综述

群智能系统的安全与隐私保护综述

严宇萍[1]，高婷[1]，谢雨晗[2]，金耀初[1]
（1. 西湖大学可信及通用人工智能实验室，浙江 杭州 310024；
2. 西安电子科技大学，陕西 西安 710071）

摘　要：群智能系统凭借其分布式架构、高自组织性和强鲁棒性等特征，在推动社会生产和生活方式智能化发展方面展现出巨大的潜力。然而，安全性与隐私保护问题已成为制约其稳定运行与广泛应用的关键因素，直接影响用户信任度与技术的规模化部署。因此，确保群智能系统的运行安全、实现数据隐私保护，并增强其在复杂环境中的抗攻击能力和鲁棒性，已成为当前亟待解决的重大问题。对此，全面描述了群智能系统的定义、特点、通用结构及其应用场景等，明确提出现阶段群智能系统涵盖数据、通信、系统可靠、鲁棒和信任管理的安全目标与相关数据、身份和意图的隐私保护目标，并分析主要攻击方法及相关防御技术。在此基础上，系统梳理当前主流的解决方案，以应对群智能系统在安全与隐私保护方面的问题。最后，深入探讨群智能系统在这一领域面临的核心挑战，并展望未来可能的发展方向，旨在为后续的研究提供理论支持与技术指导。

关键词：群体智能；群智能系统；安全防御机制；隐私保护
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025052

Security and privacy protection in swarm intelligence systems: a review

YAN Yuping[1], GAO Ting[1], XIE Yuhan[2], JIN Yaochu[1]
1. Trustworthy and General Artificial Intelligence Laboratory, Westlake University, Hangzhou 310024, China
2. Xidian University, Xi’an 710071, China

Abstract: Swarm intelligence systems are recognized for their significant potential in advancing the intelligent development of social production and lifestyle, owing to their distributed architecture, high self-organization, and strong robustness. However, security and privacy protection issues were identified as critical factors limiting their stable operation and widespread application, directly impacting user trust and the large-scale deployment of the technology. Consequently, ensuring the operational security of swarm intelligence systems, achieving data privacy protection, and enhancing their anti-attack capabilities and robustness in complex environments were highlighted as urgent challenges to be addressed. In this context, their definitions, characteristics, general structures, and application scenarios of

收稿日期：2024−12−01；修回日期：2025−03−06
通信作者：金耀初，jinyaochu@westlake.edu.cn
基金项目：国家自然科学基金资助项目（No.W2441019）
Foundation Item: The National Natural Science Foundation of China (No.W2441019)

swarm intelligence systems were comprehensively described. The security objectives covering data, communication, system reliability, robustness, and trust management, as well as the privacy protection objectives related to data, identity, and intent, were explicitly proposed. Additionally, the main attack methods and related defense techniques were analyzed. Based on this, current mainstream solutions were systematically reviewed to address the security and privacy protection issues in swarm intelligence systems. Finally, the core challenges faced by swarm intelligence systems in this field were thoroughly discussed, and potential future development directions were explored, aiming to provide theoretical support and technical guidance for subsequent research.

Key words: swarm intelligence, swarm intelligence system, security defense mechanism, privacy protection

0 引言

大量研究表明，自然界中的群体行为展示了自然界中一种“涌现”出的集体智慧，即群体智能（swarm intelligence）[1]。群体智能是一种通过大量个体的自组织和分布式协作实现的整体智能，可以在不依赖集中控制的情况下解决复杂问题[2-3]。近年来，受自然群体智能启发，研究人员开始探索如何在人工系统中模拟和应用这一集体智慧，从而催生了群智能系统（swarm intelligence system）的概念[2,4]。群智能系统通过模仿自然界中的群体行为，实现系统内个体之间的协作与自组织，能够在分布式状况下高效完成复杂任务。它们已广泛应用于群机器人[5-7]、无人机编队[8-9]、交通管理[10-11]和通信网络[12-13]等领域，为实现更强大、更灵活的分布式系统提供了全新路径。

然而，相较于模拟环境，群智能系统的实际应用环境复杂性更高、动态性更强，充满了不确定性和潜在威胁[14-15]，例如，群智能系统的个体容错性、用户与群智能系统的交互和群智能系统的安全性等问题。这些问题不仅可能导致系统的行为偏离预期，还可能对任务的成功完成造成威胁。此外，数据共享和协作是群智能系统的核心机制，但也带来了严重的数据隐私风险[16]。在此背景下，如何在设计之初构建具有安全性和隐私保护能力的群智能系统成了一个亟须解决的重要问题。

近年来，针对安全可信的群体智能的综述逐渐增多。在设计和部署可靠群体系统的领域，Hunt E R 等[17]提出了“安全机器人群体检查清单”。该清单从伦理性、合法性、责任机制以及用户–群体交互等不同角度出发，为群智能系统的安全性研究和技术实践奠定了基础。然而，该清单仅以问题的形式对关键点进行了简单列举，未对该领域的发展现状和具体技术手段进行深入分析。此外，Higgins F 等[18]首次从资源限制、物理干扰、控制、通信、认证以及密钥管理等维度详细阐述了群机器人的安全挑战。该综述并未对相关的防御技术及最新进展进行深入分析。此外，Wilson J 等[19]对群智能系统的性能、可扩展性、鲁棒性和适应性进行了详细阐述，并探讨了这些特性在建立信任中的作用。然而，该研究对群智能系统的安全和隐私保护的研究描述较少，且未涵盖所有最新技术。

综上所述，尽管以上研究在群智能系统的安全与可信性方面取得了重要进展，但对该领域的综合发展和最新技术的探讨仍显不足。本文首先从群智能系统的结构出发，分别对群智能系统在物理层、通信层、网络层和应用层的安全威胁与隐私挑战进行系统性地界定与深入分析，并提出相应的防御措施。随后，结合实际应用背景全面梳理和评估现有的主流安全与隐私保护方案，并对相关技术进行分类总结。在此基础上，提出了目前群智能系统在安全与隐私保护上面临的核心挑战，并对未来可能的

发展方向进行展望，旨在为该领域的后续研究提供理论依据和实践参考。

1　群智能系统概述

本节将全面概述群智能系统的定义、特点、系统结构和应用场景，并对其他相关概念进行区分解释，如群体智能算法和多智能体系统（multi-agent system，MAS）[20]等。

群智能系统是指通过模拟自然界中生物种群在执行特定群体行为时的自组织和协作特性，实现复杂任务的分布式系统。例如，自然界中的蚁群觅食、蜂群采蜜、鱼群聚群和鸟群迁徙等群体行为。这类系统由多个简单的个体组成，每个个体遵循一定的规则，与系统中其他个体进行局部交互，并在全局上能够实现高度复杂和智能的群体行为。

群智能系统有以下 3 个主要特点。一是分布式控制机制，群智能系统没有中心化的控制机制，相反，其采用分布式控制机制，由每个个体根据局部信息和简单规则自主决策；二是高自组织性，群智能系统具有高自组织性，其个体之间能够通过局部交互自发形成全局规模的通信机制，不需要外部干预或全局规划；三是强鲁棒性，群智能系统的正常运转基于系统内个体间的交互协同，因此对系统内单个或少量个体的故障具有天然的高鲁棒性，降低了系统崩溃的风险。

群体智能算法是群智能系统的重要组成部分。一方面，群体智能算法可以模拟群智能系统内的自组织行为，例如，使用基因调控网络模型模拟多机器人系统中的自组织行为[21]。另一方面，群体智能算法，特别是群体智能优化算法，可以解决系统内具体的优化、路径规划和任务分配等问题。自然界中的群体智能较为丰富，由此衍生出了大量的群体智能优化算法及其应用，例如，1992 年由 Dorigo 提出的蚁群优化算法（ant colony algorithm）[22]，常用于求解路径规划等问题；1995 年由 Kennedy 等提出的粒子群优化（particle swarm optimization，PSO）算法[23]，常用于求解多目标优化等问题；2014 年由 Mirjalili 等提出的灰狼优化算法[24]，常用于求解神经网络中的特征选择等问题。

群智能系统的通用结构如图 1 所示。按照层级功能，该结构可划分为 4 层，分别是物理层、通信层、网络层和应用层。每一层分别负责完成特定功能，相邻层之间相互配合，最终实现群智能系统的正常运转与目标任务的执行。

（1）物理层。提供群智能系统的硬件设施基础，用于采集数据和支持系统内外的物理交互。物理层的硬件设施种类丰富，数量较多，常见的有传感器、执行器和电源管理模块等。

（2）通信层。实现群智能系统个体之间的信息传输与交互。在通信层，群智能系统采用无线通信传输指令和反馈。常见的通信模式有局部通信模式和全局通信模式。

（3）网络层。负责群智能系统在运行过程中个体所形成的网络拓扑结构的管理，包括系统个体实时状态监测、新增和删除个体管理等。

（4）应用层。根据现实世界的生产生活需求，辅助数据分析与反馈优化等技术，执行群智能系统的既定目标任务，包括系统实时避障、动态路径规划和既定目标搜索等。

综上所述，群智能系统的构建不仅依赖于先进的群体智能算法，还需要整合物理层的硬件设施、通信层的通信机制、网络层的拓扑结构和应用层的综合设计等。这种系统级的协作与组织是实现群智能系统高效运行的关键。

群智能系统的应用场景广泛，如图 2 所示[25-30]。例如，在无人机编队中，利用群智能系统，无人机能够通过信息共享和局部感知实现动态路径规划与实时避障，极大提升灾后救援和监测效率[9]。在机器人集群中，群智能系统通过协同作业和任务优化，广泛应用于工业生产、仓储

