基于隐写术的分布式隐私保护一致性控制方法

伍益明 [1] 张润荣 [1] 徐 宏 [2, 3] 朱晨睿 [1, 3] 郑 宁 [1]

摘 要 多智能体网络 (Multi-agent network, MAN) 协同执行任务时需要个体之间频繁交换并共享信息, 这对网络安全带来了巨大风险. 考虑网络中节点状态隐私保护问题, 提出一种基于隐写术的分布式一致性控制策略. 首先, 建立网络窃听者攻击模型, 提出面向隐私保护的分布式平均一致性控制算法. 理论分析表明, 所提算法不仅有效保护节点初始状态的隐私, 而且可以通过隐写载体信息主动诱导窃听者推测得出错误结论. 其次, 通过引入概率指标, 提出一种用于量化 MAN 隐私泄露指标的模型, 实现了对网络隐私泄露程度的准确描述. 并基于该模型, 从窃听者视角, 通过权衡对网络隐私泄露的影响与付出代价成本建立一个优化问题, 据此寻找最优效益攻击策略. 最后, 通过数值仿真分析, 对比现有算法验证了所提方法的有效性和优越性.

关键词 多智能体网络, 隐私保护, 窃听攻击, 网络安全, 隐写术

引用格式 伍益明, 张润荣, 徐宏, 朱晨睿, 郑宁. 基于隐写术的分布式隐私保护一致性控制方法. 自动化学报, 2025, **51**(1): 221−232

DOI 10.16383/j.aas.c240089 **CSTR** 32138.14.j.aas.c240089

Distributed Privacy-preserving Consensus Control Based on Steganography

WU Yi-Ming[1] ZHANG Run-Rong[1] XU Hong[2, 3] ZHU Chen-Rui[1, 3] ZHENG Ning[1]

Abstract Multi-agent network (MAN) requires frequent exchange and sharing of information among individuals in collaborative task execution, which poses significant risks to network security. To address the issue of privacy protection of node states in the network, a distributed consensus control strategy based on steganography is proposed. Firstly, an eavesdropper attack model is established, and a distributed privacy-preserving average consensus control algorithm is proposed. Theoretical analysis shows that the proposed algorithm not only effectively protects the privacy of node initial states but also can actively lead eavesdroppers to draw incorrect conclusions through steganographic carrier information. Secondly, a probabilistic indicator is introduced to quantify the privacy leakage index model for MAN, achieving an accurate description of the degree of network privacy leakage. Based on this model, an optimization problem is established from the eavesdropper′s perspective by balancing the impact on network privacy leakage and the cost of attack, in order to find the optimal benefit attack strategy. Finally, through numerical simulation, the effectiveness and superiority of the proposed method are verified by comparing with existing algorithms.

Key words Multi-agent network (MAN), privacy-preserving, eavesdropping attack, network security, steganography

Citation Wu Yi-Ming, Zhang Run-Rong, Xu Hong, Zhu Chen-Rui, Zheng Ning. Distributed privacy-preserving consensus control based on steganography. *Acta Automatica Sinica*, 2025, **51**(1): 221−232

多智能体网络 (Multi-agent network, MAN) 协同控制问题在近二十年来得到了持续深入的研究, 取得一系列丰硕的成果, 并在诸多领域得到应用, 如多机器人系统[1]、智能电网[2]、无人机编队[3]、智慧交通[4] 等. MAN 一致性控制是分布式协同控制研究的关键问题之一, 是指在没有控制中心的情况下, 网络中每个个体仅使用邻居间相互传递的状态信息, 将智能体动力学方程与通信网络拓扑耦合成复杂网络, 并遵循合适的分布式控制协议, 最后在有限时间内将所有智能体状态信息达成一致.

近年来, 随着网络安全事件的频发, 关于分布式控制系统, 尤其 MAN 的隐私保护问题受到国内外学者广泛关注[5−8]. 对于每个智能体来说, 现有的分布式控制协议需要与其邻居个体共享状态信息.

收稿日期 2024-02-23 录用日期 2024-08-27
Manuscript received February 23, 2024; accepted August 27, 2024
浙江省公益技术应用研究项目 (LGF21F020011) 资助
Supported by Zhejiang Provincial Public Welfare Research Project of China (LGF21F020011)
本文责任编委 孙健
Recommended by Associate Editor SUN Jian
1. 杭州电子科技大学网络空间安全学院 杭州 310018 2. 杭州电子科技大学计算机学院 杭州 310018 3. 中国电子科技集团公司第三十二研究所 上海 201800
1. School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018 2. School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018 3. The 32nd Research Institute of China Electronics Technology Group Corporation, Shanghai 201800

毫无疑问, 如果存在一些恶意节点试图窃听他人的状态, 上述分布式协议特性将对系统带来巨大的安全隐患[9]. 而对网络中个体信息的隐私保护是十分有必要的. 例如, 一群智能体计划在某个协同决策的地点集结, 但由于一些执行任务的特定原因, 它们希望将自己的初始位置对其他个体进行保密[10]. 再例如在上述提及的智能电网应用中, 执行者通过分布式数据管控收集了所有用户的用电信息, 以此分析优化供电方案, 而如果事先没有对数据进行安全处理, 在该过程中则可以推断出单个用户的用电时间和数据, 进而推断出其居家或外出等敏感个人信息. 因此, 在实现智能体间状态一致的基础上, 研究分布式网络中如何保护各节点自身信息的隐私问题成为当前的一个研究热点.

现有的研究工作中, 针对分布式网络控制系统中防御窃听者的方法大体可以归纳为 4 类. 1) 通过人工添加噪声的方法, 即在节点信息传递过程中注入精心设计的噪声, 以混淆其真实的状态信息[5, 11−13]. 而其中最具代表性的为差分隐私算法[11], 它的基本原理是在分享的数据中加入服从某种分布的噪声, 使相邻节点不能得到原始数据的准确信息. 文献 [13] 则将这类算法有效应用于分布式智能电网的能量管理中. 2) 通过密码学的方法, 即采用端到端的加解密技术, 确保通信内容只有通信双方能够读取, 防止窃听者解析具体通信内容[14−16]. 例如文献 [16] 通过引入同态加密技术, 确保 MAN 在分布式协作中通信信息的机密性和完整性. 3) 通过产生干扰信号的方法. 例如文献 [17], 作者通过中继节点向目的地节点发送干扰信号, 以破坏窃听者对传输信号的解析能力. 值得注意的是, 与第一类方法相比, 该类方法一般通过采用带外通信的方式产生干扰信号, 即节点自身真实信息并不混入噪声. 4) 近年来新出现的基于信息拆分的方法[9, 18−20]. 即通过对网络中的个体状态信息进行拆分, 留取一部分信息在节点内部, 另一部分信息则参与共享交互. 这样即使窃听者得到了网络中传输的交互信息, 也仅能推测出拆分后的部分信息, 无法获得拆分前的完整节点信息. 根据具体对象, 这类方法可以进一步分为节点信息拆分[9] 和链路信息拆分[20].

隐写术 (Steganography)[21] 则是在多媒体信息与信号处理领域提出的一种解决通信安全的方法. 隐写术作为信息隐藏技术的重要支撑技术, 旨在将秘密信息隐藏在可公开的普通载体中进行传送, 实现隐蔽通信. 例如在谍战剧和侦探小说中, 特务用药水将情报写到纸上, 收到情报的上级再通过显影技术把情报还原出来. 将秘密信息通过可公开的载体进行传递, 且不被第三方发现, 这就是隐写术. 它与密码学方法不同, 密码学方法尽管保证了信息内容上的安全, 却暴露了通信的行为, 易引起怀疑, 给窃听者追踪的线索. 隐写术则是将某些特殊信息隐藏于正常载体之中, 从而掩盖特殊信息存在的事实, 不易被窃听者察觉. 由于其在军事、安全、工业界广泛的应用前景, 隐写术作为一类新的信息安全技术, 受到国内外研究人员越来越密切的关注[22].

当前基于密码学的隐私保护一致性算法通常将状态信息加密为无意义的密文形式, 但此举也暴露了数据的重要性. 例如, 网络中某个智能体基于密码学方法向邻居发送了一段密文信息, 这种明显的加密消息, 无论其多么难以破解, 都会引起对方的注意, 致使该智能体节点的秘密信息易遭到窃听者的怀疑和拦截. 而与之相比, 将隐写术思想用于一致性算法, 不仅可以避免窃听者怀疑, 保护秘密信息的内容安全, 而且可以确保隐蔽通信的行为安全. 此外, 现有文献鲜有从窃听者的角度, 考虑系统整体的隐私泄露程度和实施窃听攻击的成本, 进而评估成功实施攻击的效益问题. 基于上述考虑, 本文围绕隐私保护下的分布式系统一致性控制问题展开研究, 提出一种新的采用信息隐藏思想的平均一致性控制协议, 并尝试从窃听者的角度, 构建一类隐私泄露量化评估模型, 为 MAN 隐私保护设计提供更加全面的技术支撑. 本文主要贡献如下:

1) 在 MAN 协调控制领域, 尝试引入信息隐藏的思想, 提出了一种新的基于隐写术的分布式一致性控制方法.

