Chapter 8

ANOMALY DETECTION IN LIQUID PIPELINES USING MODELING, CO-SIMULATION AND DYNAMICAL ESTIMATION

Saed Alajlouni and Vittal Rao

Abstract Historically, supervisory control and data acquisition (SCADA) systems have relied on obscurity to safeguard against attacks. Indeed, external attackers lacked knowledge about proprietary system designs and software to access systems and execute attacks. The trend to interconnect to the Internet and incorporate standardized protocols, however, has resulted in an increase in the attack surface – attackers can now target SCADA systems and proceed to impact the physical systems they control. Dynamical estimation can be used to identify anomalies and attempts to maliciously affect controlled physical systems. This paper describes an intrusion detection method based on the dynamical estimation of systems. A generic water pipeline system is modeled using state space equations, and a discrete-time Kalman filter is used to estimate operational characteristics for anomaly-based intrusion detection. The effectiveness of the method is evaluated against deception attacks that target the water pipeline system. A co-simulation that integrates computational fluid dynamics software and MATLAB/Simulink is employed to simulate attacks and develop detection schemes.

Keywords: Liquid pipelines, anomaly detection, dynamical estimation

1. Introduction

SCADA systems are used to monitor and control processes in industrial environments. Typical implementations involve highly distributed operations over large geographical areas, such as the electric grid and pipeline systems. SCADA systems consist of a central control center that issues set-point values and receives measurements and alarm data from distributed controllers. In recent years, utility companies have begun to transition to the Internet and

J. Butts and S. Shenoi (Eds.): Critical Infrastructure Protection VII, IFIP AICT 417, pp. 111–124, 2013.

standard communications protocols for information exchange and remote control. SCADA systems now face a higher risk of cyber attacks, primarily because their vulnerabilities are more readily exposed [1].

Common attacks that target process control systems are denial-of-service (DoS), deception and stealth attacks. In an example DoS attack, the attacker attempts to prevent the controller from reading sensor data or prevent the actuator from receiving control commands [2]. In a deception attack, the adversary may send false control commands to actuators or inject malicious sensor readings, resulting in control actions based on the false data [8]. A stealth attack is a sequence of deception attacks, where the attacker attempts to cause damage or affect operations without being detected [1, 3]. In a stealth attack scenario, the attacker is assumed to have knowledge of the plant dynamics.

SCADA systems can be abstracted into two interrelated layers: a communications layer and a physical layer. In the case of a pipeline system, the physical layer consists of pipe sections, actuators (e.g., valves and pumps) and measurement sensors (e.g., for flow rate and pressure). The communications layer manages the flow rates and data exchange throughout the physical layer; it incorporates modems, routers, switches and the communications medium. This paper focuses on the physical layer and how dynamical estimation can be utilized to detect anomalies.

The intrusion detection approach is divided into two main steps: (i) modeling the dynamical system; and (ii) applying anomaly-based intrusion detection methods. The first step involves the derivation of a mathematical model that describes the dynamics of the physical system of interest. In the case of a SCADA system, the model can be represented using a non-linear, time-varying differential equation. In the second step, Kalman-filter-based dynamical estimation is used to predict sensor measurement values. If the estimated quantities exceed predetermined thresholds, then an error is detected, which can be attributed either to a system fault or a malicious attack.

The approach is similar to those used to address the well-known problem of "bad data" detection in power grids. However, this paper focuses on how dynamical estimation can be utilized to detect anomalies, whereas approaches for bad data detection in power grids rely on static estimation. The primary difference is that dynamical estimation relies on differential equations to estimate current values and to predict future values. Static estimation, on the other hand, uses measurements taken at various time instances to determine estimates of the current system state without predicting future states.

