On the Impacts of Redundancy, Diversity, and Trust in Resilient Distributed State Estimation

Aritra Mitra, Faiq Ghawash, Shreyas Sundaram and Waseem Abbas

***Abstract*—We address the problem of distributed state estimation of a linear dynamical process in an attack-prone environment. Recent attempts to solve this problem impose stringent redundancy requirements on the measurement and communication resources of the network. In this paper, we take a step towards alleviating such strict requirements by exploring two complementary directions: (i) making a small subset of the nodes immune to attacks, or "trusted", and (ii) incorporating diversity into the network. We define graph-theoretic constructs that formally capture the notions of redundancy, diversity, and trust. Based on these constructs, we develop a resilient estimation algorithm and demonstrate that even relatively sparse networks that either exhibit node-diversity, or contain a small subset of trusted nodes, can be just as resilient to adversarial attacks as more dense networks. Finally, given a finite budget for network design, we focus on characterizing the complexity of (i) selecting a set of trusted nodes, and (ii) allocating diversity, so as to achieve a desired level of robustness. We establish that, unfortunately, each of these problems is NP-complete.**

I. Introduction

The distributed state estimation problem, in its most basic form, concerns asymptotic reconstruction of the state of a dynamical process, via a group of sensor nodes interacting over a network [1]–[9]. Each node observes only a portion of the state dynamics and, hence, is reliant on local information exchanges with neighboring nodes for tracking the entire state. An underlying assumption that runs through almost all works on this topic is that the sensor nodes work *collaboratively* towards the common goal of state estimation. However, the recent surge of activity devoted to the security of networked control systems suggests that this may no longer be a reasonable assumption to make. Thus, it is of prime importance to design algorithms and networks that are robust to attacks on certain parts of the system.

At a high level, the literature on the security of control systems can be grouped into two categories: one where all sensor measurements are available at a single central entity, and the other where such measurements are dispersed over a network. There is a vast body of work that has studied the former, with results spanning both linear [10]–[12] and non-linear dynamics [13]; for a comprehensive survey, see [14]. The computational complexity of such problems has also been recently investigated in [15]. Regarding the networked setting of interest to us, the literature is scant, and can be broadly classified in terms of the assumptions made on the adversary model. While [16]–[19] consider attack models that are limited in scope, [20], [21] account for worst-case Byzantine adversarial attacks [22], where the adversaries can act arbitrarily. However, allowing for sophisticated attack models comes at the expense of rather stringent requirements on the communication network topology. Specifically, the guarantees provided in [20], [21] hold only when the network exhibits a sufficient amount of redundancy in both its measurement and communication resources. We are thus motivated to ask: *Can one relax the redundancy requirements on the network, and yet, tolerate a worst-case attack model?* The goal of this paper is to demonstrate that this can indeed be done.

Recently, in [23] and [24], two distinct ideas were proposed that depart from the conventional approach of increasing robustness through redundancy. In [23], the authors explored the concept of device hardening, wherein a small subset of carefully selected nodes, called *trusted nodes*, were made immune to attacks. On the other hand, in [24], the authors exploited the fact that the components of a large-scale networked control system are typically quite *diverse* in their hardware and software implementations. Such diversity, in turn, implies that the vulnerabilities of different components are not necessarily alike. The key observation here is that even if an adversary manages to breach the security of a particular type of component, its impact would remain limited to only components of that type. In the context of consensus, when the above ideas are leveraged appropriately, it has been shown that even a relatively sparse network with trusted nodes [23], or sufficient diversity [24], can still exhibit the same functional robustness as that of a highly connected, dense network.

In light of the above developments, it is natural to ask whether the ideas of *trust* and *diversity* can be adapted to solve the resilient distributed state estimation problem. Specifically, the main questions of interest to us are as follows.

- Can introducing trusted nodes and diversity into a sparse network alleviate the redundancy requirements needed for resilient distributed state estimation?
- How should one choose a set of trusted nodes, and incorporate diversity, such that the resulting network is endowed with a desired level of robustness?

This work was supported by NSF CMMI-163501 and CAREER award 1653648.

A. Mitra is with the Department of Electrical and Systems Engineering, University of Pennsylvania, PA 19104, USA. Email: amitra20@seas.upenn.edu

S. Sundaram is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. Email: sundara2@purdue.edu

F. Ghawash is with the Department of Engineering and Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway. Email: faiq.ghawash@ntnu.no

W. Abbas is with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37212, USA. Email: waseem.abbas@vanderbilt.edu

In this paper, we provide answers to the above questions by making the following contributions.

Contributions: In Section III, we introduce novel graph-theoretic constructs that formally capture the three facets of interest, namely redundancy, diversity, and trust. We then develop an attack-resilient, provably-correct filtering algorithm that exploits these facets to enable each non-compromised node to asymptotically recover the entire state, provided the graph-theoretic conditions introduced in Section III are met.

One of the assumptions typically made when dealing with Byzantine attack models is that the number of compromised nodes is bounded in some appropriate sense [20]–[32] - an assumption that we relax in Section III-C. In particular, once an adversary has managed to breach the security of a particular type of component (node), we allow it to compromise any number of nodes of that type. We show how one can account for such scenarios as long as the network is sufficiently diverse in its measurement and communication resources. In the process, we argue that one can employ diversity as a means to tackle spoofing attacks, where an attacker can impersonate the identities of multiple nodes. Next, we employ the conditions introduced in Section III to demonstrate the utility of making certain nodes trusted. Specifically, in Section IV, we prove that the absence of even a single trusted node may need to be compensated by augmenting the network with several additional measurement and communication resources, so as to achieve a desired level of robustness.

Finally, we study the problem of designing a robust network subject to cost constraints. Given a certain budget that caps the number of nodes that can be made trusted, or the amount of diversity that can be afforded, we focus on understanding (i) which nodes should be made trusted, and (ii) how one should allocate diversity, in order to achieve a desired level of robustness. In Section V, we formulate these problems as decision problems and characterize their complexity. We show that, unfortunately, each of these problems is NP-complete.

In sum, relative to our prior work [20], in this paper we (i) introduce novel graph-theoretic alternatives to redundancy that are built on the ideas of diversity and trust; (ii) establish how such alternatives can be exploited to solve the resilient distributed state estimation problem; (iii) formally demonstrate the benefits of diversity and trust; and (iv) characterize the complexity of selecting trusted nodes and incorporating diversity. Our results provide various insights into the design of resilient networks tailored to the task of state estimation.

A preliminary version of this paper appeared as [32], where we only considered the impact of making certain nodes trusted.

II. Notation, Terminology, and Problem Setup

In this section, we formally describe the various models considered throughout the paper; subsequently, we state the problem of interest. We begin by introducing relevant notation.

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 (or in-neighborhood) of the i-th node is defined as $\mathcal{N}_i \triangleq \{j \,|\, (j,i)\in\mathcal{E}\}$. A node j is said to be an out-neighbor of node i if $(i,j)\in\mathcal{E}$. The notation $|\mathcal{V}|$ is used to denote the cardinality of a set $\mathcal{V}$. The set of all eigenvalues (or 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 unstable eigenvalues by $\Lambda_U(\mathbf{A})=\{\lambda\in sp(\mathbf{A})\,|\,|\lambda|\geq 1\}$. The identity matrix of dimension r is denoted $\mathbf{I}_r$, and $\mathbb{N}_+$ is used to refer to the set of all positive integers. The terms 'communication graph' and 'network' are used interchangeably.

Plant and Observation Model: Consider a linear time-invariant dynamical process[1]

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

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. A network $\mathcal{G}=(\mathcal{V},\mathcal{E})$ of N nodes monitor the state of this system. The i-th node receives 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 define $\mathbf{C}\triangleq\begin{bmatrix}\mathbf{C}_1^T & \cdots & \mathbf{C}_N^T\end{bmatrix}^T$ and $\mathbf{y}[k]\triangleq\begin{bmatrix}\mathbf{y}_1^T[k] & \cdots & \mathbf{y}_N^T[k]\end{bmatrix}^T$ as the collective observation matrix, and collective measurement vector, respectively. In the standard distributed state estimation setup, each node i is tasked with asymptotically recovering the entire state $\mathbf{x}[k]$. We make the basic (necessary) assumption that the pair $(\mathbf{A},\mathbf{C})$ is detectable. However, for any given $i\in\mathcal{V}$, the pair $(\mathbf{A},\mathbf{C}_i)$ may not be detectable, thereby necessitating inter-node communication constrained by the structure of the network.

Remark 1. *Although observability/detectability are generic properties, the fact that we do not assume that the pair $(\mathbf{A},\mathbf{C}_i)$ is detectable is motivated by various practical settings where such an assumption is unrealistic. For instance, consider the task of environmental monitoring of a spatio-temporal process (e.g., temperature or gas concentration) that evolves over a large geographical region [33]. It is unlikely that the measurements available at one corner of the region reveal information about how the process evolves at a far-away location, i.e., the overall dynamics is not locally observable. In fact, the assumptions that (i) $(\mathbf{A},\mathbf{C}_i)$ is not necessarily detectable, and (ii) each node knows the entire state-transition matrix $\mathbf{A}$, are standard in the literature on distributed state estimation [1]–[9].* □

Adversary model: We consider a subset $\mathcal{A}\subset\mathcal{V}$ of the nodes in the network to be adversarial; the remaining regular nodes will be denoted by the set $\mathcal{R}$. To formally describe the characteristics of the adversarial set $\mathcal{A}$, we need to first lay out three key considerations. (i) What are the *capabilities* of an adversary? (ii) How *many* adversaries are there in the network? (iii) Which *type* of nodes can be compromised? Let us now elaborate on these considerations.

- *Capabilities of an adversary:* First, in terms of capabilities, we allow a compromised node to act *arbitrarily*. In particular, an adversary can send incorrect, and potentially inconsistent estimates of the state to different neighbors at the

[1] Our approach guarantees bounded mean-square error under system and measurement noise with bounded second moments.

same instant of time.[2] Furthermore, nodes in $\mathcal{A}$ can act collaboratively, and we assume that such nodes possess complete knowledge of the network topology, the system dynamics, and the algorithm employed by the non-adversarial nodes. In terms of the capabilities and knowledge of the adversarial set $\mathcal{A}$, the features we have described above are consistent with the classical worst-case Byzantine attack model [22].

• *Number of adversaries:* While considering worst-case Byzantine attack models, it is quite typical to impose certain restrictions on the number of adversaries in the network. To this end, we will use the following definitions [25].

Definition 1. *(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 2. ***(f-local adversarial model)*** *A set $\mathcal{A}$ of adversarial nodes is f-locally bounded if $\mathcal{A}$ is an f-local set.* □

From the above definitions, note that if $\mathcal{A}$ is f-locally bounded, then there are at most f adversaries in the neighborhood of each regular node. While we will mostly deal with f-locally bounded adversarial sets, ways to relax such an assumption will also be outlined as we proceed.

• *Types of nodes:* In order to explain which types of nodes can be compromised, let us first describe the diversity and trust models that we will consider.

Diversity Model: We capture node heterogeneity and, in particular, the fact that nodes have different vulnerabilities, by employing the notion of colors as suggested in [24]. Specifically, let $\Gamma = \{B_1, \ldots, B_{|\Gamma|}\}$ denote a set of colors, and let each node i be assigned a unique color $\Delta(i) \in \Gamma$, where $\Delta(\cdot)$ is a mapping from $\mathcal{V}$ to Γ. Let the node set be partitioned accordingly as $\mathcal{V} = \{\mathcal{V}_{B_1}, \ldots, \mathcal{V}_{B_{|\Gamma|}}\}$.

