

多智能体系统一致性问题中的隐私保护方法综述

黄明德, 伍益明[†], 蒋杰

(杭州电子科技大学 网络空间安全学院, 浙江 杭州 310018)

摘要: 多智能体系统(MAS)一致性算法执行中需要个体之间频繁交换并共享信息, 这对网络中智能体个体信息的隐私安全带来了巨大的风险. 本文阐述了MAS一致性问题中的隐私保护概念定义, 给出了相应的问题描述, 较系统地回顾了MAS隐私保护一致性技术的现有进展. 从方法的归类上将现有技术分为3类: 即基于噪声注入的方法、基于密码学的方法和基于状态增强的方法, 讨论了针对不同领域实际应用中的设计策略, 分析了现有技术的发展趋势, 探讨了隐私保护设计中亟待解决的问题与未来的发展方向.

关键词: 多智能体系统; 协调控制; 一致性问题; 隐私保护

引用格式: 黄明德, 伍益明, 蒋杰. 多智能体系统一致性问题中的隐私保护方法综述. 控制理论与应用, 2026, 43(4): 697–708

DOI: 10.7641/CTA.2025.40362

A survey of privacy-preserving consensus problems in multi-agent systems

HUANG Ming-de, WU Yi-ming[†], JIANG Jie

(School of Cyberspace, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China)

Abstract: In the execution of consensus algorithms in multi-agent systems (MAS), frequent information exchange and sharing among individuals are required, which poses significant risks to the privacy and security of individual information in the network. This paper introduces the concept definition of privacy-preserving in MAS consensus problems, provides corresponding problem descriptions, comprehensively reviews the existing works of privacy protection technologies for MAS consensus in a systematic manner. Based on the classification of methods, the existing technologies are divided into three categories: methods based on noise injection, methods based on cryptography, and methods based on state decomposition. Design strategies for practical applications in different domains are discussed, and the development trends of existing technologies are analyzed. The unresolved issues in privacy-preserving design and future directions are also explored.

Key words: multi-agent systems; coordinated control; consensus; privacy-preserving

Citation: HUANG Mingde, WU Yiming, JIANG Jie. A survey of privacy-preserving consensus problems in multi-agent systems. *Control Theory & Applications*, 2026, 43(4): 697–708

1 引言

多智能体系统 (multi-agent system, MAS) 协调控制是指由若干个智能体节点以及相互通信的链路构成的分布式系统, 在给定的规则或协议下进行协同, 以完成某个合作性的任务或达成共同的目标. 得益于其分布式、易扩展、效率高、自主性强等优点, MAS协调控制在近二十年来得到了广泛地研究^[1], 并在众多领域得到深入应用, 如无人机编队^[2–3]、智能电网调度^[4–5]、智慧交通控制^[6–7]、医疗信息系统^[8–9]等. 在

MAS协调控制的研究中, 一致性问题是最基本的, 也是最具挑战性的问题之一^[10]. 该问题要求系统中的成员通过局部的信息交换, 协同地更新自己的状态值和信息集合, 最终就该状态值达成一致. 其中信息交换过程和更新过程由一致性协议控制, 该协议决定了各智能体的动力学特征变化.

在MAS协调技术不断发展的同时, 对其网络安全的研究越来越得到研究者的重视. 主要原因有两方面组成: 一方面, 从个体来看, 单个智能体的计算资源

收稿日期: 2024–07–08; 录用日期: 2025–01–10.

[†]通信作者. E-mail: ymwu@hdu.edu.cn.

本文责任编辑: 邹云.

浙江省公益技术应用研究项目(LGF21F020011), 浙江省教育厅科研项目(Y202352122)资助.

Supported by the Zhejiang Provincial Public Welfare Research Project of China (LGF21F020011) and the Scientific Research Fund of Zhejiang Provincial Education Department (Y202352122).

通常十分有限,网络安全防护措施很难做到周全,因而易受到恶意攻击的影响甚至被操控;另一方面,从整体来看,由于MAS网络的开放性、节点及链路的可扩展性,导致系统在通信交互过程中同样易遭受网络攻击的威胁^[11]. 目前已有较多的研究人员针对MAS可能遭受的攻击以及相应的安全防御技术进行了研究^[12-13],提出了一些常见的网络攻击手段,这些手段一般以干扰或破坏系统协调控制为目的. 值得注意的是,有一种较为特殊的网络攻击,称为披露攻击,该攻击的目的是获取被攻击对象的隐私信息. 它的攻击途径和媒介多种多样: 直接入侵数据库、监听通讯信道、伪装成正常的智能体获取邻居信息等. 在MAS的实际应用中,需要考虑隐私保护需求的场景越来越多,小至个人出行的行踪,大至军用无人机的位置信息,都迫切需要系统具备高度可靠的隐私保护能力.

鉴于以上原因,近年来针对MAS一致性问题中隐私保护方法的研究引起了国内外研究者的广泛关注. 研究人员从该问题的各个方面开展工作,将攻击者的能力建模并分类,构建隐私保护技术框架,精心设计智能体交互依赖的通信协议以及动力学更新方程,用可量化的标准进行性能评估等. 本文将尝试对这些工作及其贡献进行综述.

2 多智能体系统一致性及其隐私保护需求

2.1 多智能体系统一致性问题

在研究MAS一致性问题时,通常使用图论符号与矩阵理论进行建模和分析. 对于一个由 n 个智能体组成的系统,其对应的网络拓扑记为 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,其中 $\mathcal{V} = \{1, 2, \dots, n\}$ 表示构成网络的节点集合(即智能体集合), $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ 表示构成网络的边集合(即通信链路集合),如果 j 到 i 存在通信链路,则 $(i, j) \in \mathcal{E}$. 这里不考虑自环,即 $(i, i) \notin \mathcal{E}$. 特别地,如果系统中通信链路是有向的,则用 $\mathcal{G}_d = (\mathcal{V}, \mathcal{E}_d)$ 来表示有向图. 权重矩阵为 $\mathbf{W} = (w_{ij})^{n \times n}$,对于 $(i, j) \in \mathcal{E}$,有 $w_{ij} > 0$,否则 $w_{ij} = 0$. 系统中每个智能体都持有一个状态值,在 t 时刻记为 $x_i(t)$, $i \in \mathcal{V}$. 该值遵循特定的动力学方程进行更新,具体当考虑连续时间系统时为

$$\dot{x}_i(t) = u_i(t), t \in \mathbb{R}^+ \cup \{0\}, \quad (1)$$

当考虑离散时间系统时为

$$x_i(t+1) = x_i(t) + u_i(t), t \in \mathbb{N}^+ \cup \{0\}, \quad (2)$$

其中 $u_i(t)$ 为需设计的控制输入. 对于具备分布式系统特性的MAS,智能体节点可获取的通常只有局部信息,即自身的历史状态信息和从一定范围内的相邻节点接收到的信息. 如无特别说明,本文中所述的邻居节点只考虑一跳,即与该智能体直接相连的其他智能体,记节点 i 的邻居集合为 $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$.

而所谓的MAS一致性,指的是系统中每个智能体

就某个状态值 $x_i(t)$ 达成一致,即

$$\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0, \forall i, j \in \mathcal{V}. \quad (3)$$

特别地,若系统最终达成一致的结果等于所有节点初始值的平均值,即

$$\lim_{t \rightarrow \infty} x_i(t) = \frac{1}{n} \sum_{j \in \mathcal{V}} x_j(0), \quad (4)$$

则称系统达成平均一致性. 多智能体系统一致性问题的一般框架由Olfati-Saber等^[14]在2004年提出,设计如下控制输入:

$$u_i(t) = \epsilon \sum_{j \in \mathcal{N}_i} w_{ij}(x_j(t) - x_i(t)), \quad (5)$$

其中 ϵ 为控制增益. 在网络拓扑为强连通平衡图条件下,智能体遵循该协议可以达成平均一致性. 此外,还有较为经典的二阶系统模型,即

$$\begin{cases} x_i(t+1) = x_i(t) + v_i(t), \\ v_i(t+1) = v_i(t) + u_i(t), \end{cases} \quad (6)$$

其中 $x_i(t)$, $v_i(t)$ 分别是位置状态和速度状态. 二阶系统控制器的一般框架由Xie等^[15]提出,具体为

$$u_i(t) = k_1 \sum_{j=1}^n a_{ij}(x_j(t) - x_i(t)) + k_2 \sum_{j=1}^n a_{ij}(v_j(t) - v_i(t)),$$

其中 $k_1 > 0$, $k_2 > 0$ 是需设定的尺度参数.

