




Resilient and privacy-preserving consensus for multi-agent systems

Mingde Huang^a, Yiming Wu^{a, }, Qiuxia Huang^{b,*}

^a School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China

^b Zhejiang Electronic Information Products Inspection and Research Institute (Key Laboratory of Information Security of Zhejiang Province), Hangzhou, 310007, China

ARTICLE INFO

Keywords:

Multi-agent systems
Network robustness
Resilient consensus
Privacy-preserving

ABSTRACT

Privacy concerns and cyber-attacks are two typical threats in networked multi-agent systems (MASs), while little research has properly addressed both. To fill this gap, we investigate a privacy-preserving consensus strategy against cyber-attacks for MASs. First, a novel network eavesdropper model and a cyber-attack model that is more strategic than existing literature are proposed. Then, a homomorphic encryption-based mean subsequence reduced (HE-MSR) consensus algorithm equipped with a privacy protection strategy is designed for each normal agent. The results reveal that the privacy of states of all normal agents and the accurate consensus are guaranteed under mild network topology conditions. Furthermore, these results are extended to the case of a time-varying MAS network environment. Finally, numerical simulations and hardware experiments on Raspberry Pi are conducted to verify the theoretical results.

1. Introduction

In the past two decades, collaborative control of multi-agent systems (MASs) has received extensive research attention and has been applied in various fields, such as smart grids [1], UAV swarm formations [2], and intelligent transportation [3]. The consensus problem is a fundamental and common research issue in the distributed cooperation of MASs [4]. It requires all agents to agree on specific values through information exchange among local neighbors [5].

However, MASs are susceptible to information disclosures and cyber-attacks due to their large-scale and spatially distributed nature. Moreover, they are commonly exposed in open environments. Considering the lightweight computation ability and defense system of each single agent, it is reasonable to assume that a certain amount of individuals will be compromised by adversarial programs. These attackers, or *adversaries*, attempt to affect consensus with malice or steal information via various media. A typical group of adversaries can pretend to be a normal agent and send forged messages to others, so as to sabotage consensus. Thus, it is meaningful to design a resilient algorithm that can protect the consensus process against adversaries.

Recently, many resilient consensus control algorithms against Byzantine attacks have been carefully constructed. Byzantine attack was derived from reference [6], demonstrating that the misbehaving agent can send arbitrary messages to any neighbor during information interaction. Some researchers explored a typical Byzantine agreement problem that the processors can decide the final result despite the presence of a linear portion of bad processors [7]. A consensus algorithm based on the eigentrust model was established to optimize the Byzantine fault tolerant rate and fulfill the scalability requirement in a large-scale blockchain [8]. Inspired

* Corresponding author.

E-mail address: hqx@zdiy.org.cn (Q. Huang).

<https://doi.org/10.1016/j.ins.2024.121843>

Received 6 May 2024; Received in revised form 6 December 2024; Accepted 29 December 2024

Table 1
Comparison with the frontier methods.

	Resiliency	Privacy	Accuracy	Robustness Requirement	Additional assumption
[20]	none	strong	\times	1-robust	incomplete eavesdropping
[25]	F -total malicious	none	\checkmark	$(F + 1, F + 1)$ -robust	none
[22]	F -total Byzantine	strong	\times	$2F + 1$ -robust	none
[24]	f -fraction deception or DoS	strong	\checkmark	p -fraction robust, $p > 2f$	security at beginning
ours	F -local Byzantine	strong	\checkmark	$F + 1$ -robust (normal network)	none

by [9], LeBlanc et al. [10] studied the resilient asymptotic consensus problem in MASs under Byzantine attacks, proposed a family of Mean-Subsequence-Reduced (MSR) algorithms, and provided the necessary communication topology conditions for implementing such algorithms. The authors unveiled that if the scale of attackers was limited by F -total or F -local model [10], sufficiently high network robustness will allow agents in the MASs to resist the impact by these attackers. Afterward, a large number of MSR-based algorithms were designed to achieve resilient consensus of MASs under an adversarial environment. Others investigated the discrete-time MASs with time delays and attacks and proposed a delay-based neighbor information resilient consensus algorithm [11]. Despite the requirement of keeping robustness at all time-steps, the authors in [12] investigated the time-varying network and designed Sliding Weighted-MSR algorithm, suggesting that regulating the topology should be cumulatively robust in each time period. The conditions needed were relieved by requiring robustness in an infinite sequence of bounded step intervals for dynamic networks [13]. Owing to the high cost and risk during data transmission, an event-trigger mechanism was developed to limit the frequency of communication. Researchers proposed Event-based-MSR algorithm to relieve communication burden while resilient consensus is achieved [14]. A secure Acceptance and Broadcasting Algorithm was introduced to reach an average consensus, by which each agent took record of neighbors' initial state and retrieved all other nodes' information via redundant voting [15].

Besides the cyber-attacks, privacy leakage, as another indicator of measuring the security performance of MASs, is also a factor that must be considered when MAS technology is applied to many practical scenarios. For example, a group of agents hopes to gather at the expected location while initial position must be preserved for some specific reasons in this mission. Without proper methods, the privacy can be leaked to attackers, causing disastrous consequences. This issue involves two aspects: eavesdropping by outsiders and privacy exposure to agents within the system. Generally, the external eavesdroppers can monitor any communication channels among agents and thus have access to all messages transmitted. Meanwhile, the internal *curious* agents will infer the privacy of target agent and collude with others that have the same goal. In other research, the concept of differential privacy was introduced into consensus problem, contributing to preserving privacy with injected noise [16]. The shift added upon message was optimized [17], demonstrating that average consensus could be certainly reached. Methods called state decomposition or network augmentation were proposed in [18,19]. Observing that homomorphic encryption works well in distributed communication, researchers turned to the Paillier cryptosystem for privacy preservation [20,21].

Several researchers have proposed consensus algorithms that effectively address both attacks and privacy leakage concerns. Specifically, the Differentially Private-MSR algorithm [22] was developed, and it incorporated the injection of noise into the MSR procedure during the updating iteration. This approach can achieve mean square consensus while preserving privacy. The resilient privacy-preserving consensus problem of MASs was addressed through a state decomposition-based approach [23]. A novel consensus algorithm that enables agents to exchange and update information by employing redundant messages was established to reinforce consensus accuracy [24]. However, such a requirement may be overly stringent in practical environments that are susceptible to cyber-attacks, where the network environment must remain secure during the initial two time-steps.

Motivated by the discussion above, we propose the Homomorphic Encryption based Mean Subsequence Reduced (HE-MSR) algorithm, which allowed the MAS to reach resilient asymptotic consensus with privacy preserved. Compared to other means, homomorphic encryption is effective in privacy preservation between several untrustworthy agents, while computation is mapped from plaintext to ciphertext. Combining it with the half-weight mechanism discussed later in Section 3, we design a transmission protocol that keeps messages secret regarding both eavesdroppers and curious agents without any loss of accuracy. Furthermore, the filtering procedure by adapting half-weighted difference is innovatively improved so that the updating process was able to filter suspected data. This algorithm requires no other conditions but a specific property in topology redundancy called network robustness, as introduced in Section 2. The results verify its effectiveness theoretically, as well as its practicality by conducting simulation experiments and hardware implementations.

Our work is compared with the previous, as detailed in Table 1. The method proposed in [20] preserved privacy, whereas resiliency was not considered. The researchers of [25] investigated resilient consensus regardless of privacy. Focusing on both privacy and resiliency, one study [22] required a stricter robustness condition to reach final consensus. Nonetheless, the method proposed in [24] ensured accurate consensus results while requiring additional assumptions to perform. Comparatively, our work enables MAS to reach an accurate consensus with the common requirement of network robustness rather than any specific conditions or assumptions. The proposed method also preserves privacy against a more threatening model of privacy-seekers than that considered in others. The main contributions of this paper are described as follows.

- 1). A new eavesdropper model, termed colluding eavesdropping nodes, is introduced to monitor all messages sent or received by any individual agent. This model poses a greater threat compared to previous models such as the ones discussed in [17] and [20]. Then, a novel privacy-preserving consensus algorithm based on homomorphic encryption is devised to defend against this type of threats.

- 2). A novel resilient consensus control strategy is established. The nodes using this control strategy can ensure the privacy of information while achieving consensus. In contrast to references [22] and [24], our proposed method requires weaker implementation conditions.
- 3). The encryption process is implemented by incorporating a quantization approach into the dynamic updating process. This method eliminates errors in convergence and thus ensures the accuracy of the consensus value.

The rest of the paper is structured as follows: In Section 2, we provide an introduction to graph theory, quantization, homomorphic encryption, and the attack model. Then, the problem formulation is presented. In Section 3, the results are demonstrated, comprising the confidential transmission protocol and the complete HE-MSR algorithm. Privacy analysis is addressed in Section 4. To validate our algorithms, experiments involving simulation and hardware implementation are performed in Section 5. Finally, conclusions are drawn in Section 6.

2. Preliminaries

2.1. Graph notions and robustness

Some notions in graph theory are employed to describe the network of MAS. The network is divided into two categories: time-invariant one and time-varying one. Generally, a time-invariant network indicates that the connection link among agents in the system is fixed, while the neighborhood in time-varying network changes over time. Since it is easier to analyze a fixed topology than a varying one, time-invariant network is first considered. An undirected, time-invariant graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, n\}$ denotes the node set and \mathcal{E} represents the set of edges connected between nodes. If two agents could communicate with each other, the edge between two corresponding nodes i and j is linked, that is, $(i, j), (j, i) \in \mathcal{E}, i \neq j$. The nodes that linked to i are called the neighbors of i , denoted by $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$. In a distributed system, each node can only obtain local information from itself and its neighbors.

From the perspective of agents' behavior in the system, the nodes are divided into two parts: normal nodes \mathcal{V}_N and adversary nodes \mathcal{V}_A . The normal nodes comply with the preset rules and protocols while adversaries attempt to violate the collaboration task of the system. The detailed definition is provided later.

The topology and the redundancy of a network are essential properties to determine its security. In this study, the concept of network robustness in [10] is introduced. Since our study focuses on undirected graph, some adaptive changes are implemented.

Definition 1 (*r-reachable*). Given an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a nonempty subset $S \subset \mathcal{V}$ is said to be an r -reachable set if $\exists i \in S$ such that $|\mathcal{N}_i \setminus S| \geq r$, where $r \in \mathbb{Z}_{\geq 0}$.

Definition 2 (*r-robustness*). A nonempty, nontrivial and undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ containing $n \geq 2$ nodes is said to be r -robust with $r \in \mathbb{Z}_{\geq 0}$, if for every pair of nonempty, disjoint subsets of \mathcal{V} , at least one of the subsets is r -reachable.

Definition 3 (*(r, s)-robustness*). A nonempty, nontrivial and undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ containing $n \geq 2$ nodes is said to be (r, s) -robust with $r \in \mathbb{Z}_{\geq 0}, n \geq s \geq 1$, if for every pair of nonempty, disjoint subsets $S_1, S_2 \subset \mathcal{V}$, at least one of the following condition holds:

- 1) $|\Psi_{S_1}^r| = |S_1|;$
 - 2) $|\Psi_{S_2}^r| = |S_2|;$
 - 3) $|\Psi_{S_1}^r| + |\Psi_{S_2}^r| \geq s,$
- (1)

where

$$\Psi_{S_k}^r = \{i \in S_k | |\mathcal{N}_i \setminus S_k| \geq r\}, k \in \{1, 2\}.$$

The following concepts are applied to a time-varying network. The time-varying graph is denoted by $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$, which suggests that the edges are changing over time. The neighboring sets of a node are legitimately dynamic, denoted by $\mathcal{N}_i[t] = \{j | (i, j) \in \mathcal{E}[t]\}$. Thus we introduce the *jointly robustness* proposed in [13].