2) 提出的一致性算法通过引入隐藏−提取规则, 使得网络窃听者无法推测节点的真实状态, 从而保证其隐私性. 相较于文献 [5, 11−13] 等人工添加噪声后导致收敛精度不可保证问题, 本文方法可以有效保证系统节点收敛至精确的平均一致, 同时算法中可任意控制传输的明文信息, 以此来迷惑和误导窃听者推测出错误结论.

3) 通过精心构造的隐私保护机制, 控制器设计中避免了采用与文献 [14−16] 等基于密码学的需消耗较高算力的数据加解密技术, 从而保证在带宽和算力有限的网络前提下, 控制信号连续以及抑制高延迟问题.

4) 构建了一种可以衡量分布式 MAN 隐私泄露的度量模型, 继而将其引入隐私窃取攻击实施问题中, 提出了一种基于 Lambda 迭代优化的最优攻击策略.

1 预备知识

1.1 图论

考虑一个由 n 个智能体组成的 MAN, 其通信

拓扑结构可以抽象为一个有向加权图 $\mathcal{G}=\{\mathcal{V},\ \mathcal{E},\ \omega\}$ 来表示. 其中 $\mathcal{V}=\{v_1,\ v_2,\ \cdots,\ v_n\}$ 是节点集, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ 是边集, $\omega=[W_{ij}]\in\mathbf{R}^{n\times n}$ 为加权邻接矩阵. 矩阵元素 W_{ij} 表示节点 v_i 到节点 v_j 的连接权重, 如果 $(v_i,\ v_j)\in\mathcal{E}\Leftrightarrow W_{ij}>0$, 否则 $W_{ij}=0$. 本文中不考虑 $\mathcal{G}$ 存在自环的情况, 即令 $W_{ii}=0$. 如果 $(v_j,\ v_i)\in\mathcal{E}$, 则 v_j 是 v_i 的入邻居, v_i 的所有入邻居集表示为 $\mathcal{N}_i^{\text{in}}=\{j:(v_j,\ v_i)\in\mathcal{E}\}$; 如果 $(v_i,\ v_j)\in\mathcal{E}$, 则 v_j 是 v_i 的出邻居, v_i 的所有出邻居集表示为 $\mathcal{N}_i^{\text{out}}=\{j:(v_i,\ v_j)\in\mathcal{E}\}$. 对于节点 v_i, 它的入度和出度分别定义为 $\deg_{\text{in}}(v_i)=\sum_{j=1}^{n}W_{ji}$ 和 $\deg_{\text{out}}(v_i)=\sum_{j=1}^{n}W_{ij}$. 如果 v_i 的入度和出度相等, 即 $\deg_{\text{in}}(v_i)=\deg_{\text{out}}(v_i)$, 则称 v_i 为平衡节点. 当 $\mathcal{G}$ 中任意节点都为平衡节点时, 此时称图 $\mathcal{G}$ 为平衡图. 在有向图 $\mathcal{G}$ 中, 对于图中任意两个节点 v_i 和 v_j, 若存在至少一条有向连接路径, 则称图 $\mathcal{G}$ 是强连通的.

1.2　隐写术

隐写术是信息隐藏研究领域的重要方向之一, 其主要思想是将秘密信息隐藏在另一非机密的载体信息中, 并通过公共信道进行传递[23]. 隐写术可以进一步分为秘密隐写和普通隐写. 前者着重于信息伪装以实现秘密信息的传递, 后者则解决信息的隐含标识[24]. 秘密信息被隐藏后, 攻击者无法从载体信息中提取或去除所隐藏的秘密信息, 甚至无法判断载体信息中是否隐藏了秘密信息.

受文献 [25] 的启发, 我们可以建立一个由六元组 $\Theta(H,\ E,\ X,\ C,\ X',\ X'')$ 表示的信息隐藏安全模型, 其中 X 是节点实时状态数据集, C 是辅助值, X' 是隐藏后的数据集, X'' 是接收的数据集. 六元组需满足以下条件:

1) 信息嵌入映射 H: $H(X)=(X',\ C)$;

2) 信息提取映射 E: $E(X'',\ C)=X$;

3) 若 $X'=X''$, 则 $E(X'',\ C)=E(X',\ C)=E(H(X))=X$ 成立.

值得注意的是, 当节点状态数据集合 X 为有限集时, 可根据上述信息隐藏安全模型中构建的双射映射 $(H,\ E)$ 的集合元素一一对应的特性, 来验证通信链路传输中有无噪声干扰和数据错误, 在一定程度上保证了发送信息与接收信息无误差.

1.3　Hadamard 变换

Hadamard 变换是在信号处理领域通常用来实现真实值在虚假值中进行隐藏和提取的一种有效方法. Hadamard 变换矩阵为一个方阵, 每个元素只能是 1 或 –1, 且每行都是互相正交的. 一个 n 阶的 Hadamard 矩阵 H 满足: $HH^{\text{T}}=nI_n$, 其中 I_n 是 $n\times n$ 的单位矩阵.

数学家詹姆斯 · 西尔维斯特给出了一种 Hadamard 变换矩阵最初的构造方法. 假设 H_n 是一个 n 阶的 Hadamard 矩阵, 则根据下式可以构造出一个 $2\,n$ 阶的 Hadamard 矩阵:

$$H_{2n}=\begin{bmatrix}H_n & H_n\\ H_n & -H_n\end{bmatrix}$$

令初始矩阵为:

$$H_1=[1]$$

采用上述西尔维斯特的方法, 就可以得到以下一系列矩阵:

$$H_2=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

$$H_4=\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -1 & 1 & -1\\ 1 & 1 & -1 & -1\\ 1 & -1 & -1 & 1\end{bmatrix}$$

$$\cdots$$

由此, 可以成功构造任何 2^k 阶 Hadamard 矩阵, 其中 k 为非负整数. 此外, 高阶的 Hadamard 变换矩阵可由两个低阶矩阵的 Kronecker 积求得, 即:

$$H_{2^k}=H_2\otimes H_{2^{k-1}}=\begin{bmatrix}H_{2^{k-1}} & H_{2^{k-1}}\\ H_{2^{k-1}} & -H_{2^{k-1}}\end{bmatrix}\tag{1}$$

其中符号“$\otimes$”表示矩阵的 Kronecker 积. 若将长度为 2^k 的实数向量 $x=[x_1,\ x_2,\ \cdots,\ x_{2^k}]$ 作为输入, 令 H_{2^k} 为 2^k 阶 Hadamard 变换矩阵, 那么对应的 Hadamard 变换与逆变换操作如下:

$$X=H_{2^k}x\tag{2}$$

$$x=\frac{1}{2^k}H_{2^k}^{\text{T}}X\tag{3}$$

2　问题描述

考虑一个由 n 个智能体组成的 MAN, 其通信拓扑结构可以抽象为一个有向加权图 $\mathcal{G}=\{\mathcal{V},\ \mathcal{E},\ \omega\}$ 来表示. 每个智能体的动态性能为:

$$x_i(t+1)=Ax_i(t)+Bu_i(t),\quad i=1,\ 2,\ \cdots,\ n\tag{4}$$

其中, A 和 B 是维度相容的常数矩阵, $x_i(t)$ 和 $u_i(t)$ 分别是节点 v_i 的状态变量和控制变量. 本文遵循与文献 [26] 相同的离散时间一致性协议, 即节点 v_i 的状态更新方程为:

$$x_i(t+1) = x_i(t) + \epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} W_{ij}\left(x_j(t) - x_i(t)\right) \quad (5)$$

其中 $x_i(t) \in \mathbf{R}$ 表示节点 v_i 在 t 时刻的状态值, $\epsilon \in (0,\ \frac{1}{\Delta})$, Δ 表示节点的最大度数. 将其写成状态空间表达形式为:

$$x(t+1) = Px(t) \quad (6)$$

其中 $x = [x_1,\ \cdots,\ x_n]^{\mathrm{T}}$, $P = I - \epsilon L$, 且 $L = [l_{ij}] \in \mathbf{R}^{n\times n}$ 表示拉普拉斯矩阵. 对于本文考虑的图 $\mathcal{G} = \{\mathcal{V},\ \mathcal{E},\ \omega\}$, 我们做出如下假设.