Compared with the electric power grid, pipeline systems typically have slower dynamics. This provides the proposed detection scheme with more time to detect and respond to malicious attacks without being concerned about system damage during the detection phase. Nevertheless, it is important to note that a critical requirement of the proposed scheme is adequate computational power to implement estimation and anomaly detection in real time. As such, the order of the model and the selected anomaly detection algorithm are both contributing factors to the feasibility of performing real-time anomaly detection. This paper

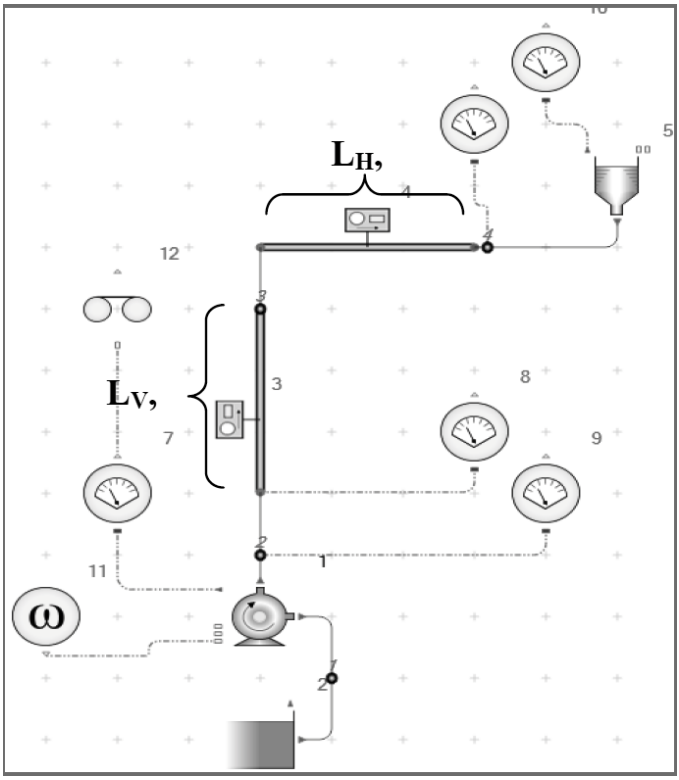

Figure 1. Generic water pipeline system.

makes the case that the availability of dynamical models, necessary computing power, and the relatively slower dynamics of pipeline systems are factors that facilitate dynamical estimation for real-time anomaly detection.

2. Modeling and Simulation

Figure 1 shows a schematic diagram of a generic water pipeline system. It consists of a water transportation system that distributes water from a well to an overhead tank. The pipeline consists of two sections: a vertical section with length (L_v) and diameter (d_v), and a horizontal section with length (L_H) and diameter (d_H). The non-linear dynamical response of the system is simulated using the Flowmaster software package [6]. This computational fluid dynamics software is used to discern the pressure surge, pressure drop, flow rate, temperature and system response times. Flowmaster transient analysis is similar to the simulation approach introduced by Miller [9].

A pipeline network may be represented as a number of modules connected at nodes. The modules consist of pipe sections and other components such as valves, tanks, accumulators and pumps. Pipeline network equations are gener-

ated after specifying the necessary parameters for each module such as friction parameters, fluid properties, elevation and initial conditions. The equations are formed by using the fact that the nodal pressure is the same for all modules connected to the same node, and the sum of the flow rates is zero at each node. The equations are solved simultaneously to yield values of nodal pressures and flow rates.

In the Flowmaster simulation, pipeline system components can be categorized as mathematically simple or complex. Equations for simple components are dependent only on the geometry of the component, while complex components involve dynamic elements [9]. Simple components include inlets from reservoirs, changes in the cross-sectional area, orifice plates, rigid pipes, dead ends and valves. Complex components include pumps, turbines, surge tanks and surge vessels. Note that there must be at least one complex component in a system in order to initiate a transient. Numerical techniques are used to solve the differential equations in order to determine the head and flow in a complex component. In the case of a pump, for example, the equations approximate flow, speed and torque.

2.1 Pipeline System Modeling

Mathematical models are required for the development of intrusion detection methods. In general, the mathematical model of a process control system is non-linear; however, the implementation of Kalman filter estimation requires a linear plant model. As such, it is desirable to discretize the mathematical relation using the first two terms of a Taylor series expansion [10] so that the mathematical model is linearized. This step is necessary because a digital controller is used to control the plant.

