Article publicat / *Published paper*:

Quevedo, J., Sanchez, H., Rotondo, D., Escobet, T., Puig, V. A two-tank benchmark for detection and isolation of cyber attacks. "IFAC-PapersOnLine", 1 Gener 2018, vol. 51, núm. 24, p. 770-775. DOI: 10.1016/j.ifacol.2018.09.662

A two-tank benchmark for detection and isolation of cyber attacks ⋆

Joseba Quevedo [*,**] **Helem Sánchez** [*,**] **Damiano Rotondo** [*,***]
Teresa Escobet [*,****] **Vicenç Puig** [*,**,***]

[*] *Research Center for Supervision, Safety and Automatic Control (CS2AC) of the Universitat Politècnica de Catalunya (UPC)*
[**] *Automatic Control Department, UPC-ESAII, Rambla de Sant Nebridi, 11, 08222 Terrassa, Spain*
[***] *Institut de Robotica i Informatica Industrial (IRI), UPC-CSIC Carrer de Llorens i Artigas, 4-6, 08028 Barcelona, Spain*
[****] *Department of Mining, Industrial and ICT Engineering, UPC, Av. de les Bases de Manresa, 61-73, 08242 Manresa, Spain*

Abstract: This paper presents a benchmark for the detection and isolation of cyber attacks, which is a non-linear controlled interconnected system based on a two tank system. In this benchmark, a malicious attacker wants to remain hidden while stealing water by altering the signals of the sensors of the levels of the tanks. It is assumed that the attacker can steal water from the tanks using extraction pumps with pre-established flow rates and, depending on the theft and the type of sensor alteration, different attack scenarios are proposed.

Keywords: Cyber attacks, cyber physical systems, benchmark, replay attacks, two-tank system.

1. INTRODUCTION

The fourth industrial revolution has brought new challenges related to connected systems, smart manufacturers and digital supply networks. In the last years, the need for a better integration of computation and physical processes has brought the scientific community to investigate *cyber physical systems* (CPSs). The term *cyber physical* refers to the presence of discrete processing and communication of information together with the physical engineered system (Jeschke et al., 2017). Nowadays, CPSs are found widely in advanced manufacturing systems, transportation networks, industrial control processes, and critical infrastructures (Pasqualetti et al., 2013). However, the increase in efficiency brought by CPSs comes at the cost of a higher risk in safety and security, due to the possibility of someone performing malicious attacks, a.k.a. *cyber attacks*. These attacks, usually motivated by terrorism, criminality or sabotage, exploit the system's vulnerabilities and result in some kind of disturbance or damage in the physical and in the cyber layers. The interconnected nature of Industry 4.0-driven operations means that cyber attacks have far more extensive effects than ever before, and digital systems, computers and their supply networks may not be prepared for this kind of risks (Armesto et al., 2016, Waslo et al., 2017). Cyber attacks are different from faults due to the fact that they do not affect only the physical layer of the CPS, but the cyber one as well.

Over the past decade, many concerns have been raised over the vulnerabilities of industrial control systems to cyber attacks. For this reason, different events and scenarios have been studied, and some diagnostic methods have been proposed and evaluated. For example, it is worth recalling the Tennessee Eastman process (Ricker, 1993) for which some experimental work was conducted in order to analyze the effects of attacks in the process control domain (Huang et al., 2009) and test the resilience against cyber physical assaults (Krotofil and Cárdenas, 2013). Remarkable attacks to water distribution facilities, such as the one which affected the Maroochy Water Services (Queensland, Australia), have contributed to motivate research on cyber security, leading to the proposal of testbeds such as the water distribution (WADI) one (Ahmed et al., 2017).

The availability of a benchmark for testing different diagnosis techniques is beneficial for finding the best strategies to handle undesired situations. Motivated by the successes of the wind turbine benchmark proposed by Odgaard et al. (2013), and later enhanced by Odgaard and Johnson (2013), and of the wind farm benchmark (Odgaard and Stoustrup, 2013), which inspired several solutions to the problems of fault diagnosis (Odgaard and Stoustrup, 2012, Simani and Castaldi, 2013, Blesa et al., 2015, Sanchez et al., 2015) and fault tolerant control (Badihi et al., 2014, Blesa et al., 2014, Odgaard and Stoustrup, 2015, Shi and Patton, 2015), this paper presents a benchmark based on a two-tank interconnected system useful for testing different schemes for detection and isolation of cyber attacks. In particular, the benchmark case study has been derived from a previously proposed fault diagnosis benchmark (Bouamama et al., 2001, 2005, Zhang, 2010) by incorporating

