A Dynamic Strategy for Cyber-Attack Detection in Large-scale Power Systems via Output Clustering

Ana Jevtić and Marija Ilić

Abstract—In this paper we are concerned with reliable operation of the electric power grid in presence of malicious cyber-attacks on measurement signals. We use the continuously changing operating conditions of the power systems to introduce an active defense method based on dynamic clustering. Our detection strategy uses a moving-target approach where information about the system’s varying operating point is first used to form dynamic clusters of measurements based on their dynamic response to disturbances. Then, similarity checks can be performed within each cluster to detect stealthy cyber-attacks. The proposed method is effective even when the attacker has extensive knowledge of the system parameters, model and detection policy at some point in time.

I. INTRODUCTION

In the recent years, there have been a number of successful attacks on cyber-physical systems [1], [2], including the well-known attack on the Ukrainian power grid in 2015 [3], and the cyber incident that disrupted grid operation in the western US in March of 2019 [4]. These events have led to the increase in awareness of the problem of securing critical infrastructure, such as the power grid, transportation systems, gas and water networks, etc. The control systems behind these critical infrastructures have long been protected by physically isolating the local control and communication networks from insecure global networks such as the Internet. However, the increasing exposure of the electric power grid to the public via smart devices with Internet connectivity, and operation of the grid through corporate networks opens new doors to malicious intrusions.

Additionally, the attackers can be very resourceful, have detailed knowledge of the system, and therefore launch highly effective and deceptive attacks. Traditional passive detection, that heavily relies on the notion of observability, is proven to be ineffective against stealthy attacks that closely mimic normal system behavior. Replay attacks [5] and zero-dynamics attacks [6] are examples of how an adversary can exploit knowledge of the system to launch attacks that can evade detection. In other words, an attacker can design an undetectable attack, by injecting signals that are not observable in the system outputs [7].

However, the defender can assume an active role in cyber-security of the system, by designing a detection strategy in a way that altogether prevents the attacker from devising a stealthy attack. This approach is referred to as active detection. In contrast to passive detectors, active detectors [5], [8]–[10] perturb the system either through topology changes, or by injecting random signals into the network, in order to expose stealthy attacks. However, these approaches assume the attacker has limited knowledge of the detection strategy, and exploit the unknown information to ensure cyber-security.

In this paper, we don’t exclude the possibility of the attacker having complete knowledge of the system, as well as parameters and detection strategy at some point in time. Instead, we propose a dynamic strategy for cyber-attack detection, that takes advantage of natural variations in power system dynamics. The operating point of power systems, and therefore its dynamic behavior, varies in time since the generation has to continually change to satisfy the current demand. This unpredictable change in demand is viewed as a disturbance in the control system, and we show that it must be taken into account when designing attack detection filters. With that in mind, we design an active and dynamic cyber-attack detection policy to detect stealthy attacks. In our design, each incoming measurement is verified before being used as a feedback signal in any control process. We perform this verification in two stages. First, we cluster all the system measurements according to their dynamic response to natural load variations during normal operation. Next, we verify each incoming measurement by running a series of similarity checks with other measurements from the same cluster.

Notation: Let \( \mathcal{I}_k \) denote a set of integers, and \( |\mathcal{I}_k| \) its cardinality. Then, \( e_{\mathcal{I}_k} \in \mathbb{R}^{n \times |\mathcal{I}_k|} \) is a matrix composed of column vectors of the identity matrix \( I_n \in \mathbb{R}^{n \times n} \) corresponding to the index set \( \mathcal{I}_k \). We use \( \|M\|_F \) to denote the Frobenius norm of matrix \( M \), defined as \( \|M\|_F = \sqrt{\text{tr}(M^T M)} \). Furthermore, \( M \) is said to be Hurwitz if every eigenvalue of \( M \) has strictly negative real part. Matrix \( M \) is said to be semistable if the zero eigenvalues of \( M \) are semisimple, and all the other eigenvalues have a negative real part. Given a stable proper transfer function of a dynamical system \( g(s) \), \( \|g(s)\|_{\mathcal{H}_2} \) is the \( \mathcal{H}_2 \)-norm of the system.

