Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A "Safe Kernel" Approach for Resilient Multi-Dimensional Consensus

Jiaqi Yan * **Yilin Mo** ** **Xiuxian Li** * **Changyun Wen** *

* *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. (e-mails: jyan004@e.ntu.edu.sg, xiuxianli@ntu.edu.sg, ecywen@ntu.edu.sg)*
** *Department of Automation and BNRist, Tsinghua University, Beijing, China. (e-mail: ylmo@tsinghua.edu.cn)*

Abstract: This paper considers the resilient multi-dimensional consensus problem in networked systems, where some of the agents might be malicious (or faulty). We propose a multi-dimensional consensus algorithm, where at each time step each healthy agent computes a "safe kernel" based on the information from its neighbors, and modifies its own state towards a point inside the kernel. Assuming that the number of malicious agents is locally (or globally) upper bounded, sufficient conditions on the network topology are presented to guarantee that the benign agents exponentially reach an agreement within the convex hull of their initial states, regardless of the actions of the misbehaving ones. It is also revealed that the graph connectivity and robustness required to achieve the resilient consensus increases linearly with respect to the dimension of the agents' state, indicating the existence of a trade-off between the low communication cost and system security. Numerical examples are provided in the end to validate the theoretical results.

Keywords: Average consensus, Resilient algorithm, Multidimensional systems.

1. INTRODUCTION

Recent advances in signal processing and cooperative control have led to growing research interests in networked systems. One of the most important focuses in such systems is the average consensus problem. Given a set of autonomous agents (such as sensors, vehicles, *etc.*), this problem seeks for a distributed protocol that the agents can utilize to reach a common decision/agreement on the average of their initial opinions (see Lynch (1996); Olfati-Saber et al. (2007)).

In the past decades, considerable attention has been paid to the development of distributed consensus algorithms (Olfati-Saber et al. (2007); Ren et al. (2007); Wei and Ozdaglar (2012)). The existing protocols, although effective in solving the problems under mild conditions, are normally based on the hypothesis that every computing agent is trustworthy and cooperate to follow the algorithms throughout the execution. Nevertheless, as the scale of the network increases, it becomes more difficult to secure every agent. One primary reason is that the widely-adopted communication infrastructures in the distributed framework make it much vulnerable to external adversaries (Mo et al. (2012)). Especially, malicious attackers can degrade the algorithm performance by manipulating the transmitted data on communication lines. On the other hand, some agents may not be willing to follow the given rules if they weigh their private interests more than the public ones. They might send out well-designed signals to manipulate the achieved solution for their benefits. It is possible that the misbehaving agents can dictate the final consensus value, or the network may fail to reach an agreement.

It is noted that the consensus problem has been widely applied to the safety-critical systems, such as transportation (Ren et al. (2007); Raffard et al. (2004)), power grids (Kar et al. (2014); Kekatos and Giannakis (2013)). Since the system failures would cause irreparable harm to economy, environment, and even public health, security and resilience are becoming priory considerations when designing the algorithms (Pasqualetti et al. (2012); Cárdenas et al. (2008)). In recent years, the secure protocols of reaching average consensus in the presence of faulty or misbehaving agents have been widely studied (Dolev et al. (1986); LeBlanc et al. (2013); Vaidya et al. (2012)). For example, Dolev et al. (1986) consider the approximate consensus problem, where the approximate, rather than exact, agreement is desired in the presence of malicious agents. They consider only complete networks. In order to overrule the effects of malicious nodes, an updating strategy, namely, Mean-Subsequence Reduced *(MSR)* algorithm, is proposed: each normal agent is required to discard the most extreme values in its neighborhood and updates the state based on the remaining values at any time. In a recent work, LeBlanc et al. (2013) generalize *MSR* to the Weighted Mean-Subsequence-Reduced *(W-MSR)* algorithm. Instead of complete networks, they attempt to analyze this algorithm in more general topologies. A novel property named network robustness is introduced, which characterizes the resilience properties of *W-MSR* in terms of the graph structure. These algorithms, under certain conditions, ensure the agreement within the range

of initial values of the normal nodes, even in an adversarial environment.

