A Practical Physical Watermarking Approach to Detect Replay Attacks in a CPS

Chuadhry Mujeeb Ahmed[a], Venkata Reddy Palleti[b,*], Vishrut Kumar Mishra[c,d]

[a] *University of Strathclyde, Glasgow, United Kingdom*
[b] *Indian Institute of Petroleum and Energy, Visakhapatnam, India*
[c] *iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design, Singapore*
[d] *Ira A. Fulton Schools of Engineering, Arizona state University*

Abstract

Cyber Physical Systems (CPSs) are considered as an integration of computing, networking, and physical processes. In such systems physical processes are monitored and controlled by the combination of computation, communication, and control technologies. Examples of CPSs include process control systems, medical devices, robots, and critical infrastructures such as water treatment plants, water distribution systems, and power grid. Often, these systems are vulnerable to attacks as the cyber components such as Supervisory Control and Data Acquisition (SCADA) workstations, Human Machine Interface (HMI), and Programmable Logic Controllers (PLCs) are potential targets for attackers. Replay attacks are easy to perform on such systems and can potentially lead to significant damages to the system. Therefore, timely detection of replay attacks is important to mitigate the attack consequences. In this work, we propose a practical watermarking technique to detect the attacks. Experiments are performed on a real WAter DIstribution system (WADI) to explore how to tackle the replay attacks using watermarking signal.

Keywords: Cyber Physical system, Water distribution system, Replay attacks, Design of watermarking signal , Attack detection

1. Introduction

Public critical infrastructures such as a water treatment plant, water distribution and power generation etc., play an important role in a nation's economy. Most modern critical infrastructures are considered as Cyber-Physical Systems (CPSs) that integrate the advanced communications, networking, sensing, and computing technologies with the traditional physical systems to provide efficient and reliable operation. In a CPS, computers and networks are embedded with physical environment to monitor and control the physical processes. Usually, the physical processes and cyber components (computers and networks) interact

*Corresponding author
Email address: venkat_palleti.che@iipe.ac.in (Venkata Reddy Palleti)

with each other in a feedback loop. Also, these infrastructures are spatially distributed and networked which makes them vulnerable to attacks. Consequently, such systems are vulnerable to either physical, cyber, or cyber-physical attacks. Physical attacks on infrastructures can cause disruptions to services and may influence the cyber components through the communication infrastructure due to feedback loops and vice-versa. Also, to increase the impact of an attack, attackers may launch cyber-physical attacks.

Malicious attacks are performed on CPS for various reasons ranging from stealthy attacks to steal resources to widespread attacks for disrupting the service provided by the system. Several such attacks on CPS have been reported in recent years. An investigation into the challenges in the security of CPS in which sensor and actuator data are compromised has been reported in [1]. The general approach in the literature is to study the effect of specific attacks against a particular system. The specific attacks include denial of service and deception attacks against a networked control systems. Denial of service attack refers to the compromise of the availability of resources by jamming communication channels [2, 3]. Deception attacks refer to the compromise of the integrity of sensor and actuator data. Specific types of deception attacks include false data injection, replay and stealthy attacks.

In [4], false data injection attacks are studied in power networks assuming an attacker has perfect model knowledge. In [5, 6] authors demonstrated the effect of replay attacks on the sensor measurements inspired by the Stuxnet example and proposed a methodology to detect such attacks. To evade detection, the attacker replays previous sensor measurements to the operator. These outputs are statistically identical to the true outputs in steady state. Furthermore the adversary requires no knowledge of the system model to generate stealthy outputs. The authors concluded that for some control systems, the classical estimation, control, failure detection strategy are not resilient to the replay attack. Recently a concept of *physical watermarking* has emerged, getting inspiration from the idea of digital watermarking. In [7] proposes a control theoretic method, called physical watermarking, to authenticate the correct operation of a control system. A known noisy signal is embedded as a watermark in the control signal to detect the replay attacks. However, the physical watermark suffer from the limitations of being practical in digital control and done theoretically in the control theory literature. This paper explores the physical watermarking technique to detect replay attacks. The attacker launch the replay attack with an intention to steal water from a pipeline without detection [8]. Therefore, attacker would be launching physical attack (leak) and cyber attack (replay) simultaneously. In practise the replay attacks are difficult to detect. Therefore, in order to detect replay attacks we embedded physical watermark into the actuator signal without compromising the demand of the consumers. Further, we design a practical watermarking technique to detect replay attack. We answered the following research questions in this work:

RQ1: *Can we tackle the replay attack using the physical watermark signal? The challenges we faced was around the design of a practical watermark technique that can be used in the real world system without disturbing the physical process.*

RQ2: *Testing of the proposed technique. There were several challenges identified during the experimentation on the WADI testbed and solutions were found along the way.*

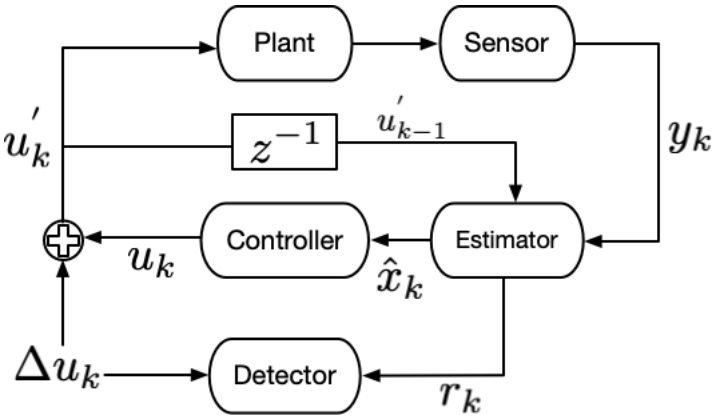

Figure 1: A closed loop feedback control system showing the addition of watermark signal Δu_k.

Organization: The remainder of this paper is organized as follows. Section 2 describes the motivation and background of the watermark and operation of WADI testbed. System model based attack detection framework was explained in Section 3. Watermark Design, implementation on WADI and K-S test Detection scheme is described in Section 4. Related work and conclusions are explained in Sections 7 and 8 respectively.

2. Background

2.1. The Idea of Physical Watermark

The idea of physical watermarking in the context of CPS authentication has been presented in [7]. The concept of watermarking is to authenticate a physical process by adding a known noisy signal and observing its impact on the system outputs. For an attacker who is unaware of this watermark signal, it is hard to truthfully imitate the real state of the system. One can consider the watermark as physical nonce. The idea of physical watermark is inspired by the traditional digital watermark embedded in digital files to verify the authenticity. Figure 1 shows a control system with a watermark signal Δu_k added into the control input. The modified control u_k' will act on the physical plant and its impact shall be contained in the sensor measurements y_k. If a detector module could not retrieve the added watermark then it means that the there is an intrusion or manipulation of the control system.

Physical watermarking can be considered as a form of challenge-response protocol commonly found in the information security literature. The watermark signal being a challenge and sensor measurements being a response. This concept can best be explained by an example. Consider the example in Figure 2, with the control signal on top and sensor output on the bottom plot. The top plot shows motorised control valve (MCV) opening in percentage being driven by a PID control generated by the controller. When plant starts we see that the valve is open to 100% resulting in a high flow rate as shown in the bottom plot. This flow rate is much higher as compared to the required consumer demand of $0.15 m^3/hr$. To adjust that, controller commands the MCV to open a fraction of fully open to control the flow rate,

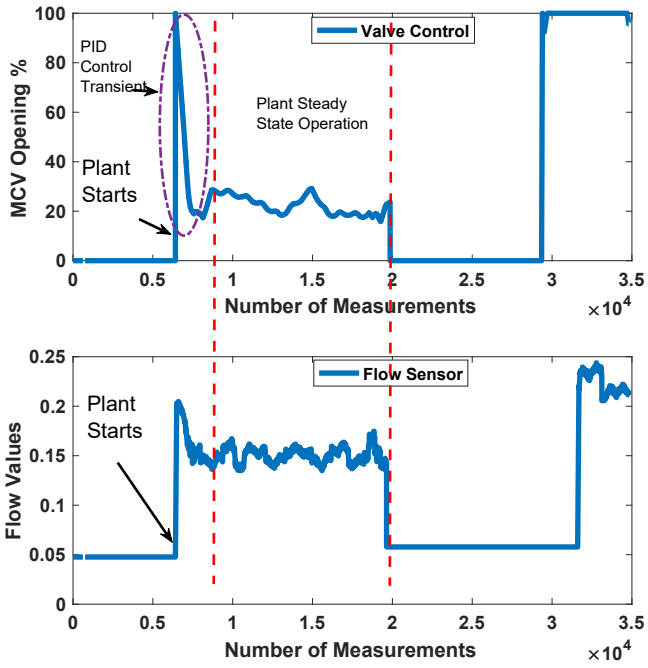

Figure 2: Normal operation for a consumer node. Top plot shows the control action to open the MCV to a particular percentage depending on the measurements of flow sensor as shown in the bottom plot. The control is executed to achieve a particular flow rate at the consumer node.

resulting in drop in flow rate as can be seen in the bottom plot. After a few adjustments we obtain the steady state with slight variations. The steady state is the area between the two dotted vertical bars. A zoomed-in version of the same region is shown in Figure 3. From Figure 3 we can observe that the control signal is quite sensitive to small perturbations in the sensor measurement. Likewise, it can be said that the variations in the sensor output can be observed based on the variation in the control. The physical watermark exploits this relationship between the input and the output.

