Detection of Cyber-Physical Attacks using Physical Model with Nonparametric EWMA Detector

Joko Supriyadi, Jazi Eko Istiyanto, Agfianto Eko Putra
Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia

***Abstract*—Industrial Control System (ICS) can suffer of cyber-physical attacks resulting in accident, damage, or financial loss. The attacks can be detected in both in physical space or cyberspace of the ICS. The detection in physical space can be based on physical models of the system. To model the physical system this study uses a data-driven modeling approach as an alternative of the analytic one. This study models the system using the dynamic mode decomposition method with control (DMDc) assuming a full state measurement. The attack detector used in some researches with predictive physical models is the cumulative sum (CUSUM), which only applies to normally distribute residual data. To detect any cyber-physical attack, this research uses a nonparametric exponentially weighted moving average (EWMA) detector. This study uses a data set from a testbed of Secure Water Treatment (SWaT). The approach used in this study was successful in detecting 8 out of 10 attacks on the first SWaT subsystem. This study demonstrates that DMDc used in this study results a better goodness of fit and the nonparametric EWMA can be used as an alternative as detector when residual data do not follow a normal distribution.**

Keywords*—Industrial control systems; cyber-physical attacks; physical model; dynamic mode decomposition method with control (DMDc); nonparametric exponentially weighted moving average (EWMA)

I. Introduction

Industrial Control System (ICS) is automation system used to control and monitor industrial functionality. It can be found in the industrial sectors and critical infrastructures, such as nuclear and thermal plants, water treatment facilities, power generation, heavy industries, and distribution systems [1].

ICS has four levels to conduct its functionalities [2]. On the field level, it has sensors and actuators. On the control level, it has a programmable logic controller (PLC). It has a human-machine interface (HMI), and on the enterprise level, it has an information technology (IT) system. The second and third levels with their cyber components conduct computation to control physical space. With this trait, ICS is an example of cyber-physical systems.

As a cyber-physical system, ICS carries the risk of cyberattacks aimed to disrupt the physical processes. A cyber-physical attack is an attack via cyberspace that targets the physical system of the ICS. Examples of such attacks are Stuxnet which attacked Iran's uranium enrichment facilities, and Black Energy 3 which struck several power transmission substations in Ukraine [3]. These attacks can have some adverse effects on factory safety or financial loss.

To detect cyber-physical attacks on ICS, researchers develop intrusion detection systems (IDS). There are three IDS categories for ICS [4], protocol analysis-based, traffic mining-based, and control process analysis-based. The former two categories are conducted in cyberspace. The third category is conducted on physical space and includes process data analysis-based, control command analysis-based, and physical model-based techniques [4].

This study is in the field of detection in physical space using a physical model, i.e. the third category in [4]. The work in [5] identify that most of researches in this field do not use input-output approach in modelling physical behavior and most use the simulation data. This research proposes to use dynamic mode decomposition method with control (DMDc) as a system identification with input-output method. It will be applied to a real data from a testbed not at simulation data.

To detect the anomalies, some researches use cumulative sum (CUSUM) that has an assumption that the distribution is normal. In this research we propose to use nonparametric exponentially weighted moving average to cover all distribution condition.

The significant contributions of this work are:

- We anticipate a full state measurement condition of physical systems. To model it, we use DMDc as a kind of a system identification approach.
- We use EWMA detector as an alternative for CUSUM detector to cover a non-normal distribution of residual.

The rest of the paper is organized as follows. In section 2 related works are presented. It posits our research in the field of detection of cyber-physical in physical space of ICS. Section 3 presents the research methods of this study. It covers the DMDc method and EWMA detector. In section 4 the results of the experiment are presented and discussed.

II. Related Works

According to Urbina *et al*. [5], there are two ways to model physical systems. The first is to develop a physical equation model that connects all the physical parameters to determine the system's dynamics. The second is to create a model based on observation through a technique known as system identification. Most researchers use the first approach, and only a few uses the second.

Researchers in [16] use input-output model developed from physical equation for a water treatment system. Their research leverages connectivity of two subsystems to detect

any cyber physical attack to upstream with behavior of the downstream of the system. Their experiment is launched directly to a real system and their procedure was conducted at the real time.

Our study is focused on the use of system identification and not model the system using physical equation. There are two options in system identification, output only model and input-output model. Researchers in [5] used both approaches to model a physical system. The system is a simulator of frequency control in the power grid. After comparing both models, researchers in [5] concludes that the input-output model has a potential in detection and motivates future researchers to use it to model physical systems. To detect any cyber-physical attacks, there is a need to model the physical model from measurement data in the form of input-output model.

Researchers in [6] used the system identification method to model a physical system with the input-output model. They used a method in a group of subspace models. They used EPANET, a software that can customize a water distribution system simulator as the physical system.

Other researchers [7] conduct modelling with subspace system identification with input-output model with CUSUM detector, bad data detector, and noiseprint. To detect the anomaly, they used CUSUM detector with false positive target around 5%.

Usually, the system identification approach used in [5,9] do not assume that the system being modeled is in full measurement. But nowadays, with the affordability of measurement cost, the system status can be measured in full. Unlike the studies, our study proposes using dynamic mode decomposition with control (DMDc) to model the physical system, assuming a complete system measurement.