Trust Model: We assume that a subset $\mathcal{T} \subseteq \mathcal{V}$ of nodes, termed trusted nodes, cannot be compromised by adversaries, i.e., $\mathcal{T} \cap \mathcal{A} = \emptyset$. Note that when $|\Gamma| = 1$, i.e., when all nodes are of the same type, we recover the setting in [32], where only the impact of trusted nodes was considered.

The models described above can be employed in a variety of practical settings to improve system resiliency. For instance, in [34], the authors demonstrate that installing diverse operating systems in a computer network prevents a single operating system-specific attack to propagate across the entire breadth of the network. Similar ideas that exploit node heterogeneity or diversity have been shown to be effective in communication systems [35] as well as in power grids [36]. The utility of trusted nodes too have been studied in the context of broadcasting in sensor networks [37].

Let us now get back to the question of which nodes can and cannot be compromised.

Assumption 1. *We make the following standing assumptions on the adversarial set $\mathcal{A}$.*

(A1) *$\mathcal{A}$ is mono-chromatic, i.e., all adversaries are of the same type or color.*

(A2) *Each trusted node is non-adversarial, i.e., $\mathcal{T} \cap \mathcal{A} = \emptyset$.*

Assumption 1 (A1) is used to capture the fact that diverse nodes have different vulnerabilities, i.e., breach of a particular type of component (node) does not necessarily imply breach of the other types. However, we can easily generalize our subsequent developments to accommodate poly-chromatic adversarial sets. Assumption 1 (A2) is more a defining property of a trusted set than an assumption on the adversary model.

It is important to emphasize the distinction between the trust and diversity models. With the former, a node precisely knows which of its neighbors cannot be compromised; with the latter, it only knows that at most one type of its neighbors can be compromised, but *does not know* which type has actually been compromised. This difference in information, while subtle, requires us to exploit trust and diversity in distinct ways. Also note that an adversary has the freedom to decide which color to compromise. However, no trusted node can be attacked.

Note that the actual number and identities of the adversarial nodes are not known to the regular nodes. However, for the purpose of analysis, we will make the following assumptions.

Assumption 2. *Each node in $\mathcal{R}$ is aware of the following.*

(B1) *An upper-bound f on the number of adversaries in its neighborhood (under an f-local adversarial model).*

(B2) *The true color of each of its neighbors, including those that are adversarial.*

(B3) *The identities of its trusted neighbors.*

Assumption 2 (B1) is a standard assumption in the distributed fault tolerant literature [20]–[32] and, in fact, we are unaware of any work that does not make such an assumption while dealing with (as we do) worst-case attack models where an adversary can act arbitrarily. The underlying rationale here is that any reliable system is typically programmed to tolerate a maximum number of component failures or attacks. Accordingly, each node is aware of the maximum number of attacks it can accommodate in its neighborhood. Assumptions 2 (B2) and (B3) are specific to our setting, and without them, it seems quite unlikely to come up with an approach that exploits diversity and trust in a meaningful way.

At this point, it is important to clarify that while incorporating trust and diversity into the network offers several merits (as we shall establish in the paper), such merits do not come for free. Indeed, ensuring that trusted nodes do not get compromised, and installing nodes of different types, might incur considerable costs; however, the point of this paper is not to contemplate how such costs compare with those for installing several redundant components. Instead, through our results, we identify the capabilities enabled by the notions of redundancy, diversity, and trust. A natural follow-up direction would be to formulate an optimization problem that tries to balance the costs associated with each of these facets; Section V provides insights into the complexity of such a problem.

We are now in position to state the problem of interest. To this end, let $\hat{\mathbf{x}}_i[k]$ represent the estimate of $\mathbf{x}[k]$ maintained by node i. Our goal in this paper will be to study how diversity and trust can be exploited to solve the following problem.

[2] Note that this essentially means that all outgoing edges/communication links associated with a compromised node can be viewed as under attack. We do not, however, explicitly consider a model where only the communication links are under attack.

Problem 1. ***(Resilient Distributed State Estimation)*** *Given an LTI system (1), a linear measurement model (2), and a time-invariant directed communication graph $\mathcal{G}$, design a set of state estimate update and information exchange rules such 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-local mono-chromatic set of Byzantine adversaries.*

While our focus will mainly be on solving the above problem, the ideas that we develop in the process will naturally extend to the case where the adversarial set is no longer f-locally bounded - a model that we will refer to as simply the *mono-chromatic Byzantine adversary model*. We will investigate such a model in Section III-C.

III. Resilient Distributed State Estimation under mono-chromatic Byzantine adversaries

A. Characterizing Sufficient Graph-theoretic Conditions

In this section, we identify certain graph-theoretic conditions that play a key role in our proposed solution to Problem 1. In particular, these topological conditions are sufficient to solve Problem 1 based on an approach that we will develop later in Section III-B. To proceed, we introduce the following notion of $(r,\Delta(\cdot),\mathcal{T})$-reachability.

Definition 3. *($(r,\Delta(\cdot),\mathcal{T})$**-reachable set**) Consider a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with a trusted node set $\mathcal{T}$, where each node $i\in\mathcal{V}$ is assigned a color $\Delta(i)$. Then, given $r\in\mathbb{N}_+\cup\{\infty\}$, and a non-empty set $\mathcal{C}\subseteq\mathcal{V}$, $\mathcal{C}$ is said to be an $(r,\Delta(\cdot),\mathcal{T})$-reachable set if $\exists i\in\mathcal{C}$ satisfying at least one of the following conditions:*

(i) ***Redundancy****: Node i has at least r neighbors outside $\mathcal{C}$, i.e., $|\mathcal{N}_i\setminus\mathcal{C}|\geq r$.*

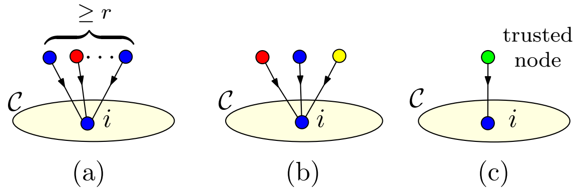
(ii) ***Diversity****: Node i has at least 3 distinct colored neighbors outside $\mathcal{C}$, i.e., there exist nodes $u,v,w\in\mathcal{N}_i\setminus\mathcal{C}$, such that $\Delta(u)\neq\Delta(v)\neq\Delta(w)\neq\Delta(u)$.*

(iii) ***Trust****: Node i has at least one trusted neighbor outside $\mathcal{C}$, i.e., $|\{\mathcal{N}_i\setminus\mathcal{C}\}\cap\mathcal{T}|\geq 1$.*

□

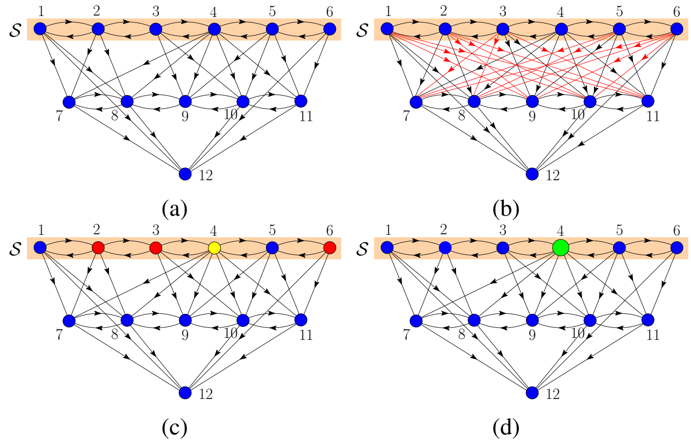
The above definition captures the ability of a set $\mathcal{C}$ of nodes to correctly process information that diffuses into the set from nodes outside it. For this to happen, intuition dictates that the lack of any one of the facets of redundancy, diversity, and trust should be compensated by the presence of at least one of the other two. More precisely, we require at least one node within $\mathcal{C}$ to either have enough neighbors outside $\mathcal{C}$, or three different types of neighbors outside $\mathcal{C}$, or a trusted neighbor outside $\mathcal{C}$.[3]

Since $|\mathcal{V}|$ is finite, observe that when $r=\infty$ in Definition 3, a set $\mathcal{C}$ can fulfill the requirements of $(\infty,\Delta(\cdot),\mathcal{T})$-reachability if and only if either condition (ii) or condition (iii) in Definition 3 is satisfied, i.e., either via diversity or trust. Henceforth, $(\infty,\Delta(\cdot),\mathcal{T})$-reachability will simply be referred to as $(\Delta(\cdot),\mathcal{T})$-reachability; this special case with $r=\infty$ will be of particular importance to us in Section III-C when we relax the f-local assumption.



(a) (b) (c)

Fig. 1: Illustration of sets that satisfy $(r,\Delta(\cdot),\mathcal{T})$-reachability as per Defn. 3, via (a) redundancy, (b) diversity, or (c) trust.



(a) (b) (c) (d)

Fig. 2: Illustration of different approaches, including redundancy (b), diversity (c), and trust (d), to improve network robustness. The set of source nodes in all figures is $\mathcal{S}=\{1,2,\ldots,6\}$. Node 4 is a trusted node in Fig. 2(d). The graphs in Figs. 2(b), 2(c), and 2(d) are all strongly $(6,\Delta(\cdot),\mathcal{T})$-robust w.r.t. $\mathcal{S}$.

Next, we introduce the key topological property required to solve Problem 1 based on our proposed approach.

Definition 4. *(**strongly** $(r,\Delta(\cdot),\mathcal{T})$**-robust graph w.r.t.** $\mathcal{S}$) Consider a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with a trusted node set $\mathcal{T}$, where each node $i\in\mathcal{V}$ is assigned a color $\Delta(i)$. Then, given $r\in\mathbb{N}_+\cup\{\infty\}$, and a set $\mathcal{S}\subseteq\mathcal{V}$, $\mathcal{G}$ is strongly $(r,\Delta(\cdot),\mathcal{T})$-robust w.r.t. $\mathcal{S}$ if for all non-empty subsets $\mathcal{C}\subseteq\mathcal{V}\setminus\mathcal{S}$, $\mathcal{C}$ is $(r,\Delta(\cdot),\mathcal{T})$-reachable.* □

The notion of strong $(r,\Delta(\cdot),\mathcal{T})$-robustness formalizes the idea that there are multiple ways to achieve a desired level of robustness in the underlying network: by creating extra links between nodes (redundancy), or by diversifying nodes (diversity), or by hardening a subset of the nodes (trust), or by a combination of these approaches. As we shall see in Section III-B, our main convergence result, namely Theorem 2, hinges upon the idea of strong $(r,\Delta(\cdot),\mathcal{T})$-robustness. Note that when all nodes are of the same color, i.e., when $\Delta(i)=\Delta(j),\forall i,j\in\mathcal{V}$, and when the trusted set $\mathcal{T}$ is empty, we recover the conventional notions of r-reachability [27], and strong r-robustness w.r.t. a set $\mathcal{S}$ [20], from Definitions 3 and 4, respectively.

Example: Consider the graph in Figure 2(a), in which all nodes have the same color (no diversity), and no node is trusted. The graph is strongly $(3,\Delta(\cdot),\mathcal{T})$-robust w.r.t. $\mathcal{S}=\{1,2,3,4,5,6\}$, where $\Delta(i)=\Delta(j),\forall i\neq j$, and $\mathcal{T}=\emptyset$. We can make such a graph strongly $(6,\Delta(\cdot),\mathcal{T})$-robust w.r.t.