The idea itself is relevant but it has some practical implementation issues. It leads to a sub-optimal control and can reduce the system efficiency. In some cases it is not possible to implement it as it is, for example, in case of on-off control in water treatment system. In this paper we make first attempt towards a practical implementation of the concept of physical water marking on part of a water distribution testbed while preserving the system performance.

2.2. WADI System Dynamics under Normal Operation

Water Distribution (WADI) plant represents a scaled-down version of a large water distribution network in a city. It is designed to supply 10 US gallons/min of filtered water [9]. WADI consists of three sub processes, namely, primary grid (P1), secondary grid (P2), and return water grid (P3). The primary grid consists of two raw water (RW) tanks of 2500

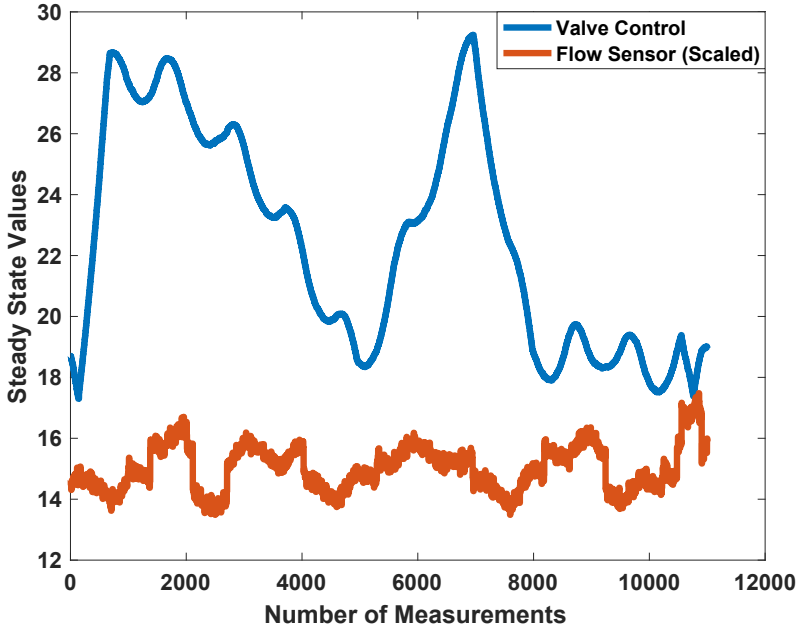

Figure 3: It might be hard to appreciate the control dynamics with respect to the flow sensor measurements from Figure 2. Therefore, the steady state dynamics are shown zoomed-in in this figure.

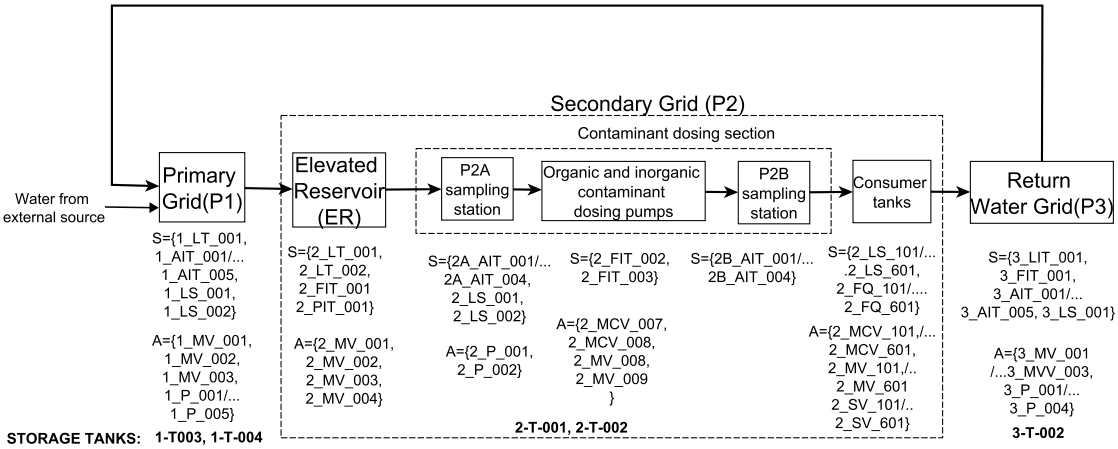

Figure 4: Three stages of the water distribution network. Sensors and Actuators are labelled in each stage of the system.

Liters each. These tanks are fed by three incoming sources including Public Utility Board (PUB), return water grid, and from a Secured Water Treatment (SWAT) plant located physically adjacent to WADI. A level sensor (1_LT_001) is installed in the primary grid to monitor the levels in the RW tanks. Water quality analyzers are installed to measure pH, turbidity, conductivity and residual chlorine.

The process $P2$ consists of two Elevated Reservoirs (ER) tanks, consumer tanks, and contamination sampling stations. A pump (1_P_003) is installed in the primary grid to lift water from raw water tanks to the ER tanks. Two level sensors, 2_LT_001 and 2_LT_002 are installed in ER tanks to measure water levels. Water flows from ER tanks into consumer tanks either via gravity or booster pump based on the preset water demand. Main inlet of the each tank is fitted with a modulating control valve (MCV) to control the inlet flow to the tank. Further, a flow meter is installed to measure the inlet flow rate. Two water quality monitoring stations are installed in the secondary grid. One station is at the immediate downstream of the reservoir and another is before the consumer tanks. These stations ensure water quality before it is sent to the consumer tanks. Once a consumer tank is filled, a level switch installed raises an alarm and water from the tank drains into the return water grid. For simplicity, we divide P2 into three parts, namely P2a, P2b, and P2c. For recycling, the return water grid pumps water to the primary grid.

2.3. Dynamics for the Consumer Tank

Water to the consumer tanks is supplied from the elevated reservoir tanks either by gravity or booster pump based on high or low demand conditions. At the inlet of each tank a flow meter and a Modulating Control Valve (MCV) is installed to measure and control the flow. Similarly, at the outlet of each tank a Motorized Valve (MV) is installed to drain the tanks whenever it is necessary. A level switch is installed in each tank to measure the water level. MCVs will have the opening from 0 to 100% and MVs operate in either fully open or close position. Figure 5 shows the controller scheme representation of a consumer tank. During the plant startup MCVs will open to 100% and therefore, maximum inflow into the tanks can be observed. A demand set point is set by the user for each consumer tank. The measured inflow and demand set point are compared, and the feed-back error is calculated. The resulting error is used in the PID controller function and the controller acts upon MCVs accordingly to minimize the error between the set point and the measured inflow. When any consumer tank becomes full, High Switch alarm will be activated and a command will be sent to the inlet MCV to fully close and outlet valve MV to open.

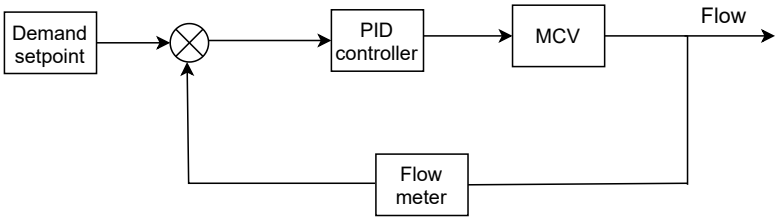

Figure 5: Controller setup in consumer tanks

3. Replay Attack and Model based Attack Detection Framework

Replay attacks tend to hide the real state of the process from the SCADA system by recording and replaying normal measurements. An experimental study is carried out in [10] to show how an attacker can hide an intentional leak in water distribution system by launching replay attacks on sensor and actuator measurements simultaneously. The experiments were performed on a real operated water distribution testbed, named as WADI. The threat model, detection frame work and experimental investigation is explained in brief here.

3.1. System Model

A system model represents the dynamics of a physical process as a mathematical model. A system model in form of LTI equations where actuator signals are considered as input and sensor measurements as output is obtained. This equation is of the form given by equation (1).