⋆ This work has been partially funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERFD) through the projects DEOCS (ref. MINECO DPI2016-76493) and SCAV (ref. MINECO DPI2017-88403-R), by AGAUR of the Generalitat de Catalunya through the Advanced Control Systems (SAC) group grant (2017 SGR 482) and by the Agència de Gestió d'Ajuts Universitaris i de Recerca. This work has been also supported by the Spanish State Research Agency through the Maria de Maeztu Seal of Excellence to IRI (MDM-2016-0656) and the grant Juan de la Cierva - Formacion (FJCI-2016-29019).

a malicious attacker who wants to steal water from the tanks while remaining hidden through an appropriate alteration of the measurements coming from the level sensors of the tanks.

The remaining of the paper is structured as follows. In Section 2, the functionality of the benchmark along with its model are described. Next, in Section 3, the attack scenarios are presented. Simulation results depicting some relevant variables in the proposed attack scenarios are shown and discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2. BENCHMARK DESCRIPTION

The benchmark [1] consists of two interconnected tanks, which are connected to each other through connecting pipes provided with a valve (see Fig. 1). The first tank, denoted as T_1, receives water from the pump P_1, which is controlled by a proportional-integral (PI) controller. The interconnecting valve V_b is regulated by an ON-OFF controller. On the other hand, the second tank, denoted as T_2, is equipped with the manual outlet valve V_o. The benchmark model has been derived from the one described in Bouamama et al. (2001) by incorporating a possible malicious attacker who has the goal of stealing water from the tanks while going unnoticed thanks to appropriate alterations of the outputs of the sensors, which hide the attacks. In the modified benchmark, it is assumed that the thief can extract water from the tanks using extraction pumps with flow rates Q_{f1} and Q_{f2}, which move the water from the tanks T_1 and T_2 to the theft tanks T_{f1} and T_{f2}, respectively. At the same time, it is assumed that the signals provided by the sensors are sent by wireless to the PI and ON-OFF controller, and the thief is able to hack these signals and modify them. Depending on the type of theft and the type of sensor alteration, different attack scenarios are obtained, as described in Section 3.

Hereafter, the model of the benchmark is described (see Table 1 for the value of the model parameters). First, the models of the different subsystems (pump, PI controller, valves and ON-OFF controller) are provided. Then, the subsystems are merged in order to obtain the global model of the plant. Note that the superscript ^{m} is used to denote variables for which a measurement is available. Additionally, the benchmark simulator provides complementary information about the amount of stolen water volumes in tanks T_{f1} and T_{f2}, denoted as V_{f1} and V_{f2}, respectively, and the real (unaltered) values of the water levels h_1 and h_2. However, these variables should be assumed not to be available to the attack detector.

2.1 Model of the pump

Q_p^m is the outflow from the pump P_1, which is assumed to be proportional to the PI controller output U_p^m. Taking into account that the flow from the pump is limited by physical constraints, modeled as a standard saturation nonlinearity, then Q_p^m is given by

$$Q_p^m(t) = \begin{cases} U_p^m(t) & if\ 0 < U_p^m(t) < Q_{p,\max} \\ 0 & if\ U_p^m(t) \le 0 \\ Q_{p,\max} & if\ U_p^m(t) \ge Q_{p,\max} \end{cases} \quad (1)$$

[1] The benchmark is available at the URL https://cs2ac.upc.edu/en/training-benchmarks/cyber-attacks-benchmark-simulator

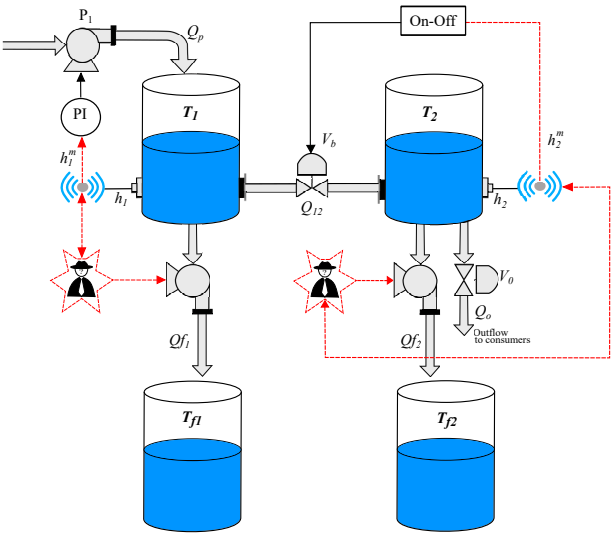

Fig. 1. Schematic diagram of the two-tank benchmark.

