Distributed Resilient Consensus on General Digraphs under Cyber-attacks

Muhammad Iqbal¹, Zhihua Qu² and Azwirman Gusrialdi¹

Abstract—This paper presents a distributed resilient consensus protocol based on competitive interaction design method to solve consensus problem on general digraphs containing directed spanning tree in the presence of cyber-attacks. The attacker aims at destabilizing the consensus dynamics by intercepting the system’s communication network. The competitive interaction method allows us to design a virtual network that protects the multi-agent systems from adversaries without requiring high network connectivity and global information about the number of adversaries. In fact, the proposed distributed consensus protocol enables agents to achieve consensus in the presence of any number of bounded adversaries and also when the virtual network is being attacked. In addition, the proposed virtual network also improves the performance of consensus algorithm in the absence of attacks. Simulations are presented to illustrate our theoretical results.

I. INTRODUCTION

Consensus problem in multi-agent systems (MAS) has been a topic of great interest for the last two decades in control community due to numerous applications such as power networks [1], satellite formation [2], and cooperative search and rescue operations [3]. Distributed control techniques, which solve consensus problem in MAS rely on neighboring information shared through communication links. The data shared on these communication links are vulnerable to cyberattacks [4], and consequently the entire MAS is at jeopardy. Therefore, resilient distributed consensus protocols need to be designed to achieve consensus in the presence of cyber-attacks.

The Adversarial robust consensus protocol (ARC-P) has shown its effectiveness in solving resilient consensus problem without requiring any assumption on the attacker’s behavior [5] together with communication link failure and noisy channels [6]. With extra infrastructure to share local information with a distant neighbor, it is shown that the ARC-P algorithm works well for a platoon of network moving on a highway [7]. However, the strategy requires knowledge on the upper bound of the maximum number of attacks and also poses a restriction on the network topology. The global information requirement about the number of attacks is relaxed in [8], but an additional assumption is added on the robustness of graph, namely \((F + 1)\)excess robustness, where \(F\) is the number of attacks on a node. Joint-agent interactions based protocol also solve the resilient consensus problem [9] without assuming the knowledge of the global information regarding the number of adversaries attacking each agent. However, the connectivity requirements are the same as that of the ARC-P. In addition, due to discarding some relevant and useful information during the consensus process to ensure safety, the state of the agents may not converge to the neighborhood of the original consensus value (i.e., the consensus value in the case of no attacks). An observer based method is another way of solving resilient consensus problem [10]–[12]. Output regulation problem for heterogeneous multi-agent systems can be solved in the presence of Denial-of-Service attack by designing an observer-based distributed resilient control strategy [10]. However, the problem is solved for undirected graph and in the presence of an exosystem which is protected. The observer based resilient consensus control strategies [11], [12] assumed that the communication links are protected in addition to the assumption that a trusted leader exist in a network.

To relax the assumptions in ARC-P, a competitive interaction design based consensus protocol was proposed in [13], [14], in which a virtual network is designed having the same number of agents as that of the physical network to protect the MAS from cyber-attacks. The virtual network can be attacked too, but its design enables agents to achieve resilient consensus in the presence of any number of attacks. However, [13], [14] assume that the attacks must be bounded, but it could destabilize the system [14], [15]. In an attacker’s perspective, stealthy and bounded attacks are always desirable as it is difficult to detect them, whereas unbounded attacks could be detected easily. In addition, stealthy attacks on the communication links cannot be mitigated by the resilient consensus strategies designed for attacks on sensors and actuators [16]. A virtual network can be designed to protect a multi-agent system from unbounded attacks as well [17], but that requires the existence of a leader which guides the dynamics of the systems or assuming that the communication channels are protected [11], [12]. In [18], a fully distributed resilient voltage containment framework is designed against unbounded attacks, which are intercepting communication channels and penetrating the local state feedback. However, it is assumed that the virtual network is protected and that there exist a leader (leader-following consensus).

In this paper, we propose a new virtual network, which
enables agents in the physical network to achieve a resilient consensus in the presence of cyber-attacks. The contributions of the paper can be summarized as below:

- The proposed virtual network enables multiple agents to achieve a resilient consensus on general digraph that contains directed spanning tree. Hence, in contrast to the virtual networks proposed in [13]–[15], [19], which are limited to strictly leader-following approach or strongly connected digraphs, this new method allows us to solve a resilient consensus for both leaderless consensus problem and leader-following consensus problem in a unified manner.
- The new virtual network shows resilience in the presence of attacks on to the virtual network.
- The gains selection for the proposed virtual system allows us to improve the performance of overall systems in terms of convergence rate and rejecting oscillations.