Normal Operation:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + v_k, \\ y_k = Cx_k + \eta_k, \qquad SystemModel. \end{cases} \tag{1}$$

$$\begin{cases} \hat{x}_{k+1} = A\hat{x}_k + Bu_k + L_k(y_k - C\hat{x}_k), \\ \hat{y}_k = C\hat{x}_k, \qquad \text{State Estimation.} \end{cases} \tag{2}$$

Where $x \in \mathbb{R}^n$ is system state vector, $A \in \mathbb{R}^{nxn}$ is state space matrix, $B \in \mathbb{R}^{nxp}$ is the control matrix, $y \in \mathbb{R}^m$ are the measured outputs, $C \in \mathbb{R}^{mxn}$ is measurement matrix, and $u \in \mathbb{R}^p$ denote the system control. η_k and v_k are the sensor and process noise, respectively. The state space matrices A, B, C capture the system dynamics and can be used to find a specific system state given an initial state. Normal data from the WADI testbed is used to obtain the system model through subspace system identification technique [11]. Equation 1 can be used to estimate the normal behavior of a physical process using the Kalman filter as given in equation (2). In case of anomalies, this can be used to determine the deviation.

3.2. Model based Anomaly Detection

The earlier studies have assessed the performance of the CUSUM model-based fault detection procedure for a variety of attacks [10]. The residual random sequence $r_k, k \in \mathbb{N}$ is defined as the difference between sensor measurements ($\bar{y}_k$) with attack value(δ_k), and the estimate of the sensor measurement.

$$r_k := \bar{y}_k - C\hat{x}_k = Ce_k + \eta_k + \delta_k. \tag{3}$$

If there are no attacks, the mean of the residual is

$$E[r_{k+1}] = CE[e_{k+1}] + E[\eta_{k+1}] = \bar{r}_{m\times 1}. \tag{4}$$

Where $\bar{r}_{m\times 1}$ denotes an $m \times 1$ matrix composed of mean of residuals under normal operation, and the co-variance is given by

$$\Sigma := E[r_{k+1} r_{k+1}^T] = CPC^T + R_2. \tag{5}$$

Two hypothesis are formulated, $\mathcal{H}_0$ the *normal operation*, i.e., no attacks, and $\mathcal{H}_1$ the *anomalous operation*, i.e., with attacks. The two hypotheses can be stated as follows.

$$\mathcal{H}_0 : \begin{cases} E[r_k] = \bar{r}_{m\times 1}, \\ E[r_k r_k^T] = \Sigma, \end{cases} \text{ or } \quad \mathcal{H}_1 : \begin{cases} E[r_k] \neq \bar{r}_{m\times 1}, \\ E[r_k r_k^T] \neq \Sigma. \end{cases}$$

Initially, experiments were conducted under normal operating conditions, i.e., without launching any attacks, to obtain a system baseline behavior. To obtain a steady state of the system, the plant is left running for some time (5-15 minutes). For a normal run of the plant, consumer tanks were emptied and the consumer demand set to a constant of $0.15 m^3/h$ for all six consumers. Sensor and actuator data was collected from WADI once the sensors and actuators reached their initial target state. Data collected from these experiments were input to the system model obtained using subspace system identification as discussed in Section 3. Based on the system model the residual (r_k) for each sensor (i) was generated using the Kalman filter based state estimation. CUmmulative SUM (CUSUM) detector was used to raise an alarm for a outlier data [12]. Further, During the normal operation of the plant the attacker chooses a set of sensors and actuators, observe and record their readings for a certain time duration in order to carry the replay attack.

3.2.1. Attack Execution

For simulating a leak, the valve 2_MCV_007 can be used to cut-off water supply. The idea is to capture the behavior of a physical attacker. An attacker can choose to cut off water supply by opening or closing the valve 2_MCV_007 in a gradual or abrupt manner from 0 % to 100%. In order to hide the physical attack, the attacker launches replay attacks simultaneously on a certain set of sensor measurements which are recorded during the normal operation of the plant. These attacks were performed strategically to avoid leak detection by choosing the upstream pressure sensor, namely 2_PIT_001, the downstream pressure sensor (2_PIT_002), and the flow sensor (2_FIT_002). The control valves for consumer tanks (MCVs) are also chosen to perform the replay attack on. As shown in [10], when the attacker uses the complete knowledge of the system and perform replay attack on all the actuators and sensors used by the detector, attack detection is not possible. For example, in [10], the authors showed that the powerful attack in which the attacker compromises all sensor and actuators was not detected and the attacker was successful in hiding the attack. As shown in Figure 6 the residual signal for two pressure sensor measurements namely, 2_PIT_001 and 2_PIT_002, stays within the upper and lower threshold of a CUSUM detector. It is evident that the attacker successfully stolen water from the system without being detected by launching replay attacks on certain set of sensor and actuator measurements. Therefore, we design a physical watermarking in which a random noise is injected into the control input to detect such kind of replay attacks.

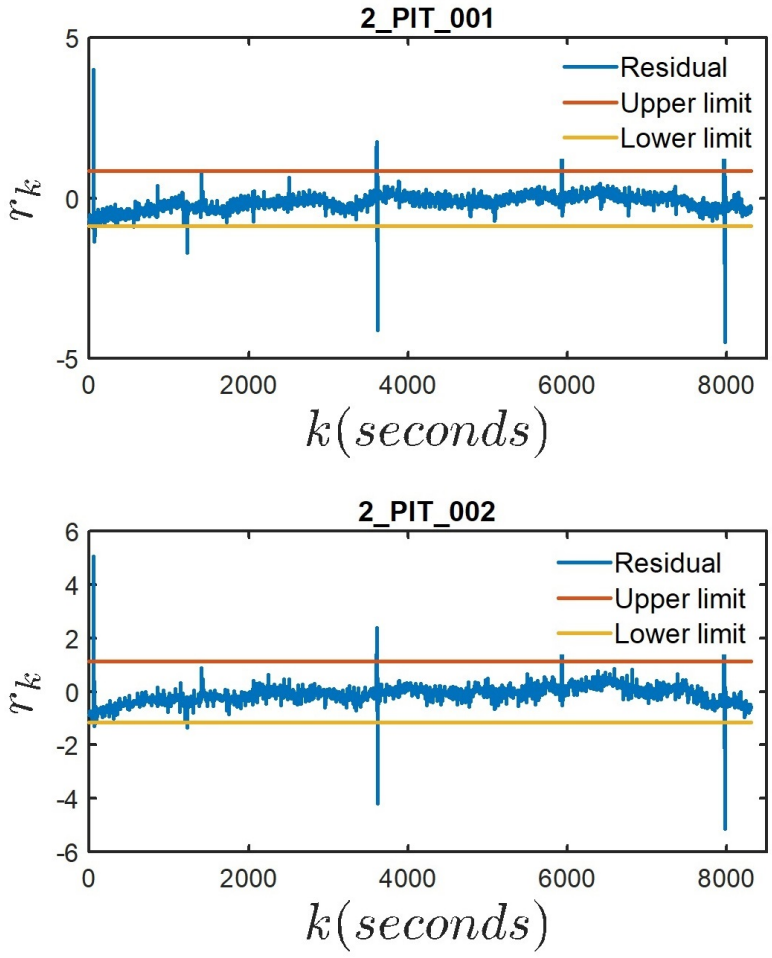

Figure 6: Residual estimation for replay attack on pressure, flow sensors and modulating control valves.

Definition 3.1. *Let us define the sensor measurements under a replay attack as* y_k^a*, control signal under replay attack as* u_k^a *and state estimate as* x_k^a *at the time step* k*. Where* $0 < k \leq T$ *for an attack time period* T*.*

Proposition 3.1. *Given the system of equations for normal system model as* (1)-(2) *and replay attack defined in above, it can be shown that replay attack would not be detected.*

Proof: The residual vector under an attack is given as,

$$r_{k+1}^a = y_{k+1}^a - \hat{y}_{k+1}. \tag{6}$$

During the replay attack for times $0 < k \leq T$ where T is the time for the readings being replayed, $y_{k+1}^a = y_{k+1}$, resulting in $r_{k+1}^a = r_{k+1}$ therefore, resulting in no detection and the alarm rate reduces to the false alarm rate of the detector being used. ■

4. Design of Watermark

The core idea of the physical watermark design is illustrated in Figure 1. A closed loop feedback control system describes the essential componenets, a controller issuing the control commands u_k, sensor output y_k, a watermarked control $u_k^{'} = u_k + \Delta u_k$. The idea is that the amount of change we inject as a watermark Δu_k, it shall affect the plant and can be captured by the sensor measurements. Moreover, the watermark being generated randomly shall not be available to a replay attacker beforehand to launch a successful attack.

