolution leads to the implementation of Cyber Physical Systems (CPS) and Internet of Things (IoT) in the industrial environment. Ubiquitous communication makes CPS susceptible to external influences, which can have a negative intention; for instance, CPS are prone to various attacks and malicious threats by different adversaries. The impact of an attack on the system can lead to anomalies and serious consequences for system parts or the system as a whole. Security mechanisms must be developed in order to timely detect different attacks and to keep the system safe and protected. In this paper, a method for sensor signal attacks detection in a continuous time controlled systems has been proposed. The method is based on Support Vector Machines (SVM) and tested on the data obtained from the Secure Water Treatment (SWaT) testbed, a scaled-down plant that produces purified water.

Keywords: *Cyber Physical Systems, Support Vector Machines, Industrial Internet of Things, Cyber Security, Industrial Control Systems*

1. INTRODUCTION

To meet the requirements concerning efficient and flexible production, manufacturing enterprises have to embrace forthcoming technologies - Internet of Things (IoT) [1] and Cyber Physical Systems (CPS) [2]. The utilization of these technologies in industrial processes leads to significant changes that are characterized as the fourth industrial revolution – Industry 4.0 [3, 4].

CPS integrate cyber capabilities into physical elements through their augmenting by cyber modules with integrated computation and communication abilities. The devices within Industry 4.0 plant (e.g., sensors and actuators) represent CPS, i.e. smart devices that can perform local processing and communicate with each other. The control of industrial plant is carried out through intensive communication between the central control system and smart devices in the plant (Figure 1) or through mutual communication between smart devices in the case of distributed control. This implies fully implemented internet communication, usually wireless [5], which allows such systems to be controlled and monitored online. The widespread communication opens up a new space for potential cyber attacks. These attacks can have fatal consequences; they can disrupt or completely disable the functioning of the system, or even influence the safety. Therefore, the development of defense techniques with a high level of protection is necessary in order to make systems attack-resilient and keep its normal functioning. Timely detection of anomalies can be crucial for preventing possible negative consequences. This represents a challenging task since, as a rule, adversaries aim to remain stealthy during the attack.

Received: May 2019, Accepted: September 2019
Correspondence to: Dušan Nedeljković
Faculty of Mechanical Engineering,
Kraljice Marije 16, 11120 Belgrade, Serbia
E-mail: dnedeljkovic@mas.bg.ac.rs
doi:10.5937/fmet2001007N

Generally, malicious cyber attacks can be classified as Denial of Service (DoS) or deception attacks [6]. Making data temporarily or permanently inaccessible, DoS attacks obstruct data from reaching their destinations and in essence they are similar to failures in communication. DoS attacks are not adapted to the system on which they act, they have a general goal to disrupt system functioning. Unlike DoS attacks that do not aim to be stealthy, deception attacks are more difficult to recognize. Deception attacks are able to manipulate data at a higher level, sending corrupted data to the system components.

In network controlled continuous time systems, controller closes control loop over the communication network through intensive communication with smart sensors and actuators. On the controller sensor input deception attack, adversary changes real plant output – sensor signal $x(t)$ to corrupted controller sensor input $x^*(t)$, whereas actuator output deception attack changes the real value of the controller output signal $u(t)$ with corrupted value $u^*(t)$ (Figure 1).

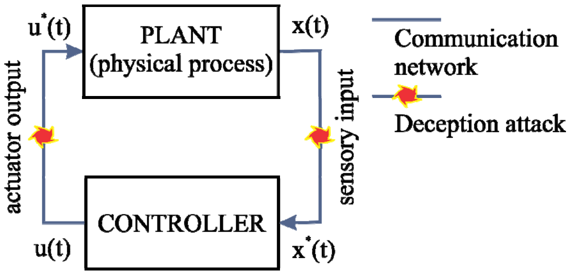

Figure 1. Model of a continuous time system under attack [6]

Deception attacks detection represents a challenging task, and a number of different techniques for solving this issue were recently proposed. All these techniques can be classified into two groups: (1) data-centric and (2) design-centric [7]. Data-centric techniques use collected system data, while design-centric techniques are based on system analytical model and its control algorithms. In [8], attack detection and identification issues are considered through the design-centric technique that uses the proposed CPS model.

