Cyber Security of Water SCADA Systems—Part I: Analysis and Experimentation of Stealthy Deception Attacks

Saurabh Amin, Xavier Litrico, Shankar Sastry, and Alexandre M. Bayen

***Abstract*— This brief aims to perform security threat assessment of networked control systems with regulatory and supervisory control layers. We analyze the performance of a proportional-integral controller (regulatory layer) and a model-based diagnostic scheme (supervisory layer) under a class of deception attacks. We adopt a conservative approach by assuming that the attacker has knowledge of: 1) the system dynamics; 2) the parameters of the diagnostic scheme; and 3) the sensor-control signals. The deception attack presented here can enable remote water pilfering from automated canal systems. We also report a field-operational test attack on the Gignac canal system located in Southern France.**

***Index Terms*— Delay systems, fault diagnosis, intrusion detection, supervisory control and data acquisition (SCADA) systems, supervisory control.**

I. Introduction

NETWORKED control systems (NCSs) are increasingly being deployed to facilitate the monitoring and control of large-scale physical infrastructures. In recent years, the traditionally hard-wired electromechanical devices in the so-called supervisory control and data acquisition (SCADA) systems are being replaced by Internet-connected embedded devices with computational capabilities. This information technology (IT) enabled modernization of NCS permits, achieving higher reliability and lower costs. Yet, recent incidents suggest significant issues with NCS security. In this brief, our aim is to perform security threat assessment of NCS/SCADA systems with regulatory and supervisory control layers. We focus on deception attacks to NCS/SCADA systems for water canal systems. We evaluate a regulatory-level proportional-integral (PI) controller [1] and a supervisory-level diagnostic scheme [2] against deception attacks to sensor measurements.

Manuscript received April 7, 2011; revised February 16, 2012; accepted July 1, 2012. Manuscript received in final form July 31, 2012. Date of publication September 14, 2012; date of current version August 12, 2013. This work was supported in part by Cemagref, the France-Berkeley Fund, the Team for Research in Ubiquitous Secure Technology, a National Science Foundation Science and Technology Center, and an MIT faculty startup grant. Recommended by Associate Editor B. Jiang.

S. Amin is with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: amins@mit.edu).

X. Litrico is with the Research and Development Center of Lyonnaise des Eaux, Bordeaux 33300, France (e-mail: xavier.litrico@lyonnaise-des-eaux.fr).

S. Sastry is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: sastry@coe.berkeley.edu).

A. M. Bayen is with the Department of Electrical Engineering and Computer Sciences and the Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720 USA (e-mail: bayen@berkeley.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCST.2012.2211873

Network-induced vulnerabilities arise in NCSs because of four factors. First, due to wider deployment of off-the-shelf IT devices, NCSs inherit the vulnerabilities of these devices, and thus are subject to correlated software and hardware failures. Second, the proprietary network protocols of traditional SCADA systems are being upgraded to open protocols, making it easier for attackers to learn about NCS operations. Third, sensor-control data is being made accessible to authorized remote users via corporate networks and the Internet. This makes NCSs subject to insider attacks. Fourth, the existence of organized cybercrime groups enhances attackers' capabilities to conduct intrusions into NCSs. Indeed, many nation states view cyber warfare as the future of armed conflict.

Security of NCSs is fundamentally different from IT security since NCSs are required to interact with the underlying physical infrastructure in real time. While recent efforts have focused on applying IT security solutions to NCSs, these solutions do not directly address the risks posed by an intelligent attacker who is capable of compromising sensor-control data [3], [4]. Moreover, existing research in automatic control assumes that the system operators are trustworthy and the sensor-control data has not been compromised. As a result, despite significant advances in the design of fault-tolerant NCSs, there is at best little understanding of their resilience to cyber incidents. Unsurprisingly, cyber security issues were not considered in the design of legacy SCADA systems.

Cyber incidents to NCS can be broadly classified as computer-based accidents, nontargeted attacks, and targeted attacks. Computer-based accidents are caused as a result of unintended IT failures. Nontargeted attacks to NCSs are similar to the incidents that any network-connected computer may suffer. Targeted attacks to NCS are the most serious class of attacks because the attackers tailor their strategies toward damaging NCS components [5], [6]. Most notably, the Stuxnet worm [7] has demonstrated serious threats to NCSs. Stuxnet has the ability to infect and reprogram the programmable logic controllers (PLCs) and hide the changes using a PLC rootkit.

This brief is organized as follows. In Section II, we discuss cyber attacks to hierarchically structured NCS/SCADA systems. We henceforth do not distinguish between NCS and SCADA systems and use these terms interchangeably. In Section III, we present a regulatory-level PI control method and the supervisory-level diagnostic scheme. The diagnostic scheme chosen here is a fault detection and isolation method based on unknown input observers (UIOs). The threat assessment of other diagnostic schemes (e.g., [8]) can be done in a similar manner [9]. In Section IV, we report results from a field-operational test attack on the Gignac water SCADA

system located in Southern France.[1] We model the attacker as an intelligent insider who is resourceful enough to obtain access to sensor-control data and who knows the diagnostic scheme. We analyze the performance of the SCADA system when the attacker obtains unauthorized access to sensor measurements and compromises them. We demonstrate that such an attack can lead to water pilfering from the canal system. Thus, we show that the existing diagnostic tools for addressing random faults may not be sufficient to diagnose cyber attacks. Finally, we summarize salient points of our analysis in Section V. In the companion brief [10], we use the insights gained in this paper to develop a robust attack diagnostic scheme based on enhanced hydrodynamic models of canal pools.

II. Cyber Attacks Against Water SCADA Systems

Cyber attacks targeting NCS/SCADA systems can be further classified as deception attacks and denial-of-service (DoS) attacks, which, respectively, result in the loss of integrity and availability of sensor-control data. Integrity for SCADA systems can be defined as their ability to maintain operational goals by preventing, detecting, or surviving deception attacks. Deception attacks can include an incorrect sensor measurement or control input, an incorrect time stamp, or a wrong identity of the sending device. An adversary can launch these attacks by obtaining the secret keys used by the sending devices, or by compromising some of the sensors and actuators. Availability for SCADA systems can be defined as the ability to maintain operational goals by preventing or surviving DoS attacks to sensor measurement and control inputs. To launch a DoS attack, the adversary can jam the communication channels, prevent field devices from sending data, or flood the communication network with random data.

