Can Replay Attacks Designed to Steal Water from Water Distribution Systems Remain Undetected?

Venkata Reddy Palleti
Indian Institute of Petroleum and Energy-Visakhapatnam
Singapore University of Technology and Design
venkat_palleti.che@iipe.ac.in

Vishrut Kumar Mishra
Singapore University of Technology and Design
kumar_mishra@sutd.edu.sg

Chuadhry Mujeeb Ahmed
Singapore University of Technology and Design,
University of Strathclyde
chuadhry@alumni.sutd.edu.sg

Aditya Mathur
Singapore University of Technology and Design
aditya_mathur@sutd.edu.sg

ABSTRACT

Industrial Control Systems (ICS) monitor and control physical processes. ICS are found in, among others, critical infrastructures such as water treatment plants, water distribution systems, and the electric power grid. While the existence of cyber-components in an ICS leads to ease of operations and maintenance, it renders the system under control vulnerable to cyber and physical attacks. An experimental study was conducted with *replay attacks* launched on an operational water distribution (WADI) plant to understand under what conditions an attacker/attack can remain undetected while stealing water. A detection method, based on an input-output Linear Time-Invariant system model of the physical process, was developed and implemented in WADI to detect such attacks. The experiments reveal the strengths and limitations of the detection method and challenges faced by an attacker while attempting to steal water from a water distribution system.

KEYWORDS

Industrial Control Systems, Water distribution systems, Water Leak, Replay attack, model-based attack detection.

1 INTRODUCTION

Public utilities, often referred to as critical infrastructure, such as a water treatment plant, water distribution system, power generation etc., play an important role in a nation's economy. Industrial Control Systems (ICS) are used for the monitoring and controlling of these infrastructures. An ICS integrates networking, sensing, and computing technologies with the physical systems to provide efficient and reliable operation. Often the physical processes and cyber components (computers and networks) interact with each other in a feedback loop. Also, some of these systems are spatially distributed rendering them vulnerable to cyber and physical attacks [20]. Several attacks on water distribution systems have been reported in recent years. Two of the often cited such attacks include one at the Maroochy Shire Water Services [36] and another at the Kemuri Water Company (KWC). Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) has emphasized the importance of understanding these attacks against critical infrastructures [18, 21].

The purpose of a water distribution system is to fulfill customer's demand while ensuring appropriate quality of the delivered water. However, due to the complex nature of these systems, they are vulnerable to attacks such as intentional contamination, leaks, etc. Among such attacks, leakage in water distribution systems is of significant importance for effective management and water quality control [16, 32]. Leaks in water distribution systems could be accidental or intentional. Accidental leaks can occur due to the high pressure in pipelines. In contrast, intentional leaks occur when an attacker wants to steal water from a pipeline [5] . Leaks can be detected through the appropriately installed pressure sensors and flow meters. In order to increase the impact of a physical attack (water leaks in this work), an attacker might compromise a subset of sensors and alter their measurements [11]. Therefore, leaks on water distribution systems can be considered as a cyber-physical attack.

In this paper, we report water leakage experiments conducted on an operational water distribution plant named WADI [8]. The leaks are simulated through a valve that bypasses water from the main pipeline. It is assumed that the primary goal of the attacker is to steal water while remaining undetected. Thus, to remain undetected during leak simulation, the attacker launches replay attacks on a strategically selected subset of sensor and actuator readings. However, the choice of sensors and actuators also depends on the knowledge and capabilities of an attacker. The objective of this paper is to experiment with different attacker models and to determine whether an attacker is successful in remaining undetected against the statistical attack detection model.

Contributions: Experimental study in a water distribution system to understand (a) the requirements for detecting cyber and physical attacks with the intention to steal water and (b) effectiveness of a Linear Time Invariant and CUSUM-based method for detecting cyber attacks.

Organization: The remainder of this paper is organized as follows. Literature related to the work reported in this paper is reviewed in Section 2. Architecture and operation of WADI used for experimentation is in Section 3. An attack detection framework, based on Linear Time Invariant model of a water distribution system together with the CUSUM detector, used in this work are described in Section 4. Water leakage experiments and results are in Section 5. Conclusions, that map to the contributions above, derived based on the analysis of the data obtained are in Section 6.

2 RELATED WORK

Supervisory control and data acquisition (SCADA) systems are widely used to monitor and control operations in critical infrastructure assets such as electrical power distribution facilities, oil and gas pipelines, water distribution systems and sewage treatment plants. Several attacks on water distribution systems have been reported in recent years [18]. In 2016, hackers breached a water utility (Kemuri water company) and manipulated systems responsible for water treatment and flow control. A famous SCADA attack incident on Maroochy Water Services in Queensland, Australia caused extensive sewage spilling [36]. It is to be noted that such attacks can have a huge impact on critical infrastructure assets leading to economic loss and service disruption on the consumers. In recent years there is a significant growth on gaining knowledge on critical infrastructure attacks, its impacts and different attack detection mechanisms. Some of these research works are discussed below.

Attack detection in canal networked systems using hydro-dynamic models has been reported in [11, 12]. An investigation into the challenges in the security of control systems when sensor and actuator data are compromised has been reported in [9, 15]. 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 system. Denial of service attack refers to the compromise of the availability of resources by jamming communication channels [10, 19]. 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 [17, 23], false data injection attacks are studied in the power networks assuming an attacker has perfect knowledge of the system model. In [24] authors have demonstrated the effect of replay attacks on the sensor measurements and proposed a methodology to detect such attacks. Various cyber attacks on networked control systems are studied in [37] on a quadruple tank testbed.

There are several works on using the model of a physical process [13, 27, 31]. Most of these works borrow ideas from fault detection literature and has also contributed towards the limitations of fault detectors to be used as attack detectors. Towards that end, secure state estimation has extensively been studied. Recently, a research work in [35] proposed a search algorithm based on Satisfiability Modulo Theory (SMT) to speed up the search of possible sensors sets. An anomaly-based methodology for the detection of a wide range of integrity attacks in cyber-physical critical infrastructure is reported in [28]. As stated previously, leakage in water distribution systems is of significant importance. [14, 40] studies leak detection and isolation in open water channels.

Research on cybersecurity in CPSs through experimental investigation of real testbeds has recently received more attention, especially in public critical infrastructures. Authors in [3] experimentally investigated the impact of various cyber attacks on an operational water treatment plant. Also, in their subsequent work [2] they designed a distributed mechanism for detecting cyber attacks on a water treatment system by systematically deriving invariants. Recently, in [6, 7, 30, 33] researchers report an investigation into the effectiveness of yet another detection scheme that uses the Kalman filter on an operational water treatment testbed system. Experimental investigation on operational and real testbeds provides how an attacker can perform attacks on these systems to achieve his goal. Attacker's intent realization depends on his/her capabilities such as network and process knowledge. In this work, an operational water distribution testbed is used to show how an attacker performs cyber-physical attacks to realize his intent. The intent of the attacker is assumed as stealing water from the pipeline of a water distribution system.

3 WATER DISTRIBUTION PLANT

Water Distribution (WADI) plant is an operational test bed supplying 10 US gallons/min of filtered water [8]. WADI is open for researchers to study and design the safe and secure large-scale cyber physical systems. It represents a scaled-down version of a large water distribution network in a city. WADI consists of three stages, namely the primary grid (P1), secondary grid (P2), and return water grid (P3). Figure 1 is an overview of the processes involved in the WADI testbed. The arrows in this figure show the flow of the water from one stage to another stage. 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. Water quality analyzers are installed in return water grid to check water quality before pumping it into the primary grid. In order to measure the water levels in tanks, level sensors, namely, 1_LT_001, 2_LT_002 and 3_LT_001 are installed in the stages P1, P2 and P3 respectively. The detailed architecture of the WADI testbed was explained in [8].

The list of all sensors (i_s) and actuators (i_a) is in Table 1 and is segregated by the different stages. The list of sensors and actuators used in our experiments is shown in the last column of the table.

In WADI, P2b is equipped with a leak detection line. Water leak can be simulated by a transparent double containment piping section fitted with a modulating control valve (2_MCV_007) that opens into the outer pipe. Valve 2_MV_008 is used to bypass water from the main pipeline section and the bypass water drains into the return water grid. Figure 2 shows a picture of the leak setup in WADI. The process flow diagram of the leak setup is represented in Figure 3. Two pressure meters, 2_PIT_001 and 2_PIT_002 are also installed in P2a and P2b, respectively. These sensors are located at, respectively, upstream and downstream of the leak point. A flow meter 2_FIT_002 is installed on the gravity line to measure the amount of water flowing into the consumer tanks.

Three Programmable Logic Controllers (PLCs) are installed to control each stage of WADI. These PLCs use National Instruments Compact RIO as RIO (Remote Input Output) devices. The communication network contains three layers namely, layer-0 (L0), layer-1 (L1) and layer-2 (L2). L0 is at the process level and connects actuators/sensors and I/O modules via RS485-Modbus protocol. L1 is the plant control network where all PLCs are connected to a central node in a star topology. The third layer L2 is a communication network between a touch panel Human-Machine Interface (HMI) and the plant control network.