However, most of the research on resilient consensus assumes that the agents' states are scalar variables, producing crucial limitations in various practical applications, such as vehicle formation control on a 2D-plane. A naive way to generalize the results on scalar system to multi-dimensional system is to apply *MSR* or *W-MSR* to each entry of the state vectors. The region that the benign agents converge to can be immediately identified as a multi-dimensional "box" limited by the minimum and maximum value of their initial states in every dimension. However, is it possible to design a resilient consensus algorithm that provides more accurate convergence results?

As proved by Su and Vaidya (2015), in the presence of the misbehaving agents, it is impossible for any distributed rule to reach the exact average of the initial states of all benign agents. As a compromise, in this paper, we aim to design a multi-dimensional consensus algorithm that converges to a convex combination of these states. In the developed algorithm, each benign agent creates a "safe kernel" and modifies its state towards a point inside the kernel. Under certain conditions on network topology, we prove that the proposed strategy guarantees the benign agents of reaching an agreement within the convex hull of their initial values, which improves the accuracy of that by simply applying the existing algorithms to each dimension. It is also noted that the "safe kernel" technique can be further extended to other consensus-based problems (*e.g.*, distributed optimization, distributed estimation). Therefore, our work acts as leverage in handling misfunctioning component in multi-dimensional spaces.

Notations: For a vector a, a_i denotes its i-th component. For set $\mathcal{S} \subset \mathbb{R}^d$, $\text{Conv}(\mathcal{S})$ denotes its convex hull, namely the set of all convex combinations of the points in $\mathcal{S}$.

2. PRELIMINARIES

We start by introducing some technical preliminaries on the graph theory, which would be applied in our further analysis.

Consider the network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of agents, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges. An edge between agent i and j is denoted by $e_{ij} \in \mathcal{E}$, indicating these two agents can communicate directly with each other. We define the neighborhood of an agent $i \in \mathcal{V}$ as

$$\mathcal{N}_i = \{j \in \mathcal{V} | e_{ij} \in \mathcal{E}\}.$$

Some definitions on the robustness of graph are discussed below (Zhang et al. (2015)):

Definition 1. (r-robust network): A network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is said to be r-robust, if for any pair of disjoint and nonempty subsets $\mathcal{V}_1, \mathcal{V}_2 \subsetneq \mathcal{V}$, at least one of the following statements hold:

(1) There exists more than one agent in $\mathcal{V}_1$, such that it has at least r neighbors outside $\mathcal{V}_1$;
(2) There exists more than one agent in $\mathcal{V}_2$, such that it has at least r neighbors outside $\mathcal{V}_2$.

Definition 2. ((r, s)-robust network): A network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is said to be (r, s)-robust, if for any pair of disjoint and nonempty subsets $\mathcal{V}_1, \mathcal{V}_2 \subsetneq \mathcal{V}$, at least one of the following statements hold:

(1) Any agent in $\mathcal{V}_1$ has at least r neighbors outside $\mathcal{V}_1$;
(2) Any agent in $\mathcal{V}_2$ has at least r neighbors outside $\mathcal{V}_2$;
(3) There are no less than s agents in $\mathcal{V}_1 \cup \mathcal{V}_2$, such that each of them has at least r neighbors outside the set it belongs to ($\mathcal{V}_1$ or $\mathcal{V}_2$).

Intuitively, the definitions of network robustness claim that for any two disjoint and nonempty subsets of agents, there are "many" agents within those sets that have a sufficient number of neighbors outsides. As we will see, the robust graph plays an important role in our analysis of achieving a resilient agreement.

3. PROBLEM FORMULATION

Consider the network modeled by an undirected and connected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{1, 2, ..., N\}$. At any time $k \geq 0$, let $x^i(k) \in \mathbb{R}^d$ denote the current state of agent i. The agents are said to reach a (distributed) consensus if and only if there exists a constant $\tilde{x}$, such that $\lim_{k\to\infty} x^i(k) = \tilde{x}$ holds for every agent i. In particular, if $\tilde{x} = 1/N \sum_{i=1}^{N} x^i(0)$, an average consensus is achieved.