A. Attack Models

Modern water SCADA systems have a hierarchical structure with regulatory control and supervisory control layers; see Fig. 1. For the ith pool, we denote the discharge (m^3/s) at the upstream end (resp. downstream end) by q_{i-1} (resp. q_i), the water-level (m) at the downstream end by y_i^d, and the offtake water withdrawal (m^3/s) at the downstream end by p_i. We will assume that q_{i-1} and q_i are the control inputs, y_i^d is the measurement, and p_i is the disturbance variable. These variables are the respective deviations around a steady state flow. Fig. 1 also illustrates possible attacks to water SCADA systems. The attack **A0** denotes attacks against the physical infrastructure or the field devices (sensors and actuators). Since such attacks require physical access, a risk-averse attacker is more likely to launch cyber attacks (**A1**–**A6**). We do not consider physical attacks in the rest of this brief.

Attacks **A1** and **A2** denote the attacks to the regulatory control layer, which interacts with the physical canal network through field devices. Attack **A1** denotes the DoS attacks on the field-area network between the PLCs and the field devices, or the deception attack on the sensor measurements y_i^d and control actuations $\mathbf{u}_i$. Attack **A2** denotes similar DoS or deception attacks on inter-PLC communication. By **A3** we mean cyber attacks to the control network which enables communication between the regulatory and the supervisory control layers. The control network transmits: 1) water level measurements y_i^d, gate openings $\mathbf{u}_i$, and discharge readings q_i from the PLCs to the supervisory control layer and 2) level set points, target offtake discharges, and controller parameters from the supervisory control layer to the PLCs. Thus, compromise of control network may affect the performance of both regulatory and supervisory control layers.

Attacks **A4** and **A5** denote the attacks to the supervisory control layer which implements state estimators for data reconciliation and observers for attack/fault diagnosis. Typically, the state estimates and diagnostic information are used to generate set points and controller parameters by using an optimization procedure or human expertise. Possible attacks here could be manipulation of state estimators and observers of attack/fault diagnostic scheme. Of course, attacks **A1**–**A3** on the regulatory layer may also affect the performance of supervisory layer, since the latter could be fed with incorrect data when the former is under attack. Finally, **A6** denotes the attacks to corporate network, e.g., malicious insiders who manage to assume canal manager's role.

We now present the model of cyber attacks specific to level sensor measurements y_i^d; attacks on control inputs can be similarly modeled. We assume that each sensor is uniquely authenticated, and has a nominal operating range $\mathcal{Y}_i$, i.e., $y_i^d(t) \in \mathcal{Y}_i,\ \forall t$. Let $\tilde{y}_i^d(t)$ denote the measurements received by the SCADA system at time t. If the ith sensor is under attack, $\tilde{y}_i^d(t) \neq y_i^d(t)$. It is reasonable to assume that the compromised measurements $\tilde{y}_i^d(t)$ also lie within $\mathcal{Y}_i$ (measurements outside this range can be detected by standard data reconciliation methods). Furthermore, once the attack is successful, the attacker is likely to continue the attack until he/she exhausts available resources or achieves the final goal. Thus, we assume the following block attack model, with duration $\mathcal{T} := [\tau^s, \tau^e]$ between the start time τ^s and stop time $\tau^e > \tau^s$

$$\tilde{y}_i^d(t) = \begin{cases} y_i^d(t), & \text{for } t \notin \mathcal{T} \\ g_i(t), & \text{for } t \in \mathcal{T}, \quad g_i(t) \in \mathcal{Y}_i \end{cases} \tag{1}$$

where $g_i(t)$ is an arbitrary (deception or DoS) attack signal.

III. Flow Models and Hierarchical System Architecture

We now describe a PI controller (regulatory layer) and a diagnostic scheme (supervisory layer) based on a reduced-order model of cascaded canal systems.

A. Model of Canal Cascade

The following frequency-domain input–output relationship has been obtained by Litrico and Fromion [11] by taking the Laplace transform of the linearized shallow water equations:

$$\hat{y}_i^d(s) = p_{i,21}(s)\hat{q}_{i-1}(s) + p_{i,22}(s)\left(\hat{q}_i(s) + \hat{p}_i(s)\right) \tag{2}$$

[1] We conducted this test attack approximately nine months before the discovery of Stuxnet.

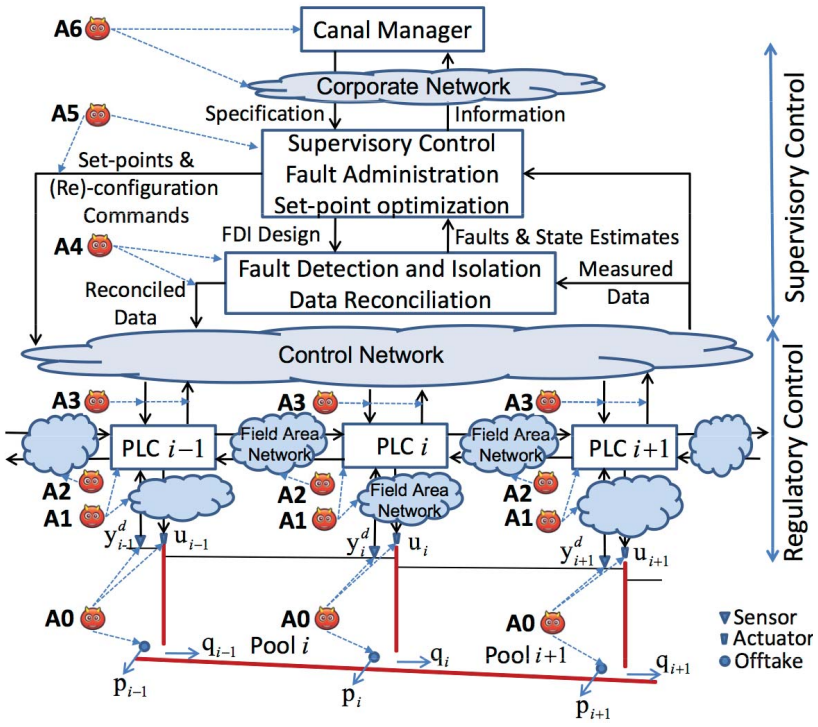

Fig. 1. Cyber attacks to hierarchical water SCADA systems.