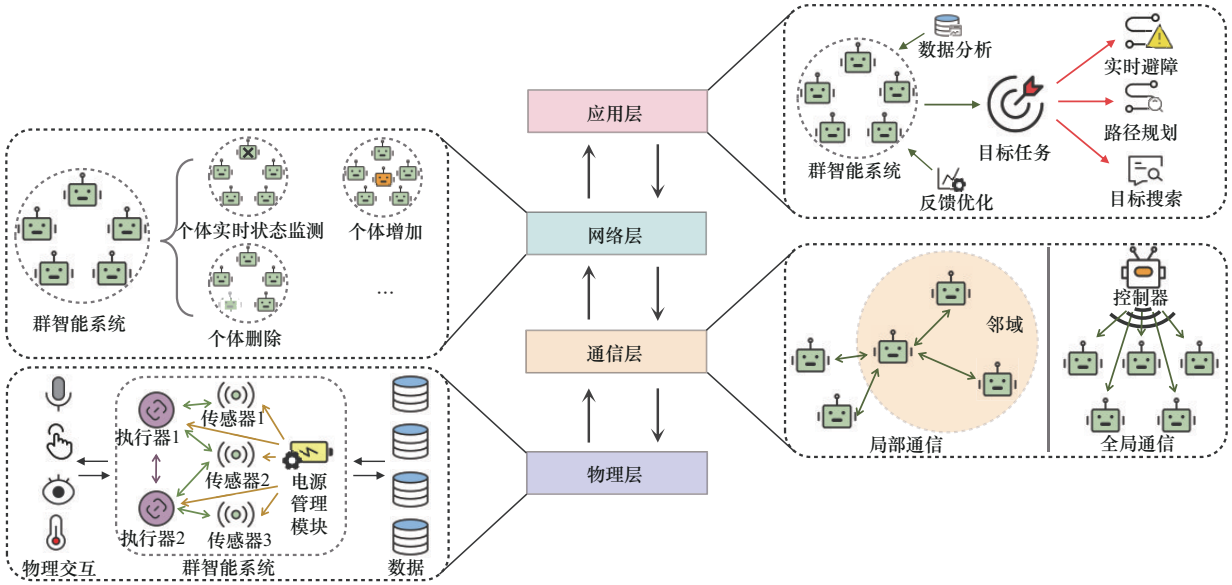

图1　群智能系统的通用结构

（a）无人机编队　（b）机器人集群　（c）智能交通系统

图2　群智能系统的应用场景

管理及农业作业[5]。此外，在智能交通系统中，借助群智能系统实现自适应交通信号调控，可有效管理交通流并减少拥堵[11]。这些应用场景表明，群智能系统将在推动各领域数字化转型及智能化发展中发挥重要作用。

与群智能系统相似的另一个常见概念是多智能体系统。多智能体系统通常指由多个具备独立感知、通信、计算和决策能力的智慧实体构成的系统[31]。多智能体系统与群智能系统最主要的区别在于个体特性。多智能体系统的每个个体都是单独的、具备高自主性的智能体，可以独立完成较为复杂的任务。而群智能系统中的个体通常是结构和功能较为简单的仿真节点或物理实体。这些个体虽然相对独立，但是其无法作为独立智能体完成较为复杂的动作或任务。

群智能系统在推动社会生产生活智能化发展方面展现出了巨大的潜力，但是其安全与隐私问题已成为影响系统运行的关键因素，直接关系到用户对系统的信任以及相关技术的广泛推广应用。因此，如何确保群智能系统在运行过程中的安全性、保障系统内数据的隐私性，并增强其在复杂环境中的抗攻击能力和鲁棒性，是当前亟须解决的重要问题。

2　群智能系统的安全威胁与隐私挑战及相应保护方法

为了保障群智能系统正常安全运行，进一步推动群智能系统在其他更多领域和实际问题中的应用，本节将从群智能系统中的安全威胁和隐私挑战出发，分别给出群智能系统中的安全和隐私保护的定义、常见的攻击方式和相应的保护方法。其中，群智能系统中安全威胁与隐私挑战的保护方法如图3所示。

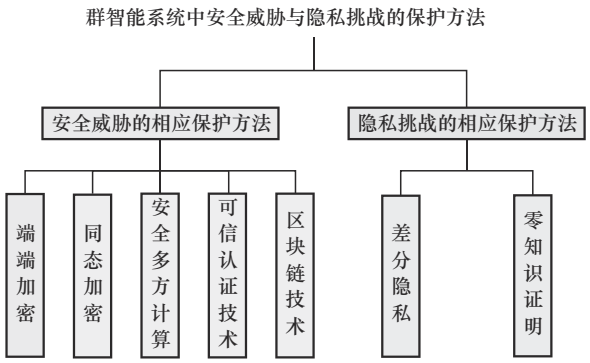

图3　群智能系统中的安全威胁与隐私挑战的保护方法

2.1　群智能系统中的安全威胁及相应保护方法

2.1.1　群智能系统中安全的定义

群智能系统的安全指的是群智能系统在运行过程中，能够抵御外部攻击和内部故障对系统物理设备、通信机制、数据完整性和一致性、网络拓扑结构和目标任务的安全性造成威胁的能力。根据群智能系统的特性，其安全性的核心目标主要涵盖数据安全性、通信安全性、系统可靠性、系统鲁棒性和系统信任管理。

数据安全性[32]包括防止敏感信息被窃取（数据保密性）、确保数据在传输或存储中未被篡改（数据完整性）、保障系统数据的正常访问和使用（数据可用性）。假设使用熵$H(x)$表示数据x的不确定性，数据保密性定义为确保攻击者的信息增益$I(A;x)$接近0，其中A是攻击者能获得的信息。数据完整性使用哈希函数$h(\cdot)$和原始数据定义完整校验，若$x'\neq x$，则$h(x')\neq h(x)$。数据可用性的定义为系统的正常工作时间U和总时间T的比例在可接受范围内。

通信安全性[33]主要保障群智能系统内部可安全正常通信，系统个体之间的通信链路不被监听、篡改或中断。通信链路上通过报文认证码（message authentication code，MAC）验证，若数据x被篡改，$\mathrm{MAC}(x)\neq\mathrm{MAC}(x')$。

系统可靠性[34]指的是群智能系统中单个或部分节点失效时，群智能系统能够继续正常运行。

系统鲁棒性[35]指的是群智能系统在面对环境干扰或恶意攻击者攻击时，能够进行系统内部的防御，保持系统稳定。系统鲁棒性体现在系统性能P在干扰下的波动幅度是否在可接受范围内，即$\Delta P=|P_o-P_a|$。其中，P_o是正常情况下的性能，P_a是受到攻击的性能。若$\Delta P\leqslant\epsilon$（$\epsilon$是被允许的波动阈值），则系统被认为具有良好的鲁棒性。

系统信任管理的核心目标在于群智能系统内部的分布式认证[36]，可帮助系统快速识别恶意节点并将其隔离。

2.1.2　群智能系统的安全威胁与攻击方式

由于群智能系统具有分布式控制机制、高自组织性和高鲁棒性的特点，其安全威胁的来源和系统内个体状态较为复杂，具有多样性。一方面，在群智能系统中，按照个体的性质，系统个体可分为正常工作节点个体和外部恶意节点个体2种。另一方面，群智能系统中个体行为可以分为诚实个体、半诚实个体和恶意个体[37]。通常情况下，外部恶意节点个体通常是恶意个体，其目标是通过主动干扰或入侵系统来破坏系统的正常运行与目标实现。正常工作节点个体通常被假设为诚实个体或半诚实个体，但由于系统的分布式控制机制和高自组织性，它们在某些情况下的动作和行为也可能成为群智能系统的安全威胁来源。

群智能系统的复杂性决定了其在系统结构的不同层级都面临着多样化的安全威胁。这些威胁涵盖了从物理设备到通信、网络和任务执行的各个方面，并相互交织，形成跨层级的安全挑战。物理层主要涉及物理硬件设备的安全问题[38]，包括传感器、执行器、无人机和机器人等。通信层是群智能系统实现正常通信功能的核心，也是攻击者的主要目标[39]。网络层负责群智能系统在运行过程中个体所形成的网络拓扑结构的管理，攻击者可能利用拓扑结构的动态特性破坏关键个体节点或相关网络链路，影响群智能系统内部个体之间的协作 [40]。应用层直接面向目标任务，攻击者可能通过篡改任务参数或既定目标，使群智能系统内的个体行为偏离预期，最终导致系统故障。同时，跨层级或多层级安全威胁问题则涉及多个层级之间的信息传递等，主要是群智能系统的信任管理问题以及系统在分布式环境下面临的安全威胁。

群智能系统中最常见的攻击是拜占庭攻击[41-43]。在拜占庭攻击中，攻击者可以通过节点的恶意行为，如矛盾消息扩散和伪造数据等，干扰分布式系统的正常运行。这种攻击表现为系统节点个体间状态的不一致、共识的失败或虚假消息的传播。拜占庭攻击的主要方式有以下3种。（1）伪造信息攻击，即恶意节点通过发送虚假的状态信息或数据，误导其他节点做出错误决策。（2）通信篡改攻击，即恶意节点选择性地丢弃或拒绝转发关键信息，阻断系统通信。（3）女巫攻击[44]，即恶意节点创建多个虚假身份，以多数权重干扰系统的共识过程。