2.2 Linear Model

For the pipeline system shown in Figure 1, a second-order, linear time-invariant state space model was derived after selecting the state vector $x = [P_t, Q]^T$, where P_t is the pressure at the bottom of the overhead tank, Q is the volumetric flow rate and T is the time. The state space model has the form $\dot{x}(t) = Ax(t) + Bu(t)$ and can be expressed as:

$$\begin{bmatrix} \dot{P}_t \\ \dot{Q} \end{bmatrix} = \begin{bmatrix} 0 & \frac{\rho g}{A_t} \\ \frac{-1}{I_{eq}} & \frac{-R_{eq}}{I_{eq}} \end{bmatrix} \begin{bmatrix} P_t \\ Q \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{I_{eq}} \end{bmatrix} P_{pump} \tag{1}$$

where I_{eq} is the fluid total inertance in the system, A_t is the cross-sectional area of the tank, R_{eq} is the total fluid resistance in the system, and P_{pump} is the pressure at the pump outlet and the input to the system.

The total inertance I_{eq} is analogous to an inductance in an electronic circuit [4]. It is expressed as:

$$I_{eq} = \frac{4\rho}{\pi}\left(\frac{L_V}{d_V^2} + \frac{L_H}{d_H^2}\right).$$

Table 1. Relationships between friction models and Reynolds Number [6].

Option	**Laminar Flow Re $\leq$ 2000**	**Transition Zone 2000 $<$ Re $<$ 4000**	**Turbulent Flow Re $\geq$ 4000**
Colebrook-White	$f_l = \frac{64}{Re}$	$f = xf_t + (1-x)f_l$	$f_t = \frac{0.25}{[\log(\frac{k}{3.7D} + \frac{5.74}{Re^{0.9}})]^2}$
Hazen-Williams	$f_l = \frac{64}{Re}$	$f = xf_t + (1-x)f_l$	$f_t = \frac{1014.2Re^{-0.148}}{C_{HW}^{1.852}D^{0.0184}}$
Fixed	f_l	f	f_t

The fluid resistance R_{eq} is analogous to a resistor in an electronic circuit [4]. It is expressed as:

$$R_{eq} = \frac{128\mu}{\pi}\left(\frac{L_V}{d_V^4} + \frac{L_H}{d_H^4}\right).$$

R_{eq} is the result of applying the Darcy-Weisbach equation to calculate the pressure drop due to friction in a pipeline. The relationship between the pressure drop due to friction P_f and the flow Q is linear when the flow is laminar. It is expressed as:

$$P_f = R_{eq} \times Q.$$

The following assumptions and operational limitations render the derived pipeline equations simple and linear:

- The fluid used is water, which is incompressible. Water is a Newtonian fluid that has a constant dynamic viscosity μ.

- The flow is laminar (i.e., the speed of flow is relatively slow). Specifically, the Reynolds number Re, a dimensionless number that indicates the type of flow in a conduit, satisfies the constraint $Re \leq 2000$. When $2000 < Re < 4000$, then the flow is transitional between laminar and turbulent flow. When $Re \geq 4000$, then the flow is turbulent. As Re changes, the frictional model of the conduit, which accounts for pressure drops due to frictional losses, changes accordingly.

  Table 1, taken from the Flowmaster V7 manual [6], shows the relation between Re and friction models. Note that, when $Re > 2000$, the relationship between pressure drop due to friction and flow rate ceases to be linear.

- A Newtonian fluid with a laminar flow has a parabolic velocity profile. However, for a relatively long pipe section, the flow profile can be approximated as a uniform velocity profile.

- Dynamic changes due to heat transfer effects are negligible.