where $\hat{f}(s) = \mathcal{L}\{f(t)\}$ is the Laplace transform of a function $f(t)$ defined for all $t \geqslant 0$, s is the Laplace variable, and $p_{i,21}(s)$ [resp. $p_{i,22}(s)$] denotes the infinite-dimensional transfer function from q_{i-1} (resp. q_i and p_i) to y_i^d. For low frequencies, these transfer functions can be approximated by the following integrator-delay (ID) model:

$$p_{i,21}(s) \approx \frac{a_i^d}{s} e^{-\underline{\tau}_i s} \qquad p_{i,22}(s) \approx -\frac{a_i^d}{s} \tag{3}$$

where a_i^d is the inverse of equivalent backwater area (m^{-2}) and $\underline{\tau}_i$ is the propagation delay (s). Using (2), the multipool representation of the canal cascade is obtained as

$$\hat{y}^d(s) = \mathcal{G}(s)\hat{q}(s) + \tilde{\mathcal{G}}(s)\hat{p}(s) \tag{4}$$

where $y^d = \left(y_1^d, \ldots, y_m^d\right)$, $q = \left(q_0, \ldots, q_m\right)$, $p = \left(p_1, \ldots, p_m\right)$, $\mathcal{G}(s) = (g_{jk}(s))$ is a $m \times (m+1)$ dimensional bidiagonal matrix, and $\tilde{\mathcal{G}}(s) = (\tilde{g}_{jk}(s))$ is a $m \times m$ dimensional diagonal matrix. We will henceforth consider a two-pool system, noting that our analysis also extends to multipool systems. Taking the inverse Laplace transform of (4) for $m = 2$, we obtain the following time-domain model with delayed inputs:

$$\dot{y}_i^d(t) = a_i^d q_{i-1}(t - \underline{\tau}_i) - a_i^d \left[q_i(t) + p_i(t)\right], \qquad i = 1, 2. \tag{5}$$

Each regulation gate is represented by the following linearized model around the steady state:

$$q_i(t) = b_i^d y_i^d(t) + k_i u_i(t), \qquad i = 1, 2 \tag{6}$$

where $u_i(t)$ is the gate opening, and the constant b_i^d (resp. k_i) denotes the gain of y_i^d (resp. $u_i(t)$). Using (5) and (6), we obtain the following system representation:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{i=0}^{2} A_i \mathbf{x}(t - \tau_i) + \sum_{i=0}^{2} B_i \mathbf{u}(t - \tau_i) \\ \mathbf{y}(t) &= C\mathbf{x}(t) \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathbf{x} &:= \left(y_1^d, y_2^d\right)^{\mathsf{T}} \in \mathbb{R}^2 \\ \mathbf{u} &:= \left(u_0, u_1, p_1, p_2\right)^{\mathsf{T}} \in \mathbb{R}^4 \\ \mathbf{y} &:= \left(y_1^d, y_2^d\right)^{\mathsf{T}} \in \mathbb{R}^2 \end{aligned}$$

and

$$\tau_0 = 0 \qquad \tau_1 = \underline{\tau}_1 \qquad \tau_2 = \underline{\tau}_2.$$

Let $\tau_{\max}$ denote the upper bound of the time delays τ_i, $i = 0, 1, 2$. The system matrices are, respectively, given by $C = \mathrm{diag}\left(1,\ 1\right)$, and

$$\begin{aligned} A_0 &= \begin{pmatrix} -a_1^d b_1^d & 0 \\ 0 & -a_2^d b_2^d \end{pmatrix} \quad A_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ A_2 &= \begin{pmatrix} 0 & 0 \\ a_2^d b_1^d & 0 \end{pmatrix} \quad B_0 = \begin{pmatrix} 0 & -a_1^d k_1 & -a_1^d & 0 \\ 0 & 0 & 0 & -a_2^d \end{pmatrix} \\ B_1 &= \begin{pmatrix} a_1^d k_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a_2^d k_1 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{8}$$

B. Regulatory Control

Following [1], we briefly discuss the structure of frequency-domain-based regulatory controllers for multipool canal cascades, and refer the reader to [11] and [12] for decentralized design of these controllers. Let y_i^r denote the set point for

pool i, which is typically obtained from the supervisory layer. The aim of the regulatory control is to regulate y_i^d to set point y_i^r. Let the output error be defined as $\epsilon_i := (y_i^r - y_i^d)$, and $y^r := \left(y_1^r, \ldots, y_m^r\right)$, $\epsilon := \left(\epsilon_1, \ldots, \epsilon_m\right)$. Let $\mathcal{K}(s)$ denote the Laplace transform of the multivariable controller $\mathcal{K}$

$$\hat{q}(s) = \mathcal{K}(s)\hat{\epsilon}(s). \tag{9}$$

From (4) and (9), we see that the control input vector $\hat{q}(s)$ and the output error ϵ are given by

$$\hat{q}(s) = \mathcal{S}_q(s)\mathcal{K}(s)\hat{y}^r(s) - \mathcal{S}_q(s)\mathcal{K}(s)\tilde{\mathcal{G}}(s)\hat{p}(s) \tag{10}$$

$$\hat{\epsilon}(s) = \mathcal{S}_\epsilon(s)\hat{y}^r(s) - \mathcal{S}_\epsilon(s)\tilde{\mathcal{G}}(s)\hat{p}(s) \tag{11}$$

where

$$\mathcal{S}_q(s) := (I + \mathcal{K}(s)\mathcal{G}(s))^{-1} (\text{resp.} \mathcal{S}_\epsilon(s) := (1 + \mathcal{G}(s)\mathcal{K}(s))^{-1})$$

denotes input (resp. output) sensitivity function.

C. Supervisory Control

We now describe a model-based diagnostic scheme for detection and isolation of unknown withdrawals from the canal offtakes. Let us consider faults $f_i(t) := \delta p_i(t)$, $i = 1, 2$, which represent the unmeasured water withdrawals occurring nonsimultaneously through the downstream canal offtakes. Extending (7) to include such faults, we obtain

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{i=0}^{2} A_i \mathbf{x}(t-\tau_i) + \sum_{i=0}^{2} B_i \mathbf{u}(t-\tau_i) + \sum_{j=1}^{2} E_j f_i(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) \end{aligned} \tag{12}$$

with A_i, B_i, C given by (8), and

$$E_1 = \begin{pmatrix} -a_1^d \\ 0 \end{pmatrix} \qquad E_2 = \begin{pmatrix} 0 \\ -a_2^d \end{pmatrix}. \tag{13}$$

The delay τ_i $(i = 1, 2)$ in (12) corresponds to the travel time of a small perturbation in the upstream flow to reach the downstream end of the ith canal pool. It can be obtained via direct system identification [1] or model reduction of (2) (see [11, Ch. 4]).