The rest of the paper is organized as follows. In Section II, we formulate the resilient consensus problem. A new virtual network is presented in Section III to solve resilient consensus problem. In Section IV, we consider cyber-attacks on both physical as well as virtual network, and show that the new virtual network mitigates the cyber-attacks. We present simulation results in Section V. The last section concludes the paper.

II. PROBLEM FORMULATION

Before presenting the problem description, we first discuss some mathematical notations and preliminaries on graph theory that we will use in subsequent discussion.

A. Notations and Preliminaries on Graph Theory

The $i$th eigenvalue of a square matrix $A$ can be written as $\lambda_i(A) = a_i + ib_i$, where $a_i = \Re\{\lambda_i(A)\}$ is the real-part of $\lambda_i(A)$, $b_i = \Im\{\lambda_i(A)\}$ is the imaginary-part of $\lambda_i(A)$, and $i = \sqrt{-1}$. Superscript “$T$” represents the transpose of a matrix or a vector.

Consider a directed graph (digraph) $G = (\mathcal{V}, \mathcal{E})$ consisting of a finite node set $\mathcal{V} = \{1, 2, \cdots, n\}$, an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Let $Q = [q_{ij}]$ be the adjacency matrix of the digraph $G$, where $q_{ij} = 1$ if node $i$ is receiving information from node $j$, and $q_{ij} = 0$ otherwise. The set $\mathcal{N}_i = \{j \mid q_{ij} = 1, \forall j \in \mathcal{V}\} \subseteq \mathcal{V}$ contains the $i$th agent neighbors. The Laplacian $L = L(G)$ of the digraph $G$ is defined as:

$$L = D - Q,$$

where $D = [d_{ij}] \in \mathbb{R}^{n \times n}$ is a diagonal matrix. The diagonal entries $d_{ii}(G)$ of the matrix $D$ represent the in-degree of the $i$th node; that is, the number of neighbors sending information to $i$th agent. The eigenvalues of $L$ are labeled in the following order: $\Re\{\lambda_1(L)\} = 0 < \Re\{\lambda_2(L)\} \leq \Re\{\lambda_3(L)\} \leq \cdots \leq \Re\{\lambda_n(L)\}$. A directed path from node $v_i$ to node $v_j$ in a digraph, is a sequence of ordered edges of the form $(i, i+k)$, where $k = 1, \cdots, l-1$. If there exists a root node, which has no parent node, such that the node has directed paths to all other nodes in the digraph, then it is called directed tree. A directed tree that connects all the nodes of a digraph is called a directed spanning tree. If a directed spanning tree is a subset of a digraph, we say that the digraph $G$ contains directed spanning tree [20].

B. Problem Formulation

We first present an assumption about the digraph $G$ that represents the network of multiple agents.

Assumption 1: The digraph $G$, representing the network topology, contains a directed spanning tree. Consider a group of $n \geq 2$ agents modeled as:

$$\dot{x}_i = -\sum_{j \in \mathcal{N}_i} (x_i - x_j),$$

where $x_i \in \mathbb{R}$ is the state of $i$th agent. In a compact vector form, we write (2) as:

$$\dot{x} = -Lx,$$

where $x \in \mathbb{R}^n$, $L$ is the Laplacian matrix associated with a digraph $G$ containing a directed spanning tree.

Let $v_1 \Lambda (v_1 \in \mathbb{R}^n)$ be the left-eigenvector of $L$, associated with a simple zero eigenvalue. Under dynamics (3) and Assumption 1, the states of (3) achieve consensus, that is $x_i \to x^*$ for all $i \in \mathcal{V}$, where $x^* = \frac{v_1^T x(0)}{v_1^T 1_n}$, $x(0) \in \mathbb{R}^n$ is the initial state vector and $1_n = [1, \cdots, 1]^T$. In the foregoing discussion, we assumed that the communication links between agents are secure. However, in practice, communication channels are vulnerable to cyber-attacks. Therefore, we consider that malicious information can be added to the communication links between agents as below:

$$\dot{x} = -L(x - d) + u,$$

where $d(t) \in \mathbb{R}^n$ is the unknown exogenous vector added to the network. Note that the communication network is attacked only, and that the attack $d(t)$ is bounded; that is $||d(t)|| \leq d^* < \infty$. In practice, the attacker aims to generate an attack that is stealthy and can destabilize the system [21] as an unbounded attack can be detected easily. Thus, we consider an attack signal $d(t)$ that can destabilize $4$, and that the attack is bounded. Moreover, unlike [5], no restrictions are made on the number of attacks.