Theorem 4.1. *Given the system model in equation* (1) *, Kalman filter* (2) *and watermarked inputs* $u_k^{'} = u_k + \Delta u_k$*, it can be shown that the residual vector is driven by the watermark signal and can be given as,* $r_{k+1} = [CA - CL_kC](x_k^a - \hat{x}_k^{wm}) + CB(u_k^a - u_k) - CB\Delta u_k + Cv_k + \eta_{k+1}$.

Proof: In the system model of eq. (1), attacker has access to the normal sensor measurements and control signals and can replay those. Assuming that an adversary has access to the system model, Kalman filter gain and other parameters of the detectors. During the replay attack using its knowledge an attacker estimates the system state as follows,

$$x_{k+1}^a = Ax_k^a + Bu_k^a + v_k \tag{7}$$

and attacker's spoofed sensor measurements as,

$$y_k^a = Cx_k^a + \eta_k \tag{8}$$

$$y_{k+1}^a = C[Ax_k^a + Bu_k^a + v_k] + \eta_{k+1} \tag{9}$$

$$y_{k+1}^a = CAx_k^a + CBu_k^a + Cv_k + \eta_{k+1} \tag{10}$$

However, using a watermaking signal in the control input u_k defender's state estimate becomes,

$$\hat{x}_{k+1}^{wm} = A\hat{x}_k^{wm} + Bu_k + B\Delta u_k + L_k(y_k^a - C\hat{x}_k^{wm}), \tag{11}$$

where Δu_k is the watermark signal.

$$\hat{y}_{k+1}^{wm} = C\hat{x}_{k+1}^{wm} \tag{12}$$

$$\hat{y}_{k+1}^{wm} = C[A\hat{x}_k^{wm} + Bu_k + B\Delta u_k + L_k(y_k^a - C\hat{x}_k^{wm})] \tag{13}$$

$$\hat{y}_{k+1}^{wm} = CA\hat{x}_k^{wm} + CBu_k + CB\Delta u_k + CL_k(Cx_k^a - C\hat{x}_k^{wm})] \tag{14}$$

$$\hat{y}_{k+1}^{wm} = CA\hat{x}_k^{wm} + CBu_k + CB\Delta u_k + CL_kCx_k^a - CL_C\hat{x}_k^{wm} \tag{15}$$

The residual vector is given as,

$$r_{k+1} = y^a_{k+1} - \hat{y}^{wm}_{k+1} \tag{16}$$

$$\begin{aligned} r_{k+1} = CAx^a_k + CBu^a_k + Cv_k + \eta_{k+1} - CA\hat{x}^{wm}_k \\ -CBu_k - CB\Delta u_k - CL_kCx^a_k + CL_kC\hat{x}^{wm}_k \end{aligned} \tag{17}$$

$$\begin{aligned} r_{k+1} = (CA - CL_kC)x^a_k - (CA - CL_kC)\hat{x}^{wm}_k + CB(u^a_k - u_k) \\ -CB(\Delta u_k) + Cv_k + \eta_{k+1} \end{aligned} \tag{18}$$

$$\begin{aligned} r_{k+1} = (CA - CL_kC)(x^a_k - \hat{x}^{wm}_k) + CB(u^a_k - u_k) \\ +Cv_k + \eta_{k+1} - CB(\Delta u_k) \end{aligned} \tag{19}$$

The last term is the watermark signal. The first term is the error. For a stable system the spectral radius of $(CA - CL_kC < 1)$ and the error converges to zero [13]. Moreover, in replay attacker an attacker chooses control command exactly same as normal process, therefore, $u^a_k = u_k$, making the second term to go to zero. Without watermark the residual vector would behave normally driven by the noise. However, with the last term added the attacker can not successfully execute the replay attack unless it can counter the effect of watermark. This concludes that with an added watermark, residual vector is driven by the watermark and can expose replay attacks. ■

4.1. Choosing Watermark Signal

Definition 4.1. *Performance Loss: The performance loss is defined as the down-gradation in the output of the system under a sub-optimal control. For example, for a water distribution network, it is important that the user demand is met. Based on the user demand a motorized valve is actuated to open to a specific value that would meet the user demand, any change in that opening might lead to a unsatisfied user demand.*

Definition 4.2. *A watermark signal which is chosen randomly in each iteration of experiment is defined as the dynamic watermark.*

In the following it is shown that how system performance is preserved on average while maintaining a control signal watermark. Figure 4 shows the block diagram of the complete WADI testbed, the portion of the testbed used in this study is the consumer tank unit that is part of secondary grid–P2. There are in total six consumer tanks representing the typical consumers in the real-world. Figure 5 shows the control implemented at each controller in that stage of the testbed, meeting the user demands. As seen in Figure 5 the key parameters for each consumer tank unit are, the demand setpoint–indicating the user demand, then the Controller would open motorized valve (MCV) as much required to meet the consumer demand, the flow meter is used to feedback the water being supplied so that the controller can adjust the control if the demand is not being met.

The objective here is to chose a watermark in the control input that can preserve the system performance. We use the uniform distribution to choose the values for the controller to add the watermark.

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{for } x < a \ \ or \ \ x > b \end{cases} \tag{20}$$

$$E(X) = \frac{1}{2}(a+b) \tag{21}$$

Equation (20) is the probability density function of the uniform distribution. The distri-

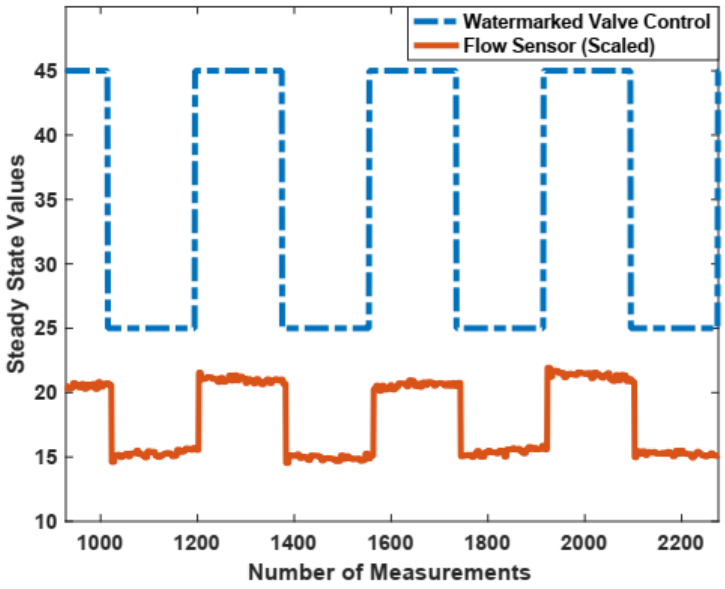

Figure 7: It is demonstrated that if the control signal is watermarked using particular control sequence then the effects of the watermark could be observed in the output signal, i.e., flow sensor. Flow sensor is scaled to display on the same plot and compare the overall pattern with the control signal.

bution explains an arbitrary outcome that lies between bounds defined by the parameters a and b. We obtain these parameters from the empirical analysis. Figure 7 shows an example experiment that lead us to define the bounds. For a fixed demand rate we measured the normal MCV control opening to fulfill the consumer demand. Then, MCV control is modified by adding incremental values in the optimal control to see the effects of added watermark on the output and the demand of the system. Ideally we should have a separate distribution model for each consumer tank but to keep the results tractable and explainable, we had taken a fixed demand pattern for all the consumers for the duration of these experiments, hence uniform distribution with same lower and upper limits. It is concluded that if a small change is made it does not reflect in the sensor output. Through experimentation, as shown in Figure 7, it is observed that if MCV opening is varied between 25% and 45%, it produces significant effects on the output sensor measurement as can be seen in the figure. Above or

below these two bounds change is not significant due to process capacity limitations, therefore, a and b are taken to be 25 and 45% respectively. Figure 8 shows the watermarking techniques in action using these bounds and randomly adding a watermark within these bounds. It can be seen that watermark introduce enough variance in the output that the effects of watermark could be observed on the output sensor measurements. To add the watermarking, we override the PID controllers for the MCVs to follow a set pattern created in Section 4.1.

