Cyber-Physical-Security Framework for Building Energy Management System

Kaveh Paridari[†], Alie El-Din Mady[‡], Silvio La Porta[§], Rohan Chabukswar[‡]
Jacobo Blanco[§], André Teixeira[*], Henrik Sandberg[†], Menouer Boubekeur[‡]

***Abstract*— Energy management systems (EMS) are used to control energy usage in buildings and campuses, by employing technologies such as supervisory control and data acquisition (SCADA) and building management systems (BMS), in order to provide reliable energy supply and maximise user comfort while minimising energy usage. Historically, EMS systems were installed when potential security threats were only physical. Nowadays, EMS systems are connected to the building network and as a result directly to the outside world. This extends the attack surface to potential sophisticated cyber-attacks, which adversely impact EMS operation, resulting in service interruption and downstream financial implications. Currently, the security systems that detect attacks operate independently to those which deploy resiliency policies and use very basic methods. We propose a novel EMS cyber-physical-security framework that executes a resilient policy whenever an attack is detected using security analytics. In this framework, both the resilient policy and the security analytics are driven by EMS data, where the physical correlations between the data-points are identified to detect outliers and then the control loop is closed using an estimated value in place of the outlier. The framework has been tested using a reduced order model of a real EMS site.**

***Index Terms*—Cyber-physical-security, energy management system, resilient control, virtual sensor, security analytics.**

I. Introduction

Automatic control of electrical components in buildings has become a necessary task for any energy management system (EMS) in order to achieve optimal performance. The aim of a modern EMS is to enhance the functionality of interactive control strategies leading towards energy efficiency and a more user friendly environment. The EMS operates several building systems, such as the supervisory control and data acquisition (SCADA), which controls the smart-grid of one or more buildings, and the building management system (BMS), which controls the building heating demand, security system, fire alarm system, etc. Heating, ventilation, and air conditioning (HVAC) is considered to be the highest source of energy consumption in the building operation, and the systems most affecting user comfort. Typically, HVACs are controlled by both the SCADA system and BMS, where SCADA manages the electrical component operations (e.g., combined heating and power (CHP)) and the BMS manages the operations of thermal components (e.g., boilers). Therefore, cyber-attacks on EMS can lead to significant financial impact. By connection of the EMS to the building communication network, the possibility of EMS cyber-attack increases. The StuxNet cyber-attack supposedly targeting a nuclear-enrichment plant (by corrupting the measurements and actuator signals) in Iran [1], and BlackEnergy malware targeting several electricity distribution companies in Ukraine [2], are concrete examples of cyber-attacks. Thus, it is crucial to make the control of EMS to be resilient against cyber crime. The existing methods for EMS cyber-security are mainly based on running tests and benchmarks to evaluate the possible cyber-attacks and their impact [3]. These methods require expert knowledge to manually perform the tests and attack assessment. There is currently no end-to-end methodology that covers the main steps in EMS cyber-security design flow. The continued rise of complexity of attacks, skills of the attackers, and failing of the traditional security applications (antivirus, Intrusion Prevention/Detection Systems, etc.) against those new type of attacks, necessitate the development of new defense systems. Targeted aggressive attacks use well-researched and well-funded multi-vector tactics to introduce stealthy and persistent malware in control infrastructure systems. At the same time, the risks related to compromise of control infrastructure are growing dramatically. The integration of old systems with new ones, and connection of the traditional SCADA systems to the Internet enlarge the attack surfaces of the systems. Furthermore, new vulnerabilities are discovered daily, which may have already been exploited by adversaries for some time. Recently there has been an increase in control systems security research [4], [5], [6]. Those work take in consideration the activity of an intelligent adversary that would increase for example the operation cost of the system [4] and the limitations of the attack detection and identification methods using linear systems in the power networks [6]. The time during which vulnerabilities remain hidden, and the time required to patch them together, leave a window large enough for adversarial system penetration.

[†]K. Paridari and H. Sandberg are with the ACCESS Linnaeus Center and the Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. Emails: paridari@kth.se, hsan@kth.se.

[‡]A. E. Mady, R. Chabukswar and M. Boubekeur are with the United Technologies Research Center, Cork, Ireland. Emails: madyaa@utrc.utc.com, chabukr@utrc.utc.com, boubekm@utrc.utc.com.

[§]S. La Porta and J. Blanco are with the EMC Research Europe, Cork, Ireland. Emails: silvio.laporta@emc.com, jacobo.blanco@emc.com.

[*]A. Teixeira is with the Delft University of Technology, Delft, the Netherlands. Email: andre.teixeira@tudelft.nl.

This work has received funding in part from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 608224, the Swedish Research Council under Grant 2013-5523, and the Swedish Civil Contingencies Agency.

978-1-5090-1772-0/16/$31.00 ©2016 IEEE

These factors highlight the importance of detecting attacks as soon as possible in order to minimise damage and impact. Analytics and response capabilities enable quick detection of cyber-attacks by checking the system behaviour at application level and responding quickly to minimise their impact. As studied in [7], the operational model must go beyond the conventional focus on distribution and generation infrastructure for fault isolation, remediation and recovery, and focus on information and a new understanding of data analysis. In addition, as discussed in [8], it requires the ability to handle processing of huge amounts of data, by using new analytics and visualization techniques. Then, one can integrate the results of that analysis with governance processes that make those results readily actionable.

Once the attack is detected, control policies which are resilient against the attacks, should be triggered. Design of control and estimation algorithms that are resilient against faults is not a new problem, but those algorithms may not be efficient against malicious cyber-attacks. For example, *virtual sensor* (VS) and *virtual actuator* concepts, have been introduced in [9] to deal with sensor and actuator failures, respectively. Attacks may be more complex than faults, and may use some information of the system to corrupt the measurements in an intelligent way, and result in worse consequences than faults. Thus, there has been a recent increase in control systems security research and design of resilient control and estimation algorithms against attacks [10], [4], [11], [12], [13], [14], [15]. In [10], the authors consider the problem of control and estimation in a networked system when the communication links are subject to disturbances (corresponding to packet losses), resulting from a DoS attack for instance. In [4], the authors consider a more intelligent jammer who plans his attacks in order to maximize a certain cost, while the objective of the controller is to minimize this same cost. The results in that study are however derived in the case of one-dimensional systems, which is the main limitation of the work. The problem of reaching consensus in the presence of malicious agents is studied in [11]. The authors characterize the number of infected nodes that can be tolerated and propose a way to overcome the effect of the malicious agents when possible. One particularity of that works is that the dynamics is part of the algorithm and can be specifically designed, rather than being given as in a physical system. The estimation and control of linear systems, when some of the sensors or actuators are corrupted by an attacker, is studied in [12]. In that work, they propose an efficient algorithm inspired from techniques in compressed sensing to estimate the state of the plant despite attacks. In that paper, the authors assume that the attacked nodes does not change over time. In addition, a general framework to model and analyse impact of attacks, is proposed in [5]. In [13], a method for state estimation in presence of attacks, for systems with noise and modeling errors is proposed. In that work, it is shown that the attacker cannot destabilize the system by exploiting the difference between the model used for state estimation and the real physical dynamics of the system. In [14], a control technique is proposed which is resilient against certain sensor attacks. In that technique, a recursive filtering algorithm, to estimate the states of the system, is implemented that takes advantage of redundancy in the information received by the controller. The main contributions of this paper are:

- A practical cyber-secure framework for EMS is proposed, which uses building physics to drive the EMS cyber-security design flow. This framework, includes security information analytics to detect attacks and resilient policy to keep the system running under the attack.
- The framework efficiency is demonstrated on a real critical attack scenario. Through simulations, it is shown that the proposed resilient control policy can recover the system from abnormal conditions (when the system has been attacked), even when there exist delay for the attack detections.

The paper is organized as follows. Section II describes a test-bed, which has been used to evaluate the feasibility of the proposed cyber-physical-security framework. The framework is proposed in Section III, which executes a reliable resilient policy whenever an attack is detected by using security information analytics. Section IV presents simulation results (based on the real data from the test-bed) and discusses performance of the proposed framework in terms of capability of attack detection and resiliency against the attacks. Final remarks and conclusions are drawn in Section V.

II. Application Domain

An EMS optimally controls all energy sources in a building in order to minimise thermal and electrical energy consumption, while maximising user comfort. Typically, an EMS employs SCADA and BMS in order to control electrical and thermal loads, respectively. Recently, smart-grid infrastructure [16] has been introduced to support both types of loads, where some equipment such as Combined Heat and Power (CHP) can be an energy source for both electrical and thermal demand. In this context, an EMS would consider an HVAC system an important contributor to energy consumption, making it a target for attacks with financial impact. In addition, attacking the EMS in a smart-grid can lead to safety risk [17] due to damage to water transport system or to the heating sources (e.g. CHP and boilers).

As a proof of concept for our framework, an EMS which controls a small size smart-grid covering several buildings at the demo-site at Cork Institute of Technology (CIT) [18] is considered. In the following sections, the main components used by the EMS to control HVAC system will be highlighted. The modelling techniques used to capture the system dynamics will also be discussed briefly. Figure 1 shows the HVAC system at the CIT demo-site, where the EMS controls two main heating sources, the boiler and the CHP, which heat up the water to a temperature set-point. This flow temperature set-point is identified using a weather compensation method [19]; an on/off controller is then used to keep the water in the heating sources at the set-point.

Considering the return temperature as an indicator of the required thermal demand in the building, then it is used to determine if the boiler and/or CHP are in operation. The header is used to deliver the heated water to each floor, where a mixing valve is used to regulate each floor flow temperature, to respect a floor flow set-point, again identified by a weather compensation method. A Proportional-Integral (PI) control algorithm is used to regulate the position of the mixing valve. At the end, supplied water to each of the floors is distributed over several radiators in each building, where radiator is controlled using an on/off controller to reach a predefined room temperature set-point.

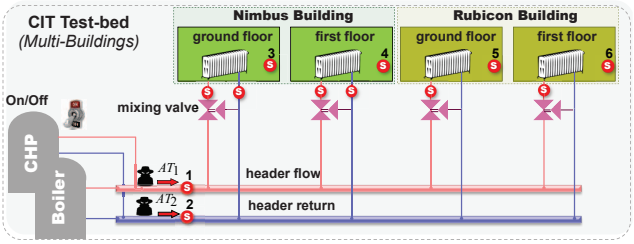

Fig. 1. Typical BMS for HVAC system

In order to evaluate the feasibility of the proposed framework, a Simulink model was developed to capture the CIT demo-site dynamics. The model was developed using *Gray Box* modelling [20], where the model structure is created based on the thermodynamic theory of each component and the model parameters are tuned using real-world data from the CIT demo-site. The model has been validated using data trend analysis against the real-world data at the CIT demo-site. According to Modeling, Analysis, Simulation & Computation (MASC) Readiness Level (MRL), model calibration against output curves trends is MRL > 4, which is used for concept and detailed design.

III. EMS Cyber-Security Framework

In this section, a cyber-physical-security framework is proposed that executes a reliable resilient policy whenever an attack is detected using security information analytics (SIA). In this framework, both the resilient policy and the SIA are designed using physical correlations between the data-points. Identifying correlation between data-points requires an expert knowledge input, which is considered to be costly in the building automation domain. To mitigate this drawback, a risk analysis step is employed to identify the sensors with the maximum impact. A risk assessment stage will be added to this framework in future work. In the following subsections, the theoretical background of SIA and resilient policy will be highlighted, with an initial risk assessment for some critical sensors at the demo site.

A. Risk Assessment

After examining the financial and safety impact of several attack scenarios on the HVAC at the demo site, the header and return temperature sensors were down-selected to be the most critical. The financial impact is linearly interpolated based on historical energy consumption and its associated cost for the demo site. As shown in Figure 2, applying a negative offset of $10°$C to the header flow temperature sensor can lead to high financial risk (10% energy degradation, about 5000 EUR per year for the demo-site). In addition, this attack scenario can force the boiler or the CHP to increase the water temperature to more than $80°$C, which leads to water circuit damaging. Another attack scenario consists of applying a positive offset of $10°$C to the return temperature sensor leads to high safety risk due to damaging of CHP.

