Distributed watermarking for secure control of microgrids under replay attacks

Alexander J. Gallo [*] **Mustafa S. Turan** [**] **Francesca Boem** [***]
Giancarlo Ferrari-Trecate [**] **Thomas Parisini** [*,****]

[*] *Imperial College London, London, UK. (emails: alexander.gallo12@imperial.ac.uk,t.parisini@gmail.com).*
[**] *École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. (e-mails: {mustafa.turan, giancarlo.ferraritrecate}@epfl.ch)*
[***] *University College London, UK. (e-mail: f.boem@ucl.ac.uk)*
[****] *University of Trieste, Italy and KIOS Research and Innovation Centre of Excellence, University of Cyprus.*

Abstract: The problem of replay attacks in the communication network between Distributed Generation Units (DGUs) of a DC microgrid is examined. The DGUs are regulated through a hierarchical control architecture, and are networked to achieve secondary control objectives. Following analysis of the detectability of replay attacks by a distributed monitoring scheme previously proposed, the need for a watermarking signal is identified. Hence, conditions are given on the watermark in order to guarantee detection of replay attacks, and such a signal is designed. Simulations are then presented to demonstrate the effectiveness of the technique.

1. INTRODUCTION

1.1 Motivation and state of the art

The interest in islanded microgrids has spiked over the past few years, following the increased penetration of renewable energy sources within the electrical network as they can offer benefits compared to traditional grids (Meng et al., 2017b). Specifically, as many of the adopted generation, storage, and end-user equipment are based on DC technology, DC microgrids (DCmGs) have attracted a lot of research activity, as can be seen from the recent survey (Meng et al., 2017b). In order to guarantee the stable and efficient behavior of the microgrid, hierarchical control architectures have been proposed (Guerrero et al., 2009), in which a decentralized primary layer ensures voltage, current and power stability (Tucci et al., 2016; Zhao and Dörfler, 2015), and secondary and tertiary control layers offer additional properties, such as power quality regulation, load sharing, and ensuring overall coordination and optimization (De Persis et al., 2016; Tucci et al., 2017; Zhao and Dörfler, 2015). It has been shown that, in order to achieve some of the objectives, the secondary and tertiary controllers require the support of a communication network (Meng et al., 2017a; Cavraro et al., 2016).

The integration of communication networks within control systems has exposed them to the possibility of being tampered by malicious agents, injecting false information within the control loop, thus altering their behavior (Cheng et al., 2017). Given the possibility of attacks, it has been recognized as necessary to introduce monitoring structures capable of evaluating whether operations are running as normal or not (see papers in (Cheng et al., 2017; Sandberg et al., 2015; Urbina et al., 2016)).

Among several types of attacks, during *replay attacks* an attacker is able to record data transmitted over a communication network, and then *replay* it, replacing actual communication signals with buffered data. This class of attacks has been shown to be particularly difficult to detect for common monitoring schemes, as they present the same statistics as the nominal behavior (Mo et al., 2015). Several techniques have been proposed in order to detect this class of attack, often by altering the characteristics of the system through the addition of a *watermark* (Mo et al., 2015; Ferrari and Teixeira, 2017). In (Mo et al., 2015), a time-varying watermark is added to the input signal of a system, in order to alter the characteristic statistics of the steady state, thus allowing the monitoring scheme to detect the presence of an attack. In (Ferrari and Teixeira, 2017), the watermark signal is added directly to the sensor measurements communicated to the monitoring scheme and controller. Other techniques have been proposed to counteract stealthy data injection attacks. For example in (Miao et al., 2017), the proposed strategy encodes the sensor measurements through a *"coding matrix"*, assumed to be unknown by the attacker, in order to prevent stealthy data injection.

All of the methods proposed to detect replay attacks rely on the possibility of implementing the monitoring scheme centrally. This, however, is not desirable for DCmGs, as all distributed generation units (DGUs) would need to communicate with a central point, which would increase the

[⋆] This work has been partially supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE). This work has also been conducted as part of: i) the research project *Stability and Control of Power Networks with Energy Storage* (STABLE-NET) which is funded by the RCUK Energy Programme (contract no: EP/L014343/1); ii) the Swiss National Science Foundation under the COFLEX project (grant number 200021_169906).

communication cost. In this preliminary work, we propose a distributed watermarking technique for microgrids.

Recently, a few works have examined attacks on microgrids, although not considering replay attacks. The susceptibility of the secondary control objectives to jamming attacks on the communication between DGUs is shown in (Danzi et al., 2016), while in (Gallo et al., 2018b) we have proposed a distributed monitoring scheme to detect the presence of attackers in the communication network, developed off the preliminary work in (Boem et al., 2017).