[3] With three distinct colored neighbors, we later devise an algorithm, namely Algorithm 1, that exploits diversity *without* requiring knowledge of which color/type of nodes has been compromised; with just two distinct colored neighbors, this is impossible.

$\mathcal{S}$ simply by adding extra links between nodes as shown in Figure 2(b). At the same time, if we have three colors, then we can assign them to nodes such that the graph becomes strongly (6,$\Delta(\cdot),\mathcal{T}$)-robust w.r.t. $\mathcal{S}$, without adding extra edges or trusted nodes, as shown in Figure 2(c). Similarly, if node 4 is a trusted node, while all the remaining nodes are of the same color, the graph again becomes strongly (6,$\Delta(\cdot),\mathcal{T}$)-robust w.r.t. $\mathcal{S}$, with no extra edges, as illustrated in Fig. 2(d).

Next, we recall the notion of source nodes [20].

Definition 5. *(**Source nodes**) For each $\lambda_j \in \Lambda_U(\mathbf{A})$, let the set $\mathcal{S}_j$ be defined as follows:*

$$\mathcal{S}_j \triangleq \{i \in \mathcal{V} | rank \begin{bmatrix} \mathbf{A} - \lambda_j \mathbf{I}_n \\ \mathbf{C}_i \end{bmatrix} = n\}. \tag{3}$$

Then, $\mathcal{S}_j$ will be called the set of source nodes for λ_j.[4] □

Let $\Omega_U(\mathbf{A}) \subseteq \Lambda_U(\mathbf{A})$ contain the set of eigenvalues of $\mathbf{A}$ for which $\mathcal{V} \setminus \mathcal{S}_j$ is non-empty. Essentially, for each unstable mode $\lambda_j \in \Omega_U(\mathbf{A})$, the source nodes $\mathcal{S}_j$ can leverage their own local measurements to estimate the portion of the state corresponding to λ_j. However, to enable each non-source node $i \in \mathcal{V} \setminus \mathcal{S}_j$ to estimate that portion, a secure medium of information flow from $\mathcal{S}_j$ to $\mathcal{V} \setminus \mathcal{S}_j$ is necessary. To this end, the concept of a Mode Estimation Directed Acyclic Graph (MEDAG) was introduced in [20]. We now suitably modify the definition of a MEDAG to account for diversity and trust.

Definition 6. *($(2f+1, \Delta(\cdot), \mathcal{T})$ **Mode Estimation Directed Acyclic Graph (MEDAG)**) Consider a mode $\lambda_j \in \Omega_U(\mathbf{A})$. Suppose there exists a spanning sub-graph $\mathcal{G}_j = (\mathcal{V}, \mathcal{E}_j)$ of $\mathcal{G}$ with the following properties for all f-local, mono-chromatic sets $\mathcal{A}$ with $\mathcal{A} \cap \mathcal{T} = \emptyset$, and $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$.*

(i) *If $i \in \{\mathcal{V} \setminus \mathcal{S}_j\} \cap \mathcal{R}$, then either $|\mathcal{N}_i^{(j)}| \geq 2f+1$; or $|\mathcal{N}_i^{(j)} \cap \mathcal{T}| \geq 1$; or $\exists u, v, w \in \mathcal{N}_i^{(j)}$ such that $\Delta(u) \neq \Delta(v) \neq \Delta(w) \neq \Delta(u)$. Here, $\mathcal{N}_i^{(j)} = \{l \in \mathcal{V} | (l, i) \in \mathcal{E}_j\}$ represents the neighborhood of node i in $\mathcal{G}_j$.*

(ii) *There exists a partition of $\mathcal{R}$ into sets $\{\mathcal{L}_0^{(j)}, \ldots, \mathcal{L}_{T_j}^{(j)}\}$, where $T_j \in \{0, \ldots, N-1\}$, $\mathcal{L}_0^{(j)} = \mathcal{S}_j \cap \mathcal{R} \neq \emptyset$, and if $i \in \mathcal{L}_q^{(j)}$ (where $1 \leq q \leq T_j$), then $\mathcal{N}_i^{(j)} \cap \mathcal{R} \subseteq \bigcup_{r=0}^{q-1} \mathcal{L}_r^{(j)}$. Furthermore, $\mathcal{N}_i^{(j)} = \emptyset, \forall i \in \mathcal{L}_0^{(j)}$.*

Then, we call $\mathcal{G}_j$ a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG for λ_j. □

Intuitively, a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG for λ_j is a sub-graph of $\mathcal{G}$ with features that facilitate reliable transmission of information from nodes in $\mathcal{S}_j$ to those in $\mathcal{V} \setminus \mathcal{S}_j$. The first such feature requires each non-source node in $\mathcal{R}$ to either have $(2f+1)$ neighbors, or a trusted neighbor, or three distinct colored neighbors in $\mathcal{G}_j$. Thus, condition (i) in Definition 6 ensures that each regular non-source node either has adequate redundancy, diversity, or trusted nodes in its local neighborhood; later, in Algorithm 1, we will see how these properties are explicitly used to correctly process information. Condition (ii) in Definition 6 states that in $\mathcal{G}_j$, the set $\mathcal{R}$ should admit a partition into levels $\{\mathcal{L}_0^{(j)}, \ldots, \mathcal{L}_{T_j}^{(j)}\}$, such that a node in a particular level q has neighbors in $\mathcal{R}$ from levels strictly lower than q, leading to an acyclic structure - a feature that we exploit in Algorithm 1 to ensure uni-directional information flow from $\mathcal{S}_j$ to $\mathcal{V} \setminus \mathcal{S}_j$, which, in turn, proves to be crucial in establishing convergence guarantees.

In Appendix A, we provide an algorithm to construct a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG and, in the process, identify the sets $\mathcal{N}_i^{(j)}, \forall i \in \mathcal{V}$. With the sets $\mathcal{N}_i^{(j)}$ in hand, one can implement the resilient distributed state estimation algorithm, namely Algorithm 1, which we will develop in the next section. Before doing so, we ask: When does a given graph contain a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG? In the next result, we show that the notion of strong $(r, \Delta(\cdot), \mathcal{T})$-robustness characterizes the existence of such sub-graphs. The proof of this result can be found in [38]; we omit it here due to space constraints.

Theorem 1. *For each $\lambda_j \in \Omega_U(\mathbf{A})$, $\mathcal{G}$ contains a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG for λ_j if and only if $\mathcal{G}$ is strongly $(2f+1, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$.* □

B. Algorithm and Analysis for f-local Mono-chromatic Byzantine Adversaries

In this section, we develop an algorithm that leverages node-diversity and the presence of trusted nodes to solve Problem 1. For clarity of exposition, we make the following assumption on the system matrix $\mathbf{A}$.

Assumption 3. $\mathbf{A}$ *has real, distinct eigenvalues.*

Although the above assumption might seem restrictive, the results that we derive subsequently can be generalized to account for system matrices with arbitrary spectrum using a more detailed technical analysis as in [20]. Since any $\mathbf{A}$ satisfying Assumption 3 can be diagonalized via an appropriate similarity transformation, we assume without loss of generality that $\mathbf{A}$ is already in diagonal form. Specifically, suppose $\mathbf{A} = diag(\lambda_1, \cdots, \lambda_n)$, where $sp(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$. Let the component of the state vector $\mathbf{x}[k]$ corresponding to eigenvalue λ_j be denoted by $x^{(j)}[k]$. Building on the general idea developed in [20], for each $\lambda_j \in \Omega_U(\mathbf{A})$, the source nodes $\mathcal{S}_j$ and the non-source nodes $\mathcal{V} \setminus \mathcal{S}_j$ employ separate update rules for estimating $x^{(j)}[k]$. In particular, the source nodes maintain local[5] Luenberger observers for estimating $x^{(j)}[k]$, while the non-source nodes rely on a resilient consensus based protocol to achieve this task. For any node i, let the set of eigenvalues it can detect be denoted by $\mathcal{O}_i$, and let $\bar{\mathcal{O}}_i = sp(\mathbf{A}) \setminus \mathcal{O}_i$. Then, the following result from [20] states that node i can estimate the components of $\mathbf{x}[k]$ corresponding to the eigenvalues in $\mathcal{O}_i$, (i.e., the locally detectable portion of $\mathbf{x}[k]$) *without* interacting with its neighbors.

Lemma 1. *Suppose Assumption 3 holds. Then, for each $i \in \mathcal{R}$, a local Luenberger observer can be constructed that ensures $\lim_{k\to\infty} |\hat{x}_i^{(j)}[k] - x^{(j)}[k]| = 0, \forall \lambda_j \in \mathcal{O}_i$, where $\hat{x}_i^{(j)}[k]$ denotes the estimate of $x^{(j)}[k]$ maintained by node i.* □

In what follows, we develop a filtering algorithm, adapted to account for node-diversity and the presence of trusted nodes, that allows each regular node to estimate the locally unde-

[4] In case $i \in \mathcal{S}_j$, we will say that "node i can detect λ_j". Each stable eigenvalue is considered detectable w.r.t. the measurements of every node.

[5] Here, by 'local', we imply that such observers can be constructed and run without any information from neighbors.

Algorithm 1 For each $i \in \mathcal{R}$, and $\lambda_j \in \bar{\mathcal{O}}_i$, steps for updating $\hat{x}_i^{(j)}[k]$

```
1: Collect the estimates $\hat{x}_l^{(j)}[k], l \in \mathcal{N}_i^{(j)}$.
2: **if** $\mathcal{N}_i^{(j)} \cap \mathcal{T} \neq \emptyset$ **then**
3:     Update $\hat{x}_i^{(j)}[k]$ as follows:
           $$\hat{x}_i^{(j)}[k+1] = \lambda_j \left( \sum_{l \in \mathcal{N}_i^{(j)} \cap \mathcal{T}} \bar{w}_{il}^{(j)} \hat{x}_l^{(j)}[k] \right), \text{where} \quad (4)$$
           $\bar{w}_{il}^{(j)} \geq 0, \forall l \in \mathcal{N}_i^{(j)} \cap \mathcal{T}$, and $\sum_{l \in \mathcal{N}_i^{(j)} \cap \mathcal{T}} \bar{w}_{il}^{(j)} = 1$.
4: **else if** $\exists$ 3 distinct colored nodes in $\mathcal{N}_i^{(j)}$ **then**
5:     Perform TrimOp1, and update $\hat{x}_i^{(j)}[k]$ as follows:
           $$\hat{x}_i^{(j)}[k+1] = \lambda_j \left( \sum_{l \in \mathcal{R}_i^{(j)}[k]} \tilde{w}_{il}^{(j)}[k] \hat{x}_l^{(j)}[k] \right), \text{where} \quad (5)$$
           $\tilde{w}_{il}^{(j)}[k] \geq 0, \forall l \in \mathcal{R}_i^{(j)}[k]$, and $\sum_{l \in \mathcal{R}_i^{(j)}[k]} \tilde{w}_{il}^{(j)}[k] = 1$.
6: **else**
7:     Perform TrimOp2, and update $\hat{x}_i^{(j)}[k]$ as follows:
           $$\hat{x}_i^{(j)}[k+1] = \lambda_j \left( \sum_{l \in \mathcal{P}_i^{(j)}[k]} w_{il}^{(j)}[k] \hat{x}_l^{(j)}[k] \right), \text{where} \quad (6)$$
           $w_{il}^{(j)}[k] \geq 0, \forall l \in \mathcal{P}_i^{(j)}[k]$, and $\sum_{l \in \mathcal{P}_i^{(j)}[k]} w_{il}^{(j)}[k] = 1$.
8: **end if**
```

tectable portion of its dynamics. To explain the steps of our algorithm, we first describe two key "trimming" operations.