Many practical applications fit into the framework of average consensus (see Ren et al. (2007); Xiao et al. (2007)). While various strategies have been developed to facilitate it, the linear algorithms have attracted much attention due to their simplicity and ease of implementation. In such strategies, every agent $i \in \mathcal{V}$ is initialized with $x^i(0)$. At each time k, it receives information from all of its neighbors, and updates its own state according to the following equation:

$$x^i(k+1) = a_i^i(k)x^i(k) + \sum_{j\in\mathcal{N}_i} a_j^i(k)x^j(k), \qquad (1)$$

The new state will then be broadcasted to its neighbors preparing for the next updating stage. The conditions under different scenarios to ensure the achievement of average consensus have been investigated widely in the literatures (see Nedic et al. (2010); Olfati-Saber et al. (2007)), the details of which are omitted here due to the space limitation.

We should note that, an implicit assumption for the effectiveness of this approach, and other distributed algorithms as well, is that all agents are reliable throughout the execution, and cooperate to achieve the desired value. However, as the number of local agents increases, certain concerns arise that might make this assumption to be violated. As discussed before, its strong dependence on the communication infrastructures creates lots of vulnerabilities for cyber attacks, where the transmitted information might be manipulated by external adversaries. Additionally, "non-participant" agent may exist, who deviates from the normal update rule and sends out self-designed information for its own benefits. Clearly, such illegal behaviors would degrade the performance of distributed protocols: they can either prevent the benign agents from reaching a consensus, or manipulate the final agreement to be false.

The security concerns lead to the study of resilient consensus protocols. By saying "resilient", we hope to achieve

the following objectives, regardless of the choice of initial states and even in the adversarial environment:

(1) *Agreement:* As k goes to infinity, it is held that $x^i(k) = \bar{x}$ with some $\bar{x} \in \mathbb{R}^d$, for any benign agent i;
(2) *Validity:* At any time and for any benign agent, its state remains in the convex hull of all benign agents' initial values.

We elucidate these conditions as below. Firstly, the states of the benign agents should converge to the same constant value even in the presence of misbehaving ones. In addition, they are not allowed to leave the convex hull of their initial states throughout the procedure. That is, they should avoid being influenced by the misbehaviors too much. It is observed that if 1D problem is considered, then the validity condition would be degraded to that "*the state of any benign agent always remains in the interval forming by the minimum and maximum of their initial states*". There has been much work proved to be effective in this simple case (*e.g.*, *MSR* proposed by Dolev et al. (1986) and *W-MSR* proposed by LeBlanc et al. (2013)). However, few research efforts have been devoted to the more general multi-dimensional systems.

A naive way to tackle this problem is by simply applying the existing scalar protocols to each component of the state vectors. Nevertheless, the region that the benign agents converge to can only be guaranteed as a multi-dimensional "box" limited by the minimum and maximum value of their initial states at every dimension, and thus the validity condition fails to be ensured in this manner. To see this, we present a 2-dimensional illustration in Fig. 1, indicating this naive algorithm cannot guarantee the convergence to a point inside the convex hull of initial states. Therefore, this paper intends to address this problem and come up with a method satisfying both Conditions 1) and 2).

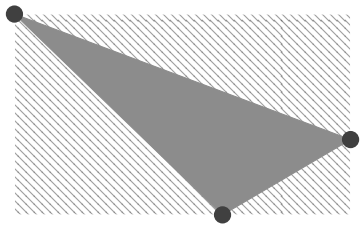

Fig. 1. A 2D illustration with agents marked with circles. The location of the node indicates its initial value. With the direct application of existing algorithms to each dimension, the final agreement is ensured to be within the rectangle represented by oblique lines. However, a better solution satisfying the validity condition of converging to the solid triangle is expected.

3.1 Attack model

We define $\mathcal{F}$ as the set of malicious/faulty agents. Any agent $i \in \mathcal{F}$ could either be the adversarial one with the value being manipulated by the attacker, or the non-participant agent who does not follow the standard updating rule. We also denote $\mathcal{B}$ as the collection of benign agents who will always follow the prescribed updating strategy. It is clear that $\mathcal{B} \cap \mathcal{F} = \varnothing$ and $\mathcal{B} \cup \mathcal{F} = \mathcal{V}$.

The faulty nodes could be characterized by the scope of threats:

(1) (F-total attack model) There are at most F misbehaving agents in the network. That is, $|\mathcal{F}| \leq F$.
(2) (F-local attack model) There are at most F misbehaving agents in the neighborhood of any benign agent. That is, $|\mathcal{F} \cap \mathcal{N}_i| \leq F$, for any agent $i \in \mathcal{B}$.

It is easy to conclude that F-total attack model is a special case of F-local one.