2.1.3 群智能系统的安全保护方法

针对群智能系统的安全威胁，常见的安全保护方法有端端加密（end-to-end encryption）、同态加密（homomorphic encryption，HE）、安全多方计算（secure multi-party computation，SMPC）、可信认证技术和区块链技术等。

（1）端端加密

端端加密[45]是一种只有参与通信的用户可以读取信息的通信系统，允许数据在从原点到终点的传输过程中始终以密文形式存在，整个传输过程中数据均受到保护。因此，端端加密可以防止潜在的窃听者获取通信双方的消息内容。在群智能系统中，端端加密能有效防止恶意攻击者窃取通信信息，保护系统节点间的通信安全。

（2）同态加密

同态加密是端端加密的一种，但其额外提供了数据处理的功能。数据经过同态加密后，可在密文空间进行特定的计算，得到的密文计算结果在进行对应的同态解密后等于直接对明文数据进行相同的计算后得到的结果。

具体地，对于加密函数 $\mathrm{enc}(\cdot)$ 和解密函数 $\mathrm{dec}(\cdot)$，如果其满足式（1），则称该加密为同态加密：

$$\mathrm{dec}\left(\mathrm{enc}(a)\odot\mathrm{enc}(b)\right)=a\oplus b \tag{1}$$

其中，a、b 为明文，$\oplus$、$\odot$ 分别对应明文和密文域上的运算。当 $\oplus$ 为加法时，称该加密为加法同态加密；当 $\oplus$ 为乘法时，则称该加密为乘法同态加密。

若加密算法只满足加法同态加密或乘法同态加密，则称该算法满足部分同态加密（partially homomorphic encryption，PHE）[46]。若加密算法同时满足加法同态加密和乘法同态加密，且次数上没有限制，则称该算法满足全同态加密（full homomorphic encryption，FHE）[47]。

（3）安全多方计算

安全多方计算[48]是指在无可信第三方的情况下，多个参与方协同计算一个约定的函数，并保证每一方仅获取自己的计算结果，任意一方无法通过计算过程中的交互数据推测出其他任意一方的输入和输出数据。其特性与群智能系统的分布式协作特性高度契合，即使某个节点被攻击成为

恶意节点，恶意攻击者也无法获取其他节点的计算数据。安全多方计算技术并非单一的安全保护技术，而是由多项安全保护技术组成的协议栈。例如，同态加密是SMPC的特殊情况。

（4）可信认证技术

在群智能系统中，为确保分布式系统中的节点为受信任的安全节点，可引进可信认证技术。目前具有代表性的认证系统主要有公钥基础设施（public key infrastructure，PKI）[49]和基于身份的加密（identity based encryption，IBE）[50]。PKI是一个标准的密钥管理平台，能够为加密和数据签名等密码服务提供必需的密钥和证书管理。IBE是一种使用用户公开信息作为公钥的加密方式，不需要通过交换密钥即可验证每个用户的签名。

（5）区块链技术

区块链的去中心化特性、数据难以篡改性和共识机制，使其与群智能系统的分布式协作机制、高自组织性高度契合。区块链技术是一种有效的保障群智能系统安全性的方案。

区块链的核心机制是通过哈希函数和链式结构实现数据难以篡改、分布式账本和共识机制，以解决去中心化信任[51]。区块链是由多个区块 $B_1,B_2,\cdots,B_n$ 链接而成的有序链表：$\text{Chain}=\{B_1,B_2,\cdots,B_n\}$。链表中的一个区块可表示如下。

$$B=\left(H_{\text{prev}},\boldsymbol{T},N\right) \tag{2}$$

其中，H_{prev} 是前一个区块的哈希值；$\boldsymbol{T}$是区块中的交易集合；N是随机数，用于满足共识机制（如PoW）。

每个区块的哈希值为：$H(\text{Block})=H\left(H_{\text{prev}}\,/\!/\,\boldsymbol{T}\,/\!/\,N\right)$。并且每个区块的哈希值依赖于前一个区块的哈希值，即 $H\left(B_i\right)=H\left(H\left(B_{i-1}\right)\left(\boldsymbol{T}_i\right)N_i\right)$ $(i=2,3,\cdots,n)$。因此，任何数据的改动都会导致整个链的哈希失效，从而可通过区块存储和共享重要数据来确保这些数据的完整性。

在分布式系统共识形成的过程中，为抵御可能面临的故障和攻击，并增强系统鲁棒性，常使用相应的安全算法进行防护，如故障容错（crash fault tolerance，CFT）算法[52]和拜占庭容错（Byzantine fault tolerance，BFT）算法[53]等。

2.2　群智能系统中的隐私挑战及相应保护方法

2.2.1　群智能系统中隐私的定义和相应挑战

群智能系统中的隐私是指系统在多节点协作、信息交互和任务执行过程中，保护个体节点、用户及环境的敏感信息不被非法访问、滥用或泄露的能力。其中，主要包含数据隐私、身份隐私和意图隐私。

数据隐私[54]指对群智能系统中采集、存储、传输和协同合作的数据信息进行保护。由于系统中的传感器采集的环境数据可能包含敏感信息，因此群智能系统的数据隐私至关重要。身份隐私[55]在于保护群智能系统的个体节点和用户的隐私，防止其被追踪或识别。意图隐私[56]指在协同合作的过程中防止系统中的个体节点行为或系统的协作模式和行为模式被外界窃取或推断，从而保障系统意图的隐私性。

目前，群智能系统的隐私挑战主要在于系统的多模态数据处理、系统分布式控制机制的动态性和协同性。多模态数据处理增加了数据保护的复杂性，因为文本数据的保护机制可能并不适用于图像数据的保护。通信系统是群智能系统的核心，其通信信道和群智能系统内的信息共享增加了隐私泄露的可能性。另外，群智能系统的个体节点通常是资源受限的，因此群智能系统如何在其隐私保护和花费的相应计算资源之间找到平衡，是群智能系统在隐私保护方面的又一大挑战。

2.2.2　群智能系统的隐私保护方法

针对群智能系统的隐私挑战，常见的隐私保护方法有差分隐私、零知识证明和联邦学习等。

（1）差分隐私

差分隐私是Dwork在2006年首次提出的一种隐私定义[57]，旨在通过在统计分析或查询结果中引入噪声，使得攻击者无法准确推断出某个特定元素是否存在于数据集中，从而有效地保护数据隐私。噪声通常来源于随机算法。随机算法是一类输入为数据集$\boldsymbol{D}$中某一元素，输出不是固定值，而是服从某一分布的算法。如果对于任意2个相邻数据集$\boldsymbol{D}$和$\boldsymbol{D}'$，即2个数据集中只有1个记录不同，以及数据集经$\mathcal{M}$运算后得到的任意可能的输出集合$\boldsymbol{S}$，随机算法$\mathcal{M}$满足式（3），则称算法$\mathcal{M}$满足经典差分隐私（亦或称为ϵ-差分隐私）：

$$\Pr[\mathcal{M}(\boldsymbol{D})\in \boldsymbol{S}]\leqslant e^{\epsilon}\cdot \Pr\left[\mathcal{M}(\boldsymbol{D}')\in \boldsymbol{S}\right] \quad (3)$$

其中，Pr表示数据集$\boldsymbol{D}$或$\boldsymbol{D}'$经算法$\mathcal{M}$运算后的结果在集合$\boldsymbol{S}$中的概率。ϵ是一个非负参数，称作隐私预算，表示隐私保护的程度。ϵ越小，隐私保护能力越强，相应数据实用性越弱。

（2）零知识证明

零知识证明[58]通常用于解决不信任的双方，即证明者和验证者，在不泄露任何额外信息的情况下，证明某个命题的有效性。零知识证明协议满足完备性、可靠性和零知识性3个性质。根据证明者和验证者之间是否存在交互，零知识证明可以分为交互式零知识证明[58]和非交互式零知识证明[59]。由于对数据隐私保护具有重要意义，零知识证明目前已广泛应用于数字签名方案、电子投票、可验证计算和用户身份验证协议等领域[60]。

（3）联邦学习

联邦学习（federated learning，FL）[61]是一种分布式的机器学习方法，旨在解决传统集中式模型训练中面临的数据隐私、数据安全以及数据孤岛等挑战。该技术通过去中心化的方式，使模型能够在多个分布式设备或服务器上协作训练，同时避免共享原始数据。因此，这种方法将隐私保护与模型训练相结合，为隐私敏感的机器学习应用提供了一种创新解决方案。

尽管联邦学习能够显著提升数据的隐私保护能力，但其本身并不能完全防御潜在的隐私威胁，例如梯度泄露攻击、数据中毒攻击等。因此，为了实现更高水平的安全性，联邦学习通常需要结合其他隐私保护技术，如同态加密、差分隐私等，共同构建一个更为稳健的安全体系。

3 群智能系统的安全与隐私保护方案

3.1 群智能系统的安全方案

群智能系统的安全性是系统稳定运行的关键，也是其在实际应用中获得广泛信任和推广的前提条件。目前对群智能系统安全防御的研究主要集中在：拜占庭防御、基于区块链的抗恶意攻击技术、强化学习方案、通信中的安全协议构建和节点安全验证等。由于一种方案可能抵御多种攻击，因此在本节中，使用防御技术对安全方案进行分类。

在群智能系统中，系统个体之间达成一致共识是系统正确执行目标任务的关键。然而，群智能系统去中心化和自适应的协作方式易受到多种类型的攻击，包括捕获与重传消息、破坏数据完整性、访问未经授权的数据以及拒绝服务攻击等。虽然群智能系统通常被认为具有高鲁棒性，但是针对群智能系统的拜占庭攻击仍然不可被忽视。