- `TrimOp1`: Suppose there exist three distinct colored nodes in $\mathcal{N}_i^{(j)}$. Then, `TrimOp1` comprises of the following steps. Node $i \in \mathcal{R}$ sorts the estimates of $x^{(j)}[k]$ received from $\mathcal{N}_i^{(j)}$ in descending order. Upon such sorting, let the indices of the nodes in $\mathcal{N}_i^{(j)}$ be $\{n_1, \ldots, n_{|\mathcal{N}_i^{(j)}|}\}$, i.e., $\hat{x}_{n_1}^{(j)}[k] \geq \hat{x}_{n_2}^{(j)}[k] \ldots \geq \hat{x}_{n_{|\mathcal{N}_i^{(j)}|}}^{(j)}[k]$.[6] Define $m \triangleq \min\{p : \Delta(n_p) \neq \Delta(n_1)\}$, and $M \triangleq \max\{p : \Delta(n_p) \neq \Delta(n_{|\mathcal{N}_i^{(j)}|})\}$. It can be easily verified that when $\mathcal{N}_i^{(j)}$ contains at least 3 distinct colored nodes, we have $M \geq m$. Accordingly, node i identifies the set $\mathcal{R}_i^{(j)}[k] = \{n_m, n_{m+1}, \ldots, n_M\}$.[7]
- `TrimOp2`: As in `TrimOp1`, node $i \in \mathcal{R}$ sorts the estimates of $x^{(j)}[k]$ received from $\mathcal{N}_i^{(j)}$ in descending order. It then removes the highest f and the lowest f estimates, i.e., it removes $2f$ estimates in all. Node i identifies the set $\mathcal{P}_i^{(j)}[k] \subset \mathcal{N}_i^{(j)} (\subseteq \mathcal{N}_i)$ of nodes whose estimates are not rejected in the above step.

Description of Algorithm 1: Based on the two trimming operations described above, we formally outline our approach in Algorithm 1. Let us now briefly discuss the key steps.

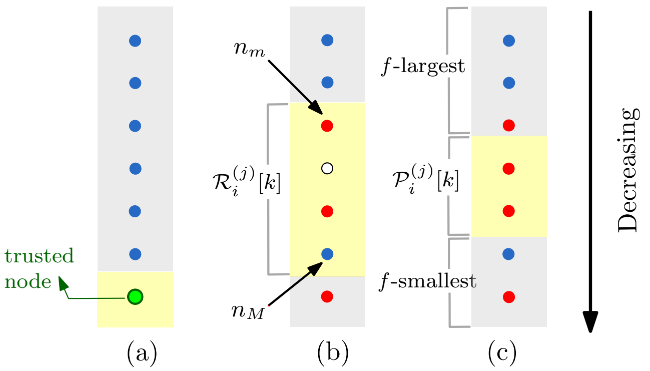


Fig. 3: Illustration of various steps in Algo. 1. (a) $\mathcal{N}_i^{(j)}$ has a trusted node (in green); node i uses only its estimate. (b) $\mathcal{N}_i^{(j)}$ has no trusted node but contains three distinct colored nodes (blue, red, and white); node i performs `TrimOp1`, and uses estimates of only nodes in $\mathcal{R}_i^{(j)}[k]$. (c) $\mathcal{N}_i^{(j)}$ neither has a trusted node nor three distinct colored nodes; node i performs `TrimOp2`, and uses estimates of only nodes in $\mathcal{P}_i^{(j)}[k]$.

For estimating $x^{(j)}[k]$, where $\lambda_j \in \bar{\mathcal{O}}_i$, each node $i \in \mathcal{R}$ first collects the estimates $x^{(j)}[k]$ received from *only* those neighbors that belong to $\mathcal{N}_i^{(j)} \subseteq \mathcal{N}_i$.[8] If there exist trusted nodes in $\mathcal{N}_i^{(j)}$ (line 2 of Algo. 1), then node i only uses the estimates of such trusted nodes, as is evident from (4). If $\mathcal{N}_i^{(j)} \cap \mathcal{T} = \emptyset$, but there exist three distinct colored nodes in $\mathcal{N}_i^{(j)}$, then node i exploits the diversity of its neighborhood to perform `TrimOp1`, and then updates $\hat{x}_i^{(j)}[k]$ via (5). Finally, if $\mathcal{N}_i^{(j)} \cap \mathcal{T} = \emptyset$, and there does not exist three distinct colored nodes in $\mathcal{N}_i^{(j)}$, then node i performs `TrimOp2` to filter out extreme estimates, and then updates $\hat{x}_i^{(j)}[k]$ via (6). We refer to the above algorithm (Algo. 1) as the Local-Filtering based Resilient Estimation (LFRE) algorithm for f-local mono-chromatic Byzantine adversaries; the steps of this algorithm are illustrated in Fig. 3. In the following key result, we establish the convergence guarantees of Algorithm 1.

Theorem 2. *Consider the system* (1) *and measurement model* (2)*, and suppose Assumption 3 holds. Let the communication graph $\mathcal{G}$ be strongly $((2f+1), \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j, \forall \lambda_j \in \Omega_U(\mathbf{A})$. Then, the LFRE algorithm for f-local mono-chromatic Byzantine adversaries solves Problem 1.* □

Proof. Consider an f-local mono-chromatic Byzantine adversarial set $\mathcal{A}$, and let $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$. Based on Lemma 1, notice that a regular node $i \in \mathcal{R}$ can asymptotically estimate each component of the state vector $\mathbf{x}[k]$ corresponding to its set of detectable eigenvalues $\mathcal{O}_i$. It remains to show that node $i \in \mathcal{R}$ can also recover $x^{(j)}[k], \forall \lambda_j \in \bar{\mathcal{O}}_i$, based on Algorithm 1. To this end, we argue that for each $\lambda_j \in \Omega_U(\mathbf{A})$, $\hat{x}_i^{(j)}[k]$ converges to $x[k]$ asymptotically for all $i \in \mathcal{R}$.

Consider any $\lambda_j \in \Omega_U(\mathbf{A})$, and notice that based on Theorem 1, there exists a sub-graph $\mathcal{G}_j$ satisfying all the properties of a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG. Specifically, the set of regular nodes $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$ can be partitioned into

[6] Here, we have suppressed the dependence of the indices n_p on i, j and k for clarity of exposition.

[7] In words, from each end, node i keeps rejecting estimates until it encounters a node with color different from that of the node with the most extreme estimate on that end. See Fig. 3(b) for an illustration of this step.

[8] Recall that $\mathcal{N}_i^{(j)}$ represents neighbors of node i in the MEDAG $\mathcal{G}_j$.

disjoint levels $\{\mathcal{L}_0^{(j)}, \ldots, \mathcal{L}_q^{(j)}, \ldots, \mathcal{L}_{T_j}^{(j)}\}$. We induct on the level number q. For $q=0$, since $\mathcal{L}_0^{(j)} = \mathcal{S}_j \cap \mathcal{R}$, it follows from Lemma 1 that for each $i \in \mathcal{L}_0^{(j)}$, $\lim_{k\to\infty} e_i^{(j)}[k] = 0$, where $e_i^{(j)}[k] = \hat{x}_i^{(j)}[k] - x^{(j)}[k]$. Next, consider a node $i \in \mathcal{L}_1^{(j)}$. We split our subsequent analysis into three separate cases.

Case 1: Suppose $\mathcal{N}_i^{(j)} \cap \mathcal{T} \neq \emptyset$. Then, based on lines 2 and 3 of Algorithm 1, node i employs the update rule (4). In this case, the error $e_i^{(j)}[k]$ evolves as follows:

$$e_i^{(j)}[k+1] = \lambda_j \left(\textstyle\sum_{l \in \mathcal{N}_i^{(j)} \cap \mathcal{T}} \bar{w}_{il}^{(j)} e_l^{(j)}[k] \right), \tag{7}$$

where we used that (i) $x^{(j)}[k+1] = \lambda_j x^{(j)}[k]$ based on the structure of the $\mathbf{A}$ matrix, and (ii) the convexity of the weights $\bar{w}_{il}^{(j)}$. Based on the fact that $\mathcal{T} \subseteq \mathcal{R}$, and property (ii) of a MEDAG in Defn. 6, we have that $\mathcal{N}_i^{(j)} \cap \mathcal{T} \subseteq \mathcal{L}_0^{(j)}$. It then follows from (7) that $\lim_{k\to\infty} e_i^{(j)}[k] = 0$.

Case 2: Suppose $\mathcal{N}_i^{(j)} \cap \mathcal{T} = \emptyset$, but there exist three distinct colored nodes in $\mathcal{N}_i^{(j)}$. Then, based on lines 4 and 5 of Algorithm 1, node i employs the update rule (5). In this case, the error $e_i^{(j)}[k]$ evolves as follows:

$$e_i^{(j)}[k+1] = \lambda_j \left(\textstyle\sum_{l \in \mathcal{R}_i^{(j)}[k]} \tilde{w}_{il}^{(j)}[k] e_l^{(j)}[k] \right), \tag{8}$$

where we have once again used that $x^{(j)}[k+1] = \lambda_j x^{(j)}[k]$, and that the weights $\tilde{w}_{il}^{(j)}[k]$ are convex. Observe that whenever $\mathcal{N}_i^{(j)}$ contains three distinct colored nodes, $\mathcal{R}_i^{(j)}[k]$ is guaranteed to be non-empty by definition. We now claim that at each time-step k, $e_l^{(j)}[k]$ lies in the convex hull of the points $e_s^{(j)}[k], s \in \mathcal{L}_0^{(j)}$, for all $l \in \mathcal{R}_i^{(j)}[k]$. To this end, fix a time-step k, and suppose that the node with the highest estimate of $x^{(j)}[k]$ in $\mathcal{N}_i^{(j)}$, namely node n_1, is regular. Then, we have that for each $l \in \mathcal{R}_i^{(j)}[k]$, $\hat{x}_l^{(j)}[k] \leq \hat{x}_{n_1}^{(j)}[k]$, where $n_1 \in \mathcal{N}_i^{(j)} \cap \mathcal{R} \subseteq \mathcal{L}_0^{(j)}$. The last inclusion follows from property (ii) in Defn. 6. Now consider the case when node n_1 is adversarial. Then, given the mono-chromaticity of the adversarial model, it must be that node n_m, as defined in Step 3, is regular, since $\Delta(n_m) \neq \Delta(n_1)$. Furthermore, based on how $\mathcal{R}_i^{(j)}[k]$ is defined in Step 3, it follows that for each $l \in \mathcal{R}_i^{(j)}[k]$, $\hat{x}_l^{(j)}[k] \leq \hat{x}_{n_m}^{(j)}[k]$, where $n_m \in \mathcal{N}_i^{(j)} \cap \mathcal{R} \subseteq \mathcal{L}_0^{(j)}$. Thus, we have established that at each time-step k, $e_l^{(j)}[k] \leq \max_{s \in \mathcal{L}_0^{(j)}} e_s^{(j)}[k]$, $\forall l \in \mathcal{R}_i^{(j)}[k]$. An identical argument reveals that at each time-step k, $e_l^{(j)}[k] \geq \min_{s \in \mathcal{L}_0^{(j)}} e_s^{(j)}[k]$, $\forall l \in \mathcal{R}_i^{(j)}[k]$. The above discussion, coupled with (8), and the fact that $\lim_{k\to\infty} e_s^{(j)}[k] = 0, \forall s \in \mathcal{L}_0^{(j)}$, readily implies that $\lim_{k\to\infty} e_i^{(j)}[k] = 0$.