Definition 4 (*Jointly r-reachable*). For the time-varying network $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$, a nonempty subset $S \subset \mathcal{V}$ is said to be an jointly r -reachable set if there exists an infinite sequence of bounded time intervals (ISBTI) $[t_l, t_{l+1})$ such that in each time interval $[t_l, t_{l+1})$, there exists $T_j \in [t_l, t_{l+1})$ and $i_j \in S$ such that $|\mathcal{N}_{i_j}[T_j] \setminus S| \geq r$.

Definition 5 (*Jointly r-robustness*). A time-varying network $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$ is said to be jointly r -robust if, for every pair of nonempty disjoint subsets of \mathcal{V} , at least one of the subsets is jointly r -reachable.

Definition 6 (*Jointly (r, s) -robustness*). A time-varying network $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$ is said to be jointly (r, s) -robust if, for every pair of nonempty disjoint subsets $S_1, S_2 \subset \mathcal{V}$, there exists an ISBTI $[t_l, t_{l+1})$ such that in each time interval $[t_l, t_{l+1})$, at least one of the following condition holds:

$$\begin{aligned} 1) \quad & |\Psi_{S_1}^r[t_l, t_{l+1})| = |S_1|; \\ 2) \quad & |\Psi_{S_2}^r[t_l, t_{l+1})| = |S_2|; \\ 3) \quad & |\Psi_{S_1}^r[t_l, t_{l+1})| + |\Psi_{S_2}^r[t_l, t_{l+1})| \geq s, \end{aligned} \quad (2)$$

where

$$\Psi_S^r[t_l, t_{l+1}) = \{i \in S \mid \exists T_{i_j} \in [t_l, t_{l+1}) \text{ s.t. } |\mathcal{N}_i^r[T_{i_j}] \setminus S| \geq r\}.$$

The robustness property is deterministic in the class of MSR-based algorithms. Our work requires the robustness of normal network (the network only consists of normal nodes and edges among them) to execute our HE-MSR algorithm.

2.2. Probabilistic quantization

The reduction of data transmission costs is necessary for the limitations in realistic communication environment. Quantization is a commonly used approach to mitigate the workload of local computation and mutual communication. Quantization relieves the trouble in restricted data flow by transferring different types of messages to a fixed length bit stream. In this study, a probabilistic quantization method [25,26] inspired the implementation in our work. The quantization function $Q : \mathbb{R} \rightarrow \mathbb{Z}$ is expressed as

$$Q(y) = \begin{cases} \lfloor y \rfloor & \text{with probability } p(y), \\ \lceil y \rceil & \text{with probability } 1 - p(y), \end{cases} \quad (3)$$

where $p(y) = \lceil y \rceil - y$, and symbols $\lceil \cdot \rceil$, $\lfloor \cdot \rfloor$ denote the closest integer no less than and no more than the parameter, respectively. Notably, $Q(y) = y$ exists with probability 1 when y is an integer (in other words, y is at quantization level). The quantizer is applied to a more refined level with the purpose of aligned with the application scenarios, expressed as

$$f(y) = \frac{Q(L_0 y)}{L_0}, \quad (4)$$

where L_0 represents a large integer and $f : \mathbb{R} \rightarrow \mathbb{Q}_{L_0}$ and \mathbb{Q}_{L_0} indicates the set of fractional numbers with denominator L_0 .

Remark 1. Deterministic quantizers, reflecting that same inputs result in the same output, have been in many studies such as [27,28]. However, the common problem with these methods is the consensus error. Although the mistake was reduced to a small value by increasing quantization accuracy, the implementation cost rose as well. Compared to deterministic quantizers, probabilistic ones have a more significant advantage that there is a positive probability for jumping out of deadlock (See more details in [25]). This property allows for accurate consensus with quantization.

Particularly, the probabilistic quantization plays multiple roles in our work. In addition to lessening the transmission overhead, outputting integer type data is necessary to realize some other computational operation. Concretely, our privacy preservation scheme introduced later is based on the Paillier cryptosystem which provided additive homomorphic encryption for integers. Therefore, this quantization method is the key to implementing our confidential protocol as well.

2.3. Homomorphic encryption

Compared to other privacy-preserving methods, cryptographic techniques demonstrate significant advantages. Firstly, they ensure that no data distortion occurs during computation. Additionally, messages transmitted through channels are always encrypted, ensuring semantic security. In our work, these advantages are utilized to form a new privacy-preserved consensus algorithm. Asymmetric cryptosystem, or public-key cryptosystem, contains two keys: the public key k_p and the private key (or the secret key) k_s . While the public key is known by all agents for encryption process, the private one is kept secret. Hence, the decryption process can only be completed by the initiator. Concretely, if *Alice* wanted to launch encrypted communication with *Bob*, it firstly generates a pair of keys (k_p^A, k_s^B) and transmits the public key k_p^A to *Bob*. Afterward, *Bob* use this key to encrypt the message x getting $E_A(x)$. After the ciphertext is received by *Alice*, it will be decrypted as $D_A(E_A(x)) = x$. Throughout the whole process, an eavesdropper *Eve* monitoring the communication channel has access to k_p^A and $E_A(x)$ except k_s^A . Therefore, *Eve* cannot directly approach x since the decryption acquires the private key. If the cryptosystem is semantically secure (suggesting that no statistical information is contained in the ciphertext), the privacy is preserved. Some well-known public-key cryptosystems include RSA [29], ElGamal [30], and Paillier [31].

Our work is based on the Paillier cryptosystem. A brief version of the Paillier cryptosystem [32] detailed in Algorithm 1 is introduced. Although the Paillier algorithm is probabilistic asymmetric, it holds the property of additive homomorphism. Homomorphism implies that some calculating operations on ciphertexts result in corresponding operations on the plaintext, e.g.,

Algorithm 1 Paillier cryptosystem.**Key Generation:**

1. Choose two large prime numbers p, q such that $p \neq q$ and they have the same bit-length;
2. Compute $n = pq$ and $g = n + 1$;
3. Compute the Euler's totient function value $\lambda = \phi(n)$ and its modular multiplicative inverse $\mu = \lambda^{-1} \bmod n$. Here, $\phi(n) = (p-1)(q-1)$;
4. The private key is $k_s = (n, g)$, and the public key is $k_p = (\lambda, \mu)$.

Encryption:

1. Generate a random integer $r \in \mathbb{Z}_n^*$, where $\mathbb{Z}_n^* = \{x | x \in \mathbb{Z}, 0 \leq x < n, \gcd(x, n) = 1\}$;
2. The ciphertext is given by $c = E(m) = g^m r^n \bmod n^2$, where $m \in \mathbb{Z}$ being a non-negative integer less than n .

Decryption:

1. The plaintext is computed by $m = L(c^\lambda \bmod n^2) \mu \bmod n$, where $L(x) = \frac{x-1}{n}$.

$$E(m_1) \oplus E(m_2) = E(m_1 * m_2),$$

where $*$, \oplus denote certain operations upon plaintexts and ciphertexts, respectively. In the Paillier cryptosystem, there is $E(m_1)E(m_2) = E(m_1 + m_2)$. Notably, the plaintext must be integer. With a known integer k , one can calculate $E(km_1) = E(m_1)^k$. Such additive homomorphism allows agents to calculate on the data without any information leakage. This feature is practical in distributed situations like MAS. Section 3 describes how the homomorphic encryption works on consensus problem to preserve the privacy.

2.4. Attack model

Our study considers the adversarial agents that misbehave to destruct the normal operation of MAS. Some typical attack models are introduced here [10].

Definition 7 (Malicious attack model). Consider a multi-agent system. A node i is a *malicious* node if it updates the state value arbitrarily, and communicate with other nodes using this illegal value at some time-step t .

Definition 8 (Byzantine Attack Model). Consider a multi-agent system. A node i is a *Byzantine* node if it violates the preset rules and sends arbitrary messages (maybe different) to its neighbors at some time-step t .

In terms of the destructive capability, malicious node is weaker than Byzantine node. In the consensus problem discussed in this paper, adversaries send fake values to neighbors, misleading them to an unexpected state and finally deviating the whole system from reaching a safe consensus value. Byzantine nodes may send different data to neighbors in ad-hoc network while malicious nodes usually work in broadcast channel. Technically, both attack models, collectively referred to as adversaries, are considered in our work.

In a realistic environment, the resources for either normal agents or adversaries are limited. Thus, it is reasonable to assume that the scope of threat was constrained by the number of adversarial nodes. The definition given in [10] is introduced in this paper.

Definition 9 (F -total adversarial model). A MAS is under F -total adversarial model if for any normal agent i , at most F neighbors are adversaries, i.e., $|\mathcal{V}_A \cap \mathcal{N}_i| \leq F$.

Definition 10 (F -local adversarial model). A MAS is under F -local adversarial model if for any normal agent i , at most F neighbors are adversaries, i.e., $|\mathcal{V}_A \cap \mathcal{N}_i| \leq F$.

Besides the cyber-attack aiming at destroying the consensus process, leakage of privacy information is another considerable issue. Although the MAS is set in an open environment, the secret information of some agents should be preserved from both internal and external eavesdroppers (such as the position of unmanned aerial vehicle). This requires the system to operate in a confidential manner, contributing to avoiding the leakage of the key. The definition of these privacy thieves is provided in this paper.

Definition 11 (Colluding eavesdropping nodes). Node j is called a curious node if j tries to infer the privacy information of other nodes. It has access to all the local information and the messages transferred in the channel between j and all its neighbors. Furthermore, a group of nodes may proceed mutual collaboration and monitor other channels in the network so that they share all the information available to infer another node's privacy. This group is called colluding eavesdropping nodes.

Remark 2. *Eavesdropping* is a concept of outsiders. Some individuals beyond the system eavesdrop communication channels among the distributed nodes. Colluding eavesdropping nodes are an enhanced version of curious nodes by endowing with them the approach to messages in other channels that they were unable to monitor.

Remark 3. The difference between attackers and eavesdropping nodes is worth noticing. Although the attackers, like Byzantine nodes, violate the updating and transmission rules trying to undermine consensus, the eavesdropping ones may act as normal nodes while attempting to infer others' privacy. Conventionally, they are called *honest but curious nodes* without the ability to eavesdrop. In other

words, the theft of other nodes' secrets is independent of the communicating and updating behaviors taken by nodes. Thus, privacy preservation requires that the initial state value cannot be leaked to any other nodes, no matter they are adversarial or honest.

2.5. Problem formulation

Our paper focuses on the resilient asymptotic consensus problem with privacy preserving in discrete-time MAS. Specifically, each normal node $i \in \mathcal{V}_N$ hold a state value $x_i[0] \in \mathbb{R}$ at the beginning. In the following time-step, agents exchange information with neighbors and updated their own state according to preset rules. Typically, there is

$$x_i[t+1] = x_i[t] + u_i[t], i \in \mathcal{V}_N, t \in \mathbb{Z}_{\geq 0}, \quad (5)$$

where $u_i[t]$ denotes the control input to be designed by distributed algorithm. The system target is to achieve consensus on the state values of all nodes, i.e., $x_i[t] = x^*, i \in \mathcal{V}_N, t \rightarrow \infty$. Notably the consensus value x^* should lay within an expected range. Generally, the interval ranged by the initial state values is considered to be safe, i.e., $[m(0), M(0)]$ where $m(t)$ and $M(t)$ indicate the minimum and maximum values at time-step t among all normal nodes.