The author in [5] used simulations as physical systems to be modeled. But there is a problem with simulations; although the physical model is detailed but the simulation is not in a real situation. Unlike theirs, our study uses a data set from Secure Water Treatment (SWaT), a physical system testbed that mimics a real water treatment system.

Some researchers use the same data set to propose their method to detect cyber-physical attacks. One of them [8] uses a graphical model-based approach that will be compared with our work.

Based on the physical model, cyber-physical attacks can be detected by monitoring an anomaly in the systems. The monitoring systems get inputs from sensors and command controls and then identify any anomalies [6]. Some researchers use CUSUM detector or bad data detector [5, 6, 7, 9, 16]. But the use of both as detector applies only to data with a normal distribution. Because not all conditions meet the normal distribution, our study proposes to use a nonparametric exponentially weighted moving average (EWMA) detector that does not depend on data distribution.

III. Research Method

In this section, we will explain the methods we used in our research. It includes details of DMDc, nonparametric EWMA, and a brief explanation of steps to apply the methods.

A. Physical System Modelling

A discrete state-space model will be used to represent the ICS control system. This model is used because it is generally more compatible with discrete digital controls. For discrete-time systems, the time may be referred as an integer index k= 0, 1, 2, ... In the system, the state as a set of variables from the system can be estimated based on the previous state. Given a linear system has several states, inputs, and output, respectively. The equation of the system can be modeled as shown in (1) and (2).

$$xk+1=Axk+Buk \quad (1)$$

$$yk+1=Cxk \quad (2)$$

In (1) and (2), $x_k \in R^n$ is the state of the system at time k, $u_k \in R^l$ is the input, and $y_k \in R^q$ is the output with n, l, and q as number of states, inputs, and outputs respectively. Meanwhile, A is a dynamic matrix, B is the control matrix, and C is the discrete-time sensor matrix. State vector x from the system can be estimated from sensor measurement y. There is a condition called full state measurement if all components of the state vector can be measured.

The system can be modeled with a system identification approach, namely by using measurement data to obtain the desired model. In this study, the dynamic mode decomposition (DMD) method [10] is used, assuming that measurements are made of all states of the system (full-state measurement).

Initially, DMD was developed for systems that do not have input, for example, control input. By taking into account control input, DMD method developed to be DMD with control (DMDc). Both DMD and DMDc methods initially were developed for large dimension systems with a dimension reduction. The next exposition will explain DMDc method based on [10] but without dimension reduction steps.

DMDc analyses the relationship between measurement vectors at $k+1$ (x_{k+1}) with the measurement at k (x_k) and control input at k (x_{k+1}). The three data are assumed to be approximately by linear operators A and B relates as shown in (3).

$$\mathbf{x}_{k+1} \approx \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \quad (3)$$

where $x_k \in R^n$, $u_k \in R^l$, $A \in R^{nxn}$, and $B \in R^{nxl}$ with n and l are number of sensors and actuators respectively. DMDc tries to find the solution for operators A and B.

If there are m measurements for n sensors in vector x we can create a matrix X with dimension $n \times m$, $X \in R^{nxm}$. We can create matrix Y and Z from matrix X, where Y is matrix X without its first column, while Z is matrix X without its last column. The relationship of matrix Y and Z with matrix A can be described with (4).

$$Y \approx AZ \quad (4)$$

Accordingly, from control vector u, we can create a matrix $\Gamma \in R^{lx(m-1)}$ contains the *1st* data to *m-1th* data for *l* actuators. So (4) can be rewritten to be (5).

$$Y \approx AZ+B\Gamma \quad (5)$$

A concatenation method can be used to assign matrix G = [A B] and matrix $\Omega = \begin{bmatrix} Z \\ \Gamma \end{bmatrix}$ to solve the (5). Thus, (5) changes to be (6) as follows.

$$Y \approx G\Omega \quad (6)$$

Then Y and Ω is stated as eigen decomposition of the matrix G as shown in (7).

$$G \approx Y\,\Omega^{\dagger} \quad (7)$$

Where † means Moore-Penrose pseudoinverse. The accurate and efficient method to find the pseudoinverse is singular value decomposition (SVD), as shown in (8).

$$\Omega = U\Sigma V^{*} \quad (8)$$

where $U \in R^{nxn}$, $\Sigma \in R^{(nxm)-1}$ and $V^{*} \in R^{(nxm)-1}$.

Using the SVD of Y, the matrix G can be approximated with (9).

$$G \approx \overline{G} = Y V \Sigma^{-1} U^{*} \quad (9)$$

where $\overline{G}$ is an approximation of the matrix G. From the matrix $\overline{G}$, the approximation of matrix A and B can be obtained by breaking operator U into two separate components given by (10)

$$[A, B] \approx [\overline{A}, \overline{B}] = \left[Y V \Sigma^{-1} U_1^{*}, Y V \Sigma^{-1} U_2^{*} \right] \quad (10)$$

where $U_1 \in R^{nxn}$, $U_2 \in R^{lxn}$, and $[U^{*}] = [U_1^{*}\, U_2^{*}]$. From (10), the dynamic systems can be constructed to be (11) as follows.

$$x_{k+1} = \overline{A}x_k + \overline{B}u_k \quad (11)$$

Equation (11) is the model we use to mimic the behavior of the physical system.