1.2 Objectives and contributions

In this paper, we consider a DCmG regulated as in (Tucci et al., 2017), and monitored as in (Gallo et al., 2018b). We introduce a distributed watermarking scheme which allows for the detection of replay attacks in the communication network. The main objectives that this fulfills are:

a. Enhance the monitoring scheme in (Gallo et al., 2018b), in order to detect replay attacks;
b. Be distributed, running attack detectors at each DGU location and using the same communication network required for secondary control in (Tucci et al., 2017);
c. Ensure the main goals of the primary and secondary controllers, i.e. voltage regulation and current sharing, respectively, are not compromised by watermarking;
d. Not be easily identifiable by a resourceful attacker.

Objective (a) is motivated by the fact that the monitoring scheme in (Gallo et al., 2018b) is vulnerable to replay attacks, as shown in Section 3. We also derive detectability conditions, fundamental for the design of the watermark, which must be fulfilled in order to guarantee detection. This analysis will allow us to design a preliminary watermark which enables the detection of replay attacks.

1.3 Paper structure

The rest of the paper is structured as follows. In Section 2 we introduce the model of the DCmG which is considered. Section 3 summarizes the monitoring scheme in (Gallo et al., 2018b),and defines the replay attack, demonstrating that it is stealthy to the considered scheme. In Section 4 we analyze the detectability properties of a generic watermarking scheme, and propose a preliminary watermark design. Some simulation results are shown in Section 5. Finally, in Section 6, we provide some concluding remarks. [1]

1.4 Notation

In the paper, the operator $|\cdot|$ applied to a set determines its cardinality, while used with matrices or vectors it defines their component-by-component absolute value. The operator $\|\cdot\|$ is used to define the matrix norm. In general, in this paper inequalities are considered component-by-component. The operators $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ define, respectively, the ceiling and floor functions.

2. CHARACTERIZATION OF DC MICROGIRD

In the following, we consider a DCmG with N distributed generation units interconnected by power lines. Each DGU is modeled as in (Tucci et al., 2016), where the interconnection power lines are taken to be purely resistive. The DCmG modeled in this framework can therefore be represented as an undirected graph, where nodes represent the DGUs, edges denote power lines, and edge weights are given by the conductance of the lines (see Fig 1). DGUs are defined as the cascade of a DC voltage source, a DC/DC converter, and a RLC filter connecting it to other DGUs at a point of common coupling (PCC). DGUs have local loads, represented as known currents I_{Li}. The dynamics of each DGU are characterized in state space as follows:

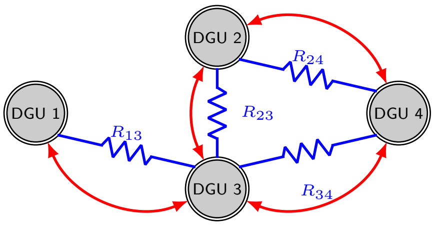

Fig. 1. Graph of DCmG. The blue lines represent the power lines connecting the DGUs, and the red arrows represent the communication graph.

$$\begin{aligned}\dot{x}_{[i]}(t) &= A_{ii}x_{[i]}(t) + B_i u_{[i]}(t) + G_i\alpha_{[i]}(t) \\ &\quad + M_i d_{[i]}(t) + \xi_{[i]}(t) + w_{[i]}(t) \quad , \qquad (1) \\ y_{[i]}(t) &= x_{[i]}(t) + \rho_{[i]}(t)\end{aligned}$$

where $x_{[i]} = [V_i, I_{ti}, v_{[i]}]^\top$ is the state of the DGU, V_i is the voltage at the PCC, I_{ti} is the terminal current from the converter, and $v_{[i]}$ is a scalar state needed to include an integrator action in the primary loop. $u_{[i]}$ and $\alpha_{[i]}$ are respectively the primary and secondary input, where the latter is defined in (3). Vector $d_{[i]} = [I_{Li}, V_{ref,i}]^\top$ is an external known input to DGU i, where $V_{ref,i}$ is a voltage reference for V_i. $\xi_{[i]} = \sum_{j\in\mathcal{N}_i} A_{ij}x_{[j]}$ is a vector modeling the interconnection with other DGUs, where $\mathcal{N}_i \subset \mathcal{N} \equiv \{1, \cdots, N\}$ is the set of the neighbors of DGU i, i.e. the DGUs that are interconnected to DGU i through power lines. Vectors $w_{[i]}$ and $\rho_{[i]}$ model unknown state and measurement noises, respectively.