Note that we do not pose any restrictions on the transmitted information of agent $i \in \mathcal{F}$, *i.e.*, the malicious agents are allowed to send out arbitrary data to their neighbors. Furthermore, they could collude among themselves to decide on the deceptive values to be communicated.

4. A RESILIENT MULTI-DIMENSIONAL CONSENSUS STRATEGY

In this section, we provide a resilient consensus algorithm. To simplify notations, we have the following definitions:

Definition 3. Consider a set $\mathcal{A} \subset \mathbb{R}^d$ with cardinality m [1]. Let $\mathcal{S}(\mathcal{A}, n)$ be the set of all its subset with cardinality $m - n$.

It is clear that the set $\mathcal{S}(\mathcal{A}, n)$ contains $\binom{m}{n}$ elements, and each of them is associated with a convex hull. The intersection of all these convex hulls plays a crucial role in our algorithm, which is defined as follows:

Definition 4. Consider the set $\mathcal{A} \subset \mathbb{R}^d$ with cardinality m. We define $\Psi(\mathcal{A}, n)$ as

$$\Psi(\mathcal{A}, n) \triangleq \bigcap_{S \in \mathcal{S}(\mathcal{A}, n)} \mathrm{Conv}(S). \tag{2}$$

Given the F-total/ F-local attack model introduced before, the designed algorithm is formally presented as follows. Each agent $j \in \mathcal{V}$ is initialized with a starting state $x^j(0) \in \mathbb{R}^d$. At any time $k > 0$, every benign agent $i \in \mathcal{B}$ updates as outlined in Algorithm 1.

Algorithm 1 Resilient consensus algorithm

1: Receive the states from all neighboring agents $j \in \mathcal{N}_i$, and collect these values in $\mathcal{X}^i(k)$.
2: Define $\mathcal{R}^i(k) \triangleq \Psi(\mathcal{X}^i(k), F)$, and denote the vertices of this set to be $\mathrm{Ver}(\mathcal{R}^i(k))$. Agent i updates its local state as:

$$x^i(k+1) = a_i^i(k) x^i(k) + \sum_{\bar{x}^j(k) \in \mathrm{Ver}(\mathcal{R}^i(k))} a_j^i(k) \bar{x}^j(k), \tag{3}$$

satisfying that each weight is lower bounded by some $\alpha > 0$, and $a_i^i(k) + \sum_{\bar{x}^j(k) \in \mathrm{Ver}(\mathcal{R}^i(k))} a_j^i(k) = 1$.
3: Transmit updated state $x^i(k+1)$ to all neighbors $j \in \mathcal{N}_i$.

Remark 1. We can interpret $\mathcal{R}^i(k)$ as the "safe kernel" (illustrated in Fig. 2), which is guaranteed to be within the convex hull forming by only benign ones, as we will

[1] To be more precise, $\mathcal{A}$ should be defined as a multi-set since we allow duplicate elements in the set, e.g., the states of m agents shall be counted as m points even if some of them may be identical.

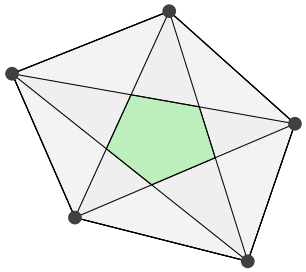

Fig. 2. A 2D illustration of "safe kernel". Suppose agent $i \in \mathcal{B}$ has 5 neighbors and each of their states is represented by the location of a circle. Let $F = 1$. The green region denotes $\mathcal{R}^i(k) = \Phi(\mathcal{X}^i(k), 1)$, namely the "safe kernel".

prove later. Intuitively, at any time, the healthy agent computes and moves its state toward a point inside the "safe kernel". As a result, the impact of malicious agents on the benign ones are limited. The proposed protocol can be implemented in a distributive fashion, as every fault-free agent is only required to access the local information, with no need to have any knowledge of the network topology or impose extra communication among agents.

5. ALGORITHM ANALYSIS

This section is devoted to proving the effectiveness of Algorithm 1. In this paper, we impose the following assumption on the network topology:

Assumption 1. For any $i \in \mathcal{V}$, it is held that $|\mathcal{N}_i| \geq (d+1)F + 1$.