2.2 Model of the PI controller

The water level of the tank T_1, denoted as h_1^m, is regulated by a PI controller, whose output is given by

$$U_p^m(t) = K_P\left(h_{1,ref} - h_1^m(t)\right) + K_I \int_0^t \left(h_{1,ref} - h_1^m(\tau)\right) d\tau \quad (2)$$

where $h_{1,ref} = 0.5\,m$ is the set-point for h_1^m, while the proportional and integral gain of the controller are chosen as $K_P = 10^{-3}\,m^{-1}$ and $K_I = 5 \cdot 10^{-6}\,(m \cdot s)^{-1}$, respectively. Note that the set-point can be modified by injecting some small perturbations, which could increase the detectability of attacks, as suggested, e.g., by Scola et al. (2018).

2.3 Model of the valve V_b

The water flow Q_{12} through the valve V_b is controlled by an ON-OFF controller. The flow can be calculated using Bernoulli's law

$$Q_{12}(t) = C_{vb} U_b^m(t) sign\left(h_1^m(t) - h_2^m(t)\right) \sqrt{\left|h_1^m(t) - h_2^m(t)\right|} \quad (3)$$

where $U_b^m \in \{0,1\}$ is the valve position provided by the controller (0 = OFF, 1 = ON), C_{vb} denotes the global hydraulic flow coefficient of the valve V_b and h_2^m is the water level of the tank T_2.

2.4 Model of the ON-OFF controller

The water level h_2^m is regulated by a switching controller, whose output is given by

$$U_b^m(t) = \begin{cases} 0 & if\ 0.09\,m \le h_2^m(t) \le 0.11\,m \\ 1 & if\ 0\,m \le h_2^m(t) \le 0.09\,m \end{cases} \quad (4)$$

2.5 Model of the valve V_o

The water outflow Q_o is controlled by a valve V_o, which is open in nominal regime

$$Q_o(t) = C_{vo}\sqrt{h_2^m(t)}\,U_o^m(t) \quad (5)$$

where C_{vo} is the global hydraulic flow coefficient of the valve V_o, and $U_o^m \in \{0,1\}$ is the valve position provided by the user (0 = CLOSED, 1 = OPENED).

Table 1. Model variables and process parameters.

Symbol	Description	Value	Units
C_{vb}	Hydraulic flow coefficient of the valve V_b	$1.5938 * 10^{-4}$	$m^{3/2}/s$
C_{vo}	Hydraulic flow coefficient of the valve V_o	$1.59640 * 10^{-4}$	$m^{3/2}/s$
$A_{i(i=1,2)}$	Cross-section of the cylindric tank T_i	$1.54 \cdot 10^{-2}$	m^2
$h_{i(i=1,2)}$	Water level in the tank T_i	*variable*	m
$h_{i,\max(i=1,2)}$	Maximum water level in the tank T_i	0.6	m
$Q_{p,\max}$	Maximum outflow from the pump P_1	0.01	m^3/s
$Q_{fi(i=1,2)}$	Flow theft from tanks T_1 and T_2 under attack	10^{-4}	m^3/s
$h_{1,ref}$	Set point of the PI level controller	0.5	m

2.6 Global model of the system

The variation of V_1 and V_2, which are the water volumes in T_1 and T_2, respectively, can be calculated as

$$\dot{V}_i(t) = A_i \dot{h}_i^m(t) = \sum Q_{in,i}(t) - \sum Q_{out,i}(t), \quad i = 1,2 \qquad (6)$$

where A_i denotes the cross-section area of the tank T_i, $\sum Q_{in,i}$ is the sum of all the water inflows into the tank T_i and $\sum Q_{out,i}$ is the sum of all the water outflows from the tank T_i.

In particular, (6) can be rewritten as

$$\dot{V}_1(t) = Q_p(t) - Q_{12}(t) - Q_{f1}(t) \qquad (7)$$

$$\dot{V}_2(t) = Q_{12}(t) - Q_o(t) - Q_{f2}(t) \qquad (8)$$

with $Q_{f1} = Q_{f2} = 0$ when no attack is performed on the system.