Different data centric techniques were developed as well. For example, an anomaly detection method based on convolutional neural networks is proposed in [9]. Research work from [10] presents methods based on deep neural networks (DNN) and one-class support vector machines. Autoregression modeling and control limits techniques deployed for PLC security monitoring in [11], also belong to the data-centric group. A comparison of the performances of linear classification methods (Logistic Regression, Lasso, Support Vectors based classification with linear kernel), decision trees, and fully connected neural networks in cyber attacks detection in networked industrial control systems is presented in [12]. The analysis has shown that among the considered techniques fully connected neural networks presented the best accuracy (around 80%).

In this work, we consider sensor signals deception attacks and we propose a method based on ε-insensitive Support Vector Regression (ε-SVR) for their detection. The performances of the method are evaluated using publically available Secure Water Treatment (SWaT) dataset [13].

The reminder of the paper is organized as follows. Section 2 briefly outlines the ε-SVR, whereas in Section 3 we present the developed method for sensor signal attacks detection. Section 4 refers to implementation and experimental evaluation of the proposed method using SWaT dataset. Finally, the conclusions and directions for future research are discussed in Section 5.

2. SUPPORT VECTOR REGRESSION

This section provides a short overview of ε insensitive support vector regression (ε-SVR) that represents a basis of the proposed method for sensor signal attacks detection. For a given training data set $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i$, $i \in [1, l]$ denotes input variables vector, and y_i represents the corresponding response value, the aim of ε-SVR is to find a function $f(\mathbf{x})$ which has a maximum ε deviation from y_i, and simultaneously tends to be as flat as possible. Errors that are less than the value of ε are not considered. The function $f(\mathbf{x})$ in the case of linear dependence is described as follows:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \tag{1}$$

The flatness of the function is achieved when $\mathbf{w}$ is small. Thus, given considerations can be formalized through the following minimization problem:

$$\begin{aligned} &\mathit{minimize} \frac{1}{2}\|\mathbf{w}\|^2 \\ &\mathit{subject\,to} \left| y_i - \langle \mathbf{w}, \mathbf{x_i} \rangle - b \right| \le \varepsilon \end{aligned} \tag{2}$$

which minimizes the norm of $\mathbf{w}$ while keeping all the data within ε tube around $f(\mathbf{x_i})$. Expression (2) implies an approximation of all pairs $(\mathbf{x_i}, y_i)$ with ε precision. As this case does not frequently happen in reality, deviations can be allowed by introduction of slack variables ξ_i, ξ_i^*, in (2) leading to the following form of the optimization problem [14]:

$$\begin{aligned} &\mathit{minimize} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &\mathit{subject\,to} \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x_i} \rangle - b \le \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x_i} \rangle + b - y_i \le \varepsilon + \xi_i^* \end{cases} \end{aligned} \tag{3}$$

where constant C corresponds the tradeoff between function flatness and the number of points that fall outside the ε-tube. The optimization problem (3) can be represented in its dual form [14]:

$$\begin{aligned} &\mathit{minimize} \\ &\left\{ \begin{array}{l} -\dfrac{1}{2}\displaystyle\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle \mathbf{x_i}, \mathbf{x_j} \rangle \\ -\varepsilon\displaystyle\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) + \displaystyle\sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{array} \right. \\ &\mathit{subject\,to} \\ &\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0, \mathit{where}\, \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{4}$$