The objective of this paper is to design a distributed resilient control law $u$ such that the system:

$$\dot{x} = -L(x - d) + u$$

achieve resilient consensus in the presence of cyber-attacks; that is

$$|x_i(t) - x^*| < \varepsilon,$$

for all $t \geq T, i \in \mathcal{V}$, where $T < \infty$ is a finite value, $\varepsilon$ is sufficiently small positive value, and $x(0) \in \mathbb{R}^n$ is the initial condition of (3).
III. VIRTUAL NETWORK FOR RESILIENT CONSENSUS

In this section, we aim to design a new virtual network that solves resilient consensus problem for digraphs containing directed spanning tree, in the presence of cyber-attacks. To this end, the dynamics of the multi-agent system (physical network) together with the virtual network, for $i$th agent is given below:

$$
x_i = -\sum_{j \in N_i} (x_i - x_j) + \beta_1 \sum_{j \in N_i} (z_i - z_j),
$$

$$
\dot{z}_i = -\beta_2 \sum_{j \in N_i} (x_i - x_j) - \beta_3 \sum_{j \in N_i} (z_i - z_j),
$$

(7)

where $z_i \in \mathbb{R}$ is the $i$th state of the virtual network, $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 > 0$ are gains. The virtual state $z_i \in \mathbb{R}$ for all $i \in \mathcal{V}$, is the auxiliary variable, internal to $i$th agent, which has no physical meaning, and thus also known as virtual state. Therefore, each $z_i$ can take any initial value. Note that the set $N_i$ for agents in the physical and virtual network is the same. The states of the physical network are shared through a communication network that can be attacked, whereas the states of the virtual network are shared using a communication network, which is different from the one used to share the physical state $x_j$, for example, a cloud network, which can be realized through network slicing methods, where a shared physical infrastructure is partitioned in multiple virtual networks [22]. Initially, we assume that the virtual network is secure, however, later in this paper, we also consider the case where the virtual network can be attacked too. Thus, every agent takes decision based on the information available from both physical as well as virtual network.

**Remark 1:** The formulation given in (7) is limited to homogeneous single-order multi-agent systems. The virtual network method can be extended for higher-order homogeneous and heterogeneous multi-agent systems with the assumption that the exact model of the physical system can be replicated at the virtual layer. We conjecture that the connectivity requirement can still be limited to a general digraph containing directed spanning tree, whereas the connectivity requirement in ARC-P based protocols is more than $(F + 1)2^m$, where $F$ denotes the number of possible attacks on a node and the superscript $m \geq 2$ denotes the dimension of an agent’s state vector [23].

Next, we discuss the information flow in the physical and virtual network. For each $i$th agent, the information available via the physical network is $\{x_j, j \in N_i\}$, whereas the information available through the virtual network is $\{\beta_1 z_j, j \in N_i\}$ and $\{\beta_2 x_j + \beta_3 z_j, j \in N_i\}$. Let $x_{ij} = x_i - x_j$ and $z_{ij} = z_i - z_j$, then the physical state of the $i$th agent takes decision based on the information $\{-x_{ij} + \beta_1 z_{ij}, j \in N_i\}$, where $z_{ij}$ is coming through the internal auxiliary state $z_i$, whereas the auxiliary state of the $i$th agent is updated using the information $\{-\beta_2 x_{ij} - \beta_3 z_{ij}, j \in N_i\}$. Note that $i$th physical agent and its corresponding virtual agent are connected internally, thus, the $i$th agent at the physical network has a copy of the virtual state $z_i$, and the $i$th virtual agent has a copy of the state $x_i$. This information flow is depicted in Fig. 1.