Since the water levels h_1 and h_2 are limited by physical constraints, thus

$$h_1(t) = h_{1,\max} \qquad if\, h_1(t) \geq 0.6m \qquad (9)$$

$$h_2(t) = h_{2,\max} \qquad if\, h_2(t) \geq 0.6m \qquad (10)$$

2.7 System measurements

It is assumed that the available measurements are given by

$$y_x^m = y_x + \varepsilon_{yx} \qquad (11)$$

where $y_x \in \{Q_p, U_p, h_1, h_2, U_b, U_o\}$ are the measured variables, and ε_{yx} denotes the corresponding measurement noise. The values of the sensors noises are provided in the file *init.m*, located in the directory *Benchmark Program Simulation*, and are obtained as uniformly distributed signals.

2.8 Analytical Redundancy Relations

The benchmark is completed by residuals designed for performing a traditional fault diagnosis based on analytical redundancy relations (Staroswiecki and Comtet-Varga, 2001), which are calculated as

$$r_1(t) = -C_{vb} U_b^m(t) sign\left(h_1^m(t) - h_2^m(t)\right) \sqrt{\left|h_1^m(t) - h_2^m(t)\right|} + Q_p^m(t) - Q_{f1}(t) - A_1 \frac{dh_1^m}{dt} \qquad (12)$$

$$r_2(t) = C_{vb} U_b^m(t) sign\left(h_1^m(t) - h_2^m(t)\right) \sqrt{\left|h_1^m(t) - h_2^m(t)\right|} - C_{vo} \sqrt{h_2^m(t)} U_o^m(t) - Q_{f_2}(t) - A_2 \frac{dh_2^m}{dt} \qquad (13)$$

$$r_3(t) = U_p^m(t) - K_P\left(h_{1,ref} - h_1^m(t)\right) - K_I \int \left(h_{1,ref} - h_1^m(\tau)\right) d\tau \qquad (14)$$

$$r_4(t) = Q_p^m(t) - \begin{cases} U_p^m(t) & if\ 0 < U_p^m(t) < Q_{p,\max} \\ 0 & if\ U_p^m(t) \leq 0 \\ Q_{p,\max} & if\ U_p^m(t) \geq Q_{p,\max} \end{cases} \qquad (15)$$

Note that a discrete-time representation of (12)-(15) is obtained by applying an Euler discretization with sampling time $T_s = 1s$.

3. ATTACK SCENARIOS AND DETECTION/ISOLATION REQUIREMENTS

In this benchmark, a number of attacks are considered, covering different attack policies. This section presents the different kinds of attacks affecting the physical and cyber layers (see Table 2), as well as the requirements for their successful detection and isolation.

Table 2. Attack scenarios in the benchmark.

Scenario	*Physical layer*	*Cyber layer*
1	×	×
2	√	×
3	√	√
4	√	√
5	√	√
6	√	×
7	√	√
8	√	√
9	√	√
10	√	√

3.1 Attack scenarios

Scenario 1 - Attackless mode: This scenario corresponds to the normal behavior of the two-tank system when nobody is stealing water.

Scenario 2 - Short-term water theft from T_1: This scenario is similar to a leakage fault, the only remarkable difference being that it is cast maliciously, with the purpose of stealing water from the tank T_1. In this scenario, a pump extracts a constant flow $Q_{f1} = 10^{-4} m^3/s$ between $t = 40s$ and $t = 80s$ without any alteration of the measurements h_1^m and h_2^m. Note that in this scenario, the residuals behave similarly to the case of a sudden leak in the original fault diagnosis benchmark.

Scenario 3 - Short-term water theft from T_1 with hiding signal added to the measurement h_1^m: In this scenario, the thief uses a pump to extract water with a constant flow $Q_{f1} = 10^{-4} m^3/s$ between $t = 40s$ and $t = 80s$ while adding a signal to the output of the level sensor in tank T_1 so that the introduced signal *hides* the theft. Thanks to the introduced signal, the water level in tank T_1 seems to remain constant, and the PI controller works as if nothing had happened providing almost the same value U_p^m as in scenario 1. In particular, the modified value of h_1^m is given by

$$h_1^m(t) = h_1(t) + \varepsilon_{h1}(t) + \frac{1}{A_1} \int_0^t Q_{f1}(\tau) d\tau \qquad (16)$$

Scenario 4 - Long-term water theft from T_1 with hiding signal added to the measurement h_1^m: This attack scenario is similar to scenario 3, but the theft duration is extended from

$40\,s$ to $120\,s$. Due to the large quantity of stolen water, the plant exhibits some physical functioning problems, since the tank T_1 is emptied out, affecting the tank T_2 due to the interconnection, and the consumption of water Q_o, which becomes zero.