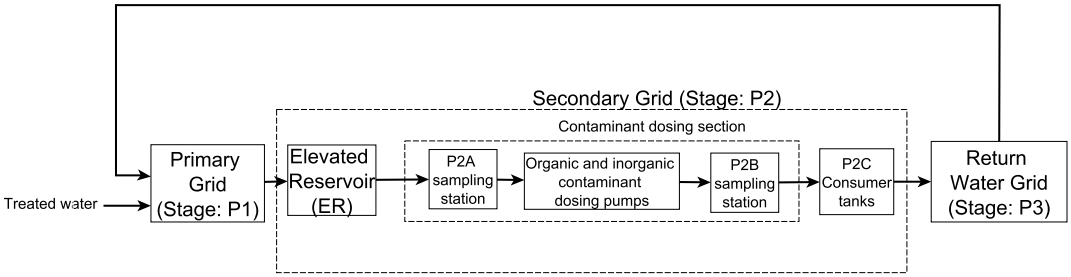


Figure 1: Three stages in WADI are shown. Solid arrows indicate flow of water and sequence of processes.

Table 1: List of sensors and actuators in WADI

Stages	Sensors (i_s)				Actuators (i_a)		
	Level	Flow	Pressure	Water quality	Valve	Pumps	Used sensors and actuators
P1	1_LT_001	1_FIT_001	None	1_AIT_001, 1_AIT_002, 1_AIT_003, 1_AIT_004, 1_AIT_005	1_MV_001, 1_MV_004, 1_MV_005	1_P_001, 1_P_002, 1_P_003, 1_P_004, 1_P_005, 1_P_006	1_LT_001, 1_FIT_001, 1_MV_001, 1_P_005
P2A	2_LT_002	2_FIT_001	2_PIT_001	2A_AIT_001, 2A_AIT_002, 2A_AIT_003, 2A_AIT_004, 2A_AIT_005	2_MV_003, 2_MV_004, 2_MV_005	None	2_LT_002, 2_FIT_001, 2_PIT_001, 2_MV_00x, where x= {1,2,3,4,6}
P2B	None	2_FIT_002, 2_FIT_003	2_PIT_002	2B_AIT_001, 2B_AIT_002, 2B_AIT_003, 2B_AIT_004, 2B_AIT_005	2_MV_008, 2_MV_009, 2_MCV_007, 2_MCV_008	2_P_003, 2_P_004	2_FIT_002, 2_FIT_003, 2_PIT_002, 2_P_003
P2C	2_LS_x01, where x= {1,…,6}	2_FQ_x01, where x= {1,…,6}	None	None	2_MCV_x01, 2_MV_x01, where x= {1,…,6}	None	2_MCV_x01, where x= {1,…,6}
P3	3_LT_001	3_FIT_001	None	3_AIT_001, 3_AIT_002, 3_AIT_003, 3_AIT_004, 3_AIT_005	3_MV_001, 3_MV_002, 3_MV_003, 3_MV_004	3_P_001, 3_P_002, 3_P_003, 3_P_004	3_MV_002

4 SYSTEM MODEL BASED ATTACK DETECTION FRAMEWORK

In this section, a system-model based attack detection framework is explored. First, a system-model is derived based on state dynamics followed by its validation. A statistical detector namely CUmulative SUM (CUSUM) is used for attack detection.

4.1 System model

A system model represents the dynamics of a physical process as a mathematical model. Figure 1 shows the physical process in WADI. The state of the system is measured using sensors. Actuators are controlled by the PLCs to affect process state. Control actions are governed by user demands labeled as consumer tanks in Figure 1. For example, for an elevated tank, the level of water is considered as its state measured by a level sensor. Such a state in a water distribution process is dynamic and is governed by the actuators at the inlet and outlet of the process. We can consider these actuation signals as inputs to the control system and sensor measurements as outputs. The goal is to obtain a system model of the following form,

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

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. 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. It is possible to predict the system state for a physical process with a precise analytic model. The system model can be used to model the normal behavior of a dynamic physical process, and in the case of anomalies, system behavior can be observed as deviating from the expected.

Figure 1 shows the infrastructure for WADI testbed. The main components are primary grid storage tanks, elevated reservoir and

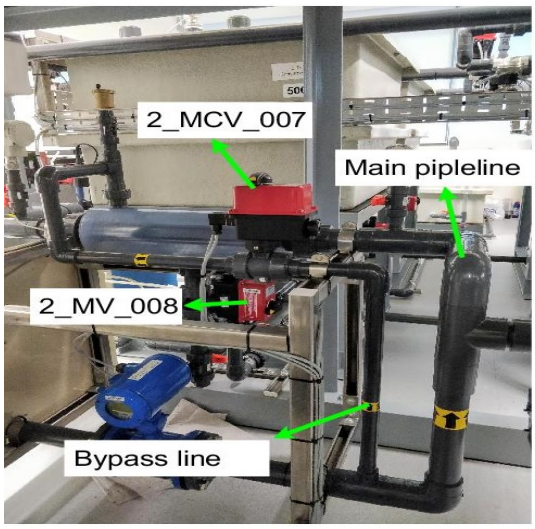


Figure 2: Leak setup in WADI.

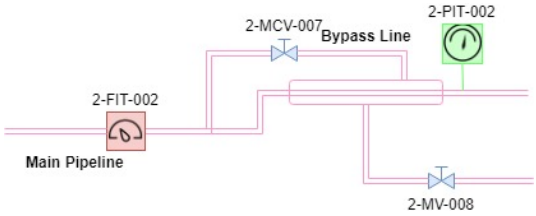


Figure 3: Process flow diagram of leak setup in WADI.

the six consumer nodes. The consumer nodes have time-varying demand patterns based on their water requirements. The controller knows the demand patterns, the water level in the tank, flow measurements and the pressures at the junctions. The data is collected for all measurable outputs and the input demands for 14 days. Measurements are recorded each second which is also used as sampling time for generating the state space model. The flows, pressures and water level measurements are used as outputs of the control system and the demands of the user nodes and the pump status are used as inputs to generate a state space model of the system. Using data collected under regular operation (no attacks) and subspace identification techniques [29], the input-output dynamical model of the WDN is approximated as a set of Linear Time Invariant (LTI) stochastic difference equations in Equation 1.

The system identification problem is to determine the system matrices A, B, C from input-output data. The model so obtained provides a good fit between measurements and simulated outputs, generated using the approximated model, with 10 states, i.e., $n = 10$; the matrices are reproduced in Appendix A. A few higher and lower order models were also identified. Eventually the model with 10 states has an acceptable trade-off between prediction error and its dimension. The quality of the identified model is validated by considering the system state evolution based on the identified state space matrices and initial state x_1. The proximity of the system evolution to the sensor measurements obtained from WADI indicates that this model is a faithful representation of the water distribution network (see Figure 4). The top pane in the figure shows the sensor readings from WADI as well as the estimated output by the model for the water level sensor using system matrices. One may observe that the modeled output is close to the sensor readings resulting in small residual shown in the bottom pane.

4.2 Attack detection framework

In this section, details of the proposed attack detection scheme are described. First, we discuss the Kalman filter based state estimation and residual generation. Next, we present a residual-based attack detection procedure, namely the CUSUM detector.

4.2.1 Kalman filter. To estimate the state of the system based on the available output y_k, we use a linear filter with the following structure:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L_k\big(\bar{y}_k - C\hat{x}_k\big), \tag{2}$$

with estimated state $\hat{x}_k \in \mathbb{R}^n$, $\hat{x}_1 = E[x(t_1)]$, where $E[\cdot]$ denotes expectation, and gain matrix $L_k \in \mathbb{R}^{n\times m}$. Define the estimation error $e_k := x_k - \hat{x}_k$. In the Kalman filter, the matrix L_k is designed to minimize the covariance matrix $P_k := E[e_k e_k^T]$ (in the absence of attacks).

From (2), we can get an overview of the system model where the Kalman filter is being used for estimation. The estimator makes an estimate at each time step based on the previous readings up to x_{k-1} and the sensor reading y_k. the estimator gives $\hat{x}_k$ as an estimate of state variable x_k. Thus, an error can be defined as,

$$e_k = \hat{x}_k - x_k, \tag{3}$$

where $\hat{x}_{k|j}$ denotes the optimal estimate for x_k given the measurements $y_1, ..., y_j$. Let P_k denote the error covariance, $Cov(e_k) = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$, and $\hat{P}_{k|j}$ the estimate of P_k given $y_1, ..., y_j$. Prediction equation for state variable using Kalman filter can be written as,

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} \tag{4}$$

$$P_{k+1|k} = AP_{k|k}A^T + Q, \tag{5}$$

where $\hat{x}_{k|k}$ is the estimate at time step k using measurements up to time k and $\hat{x}_{k+1|k}$ is $(k+1)^{th}$ prediction based on previous k measurements. Similarly, $P_{k|k}$ is the error covariance estimate until time step k. Q is the process noise covariance matrix. The next step in Kalman filter estimation is time update step using Kalman gain L_k.