B. Residual and Non Parametric EWMA

Cyber-physical attacks on industrial control systems have a potential to change sensor readings behavior. Therefore, the difference between the estimates obtained from the model and the sensor readings can be monitored. The behavior of to the difference can be used to indicate if any attack to the system. The difference is called the residual r_k, namely the difference between measurement at *k*, x_k, and the estimation $\hat{x}_k$ as shown in (12).

$$r_k = x_k - \hat{x}_k \quad (12)$$

The value of estimation $\hat{x}_k$ is determined by (11); thus, (12) can be expressed as (13) as follows.

$$r_k = x_k - (Ax_{k-1} + Bu_{k-1}) \quad (13)$$

The sensors of the system can be more than one, so with $i=\{1,\ldots, n\}$, $r_{k,i}$ is residual of sensor *i* at measurement *k*. Murguia *et al.* [9] used individual sensor residual to monitor cyber-physical attacks by making them input for CUSUM detector and bad data detector. The two detectors implicitly assume that the residual follows a normal distribution.

In this study we use a nonparametric Exponentially Weighted Moving Average (EWMA) as the residual does not follow a normal distribution. This approach is generally used to statistically monitor a product's quality or process [11, 12]. The next description will explain the approach as describe in [12].

Given X_k is the individual measurement from the continuously monitored distribution with median θ, *SN* as statistic sign can be provided by (14).

$$SN_k = \text{sign}\,(X_k-\theta) \quad (14)$$

where:

$$sign(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x = 0 \\ -1 \text{ if } x < 0 \end{cases} \quad (15)$$

The plotting statistic (Z_k) of the nonparametric EWMA as a sequentially accumulating the sign statistic SN_1, SN_2, SN_3,…defined by equation (16).

$$Z_k = \lambda SN_k + (1-\lambda)Z_{k-1} \quad (16)$$

where λ is a weighting that has a value from 0 to 1, and Z_0=0.

Two controls can be used for the value of Z, upper control limit (UCL) and lower control limit (LCL), that can reach its steady-state after a long time. They can be determined by using (17).

$$UCL = L\sqrt{\frac{\lambda}{2-\lambda}}$$
$$CL = 0 \quad (17)$$
$$LCL = -L\sqrt{\frac{\lambda}{2-\lambda}}$$

In (17) CL is the central line valued as 0. The values can be used to determine if the system is under control or not.

The choosing of parameters *L* and λ is conducted to get the targeted average run length (ARL), namely the number of average measurements before the detector detects an anomaly from a normal condition. Graham [11] has calculated the value of pair parameter *L* and λ for nonparametric EWMA with the Markov method until ARL=500.

We need to calculate the alarming rate that gives us the number of Z_k values calculated by (16) exceed UCL or LCL for a while. Then we set Z_k=0 if $Z_{k-1} \geq$ UCL or $Z_{k-1} \leq$ LCL.

C. Detection Parameter

We assume there is difference behavior between an attack condition with the normal one. In this research we use comparison of false alarm rate (FAR) or false positive rate (FPR) value between both conditions.

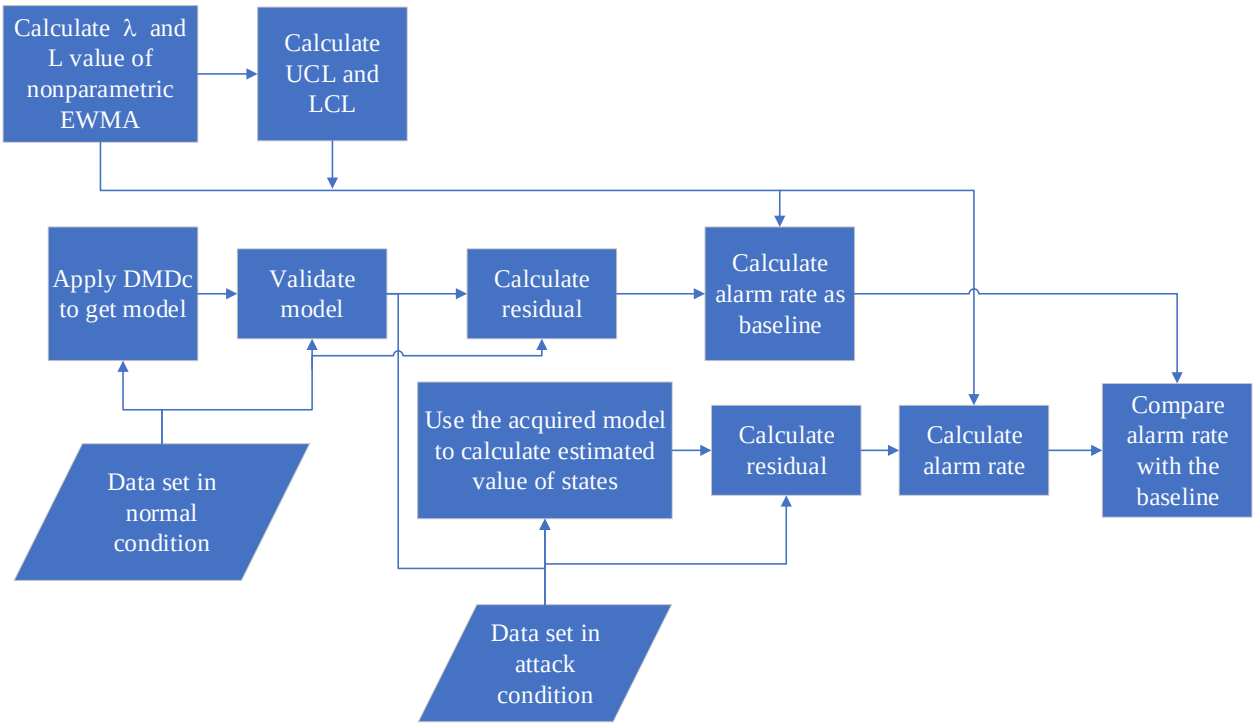

Fig. 1. Flow Chart of Detection System Set up.