2.2 隐私保护需求

自电子数据库诞生之日起,人们便开始重视其隐私安全问题^[16]. 研究人员对此进行了深入的探索,提出了各类保护信息安全的方法,如基于数学难题设计的加密系统^[17-19]、基于数据脱敏的 k -匿名算法^[20]、基于扰动的差分隐私机制^[21]等. MAS作为典型的信息物理系统(cyber-physical systems, CPS)^[22],有别于传统的数据库,是集合了存储、计算、通信等操作的动态系统,因而对其隐私保护的依赖程度和难度相较于传统数据库更大. 因此,如何将现有的隐私保护手段有效地融入MAS,或者探索新型的针对MAS的隐私保护方法成为了该领域的研究热点之一.

MAS中窃取隐私的攻击者通常可分为两种类型:外部窃听器^[23]和内部好奇节点^[24]. 二者最主要的区别在于可获取的信息不同:外部窃听器可以对系统中某条或数条通信信道进行监听,并知晓局部甚至全局的网络拓扑结构;内部好奇节点作为参与一致性的一份子,会遵循协议完成正常的交互、更新等行为,同时还会积累收集到的信息,用以对目标智能体的隐私信息进行推断. 上述两种攻击类型是实际应用中较为常见的攻击手段,都具有较高的披露隐私风险,是研究者们重点关注的对象.

在 MAS 一致性问题中, 各智能体的自身状态值, 特别是初始状态值是其重要的一项隐私信息. 而在一致性算法执行中, 智能体之间的信息交互过程是非常容易泄露隐私的阶段. 例如在标准一致性协议中, 智能体会请求邻居节点的状态信息 $x_j(t)$, 这导致隐私直接暴露在通信信道中. 此外, 还有不少可能暴露信息的要素, 如智能体状态值的变化规则、通信链路的权重等. 下文将详细阐述针对不同规模、不同能力的攻击者, 研究人员所采取的各种相应的隐私保护技术.

3 一致性问题中的隐私保护技术

目前, 在 MAS 一致性控制中主要采用的隐私保护方法可以归纳为 3 类: 基于噪声注入的方法、基于加密保护的方法和基于状态增强的方法. 上述方法从不同

的维度使原始数据处于不可观测空间, 以达到保护隐私的目的. 这 3 类方法各自的特点如表 1 所示.

3.1 基于噪声注入(掩码覆盖)的方法

噪声注入的本质是数据模糊化, 是常用的信息保护方法之一^[25]. 该方法以精心设计的噪声使传输的数据发生偏移, 既能够保证系统一致性控制的目的, 又可以避免窃听者获取有效的隐私信息. 这类方法非常适用于大多处于开放网络环境中 MAS. 通常而言, 噪声注入类方法的优劣适合以差分隐私的定义进行分析. 差分隐私要求数据集中任一个或数个数据对整体输出的影响有限, 以保证输出结果的变化情况不能反推个别数据, 起到保护隐私的效果. 这里给出差分隐私的定义.

表 1 各类隐私保护方法的特点

Table 1 Characteristics of various privacy preservation methods

	安全性	通信开销	计算开销	一致性结果	其他
噪声注入法	较高	一般	一般	存在误差的均方一致性	隐私与误差可量化平衡
加密保护法	高	高	高	误差极低的平均一致性	去中心化
状态增强法	高	一般	较高	无误差的平均一致性	无

定义 1 (差分隐私^[21]) 考虑一个随机方程 \mathcal{K} , 如果对于最多一条数据不同的任意一对数据集 D_1, D_2 以及 $S \subseteq \text{Range}(\mathcal{K})$, 满足

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{K}(D_2) \in S], \quad (7)$$

则称 \mathcal{K} 满足 ϵ -差分隐私.

差分隐私的主要实现方式是注入噪声, 其简要控制流程如图 1 所示.

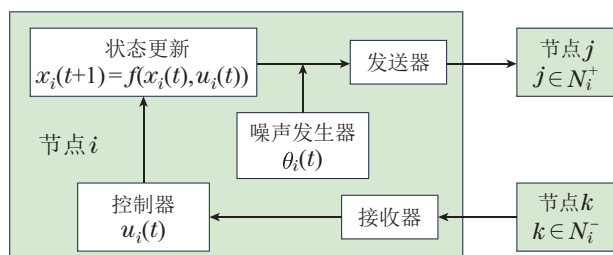


图 1 噪声注入控制流程

Fig. 1 The control process of noise injection

Huang 等^[26]在 2012 年开创性地对差分隐私一致性问题进行了探究. 文章考虑了两种通信模式下的一致性, 即客户端-服务器模式和分布式点对点模式, 分别提出了具备隐私保护能力的一致性机制. 其核心思想是智能体在发送数据前加入噪声 $\eta_i[t]$, 该噪声服从拉普拉斯分布且随时间指数衰减趋向于零, 接收者收到消息后则以一定的比例 σ 将其用于更新. 作者证明了两种机制下系统能够达成收敛, 但一致性值的精确度(与初始状态平均值的接近程度)受加入噪声的参

数影响, 并分析了隐私性能与精确度之间的权衡. 该文首次对协调控制系统中的差分隐私保护进行研究, 为研究者们提供了良好的开端. Manitara 等^[27]同样设计了一种基于注入噪声的一致性算法. 考虑到加入的噪声会影响一致性结果, 该算法巧妙地噪声进行累计补偿, 确保在若干时间步骤后将节点受到来自自身添加的噪声影响完全消除, 以完成整个系统的噪声清零. 此外, 文章定义了系统内的恶意节点(或称好奇节点). 这类节点不影响一致性过程, 但会试图窃取隐私并对其可获取的信息作出较强的假设. 文中考虑它不仅了解内部算法规则, 同时知晓网络拓扑情况、链路权重等信息. 在上述情况下, 文中作者表明只要保证智能体节点自身遵循补偿噪声协议, 或正常节点与恶意节点之间的链路经过了遵循该协议的其他邻居节点, 则该节点的隐私可以得到保护.

Mo 等^[28]在 2017 年提出了一种新的隐私保护平均一致性算法. 算法的关键点在于, 加入的噪声同时具备随时间指数衰减以及总和趋向于零的特点. 文章对算法的均方收敛速率进行了分析, 刻画了最大似然估计的协方差矩阵, 并通过表征泄露空间给出了节点隐私保护所需的条件. 具体而言, 考虑互相独立、服从高斯分布的变量集 $\{v_i(k)\}, i \in \{1, 2, \dots, n\}, k = 0, 1, \dots$. 在初始状态 $k = 0$ 时, 噪声 $w_i(0) = v_i(0)$, 在后续时间步骤 $k \geq 1$ 时, 噪声 $w_i(k) = \phi^k v_i(k) - \phi^{k-1} v_i(k-1)$, 其中 $0 \leq \phi \leq 1$. 此时智能体传输的信息为 $x_i^+(k) = x_i(k) + w_i(k)$, 即真实值和噪声之和. 现假

设好奇节点 j 会尝试从已知信息中推断节点 i 的隐私,即初始状态值.好奇节点的已知信息包括:网络的邻接矩阵 A ,节点 j 自身的初始值 $x_j(0)$,以及收到的所有邻居的数据 $y(k) = Cx^+(k)$,其中 $C = [e_{j1} \cdots e_{jm} e_j] \in \mathbb{R}^{(m+1) \times n}$,即 j 及其邻居所代表的正则基向量组成的矩阵.从统计学角度而言,如果好奇节点能够对目标信息作出无偏估计,那么意味着隐私泄露.而高斯分布可以利用极大似然估计得到最小方差无偏估计量.如果节点 j 的极大似然估计的方差为零(注意选取噪声的衰减性,导致方差单调不减),则节点 j 即可对目标信息作出无偏估计.令极大似然估计的方差矩阵为 $P(k)$,且 $P \triangleq \lim_{k \rightarrow \infty} P(k)$.如果存在向量 ζ 使得 $\zeta^T P \zeta = 0$,则称 ζ 为泄露向量.所有泄露向量张成的空间为泄露空间 D .节点 i 的隐私不会被泄露当且仅当 $e_i \notin D$.通过探究矩阵 P 及其特征向量性质,最终得出结论:当节点 i 及其邻居集 \mathcal{N}_i 不被好奇节点 j 及其邻居集 \mathcal{N}_j 所覆盖时,即 $\mathcal{N}_i \cup i \not\subseteq \mathcal{N}_j \cup j$,则节点 i 的隐私得到保护.值得注意的是,这一性质并非仅对特定分布的衰减噪声有效.实际上,对于任何独立分布的噪声,泄露空间的维数都至少为 $m+1$,其中 m 是好奇节点的邻居数量.这意味着应用不同分布的噪声在这种协议中不会透露任何超过实现平均一致性所需的信息.文中所设计的噪声注入方案及分析证明框架为后续相关研究工作起到了很好的指导作用.

