Low Complexity Resilient Consensus in Networked Multi-Agent Systems with Adversaries

Heath J. LeBlanc
heath.j.leblanc@vanderbilt.edu

Xenofon Koutsoukos
xenofon.koutsoukos@vanderbilt.edu

Institute for Software Integrated Systems
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN, USA

ABSTRACT

Recently, many applications have arisen in distributed control that require consensus protocols. Concurrently, we have seen a proliferation of malicious attacks on large-scale distributed systems. Hence, there is a need for (i) consensus problems that take into consideration the presence of adversaries and specify correct behavior through appropriate conditions on agreement and safety, and (ii) algorithms for distributed control applications that solve such consensus problems resiliently despite breaches in security. This paper addresses these issues by (i) defining the adversarial asymptotic agreement problem, which requires that the uncompromised agents asymptotically align their states while satisfying an invariant condition in the presence of adversaries, and (ii) by designing a low complexity consensus protocol, the Adversarial Robust Consensus Protocol (ARC-P), which combines ideas from distributed computing and cooperative control. Two types of omniscient adversaries are considered: (i) Byzantine agents can convey different state trajectories to different neighbors in the network, and (ii) malicious agents must convey the same information to each neighbor. For each type of adversary, sufficient conditions are provided that ensure ARC-P guarantees the agreement and safety conditions in static and switching network topologies, whenever the number of adversaries in the network is bounded by a constant. The conservativeness of the conditions is examined, and the conditions are compared to results in the literature.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems; H.1.1 [**Models and Principles**]: Systems and Information Theory—*General Systems Theory*

General Terms

Algorithms, Security, Theory

Keywords

Consensus, Multi-agent network, Resilience, Adversary, Byzantine

HSCC'12, April 17–19, 2012, Beijing, China.

1. INTRODUCTION

Due to recent improvements in computation and communication, control system design has made a shift in many applications from centralized to decentralized and distributed approaches. This trend has been fueled by the need for increased flexibility, reliability, and performance in applications such as coordination of vehicle formations [3], flocking [6], and belief propagation in Bayesian networks [15]. For these applications and many others, reaching some form of consensus is fundamental to coordination [11, 14]. However, large-scale distributed systems have many entry points for malicious attacks and intrusions. If a security breach occurs, traditional consensus algorithms will fail to produce desirable results, and therefore lack robustness [5]. Hence, there is a need for resilient consensus algorithms that guarantee correct behavior even after sustaining security breaches.

Of course, there is a long history in distributed computing of studying consensus problems in the presence of faults and adversarial processors [11, 20]. The most potentially harmful form of adversary is the Byzantine processor, which may behave arbitrarily within the limitations set by the model of computation [7]. Therefore, worst case executions must be considered. Typically, the number of processors that may be Byzantine are bounded and fundamental tight bounds have been established on the ratio of Byzantine to normal processors [1, 7], as well as on the connectivity of the graph representing the communication network [1].

From a control theoretic viewpoint, consensus in the presence of adversaries has only been considered recently, and has focused on detection and identification of misbehaving nodes in linear consensus networks [16–19, 24, 25]. While detection is clearly an important problem, these techniques require each node to have information of the network topology beyond its local neighborhood. This requirement of *nonlocal information* renders these techniques inapplicable to general time-varying networks. Further, the detection algorithms are computationally expensive and do not consider safety constraints on the states of the agents. Using these approaches, it is possible that the adversaries may drive the states of the agents outside of a predetermined safe set during the detection phase, which may not be suitable for certain safety critical applications.

In our work, we study a consensus protocol, or algorithm, that is low complexity and uses *only local information* to achieve resilience against a bounded number of adversaries in the network. In order to codify a notion of correct behavior of the uncompromised, or cooperative, agents in the presence of adversaries, we define a consensus problem that specifies formal agreement and safety conditions. The agreement condition requires that all cooperative agents asymptotically align their states. The safety condition requires that the state trajectories of the cooperative agents

remain inside the minimal hypercube formed by the initial states of the cooperative agents. This safety constraint is applicable to cases where the unsafe regions are unknown, but the minimal hypercube containing the initial states is known to be safe. Together, these conditions form the *adversarial asymptotic agreement problem*, which is a continuous-time consensus problem analogous to the Byzantine approximate agreement problem [2, 11].

For this problem, we model the networked system in continuous-time and study both static and dynamic (or switching) network topologies with directed information flow. The agents have continuous dynamics and convey state information to each other over a network that switches between a finite number of discrete topologies. In this paper, we define two types of omniscient adversaries: malicious and Byzantine. Malicious agents share the same information with each neighbor in the network and are analogous to the discrete-time malicious agents studied in [16–19, 24]. Byzantine agents are capable of conveying different information to different neighbors in the network, and are therefore more deceitful than malicious agents.

The proposed consensus protocol is the Adversarial Robust Consensus Protocol (ARC-P), which borrows ideas from computer science and cooperative control. It combines the elimination of extremal values used in Byzantine resilient consensus algorithms in distributed computing [2, 11], with the standard consensus technique in cooperative control of summing the neighboring relative states as input to an integrator agent [14].

We introduced ARC-P in [8], where we studied resilience to malicious agents in complete networks. Here we extend the study of ARC-P to more general network topologies. We present sufficient conditions on the set of possible network topologies that allow us to prove agreement for both fixed and switching topologies using a common Lyapunov function. For safety, we use an invariant set argument similar to the argument made in [8]. We also provide a necessary condition on consensus using ARC-P. Then, we relate the sufficient conditions to known necessary and sufficient conditions set forth in the literature–which have addressed different consensus problems under different models of computation. Although the sufficient conditions are conservative, we provide pathological examples in which the conditions are relaxed minimally and consensus is precluded. Finally, we illustrate the theoretical results through a simulation example.

The rest of the paper is organized as follows. Section 2 covers some preliminaries including the terminology, system model, and problem statement. ARC-P is then described in Section 3. Section 4 studies the convergence properties of ARC-P in a class of directed networks. Section 5 examines more closely the sufficient conditions given in Section 4, and illustrates the results through simulation. Section 6 gives an account of related works, and Section 7 provides conclusions and directions for future work.

2. PRELIMINARIES

2.1 Review of Graph Theory

In this section we review some fundamentals of graph theory pertinent to this paper. As is common when dealing with multi-agent networks, we model the networked multi-agent system with a (finite, simple, labelled) *digraph*, $\mathcal{D} = (\mathcal{V}, \mathcal{E})$ [12]. The *node set* $\mathcal{V} = \{1, \dots, n\}$ abstracts the n dynamic agents as *nodes*, and the *directed edge set* $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ models the information flow between the agents, which is realized either through communication or sensing. For each ordered pair $(i, j) \in \mathcal{E}$, state information flows from node i to node j. We also consider the *underlying graph* $\mathcal{G}(\mathcal{D})$, which is defined by replacing directed edges of $\mathcal{D}$ by undirected ones, resulting in the edge set $\mathcal{E}_{\mathcal{G}}$.

For local information flow, we consider the set of *in-neighbors* of node j, defined by $\mathcal{N}_j^{\text{in}} = \{i \in \mathcal{V} | (i, j) \in \mathcal{E}\}$, and the set of *inclusive in-neighbors* of node j, defined by $\mathcal{J}_j^{\text{in}} = \mathcal{N}_j^{\text{in}} \cup \{j\}$. The *in-degree* of j is denoted $d_j^{\text{in}} \triangleq |\mathcal{N}_j^{\text{in}}|$, and the *minimum in-degree* of $\mathcal{D}$ is denoted $\delta^{\text{in}}(\mathcal{D})$. Likewise, the *maximum in-degree* of $\mathcal{D}$ is denoted $\Delta^{\text{in}}(\mathcal{D})$. There are, of course, analogous definitions for *out-neighbors*, e.g., the *out-degree* of j is $d_j^{\text{out}} \triangleq |\mathcal{N}_j^{\text{out}}|$ and the *minimum out-degree* of $\mathcal{D}$ is $\delta^{\text{out}}(\mathcal{D})$.

In order to describe information flow across the network, we consider the following definitions. A *path* is a sequence of distinct vertices $i_0, i_1, \dots, i_k$ such that $(i_j, i_{j+1}) \in \mathcal{E}$, $j = 0, 1, \dots, k-1$. We use the notion of path to define different forms of connectedness. We say that $\mathcal{D}$ is *strongly connected* if for every $i, j \in \mathcal{V}$, there exists a path starting at i and ending at j. If the underlying graph is connected, then $\mathcal{D}$ is *weakly connected*. Alternatively, if the underlying graph is disconnected, then $\mathcal{D}$ is *disconnected*.

To measure the robustness and redundancy of information flow, we define a *vertex cut* as a set of vertices $\mathcal{K}$ such that the removal of $\mathcal{K}$ results in either a disconnected digraph or the trivial digraph consisting of a single node. The *connectivity* $\kappa(\mathcal{D})$ is the size of a minimal vertex cut. A digraph is said to be *k-connected* if $\kappa(\mathcal{D}) \geq k$. A simple consequence of defining connectivity in this manner is $\kappa(\mathcal{D}) = \kappa(\mathcal{G}(\mathcal{D}))$ [4].[1]

2.2 System Model

This section details the system model, with the assumptions on the cooperative agents and adversaries. To allow for time-varying, or switching, network topologies, we consider the finite set of all digraphs on n vertices, $\Gamma_n = \{\mathcal{D}_1, \dots, \mathcal{D}_d\}$. Each digraph $\mathcal{D}_k \in \Gamma_n$ has the same vertex set $\mathcal{V}$, whereas the directed edge sets $\mathcal{E}_1, \dots, \mathcal{E}_d$ are all distinct. Without loss of generality, $\mathcal{V}$ is partitioned into a set of *p cooperative agents* $\mathcal{V}_c = \{1, \dots, p\}$ and a set of *q adversaries* $\mathcal{V}_a = \{p+1, \dots, n\}$, with $q = n - p$. A *switching signal* $\sigma \colon \mathbb{R}_{\geq 0} \to \{1, \dots, d\}$ determines which digraph $\mathcal{D}_{\sigma(t)} \in \Gamma_n$ describes the network at time $t \in \mathbb{R}_{\geq 0}$. We assume a finite number of switches on any finite time interval.