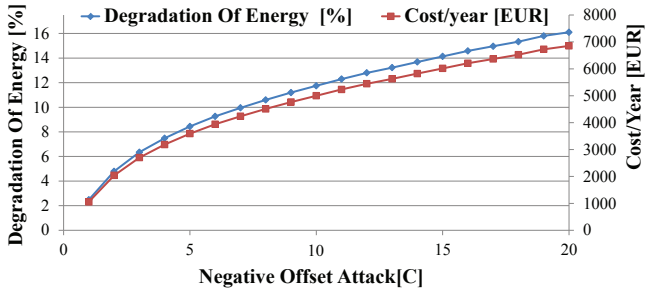

Fig. 2. Risk analysis for a financial impact of attacking a flow temperature sensor in an HVAC system

B. Resilient Control

A schematic of the proposed hierarchical control strategy for the HVAC system is illustrated in Figure 3. In this architecture, the HVAC system is represented by interconnected plants $P_i, i \in \Phi$, where $\Phi = \{1, ..., N\}$ is the set of all the plants' indices. These plants are controlled by the local controllers $K_i, i \in \Phi$, which are implemented in the BMS and signals (control and measurement signals) are sent over a communication network. In this networked control system, the controller K_i sends the control signal u_i to the plant P_i, and the plant sends the sensor measurement y_i to the controller K_i. The received control signal by the plant is called $\tilde{u}_i$, and the received sensor measurement signal by the controller is called $\tilde{y}_i$. In this scheme, the local controller K_i, receive the reference signal r_i from the control center, and the measurement signal $\tilde{y}_i$, to calculate and send the control signal u_i. Note that the two signals u_i and y_i could be modified by the *attacker*s and when they pass through the communication local network. Here, we consider attacker as a man in the middle, who can secretly alters the communication between the plants and controllers, and corrupt the signals. Thus, there are two different conditions that the controlled HVAC system works in: *normal* and *abnormal* conditions.

Normal condition: Under this condition, the BMS is healthy and there is no anomaly in the signals being sent or received by the plants and controllers ($\tilde{y}_i = y_i$ and $\tilde{u}_i = u_i, i \in \Phi$).

Abnormal condition: In this situation, the BMS is being attacked, and the attackers have the ability to alter the cyber-physical dynamics of the system through exogenous inputs.

To model the attacks here, we assume m number of the sensor measurements, y_j for $j \in \Gamma$ are under attack [13],

[15]. Here, $\Gamma \subset \Phi$ is the indices set of corrupted measurements, and the cardinality of Γ is considered to be $\mathbf{card}(\Gamma) = m$. We also define $\Gamma^C = \Phi \setminus \Gamma$ as the indices set of healthy measurements, where $\mathbf{card}(\Gamma^C) = N - m = h$.

An attack example (the attack AT) is shown in the Figure 3, in which the offset Δy_i is added to the measurement signal y_i at time k'. The time k', is called attack start time here, and is detected by the SIA, as is described in Section III-C. By having attacks to the system at time k', we have $\tilde{y}_i \neq y_i, i \in \Gamma$ for $k > k'$.

Remark 1 *Here we assume that there is no attack on the control signals, and they are received by the controllers without any changes ($\tilde{u}_i = u_i, i \in \Phi$). The setup can be easily extended to include also the attacks on the control signals (see [13] for a related study).*

As discussed in [21], attacks to the measurement signals may lead to system instability. Thus, a control strategy to increase the resiliency of the controlled system is proposed here. In this strategy, a data fusion filter (*Virtual Sensor* (VS)) and SIA, are implemented in the supervisory level controller (Control center). In this framework, VS estimates the output signals ($\hat{y}_i, i \in \Phi$) based on all the available healthy measurements, at all the times. Since the VS is running in the supervisory level, it has access to system-wide measurements ($y_i, i \in \Phi$) and can estimate the measurement signals, based on the available model of the system. After the time k', and detecting the attack by SIA, the correction signal $\hat{y}_i - (y_i + \Delta y_i)$ is being sent into the BMS, only for $i \in \Gamma$, to be added to the corrupted signal $y_i + \Delta y_i$. In this manner, the corrupted signals are being replaced by the estimated output signals $\tilde{y}_i = \hat{y}_i, i \in \Gamma$ and for the healthy signals we have $\tilde{y}_j = y_j, j \in \Gamma^C$.

The estimated output signals sent by the Control center would not be of the same quality, and may be time-delayed compared to measurements of the system under normal condition. However, the estimated data is more useful than the attacked one, and so would contribute to a more resilient control system. The other advantage of this approach is that it does not require many changes in the lower-level designs. A similar idea has been proposed in [22], in which a predictive outage compensator is designed to generate control signals when there is a communication outage and the actuator in the system does not receive the control signal.

1) Modeling and the Optimal Virtual Sensor: In the following, a mathematical formulation of the problem, under normal and abnormal conditions, is given. The plant P_i is given by

$$\begin{aligned} x_i(k+1) =& A_i x_i(k) + B_i \tilde{u}_i(k) + D_i d_i(k) \\ &+ \sum_{j \neq i} J_{ij} C_j x_j(k) + w_i(k), \\ y_i(k) =& C_i x_i(k) + \nu_i(k). \end{aligned} \tag{1}$$

where, $x_i \in \mathbb{R}^{n_i}$ is the local state vector of the plant, $\tilde{u}_i \in \mathbb{R}^{s_i}$ is the received control signal vector, and $y_i \in \mathbb{R}^{p_i}$ is the local measurement vector. In (1), d_i is a deterministic

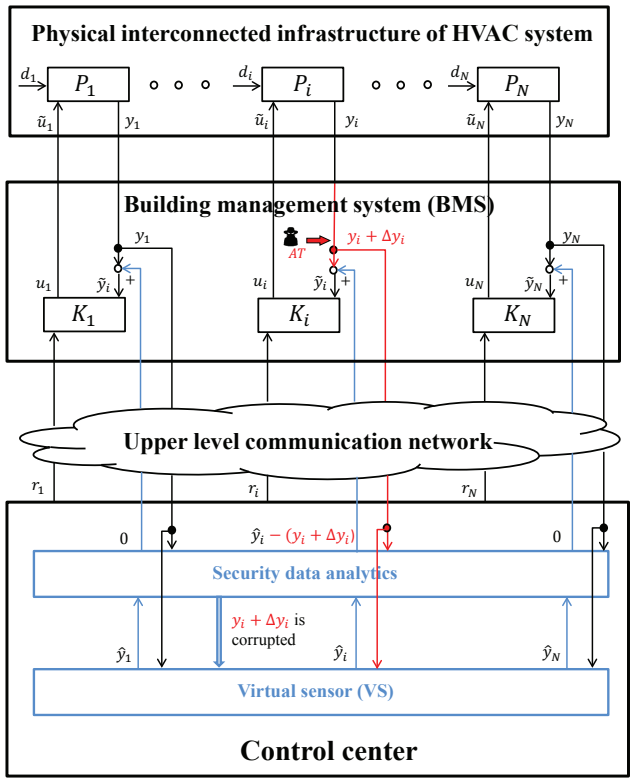

Fig. 3. Schematic of a hierarchical control system with a communication network, which is resilient against the adversarial actions on the measurements.