The remainder of this paper is structured as follows. Section II contains the mathematical formulation of the problems considered in this paper. In Section III, we present the main contribution of this paper, the moving-target cyber-attack detection method. Section IV introduces relevant component models and demonstrates the efficiency of our proposed method through a numerical example on the IEEE RTS 24-bus power system. Finally, in Section V we give some concluding remarks.
II. PROBLEM FORMULATION

In this paper, we are concerned with detection of cyber-attacks that target measurements, without affecting the physical behavior of the power system directly, but through incorrect control actions. Figure 1 depicts the block diagram representation of the system we consider. The power system contains various components, such as generators, loads, transmission lines, etc. Two types of inputs enter the power system block: controlled inputs $u(t)$ enforced by the actuators, and uncontrolled inputs or disturbances $d(t)$ that represent changes in the environment that cannot be controlled. In our context, $d(t)$ is the power consumed by the loads, which directly influences control decisions and operation of the grid. It is important to note that this kind of disturbance does not act as noise in the system, since control feedback loop is designed to drive the actuators to balance these changes in load. Thus, we assume the physical system follows:

$$\dot{x}(t) = Ax(t) + B\hat{u}(t) + Gd(t)$$

(1)

where the states of system are denoted $x \in \mathbb{R}^n$, the disturbance is $d \in \mathbb{R}^m$, the control signal issued by the Control Center (CC) is $\hat{u} \in \mathbb{R}^p$.

A large network of field sensors is deployed to monitor the operation of the power system in (1). A malicious attacker can negatively impact the system by manipulating the measurements, which is represented with the added signal $y_a(t)$ in Figure 1:

$$\hat{y}(t) = y(t) + y_a(t)$$

(2)

$$\hat{y}(t) = Cx(t)$$

(3)

where $\hat{y} \in \mathbb{R}^l$ are measurements received by the CC. Thus, a potentially manipulated measurement signal $\hat{y}$ reaches the CC, which can then issue a potentially incorrect control signal $\hat{u}$ to the power system actuators:

$$\hat{u}(t) = K\hat{y} = K[y(t) + y_a(t)]$$

(4)

Finally, the attacked system can be rewritten in closed-loop as:

$$\Sigma_a : \begin{cases}
\dot{x}(t) = Ax(t) + By_a(t) + Gd(t) \\
\hat{y}(t) = Cx(t) + y_a(t)
\end{cases}$$

(5)

where $A = A + BKC$ is the closed-loop system matrix, and $B = BK$.

Note that both signals $d(t)$ and $y_a(t)$ are unknown. We first show that this kind of disturbance, that does not represent noise but physical changes in the system that are not predictable, can not be neglected when designing a cyber-attack detection filter. Then, we propose a clustering-based detection filter that can distinguish between disturbance $d(t)$ and the attack $y_a(t)$.

A. Dynamic Detection Filter Based on Linear Observers

In this section, we explain why dynamic attack detectors based on linear observers are not suitable for attack-detection in systems that are described by (1), unless the unknown disturbance $d(t)$ is taken into account in the observer design process. Under normal conditions (no cyber-attack injected into the system, $\hat{u}(t) = u(t)$), a linear observer is designed for the system in (1), to compute the state estimate $\hat{x}(t)$ from the received measurements:

$$\dot{\hat{x}}(t) = (A - LC)\hat{x}(t) + Ly(t)$$

$$\hat{y}(t) = C\hat{x}(t)$$

(6)

The error system can then be defined as $e(t) := \hat{x}(t) - x(t)$, and the corresponding residual $r_c(t) := \hat{y}(t) - y(t)$, with the error dynamics:

$$\dot{e}(t) = (A - LC)e(t) - Gd(t)$$

(7)