The value of FPR is targeted in normal condition and considered as alarm rate of the system or components. Using the value as the baseline we judge a condition is under attack if FPR of a component is significantly more than its normal condition. With its procedure, security operator monitoring the system can detect any potential attack if the alarm rises significantly than the baseline.

D. Research Steps

Before using DMDc to model the physical system, we calculate L and λ values for specified ARL, for example, ARL=500 if we expect to get one false alarm per 500 measurements. The values were then used to calculate the value of UCL and LCL.

Parallel with previous steps; the physical system is modeled using DMDc method with the input of the normal condition data. Then the acquired model is validated by calculating its goodness of fit. If the value is satisfying, we calculate residuals as differences between model estimation and measurement. Based on the value of the residual, L, λ, UCL, and LCL, we calculate the value of Z_k. The alarm rate baseline is calculated based on the number of Z_k values that exceed UCL or LCL.

The next step is to use the model to detect cyber-physical attacks. The following procedure will be applied for every attack separately. With the acquired model from the previous steps, residuals' value is obtained by applying DMDc method on the data set in attack condition. Like previous steps in normal conditions, the value of alarm rate can be obtained using a parametric EWMA method.

To judge if the concerned attack is detected, we compare the alarm rate baseline's value of the alarming rate. We believe that alarm is detected if the alarming rate is ten or more from the baseline. Fig. 1 shows the flowchart of the research steps.

IV. Result and Discussion

In this section, we will discuss our research results based on steps explain in the previous section.

A. Physical Modelling

This study uses the data set collected from a testbed that simulates Secure Water Treatment (SWaT) [13,14]. SWaT is a testbed in a full-scale replica water treatment facility consisting of several subsystem stages. A data set containing network data and process data is generated [13] recorded every second. The data set consists of two parts:

1) Data set of the system in a normal state without attack,
2) The data set of the system under attack condition.

The SWaT system has six subsystems, as described in [13], from subsystem P1 to P6. For this research, we use the first subsystem, P1 only, namely the subsystem that takes raw water and stores it in a reservoir. This subsystem has several main components [13], namely an MV-101 valve that drains water from the source into the storage tank T.101, whose water level is measured by the sensor LIT-101. Meanwhile, a flow meter FIT-101 is used to measure the source's flow rate into the tank. A pump P-101 is used to drain water from the reservoir to the next subsystem. Also, this subsystem has a pump P-102 as a backup for the pump P-101. Fig. 2 [13] shows the components and piping of the subsystem P-1. Thus, there are two sensors and three actuators in subsystem P1 as shown in Table I.

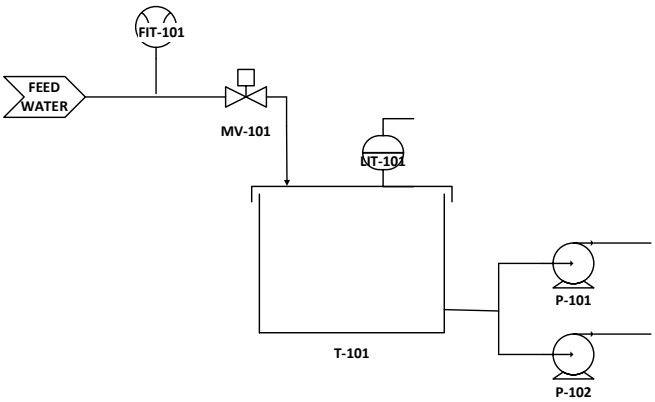

Fig. 2. Subsystem P1 of SWaT [13].

TABLE I. COMPONENTS OF THE SUBSYSTEM P1

Component	Type	Units
Flow rate FIT-101	Sensor	m3/h
Level meter LIT-101	Sensor	mm
ValveMV-101	Actuator	On/off
Pump P-101	Actuator	On/off
Pump P-102	Actuator	On/off

We use DMDc to model the subsystem P1 in the form of the discrete linear model provided by (11). From (11), the vector x contains subsystem P1 as a dynamic variable following the dynamics of water supply and demand for raw water. Thus, the vector x includes the measurement of the LIT-101 sensor. Meanwhile, the vector u in (11) contains the input into the subsystem. Among these inputs is the water supply, whose debit is measured using the FIT-101 sensor. Another information model for subsystem P1 is the pump P-101 pump and its pump P-102 as a backup.

The two pumps are modeled because the two pumps can be turned on alternately. The MV-101 valve is not modeled as it is redundant to the FIT-101. When the MV-101 valve is turned on, water flows with the discharge measured by the FIT-101 sensor. Thus, in (11), subsystem P1 can be modeled in its discrete form with (18).

$$[\mathrm{LIT}-101_{k+1}] = \bar{\mathbf{A}}\,[\mathrm{LIT}-101_k] + \bar{\mathbf{B}}\begin{bmatrix}\mathrm{FIT}-101_k \\ \mathrm{P}-101_k \\ \mathrm{P}-102_k\end{bmatrix} \quad (18)$$

In this subsystem modeling, matrix $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are matrices that will be sought with the DMDc method by using a dataset in normal conditions without attack as input. Data set for normal conditions without attack contains measurement results of sensors and actuator conditions that are monitored every second. There are 494999 pieces of datum for each sensor and actuator. The first 3.5 hours of data are not used to let the system achieves a steady-state condition. The remainder is divided into two, 70% data for modeling and 30% data for testing.