In a compact vector form, (7) can be written as:

$$
x = -L x + \beta_1 L z
$$

$$
\dot{z} = -\beta_2 L x - \beta_3 L z,
$$

(8)

where $x \in \mathbb{R}^n$, $z = [z_1, \cdots, z_n]^T \in \mathbb{R}^n$. Next, we investigate the condition on the gains $\beta_1$, $\beta_2$, and $\beta_3$ such that agents at the physical network achieve consensus in the absence of cyber-attacks. This is important to analyze as any resilient consensus protocol must solve the consensus problem in the presence and in the absence of cyber-attacks.

**Proposition 1:** Consider a MAS with a virtual systems as given in (8). Let $\beta_1, \beta_2$, and $\beta_3$ be all positive gains. Let $\lambda_i(L) = \delta_i + i\omega_i$ be the $i$th eigenvalue of the Laplacian matrix associated with the digraph $\mathcal{G}$ and $\lambda(\Omega) = \delta + i\omega$ be the eigenvalues of matrix $\Omega$, where

$$\Omega = \begin{bmatrix} -1 & \beta_1 \\ -\beta_2 & -\beta_3 \end{bmatrix}. $$

Then agents achieve consensus $x_1 = x_2 = \cdots = x_n = x^*$, if and only if the gains $\beta_1$, $\beta_2$, and $\beta_3$ are chosen such that $|\delta_2, \delta^2 > |\omega_j, \omega|^{2}$ for all $2 \leq j \leq n$.

**Proof:** Let $\xi = [x^T, z^T]^T$, (8) takes the following form:

$$\dot{\xi}(t) = M \xi,
$$

(9)

where $M = \Omega \otimes L$. The eigenvalues of $M$ are values in the set $S = \{\lambda_i(\Omega)\lambda_i(L) \mid i = 1, \cdots, n\}$ (Theorem 4.2.12 in [24]). Since the digraph $\mathcal{G}$ associated to the Laplacian matrix $L$ contains directed rooted spanning tree, thus all the eigenvalues of $L$ have positive real part except the simple zero eigenvalue. To achieve consensus, the eigenvalues of $M$ must be in the open left-half plane except the simple zero eigenvalue.
eigenvalue of the Laplacian matrix. It follows that
\[ \Re[\lambda(M)] = \Re[\lambda(\Omega L(\mathcal{L}))] = \Re[(\delta + i\omega)(\delta + i\omega)] = \delta \delta + \delta \omega i, \]
which shows that stability of (8) is guaranteed if and only if
\[ \delta \delta + \delta \omega i < 0, \text{ for } 2 \leq j \leq n \]
Since \( \delta \delta < 0 \), the above condition is equivalent to
\[ |\delta \delta| > |\delta \omega|. \]
This completes the proof.

Remark 2: To find explicit bounds for the gains \( \beta_1, \beta_2 \) and \( \beta_3 \), one can use the results [25], [26] for finding bounds on the eigenvalues of the Laplacian matrix \( \mathcal{L} \).

Remark 3: The design of \( \beta_1, \beta_2 \) and \( \beta_3 \) relies on the eigenvalues of \( \mathcal{L} \), which implies that the network topology has to be known a priori. However, the implementation is distributed because every agent is taking decisions based on relative information.

In contrast to the virtual network presented in [14], [15], one of the advantages of the virtual network proposed in this paper is that it can improve the transient performance (including improving convergence rate) of the interconnected system (8). To demonstrate this, let us compare the virtual system (8) with the one proposed in [15] for resilient leaderless consensus. In addition, let us consider the case where the communication network is undirected. The multi-agent system with the virtual system in [15] can then be written as:
\[ \begin{align*}
\dot{x} &= -\mathcal{L} x + \beta \mathcal{L} z \\
\dot{z} &= -\beta_2 \mathcal{L} x - \beta_3 \mathcal{L} z.
\end{align*} \]
(10)

Here,
\[ \Omega = \begin{bmatrix}
-1 & \beta \\
-\beta & -1
\end{bmatrix}. \]

To mitigate the affects of cyber-attacks, one needs to increase the value of \( \beta \) [15]. However, it can be observed that increasing the value of \( \beta \) for the system given in (10) increases the oscillations. To cope with this problem, we introduce three gains, namely \( \beta_1, \beta_2 \) and \( \beta_3 \) as shown in (8). The values of \( \beta_1, \beta_2 \) and \( \beta_3 \) can be chosen so that the eigenvalues of \( \Omega \) are real and negative and the imaginary part of the eigenvalues of \( M \) are not large enough compared to the negative real part. To illustrate the benefit of the proposed virtual network in reducing oscillations, the following proposition provides a guideline to choose the gains \( \beta_1, \beta_2 \) and \( \beta_3 \) for suppressing the oscillations for the case of undirected network.