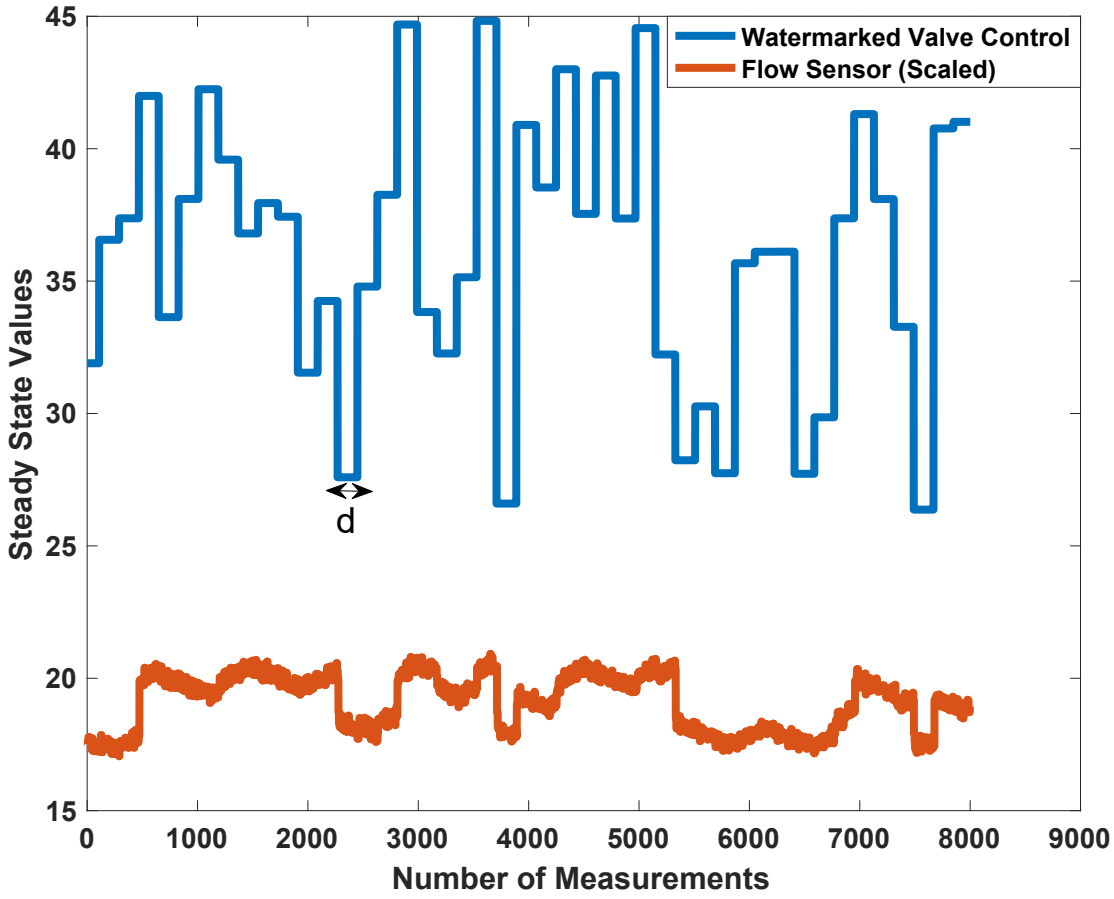

Figure 8: Randomly adding a watermark signal to the control would result in a random change in the flow measurement.

Figure 9 shows an example implementation of the watermarked control. Top plot in the figure shows time series data for three variables. The watermarked control input that varies between 25% and 45% values for MCV opening. As a result of this watermarked control input the flow sensor measurement show a variation pattern in the output, as the MCV opens to 45% the flow increases and it decreases when MCV is open 25%. Moreover, a system model is obtained mapping control input to the output and this obtained system model is used to estimate the sensor output given a particular watermarked control input.

The choice of control input between 25-45 is made empirically as this allows the demand to be fulfilled, therefore, a random value between 25 and 45 shall be chosen to inject a watermark signal. The system model helps to quantify the value of the injected watermark.

4.2. Detector: Kolmogorov-Smirnov (K-S) Test

The Kolmogorov-Smirnov (K-S) test is a non-parametric test for the equality of continuous probability distributions. It can be used in one-sample settings where a sample is compared with a reference probability distribution or in two-sample settings to compare the empirical distribution functions of two samples. In this study a two-sample K-S test is used as a detector. The K-S statistics quantifies a distance between the empirical distribution functions of two samples.

Definition 4.3. Empirical Distribution Function: *For n independent and identically distributed ordered observations of a random variable* $\mathbf{X}$*, an empirical distribution function* F_n *is defined as,*

$$F_n(x) = \frac{1}{n}\sum_{i-1}^{n} I_{[-\infty,x]}(\mathbf{X_i}), \tag{22}$$

where $I_{[-\infty,x]}(\mathbf{X_i})$ *is the indicator function which is* 1 *if* $\mathbf{X_i} \leq x$*, else it is equal to 0.*

For a given Cumulative Distribution Function (CDF) $F(x)$, the K-S statistic is given as,

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|, \tag{23}$$

where sup is the supremum of the set of distances. For the two-sample K-S test the statistic can be defined as,

$$D_{n,m} = \sup_{x\in\mathbb{R}} |F_n(x) - G_m(x)|, \tag{24}$$

where $F_n(x)$ and $G_m(x)$ are the empirical distribution functions of the first and second sample, respectively. $D_{n,m}$ is the maximum of the set of distances between the two distributions. For a large sample size the null hypothesis is rejected for the confidence level α if,

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n*m}}, \tag{25}$$

where n, m are, respectively, the sizes of the first and second samples. The value of $c(\alpha)$ can be obtained from the look up tables for different values of α, or can be calculated as follows,

$$c(\alpha) = \sqrt{-\frac{1}{2}ln(\alpha)}. \tag{26}$$

Remark: K-S test checks whether the two samples are drawn from the same distribution or not. The test statistic is based on the maximum distance between empirical distributions of the samples. If the supremum of the distance between two samples is greater than a certain threshold, the null hypothesis that both samples are from the same distribution would be rejected. Under a replay attack samples would look like the original trained model without watermark. In the absence of any attack watermark would be preserved and null hypothesis would be rejected.

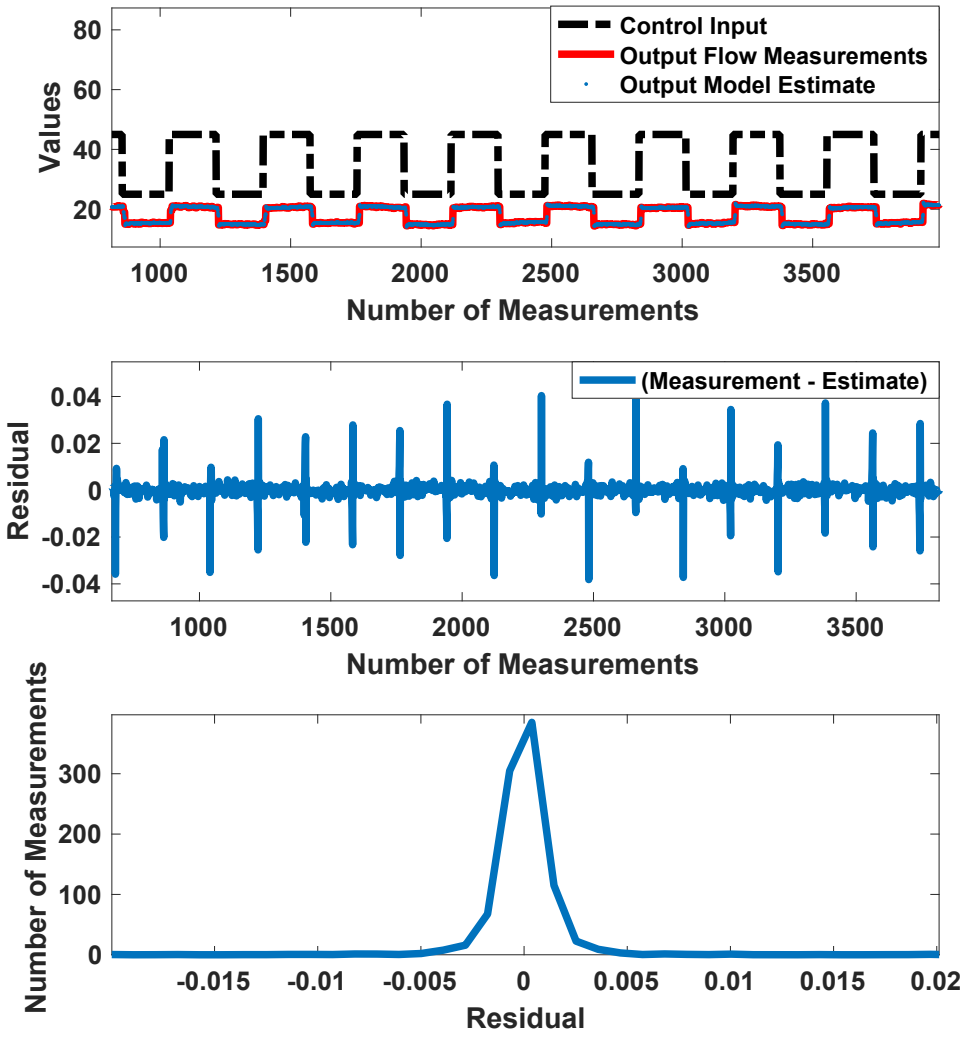

Figure 9: System model for the watermarked system.