在此基础上, Nozari等^[29]对差分隐私保护方案就精确度进行更深入的研究.作者首先提出并证明了采用噪声注入的差分隐私一致性算法不能保证状态值的收敛分布始终是精确的初始值均值,这也意味着系统的精确度与隐私性能的折中权衡是必然的.对于收敛值的精确度,文章给出了其与期望均值之间可量化的衡量标准.随后作者对系统的控制输入进行优化,通过添加拉普拉斯噪声来线性扰动状态转移函数和消息生成函数,证明了所提算法可以几乎确定性地收敛到无偏的初始状态平均值估计,并给出了收敛速率.此外,作者还对精确度与隐私保护性能之间的参数优化进行了分析,给出了确保最小收敛方差时的最佳噪声选择,并实验模拟了精确度与隐私保护程度之间的平衡关系. He等^[30]则对优化数据隐私的估计和分析进行研究,提出了一种理论框架,将接收到的局部信息用于优化邻居原始隐私的估计,称为最优分布式估计.同时,针对一般的隐私保护平均一致性算法构建了通用模型,以 (ϵ, δ) -数据隐私作为评判标准,揭示了隐私和估计准确性之间的关系,并在最优分布式估计下研究了最优噪声选择问题.随后, He等^[31]还提出了基于共识的数据安全聚合 (secure consensus-based data aggregation, SCDA) 算法.该算法针对分布式数据聚合环境中,系统完成基本的聚合任务,即一致性问

题所设计,其中各个智能体了解的信息有限,即没有全局信息或邻居的所有权重信息.在该假设下,算法通过加入随时间衰减、总和趋向于零的噪声,可以确保最终的精确平均一致性,并且隐私保护能力可以用 (ϵ, δ) -数据隐私进行量化.该算法另一个优点是复杂度低,具体的计算复杂度为 $O(n)$,通信开销为 $O(kn^2)$.在通常一些资源有限的多智能体系统实际应用中,该算法的可应用性较强.同一年, He等^[32]对上述算法进行了进一步挖掘,对加入加性噪声的通用隐保一致性算法 (general privacy-preserving average consensus, GPAC) 下的数据隐私保护能力进行研究.作者计算得出了一定暴露范围内达到最小暴露概率的噪声分布,并在此基础上,作者设计了一种可提供最优隐私保护的平均一致性算法 (optimal privacy-preserving average consensus, OPAC).具体地,每个智能体在初始时刻生成一个服从均匀分布的噪声,并将其添入初始状态值,随后通过一个保密映射函数传输给邻居.所有节点以上述接收到的值作为更新输入,在收敛过程中通过不断加入衰减、零和噪声,可最终确保系统达成平均一致性.此外, Gupta等^[33]则提出了另一种形式的分布式隐私保护机制.这种机制要求每个智能体在一致性算法执行之前先加入一个输入掩码.该掩码则由自己生成的服从高斯分布的噪声与邻居交换之后得到,且所有智能体的掩码总和为零.基于上述设计,可以确保整个系统中所有节点的状态值之和保持不变,且最终能收敛到初始值的平均值.

不局限于离散时间采样系统, Altafini^[34]对连续时间动态系统进行了研究.文中作者给出了在不影响一致性计算过程下,节点采用掩码保护状态值隐私的方法.所设计的掩码函数需要具有以下特点:1) 不等于状态值 x 且不保留 x 的邻域;2) 随 x 增大而增大;3) 与状态值 x 的差值随时间增长而减小并趋向于零.作者从齐次性和随时间指数衰减的干扰项入手,设计了包括线性掩码、仿射掩码等多种掩码.在衰减仿射掩码的保护下,虽然因掩码中非齐次项的存在而不能保证稳定,整个系统仍能以初始均值作为吸引点聚拢.2022年, Xiong等^[35]提出一种新的连续时间算法.该文章将一致性算法分为两个阶段:第1阶段为协调扰动阶段,每个智能体生成一组零和噪声,用于扰动发送给邻居的信息并补偿自身信息;第2阶段是常规收敛阶段,过程中没有任何噪声干扰.该算法同样可以达成平均一致性,且对于不唯一与好奇邻居相连的智能体可以保护隐私,也不会被任何外部窃听者获取隐私.此外,该算法的计算和通信成本较低,且额外的参数选取还能够有效避免被外部窃听者获取,因而实用性较强.

2024年, Ye等^[36]针对3种现有的一致性算法,即传统平均一致性 (conventional average consensus, CAC)

算法,推和平均一致性(push-sum average consensus, PAC)算法和有限时间平均一致性(finite-time average consensus, FAC)算法进行了隐私保护研究,提出了一种通用的隐私保护平均一致性框架,过程中集成了随机乘积变量、有限时间误差补偿和更新规则跳变等机制.在所设计的框架中,作者巧妙地利用误差补偿和更新规则跳变解决了传统方法中的矛盾,同时作者首次将隐私保护融入了有限时间平均一致性算法,保护隐私数据不被弱合谋推断(即智能体的邻居中至少有一个不参与推断)所披露.所提方法在具体执行中分为初始阶段与后续阶段.初始阶段的更新方程以保护隐私为目的,到达一定迭代次数后转变至后续阶段,该阶段的更新方程则以收敛为目的.此外,作者通过对随机变量乘积的研究,表明了该参数倒数的分布方差越大,取得的隐私保护效果则越好.考虑到网络中受限的带宽资源,Gao等^[37]提出了一种基于动态编码-解码的差分隐私控制方法.在该方法中,系统耦合了噪声注入与量化精度.具体地,作者给每一个智能体部署了编码-解码方案和可生成拉普拉斯噪声的注入器,使信息在产生偏移后可进行压缩.随后通过对噪声反馈机制的深入研究,作者给出了在不影响均方一致性结果的前提下从系统动力学的角度调整隐私级别的方法,从而大大减小数据量.文中作者对一致性、隐私级别、数据传输大小等统一进行了定量分析,以便于更为精确地衡量算法性能.

小结而言,通过主动注入噪声使通信链路中传递的信息发生偏移,是一种可靠且便捷的隐私保护方法.这类方法并不会提高信息大小的数量级,因此不会产生额外开销,计算也较为简单;但相对的,噪声的加入会不可避免地使数据产生误差,同时仍保留了一定的信息密度,并非统计意义上的绝对安全.因此,巧妙地设计通信协议和噪声类型是这类方法的重点.此外,对隐私保护效果、一致性结果偏差进行量化评估,是不可或缺的一环,也是提升系统性能的重要指标.更多采用基于噪声注入的隐私保护方法可参见文献[38-43].

3.2 基于密码学的方法

借助密码学的方法对敏感信息进行加密,是在隐私保护需求下一种最为直观的解决方案.然而,在基于分布式系统框架的MAS中想要实现数据加密主要面临两方面的挑战.一方面,分布式系统的特性使得系统设计者较难设置所谓的认证中心(certification authority, CA);另一方面,复杂的加密过程意味着较高的计算量,对通常结构简明的智能体个体而言是个不小的负担.针对上述困难,研究者们尝试给出了一些解决方案.其中,在这些方案中,为服务于不信任计算, MAS中通常采用半同态(或全同态)加密算法对信息

交换过程进行加密.其基本流程如图2所示.

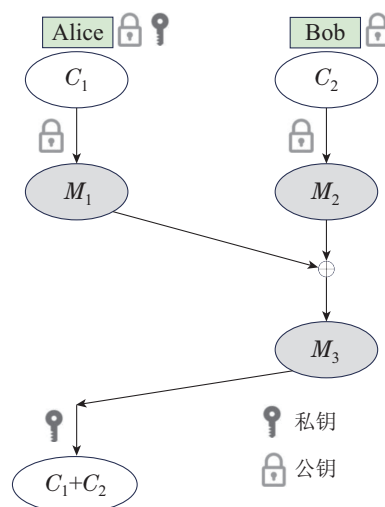


图2 半同态加密交互过程

Fig. 2 Interaction process of semi-homomorphic encryption

同态加密利用密文计算函数到明文计算函数的映射,将需要保护的信息加密后进行安全三方计算后再解密.在这种方式中,通信的另一方即使无法获知具体的信息含义,却仍旧可以得到计算结果.2014年,Lazeretti等^[44]率先探索了利用同态加密保护一致性问题中节点信息隐私的可能性.具体地,在一个多方计算的任务中,系统首先遵循八卦交流协议,即任一时刻仅有一对智能体进行信息交换,其他智能体维持原样.而在随后的一致性过程中,交互的智能体利用同态计算可以获得彼此的信息并取均值,在若干步迭代后利用模糊电路进行安全计算,通过布尔值判断收敛结果.2017年,Ruan等^[45]基于Paillier加密系统设计了一种同态加密平均一致性算法.在完全去中心化的结构下,每个智能体生成一组自己的Paillier密钥,并以此从邻居处获得密文信息.每条链路的权重都由双方智能体共同决定,该权重直接影响信息内容,这种权重生成方式确保了智能体在交互信息时不会泄露隐私.需要注意的是,由于权重的对称性,一对智能体交换信息产生的变化量对彼此是相同的(符号相反),这导致了在达成一致性后,一个智能体的所有邻居可以合谋推断出它的初始状态值.因此为防止内部好奇节点的合谋推断行为,要求被保护的智能体个体需要至少一个非好奇邻居节点.同时,Paillier加密的语义安全性保证了信息不会被外部窃听器泄露,这使得该算法的隐私保护能力较强.文章还对算法防止中间人攻击的可能性进行了探究,提出利用电子签名进行身份验证,可以有效避免系统被篡改信息.在此基础上,Ruan等^[23]将上述算法推广到了极值一致性问题中.此外,针对一致性过程中因随机耦合权重机制而产生的随机抖动会对系统性能造成一定影响的问题,Wang等^[46]进行了研究,提出了一种全新的控制框架.