5.1 Realizability

Before analyzing the resiliency of the proposed algorithm, we need to prove its realizability. First, we show that $\mathcal{R}^i(k)$ is non-empty. To this end, we shall begin with the introduction of Helly's theorem, which is a basic result of convexity theory and key supporting technique of this paper.

Theorem 1. (Helly's theorem). (Danzer et al. (1963)) Let $X_1, \cdots, X_p$ be a finite collection of convex subsets in $\mathbb{R}^d$, with $p > d$. If the intersection of every $d+1$ of these sets is nonempty, then the whole collection has a nonempty intersection. That is,

$$\bigcap_{j=1}^{p} X_j \neq \varnothing.$$

A direct result of Helly's theorem is as follows:

Corollary 1. Let $\mathcal{A}$ be a set with cardinality m in $\mathbb{R}^d$. If $m \geq n(d+1)+1$, then for any $n \leq m$, the following relation holds

$$\Psi(\mathcal{A}, n) \neq \varnothing.$$

By Corollary 1, we note that Assumption 1 guarantees that $\mathcal{R}^i(k) \neq \varnothing$ for any $i \in \mathcal{B}$ at any time.

Next we need to show the existence of $a^i_j(k)$. Notice that the number of vertices of $\mathcal{R}^i(k) = \Psi(\mathcal{X}^i(k), F)$ is upper bounded [2], so that the lower bound of the weights α exists. Therefore, one concludes that the proposed strategy is realizable.

5.2 Resiliency

Before proceeding to the main results, we shall first present some preliminary conclusions regarding $\Psi(\mathcal{A}, n)$ [cf. Definition 4]:

Proposition 1. Consider two collections of sets $\{A_i\}_{i\in I}$ and $\{B_j\}_{j\in\mathcal{J}}$. If for any $j \in \mathcal{J}$, there exists an $i \in \mathcal{I}$, such that $B_j \supseteq A_i$, then

$$\bigcap_{j\in\mathcal{J}} B_j \supseteq \bigcap_{i\in\mathcal{I}} A_i.$$

Lemma 1. Consider any set $\mathcal{A}_1$ with cardinality m_1 and $\mathcal{A}_2$ with cardinality m_2. If $\mathcal{A}_1 \subset \mathcal{A}_2$, then for any $n \leq m_1$, the following statement holds:

$$\Psi(\mathcal{A}_1, n) \subset \Psi(\mathcal{A}_2, n).$$

Lemma 2. Let $\mathcal{A}$ be a set with cardinality m in $\mathbb{R}^d$. Supposing that $m \geq (d+1)n + 1$, the following relations hold for any $n \leq m$ and any $p \in \{1, 2, .., d\}$:

(1) If no more than n elements of $\mathcal{A}$ has its p^{th} entry greater than ε, then for any $y \in \Psi(\mathcal{A}, n)$, it is held that $y_p \leq \varepsilon$;
(2) If no more than n elements of $\mathcal{A}$ has its p^{th} entry less than ε, then for any $y \in \Psi(\mathcal{A}, n)$, it is held that $y_p \geq \varepsilon$.

Now we are ready to provide our main results. For simplicity, we denote the convex hull of the states of all benign agents at time k as $\Omega(k)$. The following theorem presents the non-expansion property of $\Omega(k)$:

Theorem 2. (Validity). Consider the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. With Algorithm 1, the following relation holds for any $k \geq 0$:

$$\Omega(k+1) \subset \Omega(k), \tag{4}$$

under either F-local or F-total attack model.

Theorem 2 indicates that the proposed algorithm guarantees the validity condition of resilient consensus. That is, the healthy agents would never be out of the convex hull of their initial values, despite the influence of the misbehaving agents. In what follows, we will provide sufficient conditions on network topology, under which the agreement condition will also be satisfied. Due to the space limitation, the proof of these theorems are omitted.

Theorem 3. (Agreement: F-local). Consider the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Suppose the misbehaving agents follow an F-local attack model. If the network is with $((d+1)F+1)$-robustness, then with Algorithm 1, all the benign agents are guaranteed to achieve consensus exponentially, regardless of the actions of misbehaving agents.