Proposition 2: Let \( \mathcal{G} \) be an undirected graph. Let \( \beta_1 > 0, \beta_2 > 0 \) and \( \beta_3 \geq 2\sqrt{\beta_1 \beta_2} + 1 \), then (8) achieve consensus with no oscillations.

Proof: Let \( \mathcal{G} \) be an undirected graph that represents the interaction between agents given in (8). The eigenvalues of the system matrix \( M \) are the elements in the set \( S = \{ \lambda(\Omega) \lambda_j(\mathcal{L}) | i = 1, 2, \ldots, n \} \). Let \( \beta_1 > 0, \beta_2 > 0 \) and \( \beta_3 \geq 2\sqrt{\beta_1 \beta_2} + 1 \). Then all the eigenvalues in the set \( S \) are real and strictly less than zero (except the simple zero eigenvalue of the Laplacian matrix \( \mathcal{L} \)). Thus the result follows.

In addition, the convergence rate of (8) can be improved by increasing gain values, keeping in mind the guidelines given in Proposition 1 and Proposition 2.

IV. RESILIENT CONSENSUS IN THE PRESENCE OF CYBER-ATTACKS

In this section, we consider bounded cyber-attacks on both physical as well as virtual network. First, we consider attacks on the physical network and show that agents achieve an resilient consensus in the presence of attacks.

A. Cyber-attacks on Physical Network

The dynamics of the physical network, together with virtual system in the presence of cyber-attacks on the physical network can be written as:
\[ \begin{align*}
\dot{x} &= -\mathcal{L}(x - d) + \beta_1 \mathcal{L} z \\
\dot{z} &= -\beta_2 \mathcal{L} x - \beta_3 \mathcal{L} z.
\end{align*} \]
(11)

We then have the following theorem that shows that the resilience consensus can be guaranteed in the presence of unknown bounded attacks.

Theorem 1: Consider an interconnected system (11) where \( \mathcal{L} \) is the Laplacian matrix associated with a digraph satisfying Assumption 1, and the attack vector \( d(i) \) is bounded, that is \( |d(i)| \leq d^* \). Then for sufficiently large \( \beta_1 > 0, \beta_2 > 0 \) and \( \beta_3 > 0 \), the resilient consensus is achieved, that is, the states \( x_i \) for all \( i \in \mathcal{V} \) converge to a small neighborhood of \( x^* = v_1^T x(0) / v_1^T 1_n \).

Proof: Let \( e_x = x - 1_n v_1^T x(0) / v_1^T 1_n \) and \( e_z = z - 1_n v_1^T z(0) / v_1^T 1_n \) be error vectors. Taking the time-derivative of \( e_x \) and \( e_z \) yields
\[ \begin{align*}
\dot{e}_x &= -\mathcal{L} e_x + \beta_1 \mathcal{L} e_z + \mathcal{L} d \\
\dot{e}_z &= -\beta_2 \mathcal{L} e_x - \beta_3 \mathcal{L} e_z.
\end{align*} \]
(12a)
(12b)