为抵抗这些攻击，研究者提出了多种不同防护措施。例如，文献[62]研究了在网络中存在故障或恶意代理的情况下，通过均值子序列约简（mean subsequence reduction，MSR）算法，忽略来自潜在恶意邻居的异常值，从而降低恶意节点的影响。在此基础之上，文献[63]引入了新的图

论属性，即联合鲁棒性，将固定通信图的图论分析推广至时变通信图场景，实现具有异常行为的网络节点和时变通信图的网络化代理系统（networked agent system，NAS）鲁棒共识的问题。不仅如此，文献[64]提出了抵抗伪装攻击的加权均值子序列约简（weighted mean subsequence reduction，W-MSR）共识算法。该算法引入了接收信号的物理特征指纹分析，通过比较邻居代理信号的物理特征指纹，合法代理能够识别并隔离试图进行伪装攻击的恶意节点。然而该方法需要先验信息，并且信息传播较为受限。针对此问题，文献[65]提出了去中心化黑名单协议。该协议的核心思想是通过协作机器人共享的指控信息，执行图匹配算法生成黑名单，用于隔离恶意系统个体。为使在任务执行过程中所有节点能够检测到其邻居节点的异常行为，文献[66]基于两跳通信机制，提出了2种高效的检测恶意节点与弹性共识的分布式方案。

为防御群智能系统通信过程中的女巫攻击，文献[67]提出了一种基于无线信号分析的技术方案。该方案利用无线信号在环境中传播时的散射和吸收特性，能够提取独特的“空间指纹”。这些指纹无法被攻击者轻易篡改，因此具有高鲁棒性和抗操控性。类似地，文献[68]提出了一种轻量级的系统——ScatterID，通过为单天线机器人附加超轻、无电池的反向散射标签来缓解女巫攻击。在基于众包的智能交通系统中，文献[69]提出可利用来自传统传感基础设施的噪声信息，以及通过众包数据推断的车辆动态和邻近图来抵御虚报“鬼影”车辆的恶意攻击者造成的虚拟拥堵与路径规划干扰。

区块链技术在群智能系统安全中的应用也较为广泛。群智能系统和区块链技术的首次结合由Castelló Ferrer提出。Castelló Ferrer强调区块链是推动群体机器人系统领域取得重大进展的关键技术，并且讨论了包括安全通信、分布式决策以及创新商业模式在内的多种问题[70]。后续研究首次提供了利用区块链技术协调机器人系统的实际概念验证，包括对相关方法的详细描述、具体实现以及实验效果评估[71]。并且，文献[72]提出了一种新颖的基于区块链和代币经济的安全框架，即通过智能合约管理机器人系统个体间的代币分配，根据贡献奖励正常机器人并剔除拜占庭机器人。类似地，文献[73]使用智能合约作为群体机器人系统在集体感知场景中的“元控制器”，防止机器人伪造虚假身份。文献[74]为去中心化移动自组织网络开发了一个高吞吐量通信框架，并提出了一种利用区块链技术作为安全支撑，Pi-puck机器人群体组成的群体机器人系统。

当然，目前也有使用强化学习来检测和防御针对群智能系统攻击的相关研究。文献[75]提出了一个两阶段的入侵检测系统，包括签名检测组件和异常检测组件。其中，异常检测组件利用深度神经网络来检测偏离预期行为的命令。此外，文献[76]提出了一种对抗性深度强化学习算法，旨在增强群体动态对此类恶意干预的弹性。具体地，系统中的每个机器人都分别采用长短期记忆（long short-term memory，LSTM）人工神经网络来预测外部干预导致的间距变化，以及采用生成对抗网络（generative adversarial network，GAN）模拟和评估潜在攻击对群体动态的影响，以确保群机器人系统在对抗性环境中的稳定运行。文献[77]实现了一种基于强化学习的安全无人机系统，允许无人机自主学习目标或入侵攻击者的动态行为。

在群智能系统安全协议构建方面，文献[78]提出了一种不需要中央信任服务器的群智能系统安全架构协议。为实现系统内多个机器人的协同合作，并确保该过程的不可否认性、可追溯性和抗恶意攻击的能力，该研究提出使用协调者和状态控制者在温特尼兹栈（Winternitz stack）上记录参与其中的机器人的变更过程，并提供历史记

录和可验证日志。

在群智能系统的节点安全验证方面，研究人员陆续提出了QR-Swarm[79]、Unif-Swarm[80]、DL-Swarm以及基于DL-Swarm的CDH-Swarm等安全协议[81]。QR-Swarm协议是基于Fiat-Shamir协议构造的分布式认证协议。Unif-Swarm则是在QR-Swarm的基础上发展而来的统一零知识协议。该协议提供了一个更为通用的安全框架，适用于更多不同类型的群智能系统。DL-Swarm是Unif-Swarm协议的一种具体实现。CDH-Swarm则在可计算的Diffie-Hellman安全协议的基础上，进一步提高了协议安全的确定性。

3.2 群智能系统的隐私保护方案

群智能系统的数据泄露主要存在于系统共识形成的过程中。现有的共识算法要求系统内每个节点与其邻居交换显式状态信息，但这会造成状态隐私的泄露。为实现共识算法中的隐私保护，通常可采用基于同态加密的保护方法、基于差分隐私的保护方法、基于可观性的保护方法以及基于联邦学习的保护方法等。下面将详细阐述这些隐私保护方法的相关研究内容。

同态加密作为常见的端端加密方法，在数据传输和共识计算中有着重要作用。文献[82]提出了无向网络的方法，利用部分同态加密技术，在无聚合器的情况下实现了系统内节点间的安全交互，避免了隐私泄露，相较于传统依赖中心化聚合的方案，提高了去中心化程度和数据安全性。类似地，通过结合动态变化的量化器和Paillier密码系统，文献[83]提出了一种用于求解平均共识问题的加密控制算法，适用于通信拓扑为强连通有向图的分布式系统。该方法不仅能够有效解决平均共识问题，还能在保持计算精度的同时降低计算复杂度。同时，文献[84]通过自下而上的群体智能方法与局部规则结合同态加密技术，实现了物流行业的安全多方优化，既提升了优化效率，又增强了数据的隐私保护能力。相比于传统未加密优化方法，该方案能够在保障数据安全的同时，确保优化计算的可行性和有效性，但同时这些算法都会造成额外的计算和通信开销。

在共识形成的过程中，有诸多利用差分隐私技术来保护用户数据隐私的研究。例如，文献[85]提出了一种结合差分隐私共识算法的分布式事件触发机制，有效减少了实时通信和控制器更新的频率，在保障数据隐私的同时显著降低了系统的通信与计算开销。相比于传统的周期性通信机制，该方法提高了系统效率并减少了资源消耗，然而其可能带来精度上的问题。为消除在数据传输过程中差分隐私导致的量化误差对共识精度的影响，从而兼顾隐私保护与共识准确性，文献[86]提出了一种对数形式的动态编码－解码（logarithmic dynamic encoding-decoding，LDED）方案。除了共识形成外，差分隐私技术还可以应用于群智能系统的群体智能算法优化领域。例如，文献[87]首次尝试将差分隐私与群体智能相结合，提出了一种通用的基于差分隐私的群体智能算法框架，使算法在优化过程中既能保护个体数据隐私，又能实现性能优化。

基于可观性的隐私保护方法利用动态系统的可观性理论，研究系统动力学中状态信息的泄露机制，以防止攻击者通过观测系统的动态演化推断节点的隐私信息。在基于可观性的隐私保护方法的相关研究中，文献[88]提出了“隐私指数”的概念。隐私指数直观地表示用来恢复系统的所有初始状态所需联合的群智能系统的个体数目，并用此衡量群智能系统网络的隐私性，但为了实现更高的隐私保护，可能需要增加更多的个体联合，进而可能增加通信开销或计算复杂度，影响群智能系统的实时性和效率。另外，文献[89]受社交网络中公开意见与私人意见框架的启发，提出了一种新的迭代算法。通过设计独立的交互变量和真实状态变量，此算法可以在不引入额外隐私保护技术的情况下，自动实现对系统的隐私保

护，从而提高了隐私保护的自然性与有效性。基于可观性的隐私保护方法通常假设系统的状态可以通过观察得到一定的推断，但如果系统的动态行为发生剧烈变化，或当系统中节点数量变化较大时，隐私保护效果可能会受到影响。

联邦学习作为解决孤岛问题的有效方案，在保护群智能系统的数据隐私方面也有着重要作用，其核心思想是在不共享原始数据的情况下，通过分布式客户端协作训练全局模型。例如，文献[90]将粒子群优化算法和联邦学习进行结合，提出了一种实现隐私保护的群体智能优化算法。类似地，为提高智能移动机器人在5G网络及未来网络中的安全性和鲁棒性，文献[91]提出了一种点对点（peer-to-peer，P2P）的具有隐私感知的异步联邦学习（privacy-perceiving asynchronous federated learning，PPAFL）框架。该框架基于声誉感知的协调机制，动态协调多个智能设备组成虚拟群智能系统，确保加密的P2P联邦学习过程。然而，传统联邦学习框架仍面临投毒攻击、模型反演攻击、梯度泄露和推理攻击等安全挑战。为应对这些威胁，研究者通常结合差分隐私、安全多方计算和同态加密等技术进行防御。例如，研究者提出了一种针对联邦原型学习（federated prototype learning，FedProto）的特征图中毒攻击及双重防御机制[92]，探索了FedProto在面对特征图中毒攻击时的脆弱性，并通过全知识蒸馏与特征图筛选剔除错误特征，从而使受攻击模型的推测准确率提升了1～5倍。然而，该方案未考虑数据异构性和模型异构性的挑战。此外，为降低灾难性遗忘、异构性影响及通信资源受限等问题，研究者提出了基于知识蒸馏与回放的灾难性遗忘预防机制，并利用原型共享知识，进一步提出了基于原型学习的联邦持续学习方法[93]。尽管这些研究在隐私保护和安全性方面取得了重要进展，但如何在隐私保护与额外的通信和计算开销之间找到平衡，仍然是联邦学习领域值得深入探讨的重要问题。例如，有研究者提出了一种去中心化的联邦学习框架，结合不可链接匿名性与共互惠性原则，确保隐私与安全性。通过去中心化信誉管理系统激励节点守规，该框架在保护隐私的同时保持模型完整性，既能检测恶意更新，又降低计算成本，优于差分隐私与同态加密等方案[94]。