The next theorem elaborates a different condition for the proposed algorithm to succeed under F-total threats:

Theorem 4. (Agreement: F-total). Consider the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Suppose the misbehaving agents follow an F-total attack model. If the network is with $(dF+1, F+1)$-robustness, then with Algorithm 1, all the benign agents

[2] To see this, notice that each convex hull in $\mathcal{S}(\mathcal{X}^i(k), F)$ is a polytope limited with $|\mathcal{X}^i(k)| - F$ number of vertices and its number of facets is bounded due to the Upper Bound Theorem (Ziegler (2012)). As each vertex of $\mathcal{R}^i(k)$ is an intersection of at least d of these facets, we know that its number is upper bounded.

are guaranteed to achieve consensus exponentially, regardless of the actions of misbehaving agents.

Remark 2. By definitions, it is easy to see that a $((d+1)F+1)$-robust graph is $(dF+1, F+1)$-robust as well, but not vice versa. That is to say, the network which is able to tolerate F-local attacks could also survive the F-total ones, while the converse is not true. This observation is consistent with the fact that the F-globally bounded threats are special versions of locally bounded ones.

Based on the above results, one obtains immediately that the proposed algorithm facilitates the resilient consensus. We formally state it in the next theorem:

Theorem 5. Consider the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Suppose the network satisfies one of the following conditions:
1) under F-local attack model, and is $((d+1)F+1)$-robust,
2) under F-total attack model, and is $(dF+1, F+1)$-robust.
With Algorithm 1, all the benign agents finally achieve a consensus within the convex hull of the initial states of benign agents, regardless of the actions of misbehaving ones. That is, as $k \to \infty$,

$$x^i(k) = x^j(k) = \hat{x} \quad \text{for any } i, j \in \mathcal{B}, \tag{5}$$

where $\hat{x} \in \Omega(0)$.

Remark 3. Since the convergence of proposed algorithm does not depend on the actions of misbehaving agents, it works effectively even in the worst-case scenario, where the misbehaving agents could have full knowledge of graph topology, updating rules, *etc*, and could be able to send different data to different neighbors.

Theorem 5 indicates that under certain requirements on network topology, Algorithm 1 guarantees that all benign agents reach an agreement on a weighted average of their initial states, *i.e.*, $\hat{x} = \sum_{i\in\mathcal{B}} \gamma_i x_i(0)$ with $\gamma_i \geq 0$ and $\sum_{i\in\mathcal{B}} \gamma_i = 1$. As proved by Su and Vaidya (2015), if $\mathcal{F}$ is nonempty, it is impossible for any distributed rule to achieve the exact average of these states. Therefore, our algorithm is effective in the sense that a suboptimal result is achieved. It protects the states of the benign agents from being driven to arbitrary values, and thus could withstand the compromise of partial agents while providing a desired level of security.

6. NUMERICAL EXAMPLE

In this section, we provide numerical examples to verify the theoretical results established in the previous sections. In the example, the communication network is given by Fig. 3, in which the node set is $\mathcal{V} = \{1, 2, ..., 5\}$. It is verified that the graph is $(3, 2)$-robust. Suppose that agent 1 is compromised. It intends to prevent others from reaching a correct consensus by violating the rule in Algorithm 1 and setting its states as $x_1^1(k) = 1.5 * \sin(k/5)$ and $x_2^1(k) = k/25+1$ at any time $k > 0$. On the other hand, the benign agents are initialized with $x^2(0) = (1, 2), x^3(0) = (2, 0), x^4(0) = (1, 3), x^5(0) = (2, 4)$, and always follow (3) as updates. For simplicity, let their updating weights be $a_j^i(k) = 1/(|\operatorname{Ver}(\mathcal{R}^i(k))|+1)$ for each $j \in \operatorname{Ver}(\mathcal{R}^i(k)) \cup \{i\}$.

We test the performance of Algorithm 1 in Fig. 4. The result shows that the states of benign agents are always guaranteed within the convex hull of their initial states

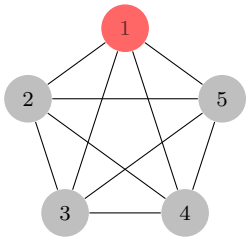

Fig. 3. Communication network.