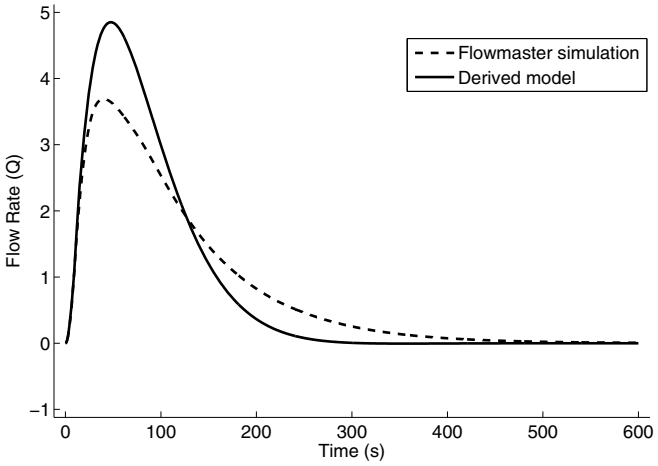

Figure 2. Normalized flow rate.

- Adiabatic changes in water density (i.e., elastic effects of water) may be ignored because heat transfer is negligible.
- Elastic effects of the pipes may be ignored because they are made of rigid steel and have high resistance to deformation.
- Pump speed increases gradually with no abrupt pressure changes. This guarantees laminar flow and preserves the linearity of the mathematical model.

2.3 Discretization

A suitable sampling period must be chosen in order to accurately capture the system dynamics after discretization. The sampling period T_s is selected such that $T_s = \frac{0.1}{|\lambda_{max}|}$, where $|\lambda_{max}|$ is the absolute maximum of the real part of the system eigenvalues.

The discretized version of Equation (1) can be expressed using zero-order hold for the input signal P_{pump}:

$$x_{k+1} = Fx_k + Gu_k \tag{2}$$

where $F = e^{AT_s}$ is the state transition matrix calculated over the sampling period T_s and $G = A^{-1}\left(e^{AT_s} - I\right)B$, where I is the identity matrix.

Using the Flowmaster software simulation as a representation of the actual non-linear dynamics of the pipeline system, the accuracy of the discretized linear model shown in Equation (2) can be evaluated. Figures 2 and 3 compare the pipeline states against the Flowmaster simulation results for the flow rate and pressure at the bottom of the overhead tank, respectively.

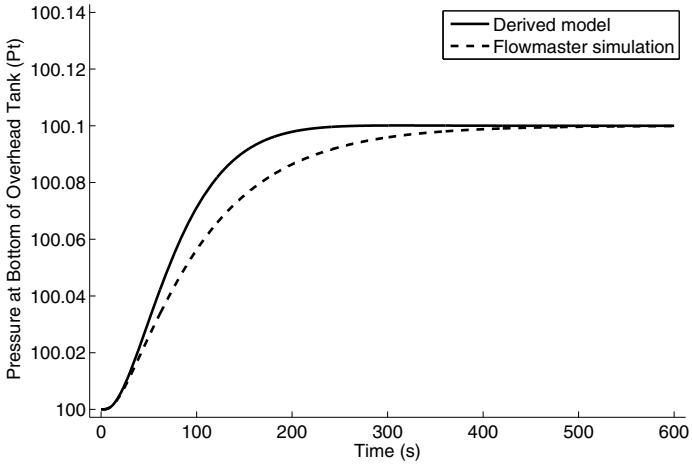

Figure 3. Normalized pressure.

Upon inspecting Figures 2 and 3, it can be seen that the discretized mathematical model in Equation (2) is sufficiently accurate for representing the dynamic behavior of the pipeline. For estimation purposes, however, the response differences between Equation (2) and the Flowmaster simulation are modeled as process disturbances (discussed in Section 4).

3. Co-Simulation

The simulation capabilities of MATLAB/Simulink and Flowmaster software can be integrated to produce a co-simulation. The purpose of a co-simulation is to perform simultaneous simulations and data exchange between a dynamical system model in MATLAB/Simulink and a fluid or electromechanical model in Flowmaster. Note that variables for the controllers and gauges can be transferred into and out of the Flowmaster model [5].