For simplicity of notation, we assume each agent's state is scalar. Collectively, $x_c(t) = [x_1(t), \dots, x_p(t)]^{\mathsf{T}} \in \mathbb{R}^p$ is the state of the cooperative agents. Likewise, the collective state of the adversaries conveyed to agent $j \in \mathcal{V}_c$ is $x_{a_j}(t) = [x_{p+1,j}(t), \dots, x_{n,j}(t)]^{\mathsf{T}} \in \mathbb{R}^q$. If $k \notin \mathcal{J}_j^{\text{in}}(t)$, then adversary k does not directly influence agent j at time t, in which case agent j does not receive $x_{k,j}(t)$. While this notation may seem overly cumbersome, it simplifies dealing with Byzantine agents. One may take the viewpoint that $x_{k,j}(t)$ is the trajectory Byzantine agent k would like to convey to agent j, but the topological constraints on the network prevent it from doing so. With this justification, we denote $x \colon \mathbb{R} \times \mathcal{V}_c \to \mathbb{R}^n$ by $x(t, j) = [x_c^{\mathsf{T}}(t), x_{a_j}^{\mathsf{T}}(t)]^{\mathsf{T}} \in \mathbb{R}^n$. Whenever the context is understood, we will drop the arguments and write x_c, x_{a_j}, and x. Finally, we denote by x_a the set of all x_{a_j}, $j \in \mathcal{V}_c$.

2.2.1 Cooperative Agents

Each cooperative agent $i \in \mathcal{V}_c$ has dynamics given by $\dot{x}_i = u_i$, where $u_i = f_{i,\sigma(t)}(x_c, x_{a_i})$ is a control input. The states of the neighboring adversaries, within x_{a_i}, are analyzed as uncertain inputs; however, because there is no prior knowledge about which agents are adversaries, the control input must treat the state infor-

[1] In [4], this form of connectivity is defined as $\kappa_1(\mathcal{D})$ and other forms of connectivity in digraphs are studied (most notably strong connectivity). For our purposes, the definition given here suffices.

mation from neighboring agents in the same manner. The dynamics of the system of cooperative agents are then defined for $t \geq 0$ by

$$\dot{x}_c = f_{c,\sigma(t)}(x_c, x_a), \quad x_c(0) \in \mathbb{R}^p, \mathcal{D}_{\sigma(t)} \in \Gamma_n, \qquad (1)$$

where $f_{c,\sigma(t)}(x_c, x_a) = [f_{1,\sigma(t)}(x_c, x_{a_1}), \ldots, f_{p,\sigma(t)}(x_c, x_{a_p})]^\mathsf{T}$. The dynamics of (1) define a switched system without impulse effects, so the trajectory of any solution is absolutely continuous [10].

2.2.2 Adversaries

The q adversaries are assumed to be designed for the purpose of disrupting the objective of the cooperative agents. It is assumed that the number of adversaries in the network is bounded above by a constant $F \in \mathbb{Z}_{\geq 0}$, so that $q \leq F$. We consider two different adversary models, defined as follows.

DEFINITION 1. *The q adversaries have continuous state trajectories; i.e., x_{a_j} is continuous for $j \in \mathcal{V}_c$. An adversary is*

(i) ***Byzantine*** *if it can convey different state trajectories to different neighbors; i.e., we may have $x_{a_i} \neq x_{a_j}$, or $x_{k,i} \neq x_{k,j}$, whenever $k \in \mathcal{V}_a \cap \mathcal{J}_i^{in}(t) \cap \mathcal{J}_j^{in}(t)$ for $i, j \in \mathcal{V}_c$;*

(ii) ***Malicious*** *if it must convey the same state trajectory to each neighbor; i.e., $x_{a_i} \equiv x_{a_j}$ for all $i, j \in \mathcal{V}_c$.*

Both classes of adversaries are assumed to be omniscient and behave in a worst case manner. Hence, the adversaries are able to carefully select their continuous state trajectories to cause maximal disruption to the consensus objective of the cooperative agents.

2.3 Problem Statement

The *adversarial asymptotic agreement problem* is defined by two conditions, agreement and safety, along with the type of adversary considered. The *agreement condition* requires that the state of the cooperative agents, x_c, converges to the agreement space, $\mathcal{A} = \text{span}\{1_p\} \subset \mathbb{R}^p$, despite the influence of the adversaries. That is, given $q \leq F$ adversaries and $x_c(0) \in \mathbb{R}^p$, then

$$x_c(t) \to \mathcal{A} \text{ as } t \to \infty. \qquad (2)$$

The *safety condition* requires that the state trajectory of each cooperative agent is contained in the interval formed by the initial states of cooperative agents and that the limit exists, despite the influence of the adversaries. That is, if we define the interval

$$\mathcal{I}_0 = [\min_{i \in \mathcal{V}_c} x_i(0), \max_{j \in \mathcal{V}_c} x_j(0)],$$

then the safety condition requires that given $q \leq F$ adversaries,

$$x_j(t) \in \mathcal{I}_0, \forall t \in \mathbb{R}_{\geq 0} \text{ and } \lim_{t \to \infty} x_j(t) \in \mathcal{I}_0 \text{ exists}, \forall j \in \mathcal{V}_c. \quad (3)$$

Equivalently, the safety condition can be stated in terms of x_c. Let $\mathcal{H}_0 = \mathcal{I}_0^p \subset \mathbb{R}^p$ denote the hypercube formed by the Cartesian product of p copies of $\mathcal{I}_0$. Then the safety condition requires $x_c(t) \in \mathcal{H}_0$ for all $t \geq 0$ and $\lim_{t \to \infty} x_c(t) \in \mathcal{H}_0$, despite the influence of the adversaries. It is important to explicitly require that the limit exists because convergence to a single point is desired.

The safety condition in (3) is similar to the validity condition defined in [8], which in turn was motivated by the validity condition of the Byzantine approximate agreement problem [2, 11]. The definition ensures that the value chosen by each normal node lies within the range of good values. This is important in applications where the values are measurements and only measurements within the range obtained by the normal nodes are considered valid. The safety condition entails this notion along with an invariant condition, which is important for safety critical applications.

3. CONSENSUS PROTOCOL

Here, we describe the Adversarial Robust Consensus Protocol (ARC-P) with respect to parameter $F \in \mathbb{Z}_{\geq 0}$. The main idea of the protocol is for each cooperative agent $i \in \mathcal{V}_c$ to sort the relative states of its inclusive in-neighbors and then remove the F largest and F smallest ones. This results in $m_i(t) = d_i^{\text{in}}(t) + 1 - 2F$ relative states if $d_i^{\text{in}}(t) \geq 2F$, which are summed to determine the first order dynamics of the agent. To make the protocol well-defined for all network topologies, i.e., whenever $d_i^{\text{in}}(t) < 2F$, in this case the agent removes all neighboring values from consideration and the input is zero. This approach adheres to the philosophy that whenever there is insufficient information to act in a way that is resilient to adversarial influence, it is best to do nothing.

In order to formally express ARC-P with parameter F, let $\xi_i(t)$ denote the vector of sorted values of the states in the inclusive in-neighborhood of node i at time $t \in \mathbb{R}_{\geq 0}$. The elements of ξ_i are denoted by $\xi_i^1, \ldots, \xi_i^{d_i^{\text{in}}(t)+1}$ and satisfy

$$\xi_i^1 \leq \xi_i^2 \leq \cdots \leq \xi_i^{d_i^{\text{in}}(t)+1}. \qquad (4)$$

Then, cooperative agent $i \in \mathcal{V}_c$ calculates $u_i = f_{i,\sigma(t)}(x_c, x_{a_i})$ at time $t \in \mathbb{R}_{\geq 0}$ by

$$f_{i,\sigma(t)}(x_c, x_{a_i}) = \begin{cases} \sum_{l=F+1}^{d_i^{\text{in}}(t)+1-F} \left(\xi_i^l(t) - x_i(t)\right) & d_i^{\text{in}}(t) \geq 2F; \\ 0 & d_i^{\text{in}}(t) < 2F. \end{cases} \qquad (5)$$

If each cooperative agent uses ARC-P, then existence and uniqueness of solutions to (1) is guaranteed $\forall t \geq 0$ since $\sigma(t)$ is piecewise constant, x_{a_i} is continuous and effectively restricted to a compact set with respect to (5) (see the discussion after Lemma 2), and $f_{i,\sigma(t)}(x_c, x_{a_i})$ is globally Lipschitz in x_c and x_{a_i} $\forall i \in \mathcal{V}_c$ [8].

Figure 1 illustrates the computation that occurs at time t for cooperative agent i whenever $d_i^{\text{in}}(t) \geq 2F$. In the figure, the state, $x_i(t)$, of the agent, whose dynamics are $\dot{x}_i(t) = u_i(t)$, is subtracted from each of the other states in its inclusive neighborhood, with each of the in-neighbors denoted x_i^j, $j = 1, 2, \ldots, d_i^{\text{in}}(t)$. The resulting relative state values are sorted and then reduced by eliminating the largest and smallest F elements. Finally, the remaining elements are summed to produce the control input $u_i(t)$ to the integrator agent. The only difference if $d_i^{\text{in}}(t) < 2F$ is that the output of the Reduce block is 0.

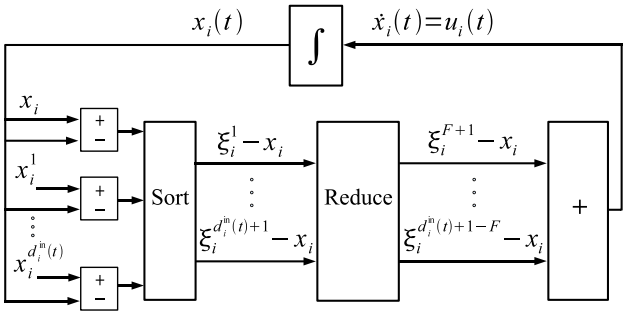


Figure 1: Synchronous data flow model of ARC-P for agent i.