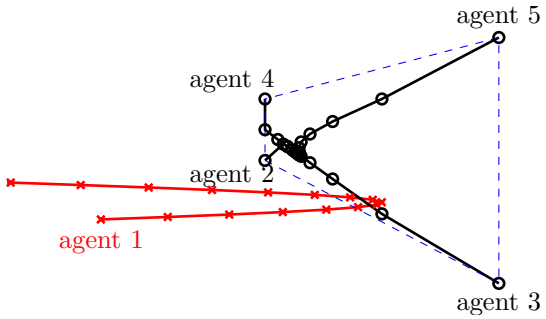

Fig. 4. The trajectory of local states under Algorithm 1, where the area surrounded by the dashed lines is the convex hull of the initial states of benign agents.

and they finally achieve a common value, which validates Theorem 4. That is, the network could tolerate a single misbehaving node in this 2-dimensional problem. Since the malicious agent is unable to affect the final agreement too much, our protocol helps to improve the system security.

7. CONCLUSION

Due to its wide applications, the problem of average consensus attracts much research interest in recent years. In this paper, we are interested in the achievement of average consensus under malicious agents in the multi-dimensional spaces. We propose a resilient distributed algorithm. Under certain network topology, the designed protocol is proved to guarantee that all benign agents exponentially reach an agreement within the convex hull of their initial states, regardless of the actions of faulty ones.

The future work involves the design of a more effective algorithm in the scenario where the network topology fails to meet the sufficient conditions. Furthermore, the theoretical analysis of the accomodation of "safe kernel" techinique to other problem settings is also a possible research direction.

REFERENCES

Cárdenas, A.A., Amin, S., and Sastry, S. (2008). Research challenges for the security of control systems. In *HotSec*.

Danzer, L., Grünbaum, B., and Klee, V. (1963). Helly's theorem and its relatives. *Proceedings of Symposia in Pure Mathematics*, 101–180.

Dolev, D., Lynch, N.A., Pinter, S.S., Stark, E.W., and Weihl, W.E. (1986). Reaching approximate agreement in the presence of faults. *Journal of the ACM (JACM)*, 33(3), 499–516.

Kar, S., Hug, G., Mohammadi, J., and Moura, J.M. (2014). Distributed state estimation and energy management in

smart grids: A consensus+innovations approach. *IEEE Journal of selected topics in signal processing*, 8(6), 1022–1038.

Kekatos, V. and Giannakis, G.B. (2013). Distributed robust power system state estimation. *IEEE Transactions on Power Systems*, 28(2), 1617–1626.

LeBlanc, H.J., Zhang, H., Koutsoukos, X., and Sundaram, S. (2013). Resilient asymptotic consensus in robust networks. *IEEE Journal on Selected Areas in Communications*, 31(4), 766–781.

Lynch, N.A. (1996). *Distributed algorithms*. Elsevier.

Mo, Y., Kim, T.H.J., Brancik, K., Dickinson, D., Lee, H., Perrig, A., and Sinopoli, B. (2012). Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1), 195–209.

Nedic, A., Ozdaglar, A., and Parrilo, P.A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4), 922–938.

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

Pasqualetti, F., Bicchi, A., and Bullo, F. (2012). Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 57(1), 90–104.

Raffard, R.L., Tomlin, C.J., and Boyd, S.P. (2004). Distributed optimization for cooperative agents: Application to formation flight. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 3, 2453–2459. IEEE.

Ren, W., Beard, R.W., and Atkins, E.M. (2007). Information consensus in multivehicle cooperative control. *IEEE Control systems magazine*, 27(2), 71–82.

Su, L. and Vaidya, N.H. (2015). Fault-tolerant distributed optimization (part iv): constrained optimization with arbitrary directed networks. *arXiv preprint arXiv:1511.01821*.

Vaidya, N.H., Tseng, L., and Liang, G. (2012). Iterative approximate byzantine consensus in arbitrary directed graphs. In *Proceedings of the 2012 ACM symposium on Principles of distributed computing*, 365–374. ACM.

Wei, E. and Ozdaglar, A. (2012). Distributed alternating direction method of multipliers. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 5445–5450. IEEE.

Xiao, L., Boyd, S., and Kim, S.J. (2007). Distributed average consensus with least-mean-square deviation. *Journal of parallel and distributed computing*, 67(1), 33–46.

Zhang, H., Fata, E., and Sundaram, S. (2015). A notion of robustness in complex networks. *IEEE Transactions on Control of Network Systems*, 2(3), 310–320.

Ziegler, G.M. (2012). *Lectures on polytopes*, volume 152. Springer Science & Business Media.