In (4), α_i and α_i^* represent Lagrange multipliers. If the condition $|f(\mathbf{x_i})\text{-}y| \ge \varepsilon$ is fulfilled, the Lagrange multipliers are non-zero. Otherwise, if the vectors are within the ε-tube, α_i and α_i^* vanish (equal zero). Vectors corresponding to non-zero α_i and α_i^* are called *support vectors*. The solution of the problem described in (3) is given by:

$$\mathbf{w} = \sum_{ns}(\alpha_i - a_i^*)\mathbf{x_i} \tag{5}$$

where ns represents the number of support vectors, and $f(\mathbf{x})$ becomes:

$$(\mathbf{x}) = \sum_{ns}\left(\alpha_i - a_i^*\right)\langle \mathbf{x_i}, \mathbf{x} \rangle + b \tag{6}$$

ε-SVR can be extended to non-linear regression. This is achieved by mapping the input vectors into high-dimensional space where the regression is linear. Since in (4), training input vectors are only present in inner product the space of higher dimension can be implicitly defined, i.e. it is only necessary to know inner product in this space and not the explicit transformation of data. Inner product can be defined using kernel $K(\mathbf{x}, \mathbf{x_i})$ (a function that satisfies the conditions of Mercer's theorem [14]) leading to the following form of relation (4):

$$\begin{aligned} &\mathit{minimize} \\ &-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x_i}, \mathbf{x_j}) \\ &\quad -\varepsilon\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \\ &\mathit{subject\,to} \\ &\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0, \mathit{where}\, \alpha_i, \alpha_i^* \in \left[0, C\right] \end{aligned} \tag{7}$$

Accordingly, the function $f(\mathbf{x})$ becomes:

$$f(\mathbf{x}) = \sum_{ns} \left(\alpha_i - \alpha_i^*\right) K\left(\mathbf{x_i}, \mathbf{x_j}\right) + b \tag{8}$$

Some of the commonly used kernels are polynomial kernel, radial basis kernel, wavelet kernel, sigmoid kernel, etc. It is necessary to select the kernel that is effective for the given application. The kernel chosen for the application in this work is radial basis function defined as follows:

$$K\left(\mathbf{x}, \mathbf{x_i}\right) = exp\left\{-\gamma \left|\mathbf{x} - \mathbf{x_i}\right|^2\right\} \tag{9}$$

where γ determines the width of the bell-shaped curve.

3. SENSOR SIGNAL ATTACKS DETECTION METHOD

The method for sensor signal attacks detection that we propose in this paper consists of offline ε-SVR training and online attack detection. During training, ε-SVR model of the sensor signal for the system operating under normal conditions is generated, whereas online attacks detection is based on the discrepancy between sensor signal values estimated using generated model and measured values (Figure 2).

In our approach, the current value of the sensor signal x_i is estimated from the buffer of previous k values x_{i-k}, ...x_{i-1} generating the independent variables vector $\mathbf{x_i}=[x_{i-k}, \dots x_{i-1}]$.

Thus, in the training phase, ordered pairs of independent and response values are defined as follows:

$$\begin{gathered}(\mathbf{x}, y) = \left(\left[x_1, \dots, x_k\right], x_{k+1}\right) \\ \left(\left[x_2, \dots, x_{k+1}\right], x_{k+2}\right), \dots, \left(\left[x_{n-k}, \dots, x_{n-1}\right], x_n\right)\end{gathered} \tag{10}$$

The buffer length k is one of the parameters that should be tuned during training in order to get as good a model as possible. Furthermore, ε-SVR parameters (ε, C, and radial basis function parameter γ) need to be set. Each combination of the parameters leads to one ε-SVR model.