In the presence of attackers, the system needs to resist the misbehavior from adversary. Thus, a secure algorithm should enable a MAS to reach consensus under certain cyber attacks, or ensure *resilient asymptotic consensus*. The definition given in [10] is introduced in this paper.

Definition 12 (Resilient asymptotic consensus). A MAS is said to reach resilient asymptotic consensus if for any initial state values of normal nodes $x_i[0]$, the following conditions are met.

1. *Safety Condition:* For any normal node i , there is $x_i[t] \in [m(0), M(0)], i \in \mathcal{V}_N, t \in \mathbb{Z}_+$.
2. *Agreement Condition:* There exists x^* such that $\lim_{t \rightarrow \infty} x_i[t] = x^*, i \in \mathcal{V}_N$.

A class of algorithms, extended by the MSR algorithm [33], were designed to address the resilient asymptotic consensus problem. The idea of MSR-based algorithm is intuitive and effective: agents should reject the messages “seem unsafe” and accept only the rest data to update its own state. The method in our paper is HE-MSR, an encrypted and quantized version. The concrete content is detailed in the next section.

Additionally, the privacy issue matters. The definition of privacy preservation is presented below.

Definition 13 (Privacy preservation). In a multi-agent system aiming at consensus problem, the initial value of a normal node is regarded as its privacy information, i.e., $x_i[0], \forall i \in \mathcal{V}_N$. The privacy of node i is preserved if $x_i[0]$ cannot be estimated by any colluding eavesdropping nodes with any accuracy.

3. Main results

This study focuses on the resilient asymptotic consensus problem in MAS with privacy preserving. As stated above, all nodes communicate with neighbors at each time-step, while the adversaries do not follow the rules and sent arbitrary messages by their strategy. Thus, a resilient updating algorithm is designed. Besides, privacy preservation is achieved by encrypted communication based on homomorphism and half-weight mechanism. Afterward, the quantization process is implemented to integrate the updating dynamics and the communication scheme.

To begin with, the first-order updating dynamics is introduced in consensus problem without adversaries. Since the system operates in a distributed manner, nodes have access to local and neighboring information. Let $w_{ij}[t]$ denote the weight from node j to node i at time-step t . If $(i, j) \notin \mathcal{E}$, then $w_{ij}[t] = 0$, otherwise $w_{ij}[t] \geq \epsilon$, ϵ is a small positive constant. Thus, $x_i[t+1] = \sum_{j \in (\mathcal{N}_i \cup \{i\})} w_{ij}[t] x_j[t]$. Let $\mathbf{W}[t]$ be the weight matrix, $\mathbf{x}[t]$ be the vector of all states at each time-step, then

$$\mathbf{x}[t+1] = \mathbf{W}[t] \mathbf{x}[t]. \quad (6)$$

Numerous researchers explored the dynamic changes under the above updating function [34]. A sufficient condition to reach asymptotic consensus is that the weight matrix $\mathbf{W}[t]$ is row stochastic and that the network topology contains a spanning tree. Furthermore, the system will converge to the average value of initial states if $\mathbf{W}[t]$ is also column stochastic. This study pursues a safe consensus value rather than the exact average.

A row stochastic weight matrix indicates that the in-weight and self-weight of each node are 1, i.e., $\sum_{j \in (\mathcal{N}_i \cup \{i\})} w_{ij}[t] = 1, i \in \mathcal{V}$. Hence, these parameters should be determined only by the node itself to realize this property locally. It is feasible for an agent to allocate the weight to each input under a desired sum. Another method is to rewrite the dynamic function as

$$x_i[t+1] = x_i[t] + \sum_{j \in \mathcal{N}_i} w_{ij}[t] (x_j[t] - x_i[t]), \quad (7)$$

which indicates the coefficients of control input (recall the equation (5)) sum to zero. In this way, the property can be satisfied while these weights keep unknown for the node. In our work, this technique is implemented to hide private information. The weight $w_{ij}[t]$ is determined by two *half-weights* given by both sides, i.e.,

$$w_{ij}[t] = \alpha_{ji}[t]\beta_{ij}[t], \quad (8)$$

where the subscript ij indicates that i decides this parameter in the communication with j . These two half-weights are bounded by $\alpha_* \leq \alpha_{ji}[t] < 1, \beta_* \leq \beta_{ij}[t] < \frac{1}{|\mathcal{N}_i|}$, where α_* and β_* denote two small positive constants. Thus, $\sum_{j \in \mathcal{N}_i} w_{ij}[t] < 1$ holds. Our algorithm separates the decision of two half-weights in different times. While $\alpha_{ij}[t]$ is chosen by i 's neighbor j , it is encrypted and hidden in the difference, thereby keeping secret to i (see Steps 6-10 in Algorithm 2). Hence, the neighbors preserve their privacy when i acquired messages from them. Note that the weights are not required to be symmetric, even though the communication between i and j is mutual. This property contributes to protecting privacy in both directions.

Subsequently, compromised nodes are considered. The MSR algorithm provides a scheme of security assurance to resist the corruption from adversaries. Although an agent receives messages from several neighbors, it checks and filters out some extreme values. Concretely, node i forms a sorted list containing the values that neighbors sent to i , denoted by $x_j^i[t], j \in \mathcal{N}_i$. Then, i removes some unsafe values, determined by the difference to its self value. With a preset parameter F , if there were less than F values larger than $x_i[t]$, all of them are removed by i ; otherwise it removes the largest F terms. Similar filtering occurs to those values smaller than $x_i[t]$. The rest of these messages are adopted for i 's updating at this time-step. The filter process is the key to defending against adversarial messages. At the cost of some legal values, agents withstand those fault values that might lead to failure of safety consensus.

Algorithm 2 Homomorphic encryption based mean subsequence reduced algorithm (HE-MSR).

Input: Initial state value $x_i[0]$ for node i , a large positive integer L_0 , two small positive constants α_*, β_* , parameter F .

```

1: Generate  $(k_p^i, k_s^i)$ 
2: Send  $k_p^i$  to  $j, j \in \mathcal{N}_i$ 
3: for  $t = 0, 1, 2, \dots$  do
4:   Encrypt message  $-x_i[t]$ 
5:   Send  $E_i(-x_i[t])$  to  $j, j \in \mathcal{N}_i$ 
6:   for  $j \in \mathcal{N}_i$  do
7:     Generate  $\alpha_{ji}[t] \in [\alpha_*, 1]$ 
8:     Do homomorphic computation and send  $E_i(x_j^i[t] - x_i[t])^{L_0 \alpha_{ji}[t]}$  to  $i$ 
9:   end for
10:  Decrypt messages from neighbors
11:  Divide them by  $L_0$ , get  $\alpha_{ji}[t](x_j^i[t] - x_i[t])$ 
12:   $C_i[t] = \{\alpha_{ji}[t](x_j^i[t] - x_i[t]) | j \in \mathcal{N}_i\}$ 
13:  for  $stepcounts = 1$  to  $F$  do
14:     $min\_candidate = \min(C_i[t]),$ 
15:     $max\_candidate = \max(C_i[t])$ 
16:    if  $min\_candidate < 0$  then
17:      Remove  $min\_candidate$  from  $C_i[t]$ 
18:    end if
19:    if  $max\_candidate > 0$  then
20:      Remove  $max\_candidate$  from  $C_i[t]$ 
21:    end if
22:   $R_i[t] = C_i[t]$ 
23:  Generate  $\beta_{ij}[t] \in [\beta_*, \frac{1}{|R_i[t]|+1}]$ 
24:

```

$$x_i[t+1] = x_i[t] + f\left(\sum_{j \in R_i[t]} \alpha_{ji}[t]\beta_{ij}[t](x_j^i[t] - x_i[t])\right) \quad (9)$$

25: **end for**

The existing studies based on the MSR algorithm required the node to judge unsafe messages by the exact value of neighbors' state $x_j[t]$. In other words, node i had access to the private information of its neighbors, contradicting the possibility of privacy preservation. To handle this, we modify the criterion. Besides, those half-weighted differences with relatively high absolute value unsafe messages are considered to achieve similar effect of the MSR-algorithm. The relationship between neighboring states is reflected by the signal of the difference. Thus, the filtering process is easily fulfilled under such mechanism.

By rigorously designing the process, we establish the method HE-MSR. This was described in Algorithm 2.

The proposed algorithm is distributed for all honest nodes in the systems. Without loss of generality, it is described from the view of node i . Before the state updating began, Steps 1 and 2 ask all nodes to generate a pair of keys according to Algorithm 1 and share the public keys with neighbors. In each iteration, node i executes Steps 4-11 to obtain half-weighted differences from neighbors under encrypted communication. Then Steps 12-21 indicate the filtering procedures similar to MSR algorithm: node i judges those messages with high absolute value as unsafe ones and discard them. Finally, the rest safe messages are utilized to update i 's state by multiplying the other half-weight decided by i .

As expressed in the dynamic function (9), if the initial values are at the quantization level, the state values are also the same. The assumption is realistic since L_0 could be designed to meet the application environment.

The optional range of weight parameters is purposeful. α_{ji} is limited to be at quantization level such that $L_0\alpha_{ji}$ is an integer, thus feasible for the exponentiation in Paillier process. Although node i has no information about α_{ji} , it filters the messages and counts the number of “safe” neighbors $|R_i[t]|$. Therefore, β_{ij} is upper bounded and $\sum_{j \in R_i[t]} \alpha_{ji}[t]\beta_{ij}[t] < 1$ held.

Remark 4. The half-weight mechanism here achieves two targets at once. The first is to keep privacy secret when a neighbor is responding to node i considering that i might be an honest-but-curious node. Combined with homomorphic encryption, no privacy leakage takes place in the whole communication process, no matter facing external or internal adversaries. The second is that it conforms to the kernel of filtering process in the MSR algorithm. The criteria for determining safety or suspicion is not the value of one neighbor's state but the actual received information, that is the half-weighted difference $\alpha_{ji}[t](x_j^i[t] - x_i[t])$ for node i . Therefore, this mechanism is the key to merging both targets.

To verify the efficiency of our algorithm, we need to prove the safety condition held first.

Lemma 1. For an undirected, time-invariant network, which is modeled by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, each node in the set of normal nodes \mathcal{V}_N follows the HE-MSR algorithm with parameter F . Assume that the attackers submit to F -total or F -local adversarial model. Let $M[t], m[t]$ denote the maximum and minimum states value of normal nodes in the network at time step t , respectively. Then $M[t]$ is non-increasing and $m[t]$ is non-decreasing as t steps forward, i.e., $M[t+1] \leq M[t], m[t+1] \geq m[t]$.

Proof. The proof is similar to that in [10]. This study focuses on the dynamics without quantization process.