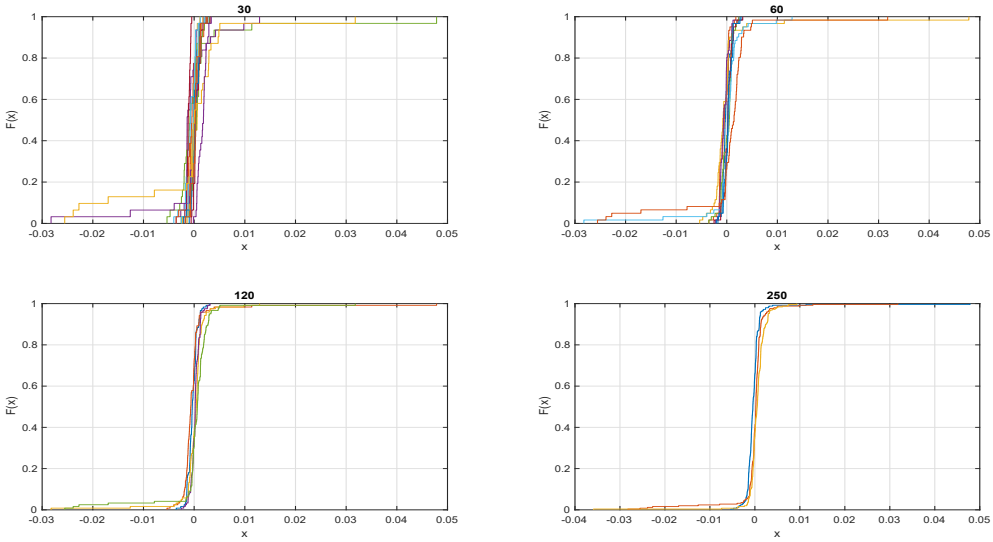

Figure 10: CDF for different chunks of data for the residual signal obtained from the system model under the normal operation of the plant.

5. Evaluation

5.1. Performance Metrics

Following metrics are used while assessing the effectiveness of the attack detection procedure.

- *True Positive Rate (TPR):* This rate is defined as labeling the data as an attack when the anomaly actually existed. This is a correct prediction of an attack.
- *False Positive Rate (FPR)*: This is defined as raising an alarm when the data received is attack free. This is a false alarm.
- *False Negative Rate (FNR)*: This is defined as not raising an alarm when the the anomaly actually existed. This is a incorrect prediction of an attack.
- *True Negative Rate (TNR)*: This is defined as labeling the data as normal when the operation is actually normal.This is a correct prediction of a normal operation.

Ideally, FPR should be as small as possible and TPR as high as possible. Both TPR and FPR being ratios range between 0 and 1.

5.2. Chunk Size vs Accuracy on Normal Operation

In the following, the empirical distributions for the residual vector will be derived and a reference model without watermarking operation is trained. A trade-off between the speed of detection and detection performance is desired. Empirical distributions for different chunk

size of a example flow sensor residuals of a consumer tank are shown in Figure 10. This experiment is performed to establish a trade-off between the number of samples required and the accuracy of detection. Sensor measurements and the system model is used to calculate the residual vector on which the proposed technique is applied. Normal plant operation data is used in this experiment. A chunk size of $10, 30, 60, 120, 250, 500$ is taken. Figure 10 shows the visual inspection of the data for different chunk sizes. Empirical CDF is plotted for data samples from each chunk. The essence of choosing a right chunk size means that the number of samples in a chunk are enough to capture the variation in the data. A chunk size of 10 means that there are 10 samples of data in each chunk and as shown in Figure 10, several chunk's CDF does not correlate to each other, making it non-substantial to be used with a K-S test. Since this is the normal data, it is desired to obtain a chunk size that indicate that the samples are drawn from the same distribution, which in fact are drawn from the same distribution. From Figure 10, it can be seen that for a small chunk size the distributions could not be approximated very well and from one chunk to the other, the distance between the distributions of two samples is higher resulting in increased false alarms. At the larger number of samples being used, the empirical CDF is smooth and true negative rate is higher but then it needs to wait for additional time to make a decision. A good trade off is made somewhere in the middle. From Figure 10 it is evident that the chunk size of 120 onward gives the best result, however, it is determined that the a chunk size of 250 gives the best result in terms of accuracy and amount of samples required to detect an attack.

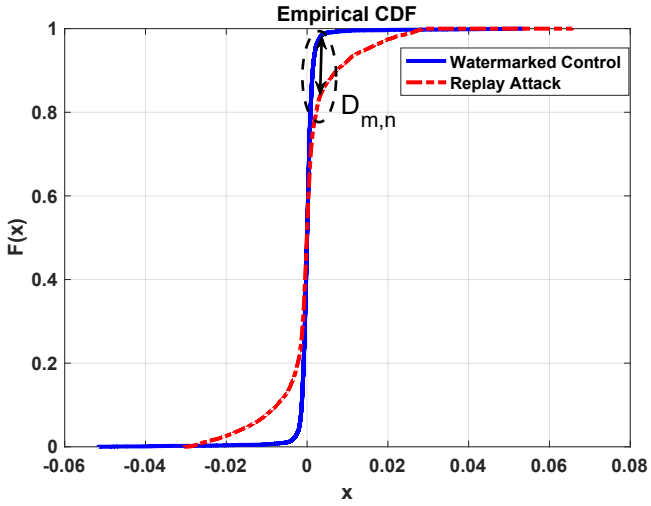

Figure 11: CDF for the replay attack and the watermarked signal.

5.3. Replay Attack Detection

Replay attack detection results are shown in the Table 1 for different chunk size of the data. It is observed that for a chunk size of 250 we are able to achieve 100% TPR and from

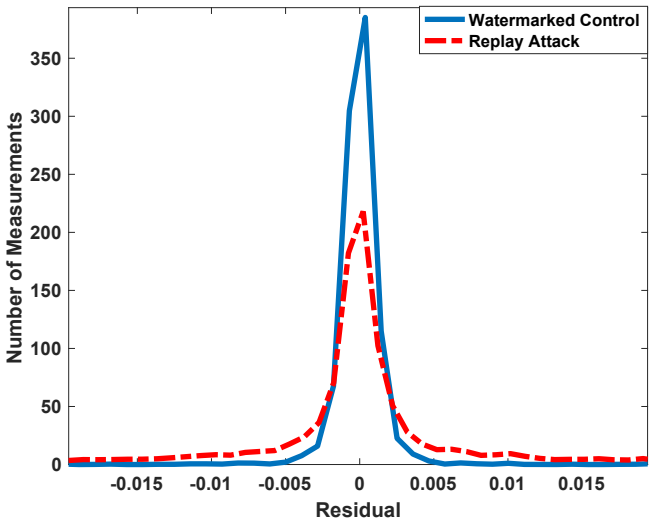

Figure 12: Watermarked control signal vs replay attack distributions.

Figure 10, it is the same chunk size for which we had minimum false alarm rate. Table 1 shows the results for six consumer units, C1 to C6, where the replay attacks were launched.

For visual inspection, consider the plot in Figure 11, it shows the CDF for the scenario of added watermark signal and replay attack, both being from different distributions, $D_{m,n}$ can be seen as significant to result in detection. One sample is taken from the replayed data and the second sample from the watermarked control signal. It is observed that the distance metric $D_{m,n}$ is greater as compared to the plot of normal data, pointing out that these two distributions are drawn from two different distributions. This is the key intuition to detect replay attacks in the presence of a watermark signal. The replay attack would record the normal data and then replay it hence not showing the evidence of the existence of the watermark signal and would expose itself. Figure 12 shows the same result in the form of PDF for the watermarked control case and the replayed data case, for the residuals and their distributions.