Start from (5), the matrix Y is obtained as sensor measurement data from the 2nd to m^{th} measurement, the matrix Z the sensor measurement data from the 1st to $m\text{-}1^{th}$ measurement, and the matrix Γ as the actuator data from the 1^{st} to $m\text{-}1^{th}$ measurement. Furthermore, by following the procedure as described in (6) to (10), the values of a matrix $\bar{\mathrm{A}}$ and $\bar{\mathrm{B}}$ in (18) is obtained as follows.

$$\bar{\mathbf{A}} = [1.000]$$

$$\bar{\mathbf{B}} = [0.195 \;\; -0.461 \;\; 0.392]$$

With the values of the matrix $\bar{\mathrm{A}}$ and $\bar{\mathrm{B}}$, the value of the inflow of water from the source and the water level at $k + 1$ can be estimated based on the measurement value at time k. The comparison of the level value of water in the tank measured by the LIT-101 sensor with the model's estimation results is shown in Fig. 3.

The estimation results are then compared with the actual measurement value to determine the model's level of suitability (goodness of fit) with the measurement results. Based on calculations using data for testing (30% of the overall data), the goodness of fit of the sensor LIT-101 is 99.7%. The value of goodness of fit is better than result in [6] that use another method of system identification. They use simulation data from EPANET, software to simulate a water system. With their method they achieve goodness of fit of 70%.

The value of goodness of fit shows a good agreement between model and measurement. Since the data set to test the model is independent of the training data set, it can be concluded that the model is not overfitted to the training data.

Meanwhile, Fig. 4 shows the error fluctuation in the form of the difference between the measurement results and the model estimation results. The error fluctuation is calculated with (13) as a residual that will be used to monitor the anomalies of subsystem P1.

B. Nonparametric EWMA

The residual as the difference between the estimation results from the model and the measurement at each time k resulted has a non-normal distribution. Therefore, the use of the CUSUM detector cannot be legitimized, and thus the nonparametric EWMA detector as described previously is used in this study.

The EWMA detector has a parameter λ and L as shown in (16) and (17). For this study we choose ARL_0=500. For this ARL value we choose a low value of λ (λ=0.01) to get a smooth change of Z_k sequences. For this value λ, Graham [11] get value of L = 1.990 as his calculation with Markov method.

With the value of L and λ, we calculate the control limits for UCL and LCL for a normal condition of their residual value of sensor LIT-101 with (17). The acquired values for UCL and LCL are 0.141067 and -0.141067 respectively.

With the acquired value of the parameters, the value of Z is sequentially calculated by (14). The value of Z_k fluctuates between the value of UCL and LCL as shown in Fig. 5. When it hits one of them the next value will be set to zero and the sequence starts again.

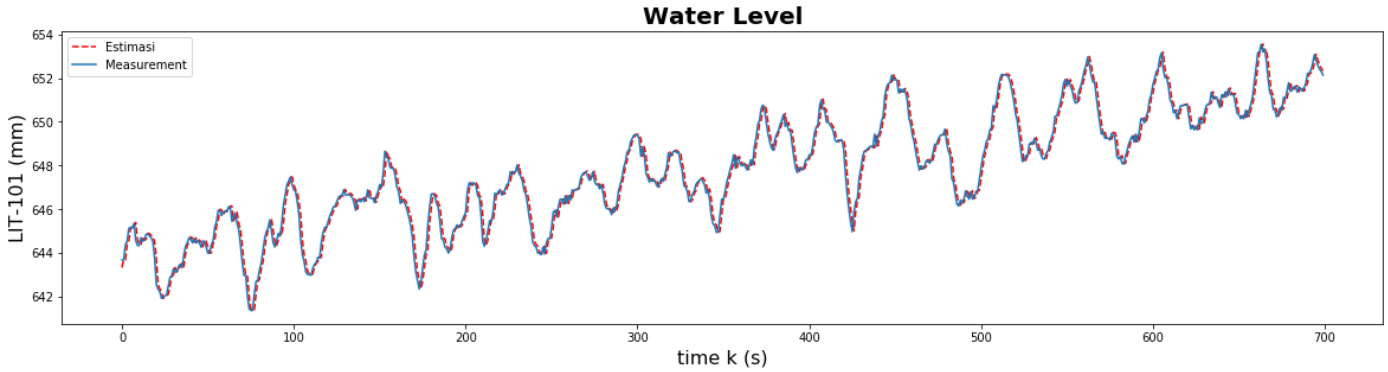

Fig. 3. The Dynamics of the Water Level from Measurement and Estimation.