假设 1. 假设图 $\mathcal{G}$ 是强连通且平衡的. 换句话说, 对于任意 $v_i \neq v_j \in \mathcal{V}$, 从节点 v_i 到节点 v_j 至少存在一条有向路径, 且图中所有节点的出入度相等.

从文献 [26] 可知, 在满足假设 1, 且 $\epsilon \in (0,\ \frac{1}{\Delta})$, 系统中节点采用上述更新规则 (5) 下, 最后所有状态值可以取得平均一致. 即对于任意节点 v_i, 随着时间的推演, 其状态值收敛到所有节点初始值的平均值:

$$\lim_{t\to\infty} x_i(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(0) \quad (7)$$

2.1 窃听者模型

在 MAN 协同控制的研究工作中, 设计者通常利用节点的本地信息设计分布式控制协议, 而其中节点的状态信息需要在相邻智能体之间进行交换. 这种分布式的通过无线或有线通信渠道进行的信息交换和更新为窃听者提供了获取个体信息的机会. 如果窃听者记录了每次迭代中传输的所有状态信息, 那么节点的身份和其对应最初的状态值将被完全公开. 更有甚者, 他们可以通过窃取通信链路上的信息掌握节点内部控制机理并据此对受控系统实施精准攻击.

现有文献中对窃听者的模型大体可以分为 2 类: 内部窃听者[9, 18–20] 和外部窃听者[12, 16, 27–28]. 内部窃听者, 也被称为好奇节点, 是指网络中遵循控制协议执行状态更新, 但同时试图通过接收到的信息计算推测邻居节点初始状态值的一类智能体节点. 而外部窃听者是指能够获取整个 MAN 拓扑结构信息 (包括节点间链路权重) 的窃听者, 同时这类窃听者也能够获取多个节点间通信链路上所交换的信息.

显然, 相较于只收集局部信息的内部窃听者, 外部窃听者收集掌握更多的系统信息, 它对于网络隐私的窃取能力也远大于内部窃听者. 因此, 为最大化考验本文所设计算法的隐私保护能力, 在窃听者模型上本文将重点考虑外部窃听者. 同时, 更进一步, 在现有文献的基础上, 根据是否知晓 MAN 内部采用的隐私保护机制, 本文又将外部窃听者分为 I 型窃听者和 II 型窃听者. 具体的定义分别如下:

定义 1 (I 型窃听者). 这类窃听者是现有文献中描述的外部窃听者, 他们通过黑客技术入侵系统, 掌握节点的一致性更新协议、网络通信拓扑及对应的链路权重, 据此推断被窃听网络中节点的初始状态值.

定义 2 (II 型窃听者). II 型窃听者不仅拥有 I 型窃听者所有能力, 同时知晓被窃听系统内部采用的隐私保护策略, 从而将采取针对性措施来推断节点的真实状态信息.

由于基于 MAN 技术的多数实际应用, 如无人机集群、多机器人系统、车联网等, 通常部署在户外, 依靠开放和共享的无线网络传输数据, 致使外部窃听者很容易截获网络中共享的数据. 而对于获取数据后, 窃听者是否知晓数据背后的隐私保护策略, 现有大多数文献中, 作者并不明确, 而是通过直接构建观测器进行隐私性能分析. 有别于此, 本文在后续的分析讨论中, 针对 I 型窃听者, 对所提算法采用与目前其他文献一致的方法, 即构建观测器的方法进行隐私性能分析, 而对于 II 型窃听者, 则通过与之对应的隐私破解方法进行分析.

3 主要结果

为便于后续控制算法的描述与理解, 首先对算法中将会用到的干扰信号矩阵 Q 进行定义.

定义 3 (干扰信号矩阵 Q). 每个节点 v_i 根据 $\mathcal{N}_i^{\text{out}}$ 产生一组随机数 $q_i = [q_i^i,\ q_i^j]^{\mathrm{T}}$, $j = 1,\ 2,\ \cdots,\ n$, 满足:

1) 如果节点 $j \in \mathcal{N}_i^{\text{out}}$, $q_i^j \in (0,\ 1)$;

2) 如果节点 $j \notin \mathcal{N}_i^{\text{out}}$, $q_i^j = 0$;

3) $q_i^i = -\sum_{j\in\mathcal{N}_i^{\text{out}}} q_i^j m$, $m \in \mathbf{Z}$ 且 $m \neq 0$ 或 1.

将所有节点产生的随机数 q_i 按列依次放入干扰信号矩阵 Q, 满足 $Q = [q_i] \in \mathbf{R}^{n\times n}$.

本文中, 我们给出如下假设:

假设 2. 考虑的 MAN 在初始时刻的网络环境是安全的. 即在 $t = 0$ 时, 外部窃听者未开始实施窃听攻击.

假设 3. 网络中 Hadamard 矩阵的初始矩阵和根据通信频次转化 Hadamard 矩阵的变化次数是先验全局已知的.

3.1 基于隐写术的平均一致性算法设计

为了实现窃听攻击下具备隐私保护的 MAN 一

致性控制, 本文提出一种基于隐写术的分布式平均一致性算法. 该算法的基本示意图如图 1 所示.

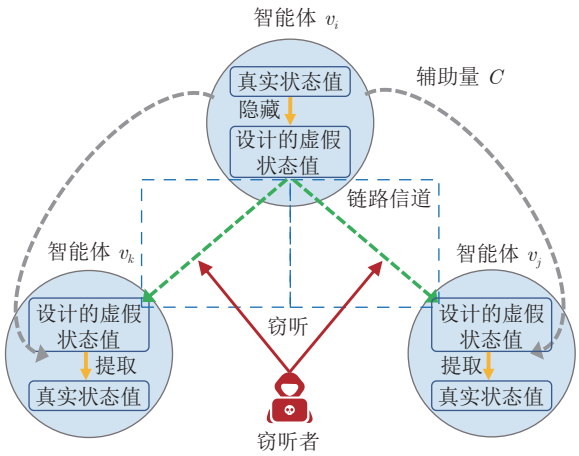


图 1　基于隐写术的一致性算法示意图

Fig. 1　Schematic diagram of consensus algorithm based on steganography

首先, 网络中每个节点在第一次信息发送前产生一组干扰信号, 每组干扰信号由产生节点私有. 不失一般性, 将节点 v_i 第一次发送给邻居 v_j 的消息表示为 $y_i^j(0)$, q_i^j 为 v_i 向 v_j 传递初始信息时节点 v_i 使用的干扰信号, q_i^i 则是 v_i 用于产生虚假初值的信号. 随后, 每个节点根据詹姆斯 · 西尔维斯特的 Hadamard 矩阵构造方法, 产生多重 Hadamard 变换与逆变换因子, 作为其真实状态值的隐藏–提取规则.

网络中的每个节点既作为信息接收方又作为信息发送方, 作为发送方利用隐藏规则将真实状态值 $x_i(t)$ 写入到虚假状态值 $x_i'(t)$ 中, 并在通信链路中发送该虚假状态值给邻居节点; 作为接收方则通过提取规则从虚假状态值中提取出真实状态值 $x_i(t)$, 再利用更新规则分别对提取得到的真实状态值和原始虚假状态值进行更新. 每一轮持续上述方法直至算法收敛. 所提出算法的具体步骤如算法 1 所示.

算法 1. 基于隐写术的平均一致性算法

输入. 节点 v_i 初始状态值 $x_i(0)$, $i = 1, 2, \cdots, n$.

步骤 1. 初始化阶段. 当 $t = 0$ 时:

1) 节点 v_i 产生一组随机数 q_i^j, 并计算 q_i^i 使其满足定义 3.