It is well known that, if $(A, C)$ is observable, $L$ can be chosen such that $(A - LC)$ is Hurwitz, making the error system exponentially stable [11] in absence of the disturbance $d(t)$, i.e. when $d(t) = 0$. In other words, an observer of this form will successfully estimate the state of the system under normal conditions only when the system does not experience a disturbance. This assumption is made frequently in the existing literature on cyber-attack detection, and any unknown disturbance $d(t)$ that enters the system is neglected. However, in presence of the disturbance $d(t) \neq 0$, both the error and the residual will converge to nonzero values, even when the system is attack-free. If we take into account the disturbance $d(t)$, and consider the attacked system, the error dynamics of the observer become:

$$\dot{e}(t) = (A - LC)e(t) + [(L - B)y_a(t) - Gd(t)]$$

$$r_c(t) = Ce(t) - y_a(t)$$

Note that both the disturbance $d(t)$ and attack signal $y_a(t)$ enter the dynamics of the error system and residual. Therefore, using a detection filter based on such observers to detect cyber-attacks on measurements may be ineffective for several reasons. First, the effect of the disturbance $d(t)$ in the error dynamics may frequently trigger false alarms. Second, it will be ineffective against stealthy attacks, since the signals $d(t)$ and $y_a(t)$ may not be distinguishable. In the rest of this paper, we will show how one can take advantage of the system’s structure to expose stealthy attacks in a dynamic fashion by designing a new residual and a corresponding detection algorithm.

Fig. 1: Block diagram of attacked power system in (5). The attacker injects signal $y_a$ into measurements $y$ in order to manipulate the system.

- Under certain assumptions on the disturbance, i.e. if it assumed to be white noise, an observer can be designed so that the residual converges to zero in expectation, i.e. $E[r_c(t)] = 0$. 

4232
B. Proposed Clustering-based Detection Filter

In order to be able to discriminate between effects of the signals \( d(t) \) and \( y_i(t) \), we will first consider the system in (5) in absence of attacks, i.e. for \( y_i(t) \equiv 0 \):

\[
\Sigma : \begin{align*}
\dot{x}(t) &= Ax(t) + Gd(t) \\
y(t) &= Cx(t)
\end{align*} \tag{8}
\]

These equations represent the normal behavior of the system. To quantify the effects of disturbances on measurement signals, we define the clusters \( \mathcal{I}_k \) as subsets of measurements that have a similar dynamic response to \( d(t) \). More specifically, measurements \( i, j \) belonging to the cluster \( \mathcal{I}_k \) are approximately proportional \( a_i y_i(t) \approx a_j y_j(t) \approx \cdots \approx z^{(k)}(t) \), where \( a_i, a_j, \ldots \) are some constant coefficients. This relation of measurements within the same cluster can also be written as:

\[
y^{(k)}(t) = \begin{bmatrix} p_1^{(k)} & \ldots & p_{|\mathcal{I}_k|}^{(k)} \end{bmatrix}^T z^{(k)}(t) \tag{9a}
\]

such that \( ||y^{(k)}(t) - \hat{y}^{(k)}(t)|| \leq \theta, \quad \theta \geq 0 \) \tag{9b}

where \( p_i^{(k)} = a_i^{-1} \) for all \( i \in \mathcal{I}_k \), and \( y^{(k)} \) is a subset of measurements \( y \) belonging to cluster \( \mathcal{I}_k \), \( y^{(k)} = (e_{\mathcal{I}_k})^T y \), and all its elements are approximately proportional to a single scalar variable \( z^{(k)} \). Parameter \( \theta \) is a design parameter, which will be discussed in later sections. Thus, we use terms cluster and clustering in the sense of model order reduction method in this work. We aim to estimate the full system state not only based on the received measurements and full system model, but also using the reduced order model which contains and encodes more detailed information (e.g. as in (9a)) that can be used to detect stealthy attacks. Specifically, we use the knowledge of the fact that incoming measurements \( \hat{y}^{(k)}(t) = y^{(k)}(t) + y_i^{(k)}(t) \) belonging to cluster \( k \) also have the property in (9a) only if \( y_i^{(k)} \equiv 0 \). Therefore, we define the set of residuals \( r_{i,j}(t) \) that exploit this property as:

\[
r_{i,j}(t) = ||p_j^{(k)} y_i(t) - p_i^{(k)} \hat{y}_j(t)||, \quad \forall i, j \in \mathcal{I}_k \tag{10}
\]

The residuals \( r_{i,j} \) defined above will have a value larger than some threshold \( \varepsilon \) only in presence of cyber-attacks. The appropriate choice for the threshold \( \varepsilon \) will be discussed in later sections. In the following sections, we provide a method for finding the boundaries of clusters used for attack detection and design a moving-target cyber-attack detection algorithm.