For node i , if it removes all messages sent by adversarial neighbors, all the nodes in $R_i[t]$ are normal nodes. Note that the dynamic equation is a convex combination of these normal states. Thus,

$$x_i[t] + \sum_{j \in R_i[t]} \alpha_{ji}[t]\beta_{ij}[t](x_j^i[t] - x_i[t]) \leq \max_{j \in \mathcal{V}_N} x_j[t] = M[t]. \quad (10)$$

If some improper messages are kept by node i , $\alpha_{ji}(x_j^i[t] - x_i[t]) \leq \alpha_{ki}(x_k^i[t] - x_i[t])$, where j denotes a adversarial neighbor and k is a normal neighbor. Therefore, the above inequality still holds. Similarly, the lower bound is the smallest state value of normal nodes, i.e.,

$$m[t] \leq x_i[t] + \sum_{j \in R_i[t]} \alpha_{ji}[t]\beta_{ij}[t](x_j^i[t] - x_i[t]) \leq M[t]. \quad (11)$$

As the state of normal nodes is always at quantization level, including $m[t]$ and $M[t]$, the state $x_i[t+1]$ being a quantization result also lies in this range. Consequently, all normal states at time-step t are bounded by $[m[t], M[t]]$, deriving that $M[t+1] \leq M[t], m[t+1] \geq m[t]$. \square

Next, the resilient asymptotic consensus is proved following the HE-MSR algorithm under sufficient or necessary conditions.

Theorem 1. For an undirected, time-invariant network, which is modeled by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, each node in the set of normal nodes \mathcal{V}_N follows HE-MSR algorithm with parameter F . Assume that the attackers submit to F -total malicious model. Then, the system will reach resilient asymptotic consensus if and only if the network topology satisfies $(F+1, F+1)$ -robustness.

Proof. We prove the necessity and the sufficiency separately.

(Necessity) If the network is not $(F+1, F+1)$ -robust, there exist two nonempty, disjoint subsets S_1, S_2 that none of conditions in (1) hold. Suppose at time step t , the nodes in S_1 and S_2 hold the minimum and maximum state value respectively, i.e., $x_{i_1}[t] = m[t], x_{i_2}[t] = M[t], \forall i_1 \in S_1, i_2 \in S_2$. Notably, $|\Psi_{S_1}^r| + |\Psi_{S_2}^r| \leq F$. That is to say, there are at most F nodes in $S_1 \cup S_2$ having at least $F+1$ neighbors outside their belonging set. Assume that these nodes are malicious ones, submitting to the F -total model.

A node i in the subset S_1 is considered in the next time step. According to the algorithm, it receives messages and executes the function

$$x_i[t+1] = x_i[t] + f\left(\sum_{j \in \mathcal{N}_i \cap S_1} c_{ij}[t]w_{ij}[t](x_j^i[t] - x_i[t]) + \sum_{j \in \mathcal{N}_i \setminus S_1} c_{ij}[t]w_{ij}[t](x_j^i[t] - x_i[t])\right), \quad (12)$$

where $c_{ij}[t]$ represents the label that takes value from $\{0, 1\}$ to discard or keep messages. Let all malicious nodes keep state value unchanged. Node i has at most F outside neighbors, i.e., $|\mathcal{N}_i \setminus S_1| \leq F$, and the neighbors in S_1 have the same value as i does. The non-zero messages received from outside neighbors are discarded due to HE-MSR algorithm. Since $c_{ij}[t] = 0, j \in \mathcal{N}_i \setminus S_1$, $x_i[t+1] = x_i[t]$ holds. The iteration in time step $t+1$ makes no alteration on the value of node i . Similarly, all nodes in S_1 and S_2 keep their values unchanged if malicious nodes keep state constant. Under such a condition, consensus cannot be reached in this system.

(Sufficiency) It is proved by contradiction. As suggested in Lemma 1, $M[t]$ and $m[t]$ denote the maximum and minimum values of normal nodes, respectively, and they are monotone as t increases. Since $M[t] \leq M[0], m[t] \geq m[0]$ and $M[t] \geq m[t]$, both of them have limits as t goes to infinity. Let

$$A_M = \lim_{t \rightarrow \infty} M[t], A_m = \lim_{t \rightarrow \infty} m[t], A_M - A_m = 2\epsilon_0.$$

If the system cannot reach consensus, $\epsilon_0 > 0$. Assume that at time step t_δ , the normal state values are bounded by

$$M[t_\delta] \leq A_M + \delta, m[t_\delta] \geq A_m - \delta, 0 < \delta < \frac{(\beta_* \alpha_*)^{|\mathcal{V}_N|}}{1 - (\beta_* \alpha_*)^{|\mathcal{V}_N|}} \epsilon_0.$$

Now we focus on two subsets including nodes with relatively high or low values, denoted by $X_M(t, \epsilon_k) = \{i \in \mathcal{V} | x_i[t] > A_M - \epsilon_k\}$, $X_m(t, \epsilon_k) = \{i \in \mathcal{V} | x_i[t] < A_m + \epsilon_k\}$, where $k = 0, 1, 2, \dots$. Note that A_M, A_m are the limit of $M[t], m[t]$, and there must be normal nodes of which state values are larger than $A_M - \epsilon_k$ and $A_m + \epsilon_k$ separately if $\epsilon_k > 0$. In other words, $\mathcal{V}_N \cap (X_M(t, \epsilon_k) \cup X_m(t, \epsilon_k)) \neq \emptyset$. Since the network is $(F + 1, F + 1)$ -robust, at least one condition in (1) holds for $X_M(t, \epsilon_k)$ and $X_m(t, \epsilon_k)$. Whatever it is, at least one normal node has $F + 1$ outside neighbors.

Without loss of generality, suppose that $i \in X_M(t_\delta, \epsilon_0)$ satisfies $i \in \mathcal{V}_N$, $|\mathcal{N}_i \setminus X_M(t_\delta, \epsilon_0)| \geq F + 1$. This node receives at least $F + 1$ messages from neighbors that the state value is no larger than $A_M - \epsilon_0$. Thus at least one message that is negative and sent by node $j \in \mathcal{N}_i \setminus X_M(t_\delta, \epsilon_0)$ remains in $R_i[t_\delta]$. Thus

$$\beta_{ij} \alpha_{ji} (x_j^i[t_\delta] - x_i[t_\delta]) \leq \beta_{ij} \alpha_{ji} (A_M - \epsilon_0 - x_i[t_\delta]) \leq \beta_* \alpha_* (A_M - \epsilon_0 - x_i[t_\delta]).$$

Some of neighbors in $X_M(t_\delta, \epsilon_0)$ may be malicious. While node i labels and removes at most F largest differences, there is a condition that all information from malicious nodes is abandoned (since the number of them is not larger than F). If that happens, it can be derived that

$$\beta_{ij} \alpha_{ji} (x_j^i[t_\delta] - x_i[t_\delta]) \leq \beta_{ij} \alpha_{ji} (A_M + \delta - x_i[t_\delta]), j \in \mathcal{V}_N \cap (\mathcal{N}_i \cap X_M(t_\delta, \epsilon_0)).$$

From another perspective, if some messages from malicious neighbors are remained, some messages from normal ones must be considered “unsafe” and removed. Thus,

$$\beta_{ik} \alpha_{ki} (x_k^i[t_\delta] - x_i[t_\delta]) \leq \beta_{ij} \alpha_{ji} (x_j^i[t_\delta] - x_i[t_\delta]) \leq \beta_{ij} \alpha_{ji} (A_M + \delta - x_i[t_\delta]),$$

where k refer to a malicious node of which the message is remained; j represents the normal node that sends the largest difference to i . Let $r_M = |\mathcal{R}_i[t_\delta] \cap X_M(t_\delta, \epsilon_0)|$, then

$$\begin{aligned} & x_i[t_\delta] + \sum_{j \in \mathcal{R}_i[t_\delta]} \beta_{ij} [t_\delta] \alpha_{ji} [t_\delta] (x_j^i[t_\delta] - x_i[t_\delta]) \\ & \leq x_i[t_\delta] + \beta_* \alpha_* (A_M - \epsilon_0 - x_i[t_\delta]) + \sum_{j \in \mathcal{R}_i[t_\delta] \cap X_M(t_\delta, \epsilon_0)} \beta_{ij} [t_\delta] \alpha_{ji} [t_\delta] (x_j^i[t_\delta] - x_i[t_\delta]) \\ & \leq x_i[t_\delta] + \beta_* \alpha_* (A_M - \epsilon_0 - x_i[t_\delta]) + r_M \max_{j \in \mathcal{V}_N \cap (\mathcal{N}_i \cap X_M(t_\delta, \epsilon_0))} \beta_{ij} \alpha_{ji} (A_M + \delta - x_i[t_\delta]) \\ & = (1 - \beta_* \alpha_* - r_M \beta_{ij} \alpha_{ji}) x_i[t_\delta] + \beta_* \alpha_* (A_M - \epsilon_0) + r_M \beta_{ij} \alpha_{ji} (A_M + \delta). \end{aligned} \quad (13)$$

The last expression is a convex combination of three positive variables. $x_i[t_\delta] \leq A_M + \delta$, and the maximum value of r.h.s. of equation (13) is attained when each variable takes its own maximum, that is

$$(1 - \beta_* \alpha_*) (A_M + \delta) + \beta_* \alpha_* (A_M - \epsilon_0) = A_M + (1 - \beta_* \alpha_*) \delta - \beta_* \alpha_* \epsilon_0. \quad (14)$$

Then, the l.h.s. of equation (13) is no larger than equation (14), and the upper bound of $x_i[t_\delta + 1]$ is

$$x_i[t_\delta + 1] = x_i[t_\delta] + f\left(\sum_{j \in \mathcal{R}_i[t_\delta]} \beta_{ij} [t_\delta] \alpha_{ji} [t_\delta] (x_j^i[t_\delta] - x_i[t_\delta])\right) \leq A_M + (1 - \beta_* \alpha_*) \delta - \beta_* \alpha_* \epsilon_0, \quad (15)$$

which holds with positive possibility. Similarly, if there is an honest node $i \in X_m(t_\delta, \epsilon_0)$ that has $F + 1$ outside neighbors, it can be derived that

$$x_i[t_\delta + 1] \geq A_m - (1 - \beta_* \alpha_*) \delta + \beta_* \alpha_* \epsilon_0. \quad (16)$$

These bounds hold not only for normal node $i \in X_M(t_\delta, \epsilon_0) \cup X_m(t_\delta, \epsilon_0)$, but also for other normal nodes $i_1 \in \mathcal{V}_N \setminus (X_M(t_\delta, \epsilon_0) \cup X_m(t_\delta, \epsilon_0))$ by removing the equal sign. This is because the coefficient of node i_1 's own state value is positive, resulting in smaller coefficient for $A_M + \delta$ or $A_m - \delta$ and thus a smaller convex hull.