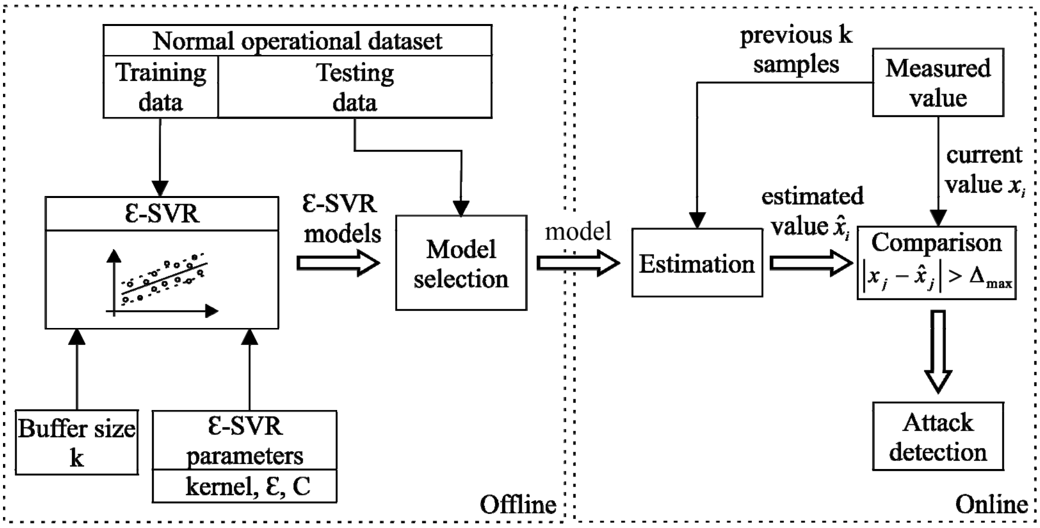

Figure 2. Overview of the method for sensor signal attacks detection

For the selection of the most appropriate model, two criteria are employed: (1) number of support vectors (SV), and (2) the model accuracy over the whole dataset including test data. To keep the model as simple as possible and to reduce the computational complexity and the online estimation latency thereof, the number of the support vectors should be as low as possible. On the other hand, the reduction of the number of support vectors can lead to the inaccurate model. Thus, to preserve the quality of the estimation, we also consider the number (Δ_{out}) of estimated values over the whole dataset whose absolute errors with respect to response values exceeded the threshold m. The threshold is based on the mean absolute deviation between the real x_i and the estimated values $\left|x_j - \hat{x}_j\right| > \Delta_{max}$:

$$m = 100 \frac{1}{n} \sum_{i=k+1}^{n} \left|x_i - \hat{x}_i\right| \tag{11}$$

Once appropriate model is selected during offline training phase, the online attack detection is carried out based on the difference of estimated and measured values. If the absolute error between measured sensor value and estimated value exceeds the detection threshold, the attack or another anomaly in the system performance is present. The attack detection threshold (Δ_{max}) represents the maximum absolute error between estimated and measured signal values on the whole dataset increased by 10%:

$$\begin{gathered}\Delta_{max} = 1.1 \cdot max\left|x_i - \hat{x}_i\right| \\ for\, i = k+1, \dots, n\end{gathered} \tag{12}$$

Thus, during online detection, the attack occurs when the absolute difference between measured (x_j) and estimated $\hat{x}_j$ sensor signal value crosses over Δ_{max}:

$$\left|x_j - \hat{x}_j\right| > \Delta_{max} \tag{13}$$

4. IMPLEMENTATION AND EXPERIMENTAL EVALUATION OF THE PROPOSED METHOD

To test the performances of the proposed method, we have used the publicly available SWaT dataset. In this section, before presenting the results of the method implementation, we briefly overview the dataset and the system that was used as a platform for its generation.

4.1 Secure Water Treatment (SWaT) testbed

SWaT is functional scaled down water treatment plant at the Singapore University of Technology and Design, which is capable to produce 5 gallons of purified water per minute. It was built to provide data from the real system useful for research mainly in the field of cyber-security [15]. The whole process of water purification is divided into 6 serially interconnected stages (P1 to P6), as shown in Figure 3. Each of the stages contains a number of sensors and actuators, from which the required signals are obtained. Sensors and actuators are connected via wired or wireless links to the appropriate PLCs (Programmable Logic Controller). Further, all PLCs within the SWaT are connected to the SCADA (Supervisory Control and Data Acquisition) system.