2) 生成虚假初始状态. 节点 v_i 根据拉普拉斯矩阵和干扰矩阵生成的信号 q_i^j 按照下式生成中间值 $y_i^j(0)$, 并将其发送给邻居节点 v_j

$$y_i^j(0) = -l_{ji}x_i(0) + q_i^j \tag{8}$$

节点 v_i 接收相邻节点发送的信息后, 按下式生成虚假初始值 $x_i'(0)$:

$$x_i'(0) = x_i(0) + \sum_{j\in\mathcal{N}_i^{\text{in}}} \left(y_j^i(0) + l_{ij}x_i(0)\right) + q_i^i \tag{9}$$

步骤 2. 迭代更新阶段. 当 $t = 0, 1, 2, \cdots$

3) 发送方 v_i 按下式得到虚假值和真实值的差值 $e_i(t)$:

$$e_i(t) = x_i(t) - x_i'(t) \tag{10}$$

采用 Hadamard 变换对差值 $e_i(t)$ 进行隐藏: 每个节点随机生成一个正整数 $k_i(t)$, 对应得到变换次数 $2^{k_i(t)}$, 生成 $2^{k_i(t)}$ 阶 Hadamard 矩阵 $H_{2^{k_i(t)}}$. 将差值 $e_i(t)$ 分成 $2^{k_i(t)}$ 等份并生成向量维度为 $2^{k_i(t)}$ 的实数向量 $\bar{x}_i(t) = [e_i(t)/2^{k_i(t)}, \cdots, e_i(t)/2^{k_i(t)}]$, 再通过 Hadamard 变换公式进行 $2^{k_i(t)}$ 次变换得到:

$$z_i(t) = \left(H_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \bar{x}_i^{\mathrm{T}}(t) \tag{11}$$

最后, 将变换结果 $z_i(t)$ 和虚假值 $x_i'(t)$ 以频次为 $k_i(t)$ 的形式发送给邻居节点.

4) 接收方 v_j 冗余收到的数据 $(z_i(t), x_i'(t))$, 分析频次得到变换次数 $2^{k_i(t)}$. 并据此生成 $2^{k_i(t)}$ 阶 Hadamard 矩阵 $H_{2^{k_i(t)}}$, 再通过逆变换公式进行 $2^{k_i(t)}$ 次逆变换求得实数向量 $\left[e_i(t)/2^{k_i(t)}, \cdots, e_i(t)/2^{k_i(t)}\right]$:

$$\bar{x}_i^{\mathrm{T}}(t) = \left(\frac{1}{2^{k_i(t)}} H_{2^{k_i(t)}}^{\mathrm{T}}\right)^{2^{k_i(t)}} z_i(t) \tag{12}$$

$$\mathbf{1}_n \bar{x}_i^{\mathrm{T}}(t) = e_i(t) \tag{13}$$

其中 $\mathbf{1}_n = [1, 1, \cdots, 1]$. 最后通过式 (10) 即可得到 v_i 真实状态值 $x_i(t)$.

5) 所有节点分别根据状态更新方程对真实状态值 $x_i(t)$ 和虚假状态值 $x_i'(t)$ 进行更新, 得到各自下一时刻的状态值, 即:

$$x_i(t+1) = x_i(t) + \epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} W_{ij}\left(x_j(t) - x_i(t)\right) \tag{14}$$

$$x_i'(t+1) = x_i'(t) + \epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} W_{ij}\left(x_j'(t) - x_i'(t)\right) \tag{15}$$

6) 重复以上迭代过程, 直至节点状态值收敛.

值得注意的是, 在上述算法中, 节点同步执行了 2 种一致性更新过程, 即分别以真实状态值作为输入的更新和以虚假状态值作为输入的更新. 目的是混淆窃听者, 使得窃听者通过链路数据得到表层虚假状态值作为输入的一致性进程, 从而隐藏保护了真实状态值一致性收敛的全过程.

注 1. 在上述算法步骤 2 中, 每个节点随机生成了任意正整数 $k_i(t)$, 用于 Hadamard 矩阵的计算. 而在实际应用中, 考虑智能体节点的计算和通信资源受限问题, 需要给随机数 $k_i(t)$ 设置一个上界值, 以保证节点计算和通信的效率.

根据算法 1, 通过对算法中步骤 1 和步骤 2 的参数设计, 将使得系统具备以下性质.

定理 1. 设计者可以通过设置干扰信号矩阵 Q 中的变量 m, 控制初始虚假值均值与初始真实值均值的偏离程度 N.

证明. 令 $x(t) = [x_1(t),\ x_2(t),\ \cdots,\ x_n(t)]^{\mathrm{T}}$, $x'(t) = [x'_1(t),\ x'_2(t),\ \cdots,\ x'_n(t)]^{\mathrm{T}}$. 根据式 (8) 和 (9) 得到:

$$x'(0) = x(0) - Lx(0) + Q^{\mathrm{T}}\mathbf{1}_n^{\mathrm{T}} \tag{16}$$

其中, L 为拉普拉斯矩阵, Q 为干扰信号矩阵, $\mathbf{1}_n = [1,\ 1,\ \cdots,\ 1]$. 进一步对式 (16) 左乘 $\mathbf{1}_n$, 得:

$$\mathbf{1}_n x'(0) = \mathbf{1}_n x(0) + \mathbf{1}_n Q^{\mathrm{T}}\mathbf{1}_n^{\mathrm{T}} \tag{17}$$

上式中, $\mathbf{1}_n x(0)$ 即为真实初值的总和, $\mathbf{1}_n x'(0)$ 为虚假初值的总和. 计算式 (17), 得出虚假值初值总和与真实值初值总和的差值为:

$$\mathbf{1}_n x'(0) - \mathbf{1}_n x(0) = \mathbf{1}_n Q^{\mathrm{T}}\mathbf{1}_n^{\mathrm{T}} = \frac{m-1}{m}\sum_{i=1}^{n} q_i^i$$

从而

$$N = \frac{\mathbf{1}_n x(0) - \mathbf{1}_n x'(0)}{n} = \frac{\mathbf{1}_n Q^{\mathrm{T}}\mathbf{1}_n^{\mathrm{T}}}{n} = \frac{(m-1)\sum\limits_{i=1}^{n} q_i^i}{mn}$$

即说明了设计者可以通过调整 m 的值, 对偏离程度 N 进行控制. □

上述定理指明, 本文提出的基于隐写术的隐私保护方法, 在保护节点初始状态信息隐私的同时, 也可以对节点间交互的信息进行设计, 用以误导窃听者.

3.2 系统收敛性能分析

本节对系统的收敛性进行分析. 本质上, 算法 1 通过在传统更新协议 (5) 中融入隐写术, 即结合隐藏–提取机制来实现节点信息的隐私保护. 针对 MAN 设计的分布式控制协议, 可得如下结论.

定理 2. 考虑一阶离散时间 MAN 式 (4). 如果假设 1 ~ 3 成立且网络中所有节点按照算法 1 进行信息处理. 网络中所有节点则可以通过邻居间交换虚假状态值信息, 使得其真实状态值最终达成精确的平均一致.

证明. 根据算法 1 采用的隐藏–提取规则可知, 对于网络中任意节点 v_i, 先将真实状态值和生成的虚假状态值的差值 $e_i(t)$ 进行 $2^{k_i(t)}$ 等分, 转化为长度为 $2^{k_i(t)}$ 的实数向量 $\bar{x}_i(t)$, 再进行多重 Hadamard 变换并将最终隐藏结果发送给邻居节点. 其中变换矩阵为 $2^{k_i(t)}$ 阶 Hadamard 矩阵, 变换次数为 $2^{k_i(t)}$. 此时, 得到状态变换结果 $z_i(t)$ 为

$$z_i(t) = \left(H_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \bar{x}_i^{\mathrm{T}}(t) \tag{18}$$

接收方收到邻居节点 v_i 发送的虚假值信息 $x'_i(t)$. 根据信息接收冗余量 (或接收频次), 得到信息隐藏安全模型六元组 Θ 中的辅助值 C, 即逆变换次数 $2^{k_i(t)}$.

根据 n 阶 Hadamard 矩阵 H 的性质: $HH^{\mathrm{T}} = nI_n$, 有 $(1/2^{k_i(t)})H^{\mathrm{T}}_{2^{k_i(t)}}H_{2^{k_i(t)}} = I_{2^{k_i(t)}}$. 再将来自 v_i 的 $x'_i(t)$ 代入逆变换式 (3) 进行还原得到实数向量 $\bar{x}_i(t)$:

$$\begin{aligned}
&\left(\frac{1}{2^{k_i(t)}}H^{\mathrm{T}}_{2^{k_i(t)}}\right)^{2^{k_i(t)}} x'_i(t) = \\
&\left(\frac{1}{2^{k_i(t)}}H^{\mathrm{T}}_{2^{k_i(t)}}\right)^{2^{k_i(t)}} z_i(t) = \\
&\left(\frac{1}{2^{k_i(t)}}H^{\mathrm{T}}_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \left(H_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \bar{x}_i^{\mathrm{T}}(t) = \\
&\left(\frac{1}{2^{k_i(t)}}H^{\mathrm{T}}_{2^{k_i(t)}}H_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \bar{x}_i^{\mathrm{T}}(t) = \\
&\left(I_{2^{k_i(t)}}\right)^{2^{k_i(t)}} \bar{x}_i^{\mathrm{T}}(t) = \bar{x}_i^{\mathrm{T}}(t)
\end{aligned} \tag{19}$$

最后, 接收方通过式 (10) 即可求得节点 v_i 的真实状态值 $x_i(t)$. 接下来, 可以根据网络拓扑条件假设 1 和节点信息, 采用文献 [26] 中的分析证明方法, 得出系统最终可以实现精确的平均一致. □

3.3 隐私保护性能分析

本节考虑存在外部窃听者的情形下, 对所建立网络的隐私保护性能进行分析. 本文沿用文献 [5] 中对节点初始值隐私保护的定义, 具体如下:

定义 4 (节点初始状态值的隐私保护)[5]. 对于一个 MAN, 如果窃听者无法以任何确定的精度估计网络中节点 v_i 的初始状态信息 $x_i(0)$, 则称 v_i 的初始状态值具备隐私保护.