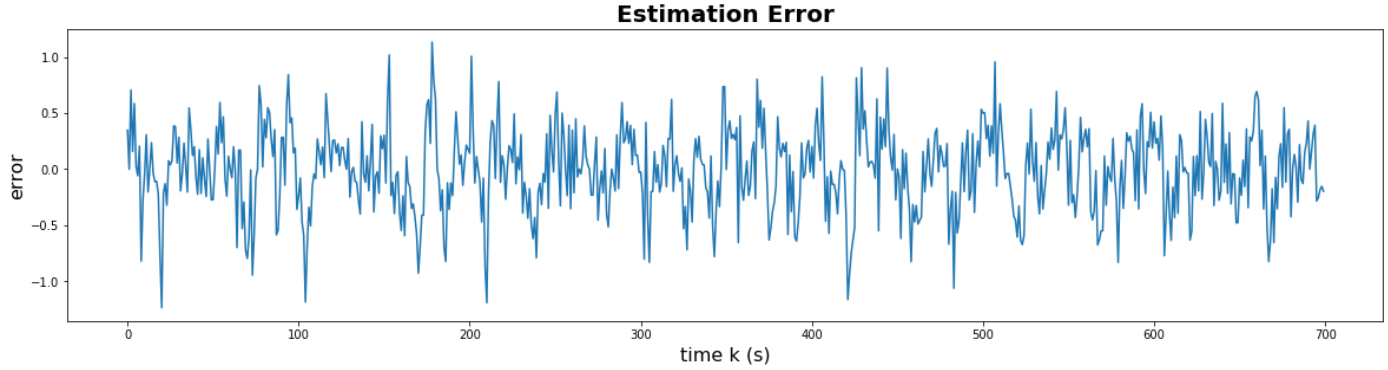

Fig. 4. Error Fluctuation of the Estimation of Water Level.

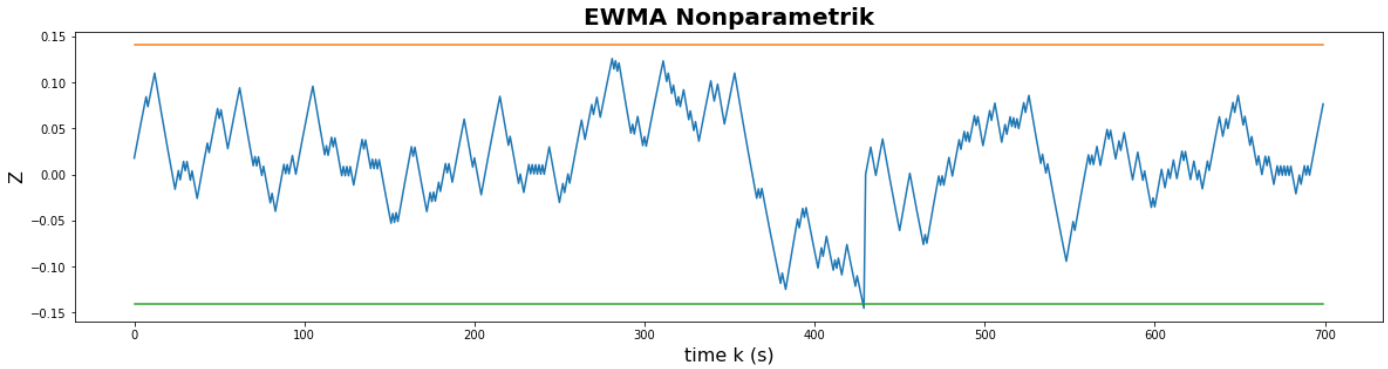

Fig. 5. The Fluctuation of Value Z with UCL and LCL as the Control.

The Z_k value passes UCL and LCL with a ratio of 1 in 489 data (alarm rate equals 0.00204). In other words, about 2.2% deviations from the ARL calculation theoretically using the Markov method of ARL_0=500 as the target. From the perspective of detection, the ratio is an expected false alarm rate in a normal condition. We use the value of the ratio as the baseline where the alarm rate of attack conditions will be compared, as we will explain in the next subsection.

The benefit of using nonparametric EWMA than CUSUM as in [5, 6, 9] is there is no need to assume the distribution of residual *r* as normal. As shown in the next subsection, this approach can be used to detect the cyber-physical attack as other detector.

C. Detection of Cyber-physical Attacks

To test the proposed detection design, we also use data set in [14] that contains sensors and actuators data from the SWAT under attack conditions. The attacks result in the data set use attack and attacker models developed by Adepu and Mathur [15]. The attacker model relates to the attacker's intention with the systems' components, properties, and performance.

The attack model captures the space of potential attacks aimed at achieving a specific set of goals. For example, an attack has a plan to make a tank overflow. The attack needs to consider some points of the system involved, for example, pumps or valves. A suitable procedure to launch the attack consists of identifying the tag in the programmable logic controller (PLC) that should be manipulated and a step to

compromise the link between PLC and supervisory control and data acquisition (SCADA) a step to conduct the manipulation of the tag.

There were 41 attack scenarios launched to the SWaT testbed, but only 36 attacks had physical impacts and were recorded in the data set. From the 36 attacks, there were 10 attacks involving the subsystem P1 in the data set as shown in Table II. All of the attack scenarios were launched to subsystem P1 only, but scenario 26 launched to subsystem P1 together with subsystem P3.

The system model and the nonparametric EWMA detector obtained from the previous section are used to detect cyber-physical attacks based on the data set [12]. The model is used to estimate the LIT-101 sensor reading value at time $k + 1$ with the time k. The estimation result is compared with the sensor reading value, which produces a residual value r_k. Then the residual value is used to detect an attack with an indication if the Z_k value exceeds the UCL or LCL values.