Data from all sensors and actuators were recorded every second for 11 days. The data is generated in 2 cases: (1) under normal conditions (without attack), and (2) when the attacks are present. A total of 41 attacks have been created, of which 36 have a physical impact on the system. Regarding the location of the attack, they can be simultaneously focused on a single or on multi-points within one or more stages, and each of the attacks is labeled along timeline. Attacks are of different types, duration, intensities, and therefore impact the parts of the system and/or the system as a whole.

In the focus of this paper are anomalies/attacks affecting the LIT101 sensor, a water level sensor on raw water tank, whose location is at the first stage, as shown in Figure 3. On the LIT101 sensor signal attack, the water level in the tank can come out of the permitted range (underflow/overflow), or other damages to the system such as pump P101 damage and UF Feed Tank water level underflow/overflow can occur.

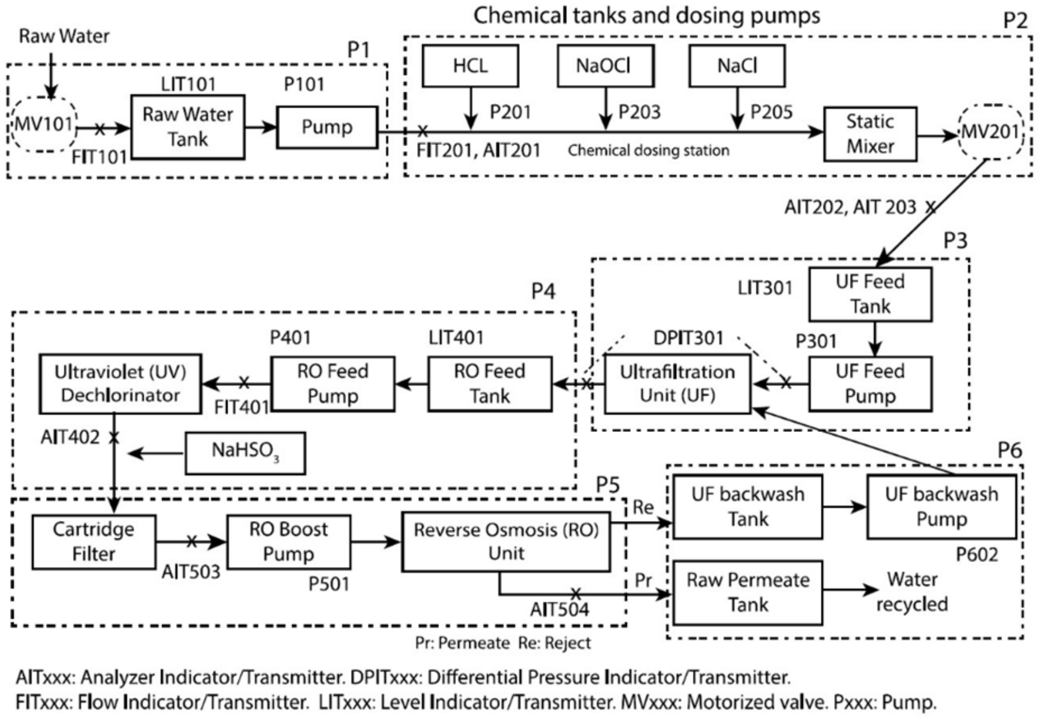

Figure 3. SWaT testbed process overview [13]

4.2 Experimental results

Described method for attacks detection was implemented in MATLAB and tested using LIT101 sensor signal data, a part of which (for the normal system performance) is shown in Figure 4. Before using SWaT data in offline training, it is necessary to cut off the part that relates to the operation of the system before establishing a stable operating mode as presented in red line in Figure 4.

In the case of LIT101 sensor, this means that the tank should get into the continuous mode of charging and discharging. Therefore, the first 15400 data records were removed, leaving a normal operation dataset with a total of 481400 records. 10% of normal operation dataset are used for ε-SVR training. A number of models for different parameters' values were tested. The experiments have shown that the error cost parameter C and kernel parameter γ do not have significant influence on the model accuracy. On the other hand, we have varied the buffer size k in the range of 2 to 10 and the ε between 0.03 and 1. The best result for both criteria (707 support vectors and Δ_{out}=0) was achieved for parameter values ε=1 and k=2.