In the context of the pipeline system, co-simulation facilitates a Flowmaster simulation of the non-linear behavior of the pipeline system and the ability to transfer sensor measurements to MATLAB/Simulink to perform state estimation. In addition, the MATLAB/Simulink environment can be used to design a digital controller that sends control signals to the pipeline process in Flowmaster. Malicious attacks on the pipeline system can be simulated using a real-time, co-simulation process between MATLAB/Simulink and Flowmaster. For example, during a co-simulation, MATLAB/Simulink can be programmed to inject malicious data into the true sensor measurements that originate from the Flowmaster simulation. The flexibility enables the simulation of attack scenarios involving DoS, deception and stealth attacks that target sensor measurements.

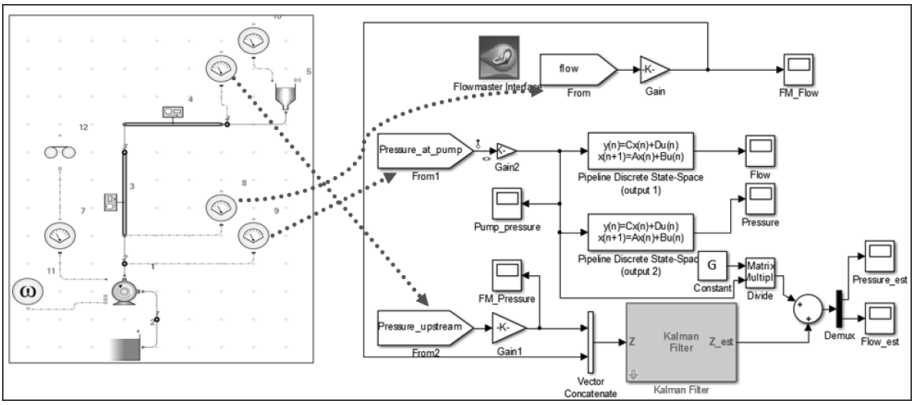

Figure 4. Variable exchange during the co-simulation process.

Figure 4 illustrates the co-simulation process. The figure shows how measurements from the Flowmaster software can be transferred to MATLAB/Simulink to conduct dynamical estimation and to simulate malicious attacks.

4. Anomaly Detection

Statistical methods have been used very effectively to detect anomalies in dynamical systems. For example, Cardenas, *et al.* [1] have used hypothesis testing in an evaluation of the Tennessee-Eastman plant. Their results demonstrate the ability to model changes in behavior of the underlying physical system. Interested readers are referred to [7] for an overview of hypothesis testing schemes that can be used in anomaly detection.

The discrete Kalman filter is used to estimate pipeline states at each sampling time. When one or more sensor measurements are compromised by the injection of malicious data, Kalman filtering and prediction results can be used to detect anomalies. Note that successful detection requires the careful selection of alarm threshold levels, which will be examined in future work.

As stated previously, the response differences between the discretized linear mathematical model and the Flowmaster simulation can be modeled as process disturbances. To incorporate a process disturbance, the pipeline state space model specified by Equation (2) is updated as follows:

$$x_{k+1} = Fx_k + Gu_k + w_k$$

$$z_{k+1} = Hx_k + v_k$$

where z_{k+1} is the sensor measurement equation, H is the measurement matrix that relates system states to sensor measurements through a linear relationship, w_k is the process disturbance, and v_k is the sensor measurement noise. Note that both w_k and v_k are assumed to be Gaussian white processes with zero means.

Table 2. Co-simulation cases for pipeline system under attack.

Simulation Case	Flow Rate Integrity	Pressure Integrity
Case I	Not Compromised	Compromised
Case II	Compromised	Not Compromised
Case III	Compromised	Compromised

Since the process disturbance covariance $E[w_k, w_k^T] = Q$ and the measurement noise covariance $E[v_k, v_k^T] = R$ denote the statistical expectation, the Kalman filter equations may be written as [11]:

$$P_{k+1|k} = FP_{k|k}F^T + Q$$

$$K_{k+1} = P_{k+1|k}H^T\left[HP_{k+1|k}H^T + R\right]^{-1}$$

$$\hat{x}_{k+1|k} = F\hat{x}_{k|k} + Gu_k$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}[z_{k+1} - H\hat{x}_{k+1|k}]$$

$$P_{k+1|k+1} = [I - K_{k+1}H]P_{k+1|k}$$

where $\hat{x}_{k+1|k+1}$ is the estimated state vector, $\hat{x}_{k+1|k}$ is the predicted state vector, $P_{k+1|k+1}$ is the covariance of the estimation error, and $\hat{z}_k$ denotes the estimated sensor measurements with $\hat{z}_k = H\hat{x}_{k|k}$. The filter initialization is expressed as:

$$\hat{x}_{0|0} = E[x_0]$$

$$P_{0|0} = E\left[\left(x_0 - \hat{x}_{0|0}\right)\left(x_0 - \hat{x}_{0|0}\right)^T\right].$$