Since the directed graph \( \mathcal{G} \) has a global reachable node, thus the Laplacian matrix \( \mathcal{L} \) of the graph \( \mathcal{G} \) has a simple zero eigenvalue and all the other eigenvalues have positive real part. To factor out the dynamics associated with the simple zero eigenvalue, we use the following transformation:
\[ e_x \mapsto [\bar{e}_x \bar{e}_z]^T = T e_x, \quad e_z \mapsto [\bar{e}_z \bar{e}_z]^T = T \bar{e}_z, \]
and
\[ d \mapsto [\bar{d} \bar{d}]^T = T d, \]
where \( \bar{e}_x, \bar{e}_z, \bar{d} \in \mathbb{R}^n, \quad \bar{e}_x, \bar{e}_z, \bar{d} \in \mathbb{R}^n, \quad T = [v_1 v_2 v_3 \cdots v_n] \) with \( v_1 \) is a left-eigenvector associated with \( \lambda_1(\mathcal{L}) \). After coordinate transformation, (12a) can be written as:
\[ \begin{align*}
\dot{\bar{e}}_x &= -T \mathcal{L} T^{-1} \bar{e}_x + \beta_1 T \mathcal{L} T^{-1} \bar{e}_z + T \mathcal{T}^{-1} \bar{d}.
\end{align*} \]
(13)
The Jordan form of \(-\mathcal{L}\) can be written as:
\[
-\mathcal{T}_L \mathcal{T}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & J_r \end{bmatrix},
\]
where \(J_r \in \mathbb{R}^{(n-1) \times (n-1)}\) is a Hurwitz matrix. Therefore, (13) can be written as
\[
\begin{align}
\dot{\tilde{e}}_x &= 0, \quad (14a) \\
\dot{\tilde{e}}_z &= J_r \tilde{e}_z - \beta_1 J_r \tilde{e}_z - J_r \tilde{d}. \quad (14b)
\end{align}
\]
Similarly, we have
\[
\begin{bmatrix}
\dot{\tilde{e}}_x \\
\dot{\tilde{e}}_z
\end{bmatrix} = -\beta_2 T \mathcal{L} \mathcal{T}^{-1} \begin{bmatrix} \tilde{e}_x \\
\tilde{e}_z
\end{bmatrix} - \beta_3 T \mathcal{L} \mathcal{T}^{-1} \begin{bmatrix} \tilde{e}_x \\
\tilde{e}_z
\end{bmatrix} \quad (15)
\]
and
\[
\begin{align}
\dot{\tilde{e}}_x &= 0, \quad (16a) \\
\dot{\tilde{e}}_z &= \beta_2 J_r \tilde{e}_z + \beta_3 J_r \tilde{e}_z. \quad (16b)
\end{align}
\]
All the solutions of the system given in (13) and (15) are defined for all \(t \geq 0\) as \(\tilde{d}\) is bounded. The dynamics of \(\tilde{e}_x\) and \(\tilde{e}_z\) can be ignored as they are not appearing in (14a) and in (16b). Now, rewriting the subsystems \(\tilde{e}_x\) and \(\tilde{e}_z\) in the following form:
\[
\dot{\tilde{\xi}}_e = \Xi \tilde{\xi}_e + \Delta(\tilde{d}),
\]
where \(\tilde{\xi}_e = [\tilde{e}_x^T, \tilde{e}_z^T]^T, \Delta(\tilde{d}) = [-(J_r \tilde{d})^T, 0^T]^T, \) and
\[
\Xi = \begin{bmatrix}
J_r & -\beta_1 J_r \\
\beta_2 J_r & \beta_3 J_r
\end{bmatrix}
\]
The solution of (17) can be written as:
\[
\tilde{\xi}_e(t) = \exp(\Xi t) \tilde{\xi}_e(0) + \int_0^t \exp(\Xi(t-\tau)) \Delta(\tilde{d}) d\tau. \quad (19)
\]
Next, we show \(\lim_{t \to \infty} \| \tilde{\xi}_e(t) \| \leq \varepsilon\), where \(\varepsilon > 0\) can be made small by choosing \(\beta_1, \beta_2\) and \(\beta_3\). Thus, we have
\[
\lim_{t \to \infty} \| \tilde{\xi}_e(t) \| = \lim_{t \to \infty} \left\| \int_0^t \exp(\Xi(t-\tau)) \Delta(\tilde{d}) d\tau \right\|.
\]
By using Cauchy-Schwartz inequality, we produce the following inequality from (20):
\[
\lim_{t \to \infty} \| \tilde{\xi}_e(t) \| \leq \lim_{t \to \infty} \left\| \int_0^t \exp(\Xi(t-\tau)) \Delta(\tilde{d}) d\tau \right\| \leq \left\| \Xi^{-1} \Delta(\tilde{d}) \right\|.
\]
The first part on the right-hand side of (21) converges to zero; that is, \(\lim_{t \to \infty} \left\| \int_0^t \exp(\Xi(t-\tau)) \Delta(\tilde{d}) d\tau \right\| = 0\), because \(J_r\) is Hurwitz. Therefore, we have
\[
\lim_{t \to \infty} \| \tilde{\xi}_e(t) \| \leq \left\| \Xi^{-1} \Delta(\tilde{d}) \right\|,
\]
and by Banachiewics inversion formula [27], we have
\[
\begin{align}
M_1 &= J_r - \beta_1 \beta_2 (\beta_3 J_r + \beta_3 \beta_2 J_r)^{-1}, \\
M_2 &= \beta_1 (\beta_3 J_r + \beta_3 \beta_2 J_r)^{-1}, \\
M_3 &= -\beta_2 (\beta_3 J_r + \beta_3 \beta_2 J_r)^{-1}, \\
M_4 &= (\beta_3 J_r + \beta_3 \beta_2 J_r)^{-1}.
\end{align}
\]
After algebraic manipulation, we can write
\[
\begin{align}
M_1 &= \frac{\beta_3}{\beta_2 \beta_3 + \beta_1 \beta_2} J_r^{-1}, \\
M_2 &= \frac{\beta_1}{\beta_2 \beta_3 + \beta_1 \beta_2} J_r^{-1}, \\
M_3 &= \frac{\beta_2}{\beta_2 \beta_3 + \beta_1 \beta_2} J_r^{-1}, \\
M_4 &= \frac{1}{\beta_2 \beta_3 + \beta_1 \beta_2} J_r^{-1}.
\end{align}
\]
Thus by choosing \(\beta_1 \beta_2 + \beta_3\) large enough, one can make \(\tilde{e}_x\) and \(\tilde{e}_z\) to converge to a small neighborhood around zero. Hence, from (13) and (15), the states \(x(t)\) converge to a small neighborhood \(x^{+} 1_n\). This completes the proof.