Our diagnostic scheme is based on a bank of two UIO, i.e., observer 1 (resp. observer 2) is designed to be insensitive to $f_1(t)$ (resp. $f_2(t)$). We build on past results on the design of UIO [13], [14], and utilize delay-dependent stability conditions for time-delay systems [15]. While we detail the diagnostic scheme for a simple canal system as in [2], we focus on its evaluation against remote deception attacks (see Section IV).

The residual $\mathbf{r}_j$ of the jth observer is defined as follows:

$$\mathbf{r}_j(t) := \mathbf{y}_j(t) - C\hat{\mathbf{x}}_j(t) \tag{14}$$

where $\hat{\mathbf{x}}_j(t)$ is the jth observer's output denoting the state of the following fault model:

$$\begin{aligned} \dot{\mathbf{x}}_j(t) &= \sum_{i=0}^{2} A_i \mathbf{x}_j(t-\tau_i) + \sum_{i=0}^{2} B_i \mathbf{u}_j(t-\tau_i) \\ &\quad + E_j f_j(t) + E_{-j} f_{-j}(t) \\ \mathbf{y}_j(t) &= C\mathbf{x}_j(t). \end{aligned} \tag{15}$$

TABLE I
DECISION TABLE FOR A TWO-POOL SYSTEM UNDER OFFTAKE WITHDRAWALS

If	$\|\mathbf{r}_1\|$	$\|\mathbf{r}_2\|$
$f_1 \neq 0$	≈ 0	$\neq 0$
$f_2 \neq 0$	$\neq 0$	≈ 0

Here, the matrices A_i, B_i $i = 0, 1, 2$, and C are given by (8), and vectors E_j, E_{-j} are given by (13) with $-j := (3 - j)$, $j = 1, 2$. The following (full-order) model:

$$\begin{aligned} \dot{\mathbf{z}}_j(t) &= \sum_{i=0}^{2} F_{ij} \mathbf{z}_j(t-\tau_i) + \sum_{i=0}^{2} T_j B_i u_j(t-\tau_i) \\ &\quad + \sum_{i=0}^{2} G_{ij} \mathbf{y}_j(t-\tau_i) \\ \hat{\mathbf{x}}_j(t) &= \mathbf{z}_j(t) + N_j \mathbf{y}_j(t) \end{aligned} \tag{16}$$

with initial state $\mathbf{z}_j(\theta) = \rho(\theta)$, $\forall \theta \in [-\tau_{\max}, 0]$, describes the dynamics of the jth observer for the fault model (15), and F_{ij}, G_{ij}, $i = 0, 1, 2$, T_j, and N_j are unknown parameter matrices with real-valued elements. The design of observers is based on the following proposition.

Proposition 1: If the parameter matrices F_{ij}, G_{ij}, $i = 0, 1, 2$, T_j, and N_j in the jth observer (16) are such that the residuals $\mathbf{r}_j(t) = (\mathbf{y}_j(t) - C\hat{\mathbf{x}}_j(t))$, $j = 1, 2$ satisfy the properties:

1) $\mathbf{r}_j(t)$ is insensitive to $f_j(t)$;
2) $\mathbf{r}_j(t)$ asymptotically converges to zero if $f_{-j}(t) = 0$ for every t;
3) $\|\mathbf{r}_j(t)\| \neq 0$ when $f_{-j}(t) \neq 0$

then the diagnosis of faults can be achieved using the decision rule presented in Table I. □

In order to achieve this observer design objective, we observe from (14)–(16) that the residual $\mathbf{r}_j(t)$ can be written as output of the error dynamic

$$\begin{aligned} \dot{\mathbf{e}}_j(t) &= \sum_{i=0}^{2} F_{ij} \mathbf{e}_j(t-\tau_i) + T_j E_j f_j(t) + T_j E_{-j} f_{-j}(t) \\ &\quad - \sum_{i=0}^{2} \left(F_{ij} + \bar{G}_{ij} C - T_j A_i\right) \mathbf{x}_j(t-\tau_i) \\ \mathbf{r}_j(t) &= C\mathbf{e}_j(t) \end{aligned} \tag{17}$$

where $\mathbf{e}_j(t) := \mathbf{x}_j(t) - \hat{\mathbf{x}}_j(t)$, and

$$\bar{G}_{ij} := (G_{ij} - F_{ij} N_j), \qquad i = 0, 1, 2 \tag{18}$$

$$T_j := (\mathrm{I}_n - N_j C). \tag{19}$$

Consider the following conditions:

$$F_{ij} = T_j A_i - \bar{G}_{ij} C, \qquad i = 0, 1, 2 \tag{20}$$

$$T_j E_j = 0 \tag{21}$$

$$\dot{\mathbf{e}}_j(t) = \sum_{i=0}^{2} F_{ij} \mathbf{e}_j(t-\tau_i) \quad \text{is asymptotically stable.} \tag{22}$$

Let (18)–(22) hold, and note from (13) that E_1 and E_2 are linearly independent. Then it can be concluded that $T_j E_{-j} \neq 0$, $j = 1, 2$. Thus, $\mathbf{r}_1$ and $\mathbf{r}_2$ satisfy the conditions of Proposition 1.

The computation of observer parameter matrices F_{ij}, G_{ij}, T_j, and N_j for $i = 0, 1, 2$ and $j = 1, 2$ proceeds in two steps. (For simplicity, we will henceforth omit the observer index j.)