定理 3. 考虑存在 I 型窃听者情形下的 MAN 式 (4). 如果假设 1 ~ 3 成立且网络中所有节点按照算法 1 进行状态值更新, 则该系统的虚假初始信息可以被 I 型窃听者得到, 即窃听者可以推断得到网络中任何一个节点的初始状态值 $x'_i(0)$.

证明. 根据 I 型窃听者的定义, 它能够获取整个网络拓扑结构信息和节点的更新协议, 同时它能够获取节点间通信链路上所传输的所有数据. 不失一般性, 假设网络中任意节点 v_i 的初始状态值为该窃听者的窃听目标. 针对 v_i, 窃听者在该网络中窃听得到的所有信息可以表示为:

$$I_i = \bigcup_{t=1}^{T} \{x_i'(t),\, W_{ij},\, \epsilon \mid \forall v_i,\, v_j \in \mathcal{V}\}$$

窃听者在任意时刻 t_a, $t_a \in [1,\, T]$, 收集得到 $1 \sim t_a$ 时刻所有链路发送的消息. 根据收集的信息和节点 v_i 更新协议, 窃听者可以按下式逐步推测每个节点的初始值:

$$\begin{aligned} x_i'(t_a) = {} & x_i'(t_a - 1) + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(t_a - 1) - x_i'(t_a - 1)\right) = \\ & x_i'(t_a - 2) + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(t_a - 2) - x_i'(t_a - 2)\right) + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(t_a - 1) - x_i'(t_a - 1)\right) = \\ & x_i'(0) + \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(0) - x_i'(0)\right) + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(1) - x_i'(1)\right) + \cdots + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(t_a - 1) - x_i'(t_a - 1)\right) \end{aligned} \tag{20}$$

其中, $x_i'(t_a)$ 表示在 t_a 时刻窃听得到的节点 v_i 的传输值. 令

$$\begin{aligned} & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(1) - x_i'(1)\right) + \cdots + \\ & \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(t_a - 1) - x_i'(t_a - 1)\right) = T_i \end{aligned}$$

将上式进行整理可得:

$$x_i'(t_a) = x_i'(0) + \epsilon \sum_{v_j \in \mathcal{N}_i^{\text{in}}} W_{ij} \left(x_j'(0) - x_i'(0)\right) + T_i \tag{21}$$

对于每个节点 v_i, $i = 1,\, 2,\, \cdots,\, n$, 代入式 (21), 联立得:

$$\begin{aligned} & x_1'(t_a) = x_1'(0) + \epsilon \sum_{v_j \in \mathcal{N}_1^{\text{in}}} W_{1j} \left(x_j'(0) - x_1'(0)\right) + T_1 \\ & x_2'(t_a) = x_2'(0) + \epsilon \sum_{v_j \in \mathcal{N}_2^{\text{in}}} W_{2j} \left(x_j'(0) - x_2'(0)\right) + T_2 \\ & \qquad \vdots \\ & x_n'(t_a) = x_n'(0) + \epsilon \sum_{v_j \in \mathcal{N}_n^{\text{in}}} W_{nj} \left(x_j'(0) - x_n'(0)\right) + T_n \end{aligned} \tag{22}$$

由于窃听者已收集了 $1 \sim t_a$ 时刻所有链路发送的状态值 $x_i'(t)$, 据此 T_i 可以通过计算得知. 因此, 上述联立式中, 未知数个数与方程式个数相等, 均为 n. 由此窃听者即可通过求解上述联立式, 得出任意节点的初始状态值 $x_i'(0)$. □

值得注意的是, 由于 I 型窃听者并不知晓系统采用了隐私保护策略, 仅根据网络中传输的信息和一致性更新协议 (5), 最终推断出的初始值是系统根据算法 1 产生的虚假初始状态值.

相较于 I 型窃听者, II 型窃听者进一步知晓了系统的隐私保护策略. 在对系统初始值的窃取过程中, II 型窃听者因为知晓算法 1 的规则, 它有两种方法对网络中节点初始值进行推断. 第一种方法是在 $t = 0$ 初始化阶段, 直接从算法产生的虚假初始值中推断出真实初始值; 第二种方法则是通过对整个时间段 T 的窃听, 尝试将隐藏在虚假状态值中的真实状态值进行还原, 进而再根据定理 3 中 I 型窃听者的计算方法, 求解得到节点初始时刻的真实状态值.

定理 4. 考虑存在 II 型窃听者情形下的 MAN 式 (4). 如果假设 1 ~ 3 成立且网络中所有节点按照算法 1 进行状态更新, 则该系统真实初始信息是无法被 II 型窃听者获取的, 即窃听者无法得知网络中任何一个节点的真实初始状态值 $x_i(0)$.

证明. 不失一般性, 假设窃听者的目标为网络中任意节点 v_i 的初始值信息 $x_i(0)$. 网络中所有节点按照算法 1 进行信息交互处理. 在 $0 \sim T$ 时间段内, 窃听者针对节点 v_i 收集得到的信息集合 I_i 为

$$I_i = \bigcup_{t=0}^{T} \left\{I_i^{\text{send}}\,(t) \cup I_i^{\text{receive}}\,(t)\right\}$$

其中, $I_i^{\text{send}}(t)$ 和 $I_i^{\text{receive}}(t)$ 分别表示节点 v_i 在 t 时刻发送和接收的消息集合. 具体为

$$\begin{aligned} & I_i^{\text{send}}\,(t) = \left\{y_i^m(0),\, x_i'(t),\, W_{im},\, \epsilon \mid m \in \mathcal{N}_i^{\text{out}}\right\} \\ & I_i^{\text{receive}}\,(t) = \left\{y_n^i(0),\, x_n'(t),\, W_{ni},\, \epsilon \mid n \in \mathcal{N}_i^{\text{in}}\right\} \end{aligned}$$

首先, 证明第一种方法对本文算法的无效性. 考虑在初始化阶段, 即 $t = 0$ 时刻, 窃听者对链路中发送的消息进行收集并分析. 根据式 (8), 有

$$x_i(0) = -\frac{y_i^m(0) - q_i^m}{l_{mi}} \tag{23}$$

尽管窃听者通过已知的网络拓扑结构, 求得了拉普拉斯矩阵中的值 l_{mi}, 同时, 窃听获取了链路中传输值 $y_i^m(0)$. 但基于所设计的干扰矩阵 Q, 我们知道干扰信号 q_i^m 是由节点 v_i 为其出邻居 v_m 唯一生成的, 且为节点 v_i 内部私有, 并不在网络中参与交互. 窃听者无法获知式 (23) 中的 q_i^m, 进而无法推断节点初始状态值 $x_i(0)$. 类似地, 用同样的方法分析

窃听者仅根据节点 v_i 接收到的信息 $y_n^i(0)$, 同样可以得出窃听者无法推断节点 v_i 的初始状态值.

接下来, 证明第二种方法对本文提出算法的无效性. 根据定义, II 型窃听者获知了系统隐私保护机制, 即知晓系统采用了 Hadamard 变换与逆变换来进行真实状态值的隐藏和提取操作. 尽管在 $0 \sim T$ 时间段内, 窃听者收集得到了网络通信链路中所有的交互信息以及通信拓扑信息, 但是所构建的信息隐藏安全模型六元组 Θ 中的辅助值 $C = 2^{k_i(t)}$ 并未出现在信息交互数据中. 而是通过带外信息 (即频次) 的方式告知邻居节点.

所以基于现有数据, 窃听者即使掌握了隐藏和提取映射 H 和 E, 因 C 的缺失, 也无法在信息中有效提取隐藏数据, 即无法从虚假值中提取真实状态值 $x_i(t)$, 进而无法根据定理 3 中 I 型窃听者的计算方法, 求解得到节点初始时刻的真实状态值 $x_i(0)$.