Case 3: Suppose $\mathcal{N}_i^{(j)} \cap \mathcal{T} = \emptyset$, and there does not exist three distinct colored nodes in $\mathcal{N}_i^{(j)}$. Then, based on property (i) of a MEDAG in Defn. 6, it must be that $|\mathcal{N}_i^{(j)}| \geq (2f+1)$. In this case, based on lines 6 and 7 of Algorithm 1, node i employs the update rule (6), which corresponds precisely to the resilient filtering algorithm developed in [20] for f-local Byzantine adversarial models. Thus, for this case, the fact that $\lim_{k\to\infty} e_i^{(j)}[k] = 0$ follows directly from arguments in [20].

This completes the analysis for the base case $q=1$. Using arguments similar to those for the base case, and a simple inductive reasoning as in [20], one can establish that the result holds for all levels $q \in \{1, \ldots, T_j\}$. □

Utility of Theorem 2: In Theorem 2, by identifying appropriate graph-theoretic conditions, we formalize how the ideas of trust and diversity serve as viable alternatives to redundancy, when it comes to solving Problem 1. These alternatives may prove to be particularly useful when the number of adversaries f in the neighborhood of each regular node can be large, which would require a large number of redundant links and nodes under traditional approaches. In such cases, a small number of trusted nodes, or nodes of just three different types, can fulfill the requirements for solving Problem 1, without requiring substantial redundancy.

On a related note, if a regular node does not have a proper sense of the number f, and happens to underestimate it, then the approach developed in [20] is no longer guaranteed to work. Algorithm 1 provides a way to overcome such a situation by exploiting the mechanisms of trust and diversity.

C. Resilient Distributed State Estimation Under Mono-chromatic Byzantine Adversaries

Thus far, we have studied the case where the adversarial set $\mathcal{A}$ is f-local, the idea being to account for scenarios where the adversary is resource-limited, and/or faces an increasing risk of getting detected with each component it compromises. In this section, we turn our attention to a more powerful adversary model where such considerations no longer apply. Specifically, we will see how the developments in the previous section can enable us to relax the assumption of f-locality typically made in the literature on resilient distributed algorithms [20]–[32], and allow an adversary to compromise an *arbitrary* number of nodes of a particular type. Our philosophy here is as follows: once an adversary has figured out a way to breach the security of a particular type of component (node), it is in its interest to compromise more (if not all) nodes of that type, if this does not incur any additional resource or risk on its part.

The appropriate concept that we need here is $(\Delta(\cdot), \mathcal{T})$ reachability - a special case of $(r, \Delta(\cdot), \mathcal{T})$-reachability in Defn. 3 with $r = \infty$, where the reachability condition can only be satisfied via diversity or trust. The more stringent concept of $(\Delta(\cdot), \mathcal{T})$-reachability seeks to make up for the inadequacy of the traditional notion of redundancy in coping with a general mono-chromatic Byzantine adversarial model. Indeed, once f-locality is relaxed, a node may have direct or indirect paths from several informative nodes and, yet, fall short of estimating the state dynamics. In particular, an adversary can compromise all such informative nodes if they are of the same type, and not a part of the trusted set $\mathcal{T}$. This highlights the importance of incorporating diversity and/or trust into the measurement and communication structure of the network as alternatives to incorporating redundancy.

Note that a strongly $(\Delta(\cdot), \mathcal{T})$-robust graph w.r.t. $\mathcal{S}$ and a $(\Delta(\cdot), \mathcal{T})$ MEDAG are simply special cases of Defn.'s 4 and 6, respectively, where the redundancy parameter is ∞. Then, following identical arguments as in Theorem 1, one can establish that for each $\lambda_j \in \Omega_U(\mathbf{A})$, $\mathcal{G}$ contains a $(\Delta(\cdot), \mathcal{T})$ MEDAG for λ_j if and only if $\mathcal{G}$ is strongly $(\Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$. To estimate its locally undetectable portion of the state, suppose each node $i \in \mathcal{R}$ executes *only* lines 1-5 of Algorithm 1 in Section III-B to update $\hat{x}_i^{(j)}[k], \forall \lambda_j \in \bar{\mathcal{O}}_i$, i.e., `TrimOp2` is never performed. Let us call this algorithm the LFRE algorithm for mono-chromatic Byzantine adversaries. We then have the following result; we omit its proof since it is similar to that of Theorem 2.

Theorem 3. *Consider the system* (1) *and measurement model* (2)*, and suppose Assumption* 3 *holds. Let the communication graph $\mathcal{G}$ be strongly $(\Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j, \forall \lambda_j \in \Omega_U(\mathbf{A})$. Then, the LFRE algorithm for mono-chromatic Byzantine adversaries solves the variant of Problem* 1 *corresponding to a mono-chromatic Byzantine adversary model.* □

Remark 2. ***(Implications for Countering Spoofing Attacks)****: Recently, in the context of multi-robot coordination, the authors in* [39], [40] *propose methods to tackle the so called "Sybil attack", where an attacker spoofs or impersonates the identities of existing agents to gain a disproportionate advantage in the network. The methods developed in* [39], [40] *are based on analyzing the physics of wireless signals. Since such signals are invariably corrupted by environment and channel noise, the guarantees in* [39], [40] *are of a probabilistic nature. In contrast, we claim that the ideas developed in this section can provide deterministic guarantees in the face of spoofing attacks. The key enabling observation here is that even if an adversary generates multiple identities of an existing regular node, each such identity would share the same digital signature as that of the node being replicated. In other words, the node being spoofed along with its replicated identities would all be of the same type, or color. Thus, regardless of the number of fake identities, as long as the conditions in Theorem* 3 *hold, our techniques would go through.* □

IV. Augmenting Strong-Robustness via Trust

In the previous section, we argued that the traditional notion of redundancy, by itself, proves to be ineffective in coping with a mono-chromatic Byzantine adversary model, thereby necessitating the presence of trusted nodes and/or node-diversity. In this section, we revert back to the f-local mono-chromatic Byzantine adversary model, and demonstrate how incorporating trusted nodes into the network complements redundancy, and helps to significantly augment the strong-robustness property in Definition 4.[9] In the absence of any trusted nodes and node-diversity, solving Problem 1 based on the approach developed in [20] requires the graph $\mathcal{G}$ to be strongly $(2f+1)$-robust w.r.t. $\mathcal{S}_j, \forall \lambda_j \in \Omega_U(\mathbf{A})$. In the following result, we isolate the impact of trusted nodes, and show how their presence can relax the graph-theoretic conditions in [20].

Theorem 4. *Consider a graph $\mathcal{G}$ where all nodes are of the same color, with a non-empty trusted node set $\mathcal{T}$. Suppose the largest integer p for which $\mathcal{G}$ is strongly $(p, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$, is finite, and given by r. Let $\bar{\mathcal{G}}$ be a graph obtained from $\mathcal{G}$ by replacing each trusted node $\tau \in \mathcal{T}$ with a set of r nodes such that each of the r nodes have (i) the same measurements as τ, and (ii) the same in- and out-neighborhood as τ in $\mathcal{G}$. Let $\bar{\mathcal{S}}_j$ denote the new source node set for λ_j in $\bar{\mathcal{G}}$. Then, the largest integer p for which $\bar{\mathcal{G}}$ is strongly p-robust w.r.t. $\bar{\mathcal{S}}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$, is r.* □

Proof. Let the node set $\mathcal{V}$ of $\mathcal{G}$ be partitioned as $\mathcal{W} \cup \mathcal{T}$, where $\mathcal{T} = \{\tau_1, \dots, \tau_{|\mathcal{T}|}\}$ constitutes the set of trusted nodes in $\mathcal{G}$. Let each trusted node τ_i in $\mathcal{G}$ be replaced by the set of r nodes $\{\tau_i^1, \cdots, \tau_i^r\}$ in $\bar{\mathcal{G}}$. The node set of $\bar{\mathcal{G}}$ is then $\bar{\mathcal{V}} = \mathcal{W} \cup \mathcal{T}'$, where $\mathcal{T}' = \{\tau_1^1, \dots, \tau_1^r, \dots, \tau_{|\mathcal{T}|}^1, \dots, \tau_{|\mathcal{T}|}^r\}$.

We first argue that $\bar{\mathcal{G}}$ is strongly r-robust w.r.t. $\bar{\mathcal{S}}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$. To this end, consider any $\lambda_j \in \Omega_U(\mathbf{A})$. Let us observe the following simple facts that are a direct consequence of the way measurements are allocated to nodes in $\mathcal{T}'$: (i) $\bar{\mathcal{S}}_j = \{\mathcal{S}_j \cap \mathcal{W}\} \cup \{\{\tau_i^l\}_{l=1}^r : \tau_i \in \mathcal{S}_j\}$, and (ii) $\bar{\mathcal{V}} \setminus \bar{\mathcal{S}}_j = \{\{\mathcal{V} \setminus \mathcal{S}_j\} \cap \mathcal{W}\} \cup \{\{\tau_i^l\}_{l=1}^r : \tau_i \in \{\mathcal{V} \setminus \mathcal{S}_j\}\}$, where $\bar{\mathcal{V}}$ is the node set of $\bar{\mathcal{G}}$. Consider any non-empty set $\bar{\mathcal{C}} \subseteq \bar{\mathcal{V}} \setminus \bar{\mathcal{S}}_j$. To prove that $\bar{\mathcal{G}}$ is strongly r-robust w.r.t. $\bar{\mathcal{S}}_j$, we need to establish that $\bar{\mathcal{C}}$ is r-reachable in $\bar{\mathcal{G}}$. To this end, consider the non-empty set $\mathcal{C} = (\bar{\mathcal{C}} \cap \mathcal{W}) \cup \{\tau_i : \tau_i^l \in \bar{\mathcal{C}}$ for some $l \in \{1, \cdots, r\}\}$. Based on the above discussion, it is clear that $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S}_j$. Then, based on the hypothesis of the theorem, $\mathcal{C}$ is $(r, \Delta(\cdot), \mathcal{T})$-reachable in $\mathcal{G}$. Since all nodes are of the same color in $\mathcal{G}$, it must be that $\mathcal{C}$ contains at least one node that either has r neighbors outside $\mathcal{C}$, or a trusted neighbor outside $\mathcal{C}$ (see Defn. 3). Given the way $\bar{\mathcal{G}}$ is constructed, it is not hard to verify that this necessarily implies r-reachability of $\bar{\mathcal{C}}$ in $\bar{\mathcal{G}}$.