Scenario 5 - Long-term water theft from T_1 with small signal added to the measurement h_1^m: In this scenario, the thief will steal water as in the previous scenarios while adding a signal that deceives the PI controller to force more water to be pumped inside the system while making harder to detect the theft. In particular, the modified value of h_1^m is given by

$$h_1^m(t) = h_1(t) + \varepsilon_{h1}(t) + \frac{1}{A_1}\int_0^t 0.5 Q_{f1}(\tau)d\tau \qquad (17)$$

Scenario 6 - Short-term water theft from T_2: This attack scenario is similar to scenario 2, but it affects T_2 instead of T_1.

Scenario 7 - Short-term water theft from T_2 with hiding signal added to the measurement h_2^m: This attack scenario is similar to scenario 3, but it affects T_2 instead of T_1. In this case, the thief applies a constant signal $Q_{f2} = 10^{-4}\,m^3/s$ while adding a signal to the output of the level sensor in tank T_2, which forces the ON-OFF controller to act on the interconnecting valve V_b as if nothing had happened. In particular, the modified value of h_2^m is given by

$$h_2^m(t) = h_2(t) + \varepsilon_{h2}(t) + \frac{1}{A_2}\int_0^t Q_{f2}(\tau)d\tau \qquad (18)$$

Scenario 8 - Long-term water theft from T_2 with hiding signal added to the measurement h_2^m: This scenario is similar to scenario 4, but the pump corresponding to Q_{f2} is used by the thief instead of the one corresponding to Q_{f1}.

Scenario 9 - Long-term water theft from T_2 with small signal added to the measurement h_2^m: This scenario is similar to scenario 5, but the thief steals water from the tank T_2 and the introduced signal is meant to deceive the ON-OFF controller instead. In this case, the modified value of h_2^m is given by

$$h_2^m(t) = h_2(t) + \varepsilon_{h2}(t) + \frac{1}{A_2}\int_0^t 0.5 Q_{f2}(\tau)d\tau \qquad (19)$$

Scenario 10 - Replay attack: In this scenario, the thief steals water when the plant has reached its steady-state. However, before doing so, he/she records the measurements coming from the sensors without stealing water from the tanks. Then, in a subsequent phase of the attack, the thief steals water while replacing the real data with the recorded one. This type of attack is very hard to detect, if not impossible, and for this reason alternative approaches must be employed, see e.g. (Mo and Sinopoli, 2009, Zhu and Martínez, 2014). More specifically, the water is stolen from $t = 160\,s$ to $t = 200\,s$, while measurements recorded in the $50\,s$ previous to the attack are used to deceive the controller and the supervision system. At time $t = 200\,s$, the replay attack ends and the controller and the supervision system are able to see the real data coming from the system.

3.2 Detection and isolation requirements

The effectiveness of different cyber attack detection and isolation techniques can be assessed using the proposed benchmark by comparing different performance indices. Such a comparison would be performed by applying Monte Carlo studies with a sufficient high number of simulations, each one of which corresponding to a different realisation of the measurement noise, independent from the previous ones. In particular, the effectiveness of the proposed techniques would be tested by checking their ability to provide information both about the effect of the attack being performed (detection) and the exact nature and location of the attack (isolation). Typical performance indices are:

- **Attack detection time delay t_{AD}:** Time needed by the attack detection method to detect the presence of an attack, calculated from the time at which the attack begins;
- **Attack isolation time delay t_{AI}:** Time needed by the attack detection method to isolate whether the attack affects T_1 or T_2, calculated from the time at which the attack begins;
- **False negative attack detection rate r_{FNAD}:** Percentage of time during which the system is under attack, but the detection method determines an attackless situation;
- **False positive attack detection rate r_{FPAD}:** Percentage of time during which the system is not being attacked, but the detection method determines that the system is under attack;
- **Wrong isolation rate r_{WIR}:** Percentage of time during which the isolation method provides a wrong information about which tank is being affected by the attack, calculated with respect to the overall time during which an isolation information is provided;
- **Attack estimation accuracy for tank i $a_{AE,i}$:** This index takes into account the difference between the real stolen water volume V_{fi}, $i = 1,2$, and the estimated stolen water volume $\hat{V}_{fi}$, $i = 1,2$, and it is calculated as

$$a_{AE,i} = \frac{V_{fi} - \hat{V}_{fi}}{V_{fi}} \qquad (20)$$

All these indices should be kept positive and as low as possible, in order to allow for the activation of policies or strategies to protect the system from the attacks. In addition, another requirement is the robustness of the proposed methods towards uncertainties in the model (e.g. due to unknown changes in the values listed in Table 1).

