2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Secure Distributed Observers for a Class of Linear Time Invariant Systems in the Presence of Byzantine Adversaries

Aritra Mitra and Shreyas Sundaram

***Abstract*— We study the problem of distributed state estimation of a linear time-invariant system by a network of nodes, some of which are subject to adversarial attacks. We develop a secure distributed estimation strategy subject to an *f-locally bounded Byzantine* adversary model, where a compromised node can arbitrarily deviate from the rules of any prescribed algorithm. Under such a threat model, we provide sufficient conditions guaranteeing the success of our estimation strategy. Our method relies on the construction of a subgraph, which we call a Mode Estimation Directed Acyclic Graph (MEDAG), for each unstable and marginally stable eigenvalue of the plant. We provide a distributed algorithm for constructing a MEDAG and characterize graph topologies for which the construction algorithm is guaranteed to succeed. Our approach provides fundamental insights into the relationship between the dynamics of the system, the measurement structure of the nodes, and the underlying graph topology.**

I. Introduction

The distributed estimation problem consists of a dynamical system (or plant) together with a network of nodes (or observers) that each aim to estimate the state of the plant using local measurements and information exchanges with neighbors. This widely studied problem is broadly tackled using two main approaches, namely: Kalman-filter based techniques [1]–[3], and LTI observer based techniques [4]–[6]. For example, in [4], the authors propose a scalar-gain estimator which requires the system dynamics and the network topology to jointly satisfy certain conditions. Alternate approaches that work under the broadest observability assumptions were proposed in [5]–[7].

Recently, distributed algorithms for various applications such as consensus [8]–[10], broadcast [11], optimization [12], [13] and fault detection [14] have been investigated from the perspective of security. The main challenge in such problems is to come up with strategies which are guaranteed to work (subject to certain conditions on the underlying graph topology) under carefully crafted adversarial attacks. There is a very limited literature dealing with secure distributed state estimation (in the context that we are considering here). In [15] and [16], the authors employ a metric known as 'belief divergence', which provides a measure of how much a received state estimate deviates from the other received estimates in the neighborhood of a given node. Based on this metric, the concerned node assigns 'trust' values to each of its neighbors. A similar scheme is used in [17]. However, these works do not provide formal proofs of convergence, nor impose any conditions on the graph topology which guarantee success of their schemes.

In this work, we provide a secure distributed state estimation algorithm with provable guarantees. In particular, we develop a secure state estimation scheme where each regular (non-adversarial) node requires knowledge of only the plant dynamics, its neighborhood, and an upper bound on the number of adversaries in its neighborhood. On the other hand, an adversarial node is endowed with complete knowledge of the system model (network topology and plant dynamics) and is allowed to deviate arbitrarily from the prescribed algorithm. We provide a sufficient condition on the graph topology which guarantees the success of our proposed technique.

The authors are with the School of Electrical and Computer Engineering at Purdue University. Email: {mitra14,sundara2}@purdue.edu

II. System Model

A. Notation

A directed graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \cdots, N\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the edges. An edge from node j to node i, denoted by (j, i), implies that node j can transmit information to node i. The neighborhood of the i-th node is defined as $\mathcal{N}_i \triangleq \{j \,|\, (j,i) \in \mathcal{E}\}$. A node i is said to be an out-going neighbor of node j if $(j,i) \in \mathcal{E}$. The notation $|\mathcal{V}|$ is used to denote the cardinality of a set $\mathcal{V}$. Throughout the rest of this paper, we use the terms 'nodes' and 'observers' interchangeably.

The set of all eigenvalues (modes) of a matrix $\mathbf{A}$ is denoted by $sp(\mathbf{A}) = \{\lambda \in \mathbb{C} \,|\, det(\mathbf{A} - \lambda \mathbf{I}) = 0\}$ and the set of all marginally stable and unstable eigenvalues of $\mathbf{A}$ is denoted by $\Lambda_U(\mathbf{A}) = \{\lambda \in sp(\mathbf{A}) \,||\lambda| \geq 1\}$. We use $a_{\mathbf{A}}(\lambda)$ and $g_{\mathbf{A}}(\lambda)$ to denote the algebraic and geometric multiplicities, respectively, of an eigenvalue $\lambda \in sp(\mathbf{A})$. An eigenvalue λ is said to be simple if $a_{\mathbf{A}}(\lambda) = g_{\mathbf{A}}(\lambda) = 1$.

B. Problem Formulation

Consider the linear dynamical system

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k], \tag{1}$$

where $k \in \mathbb{N}$ is the discrete-time index, $\mathbf{x}[k] \in \mathbb{R}^n$ is the state vector and $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the system matrix. The system is monitored by a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of N nodes. The i-th node has a measurement of the state, given by

$$\mathbf{y}_i[k] = \mathbf{C}_i \mathbf{x}[k], \tag{2}$$

where $\mathbf{y}_i[k] \in \mathbb{R}^{r_i}$ and $\mathbf{C}_i \in \mathbb{R}^{r_i \times n}$. We denote $\mathbf{y}[k] = \begin{bmatrix}\mathbf{y}_1^T[k] & \cdots & \mathbf{y}_N^T[k]\end{bmatrix}^T$, and $\mathbf{C} = \begin{bmatrix}\mathbf{C}_1^T & \cdots & \mathbf{C}_N^T\end{bmatrix}^T$.

Each node is tasked with estimating the entire system state $\mathbf{x}[k]$. In particular, let $\hat{\mathbf{x}}_i[k]$ denote the state estimate

978-1-5090-1837-6/16/$31.00 ©2016 IEEE

of node i, which it updates at each time-step k based on information received from its neighbors and its local measurements (if any). We refer to the network of nodes maintaining and updating these estimates as a *distributed observer*. In accordance with the terminology established in [5], [6], consider the following definition.

Definition 1: (**Omniscience**) A distributed observer achieves *omniscience* if $\lim_{k\to\infty} ||\hat{\mathbf{x}}_i[k] - \mathbf{x}[k]|| = 0, \forall i \in \{1, \cdots, N\}$, i.e., the state estimate maintained by each node asymptotically converges to the true state of the plant. □

In this paper, we allow for the possibility that certain nodes in the network are compromised by an adversary, and *do not* follow their prescribed state estimate update rule. We will use the following adversary model in this paper.