Assumption 1. The unknown process noise $w_{[i]}$ and measurement noise $\rho_{[i]}$ are bounded for all time:

$$|w_{[i]}(t)| \leq \bar{w}_{[i]}, |\rho_{[i]}(t)| \leq \bar{\rho}_{[i]}, \forall t, \qquad (2)$$

where $\bar{w}_{[i]}$ and $\bar{\rho}_{[i]}$ are known. △

The primary input $u_{[i]} = V_{ti} = K_i y_{[i]}$ is given by a decentralized feedback controller (Tucci et al., 2016), and guarantees voltage stability and reference tracking in the whole DCmG. The secondary control input $\alpha_{[i]}$ allows for current to be shared among DGUs, and is the result of the following consensus-based protocol:

$$\dot{\alpha}_{[i]}(t) = -\sum_{j\in\mathcal{N}_i} [0\; k_I\; 0] \left(\frac{y_{[i]}(t)}{I^s_{ti}} - \frac{y^c_{[j,i]}(t)}{I^s_{tj}} \right), \qquad (3)$$

where the scalars $I^s_{ti} > 0, \forall i \in \mathcal{N}$ are design parameters which allow current sharing at different rates. Scalar $k_I > 0$ is the consensus weight, common to all DGUs. Vector $y^c_{[j,i]}(t)$ represents the output measurement which DGU j transmits to DGU i, as defined in (4) below. The matrices in (1) are defined as in (Tucci et al., 2016).

[1] Due to space limitations, all proofs are reported in (Gallo et al., 2018a)

In order to operate the consensus-based secondary controller described in (3), it is necessary to introduce a communication network connecting different DGUs, assumed to have the same topology as the DCmG. For ease of analysis, in this preliminary paper, we introduce the following assumption:

Assumption 2. The communication network is ideal, i.e. information is exact and without time delays. Hence communicated data is:

$$y^c_{[i,j]}(t) = y_{[i]}(t) \quad , \tag{4}$$

i.e. is equal to the measurement vectors. △

The introduction of a communication network exposes the system to malicious attacks. In the following section we describe how we model the attack strategy, and define the type of attack motivating this work.

3. CYBER-ATTACKS AND MONITORING SCHEME

Our goal is to equip all DGUs with a monitoring scheme, allowing to check whether the information received from each neighbor is corrupted by an attack or not. In order to formally introduce the attack in the communication value $y^c_{[i,j]}(t)$, we redefine (4) as the following:

$$y^c_{[i,j]}(t) = y_{[i]}(t) + \beta_{ij}(t - T_a)\phi_{i,j}(t) \tag{5}$$

where $\phi_{i,j}(\cdot)$ is the attack function - designed by the attacker according to its objectives and available resources - and $\beta_{ij}(t - T_a)$ is a step function, representing attack activation at time T_a.

In order to detect the action of an attacker, a distributed attack detection scheme based on multiple Unknown Input Observers (UIOs) has been introduced (Gallo et al., 2018b). This monitoring strategy is capable of detecting whether the communication is subject to an attack, based on limited knowledge of the neighbors' dynamics, and on information regarding the bounds on the disturbance in Assumption 1. We will now briefly summarize the considered monitoring strategy.

3.1 Monitoring strategy

In the scheme proposed in (Gallo et al., 2018b), each DGU estimates the state of each of its neighbors. The error between the estimate and the received measurements vector is then compared to a time-varying threshold, designed based on the bounds in Assumption 1, to determine whether the communication network is secure. In the remainder, we will present the estimation scheme in DGU i for the state of its neighbor j. The UIO framework allows to avoid communication of some information, as any transmitted information may be subject to attacks.

In order to exploit the UIO estimator, we rewrite the dynamics of DGU j as follows:

$$\begin{aligned} \dot x_{[j]}(t) &= A_{Kj} x_{[j]}(t) + \bar E_j \bar d_{[j]}(t) + \tilde w_{[j]}(t) \\ y_{[j]}(t) &= x_{[j]}(t) + \rho_{[j]}(t) \end{aligned} \tag{6}$$

where $A_{Kj} = A_{jj} + B_j K_j$, $\tilde w_{[j]}(t) = w_{[j]}(t) + B_j K_j \rho_{[j]}(t)$, and $\bar E_j \bar d_{[j]}$ represents the inputs to DGU j which are unknown to the UIO in DGU i. Specifically, $\bar d_{[j]} = \widehat E_j \hat d_{[j]}$ is a linear combination of the vector of variables:

$$\hat d_{[j]} = \left[d^\top_{[j]}(t), \alpha_{[j]}(t), x^\top_{[k_1]}(t), \dots, x^\top_{[k_{|\mathcal N_j|}]}(t) \right]^\top ,$$

while $\bar E$ is full column rank and derived from matrices in (1), and is defined in (Gallo et al., 2018b) along with $\widehat E_j$.