Step 1: To check that the parameter matrices in (15) satisfy Proposition 1, we express (19)–(21) as

$$S\Theta = \Psi \tag{23}$$

where $S = \begin{pmatrix} T & N & F_0 & \bar{G}_0 & F_1 & \bar{G}_1, & F_2 & \bar{G}_2 \end{pmatrix}$, $\Psi = \begin{pmatrix} \mathsf{I}_n & 0 & 0 & 0 & 0 \end{pmatrix}$, and

$$\Theta = \begin{pmatrix} \mathsf{I}_n & E & A_0 & A_1 & A_2 \\ C & 0 & 0 & 0 & 0 \\ 0 & 0 & -\mathsf{I}_n & 0 & 0 \\ 0 & 0 & -C & 0 & 0 \\ 0 & 0 & 0 & -\mathsf{I}_n & 0 \\ 0 & 0 & 0 & -C & 0 \\ 0 & 0 & 0 & 0 & -\mathsf{I}_n \\ 0 & 0 & 0 & 0 & -C \end{pmatrix}.$$

Under the condition that $\operatorname{rank}(CE) = \operatorname{rank}(E)$, the general solution of (23) is

$$S = \Psi\Theta^{+} - K(I - \Theta\Theta^{+}) \tag{24}$$

where K is an arbitrary matrix of appropriate dimension, and Θ^{+} is the generalized inverse matrix of Θ. By inserting the solution (24) in (20), the matrices F_i can now be expressed as

$$F_i = \chi_i - K\beta_i, \qquad i = 0, 1, 2 \tag{25}$$

where

$$\chi_0 = \Psi\Theta^{+} \begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}}$$
$$\beta_0 = \tilde{\Theta} \begin{pmatrix} A_0 & 0 & 0 & -C & 0 & 0 & 0 & 0 \end{pmatrix}^{\mathsf{T}}$$

with $\tilde{\Theta} = (I - \Theta\Theta^{+})$; similarly for χ_1, χ_2, β_1, and β_2. Now, we can rewrite (22) as

$$\dot{\mathbf{e}}(t) = \sum_{i=0}^{2} (\chi_i - K\beta_i)\mathbf{e}(t - \tau_i) \quad \text{is asymptotically stable.} \tag{26}$$

Step 2: The following result enables computation of matrix K such that (26) holds.

Proposition 2: System in (26) is asymptotically stable if for some scalars $\epsilon_0, \ldots, \epsilon_6$, $\bar{\epsilon}_1, \bar{\epsilon}_2$, there exist matrices $S_i > 0$, $Z_i > 0$, $Q_i > 0$, $R_i > 0$, U_i, W_i, $i = 1, 2$, and matrices H_j, $j = 1, \ldots, 6$, U and $P > 0$ such that the following linear matrix inequalities (LMIs) are satisfied:

$$\begin{pmatrix} \Phi & h_1\bar{H}_1 & h_2\bar{H}_2 \\ * & -h_1\bar{Z}_1 & 0 \\ * & * & -h_2\bar{Z}_2 \end{pmatrix} < 0 \qquad \begin{pmatrix} Q_k & U_k \\ U_k^{\mathsf{T}} & R_k \end{pmatrix} \geqslant 0$$

$$\bar{Z}_k \coloneqq \begin{pmatrix} S_k & W_k \\ * & Z_k \end{pmatrix} \qquad \bar{H}_k \coloneqq \begin{pmatrix} -\bar{\epsilon}_k(P\chi_1 - U\beta_1)^{\mathsf{T}} & H_1 \\ -\bar{\epsilon}_k(P\chi_2 - U\beta_2)^{\mathsf{T}} & H_2 \\ -\bar{\epsilon}_k(P\chi_3 - U\beta_3)^{\mathsf{T}} & H_3 \\ \bar{\epsilon}_k P & H_4 \\ 0 & H_5 \\ 0 & H_6 \end{pmatrix}$$

for $k = 1, 2$, and $\Phi = (\phi_{ij})$ is a symmetric matrix with following block elements:

$$\begin{aligned}
\phi_{11} &= \sum_{i=1}^{2}(Q_i + h_i S_i) + \epsilon_1 \operatorname{sym}(P\chi_0 - U\beta_0) + 2\operatorname{sym}(H_1) \\
\phi_{12} &= \epsilon_1(P\chi_1 - U\beta_1) + \epsilon_2(P\chi_0 - U\beta_0)^{\mathsf{T}} + 2H_2^{\mathsf{T}} - H_1 \\
\phi_{13} &= \epsilon_3(P\chi_0 - U\beta_0)^{\mathsf{T}} + \epsilon_1(P\chi_2 - U\beta_2) + 2H_3^{\mathsf{T}} - H_1 \\
\phi_{14} &= P + \sum_{i=1}^{2}(U_i + h_i W_i) + \epsilon_4(P\chi_0 - U\beta_0)^{\mathsf{T}} + 2H_4^{\mathsf{T}} - \epsilon_1 P \\
\phi_{15} &= 2H_5^{\mathsf{T}} + \epsilon_5(P\chi_0 - U\beta_0)^{\mathsf{T}} \\
\phi_{16} &= 2H_6^{\mathsf{T}} + \epsilon_6(P\chi_0 - U\beta_0)^{\mathsf{T}} \\
\phi_{22} &= -Q_1 - \operatorname{sym}(H_2) + \epsilon_2 \operatorname{sym}(P\chi_1 - U\beta_1) \\
\phi_{23} &= -H_3^{\mathsf{T}} - H_2 + \epsilon_2 \operatorname{sym}(P\chi_2 - U\beta_2) + \epsilon_3 \operatorname{sym}(P\chi_1 - U\beta_1)^{\mathsf{T}} \\
\phi_{24} &= -H_4^{\mathsf{T}} + \epsilon_4 \operatorname{sym}(P\chi_1 - U\beta_1)^{\mathsf{T}} - \epsilon_2 P \\
\phi_{25} &= -U_1 - H_5^{\mathsf{T}} + \epsilon_5 \operatorname{sym}(P\chi_1 - U\beta_1)^{\mathsf{T}} \\
\phi_{26} &= -H_6^{\mathsf{T}} + \epsilon_6(P\chi_1 - U\beta_1)^{\mathsf{T}} \\
\phi_{33} &= -Q_2 + \epsilon_3 \operatorname{sym}(P\chi_2 - U\beta_2) - \operatorname{sym}(H_3) \\
\phi_{34} &= -\epsilon_3 P + \epsilon_4(P\chi_2 - U\beta_2)^{\mathsf{T}} - H_4^{\mathsf{T}} \\
\phi_{35} &= +\epsilon_5(P\chi_2 - U\beta_2)^{\mathsf{T}} - H_5^{\mathsf{T}} \\
\phi_{36} &= -U_2 + \epsilon_6(P\chi_2 - U\beta_2)^{\mathsf{T}} - H_6^{\mathsf{T}} \\
\phi_{44} &= \sum_{i=1}^{2}(R_i + h_i Z_i) - \epsilon_4 \operatorname{sym}(P) \\
\phi_{45} &= -\epsilon_5 P^{\mathsf{T}} \\
\phi_{46} &= -\epsilon_6 P^{\mathsf{T}} \\
\phi_{55} &= -R_1 \\
\phi_{56} &= 0 \\
\phi_{66} &= -R_2
\end{aligned}$$

where $h_i = \tau_i$, and $\operatorname{sym}(M) \coloneqq M + M^{\mathsf{T}}$. The parameter matrix K is given by $K = P^{-1}U$. □

The proof is analogous to the proof of Proposition 4 in the companion paper [10].