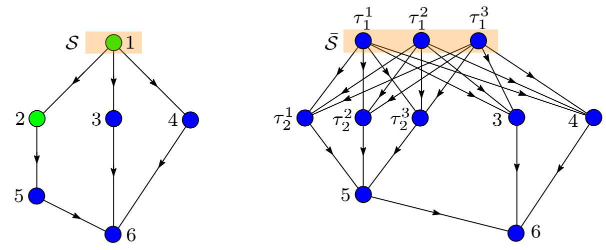
We now claim that $\bar{\mathcal{G}}$ is *not* strongly $(r+1)$-robust w.r.t. $\bar{\mathcal{S}}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$. To see this, note from the hypothesis of the theorem that there must exist some mode $\lambda_q \in \Omega_U(\mathbf{A})$, such that $\mathcal{G}$ is not strongly $(r+1, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_q$. This in turn implies the existence of a non-empty set $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S}_q$ that is not $(r+1, \Delta(\cdot), \mathcal{T})$-reachable in $\mathcal{G}$. Now consider the non-empty set $\bar{\mathcal{C}} = \{\mathcal{C} \cap \mathcal{W}\} \cup \{\{\tau_i^l\}_{l=1}^r : \tau_i \in \mathcal{C}\}$. It is easy to verify that $\bar{\mathcal{C}} \subseteq \bar{\mathcal{V}} \setminus \bar{\mathcal{S}}_q$, and that $\bar{\mathcal{C}}$ is not $(r+1)$-reachable in $\bar{\mathcal{G}}$. This establishes our claim. □

Implication of Theorem 4: Roughly speaking, the above result suggests that replacing even a single trusted node may require allocating additional measurement and communication resources to the network so as to preserve the same level of robustness. In particular, as revealed by Theorem 4, the number of such additional resources scales linearly with the desired level of resilience r. Thus, for cases where r is large, i.e., several adversaries need to be tolerated, our result identifies the potential benefits that can be reaped by making a small fraction of the nodes trusted. An illustration of the above result is given in Figure 4. $\mathcal{G}$ has two trusted nodes $\mathcal{T} = \{1, 2\}$, and $\mathcal{S} = \{1\}$ is the source node. In $\bar{\mathcal{G}}$, each trusted node is replaced by three nodes, each of which has the same in- and out-neighborhood as the corresponding trusted node in $\mathcal{G}$. Moreover, each of the nodes τ_1^1, τ_1^2, and τ_1^3 have the same measurements as node 1 in $\mathcal{G}$.

V. On the Complexity of Incorporating Diversity and Trust

In practice, hardening sensors against attacks (i.e., making

[9] Similar conclusions can be drawn when it comes to incorporating diversity; we omit such a discussion in the interest of space.



(a) $\mathcal{G}$ (b) $\bar{\mathcal{G}}$

Fig. 4: (a) $\mathcal{G}$ is strongly $(3, \Delta(.), \mathcal{T})$-robust w.r.t. $\mathcal{S} = \{1\}$. (b) $\bar{\mathcal{G}}$ is strongly 3-robust w.r.t. $\bar{\mathcal{S}} = \{\tau_1^1, \tau_1^2, \tau_1^3\}$.

nodes trusted), and implementing several variants of nodes (i.e., making the network diverse), comes at a cost. Thus, it is natural to consider the design problem of (i) finding a trusted set of minimum cardinality; and/or (ii) finding the minimum number of colors, and the corresponding allocation of colors to nodes, so as to make the resulting network strongly-robust to a desired extent. In what follows, we separately explore the complexity of each of these problems.

A. *On the Complexity of Selecting Trusted Nodes*

To isolate the complexity of selecting trusted nodes, we consider a scenario where all nodes are of the same color (i.e., $\Delta(i) = \Delta(j), \forall i, j \in \mathcal{V}$). To proceed, we formally state the problem of interest and then characterize its complexity.

Problem 2. *(**Trusted Strong-Robustness Augmentation (TSRA**)) Given a system model* (1)*, a measurement model* (2)*, a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where all nodes are of the same color (i.e., $\Delta(i) = \Delta(j), \forall i, j \in \mathcal{V}$), and positive integers r, t, does there exist a set of trusted nodes $\mathcal{T}$ of cardinality t, such that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$?*

To characterize the complexity of the TSRA problem, we will provide a reduction from the NP-hard Set Cover (SC) problem, defined as follows.

Definition 7. *(**Set Cover (SC)**) Given a collection of elements $\mathcal{U} = \{1, \ldots, p\}$, a set of subsets $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ of $\mathcal{U}$, and a positive integer t, do there exist t subsets in $\mathcal{F}$ whose union is $\mathcal{U}$?* □

Theorem 5. *The TSRA problem is NP-complete.* □

The proof of the above result is presented in Appendix B. We now briefly describe a simple greedy heuristic, namely Algorithm 2, that finds a potentially sub-optimal set of trusted nodes in polynomial time.

Greedy Heuristic for Selecting Trusted Nodes: Consider the setup in Problem 2, and suppose we need to find a set of trusted nodes $\mathcal{T}$ such that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$. We proceed as follows. Fix a $\lambda_j \in \Omega_U(\mathbf{A})$, and suppose each node $i \in \mathcal{V} \setminus \mathcal{S}_j$ is reachable from $\mathcal{S}_j$ (since otherwise, there is no hope of achieving the desired property). Our proposed greedy algorithm proceeds in rounds l, where in each round precisely one node is made trusted, if needed. Two lists are maintained and updated each round: a list of "active" nodes $\mathcal{W}_j(l)$, and a list of trusted nodes $\mathcal{T}_j(l)$, with $\mathcal{W}_j(0)$ initially set to $\mathcal{S}_j$, and $\mathcal{T}_j(0)$ to $\emptyset$. At the beginning of round l, where $l \geq 1$, each node in $\mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)$ is a candidate for being made trusted in that round. For each such candidate node $v \in \mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)$, we run a virtual bootstrap percolation[10] process by making node v trusted temporarily, and computing the number of new nodes it activates in the process. Here, an inactive node gets activated if it either has at least r active neighbors, or a trusted active neighbor. Let $\delta(v)$ denote the new nodes activated by node v. Having run this virtual percolation process separately for each $v \in \mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)$, we greedily pick $\tau(l) \in \operatorname{argmax}_{v \in \mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)} |\delta(v)|$ to be trusted in round l, i.e., we pick the node that activates the maximum number of new nodes. Subsequently, we update $\mathcal{W}_j(l) = \mathcal{W}_j(l-1) \cup \delta(\tau(l))$, and $\mathcal{T}_j(l) = \mathcal{T}_j(l-1) \cup \tau(l)$. Let $\bar{l}_j$ be the smallest integer such that $\mathcal{W}_j(\bar{l}_j) = \mathcal{V}$. We then say that the greedy algorithm described above terminates in round $\bar{l}_j$. It is easy to see that $\bar{l}_j \leq N-1$, and that on termination, $\mathcal{T}_j(\bar{l}_j)$ is such that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T}_j(\bar{l}_j))$-robust w.r.t. $\mathcal{S}_j$. Thus, we can run the above greedy heuristic for each $\lambda_j \in \Omega_U(\mathbf{A})$, and obtain the desired trusted set $\mathcal{T} = \cup_{\lambda_j \in \Omega_U(\mathbf{A})} \mathcal{T}_j(\bar{l}_j)$.

Algorithm 2 Greedy heuristic to identify trusted set for each $\lambda_j \in \Omega_U(\mathbf{A})$

```
Initialization: Set $l = 0$, initial active set $\mathcal{W}_j(0) = \mathcal{S}_j$, and initial trusted set $\mathcal{T}_j(0) = \emptyset$.
while $\mathcal{W}_j(l) \neq \mathcal{V}$ do
    Set $l = l + 1$.
    for $v \in \mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)$ do
        Run a virtual bootstrap percolation process with node $v$ made trusted temporarily.
        Identify set of new nodes $\delta(v)$ activated by node $v$.
    end for
    Greedily pick any node $\tau(l)$ satisfying
        $$\tau(l) \in \operatorname*{argmax}_{v \in \mathcal{W}_j(l-1) \setminus \mathcal{T}_j(l-1)} |\delta(v)|.$$
    Update $\mathcal{W}_j(l) = \mathcal{W}_j(l-1) \cup \delta(\tau(l))$.
    Update $\mathcal{T}_j(l) = \mathcal{T}_j(l-1) \cup \tau(l)$.
end while
Return $\mathcal{T}_j(l)$.
```

A rigorous theoretical characterization of the performance of the above greedy heuristic is beyond the scope of this paper. However, it is not too hard to verify that this heuristic does output a trusted set of optimal size for simple graphs such as star graphs, directed trees, rings, and complete graphs.

B. *On the Complexity of Allocating Diversity*

We now turn our attention to the problem of allocating colors to the nodes from a set of specified cardinality so

[10] Given a graph $\mathcal{G}$ and a threshold $r \geq 2$, bootstrap percolation is a process of spread of activation where one starts off with an initially active set. The process then evolves over the network in rounds, where in each round an inactive node becomes active if and only if it has at least r active neighbors; here, we modify the activation rule to suit our purpose.

as to achieve a certain level of strong-robustness. To isolate the challenges associated with this problem, our subsequent analysis will focus exclusively on scenarios where the trusted set $\mathcal{T}$ is empty. Next, we formally state the problem of interest.

Problem 3. *(**q-Colored Strong-Robustness Augmentation (q-CSRA)**) Given a system model (1), a measurement model (2), a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an empty trusted set $\mathcal{T}$, and positive integers r, q, does there exist an allocation $\Delta : \mathcal{V} \rightarrow \{1, \ldots, q\}$, such that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$?*

Let us note that when $q < 3$, the q-CSRA problem as stated above boils down to checking whether the given graph $\mathcal{G}$ is strongly r-robust w.r.t. $\mathcal{S}_j$, $\forall \lambda_j \in \Omega_U(\mathbf{A})$. In [20], by exploiting a connection to the process of bootstrap percolation, it was shown that this can be done in polynomial-time. Thus, the complexity of the q-CSRA problem remains to be characterized only when $q \geq 3$. In the remainder of this section, we establish that the 3-CSRA problem is computationally hard by providing a reduction from the NP-complete 3-Disjoint Set Cover (3-DSC) problem, defined as follows [41].

Definition 8. *(**3-Disjoint Set Cover (3-DSC)**) Given a collection of elements $\mathcal{U} = \{1, \ldots, p\}$, and a set of subsets $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ of $\mathcal{U}$, can $\mathcal{F}$ be partitioned into three disjoint collections of subsets, such that the union of the subsets within each such collection covers $\mathcal{U}$?* □

Theorem 6. *The 3-CSRA problem is NP-complete.* □

The proof of the above result is presented in Appendix C. At the moment, we do not have a clean heuristic algorithm to allocate diversity; we reserve this as future work.

VI. Conclusion

We introduced novel graph-theoretic constructs to study the impacts of redundancy, diversity, and trust in the context of resilient distributed state estimation. We then proposed an attack-resilient algorithm that appropriately leverages each of the three above facets, and provides provable guarantees. Roughly speaking, we established that even relatively sparse networks that are either diverse, or contain a small subset of trusted nodes, can exhibit the same functional robustness as densely connected networks. Finally, we separately studied the complexity of (i) selecting a trusted node set, and (ii) allocating diversity, in order to achieve a prescribed level of robustness. Our analysis revealed that each of these problems is NP-complete; in the future, we plan to explore approximation algorithms with provable guarantees for these problems. We are also interested in scenarios where the state transition matrix $\mathbf{A}$ is not known exactly; see, for instance, [42].

References

[1] 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.

[2] V. Ugrinovskii, "Conditions for detectability in distributed consensus-based observer networks," *IEEE Trans. on Autom. Control*, vol. 58, no. 10, pp. 2659–2664, 2013.