UIO state and state estimate are (Chen et al., 1996):

$$\begin{aligned} \dot z_{[j,i]}(t) &= F_j z_{[j,i]}(t) + S_j B \bar u_{[j]}(t) + \widehat K_j y^c_{[j,i]}(t) \\ \hat x_{[j,i]}(t) &= z_{[j,i]}(t) + H_j y^c_{[j,i]}(t) \end{aligned} \tag{7}$$

where the matrices are defined as in (Gallo et al., 2018b), and are such that F_j is Hurwitz stable, and $S_j \bar E_j = 0$. The residual is defined as $r_{[j,i]}(t) = y^c_{[j,i]}(t) - \hat x_{[j,i]}(t)$, and, given stability of (7) and Assumption 1, is bounded, i.e.

$$\left| r_{[j,i]}(t) \right| \le \bar r_{[j,i]}(t) \tag{8}$$

holds for all $t \ge 0$, where

$$\bar r_{[j,i]}(t) = \bar e_{[j,i]}(t) + \bar\rho_{[j]} . \tag{9}$$

is the detection threshold. In (9) , $\bar e_{[j,i]}(t)$ is the time-varying bound on the estimation error $e_{[j,i]}(t) = x_{[j]}(t) - \hat x_{[j,i]}(t)$, defined as:

$$\begin{aligned} \bar e_{[j,i]}(t) &= \kappa e^{-\mu t} \left[\bar e_{[j,i]}(0) + |H_j| \bar\rho_{[j]} \right] + |H_j| \bar\rho_{[j]} \\ &+ \int_0^t \kappa e^{-\mu(t-\tau)} \left[|S_j| \bar w_{[j]} + |S_j B_j K_j - \widehat K_j| \bar\rho_{[j]} \right] d\tau \end{aligned} \tag{10}$$

where scalars $\kappa, \mu > 0$ are such that $\|e^{F_j t}\| \le \kappa e^{-\mu t}$. This bound, for suitably defined $\bar e_{[j,i]}(0)$, guarantees that $|e_{[j,i]}(t)| \le \bar e_{[j,i]}(t), \forall t \ge 0$. Given these bounds on the estimation error and residual, an attack is detected if (8), used as a detection test, is not satisfied.

The detectability conditions of this scheme have been studied in (Gallo et al., 2018b). In the following section we show that it may fail to detect *replay attacks.*

3.2 Replay attack

As anticipated in the introduction, we focus specifically on the class of replay attacks, as even without any knowledge of the system model and much computational resources, they can be undetectable (Mo et al., 2015). In the scenario we are considering, a malicious agent is able to eavesdrop the communication between DGU j and DGU i, from some unknown time $t = T_0$, and thus to start storing this data in a memory buffer up to a time $T_a > T_0$. At time $t = T_a$, the attacker starts injecting the following attack function into the communicated signal (5):

$$\phi_{[j,i]}(t) = -y^c_{[j,i]}(t) + y^c_{[j,i]}(t - nT), \tag{11}$$

i.e. it replaces the current transmitted measurements of the output of DGU j with past recorded measurements. The integer $n = \lceil (t - T_a)/T \rceil$ represents the periodicity of the signal the attacker injects, where $T \le T_a - T_0$ is the period of the repeated data, as decided by the attacker.

These attacks are particularly deceptive if data is recorded when the state of the system is in a quasi stationary régime. In Prop. 1, we give a preliminary result showing a sufficient condition for the attack to be stealthy.

Proposition 1. (Stealthy Replay Attacks). If

$$|e^a_{[j,i]}(T_a)| \le \bar e_{[j,i]}(T_a) \tag{12}$$

is satisfied, where $e^a_{[j,i]}(T_a) = x(T_a - T) - \hat x_{[j,i]}(T_a)$, then inequality (8) holds for all $t \in [T_a, T_a + T)$, and the attack will not be detected. □

Remark 1. The objective of the malicious agent executing a replay attack is to hide any change in operating conditions in DGU j from DGU i, i.e. the changes caused by a load change in DGU j or one of its neighbors. By doing so, the attack is able to alter the equilibrium which is reached through consensus, thus impeding current sharing, or it may even be able to make it impossible to reach consensus.

In the following we will present a detection strategy based on *watermarking* to detect the presence of replay attacks in the communication network.