Adversary Model: The set of nodes $\mathcal{V}$ is partitioned into two subsets: $\mathcal{R}$ comprising of *regular nodes*, and $\mathcal{A} = \mathcal{V} \setminus \mathcal{R}$ comprising of *adversarial nodes*. We consider the *Byzantine fault model* where an adversarial node can arbitrarily deviate from the rules of any prescribed algorithm, and can transmit different state estimates to different neighbors at the same time step [18]. In addition, the adversarial nodes possess complete knowledge about the graph topology and the plant dynamics, i.e., an adversarial node knows the measurements of the normal nodes at every time step. We endow such privileges to the adversaries with the aim of providing resilience to worst-case (potentially coordinated) behavior.

In the literature dealing with distributed fault-tolerant algorithms, it is a common assumption to assign an upper bound f to the total number of adversarial nodes in the network. This is known as the *f-total adversarial model*. However, to allow for a large number of adversaries in large scale networks, we will consider a *locally bounded* fault model, taken from [19], [20], defined as follows.

Definition 2: (**f-local set**) A set $\mathcal{C} \subset \mathcal{V}$ is *f-local* if it contains at most f nodes in the neighborhood of the other nodes, i.e., $|\mathcal{N}_i \cap \mathcal{C}| \leq f$, $\forall i \in \mathcal{V} \setminus \mathcal{C}$. □

Definition 3: (**f-local adversarial model**) A set $\mathcal{A}$ of adversarial nodes is *f-locally bounded* if $\mathcal{A}$ is an f-local set. □

We study the following problem in this paper.

Problem 1: **(Secure Omniscience Achieving Problem)** Given a system (1), a set of nodes interconnected by a graph $\mathcal{G}$, and an observation model at each node given by (2), formulate a state estimation scheme so that $\lim_{k\to\infty} ||\hat{\mathbf{x}}_i[k] - \mathbf{x}[k]|| = 0, \forall i \in \mathcal{R}$, *regardless* of the actions of any f-locally bounded set of Byzantine adversaries.

III. Secure Distributed Estimation

In order to establish the key ideas for dealing with adversarial behavior while minimizing notational complexity, we make the following assumption in the rest of the paper.

Assumption 1: $\mathbf{A}$ has real and simple eigenvalues.

A direct consequence of having simple eigenvalues is that we can diagonalize $\mathbf{A}$ by using the coordinate transformation matrix $\mathbf{V} = \begin{bmatrix}\mathbf{v}^{(1)} & \cdots & \mathbf{v}^{(n)}\end{bmatrix}$, where $\mathbf{v}^{(1)}, \cdots, \mathbf{v}^{(n)}$ are n linearly independent eigenvectors of $\mathbf{A}$. With $\mathbf{z}[k] = \mathbf{V}^{-1}\mathbf{x}[k]$, the dynamics (1) are transformed into the form[1]

$$\begin{aligned} \mathbf{z}[k+1] &= \mathbf{M}\mathbf{z}[k] \\ \mathbf{y}_i[k] &= \bar{\mathbf{C}}_i\mathbf{z}[k], \quad \forall i \in \{1, \cdots, N\} \end{aligned} \tag{3}$$

where $\mathbf{M} = \mathbf{V^{-1}AV}$ is a diagonal matrix, and $\bar{\mathbf{C}}_i = \mathbf{C}_i\mathbf{V}$. We denote the eigenvalues of $\mathbf{M}$ by $\lambda_1, \cdots, \lambda_n$. For each node i, we denote the detectable and undetectable eigenvalues by the sets $\mathcal{O}_i$ and $\mathcal{UO}_i$, respectively. We define $\rho_i = |\mathcal{O}_i|$. Next, we introduce the notion of *source nodes*.

Definition 4: (**Source nodes**) For each $\lambda_j \in \Lambda_U(\mathbf{M})$, the set of nodes that can detect λ_j is denoted by $\mathcal{S}_j$, and called the set of *source nodes* for λ_j. □

We develop an estimation scheme which enables each regular node to estimate $\mathbf{z}[k]$ (from which they can obtain $\mathbf{x}[k] = \mathbf{V}\mathbf{z}[k]$). For each $\lambda_j \in \Lambda_U(\mathbf{M})$, our estimation scheme relies on separate strategies for nodes in $\mathcal{S}_j$ and $\mathcal{V} \setminus \mathcal{S}_j$. In particular, each node in $\mathcal{S}_j$ employs a Luenberger observer for estimating $z_j[k]$ (the component of $\mathbf{z}[k]$ corresponding to the eigenvalue λ_j), while the nodes in $\mathcal{V} \setminus \mathcal{S}_j$ rely on a secure consensus algorithm for asymptotically estimating that state.[2] Next, we discuss these ideas in detail.

A. Design of Luenberger Observers

Consider a regular node i. Let $\mathcal{O}_i = \{\lambda_{n_1}, \lambda_{n_2}, \cdots, \lambda_{n_{\rho_i}}\}$ (recall $\rho_i = |\mathcal{O}_i|$) be the set of detectable eigenvalues for node i. Define $\mathbf{J}_i \triangleq diag(\lambda_{n_1}, \lambda_{n_2}, \cdots, \lambda_{n_{\rho_i}})$ and $\mathbf{z}_{\mathcal{O}_i}[k] \triangleq \left[z_{n_1}[k], z_{n_2}[k], \cdots, z_{n_{\rho_i}}[k]\right]^T$. Let $\bar{\mathbf{c}}^i_{n_j}$ denote the column of $\bar{\mathbf{C}}_i$ corresponding to the eigenvalue $\lambda_{n_j} \in \mathcal{O}_i$. Define the observation matrix $\bar{\mathbf{C}}_{\mathcal{O}_i} \in \mathbb{R}^{r_i \times \rho_i}$ as $\bar{\mathbf{C}}_{\mathcal{O}_i} \triangleq \left[\bar{\mathbf{c}}^i_{n_1}, \bar{\mathbf{c}}^i_{n_2}, \cdots, \bar{\mathbf{c}}^i_{n_{\rho_i}}\right]$. Let $\hat{z}^i_{n_j}[k]$ denote the i-th node's estimate of component $z_{n_j}[k]$ (corresponding to the eigenvalue λ_{n_j}) of the state vector $\mathbf{z}[k]$. Define $\hat{\mathbf{z}}_{\mathcal{O}_i}[k] \triangleq \left[\hat{z}^i_{n_1}[k], \hat{z}^i_{n_2}[k], \cdots, \hat{z}^i_{n_{\rho_i}}[k]\right]^T$. Note that under Assumption 1, if λ_j is not detectable, then its corresponding column in $\bar{\mathbf{C}}_i$ is a zero vector [21]. Consider the following Luenberger observer at node i:

$$\hat{\mathbf{z}}_{\mathcal{O}_i}[k+1] = \mathbf{J}_i\hat{\mathbf{z}}_{\mathcal{O}_i}[k] + \mathbf{L}_i\left(\mathbf{y}_i[k] - \bar{\mathbf{C}}_{\mathcal{O}_i}\hat{\mathbf{z}}_{\mathcal{O}_i}[k]\right), \tag{4}$$

where $\mathbf{L}_i \in \mathbb{R}^{\rho_i \times r_i}$ is an observation gain matrix at node i. From the definitions of $\mathbf{J}_i$ and $\bar{\mathbf{C}}_{\mathcal{O}_i}$, it follows that the pair $(\mathbf{J}_i, \bar{\mathbf{C}}_{\mathcal{O}_i})$ is detectable. Thus, $\mathbf{L}_i$ can be chosen so that $(\mathbf{J}_i - \mathbf{L}_i\bar{\mathbf{C}}_{\mathcal{O}_i})$ is Schur stable, and $\lim_{k\to\infty} ||\hat{\mathbf{z}}_{\mathcal{O}_i}[k] - \mathbf{z}_{\mathcal{O}_i}[k]|| = 0$. This leads to the following straightforward result which we shall use in our subsequent development.

Lemma 1: Suppose Assumption 1 holds. Then, for each regular node $i \in \mathcal{R}$ and each $\lambda_j \in \mathcal{O}_i$, the observer given by equation (4) ensures that $\lim_{k\to\infty} |\hat{z}^i_j[k] - z_j[k]| = 0$. □

The above result shows that a node does not have to rely on information exchange with neighbors in order to estimate certain subsets of the state. In the following section, we

[1] As this only relies on the knowledge of the system matrix $\mathbf{A}$ (which is assumed to be known by all the nodes), all of the nodes can do this in a distributed manner (e.g., by using an agreed-upon convention for ordering the eigenvalues and corresponding eigenvectors).

[2] Our strategies will apply identically to each unstable and marginally stable eigenvalue of $\mathbf{M}$. Thus, we focus on a generic $\lambda_j \in \Lambda_U(\mathbf{M})$.

describe a secure consensus based strategy for estimating the portion of the state that is not locally detectable.

B. Consensus Based Secure State Estimate Update Rule

Consider an unstable (or marginally stable) eigenvalue $\lambda_j \in \mathcal{UO}_i$. For such an eigenvalue, node i has to rely on the information received from its neighbors (some of whom might be adversarial) in order to estimate $z_j[k]$. To this end, we propose a secure consensus algorithm that requires each regular node $i \in \mathcal{V} \setminus \mathcal{S}_j$ to update its estimate of $z_j[k]$ using the following two stage filtering strategy:

1) At each time-step k, each regular node i collects the state estimates of $z_j[k]$ received from *only* those neighbors which belong to a certain subset $\mathcal{N}_j^i \subseteq \mathcal{N}_i$ (to be defined later), and ranks them from largest to smallest.[3]
2) Node i removes the largest and smallest f estimates (i.e., removes $2f$ estimates in all), and updates its own state estimate using the following rule:[4]

$$\hat{z}_j^i[k+1] = \lambda_j \sum_{l \in \mathcal{M}_j^i[k]} w_{il}^j[k] \hat{z}_j^l[k], \tag{5}$$

where $\mathcal{M}_j^i[k] \subset \mathcal{N}_j^i (\subseteq \mathcal{N}_i)$ is the set of nodes from which node i chooses to accept estimates of $z_j[k]$ at time-step k, after removing the f largest and f smallest estimates from $\mathcal{N}_j^i$. Node i assigns the weight $w_{il}^j[k]$ to the l-th node at the k-th time-step for estimating $z_j[k]$. The weights are nonnegative and chosen to satisfy $\sum_{l \in \mathcal{M}_j^i[k]} w_{il}^j[k] = 1, \forall \lambda_j \in \mathcal{UO}_i$.

We refer to the above algorithm as the Local-Filtering based Secure Estimation (LFSE) algorithm. For implementing this algorithm, a regular node i needs to construct the set $\mathcal{N}_j^i$, $\forall \lambda_j \in \mathcal{UO}_i$, based on the relative positions of its neighbors (with respect to its own position) in the graph $\mathcal{G}$. We will provide the exact definition of $\mathcal{N}_j^i$, and a distributed algorithm for constructing such a set in the next sections.

Remark 1: The strategy of disregarding the most extreme values in one's neighborhood, and using a convex combination of the rest for performing linear updates, has been adopted for secure distributed consensus in [9], [10]. The secure distributed estimation problem can be thought of as a class of secure consensus problems, where the consensus value itself needs to follow a certain trajectory dictated by the given plant dynamics. □

Summary of the Secure Estimation Scheme: We briefly summarize the secure estimation scheme as follows.

1) All nodes perform the coordinate transformation $\mathbf{z}[k] = \mathbf{V}^{-1}\mathbf{x}[k]$; a regular node i identifies its detectable and undetectable eigenvalues ($\mathcal{O}_i$ and $\mathcal{UO}_i$).
2) Each regular node i uses the Luenberger observer (4) to estimate the states $\mathbf{z}_{\mathcal{O}_i}[k]$ corresponding to its detectable eigenvalues.
3) For each undetectable eigenvalue $\lambda_j \in \mathcal{UO}_i$, each regular node i follows the LFSE algorithm governed by equation (5) for updating $\hat{z}_j^i[k]$.