综上, 窃听者无论采用上述两种方法中的哪一种, 在本文提出的算法框架下, 均无法获得节点的真实初始状态值. □

尽管定理 4 只给出了网络中节点初始状态信息可以得到有效保护的证明, 不难发现, 由于窃听者无法获取辅助量 C, 分析算法步骤, 其同样无法推断后续时间序列的状态值信息.

注 2. 值得注意的是, 如果考虑内部诚实且好奇节点, 本文所提算法的有效性将取决于文中的假设 3 是否对这类节点有效. 如果好奇节点同样具备假设 3 中的能力, 那么显然它可以通过提取映射获得邻居节点的真实状态.

3.4 隐私泄露模型与最优窃听攻击设计

本节尝试对 MAN 隐私泄露程度进行分析建模, 并构建了一种面向窃听攻击的成本函数. 随后, 从攻击者的角度, 评估网络中节点的隐私泄露程度, 并以此寻求最优效益下的窃听概率分配策略.

为了度量节点的隐私泄露程度, 受文献 [28] 的启发, 本文引入概率指标 ψ_i, 表示节点 v_i 的所有出边中至少有一条边遭遇窃听的概率, 并将 ψ_i 确立为节点 v_i 的隐私泄露值.

定义窃听矩阵 $\Phi \in \mathbf{R}^{n\times n}$, 其表达式为

$$\Phi[i,\ j] = \begin{cases} \lambda_{ij}, & (v_i,\ v_j) \in \mathcal{E} \\ 0, & \text{否则} \end{cases}$$

其中 $\lambda_{ij} \in [0,\ 1]$ 为链路 ε_{ij} 被窃听的概率. 定义

$$W_i = 1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1}^{n} (1 - \Phi[t,\ m]) \tag{24}$$

为节点 v_i 的所有出邻居节点的入边至少有一条遭遇窃听的概率. 相应地, 定义

$$V_i = 1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1,\ t\neq i}^{n} (1 - \Phi[t,\ m]) \tag{25}$$

为节点 v_i 的所有出邻居节点的入边除 v_i 自身外至少有一条遭遇窃听的概率.

根据上述定义和窃听行为建模, 可以得到以下结论.

定理 5. 考虑 MAN 式 (4) 和窃听矩阵 Φ. 网络中各节点 v_i, $i = 1,\ 2,\ \cdots,\ n$, 的隐私泄露度量值 ψ_i 可通过下式获得

$$\psi_i = 1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) \tag{26}$$

证明. 首先, 将式 (24) 中 $t = i$ 部分拆分出来, 可得

$$\begin{aligned} &W_i = 1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1}^{n} (1 - \Phi[t,\ m]) = \\ &1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \left(\prod_{t=1,\ t\neq i}^{n} (1 - \Phi[t,\ m])(1 - \Phi[i,\ m]) \right) = \\ &1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1,\ t\neq i}^{n} (1 - \Phi[t,\ m]) \prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) \end{aligned} \tag{27}$$

进而, 根据 W_i 和 V_i 的定义, 有

$$\begin{aligned} &\psi_i = W_i - V_i \left(\prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) \right) = \\ &1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1,\ t\neq i}^{n} (1 - \Phi[t,\ m]) \prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) - \\ &\left[1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} \prod_{t=1,\ t\neq i}^{n} (1 - \Phi[t,\ m]) \right] \times \\ &\left(\prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) \right) = \\ &1 - \prod_{m\in\mathcal{N}_i^{\text{out}}} (1 - \Phi[i,\ m]) \end{aligned}$$

□

对节点链路窃听需要消耗窃听者一定的资源, 站在窃听者的角度, 要获取网络中单个节点的隐私泄露值 ψ_i, 对应需付出的成本定义为 $C_i(\psi_i)$. 考虑窃听者的资源有限, 网络窃听代价通常用成本函数来刻画, 成本函数是一个递增的严格凸函数, 两种常用的成本函数是分段线性函数和二次函数. 不失

一般性, 本文采用如下二次函数作为窃听成本函数:

$$C_i\left(\psi_i\right)=\alpha_i\psi_i^2+\beta_i\psi_i+\gamma_i \tag{28}$$

其中, α_i 、β_i 、γ_i 为可根据实际场景设定的中间参数, ψ_i 为节点的隐私泄露值. 窃听者对网络进行窃听的总成本 μ 为对所有单个节点窃听成本的累加值, 即

$$\mu=\sum_{i=1}^{n}C_i\left(\psi_i\right) \tag{29}$$

最后, 从攻击者的角度, 根据最小窃听成本要求, 在满足为所有节点分配的窃听概率的和为 1 条件下, 提出以下最优分配策略的优化问题:

$$\begin{aligned}&\min\mu=\sum_{i=1}^{n}C_i(\psi_i)\\&\text{s.t.}\quad\psi_1+\psi_2+\cdots+\psi_n=1,\\&\qquad 0\le\psi_i\le 1,\qquad i=1,\,2,\,\cdots,\,n\end{aligned} \tag{30}$$

根据窃听者对网络中各个节点的窃听投入分配与调整, 给出一种基于 Lambda 迭代的优化方法进行求解问题 (30). 具体地, 首先根据各个节点连接链路的数目进行排序, 节点 v_i 在重新降序排序后的排名用 $\text{rank}(v_i)$ 表示, 确定每个节点不同的成本曲线, 使得节点的出链路越多, 分配的窃听概率越大. 给定总窃听概率 $\sum_{i=1}^{n}\psi_i=1$, 初始化后进行迭代, 计算灵敏度 s_i, 并根据拉格朗日乘数和灵敏度计算每个节点的最小成本窃听概率 $\psi_{i_{\text{new}}}$; 判定窃听概率是否超出了最大界值或最小界值, 若超出则将该值设置为对应最新的最大界值或最小界值; 更新总窃听概率 ψ_{total} 和总成本 μ; 根据更新的拉格朗日乘数 λ, 判断是否达到收敛条件, 如果达到则跳出循环, 否则继续迭代; 最后输出结果, 包括每个节点分配的窃听概率和总成本. 所提算法的具体步骤如算法 2 所示.

算法 2. 基于 Lambda 迭代的优化算法

输入. 优化目标 (30).

输出. 节点的窃听概率 $\psi_{i_{\text{new}}}$, $i=1,\,2,\,\cdots,\,n$, 窃听者总成本 μ.

步骤 1. 初始化阶段. 当 $k=0$ 时:

1) 对网络中所有节点的出链路个数进行降序排列, 且从大到小进行重新标记. 令 $\text{rank}(v_i)$ 表示节点 v_i 在重新降序排序后的排名, 据下式构建每个节点的成本曲线:

$$C_i\left(\psi_i\right)=\psi_i^2+\frac{\text{rank}\left(v_i\right)}{n}\psi_i \tag{31}$$

2) 设置初始化总窃听概率 $\psi_{\text{total}}=\sum_{i=1}^{n}\psi_i=1$; 拉格朗日乘数 $\lambda=0$, 迭代次数 max_{iter} 和容许误差 ε.

步骤 2. 更新迭代阶段. 当 $k=0,\,1,\,\cdots,\,max_{\text{iter}}$ 时:

3) 计算灵敏度: $s_i=\frac{\partial C_i(\psi_i)}{\partial\psi_i}$;

4) 计算最小成本窃听概率: $\psi_{i_{\text{new}}}=\frac{\lambda-s_i}{2}$;

5) 检查窃听概率是否超过了上下界值. 如果 $\psi_{i_{\text{new}}}<0$, 则令 $\psi_{i_{\text{new}}}=0$; 如果 $\psi_{i_{\text{new}}}>1$, 则令 $\psi_{i_{\text{new}}}=1$;

6) 更新总窃听概率: $\psi_{\text{total}}=\sum_{i=1}^{n}\psi_{i_{\text{new}}}$;

7) 更新总成本: $\mu=\sum_{i=1}^{n}C_i\left(\psi_{i_{\text{new}}}\right)$;

8) 更新拉格朗日函数: $\lambda=\lambda+0.1\times\left(1-\psi_{\text{total}}\right)$;

9) 判断是否达到收敛条件. 如果 $\left|\psi_{i_{\text{new}}}-\psi_i\right|<\varepsilon$, 则算法结束; 否则, 令 $k=k+1$, 回到进程 3).