III. CYBER-ATTACK DETECTION METHODOLOGY

In order to address the problem of detecting cyber-attacks on measurements in a large power system network, we propose a methodology that consists of two steps. In the first step, we group the state measurements \( y(t) \) into clusters, based on similarity of their dynamic response to input \( d(t) \) in normal operating conditions. Then, in step two, we identify the attacked measurements through a series of consistency checks within each cluster. This procedure can be written compactly as:

- group the indices \( i \) of all measurements \( y_i(t) \) with similar dynamic responses into index sets \( \mathcal{I}_k, \quad k \in \{1, \ldots, K\} \)
- for a given threshold \( \varepsilon > 0 \), check all incoming measurements \( \hat{y}_i(t) \) for consistency. If there exist \( i, j \in \mathcal{I}_k \) such that following condition is satisfied

\[
r_{i,j}(t) = ||p_j^{(k)} y_i(t) - p_i^{(k)} \hat{y}_j(t)|| \geq \varepsilon \tag{11}
\]

flag measurements \( \hat{y}_i \) and \( \hat{y}_j \) as attacked measurements.

In the following subsections, we will introduce each of these steps in detail.

A. Measurement clustering

In this section, we introduce an aggregation method for clustering measurements according to their dynamic response to the external input \( d(t) \), inspired by [12], [13]. We begin with the definition of a cluster.

**Definition.** Let \( \mathbb{L} = \{1, \ldots, l\} \) be the set of measurement indices, and \( \mathbb{K} = \{1, \ldots, K\} \) the set of cluster indices. Then, clusters \( \mathcal{I}_k, \quad k \in \mathbb{K} \), are defined as disjoint subsets of \( \mathbb{L} \) that cover all the elements in \( \mathbb{L} \). i.e., \( \bigcup_{k \in \mathbb{K}} \mathcal{I}_k = \mathbb{L} \).

With this definition in mind, we aim to partition the set \( \mathbb{L} \) into clusters \( \mathcal{I}_k \) such that

\[
p_j^{(k)} g_i(s) = p_i^{(k)} g_j(s), \quad \forall i, j \in \mathcal{I}_k \tag{12}
\]

where \( g_i \) is the \( i \)-th element of \( g(s) = (sI_n - \mathcal{A})^{-1}G \), an input-output transfer matrix of the system in (8). We can rewrite this condition for cluster formation in a more compact way as follows.

**Definition.** A set of measurements \( \{y_i\} \) should form a cluster \( \mathcal{I}_k \) if there exists a scalar function \( \hat{g}(s) \) such that:

\[
(e_{\mathcal{I}_k}^n)^T g(s) = p^{(k)} g^{(k)} \hat{g}(s) \tag{13}
\]

This definition provides intuition on the meaning of clustering in our application, but is not practical for designing a procedure that would form such clusters, as it is impractical to perform similarity checks on functions. To get around this problem, we will derive an equivalent matrix-based condition equivalent to (13), based on the notion of reachability. To that end, we first derive reachability Gramian of a semistable system (8). The reachability Gramian is defined as [11]

\[
W_c = \int_0^\infty e^{A t} G G^T e^{A^T t} dt \tag{14}
\]