Now consider $\epsilon_k = \beta_* \alpha_* \epsilon_{k-1} - (1 - \beta_* \alpha_*) \delta$, $k = 1, 2, \dots, |\mathcal{V}_N| - 1$. There is at least one normal node i in $X_M(t_\delta, \epsilon_0) \cup X_m(t_\delta, \epsilon_0)$ and its bound is given by equation (14) and (16). Thus, $A_m + \epsilon_1 \leq x_i[t_\delta + 1] \leq A_M - \epsilon_1$ which holds with positive possibility. This indicates that $i \notin X_M(t_\delta + 1, \epsilon_1) \cup X_m(t_\delta + 1, \epsilon_1)$. Although the normal nodes not in $X_M(t_\delta, \epsilon_0) \cup X_m(t_\delta, \epsilon_0)$ are also limited by the bounds, they will not enter $X_M(t_\delta, \epsilon_1) \cup X_m(t_\delta, \epsilon_1)$. Thus, it is expressed as

$$|\mathcal{V}_N \cap (X_M(t_\delta + 1, \epsilon_1) \cup X_m(t_\delta + 1, \epsilon_1))| \leq |\mathcal{V}_N \cap (X_M(t_\delta, \epsilon_0) \cup X_m(t_\delta, \epsilon_0))| - 1, \quad (17)$$

which holds with positive possibility. By recursion, this equation holds for $\epsilon_k, k = 1, 2, \dots, K$ where $0 < K \leq |\mathcal{V}_N|$, $K \in \mathbb{Z}$. At time step $t_\delta + K$, it is possible that no normal node exists in $X_M(t_\delta + K, \epsilon_K) \cup X_m(t_\delta + K, \epsilon_K)$. Meanwhile, since $\delta < \frac{(\beta_* \alpha_*)^{|\mathcal{V}_N|}}{1 - (\beta_* \alpha_*)^{|\mathcal{V}_N|}} \epsilon_0$, it can be obtained that

$$\begin{aligned}
\epsilon_K &= \beta_* \alpha_* \epsilon_{K-1} - (1 - \beta_* \alpha_*) \delta \\
&= (\beta_* \alpha_*)^2 \epsilon_{K-2} - (\beta_* \alpha_* + 1)(1 - \beta_* \alpha_*) \delta \\
&= \dots \\
&= (\beta_* \alpha_*)^K \epsilon_0 - (1 - (\beta_* \alpha_*)^K) \delta \\
&\geq (\beta_* \alpha_*)^{|\mathcal{V}_N|} \epsilon_0 - (1 - (\beta_* \alpha_*)^{|\mathcal{V}_N|}) \delta \\
&> 0.
\end{aligned} \tag{18}$$

Thus no normal node reaches the limit of maximum or minimum value A_M, A_m , contradicting the assumption that $A_M - A_m = 2\epsilon_0 > 0$. Hence $A_M = A_m$. \square

The above theorem provides necessary and sufficient conditions for HE-MSR applying on MAS including malicious nodes. The following theorem is proposed to withstand more destructive Byzantine nodes.

Theorem 2. *For an undirected, time-invariant network, which is modeled by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, each node in the set of normal nodes \mathcal{V}_N follows HE-MSR algorithm with parameter F . Assume that the attackers submit to F -total or F -local Byzantine model. Then, the system will reach resilient asymptotic consensus if and only if the normal network topology satisfies $(F + 1)$ -robustness.*

Proof. (Necessity) Suppose that the normal network is not $(F + 1)$ -robust. Then, there exist two nonempty, disjoint subsets $S_1, S_2 \subset \mathcal{V}_N$, such that they are not r -reachable. At time step t , suppose that the nodes in S_1 and S_2 hold the minimum and maximum state values respectively, i.e., $x_{i_1}[t] = m[t], x_{i_2}[t] = M[t], \forall i_1 \in S_1, i_2 \in S_2$. A node i in the subset S_1 is considered in the next time step. Let all Byzantine nodes hold state value $m[t]$ towards node i , or $x_j^i[t] = m[t], \forall j \in \mathcal{V}_A \cap \mathcal{N}_i$. Thus $x_j^i[t] - x_i[t] = 0$ holds for all Byzantine neighbors j . Next, node i does not receive any negative message since itself keeps the minimum value among all normal nodes. Besides, the positive messages are from normal nodes in $\mathcal{N}_i \setminus S_1$, of which the amount is at most F . Hence, node i discards all these messages according to the algorithm. Particularly, the iteration in time step $t + 1$ makes no alteration on the value of node i . Similarly, all nodes in S_1 and S_2 keep their values unchanged if Byzantine nodes proceed proper attack. Under such a condition, consensus cannot be reached in this system.

(Sufficiency) Using the notations and conclusion of Lemma 1, A_M, A_m are the limitations of $M[t], m[t]$ as t goes to infinity. To prove by contradiction, it is assumed that $A_M > A_m$. Suppose that the limitations are reached at different values at a finite time step t_σ , i.e. $M[t_\sigma] = A_M > m[t_\sigma] = A_m$. Let $X_M[t]$ and $X_m[t]$ denote the set of normal nodes holding the state value of A_M and A_m at time step t , respectively. We focus on the nodes in $X_M[t]$ and $X_m[t]$ when $t > t_\sigma$.

Following the robustness of the network, $X_M[t]$ and $X_m[t]$ are nonempty, disjoint subsets of \mathcal{V} , and thus at least one node has at least $F + 1$ neighbors outside its belonging set ($X_M[t]$ or $X_m[t]$). Without loss of generality, it is assumed that node $i \in X_M[t]$ has at least $F + 1$ neighbors in $\mathcal{V}_N \setminus X_M[t]$, or

$$|\mathcal{N}_i \cap (\mathcal{V}_N \setminus X_M[t])| \geq F + 1.$$

Thus at least $F + 1$ negative messages are received by i at t . Since $x_i[t + 1] = x_i[t] + f(\sum_{k \in R_i[t]} \beta_{ik} \alpha_{ki} (x_k[t] - x_i[t]))$, it can be derived that

$$\sum_{k \in R_i[t]} \beta_{ik} \alpha_{ki} (x_k[t] - x_i[t]) \leq \sum_{k \in R_i[t]} \beta_{ik} \alpha_{ki} \left(-\frac{1}{Q}\right) \leq \beta_* \alpha_* \left(-\frac{1}{Q}\right).$$

The quantified value of the right side is zero with probability $1 - \beta_* \alpha_*$. Since the left side is not larger than the right side, the possibility for its quantified value being zero is less than $1 - \beta_* \alpha_*$ (reaches zero if the left side is less than or equal to $-\frac{1}{Q}$), expressed as

$$\begin{aligned}
\mathbb{P}\{x_i[t + 1] - x_i[t] = 0\} &\leq 1 - \beta_* \alpha_*, \\
\mathbb{P}\{x_i[t + 1] \leq A_M - \frac{1}{Q}\} &\geq \beta_* \alpha_*.
\end{aligned} \tag{19}$$

In other words, the normal node i in $X_M[t]$ no longer keeps the maximum value and is removed from $X_M[t + 1]$ with positive probability. Similarly, if a normal node $j \in X_m[t]$ has at least $F + 1$ neighbors outside $X_m[t]$, it will be removed from $X_m[t + 1]$ with positive probability.

The following part focuses on normal nodes not inside $X_M[t]$ or $X_m[t]$. Suppose node $h \in \mathcal{V}_N \setminus X_M[t]$. The updating function without quantization is a convex combination of $x_k[t], k \in (\{h\} \cup R_h[t])$ with positive coefficients. Hence, equation (19) also holds for h as well. This suggests that with positive probability, a normal node in $\mathcal{V}_N \setminus X_M[t]$ will not enter $X_M[t + 1]$. Similarly, this deduction holds for $X_m[t]$.

With positive probability, the number of nodes in $X_M[t]$ or $X_m[t]$ decreases as t increases, expressed as

$$\begin{aligned}
\mathbb{P}\{|X_M[t + 1]| < |X_M[t]|\} &> 0, \\
\mathbb{P}\{|X_m[t + 1]| < |X_m[t]|\} &> 0.
\end{aligned} \tag{20}$$

Algorithm 3 Dynamically adapted HE-MSR.

Input: Initial state value $x_i[0]$ and the list of neighbors $\mathcal{N}_i[t]$ for node i , a large positive integer L_0 , two small positive constants α_*, β_* , parameter F .

```

1: Generate  $(k_p^l, k_s^l)$ 
2: for  $t = 0, 1, 2, \dots$  do
3:   Encrypt message  $-x_i[t]$ 
4:   Send  $k_p^l$  and  $E_i(-x_i[t])$  to  $j, j \in \mathcal{N}_i[t]$ 
5:   for  $j \in \mathcal{N}_i$  do
6:     Generate  $\alpha_{ji}[t] \in [\alpha_*, 1]$ 
7:     Do homomorphic computation and send  $E_i(x_j^l[t] - x_i[t])^{L_0 \alpha_{ji}[t]}$  to  $i$ 
8:   end for
9:   Decrypt messages from neighbors
10:  Divide them by  $L_0$ , get  $\alpha_{ji}[t](x_j^l[t] - x_i[t])$ 
11:   $C_i[t] = \{\alpha_{ji}[t](x_j^l[t] - x_i[t]) | j \in \mathcal{N}_i\}$ 
12:  for  $stepcounts = 1$  to  $F$  do
13:     $min\_candidate = \min(C_i[t])$ ,
     $max\_candidate = \max(C_i[t])$ 
14:    if  $min\_candidate < 0$  then
15:      Remove  $min\_candidate$  from  $C_i[t]$ 
16:    end if
17:    if  $max\_candidate > 0$  then
18:      Remove  $max\_candidate$  from  $C_i[t]$ 
19:    end if
20:  end for
21:   $R_i[t] = C_i[t]$ 
22:  Generate  $\beta_{ij}[t] \in [\beta_*, \frac{1}{|R_i[t]|+1}]$ 
23:   $x_i[t+1] = x_i[t] + f(\sum_{j \in R_i[t]} \alpha_{ji}[t] \beta_{ij}[t] (x_j^l[t] - x_i[t]))$ 
24: end for

```

Therefore, it is possible for $t \geq t_\sigma + |\mathcal{V}_N|$, it is possible that $|X_M[t]| = 0$ or $|X_m[t]| = 0$. This possibility contradicts with the assumption that the limit for the maximum and minimum values of normal nodes are A_M and A_m , respectively, where $A_M > A_m$. Thus, the sufficiency is proved. \square

In real application for MAS, the connection between agents generally changes over time under a dynamic and open environment. Considering the formation of unmanned aerial vehicles, the movement of agents leads to changes in topology, which might affect the communication channel. Therefore, the system has to build a new logical network adapting to a real situation that is time-varying. Thus, the theorem is extended to situations that network topology is time-varying. Algorithm 3 is proposed. Theorem 3 and Theorem 4 are given below. For sake of concision, the proof is given in appendix.

Theorem 3. For an undirected, time-varying network, which is modeled by $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$, each node in the set of normal nodes \mathcal{V}_N follows dynamically adapted HE-MSR Algorithm 3 with parameter F . Assume that the attackers submit to F -total malicious model. Then, the system will reach resilient asymptotic consensus if and only if the network topology satisfies jointly $(F + 1, F + 1)$ -robustness.

Theorem 4. For an undirected, time-varying network, which is modeled by $\mathcal{G}[t] = (\mathcal{V}, \mathcal{E}[t])$, each node in the set of normal nodes \mathcal{V}_N follows dynamically adapted HE-MSR Algorithm 3 with parameter F . Assume that the attackers submit to F -total or F -local Byzantine model. Then, the system will reach resilient asymptotic consensus if and only if the normal network topology satisfies jointly $(F + 1)$ -robustness.

Remark 5. Although there are more communicating actions in our algorithm than a normal consensus protocol, they do not boost the complexity of computation process since the cost is increased linearly. Besides, the systems keep favorable scalability. An increase in the number of neighbors will accumulate for higher robustness of the whole system, defending against more threatening attacks. The time cost in each iteration and a comparative experiment regarding a method without privacy preserving functions are exhibited and briefly discussed in Section 5.

4. Privacy analysis

In this section, the privacy preservation of the proposed algorithm is validated. The risk of leakage can be analyzed from two parts: the external eavesdropper and the internal betrayal. The risk from an eavesdropper that monitors communication channels is eliminated with the semantic security provided by Paillier cryptosystem [35]. To address the issue with eavesdropping nodes, our study introduced the framework of indistinguishability proof referred to [20]. The privacy is technically protected as long as the eavesdropping nodes cannot distinguish between two different initial states.