The Z_k value is calculated using (16) and (17) for each attack as described in Table II. As an illustration, Fig. 6 show the behavior of Z_k when subsystem P1 is under attack number 2 condition. Differ from the normal condition as shown in Fig. 5 that show a random behavior, the value of Z_k in attack no 2 as shown in Fig. 6 is systematically pushed to hit its lower threshold LCL. The change of behavior is an indication that the system is not in normal condition and the cyber-physical attack is a probable cause. As explained in the previous subsection, the baseline of false alarm rate (FAR) or False Positive Rate (FPR) for sensor LIT-101 calculated in normal conditions based on the normal condition dataset is 0.0020 (0.2%). To judge if any attacks can be detected, we compare the alarming rate in any attack conditions with the baseline.

TABLE II. DESCRIPTION OF ATTACKS SCENARIO

No.	Attack Number	Description of the attack	Duration (s)
1	1	Open MV-101	940
2	2	Turn on P-102	1407
3	3	Increase LIT-101 by 1 mm per second	383
4	21	Keep MV-101 on continuously; Value of LIT-101 set as 700 mm	701
5	26	P-101 is turned on continuously; a set value of LIT-301 as 801 mm	1445
6	30	Turn P-101 on continuously: Turn MV-101 on continuously; Set value LIT 101 as 700 mmP-101	1171
7	33	Set LIT-101to above H	444
8	34	Turn P-101off	101
9	35	Turn P-101off; keep P-102 off	481
10	36	Set LIT-101 less than LL	474

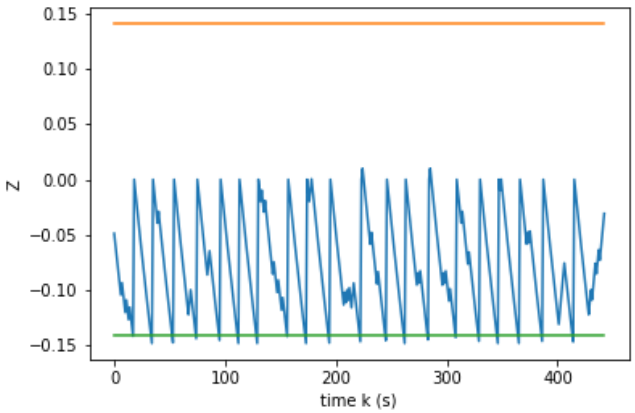

Fig. 6. The Fluctuation of Value *Zk* when System is under Attack no.2 Condition.

TABLE III. ALARM RATE FOR ATTACKS INVOLVING SUBSYSTEM P1

No.	Attack Number	Alarm rate (compared to baseline)	Judgment	Detection in [8]
1	1	0.0021 (1.40)	Not detected	Detected
2	2	0.0452 (22.09)	Detected	Detected
3	3	0.0470 (22.99)	Detected	Not detected
4	21	0.0585 (28.62)	Detected	Detected
5	26	0.0505 (24.72)	Detected	Detected
6	30	0.0564 (27.58)	Detected	Detected
7	33	0.0495 (24.24)	Detected	Detected
8	34	0.0297 (14.53)	Detected	Detected
9	35	0.0 (0)	Not detected	Detected
10	36	0.0570 (27.89)	Detected	Detected

We choose if the alarm rate is more than 1 order (multiply by 10) compared to the baseline rate of 0.0020. The choice is based on consideration that person operate and monitor the security system can detect the anomaly in 1 order. Based on this procedure, 8 from 10 attacks can be detected, as shown in Table III. The alarm rates for detected attacks as shown in Table III are 14 to 28 times greater than the baseline of 0.0020.

The results show that our approach can detect all attacks involve sensors as targets (Scenario 3, 21, 26, 30, 33, and 36). This success can be understood because our approach monitors the value of water level. Thus, the discrepancy between sensor reading and model estimation can increase or decrease the Z value systematically.

The other four attacks (scenarios 1, 2, 34, 35) involve actuators only. It is just two (scenarios 2 and 34) of the four attacks that can be detected. Attack scenario 1 opens valve MV-101 when it should be close. The attack cannot be

detected because the level sensor will respond normally with raw water flow. The other three attacks involve one or both pumps. The three attacks cannot be interpreted easily because of a long normal condition pump P-102 as a backup was never turned on.

It is to be noted that the attacks duration is vary from around 1 to 3 times ARL as shown in Table I, except the attack no. 34 is only 100 seconds. Because of the random nature of the residual *r*, it can be understood that the alarm rate of attack no. 35 is zero. But for the other attacks that our set up can detect them their short durations does not affect the detectability.

As a matter of comparison, Table III contains the research results [8] that used the same SWaT data set but with a different approach. The research uses a novel graphical model-based approach for anomaly detection. From the table, it can be concluded that their system is better than ours to detect attacks on actuators. But their system cannot detect scenario three that ours can detect. The scenario can be under the radar of detection because it increases the reading of the sensor gradually. Based on the consideration our approach has a promising result to be leveraged to detect stealthy attacks.

V. Conclusion

In this study, the physical modeling is successfully carried out using the DMDc method with a goodness of fit of 99.7%. It shows that the model has a good agreement with the measurement. The high value of goodness of fit gives the model a potential to be used to detect any anomaly caused by cyber-attack deviating the sensor measurement or actuator signal.