$$L_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} \tag{6}$$

$$\bar{x}_{k+1|k} = \hat{x}_{k+1|k} + L_k(y_k - C\hat{x}_{k+1|k}) \tag{7}$$

$$\bar{P}_{k+1|k} = (I - L_kC)\hat{P}_{k+1|k}, \tag{8}$$

where $\bar{x}_{k+1|k}$ and $\bar{P}_{k+1|k}$, are the updates for the $k+1$ time step using measurements y_i from the i^{th} sensor and Kalman gain L_k. R is the measurement noise covariance matrix. The initial state can be selected as $x_0 = x_0$ with $P_0 = E[(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T]$. Kalman gain L_k is updated at each time step but after a few iterations it converges and operates in a steady state. Kalman filter is an iterative estimator and $\hat{x}_{k|k}$ in Eqn. (4) comes from $\bar{x}_{k-1|k}$ in Eqn. (7). It is assumed that

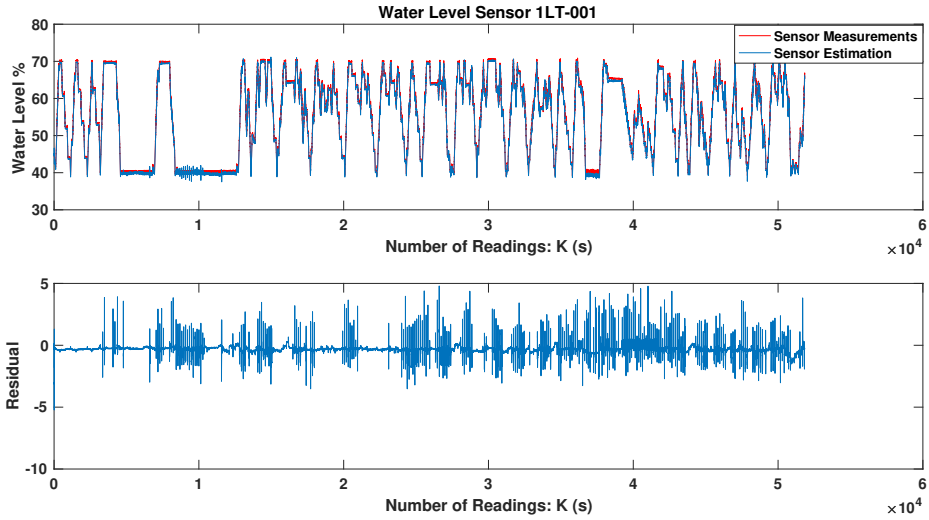


Figure 4: System Model Validation. The top pane shows the sensor readings from WADI as well as the estimated output by the model for the water level sensor using system matrices. It can be observed that the modeled output is very close to sensor readings, resulting in small residual shown in the bottom pane.

the system is in a steady state before attacks are launched. Kalman filter gain is represented by L in steady state.

4.2.2 Residuals and hypothesis testing. The model based fault detection is common in fault detection literature [39]. Using residual vector for fault detection focuses on a particular structure of the faults. However, an intelligent adversarial attack is challenging to detect. In this study, the focus is on the performance of the residual based anomaly detection scheme against strategically designed attacks, such as for example, replay attacks [6, 25]. In this work, we assess the performance of the CUSUM model-based fault detection procedure for a variety of attacks. These procedures rely on a state estimator, e.g., Kalman filter, to predict system state. The estimated state is compared with sensor measurements $\bar{y}_k$ which may have been compromised. The difference between the two should stay within a certain threshold under normal operation, otherwise, an alarm is triggered. The residual random sequence $r_k, k \in \mathbb{N}$ is defined as

$$r_k := \bar{y}_k - C\hat{x}_k = Ce_k + \eta_k + \delta_k. \tag{9}$$

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

$$E[r_{k+1}] = CE[e_{k+1}] + E[\eta_{k+1}] = \bar{r}_{m\times 1}. \tag{10}$$

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

$$\Sigma := E[r_{k+1}r_{k+1}^T] = CPC^T + R_2. \tag{11}$$

For this residual, we identify two hypothesis to be tested, $\mathcal{H}_0$ the *normal mode*, i.e., no attacks, and $\mathcal{H}_1$ the *faulty mode*, i.e., with attacks. In our case study, the pressure at the nodes and the water level in the tank are the outputs of the system. The residuals are obtained using this data together with the state estimates. Thus we can state the two hypotheses as follows.

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

In the following, we formulate the hypothesis testing in a more formal manner using existing change detection techniques based on the statistics of the residuals.

4.2.3 Cumulative Sum (CUSUM) detector. The residual vector is given as an input to the CUSUM procedure, also known as the stateful detector. The input to the CUSUM procedure can be considered as a *distance measure*, i.e., a measure of how far the estimate is from the expected measurements. For this work, all the sensors and actuators are considered together while creating a system model. The Kalman filter gain is presented in form of a matrix as all the sensors together constitute the system outputs. A dedicated detector for each sensor is designed. The index i denotes the sensor/detector, $i \in \mathcal{I} := \{1, 2, \ldots, m\}$, where m is the number of sensors. Thus, the attacked output vector can be partitioned as, $\bar{y}_k = \text{col}(\bar{y}_{k,1}, \ldots, \bar{y}_{k,m})$ where $\bar{y}_{k,i} \in \mathbb{R}$ denotes the i-th entry of $\bar{y}_k \in \mathbb{R}^m$; then

$$\bar{y}_{k,i} = C_i x_k + \eta_{k,i} + \delta_{k,i}, \tag{12}$$

with C_i denotes the i-th row of C, and $\eta_{k,i}$ and $\delta_{k,i}$ denote the i-th entries of η_k and δ_k, respectively. δ_k is the value added by the attacker to the sensor measurements to achieve a specific attack objective. The residual vector for each sensor can be given as,

$$r_{k,i} = C_i e_k + \eta_{k,i} + \delta_{k,i}. \tag{13}$$

The standard CUSUM [26] procedure is explained using the following equations.

CUSUM: $S_{0,i}^- = 0, \;\; S_{0,i}^+ = 0, \;\; \tilde{k}_i^+ = 0, \;\; \tilde{k}_i^- = 0,$

$$\begin{cases} S_{k,i}^+ = \max(0, S_{k-1,i}^+ + r_{k,i} - \bar{T}_i - \kappa_i), & \text{if } S_{k-1,i}^+ \leq \tau_i^+, \\ S_{k,i}^+ = 0 \;\text{ and }\; \tilde{k}_i^+ = \tilde{k}_i^+ + 1, & \text{if } S_{k-1,i}^+ > \tau_i^+. \end{cases} \tag{14}$$

$$\begin{cases} S^{-}_{k,i} = \min(0, S^{-}_{k-1,i} + r_{k,i} - \bar{T}_i + \kappa_i), & \text{if } S^{-}_{k-1,i} \geq \tau^{-}_i, \\ S^{-}_{k,i} = 0 \text{ and } \tilde{k}^{-}_i = \tilde{k}^{-}_i + 1, & \text{if } S^{-}_{k-1,i} < \tau^{-}_i. \end{cases} \quad (15)$$

Design parameters: Bias $\kappa_i > 0$; threshold $\tau_i > 0$.
Output: $Alarm(s) = \tilde{k}^{+}_i + \tilde{k}^{-}_i$.

From (14)-(15), it can be observed that $S^{+}_{k,i}$ and $S^{-}_{k,i}$ accumulate the distance measure $r_{k,i}$ over time to measure how far are the values of the residual from the target mean ($\bar{T}_i$). To tune the CUSUM detector there is also a slack variable κ chosen to be $\frac{1}{2} * \sigma_i$ in this study. $\tau_i = \pm\Gamma * \sigma_i$, where Γ is a multiplier to the standard deviation (σ) and usually taken between 3 and 5 [26]. An alarm is raised when this accumulation becomes greater or less than a chosen threshold τ_i. The sequence $S_{k,i}$ is reset to zero each time it becomes negative or larger than τ_i. If r_k, i is tightly bounded and κ_i is not sufficiently large, the CUSUM sequence $S_{k,i}$ grows unbounded until the threshold τ_i is reached, no matter how large τ_i is set. In order to prevent such drifts, the slack variable κ_i must be selected properly based on the statistical properties of the distance measure. Once κ is chosen, the threshold τ_i must be selected to achieve a required false alarm rate $\mathcal{A}^{*}_i$. $\mathcal{A}_i \in [0, 1]$ denotes the *false alarm rate* for the CUSUM procedure defined as the expected proportion of observations which are false alarms [1, 38]. In this study a false alarm rate of 5% is considered under the normal operation. Table 2 shows a list of CUSUM parameters for all the sensors used in this study. Following metrics are used while assessing the effectiveness of the attack detection procedure.

- *True Positive Rate (TPR)*: This rate is defined as the ratio of data detected as attack to the total number of observations when the anomaly actually existed.
- *False Positive Rate (FPR)*: This is defined as the ratio of data detected as attack to the total number of observations when the plant was under 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.