disturbance vector, and w_i and ν_i are vectors of process and measurement white noise of the plant, respectively. Here, the matrices A_i, B_i, C_i, D_i and J_{ij} have dimensions conformably with the vectors, and the matrix J_{ij} captures the interaction between the plants P_i and P_j. The local control signal $u_i(k)$ is given by the controller K_i,

$$\begin{aligned} z_i(k+1) &= E_i z_i(k) + F_i \left(\tilde{y}_i(k) - r_i(k)\right), \\ u_i(k) &= H_i z_i(k), \end{aligned} \tag{2}$$

where, $z_i \in \mathbb{R}^{q_i}$ is the local state vector of the controller, $u_i \in \mathbb{R}^{s_i}$ is the local control signal vector calculated by the controller, $\tilde{y}_i \in \mathbb{R}^{p_i}$ is the received measurement vector, and $r_i \in \mathbb{R}^{p_i}$ is the reference signal sent by the control center. Here, the matrices E_i, H_i and F_i have dimensions conformably with the vectors. Thus, the closed-loop interconnected system evolves as

$$\begin{aligned} \begin{bmatrix} x(k+1) \\ z(k+1) \end{bmatrix} &= \underbrace{\begin{bmatrix} [A] + J[C] & [B][H] \\ [F][C] & [E] \end{bmatrix}}_{A_{cl}} \begin{bmatrix} x(k) \\ z(k) \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} [D] & 0 \\ 0 & [-F] \end{bmatrix}}_{B_{cl}} \begin{bmatrix} d(k) \\ r(k) \end{bmatrix} + \underbrace{\begin{bmatrix} I & 0 \\ 0 & [F] \end{bmatrix}}_{M} \begin{bmatrix} w(k) \\ \nu(k) \end{bmatrix} \\ y(k) &= \underbrace{\begin{bmatrix} [C] & 0 \end{bmatrix}}_{C_{cl}} \begin{bmatrix} x(k) \\ z(k) \end{bmatrix} + \nu(k), \end{aligned} \tag{3}$$

where, $[A]$ represents a block-diagonal matrix with A_i as the i-th diagonal block, and J is a matrix with J_{ij} as the ij-th block (diagonal blocks are zero). Note that $x = [x_1^\top ... x_N^\top]^\top$, $z = [z_1^\top ... z_N^\top]^\top$, $y = [y_1^\top ... y_N^\top]^\top$, $d = [d_1^\top ... d_N^\top]^\top$, $r = [r_1^\top ... r_N^\top]^\top$, $w = [w_1^\top ... w_N^\top]^\top$ and $\nu = [\nu_1^\top ... \nu_N^\top]^\top$. Let the expectations to be given as

$$\mathbf{E}\begin{bmatrix} w(k) \\ \nu(k) \end{bmatrix} = 0,\ \mathbf{E}\begin{bmatrix} w(k) \\ \nu(k) \end{bmatrix}\begin{bmatrix} w(l) \\ \nu(l) \end{bmatrix}^\top = \underbrace{\begin{bmatrix} R_1 & R_{12} \\ R_{21} & R_2 \end{bmatrix}}_{R}\delta_{kl},$$

where, R_1 and R_2 are the covariance of w and ν, respectively, and $R_{12} = R_{21}^\top$ is the constant cross covariance between w and ν. Under the normal and abnormal conditions, the VS generates an estimate of the outputs $(C_i x_i(k), i \in \Phi)$ of the controlled HVAC system, differently.

Assumption 1 *We assume here that, because of communication and computation delays, at the given time k, only the measurements till time $k-1$ are available.*

Estimates under normal condition: Under this condition, the optimal estimator for the states in (3), is the Kalman filter (see [23]).

$$\begin{aligned}\begin{bmatrix} \hat{x}(k+1|k) \\ \hat{z}(k+1|k) \end{bmatrix} &= A_{cl}\begin{bmatrix} \hat{x}(k|k-1) \\ \hat{z}(k|k-1) \end{bmatrix} + B_{cl}\begin{bmatrix} d(k) \\ r(k) \end{bmatrix} \\ &+ K_{cl}\underbrace{\left[y(k) - C_{cl}\begin{bmatrix} \hat{x}(k|k-1) \\ \hat{z}(k|k-1) \end{bmatrix}\right]}_{\epsilon(k)},\end{aligned} \tag{4}$$

where,

$$\begin{aligned}K_{cl} &= \left(A_{cl}PC_{cl}^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right)\left(C_{cl}PC_{cl}^\top + R_2\right)^{-1}, \\ P &= A_{cl}PA_{cl}^\top + MRM^\top - \left(A_{cl}PC_{cl}^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right) \\ &\times \left(C_{cl}PC_{cl}^\top + R_2\right)^{-1}\left(A_{cl}PC_{cl}^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right)^\top.\end{aligned} \tag{5}$$

Here, $\epsilon(k)$ is the optimal one-step ahead prediction error of the VS, K_{cl} is the stationary Kalman gain, and P is the stationary error covariance matrix. Note that by $\hat{x}(k|k-1)$ we mean an estimation of $x(k)$, given all the measurements $y_i(k)$ up until time $k-1$, and the optimal one step ahead prediction of $y(k)$ is $\hat{y}(k) = C_{cl}\hat{x}(k|k-1)$.

Assumption 2 *There exists redundancy in the information received by the VS, and the system remains observable by using $y_i, i \in \Gamma^C$ for the estimation.*

Estimates under abnormal condition: Under this condition, as it was mentioned before, some of the measurement signals in BMS are corrupted ($\tilde{y}_i \neq y_i, i \in \Gamma$) after the time k' and are detected by SIA. Thus, SIA informs the VS that the measurement being sent by sensor i is corrupted (see Figure 3) and should not be used for updating and predicting the states of the system.