The difference between the model's predictions and the actual sensor measurement results is monitored with a nonparametric EWMA detector to detect anomalies resulting from cyber-physical attacks. From 10 attacks conducted to the subsystem, 8 of them can be detected using method used in this research. Compared with the baseline in normal condition, the alarm rates of detected attacks are 14 to 28 times greater. It shows that this method is successful in detecting most attacks, especially on sensors.

As observed in [5] there is a lack of use of input-output models in security field. Besides that, most of researches in security using physical-model use physical equation and not many researches use data-driven method. The other lack is use of real data or data drawn from testbeds, and mostly use simulation data. Our research fills the lacks by using DMDc as input-output model and it is a kind of system identification using data to model the physical behavior of system. Our research also uses the data collected from a test bed.

When we start our research, we plan to use CUSUM as the detector to detect anomalies as proposed in [5, 6, 9]. But we find that in our research the distribution of residual is not normal. Then the nonparametric EWMA is used to detect the anomalies. This research show that DMDc as a system identification combined with EWMA as detector can be used to detect cyber-physical attacks.

In future, this research can be continued by utilizing the interrelationships between subsystems. In this case, attacks on the upstream subsystems may be detected by the downstream subsystems. With the approach the detection probability may be increased because of the double detection, in the subsystem itself and in its downstream. Differ from [16] we use system identification method with DMDc instead of physical equation.

References

[1] Deval Bhamare , Maede Zolanvari , Aiman Erbad , Raj Jain , Khaled Khan, Nader Meskin , “Cybersecurity for Industrial Control Systems: A Survey”, Computers & Security, Vol. 89, February 2020.

[2] M., Krotofil, K., Kursawe, and D. Gollmann, “Securing Industrial Control Systems”, in C. Alcaraz, Security and Privacy Trends in the Industrial Internet of Things, Springer, Switzerland, pp. 3-27, 2019.

[3] R. M. Lee, M. J. Assante, T. Conway, “Analysis of the Cyber Attack on the Ukrainian Power Grid”, SANS, Washington, DC, 2016.

[4] Y. Hu, A. Yang, H. Li, Y. Sun, and L. Sun, “A survey of intrusion detection on industrial control systems,” International Journal of Distributed Sensor Networks, vol. 14, no. 8, 2018.

[5] D. I., Urbina, J. Giraldo, A. A. Cardenas, J. Valente, M. Faisal, N. O., Tippenhauer, J. Ruths, R. Candell, and H. Sandberg, “Survey and New Directions for Physics-Based Attack Detection in Control Systems”, National Institute of Standards and Technology (NIST), US Department of Commerce, 2016.

[6] C.M. Ahmed, C. Murguia, and J. Ruths, 2017, “Model-based Attack Detection Scheme for Smart Water Distribution Networks”, In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security - ASIA CCS ’17, Abu Dhabi, United Arab Emirates, pp. 101–113, 2017.

[7] Surabhi Athalye, Chuadhry Mujeeb Ahmed, and Jianying Zhou, “A Tale of Two Testbeds: A Comparative Study of Attack Detection Techniques in CPS”, In: Rashid, A., Popov, P. (eds) Critical Information Infrastructures Security, CRITIS 2020, Lecture Notes in Computer Science, vol 12332.

[8] Q. Lin, S. Adepu, S. Verwer, and A. Mathur. “TABOR: A Graphical Model-based for Anomaly Detection in Industrial Control Systems”, In: Proceedings of the 2018 Asia Conference on Computer and Communications Security, Incheon, Republic of Korea, pp. 525-536, 2018.

[9] C. Murguia and J. Ruths, “Characterization of a CUSUM Model-Based Sensor Attack Detector”, In Proceedings of IEEE 55th Conference on Decision and Control (CDC), Las Vegas, USA, pp. 1303–1309, 2016.

[10] J. L. Proctor, S. L. Brunton, and J. N. Kutz, “Dynamic Mode Decomposition with Control”, SIAM Journal on Applied Dynamic Systems, Vol. 15, No. 1, pp. 142–161, 2016.

[11] M. A. Graham, S. Chakraborti, and SW Human, “Nonparametric EWMA Sign Chart for Location Based on Individual Measurements”, Quality Engineering, Vol. 23, Issue 3, pp 227-241, 2011.

[12] S. Chakraborti, and M. A. Graham, Nonparametric Statistical Process Control, Wiley, New York, NY, 2019.

[13] K. M. Aung, Secure Water Treatment Testbed (SWaT): An Overview, Technical Report, Singapore University of Technology and Design, Singapore, 2015.

[14] J. Goh, S. Adepu, S., K.N. Junejo, and A. Mathur, “A Dataset to Support Research in the Design of Secure Water Treatment Systems”, In Proceedings of International Conference on Critical Information Infrastructures Security (CRITIS), Paris, France, pp. 88- 99, 2016.

[15] S. Adepu, A. Mathur, “An Investigation into The Response of a Water Treatment System to Cyber-Attacks”, In Proceedings of the 17th IEEE High Assurance Systems Engineering Symposium, Orlando, FL, USA, pp. 141-148, 2016.

[16] Qadeer, R., Murguia, C., Ahmed, C. M., and Ruths, J., “Multistage Downstream Attack Detection in a Cyber Physical System”, Cyber ICPS Workshop 2017 in Conjunction with ESORICS 2017, 14-15 Nopember 2017, 177-185.