3.3　群智能系统隐私与安全相关数据集

为全面验证群智能系统的隐私与安全性，研究者们针对不同类型的攻击开发了多样化的验证数据集。这些数据集根据其针对的攻击类型可划分为三大类：通用数据集、联邦学习安全与隐私数据集以及区块链和分布式系统相关数据集。

3.3.1　通用数据集

在通用数据集领域，研究者们构建了多个具有代表性的数据集，涵盖了从传统攻击到现代网络威胁的广泛场景。其中，DARPA98数据集[95]作为开创性工作，首次为军事网络测试提供了入侵检测系统的性能基准，包含了拒绝服务（denial of service，DoS）攻击、探测攻击和权限提升攻击等多种攻击类型。这一开创性工作直接推动了KDD99数据集[96]的开发，后者对DARPA98中的原始流量进行了系统处理，包含了更全面的正常和恶意流量样本。然而，KDD99数据集存在显著的记录冗余问题，这一问题在NSL-KDD数据集[97]中得到了有效解决。NSL-KDD数据集通过精心设计的预处理流程消除了冗余记录，为入侵检测系统和异常检测算法提供了更贴近现实的评估环境。尽管如此，这些早期数据集在应对现代网络流量分析时仍显不足。

为弥补这一缺陷，研究者们开发了多个更现代化的数据集。Kyoto 2006+数据集[98]采集了2006年至2009年间京都大学的真实网络流量，虽然新增了10个有助于入侵检测系统（intrusion detection system，IDS）研究的属性，但在人工标注和流量多样性方面仍存在局限。相比之下，

UNSW-NB15 数据集[99]提供了更全面的攻击场景，包含9种攻击类别（如后门攻击、DoS、漏洞利用等），并以数据包格式和双向流格式提供了约250万条流量记录，其中恶意流量占比2.8%。

在真实网络流量采集方面，UNIBS数据集[100]和CTU-13数据集[101]提供了独特的视角。UNIBS数据集通过校园网络边缘路由器收集了20个工作站的行为数据，为研究校园网络通信模式提供了宝贵资源。CTU-13数据集则专注于僵尸网络研究，包含了13个精心设计的僵尸网络样本场景，每个场景都混合了僵尸网络流量、正常流量、命令与控制流量以及背景流量，并提供了详细的恶意软件攻击类型标注。

针对特定攻击类型的研究，TUIDS数据集[102]和分布式拒绝服务（distributed denial of service，DDoS）2016数据集[103]提供了专业化的数据支持。TUIDS数据集涵盖了僵尸网络、DoS/DDoS、探测等多种恶意流量类型，虽然未公开发布原始流量大小，但其包含的25万条标注流量记录仍具有重要研究价值。DDoS 2016数据集则专注于DDoS攻击研究，包含了UDP泛洪攻击、Smurf攻击等多种攻击类型，尽管其流量是通过模拟器生成的，但仍为DDoS攻击防御研究提供了重要参考。

在应用层攻击研究方面，CIC-DDoS数据集[104]专注于8种不同的应用层DDoS攻击，特别是HTTP DDoS攻击，其采用的ISCX 2012数据集的正常流量为研究提供了可靠的基准。最新开发的TII-SSRC-23数据集[105]则代表了数据集发展的新方向，作为一个异构集合，它涵盖了8种主要流量类型（音频、背景流量、文本等）和32个子类型，为研究复杂网络环境下的安全防御提供了全面的数据支持。

3.3.2 联邦学习安全与隐私数据集

根据联邦学习针对的安全威胁类型，这些数据集可以划分为以下几类：投毒攻击数据集、模型反演攻击数据集和成员推理攻击数据集等。

在投毒攻击数据集领域，研究者们通过模拟恶意客户端上传篡改的模型参数或梯度数据，生成了多种实验数据。例如，LEAF Benchmark [106]是一个联邦学习基准测试框架，支持生成投毒攻击的实验数据。这些数据集通常包含正常客户端和恶意客户端的训练数据，恶意客户端可能通过上传错误数据或发起标签翻转攻击来破坏全局模型的训练过程。此外，拜占庭鲁棒性联邦学习数据集[107]提供了多种拜占庭攻击场景的数据，包括梯度篡改和模型投毒攻击，为研究联邦学习的鲁棒性提供了重要参考。常见的公开数据集，如MNIST[108]、CIFAR-10[109]等也常被用于模型反演攻击、成功推理攻击等的研究，并评估其隐私保护能力。

3.3.3 区块链和分布式系统相关数据集

为验证分布式系统和区块链网络在拜占庭攻击下的鲁棒性与安全性，研究者们开发了多种专门的数据集和实验环境。这些数据集通过模拟恶意节点的行为，为拜占庭容错算法和防御机制的研究提供了重要支持。

研究者们通过模拟分布式系统中的恶意节点行为，生成了多种实验数据。例如，实用拜占庭容错协议（practical Byzantine fault tolerant protocol，PBFT）实验数据集[110]基于经典的PBFT，包含了正常节点和拜占庭节点的通信日志和状态变化数据。拜占庭节点可能通过发送错误消息、延迟响应或故意破坏共识来干扰系统运行。此外，SWIM数据集[111]提供了分布式系统中的故障注入功能，支持模拟节点崩溃、消息丢失等拜占庭行为，为研究分布式系统的容错能力提供了重要数据支持。

在与区块链相关拜占庭攻击数据集方面，研究者们针对区块链网络中的拜占庭行为开发了多种数据集。例如，BlockSim[111]是一个区块链模

拟框架，支持生成拜占庭攻击场景的数据，包括双花攻击、自私挖矿和分叉攻击等。这些数据集通常包含恶意节点的交易行为、共识消息和网络延迟数据，为区块链安全研究提供了丰富的实验环境。

4　挑战与未来展望

4.1　群智能系统的挑战

针对上述有关群智能系统的安全与隐私保护领域的分析，本节分析未来群智能系统在这一领域可能面临的挑战，包括但不限于系统通信链路的安全性、区块链技术的适配性问题、系统协作与隐私保护的矛盾，以及人机协作与伦理问题等。

4.1.1　系统通信链路的安全性

群智能系统中的节点通过频繁地交换数据来进行协同决策。因此，通信中的数据可能受到窃取、篡改、重放和DoS等攻击的威胁。尤其是在开放或者不受信任的网络环境中，无线通信系统易受恶意干扰信号的影响，导致信息的完整性、机密性和真实性难以保证。目前，已有研究提出了一系列异常检测方案和广播通信方法，但如何设计高效的通信和加密机制，在保证数据机密性的同时考虑低时延和节点资源受限的现状是群智能系统在实际应用中可能面临的挑战。同时，DoS与通信堵塞会造成系统内合法节点无法正常通信。因此，如何通过流量控制和带宽分配算法来抵御DoS，以及在资源有限的情况下实现攻击流量的快速识别和过滤，动态调整通信路径与数据传输策略，提高通信链路的可靠性与容错性，以及保证数据的高效传输，都是待解决的问题。

4.1.2　区块链技术的适配性问题

区块链技术凭借其去中心化、难以篡改、公开可追溯等特性，为提升群智能系统的安全性、可靠性和用户信任提供了重要支撑。然而，将区块链技术与群智能系统进行结合仍面临诸多挑战，主要体现在计算资源、网络通信、数据隐私与系统性能等方面。

从计算与存储资源受限的角度分析，由于群智能系统中的个体往往计算能力和存储空间有限，因此，如何通过轻量化区块链设计以及引入资源友好的共识机制来降低计算消耗和存储负担，成为将区块链技术用于群智能系统的关键挑战。从网络通信开销与实时性问题的角度分析，区块链系统节点间频繁进行数据同步与共识确认不仅存在时延，而且将导致系统通信流量与带宽消耗大幅增加。这与群智能系统在任务协同与决策中对高实时性的要求存在矛盾，因此，如何设计具有低时延的轻量级共识机制，提高数据同步效率和系统鲁棒性，成为另一个亟待解决的问题。在数据隐私保护方面，区块链的公开性与群智能系统的数据隐私保护需求存在冲突，因此，如何通过隐私保护技术，实现系统内数据的安全共享与验证，并通过链下存储与链上验证机制减少敏感数据的直接上链，是区块链技术与群智能系统进行结合研究的重要方向。在恶意攻击与恶意节点防范方面，恶意节点可能扰乱系统的协同运行。同时，智能合约的潜在漏洞可能被攻击者利用，导致系统的安全性和稳定性受损，因此，需要增强系统的拜占庭容错能力，并通过智能合约的自动化安全审查与防护机制，全面提升系统的安全防护能力。