From a complexity standpoint, ARC-P consists of low complexity operations in both time and space, including sort, reduce, and sum methods (see Figure 1). The worst performing subroutine of ARC-P is the sort method. But, if quicksort is used, it is worst-case quadratic in time and linear in space, with respect to the size of the inclusive in-neighborhood. Therefore, ARC-P is also worst-case quadratic in time and linear in space, and hence low complexity.

4. ANALYSIS

This section details the analysis of ARC-P with parameter F. We begin by introducing a function that characterizes the maximum disagreement amongst the cooperative agents' states. Define $\Psi\colon \mathbb{R}^p \to \mathbb{R}$ by

$$\Psi(x_c) = \max_{k\in\mathcal{V}_c}\{x_k\} - \min_{k\in\mathcal{V}_c}\{x_k\}. \tag{6}$$

The function Ψ has several attractive properties: (i) it is nonnegative with $\Psi(x_c) = 0$ for all $x_c \in \mathcal{A}$ and $\Psi(x_c) > 0\ \forall x_c \notin \mathcal{A}$, (ii) it is Lipschitz, (iii) it is increasing away from $\mathcal{A}$ in the sense that $\Psi(y_1) > \Psi(y_2)\ \forall y_1, y_2 \in \mathbb{R}^p$ satisfying $\text{dist}(y_1, \mathcal{A}) > \text{dist}(y_2, \mathcal{A})$, and (iv) it is radially unbounded away from $\mathcal{A}$ in the sense that $\Psi(y) \to \infty$ as $\text{dist}(y, \mathcal{A}) \to \infty$. These properties make Ψ an excellent Lyapunov candidate for proving global convergence to $\mathcal{A}$. Ψ has been used to prove convergence of asynchronous consensus algorithms whenever all nodes are cooperative [26].

But, one issue with Ψ is that it is not everywhere differentiable. Therefore, to study the monotonicity of $\psi(t) = \Psi(x_c(t))$, we consider the *upper-right Dini derivative* $D^+\psi(t)$ of ψ at t, defined by

$$D^+\psi(t) = \limsup_{h\to 0^+} \frac{\psi(t+h) - \psi(t)}{h},$$

and the *upper-directional derivative* of Ψ with respect to (1):

$$D^+\Psi(x_c, x_a) = \limsup_{h\to 0^+} \frac{\Psi(x_c + h f_{c,\sigma(t)}(x_c, x_a)) - \Psi(x_c)}{h}.$$

The motivation for considering the upper directional derivative of Ψ is that $D^+\psi(t) = D^+\Psi(x_c(t), x_a(t))$ for almost all t along solutions of (1) since Ψ is locally Lipschitz [21]. In this case,

$$D^+\Psi(x_c, x_a) = \limsup_{h\to 0^+} \frac{N_1(h)}{h} + \limsup_{h\to 0^+} \frac{N_2(h)}{h}, \tag{7}$$

with
$$N_1(h) = \max_{i\in\mathcal{V}_c}\{x_i + h f_{i,\sigma(t)}(x_c, x_{a_i})\} - \max_{i\in\mathcal{V}_c}\{x_i\},$$
$$N_2(h) = \min_{i\in\mathcal{V}_c}\{x_i\} - \min_{i\in\mathcal{V}_c}\{x_i + h f_{i,\sigma(t)}(x_c, x_{a_i})\}.$$

4.1 Preliminary Results

At this point, we derive some preliminary results that hold for all network topologies. We begin with a fundamental result for ARC-P that bounds $\dot{x}_c$ to a time-dependent compact convex set, which includes the origin.

LEMMA 1. *Consider the cooperative agent $i \in \mathcal{V}_c$ executing ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most $F < n$ malicious or Byzantine agents. Then, for $t \in \mathbb{R}_{\geq 0}$ and $\mathcal{D}_{\sigma(t)} \in \Gamma_n$*

$$m_i(t)(x_{i,lmin} - x_i) \leq f_{i,\sigma(t)}(x_c, x_{a_i}) \leq m_i(t)(x_{i,lmax} - x_i),$$

where $x_{i,lmin}(t) = \min\{x_j | j \in \mathcal{J}_i^{in}(t) \cap \mathcal{V}_c\}$ and $x_{i,lmax}(t) = \max\{x_j | j \in \mathcal{J}_i^{in}(t) \cap \mathcal{V}_c\}$ are defined for $t \in \mathbb{R}_{\geq 0}$, and

$$m_i(t) = \begin{cases} d_i^{in}(t) + 1 - 2F & \text{if } d_i^{in}(t) \geq 2F; \\ 1 & \text{otherwise.} \end{cases}$$

PROOF. If $d_i^{\text{in}}(t) < 2F$, $f_{i,\sigma(t)}(x_c, x_{a_i}) = 0$ and the result follows. Therefore, assume $d_i^{\text{in}}(t) \geq 2F$. Since there are at most F adversaries, we know $x_{i,\text{lmin}} \leq \xi_i^{F+1}$ and $\xi_i^{d_i^{\text{in}}(t)+1-F} \leq x_{i,\text{lmax}}$. Hence, (4) implies

$$m_i\left(x_{i,\text{lmin}} - x_i\right) \leq \sum_{l=F+1}^{d_i^{\text{in}}(t)+1-F} (\xi_i^l - x_i) \leq m_i\left(x_{i,\text{lmax}} - x_i\right).$$

□

While Lemma 1 restricts the behavior of $\dot{x}_c$ almost everywhere, the next result restricts the feasible trajectories of $x_c(t)$. It shows that the minimal hypercube $\mathcal{H}_0$ formed by the initial values of the cooperative agents is *robustly positively invariant*.

DEFINITION 2. *The set $\mathcal{S} \subset \mathbb{R}^p$ is **robustly positively invariant** for the system given by (1) if for all $x_c(0) \in \mathcal{S}$, $x_{a_i}(t) \in \mathbb{R}^q$ $i \in \mathcal{V}_c$, and $t \geq 0$, the solution satisfies $x_c(t) \in \mathcal{S}$.*

LEMMA 2. *Suppose the cooperative agents in $\mathcal{V}_c$ execute ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most $F < n$ malicious or Byzantine agents. Then, for every $\mathcal{D}_{\sigma(t)} \in \Gamma_n$ the hypercube*

$$\mathcal{H}_0 = \{y \in \mathbb{R}^p | x_{0,min} \leq y_i \leq x_{0,max},\ i = 1, 2, \ldots, p\},$$

where $x_{0,min} = \min_{i\in\mathcal{V}_c}\{x_i(0)\}$ and $x_{0,max} = \max_{i\in\mathcal{V}_c}\{x_i(0)\}$, is robustly positively invariant for the system (1).

PROOF. Since $\mathcal{H}_0$ is compact and any solution of (1) using (5) is continuous with $x_c(0) \in \mathcal{H}_0$, we must show that $f_{c,\sigma(t)}(x_c, x_a)$ is not directed outside of $\mathcal{H}_0$, whenever $x_c(t) \in \partial\mathcal{H}_0$, for all $\mathcal{D}_{\sigma(t)} \in \Gamma_n$ and $x_{a_i}(t) \in \mathbb{R}^q$ for $i \in \mathcal{V}_c$. The boundary $\partial\mathcal{H}_0$ is given by

$$\partial\mathcal{H}_0 = \{y \in \mathcal{H}_0 | \exists i \in \{1, 2, \ldots, p\} \text{ s.t. } y_i \in \{x_{0,\min}, x_{0,\max}\}\}.$$

Fix $x_c(t) \in \partial\mathcal{H}_0$. Let e_j denote the j-th canonical basis vector and denote $\mathcal{I}_{x_c,\min}, \mathcal{I}_{x_c,\max} \subseteq \{1, 2, \ldots, p\}$ as the sets defined by

$$j \in \mathcal{I}_{x_c,\min} \Leftrightarrow x_j = x_{0,\min} \text{ and } k \in \mathcal{I}_{x_c,\max} \Leftrightarrow x_k = x_{0,\max}.$$

Then, from the geometry of the hypercube, we require

$$e_j^{\mathsf{T}} f_{c,\sigma(t)}(x_c, x_a) \geq 0 \quad \forall j \in \mathcal{I}_{x_c,\min},$$
$$e_k^{\mathsf{T}} f_{c,\sigma(t)}(x_c, x_a) \leq 0 \quad \forall k \in \mathcal{I}_{x_c,\max}.$$

These conditions are true for all $\mathcal{D}_{\sigma(t)} \in \Gamma_n$ and $x_{a_i}(t) \in \mathbb{R}^q$ with $i \in \mathcal{V}_c$ by Lemma 1, in which the lower bound is used for $j \in \mathcal{I}_{x_c,\min}$ since $x_j = x_{j,\text{lmin}} = x_{0,\min}$, and the upper bound is used for $k \in \mathcal{I}_{y,\max}$ since $x_k = x_{k,\text{lmax}} = x_{0,\max}$. □

The argument made in Lemma 1 implies that any time an adversary is outside of $\mathcal{I}_t = [\min_{i\in\mathcal{V}_c}\{x_i(t)\}, \max_{i\in\mathcal{V}_c}\{x_i(t)\}]$, its influence is guaranteed to be removed by its cooperative neighbors, and therefore has the same effect as if it were on the boundary of $\mathcal{I}_t$. Using Lemma 2 we conclude $\mathcal{I}_t \subseteq \mathcal{I}_0$, $\forall t \geq 0$. Hence, each adversary is effectively restricted to the compact set $\mathcal{I}_0$, with respect to (1). This fact enables us to allow adversary states in $\mathbb{R}^q$ rather than explicitly restricting them to a compact set, while still ensuring existence and uniqueness of solutions. Next, we derive an explicit equation for $D^+\Psi(x_c, x_a)$, valid at a fixed time $t \geq 0$.