4.1 Attack Scenario

For demonstration purposes, we consider a deception attack scenario where the attacker is able to compromise at least one sensor measurement in the pipeline system. In addition, we assume that the attacker has knowledge of the maximum and minimum sensor readings that are pre-defined by the pipeline SCADA system and the alarms that are triggered when the sensor readings are outside the permissible ranges.

The two sensor measurements that are compared against the estimation results are the flow rate and the pressure at the bottom of the storage tank. Under normal operating conditions, the dynamical estimation of each sensor reading based on one or both sensor readings yields results that are comparable with the actual sensor readings. However, there are clear differences between the dynamical estimation results and sensor readings in the case of integrity attacks. Table 2 summarizes the conditions underlying the simulated attack scenarios.

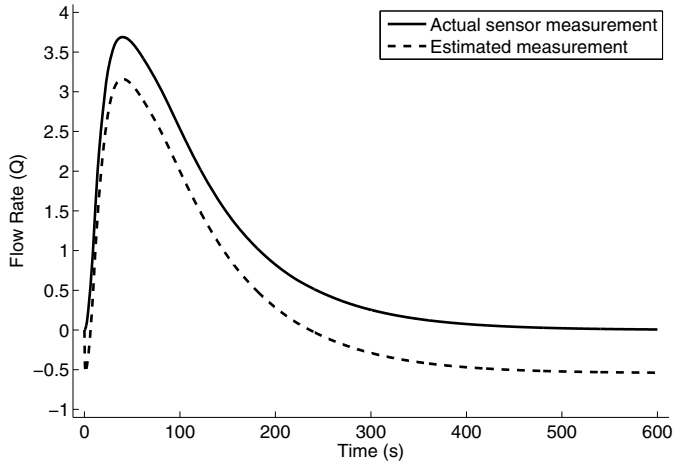

Figure 5. Case I: Flow rate observations.

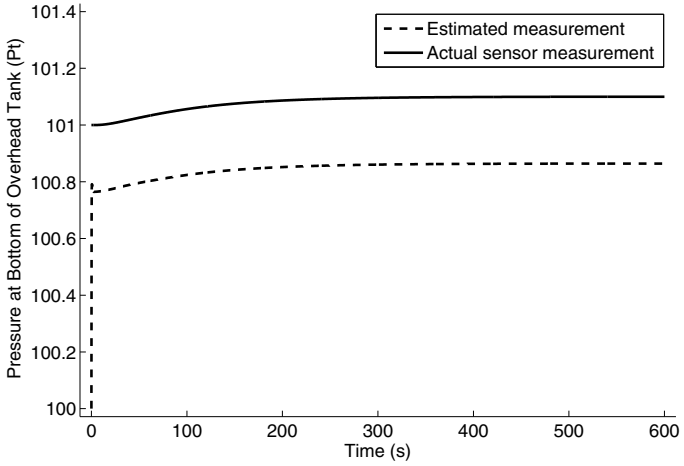

Figure 6. Case I: Pressure observations.

4.2 Anomaly Detection Results

Figures 5 through 10 show the results obtained for the flow rate and pressure measurements using the estimation algorithm. Figure 5 shows the differences in the flow rate for Case I. Figure 6 shows the differences observed in the same attack scenario for the pressure readings at the bottom of the overhead tank.