In the next section, we analyze the proposed secure estimation strategy.[5]

IV. Analysis of the Secure Distributed Estimation Strategy

In this section, we provide our main result concerning the asymptotic convergence of the state estimates of the regular nodes to the true state of the plant. To this end, we first introduce the following definition.

Definition 5: (**Mode Estimation Directed Acyclic Graph (MEDAG)**) For each eigenvalue $\lambda_j \in \Lambda_U(\mathbf{M})$, suppose there exists a spanning subgraph $\mathcal{G}_j = (\mathcal{V}, \mathcal{E}_j)$ of $\mathcal{G}$ with the following properties.

(i) If $i \in \{\mathcal{V} \setminus \mathcal{S}_j\} \cap \mathcal{R}$, then $|\mathcal{N}_j^i| \geq 2f+1$, where $\mathcal{N}_j^i = \{l | (l,i) \in \mathcal{E}_j\}$.[6]
(ii) There exists a partition of $\mathcal{R}$ into the sets $\{\mathcal{L}_0^j, \mathcal{L}_1^j, \cdots, \mathcal{L}_{T_j}^j\}$, where $\mathcal{L}_0^j = \mathcal{S}_j \cap \mathcal{R}$, and if $i \in \mathcal{L}_m^j$ (where $1 \leq m \leq T_j$), then $\mathcal{N}_j^i \cap \mathcal{R} \subseteq \bigcup_{r=0}^{m-1} \mathcal{L}_r^j$.

Then, we call $\mathcal{G}_j$ a *Mode Estimation Directed Acyclic Graph (MEDAG)* for $\lambda_j \in \Lambda_U(\mathbf{M})$. □

We say that a regular node $i \in \mathcal{L}_m^j$ belongs to level m. The implication of the first property of a MEDAG is that the set $\mathcal{M}_j^i[k]$ (recall that a regular node $i \in \mathcal{V} \setminus \mathcal{S}_j$ only uses estimates from $\mathcal{M}_j^i[k] \subset \mathcal{N}_j^i$ for updating $\hat{z}_j^i[k]$ at time-step k) is non-empty $\forall i \in \{\mathcal{V} \setminus \mathcal{S}_j\} \cap \mathcal{R}$. The second property implies that a regular node i in level m (where $1 \leq m \leq T_j$) accepts estimates from only those regular nodes which belong to levels 0 to $m-1$ (and which belong to $\mathcal{M}_j^i[k] \subset \mathcal{N}_j^i$).[7] We later prove that $\mathcal{G}_j$ contains no directed cycles consisting only of regular nodes. The significance of this acyclic structure will be apparent during the convergence analysis.

We now provide a lemma that will be required for proving our main result.

Lemma 2: Let Assumption 1 hold. Suppose that $\mathcal{G}$ contains a MEDAG $\mathcal{G}_j$ for each $\lambda_j \in \Lambda_U(\mathbf{M})$, and let $\mathcal{N}_j^i$ be the neighbors of node i in $\mathcal{G}_j$. Then, for each regular node $i \in \mathcal{R}$ and each $\lambda_j \in \mathcal{UO}_i$, the LFSE dynamics described by equation (5) ensures that $\lim_{k\to\infty} |\hat{z}_j^i[k] - z_j[k]| = 0$. □

Proof: As $\mathcal{G}$ contains a MEDAG for each $\lambda_j \in \Lambda_U(\mathbf{M})$, the sets $\{\mathcal{L}_0^j, \mathcal{L}_1^j, \cdots, \mathcal{L}_p^j, \cdots \mathcal{L}_{T_j}^j\}$ form a partition of the set $\mathcal{R}$. We prove by induction on the level number p. For $p = 0$, by definition of the set $\mathcal{L}_0^j$, all the regular nodes in $\mathcal{L}_0^j$ belong to the set $\mathcal{S}_j$, i.e., $\lambda_j \in \mathcal{O}_i$ for each regular node in $\mathcal{L}_0^j$. By Assumption 1 and Lemma 1, each regular node in level 0 can estimate $z_j[k]$ asymptotically. Notice that for any regular node i belonging to a level p, where

[3] If node i does not receive an estimate from some *adversarial* node $l \in \mathcal{A} \cap \mathcal{N}_j^i$ at a given time-step (every regular node in $\mathcal{N}_j^i$ will always transmit an estimate to node i at every time-step), it simply assigns a 0 value to node l's estimate (without loss of generality).

[4] Notice that in the update rule (5), a regular node $i \in \mathcal{V} \setminus \mathcal{S}_j$ does not use its own estimate value, i.e., it assigns a zero self-weight to itself.

[5] It should be noted that the estimation strategy described previously is developed for all nodes, as one does not know which nodes belong to the set $\mathcal{A}$. The adversarial nodes may or may not choose to follow the strategy.

[6] Given a regular node $i \in \mathcal{V} \setminus \mathcal{S}_j$, and an undetectable eigenvalue λ_j, the set $\mathcal{N}_j^i$ constructed specifically for λ_j is a certain subset of the original neighborhood of node i in $\mathcal{G}$.

[7] A regular node i also listens to adversarial nodes (if any) in $\mathcal{M}_j^i[k]$.