4.1.3　系统协作与隐私保护的矛盾

群智能系统的数据共享与其隐私保护之间存在不可避免的矛盾。若系统节点间数据共享程度过低，可能会导致系统无法充分获取全局信息，影响协同决策效果，降低系统性能。若系统节点间数据共享程度过高，可能导致隐私泄露风险增加，因此，如何在系统隐私保护与其任务协作效果之间取得平衡，是群智能系统在隐私保护方面的核心难题。

另外，传统的数据保护方法存在较大的局限性，例如，不适合的加密方式可能带来额外的计算消耗和存储代价，差分隐私在实时决策场景中会对决策性能产生消极影响。因此，如何在保持数据有效性的同时，最大限度地减少系统隐私泄露的风险是其中的又一重大挑战。

4.1.4 人机协作与伦理问题

随着群智能系统的广泛应用，其与人类用户的协作日益增多，尤其在复杂任务中，人机协同已成为提升任务执行效率的重要手段。同时，人机交互过程中不可避免地存在伦理与社会问题，主要在决策透明性和不当行为的责任划分等方面。这些问题不仅对相关技术提出了更高要求，同时也应受到社会伦理规则和相关法律规范的共同约束。

当群智能系统做出决策时，往往存在人类难以追溯决策依据的问题，因此，如何通过可解释人工智能技术，使群智能系统的决策过程透明、可追溯，并能够以直观的方式向人类解释其行为，成为亟待解决的关键问题。此外，当群智能系统在执行任务过程中出现错误，往往难以明确责任在于技术故障、算法设计缺陷，还是人为操作不当。因此，确保系统在关键决策中遵循可控性和伦理原则，通过技术手段减少不当行为带来的负面影响，也是一项重大的挑战。

4.2 群智能系统的未来展望

通过上述对群智能系统的安全与隐私保护领域的相关分析，本节对未来群智能系统这一领域可能的发展方向进行描述，包括但不限于多模态融合与智能协作、群智能系统的可信与隐私保护方案、人机交互中的安全研究等。

4.2.1 多模态融合与智能协作

随着多模态感知技术的不断突破，未来的群智能系统将进入一个数据高度融合和任务智能协作的新时代。这种转变依赖于多模态数据处理与信息融合，使系统能够全面感知环境与任务，并实现高效决策，从而推动各个领域协同应用的广度和深度，为复杂场景提供更加精准、实时、智能的解决方案。主要包括以下3个方面。（1）多模态数据的深度融合与信息互补。通过突破单一感知模式的局限，将视觉、语音、触觉、温度及环境监测等多种传感器采集的数据进行结合，形成强大的多模态感知能力。（2）跨模态系统节点协同。群智能系统各节点之间可以利用不同模态的数据进行实时的信息共享与反馈，形成高效的协作机制。（3）强化多模态深度学习与智能推理能力。通过引入多模态深度学习，群智能系统能够不断从新的数据和任务场景中学习，并将所学知识迁移到不同任务中，以此提升协作效率和泛化能力。通过这3个方面的深度融合，未来的群智能系统将在复杂应用场景中发挥更大的作用。

4.2.2 群智能系统的可信与隐私保护

未来，群智能系统对数据安全与隐私保护的需求将日益迫切。为此，可信与隐私保护方案的发展将围绕3个核心方向展开：数据隐私安全共享、系统可信执行和动态信任管理。

然而，在这一过程中，必然会出现隐私保护与系统性能、数据可用性之间的权衡，因此，需要通过技术创新来寻找平衡。在此情况下，单一的方案往往无法满足对可信性和隐私保护的全面需求，因此必须采用多种技术方案的协同协作，如联邦学习、安全多方计算、差分隐私、可信执行环境（trusted execution environment，TEE）和区块链技术等，以实现数据安全共享与系统可信执行的有机结合。这种多层次、跨领域的方案设计有助于解决群智能系统的隐私保护与计算性能、数据共享、数据安全之间的平衡问题，为群智能系统构建一个安全、透明和高效的运行框架。

4.2.3 人机交互中的安全研究

为了增强人机协作的信任与可控性，未来的群智能系统将更加强调决策过程的透明化和结果

的可解释性。这包括开发适配于群智能系统的追溯机制和透明化工具，以确保节点间的协同行为可追踪，决策结果可验证。

此外，针对群智能系统与人类协作所带来的伦理问题，可逐步建立更清晰的伦理审查与责任划分机制。这包括制定责任归属框架和相应法律规范，明确群智能系统的设计者、开发者、操作人员和用户在不同场景下的相关责任，从而减少不必要的纠纷与冲突。

5　结束语

群智能系统以其分布式架构、高自组织性和强鲁棒性的特点，展示出了在推动社会生产与生活方式智能化变革中的显著优势。然而，其安全与隐私保护问题已成为制约其稳定运行和广泛应用的重要瓶颈。如何保障系统运行的安全性、实现高效且可靠的隐私保护，并增强其在复杂环境中的抗攻击性和鲁棒性，是目前群智能系统研究中须解决的核心课题。

本文首先从群智能系统的定义、特点和通用结构入手，系统描述了群智能系统的基础属性和典型应用场景，并明确指出安全与隐私保护的核心目标。通过分析群智能系统可能面临的主要安全威胁，总结了常见防御技术。同时，本文从系统架构的多个层面出发，梳理了当前主流的安全与隐私保护方案。这些保护机制的提出，为群智能系统的运行安全与隐私保护提供了有力的技术支持。

未来，随着群智能系统应用领域的不断扩展，其运行环境中的不确定性及安全威胁将更为复杂，如何构建更高效、灵活的防护机制将成为重要研究方向。在技术层面，需要探索新兴技术与传统方法的结合，以建立更加全面和灵活的安全防护机制；在非技术层面，也应引入更多法律、伦理以及监管措施，构建明确的安全和隐私保护框架，以平衡科技发展与社会利益之间的关系。

综上所述，群智能系统的安全与隐私保护是一个多维复杂的研究领域，需要持续关注与深入研究。群智能系统领域的相关研究不仅能够为智能交通、工业监控、智能医疗等领域提供坚实的理论和技术支撑，还将推动新一代智能系统的优化与广泛应用。通过持续深入的多维度研究，该领域必将迎来更加安全、可靠和高效的技术突破与行业落地，为社会生产生活方式的智能化变革注入强大的动能。

参考文献：

[1] KENNEDY J. Swarm intelligence[M]// Handbook of nature-inspired and innovative computing: integrating classical models with emerging technologies. New York: Springer, 2006: 187-219.

[2] KENNEDY J , EBERHART R C,SHI Y H. Swarm Intelligence [M]. San Francisco: Morgan Kaufmann, 2001.

[3] HASSANIEN A E, EMARY E. Swarm intelligence: principles, advances, and applications[M]. Boca Raton: CRC Press, 2016.

[4] BONABEAU E, DORIGO M, THERAULAZ G. Swarm intelligence: from natural to artificial systems[M]. New York : Oxford University Press, 1999.

[5] BRAMBILLA M, FERRANTE E, BIRATTARI M, et al. Swarm robotics: a review from the swarm engineering perspective[J]. Swarm Intelligence, 2013, 7: 1-41.

[6] DORIGO M, THERAULAZ G, TRIANNI V. Reflections on the future of swarm robotics[J]. Science Robotics, 2020, 5(49): eabe4385.

[7] ŞAHIN E. Swarm robotics: from sources of inspiration to domains of application[C]//Proceedings of the SAB 2004 International Workshop. Berlin: Springer, 2005: 10-20.

[8] CHEN W, LIU J J, GUO H Z, et al. Toward robust and intelligent drone swarm: challenges and future directions[J]. IEEE Network, 2020, 34(4): 278-283.

[9] TOSATO P, FACINELLI D, PRADA M, et al. An autonomous swarm of drones for industrial gas sensing applications[C]//Proceedings of the 2019 IEEE 20th International Symposium on “A World of Wireless, Mobile and Multimedia Networks"

(WoWMoM). IEEE, 2019: 1-6.

[10] GARCÍA-NIETO J, ALBA E, CAROLINA OLIVERA A. Swarm intelligence for traffic light scheduling: Application to real urban areas[J]. Engineering Applications of Artificial Intelligence, 2012, 25(2): 274-283.

[11] HOAR R, PENNER J, JACOB C. Evolutionary swarm traffic: if ant roads had traffic lights[C]//Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02. Piscataway: IEEE Press, 2002: 1910-1915.

[12] CAMPION M, RANGANATHAN P, FARUQUE S. UAV swarm communication and control architectures: a review[J]. Journal of Unmanned Vehicle Systems, 2019, 7(2): 93-106.

[13] KASSABALIDIS I, EL-SHARKAWI M A, MARKS R J, et al. Swarm intelligence for routing in communication networks[C]//Proceedings of the GLOBECOM'01. IEEE Global Telecommunications Conference. Piscataway: IEEE Press, 2001: 3613-3617.

[14] SCHRANZ M, DI CARO G A, SCHMICKL T, et al. Swarm intelligence and cyber-physical systems: concepts, challenges and future trends[J]. Swarm and Evolutionary Computation, 2021, 60: 100762.

[15] HIGGINS F, TOMLINSON A, MARTIN K M. Threats to the swarm: security considerations for swarm robotics[J]. International Journal on Advances in Security, 2009, 2(2&3):288-297.

[16] SHAMMAR E, CUI X H, AL-QANESS M A A. Swarm learning: a survey of concepts, applications, and trends[J]. arXiv preprint, 2024, arXiv: 2405.00556.