Table 2: CUSUM design parameters

Sensors	Design parameters		
	Bias (κ)	Threshold (upper limit) τ^{+}	Threshold (Lower limit) τ^{-}
1_LT_001	0.1231	0.9852	-0.9852
2_LT_002	0.0873	0.6988	-0.6988
2_PIT_001	0.106	0.848	-0.848
2_PIT_002	0.1412	1.13	-1.13
1_FIT_001	0.0186	0.1488	-0.1488
2_FIT_001	0.0361	0.2888	-0.288
2_FIT_002	0.007	0.0564	-0.0564
2_FIT_003	0.0155	0.1244	-0.1244

5 WATER LEAKAGE EXPERIMENTS

5.1 Attacker model and attack Design

A generalized CPS attacker model was considered [3] for the design of attacks. An attack in this model consists of a finite set of intent, e.g., goals, and targeted components of the CPS. The intents may include damage/degrade the performance of an instrument, and/or reduce the production of a process plant, etc. In this work, it is assumed that the attacker's intention is to fully or partially cut-off water supply to the consumers. The attacker's objective could be manifold–either to cause service disruption or to steal water [11]. To realize these objectives the attacker performs a range of cyber-physical attacks on one of the main pipeline sections of a functional water distribution network. The system model developed in Section 4 is used to detect attacks in all of the following attack scenarios.

5.2 Attack scenarios

Following three types of attack scenarios are considered depending on the knowledge and capabilities of the attacker.

- Naive Attack (NA): In this attack scenario, an attacker is assumed who has limited knowledge about the water distribution system. However, the attacker has physical access and would try to leak the water from the physical water link.
- Cyber-Physical Attack (CPA): A cyber and physical attacker is assumed to have access to the cyber and physical domains. Besides physically cutting off the water supply he also replays the previous sensor readings to stay hidden.
- Powerful Attack (PA): In this scenario, an attacker knows the system dynamics, the state space system model, the control inputs and outputs and the model based attack detection technique. The attacker is capable of executing both cyber and physical attacks. Therefore, to increase the impact of the attack, and remain undetected, the attacker chooses to launch cyber domain (replay) attack not only on sensors but also on actuators while physically cutting-off the water.

5.3 Attack tool

WADI uses a multi-layered network comprising different protocols at different levels and between different devices. In the experiments described here, the focus is on the National Instruments Publish-Subscribe Protocol (NI-PSP). NI-PSP is widely used in the WADI network and provides access to all data on the network. A tool named NiSploit [4] was used in the experiments to launch attacks on WADI. NiSploit uses custom LabVIEW Virtual Instruments (VIs) that can communicate with shared variables [22] in different PLCs using NI-PSP.

5.4 Normal plant operation

Experiments were conducted under normal operating conditions, i.e., without launching any attacks, to obtain a system baseline behavior. 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.The same set of initial conditions have been used for all the experiments reported in this paper. 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 4. Based on the system model the residual (r_k) for each sensor (i) was generated using the Kalman filter based state estimation. Figure 6a shows the residual vector under normal operation of the plant for the pressure sensors. For the sake of brevity, plots are omitted for the remainder of the sensors; a complete list of threshold and bias selection for CUSUM is included in Table 2. These residuals are tested against the calculated thresholds for CUSUM procedure as explained in the previous section.

During normal operation, False Positive (FP) rates ($\mathcal{A}_i^*$) are shown in Table 3. Three runs of normal operation were experimented with to validate the system model. As there were no attacks, any alarm raised by the CUSUM detector is considered as a false positive. For example in Run 1 for 1_LT_001, out of 7200 data points, 27 points (0.0379) are flagged as alarms. It is observed that these experiments have similar FP rates for different runs. Under normal operation, this false alarm rate validates the stability of system model and the CUSUM based detector.

Table 3: False positive rate under normal experiments

Sensors	False Positive Rate		
	Run 1	Run 2	Run 3
1_LT_001	0.0379	0.0428	0.0399
2_LT_002	0.0372	0.0370	0.0646
2_PIT_001	0.1049	0.0345	0.0501
2_PIT_002	0.1145	0.0492	0.1152
1_FIT_001	0.0089	0.0041	0.0044
2_FIT_001	0.0193	0.0142	0.0211
2_FIT_002	0.0609	0.0359	0.1177
2_FIT_003	0.0582	0.0238	0.0655

5.5 Leak execution (Naive attack)

A naive attacker performs an attack to steal water from one of the main pipeline sections. Since the attacker has only physical access to a portion of the plant, i.e., a water distribution pipe, the attacker executes four different types of attacks in an effort to remain undetected. Note that valve 2_MCV_007 can be used to cut-off water supply. Steps described in Part (a) of Figure 5 are used for introducing a leak in the plant.

Before starting the leak process, for the sake of conserving water, another valve labeled as 2_MV_008 (shown in Figure 2) is kept open so that the leaked/stolen water can drain into the return water grid of the plant. 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. The change could be observed by the operator or using pre-installed fault detectors. Thus, an attacker can launch any one of the following attacks labeled as $L1$, $L2$, $L3$ or $L4$. For each of these attacks, the CUSUM based attack detection is used to detect the abnormal behavior of the system.

(1) *Gradual increase and abrupt decrease (L1)*: In this attack scenario, the leak process begins by gradually opening the valve 2_MCV_007. The position of the valve was increased by 5% every 60 seconds until it reached 95%. Once it reached 95%, the system was run for 10 minutes and then the valve closed fully and abruptly.
(2) *Gradual increase and gradual decrease (L2)*: This experiment was performed by gradually opening valve 2_MCV_007. The position of the valve was increased by 5% every 60 *seconds* until it reached 95%. At this position, the system was allowed to run for 10 *minutes*. The valve was then closed gradually i.e., the position was decreased by 5% every 60 *seconds* until it reached 0%.
(3) *Abrupt increase gradual decrease (L3)*: In this experiment the valve (2_MCV_007) was opened abruptly, i.e, the position of the valve (2_MCV_007) set to 100%. The plant was left running for 10,minutes and then the valve was closed gradually. The position was decreased by 5% every 60 seconds until it reached 0%, i.e., until completely closed.
(4) *Abrupt increase and abrupt decrease (L4)*: In this leak experiment, the leak process started by opening the valve (2_MCV_007) completely, i.e, the position of valve 2_MCV_007 was set to 100%. The system continued to run for 10 *minutes* in this state and then the valve was immediately closed completely, i.e, the position of the valve (2_MCV_007) set to 0%.

Results and discussion. Table 4 shows the FPs and TPs for the four leak experiments. It can be seen that sensors 2_PIT_001, 2_PIT_002, 2_FIT_001, and 2_FIT_002 show the high true positive rate. This is due to the fact that these sensors are upstream and downstream from the leak point and are most affected by the leak. There are other sensors in the table which were never affected by the attack; for example, level sensors 1_LT_001 and 2_LT_002. However, these level sensors are included for comparison with the sensors under attack. The affected sensors also show a slightly higher FPs due to the data points that are physically located following the leak attack and carry the effect forward. For example, in experiment L1, the attack start and end points are at k = 2103 and k = 3969, respectively. Where, k is the time at which sensor measurements are recorded. Each sensor measurement is stored at each second. The calculated values of FP and TP for 2_PIT_001 are 0.3641 and 0.8757, respectively. Visual impact of an attack is shown in Figure 6. By comparing Figure 6a (normal operation) with Figure 6b (under attack), one could observe that the sensor residuals deviate significantly from the normal thresholds.

Though the attacker attempts to steal water in four different ways, the attack detection mechanism is able to detect all attacks. The TP and FP alarm rates are found to be identical in all these cases. Therefore, it can be concluded that different procedures for performing the physical attacks has the same impact on the attack detection model. As it is not possible to successfully hide the physical attack alone, the attacker proceeds to perform a cyber attack simultaneously as explained in the next section.

5.6 Leak attacks with replay

Two attacks were executed in this set of experiments. In the first attack, referred to as Replay_1 and Replay_2, the attacker launches a physical attack on the system to steal water while simultaneously launching a cyber-attack to spoof the sensor measurements. The