$1 \leq p \leq T_j$, we have $\lambda_j \in \mathcal{UO}_i$. Consider a regular node i in $\mathcal{L}_1^j$. We partition the set $\mathcal{N}_j^i$ into the sets $\mathcal{U}_j^i[k]$, $\mathcal{J}_j^i[k]$, and $\mathcal{M}_j^i[k]$, such that the sets $\mathcal{U}_j^i[k]$ and $\mathcal{J}_j^i[k]$ contain f nodes each, with the highest and lowest estimate values (for $z_j[k]$) respectively, transmitted to node i at time-step k, and $\mathcal{M}_j^i[k]$ contains the rest of the nodes in $\mathcal{N}_j^i$. According to the LFSE dynamics (5), node i only uses estimates from the set $\mathcal{M}_j^i[k]$ to update its own estimate $\hat{z}_j^i[k]$. Notice that based on the properties of a MEDAG, the set $\mathcal{M}_j^i[k]$ is non-empty. Let the error in estimation of $z_j[k]$ for node i be denoted by $e_j^i[k] \triangleq \hat{z}_j^i[k] - z_j[k]$. Subtracting $z_j[k+1]$ from both sides of equation (5), and noting that $z_j[k+1] = \lambda_j z_j[k]$ (based on the decoupled dynamics given by (3)), we obtain

$$e_j^i[k+1] = \lambda_j \sum_{l \in \mathcal{M}_j^i[k]} w_{il}^j[k] e_j^l[k], \tag{6}$$

where we used the fact that $\sum_{l \in \mathcal{M}_j^i[k]} w_{il}^j[k] = 1$. Now, consider the following two cases. (i) $\mathcal{M}_j^i[k] \cap \mathcal{A} = \emptyset$, *i.e., there are no adversarial nodes in the set* $\mathcal{M}_j^i[k]$: in this case, all the nodes in the set $\mathcal{M}_j^i[k]$ are regular and belong to $\mathcal{S}_j$ (as $\mathcal{N}_j^i \cap \mathcal{R} \subseteq \mathcal{L}_0^j = \mathcal{S}_j \cap \mathcal{R}$). (ii) $\mathcal{M}_j^i[k] \cap \mathcal{A}$ *is non-empty, i.e., there are some adversarial nodes in the set* $\mathcal{M}_j^i[k]$: based on the f-local adversarial model, it is apparent that each of the sets $\mathcal{U}_j^i[k]$ and $\mathcal{J}_j^i[k]$ contain at least one regular node. Let r and q be two such regular nodes belonging to $\mathcal{U}_j^i[k]$ and $\mathcal{J}_j^i[k]$ respectively. Based on the definitions of the sets $\mathcal{U}_j^i[k]$, $\mathcal{J}_j^i[k]$, and $\mathcal{M}_j^i[k]$, we have $\hat{z}_j^q[k] \leq \hat{z}_j^l[k] \leq \hat{z}_j^r[k]$, and hence $e_j^q[k] \leq e_j^l[k] \leq e_j^r[k]$, for any node $l \in \mathcal{M}_j^i[k]$. Thus, for any $l \in \mathcal{M}_j^i[k]$, we can express $e_j^l[k]$ as a convex combination of the errors $e_j^q[k]$ and $e_j^r[k]$. Analyzing each of the two cases, and referring to equation (6), we infer that at every time-step k, the estimation error $e_j^i[k+1]$ is expressible as a convex combination of the errors of regular nodes in level 0, i.e., regular nodes belonging to the set $\mathcal{S}_j$. Based on Lemma 1, we have that $\lim_{k\to\infty} e_j^l[k] = 0$, $\forall l \in \mathcal{S}_j \cap \mathcal{R}$. Thus, we conclude that $\hat{z}_j^i[k]$ converges asymptotically to $z_j[k]$ for any regular node i in $\mathcal{L}_1^j$. Next, suppose the result holds for all levels from 0 to p (where $1 \leq p \leq T_j - 1$). It is easy to see that the result holds for all regular nodes in $\mathcal{L}_{p+1}^j$ as well, by noting the following.

- A regular node $i \in \mathcal{L}_{p+1}^j$ has $\mathcal{N}_j^i \cap \mathcal{R} \subseteq \bigcup_{m=0}^p \mathcal{L}_m^j$.
- For each $i \in \mathcal{L}_{p+1}^j$, it holds that every estimate of $z_j[k]$ received (and used for state estimate update) is either from a regular node belonging to some level from 0 to p, or from an adversarial node. In the latter case, based on the LFSE dynamics, this value is sandwiched between the values of two regular nodes (belonging to some level from 0 to p). Since the values of regular nodes in levels $\bigcup_{m=0}^p \mathcal{L}_m^j$ asymptotically converge to $z_j[k]$ (based on our induction hypothesis), the value $\hat{z}_j^i[k]$ will also converge to $z_j[k]$ asymptotically.

Our argument was general and hence holds for any $\lambda_j \in \Lambda_U(\mathbf{M})$. We arrive at the conclusion that any node $i \in \mathcal{R}$ can asymptotically estimate $z_j[k]$ for any eigenvalue $\lambda_j \in \mathcal{UO}_i$. ■

We now state and prove our main result which provides sufficient conditions for achieving secure omniscience.

Theorem 1: Let Assumption 1 hold and suppose that the network $\mathcal{G}$ contains a MEDAG for each $\lambda_j \in \Lambda_U(\mathbf{M})$. Then, the distributed estimation strategy governed by the Luenberger observer based dynamics (4) and the LFSE dynamics (5) achieves secure omniscience under the f-locally bounded Byzantine adversary model. □

Proof: Assumption 1 implies a one-to-one correspondence between the eigenvalues of $\mathbf{M}$ and the components of the transformed state vector $\mathbf{z}[k]$. Accordingly, for each regular node i, the state vector $\mathbf{z}[k]$ can be partitioned into two components $\mathbf{z}_{\mathcal{O}_i}[k]$ and $\mathbf{z}_{\mathcal{UO}_i}[k]$, corresponding to the detectable and undetectable eigenvalues of node i, respectively. As Assumption 1 holds, Lemma 1 holds, and hence $\hat{\mathbf{z}}_{\mathcal{O}_i}[k]$ converges to $\mathbf{z}_{\mathcal{O}_i}[k]$ asymptotically. As Assumption 1 holds and a MEDAG exists for each $\lambda_j \in \mathcal{UO}_i$ (notice that $\mathcal{UO}_i \subseteq \Lambda_U(\mathbf{M})$), Lemma 2 also holds. Consequently, node i can asymptotically estimate each component of the state $\mathbf{z}[k]$ associated with eigenvalues in $\mathcal{UO}_i$, i.e., node i can estimate $\mathbf{z}_{\mathcal{UO}_i}[k]$ asymptotically. Combining these results, we conclude that node i can asymptotically estimate the entire state $\mathbf{z}[k]$ (and hence also $\mathbf{x}[k] = \mathbf{V}\mathbf{z}[k]$). ■

Having established that secure omniscience can be achieved by each of the regular nodes, we now present a distributed algorithm for constructing a MEDAG for each $\lambda_j \in \Lambda_U(\mathbf{M})$.