LEMMA 3. *Fix $t \geq 0$ and $x_c(t) \in \mathbb{R}^p$. Suppose each cooperative agent in $\mathcal{V}_c$ executes ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most $F < n$ Byzantine or malicious agents. Let $\mathcal{D}_{\sigma(t)} \in \Gamma_n$ and define $\mathcal{S}_{min}(t), \mathcal{S}_{max}(t)\colon \mathbb{R} \to \{1, \ldots, p\}$ by*

$$j \in \mathcal{S}_{min}(t) \Leftrightarrow x_j(t) = \min_{i\in\mathcal{V}_c}\{x_i(t)\},$$
$$k \in \mathcal{S}_{max}(t) \Leftrightarrow x_k(t) = \max_{i\in\mathcal{V}_c}\{x_i(t)\}.$$

Fix $j_t \in \mathcal{S}_{min}(t)$ such that

$$f_{j_t,\sigma(t)}(x_c, x_{a_{j_t}}) \leq f_{j,\sigma(t)}(x_c, x_{a_j}), \quad \forall j \in \mathcal{S}_{min}(t).$$

Likewise, fix $k_t \in \mathcal{S}_{max}(t)$ such that

$$f_{k_t,\sigma(t)}(x_c, x_{a_{k_t}}) \geq f_{k,\sigma(t)}(x_c, x_{a_k}), \quad \forall k \in \mathcal{S}_{max}(t).$$

Then, at time t, we have

$$D^+\Psi(x_c(t), x_a(t)) = f_{k_t,\sigma(t)}(x_c, x_{a_{k_t}}) - f_{j_t,\sigma(t)}(x_c, x_{a_{j_t}}), \tag{8}$$

and $D^+\Psi(x_c, x_a) \leq 0$ for all $t \geq 0$.

PROOF. Let $m_{t,\max} = \max_{i\in\mathcal{V}_c}\{m_i(t)\}$, where $m_i(t)$ is defined in Lemma 1. Lemma 1 implies that

$$-m_{t,\max}\Psi(x_c) \le f_{i,\sigma(t)}(x_c, x_{a_i}) \le m_{t,\max}\Psi(x_c)$$

holds $\forall i \in \mathcal{V}_c$. If $x_c(t) \notin \mathcal{A}$, then there exists $\epsilon_{\min} > 0$ such that $x_i - x_j \ge \epsilon_{\min} > 0$ for all $j \in \mathcal{S}_{\min}(t)$ and $i \in \mathcal{V}_c \setminus \mathcal{S}_{\min}(t)$. Similarly, there exists $\epsilon_{\max} > 0$ such that $x_k - x_i \ge \epsilon_{\max}$ for all $k \in \mathcal{S}_{\max}(t)$ and $i \in \mathcal{V}_c \setminus \mathcal{S}_{\max}(t)$. Then, by letting $\epsilon = \min\{\epsilon_{\min}, \epsilon_{\max}\}$ and taking $h \le \epsilon/(2m_{t,\max}\Psi(x_c(t)))$, we may write

$$\begin{aligned} x_i + hf_{i,\sigma(t)}(x_c, x_{a_i}) &\ge x_i - hm_{t,\max}\Psi(x_c(t)) \\ &\ge x_i - \epsilon/2 \\ &\ge x_j + \epsilon/2 \\ &\ge x_j + hm_{t,\max}\Psi(x_c(t)) \\ &\ge x_j + hf_{j,\sigma(t)}(x_c, x_{a_j}) \\ &\ge x_{j_t} + hf_{j_t,\sigma(t)}(x_c, x_{a_{j_t}}) \end{aligned}$$

for all $i \in \mathcal{V}_c \setminus \mathcal{S}_{\min}(t)$, $j \in \mathcal{S}_{\min}(t)$. Therefore, at time t

$$\min_{i\in\mathcal{V}_c}\{x_i + hf_{i,\sigma(t)}(x_c, x_{a_i})\} = x_{j_t} + hf_{j_t,\sigma(t)}(x_c, x_{a_{j_t}}).$$

Following a similar argument, we deduce

$$\max_{i\in\mathcal{V}_c}\{x_i + hf_{i,\sigma(t)}(x_c, x_{a_i})\} = x_{k_t} + hf_{k_t,\sigma(t)}(x_c, x_{a_{k_t}}).$$

Combining this with (7), gives (8). On the other hand, if $x_c(t) \in \mathcal{A}$ then both $\Psi(x_c)$ and $D^+\Psi(x_c, x_a)$ are zero. Finally, applying Lemma 1 with $x_{j_t,\text{lmin}} = x_{j_t}$ and $x_{k_t,\text{lmax}} = x_{k_t}$ shows that $D^+\Psi(x_c, x_a) \le 0$. □

Notice in (8) that the agents acting as k_t and j_t may change with time. It will be important to show in the convergence argument that bounds on $D^+\Psi(x_c, x_a)$ hold for all $t \in \mathbb{R}_{\ge 0}$, regardless of which cooperative agents fill the roles of k_t and j_t. Next, we show that $\Psi(x_c)$ is bounded by scaled versions of $\text{dist}(x_c, \mathcal{A}) = \inf_{y\in\mathcal{A}} ||x_c - y||_2$. For this argument, we use the following properties of the min and max functions. If $\alpha \in \mathbb{R}$, then

$$\begin{aligned} \min_{i\in\mathcal{V}_c}\{x_i + \alpha\} &= \min_{i\in\mathcal{V}_c}\{x_i\} + \alpha \\ \max_{i\in\mathcal{V}_c}\{x_i + \alpha\} &= \max_{i\in\mathcal{V}_c}\{x_i\} + \alpha. \end{aligned} \tag{9}$$

LEMMA 4. *Given $x_c \in \mathbb{R}^p$, $\Psi(x_c)$ is bounded by*

$$\frac{1}{\sqrt{p}}\textit{dist}(x_c, \mathcal{A}) \le \Psi(x_c) \le 2\textit{dist}(x_c, \mathcal{A}), \tag{10}$$

PROOF. Consider the decomposition of x_c: $x_c = v_{\mathcal{A}} + v_{\mathcal{A}^\perp}$, in which $v_{\mathcal{A}} \in \mathcal{A}$ and $v_{\mathcal{A}^\perp} \in \mathcal{A}^\perp$. Given this decomposition, we conclude $\|v_{\mathcal{A}^\perp}\|_2 = \text{dist}(x_c, \mathcal{A})$ and $\exists\gamma \in \mathbb{R}$ such that $v_{\mathcal{A}} = \gamma 1_p$. Because of this, we can use (9) to write

$$\Psi(x_c) = \max_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\} - \min_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\}, \tag{11}$$

in which $(v_{\mathcal{A}^\perp})_i$ is the i-th element of $v_{\mathcal{A}^\perp}$. From this, we obtain the upper bound

$$\Psi(x_c) \le \max_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\} + |\min_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\}| \le 2\|v_{\mathcal{A}^\perp}\|_2.$$

On the other hand, since $v_{\mathcal{A}^\perp} \in \mathcal{A}^\perp$, $\sum_{i=1}^p (v_{\mathcal{A}^\perp})_i = 0$, so that $\max_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\} \ge 0$ and $\min_{i\in\mathcal{V}_c}\{(v_{\mathcal{A}^\perp})_i\} \le 0$. From this and (11) we conclude $\Psi(x_c) \ge |(v_{\mathcal{A}^\perp})_j|$ for all $j \in \mathcal{V}_c$. Thus, we obtain the lower bound

$$\frac{1}{\sqrt{p}}\|v_{\mathcal{A}^\perp}\|_2 \le \frac{1}{\sqrt{p}}\sqrt{p\Psi^2(x_c)} = \Psi(x_c).$$

□

In the sequel, we first consider fixed network topology, and prove exponential convergence of $x_c(t)$ to $\mathcal{A}$ for a subset of network topologies by using properties of $\Psi(x_c)$ and $D^+\Psi(x_c, x_a)$. We then combine the agreement result with an invariant set argument to prove safety. Afterwards, we prove a necessary condition, and then generalize the results to the case of switching topology by using Ψ as a common Lyapunov function.

4.2 Fixed Topology

In this section, we assume that $\sigma(t) \equiv s$ and $\mathcal{D}_s$ belongs to $\Gamma_{M,F} \subset \Gamma_n$ or $\Gamma_{B,F} \subset \Gamma_n$ whenever the adversaries are, respectively, malicious or Byzantine. When dealing with malicious agents, we consider the following class of digraphs with restricted in-degrees or out-degrees, defined by

$$\Gamma_{M,F} = \{\mathcal{D}_k \in \Gamma_n | \text{ at least one of } M1_F \text{ and } M2_F \text{ holds}\}, \tag{12}$$

where

$M1_F$: $\delta^{\text{in}}(\mathcal{D}_k) \ge \lfloor n/2 \rfloor + F$;

$M2_F$: $\exists\mathcal{S} \subseteq \mathcal{V}(\mathcal{D}_k), |\mathcal{S}| \ge 2F+1$, such that $d_i^{\text{out}} = n-1, \forall i \in \mathcal{S}$.

When dealing with Byzantine agents, we require stronger assumptions on the in-degrees and out-degrees. In this case, we define

$$\Gamma_{B,F} = \{\mathcal{D}_k \in \Gamma_n | \text{ at least one of } B1_F \text{ and } B2_F \text{ holds}\}, \tag{13}$$

where

$$B1_F\text{: } \delta^{\text{in}}(\mathcal{D}_k) \ge \begin{cases} n/2 + \lfloor 3F/2 \rfloor & \text{if } n \text{ is even and } F \text{ is odd;} \\ \lfloor n/2 \rfloor + \lceil 3F/2 \rceil & \text{otherwise.} \end{cases}$$

$B2_F$: $\exists\mathcal{S} \subseteq \mathcal{V}(\mathcal{D}_k), |\mathcal{S}| \ge 3F+1$, such that $d_i^{\text{out}} = n-1, \forall i \in \mathcal{S}$.