[3] T. Kim, H. Shim, and D. D. Cho, "Distributed luenberger observer design," in *Proc. of the 55th IEEE Decision and Control Conference*, 2016, pp. 6928–6933.

[4] S. Park and N. C. Martins, "Design of distributed LTI observers for state omniscience," *IEEE Trans. on Autom. Control*, vol. 62, no. 2, pp. 561–576, 2017.

[5] A. Mitra and S. Sundaram, "Distributed observers for LTI systems," *IEEE Trans. on Autom. Control*, vol. 63, no. 11, pp. 3689–3704, 2018.

[6] L. Wang and A. S. Morse, "A distributed observer for a time-invariant linear system," *IEEE Trans. on Autom. Control*, vol. 63, no. 7, 2018.

[7] W. Han, H. L. Trentelman, Z. Wang, and Y. Shen, "A simple approach to distributed observer design for linear systems," *IEEE Trans. on Autom. Control*, vol. 64, no. 1, pp. 329–336, 2019.

[8] F. F. Rego, A. P. Aguiar, A. M. Pascoal, and C. N. Jones, "A design method for distributed Luenberger observers," in *Proc. of the 56th IEEE Conference on Decision and Control*, 2017, pp. 3374 – 3379.

[9] Á. R. del Nozal, P. Millán, L. Orihuela, A. Seuret, and L. Zaccarian, "Distributed estimation based on multi-hop subspace decomposition," *Automatica*, vol. 99, pp. 213–220, 2019.

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

[11] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. of the American Control Conference*. IEEE, 2015, pp. 2439–2444.

[12] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, 2017.

[13] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "A multi-observer based estimation framework for nonlinear systems under sensor attacks," *Automatica*, vol. 119, p. 109043, 2020.

[14] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[15] Y. Mao, A. Mitra, S. Sundaram, and P. Tabuada, "When is the secure state-reconstruction problem hard?" in *Proc. of the Conference on Decision and Control*, 2019, pp. 5368–5373.

[16] M. Deghat, V. Ugrinovskii, I. Shames, and C. Langbort, "Detection and mitigation of biasing attacks on distributed estimation networks," *Automatica*, vol. 99, pp. 369–381, 2019.

[17] J. Kim, J. G. Lee, C. Lee, H. Shim, and J. H. Seo, "Local identification of sensor attack and distributed resilient state estimation for linear systems," in *Proc. of the 57th IEEE Conference on Decision and Control*, 2018, pp. 2056–2061.

[18] A. Mustafa and H. Modares, "Secure event-triggered distributed kalman filters for state estimation," *arXiv preprint arXiv:1901.06746*, 2019.

[19] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "Secure distributed filtering for unstable dynamics under compromised observations," *arXiv:1903.07345*, 2019.

[20] A. Mitra and S. Sundaram, "Byzantine-resilient distributed observers for LTI systems," *Automatica*, vol. 108, p. 108487, 2019.

[21] A. Mitra, J. A. Richards, S. Bagchi, and S. Sundaram, "Resilient distributed state estimation with mobile agents: overcoming Byzantine adversaries, communication losses, and intermittent measurements," *Autonomous Robots*, vol. 43, no. 3, pp. 743–768, 2019.

[22] 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.

[23] W. Abbas, A. Laszka, and X. Koutsoukos, "Improving network connectivity and robustness using trusted nodes with application to resilient consensus," *IEEE Transactions on Control of Network Systems*, 2017.

[24] F. Ghawash and W. Abbas, "Leveraging diversity for achieving resilient consensus in sparse networks," *IFAC-PapersOnLine*, 2019.

[25] A. Pelc and D. Peleg, "Broadcasting with locally bounded Byzantine faults," *Information Processing Letters*, vol. 93, pp. 109–115, 2005.

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

[27] 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.

[28] S. M. Dibaji and H. Ishii, "Resilient consensus of second-order agent networks: Asynchronous update rules with delays," *Automatica*, vol. 81, pp. 123–132, 2017.

[29] J. Usevitch and D. Panagou, "Resilient leader-follower consensus to arbitrary reference values," in *Proc. of the Annual American Control Conference*. IEEE, 2018, pp. 1292–1298.

[30] S. Sundaram and B. Gharesifard, "Distributed optimization under adversarial nodes," *IEEE Trans. on Autom. Control*, vol. 64, no. 3, pp. 1063–1076, 2019.

[31] L. Su and N. H. Vaidya, “Fault-tolerant multi-agent optimization: optimal iterative distributed algorithms,” in *Proc. of the ACM Symp. on Principles of Distributed Comp.*, 2016, pp. 425–434.

[32] A. Mitra, W. Abbas, and S. Sundaram, “On the impact of trusted nodes in resilient distributed state estimation of LTI systems,” in *Proc. of the IEEE Conference on Decision and Control*, 2018, pp. 4547–4552.

[33] K. M. Lynch, I. B. Schwartz, P. Yang, and R. A. Freeman, “Decentralized environmental modeling by mobile sensor networks,” *IEEE Trans. on Robotics*, vol. 24, no. 3, pp. 710–724, 2008.

[34] A. J. O’Donnell and H. Sethu, “On achieving software diversity for improved network security using distributed coloring algorithms,” in *Proc. of the 11th ACM conference on Computer and communications security*, 2004, pp. 121–131.

[35] A. Newell, D. Obenshain, T. Tantillo, C. Nita-Rotaru, and Y. Amir, “Increasing network resiliency by optimally assigning diverse variants to routing nodes,” *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 6, pp. 602–614, 2014.

[36] M. Touhiduzzaman, A. Hahn, and A. K. Srivastava, “A diversity-based substation cyber defense strategy utilizing coloring games,” *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5405–5415, 2018.

[37] L. Tseng, Y. Wu, H. Pan, M. Aloqaily, and A. Boukerche, “Reliable broadcast in networks with trusted nodes,” in *Proc. of the Global Communications Conference*. IEEE, 2019, pp. 1–6.

[38] A. Mitra, F. Ghawash, S. Sundaram, and W. Abbas, “On the impacts of redundancy, diversity, and trust in resilient distributed state estimation,” *arXiv preprint arXiv:2001.07056*, 2020.

[39] S. Gil, S. Kumar, M. Mazumder, D. Katabi, and D. Rus, “Guaranteeing spoof-resilient multi-robot networks,” *Autonomous Robots*, vol. 41, no. 6, pp. 1383–1400, 2017.

[40] V. Renganathan and T. Summers, “Spoof resilient coordination for distributed multi-robot systems,” in *Proc. of the International Symposium on Multi-Robot and Multi-Agent Systems*, 2017, pp. 135–141.

[41] M. Cardei and D.-Z. Du, “Improving wireless sensor network lifetime through power aware organization,” *Wireless networks*, vol. 11, no. 3, pp. 333–340, 2005.

[42] P. Duan, Z. Duan, G. Chen, and L. Shi, “Distributed state estimation for uncertain linear systems: A regularized least-squares approach,” *Automatica*, vol. 117, p. 109007, 2020.

Appendix A

Construction of a $(2f+1, \Delta(\cdot), \mathcal{T})$ **MEDAG**: We briefly discuss an algorithm that can be used to construct a $(2f+1, \Delta(\cdot), \mathcal{T})$ MEDAG. Suppose we are given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a trusted node set $\mathcal{T}$, where each node $i \in \mathcal{V}$ is assigned a color $\Delta(i)$. For each $\lambda_j \in \Omega_U(\mathbf{A})$, our objective is to construct a sub-graph $\mathcal{G}_j$ satisfying the conditions in Defn. 6 and, in the process, to identify the sets $\mathcal{N}_i^{(j)}$, $\forall i \in \mathcal{V}$. The MEDAG construction algorithm requires each node i to maintain a counter $c_i(j)$ and a list of indices $\mathcal{N}_i^{(j)}$ for each $\lambda_j \in \Omega_U(\mathbf{A})$. These parameters are initialized with $c_i(j) = 0$ and $\mathcal{N}_i^{(j)} = \emptyset$, for each $i \in \mathcal{V}$. Subsequently, the algorithm proceeds in rounds where in round zero, each node in $\mathcal{S}_j$ broadcasts the message “1” to its out-neighbors, sets $c_i(j) = 1$, maintains $\mathcal{N}_i^{(j)} = \emptyset$ for all future rounds, and goes to sleep. A node $i \in \mathcal{V} \setminus \mathcal{S}_j$ waits until it either receives “1” from at least $(2f+1)$ distinct neighbors, or from at least three distinct colored neighbors, or from at least one trusted neighbor. When any one of these conditions is eventually met, it sets $c_i(j) = 1$, appends the labels of each of the neighbors from which it received “1” to $\mathcal{N}_i^{(j)}$, broadcasts the message “1” to its out-neighbors, and goes to sleep. The MEDAG construction algorithm “*terminates for* λ_j” if there exists $T_j \in \mathbb{N}_+$ such that $c_i(j) = 1$ $\forall i \in \mathcal{V}$, for all rounds following round T_j. The **objective** of the algorithm is to return a set of sets $\{\mathcal{N}_i^{(j)}\}$, where $\lambda_j \in \Omega_U(\mathbf{A})$, $i \in \mathcal{V}$.

Appendix B

Proof. (**Proof of Theorem 5**) We first argue that TSRA $\in$ NP. To see this, notice that for “yes” instances of the problem, the set of trusted nodes $\mathcal{T}$ of size t yields a certificate w.r.t. the MEDAG construction algorithm described in Section III-A. Specifically, based on Theorem 1, for each $\lambda_j \in \Omega_U(\mathbf{A})$, the MEDAG construction algorithm terminates if and only if $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}_j$; thus, such an algorithm can be used to verify the desired graph property. That this verification algorithm has polynomial-time complexity follows from an analogous argument made in [20, Proposition 2].

Next, we establish that TSRA is NP-hard. To this end, given an instance of SC, we first construct an instance of TSRA as follows. We consider a scalar unstable dynamical system $x[k+1] = \lambda x[k]$, and construct an associated communication graph $\mathcal{G}$ with node set $\mathcal{V} = \bar{\mathcal{U}} \cup \bar{\mathcal{F}}$, where $\bar{\mathcal{U}} = \{u_1, \ldots, u_p\}$, and $\bar{\mathcal{F}} = \{f_1, \ldots, f_m\}$. For each $i \in \{1, \ldots, p\}$, node $u_i \in \bar{\mathcal{U}}$ corresponds to element i of $\mathcal{U}$, and for each $j \in \{1, \ldots, m\}$, node $f_j \in \bar{\mathcal{F}}$ corresponds to subset $\mathcal{F}_j \in \mathcal{F}$. If $i \in \mathcal{F}_j$, then a directed edge is added from node f_j to node u_i in $\mathcal{G}$. Each node $f_j \in \bar{\mathcal{F}}$ is allocated a non-zero measurement of the state $x[k]$. The cardinality of the trusted set $\mathcal{T}$ is set to t, and the desired level of strong-robustness is given by $r = |\mathcal{F}|$. Clearly, given any instance of SC, the above TSRA instance can be constructed in polynomial-time. We now argue that the answer to any given instance of SC is “yes” if and only if the answer to the constructed instance of TSRA is “yes”.