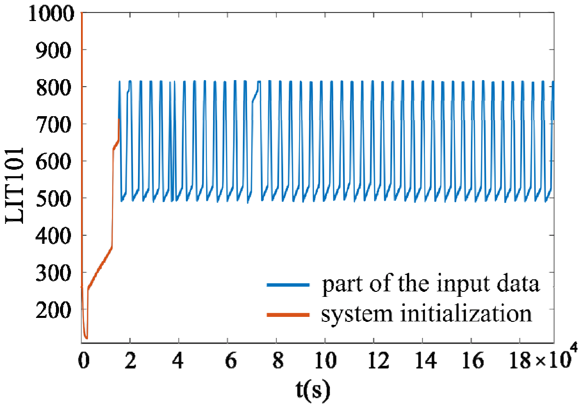

Figure 4. Part of the input data under normal conditions

The online application of the method employing the selected model was simulated using the data obtained during system performance under attack. This dataset contains 449919 records. Our method was able to appropriately detect all five attacks on the LIT101 sensor, as presented in Figure 5. In Figure 5, the signal values for LIT101 from the SWaT dataset during attacks on LIT101, and their predicted (estimated) values are shown in blue and red line, respectively.

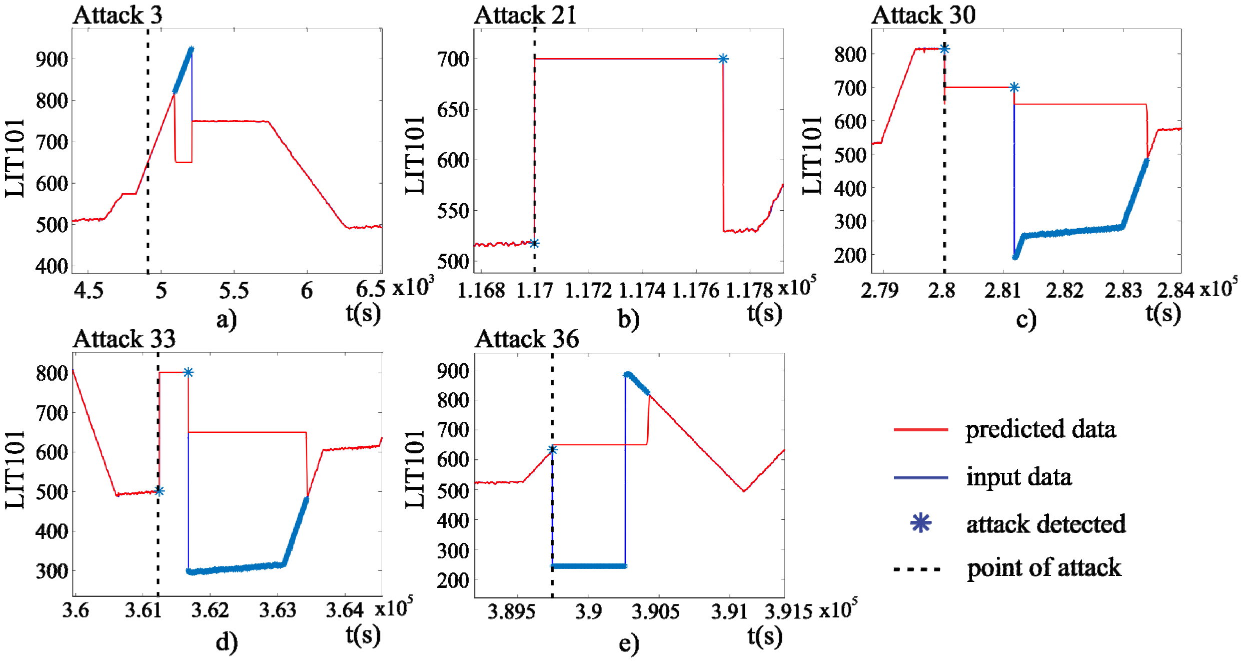

Figure 5. Detected attacks on the LIT101 sensor

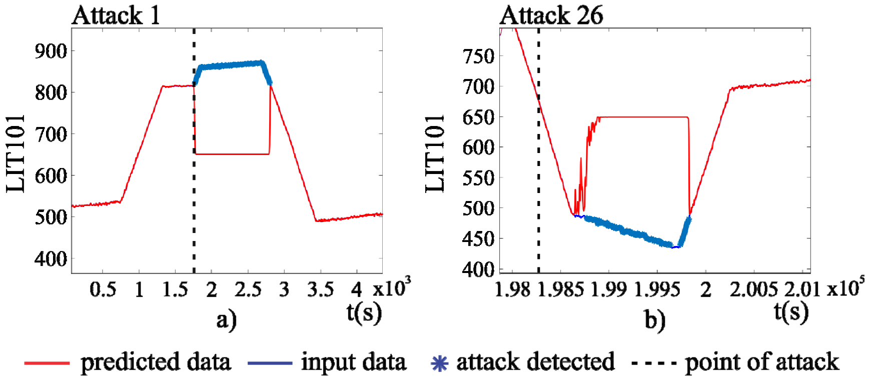

Figure 6. Detected attacks on the adjacent devices

A blue asterix markers indicate the moments in which the condition in (13) is fulfilled, and in which the attack is detected. The set of detected attacks includes[1]: (1) single stage single point attacks on LIT101 (attack 3 – Figure 5a, attack 33 - Figure 5d and attack 36 – Figure 5e), (2) single stage multi points attack 21 on raw water inlet motor valve MV101 and LIT101 (Figure 5b), and (3) multi stage multi point attack 30 on LIT101, raw water pump P101 and feed water tank inlet motor valve MV201 (Figure 5c).

In addition to the LIT101 sensor signal attacks, the method detected two attacks on the adjacent devices that had the influence on the water level of raw water tank (Figure 6).

In particular, single stage single point attack on raw water inlet motor valve MV101 (attack 1 – Figure 6a) and multi point multi stage attack on P101 and LIT301 (attack 26 – Figure 6b) were detected. On the other hand, the method presents false positives neither on dataset with attacks nor on dataset obtained during normal system functioning.

5. CONCLUSSION