4　数值仿真实验

考虑一个由 5 个智能体节点组成的 MAN, 其通信图如图 2 所示. 为方便起见, 将 2 个节点之间有向链路权值设置为 1, 否则为 0. 不难验证, 所给出的通信图满足强连通和平衡图要求, 即满足本文假设 1. 实例中, 随机赋予每个节点的初始状态值为: $x_1(0)=20$, $x_2(0)=18$, $x_3(0)=30$, $x_4(0)=48$, $x_5(0)=1$. 此时不难求出初始值均值 $\bar{x}(0)=23.4$. 网络中的每个节点在 $t=0$ 时刻开始执行本文提出的算法 1. 实例中, 将控制增益设置为 $\epsilon=0.1$, 干扰矩阵 Q 中设置变量 $m=20$, 根据定理 1, 对应产生虚假值与真实值最终偏离值为 7.6. 此时, 生成的虚假初始状态值分别为 $x_1'(0)=-3.75$, $x_2'(0)=25.25$, $x_3'(0)=9.5$, $x_4'(0)=-9.5$, $x_5'(0)=57.5$.

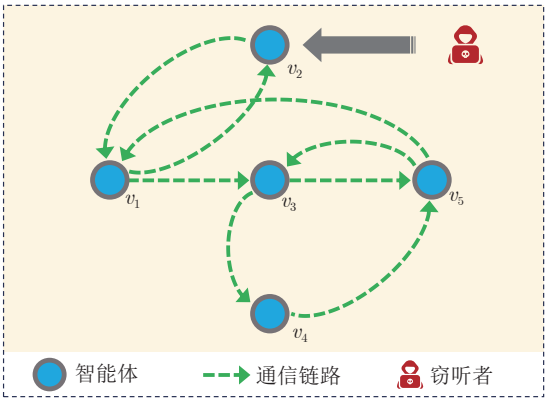


图 2　5 个节点组成的 MAN 通信图

Fig. 2　Communication graph of MAN with 5 nodes

网络中各个节点的真实状态值与虚假状态值变化轨迹如图 3 所示. 其中, 实线表示真实状态值的变化轨迹, 虚线表示各节点虚假状态值的变化轨迹, 加粗虚线 *R-avg* 和 *F-avg* 分别表示真实初始值的平均值和虚假初始值的平均值. 从实验结果可以看出, 节点在采用算法 1 的情况下, 它们的真实状态值和虚假状态值均能完成一致性收敛, 真实状态值收敛至精确的初始值均值, 而虚假状态值收敛至用户预

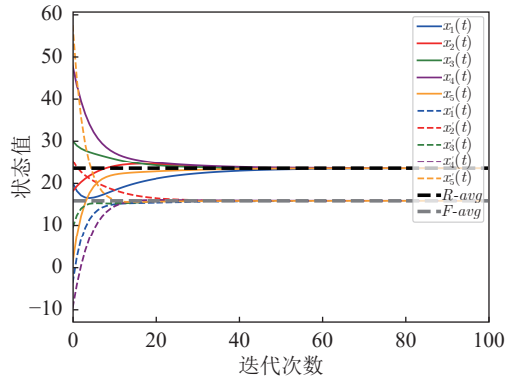


图 3 实例中 5 个节点的真实状态值和虚假状态值轨迹变化曲线

Fig. 3 The trajectory curves of the real and fake state value of 5 nodes in the instance

设的均值 15.8. 该实验结果验证了定理 1, 即可通过预设变量 m 的值来控制虚假值最终收敛状态的偏离程度. 同时也验证了定理 2, 即系统节点在算法 1 下, 所有节点的状态值最终实现平均一致.

接下来, 通过实例验证本文所提算法的隐私保护能力. 具体地, 在图 2 所示的 5 个节点组成的网络中, 考虑存在外部窃听者, 其目的是通过窃听收集到的信息推测节点 v_2 的真实初始状态 $x_2(0)$. 本实例中, 窃听者通过构建以下观测器来对节点 v_2 的初始状态值进行推测:

$$z(t+1)=z(t)+x_2^+(t+1)-\left(x_2^+(t)+\epsilon\sum_{v_j\in\mathcal{N}_2^{\rm in}}W_{2j}\times\left(x_j^+(t)-x_2^+(t)\right)\right)\tag{32}$$

观测器初始状态设置为节点 v_2 第一次发送的消息值, 即 $z(0)=x_2^+(0)$. 上述观测器的基本原理是基于累积的当前节点传输值与观测器预测更新值之间的差值调整估计结果. 窃听者一旦获得足够的传输信息, 通过这类观测器则可以成功破除基于添加人为噪声的隐私保护算法, 如文献 [5, 11–13, 29–30] 等中的算法, 得到节点的初始状态值.

作为与人为添加噪声方法的对比, 让图 2 中 5 个节点首先采用现有文献 [5] 中的隐私保护一致性算法, 而窃听者采用观测器 (32) 推测节点 v_2 的初始值. 实例中, 将随机噪声添入的步时设置为 40, 实验结果如图 4 所示. 时至 $t=40$, 观测器消除了之前累计 40 次算法迭代中添入的随机值的影响. 此时从图中可以看到, 窃听者推测值 (标星号虚线) 与真实初始值 (黑色虚线) 重合, 即窃听者通过上述观测器成功推断出节点 v_2 的初始状态值 $x_2(0)$.

随后, 让图 2 中 5 个节点采用本文提出的算法 1 进行状态值更新. 窃听者同样采用观测器 (32) 尝试推测节点 v_2 的初始值. 实验结果如图 5 所示. 图中标星号虚线为窃听者的推测值, 虽然窃听者通过观测器推测出了节点初始值, 但该数值为算法 1 处理后的载体信息 (虚假初始值 $x_2'(0)$), 并非节点的隐藏信息 (真实状态初值 $x_2(0)$). 上述结果验证了定理 3 和定理 4, 即通过采用隐写术机制, 节点 v_2 的真实初始值 (黑色虚线) 在本文算法下得到了隐私保护, 而其虚假状态值则被窃听者观测获取.

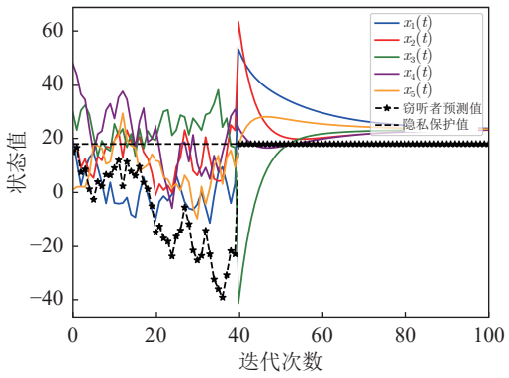


图 4 窃听者对采用文献 [5] 算法的节点 v_2 初始状态值的观测结果

Fig. 4 The observation results of eavesdropper on the initial state value of node v_2 using the algorithm in reference [5]

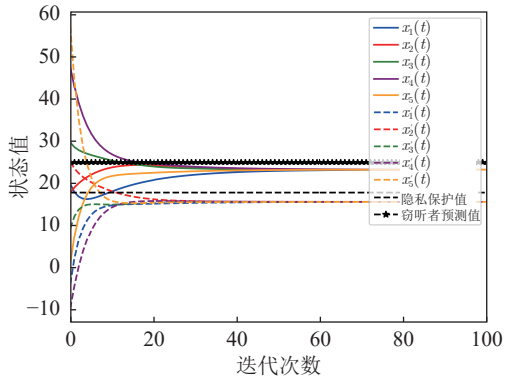


图 5 窃听者对采用本文算法的节点 v_2 初始状态值的观测结果

Fig. 5 The observation results of eavesdropper on the initial state value of node v_2 using the algorithm proposed in this paper

5 结束语

针对 MAN 一致性控制过程中节点信息隐私泄露问题, 本文提出了一种具备隐私保护的分布式平均一致性控制方法. 根据网络中窃听者的能力范围, 定义了两类隐私窃听攻击模型, 并在此基础上设计基于隐写术方法的一致性控制策略, 在确保节点初

始状态值隐私得到保护的同时, 实现了网络中所有节点的平均一致. 此外, 给出了一种有效衡量分布式 MAN 隐私泄露的度量模型, 并将其引入隐私窃听攻击实施问题中, 提出了一种基于 Lambda 迭代优化的最优窃听策略. 最后通过数值仿真实验验证了本文所提方法的有效性. 值得注意的是, 本文在窃听者攻击策略设计中, 针对的是全局网络的隐私保护问题, 而在实际应用环境中, 窃听者如果对于某个节点的隐私数据感兴趣, 针对该节点及其邻居节点的攻击资源分配问题同样值得考虑. 因此, 针对单个节点隐私泄露最优攻击策略的研究, 是未来进一步研究多智能体隐私保护的重要方向.