In this case, the Kalman filter is no longer stationary and should use a time-varying gain. Note that the filter uses the healthy measurements $y_j, j \in \Gamma^C$ for the estimation. Thus, the estimator takes the prediction step for the state of the system as

$$\begin{aligned}\begin{bmatrix} \hat{x}(k+1|k) \\ \hat{z}(k+1|k) \end{bmatrix} &= A_{cl}\begin{bmatrix} \hat{x}(k|k-1) \\ \hat{z}(k|k-1) \end{bmatrix} + B_{cl}\begin{bmatrix} d(k) \\ r(k) \end{bmatrix} \\ &+ K'_{cl}(k)\left[y'(k) - C'_{cl}\begin{bmatrix} \hat{x}(k|k-1) \\ \hat{z}(k|k-1) \end{bmatrix}\right],\end{aligned} \tag{6}$$

where, $y'(k)$ a vector of healthy measurements $y_i(k), i \in \Gamma^C$, and the time-varying Kalman gain is given by

$$\begin{aligned}K'_{cl}(k) =& \left(A_{cl}P'(k)C_{cl}'^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right) \\ &\times \left(C'_{cl}P'(k)C_{cl}'^\top + R_2\right)^{-1}, \\ P'(k) =& A_{cl}P'(k-1)A_{cl}^\top + MRM^\top - \\ &\left(A_{cl}P'(k-1)C_{cl}'^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right) \\ &\times \left(C'_{cl}P'(k-1)C_{cl}'^\top + R_2\right)^{-1} \\ &\times \left(A_{cl}P'(k-1)C_{cl}'^\top + M\begin{bmatrix} R_{12} \\ R_2 \end{bmatrix}\right)^\top.\end{aligned} \tag{7}$$

Here, C'_{cl} is constructed from the matrix C_{cl} by removing the rows related to $C_i, i \in \Gamma$ in that, and $P'(k)$ is the time-varying error covariance matrix. Note that immediately after k', $P'(k)$ in (7) would be initialized by $P_{k'} = P$ in which P is given by (5), and will update in the next time steps.

Finally, the VS prediction is $\hat{y}(k) = [C]\hat{x}(k|k-1)$, and it feeds the estimation $\hat{y}(k)$ into the BMS to replace the corrupted measurement signals, which means $\tilde{y}_i(k) = \hat{y}_i(k)$, $i \in \Gamma$.

Remark 2 *Considering Assumption 1, the output estimate $\hat{y}(k) = [C]\hat{x}(k|k-1)$ is calculated by the VS, based on the state prediction $\hat{x}(k|k-1)$. Note that, in the cases where the delay is less than one sampling period, the VS can do a better estimation using $\hat{y}(k) = [C]\hat{x}(k|k)$, based on the updated state estimate $\hat{x}(k|k)$.*

2) Implementation Using System Identification: The proposed resilient policy does not need to estimate the lower-level states, since only the estimated outputs are used. Therefore, in practice, one can identify a model that is able to explain the covariance of the outputs, in cases where a model like (3) is not available. To identify a linear model of the controlled HVAC system, which is described in Section II, a subspace identification followed by a prediction error method [24] is applied. In this identification, the external temperature (which is the disturbance d to the system) is considered as the input, and the temperature of header flow, header return, Nimbus building ground floor, Nimbus building first floor, Rubicon building ground floor and Rubicon building first floor are considered as the outputs, respectively ($N = 6$). The system modeling in this manner results in a simple linear third order system, in the innovation form [25]:

$$\begin{aligned}x_s(k+1) &= A_s x_s(k) + B_s u_s(k) + K_s\epsilon(k) \\ y_s(k) &= C_s x_s(k) + D_s u_s(k) + \epsilon(k),\end{aligned} \tag{8}$$

where, x_s, u_s and y_s are the system state, input and output, respectively. The matrices A_s, B_s, C_s and D_s are the system matrices with appropriate dimensions, and the innovation $\epsilon(k)$ is white noise and independent of past input and output data [25].

Comparing the represented systems in (3)-(4) and (8), the outputs y and y_s should be close, assuming a successful system identification. Since the identified model in (8) is linear also, we can reuse the expressions (4)-(7) to construct $\hat{y}$, having the matrices A_{cl}, B_{cl}, C_{cl}, D_{cl} and K_cl to be substituted by the matrices A_s, B_s, C_s, D_s and K_s, respectively. Thus, the VS in the supervisory level, by having access to system-wide measurements $(y_1, ..., y_6)$, can estimate the measurement signals based on the available model of the system in (8).

C. Security Information Analytics

Intelligence-driven security systems understand what good behaviour is within an IT environment by monitoring and learning a variety of machine and human activities. Analytics solutions often rely on logs and configuration information as data sources. Similarly, capabilities such as network packet-capture are important in establishing normal behaviour in IT infrastructures. These techniques help organizations to learn what is typical within an IT environment so that future deviations from the norm (which often indicate problems), can be identified and investigated. Analysis systems capture and analyse terabytes of rapidly evolving real-time data from multiple sources, by using different methods of detection. For example, data can be captured and analysed for potential security issues as it traverses the network. This analysis, identifies suspicious activities of the attackers (which are done by the tools, services, communications and techniques), that do not depend on logs, events, or signatures from other security systems. Processing of these information flows happens as they occur. It means that the suspicious activities are spotted while there is still time for security teams to stop the attacks in progress. To do this, the SIA system works at the application-level, and it will check the entire system and each component's behaviours, and provide analysts with an overview of the security status of the control system. For example, measurement data is examined to detect potential anomalous behaviour. Then, the analytics system can operate on live data streaming from the system or used offline for investigation. The SIA system feeds the resilient control system with information indicating which measurements are corrupted. SIA consists of a set of outlier detection algorithms and a web application that allows analysts to examine the results. It uses two detection methods: the static outlier detection, which relies on preconceived rules to detect outliers; and the machine-learning (ML) outlier detection, which rely solely on the data to determine normal behaviour and classify outlier behaviour. There are two static detectors: a threshold-based outlier detector, which enforces the expected operational limits on single variables measured by the sensors; and a rule-based detector, which examines the behaviour of the system based on physical laws. Finally, a single dynamic machine-learning outlier detector which uses a ML algorithm to learn the normal behaviour of the system and finds anomalies in new data is used. The results from these three algorithms can be combined, in particular the ML component and the rule-based detector. Some work remains to be done in this area to determine the best approach, however there are several possibilities. Firstly, the overlap in results can be used to reduce false positives and determine which measurement are most important. Depending on the level of confidence in each detector, an anomaly score could be calculated for each timestamp. This would provide a ranking of incident severity for analysts to act upon.