[17] HUNT E R, HAUERT S. A checklist for safe robot swarms[J]. Nature Machine Intelligence, 2020, 2(8): 420-422.

[18] HIGGINS F, TOMLINSON A, MARTIN K M. Survey on security challenges for swarm robotics[C]//Proceedings of the 2009 Fifth International Conference on Autonomic and Autonomous Systems. Piscataway: IEEE Press, 2009: 307-312.

[19] WILSON J, CHANCE G, WINTER P, et al. Trustworthy swarms[C]//Proceedings of the First International Symposium on Trustworthy Autonomous Systems. New York: ACM, 2023: 1-11.

[20] FERBER J, WEISS G. Multi-agent systems: an introduction to distributed artificial intelligence[M]. Boston: Addison-Wesley, 1999.

[21] GUO H L, MENG Y, JIN Y C. Analysis of local communication load in shape formation of a distributed morphogenetic swarm robotic system[C]//Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway: IEEE Press, 2010: 1-8.

[22] DORIGO M. Optimization, learning and natural algorithms[D]. Milano: Politecnico di Milano, 1992.

[23] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of the ICNN'95 - International Conference on Neural Networks. Piscataway: IEEE Press, 1995: 1942-1948.

[24] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61.

[25] AUER L, FEICHTNER A, STEINHÄUSLER F, et al. Swarm-technology for large-area photogrammetry survey and spatially complex 3D modelling[J]. International Journal of Latest Research in Engineering and Technology, 2018, 4(9): 33-39.

[26] Ball J B. Countering swarms: strategic considerations and opportunities in drone warfare[J]. Joint Force Quarterly 107, 2022(4): 4-14.

[27] YANG G Z, BELLINGHAM J, DUPONT P E, et al. The grand challenges of *Science robotics*[J]. Science Robotics, 2018, 3(14): eaar7650.

[28] RUBENSTEIN M, CORNEJO A, NAGPAL R. Programmable self-assembly in a thousand-robot swarm[J]. Science, 2014, 345 (6198): 795-799.

[29] PRAKASH J, MURALI L, MANIKANDAN N, et al. A vehicular network based intelligent transport system for smart cities using machine learning algorithms[J]. Scientific Reports, 2024, 14(1): 468.

[30] ENGLUND C, AKSOY E E, ALONSO-FERNANDEZ F, et al. AI perspectives in Smart Cities and Communities to enable road vehicle automation and smart traffic control[J]. arXiv preprint, 2021, arXiv: 2104.03150.

[31] ZHANG D, FENG G, SHI Y, et al. Physical safety and cyber security analysis of multi-agent systems: a survey of recent advances[J]. IEEE/CAA Journal of Automatica Sinica, 2021, 8(2): 319-333.

[32] SAMONAS S, COSS D. The CIA strikes back: redefining confidentiality, integrity and availability in security[J]. Journal of Information System Security, 2014,10(3):21-45.

[33] STAVROULAKIS P, STAMP M. Handbook of information and communication security[M]. Berlin: Springer, 2010.

[34] GU Q J, LIU P. Denial of service attacks[M]//Handbook of computer networks: distributed networks, network planning, control, management, and new trends and applications. Hcbokel Wiley, 2007: 454-468.

[35] GRIBBLE S D. Robustness in complex systems[C]//Proceedings Eighth Workshop on Hot Topics in Operating Systems. Piscataway: IEEE Press, 2002: 21-26.

[36] OMETOV A, BEZZATEEV S, MÄKITALO N, et al. Multifactor authentication: a survey[J]. Cryptography, 2018, 2(1): 1.

[37] SARKAR K R. Assessing insider threats to information security using technical, behavioural and organisational measures[J]. Information Security Technical Report, 2010, 15(3): 112-133.

[38] YOUSEF K M A, ALMAJALI A, GHALYON S A, et al. Analyzing cyber-physical threats on robotic platforms[J]. Sensors, 2018, 18(5): 1643.

[39] LU Z, LU X, WANG W Y, et al. Review and evaluation of security threats on the communication networks in the smart grid[C]// Proceedings of the 2010 - MILCOM 2010 Military Communications Conference. Piscataway: IEEE Press, 2010: 1830-1835.

[40] BACHER R, GLAVITSCH H. Network topology optimization with security constraints[J]. IEEE Transactions on Power Systems, 1986, 1(4): 103-111.

[41] BOUHATA D, MOUMEN H, MAZARI J A, et al. Byzantine fault tolerance in distributed machine learning: a survey[J]. arXiv preprint, 2022, arXiv: 2205.02572.

[42] GUERRAOUI R, GUPTA N, PINOT R. Byzantine machine learning: a primer[J]. ACM Computing Surveys, 2024, 56(7): 1-39.

[43] ZHANG G R, PAN F, MAO Y H, et al. Reaching consensus in the Byzantine empire: a comprehensive review of BFT consensus algorithms[J]. ACM Computing Surveys, 2024, 56(5): 1-41.

[44] DOUCEUR J R. The sybil attack[C]// Proceedings of the First International Workshop(IPTPS 2002). Berlin: Springer, 2002: 251-260.

[45] NABEEL M. The many faces of end-to-end encryption and their security analysis[C]//Proceedings of the 2017 IEEE International Conference on Edge Computing (EDGE). Piscataway: IEEE Press, 2017: 252-259.

[46] ROTHBLUM R. Homomorphic encryption: from private-key to public-key[C]//Proceedings of the 8th Theory of Cryptography Conference(TCC 2011). Berlin: Springer, 2011: 219-234.

[47] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) Fully homomorphic encryption without bootstrapping[J]. ACM Transactions on Computation Theory, 2014, 6(3): 1-36.

[48] GOLDREICH O. The foundations of cryptography: volume II basic applications[M]. New York: Cambridge University Press, 2004.

[49] PERLMAN R. An overview of PKI trust models[J]. IEEE Network, 1999, 13(6): 38-43.

[50] BONEH D, FRANKLIN M. Identity-based encryption from the Weil pairing[C]//Proceedings of the 21st Annual International Cryptology Conference. Berlin: Springer, 2001: 213-229.

[51] ZHENG Z B, XIE S A, DAI H N, et al. Blockchain challenges and opportunities: a survey[J]. International Journal of Web and Grid Services, 2018, 14(4): 352-375.

[52] LIU S Y, VIOTTI P, CACHIN C, et al. XFT: practical fault tolerance beyond crashes[C]//Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). Berkeley: USENIX Association, 2015: 485-500.

[53] CASTRO M, LISKOV B. Practical Byzantine fault tolerance and proactive recovery[J]. ACM Trans Comput Syst, 2021, 20(4): 398-461.

[54] JAIN P, GYANCHANDANI M, KHARE N. Big data privacy: a technological perspective and review[J]. Journal of Big Data, 2016, 3: 25.

[55] CAPURRO R, ELDRED M, NAGEL D. Digital whoness: identity, privacy and freedom in the cyberworld[M]. Frankfurt: Ontos Verlag, 2013.

[56] GERBER N, GERBER P, VOLKAMER M. Explaining the privacy paradox: a systematic review of literature investigating privacy attitude and behavior[J]. Computers & Security, 2018, 77: 226-261.

[57] DWORK C. Differential privacy[C]//Proceedings of the 33rd International Colloquium(ICALP 2006). Berlin: Springer, 2006: 1-12.

[58] FIEGE U, FIAT A, SHAMIR A. Zero knowledge proofs of identity[C]//Proceedings of the Nineteenth Annual ACM Conference on Theory of Computing - STOC '87. New York: ACM, 1987: 210-217.

[59] DE SANTIS A, MICALI S, PERSIANO G. Non-interactive zero-knowledge proof systems[M]//Advances in Cryptology - CRYPTO '87. Berlin: Springer, 1988: 52-72.

[60] PARTALA J, NGUYEN T H, PIRTTIKANGAS S. Non-interactive zero-knowledge for blockchain: a survey[J]. IEEE Access, 2020, 8: 227945-227961.

[61] ZHANG C, XIE Y, BAI H, et al. A survey on federated learning[J]. Knowledge-Based Systems, 2021, 216: 106775.

[62] WANG Y, ISHII H. Resilient consensus through event-based communication[J]. IEEE Transactions on Control of Network Systems, 2020, 7(1): 471-482.

[63] WEN G H, LV Y Z, ZHENG W X, et al. Joint robustness of time-varying networks and its applications to resilient consensus[J]. IEEE Transactions on Automatic Control, 2023, 68(11): 6466-6480.

[64] RENGANATHAN V, SUMMERS T. Spoof resilient coordination for distributed multi-robot systems[C]//Proceedings of the 2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS). Piscataway: IEEE Press, 2017: 135-141.

[65] WARDEGA K, VON HIPPEL M, TRON R, et al. Byzantine resilience at swarm scale: a decentralized blocklist protocol from inter-robot accusations[J]. arXiv preprint, 2013, arXiv: 2301.06977.

[66] YUAN L W, ISHII H. Secure consensus with distributed detection via two-hop communication[J]. Automatica, 2021, 131: 109775.

[67] GIL S, KUMAR S, MAZUMDER M, et al. Guaranteeing spoof-resilient multi-robot networks[J]. Autonomous Robots, 2017, 41(6): 1383-1400.

[68] HUANG Y, WANG W, JIANG T, et al. Detecting colluding sybil attackers in robotic networks using backscatters[J]. IEEE/ACM Transactions on Networking, 2021, 29(2): 793-804.

[69] SHOUKRY Y, MISHRA S, LUO Z T, et al. Sybil attack resilient traffic networks: a physics-based trust propagation approach[C]//Proceedings of the 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS). Piscataway: IEEE Press, 2018: 43-54.