It follows from these definitions that $\Gamma_{B,F} \subseteq \Gamma_{M,F}$. Additionally, the conditions in (12) and (13) implicitly bound the maximum number of adversaries F by a function of the total number of agents n. Specifically, property $M1_F$ implies

$$n - 1 \ge \delta^{\text{in}}(\mathcal{D}_s) \ge \lfloor n/2 \rfloor + F \implies F \le \lceil n/2 \rceil - 1.$$

Similarly, property $M2_F$ implies

$$n \ge |\mathcal{S}| \ge 2F + 1 \implies 2F \le n - 1.$$

In either case, $n > 2F$. Analogously, the properties $B1_F$ and $B2_F$ imply $n > 3F$. Therefore, it follows that $F \le \lceil n/2 \rceil - 1$ (or $F \le \lceil n/3 \rceil - 1$) whenever $\mathcal{D}_s \in \Gamma_{M,F}$ (or $\mathcal{D}_s \in \Gamma_{B,F}$). A consequence of this is that $\delta^{\text{in}}(\mathcal{D}_s) \ge 2F$ for all $\mathcal{D}_s \in \Gamma_{M,F}$ (or $\mathcal{D}_s \in \Gamma_{B,F}$). Hence, in this case, (8) may be rewritten using (5) as

$$D^+\Psi(x_c, x_a) = \sum_{m=F+1}^{d_{k_t}^{\text{in}}(t)+1-F} (\xi_{k_t}^m - x_{k_t}) - \sum_{l=F+1}^{d_{j_t}^{\text{in}}(t)+1-F} \left(\xi_{j_t}^l - x_{j_t}\right). \tag{14}$$

This equation is the basis of the agreement argument below.

4.2.1 Agreement

Here we combine Lemmas 3 and 4 with the assumption $\mathcal{D}_s \in \Gamma_{M,F}$ or $\mathcal{D}_s \in \Gamma_{B,F}$ for malicious or Byzantine adversaries, respectively, in order to show global exponential convergence of x_c to $\mathcal{A}$.

THEOREM 1. *Suppose each cooperative agent in $\mathcal{V}_c$ executes ARC-P with parameter $F \in \mathbb{Z}_{\ge 0}$ and at most (i) F malicious agents with $\mathcal{D}_s \in \Gamma_{M,F}$, or (ii) F Byzantine agents with $\mathcal{D}_s \in \Gamma_{B,F}$. Then x_c globally exponentially converges to the agreement space $\mathcal{A}$, and therefore the agreement condition (2) is satisfied. Moreover, the convergence to the agreement space is bounded by*

$$\textit{dist}(x_c(t), \mathcal{A}) \le 2\sqrt{p}\,\textit{dist}(x_c(0), \mathcal{A})e^{-t}. \tag{15}$$

PROOF. (i) Fix $t \geq 0$ and consider (14). Since there are at most F adversaries, each term in the first sum is nonpositive and each term in the second sum is nonnegative. If at least one of the sorted values in the second sum is greater than or equal to any of the values in the first, say $\xi_{k_t}^{m'} \leq \xi_{j_t}^{l'}$, then

$$D^+\Psi(x_c, x_a) \leq -\Psi(x_c), \tag{16}$$

since, in this case,

$$D^+\Psi(x_c, x_a) = \sum_{\substack{m=F+1 \\ m\neq m'}}^{d^{\text{in}}_{k_t}(t)+1-F} (\xi_{k_t}^m - x_{k_t}) - \sum_{\substack{l=F+1 \\ l\neq l'}}^{d^{\text{in}}_{j_t}(t)+1-F} (\xi_{j_t}^l - x_{j_t}) + \left(\xi_{k_t}^{m'} - \xi_{j_t}^{l'}\right) - \Psi(x_c) \leq -\Psi(x_c).$$

A sufficient condition for this to hold, given that *all* agents convey the same values to all neighbors, is to ensure there is a common value in the two sums, e.g., $\xi_{k_t}^{m'} = \xi_{j_t}^{l'}$. This is guaranteed if $|\mathcal{J}_{j_t}^{\text{in}} \cap \mathcal{J}_{k_t}^{\text{in}}| > 2F$, which is obviously true if property $M2_F$ holds. If only property $M1_F$ holds, it must also be the case, since otherwise, we reach the contradiction

$$n \geq |\mathcal{J}_{j_t}^{\text{in}} \cup \mathcal{J}_{k_t}^{\text{in}}| = |\mathcal{J}_{j_t}^{\text{in}}| + |\mathcal{J}_{k_t}^{\text{in}}| - |\mathcal{J}_{j_t}^{\text{in}} \cap \mathcal{J}_{k_t}^{\text{in}}| \geq 2(\lfloor \frac{n}{2} \rfloor + F + 1) - 2F \geq n + 1.$$

Therefore, (16) holds for all $t \geq 0$, and hence it can be shown that

$$\Psi(x_c(t)) \leq \Psi(x_c(0))e^{-t}.$$

Finally, using (10), we conclude (15). Thus, we have shown global exponential convergence of x_c to $\mathcal{A}$.

(ii) The argument is identical to (i), except here to ensure there exists m' and l' such that $\xi_{k_t}^{m'} = \xi_{j_t}^{l'}$ and thereby guarantee (16), we need $|\mathcal{J}_{j_t}^{\text{in}} \cap \mathcal{J}_{k_t}^{\text{in}}| > 3F$. This is required if there are F Byzantine agents in the intersection because of the following argument. Suppose F of the cooperative agents' states are strictly greater than F other cooperative agents in the intersection. Then there are $3F$ agents in the intersection, and the adversaries may create $2F$ different values all strictly between these two sets of cooperative agent states. Thus, at least one more cooperative agent in the intersection is necessary to ensure a common value. Analogously to (i), property $B2_F$ guarantees $|\mathcal{J}_{j_t}^{\text{in}} \cap \mathcal{J}_{k_t}^{\text{in}}| > 3F$ by construction, and so does property $B1_F$. Otherwise, we reach the contradiction

$$n \geq \begin{cases} 2(\frac{n}{2} + \lfloor \frac{3F}{2} \rfloor + 1) - 3F \geq n+1 & n \text{ even \& } F \text{ odd;} \\ 2(\lfloor \frac{n}{2} \rfloor + \lceil \frac{3F}{2} \rceil + 1) - 3F \geq n+1 & \text{otherwise.} \end{cases}$$ □

Notice in the proof of Theorem 1 that it is not necessary that there exists a common in-neighbor in the reduced set of in-neighbors of j_t and k_t to show (16), and therefore (15). All that is required is that there exist $\xi_{k_t}^{m'}$ and $\xi_{j_t}^{l'}$ such that $\xi_{k_t}^{m'} \leq \xi_{j_t}^{l'}$. However, because this must hold globally (i.e., for all $x_c(0) \in \mathbb{R}^p$ and $x_{a_i}(t) \in \mathbb{R}^p$ for $i \in \mathcal{V}_c$) and for all $t \geq 0$, it is untenable to depend on the values in those neighborhoods without insisting that there is a cooperative agent as a common in-neighbor.

4.2.2 Safety

In this section, we verify that the safety condition (3) holds by using an invariant set argument.

THEOREM 2. *Suppose each cooperative agent in $\mathcal{V}_c$ executes ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most (i) F malicious agents with $\mathcal{D}_s \in \Gamma_{M,F}$, or (ii) F Byzantine agents with $\mathcal{D}_s \in \Gamma_{B,F}$. Then the safety condition (3) is satisfied.*

PROOF. Lemma 1 implies that for each $i \in \mathcal{V}_c$

$$-(n-2F)\Psi(x_c) \leq f_{i,\sigma(t)}(x_c, x_{a_i}) \leq (n-2F)\Psi(x_c). \tag{17}$$

It was shown in the proof of Theorem 1 that under either (i) or (ii), $\lim_{t\to\infty} \Psi(x_c(t)) = 0$. Hence, (17) implies

$$\lim_{t\to\infty} f_{i,\sigma(t)}(x_c, x_a) = 0,$$

and thus $\lim_{t\to\infty} x_i(t)$ exists. Since $\mathcal{H}_0$ is compact, Lemma 2 implies the result. □

4.3 Necessary Condition

Next, we consider the following necessary condition for ARC-P to achieve agreement in networks with fixed topology.

THEOREM 3. *Consider a networked multi-agent system that executes ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most $F < n$ malicious or Byzantine agents. If the agreement condition is satisfied, then $\delta^{in}(\mathcal{D}_s) \geq 2F$.*

PROOF. The case $F = 0$ is vacuously true, so assume $F \geq 1$. Suppose $\exists i \in \mathcal{V}_c$ with $d_i^{\text{in}} < 2F$ and $\exists \epsilon > 0$ such that $x_j(0) - x_i(0) > \epsilon \; \forall j \in \mathcal{V}_c \setminus \{i\}$. If $d_i^{\text{in}} \geq F - 1$, let $F - 1$ of i's in-neighbors be adversaries with values smaller than $x_i(0)$. Then, $\dot{x}_i \equiv 0$ since both the cooperative and adversary values are removed. On the other hand, using Lemma 2 while treating i as an adversary, ensures $x_j(t) - x_i(t) > \epsilon \; \forall j \in \mathcal{V}_c \setminus \{i\}$ and $t \geq 0$. □

Recall that the sufficient conditions imply the necessary condition, $\delta^{\text{in}}(\mathcal{D}_s) \geq 2F$. However, the converse is clearly not true. The question then arises, are the sufficient conditions also necessary? The answer is no, but we delay further discussion of the conservativeness of the sufficient conditions until Section 5. Next, we study the sufficient conditions under switching network topologies.

4.4 Switching Topology

Switching network topologies can arise from a number of factors: temporary removal of edges due to lossy communication channels, the addition or loss of edges caused by mobile agents, and so on. The results of the previous sections may be extended to switching topologies in a straightforward manner by assuming $\mathcal{D}_{\sigma(t)} \in \Gamma_{M,F}$ or $\mathcal{D}_{\sigma(t)} \in \Gamma_{B,F}$ for $t \geq 0$ whenever the adversaries are malicious or Byzantine, respectively. It is shown in Theorem 1 that Ψ is a Lyapunov function for each possible digraph $\mathcal{D}_s \in \Gamma_{M,F}$ or $\mathcal{D}_s \in \Gamma_{B,F}$. Further, the upper bound on convergence of x_c to $\mathcal{A}$ (15) holds globally and for each digraph $\mathcal{D}_s \in \Gamma_{M,F}$ or $\mathcal{D}_s \in \Gamma_{B,F}$. Therefore, Ψ is a common Lyapunov function, thus proving global exponential convergence of x_c to $\mathcal{A}$ for the switched system (1). On the other hand, Lemma 2 and (17) hold for all network topologies. Therefore, the same argument used in the proof of Theorem 2 may be used for the case of switching topologies. Hence, we have the following result.