4. SIMULATION RESULTS

This section provides the plots of some relevant variables that give more insight about the behavior of the benchmark in the proposed attack scenarios.

Fig. 2 shows a comparison between the water levels in the tanks in scenarios 1 and 3. It can be seen that due to the hiding signal introduced by the thief, h_1^m in scenario 3 (yellow dashed line) matches h_1 in scenario 1 (blue solid line), which misleads in determining whether someone is stealing water from the tank or not. It can be seen that the real value of h_1 in scenario 3 (red solid line) plummets to a much lower value. On the other hand, Fig. 3 shows the evolution of the residuals r_1 and r_2, which are calculated as in (12)-(13). Notably, r_1 and r_2 are useful for evaluating the presence of leakages in tanks T_1 and T_2, respectively. In fact, they behave as expected (taking a value different than zero from $t = 40\,s$ to $t = 80\,s$) in scenario 2 (red solid line) which, as stated in Section 3.1, is similar to a leakage fault. However, their values in scenario 3 (yellow dashed line) are approximately zero during the attack, which prevents traditional fault detection and isolation (FDI) algorithms from detecting correctly the presence of the attack.

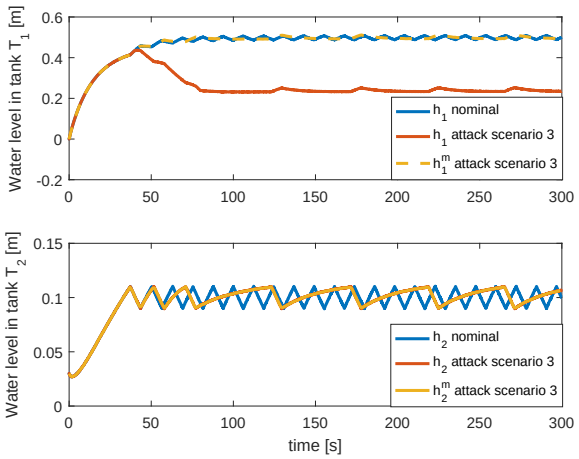

Fig. 2. Water level signals in scenarios 1 and 3.

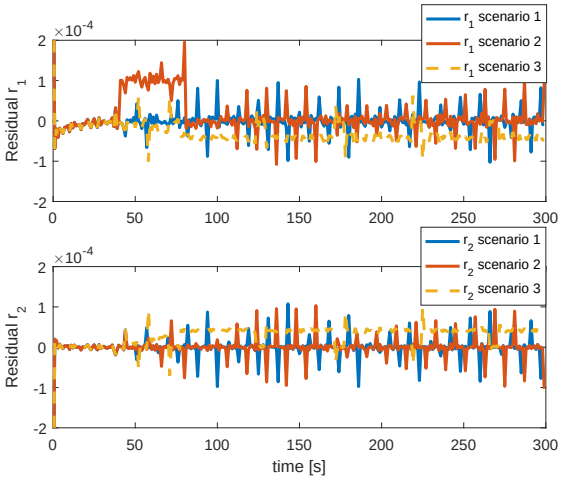

Fig. 3. Residuals in scenarios 1, 2 and 3.

The water level responses in scenario 5 are shown and compared to the ones in scenario 1 in Fig. 4. It can be seen that the signal h_1^m sent to the PI controler is similar, although not identical, to the real h_1 in scenario 1. This makes it harder to diagnose the water theft, as shown by the residuals depicted in Fig. 5.

Finally, Fig. 6 presents the water level responses in scenarios 1 and 10 (replay attack). This is the hardest attack to detect, since the measured water levels in both tanks h_1^m and h_2^m are almost identical to the water levels h_1 and h_2 in the attackless scenario.

As a consequence, the residuals in scenario 10 during the replay attack are undistinguishable from the ones when no attack is being performed (see Fig. 7). When the replay attack ends, one can see a spike in the residuals due to the discontinuity between the received data during and after the replay attack. It is evident that alternative and innovative approaches must be investigated in order to detect earlier this kind of attacks.