[70] FERRER E C. The blockchain: a new framework for robotic swarm systems[J]. arXiv preprint, 2016, arXiv:1608.00695.

[71] STROBEL V, FERRER E C, DORIGO M. Managing Byzantine robots via blockchain technology in a swarm robotics collective decision making scenario[C]//Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems(AAMAS '18). Richland: International Foundation for Autonomous Agents and Multiagent Systems, 2018: 541-549.

[72] STROBEL V, PACHECO A, DORIGO M. Robot swarms neutralize harmful Byzantine robots using a blockchain-based token economy[J]. Science Robotics, 2023, 8(79): eabm4636.

[73] STROBEL V, CASTELLÓ FERRER E, DORIGO M. Blockchain technology secures robot swarms: a comparison of consensus protocols and their resilience to Byzantine robots[J]. Frontiers in Robotics and AI, 2020, 7: 54.

[74] Pacheco A, Strobel V, Dorigo M. A blockchain-controlled physical robot swarm communicating via an Ad-Hoc network[C]// Proceedings of the 12th International Conference(ANTS 2020). Cham: Springer, 2020: 3-15.

[75] JONES A, STRAUB J. Using deep learning to detect network intrusions and malware in autonomous robots[C]//Proceedings of the Cyber Sensing 2017. Anaheim: SPIE, 2017, 10185: 45-50.

[76] ABOUELYAZID M. Adversarial deep reinforcement learning to mitigate sensor and communication attacks for secure swarm robotics[J]. Journal of Intelligent Connectivity and Emerging Technologies, 2023, 8(3): 94-112.

[77] MASADEH A, ALHAFNAWI M, SALAMEH H A B, et al. Reinforcement learning-based security/safety UAV system for intrusion detection under dynamic and uncertain target movement[J]. IEEE Transactions on Engineering Management, 2022, 71: 12498-12508.

[78] SHAFARENKO A. A zero-trust swarm security architecture and protocols[J]. IACR Cryptology ePrint Archive, 2024: 1176.

[79] COGLIANI S, FENG B, FERRADI H, et al. Public key-based lightweight swarm authentication[M]//Cyber-Physical systems security. Cham: Springer, 2018: 255-267.

[80] MAIMUȚ D A, TEȘELEANU G. A generic view on the unified zero-knowledge protocol and its applications[C]//Information Security Theory and Practice. Cham: Springer, 2020: 32-46.

[81] TESELEANU G. Lightweight swarm authentication[C]//Proceedings of the 14th International Conference(SecITC 2021). Cham: Springer, 2022:248-259.

[82] RUAN M H, GAO H, WANG Y Q. Secure and privacy-preserving consensus[J]. IEEE Transactions on Automatic Control, 2019, 64(10): 4035-4049.

[83] KISHIDA M. Encrypted average consensus with quantized control law[C]//Proceedings of the 2018 IEEE Conference on Deci-

sion and Control (CDC). Piscataway: IEEE Press, 2018: 5850-5856.

[84] GOJKOVIĆ M, SCHRANZ M. Preserving privacy in logistics by using swarm intelligence from the bottom-up[C]//Proceedings of the 2024 IEEE 12th International Conference on Intelligent Systems (IS). Piscataway: IEEE Press, 2024: 1-7.

[85] WANG X, HE J P, CHENG P, et al. Privacy preserving average consensus with different privacy guarantee[C]//Proceedings of the 2018 Annual American Control Conference (ACC). Piscataway: IEEE Press, 2018: 5189-5194.

[86] CHEN W, WANG Z D, HU J, et al. Differentially private average consensus with logarithmic dynamic encoding-decoding scheme[J]. IEEE Transactions on Cybernetics, 2023, 53(10): 6725-6736.

[87] ZHANG Z Q, ZHU H, XIE M Y. Differential privacy may have a potential optimization effect on some swarm intelligence algorithms besides privacy-preserving[J]. Information Sciences, 2024, 654: 119870.

[88] RAMOS G, PEQUITO S. Designing communication networks for discrete-time consensus for performance and privacy guarantees[J]. Systems & Control Letters, 2023, 180: 105608.

[89] ZHANG J, LU J Q, HADJICOSTIS C N. Average consensus for expressed and private opinions[J]. IEEE Transactions on Automatic Control, 2024, 69(8): 5627-5634.

[90] TORRA V, GALVÁN E, NAVARRO-ARRIBAS G. PSO + FL = PAASO: particle swarm optimization + federated learning = privacy-aware agent swarm optimization[J]. International Journal of Information Security, 2022, 21(6): 1349-1359.

[91] ZHOU X K, LIANG W, WANG K I, et al. Decentralized P2P federated learning for privacy-preserving and resilient mobile robotic systems[J]. IEEE Wireless Communications, 2023, 30(2): 82-89.

[92] 王瑞锦, 王金波, 张凤荔, 等. 联邦原型学习的特征图中毒攻击和双重防御机制[J]. 软件学报, 2025, 36 (3): 1355-1374.
WANG R J, WANG J B, ZHANG F L, et al. Feature map poisoning attack and dual defense mechanism for federated prototype learning[J]. Journal of Software, 2025, 36(3):1355-1374.

[93] 张浩东, 杨柳, 于剑, 等. 基于原型学习的联邦持续学习方法[J]. 中国科学: 信息科学, 2024, 54(10): 2428-2442.
ZHANG H D, YANG L, YU J, et al. Federated continual learning based on prototype learning[J]. Scientia Sinica (Informationis), 2024, 54(10): 2428-2442.

[94] DOMINGO-FERRER J, BLANCO-JUSTICIA A, MANJÓN J, et al. Secure and privacy-preserving federated learning via co-utility[J]. IEEE Internet of Things Journal, 2021, 9(5): 3988-4000.

[95] LIPPMANN R, HAINES J W, FRIED D J, et al. The 1999 DARPA off-line intrusion detection evaluation[J]. Computer Networks, 2000, 34(4): 579-595.

[96] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Piscataway: IEEE Press, 2009: 1-6.

[97] DHANABAL L, SHANTHARAJAH D S P. A study on NSL-KDD dataset for intrusion detection system based on classification algorithms[J]. International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4(6): 446-452.

[98] SONG J, TAKAKURA H, OKABE Y, et al. Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation[C]//Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security. New York: ACM, 2011: 29-36.

[99] MOUSTAFA N, SLAY J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS). Piscataway: IEEE Press, 2015: 1-6.

[100] GRINGOLI F, SALGARELLI L, DUSI M, et al. GT: picking up the truth from the ground for internet traffic[J]. ACM SIGCOMM Computer Communication Review, 2009, 39(5): 12-18.

[101] GARCÍA S, GRILL M, STIBOREK J, et al. An empirical comparison of botnet detection methods[J]. Computers & Security, 2014, 45: 100-123.

[102] BHUYAN M H, BHATTACHARYYA D K, KALITA J K. Towards generating real-life datasets for network intrusion detection[J]. International Journal of Network Security, 2015, 17(6): 683-701.

[103] ALKASASSBEH M, AL-NAYMAT G, A A B, et al. Detecting distributed denial of service attacks using data mining techniques[J]. International Journal of Advanced Computer Science and Applications, 2016, 7(1):436-445.

[104] SHARAFALDIN I, LASHKARI A H, HAKAK S, et al. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy[C]//Proceedings of the 2019 International Carnahan Conference on Security Technology (ICCST). IEEE, 2019: 1-8.

[105] HERZALLA D, LUNARDI W T, ANDREONI M. TII-SSRC-23 dataset: typological exploration of diverse traffic patterns for intrusion detection[J]. IEEE Access, 2023, 11: 118577-118594.

[106] CALDAS S, DUDDU S M K, WU P, et al. LEAF: a benchmark for federated settings[J]. arXiv preprint, 2018, arXiv:1812.01097.

[107] CAO X Y, FANG M H, LIU J, et al. FLTrust: Byzantine-robust federated learning via trust bootstrapping[J]. arXiv preprint, 2020, arXiv:2012.13995.

[108] DENG L. The MNIST database of handwritten digit images for machine learning research[J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.

[109] KRIZHEYSKY A, SUTSKEVER I, HINTON G E. Imagevet classification with deep convolutional neural networks[J]. Communication of the ACM, 2017, 60(6): 84-90 .

[110] CASTRO M, LISKOV B. Practical Byzantine fault tolerance[C]// Proceedings of the Third Symposium on Operating Systems Design and Implementation. Berkeley: USENIX Association, 1999: 173-186.

[111] ALHARBY M, van MOORSEL A. Blocksim: an extensible simulation tool for blockchain systems[J]. Frontiers in Blockchain, 2020, 3: 28.

[作者简介]

严宇萍（1995–），女，博士，西湖大学可信及通用人工智能实验室在站博士后，主要研究方向为安全与隐私保护的机器学习与优化、具身智能安全、区块链等。

高婷（1998–），女，西湖大学可信及通用人工智能实验室科研助理，主要研究方向为群体智能及其安全、具身智能安全等。

谢雨晗（2003–），女，西安电子科技大学在读，主要研究方向为具身智能安全等。

金耀初（1966–），男，博士，西湖大学可信及通用人工智能实验室讲席教授，主要研究方向为人工智能与计算智能的理论、算法和工程应用研究，特别是数据驱动的优化、多目标优化、演化机器学习、安全与隐私保护的机器学习与优化、图神经网络组合优化、演化发育通用人工智能及形态发育自组织机器人等。