COROLLARY 1. *Suppose each cooperative agent in $\mathcal{V}_c$ executes ARC-P with parameter $F \in \mathbb{Z}_{\geq 0}$ and at most (i) F malicious agents with $\mathcal{D}_{\sigma(t)} \in \Gamma_{M,F}$ for all $t \in \mathbb{R}_{\geq 0}$, or (ii) F Byzantine agents with $\mathcal{D}_{\sigma(t)} \in \Gamma_{B,F}$ for all $t \in \mathbb{R}_{\geq 0}$. Then the agreement condition (2) is satisfied with the convergence to the agreement space bounded by (15), and the safety condition (3) is satisfied. Therefore, under these conditions, ARC-P **solves** the adversarial asymptotic agreement problem in the presence of (i) malicious and (ii) Byzantine agents.*

So far we have studied explicit switching in the network topology when the range of the switching signal is appropriately restricted

(i.e., $\mathcal{D}_{\sigma(t)} \in \Gamma_{M,F}$ or $\mathcal{D}_{\sigma(t)} \in \Gamma_{B,F}$ for all $t \in \mathbb{R}_{\geq 0}$). But, even in fixed network topology, the algorithm ARC-P may be viewed as the linear consensus protocol of [14] with state-dependent switching. In ARC-P, the sort and reduce functions effectively remove the influence of a subset of neighbors based on the state values of those neighbors. The remaining relative states are summed as input to the integrator in the same manner as all of the neighbors are in the linear consensus protocol of [14], which justifies the analogy. Hence, the results of Section 4.2 provide new insight into the convergence of the protocol of [14] with state-dependent switching.

5. EXAMINATION OF CONDITIONS

In this section, we examine the conditions $M1_F$ and $M2_F$ that define $\Gamma_{M,F}$ and $B1_F$ and $B2_F$ that define $\Gamma_{B,F}$. Important questions arise with regard to these properties: (i) How do these conditions relate to known conditions on the maximum number of Byzantine processors in the network [1, 7]; (ii) How do they relate to conditions on the connectivity of the network when reaching agreement with Byzantine processors [1], or detecting and isolating malicious agents [18,24]; (iii) How conservative are the conditions with respect to achieving the adversarial agreement problem using ARC-P; and (iv) How applicable are the conditions to networks of interest? The first question has been answered in Section 4.2, where we showed that $B1_F$ and $B2_F$ imply $n > 3F$, which is a necessary condition when dealing with Byzantine behavior of finite automata in synchronous networks [1,7]. The rest of this section is devoted to addressing the remaining questions.

To address (ii), we show that $M1_F$ and $M2_F$–and therefore also $B1_F$ and $B2_F$–imply $\kappa(\mathcal{D}) \geq 2F + 1$, which is a necessary and sufficient condition for the existence of an algorithm that can (a) ensure agreement of the nonfaulty nodes in the presence of at most F Byzantine nodes in synchronous networks [1], or (b) detect and isolate up to F malicious nodes in linear consensus networks [18, 24].

THEOREM 4. *If $F \in \{0, 1, \ldots, \lfloor n/2 \rfloor - 1\}$ and the digraph satisfies (i) $M1_F$ or (ii) $M2_F$, then $\mathcal{D}$ is $2F + 1$-connected.*

PROOF. (i) Fix $F \in \{0, 1, \ldots, \lfloor n/2 \rfloor - 1\}$ and consider the underlying graph $\mathcal{G}$, which must satisfy $\delta(\mathcal{G}) \geq \lfloor n/2 \rfloor + F$. By Menger's Theorem, $\kappa(\mathcal{G}) \geq 2F + 1$ is equivalent to $\mathcal{G}$ having at least $2F + 1$ vertex-disjoint paths between any distinct vertices $i, j \in \mathcal{V}$. Indeed, this is the case if $|\mathcal{J}_i \cap \mathcal{J}_j| \geq 2F + 2$ for all $i, j \in \mathcal{V}$. On the other hand, we know that $|\mathcal{J}_i \cap \mathcal{J}_j| \geq 2F+1$ (c.f. the proof of Theorem 1). From this we conclude that if $(i, j) \notin \mathcal{E}_\mathcal{G}$ then there are at least $2F + 1$ vertex-disjoint paths between i and j. Therefore, assume there exists $i, j \in \mathcal{V}$ such that $(i, j) \in \mathcal{E}_\mathcal{G}$ and $|\mathcal{J}_i \cap \mathcal{J}_j| = 2F + 1$. In this case, there are $2F$ vertex-disjoint paths accounted for with vertices in $\mathcal{J}_i \cap \mathcal{J}_j$. But, because $F \leq \lfloor n/2 \rfloor - 1$, we know

$$|\mathcal{J}_i|, |\mathcal{J}_j| \geq \lfloor n/2 \rfloor + F + 1 \geq 2F + 2,$$

which means there exists $i' \in \mathcal{J}_i \setminus \mathcal{J}_i \cap \mathcal{J}_j$ and $j' \in \mathcal{J}_j \setminus \mathcal{J}_i \cap \mathcal{J}_j$. If $(i', j') \in \mathcal{E}_\mathcal{G}$, then i, i', j', j is the last vertex-disjoint path necessary to conclude $2F + 1$-connectivity. If $(i', j') \notin \mathcal{E}_\mathcal{G}$, then we know that $|\mathcal{J}_{i'} \cap \mathcal{J}_{j'}| \geq 2F + 1$, and there are at most $2F - 1$ vertices in $(\mathcal{J}_{i'} \cap \mathcal{J}_{j'}) \cap (\mathcal{J}_i \cap \mathcal{J}_j)$ because i and j cannot be in $\mathcal{J}_{i'} \cap \mathcal{J}_{j'}$. Hence, there exists $m \in \mathcal{J}_{i'} \cap \mathcal{J}_{j'} \setminus \mathcal{J}_i \cap \mathcal{J}_j$, so that i, i', m, j', j is the last vertex-disjoint path necessary to conclude $2F + 1$-connectivity.

(ii) Any vertex cut must contain at least $2F+1$ vertices, because otherwise a vertex remains in $\mathcal{S}$ adjacent to all other vertices. □

To address the conservativeness of the conditions with respect to convergence of ARC-P, we show that we can do no better using traditional metrics such as in-degree, out-degree, or connectivity. We do this by demonstrating that minimally relaxing these conditions leads to pathological examples with high connectivity in which ARC-P does not achieve agreement.

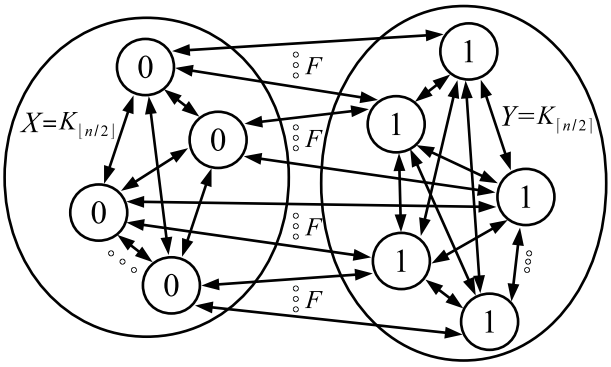


Figure 2: Relax $M1_F$ with $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + F - 1$.

Example 1 [Relax $M1_F$ with $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + F - 1$]. Consider the network topology in Figure 2, in which $K_{\lceil n/2 \rceil}$ is the complete digraph on $\lceil n/2 \rceil$ vertices, and each vertex in X has exactly F neighbors in Y and each vertex in Y has either $F - 1$ or F neighbors in X. Now, assume there are no adversaries and let all states in X have value 0 and all states in Y have value 1. Then, by (5), all agents in X will remove the influence of their neighbor in Y and vice versa. Hence, no consensus is reached, and no agent even changes its state. Furthermore, this graph is $(\lfloor n/2 \rfloor + F - 1)$-connected, which for large n may be much larger than $\kappa(\mathcal{D}) \geq 2F + 1$.

From this example, we see that reducing the minimum in-degree by just one from $M1_F$ is not sufficient for global convergence of x_c to $\mathcal{A}$. Additionally, in this example, the connectivity is very high. This suggests that the minimum in-degree and connectivity are not appropriate metrics to use in characterizing the network toplogies in which ARC-P achieves agreement. The following example demonstrates that the minimum out-degree is also inadequate and further emphasizes the inadequacy of connectivity. Here, the number of nodes in $\mathcal{S}$ from $M2_F$ is reduced by one.

Example 2 [Relax $M2_F$ with $|\mathcal{S}| = 2F$ and $d_i^{\text{out}} = n-2, \forall i \in \mathcal{V} \setminus \mathcal{S}$, so that $\delta^{\text{out}}(\mathcal{D}) = n-2$]. Consider the example of Figure 3, which has $|\mathcal{S}| = 2F$, with $\mathcal{S} = \mathcal{S}' \cup \{j\}$ and $d_i^{\text{in}} = n-2, \forall i \in \mathcal{V} \setminus \mathcal{S}$, so that $\delta^{\text{out}}(\mathcal{D}) = n-2$. Since $d_j^{\text{in}} = 2F-1$, this example does not satisfy the necessary condition of Theorem 3. The argument in the proof shows that the agreement condition is not satisfied. Since the underlying graph is complete, this digraph is $(n-1)$-connected, which emphasizes the inadequacy of connectivity in characterizing the convergence properties of ARC-P.