This paper proposed a method for sensor signals attack detection in remotely controlled continuous time systems that is based on the prediction of sensor signal values using ε-insensitive SVR. The method is data centric and it is based on ε-SVR model obtained from the sensor signals values acquired during normal system operation. For the evaluation, we used the raw water tank level sensor - LIT101 signals from SWaT testbed dataset. As presented in the paper, the method was able not only to efficiently detect the attacks on the sensor signals, but also some of the attacks on the adjacent actuators, without false positives. It is worth noting that there is a low latency of the detection with respect to the

[1] Attacks are labeled as presented in [13]

moment of the attack occurrence in the signal – it is at the level of 1 sample. During implementation, a number of parameters should be set. In this paper we chose their values by trial and error. A method for optimization of these parameters represents an avenue for future work. Furthermore, future work will address the attacks on actuators as well as the methods for the classification of the attacks according to the point of the attack.

ACKNOWLEDGMENT

The research in this paper was supported by the Ministry of Education, Science and Technological Development of the Serbian Government, Grant TR-35004. This paper was also presented at YOUng ReseearcherS Conference, YOURS 2019, held in Belgrade, where it was chosen as one of the best papers and was awarded with the publication in the journal FME Transactions.

REFERENCES

[1] Atzori, L., Iera, A. and Morabito, G.: The Internet of Things: A survey, Computer Networks, Vol. 54, No. 15, pp. 2787–2805, 2010.

[2] Wang, L., Törngren, M. Onori, M.: Current status and advancement of cyber-physical systems in manufacturing, Journal of Manufacturing Systems, Vol. 37, No. 2, pp. 517–527, 2015.