5.4. Challenges

Limited Choice of Parameters for the Uniform Distribution. An important design parameter is the bounds a and b of the uniform distribution to chose the watermark to be added from. It is stated previously that we had chosen these bounds based on the empirical observations. This part elaborates on those experiments and the limitations we faced for our choice of the bounds. When the position of MCV value is less than 20% the variation in the flow is observed to be low and not distinguishable from adding a watermark to normal operation. Similarly for position of MCV greater than 45% the flow was close to maximum with very little variation and the flow reaches maximum value ($0.15m^3/h$) at MCV position of 65%. This means that the design of watermark in this case is limited by the physical

Consumer / Chunk Size	10	30	60	120	250	500
C1: TPR	87.97%	88.59%	89.86%	100%	100%	100%
FNR	12.03%	11.41%	10.14%	0%	0%	0%
C2: TPR	78.86%	83.43%	90.91%	97.56%	100%	100%
FNR	21.14%	16.57%	9.09%	2.44 %	0%	0%
C3: TPR	66.70%	67.78%	71.62%	91.78%	100%	100%
FNR	33.30%	32.22%	28.38%	8.22%	0%	0%
C4: TPR	83.43%	88.59%	90.95%	98.56%	100%	100%
FNR	16.57%	11.41%	9.05%	1.44%	0%	0%
C5:TPR	68.60%	80.54%	91.22%	89.04%	100%	100%
FNR	31.40%	19.46%	8.78%	10.96%	0%	0%
C6:TPR	53.23%	50%	57.43%	60.27%	70.59%	81.25%
FNR	46.77%	50%	42.56%	39.72%	29.41%	18.75%

Table 1: K-S test for Replay Attack. Chunk size vs attack detection accuracy for all Consumers in the WADI testbed.

capacity of the process, this is an important insight towards a practical watermark design. Therefore, we had to randomly choose the position of MCVs for watermarking between 25% and 45%.

As shown by Figure 8, the change in MCV position need to be sufficiently high to ensure that it can be observed on the flow meter. Since for small changes the change in flow rate cannot be observed. Further the time between two successive changes in MCV position need to be adequate. If enough time is not given, the flow doesn't reach a steady state before MCV changes again making the flow unpredictable.

Keeping the System Performance Intact Over a Desired Period of Time. From the design of watermark, recall that the watermark is chosen between bounds to achieve the system performance as the normal process on an average. A key consideration in this regard is the time frame we have to implement the watermark and fulfill the user demand. In the case of testbed used in this study, demand from users is assumed at full capacity meaning we get the maximum time to implement and achieve desired system performance in average sense. The consumer tanks becomes full in 150 minutes from the start of the system and the systems take 30 minutes to reach a steady initial state. This leaves us with only 120 minutes of time for total experiment cycle, meaning from an empty consumer tank to the complete filling of consumer tank using a constant demand throughout this period of time. Therefore watermarking experimentation is quite challenging to perform different scenarios within 120 minutes of duration. While deciding the watermark, it needs to be ensured that the consumer demand is met and the MCV position pattern is kept random.

Chunk Size Vs Accuracy. As described earlier, deciding chunk size to perform detection is important as it is desired to perform detection as fast as possible. we found that 250 samples gave us the best result but this parameter can be system dependent and need to be part of design for a specific process domain.

6. Discussion

6.1. Multi Input and Output System

Let's consider an example of a LTI system model with two system states (x_k^1, x_k^2) and two control inputs(u_k^1, u_k^2),

$$\begin{bmatrix} x_{k+1}^1 \\ x_{k+1}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} u_k^1 \\ u_k^2 \end{bmatrix} \tag{27}$$

$$\begin{matrix} x_{k+1}^1 \\ x_{k+1}^2 \end{matrix} = \begin{matrix} a_{11}x_k^1 + a_{12}x_k^2 + b_{11}u_k^1 + b_{12}u_k^2 \\ a_{21}x_k^1 + a_{22}x_k^2 + b_{21}u_k^1 + b_{22}u_k^2 \end{matrix} \tag{28}$$

The two system states are labeled as x_k^1 and x_k^2. It can be observed that for a joint model the control has effects on both system states. This means there is an obvious relationship through control signal and then inject respective watermark. Consider the following cases:

Case 1: Watermark injected in U1

$$\begin{bmatrix} x_{k+1}^1 \\ x_{k+1}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} u_k^1 + \Delta u_k^1 \\ u_k^2 \end{bmatrix} \tag{29}$$

$$\begin{matrix} x_{k+1}^1 \\ x_{k+1}^2 \end{matrix} = \begin{matrix} a_{11}x_k^1 + a_{12}x_k^2 + b_{11}(u_k^1 + \Delta u_k^1) + b_{12}u_k^2 \\ a_{21}x_k^1 + a_{22}x_k^2 + b_{21}(u_k^1 + \Delta u_k^1) + b_{22}u_k^2 \end{matrix} \tag{30}$$

Case 2: Watermark injected in U2

$$\begin{bmatrix} x_{k+1}^1 \\ x_{k+1}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} u_k^1 \\ u_k^2 + \Delta u_k^2 \end{bmatrix} \tag{31}$$

$$\begin{matrix} x_{k+1}^1 \\ x_{k+1}^2 \end{matrix} = \begin{matrix} a_{11}x_k^1 + a_{12}x_k^2 + b_{11}u_k^1 + b_{12}(u_k^2 + \Delta u_k^2) \\ a_{21}x_k^1 + a_{22}x_k^2 + b_{21}u_k^1 + b_{22}(u_k^2 + \Delta u_k^2) \end{matrix} \tag{32}$$

From the case 1 and case 2 above we can observe that the injected watermark in either of the control signal will have an effect on both system states.

Case 3: Watermark injected in both control signals simultaneously

$$\begin{matrix} x_{k+1}^1 \\ x_{k+1}^2 \end{matrix} = \begin{matrix} a_{11}x_k^1 + a_{12}x_k^2 + b_{11}(u_k^1 + \Delta u_k^1) + b_{12}(u_k^2 + \Delta u_k^2) \\ a_{21}x_k^1 + a_{22}x_k^2 + b_{21}(u_k^1 + \Delta u_k^1) + b_{22}(u_k^2 + \Delta u_k^2) \end{matrix} \tag{33}$$

From above equations we can see that the modeling might be complex with multiple inputs and outputs but it results in better security due to the fact that there are more vectors to randomize and profile their effect on the output making it harder for an attacker to predict the random watermark.

6.2. Limitations

Control Limits: It is known that the watermarking techniques can not be applied to all types of control, especially in case of a bang-bang control where the control action rapidly ON/OFF [14, 7]. The results from [14] show that in theory any arbitrary noise signal can be generated to be added to control but in real-world implementation of such a control is not explored. As we acknowledge this limitation of watermarking techniques and also that some systems might not support a range of 25% to 45%, it is important to highlight that our work is the first work to demonstrate a practical design and implementation of such a watermark on a real-world system.

Overhead: An overhead of the proposed watermarking technique is that it would increase the power consumption and component degradation due to frequent change in the control. For example, consider Figure 8, where d represents the duration for which a watermark signal is active. It can be seen that if d is large, MCV changes position less frequently and resulting in lower energy consumption (E) and vice versa if d is small. Therefore, it can expressed as,

$$E \propto \frac{1}{d} \tag{34}$$

If d is larger the energy consumption is low but security introduced by watermarking also reduce as an attacker can observe the watermark for longer duration and have a better chance to learn the watermark. We need to find an appropriate tradeoff, however, this won't be problem in modern equipment due to lower energy consumption and higher number of cycles, as an example, the pneumatic industrial valves consume energy as low as 0.1 Watt with a life of over 200 million cycles [15].

Practical Aspects: We have tested the proposed method on a water distribution testbed over a few weeks time. Moreover, being used in a testbed for few weeks is different from being used in a real-world production system of physical plants with possibly more harsh environment. It has been discussed in earlier work [16] that measurement noise profile of the devices might change over time due to wear and tear. However, as seen in Figure 8 if we can drive the output response far from the normal process noise then a higher accuracy can be achieved. Watermark can be exploited to achieve such a behaviour but again this would be limited by the limits of the control and bounds on watermark as discussed in this section.

6.3. Future Work

Considering a threat model in which an attacker can collect data with injected watermark and then replay that distribution is not considered in this work but this shall be mentioned in potential limitations and possible defenses shall be mentioned. There are two lines of arguments that can be taken here: 1) Considering the multiple input and multiple output modeling in newly added discussion section, it can be seen that a combination of multiple processes can generate more possible combinations of watermark to be added and making it harder for the attacker. We plan to tackle this problem in the future work. 2) From literature we have found [17] which propose another argument to tackle this problem. The

argument is that although an attacker can learn the pattern but can not figure out beforehand what exactly will be the watermark from that distribution to be added at each time instant. Therefore, an adversary needs live monitoring and response to detect a new watermark, which is possible but it would need n samples to detect and respond to the injected watermark, where n is the smallest number of samples to learn a pattern. Therefore, time spent on collecting those samples shall expose the attackers.

7. Related Work

CPSs play an important role in critical infrastructures, such as water transportation, power, oil and gas. Modern infrastructure are prone to cyber-physical attacks and successful attacks may bring huge damages to critical infrastructure, human lives and properties, and even threaten the national security. Maroochy water breach in 2000 [18], Stuxnet malware in 2010 [19], Ukraine power outage in 2015 [20] and other security incidents, motivates the researchers to focus more on CPS security. In the recent years, researchers have focused on the design of watermarking technique to detect replay attacks.