新框架包括了一阶系统的平均一致性控制器, 以及为二阶系统设计的集合控制器. 文中作者利用隐私保护关系测试协议进行了安全对比, 并基于邻居之间关系信息设计控制器, 巧妙地避免了邻居关系的不连续变化问题, 也即规避了随机抖动问题.

2018年, Hadjicostis^[47]对同态加密在一致性问题中的应用进行了新的探索. 作者针对比率一致性(推和算法)的过程进行了优化, 优化后的算法适用于整数状态值的动态变化. 同时, 作者将同态加密算法应用在状态值的交换过程中, 通过一个可信任节点发布公钥, 确保智能体在并行更新两个状态值时保持信息的私密. 在若干次迭代后, 计算两个状态值的比值即可得到节点真实状态值, 更新过程的结束可由信任节点解密后广播完成. 随后, 在文献[48]中, 作者对上述算法进行了进一步扩展, 将对设置可信任节点的条件放宽至一组节点中至少有一个可信任节点即可. Yin等^[49]在2020年提出了一种精确的隐私保护平均一致性算法(accurate privacy preserving average consensus, APPAC). 该算法要求各智能体生成一组公钥和私钥, 在信息交互中额外加入一组随机噪声, 并在更新时对自己的状态值进行噪声补偿, 以此达到全局状态值之和不变的效果, 同时保护了交互双方的隐私不被披露. Feng等^[50]提出了一种混合匿名加密通信算法(anonymous mixed encryption communication, AMEC). 该算法的核心思想在于智能体的匿名性, 即智能体使用匿名(代号)作为标识, 而不被邻居了解自己的真实身份, 因此在发送或接收消息时, 智能体仅了解对方的代号, 并没有知晓数据的真实来源或去向信息. 此外, 发送的消息由两个智能体节点的状态值之差决定(或称状态差分), 并用公钥加密算法进行加密, 这种处理方式降低了系统的计算成本, 但仍旧具有足够的保密性.

Yang等^[51]在2022年提出了一种具有对抗性交互的离散非线性MAS双边一致性方案. 这里的双边一致性是指, 对于一类具有平衡结构的通信图, 网络中存在两类具有对抗性质的智能体集合, 同一集合内的智能体达成一致性, 而不同集合之间的一致性目标相反. 类似地, 算法利用同态加密对交互过程进行了隐私保护处理, 并将相关权重与差值量化为整数, 以满足实现同态加密的条件. 在此基础上, 算法融入了事件触发机制, 控制了智能体更新的次数, 以降低系统的运算负荷. 实验结果印证了所提算法的隐秘性, 同时确保了有界双边一致性(存在量化过程产生的误差).

Yan等^[52]在2023年设计了一种双层加密方案. 文章针对利用远程控制器连接的多智能体系统网络, 分析了单一的同态加密算法的不足, 并在此基础上增加了一层加密过程, 使得任意两个智能体以及控制器之间(三者互不信任)可以进行安全、隐秘的通信. 基于该框架, 作者给出了算法保证系统实现渐进一致性结

果的控制输入函数的条件. 随后, 采用该算法作者对含有100个智能体的网络进行仿真实验, 实验结果显示每一轮更新用时约29 ms, 验证了所提算法具有较高的执行效率.

小结而言, 基于密码学的隐私保护一致性方法在3类方法中提供了最严格意义的安全性, 甚至包括了语义安全. 但相对应的, 这类方法的计算和通信开销往往巨大(具体依赖于选取的密钥大小), 同时在分布式网络中要求节点双向通信交互信息, 因此不适用于一些计算资源或网络资源受限的应用场合. 这类方法主要应用于对隐私安全需求很高, 同时对计算效率允许该加密方案运行的现实场景.

3.3 基于状态增强的方法

状态增强是近年来提出并得到快速发展的一种节点信息隐私保护方法. 该方法的核心思想是将一个智能体节点(状态)抽象成两个或更多的虚拟子节点(状态), 将其中一个子节点(状态)保持隐秘, 即不参与节点间的信息交互, 而另外的子节点则与相邻智能体节点(或虚拟节点)通信, 从而保证其他个体无法获取该智能体的全部信息, 以此实现状态信息的隐私保护. 从数学意义上分析, 这类方法提高了目标的维数, 但保持其观测空间大小不变, 从而无法确定关键信息. 本文以 $C(0)$, $C'(0)$ 表示两种不同的初始状态, 以 $I_e(t)$, $I'_e(t)$ 表示窃听者 e 在 t 时刻对应可获取的信息, 此时想要证明初始状态具备隐私保护, 则只需证明 $\exists C(0) \neq C'(0)$, 使得

$$I_e(t) = I'_e(t), \forall t \in \mathbb{R}^+ \cup \{0\}, \quad (8)$$

基于节点状态分解的方法示意图如图3所示.

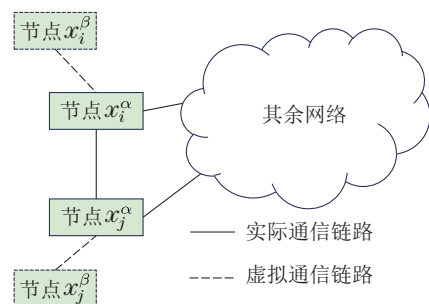


图3 状态分解示意图

Fig. 3 Overview of state decomposition

Wang^[24]在2019年针对一致性算法隐私保护问题, 首次提出了状态分解的方法. 该方法的特点是要求智能体将自身状态值拆分成两个子状态 x_i^α 和 x_i^β , 并令两个子状态初值的和满足该智能体两倍的初始值, 即 $x_i^\alpha[0] + x_i^\beta[0] = 2x_i[0]$. 在通信过程中, 只有代表 α 的子节点会与其他相邻节点交换信息, 而 β 子节点只和自己对应的 α 子节点互相影响并更新. 在不公开地选取链路权重的前提下, 智能体节点遵循作者设计的协

议,无论对内部好奇节点还是对外部窃听器,系统最终可以达成平均一致性,且节点的初始状态隐私得到保护.可以说,作者提出的基于这类状态分解的方法为研究隐私保护一致性算法设计问题开辟了一条新的道路.2021年,Wang等^[53]设计了一种新的节点分解策略.在新的策略下,每个智能体基于自己的邻居数目决定分解成虚拟节点的个数.文中作者设计了两种一致性协议,第1种是基于同态加密的平均一致性协议,另一种则是所谓的规模一致性协议,其中后者协议中各个虚拟节点最终的收敛值会受到本地邻居数量和权重的影响,因而无法得到精确的平均值.Zhang等^[54]在2023年从边的角度提出了另一种状态分解方法.相较于先前针对节点的分解,文中作者设计了一种针对通信图中边进行分解的逻辑状态增强算法,即每个智能体都会根据自己实际的通信链路一一对应地生成虚拟节点.这种算法相较于节点拆分方法,节点的状态变量更多,数据相关性更小,进而隐私泄露的可能性也更低.该算法另一个优点是能够同时保护系统抵御内部好奇节点和外部窃听者的攻击.Chen等^[55]在2023年设计了一种适用于有向网络的基于推和一致性的隐私保护方法.具体地,网络中节点将自身状态值分为4个子状态值,节点利用推和算法的双状态值并行更新,其稳态结果符合转移矩阵左特征向量的特点.由于隐秘节点的存在,只要智能体的邻居中有至少一个不参与隐私推断,那么该智能体的隐私可保证不会被泄露.Ramos等^[56]在2024年对上述问题进行了进一步研究.对于有向网络,文章提出了一种扩展虚拟子网络的方法,该方法要求每个节点额外分配3个虚拟节点,并同时设计通信链路,但该方法仍受限于计算左特征向量的高计算量.对于无向网络,文章提出了一种额外分配4个虚拟节点的扩展方法,该方法则易于计算特征向量,具有较高的实用性.