In Case II, launching a deception attack on the flow measurement reading produces no significant difference between the estimated and actual flow rates (Figure 7). The attack, however, produces a significant difference between

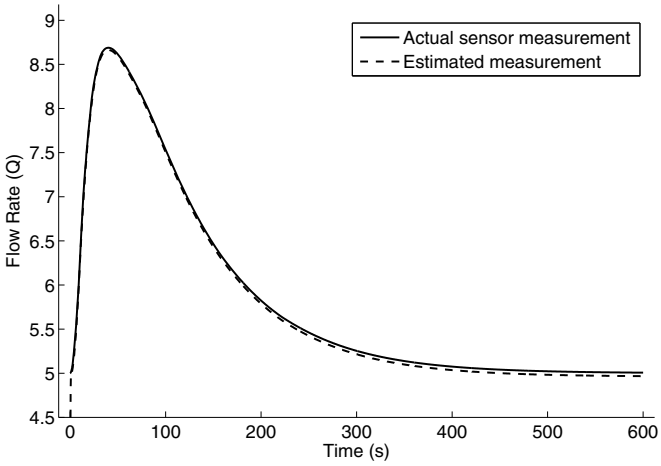

Figure 7. Case II: Flow rate observations.

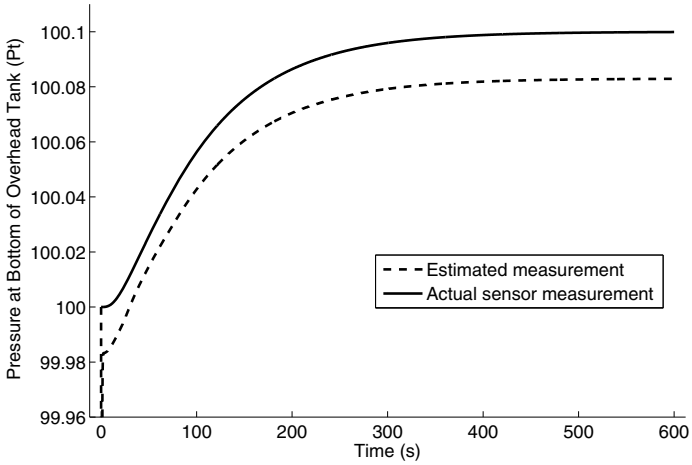

Figure 8. Case II: Pressure observations.

the estimated pressure and the actual pressure readings as shown in Figure 8. The results are due to the characteristics of the particular pipeline system and emphasize the importance of multiple measurement points.

In Case III, the integrity of both measurements are compromised. The results, shown in Figures 9 and 10, reveal significant differences between the estimated and actual sensor measurements.

The results demonstrate notable differences between the estimates under normal operating conditions and those when sensor readings are compromised. Note that the modeled pipeline system is a non-linear dynamical system sim-

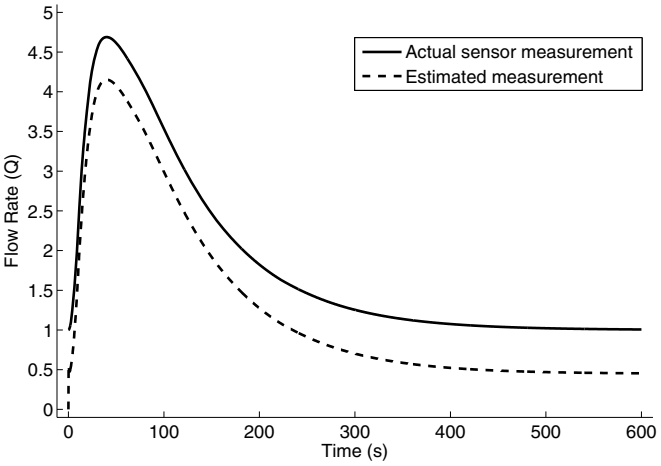

Figure 9. Case III: Flow rate observations.