Theorem 5. Consider a multi-agent system aiming at a consensus problem. If the normal nodes exchange messages with neighbors following Algorithm 2 or 3, the privacy information of all normal nodes is preserved against any curious node or colluding eavesdropping nodes (with finite number).

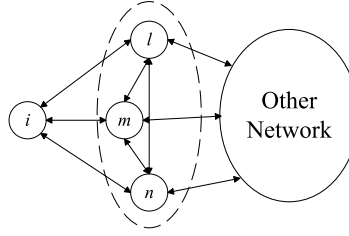


Fig. 1. Part of a network. Node i is connected to a group of colluding eavesdropping nodes.

Proof. Without loss of generality, this study focuses on the condition of one normal node i . If i could keep its privacy not leaked in the presence of colluding eavesdropping nodes, the whole system is also privacy preserved.

A network topology is illustrated in Fig. 1 to keep consistency with the robustness demand for applying HE-MSR Algorithm. Node i has three neighbors, allowing it to run the HE-MSR Algorithm with $F = 1$ (see [10] for the details of the relationship between the minimum degree of a node and the r -robustness of a graph). Assume that all neighboring nodes of i , namely l, m, n , are eavesdropping nodes and form a group to infer i 's privacy in collusion. This assumption submits to the definition of colluding eavesdropping nodes, and the group possesses the largest scope to eavesdrop the target node since all messages sent from and to i are monitored. Following the convention of cryptography analysis, node i is called *Alice* and the group is called *Eve*. Furthermore, the variation of possible privacy information, namely the initial value of *Alice* $x_A[0]$, is not distinguishable to *Eve*. Hence, the eavesdropping ones cannot estimate the initial value with any accuracy.

$I_E[t]$ denotes the information obtained by Eve at time-step t . The Paillier cryptosystem provides semantic security, revealing that *Eve* learns nothing from messages encrypted by *Alice* or other nodes. Therefore, $I_E[t]$ contains the local information and the messages decrypted by these three nodes, i.e.,

$$I_E[t] = \{x_e[t], \alpha_{eA}[t], \beta_{eA}[t], \alpha_{Ae}[t](x_A[t] - x_e[t]) \mid \forall e \in \{l, m, n\}\}. \quad (21)$$

As time-step goes forward, the accessible information collected by *Eve* is $I_E = \bigcup_{k=0}^{\infty} I_E[k]$.

To verify that *Eve* is unable to distinguish the privacy value from a variation of possibilities, our analysis demonstrates that different initial values result in the same information set for *Eve*, i.e., $x'_A[0] \neq x_A[0], I'_E = I_E$. All information approachable for *Eve* is I_E , and I_E is the same outcome under the premise of different privacy information $x_A[0]$. Thus, *Eve* cannot infer or estimate $x_A[0]$ with any accuracy.

To prove $I'_E = I_E$, it suffices to prove $I'_E[k] = I_E[k], k = 0, 1, \dots$, where all parameters decided only by *Eve* remain the same, involving the initial value of *Eve* and the half-weight $\alpha_{eA}[k], \beta_{eA}[k]$. At time-step $k = 0$, $I_E[0] = \{x_e[0], \alpha_{eA}[0], \beta_{eA}[0], c_{Ae}[0] \mid \forall e \in \{l, m, n\}\}$. The value of candidate messages $c_{Ae}[0]$ is affected by *Alice* as the equation $c_{Ae}[0] = \alpha_{Ae}[0](x_A[0] - x_e[0])$ holds. Thus

$$\alpha'_{Ae}[0] = \frac{x_A[0] - x_e[0]}{x'_A[0] - x_e[0]} \cdot \alpha_{Ae}[0], \forall e \in \{l, m, n\}. \quad (22)$$

Hence, $c'_{Ae}[0] = c_{Ae}[0], I'_E[0] = I_E[0]$. Furthermore, the dynamic updating of *Eve* remains the same since *Eve* has got same candidate information $C_e[t]$. In other words, $x'_e[1] = x_e[1]$. The next time-step emphasizes the relationship between $I'_E[1]$ and $I_E[1]$. The state values under two conditions are equal, and the difference between $c'_{Ae}[1]$ and $c_{Ae}[1]$ is the only way to distinguish $I'_E[1]$ from $I_E[1]$. Since $c_{Ae}[1] = \alpha_{Ae}[1](x_A[1] - x_e[1])$, the key is ensuring that the state value of *Alice* is the same under different conditions. The dynamic function of *Alice* at $k = 0$ is analyzed. According to HE-MSR Algorithm with $F = 1$, *Alice* collects three candidate messages from l, m, n and abandons some of them (at least one and at most two). Let σ denote the count of messages remaining, i.e., $\sigma = |R_A[0]|$. Then

$$\beta'_{Ae}[0] = \frac{\frac{x_A[0] - x'_A[0]}{\sigma \alpha_{eA}[0]} + \beta_{Ae}(x_e[0] - x_A[0])}{x_e[0] - x'_A[0]}, \forall e \in R_A[0]. \quad (23)$$

With such a choice of half-weight decided by *Alice*, the updating functions are calculated, yielding $x'_A[1] = x_A[1]$. Then, $c'_{Ae}[1] = c_{Ae}[1]$ if $\alpha'_{Ae}[1] = \alpha_{Ae}[1]$. Now $I'_E[k] = I_E[k]$ for $k = 0$ and 1 . For the rest of parameters, just let

$$\alpha'_{Ae}[k] = \alpha_{Ae}[k], \beta'_{Ae}[k] = \beta_{Ae}[k], \forall e \in \{l, m, n\}, \quad (24)$$

resulting in $I'_E[k] = I_E[k], k = 2, 3, \dots$. Therefore, with the parameters following equations (22), (23) and (24), it is concluded that a variation of different private values of *Alice* could provoke same accessible information for *Eve*, indicating that *Eve* cannot infer or estimate the privacy information with any accuracy even if *Eve* eavesdrops all communication channel connected to *Alice*. \square

Our algorithm protects the privacy even if all neighbors of the target node are colluding. This is theoretically achieved by the asymmetric half-weight mechanism that eavesdropping nodes have no idea of the change in target's state. Compared to methods in [20] where at least one legitimate neighbor is required, HE-MSR can preserve privacy under more severe conditions.

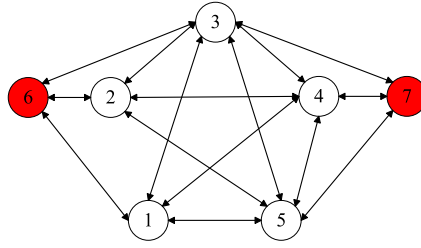


Fig. 2. The network topology of a MAS with 5 normal agents (depicted by white nodes) and 2 Byzantine agents (depicted by red nodes).

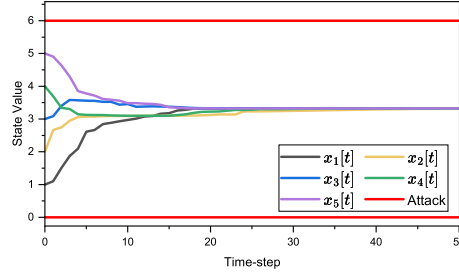


Fig. 3. The evolution of normal nodes' state value. The red lines symbolize the value that Byzantine nodes pretend to be.

Remark 6. It is subtle to claim that eavesdropping nodes cannot infer the private information *with any accuracy*. To keep the states not deviating to unexpected result, the algorithm ensures that all states lie within the safety interval. This seems to provide the eavesdropping nodes a range to make estimation. However, the exact interval is unknown to colluding nodes because the state of target is not obtained. The estimation range is meaningless, if some eavesdropping nodes begin with intentionally small or large value, such as 0 and 2π in an angle-consensus process. Therefore, the privacy preservation still works.

5. Simulation and hardware experiments

In this section, we verify our method by both simulation and hardware experiments. Firstly, the effectiveness and advantages of HE-MSR algorithm is shown in numerical simulations with comparison to some existed approaches. Then a MAS based on several Raspberry Pi and cell phones is constructed. The proposed algorithm is implemented on such devices to show its practicality.

5.1. Simulation experiments

A MAS contains 7 agents where the connection network topology is illustrated in Fig. 2. Suppose that the system was under 2-local Byzantine attack, indicated by red background. As verified, the normal network topology was 3-robust. The initial values for the five normal nodes were set to be $x_i[0] = i, i = 1, 2, 3, 4, 5$ and the large integers were chosen as $L_0 = 10^5$, so as to perform HE-MSR algorithm on the system with parameter $F = 2$. The lower bound of half-weights is set to be $\alpha_* = 0.01, \beta_* = 0.01$. The bit-length of keys generated and used by each node was 128-bit long.

The convergence process, exhibited in Fig. 3, suggests that the five normal agents reached consensus in finite time-steps with state values staying within safety interval $[1, 5]$. The messages transmitted between neighbors are displayed in Fig. 4. The messages encrypted by the public key of Agent 1 were only drawn for concision's sake. Since the semantic security was provided by Paillier encryption, no statistical information or distinguishable characteristics was accessible in this plot. Therefore, outside eavesdroppers cannot infer any information by monitoring transmission channel.

A comparative experiment was conducted to convince that the algorithm obtained privacy-preservation against colluding eavesdropping nodes, illustrated in Fig. 5. Suppose that a group of eavesdropping nodes are cooperating on inferring the privacy of their common neighbor Agent 1. If we modified the HE-MSR algorithm to a version with symmetric half-weight, i.e., $\alpha_{ij} = \beta_{ij}$, eavesdropping nodes have the knowledge of all updating messages accepted by Agent 1. Thus the state variation $\Delta x_1[t] = x_1[t+1] - x_1[t]$ was obtained. As the consensus was asymptotically reached, the initial state was deduced by removing all variation, described as

$$\begin{aligned}
 \lim_{t \rightarrow \infty} z_1[t+1] &= \lim_{t \rightarrow \infty} x_j[t+1] - \sum_{k=0}^t \Delta x_1[k] \\
 &= \lim_{t \rightarrow \infty} x_1[t+1] - \sum_{k=0}^t \Delta x_1[k] \\
 &= x_1[0],
 \end{aligned} \tag{25}$$

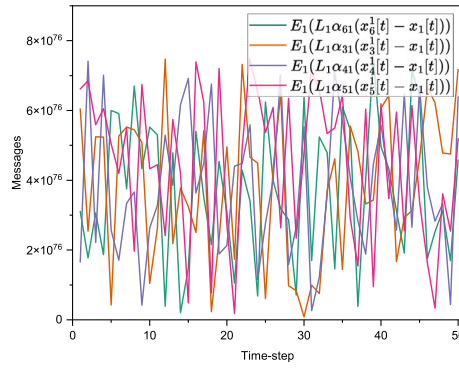


Fig. 4. The messages transmitted between Agent 1 and its neighbors. The messages in the channel are a mess and no statistical information is revealed due to semantic security.

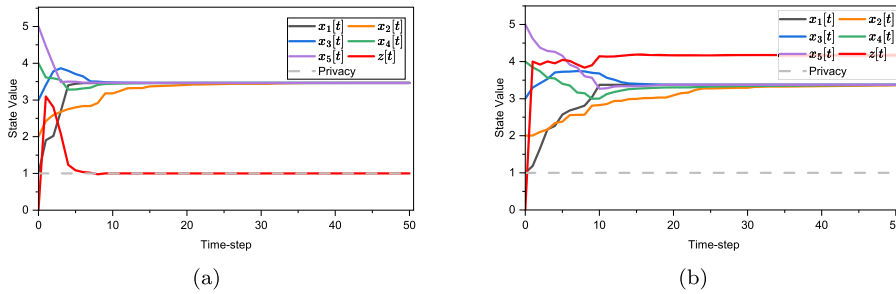


Fig. 5. The result of executing HE-MSR algorithm:(a) without and (b) with asymmetric half-weight.