When \( \mathcal{A} \) is Hurwitz, the above integral converges, and \( W_c \) can also be found as a solution of the Lyapunov equation \( \mathcal{A} W_c + W_c \mathcal{A}^T + GG^T = 0 \). However, in power systems, the system matrix \( \mathcal{A} \) has an inherent structural singularity, as a direct consequence of power conservation law. Due to semistability of the system matrix \( \mathcal{A} \), the integral in (14) may not converge. To compute the reachability Gramian of a semistable system, we first consider the decomposition of \( \mathcal{A} \) where 0 = \( \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \)

\[
\mathcal{A} = U \Lambda U^{-1} = [u_{\max} \bar{U}] \begin{bmatrix} 0 & \bar{A} \\ \bar{U}^T \end{bmatrix}
\]

4233
where \( u_{\text{max}} \) and \( v_{\text{max}} \) are the right and left eigenvectors corresponding to the largest eigenvalue \( \lambda_1 = 0 \), and \( \bar{\Lambda} \) is diagonal and Hurwitz. Let \( \bar{\Lambda} = \bar{V}' \bar{A} \bar{G} \) and \( \bar{G} = \bar{V}' G \), defined as the stable subspace of \( \Sigma \). Then, the reachability Gramian of the stable subspace is the solution of \( \bar{\Lambda} \bar{W}_c + \bar{W}_c \bar{A}' + \bar{G} \bar{G}' = 0 \). Substituting \( \bar{\Lambda} \) and \( \bar{G} \) into (14) yields

\[
W_c = \bar{V} \bar{W}_c \bar{V}'
\]

(15)

where \( W_c \) is the reachability Gramian of the semistable system \( \Sigma \) and contains information on the degree of reachability of states with respect to the input \( d(t) \). In the following theorem we find the condition equivalent to (13) using the reachability Gramian \( W_c \) of the semistable system \( \Sigma \).

**Theorem.** Consider the reachability Gramian \( W_c \) in (15) of the semistable system \( \Sigma \) in (8). Furthermore, let the Cholesky factorization of \( W_c \) be given by \( W_c = W_L W_L' \), and \( \Phi = C W_L \). Then, the condition in (13) is equivalent to

\[
(\bar{e}_s T) \Phi = p(k) T \phi
\]

(16)

where \( \phi \in \mathbb{R}^{1 \times n} \) is a constant vector.

**Proof.** Omitted due to space limitation.

However, in real systems, the identity in (13) is almost never the case. Therefore, we relax the strict equality, and require

\[
\| p_j(k) g_i(s) - p_i(k) g_j(s) \|_{\mathcal{H}_2} \leq \varepsilon, \quad \forall i, j \in \mathcal{I}_k
\]

(17)

to hold for each cluster. Equivalently, we can check for linear dependence between rows of matrix \( \Phi \):

\[
\| p_i \Phi_i - p_j \Phi_j \| \leq \theta \quad \forall i, j \in \mathcal{I}_k
\]

(18)

where \( \theta > 0 \) and \( \phi_i \) is the \( i \)-th row of \( \Phi \). Here, \( \theta \) is a parameter that allows us to control the coarseness of clustering. In other words, it allows us to find outputs that have a “similar”, instead of equal, response, which relaxes the condition (13). However, the choice of \( \theta \) is not trivial, as it introduces a trade-off between accuracy of the approximation and size of clusters. Ideally, \( \theta \) should be chosen as a smallest value for which each cluster contains at least two measurements.

Finally, the procedure for finding the cluster boundaries of the set of measurements is summarized in the algorithm below.

### Algorithm 1: Clustering algorithm

- **Initialize** cluster index \( k = 0 \), and cluster set \( \mathcal{K} = \emptyset \)
- **Repeat**
  - Choose measurement index \( i \in \mathcal{L} \) that hasn’t been assigned to a cluster yet, and add it to cluster \( \mathcal{I}_{k+1} \)
  - Set \( k \leftarrow k + 1 \), \( \mathcal{K} \leftarrow \{ \mathcal{K}, k \} \)
  - Find all \( j \in \mathcal{L} \) that haven’t been assigned to a cluster yet and that satisfy (18) and add them to \( \mathcal{I}_{k+1} \)
  - until all measurements are assigned to a cluster, i.e. \( \bigcup_{k \in \mathcal{K}} \mathcal{I}_k = \mathcal{L} \)

Next, we introduce the moving-target cyber-attack detection method, based on this clustering procedure.