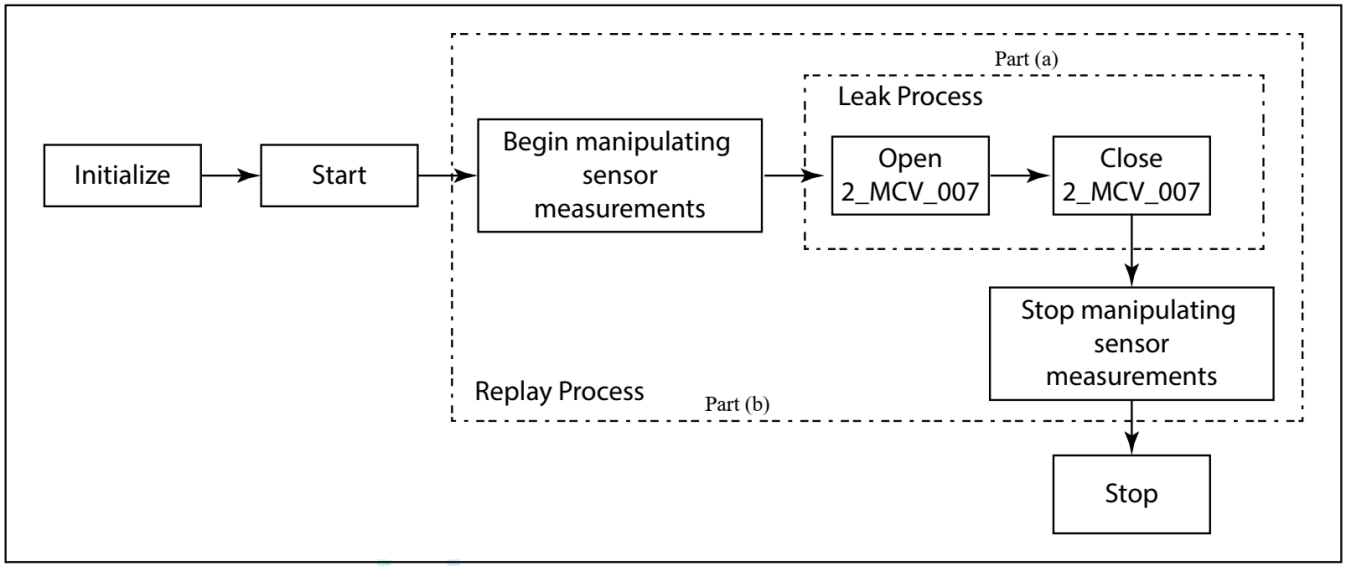
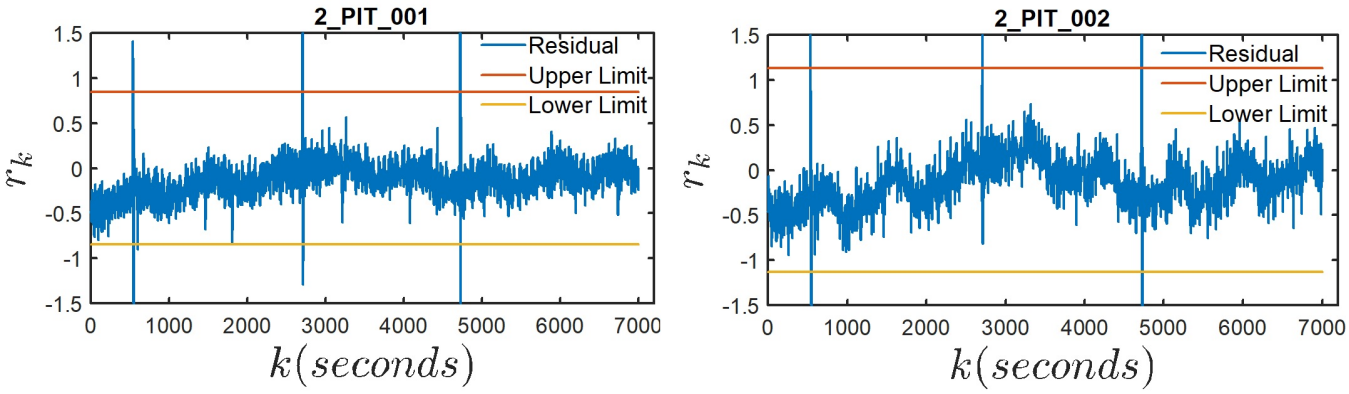
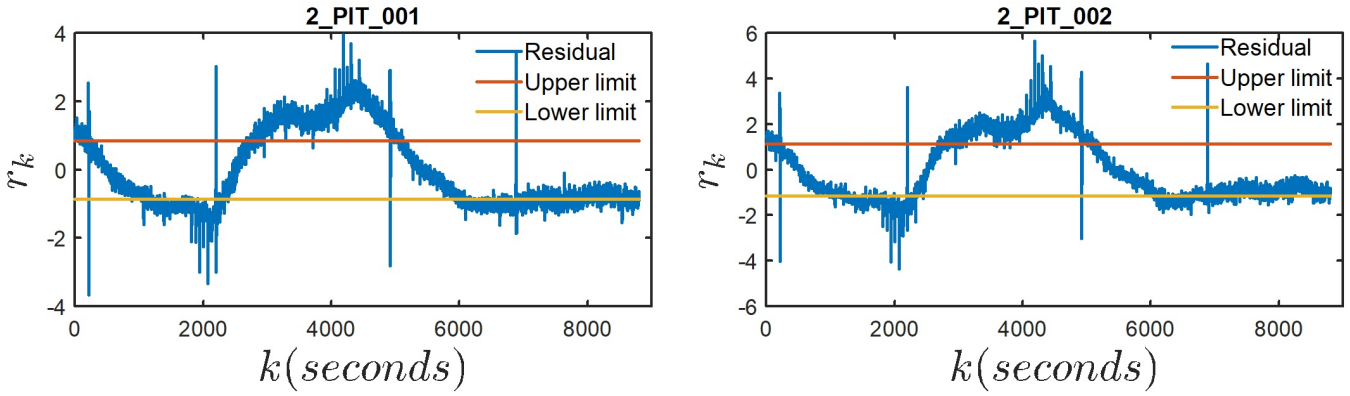


Figure 5: Steps in performing leak and replay attack experiments.



(a) Normal operation: Residual estimation for Run 2 of the normal experiments. Figure also shows the upper and lower limits of the CUSUM detector for each sensor.



(b) Residual under Leak execution

Figure 6: Comparison of residual estimates between normal and naive attack experiment *L2*.

Table 4: Alarm rates under leak experiments

Sensors	L1		L2		L3		L4	
	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR
1_LT_001	0.0038	0.0086	0.0811	0.0721	0.0026	0.0019	0.0028	0.0052
2_LT_002	0.0064	0.0112	0.0472	0.0527	0.0063	0.0142	0.0056	0.0117
2_PIT_001	0.3641	0.8757	0.2192	0.8023	0.2763	0.7389	0.2845	0.6715
2_PIT_002	0.3979	0.9738	0.4303	0.8768	0.3451	0.9062	0.2979	0.9935
1_FIT_001	0.0160	0.0204	0.0729	0.0235	0.0156	0.0210	0.0142	0.0262
2_FIT_001	0.2328	0.0142	0.0262	0.0165	0.3804	0.5298	0.2753	0.7353
2_FIT_002	0.3560	0.8891	0.3994	0.6033	0.2802	0.76501	0.2874	0.7227
2_FIT_003	0.2566	0.4826	0.2569	0.7727	0.2355	0.4587	0.2194	0.6082

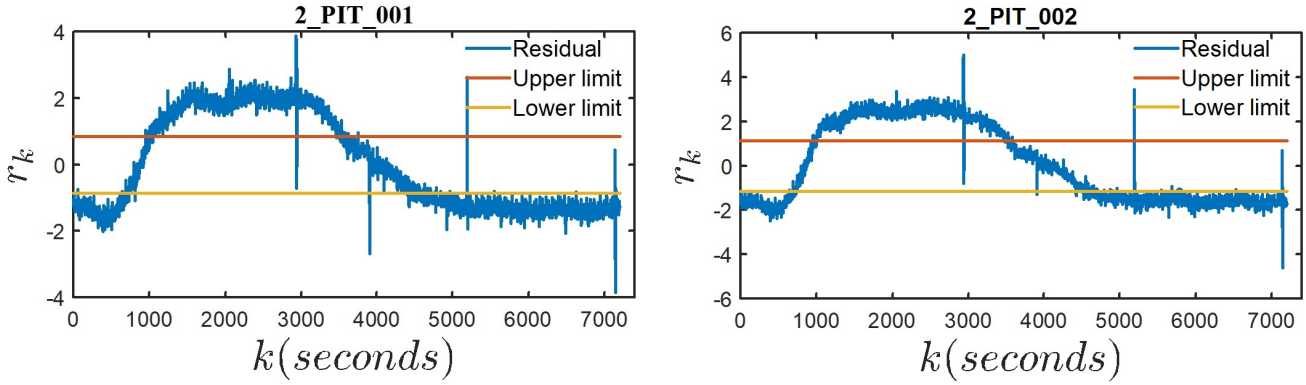


Figure 7: Residual estimates under Replay_1 attack

cyber attack was launched as a replay attack on the sensor measurements. The attacker compromised a set of sensors, observed and recorded their readings for a certain time duration and replayed the recorded measurements while carrying out the attack. These attacks were performed strategically to avoid leak detection. The target replay attack points were chosen based on the attacker's knowledge of the process. Initially, the attacker chooses the upstream pressure sensor, namely 2_PIT_001 and the downstream pressure sensor (2_PIT_002), to perform the replay attack (Replay_1). Then the attacker launches the replay attacks on both pressure sensors and the flow sensor (2_FIT_002) labeled as Replay_2 in Table 5.

The replay attack procedure on sensors is as follows. As shown in part (b) of Figure 5, initially, the replay attack is launched at time k_s. This is followed by the launch of the actual physical attack at time $k_s+\Delta k$, where Δk is the time difference between the launch of replay and leak attacks. These attacks are kept running simultaneously until time k_e, where k_e denotes the end of the leak attack. Even after the leak attack has ended, the attacker continues to replay the sensors for a duration of k_{af} to hide the after-effect of the attack.

Results and discussion. The alarm rates for the replay attacks are listed in Table 5. From the table it is observed that even after performing the replay attack on sensors 2_PIT_001, 2_PIT_002, and 2_FIT_002, the alarm rates are high implying that the CUSUM based detector can detect the attacks. This is due to the fact that the attacker chooses to replay the sensor measurements. However, the model considers actuation signals as input and sensor measurements as outputs as mentioned in Section 4. A visual inspection of Figures 7 and 8 reveals the residual vector of sensors during different replay attack scenarios. From this figure it can be seen that the residuals are out of bounds of the normal operation. As a result, the attacker cannot remain undetected by only replaying the sensor measurements. It can be observed from the results so far that replaying only sensor data is not adequate to hide cyber and physical attacks from the model based detectors.