4. CYBER-ATTACK DETECTION METHOD

To make a replay attack detectable, similar to the intuition behind sensor watermarking (Ferrari and Teixeira, 2017), we add a time varying signal $\Delta_{[i,j]}(t)$ to the measurements communicated from DGU i to DGU j. For this preliminary work the following is assumed to not bias the performance of the consensus scheme (objective (c) in Section 1.2):

Assumption 3. Watermark $\Delta_{[i,j]}(t)$ added to $y_{[i,j]}(t)$ is known exactly by both DGU i and DGU j for all t. △

4.1 Watermark signal

With the addition of the watermark, the communicated measurement (4) becomes:

$$y^c_{[i,j]}(t) = y_{[i]}(t) + \Delta_{[i,j]}(t). \tag{13}$$

Given Assumption 3, once $y^c_{[j,i]}(t)$ is received by DGU i, the known watermark is subtracted, as to achieve exact consensus. It is clear to see that the decoded information at DGU i, $\hat{y}_{[j,i]}$, is the measurement of DGU j:

$$\hat{y}^c_{[j,i]}(t) = y^c_{[j,i]}(t) - \Delta_{[j,i]}(t) = y_{[j]}(t),$$

which is then used in the dynamics of the secondary input (3), as well as in the computation of the estimates (7) and in the evaluation of the residual in inequality (8).

Analyzing the effect of the watermark on the UIO estimators, it appears evident that under normal operating conditions, the value of $\hat{y}^c_{[j,i]}(t)$ will be the same as if the watermark weren't present. Hence, given analysis in (Gallo et al., 2018b), the residual $r_{[j,i]}(t) = \hat{y}^c_{[j,i]}(t) - \hat{x}_{[j,i]}(t)$ does not exceed its bound, avoiding false alarms, as (8) always holds. We now analyze the residual under replay attack.

4.2 Detectability analysis

For $t \geq T_a$, as previously mentioned, the information received by DGU i will be the buffered measurement $y^c_{[j,i]}(t - nT)$, and hence decoded data is:

$$\hat{y}^c_{[j,i]}(t) = y_{[j,i]}(t - nT) + \delta_{[j,i]}(t) \tag{14}$$

where $\delta_{[j,i]}(t)$ is defined as

$$\delta_{[j,i]}(t) \coloneqq \Delta_{[j,i]}(t - nT) - \Delta_{[j,i]}(t). \tag{15}$$

We also redefine the state estimation error in the presence of the watermark as $\epsilon^a_{[j,i]}(t) = x_{[j]}(t - T) - \hat{x}_{[j,i]}(t)$, and note that, for time instance $t = T_a$:

$$\epsilon^a_{[j,i]}(T_a) = e^a_{[j,i]}(T_a) - H_j \delta_{[j,i]}(T_a), \tag{16}$$

To verify detectability conditions of the replay attack, it is necessary to analyze the residual under replay attacks, which must then be compared with the detection threshold $\bar{r}_{[j,i]}(t)$. Given T_a and T, for $t \geq T_a$ residual is:

$$r_{[j,i]}(t) = \epsilon^a_{[j,i]}(t) + \rho_{[j]}(t - nT) + \delta_{[j,i]}(t).$$

For the value of $\epsilon^a_{[j,i]}(t)$, we analyze its dynamics starting from (1) and (7). In the first period over which the replay attack is active, i.e. for $t \in [T_a, T_a + T)$, the dynamics of the estimation error are:

$$\begin{aligned} \dot{\epsilon}^a_{[j,i]}(t) &= F_j \epsilon^a_{[j,i]}(t) + S_j \tilde{w}_{[j]}(t) + \\ &- \widetilde{K}_j \left(\rho_{[j]}(t) + \delta_{[j,i]}(t)\right) - H_j \left(\dot{\rho}_{[j]}(t) + \dot{\delta}_{[j,i]}(t)\right). \end{aligned} \tag{17}$$

Using integration by parts as in (Gallo et al., 2018b), the explicit solution of these dynamics are:

$$\begin{aligned} \epsilon^a_{[j,i]}(t) &= e^{F_j(t-T_a)} \Big(\Big[e^a_{[j,i]}(T_a) - H_j \delta_{[j,i]}(T_a) \Big] + \\ &+ H_j \rho_{[j]}(T_a - T) + H_j \delta_{[j,i]}(T_a) \Big) - H_j \left(\rho_{[j]}(t - T) + \right. \\ &\left. + \delta_{[j,i]}(t) \right) + \int_{T_a}^{t} e^{F_j(t-\tau)} \Big[S_j w_{[j]}(\tau - T) + \\ &+ (S_j B_j K_j - \widehat{K}_j) \rho_{[j]}(\tau - T) - \widehat{K}_j \delta_{[j,i]}(\tau) \Big] \, d\tau. \end{aligned} \tag{18}$$

We analyze the case presented in Prop. 1, where the attack is stealthy in the absence of the watermark. It is possible to formulate the following detectability condition:

Proposition 2. If, for some $t = T_d \geq T_a$, the inequality

$$\left| S_j \delta_{[j,i]}(t) - \int_{T_a}^{t} e^{F_j(t-\tau)} \left[\widehat{K}_j \delta_{[j,i]}(\tau) \right] d\tau \right| > 2 \bar{r}_{[j,i]}(t) \tag{19}$$

holds, then detection of the replay attack is guaranteed. □

4.3 Watermark design

Now, we design a watermark signal $\Delta_{[j,i]}(t)$, that allows the detection of a replay attack through (19), with the additional aim of obstructing the attacker from identifying the watermark from analysis of the communicated signal (13). In qualitative terms, we note that these two objectives conflict with each other, as the first would benefit from large amplitude changes over time, which may, however, aid the identification of the watermark. Furthermore, we notice that the watermark enters condition in (19) through $\delta_{[j,i]}(t)$, rather than $\Delta_{[j,i]}(t)$ itself, thus making it challenging to design a watermark satisfying (19) independently of T_a and T, which are unknown to all but the attacker. For this work, we introduce the following assumption, to simplify watermark design.

Assumption 4. The attack period is upper bounded by some known quantity, i.e. $T \leq \bar{T}$. △

To justify that this assumption is mild, we first note that if the attacker records data while the DGUs are in steady state, the attack will be stealthy to the considered monitoring scheme, following Prop. 1. We also stress that we are considering a network of DGUs, whose steady state is determined by the load currents in all subsystems. It is usually possible to define an upper bound $\bar{T}$ as the maximum period between load changes within the DCmG, which can be evaluated empirically, or estimated *a priori*.

Using this assumption, we then propose a sawtooth signal of period $2\bar{T}$ as the watermark:

$$\Delta_{[j,i]}(t) = c_{[j,i]}(t - 2\nu\bar{T}), \tag{20}$$

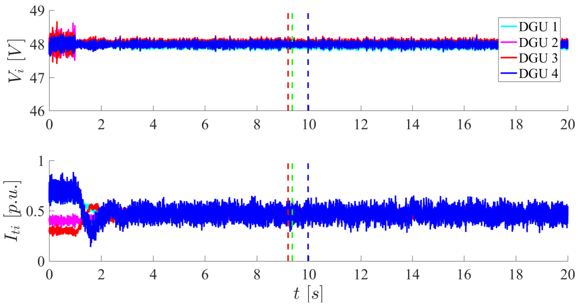

Fig. 2. States V_i and I_{ti} of DGUs 1-4. Vertical lines show initial time of attack on DGU 4 (red), and detection of replay attack in communication from DGU 2 and 3 (in blue and green, respectively).

where $\nu = \lfloor t/2\bar{T} \rfloor$, and $c_{[j,i]}$ is the slope. The period $2\bar{T}$ is selected to avoid having $\delta_{[j,i]}(t) = 0$ for any $T \in (0, \bar{T}]$. Indeed, $\delta_{[j,i]}(t)$ will be a square wave of period $2\bar{T}$.

AssIn the following we show through simulation how it is possible to tune the proposed watermark such that it is difficult to identify by the attacker, while nonetheless enabling detection of a replay attack which is stealthy for the monitoring scheme proposed in (Gallo et al., 2018b).

5. SIMULATION RESULTS

The proposed detection scheme with watermark is validated through simulations in MATLAB. A DC microgrid consisting of 4 DGUs, interconnected as in Fig 1, is considered, where the parameters and matrices for DGUs and UIOs are taken as in (Gallo et al., 2018b). Process and measurement noises are drawn from uncorrelated uniform distributions satisfying Assumption 1, where $\bar{w}_{[i]} = [0.1, 0.1, 0.1]^\top$ and $\bar{\rho}_{[i]} = [0.01, 0.01, 0.01]^\top$. These noises induce variations in the electrical signals comparable to those seen in real micrgrids, due to the converter operations and measurement noises.

Before time $t = 1s$, all the DGUs are disconnected and do not communicate. At time $t = 1s$, neighboring DGUs connect to each other to create the microgrid topology given in Fig 1. The current loads at the PCCs of DGUs 1, 2, and 3 are considered to periodically change from $6A$ to $6.2A$, $4A$ to $4.25A$, and $3A$ to $3.15A$, respectively, to show the dynamic nature of loads in a microgrid. The attacker chooses $T_a = 9.2s$ to start replaying the data it recorded starting at $t = 7.4s$, i.e. choosing attack period $T = \bar{T}$ =1.8s, so that the data is recorded in steady state.

Each DGU adds a watermark signal as in (20) to its measurements communicated to its neighbors. Added watermarks are the same for each communication link, i.e., $c_{[i,j]} = c_{[i,k]}$ $\forall i \in \mathcal{N}, \forall j, k \in \mathcal{N}_i$, to prevent attacker from identifying the watermark from the differences between outgoing communication from the same DGU. The slopes $c_{[j,i]}$ of the watermarks take constant values from $10^{-3.2}$ to $10^{-3.5}$ for each DGU.