5. CONCLUSIONS

In this paper, a benchmark model for the detection of cyber attacks has been presented. The benchmark consists in a nonlinear system made up by two interconnected tanks. A malicious attacker, who has the goal of stealing water from the tanks while going unnoticed thanks to appropriate alterations

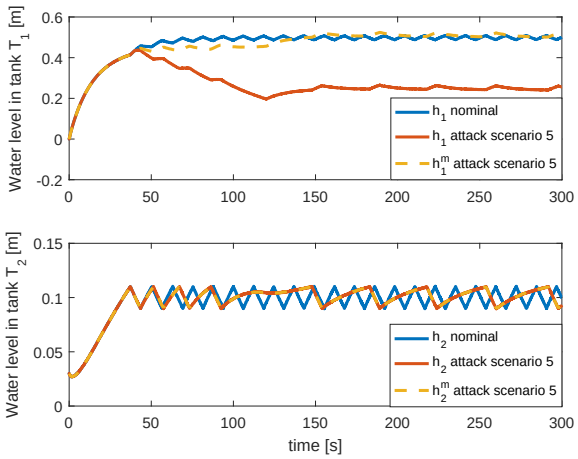

Fig. 4. Water level signals in scenarios 1 and 5.

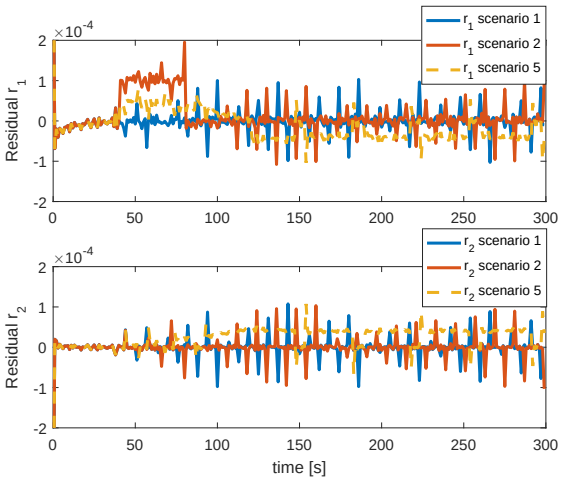

Fig. 5. Residuals in scenarios 1, 2 and 5.

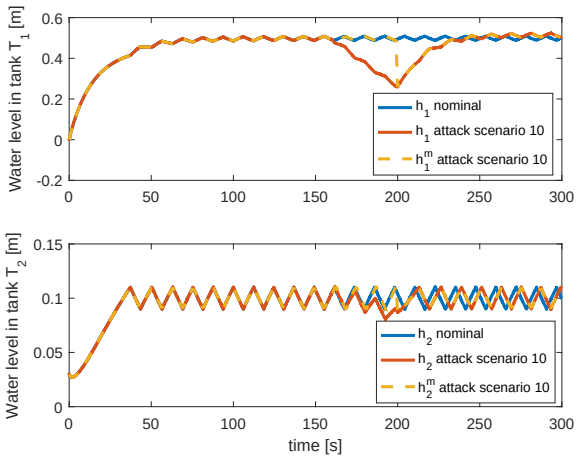

Fig. 6. Water level signals in scenarios 1 and 10.

of the outputs of the sensors, has been incorporated. Depending on the type of theft and the type of sensor alteration, different attack scenarios have been obtained, which simulate short-term and long-term water theft from both tanks. Simulation scenarios have been provided and discussed, highlighting some of the difficulties with detecting and isolating the proposed cyber attacks

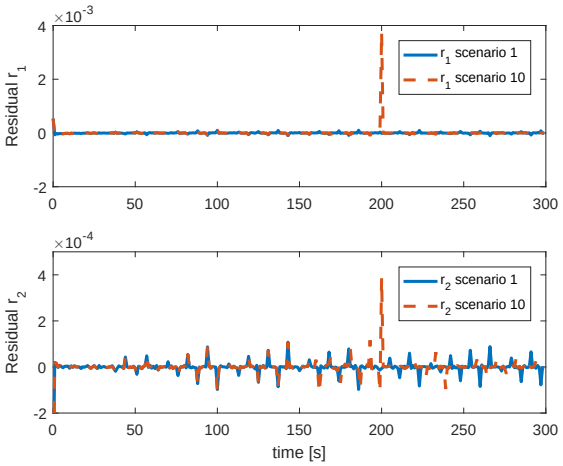

Fig. 7. Residuals in scenarios 1 and 10.

using traditional FDI techniques. The authors hope that the proposed benchmark will be useful to the scientific community for testing different kinds of cyber attack detection and isolation schemes. Furthermore, the authors encourage the design of secure controllers that improve the resilience of the system against these attacks.

REFERENCES

C. M. Ahmed, V. R. Palleti, and A. P. Mathur. WADI: a water distribution testbed for research in the design of secure cyber physical systems. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, pages 25–28. ACM, 2017.

L. Armesto, L. Arnal, J. Dols, V. Girbés, and J.C. Peris. Proyecto safebus: Sistemas avanzados de seguridad integral en autobuses. *Revista Iberoamericana de Automática e Informática industrial*, 13(1):103–114, 2016. ISSN 1697-7920.