Table 5: Alarm rates under replay experiments

Sensors	Replay_1		Replay_2		Replay_3	
	FPR	TPR	FPR	TPR	FPR	TPR
1_LT_001	0.0461	0.0689	0.0504	0.0815	0.0311	0.0625
2_LT_002	0.0427	0.0729	0.0295	0.0302	0.0243	0.0162
2_PIT_001	0.2347	0.6889	0.1598	0.4081	0.0822	0.2196
2_PIT_002	0.2416	0.8374	0.3250	0.7238	0.1736	0.0336
1_FIT_001	0.0648	0.0485	0.0813	0.0633	0.0033	0.0037
2_FIT_001	0.2718	0.7066	0.3325	0.5611	0.0617	0.1079
2_FIT_002	0.2064	0.6638	0.3184	0.5448	0.3320	0.3727
2_FIT_003	0.1645	0.6330	0.1503	0.4435	0.1299	0.0772

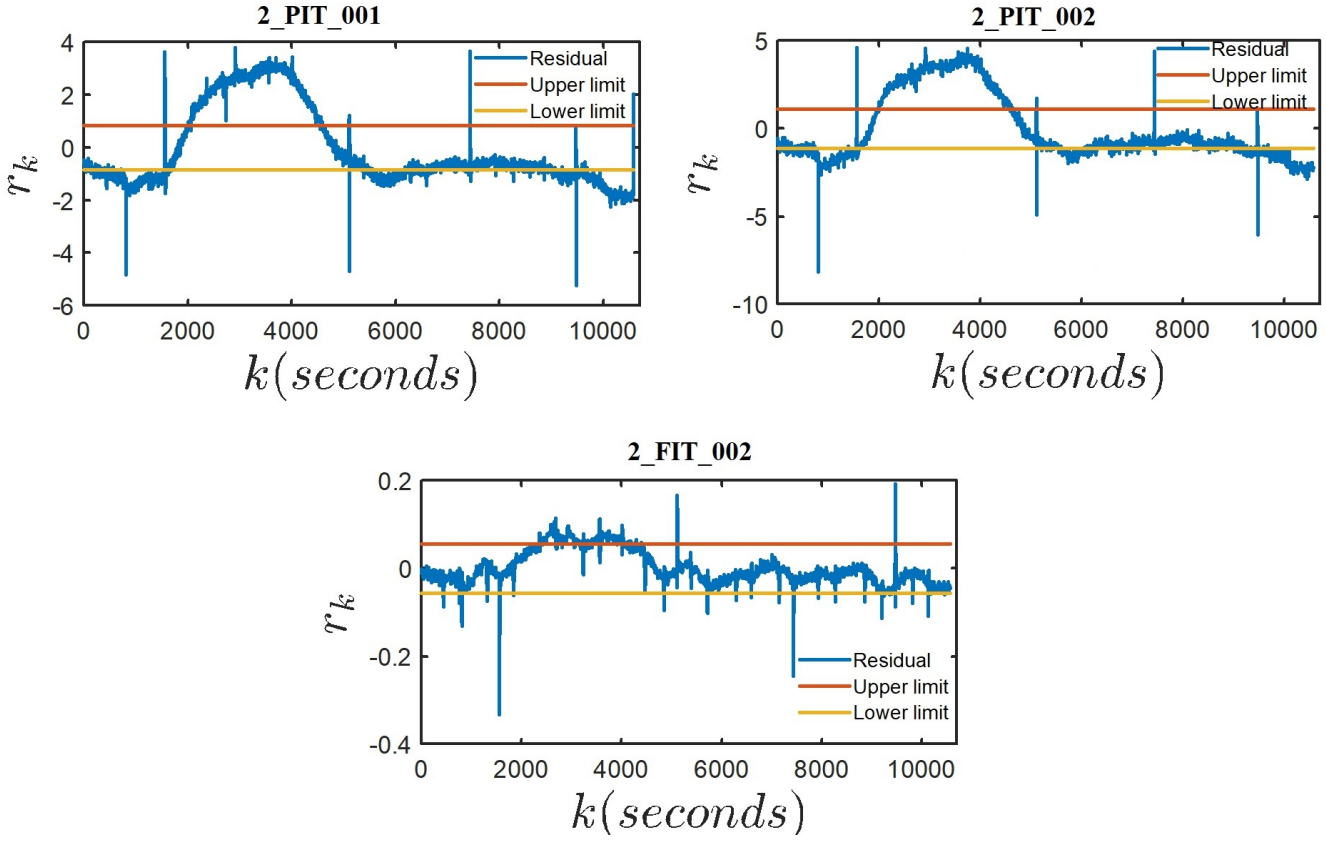


Figure 8: Figure shows estimates of residuals for replay_2 attack

5.7 Powerful attack

A powerful attacker with complete knowledge of the system model and detection mechanism can identify the sensors and actuators on which replay attack needs to be launched. In the previous section we observed that only replaying sensor measurements could not hide the attack from a CUSUM detector. A powerful attacker with complete knowledge of system model, inputs, and outputs, should launch a replay attack not only on sensor (output) measurements but also on the actuators (valves). For this experiment, an attacker launches a replay attack on Modulating Control Valves (MCVs), together with the sensor measurements, as described in Section 5.6.

MCVs in WADI control the input flow to the consumer tanks. The opening position of an MCV is decided based on the current consumer demand. The system model uses this information as one of the inputs. Therefore the attacker chooses to launch replay on these valves, together with other sensors, to mask the attack. Figure 9 shows the residual estimates when the replay is performed on sensors and actuators. It is observed that the residuals do not deviate from the normal operation and are within the thresholds of the detector. Therefore the attack is not detected. Also, the FP and TP rates are low as compared to other two attack scenarios as shown in column labelled "Replay_3" in Table 5. However, it is important to mention that the false alarm rate for the pressure sensors is quite high even during the normal operation as shown in Table 3. The idea is to show how sensor instability can cause issues in designing the detectors. However, it is recommended to design process specific heuristics, for example, based on the frequency of false alarms to reduce the false positive rate and make it viable for the production in the field.

The attacker remains undetected by executing such a powerful attack There exists a strong interdependence between MCVs and sensor measurements. Therefore, in this attack scenario, the attacker succeeds in stealing water without being detected.

6 SUMMARY, CONCLUSIONS AND FUTURE WORK

6.1 Summary

Replay attacks were launched on an operational water distribution system (WADI) to understand the effectiveness of a model-based method in detecting such attacks aimed at stealing water. Three attack scenarios, namely naive, cyber-physical, and powerful attacks were considered. From the experiments, it was observed that the powerful attacker succeeds in stealing water without being detected. It is also observed that if the system model was developed using only the sensor measurements, then the cyber-physical attacker can easily remain undetected. However, as the model is built

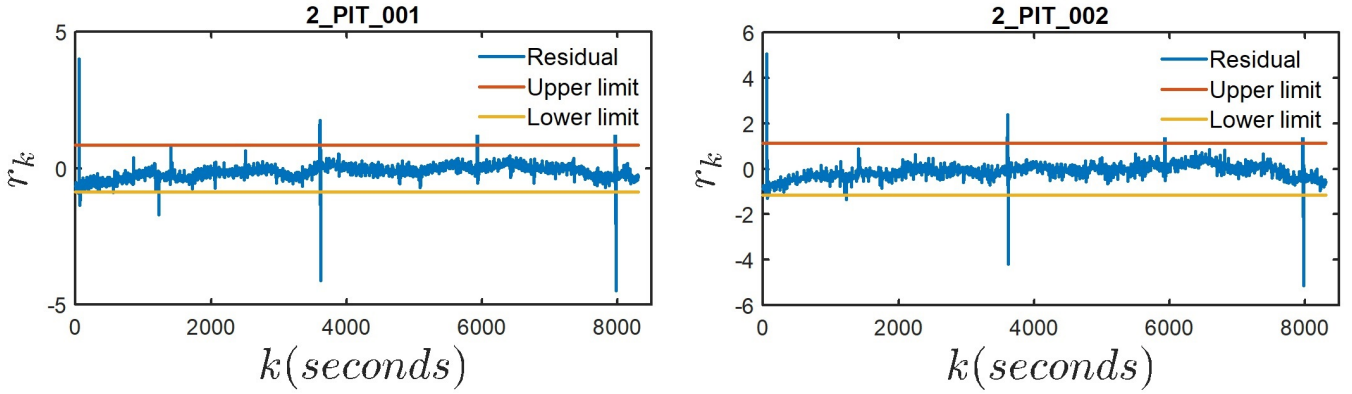


Figure 9: Residual estimation for replay attack on pressure, flow sensors and modulating control valves.

using both sensor and actuator measurements, the cyber-physical attacker fails to mask the attack.