Remark 4: The error \(\tilde{\xi}_e(t)\) can be bounded by \(\varepsilon > 0\), that is \(\| \tilde{\xi}_e(t) \| < \varepsilon, t > T\) for smaller values of \(\beta_1, \beta_2\) and \(\beta_3\) if \(\| \Delta(\tilde{d}) \|\) is smaller. However, for smaller value of \(\varepsilon\) and larger \(\| \Delta(\tilde{d}) \|\), from (22) high gain values of \(\beta_1, \beta_2\) and \(\beta_3\) must be chosen to reduce \(\lim_{t \to \infty} \| \tilde{\xi}_e(t) \| < \varepsilon\), which can amplify noise.

In the next subsection, we consider that the virtual network can also be attacked.

B. Cyber-attacks on Virtual Network

Next, we consider attack on the virtual network given in (8), where the adversary compromises the information \(\beta_1 z_j\) and \((\beta_3 x_j + \beta_2 z_j)\) being exchanged via the virtual network. Hence, (8) takes the following form:
\[
\begin{align}
\dot{x} &= -\mathcal{L} x + \beta_1 L z + \mathcal{L} d_1 \\
\dot{z} &= -\beta_2 L x - \beta_3 L z + \mathcal{L} d_2,
\end{align}
\]
where \(d_1 \in \mathbb{R}^n\) and \(d_2 \in \mathbb{R}^n\) are bounded attacks on physical and virtual network, respectively. Note that the communication network between physical and virtual network is safe, because, communication between ith agent and its corresponding virtual agent is an internal signal. In the following, we show that agents in the physical network achieve the resilient consensus in the presence of adversarial attacks on both physical and virtual network.

Theorem 2: Consider an interconnected system (25) where \(\mathcal{L}\) is the Laplacian matrix associated with a digraph satisfying Assumption 1, and the attack vectors \(d_1(t), d_2(t)\) are bounded, that is \(\| d_1(t) \| \leq d_1^*\) and \(\| d_2(t) \| \leq d_2^*\), where \(d_1^*\) and \(d_2^*\) are constant values. Then for sufficiently large \(\beta_1 > 0, \beta_2 > 0\) and \(\beta_3 > 0\), the multi-agent system in (25) solve the resilient consensus problem, that is, the state vector \(x(t) \in \mathbb{R}^n\) converges to a small neighborhood of \(x^{+} 1_n\).