小结而言,基于状态增强的方法利用了互动信息的不确定性,通过提高自己状态的维数来掩盖真实信息,是一种较为新颖的隐私保护方法.该类方法除了线性加大了计算开销外,并没有明显的额外支出.值得注意的是,利用这类方法要求被保护的智能体节点至少有一个可安全交互的邻居节点,否则该节点自身的累计变化量可以用于推断初始值,从而暴露隐私.

4 一致性问题中的延伸

一致性问题并不局限于一阶线性系统,同样还存在于更复杂的系统动态和环境条件,如二阶系统、多维向量状态、网络延迟、恶意攻击等,这些系统动态和环境条件同样在国内外学者设计一致性协议时得到了考虑.

4.1 高阶系统与多维状态系统

高阶和多维的MAS有着许多应用场合,更切合实

际需求,包括但不限于无人机集群编队、经济调度问题、智能交通管控等.在一些场景中,隐私信息是保护系统安全的重要一环.例如,一群无人机编队过程中需要基于某些协同决策进行耦合运动,但又出于任务的特殊性需要保护自己的原始位置不被泄露,那就有必要在该协同过程中采取隐私保护措施.

Chen等^[57]针对二阶系统提出了一种隐私保护一致性协议.文章考虑了离散时间周期采样的二阶动态系统,其动力学方程为

$$\dot{x}_i(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x_i(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} u_i(t), \quad (9)$$

其中 $x_i(t) = [x_{1i}(t) \ x_{2i}(t)]^T$.受文献[28]的启发,作者设计了零和衰减噪声 $\theta_i(k)$,并将其加入控制输入中,给出了收敛速率分析,并采用文献[32]中提出的隐私定义进行量化,确保了协议的有效性.在此基础上,文献[58]对领导者-跟随者模型进行了研究,所构建的系统要求每个跟随者达到与领导者之间预期的相对距离 Δx_i ,并以差分隐私保护智能体的初始位置信息.作者采用文献[26]中的精确度表征了收敛结果的准确性,并给出了噪声方差参数与偏差概率下界之间的平衡性.Zhang等^[59]针对二阶系统中通信容量受限的情况进行了研究,提出了一种量化通信的二阶一致性算法.作者基于动态编码-解码策略,对注入噪声后的状态值进行量化,最终使系统达成均方一致性.文中作者证明了算法起到差分隐私的保护效果,并且针对量化器饱和和不确定性,给出了存储容量与不饱和概率之间的关系.Tian等^[60]对具有特殊拓扑结构的多智能体系统一致性进行研究,考虑一个有符号的二分图,如果任一部分内部的链路权重大于等于零,两部分之间的链路权重小于等于零,则称之为结构平衡的.对于这种图,作者给出了达成二分一致性的充要条件,并通过注入衰减噪声起到差分隐私的保护效果;而对与非结构平衡图,在保护系统隐私的情况下,给出了能保证系统达到区间二分一致性或稳定性的充要条件.密码学算法也可以应用于二阶系统的隐私保护,如Fang等^[61]考虑了离散采样的时变网络,在基本的一致性动力学方程基础上,作者设计了具有保密性的通信协议,并证明了在系统收敛速率足够快时,可以完整地保护隐私信息.

此外,Pan等^[62]对状态值为向量的特定情况进行了研究.相较于一般的标量状态值,向量维数多,且边权重均为矩阵值,其收敛条件更为复杂.另外,向量值以转移矩阵更新,这意味着不同项之间存在的相关性可能被探听.因此,文章对状态值进行增强,在原本的向量上增加3个维度,并基于一组标准正交基构建权重值,从而保护隐私,避免泄露相关性的效果,同时设计了周期轮换权重矩阵机制,以便于达成收敛条件.

同年, Wang等^[63]从差分隐私的角度为该问题提供了另一种解决方法. 文中作者设计了一种特殊的噪声, 将其添加到智能体在每个迭代过程中的输出信息中, 同时控制输入增益矩阵, 以保证达到均方一致性, 并给出了收敛速率的下界. 在分析隐私时, 考虑到不同项之间的相关性, 采用了不同的方法, 证明并计算得到隐私预算. Yu等^[64]则利用RSA(Rivest-Shamir-Adleman)加密算法对多维状态一致性过程进行了隐私保护.

4.2 分布式优化问题

分布式优化问题是一致性问题的重要延展. 在分布式优化问题中, 每个本地节点存在一个代价函数 $f_i(x)$, 求得的解要使全局代价最小, 即

$$\arg \min \sum_{i \in [N]} f_i(x_i), \quad (10)$$

其中状态值 x_i 处于优化域 $\mathcal{X} \subseteq \mathbb{R}^n$ 内.

Huang等^[65]在2015年对分布式优化过程进行了隐私保护研究. 率先提出了一种解决隐私保护分布式优化问题的通用框架, 其中每个智能体对当前最优解进行估计, 在通信时加入噪声并不断更新估计, 最终令状态值实现一致. 文章设计加入的噪声为衰减的拉普拉斯噪声, 并依此分析了差分隐私的隐私预算, 证明了该算法期望能收敛到最优解, 且精度与隐私预算平方的反比同阶. 在此基础上, Xiong等^[66]利用次梯度重缩放技术进行深入研究. 算法通过加入噪声以保护数据隐私, 并且在数据更新时以次梯度修正下一时刻的状态值, 这使得系统能够更快地达到最优状况. 文章分析了差分隐私预算, 并计算了对于强凸和一般凸的成本函数具有 $O(\log T)$ 和 $O(\sqrt{T})$ 的期望遗憾界. Lu等^[67]则设计了基于同态加密的分布式优化算法. 系统采取的是基于梯度的分布式优化算法, 各智能体将该时刻状态值发送给系统操作者, 并由系统操作者计算出梯度返回给智能体用于更新. 文章提出了两种加密算法, 分别是私钥安全计算与公钥安全计算, 共同点是使系统操作者能够对数据进行同态计算, 从而不泄露隐私. 文章还研究了一类二次联合函数的输入输出推断问题, 并通过对权重矩阵的零空间进行结构分析, 提供了安全的充分条件. 还有不少相关工作对优化问题中的隐私保护进行了深入研究^[68-69].

4.3 对复合网络攻击的防范

在实际应用中, MAS遭受的攻击可能不止于隐私窃取, 还有其他蓄意破坏一致性过程的网络攻击, 如拜占庭攻击^[70]、DoS攻击、虚假数据注入攻击等. 拜占庭攻击会使系统中的部分智能体不遵循一致性协议进行活动, 并任意地发送数据给邻居以干扰其正常更新. Fiore等^[71]在2019年对这种情况进行了讨论. 文章提出了一种基于满足差分隐私要求的中间子序列

缩减(mean subsequence reduced, MSR)算法, 在过滤不安全信息的同时, 适量加入的噪声使智能体的隐私得到保护. 在网络拓扑满足一定鲁棒性的前提下, 系统可以达成均方收敛, 并给出了准确度的上下界以及隐私预算. Wang等^[72]则利用同态加密技术进行了相关研究, 当网络中存在一定规模的恶意节点时, 智能体通过发送信息时附加的时间戳来判断攻击是否发生, 并拒绝被恶意篡改的数据. 系统采用推和算法进行一致性, 并且通信过程受到同态加密的保护, 以保障隐私不被泄露. Hu等^[73]对抵御DoS攻击的方法进行探索, 分析了DoS攻击可能发生的环境与规模, 设计了节点的状态分解策略与控制器输入, 定义了观测矩阵, 以便对于节点状态值不便准确测量的情况进行适应. 通过对网络的鲁棒性和连通性进行计算, 确保在遭受DoS攻击后仍能保证强连通性, 并在不泄露隐私的情况下达成均值一致性. 对于虚假数据注入攻击, Yang等^[74]提出了一种自适应动态事件触发控制方案. 该文巧妙地构建了动态事件触发条件, 减小了数据传输量并且降低了隐私泄露的风险, 并向状态值加入了时变偏移量, 保证真实信息的隐蔽. 借助观测器和补偿器, 智能体引入了自适应辅助变量以补偿非对称偏移引起的残差, 并借由自适应观测误差项设计了攻击补偿器, 消除了注入攻击的影响. 此外, 该文还将适用范围拓展到了有向图中, 利用了半正定矩阵的性质框定了动态参数的范围. 更进一步地, Ying等^[75]考虑了更复杂的网络攻击情况, 研究给出了对应的安全一致性协议设计方法. 假设网络中同时存在欺骗攻击、DoS攻击和披露攻击, 其中前两种攻击受规模限制, 披露攻击则存在于任意信道以及好奇节点中, 文章设计了一种隐私保护自适应弹性算法, 具体利用状态分解保护节点隐私, 同时利用节点对外交换邻居信息在有限迭代次数内保证信息的全覆盖, 当网络通信图满足文章给出的鲁棒性要求时, 系统中所有正常节点的状态值即可达到平均一致性. 更多的相关工作可参见文献[76-77].