References

1 Zhao J N, Zhu K Y, Hu H J, Yu X, Li X W, Wang H S. Formation control of networked mobile robots with unknown reference orientation. *IEEE/ASME Transactions on Mechatronics*, 2023, **28**(4): 2200−2212

2 Gough M B, Santos S F, AlSkaif T, Javadi M S, Castro R, Catalao J P S. Preserving privacy of smart meter data in a smart grid environment. *IEEE Transactions on Industrial Informatics*, 2021, **18**(1): 707−718

3 Dong X, Hu G. Time-varying output formation for linear multiagent systems via dynamic output feedback control. *IEEE Transactions on Control of Network Systems*, 2017, **4**(2): 236−245

4 Xiao N, Wang X H, Xie L H, Wongpiromsarn T, Frazzoli E, Rus D. Road pricing design based on game theory and multiagent consensus. *IEEE/CAA Journal of Automatica Sinica*, 2014, **1**(1): 31−39

5 Mo Y, Murray R M. Privacy preserving average consensus. *IEEE Transactions on Automatic Control*, 2017, **62**(2): 753−765

6 Qin J, Ma Q, Shi Y, Wang L. Recent advances in consensus of multi-agent systems: A brief survey. *IEEE Transactions on Industrial Electronics*, 2017, **64**(6): 4972−4983

7 Zhu Jing, Wang Fei-Yue, Wang Ge, Tian Yong-Lin, Yuan Yong, Wang Xiao, et al. Federated control: A distributed control approach towards information security and rights protection. *Acta Automatica Sinica*, 2021, **47**(8): 1912−1920
(朱静, 王飞跃, 王戈, 田永林, 袁勇, 王晓, 等. 联邦控制: 面向信息安全和权益保护的分布式控制方法. 自动化学报, 2021, **47**(8): 1912−1920)

8 Ying Chen-Duo, Wu Yi-Ming, Xu Ming, Zheng Ning, He Xiong-Xiong. Privacy-preserving average consensus control formulti-agent systems under deception attacks. *Acta Automatica Sinica*, 2023, **49**(2): 425−436
(应晨铎, 伍益明, 徐明, 郑宁, 何熊熊. 欺骗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, 2023, **49**(2): 425−436)

9 Wang Y, Lu J, Zheng W X, Shi K. 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

10 Gao L, Deng S, Ren W. Differentially private consensus with an event-triggered mechanism. *IEEE Transactions on Control of Network Systems*, 2019, **6**(1): 60−71

11 Liu X K, Zhang J F, Wang J. Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems. *Automatica*, 2020, **12**: Article No. 109283

12 He J, Cai L, Cheng P, Pan J, Shi L. Consensus-based data-privacy preserving data aggregation. *IEEE Transactions on Automatic Control*, 2019, **64**(12): 5222−5229

13 Zhao C, Chen J, He J, Cheng P. Privacy-preserving consensus-based energy management in smart grids. *IEEE Transactions on Signal Processing*, 2018, **66**(23): 6162−6176

14 Pan D, Ding D, Ge X, Han Q L, Zhang X M. Privacy-preserving platooning control of vehicular cyber-physical systems with saturated inputs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, **53**(4): 2083−2097

15 Gao C, Wang Z, He X, Dong H. Fault-tolerant consensus control for multiagent systems: An encryption-decryption scheme. *IEEE Transactions on Automatic Control*, 2022, **67**(5): 2560−2567

16 Hadjicostis C N. Privary preserving distributed average consensus via homomorphic encryption. In: Proceedings of the IEEE Conference on Decision and Control (CDC). Miami, USA: IEEE, 2018. 1258−1263

17 Lee K, Hong J P, Choi H H, Levorato M. Adaptive wireless-powered relaying schemes with cooperative jamming for two-hop secure communication. *IEEE Internet of Things Journal*, 2018, **5**(4): 2793−2803

18 Wang Y. Privacy-preserving average consensus via state decomposition. *IEEE Transactions on Automatic Control*, 2019, **64**(11): 4711−4716

19 Chen X, Huang L, Ding K, Dey S, Shi L. Privacy-preserving push-sum average consensus via state decomposition. *IEEE Transactions on Automatic Control*, 2023, **68**(12): 7974−7981

20 Zhang J, Lu J, Chen X. Privacy-preserving average consensus via edge decomposition. *IEEE Control Systems Letters*, 2022, **6**: 2503−2508

21 Anderson R J, Petitcolas F A P. On the limits of steganography. *IEEE Journal on Selected Areas in Communications*, 1998, **16**(4): 474−481

22 Wang Z, Byrnes O, Wang H, Sun R, Ma C, Chen H, et al. Data hiding with deep learning: A survey unifying digital watermarking and steganography. *IEEE Transactions on Computational Social Systems*, 2023, **10**(6): 2985−2999

23 Mazurczyk W, Caviglione L. Steganography in modern smartphones and mitigation techniques. *IEEE Communications Surveys & Tutorials*, 2015, **17**(1): 334−357

24 Shen Chang-Xiang, Zhang Huan-Guo, Feng Deng-Guo, Cao Zhen-Fu, Huang Ji-Wu. Overview of information security. *Science in China (Series E)*, 2007, **37**(2): 129−150
(沈昌祥, 张焕国, 冯登国, 曹珍富, 黄继武. 信息安全综述. 中国科学 E 辑, 2007, **37**(2): 129−150)

25 Xiao X, Sun X, Yang L, Chen M. Secure data transmission of wireless sensor network based on information hiding. In: Proceedings of the 4th Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services. Philadelphia, USA: IEEE, 2007. 1−6

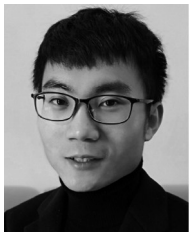
26 Olfati-Saber R, Fax J A, Murray R. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 2007, **95**(1): 215−233

27 Zhao B, Zhang Y. Secure encoding strategy for consensus of multi-agent systems in the presence of eavesdropper. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022, **69**(8): 3420−3424

28 Wang A, Liu W, Li T, Huang T. Privacy-preserving weighted average consensus and optimal attacking strategy for multi-agent networks. *Journal of the Franklin Institute*, 2021, **358**(6): 3033−3050

29 Charalambous T, Manitara N E, Hadjicostis C N. Privacy-preserving average consensus over digraphs in the presence of time delays. In: Proceedings of the 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Monticello, USA: IEEE, 2019. 238−245

30 Manitara N E, Rikos A I, Hadjicostis C N. Privacy-preserving distributed average consensus in finite time using random gossip. In: Proceedings of the 2022 European Control Conference (ECC). London, UK: IEEE, 2022. 1282−1287



伍益明 杭州电子科技大学网络空间安全学院副教授. 主要研究方向为网络安全, 分布式系统安全控制, 集群智能. 本文通信作者.
E-mail: ymwu@hdu.edu.cn
(**WU Yi-Ming** Associate professor at the School of Cyberspace, Hangzhou Dianzi University. His research interest covers cyber security, distributed system secure control, and swarm intelligence. Corresponding author of this paper.)

张润荣 杭州电子科技大学网络空间安全学院硕士研究生. 主要研究方向为多智能体系统网络安全和隐私保护.
E-mail: 212270030@hdu.edu.cn
(**ZHANG Run-Rong** Master student at the School of Cyberspace, Hangzhou Dianzi University. Her research interest covers cyber security for multi-agent systems and privacy-preserving.)

徐 宏 中国电子科技集团公司第三十二研究所高级工程师. 杭州电子科技大学计算机学院博士研究生. 主要研究方向为计算机软件架构, 信号处理和群体机器人技术.
E-mail: frankxuh@126.com
(**XU Hong** Senior engineer at the 32nd Research Institute of China Electronics Technology Group Corporation. Ph.D. candidate at the School of Computer Science and Technology, Hangzhou Dianzi University. His research interest covers computer software architecture, signal processing, and swarm robotics.)

朱晨睿 杭州电子科技大学网络空间安全学院博士研究生. 主要研究方向为无人集群, 边缘计算和资源调度.
E-mail: zzzcr2022@gmail.com
(**ZHU Chen-Rui** Ph.D. candidate at the School of Cyberspace, Hangzhou Dianzi University. His research interest covers UAV swarm, edge computing, and resource scheduling.)

郑 宁 杭州电子科技大学网络空间安全学院研究员. 主要研究方向为信息安全, 信息管理系统.
E-mail: nzheng@hdu.edu.cn
(**ZHENG Ning** Professor at the School of Cyberspace, Hangzhou Dianzi University. His research interest covers information security and information management systems.)