Remark 3: The error dynamics (17) does not depend on the regulatory-layer control input **u**. Thus, the behavior of diagnostic scheme is not affected by changes in **u**.

IV. Field Operational Test Attacks

In this section, we present results from a field operation test in which deception attacks were implemented on the Gignac canal network. This canal network is located about 40 km northwest of Montpellier in Southern France. The main canal network is comprised of an 8-km feeder canal which emerges from the Hérault river, and bifurcates at the diversion structure Partiteur into two branches: left branch (27 km) and right branch (15 km). The design flow of the canal is about 3.5 m^3/s. The SCADA system's centralized control station communicates with field devices in real time over the Internet and public switched telephone networks.

A. Field Operational Test Setup

In our field operational test, we consider a two-pool system situated on the canal branch which diverts from the

Partiteur device to the right bank of the Hérault river. The first pool is the 4.8-km canal reach between the Partiteur device and the Avencq cross-regulator, and the second pool is the 5.2-km reach between the Avencq and Lagarel cross-regulators. The ID model parameters for the respective pools are given by $a_1^d = 1.105\times10^{-4}$ m^{-2}, $a_2^d = 2.597\times10^{-5}$ m^{-2}, $\underline{\tau}_1 = 45$ min, $\underline{\tau}_2 = 40$ min. During our experiment, the Lagarel gate was submerged and, therefore, the linearized gate equation (6) for free-flow gate cannot be used. The linearized gate equation for submerged gate is given by

$$q_i(t) = b_i^d y_i^d(t) + b_{i+1}^u y_{i+1}^u(t) + k_i u_i(t), \qquad i = 1, 2 \quad (27)$$

where b_{i+1}^u is the gain from the water level y_{i+1}^u downstream of the gate. Both Avencq and Lagarel regulators are equipped with motorized gates and level sensors, and communicate with the base station via radio links. The discharge required to regulate the upstream water level in response to perturbations caused by offtake withdrawals is achieved by the slave controller (PLC) via movement of a 1-m-wide sluice gate; see Fig. 2. We now implement a deception attack (**A**1 in Fig. 1) on the Avencq regulator. We assume that the attacker has the knowledge of: 1) the approximate system dynamics; 2) the parameters of diagnostic scheme; and 3) the sensor-control signals. The attacker's intent is to steal water from the canal system by attacking the downstream level sensor measurements y^d.[2] Our goal is to synthesize a stealthy attack, i.e., an attack that evades detection by the SCADA system. Such a powerful attack can be conducted in practice by a malicious insider or a computer hacker who is able to bypass existing IT security mechanisms and adaptively manipulate sensor measurements. Recall that performance of PI controller (resp. UIO-based diagnostic scheme) essentially depends on the output error ϵ (resp. observer residual $\mathbf{r}$) as defined in (9) in Section III-B [resp. (14) in Section III-C]. Thus, a stealthy attack should manipulate these quantities to avoid detection.

B. Effect of Cyber Attack on Regulatory Control

We assess the performance loss of regulatory control under compromise of y^d at the Avencq gate, first, in simulation and, then, in a field operational test on the Gignac canal. In our setting, the steady-state water level is $\bar{Y} = 79$ cm. The regulatory control aim is to stabilize y^d (which is the deviation from $\bar{Y}$) to 0, i.e., a set point $y^r = 0$ cm. The upstream water level is measured every 2 min, and a PI controller

$$\kappa(s) = k\left(1 + \frac{1}{Ts}\right)$$

with the proportional gain $k = -2.9$ and the integral time $T = 360$ s is used to regulate y^d at Avencq gate. Now, if the attack signal $g(t)$ in the attack model (1) is chosen such that error under attack $\tilde{\epsilon} = (\tilde{y}^d - y^r)$ is close to zero, then from (9)–(11), we conclude that the regulatory controller will not react correctly to reject water level deviations from the set point y^r. Thus, $g(t) \approx y^r(t)$ achieves a stealthy attack for regulatory control layer.

[2]As shown in the companion paper [10], downstream level measurements provide a more critical diagnosis of faults/attacks.

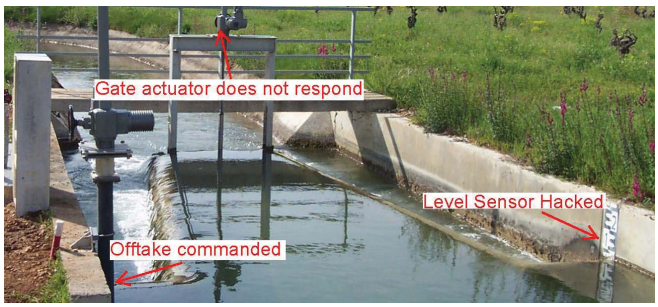

Fig. 2. Level sensor of Avencq station manipulated to disable the controller to correctly actuate the sluice gate.

We now demonstrate the feasibility of deception attacks with a field operational test on Avencq. The experiment was performed on October 12, 2009. We carried out the attack by directly modifying the sensor measurements sent from the SCADA system's real-time software interface to the MATLAB code which implemented the PI controller. At the start of experiment, the PI controller reacts by changing set points every few minutes and then letting the water level stabilize close to set point in a closed loop. At $t = 90$ min, the offtake is opened and the attacker injects false data to water level measurement; see Fig. 3.

At around $t = 184$ min, the offtake was fully opened and then at around $t = 190$ min it is fully closed by a physical intervention at the Avencq cross-regulator; see Fig. 3(a). This effect is captured in the sudden drop in the actual water level as shown in Fig. 3(c). From $t = 190$ min until $t = 510$ min, the attacker continues the attack; see Fig. 3(b). This results in open-loop response of actual water level, i.e., the PI controller fails to react to the actual perturbation. A residual error still remains after the end of the attack, and the PI controller reacts to this error as seen in Fig. 3(d). This may signal an *a posteriori* detection; however, it may still be difficult to distinguish between a residual error resulting from an attack and an error resulting from random disturbances in y^d. The amount of water the attacker manages to withdraw from the offtake between $t = 90$ min and $t = 190$ min can be computed by integrating the gate discharge equation

$$Q(t) = C_g L_g U\sqrt{2g\left(y^d(t) + \bar{Y}\right)}$$

where $C_g \approx 0.6$ denotes the discharge coefficient, $L_g = 1$ m the gate width, $U = 0.03$ m the offtake opening, and $(y^d(t) + \bar{Y})$ the actual water level.