### B. Detection algorithm

In this section, we will introduce our cyber-attack detection algorithm that employs clusters of measurement found using Algorithm 1. Two properties of this clustering method are key in our cyber-attack detection filter design. Firstly, we know that, once clustering is performed on the system in normal operating conditions, the outputs within the clusters will be approximately proportional to each other at all times. That enables us to perform quick consistency checks to ensure the safety and reliability of the system. Secondly, the result of clustering will change over time as operating conditions change, which means our proposed detection filter will behave as a moving-target.

In the analysis in Section III-A, we have shown that clusters can be formed such that measurements \( i, j \) within the cluster \( \mathcal{I}_k \) are approximately proportional, i.e. \( a_i y_i(t) \approx \cdots \approx z^{(k)}(t) \). Then, we derived a matrix-based condition to find such clusters. Next, we show that the original system outputs \( y = C x \) can be approximated by \( \hat{y} = \Pi T z \), where \( z = \Pi y = \Pi C x \). The clustering matrix \( \Pi \in \mathbb{R}^{K \times n} \) is defined as:

\[
\Pi := \text{Diag}(p_1, p_2, \ldots, p_K) E \in \mathbb{R}^{K \times n}
\]

(19)

where \( p_i \) are row vectors of clustering coefficients, and \( E \) is a permutation matrix defined as \( E = [e_1, 1, \ldots, e_n] \). The input-output transfer matrix associated with \( \hat{y} \) can then be defined as \( \hat{y}(s) = \Pi T \Pi (sI_n - A)^{-1} \). The following theorem establishes that \( \hat{y} \) is a good approximation of \( y \), and that it can be used in our cyber-attack detection methodology.

**Theorem.** Consider a semistable linear system in (8). Consider also a clustering-based approximation \( \hat{y} \) obtained using the aggregation matrix \( \Pi \). Then, the error system of the approximation \( g_c(s) = g(s) - \hat{y}(s) \) is asymptotically stable.

**Proof.** Omitted due to space limitation.

Now, we introduce the moving-target detection algorithm based on dynamic clustering of system outputs. Firstly, the moving-target nature of our proposed method stems from the natural fluctuations occurring in power systems. As the underlying power system model is nonlinear, the system matrices \( A \) and \( G \) are only valid around a certain operating point \( x^0 \), and we will denote them with \( \mathcal{A}(x^0) \) and \( \mathcal{G}(x^0) \). As a result, cluster boundaries have to be recomputed approximately every one hour, for the current operating conditions \( x^0 \). In this method, additional computations are traded for increased difficulty for the attacker. In other words, even if we assume the attacker has complete knowledge of the system and the detection strategy at one point in time, that knowledge will eventually become unusable to construct stealthy attacks, as our detection strategy is dynamic.

Once the new operating point is received from the Control Center, the linearized system matrices and clusters need to be recomputed. Then, at every time step of the control
processes, the incoming measurements must be verified using residuals of the clustering-based detection filter given in (10) before the control action is computed and performed. The proposed cyber-attack detection algorithm, based on dynamic clustering, is outlined in Algorithm 2.

Algorithm 2: Clustering-based Cyber-attack Detection Method

```
repeat
  Get new operating conditions $x^0$.
  Compute matrices $A(x^0), G(x^0)$, and find cluster sets $I_k$ according to Algorithm 1.
  for every time-step of the control process do
    for all measurements $y_i, i \in I_k$ do
      if condition in (11) is satisfied then
        → cyber-attack detected
      end if
    end for
  end for
until new operating conditions $x^0$ are received.
```

IV. Numerical Examples

We begin this section by providing necessary power system component models, and deriving the standard state space model of the interconnected power system in form given in (8). Then, we demonstrate the performance of our method on the IEEE RTS 24-bus system.