6.2 Conclusions

Analysis of data acquired during experimentation on WADI offers evidence in support of the following claims. (a) The LTI and CUSUM framework is effective in detecting attacks launched by an attacker with incomplete knowledge of the system and of the detection mechanism, and (b) an attacker could remain undetected, and successfully steal water, when the attacks are designed strategically using a complete knowledge of the water distribution system and the detection framework. The experimental evidence does point to the need for a more advanced method for detecting attacks aimed at stealing water in water treatment systems coupled with design enhancements that allow better detection rates.

6.3 Future Work

Effects of randomness in user demands: Consumers' demands are kept constant in this study to avoid variation in the data due to variations in the consumer demand and to highlight the replay attack. Random consumer demands make it harder for the attacker to launch a successful replay attack. One can use this randomness to deceive an attacker and make it even harder to launch a replay attack.

Sensor ageing effects: Another issue is sensor/actuator wear and tear. Although this issue is not considered in the study reported here, it can play a significant role in similar future studies.

Speed of detection: The standard CUSUM method was used such that when an attacker makes a small change in a particular direction of the process, it will take some time for the effect to accumulate and cross the CUSUM threshold. There have been works to detect the changes faster than the standard CUSUM detector [34]. The quickest change detection with observation scheduling in the min-imax setting can be used in future work to include the speed of detection as a parameter.

ACKNOWLEDGMENTS

This work was supported in part by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2015NCR-NCR003-001) and administered by the National Cybersecurity R&D Directorate.

REFERENCES

[1] B.M. Adams, W.H. Woodall, and C.A. Lowry. 1992. *The use (and misuse) of false alarm probabilities in control chart design.* Physica, Heidelberg. 155–168 pages.
[2] Sridhar Adepu and Aditya Mathur. 2016. Distributed detection of single-stage multipoint cyber attacks in a water treatment plant. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. ACM, 449–460.
[3] Sridhar Adepu and Aditya Mathur. 2016. Generalized attacker and attack models for Cyber Physical Systems. In *Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual*, Vol. 1. IEEE, 283–292.
[4] S. Adepu, G. Mishra, and A. Mathur. 2017. Access Control in Water Distribution Networks: A Case Study. In *2017 IEEE International Conference on Software Quality, Reliability and Security (QRS)*. 184–191.
[5] Anand Agrawal, Chuadhry Mujeeb Ahmed, and Ee-Chien Chang. 2018. Poster: Physics-Based Attack Detection for an Insider Threat Model in a Cyber-Physical System. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIA CCS '18)*. Association for Computing Machinery, New York, NY, USA, 821–823.
[6] C. M. Ahmed, S. Adepu, and A. Mathur. 2016. Limitations of state estimation based cyber attack detection schemes in industrial control systems. In *2016 Smart City Security and Privacy Workshop (SCSP-W)*. 1–5.
[7] Chuadhry Mujeeb Ahmed, Carlos Murguia, and Justin 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)*. ACM, New York, NY, USA, 101–113. https://doi.org/10.1145/3052973.3053011
[8] Chuadhry Mujeeb Ahmed, Venkata Reddy Palleti, and Aditya P. Mathur. 2017. 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)*. 25–28.
[9] Chuadhry Mujeeb Ahmed and Jianying Zhou. 2020. Challenges and Opportunities in CPS Security: A Physics-based Perspective. arXiv:cs.CR/2004.03178
[10] S. Amin, A.A. Càrdenas, and S. S Sastry. 2009. 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*. 31–45.
[11] S. Amin, X. Litrico, S. Sastry, and A.M. Bayen. 2013. Cyber Security of Water SCADA Systems; Part I: Analysis and Experimentation of Stealthy Deception Attacks. *IEEE Transactions on Control Systems Technology* 21, 5 (2013), 1963–1970.
[12] S. Amin, X. Litrico, S.S. Sastry, and A.M. Bayen. 2013. Cyber Security of Water SCADA Systems; Part II: Attack Detection Using Enhanced Hydrodynamic Models. *IEEE Transactions on Control Systems Technology* 21, 5 (2013), 1679–1693.
[13] Cheng-Zong Bai and Vijay Gupta. 2014. On Kalman filtering in the presence of a compromised sensor: Fundamental performance bounds. In *2014 American control conference*. IEEE, 3029–3034.

[14] Nadia Bedjaoui and Erik Weyer. 2011. Algorithms for leak detection, estimation, isolation and localization in open water channels. *Control Engineering Practice* 19, 6 (2011), 564 – 573. http://www.sciencedirect.com/science/article/pii/S0967066110001498 SAFEPROCESS 2009.

[15] Alvaro A. Cárdenas, Saurabh Amin, and Shankar Sastry. 2008. 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, Article 6, 6 pages.

[16] Andrew F. Colombo and Bryan W. Karney. 2002. Energy and Costs of Leaky Pipes: Toward Comprehensive Picture. *Journal of Water Resources Planning and Management* 128, 6 (2002), 441–450.

[17] R. Deng, G. Xiao, and R. Lu. 2017. Defending Against False Data Injection Attacks on Power System State Estimation. *IEEE Transactions on Industrial Informatics* 13, 1 (Feb 2017), 198–207.

[18] Department of Homeland Security [n.d.]. ICS-CERT Advisories https://ics-cert.us-cert.gov/advisories.

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

[20] Abdulmalik Humayed, Jingqiang Lin, Fengjun Li, and Bo Luo. 2017. Cyber-Physical Systems Security - A Survey. *CoRR* abs/1701.04525 (2017). arXiv:1701.04525 http://arxiv.org/abs/1701.04525

[21] ICS-CERT-FY2015 [n.d.]. ICS-CERT Advisories. https://ics-cert.us-cert.gov/sites/default/files/Annual_Reports/FY2015_Industrial_Control_Systems_Assessment_Summary_Report_S508C.pdf.

[22] National Instruments. 2018. *Using the LabVIEW Shared Variable.* http://www.ni.com/white-paper/4679/en/

[23] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong. 2017. A Review of False Data Injection Attacks Against Modern Power Systems. *IEEE Transactions on Smart Grid* 8, 4 (July 2017), 1630–1638.

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

[25] Y. Mo, S. Weerakkody, and B. Sinopoli. 2015. Physical Authentication of Control Systems: Designing Watermarked Control Inputs to Detect Counterfeit Sensor Outputs. *IEEE Control Systems Magazine* 35, 1 (Feb 2015), 93–109. https://doi.org/10.1109/MCS.2014.2364724

[26] D.C. Montgomery. 2009. *Introduction to Statistical Quality Control.* Wiley.

[27] C. Murguia and J. Ruths. 2016. Characterization of a CUSUM model-based sensor attack detector. In *2016 IEEE 55th Conference on Decision and Control (CDC)*. 1303–1309. https://doi.org/10.1109/CDC.2016.7798446

[28] S. Ntalampiras. 2015. Detection of Integrity Attacks in Cyber-Physical Critical Infrastructures Using Ensemble Modeling. *IEEE Transactions on Industrial Informatics* 11, 1 (Feb 2015), 104–111.

[29] P. Van Overschee and B. De Moor. 1996. Subspace Identification for Linear Systems: theory, implementation, applications. *Boston: Kluwer Academic Publications* (1996).

[30] Venkata Reddy Palleti, Yu Chong Tan, and Lakshminarayanan Samavedham. 2018. A mechanistic fault detection and isolation approach using Kalman filter to improve the security of cyber physical systems. *Journal of Process Control* 68 (2018), 160 – 170.

[31] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. 2013. Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control* 58, 11 (2013), 2715–2729.

[32] R. Puust, Z. Kapelan, D. A. Savic, and T. Koppel. 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal* 7, 1 (2010), 25–45.

[33] Rizwan Qadeer, Carlos Murguia, Chuadhry Mujeeb Ahmed, and Justin Ruths. 2018. Multistage Downstream Attack Detection in a Cyber Physical System. In *Computer Security*. Springer International Publishing, Cham, 177–185.

[34] X. Ren, K. H. Johansson, and L. Shi. 2017. Quickest Change Detection With Observation Scheduling. *IEEE Trans. Automat. Control* 62, 6 (June 2017), 2635–2647. https://doi.org/10.1109/TAC.2016.2609998

[35] Yasser Shoukry, Michelle Chong, Masashi Wakaiki, Pierluigi Nuzzo, Alberto Sangiovanni-Vincentelli, Sanjit A Seshia, Joao P Hespanha, and Paulo Tabuada. 2018. SMT-based observer design for cyber-physical systems under sensor attacks. *ACM Transactions on Cyber-Physical Systems* 2, 1 (2018), 1–27.

[36] Jill Slay and Michael Miller. 2008. Lessons Learned from the Maroochy Water Breach. In *Critical Infrastructure Protection*, Eric Goetz and Sujeet Shenoi (Eds.). Springer US, Boston, MA, 73–82.

[37] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. 2012. Attack Models and Scenarios for Networked Control Systems. In *Proceedings of the 1st International Conference on High Confidence Networked Systems (HiCoNS '12)*. 55–64.

[38] C.S. van Dobben de Bruyn. 1968. *Cumulative sum tests : theory and practice.* London : Griffin.

[39] Xiukun Wei, Michel Verhaegen, and Tim van Engelen. 2010. Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques. *International Journal of Adaptive Control and Signal Processing* 24, 8 (2010), 687–707. https://doi.org/10.1002/acs.1162

[40] Erik Weyer and Georges Bastin. 2008. Leak detection in open water channels. *IFAC Proceedings Volumes* 41, 2 (2008), 7913 – 7918. 17th IFAC World Congress.