C. Effect of Cyber Attack on Supervisory Diagnostic Scheme

In order to assess the effect of attacks on the diagnostic scheme proposed in Section III-C, we collected the 15-min archived data from the Gignac SCADA system. The data includes the upstream and downstream water levels, gate openings, and discharges. We approach the design of UIOs (16) by considering the withdrawal through pool 1 offtake during 90–190 min as an unknown input. The LMI conditions in Proposition 2 are found to be feasible, and the observer parameter matrices are obtained from the two-step procedure

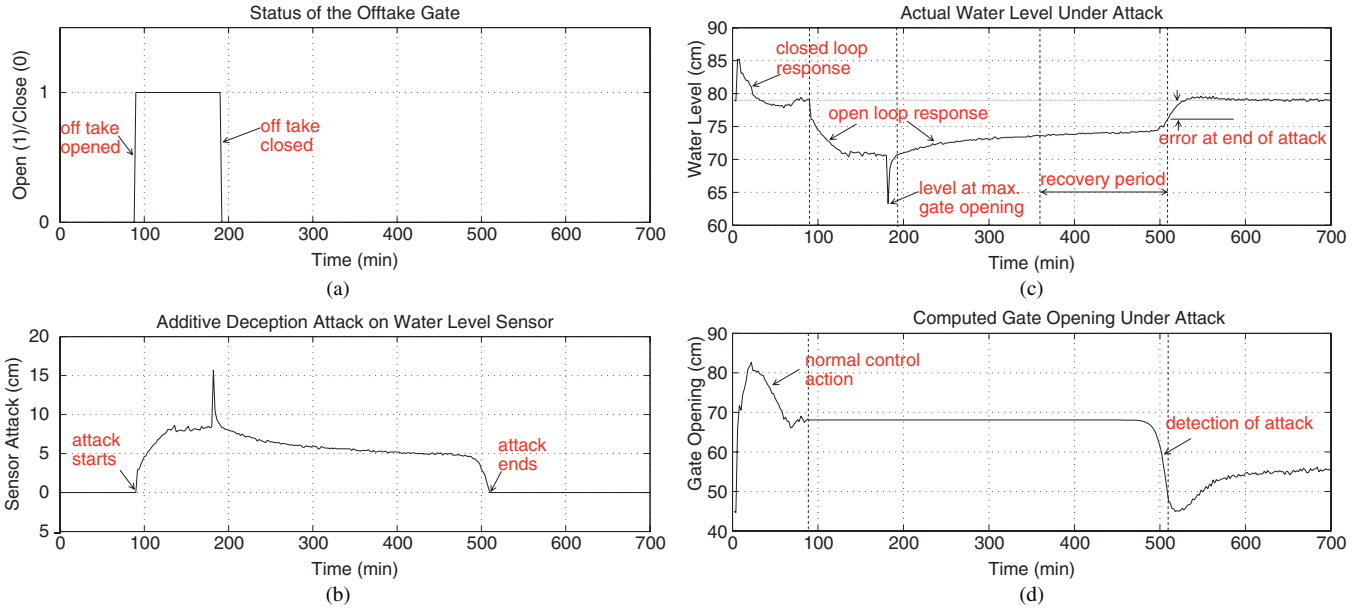

Fig. 3. Performance of local upstream PI controller at Avencq cross-regulator under attack. (a) Offtake opening. (b) Additive deception attack. (c) Actual water level. (d) Gate opening.

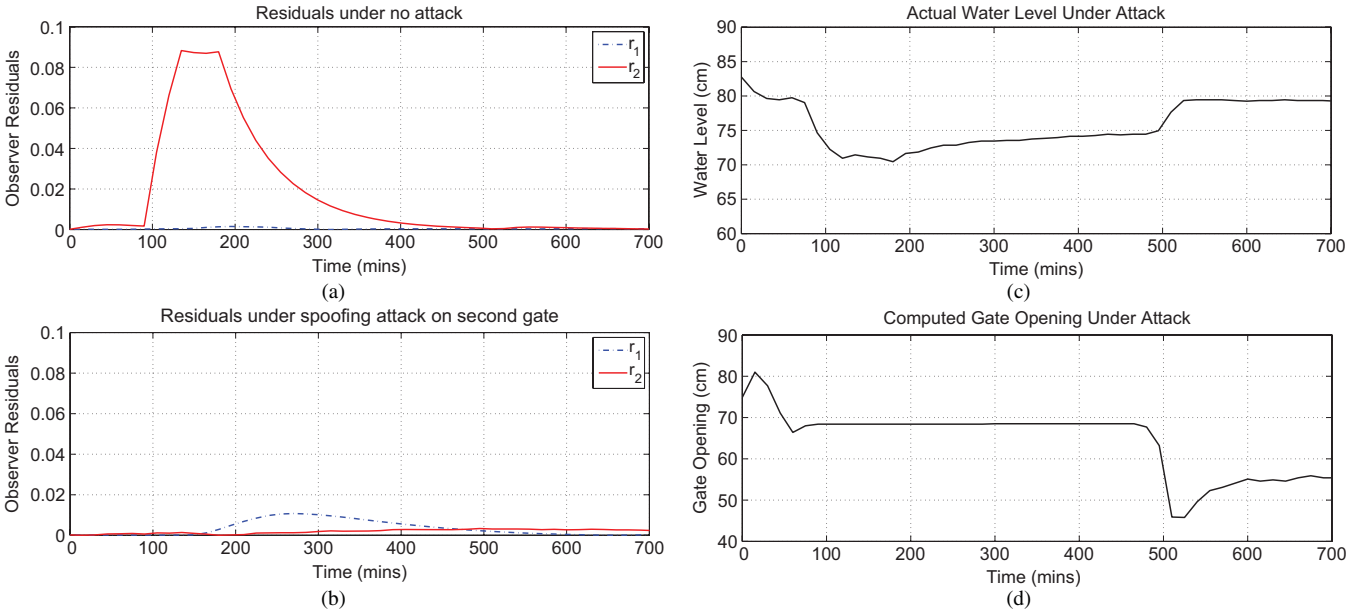

Fig. 4. Performance of the supervisory diagnostic scheme. (a) Observer residuals under no attack. (b) Observer residuals under attack. (c) Actual water level under attack. (d) Gate opening under attack.

in Section III-C. We note that

$$T_1E_1 = 10^{-15}\begin{pmatrix}0.247\\-112\end{pmatrix} \approx 0 \qquad T_1E_2 = \begin{pmatrix}-0.001\\-0.7595\end{pmatrix} \neq 0$$

$$T_2E_1 = \begin{pmatrix}-0.729\\0.0\end{pmatrix} \neq 0 \qquad T_2E_2 = 10^{-15}\begin{pmatrix}0.00\\-0.125\end{pmatrix} \approx 0.$$

Thus, the residuals satisfy the conditions of Proposition 1, and correct diagnosis is achieved (see Table I).