1) Machine-learning Outlier Detection: This outlier detector uses a one-class support vector machine (OC-SVM) that learns the normal behaviour of the system to look for anomalous data. Many ML problems involve the labelling of measurements into groups. These algorithms can be divided into supervised and unsupervised, depending on whether they work on labelled data. In many anomaly detection problems, the data available is highly unbalanced, meaning that it contains mostly data associated with one class, usually normal behaviour. This occurs when obtaining examples of anomalous data is either time-consuming, infeasible, or expensive. To treat such scenarios, standard ML algorithms have been adopted to work with a single class. A OC-SVM [26], [27] is used in this case. This algorithm tries to create a hyperplane that maximally separates the data from areas of phase-space which are not populated.

This also has the advantage of not assuming any structure for the anomalous behaviour. By avoiding searching for specific attacks the approach remains capable of detecting previously unobserved attack patterns.

Using the historical training data, a model is learned for each sensor in question. These models are then used to predict outliers against new unseen data. In this case, the algorithm produces a simple binary decision and no outlier score is assigned.

2) Threshold-based Outlier Detection: In this detection method, a threshold-based detector compares the measured values against expected operational limits. A measurement is deemed anomalous if this threshold is violated.

3) Rule-Based Detection: A rule-based detector verifies that the measured sensor values obey the physical laws and stay within the statistical boundaries. These boundaries can be violated due to several reasons such as asynchronous measurements, system and sensor noise, quantisation, and etc. These sources of statistical errors can be estimated beforehand to generate a historical baseline, after which any anomaly out of this baseline can be deemed malicious.

Static rules: These rules contain only the current measurements from the sensors. While static rules are the simplest to be implemented, they only apply to the systems or subsystems that exhibit no dynamics.

Dynamic rules: These rules take into account system dynamics to estimate the current measurements. Implementing dynamic rules requires estimating and keeping track of hidden system states, but it can be applied to systems or

subsystems that exhibit linear time-invariant dynamics. Once the residue $\epsilon(k)$ is calculated here, the further analysis and detection is similar to that of the static rules above.

For example, in the case of the CHP and boiler systems shown in Figure 1, the system measures the header flow temperature (y_1), and the header return temperature (y_2). The system dynamics are based on the number of boilers (N_b), which are currently operational and we have $N_b(t) = 1$ or $N_b(t) = 2$. The boilers' water temperature (T_i), are the hidden states here. Applying conservation of energy to this system, discrete-time equations for the temperature of the water in the boilers can be deduced:

$$\begin{aligned} T_i(k+1) = \frac{Z_i(k) Q_b}{C_{P_W} M_i} + \frac{\Delta M_i}{N_b(k) M_i} \\ \times (y_2(k) - \nu_2(k) - T_i(k)) + w_i(k), \end{aligned} \tag{9}$$

where, $w(t)$ and $\nu(t)$ are process and measurement white noise, respectively. Here, Z_i indicates the operational state of each boiler ($Z_i = 1$ if the boiler is on, and $Z_i = 0$ if it is off). Q_b is the energy used by the boiler to heat up the water, ΔM_i is the mass flow into the boilers during each time slot, M_i is the mass capacity of each boiler, and C_{P_W} is the heat capacity of the water (all are assumed constant). The temperature of the header flow is given by

$$y_1 = \frac{1}{N_b(k)} \sum_i Z_i(k) T_i(k). \tag{10}$$

These equations are not linear in the current form. However, they are linear in each operating condition of number of operating boilers (N_b). For example, if the system has 2 boilers (as is the case in the test bed), and one of them is always operational, there are only two operating conditions: $N_b = 1$ and $N_b = 2$. If the operating condition is known, these equations can be written as a system using the boiler temperatures as the hidden states:

$$\begin{aligned} x_b(k+1) = A_b x_b(k) + B_b u_b(k) + w(k), \\ y_b(k) = C_b x_b(k) + D_b u_b(k) + \nu(k), \end{aligned} \tag{11}$$

in which

$$\begin{aligned} x_b = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, \; u_b = \begin{bmatrix} Z_1 Q_b \\ Z_2 Q_b \\ y_2 - \nu_2 \end{bmatrix}, \; y_b = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \\ A_b = \begin{bmatrix} 1 - \frac{\Delta M_1}{N_b M_1} & 0 \\ 0 & 1 - \frac{(N_b-1)\Delta M_2}{N_b M_2} \end{bmatrix}, \\ B_b = \begin{bmatrix} \frac{1}{C_{P_W} M_1} & 0 & \frac{\Delta M_1}{N_b M_1} \\ 0 & \frac{1}{C_{P_W} M_2} & \frac{\Delta M_2}{N_b M_2} \end{bmatrix}, \\ C_b = \begin{bmatrix} \frac{1}{N_b} & \frac{N_b - 1}{N_b} \\ 0 & 0 \end{bmatrix}, \; D_b = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{12}$$

Given the system above, a Kalman filter with gain K can be used to estimate the hidden states and the next measurements.

Remark 3 *In contrast to the Kalman filter in Section III-B.1, which makes a system-wide state estimate ($y_1, ..., y_6$), the filter here makes the estimate based on a subset of measurements (y_1 and y_2).*

IV. Results

In this section, the proposed framework is evaluated through simulation results (based on the data from the CIT test-bed model).

A. Security Information Analytics

In the simulations and for the scenarios described in Section III-A we focus on the rule-based detection method due the simulated dataset does not contains threshold outliers and the amount is not enough to properly train the Machine Learning component. As it is shown in Figure 1, the attacks are considered to be on the header flow and header return temperature measurements (the attacks AT_1 and AT_2 on the signals S_1 and S_2, respectively). Thus, three data-sets are examined with the following characteristics:

- The header flow temperature is manipulated (AT_1)
- The header return temperature is manipulated (AT_2)
- Both header flow and return temperatures are manipulated (AT_1 and AT_2)

The aim of the simulation is to understand how the SIA reacts in different scenarios. Measurements were simulated every second over a period of 24 hours leading to a 86400 samples. A variation on the data resolution will increase the detection time, reducing the accuracy of the calculation and the number of outliers per hour. In our simulation with one second of data resolution, the attack will be detected after one second of starting. For that reason the data resolution must taken in consideration to properly configure the tool in relation to the analyzed system. The more variables and sensors the SIA can check, the smaller the detection time will be due to the correlation of events and the amount of outliers generated. All the three scenarios were flagged as containing a substantial amount of anomalies. Figure 4 shows the number of outliers per hour for each scenario. The distribution of outliers are similar for the cases where the attacks AT_1 and AT_2 manipulate the measurements, individually. As it is illustrated, the attacks in all the scenario start at 2:00 and there is a substantial amount of outlier activity detected. The large outlier multiplicity is also the result of the granularity of the measurement.