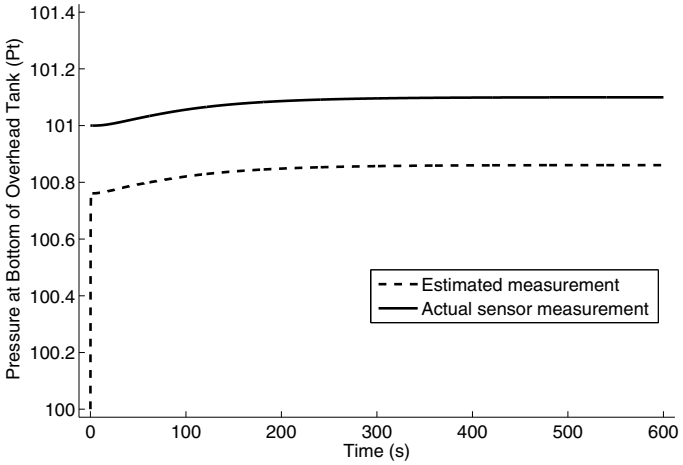

Figure 10. Case III: Pressure observations.

ulated in Flowmaster. If the nominal operating trajectory of the non-linear system changes such that the flow becomes turbulent or transitional between laminar and turbulent, then the derived linear model in Equation (1) becomes a less accurate representation of the actual pipeline dynamics. However, small changes in the nominal operating point can be modeled as process disturbances in the Kalman filter equations. In other words, as the nominal operating point of the non-linear system changes slightly, the process disturbance covariance matrix Q can be increased to account for modeling differences between the linear equations and the non-linear operating point. Increasing Q involves the

trade-off of making the Kalman estimator less sensitive to detecting anomalies, specifically when the attacker tries to shift the behavior of the system over a relatively long period of time. Indeed, to maintain sufficient accuracy, the linear model used in the Kalman estimation should be updated to reflect changes in the point of operation of the non-linear system. Continuous updates can be readily implemented using an extended Kalman filter [11].

5. Conclusions

The anomaly-based intrusion detection method for process control systems presented in this paper uses dynamical state estimation techniques. The water pipeline example demonstrates the utility of the method and the application of co-simulation techniques. The integration of Flowmaster and MATLAB/Simulink software is a promising approach for simulating attacks when developing and refining anomaly detection techniques. The experimental results reveal significant differences between normal operating conditions and scenarios involving the injection of sensor data. The experimental results also demonstrate the feasibility of developing practical intrusion detection algorithms based on dynamical state estimation.

Acknowledgement

This research was supported by the National Science Foundation under MRI Grant No. ECCS-1040161.

References

[1] A. Cardenas, S. Amin, Z. Lin, Y. Huang, C. Huang and S. Sastry, Attacks against process control systems: Risk assessment, detection and response, *Proceedings of the Sixth ACM Symposium on Information, Computer and Communications Security*, pp. 355–366, 2011.

[2] A. Cardenas, S. Amin and S. Sastry, Research challenges for the security of control systems, *Proceedings of the Third USENIX Conference on Hot Topics in Security*, article no. 6, 2008.

[3] G. Dan and H. Sandberg, Stealth attacks and protection schemes for state estimators in power systems, *Proceedings of the First IEEE Conference on Smart Grid Communications*, pp. 214–219, 2010.

[4] C. De Silva, *Mechatronics: An Integrated Approach*, CRC Press, Boca Raton, Florida, 2005.

[5] Flowmaster, FlowmasterLink for MATLAB V2.0.1, Schaumburg, Illinois (`www.flowmaster.com/flowmaster_flowmasterlink_matlab.html`).

[6] Flowmaster, Flowmaster V7 Overview, Schaumburg, Illinois (`www.flowmaster.com/flowmaster_overview.html`).

[7] T. Kailath and H. Poor, Detection of stochastic processes, *IEEE Transactions on Information Theory*, vol. 44(6), pp. 2230–2259, 1998.

[8] Y. Liu, P. Ning and M. Reiter, False data injection attacks against state estimation in electric power grids, *Proceedings of the Sixteenth ACM Conference on Computer and Communications Security*, pp. 21–32, 2009.

[9] D. Miller, *Internal Flow Systems*, British Hydromechanics Research Association, Cranfield, England, 1990.

[10] W. Rugh, *Linear System Theory*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[11] D. Simon, *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*, John Wiley, Hoboken, New Jersey, 2006.