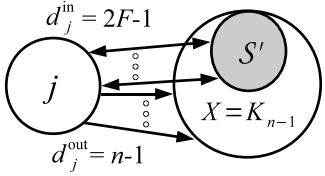


Figure 3: Relax $M2_F$ with $|\mathcal{S}| = 2F$ and $\delta^{\text{out}}(\mathcal{D}) = n - 2$.

Example 3 [Relax $B1_F$ with $\delta^{\text{in}}(\mathcal{D}) = n/2 + \lfloor 3F/2 \rfloor - 1$ if n is even and F is odd, and $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + \lceil 3F/2 \rceil - 1$ otherwise]. Consider the digraph shown in Figure 4. In the figure, the digraph is partitioned into 3 cliques (i.e., complete subdi-

graphs), $\mathcal{D} = X_1 \cup X_2 \cup X_3$, and each clique has $\lfloor n/2 \rfloor - \lfloor F/2 \rfloor$, F, and $\lceil n/2 \rceil - \lceil F/2 \rceil$ nodes, respectively. For clarity, we do not show edges internal to the cliques. We only show one representative node from the sets X_1 and X_3, but all nodes in each of these sets have F in-neighbors in each of the other two sets–which is possible since $n > 3F$. This leads to an in-degree of $d_i^{\text{in}} = \lfloor n/2 \rfloor + \lceil 3F/2 \rceil - 1$ for each $i \in X_1$, and an in-degree of $d_j^{\text{in}} = \lceil n/2 \rceil + \lfloor 3F/2 \rfloor - 1$ for each $j \in X_3$. On the other hand, the nodes in X_2 exchange information bidirectionally with all other nodes, so that $d_k^{\text{in}} = d_k^{\text{out}} = n - 1$ for all $k \in X_2$. Therefore, the minimum in-degree depends on the parity of n and F. If they have the same parity, $|X_1| = |X_3|$, and $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + \lceil 3F/2 \rceil - 1$. If n is odd and F is even, $|X_3| = |X_1| + 1$, and $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + \lceil 3F/2 \rceil - 1$. But, if n is even and F is odd, $|X_3| = |X_1| - 1$, and $\delta^{\text{in}}(\mathcal{D}) = n/2 + \lfloor 3F/2 \rfloor - 1$, which means, in any case, $B1_F$ is minimally relaxed.

To show that ARC-P may not achieve agreement in this digraph, let each node in X_1 and X_3 have initial value 1 and 3, respectively. Suppose all nodes in X_2 are Byzantine, and they transmit a constant trajectory of 1 to nodes in X_1 and 3 to nodes in X_3. Then nodes in X_1 remove the influence from their F neighbors in X_3 and vice versa, so that agreement fails.

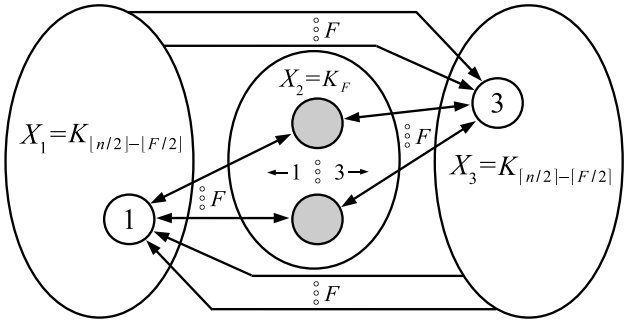


Figure 4: Relax $B1_F$ with $\delta^{\text{in}}(\mathcal{D}) = n/2 + \lfloor 3F/2 \rfloor - 1$ if n is even and F is odd, and $\delta^{\text{in}}(\mathcal{D}) = \lfloor n/2 \rfloor + \lceil 3F/2 \rceil - 1$ otherwise.

Example 4 [Relax $B2_F$ with $|\mathcal{S}| = 3F$ and $\mathcal{S} = S_1 \cup S_2 \cup S_3$]. Consider the digraph in Figure 5. In this example, $\mathcal{S} = S_1 \cup S_2 \cup S_3$, with $|S_i| = F$ for $i = 1, 2, 3$. The remaining nodes in $\mathcal{V} \setminus \mathcal{S}$ form a clique, K_{n-3F}. Nodes in S_1 and S_3 have value 1 and 3, respectively, and nodes in $\mathcal{V} \setminus \mathcal{S}$ have value 2. Nodes in S_2 are Byzantine and send values 1, 2, and 3, respectively, to nodes in S_1, $\mathcal{V} \setminus \mathcal{S}$, and S_3. Clearly, as in the previous examples, the cooperative nodes do not reach agreement, but remain fixed at their initial values.

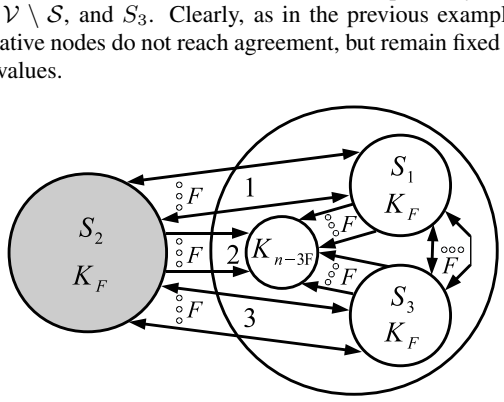


Figure 5: Relax $B2_F$ with $|\mathcal{S}| = 3F$ and $\mathcal{S} = S_1 \cup S_2 \cup S_3$.

Although this section is replete with pathological examples in which ARC-P fails to achieve agreement–even when the networks have high minimum degrees and high connectivity–the news is not all bad. First, we now know that the sufficient conditions studied in Section 4.2 are the best we can have using minimum degrees and connectivity. Second, we can discern a pattern in the various examples. A common property is that there are pairs of subsets with high connectivity within the subsets, but nodes in each subset have relatively few in-neighbors outside of their subsets. Therefore, new toplogical conditions for digraphs that deal with (a) pairs of subsets of nodes and (b) the number of nodes with "enough" in-neighbors outside of their respective subset will be crucial to better understanding the convergence properties of ARC-P. Finally, we end the section by demonstrating the results with the following example.

Example 5: [Morale dynamics on fixed topology with single Byzantine agent] Consider a variation of the Byzantine generals problem in which the loyal generals attempt to improve the morale of their troops and reach consensus on the level of morale despite the influence of a subset of Byzantine generals. In addition, the troops have no knowledge of the goal of the generals. For the purposes of this example, the state value represents the level of morale. The sign of the value indicates either good (positive) or bad (negative) morale and the magnitude signifies the relative levels of morale. Here, we assume that the morale dynamics of each node behave as an integrator with the input (influence) either given by ARC-P, as in (5), or simply by the sum of relative morale values:

$$\dot{x}_i(t) = \sum_{j \in \mathcal{N}_i} \left(x_j(t) - x_i(t) \right), \; x_i(0) = x_{0_i}, \tag{18}$$

where $x_i(t)$ is the morale value of node i and x_{0_i} is the initial morale value of node i. We refer to the influence rule of (18) as the linear consensus protocol (LCP), which is a special case of the weighted sum of relative states studied extensively in the literature [14], and has been compared with ARC-P in the special case of complete networks in [8].

Each general is able to continuously influence all of the troops and the other generals, and the generals can provide different influence to different individuals. The influence network is shown in Figure 6, in which nodes 17 through 20 form a clique and are the generals (shown as squares). The other nodes are the troops (shown as circles). Troop i has initial morale $-i$, for $i = 1, \ldots, 16$, and the generals have initial morale of 1, 2, 3, and 4, respectively, for nodes 17, 18, 19, and 20.

The central question of this example is whether either LCP or ARC-P can ensure that the troops reach asymptotic consensus on a positive morale given that it is possible that one of the generals is Byzantine (i.e., $F = 1$). Observe that the network of Figure 6 satisfies $B2_F$ whenever $F = 1$, with $\mathcal{S} = \{17, 18, 19, 20\}$, and can therefore sustain the compromise of a single node as Byzantine whenever the troops and loyal generals use ARC-P. In this case, we choose node 20 to be the Byzantine general. In order to elude detection, the Byzantine general conveys a morale trajectory that satisfies the preassigned strategy–either ARC-P or LCP–to the other generals. But, to the troops, the Byzantine general conveys a highly negative morale of -87.5. The results for LCP and ARC-P are shown in Figure 7. The Byzantine morale trajectory shown in the figures is the one conveyed to the other generals. Using LCP, the troops reach consensus at a negative morale of -20 and the generals reach consensus at 2.5, whereas with ARC-P the troops reach consensus at the same value of the other generals at 2.5.

This example illustrates an important property of ARC-P: *It only requires local information for resilience against adversaries*. In contrast, without nonlocal information, the detection and identification techniques of [16–19, 22–25] would not successfully detect the Byzantine general. This is because from the perspective of the

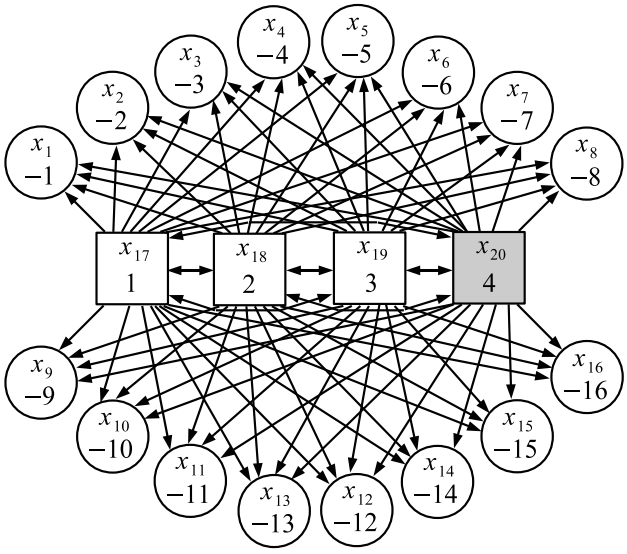


Figure 6: Influence network in which square nodes are generals and circular nodes are troops. Node 20 is Byzantine.