Authors in [5, 6] studied the problem of replay attack detection and proposed a physical watermarking scheme in which an authenticating signal is introduced into the control system. This technique enables the detection of replay attacks, however, the watermark signal may degrade the controller performance and may lead to sub-optimal solution. Therefore, [6] also studied the control performance loss and find the optimal trade-off between detection rate and performance loss. Authors in [21, 22] proposed and additive based watermark signal generated by dynamical systems. An optimization problem is formulated to give a loss effective watermark signal with a specified detection rates by tuning the design parameters. Further , [23, 14] provides a comprehensive procedure for dynamic watermarking signal. [24] proposed an algorithm which can generate simultaneous watermarking signal and design parameters for attack detection and it is shown that the algorithm converges to the optimal one. In [25, 26] authors identify the limitations of the detection schemes proposed by [5] and a multi watermark based detection technique is proposed to overcome those limitations. A periodic watermarking strategy was proposed in [27] by including the control loss performance when a random signal is injected into a control system. In contrast to the other works, authors in [28] proposed a multiplicative sensor watermarking. In this scheme each sensor is watermarked and a watermark remover is embedded to reconstruct the original sensor measurements. It is shown that this scheme is effective and does not degrade the controller performance in the absence of attacks. Our work is first kind of study taking the idea of watermark from theory to practise and solving several challenges along the way. It is demonstrated that the design of practical watermark is possible but needs domain knowledge and system expertise.

8. Conclusions

Stealthy and replay attacks on critical water cyber physical systems have gone undetected for long [10, 8]. Physical watermarking [7] has been proposed to detect the replay attacks,

however, it is challenging to come up with a practical design. It is demonstrated in this work, that a practical watermarking design can be achieved but presents with the new challenges. Design of practical physical watermarking requires domain knowledge, accurate system models and process parameters, so that we can chose a random watermark without affecting the system performance. Experimental results provided an understanding of these parameters, resulting in a successful design, that is evaluated on the live water distribution network. We can extend this work assuming the more intelligent attacker who has the knowledge of injected water mark signal into the system. In such a case more efficient algorithms are needed to counter the attacker knowledge and it will be considered as part of the future research work.

References

[1] A. A. Cárdenas, S. Amin, S. Sastry, Research challenges for the security of control systems, in: Proceedings of the 3rd Conference on Hot Topics in Security, HOTSEC'08, USENIX Association, Berkeley, CA, USA, 2008, pp. 6:1–6:6.

[2] S. Amin, A. Càrdenas, S. S. Sastry, Safe and secure networked control systems under denial-of-service attacks, in: Hybrid Systems: Computation and Control. Proc. 12th Intl. Conf. (HSCC), LNCS, Vol. 5469, Springer-Verlag, 2009, pp. 31–45.

[3] A. Gupta, C. Langbort, T. Basar, Optimal control in the presence of an intelligent jammer with limited actions, in: 49th IEEE Conference on Decision and Control (CDC), 2010, pp. 1096–1101. doi:10.1109/CDC.2010.5717544.

[4] G. Liang, J. Zhao, F. Luo, S. R. Weller, Z. Y. Dong, A review of false data injection attacks against modern power systems, IEEE Transactions on Smart Grid 8 (4) (2017) 1630–1638.

[5] Y. Mo, B. Sinopoli, Secure control against replay attacks, in: 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2009, pp. 911–918.

[6] Y. Mo, R. Chabukswar, B. Sinopoli, Detecting integrity attacks on scada systems, IEEE Transactions on Control Systems Technology 22 (4) (2014) 1396–1407. doi:10.1109/TCST.2013.2280899.

[7] Y. Mo, S. Weerakkody, B. Sinopoli, Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs, IEEE Control Systems Magazine 35 (1) (2015) 93–109. doi:10.1109/MCS.2014.2364724.

[8] S. Amin, X. Litrico, S. Sastry, A. M. Bayen, Cyber security of water scada systems–part 1: Analysis and experimentation of stealthy deception attacks, IEEE Transactions on Control Systems Technology 21 (5) (2012) 1963–1970.

[9] C. M. Ahmed, V. R. Palleti, A. P. Mathur, WADI: A water distribution testbed for research in the design of secure cyber physical systems, in: Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWATER '17, 2017, pp. 25–28.

[10] V. R. Palleti, V. K. Mishra, C. M. Ahmed, A. Mathur, Can replay attacks designed to steal water from water distribution systems remain undetected?, ACM Trans. Cyber-Phys. Syst. 5 (1). doi:10.1145/3406764.
URL `https://doi.org/10.1145/3406764`

[11] P. V. Overschee, B. D. Moor, Subspace identification for linear systems: theory, implementation, applications, Boston: Kluwer Academic Publications.

[12] C. Murguia, J. Ruths, Characterization of a cusum model-based sensor attack detector, in: 55th IEEE Conference on Decision and Control Conference (CDC), 2016.

[13] K. J. Aström, B. Wittenmark, Computer-controlled Systems (3rd Ed.), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1997.

[14] B. Satchidanandan, P. R. Kumar, On the design of security-guaranteeing dynamic watermarks, IEEE Control Systems Letters 4 (2) (2020) 307–312. doi:10.1109/LCSYS.2019.2925278.

[15] R. F. Bullers, How much coal does that valve burn?, https://content2.smcetech.com/pdf/BP5_How_Much_Coal_Does_That_Valve_Burn.pdf (January 2011).

[16] C. M. Ahmed, M. Ochoa, J. Zhou, A. P. Mathur, R. Qadeer, C. Murguia, J. Ruths, Noiseprint: Attack detection using sensor and process noise fingerprint in cyber physical systems, in: Proceedings of the 2018 on Asia Conference on Computer and Communications Security, ASIACCS '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 483–497. doi:10.1145/3196494.3196532.
URL https://doi.org/10.1145/3196494.3196532

[17] S. V. Radhakrishnan, A. S. Uluagac, R. Beyah, Gtid: A technique for physical device and device type fingerprinting, IEEE Transactions on Dependable and Secure Computing 12 (5) (2015) 519–532. doi:10.1109/TDSC.2014.2369033.

[18] J. Slay, M. Miller, Lessons learned from the maroochy water breach, in: E. Goetz, S. Shenoi (Eds.), Critical Infrastructure Protection, Springer US, Boston, MA, 2008, pp. 73–82.

[19] S. Weinberger, Computer security: Is this the start of cyberwarfare?, Nature 174.

[20] D. E. Whitehead, K. Owens, D. Gammel, J. Smith, Ukraine cyber-induced power outage: Analysis and practical mitigation strategies, in: 2017 70th Annual Conference for Protective Relay Engineers (CPRE), 2017, pp. 1–8. doi:10.1109/CPRE.2017.8090056.

[21] A. Khazraei, H. Kebriaei, F. R. Salmasi, A new watermarking approach for replay attack detection in lqg systems, in: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 2017, pp. 5143–5148.

[22] A. Khazraei, H. Kebriaei, F. R. Salmasi, Replay attack detection in a multi agent system using stability analysis and loss effective watermarking, in: 2017 American Control Conference (ACC), 2017, pp. 4778–4783.

[23] B. Satchidanandan, P. R. Kumar, Dynamic watermarking: Active defense of networked cyber-physical systems, Proceedings of the IEEE 105 (2) (2017) 219–240.

[24] H. Liu, Y. Mo, J. Yan, L. Xie, K. H. Johansson, An online approach to physical watermark design, IEEE Transactions on Automatic Control 65 (9) (2020) 3895–3902. doi:10.1109/TAC.2020.2971994.

[25] J. Rubio-Hernan, L. De Cicco, J. Garcia-Alfaro, Revisiting a watermark-based detection scheme to handle cyber-physical attacks, in: 2016 11th International Conference on Availability, Reliability and Security (ARES), 2016, pp. 21–28.

[26] J. Rubio-Hernan, L. De Cicco, J. Garcia-Alfaro, On the use of watermark-based schemes to detect cyber-physical attacks, EURASIP Journal on Information Security 2017 (1) (2017) 1–25.

[27] C. Fang, Y. Qi, P. Cheng, W. X. Zheng, Cost-effective watermark based detector for replay attacks on cyber-physical systems, in: 2017 11th Asian Control Conference (ASCC), 2017, pp. 940–945. doi:10.1109/ASCC.2017.8287297.

[28] R. M. Ferrari, A. M. Teixeira, Detection and isolation of replay attacks through sensor watermarking, IFAC-PapersOnLine 50 (1) (2017) 7363–7368, 20th IFAC World Congress.