V. Distributed MEDAG Construction Algorithm

Recall that the filtering algorithm for secure consensus required a node $i \in \mathcal{V} \backslash \mathcal{S}_j$ to accept estimates from neighbors belonging to the set $\mathcal{M}_j^i[k]$, which was a subset of $\mathcal{N}_j^i$ (the neighbor set of node i in the MEDAG $\mathcal{G}_j$). In this section, we present a distributed algorithm (Algorithm 1) for constructing a MEDAG for each unstable and marginally stable eigenvalue $\lambda_j \in sp(\mathbf{M})$. The construction of these MEDAGs constitutes the initialization phase of our design, which can then be followed up by the estimation phase described earlier. Algorithm 1 requires every node $i \in \mathcal{R}$ to maintain a counter $c_j(i)$ and a list of indices $\mathcal{N}_j^i$ for each $\lambda_j \in \Lambda_U(\mathbf{M})$. The nodes in $\mathcal{N}_j^i \subseteq \mathcal{N}_i$ will be the parents of node i in the DAG constructed for the estimation of $z_j[k]$. We start with $c_j(i) = 0$ and $\mathcal{N}_j^i = \emptyset$ for each regular node. Each regular source node in $\mathcal{S}_j$ broadcasts a predefined (and arbitrary) message m_j to its out-going neighbors, sets $c_j(i) = 1$, maintains $\mathcal{N}_j^i = \emptyset$ for all future time, and goes to sleep. Each regular node $i \in \mathcal{V} \backslash \mathcal{S}_j$ waits until it has received m_j from at least $2f+1$ distinct neighbors, at which point it sets $c_j(i) = 1$, appends the labels of each of the neighbors from which it received m_j to $\mathcal{N}_j^i$, broadcasts m_j to its out-going neighbors, and goes to sleep. When the algorithm terminates, if we have $c_j(i) = 1$, $\forall i \in \mathcal{R}$, we say that the MEDAG construction algorithm "*terminates*" for λ_j. *The objective of the algorithm is to return* $\mathcal{N}_j^i$ *for every unstable and marginally stable eigenvalue* $\lambda_j \in sp(\mathbf{M})$*, and* $i \in \mathcal{R}$.

Notice that during the construction of a MEDAG for an eigenvalue $\lambda_j \in \Lambda_U(\mathbf{M})$, an adversarial node can misbehave

Algorithm 1 MEDAG Construction Algorithm

For each eigenvalue $\lambda_j \in \Lambda_U(\mathbf{M})$ **do**:
Initialization : Initialize $c_j(i) = 0, \quad \mathcal{N}_j^i = \emptyset, \quad \forall i \in \mathcal{R}$. Each node determines whether it belongs to the set $\mathcal{S}_j$.
Source nodes transmit : Each regular node in $\mathcal{S}_j$ updates its counter value $c_j(i) = 1$, and transmits a message m_j (e.g. "1") to its out-going neighbors. Following this step, it does not listen to any other node, i.e., $\mathcal{N}_j^i = \emptyset$ and $c_j(i) = 1, \forall i \in \mathcal{S}_j \cap \mathcal{R}$ for the remainder of the algorithm.
Non-source nodes receive : Each regular node $i \in \mathcal{V} \setminus \mathcal{S}_j$ does the following:

- If $c_j(i) = 0$, *and* node i has received m_j from at least $2f + 1$ distinct neighbors (not necessarily simultaneously) it updates $c_j(i)$ to 1, appends the labels of the neighbors from which it received m_j to $\mathcal{N}_j^i$, and transmits m_j to its out-going neighbors.
- If $c_j(i) = 1$, it does not change $c_j(i)$, and does not listen to (discards) the information received from its neighbors, i.e., it does not update $\mathcal{N}_j^i$.

Return : A set of sets $\{\mathcal{N}_j^i\}, \lambda_j \in \Lambda_U(\mathbf{M}), i \in \mathcal{R}$.

in any of the following ways. (i) It can choose to transmit any message other than the true message m_j. (ii) It can transmit the true message, but out of turn, i.e., before receiving m_j from at least $(2f + 1)$ neighbors. (iii) It can choose not to transmit a message at all. In the next section, we shall detail graph conditions which guarantee the termination of the MEDAG construction algorithm under arbitrary adversarial behavior. For the following discussion, we characterize the properties of the output of Algorithm 1 if it terminates. Consider the spanning subgraph $\mathcal{G}_j = (\mathcal{V}, \mathcal{E}_j)$ of the original graph $\mathcal{G}$, induced by the sets $\{\mathcal{N}_j^i\}$, $i \in \mathcal{R}$, returned by Algorithm 1. We now present certain results that show that $\mathcal{G}_j$ satisfies all the properties of a MEDAG as outlined in Definition 5.