Voltages and currents of each DGU of the network are shown in Fig. 2, whereas Fig. 3 shows the comparison of the residuals of DGU 4's estimates of its neighbors, compared to their respective thresholds. It can be seen that detection occurs for both communicated measurements, although the effect of the replay attack on the DCmG is not noticeable from Fig. 2. The relatively late detection based on the residual of the UIO estimating the states of DGU 3 is due to the smaller watermark slope $c_{[3,i]}$. We also show, Fig. 4, that if the watermark were not present, the detection of the attack would not occur.

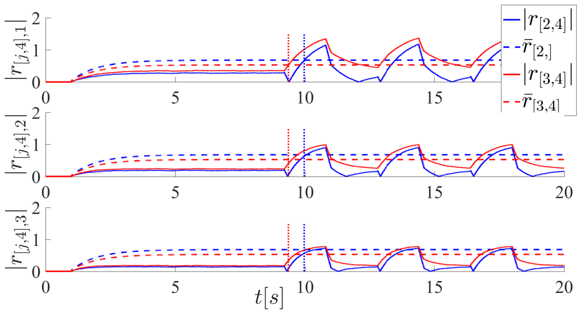

Fig. 3. Comparison of residuals and thresholds, in solid and dotted lines, respectively, of monitors in DGU 4 with watermarking. Data regarding estimation of state of DGU 2 is presented in blue, while data referring to DGU 3 is in red. Detection occurs as one component of the residual exceeds its threshold (vertical lines).

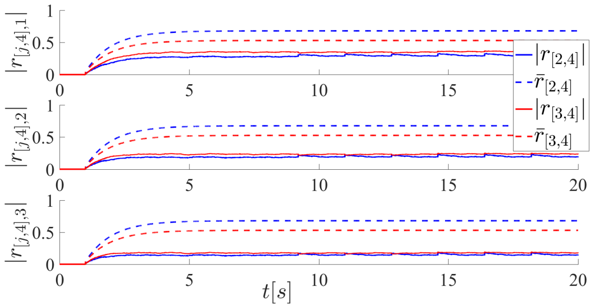

Fig. 4. Comparison of $r_{[j,4]}$ and $\bar{r}_{[j,4]}$ of estimators in DGU 4 without watermark (colors used as in Fig. 3). As can be seen, the attack is not detected.

Enabling Detection: We now provide some insight on the design of the watermark. Specifically, we note that in (18), while $w_{[j]}$ and $\rho_{[j]}$ are multiplied by S_j and $(S_j K_j B_j - \widehat{K}_j)$, respectively, $\delta_{[j,i]}(t)$ is scaled by $\widehat{K}_j$. This is a design parameter of the UIO which, as seen in (Gallo et al., 2018b), can be made to have a larger absolute value than the other scaling matrices. Hence, even with $c_{[j,i]}$ small, $\Delta_{[j,i]}(t)$ will have a large impact on $r_{[j,i]}$ when subject to attack.

Identifiability of the watermark: We now focus on the identifiability of the watermark by the attacker from the communicated data. The methods which have been considered possible to be used by the attacker are: by inspection, and through frequency spectrum analysis. In terms of identifiability by inspection, we note that $y^c_{[j,i]} = x_{[j]} + \rho_{[j]} + \Delta_{[j,i]}$, and that the maximum value of a watermark signal over the whole microgrid, $2\bar{T}10^{-3.2}$, is less than a quarter of the bound on the measurement noise of the corresponding DGU, and hence the discontinuity occurring every period will be masked by noise. On the

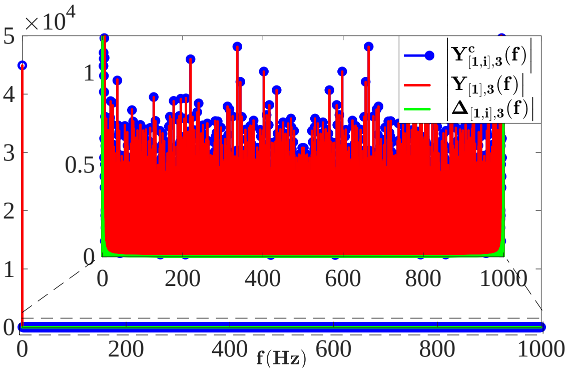

Fig. 5. Comparison of FFTs of $y^c_{[1,i]}$ (in blue), $y_{[1]}$ (in red), and the $\Delta_{[1,i]}$ (in green) of $v_{[1]}$ of DGU 1 from $t \in [5s, 9s]$ (i.e. during quasi stationary régime).

other hand, if the watermark causes significant changes in the frequency domain characteristics of the communicated outputs, it may be possible for the attacker to identify it. Hence, in Fig. 5 we compare the frequency domain of the measurements of $v_{[1]}$ with and without the watermark and to the watermark itself, by analyzing their fast Fourier transforms (FFTs) for $t \in [5s, 9s]$, One can see from this figure that the FFTs of the communicated and actual outputs are very similar and that of the watermark signal is incomparably small. Hence we stipulate that it is difficult for the attacker to identify the watermark signal from a spectral analysis of the communicated outputs.