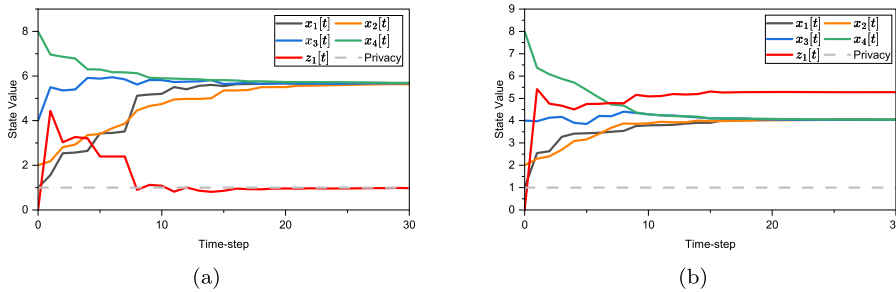


Fig. 6. A comparative experiment regarding method in [20]. (a) Shows that agent 1's initial state is exposed when all neighbors of an agent are colluding eavesdropping nodes. Under HE-MSR algorithm, the privacy keeps secret in (b).

where j denotes a curious neighbor. Therefore, the privacy was exposed as the system converged. On the contrary, the complete HE-MSR algorithm preserved the privacy well as is depicted.

The next section explains the advantages of HE-MSR algorithm over some existing MAS consensus approaches. Concretely, it was compared with a privacy-preserving approach [20] through simulation experiments. The analysis was conducted to evaluate that our resistance against cyber-attack was stronger than the methods in [25]. Furthermore, the prerequisites to carry out our methods were less required regarding the MSR-class algorithms in [24]. This was reflected in the robustness condition for network topology.

Firstly, the network topology [20], where four nodes were connected in the form of a ring, was analyzed. The authors proposed a method called Confidential Average Consensus (CAC), which ensured privacy preservation through a confidential interaction protocol implemented with homomorphic encryption. If a node had at least one legitimate neighbor that will not pry into its initial state, then the privacy is protected. If the scope of curious nodes were enhanced so that all messages from the target were monitored, eavesdroppers can easily infer the privacy with equation (25), as demonstrated in Fig. 6(a). On the contrary, our method protected privacy well even if all neighbors were eavesdropping nodes, as depicted in Fig. 6(b).

The randomized quantized algorithm (RQ-MSR) allows the system to reach resilient consensus under synchronous networks [25]. This requires the network topology to be $(F + 1, F + 1)$ -robust to defend against F -local malicious model. In contrast, our method requires the normal network topology being $(F + 1)$ -robust and is resistant to F -local Byzantine model with privacy preserved. Although the robustness condition is more demanded, the HE-MSR algorithm defends against stronger cyber-attack and conceal

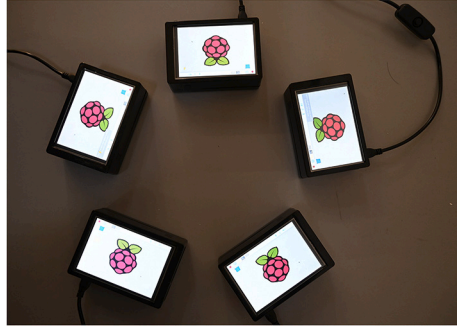


Fig. 7. 5 Raspberry Pis work as normal agents defending against 2 Byzantine agents (PCs).

privacy information. Thus our algorithm is more preferable in real application environment where the MAS needs higher security insurance.

The works mentioned in the above two papers have limited protection ability against the combination of Byzantine attack and privacy stealing agents. Next, another work with higher security [24] is presented. The main result of that article is privacy-preserving adaptive resilient consensus algorithm (PPARCA). The key to hide secret information is decomposing the real state into two sub-states. With one sub-state x_i^a keeping private, the eavesdropping nodes can infer the real states.

With resiliency being considered, PPARCA requires significant prerequisites to operate correctly. This algorithm rules that all agents exchange their states at the beginning. In the following process, the agents distinguished messages from corrupted ones by voting, with the help of this basic information of initial sub-states from neighbors. Thus the assumption that the environment is completely secure at the first two time-steps is essential to the whole algorithm. However, this condition is commonly unrealistic in a dynamic and open MAS environment. Although additional agents join in the system and new consensus process is launched, this “secure beginning” premise may fail. Consequently, it limits a strict condition to implement the algorithm.

Besides, the robustness requirement for PPARCA is not loose. The cyber attack considered is the combination of deception and DoS attack, allowing to corrupt data package or simply discard it. This can be regarded as a specific type of Byzantine attack since Byzantine node can send data exceeding the processing range of normal agent to play the role of DoS attack. Thus, an algorithm that defends against Byzantine attack also works well under deception and DoS attack. Regarding the robustness requirement, PPARCA takes effect in p -fraction robust network with $p > 2f$ under f -fraction local cyber attacks. In the HE-MSR algorithm, the normal network should be $(F + 1)$ -robust under the F -local model. As revealed by converting the fraction to number, the network is at least $(2F + 1)$ -robustness to operate PPARCA algorithm under F -local model. This is more demanding compared to using HE-MSR. For example, the fraction of the network illustrated in Fig. 2 is $f = \frac{2}{5}$. It is verified that the network is not $\frac{4}{5}$ -fraction robust, indicating the topology is not sufficient to run PPARCA. HE-MSR is applicable on this network. Consequently, our method has remarkable advantages regarding operating conditions.

5.2. Hardware experiments

This section demonstrates the effectiveness and practicality of our algorithm through hardware experiments. Raspberry Pi board is an ARM based microcomputer motherboard and is convenient for programming use. In this experiment, a MAS was built with several Raspberry Pi boards and PC devices, which is shown in Fig. 7.

The boards used in this study are Raspberry Pi 4 Model B with 1.5GHz, 4-core BroadcomBCM2711B0 (Cortex A-72) CPU, 4GB RAM and 802.11ac (2.4/5GHz) wireless connection. A single board can perform local computation and launch connections to others, thus being qualified as an agent. For consistency, the network topology in Fig. 2 is applied. To form an autonomous and distributed network communication among these devices, the wireless connection mode of these boards was set to be ad-hoc mode. After a network segment was preset, all boards entered the same cell and had access to each other. Communication was achieved by sending sockets under UDP protocol between any two boards. Each board initially stored the IP address of logical neighbors representing the communication links.

The algorithm was coded in Python. The “phe” package provided a light implementation of the Paillier cryptosystem. In the whole system, 2 PC devices played the role of two Byzantine nodes, while 5 boards acted as normal agents. They pretended to be two agents and aimed at the deviation of the system’s state value from safety interval. Concretely, the Byzantine nodes received the requests from normal nodes and replied with a corrupted message. Additionally, they asked neighbors for state information with a pair of encryption keys and tried to deduce the privacy of targets. The 5 normal agents conducted HE-MSR to reach asymptotic consensus. To achieve messages from different agents, each board launched several threads in parallel. When an agent i received a request from agent j , it homomorphically computed the response message with the public key of j (i knows the sender’s address and binds the public key to it). After j collected all responses in one iteration, it filtered and discarded unsafe messages to update itself. As the time-step increased, the states of all normal agents reached the same. There existed the situation that agents were not synchronous on computation, probably caused by performance gaps and inconsistent thread processing sequences between different devices. Thus, a synchronization-trigger should be set for each agent.

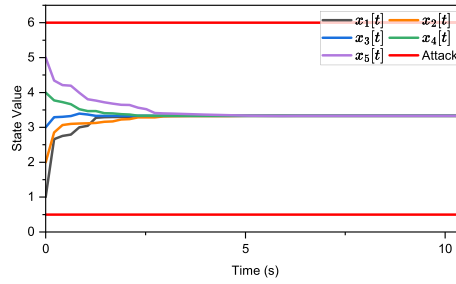


Fig. 8. Normal Raspberry Pis reach consensus in several seconds.

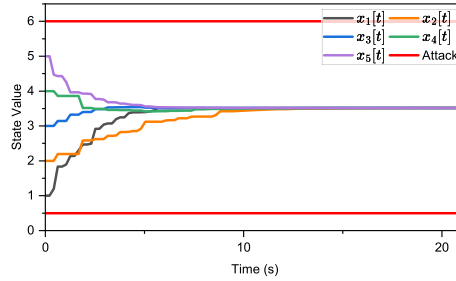


Fig. 9. Pis under time-varying network reach consensus.

The first experiment results are presented in Fig. 8. The initial states for normal agents were integers from 1 to 5. The key length for Paillier cryptosystem was 64-bit. The message transmitted in a channel at the same time was within 32 bytes, which was applicable for real-world MAS. The system achieved consensus in less than 10 seconds by overcoming the interference from Byzantine nodes. Actually, it took about 210 ms for each iteration, which can be significantly lessened by coding in basic programming language such as C. Besides, a comparative experiment that the system reached consensus was executed following the MSR algorithm (without any privacy-preserving mechanism), which was coded in python as well. It turned out that each iteration took 180 ms for communicating and updating, revealing that the proposed algorithm did not remarkably increase time overhead compared to the MSR algorithm. Thus, it is applicable in real environment.

Additionally, an experiment considering time-varying network was conducted. The linking environment was set to be periodically altered for every 3 iterations to build a MAS with dynamic communicating network using the devices stated above. These three conditions satisfied that the normal network topologies were 1-robust, 2-robust and 3-robust, respectively. This dynamic network was jointly 3-robust, allowing to resist 2-local Byzantine attack. As exhibited in Fig. 9, the system reached consensus in finite time, verifying the feasibility of our algorithm in such conditions.

6. Conclusion

A novel algorithm HE-MSR that allows MAS to reach resilient asymptotic consensus with privacy preservation was proposed in this paper. Technically, a confidential transmission protocol was designed based on homomorphism encryption and asymmetric half-weight mechanism to conceal the privacy from both external eavesdroppers and internal betrayals. Moreover, the ideas of MSR were integrated into the update process to provide resiliency against Byzantine attackers. Compared to previous studies, our method has several advantages. The proposed method considered higher scope of adversaries to snoop privacy and sabotage consensus. Besides, executing our algorithm required less strict conditions compared to those with similar effects. The practicability of HE-MSR was demonstrated by simulation results and hardware experiments.

Despite the achievement obtained in this paper, there are still some issues to be addressed. The computation overhead is still challenging for small-sized agent. Moreover, the consensus result may deviate from the average value of initial states, which should be tackled to be applied in specific conditions. Furthermore, HE-MSR allowed the system to converge in finite time, the analysis of the convergence rate is still worth further investigation though. Future research can focus on these directions to improve privacy preservation and resilient consensus algorithms.

CRedit authorship contribution statement

Mingde Huang: Writing – original draft, Methodology. **Yiming Wu:** Writing – review & editing, Supervision. **Qiuxia Huang:** Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Research Fund of Key Laboratory of Information Security of Zhejiang Province (No. KF201902) and in part by the Scientific Research Fund of Zhejiang Provincial Education Department (No. Y202352122).

Appendix A. Proof of Theorem 3

The proof follows a similar approach to that of Theorem 1.