Suppose the answer to the SC instance is “yes”. Thus, there exists a set of t subsets of $\mathcal{F}$ whose union is $\mathcal{U}$. Without loss of generality, let these subsets be $\{\mathcal{F}_1, \ldots, \mathcal{F}_t\}$. Let the set of trusted nodes $\mathcal{T}$ be $\{f_1, \ldots, f_t\}$. We first observe that the set of source nodes $\mathcal{S}$ (the set of nodes that can detect λ) of $\mathcal{G}$ is precisely the set $\bar{\mathcal{F}}$. Thus, $\mathcal{T} \subseteq \mathcal{S}$. To establish that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$, we pick a non-empty subset $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S} = \bar{\mathcal{U}}$. Since $\{\mathcal{F}_1, \ldots, \mathcal{F}_t\}$ cover $\mathcal{U}$, $\mathcal{N}_{u_i} \cap \mathcal{T} \neq \emptyset, \forall u_i \in \bar{\mathcal{U}}$. Thus, $\mathcal{C}$ is $(r, \Delta(\cdot), \mathcal{T})$-reachable, and the answer to the constructed instance of TSRA is “yes”.

To show the converse, we proceed via contraposition. Suppose the answer to the SC instance is “no”. In other words, no t subsets of $\mathcal{F}$ cover $\mathcal{U}$. Consider any set of trusted nodes $\mathcal{T}$ of cardinality t. Let $\mathcal{M} = \bar{\mathcal{F}} \cap \mathcal{T}$. We first consider the case when $\mathcal{M}$ is non-empty. In this case, there exists at least one node $u_i \in \bar{\mathcal{U}}$ that has neighbors (if any) only in $\bar{\mathcal{F}} \setminus \mathcal{M}$. Noting that the source set $\mathcal{S} = \bar{\mathcal{F}}$, we consider the non-empty set $\mathcal{C} = \{u_i\}$ contained in $\mathcal{V} \setminus \mathcal{S}$. Since $r = |\bar{\mathcal{F}}|$, it follows that u_i neither has a trusted neighbor nor has at least r neighbors. Thus, $\mathcal{C}$ is not $(r, \Delta(\cdot), \mathcal{T})$-reachable. For analyzing the case when $\mathcal{M}$ is empty, we observe that there must exist at least one node $u_i \in \bar{\mathcal{U}}$ such that $\mathcal{N}_{u_i} \subset \bar{\mathcal{F}}$; else, each $\mathcal{F}_j \in \mathcal{F}$ would cover $\mathcal{U}$, and the answer to SC would be trivially “yes”, leading to a contradiction. It then follows that $\mathcal{C} = \{u_i\}$ is not $(r, \Delta(\cdot), \mathcal{T})$-reachable. Thus, $\mathcal{G}$ is not strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$, regardless of the way t trusted nodes are picked in $\mathcal{G}$, and the answer to the constructed TSRA instance is “no”. □

Appendix C

Proof. (**Proof of Theorem 6**) The fact that CSRA $\in$ NP follows an analogous argument as in Theorem 5. In particular, given any "yes" instance of the problem, the associated allocation Δ yields a certificate w.r.t. the MEDAG construction algorithm in Sec. III-A that acts as a polynomial-time verifier.

Given an instance of 3-DSC, we construct an instance of 3-CSRA in a manner identical to that in the proof of Theorem 5, and adhere to the notation used in that proof. Note however that unlike TSRA, the cardinality t of the trusted set $\mathcal{T}$ plays no role in 3-CSRA, and hence requires no specification while constructing the instance of 3-CSRA. It is easy to see that given any instance of 3-DSC, the above 3-CSRA instance can be constructed in polynomial-time. We now argue that the answer to any given instance of 3-DSC is "yes" if and only if the answer to the constructed instance of 3-CSRA is "yes". Throughout the proof, we will assume that $|\mathcal{F}| \geq 3$, as otherwise, the answer to 3-DSC is trivially "no".

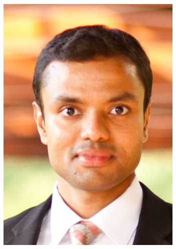
Suppose the answer to the 3-DSC instance is "yes". Thus, $\mathcal{F}$ can be partitioned into 3 disjoint set covers of $\mathcal{U}$. Let these partitions be denoted $\mathcal{P}_1 = \{\mathcal{F}_{i_1}, \ldots, \mathcal{F}_{i_{p_1}}\}$, $\mathcal{P}_2 = \{\mathcal{F}_{j_1}, \ldots, \mathcal{F}_{j_{p_2}}\}$, and $\mathcal{P}_3 = \{\mathcal{F}_{k_1}, \ldots, \mathcal{F}_{k_{p_3}}\}$, where $p_i = |\mathcal{P}_i|, i \in \{1,2,3\}$. Let the corresponding sets of nodes in $\bar{\mathcal{F}}$ be denoted $\bar{\mathcal{P}}_1$, $\bar{\mathcal{P}}_2$ and $\bar{\mathcal{P}}_3$. Consider the following allocation of colors to the nodes in $\bar{\mathcal{F}}$: $\Delta(f_{i_s}) = 1, \forall i_s \in \bar{\mathcal{P}}_1$, $\Delta(f_{j_s}) = 2, \forall j_s \in \bar{\mathcal{P}}_2$, and $\Delta(f_{k_s}) = 3, \forall k_s \in \bar{\mathcal{P}}_3$. The assignment of colors to the nodes in $\bar{\mathcal{U}}$ is arbitrary, i.e., each $u_i \in \bar{\mathcal{U}}$ is assigned any one of the three colors. Noting that the set of source nodes $\mathcal{S}$ is precisely the set $\bar{\mathcal{F}}$, we claim that $\mathcal{G}$ is strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$. To see this, pick any non-empty subset $\mathcal{C} \subseteq \mathcal{V} \setminus \mathcal{S} = \bar{\mathcal{U}}$. Since $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ each cover $\mathcal{U}$, it follows that every $u_i \in \bar{\mathcal{U}}$ has a neighbor in each of the sets $\bar{\mathcal{P}}_1$, $\bar{\mathcal{P}}_2$, and $\bar{\mathcal{P}}_3$, i.e., each $u_i \in \bar{\mathcal{U}}$ has 3 distinct colored neighbors. Thus, $\mathcal{C}$ is $(r, \Delta(\cdot), \mathcal{T})$-reachable, and the answer to the constructed instance of 3-CSRA is "yes".

We now establish the converse. Suppose the answer to the 3-DSC instance is "no". In other words, no matter how one partitions $\mathcal{F}$ into 3 disjoint collections of subsets, not all three such collections can each cover $\mathcal{U}$. We first argue that $\mathcal{G}$ cannot be made strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$, if one uses fewer than three colors to color the set $\bar{\mathcal{F}}$. To see this, note that if fewer than three colors are used to color $\bar{\mathcal{F}}$, then $\mathcal{G}$ will be strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$ if and only if each $f_j \in \bar{\mathcal{F}}$ is a neighbor of every $u_i \in \bar{\mathcal{U}}$, since each u_i would need to have precisely $r = |\bar{\mathcal{F}}|$ neighbors to meet the $(r, \Delta(\cdot), \mathcal{T})$-reachability requirement (recall that $\mathcal{T} = \emptyset$). However, that would imply $\mathcal{F}_j = \mathcal{U}, \forall \mathcal{F}_j \in \mathcal{F}$. This in turn would collapse the size of the set $\mathcal{F}$ to just 1 (since all its elements would be identical), contradicting the fact that $|\mathcal{F}| \geq 3$.

Next, consider any allocation of these colors to the nodes in $\bar{\mathcal{F}}$, where each of the three colors is used at least once. Such a coloring naturally partitions $\bar{\mathcal{F}}$ into 3 disjoint non-empty sets, say $\bar{\mathcal{P}}_1$, $\bar{\mathcal{P}}_2$ and $\bar{\mathcal{P}}_3$. Since the answer to 3-DSC is "no", there must exist some node $u_i \in \bar{\mathcal{U}} = \mathcal{V} \setminus \mathcal{S}$, such that u_i contains at most 2 distinct colored neighbors from $\bar{\mathcal{F}}$. Since $\bar{\mathcal{P}}_1$, $\bar{\mathcal{P}}_2$ and $\bar{\mathcal{P}}_3$ are each non-empty, and $r = |\bar{\mathcal{F}}| = |\bar{\mathcal{P}}_1| + |\bar{\mathcal{P}}_2| + |\bar{\mathcal{P}}_3|$, it follows that $|\mathcal{N}_{u_i}| < r$. Consequently, $\{u_i\} \in \mathcal{V} \setminus \mathcal{S}$ is not $(r, \Delta(\cdot), \mathcal{T})$-reachable, and there does not exist any allocation $\Delta : \mathcal{V} \rightarrow \{1,2,3\}$ that renders $\mathcal{G}$ strongly $(r, \Delta(\cdot), \mathcal{T})$-robust w.r.t. $\mathcal{S}$. The answer to the constructed instance of 3-CSRA is thus "no". This completes the proof. □

Aritra Mitra is a Postdoctoral Researcher in the Department of Electrical and Systems Engineering, University of Pennsylvania. He received the Ph.D. degree from Purdue University, USA, the M.Tech. degree from the Indian Institute of Technology Kanpur, India, and the B.E. degree from Jadavpur University, Kolkata, India, in 2020, 2015, and 2013, respectively, all in Electrical Engineering. His current research interests include distributed learning and optimization, statistical signal processing, networked control systems, and secure control. He was a recipient of the University Gold Medal at Jadavpur University and the Academic Excellence Award at IIT Kanpur.

Faiq Ghawash received the B.Sc. degree in Electrical Engineering from University of Engineering and Technology, Lahore, Pakistan and the M.Sc. degree in Electrical engineering from Information Technology University, Lahore, Pakistan, in 2017 and 2019 respectively. He is currently a Ph.D. research fellow at the Department of Engineering and Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway. His research focuses on optimal control, model predictive control and network control systems.



Shreyas Sundaram is an Associate Professor in the School of Electrical and Computer Engineering at Purdue University. He received his MS and PhD degrees in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2005 and 2009, respectively. He was a Postdoctoral Researcher at the University of Pennsylvania from 2009 to 2010, and an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Waterloo from 2010 to 2014. He is a recipient of the NSF CAREER award, and an Air Force Research Lab Summer Faculty Fellowship. At Purdue, he received the Hesselberth Award for Teaching Excellence and the Ruth and Joel Spira Outstanding Teacher Award. At Waterloo, he received the Department of Electrical and Computer Engineering Research Award and the Faculty of Engineering Distinguished Performance Award. He received the M. E. Van Valkenburg Graduate Research Award and the Robert T. Chien Memorial Award from the University of Illinois, and he was a finalist for the Best Student Paper Award at the 2007 and 2008 American Control Conferences. His research interests include network science, analysis of large-scale dynamical systems, fault-tolerant and secure control, linear system and estimation theory, game theory, and the application of algebraic graph theory to system analysis.

Waseem Abbas is a Research Assistant Professor in the Electrical Engineering and Computer Science Department at the Vanderbilt University, Nashville, TN, USA. Previously, he was an Assistant Professor at the Information Technology University Lahore in Pakistan, and a postdoctoral research scholar at the Vanderbilt University between 2014 and 2017. He received Ph.D. (2013) and M.Sc. (2010) degrees, both in Electrical and Computer Engineering, from Georgia Institute of Technology, Atlanta, GA, and was a Fulbright scholar from 2009 till 2013. His research interests are in the areas of resilience and robustness of network control systems, cyber-physical systems, and graph-theoretic methods in complex networks.