A. Power system modeling

In this section, we introduce the power system component models used to derive the system matrices in (8). We model the loads as dynamic using the structure-preserving load model [14], alongside the well-known generator model with governor control [15]. Since the load side is more exposed to the public through their Internet connectivity [16] (through devices such as smart meters, electric vehicles and other smart appliances), modeling loads as dynamical components when studying the impact of cyber-attacks on power systems becomes highly necessary.

We consider a power system with $n_G$ generators and $n_L$ loads, and denote the set of generator buses by $G$, and the set of load buses by $L$. The mechanical dynamics of generators with governor control and aggregate loads at the substation level are given by:

\[
\begin{align*}
J_i \omega_i + D_i \dot{\omega}_i &= P_{T,i} - P_i + e_{T,i} a_i, & i \in G \\
J_i \dot{\omega}_i + D_i \omega_i &= -P_i - L_i, & i \in L \\
T_{u,i} \dot{P}_{T,i} &= -P_{T,i} + K_{T,i} a_i, & i \in G \\
T_{g,i} \dot{a}_i &= -r_i a_i - (\omega_i - \omega^{ref}), & i \in G
\end{align*}
\]

(20)

For each bus $i$, state variable $\omega_i$ denotes its frequency, $P_i$ the net real power injected into the network, and parameters $J_i, D_i$ the inertia and damping. At load buses $i \in L$, $L_i$ is defined as actual mechanical power consumed by the load. At generator buses $i \in G$, there are additional controller dynamics, namely, governor dynamics. States $P_{T,i}$ and $a_i$ denote the mechanical power of the generator and the turbine valve position respectively, and $e_{T,i}$ is a parameter of the turbine. The governor’s and turbine’s time constants are denoted by $T_{u,i}$ and $T_{g,i}$ while $K_{T,i}$ and $r_i$ are control gains. Finally, $\omega^{ref}$ is the frequency reference provided by the higher control layer. In order to derive the interconnected system, we treat $P_i$ as a coupling state variable whose dynamics can be obtained by differentiating the DC power flow equation, and expressed in matrix form as:

\[
\begin{bmatrix}
\dot{P}_G \\
\dot{P}_L
\end{bmatrix} = Y_{bus} \omega, \quad \text{where } Y_{bus} = \begin{bmatrix} Y_{GG} & Y_{GL} \\
Y_{LG} & Y_{LL} \end{bmatrix}
\]

(21)

where $P_G := [P_{i}]_{i \in G}$ and $P_L := [P_{i}]_{i \in L}$, and $Y_{bus}$ is the admittance matrix of a lossless transmission network. Further, $\omega$ can also be partitioned as $\omega := [\omega_G \ \omega_L]^T$ where $\omega_G := [\omega_{i}]_{i \in G}$ and $\omega_L := [\omega_{i}]_{i \in L}$. Finally, the linearized closed-loop power system model described by equations (20) and (21) can be expressed in form (8), with the state vector $x := [\omega_G, \omega_L, P_G, P_L, a_i]^T$, and $d = [L_i]_{i \in L}$.

In the following sections, we present numerical simulation examples performed on the IEEE RTS 24-bus power system to illustrate the performance of our proposed cyber-attack detection method.

B. Test system and illustrative scenarios

The IEEE RTS 24-bus system [17] consists of 10 generators equipped with governor control and 14 loads. The interconnected system is modeled using equations (20)-(21), where the dimension of $x$ is 68. For this system, we first consider two scenarios with different loading conditions, to demonstrate how boundaries of clusters change when operating conditions vary. In both scenarios, from time $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ s loading returns to nominal value. However, in Scenario 1, the system is at high loading condition, and in Scenario 2, the system is at low loading condition. The net real power injection of generator 8, $P_{G8}$, and its mechanical power output $P_{T8}$ are clustered with respective states of generators 3 and 10 under Scenario 1, and generators 4 and 5 under Scenario 2. In these experiments, the clustering procedure resulted in 21 clusters under Scenario 1, and 23 under Scenario 2.