Less outliers are detected in the combined attack scenario, compared to the single attacks scenarios, because the rules check the internal temperature of the boilers against the incoming temperature and outgoing temperature. By modifying both incoming and outgoing temperatures, the behaviour of the system is similar to normal operations, and outliers are not detected as often. The individual AT_1 and AT_2 attacks are less subtle and so result in a much larger number of outliers being detected. The attack stops at 23:00 in the AT_1 scenario so the number of outlier falls to 0, instead in AT_2 it stops at 21:30 and the number of outliers per hour falls to around 1000.

B. Resilient Control

The performance of the proposed resilient control for the BMS, is evaluated in this section. In the simulation results shown in the Figure 5, the temperature of header flow,

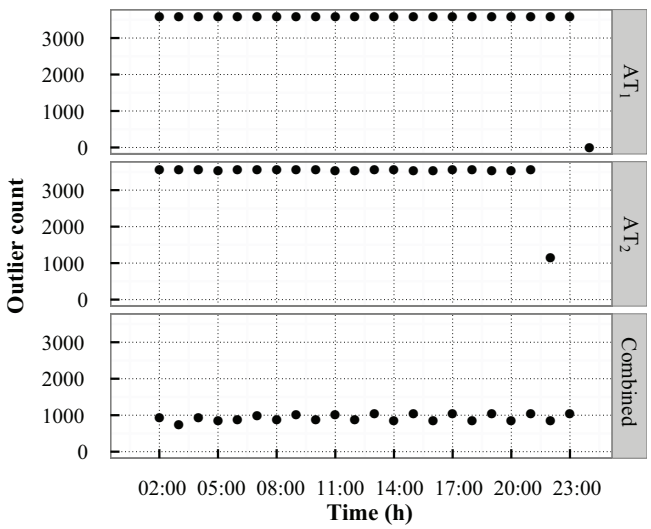

Fig. 4. Number of outliers found per hour. Shown are the attacks on AT_1, attacks on AT_2, and combined attacks on AT_1 and AT_2.

header return, the ground and first floor of Nimbus building, and the ground and first floor of the Rubicon building are under consideration. These temperatures are corresponding to the six measurements of the sensors S_1-S_6 in Figure 1, respectively. The results are shown for three different cases of healthy BMS, attacked BMS, and attack-resilient BMS. In the simulations, the worst attack scenario (the combined attacks AT_1 and AT_2, which is described in Section IV-A) is considered to start at the k' =7200s (at 2:00 am). It means that after the k', the measurements of the header flow and header return temperatures are manipulated by adding 15°C to each of them, and are fed to the respective controllers. In this attack scenario, the attack is considered to be detected by the SIA immediately after k', and the corrupted measurements are replaced with their estimates that are sent by the VS. As it is illustrated in Figure 1, the multi-attack on the measurements leads to high safety risk due to damaging of CHP in the attacked BMS, since return temperature is below 65°C after the attack. In this attack scenario, the attack-resilient BMS is robust against the multi-attack and have the same performance as the healthy BMS.

In Figure 6, measurements of the header flow and header return temperatures in the healthy BMS are shown, and compared with the estimates of the outputs of the attack resilient BMS. Note that the estimates of the outputs are computed by the VS, and one can see that the VS accurately estimates the measurements of the healthy system in the presence of the attack.

Some major factors such as communication delay, large amount of data, and time-consuming security analysis algorithms, can affect the real-time attack detections. To investigate performance of the proposed resilient control in these situations, the following scenario is considered. Assume that the same attack as before, starts at time k' =7200s, and is detected at 7500s (with five minutes delay). As it is shown

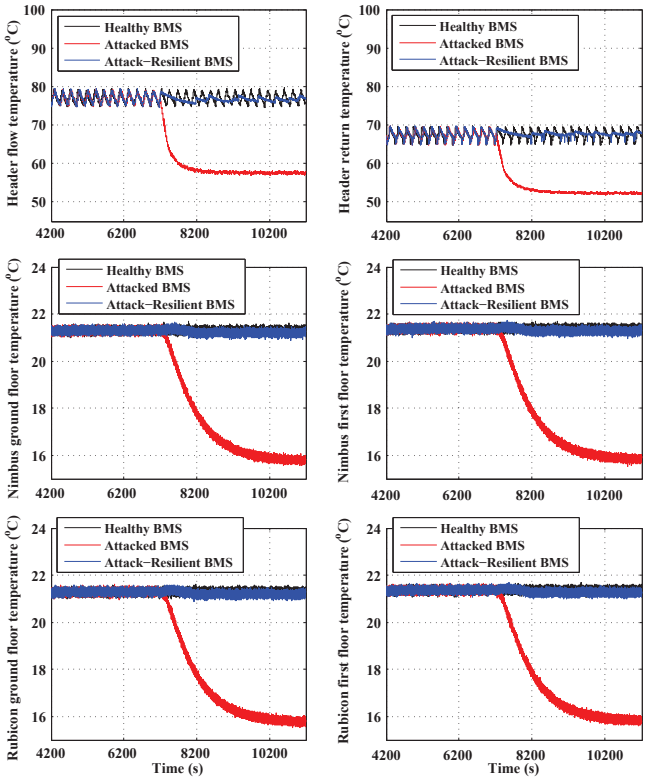

Fig. 5. Performance comparison of the Healthy BMS, Attacked BMS and Attack-Resilient BMS, in the presence of attack on the header flow and return temperature measurements

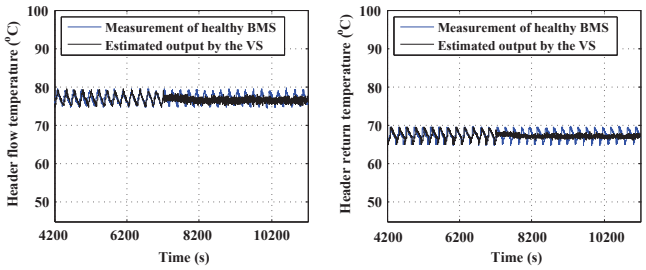

Fig. 6. Measurements of the header flow and header return temperatures in the healthy BMS in comparison with the estimation of the outputs (which are computed by the VS) in the attack-resilient BMS.

in Figure 7, the attack-resilient BMS has the same outputs as the attacked BMS until attack detection (7500s), but it can recover the system to return to the normal conditions after that. We have done other simulations with different delays for the attack detections, and in all the cases, the attack-resilient BMS has recovered the system to return to the normal conditions after the attack detection.