Proposition 1: Based on Algorithm 1, the spanning subgraph $\mathcal{G}_j$ of the original graph $\mathcal{G}$ contains no directed cycles where every node belongs to $\mathcal{R}$. □

Proof: We prove by contradiction. Let there exist a directed cycle $v_i P v_i$, where v_i and the nodes in P belong to $\mathcal{R}$. The path P originates from v_i when node i transmits m_j to its out-going neighbor on path P and updates its counter $c_j(i)$ from 0 to 1 at time-step $t = k$. Let the last vertex on path P be v_l. Clearly, node i receives information from node l at a time instant $t > k$. As a directed edge exists from node l to node i, it is apparent that node i chooses to listen to node l even when $c_j(i)$ is set to 1. This goes against the rules to be followed by a regular node i as described by Algorithm 1. Thus, we reach a contradiction. ■

During the MEDAG construction algorithm, suppose a regular node i updates its counter value $c_j(i)$ from 0 to 1, and transmits the message m_j at time-step $t = k$. Then, we say that node i belongs to level k (in the subgraph $\mathcal{G}_j$ corresponding to λ_j), denoted by the set $\mathcal{L}_k^j$. We say that node i belongs to $\mathcal{L}_0^j$ if $i \in \mathcal{S}_j \cap \mathcal{R}$.[8]

Proposition 2: If Algorithm 1 terminates for $\lambda_j \in \Lambda_U(\mathbf{M})$, the sets $\{\mathcal{L}_0^j, \mathcal{L}_1^j, \cdots, \mathcal{L}_{T_j}^j\}$ form a partition of the set $\mathcal{R}$ in $\mathcal{G}_j$, where T_j denotes the smallest integer such that at time-step T_j, we have $c_j(i) = 1, \forall i \in \mathcal{R}$. □

Proof: Suppose Algorithm 1 terminates for $\lambda_j \in \Lambda_U(\mathbf{M})$. Then, each regular node must update its counter and transmit m_j at some time-step. Thus, it is obvious that $\bigcup_{m=0}^{T_j} \mathcal{L}_m^j = \mathcal{R}$. Also, it is apparent that a regular node cannot update its counter from 0 to 1 and transmit m_j at two distinct time instants (goes against the rules of Algorithm 1). Thus, $\mathcal{L}_m^j \cap \mathcal{L}_n^j = \emptyset, \forall m \neq n$. This completes the proof. ■

Theorem 2: If the MEDAG construction algorithm terminates for $\lambda_j \in \Lambda_U(\mathbf{M})$, then there exists a subgraph $\mathcal{G}_j$ satisfying all the properties of a MEDAG. □

Proof: The result follows trivially from the way $\mathcal{G}_j$ and the sets $\mathcal{L}_m^j$ (for $0 \leq m \leq T_j$) are defined, and from the results of Propositions 1 and 2. ■

It follows from the above theorem that if the MEDAG construction algorithm terminates for every $\lambda_j \in \Lambda_U(\mathbf{M})$, then based on Theorem 1, secure omniscience can be achieved by our proposed estimation scheme.

Remark 2: Our overall estimation scheme can be broadly decomposed into two phases, namely, the initialization phase and the estimation phase. The initialization phase involves the construction of a MEDAG for each $\lambda_j \in \Lambda_U(\mathbf{M})$, and needs to be implemented just once. Once the initialization phase ends, one can implement the estimation phase, summarized in Section III. A merit of the proposed scheme is that each phase of the design admits a fully distributed implementation even under adversarial behavior. □

VI. Feasible Graph Topologies

In this section, we characterize a set of feasible graph topologies which guarantee the termination of the MEDAG construction algorithm for each $\lambda_j \in \Lambda_U(\mathbf{M})$. To this end, we borrow the following definition from [8], [9].

Definition 6: (r-**reachable set**) For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a set $\mathcal{S} \subset \mathcal{V}$, we say $\mathcal{S}$ is an *r-reachable set* if there exists an $i \in \mathcal{S}$ such that $|\mathcal{N}_i \setminus \mathcal{S}| \geq r$, where $r \in \mathbb{N}_+$. □

Thus, if a set $\mathcal{S}$ is r-reachable, then it contains a node which has at least r neighbors outside $\mathcal{S}$. We modify the notion of a *strongly-r robust graph* [8] as follows.

Definition 7: (**strongly** r-**robust graph** *w.r.t.* $\mathcal{S}_j$) For $r \in \mathbb{N}_+$ and $\lambda_j \in \Lambda_U(\mathbf{M})$, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is *strongly r-robust w.r.t. to the set of source nodes $\mathcal{S}_j$*, if for any non-empty subset $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S}_j$, $\mathcal{C}$ is r-reachable. □

Lemma 3: The MEDAG construction algorithm terminates for $\lambda_j \in \Lambda_U(\mathbf{M})$ if $\mathcal{G}$ is strongly $(3f+1)$-robust w.r.t. $\mathcal{S}_j$. □

Proof: We prove by contradiction. Consider any $\lambda_j \in \Lambda_U(\mathbf{M})$ and let $\mathcal{G}$ be strongly $(3f + 1)$-robust w.r.t. the set of source nodes $\mathcal{S}_j$. Suppose that the MEDAG construction algorithm for λ_j does not terminate. Let $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S}_j$ be the set

[8]Note that our strategy allows even some of the source nodes in $\mathcal{S}_j$ to be adversarial.

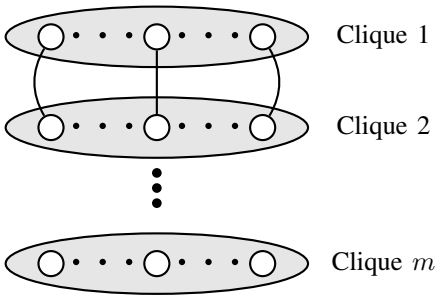

Fig. 1. Illustration of a feasible graph topology. Each clique has size $(3f+1)$. Each node in clique p (where $2 \leq p \leq m-1$) is connected to every node in cliques $p-1$ and $p+1$.

of *regular* nodes which never update their counter $c_j(i)$ from 0 to 1, where $i \in \mathcal{C}$. As $\mathcal{G}$ is strongly $(3f+1)$-robust w.r.t. $\mathcal{S}_j$, it follows that $\mathcal{C}$ is $(3f+1)$-reachable, i.e., there exists a node $i \in \mathcal{C}$ which has at least $3f+1$ neighbors outside $\mathcal{C}$. Under the f-local adversarial model, at least $2f+1$ of them are regular nodes with $c_j(i) = 1$. Thus, at least $2f+1$ regular nodes must have transmitted m_j to node i. Thus, based on the rules of Algorithm 1, node i must have updated $c_j(i)$ from 0 to 1 at some point of time, leading to a contradiction. ■

We now present the main result of this section, which presents a connection between the graph topology and the solution of the secure omniscience achieving problem.