[3] Kagermann, H., Wahlster, W. and Helbig, J., Recommendations for implementing the strategic initiative INDUSTRIE 4.0, 2013. [Online] Available: http://www.acatech.de

[4] Putnik, G. D., Ferreira, L., Lopes, N. and Putnik, Z.: What is a Cyber-Physical System: Definitions and models spectrum, FME Transactions, Vol. 47, No. 4, pp. 663–674, 2019.

[5] Oliveira e Sá, J., Pereira, J. L. Cacho, J.: Internet of Things evolution: A European Perspective, FME Transactions, Vol. 47, No. 4, pp. 739-748, 2019.

[6] Mitrović, S., Dimić, Z. and Jakovljević, Ž.: Distributed control of manufacturing resources - security related issues, in: *Proceedings of MMA 2018 Conference*, Sep. 2018., pp. 195-198

[7] Umer, M. A., Mathur, A., Junejo, K. N. and Adepu, S.: Integrating Design and Data Centric Approaches to Generate Invariants for Distributed Attack Detection, in: *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy*, Nov. 2017., pp. 131–136

[8] Pasqualetti, F., Dörfler, F. Bullo, F.: Attack Detection and Identification in Cyber-Physical Systems, IEEE Transactions on automatic control, Vol. 58, No. 11, pp. 2715–2729, 2013.

[9] Kravchik, M. Shabtai, A.: Detecting Cyber Attacks in Industrial Control Systems Using Convolutional Neural Networks, in: *Proceedings of CPS-SPC 18 conference*, Oct. 2018., pp. 72-83

[10] Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C. M. and Sun, J.: Anomaly detection for a water treatment system using unsupervised machine learning, in: *Proceedings of IEEE International Conference on Data Mining*, Nov. 2017., pp. 1058–1065

[11] Hadžiosmanović, D., Sommer, R., Zambon, E. and Hartel, P. H.: Through the eye of the PLC: semantic security monitoring for industrial processes, in: *Proceding of 30th Annual Computer Security Applications Conference*, Dec. 2014., pp. 126–135

[12] Sokolov, A. N., Pyatnitsky, I. A. and Alabugin, S. K.: Applying methods of machine learning in the task of intrusion detection based on the analysis of industrial process state and ICS networking, FME Transactions, Vol. 47, No. 4, pp. 782-789, 2019.

[13] Centre for Research in Cyber Security, Singapore University of Technology and Design: Secure Water Treatment (SWaT), Available: http://itrust.sutd.edu.sg/research/testbeds/secure-water-treatment-swat/

[14] Smola, A. J. and Schölkopf, B.: A tutorial on support vector regression, Statistics and Computing, Vol. 14, No. 3, pp. 199–222, 2004.

[15] Goh, J., Adepu, S., Junejo, K. N. and Mathur, A.: 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*, Oct. 2016., pp. 88–99

ДЕТЕКЦИЈА НАПАДА НА СЕНЗОРСКЕ СИГНАЛЕ У ИНДУСТРИЈСКИМ КОНТРОЛНИМ СИСТЕМИМА

Д. Недељковић, Ж. Јаковљевић, З. Миљковић

У циљу повећања продуктивности и ефикасности производње, четврта индустријска револуција води ка имплементацији кибернетско физичких система и интернета ствари у индустријском окружењу. Свеобухватна комуникација чини кибернетско физичке системе подложним на спољашње утицаје, који често могу имати негативну намеру, нпр. напади и сметње проистекли од различитих узрочника. Утицај напада на систем може довести до аномалија и озбиљних последица по делове система или систем у целости. Стога, одбрамбени механизми за правовремену детекцију напада морају бити развијени, како би се систем заштитио и одржала његова функционалност. У овом раду, предложен је метод за детекцију напада на сензорске сигнале у континуално управљаним системима. Метод је базиран на машинама са носећим векторима, а тестиран на скупу података из система за прераду воде.