C. Implementation of the attack detection methodology and result analysis

We consider Scenario 3 in the remainder of this section to demonstrate attack-detection capabilities of our proposed method:

- Scenario 3: the system is at high loading condition. From time $t = 0$ to 20 s, the loading is nominal. At time $t = 20$ s, load at bus 3 increases by 0.5 p.u., and at time $t = 200$ s loading returns to nominal value; at time $t = 125$ s, a sequence of 6 scaling attacks are launched on the measurement $P_{G8}$, each lasting 5 seconds, with total duration of the attack $T_a = 55$ s.

where a scaling attack can be represented in terms of system in (3) as $\tilde{y} = y + ya$, where $ya = k \cdot y$, and $k = 0.1$. 4235
We use Scenario 3 to demonstrate the performance of our detection filter. In Figure 2a we compare the attacked measurement $P_{G,8}$ (middle plot) with other measurements belonging to the same cluster (top plot), to obtain the residual in the bottom plot, which only crosses the chosen detection threshold for each of the attacks. In Figure 2b we show cluster measurements (top) and residuals (bottom) in absence of cyber-attacks. Note that the residual does not cross the detection threshold even when system conditions change, i.e., when there is a disturbance in the system.

**Fig. 2:** Group of measurements belonging to one of the clusters under Scenario 3.

**V. CONCLUSION**

In this paper, we presented a moving-target defense algorithm that is based on dynamic clustering to detect false data injection attacks on power system measurements. Our strategy is dynamic and uses information about the system’s changing operating point to define clusters of measurements that have similar dynamic response and then carries out detection through similarity checks on the measurements belonging to the same cluster. We numerically showed the performance of our proposed detection algorithm and its ability to successfully detect cyber-attacks through an example on the IEEE 24-bus power system.

**REFERENCES**

[1] Nicolas Falliere, Liam O Murchu, and Eric Chien. W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*, 5(6):29, 2011.

[2] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *International Conference on Critical Infrastructure Protection*, pages 73–82. Springer, 2007.

[3] Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. 2016.

[4] Blake Sobczak. ‘Cyber event’ disrupted U.S. grid networks - DOE, *E & E News*, April 30, 2019.

[5] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 911–918. IEEE, 2009.

[6] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. Attack models and scenarios for networked control systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, pages 55–64. ACM, 2012.

[7] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control*, 58(11):2715–2729, 2013.

[8] Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2017.

[9] André Teixeira, Iman Shames, Henrik Sandberg, and Karl H Johansson. Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1806–1813. IEEE, 2012.

[10] Taouba Rhouma, Jean-Yves Keller, Karim Chaib, Dominique Sauter, and Mohamed Naceur Abdelkrim. Coding control signals and switching lqg controller for secure fault-tolerant control against stealthy false data injection. In *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, pages 750–755. IEEE, 2016.

[11] Chi-Tsiong Chen. *Linear System Theory and Design*. Oxford University Press, Inc., New York, NY, USA, 3rd edition, 1998.

[12] Takayuki Ishizaki, Kenji Kashima, Antoine Girard, Jun ichi Imura, Luonan Chen, and Kazuyuki Aihara. Clustered model reduction of positive directed networks. *Automatica*, 59:238 – 247, 2015.

[13] Takayuki Ishizaki, Kenji Kashima, Jun ichi Imura, and Kazuyuki Aihara. Model order reduction for mimo linear dynamical networks via reaction-diffusion transformation, pages 5019–5024, 2011.

[14] Arthur R. Bergen and David J. Hill. Structure preserving model for power system stability analysis. *IEEE Trans. Power Appar. Syst.*, 1981.

[15] Marija D. Ilic and John Zaborszky. *Dynamics and Control of Large Electric Power Systems*. John Wiley & Sons, Inc., 2000.

[16] Richard J. Campbell. Cybersecurity issues for the bulk power system. Technical report, Congressional Research Service Report, 2015.

[17] Christos Ordoudis, Pierre Pinson, Juan Miguel Morales González, and Marco Zugno. An updated version of the ieee rts 24-bus system for electricity market and power system operation studies. 2016.