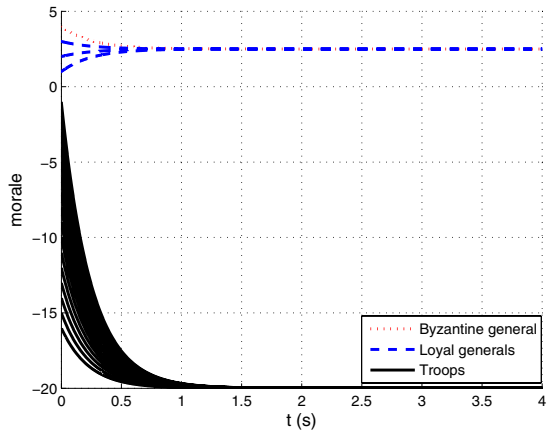
loyal generals, the Byzantine general behaves as it should and they receive no feedback from the troops. From the perspective of the troops, the Byzantine general appears to be influenced by no other node. Hence, without prior knowledge of at least some nonlocal aspects of the network topology, the Byzantine general remains undetected.

6. RELATED WORK

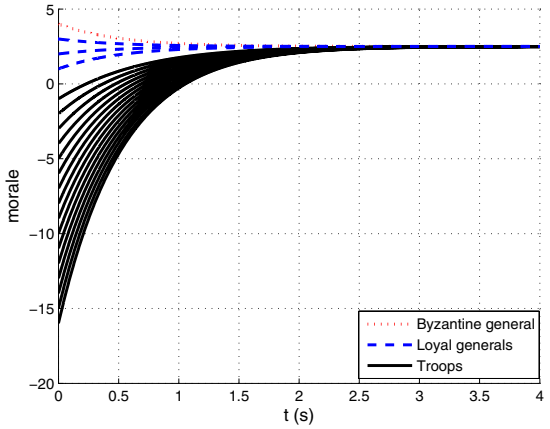
The research most closely related to this work is [16–19, 22–25]. In [16], the issue of detecting and identifying a single misbehaving agent using a linear iterative strategy in discrete-time synchronous networks is introduced. Then, Sundaram and Hadjicostis show in [22] that $\kappa(G) \geq 2F+1$ is a necessary condition for detecting and identifying up to F malicious agents using linear iterations in synchronous networks. In the companion paper [23], $\kappa(G) \geq 2F+1$ is shown to be sufficient for the problem. In this case, the linearity of the protocol is exploited so that every node is able to calculate the initial values exactly, and thus any function of the initial states, in at most n steps. The results of [22, 23] are generalized in [24] to characterize under which conditions any subset of nodes can obtain all of the initial values.

The authors of [16], later extend the analysis done in [22, 23] by characterizing the type of behavior of the malicious agents that is most troublesome to the linear network and by characterizing the network connectivity required to tolerate both malicious agents and non-colluding agents in [17]. A computationally expensive but exact algorithm is presented in [17] to detect and identify up to F malicious agents in networks with connectivity at least $2F+1$. This exact algorithm requires each node to know the topology of the entire network. In [19], two approaches are considered to reduce the computational complexity and require only partial network information. The first assumes the network is comprised of weakly interconnected subcomponents and restricts the behavior of the misbehaving nodes. The second imposes a hierarchical structure to detect and isolate the malicious agents. These results are combined and extended in [18].

In [25], the authors study detection and identification of cyber attacks on networked control systems modeled as continuous-time



(a) LCP.



(b) ARC-P.

Figure 7: Byzantine general attempts to reduce morale of the troops. Byzantine morale shown is the one conveyed to other generals. Byzantine morale conveyed to troops is -87.5. ARC-P succeeds in the goal of improving the morale while reaching consensus, but LCP fails.

linear systems. Attacks on nodes and on their outgoing communication channels are both studied, and it is shown that from the perspective of other nodes, the two cases are indistinguishable. As in [18], unknown input observers are used for the FDI scheme. The approach is demonstrated on a network of nodes using the linear consensus protocol of (18) augmented with the FDI scheme, and on the swing equation of a power network.

There are several differences between the related works and this paper. First, the aforementioned works require nonlocal information on the network topology to ensure consensus. ARC-P requires *only local information*. Second, the computational burden of the FDI algorithms is greater than ARC-P, which is low complexity. Third, we study directed information flow in both fixed and switching topologies. The FDI schemes would not be able to handle this case because of the nonlocal information required on the network topology. Fourth, the other works do not consider safety conditions and are therefore not suitable for safety critical applications. Lastly, we study both malicious and Byzantine agents, whereas the aforementioned works do not consider Byzantine agents.

Finally, the reader may wonder how this paper relates to robust

consensus algorithms designed to withstand outliers [9, 13]. The problem of robust consensus to outliers does not assume a threat model, such as malicious or Byzantine nodes. Instead, some measurements may be statistical outliers caused by noisy measurements and the goal is to reach consensus on the measurements in a manner that reduces the error introduced by the outliers. In these works the nodes with outlier measurements are cooperative in the consensus process. Therefore, such techniques are not designed to work in the presence of adversaries.

7. CONCLUSIONS

In this paper, we have studied a low complexity protocol (algorithm), ARC-P, for reaching consensus in networked multi-agent systems with adversaries. We formulated a consensus problem, the adversarial asymptotic agreement problem, appropriate for distributed control applications. We defined two different models for adversaries depending on how information is conveyed. Malicious agents must convey the same information to each neighbor, whereas Byzantine agents may convey different information to each neighbor. We analyzed the convergence properties of ARC-P in directed networks with fixed and switching topologies in the presence of malicious and Byzantine agents, while restricting the range of the switching signal so that each topology satisfies sufficient conditions on the in-degrees and out-degrees of nodes in the network. Finally, we examined the conservativeness of the conditions.

Based on the examples in Section 5, it is clear that traditional graph theoretic metrics like minimum degree and connectivity are not suitable for characterizing under which conditions ARC-P ensures agreement. Therefore, to ascertain conditions which are both necessary and sufficient, new graph theoretic metrics are needed.

8. ACKNOWLEDGMENTS

The authors would like to thank Shreyas Sundaram for suggesting the example of Figure 2. This work is supported in part by the National Science Foundation (CNS-1035655, CCF-0820088), the U.S. Army Research Office (ARO W911NF-10-1-0005), and Lockheed Martin.

9. REFERENCES

[1] D. Dolev. The Byzantine generals strike again. *Journal of Algorithms*, 3(1):14 – 30, 1982.

[2] 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*, 33(3):499 – 516, 1986.

[3] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Trans. on Aut. Control*, 49(9):1465 – 1476, 2004.

[4] D. Geller and F. Harary. Connectivity in digraphs. In *Recent Trends in Graph Theory*, volume 186 of *Lect. Notes in Math.*, pages 105–115. Springer Berlin / Heidelberg, 1971.

[5] V. Gupta, C. Langbort, and R. Murray. On the robustness of distributed algorithms. In *IEEE Conf. on Decision and Control*, Dec. 2006.

[6] T. T. Johnson and S. Mitra. Safe flocking in spite of actuator faults using directional failure detectors. *Journal of Nonlinear Sys. and App.*, 2(1-2):73–95, 2011.

[7] L. Lamport, R. Shostak, and M. Pease. The Byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, 1982.

[8] H. J. LeBlanc and X. D. Koutsoukos. Consensus in networked multi-agent systems with adversaries. In *Proc. of the 14th Int. Conf. on Hybrid systems: Computation and Control*, (HSCC '11), pages 281–290, Chicago, IL, 2011.

[9] J. Li, E. Elhamifar, I.-J. Wang, and R. Vidal. Consensus with robustness to outliers via distributed optimization. In *IEEE Conf. on Decision and Control*, pages 2111–2117, Dec. 2010.

[10] D. Liberzon. *Switching in Systems and Control*. Birkhauser, Boston, MA, USA, 2003.

[11] N. A. Lynch. *Distributed Algorithms*. Morgan Kaufmann Publishers Inc., San Francisco, California, 1997.

[12] M. Mesbahi and M. Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, Princeton, New Jersey, 2010.

[13] E. Montijano, S. Martínez, and S. Sagués. De-RANSAC: robust distributed consensus in sensor networks. *IEEE Transactions on Systems, Man, and Cybernetics: part B*. Submitted, May 2010.

[14] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[15] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. In *Networked Embedded Sensing and Control*, volume 331 of *Lecture Notes in Control and Information Sciences*, pages 169–182. Springer Berlin / Heidelberg, 2006.

[16] F. Pasqualetti, A. Bicchi, and F. Bullo. Distributed intrusion detection for secure consensus computations. In *IEEE Conf. on Decision and Control*, pages 5594 –5599, Dec. 2007.

[17] F. Pasqualetti, A. Bicchi, and F. Bullo. On the security of linear consensus networks. In *IEEE Conf. on Decision and Control*, pages 4894 – 4901, Dec. 2009.

[18] F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Trans. on Aut. Control*, 57(1):90–104, Jan. 2012.

[19] F. Pasqualetti, R. Carli, A. Bicchi, and F. Bullo. Identifying cyber attacks under local model information. In *IEEE Conf. on Decision and Control*, pages 5961–5966, Dec. 2010.

[20] M. Pease, R. Shostak, and L. Lamport. Reaching agreement in the presence of faults. *J. ACM*, 27(2):228–234, 1980.

[21] N. Rouche, P. Habets, and M. Laloy. *Stability Theory by Liapunov's Direct Method*, volume 22 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1977.

[22] S. Sundaram and C. Hadjicostis. Distributed function calculation via linear iterations in the presence of malicious agents; part I: Attacking the network. In *American Control Conf.*, pages 1350 –1355, June 2008.

[23] S. Sundaram and C. Hadjicostis. Distributed function calculation via linear iterations in the presence of malicious agents; part II: Overcoming malicious behavior. In *American Control Conf.*, pages 1356 –1361, June 2008.

[24] S. Sundaram and C. Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Trans. on Aut. Control*, 56(7):1495 –1508, July 2011.

[25] A. Teixeira, H. Sandberg, and K.H. Johansson. Networked control systems under cyber attacks with applications to power networks. In *American Control Conf.*, pages 3690–3696, July 2010.

[26] J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Dept of EECS, MIT, 1984.