V. Conclusion

The paper presented a cyber-security framework as applicable to a building Energy Management System. The framework uses the physics of the system to drive the security information analytics and resilient policy. The framework efficiency was demonstrated on a real critical attack scenario, where the security information analytics algorithm triggers

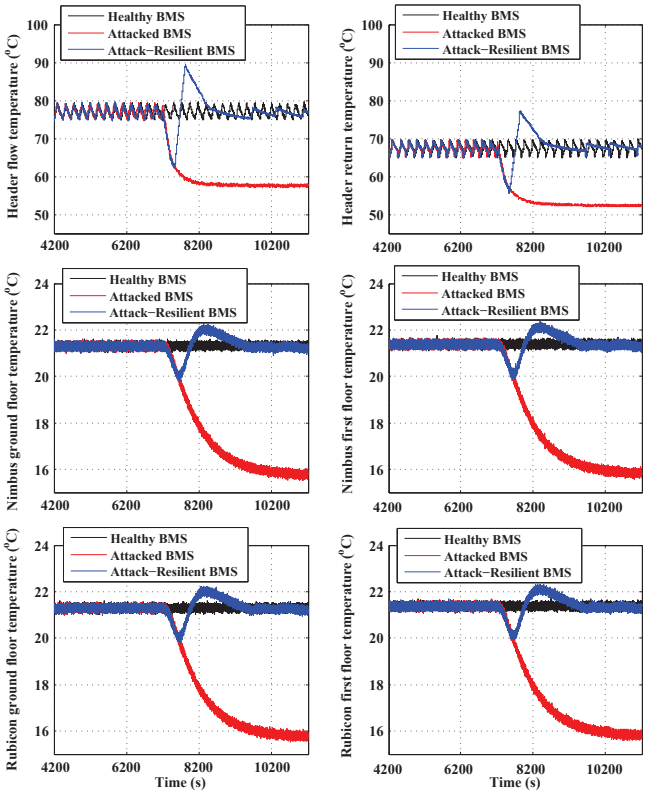

Fig. 7. Performance comparison of the Healthy BMS, Attacked BMS and Attack-Resilient BMS, in the presence of delay in attack detection (the attack starts at time 7200s, and is detected at 7500s).

the resilient control to recover from the attack. Simulation results show that the proposed resilient control policy can recover the system from abnormal conditions, even when there exist delay for the attack detections. As a future work, this framework is planned to be extended to consider the risk assessment as part of the resilient policy.

REFERENCES

[1] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[2] [Online]. Available: http://www.securityweek.com/blackenergy-group-uses-destructive-plugin-ukraine-attacks

[3] S. Gold, "The scada challenge: securing critical infrastructure," *Network Security*, vol. 2009, no. 8, pp. 18–20, 2009.

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

[5] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135 – 148, 2015.

[6] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *Automatic Control, IEEE Transactions on*, vol. 58, no. 11, pp. 2715–2729, 2013.

[7] M. Govindarasu, A. Hann, and P. Sauer, "Cyber-physical systems security for smart grid," *The Future Grid to Enable Sustainable Energy Systems, PSERC Publication*, 2012.

[8] A. A. Cardenas, P. K. Manadhata, and S. P. Rajan, "Big data analytics for security," *IEEE Security & Privacy*, no. 6, pp. 74–76, 2013.

[9] J. L. M. S. M Blanke, M Kinnaert, *Diagnosis and Fault-Tolerant Control*. Berlin Heidelberg: Springer, 2006.

[10] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[11] S. Sundaram and C. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *Automatic Control, IEEE Transactions on*, vol. 56, no. 7, pp. 1495–1508, 2011.

[12] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *Automatic Control, IEEE Transactions on*, vol. 59, no. 6, pp. 1454–1467, 2014.

[13] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *Cyber-Physical Systems (ICCPS), 2014 ACM/IEEE International Conference on*, 2014, pp. 163–174.

[14] N. Bezzo, J. Weimer, M. Pajic, O. Sokolsky, G. Pappas, and I. Lee, "Attack resilient state estimation for autonomous robotic systems," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, 2014, pp. 3692–3698.

[15] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *Control Systems, IEEE*, vol. 35, no. 1, pp. 110–127, 2015.

[16] K. Park, Y. Kim, S. Kim, K. Kim, W. Lee, and H. Park, "Building energy management system based on smart grid," *Telecommunications Energy Conference (INTELEC), 2011 IEEE 33rd International*, pp. 1–4, 2011.

[17] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 81–85, 2010.

[18] V. Valdivia, S. OConnell, F. Gonzalez-Espin, A. E. din Mady, K. Kouramas, L. D. Tommasi, H. Wiese, B. C. Villaverde, R. Foley, M. Cychowski, L. Hertig, D. Hamilton, and D. Pesch, "Sustainable building integrated energy test-bed," *Power Electronics for Distributed Generation Systems (PEDG), 2014 IEEE 5th International Symposium on*, pp. 1–6, 2010.

[19] S. Timlin, "Improving automation routines for automatic heating load detection in buildings," *Journal of Sustainable Engineering Design*, vol. 1, no. 2, 2012.

[20] B. Sohlberg, "Grey box modelling," *Supervision and Control for Industrial Processes, Part of the series Advances in Industrial Control*.

[21] A. Teixeira, K. Paridari, H. Sandberg, and K. Johansson, "Voltage control for interconnected microgrids under adversarial actions," in *IEEE International Conference on Emerging Technology and Factory Automation*, 2015.

[22] E. Henriksson, H. Sandberg, and K. Johansson, "Reduced-order predictive outage compensators for networked systems," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, 2009, pp. 3775–3780.

[23] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[24] L. Ljung, "Prediction error estimation methods," *Circuits, Systems and Signal Processing*, vol. 21, no. 1, pp. 11–21, 2002.

[25] S. J. Qin, "An overview of subspace identification," *Computers and Chemical Engineering*, vol. 30, no. 12, pp. 1502–1513, 2006.

[26] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000.

[27] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, July 2001.