5 MAS隐私保护技术的实际应用

目前MAS技术已经广泛应用于各个领域, 其中不乏对隐私有保密需求的场景, 如信息物理系统(CPS)、多机器人编队控制、智能电网、智慧交通等. 在许多商业化管理或多方计算的场景中, 分布式优化是解决问题的主要手段, 其中的计算过程与状态信息往往具有很高的价值, 如智能电网资源管理问题中, 各个发电厂和用电部门需要协调分配电力资源的生产与使用, 以求解分布式优化问题的手段进行系统性控制. 如果没有足够的对信息的隐私保护手段, 那么可能存在个别参与者甚至外来者, 根据传输的数据推断出部分电厂或个人用户的用电情况, 进而获取更敏感的信

息,如电厂发电高峰、用户外出情况等.研究人员针对各种实际环境进行了可应用的技术研究,并通过仿真实验或硬件实验验证了可行性.

Taheri等^[78]对CPS中的隐私保护问题进行研究.具体考虑一种具有同构智能体的CPS,其内部的通信可以视作一个MAS.作者提出了一种等距同构方法,通过该方法将智能体的动力学信息映射到一组新的基,并且保留了原先数据的范数关系.随后为控制器设计隐秘通信协议,基于转换后的输出测量和控制器状态交互,从而保护隐私免于被外部窃听器或内部好奇节点披露.Huo等^[79]则针对CPS中的分布式优化问题进行了隐私保护研究.对于强耦合优化问题,文章给出了一种新范式,具体将半同态加密与分布式优化相结合,以此同时保护智能体和运营商的隐私,并且考虑了运营商不可信时的情况.通过分析来自3种攻击者的披露信息可能,确保系统中的每个成员得到隐私保护,并在真实硬件实验中验证了上述结论.目前,CPS中的隐私保护手段以信息加密和增加映射等方法为主,这类方法可以有效地防止存在好奇节点时隐私可能泄露的情况.不仅如此,另一个显著优点在于保护隐私的同时保留了数据的精确度,对于精密数据的一致性具备较高的实用性.

Zhang等^[80]将隐私保护技术应用于多机器人编队控制领域.考虑到编队过程具有动态变化的参考信号,作者基于代数图论与李雅普诺夫控制理论,设计了加入观测值辅助量的动态更新方程,从而确保机器人的相对位置估计等状态值达到平均一致性,并利用状态分解法保护隐私.该算法在二维平面上具有良好的实验效果.Yue等^[81]则综合多种技术提出了一种连续时间编队算法.文章采用了局部、确定性的时变输出映射函数对状态值进行编码,以保证传输的数据不泄露隐私;利用有限时间稳定性理论,约束与计算编队控制的收敛性能;设计了基于事件触发的控制器,通过合适的触发机制,降低网络中的通信负担,易于实际应用,同时不会对收敛性能造成明显影响.Liang等^[82]则对海面舰只网络(the networked marine surface vehicles, NMSVs)进行了研究.文章提出了一种包含分布式脉冲估计器算法和局部非线性控制算法的网络-物理框架.其中在网络层作者设计了基于同态加密的信息传输协议,实现了对虚拟领航者状态的分布式估计;在物理层则根据估计器信息实现了编队控制,同时具备一定的抗扰动能力.在满足闭环系统收敛和稳定的条件下,该框架为NMSVs在外部扰动和披露攻击下的编队控制提供了有效的解决方案.

智能电网分布式调度问题是另一个重要的研究领域.这类问题映射到MAS中,每个节点代表一个发电站,其发电量为 P_i ,对应的成本函数 C_i 通常为一个二次函数,要求系统在满足额定总发电量的情况下开销

最小,即

$$\begin{aligned} \min \sum_i C_i(P_i), \\ \text{s.t. } \sum_i P_i = P. \end{aligned} \quad (11)$$

Zhao等^[83]在2018年对上述问题进行研究,提出了相应的隐私保护方案.文章设计了两种保护节点隐私的方法,其中一种是引入仅通信双方知道的秘密函数进行交互,并且在本地更新中加入噪声;另一种则是在通信时向输出数据中注入噪声,本地更新则采用真实值.两种方法都可以达成渐进一致性,收敛到最优解,并提供了不同程度的隐私保护性能.此外,文章考虑的模式更加复杂,包括消费者的耗电量以及其消费收益函数等,并对相关信息的隐私性进行了分析.Chen等^[84]提出了一种基于同态加密的分布式调度方案.考虑到同态加密仅作用于整型数据,作者首先构建了一个分级量化器,以便于进行加密操作,同时起到限制数据流大小的效果,随后利用矩阵范数性质推导出量化输出的有界条件,确保算法的精确收敛,同时规定了隐秘传输协议,保证了隐私的安全.Tu等^[85]则针对该问题设计了一种基于状态分解的分布式平均观测器.该观测器采用状态分解的方式保护个体隐私,并分析证明了其收敛性和隐私保护能力.Hu等^[86]则针对二级控制层提出了将输出掩码与节点分解机制耦合的解决方案,其中动态掩码插入在信息传输之前,以掩盖初始值,同时保持了对吸引子的收敛;节点分解机制则将每个分布式电源分解为两个子智能体(子智能体内部不需要掩码隐蔽,因此不额外增加开销),并借此消除了对通信拓扑结构的约束.文章所提出的算法能够满足二阶系统精确共识,且适用的通信拓扑方案非常多,对实际应用意义较大.对于智能电网分布式调度问题的研究工作已经相当全面,从效率和成本的角度考虑,基于噪声注入或状态分解的方法更适合大面积、复杂的电网进行协调一致性控制,而加密类的方法则能提供更高程度的隐私保护,实际应用中应视需求决定具体方案.更多的工作可参见文献[87-89]等.

6 MAS隐私保护技术的未来方向

尽管目前已有相当一部分的隐私保护手段可应用于MAS一致性问题中,但随着信息系统的发展,人们对MAS中个体隐私保护的要求会越来越高.数据隐私保护作为网络安全的重要一环,仍应当是该领域的研究重点.鉴于MAS技术应用场景的不断增多和复杂化,相对应的,隐私保护技术也需要适应新环境.现有的方法仍存在以下几点待解决的问题,值得学者们深入研究:

1) 目前的研究主要针对网络层的信息交互,环境条件较为理想.在实际应用场景中,外部环境因素往

往会对信息传输产生一定的影响,如硬件响应时存在的通信时延、链路通信时的电波扰动等。这些因素会直接影响到数据传输的实时性和有效性,可能导致数据的观测值受到不同程度的影响,进而破坏现有的隐私保护方法,因此需要专门针对这些环境因素寻找合适的补偿和克服方法;

2) 在现有的工作中, MAS多以同构模型构建,而实际应用中有些系统存在异构智能体,如领导者-追随者模型、服务端-客户端模型等。异构MAS中智能体的动力学方程不同,随之而来的是隐私保护实施方法以及需求的多样化。如何让异构MAS在完成一致性目标的同时,根据其需求提供不同的隐私保护级别,是具有较高研究价值的方向;

3) 面向更加复杂任务的MAS一致性控制隐私保护方法设计问题研究不足。例如无人机编队中的一致性控制问题,目前研究居多聚焦于二维平面编队,而针对三维空间的相应算法研究仍有待深入。同样基于一致性的电网能源调度的优化问题也存在更多的实际场景的约束,包括传输时的能量损耗、更复杂的优化目标等。因此,如何将现有技术推陈出新,以适用在更艰巨的任务上,也是一项极具潜力的研究工作。

7 结束语

本文对MAS一致性问题中的隐私保护技术相关工作进行了综述。考虑MAS一致性控制在实际应用领域中系统内部信息所面临的隐私泄露风险,将现有的研究工作按技术实现手段分为了3类,即基于噪声注入的方法、基于密码学的方法和基于状态增强的方法。随着MAS一致性控制技术在复杂多变的环境下应用需求增长,研究人员们提出了较多注重实用性、安全级别、效率等综合性能的隐私保护方法。可以预见的是,未来MAS技术将广泛地应用到众多领域中,相应的隐私保护技术也需要得到进一步研究发展,使其更加成熟和完善,更好地服务于实际应用场景。

参考文献:

- [1] YANG R, LIU L, FENG G. An overview of recent advances in distributed coordination of multi-agent systems. *Unmanned Systems*, 2022, 10(3): 307 – 325.
- [2] MUSLIMOV T Z, MUNASYPOV R A. Consensus-based cooperative control of parallel fixed-wing UAV formations via adaptive backstepping. *Aerospace Science and Technology*, 2021, 109: 106416.
- [3] ZHOU P, CHEN B M. Semi-global leader-following consensus-based formation flight of unmanned aerial vehicles. *Chinese Journal of Aeronautics*, 2022, 35(1): 31 – 43.
- [4] LI Z, CHEN G. Fixed-time consensus based distributed economic generation control in a smart grid. *International Journal of Electrical Power & Energy Systems*, 2022, 134: 107437.
- [5] AHMED I, REHAN M, HONG K S, et al. A consensus-based approach for economic dispatch considering multiple fueling strategy of electricity production sector over a smart grid. *The 13th Asian Control Conference*. Jeju, Korea: IEEE, 2022: 1196 – 1201.
- [6] MISBAH S, SHAHID M F, SIDDIQUI S, et al. Optimizing smart transportation systems with blockchain-based consensus mechanisms: A novel approach. *The 25th International Multitopic Conference*. Lahore, Pakistan: IEEE, 2023: 1 – 6.
- [7] YU G, WONG P K, HUANG W, et al. Distributed adaptive consensus protocol for connected vehicle platoon with heterogeneous time-varying delays and switching topologies. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(10): 17620 – 17631.
- [8] YANG S, ZHU F, LING X, et al. Intelligent health care: Applications of deep learning in computational medicine. *Frontiers in Genetics*, 2021, 12: 607471.
- [9] IQBAL S, ALTAF W, ASLAM M, et al. Application of intelligent agents in health-care: Review. *Artificial Intelligence Review*, 2016, 46(1): 83 – 112.
- [10] AMIRKHANI A, BARSHOOI A H. Consensus in multi-agent systems: A review. *Artificial Intelligence Review*, 2022, 55(5): 3897 – 3935.
- [11] DING Lifu, YAN Gangfeng. A survey of the security issues and defense mechanisms of multi-agent systems. *CAAI Transactions on Intelligent Systems*, 2020, 15(3): 425 – 434. (丁俐夫, 颜钢锋. 多智能体系统安全性问题及防御机制综述. 智能系统学报, 2020, 15(3): 425 – 434.)
- [12] HEDIN Y, MORADIAN E. Security in multi-agent systems. *Procedia Computer Science*, 2015, 60: 1604 – 1612.
- [13] OWOPUTI R, RAY S. Security of multi-agent cyber-physical systems: A Survey. *IEEE Access*, 2022, 10: 121465 – 121479.
- [14] OLFATI-SABER R, MURRAY R M. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 2004, 49(9): 1520 – 1533.
- [15] XIE D, WANG S. Consensus of second-order discrete-time multi-agent systems with fixed topology. *Journal of Mathematical Analysis and Applications*, 2012, 387(1): 8 – 16.
- [16] JAIN P, GYANCHANDANI M, KHARE N. Big data privacy: A technological perspective and review. *Journal of Big Data*, 2016, 3: 1 – 25.
- [17] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 1978, 21(2): 120 – 126.
- [18] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 1985, 31(4): 469 – 472.
- [19] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes. *International Conference on the Theory and Applications of Cryptographic Techniques*. Berlin, Heidelberg: Springer, 1999: 223 – 238.
- [20] SWEENEY L. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 2002, 10(5): 557 – 570.
- [21] DWORK C. Differential privacy. *International Colloquium on Automata, Languages, and Programming*. Berlin, Heidelberg: Springer, 2006: 1 – 12.
- [22] BODKHE U, MEHTA D, TANWAR S, et al. A survey on decentralized consensus mechanisms for cyber physical systems. *IEEE Access*, 2020, 8: 54371 – 54401.
- [23] RUAN M, GAO H, WANG Y. Secure and privacy-preserving consensus. *IEEE Transactions on Automatic Control*, 2019, 64(10): 4035 – 4049.
- [24] WANG Y. Privacy-preserving average consensus via state decomposition. *IEEE Transactions on Automatic Control*, 2019, 64(11): 4711 – 4716.
- [25] BRUNTON F, NISSENBAUM H. Obfuscation: A user's guide for privacy and protest. Peachtree City, Georgia: Mit Press, 2015: 154 – 156.

- [26] HUANG Z, MITRA S, DULLERUD G. Differentially private iterative synchronous consensus. *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*. New York, USA: ACM, 2012: 81 – 90.
- [27] MANITARA N E, HADJICOSTIS C N. Privacy-preserving asymptotic average consensus. *European Control Conference*. Zurich, Switzerland: IEEE, 2013: 760 – 765.
- [28] MO Y, MURRAY R M. Privacy preserving average consensus. *IEEE Transactions on Automatic Control*, 2017, 62(2): 753 – 765.
- [29] NOZARI E, TALLAPRAGADA P, CORTÉS J. Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design. *Automatica*, 2017, 81: 221 – 231.
- [30] HE J, CAI L, GUAN X. Preserving data-privacy with added noises: Optimal estimation and privacy analysis. *IEEE Transactions on Information Theory*, 2018, 64(8): 5677 – 5690.
- [31] HE J, CAI L, CHENG P, et al. Consensus-based data-privacy preserving data aggregation. *IEEE Transactions on Automatic Control*, 2019, 64(12): 5222 – 5229.
- [32] HE J, CAI L, ZHAO C, et al. Privacy-preserving average consensus: Privacy analysis and algorithm design. *IEEE Transactions on Signal and Information Processing over Networks*, 2019, 5(1): 127 – 138.
- [33] GUPTA N, KATZ J, CHOPRA N. Privacy in distributed average consensus. *IFAC-PapersOnLine*, 2017, 50(1): 9515 – 9520.
- [34] ALTAFINI C. A dynamical approach to privacy preserving average consensus. *IEEE 58th Conference on Decision and Control*. Nice, France: IEEE, 2019: 4501 – 4506.
- [35] XIONG Y, LI Z. Privacy-preserved average consensus algorithms with edge-based additive perturbations. *Automatica*, 2022, 140: 110223.
- [36] YE F, CAO X, CHOW M Y, et al. Privacy-preserving average consensus: Fundamental analysis and a generic framework design. *IEEE Transactions on Information Theory*, 2024, 70(4): 2870 – 2885.
- [37] GAO C, WANG Z, HE X, et al. Differentially private consensus control for discrete-time multiagent systems: Encoding-decoding schemes. *IEEE Transactions on Automatic Control*, 2024, 69(8): 5554 – 5561.
- [38] REZAZADEH N, KIA S S. Privacy preservation in a continuous-time static average consensus algorithm over directed graphs. *Annual American Control Conference*. Milwaukee, WI, USA: IEEE, 2018: 5890 – 5895.
- [39] ALTAFINI C. A system-theoretic framework for privacy preservation in continuous-time multiagent dynamics. *Automatica*, 2020, 122: 109253.
- [40] CHEN W, WANG Z, HU J, et al. Differentially private average consensus with logarithmic dynamic encoding-decoding scheme. *IEEE Transactions on Cybernetics*, 2023, 53(10): 6725 – 6736.
- [41] LIU X K, ZHANG J F, WANG J. Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems. *Automatica*, 2020, 122: 109283.
- [42] ZHANG J, LU J, LOU J. Privacy-preserving average consensus via finite time-varying transformation. *IEEE Transactions on Network Science and Engineering*, 2022, 9(3): 1756 – 1764.
- [43] LU Y, ZHU M. On privacy preserving data release of linear dynamic networks. *Automatica*, 2020, 115: 108839.
- [44] LAZZERETTI R, HORN S, BRACA P, et al. Secure multi-party consensus gossip algorithms. *IEEE International Conference on Acoustics, Speech and Signal Processing*. Florence, Italy: IEEE, 2014: 7406 – 7410.
- [45] RUAN M, AHMAD M, WANG Y. Secure and privacy-preserving average consensus. *Proceedings of the 2017 Workshop on Cyber-physical Systems Security and Privacy*. New York, USA: ACM, 2017: 123 – 129.
- [46] WANG H, LI D, GUAN Z, et al. Consensus control based on privacy-preserving two-party relationship test protocol. *IEEE Control Systems Letters*, 2023, 7: 2185 – 2190.
- [47] HADJICOSTIS C N. Privacy preserving distributed average consensus via homomorphic encryption. *IEEE Conference on Decision and Control*. Miami, FL, USA: IEEE, 2018: 1258 – 1263.
- [48] HADJICOSTIS C N, DOMÍNGUEZ-GARCÍA A D. Privacy-preserving distributed averaging via homomorphically encrypted ratio consensus. *IEEE Transactions on Automatic Control*, 2020, 65(9): 3887 – 3894.
- [49] YIN T, LV Y, YU W. Accurate privacy preserving average consensus. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2019, 67(4): 690 – 694.
- [50] FENG Y, WANG F, DUAN F, et al. Anonymous privacy-preserving consensus via mixed encryption communication. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022, 69(8): 3445 – 3449.
- [51] YANG Z, YU L, LIU Y, et al. Event-triggered privacy-preserving bipartite consensus for multi-agent systems based on encryption. *Neurocomputing*, 2022, 503: 162 – 172.
- [52] YAN Y, CHEN Z, VARADHARAJAN V. Consensus of networked control multi-agent systems using a double-layer encryption scheme. *Journal of Automation and Intelligence*, 2023, 2(4): 218 – 226.
- [53] WANG Y, LU J, ZHENG W X, et al. Privacy-preserving consensus for multi-agent systems via node decomposition strategy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2021, 68(8): 3474 – 3484.
- [54] ZHANG J, LU J, CHEN X. Privacy-preserving average consensus via edge decomposition. *IEEE Control Systems Letters*, 2022, 6: 2503 – 2508.
- [55] CHEN X, HUANG L, DING K, et al. Privacy-preserving push-sum average consensus via state decomposition. *IEEE Transactions on Automatic Control*, 2023, 68(12): 7974 – 7981.
- [56] RAMOS G, AGUIARZ A P, KARX S, et al. Privacy preserving average consensus through network augmentation. *IEEE Transactions on Automatic Control*, 2024, 69(10): 6907 – 6919.
- [57] CHEN Z, FU B, WU Z, et al. Privacy preserving second-order consensus for wasns. *The 37th Chinese Control Conference*. Wuhan, China: IEEE, 2018: 7236 – 7241.
- [58] MA M, ZHAO C, HE J. Differentially private discrete-time second-order consensus under directed topologies. *IEEE 16th International Conference on Control & Automation*. Singapore: IEEE, 2020: 1118 – 1123.
- [59] ZHANG W, WANG B C, LIANG Y. Differentially private consensus for second-order multiagent systems with quantized communication. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 35(4): 5523 – 5535.
- [60] TIAN R, ZUO Z, HAN Q, et al. Differential privacy for second-order bipartite consensus over signed digraph. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024, 54(6): 3652 – 3664.
- [61] FANG W, ZAMANI M, CHEN Z. Secure and privacy preserving consensus for second-order systems based on paillier encryption. *Systems & Control Letters*, 2021, 148: 104869.
- [62] PAN L, SHAO H, LU Y, et al. Vector-valued privacy-preserving average consensus. *ArXiv Preprint*, 2022, arXiv: 2209.10786.
- [63] WANG Y, LAM J, LIN H. Consensus of linear multivariable discrete-time multiagent systems: Differential privacy perspective. *IEEE Transactions on Cybernetics*, 2022, 52(12): 13915 – 13926.
- [64] YU L, YU W, LV Y. Multi-dimensional privacy-preserving average consensus in wireless sensor networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021, 69(3): 1104 – 1108.