Proof: Let \(e_x = x - 1_n v_1^T x(0)/v_1^T 1_n\) and \(e_z = z - 1_n v_1^T z(0)/v_1^T 1_n\). The time-derivative of \(e_x\) and \(e_z\) yields
\[
\begin{align}
\dot{e}_x &= -\mathcal{L} e_x + \beta_1 L e_x + \mathcal{L} d_1 \\
\dot{e}_z &= -\beta_2 L e_x - \beta_3 L e_z + \mathcal{L} d_2.
\end{align}
\]
To factor out the dynamics associated with the simple zero eigenvalue of \(\mathcal{L}\), we again use the coordinate transformation.
matrix $T$, and following the same steps in the proof of Theorem 1, we have the following reduced-order system:

$$\ddot{\xi}_r = \Xi \xi_r + \Delta(d_1, d_2),$$

(27)

where $\Delta(d_1, d_2) = [-(J_1 \dot{x}_1)^T, -(J_2 \dot{x}_2)^T]^T$. Now again following the same steps as given in the proof of Theorem 1, the solution of (27) can be bounded as below:

$$\lim_{t \to \infty} \|\xi_r(t)\| \leq \lim_{t \to \infty} \|\Xi^{-1} \Delta(d_1, d_2)\|.$$  

(28)

Thus by choosing $\beta_1 \beta_2 + \beta_1$ large enough, $\tilde{e}_r, \tilde{e}_z$ converge to a small neighborhood around zero. Thus, from (25), the states $x(t)$ converge to a small neighborhood $x^{T}1_n$. This completes the proof.

Remark 5: The Proof of Theorem 2 follows the same path as the Proof of Theorem 1, however, $\Delta(\ddot{d})$ and $\Delta(d_1, d_2)$ are different. For (28), it is suggested to use $\ddot{\beta}_2$.

V. SIMULATION RESULTS

In this section, we present numerical examples to show that the multi-agent systems together with the virtual system achieve consensus. Then we demonstrate that in the case of cyber-attacks on to the physical network, the design of virtual network enable agents to mitigate the effects of any number of attacks on the physical system. Last, we demonstrate that both physical and virtual networks are attacked. Still with the help of virtual system the agents in the physical network achieve resilient consensus.

First, we demonstrate that (8) achieve consensus, in the absence of cyber-attacks for $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 > 0$, where $\beta_3 > 2\sqrt{\beta_1 \beta_2} + 1$. Let

$$\mathcal{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}. $$

(29)

In Fig. 2, the trajectories of agents are oscillating while achieving consensus, because the value of $\beta_3 < 2\sqrt{\beta_1 \beta_2} + 1$ for $\beta_1 = \beta_2$. Now, for $\beta_3 > 2\sqrt{\beta_1 \beta_2} + 1$, we see in Fig. 3 that the oscillations almost do not exist. The gains can be selected in other ways to reduce the oscillations as we have three degree of freedom (three gains to adjust). Here, we demonstrate one way to select the gains. According to Proposition 1, if $\beta_3 < 2\sqrt{\beta_1 \beta_2} + 1$, then $\Delta \tilde{d}^T < |\omega_0 \omega_0|$, and the state trajectories of (8) diverge as shown in Fig. 4. The instability in Fig. 4 is demonstrated using gains $\beta_1 = \beta_2 = \beta_3 = 2$. Next, we demonstrate how agents achieve resilient consensus in the presence of malicious information in the physical network. One can always design a destabilizing bounded attack, which can destabilize a multi-agent system. For example consider the dynamics of an attacker as given below:

$$\dot{d}(t) = F_d \tilde{d} + B_\omega \tilde{\omega},$$

(30)

where

$$F_d = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & -0.7 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } B_\omega = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -8 & -2 & -6 & -6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

If the multi-agent system given in (4) is subjected to the attack given (30), the trajectories of agents diverge as shown in Fig. 5. To protect the multi-agent network, we design a virtual network as given in (7) that mitigate the cyber-attacks on the multi-agent network. Note that the attack is on the communication network, not the nodes. By choosing the gains large enough $\beta_1 = 20$, $\beta_2 = \beta_1^6$ and $\beta_3 = \beta_1^2$, we
The cooperative resilient consensus method in [14] employs generic dynamic models ($L_{oo}$ and $L_2$) for attacks, whereas the attack model in this paper is limited to $L_{oo}$. This simplification allows us to derive much explicit and simpler results and mostly importantly, to avoid the limit on how many nodes could be attacked. In comparison, the results in [14] requires that the number of nodes under attack is less than half of the total nodes. In general, the interaction between attack and defense involves a dynamic game for which information superiority is crucial [28].

**REFERENCES**