A STATE SPACE MATRICES FOR THE SYSTEM MODEL

In what follows, we present the state space matrices (A, B, C) and Kalman gain K, obtained using sub-space system identification. A 10^{th} order model is used. Therefore, we have system matrix A_2 as a 10 x 10. For a 15 inputs, matrix dimensions for B are 10 x 15. For 8 outputs, the dimensions for matrix C are, 8 x 10. Using these state space matrices and system model of (1), one can find the dynamics of the system evolution.

$$A = \begin{pmatrix} 0.995 & -1.17\times10^{-4} & 8.26\times10^{-5} & -2.54\times10^{-5} & -0.0015 & 8.7181\times10^{-5} & -0.000117 & - \\ -0.0000124 & 0.999 & -1.412\times10^{-4} & 6.892\times10^{-5} & -2.42\times10^{-5} & -2.72\times10^{-4} & 0.0015 & -5 \\ -0.0038 & -7.641\times10^{-4} & 0.997 & 0.0033 & 0.0078 & -0.0319 & 0.1475 & \\ 0.0000679 & 7.47\times10^{-4} & -2.13\times10^{-5} & 0.9936 & 1.9822\times10^{-4} & 0.0073 & -0.0410 & \\ 0.0209 & -0.0018 & -0.0324 & 0.028 & 0.8825 & 0.0160 & -0.0359 & \\ 0.0026 & -0.0063 & 0.0178 & -0.0260 & 0.0158 & 0.9237 & 0.0210 & \\ 0.0053 & -0.0045 & -0.0249 & 0.0491 & 0.0052 & 0.0077 & 0.888 & \\ 0.0064 & -0.0035 & -0.0169 & 0.0489 & 0.0030 & 0.0148 & -0.0762 & \\ -0.0223 & -0.0152 & -0.1805 & 0.1656 & 0.2589 & -0.0095 & -0.1766 & \\ -0.000765 & 0.0024 & 0.0303 & 0.0179 & -0.019 & -0.0326 & 0.175 & \end{pmatrix}$$

$$B = \begin{pmatrix}
-4.18\times10^{-5} & -3.46\times10^{-5} & -5.68\times10^{-6} & 0.000126 & 0.000126 & 2.44\times10^{-5} & -0.00014 & -5.47110\times10^{-5} & 8.12\times10^{-6} & 4.18\times10^{-7} & -3.16\times10^{-7} & -7.96\times10^{-7} & -2.2\times-7 & -3.52\times10^{-7} & -4.53\times10^{-7} \\
-9.268\times10^{-6} & -1.378\times10^{-5} & 5.084\times10^{-6} & 9.178\times10^{-5} & 9.178\times10^{-5} & 5.928\times10^{-6} & -0.00013 & -2.448\times10^{-5} & 3.108\times10^{-6} & 2.918\times10^{-8} & -9.748\times10^{-8} & -2.408\times10^{-7} & -1.2448\times10^{-7} & -1.468\times10^{-7} & -1.538\times10^{-7} \\
-0.000364 & -0.00059 & 0.00018 & 0.0085 & 0.0085 & 0.00047 & -0.011 & -0.00183 & 0.00018 & -2.108\times10^{-6} & -7.7091\times10^{-6} & -1.411\times10^{-5} & -8.212\times10^{-6} & -8.27\times10^{-6} & -1.037\times10^{-5} \\
-0.0004 & -0.00017 & -4.82\times10^{-6} & -0.0014 & -0.0014 & 1.46\times10^{-5} & 0.0019 & 7.38\times10^{-5} & 2.89\times10^{-5} & 7.02\times10^{-6} & 2.16\times10^{-7} & -2.65\times10^{-6} & 1.8\times10^{-6} & 1.97\times10^{-7} & -1.24\times10^{-7} \\
-0.0009 & -0.001 & -0.0015 & 0.0039 & 0.0039 & 0.00032 & -0.0046 & -0.0015 & 0.000287 & 1.296\times10^{-5} & -1.708\times10^{-6} & -3.70\times10^{-5} & -1.64\times10^{-5} & -1.99\times10^{-5} & -2.52\times10^{-5} \\
-0.0017 & -0.0025 & 1.58\times10^{-5} & 0.0016 & 0.0016 & 0.00020 & 0.00050 & 0.00014 & -4.42\times10^{-5} & -9.73\times10^{-6} & 5.53\times10^{-6} & 1.139\times10^{-5} & 3.16\times10^{-6} & 3.56\times10^{-6} & 6.069\times10^{-6} \\
-0.0013 & -0.00086 & -1.84\times10^{-6} & -0.0035 & -0.0035 & -7.24\times10^{-5} & 0.0059 & -0.00011 & 5.96\times10^{-5} & 9.49\times10^{-6} & -1.05\times10^{-6} & -5.30\times10^{-6} & 1.63\times10^{-7} & -9.30\times10^{-7} & -1.24\times10^{-6} \\
-0.0020 & -0.0010 & -0.00015 & 0.00036 & 0.00036 & 3.085\times10^{-5} & 0.0011 & -0.00087 & 0.00025 & 1.835\times10^{-5} & -5.015\times10^{-6} & -1.815\times10^{-5} & -1.185\times10^{-7} & -5.325\times10^{-5} & -7.245\times10^{-6} \\
-0.0051 & -0.0037 & 0.011 & 0.010 & 0.010 & 0.0027 & -0.022 & -0.0067 & 0.0013 & 7.93\times10^{-5} & -7.55\times10^{-6} & -0.00016 & -6.25\times10^{-5} & -8.038\times10^{-6} & -0.00010 \\
-0.0053 & -0.0035 & -0.0019 & 0.013 & 0.013 & 0.00032 & -0.016 & -0.0056 & 0.00075 & 4.69\times10^{-5} & -7.23\times10^{-6} & -4.084\times10^{-5} & 6.88\times10^{-6} & -9.60\times10^{-6} & -1.290\times10^{-5}
\end{pmatrix}$$

$$C = \begin{pmatrix}
-475.04 & -3976.80 & 46.14 & 21.01 & 1.82 & 1.95 & 0.34 & 2.11 & -0.12 & -0.26 \\
1519.69 & -102.13 & -17.89 & 4.54 & 3.000 & 0.344 & -1.33 & 0.14 & 0.65 & 1.80 \\
348.066 & -308.87 & 609.891 & -3742.44 & -19.061 & -426.50 & 2586.86 & -646.60 & -29.84 & -251.44 \\
372.86 & -569.67 & 225.67 & -4452.78 & -44.44 & -168.80 & 1002.75 & 319.95 & -22.82 & -192.38 \\
-1.407 & -1.68 & -0.703 & 0.037 & -7.706 & -42.83 & -6.95 & -0.042 & 0.62 & -0.012 \\
10.648 & 0.57 & -0.99 & 2.01 & -84.934 & 14.11 & 3.13 & 0.13 & -8.45 & 0.12 \\
-1.86 & -3.66 & -31.197 & -8.68 & -0.0443 & -0.377 & 2.38 & -4.58 & -0.032 & -0.59 \\
2.30 & 12.54 & 0.23 & 102.90 & 1.62 & -0.59 & 1.68 & 59.57 & 0.388 & 6.28
\end{pmatrix}$$

$$K = \begin{pmatrix}
-1.02\times10^{-5} & 0.000319 & -2.03\times10^{-6} & 8.47\times10^{-6} & -0.000103 & -2.34\times10^{-5} & -6.43\times10^{-5} & 2.26 \\
-8.66\times10^{-5} & -3.36\times10^{-5} & -4.67\times10^{-6} & 7.71\times10^{-6} & -8.57\times10^{-5} & -5.35\times10^{-5} & -2.76\times10^{-5} & 2.34 \\
-7.05\times10^{-6} & 4.16\times10^{-5} & -0.000435 & 0.000695 & -0.00514 & -0.00133 & -0.00436 & 0.0 \\
1.29\times10^{-5} & -8.43\times10^{-5} & 1.71\times10^{-5} & -0.000105 & 0.00173 & 0.000154 & 0.00118 & 0.0 \\
-0.000219 & -0.0001582 & 3.77\times10^{-5} & 0.000126 & -0.00248 & -0.00427 & 0.00256 & 0.0 \\
0.0001308 & 4.38\times10^{-5} & -5.63\times10^{-5} & -5.99\times10^{-5} & -0.0121 & 0.000444 & -0.00192 & -0.0 \\
4.17\times10^{-5} & -9.79\times10^{-5} & 0.000341 & -0.000183 & -1.97\times10^{-5} & 0.000251 & 0.00246 & 0.00 \\
6.64\times10^{-5} & -0.000446 & -0.000273 & 0.000525 & 0.00269 & 8.80\times10^{-5} & 0.00376 & 0.0 \\
-0.000581 & 0.00117 & 0.000173 & 0.000429 & -0.00496 & -0.00931 & -0.00938 & -0.0 \\
0.000131 & 0.000527 & -0.000214 & 2.90\times10^{-5} & -0.0129 & -0.0010 & -0.0025 & -0.
\end{pmatrix}$$