(Necessity) If the network is not $(F + 1, F + 1)$ -robust, there exist two nonempty, disjoint subsets S_1, S_2 that for $t \geq \bar{t}$, we have $|\Psi_{S_1}^r[t_l, \infty)| + |\Psi_{S_2}^r[t_l, \infty)| \leq F$. In other words, at most F nodes in $S_1 \cup S_2$ have at least $F + 1$ neighbors outside their belonging set at some time steps $t > \bar{t}$. Assume that these nodes are malicious ones, submit to F -total model. Suppose that at time step t , the nodes in S_1 and S_2 hold the minimum and maximum state values, respectively, i.e., $x_{i_1}[t] = m[t], x_{i_2}[t] = M[t], \forall i_1 \in S_1, i_2 \in S_2$. Now that any normal node in $S_1 \cup S_2$ has at most F neighbors outside. As a result, any non-zero message is discarded according to the algorithm. Thus $x_i[t] = x_i[\bar{t}], \forall i \in \mathcal{V}_N \cap (S_1 \cup S_2), t \geq \bar{t}$. The consensus cannot be reached.

(Sufficiency) Using the notations in proof of Theorem 1, we can prove this conclusion by contradiction as well. Modify the value range of δ to

$$0 < \delta < \frac{(\beta_* \alpha_*)^{T|\mathcal{V}_N|}}{1 - (\beta_* \alpha_*)^{T|\mathcal{V}_N|}} \epsilon_0,$$

where T denotes the max length of intervals $[t_l, t_{l+1})$. Represented by t_δ^* the time step that for $t \geq t_\delta^*$, we have $M[t] \leq A_M + \delta, m[t] \geq A_m - \delta$. For two nonempty, disjoint subsets $X_M(t_\delta^*, \epsilon_0), X_m(t_\delta^*, \epsilon_0)$, it is intuitive that at least one normal node inside has at least $F + 1$ outside neighbors at a certain time instant in each time interval $[t_l, t_{l+1})$. Suppose that T_{l1} is the first time step after t_δ^* that one normal node in the union of two subsets has $F + 1$ outside neighbors. Obviously, $T_{l1} - t_\delta^* \leq T$. Then the number of normal nodes in $X_M(T_{l1} + 1, \epsilon_1) \cup X_m(T_{l1} + 1, \epsilon_1)$ is one less than that in $X_M(T_{l1}, \epsilon_0) \cup X_m(T_{l1}, \epsilon_0)$ with positive possibility. Recursively, we have

$$\mathcal{V}_N \cap (X_M(T_{lK} + 1, \epsilon_K) \cup X_m(T_{lK} + 1, \epsilon_K)) = 0.$$

While $\epsilon_K > 0$ still holds, no normal nodes keep state value as A_M or A_m , thus contradicting the assumption.

Appendix B. Proof of Theorem 4

The proof follows a similar approach to that of Theorem 2.

(Necessity) Suppose that the normal network is not jointly $(F + 1)$ -robust. Then there exist two nonempty, disjoint subsets $S_1, S_2 \subset \mathcal{V}_N$, such that at least one of them is jointly $(F + 1)$ -reachable in limited time instants. Let \bar{t} denote the maximum value of these time steps. For $t > \bar{t}$, each node in $S_1 \cup S_2$ has at most F neighbors outside. The same assumption above that the nodes in S_1 and S_2 hold values $m[t]$ and $M[t]$, respectively, is applied. At time step \bar{t} , the nodes in S_1 and S_2 keep the minimum and maximum value among all normal nodes. The consensus is not reached by analogous attack procedures.

(Sufficiency) With the notations and conclusion of Lemma 1, we assume that A_M, A_m are the limitations of $M[t], m[t]$ as t goes to infinity. Assume that at a finite time step t_σ , the limitations are reached, i.e. $M[t_\sigma] = A_M > m[t_\sigma] = A_m$. We focus on the nodes in $X_M[t]$ and $X_m[t]$ when $t > t_\sigma$.

Since the network is jointly $(F + 1)$ -robust, there exists an ISBTI $[t_l, t_{l+1})$ that $X_M[t]$ or $X_m[t]$ is $(F + 1)$ -reachable at a certain time instant in each interval. The time instants are denoted by $T_l \in [t_l, t_{l+1})$. Without loss of generality, assume that $i \in X_M[t]$ has at least $F + 1$ neighbors outside. Thus, at least $F + 1$ negative messages are received by i at T_l . According to the algorithm, at least one of them is reserved, allowing in equations (19) and (20) hold.

In the time interval $[t_l, t_{l+1})$, there is one time instant that the number of nodes holding maximum or minimum values decreases. At the rest of time steps, the number keeps the same with positive possibility. Since the time intervals form an infinite sequence, it is possible that no normal nodes are inside $X_M[t]$ or $X_m[t]$ at a finite time step, contradicting the assumption that the limit for the maximum and minimum values of normal nodes are A_M and A_m where $A_M > A_m$. Thus, sufficiency has been proved.

Data availability

No data was used for the research described in the article.

References

- [1] S. Yang, S. Tan, J.-X. Xu, Consensus based approach for economic dispatch problem in a smart grid, *IEEE Trans. Power Syst.* 28 (2013) 4416–4426.

- [2] Y. Kuriki, T. Namerikawa, Consensus-based cooperative formation control with collision avoidance for a multi-UAV system, in: 2014 American Control Conference, 2014, pp. 2077–2082.
- [3] L.E. Beaver, et al., Constraint-driven optimal control of multiagent systems: a highway platooning case study, *IEEE Control Syst. Lett.* 6 (2021) 1754–1759.
- [4] H. Cai, Y. Su, J. Huang, *Cooperative Control of Multi-Agent Systems*, Springer-Verlag, Cham, 2022.
- [5] A. Amirkhani, A.H. Barshooi, Consensus in multi-agent systems: a review, *Artif. Intell. Rev.* 55 (2022) 3897–3935.
- [6] L. Lamport, R. Shostak, M. Pease, The Byzantine Generals Problem, *Association for Computing Machinery*, 2019, pp. 203–226.
- [7] V. King, J. Saia, Byzantine agreement in expected polynomial time, *J. ACM* 63 (2016) 1–21.
- [8] S. Gao, T. Yu, J. Zhu, W. Cai, T-pbft: an eigentrust-based practical byzantine fault tolerance consensus algorithm, *China Commun.* 16 (2019) 111–123.
- [9] N.H. Vaidya, L. Tseng, G. Liang, Iterative approximate byzantine consensus in arbitrary directed graphs, in: *Proceedings of the 2012 ACM Symposium on Principles of Distributed Computing*, 2012, pp. 365–374.
- [10] H.J. LeBlanc, H. Zhang, X. Koutsoukos, S. Sundaram, Resilient asymptotic consensus in robust networks, *IEEE J. Sel. Areas Commun.* 31 (2013) 766–781.
- [11] Y. Wu, X. He, S. Liu, L. Xie, Consensus of discrete-time multi-agent systems with adversaries and time delays, *Int. J. Gen. Syst.* 43 (2014) 402–411.
- [12] D. Saldana, A. Prorok, S. Sundaram, M.F. Campos, V. Kumar, Resilient consensus for time-varying networks of dynamic agents, in: *2017 American Control Conference*, 2017, pp. 252–258.
- [13] G. Wen, Y. Lv, W.X. Zheng, J. Zhou, J. Fu, Joint robustness of time-varying networks and its applications to resilient consensus, *IEEE Trans. Autom. Control* 68 (2023) 6466–6480.
- [14] Y. Wang, H. Ishii, Resilient consensus through event-based communication, *IEEE Trans. Control Netw. Syst.* 7 (2020) 471–482.
- [15] S.M. Dibaji, M. Safi, H. Ishii, Resilient distributed averaging, in: *2019 American Control Conference*, 2019, pp. 96–101.
- [16] Z. Huang, S. Mitra, G. Dullerud, Differentially private iterative synchronous consensus, in: *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, 2012, pp. 81–90.
- [17] Y. Mo, R.M. Murray, Privacy preserving average consensus, *IEEE Trans. Autom. Control* 62 (2017) 753–765.
- [18] Y. Wang, Privacy-preserving average consensus via state decomposition, *IEEE Trans. Autom. Control* 64 (2019) 4711–4716.
- [19] G. Ramos, A.P. Aguiar, S. Kar, S. Pequito, Privacy-preserving average consensus through network augmentation, *IEEE Trans. Autom. Control* 69 (2024) 6907–6919.
- [20] M. Ruan, H. Gao, Y. Wang, Secure and privacy-preserving consensus, *IEEE Trans. Autom. Control* 64 (2019) 4035–4049.
- [21] C.N. Hadjicostis, A.D. Domínguez-García, Privacy-preserving distributed averaging via homomorphically encrypted ratio consensus, *IEEE Trans. Autom. Control* 65 (2020) 3887–3894.
- [22] D. Fiore, G. Russo, Resilient consensus for multi-agent systems subject to differential privacy requirements, *Automatica* 106 (2019) 18–26.
- [23] Y. Zhang, Z. Peng, G. Wen, J. Wang, T. Huang, Privacy preserving-based resilient consensus for multiagent systems via state decomposition, *IEEE Trans. Control Netw. Syst.* 10 (2022) 1172–1183.
- [24] C. Ying, N. Zheng, Y. Wu, M. Xu, W.-A. Zhang, Privacy-preserving adaptive resilient consensus for multiagent systems under cyberattacks, *IEEE Trans. Ind. Inform.* 20 (2023) 1630–1640.
- [25] S.M. Dibaji, H. Ishii, R. Tempo, Resilient randomized quantized consensus, *IEEE Trans. Autom. Control* 63 (2017) 2508–2522.
- [26] T.C. Aysal, M.J. Coates, M.G. Rabbat, Distributed average consensus with dithered quantization, *IEEE Trans. Signal Process.* 56 (2008) 4905–4918.
- [27] P. Frasca, R. Carli, F. Fagnani, S. Zampieri, Average consensus on networks with quantized communication, *Int. J. Robust Nonlinear Control* 19 (2009) 1787–1816, IFAC-Affiliated Journal.
- [28] A. Kashyap, T. Başar, R. Srikant, Quantized consensus, *Automatica* 43 (2007) 1192–1203.
- [29] R.L. Rivest, A. Shamir, L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Commun. ACM* 21 (1978) 120–126.
- [30] T. Elgamal, A public key cryptosystem and a signature scheme based on discrete logarithms, *IEEE Trans. Inf. Theory* 31 (1985) 469–472.
- [31] P. Paillier, Public-key cryptosystems based on composite degree residuosity classes, in: *International Conference on the Theory and Applications of Cryptographic Techniques*, 1999, pp. 223–238.
- [32] D. Catalano, R. Gennaro, N. Howgrave-Graham, P.Q. Nguyen, Paillier’s cryptosystem revisited, in: *Proceedings of the 8th ACM Conference on Computer and Communications Security*, 2001, pp. 206–214.
- [33] R. Kieckhafer, M. Azadmanesh, Reaching approximate agreement with mixed-mode faults, *IEEE Trans. Parallel Distrib. Syst.* 5 (1994) 53–63.
- [34] C.N. Hadjicostis, A.D. Domínguez-García, T. Charalambous, et al., Distributed averaging and balancing in network systems: with applications to coordination and control, *Found. Trends® Syst. Control* 5 (2018) 99–292.
- [35] O. Goldreich, *Foundations of Cryptography, Basic Applications*, vol. 2, Cambridge University Press, 2009.