6. CONCLUSIONS

In this work we have presented a distributed watermarking strategy to support a monitoring scheme used to validate information transmitted between DGUs of a DCmG. Analysis of the monitoring scheme under replay attack, without the additive watermark, shows that as long as data is recorded in steady state, the attack is undetectable. We then introduce the watermark, and we derive a condition on the watermark to guarantee detection. Finally, we propose a preliminary watermark signal design, showing its effectiveness through simulation. As future work, we plan to consider non-ideal communication networks, and to propose a more refined watermarking signal.

REFERENCES

Boem, F., Gallo, A.J., Ferrari-Trecate, G., and Parisini, T. (2017). A distributed attack detection method for multi-agent systems governed by consensus-based control. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 5961–5966. IEEE.

Cavraro, G., Bolognani, S., Carli, R., and Zampieri, S. (2016). The value of communication in the voltage regulation problem. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 5781–5786.

Chen, J., Patton, R.J., and Zhang, H.Y. (1996). Design of unknown input observers and robust fault detection filters. *International Journal of control*, 63(1), 85–105.

Cheng, P., Shi, L., and Sinopoli, B. (2017). Guest editorial special issue on secure control of cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 4(1), 1–3.

Danzi, P., Stefanovic, C., Meng, L., Guerrero, J.M., and Popovski, P. (2016). On the impact of wireless jamming on the distributed secondary microgrid control. In *2016 IEEE Globecom Workshops (GC Wkshps)*, 1–6. IEEE.

De Persis, C., Weitenberg, E., and Dörfler, F. (2016). A power consensus algorithm for DC microgrids. *Automatica*. Submitted.

Ferrari, R.M. and Teixeira, A.M. (2017). Detection and isolation of replay attacks through sensor watermarking. *IFAC-PapersOnLine*, 50(1), 7363–7368.

Gallo, A.J., Turan, M.S., Boem, F., Ferrari-Trecate, G., and Parisini, T. (2018a). Distributed watermarking for secure control of microgrids under replay attacks. *ArXiv e-prints arXiv:1805.00737*. Available on arxiv.org.

Gallo, A.J., Turan, M.S., Nahata, P., Boem, F., Parisini, T., and Ferrari-Trecate, G. (2018b). Distributed cyber-attack detection in the secondary control of dc microgrids. In *European Control Conference, 2018*.

Guerrero, J.M., Vásquez, J.C., and Teodorescu, R. (2009). Hierarchical control of droop-controlled DC and AC microgrids: a general approach towards standardization. In *2009 35th Annual Conference of IEEE Industrial Electronics*, 4305–4310.

Meng, L., Shafiee, Q., Ferrari-Trecate, G., Karimi, H., Fulwani, D., Lu, X., and Guerrero, J.M. (2017a). Review on control of DC microgrids and multiple microgrid clusters. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 5(3), 928–948.

Meng, L., Shafiee, Q., Ferrari-Trecate, G., Karimi, H., Fulwani, D., Lu, X., and Guerrero, J.M. (2017b). Review on control of dc microgrids and multiple microgrid clusters. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 5(3), 928–948.

Miao, F., Zhu, Q., Pajic, M., and Pappas, G.J. (2017). Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Transactions on Control of Network Systems*, 4(1), 106–117.

Mo, Y., Weerakkody, S., and Sinopoli, B. (2015). Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1), 93–109.

Sandberg, H., Amin, S., and Johansson, K.H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems*, 35(1), 20–23.

Tucci, M., Meng, L., Guerrero, J.M., and Ferrari-Trecate, G. (2017). Plug-and-play control and consensus algorithms for current sharing in DC microgrids. In *Proceedings of the 20th IFAC World Congress*, 12951–12956.

Tucci, M., Riverso, S., Vasquez, J.C., Guerrero, J.M., and Ferrari-Trecate, G. (2016). A decentralized scalable approach to voltage control of DC islanded microgrids. *IEEE Transactions on Control Systems Technology*, 24(6), 1965–1979.

Urbina, D.I., Urbina, D.I., Giraldo, J., Cardenas, A.A., Valente, J., Faisal, M., Tippenhauer, N.O., Ruths, J., Candell, R., and Sandberg, H. (2016). *Survey and new directions for physics-based attack detection in control systems*. US Department of Commerce, National Institute of Standards and Technology.

Zhao, J. and Dörfler, F. (2015). Distributed control and optimization in DC microgrids. *Automatica*, 61, 18–26.