Indeed, Fig. 4(a) shows that, under no attack on sensor measurements, the residual of observer 2 is sensitive to fault occurring in the form of lateral withdrawal in pool 1. However, when Avencq's sensor measurements are compromised and false data $g = 0$ is injected, the observer residuals no longer indicate a correct diagnosis as shown in Fig. 4(b). The actual upstream water level at the Avencq regulator and the computed gate opening are shown in Fig. 4(c) and (d), respectively. Notice that spike in the actual water level in Fig. 3(c) is not captured by the 15-min archived data that is stored by the SCADA system [Fig. 4(c)]. Thus, this spike is not visible to the diagnostic scheme, i.e., it is not reflected in the residual computed from the observers.

V. Conclusion

In this brief, we conducted security threat assessment of hierarchically structured water SCADA systems, and presented

results from a field operational test on the Gignac water SCADA system. We studied the effect of stealthy deception attacks on a PI control scheme (regulatory layer) and a UIO-based diagnostic scheme (supervisory layer). Both schemes used downstream sensor measurements. This is typically the case when offtakes are located at the downstream end of the canal pools. We find that, although the diagnostic scheme works well for nonsimultaneous random withdrawals [i.e., fault model (12)], it is not robust to the deception attack (1). Our field operational test demonstrates that such attacks can be stealthy, i.e., they can bypass detection by the SCADA system. The characterization of stealthy attacks is important because it can guide the deployment of IT-specific security mechanisms and enable the design of better attack diagnostic schemes.

Our analysis can be extended to the case when multiple sensor measurements y_i^d are subject to attacks. A possible stealthy attack strategy is to first compromise the most downstream sensor measurement y_m^d, and systematically proceed to compromise upstream sensor measurements $y_{m-1}^d, y_{m-2}^d, \ldots, y_1^d$. An interesting research question is then to characterize the relation between the resources required by the attacker and the impact of the resulting attack. Such analyses can ultimately lead to a rigorous framework for security threat assessment of NCS/SCADA systems.

ACKNOWLEDGMENT

The authors would like to thank C. Hugodot, Director of the Canal de Gignac, Gignac, France, and D. Dorchies, Cemagref Research Staff Member for their help during the field operational test attack. The authors are also grateful to the anonymous reviewers for their valuable feedback. The field operational test attack was conducted when X. Litrico was with Cemagref, Unité Mixte de Recherche G-EAU, Montpellier, France.

REFERENCES

[1] X. Litrico, P.-O. Malaterre, J.-P. Baume, P.-Y. Vion, and J. Ribot-Bruno, "Automatic tuning of PI controllers for an irrigation canal pool," *J. Irrigat. Drainage Eng.*, vol. 133, no. 1, pp. 27–37, 2007.

[2] D. Koenig, N. Bedjaoui, and X. Litrico, "Unknown input observers design for time-delay systems application to an open-channel," in *Proc. IEEE 44th Conf. Decision Control*, Dec. 2005, pp. 5794–5799.

[3] A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proc. 3rd Conf. Hot Topics Security*, Jul. 2008, pp. 1–6.

[4] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Security*, 2009, pp. 21–32.

[5] J. Slay and M. Miller, "Lessons learned from the Maroochy Water breach," in *Critical Infrastructure Protection*. New York: Springer-Verlag, Nov. 2007, pp. 73–82.

[6] U. Attorney. (2007, Nov.). *Willows Man Arrested for Hacking into Tehama Colusa Canal Authority Computer System* [Online]. Available: http://www.usdoj.gov/usao/cae/press_releases/

[7] N. Falliere, L. Murchu, and E. Chien, *W32.Stuxnet Dossier*. Mountain View, CA: Symantec, Sep. 2010.

[8] N. Bedjaoui and E. Weyer, "Algorithms for leak detection, estimation, isolation and localization in open water channels," *Control Eng. Practice*, vol. 19, no. 6, pp. 564–573, Jun. 2011.

[9] A. A. Cárdenas, S. Amin, Z.-Y. Lin, Y.-L. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proc. 6th ACM Symp. Inform. Comput. Commun. Security*, Mar. 2011, pp. 355–366.

[10] S. Amin, X. Litrico, S. Sastry, and A. Bayen, "Cyber security of water SCADA systems: II attack detection using an enhanced hydrodynamic model," *IEEE Trans. Control Syst. Technol.*, 2012, doi:10.1109/TCST.2012.2211874.

[11] X. Litrico and V. Fromion, *Modeling and Control of Hydrosystems*. New York: Springer-Verlag, 2009.

[12] M. Cantoni, E. Weyer, Y. Li, S.-K. Ooi, I. Mareels, and M. Ryan, "Control of large-scale irrigation networks," *Proc. IEEE*, vol. 95, no. 1, pp. 75–91, Jan. 2007.

[13] G. Conte and A. Perdon, "Unknown input observers and residual generators for linear time delay systems," in *Current Trends in Nonlinear Systems and Control*, Cambridge, MA: Birkhaüser, 2006, pp. 15–33.

[14] M. Darouach, M. Zasadzinski, and S. Xu, "Full-order observers for linear systems with unknown inputs," *IEEE Trans. Autom. Control*, vol. 39, no. 3, pp. 606–609, Mar. 1994.

[15] C. Lin, Q.-G. Wang, and T. Lee, "A less conservative robust stability test for linear uncertain time-delay systems," *IEEE Trans. Autom. Control*, vol. 51, no. 1, pp. 87–91, Jan. 2006.