H. Badihi, Y. Zhang, and H. Hong. Fuzzy gain-scheduled active fault-tolerant control of a wind turbine. *Journal of the Franklin Institute*, 351(7):3677–3706, 2014.

J. Blesa, D. Rotondo, V. Puig, and F. Nejjari. Fdi and ftc of wind turbines using the interval observer approach and virtual actuators/sensors. *Control Engineering Practice*, 24: 138–155, 2014.

J. Blesa, P. Jiménez, D. Rotondo, F. Nejjari, and V. Puig. An interval nlpv parity equations approach for fault detection and isolation of a wind farm. *IEEE Transactions on Industrial Electronics*, 62(6):3794–3805, 2015.

B. O. Bouamama, R M. Alaoui, P. Taillibert, and M. Staroswiecki. Diagnosis of a two-tank system. Technical report, Internal report of CHEM-project USTL Lille, France, 2001.

B. O. Bouamama, A. K. Samantaray, K. Medjaher, M. Staroswiecki, and G. Dauphin-Tanguy. Model builder using functional and bond graph tools for fdi design. *Control Engineering Practice*, 13(7):875–891, 2005.

Y.-L. Huang, A. A. Cárdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry. Understanding the physical and economic consequences of attacks on control systems. *International Journal of Critical Infrastructure Protection*, 2(3):73–83, 2009.

S. Jeschke, C. Brecher, T. Meisen, D. Özdemir, and T. Eschert. Industrial internet of things and cyber manufacturing systems. In *Industrial Internet of Things*, pages 3–19. Springer, 2017.

M. Krotofil and A. A. Cárdenas. Resilience of process control systems to cyber-physical attacks. In *Nordic Conference on Secure IT Systems*, pages 166–182. Springer, 2013.

Y. Mo and B. Sinopoli. Secure control against replay attacks. In *47th Annual Allerton Conference on Communication, Control, and Computing*, pages 911–918. IEEE, 2009.

P. F. Odgaard and K. E. Johnson. Wind turbine fault detection and fault tolerant control: an enhanced benchmark challenge. In *American Control Conference (ACC), 2013*, pages 4447–4452. IEEE, 2013.

P. F. Odgaard and J. Stoustrup. Results of a wind turbine fdi competition. *IFAC Proceedings Volumes*, 45(20):102–107, 2012.

P. F. Odgaard and J. Stoustrup. Fault tolerant wind farm control: A benchmark model. In *Control Applications (CCA), 2013 IEEE International Conference on*, pages 412–417. IEEE, 2013.

P. F. Odgaard and J. Stoustrup. A benchmark evaluation of fault tolerant wind turbine control concepts. *IEEE Transactions on Control Systems Technology*, 23(3):1221–1228, 2015.

P. F. Odgaard, J. Stoustrup, and M. Kinnaert. Fault-tolerant control of wind turbines: A benchmark model. *IEEE Transactions on Control Systems Technology*, 21(4):1168–1182, 2013.

F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

N. L. Ricker. Model predictive control of a continuous, nonlinear, two-phase reactor. *Journal of Process Control*, 3(2): 109–123, 1993.

H. Sanchez, T. Escobet, V. Puig, and P. F. Odgaard. Fault diagnosis of an advanced wind turbine benchmark using interval-based arrs and observers. *IEEE Transactions on Industrial Electronics*, 62(6):3783–3793, 2015.

I. R. Scola, G. Besançon, and D. Georges. Optimizing kalman optimal observer for state affine systems by input selection. *Automatica*, 93:224–230, 2018.

F. Shi and R. Patton. An active fault tolerant control approach to an offshore wind turbine model. *Renewable Energy*, 75: 788–798, 2015.

S. Simani and P. Castaldi. Data-driven and adaptive control applications to a wind turbine benchmark model. *Control Engineering Practice*, 21(12):1678–1693, 2013.

M. Staroswiecki and G. Comtet-Varga. Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37(5):687–699, 2001.

R. Waslo, T. Lewis, R. Hajj, and R. Carton. *Industry 4.0 and cybersecurity: Managing risk in an age of connected production*. Deloitte University Press, 2017.

X. Zhang. Structural analysis for diagnosis of a two-tank system. In *Pervasive Computing and Applications (ICPCA), 2010 5th International Conference on*, pages 273–276. IEEE, 2010.

M. Zhu and S. Martínez. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Transactions on Automatic Control*, 59(3):804–808, 2014.