- [65] HUANG Z, MITRA S, VAIDYA N. Differentially private distributed optimization. *Proceedings of the 16th International Conference on Distributed Computing and Networking*, New York, USA: ACM, 2015: 1 – 10.
- [66] XIONG Y, XU J, YOU K, et al. Privacy-preserving distributed on-line optimization over unbalanced digraphs via subgradient rescaling. *IEEE Transactions on Control of Network Systems*, 2020, 7(3): 1366 – 1378.
- [67] LU Y, ZHU M. Privacy preserving distributed optimization using homomorphic encryption. *Automatica*, 2018, 96: 314 – 325.
- [68] YE Y, CHEN H, XIAO M, et al. Privacy-preserving incremental ADMM for decentralized consensus optimization. *IEEE Transactions on Signal Processing*, 2020, 68: 5842 – 5854.
- [69] RIKOS A I, NYLÖF J, GRACY S, et al. Distributed optimal allocation with quantized communication and privacy-preserving guarantees. *IFAC-PapersOnLine*, 2022, 55(41): 64 – 70.
- [70] LAMPOR T L, SHOSTAK R, PEASE M. The Byzantine generals problem. *Concurrency: The Works of Leslie Lamport*, 2019: 203 – 226.
- [71] FIORE D, RUSSO G. Resilient consensus for multi-agent systems subject to differential privacy requirements. *Automatica*, 2019, 106: 18 – 26.
- [72] WANG D, ZHENG N, XU M, et al. Resilient privacy-preserving average consensus for multi-agent systems under attacks. *The 16th International Conference on Control, Automation, Robotics and Vision*. Shenzhen, China: IEEE, 2020: 1399 – 1405.
- [73] HU Qinling, ZHENG Ning, XU Ming, et al. Privacy-preserving average consensus control for multi-agent systems under DoS attacks. *Acta Automatica Sinica*, 2022, 48(8): 1961 – 1971. (胡沁伶, 郑宁, 徐明, 等. DoS攻击下具备隐私保护的多智能体系统均值趋同控制. *自动化学报*, 2022, 48(8): 1961 – 1971.)
- [74] YANG Y, LI J, WANG X, et al. Adaptive dynamic average consensus scheme with preserving privacy and against false data injection attacks: Dynamic event-triggered mechanism. *IEEE Transactions on Vehicular Technology*, 2024, 73(6): 7826 – 7837.
- [75] YING C, ZHENG N, WU Y, et al. Privacy-preserving adaptive resilient consensus for multiagent systems under cyberattacks. *IEEE Transactions on Industrial Informatics*, 2023, 20(2): 1630 – 1640.
- [76] ZHANG Y, PENG Z, WEN G, et al. Privacy preserving-based resilient consensus for multiagent systems via state decomposition. *IEEE Transactions on Control of Network Systems*, 2022, 10(3): 1172 – 1183.
- [77] XU Ming, ZHANG Baojun, WU Yiming, et al. Cyber attacks and privacy protection distributed consensus algorithm for multi-agent systems. *Journal on Communications*, 2023, 44(3): 117 – 127. (徐明, 张保俊, 伍益明, 等. 面向网络攻击和隐私保护的多智能体系统分布式一致性算法. *通信学报*, 2023, 44(3): 117 – 127.)
- [78] TAHERI M, KHORASANI K, SHAMES I, et al. Towards privacy preserving consensus control in multi-agent cyber-physical systems subject to cyber attacks. *European Control Conference*. Delft, Netherlands: IEEE, 2021: 939 – 945.
- [79] HUO X, LIU M. Encrypted decentralized multi-agent optimization for privacy preservation in cyber-physical systems. *IEEE Transactions on Industrial Informatics*, 2021, 19(1): 750 – 761.
- [80] ZHANG K, LI Z, WANG Y, et al. Privacy-preserving dynamic average consensus via state decomposition: Case study on multi-robot formation control. *Automatica*, 2022, 139: 110182.
- [81] YUE J, QIN K, SHI M, et al. Event-trigger-based finite-time privacy-preserving formation control for multi-uav system. *Drones*, 2023, 7(4): 235.
- [82] LIANG C D, GE M F, XU J Z, et al. Secure and privacy-preserving formation control for networked marine surface vehicles with sampled-data interactions. *IEEE Transactions on Vehicular Technology*, 2021, 71(2): 1307 – 1318.
- [83] ZHAO C, CHEN J, HE J, et al. Privacy-preserving consensus-based energy management in smart grids. *IEEE Transactions on Signal Processing*, 2018, 66(23): 6162 – 6176.
- [84] CHEN W, LIU L, LIU G P. Privacy-preserving distributed economic dispatch of microgrids: A dynamic quantization-based consensus scheme with homomorphic encryption. *IEEE Transactions on Smart Grid*, 2022, 14(1): 701 – 713.
- [85] TU H, DU Y, YU H, et al. Privacy-preserving robust consensus for distributed microgrid control applications. *IEEE Transactions on Industrial Electronics*, 2023, 71(4): 3684 – 3697.
- [86] HU J, SUN Q, WANG R, et al. An improved privacy-preserving consensus strategy for AC microgrids based on output mask approach and node decomposition mechanism. *IEEE Transactions on Automation Science and Engineering*, 2022, 21(1): 642 – 651.
- [87] YAN Y, CHEN Z, VARADHARAJAN V, et al. Distributed consensus-based economic dispatch in power grids using the paillier cryptosystem. *IEEE Transactions on Smart Grid*, 2021, 12(4): 3493 – 3502.
- [88] WANG A, LIU W, DONG T, et al. DisEHPPC: Enabling heterogeneous privacy-preserving consensus-based scheme for economic dispatch in smart grids. *IEEE Transactions on Cybernetics*, 2020, 52(6): 5124 – 5135.
- [89] WANG Z, WANG J, LA SCALA M, et al. Real-time privacy-preserving average consensus and its application to secondary control for ac microgrid. *IEEE Transactions on Industrial Informatics*, 2024, 20(7): 9655 – 9669.

作者简介:

黄明德 硕士研究生, 目前研究方向为多智能体系统一致性控制,

E-mail: huang_mingde@hdu.edu.cn;

伍益明 副教授, 博士生导师, 目前研究方向为自主无人系统网络安全、分布式系统数据隐私保护、集群智能等, E-mail: ymwu@hdu.edu.cn;

蒋杰 硕士研究生, 目前研究方向为多智能体系统一致性控制, E-mail: jiangjie@hdu.edu.cn.