Theorem 3: If $\mathcal{G}$ is strongly $(3f+1)$-robust w.r.t. $\mathcal{S}_j, \forall \lambda_j \in \Lambda_U(\mathbf{M})$, then secure omniscience can be achieved using the proposed estimation strategy. □

Proof: From Lemma 3, it follows that if $\mathcal{G}$ is strongly $(3f+1)$-robust w.r.t. $\mathcal{S}_j$ for every $\lambda_j \in \Lambda_U(\mathbf{M})$, then the MEDAG construction algorithm terminates for every unstable and marginally stable eigenvalue λ_j. As a result, from Theorem 2, it follows that a MEDAG exists for every $\lambda_j \in \Lambda_U(\mathbf{M})$. Finally, based on the result of Theorem 1, the existence of a MEDAG for every $\lambda_j \in \Lambda_U(\mathbf{M})$ implies that secure omniscience can be achieved using our proposed estimation strategy. This completes the proof. ■

As an illustration of a feasible graph topology, consider the undirected graph $\mathcal{G}$ shown in Figure 1. $\mathcal{G}$ comprises of m cliques of size $(3f+1)$ each. Further, each node in clique p is connected to every node in cliques $p-1$ and $p+1$ (where $2 \leq p \leq m-1$). For each $\lambda_j \in \Lambda_U(\mathbf{M})$, let the set of source nodes $\mathcal{S}_j$ correspond to one of the m cliques of $\mathcal{G}$. Then, it can be easily verified that $\mathcal{G}$ is strongly $(3f+1)$-robust w.r.t. any $\mathcal{S}_j$, where $\lambda_j \in \Lambda_U(\mathbf{M})$.

VII. Conclusion

We proposed a secure distributed state estimation algorithm for a class of LTI systems subject to worst-case adversarial attacks, and established sufficient conditions for the success of the algorithm. We introduced the notion of a *Mode Estimation Directed Acyclic Graph (MEDAG)*, showed that these MEDAGs play a key role in guaranteeing asymptotic convergence, and presented a distributed algorithm for constructing them. Finally, we characterized graph topologies which guarantee the termination of the MEDAG construction algorithm, and in turn, the success of our overall secure estimation scheme. Future work in this area would aim towards extending our method to systems with more general plant dynamics and obtaining a single necessary and sufficient topological condition for such systems.

References

[1] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. of the 44th IEEE Conference on Decision and Control and European Control Conference*, 2005, pp. 8179–8184.

[2] ——, "Distributed Kalman filtering for sensor networks," in *Proc. of the 46th IEEE Conference on Decision and Control*, 2007, pp. 5492–5498.

[3] U. A. Khan and A. Jadbabaie, "On the stability and optimality of distributed Kalman filters with finite-time data fusion," in *Proc. of the American Control Conference*, 2011, pp. 3405–3410.

[4] U. Khan, S. Kar, A. Jadbabaie, and J. M. Moura, "On connectivity, observability, and stability in distributed estimation," in *Proc. of the 49th IEEE Conference on Decision and Control*, 2010, pp. 6639–6644.

[5] S. Park and N. C. Martins, "Necessary and sufficient conditions for the stabilizability of a class of LTI distributed observers," in *Proc. of the 47th IEEE Conference on Decision and Control*, 2012, pp. 7431–7436.

[6] ——, "Design of distributed LTI observers for state omniscience," *IEEE Transactions on Automatic Control*, to appear.

[7] A. Mitra and S. Sundaram, "Distributed observers for LTI systems," *arXiv preprint arXiv:1608.01429*, 2016.

[8] H. Zhang and S. Sundaram, "Robustness of information diffusion algorithms to locally bounded adversaries," in *Proc. of the American Control Conference*, 2012, pp. 5855–5861.

[9] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 766–781, 2013.

[10] N. H. Vaidya, L. Tseng, and G. Liang, "Iterative approximate byzantine consensus in arbitrary directed graphs," in *Proc. of the 2012 ACM Symposium on Principles of Distributed Computing*. ACM, 2012, pp. 365–374.

[11] L. Tseng, N. Vaidya, and V. Bhandari, "Broadcast using certified propagation algorithm in presence of byzantine faults," *arXiv preprint arXiv:1209.4620*, 2012.

[12] S. Sundaram and B. Gharesifard, "Consensus-based distributed optimization with malicious nodes," in *Proc. of the 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015, pp. 244–249.

[13] L. Su and N. Vaidya, "Byzantine multi-agent optimization: Part i," *arXiv preprint arXiv:1506.04681*, 2015.

[14] I. Shames, A. M. Teixeira, H. Sandberg, and K. H. Johansson, "Distributed fault detection for interconnected second-order systems," *Automatica*, vol. 47, no. 12, pp. 2757–2764, 2011.

[15] I. Matei, J. S. Baras, and V. Srinivasan, "Trust-based multi-agent filtering for increased smart grid security," in *Proc. of the 20th Mediterranean Conference on Control & Automation*, 2012, pp. 716–721.

[16] T. Jiang, I. Matei, and J. Baras, "A trust based distributed Kalman filtering approach for mode estimation in power systems," in *Proc. of the First Workshop on Secure Control Systems*, 2010.

[17] U. Khan and A. M. Stankovic, "Secure distributed estimation in cyber-physical systems," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 5209–5213.

[18] D. Dolev, N. A. Lynch, S. S. Pinter, E. W. Stark, and W. E. Weihl, "Reaching approximate agreement in the presence of faults," *Journal of the ACM (JACM)*, vol. 33, no. 3, pp. 499–516, 1986.

[19] C.-Y. Koo, "Broadcast in radio networks tolerating byzantine adversarial behavior," in *Proc. of the Twenty-Third Annual ACM Symposium on Principles of Distributed Domputing*. ACM, 2004, pp. 275–282.

[20] A. Pelc and D. Peleg, "Broadcasting with locally bounded byzantine faults," *Information Processing Letters*, vol. 93, no. 3, pp. 109–115, 2005.

[21] C.-T. Chen, *Linear System Theory and Design*. Oxford University Press, Inc., 1995.