智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 多智能体系统安全性问题及防御机制综述

丁俐夫，颜钢锋

引用本文:
丁俐夫, 颜钢锋. 多智能体系统安全性问题及防御机制综述[J]. 智能系统学报, 2020, 15(3): 425–434.
DING Lifu, YAN Gangfeng. A survey of the security issues and defense mechanisms of multi–agent systems[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(3): 425–434.

在线阅读 View online: https://dx.doi.org/10.11992/tis.201812015

DOI: 10.11992/tis.201812015

# 多智能体系统安全性问题及防御机制综述

丁俐夫[1], 颜钢锋[2]
(1. 浙江大学 电气工程学院, 浙江 杭州 310027; 2. 浙江大学 华南工业技术研究院, 广东 广州 510760)

**摘 要:** 多智能体系统作为分布式人工智能的重要分支, 已成为解决大型、复杂、分布式及难预测问题的重要手段。在开放网络中, 多智能体系统仍面临许多安全问题, 潜在的安全威胁很可能影响其实际应用的稳定性、快速性和准确性。基于目前已知的多智能体系统通用模型, 介绍了多智能体系统通信协议、访问控制和协调机制中潜在的安全问题, 规范了多智能体系统安全性问题的研究体系, 总结了系统设计过程中可行的防御技术和隐私保护技术, 最后展望了多智能体系统安全研究的发展方向。

**关键词:** 多智能体系统; 分布式人工智能; 安全威胁; 防御机制; 网络安全; 通信协议; 访问控制; 协调机制
**中图分类号:** TP18; TP309 **文献标志码:** A **文章编号:** 1673−4785(2020)03−0425−10

**中文引用格式: 丁俐夫, 颜钢锋. 多智能体系统安全性问题及防御机制综述 [J]. 智能系统学报, 2020, 15(3): 425–434.**
**英文引用格式: DING Lifu, YAN Gangfeng. A survey of the security issues and defense mechanisms of multi-agent systems[J]. CAAI transactions on intelligent systems, 2020, 15(3): 425–434.**

# A survey of the security issues and defense mechanisms of multi-agent systems

DING Lifu[1], YAN Gangfeng[2]
(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; 2. Huanan Industrial Technology Research Institute, Zhejiang University, Guangzhou 510760, China)

**Abstract**: Currently, multi-agent systems are an important branch of distributed artificial intelligence and an important means of solving large, complex, distributed, and difficult-to-predict problems. However, in an open network, multi-agent systems still face many security issues. Potential security threats are likely to affect the stability, speed, and accuracy of the practical applications of multi-agent systems. Based on the general models of currently known multi-agent systems, the potential security problems in multi-agent system communication protocol, access control, and coordination mechanism are introduced. Research on multi-agent system security problems is standardized, and feasible defense and privacy protection technologies in the system design process are summarized. Finally, the development trend of multi-agent system security research is analyzed.

**Keywords**: multi-agent system; distributed artificial intelligence; security threat; defense mechanism; network security; communication protocol; access control; coordination mechanism

人工智能已经成为当前计算机研究最热门和应用最广泛的领域之一。在人工智能的不断发展中, 逐渐产生了一个前沿领域, 即智能体理论。智能体所具有的自治性、社交性和灵活性[1], 使其成为许多复杂系统的核心技术。随着计算机技术和人工智能的发展, 计算机需要处理的问题在复杂度和综合性都有所提高, 而单个智能体的计算能力和计算资源能力都有限, 无法满足分布式人工智能 (distributed artificial intelligence, DAI) 的需求。因此, 多智能体系统 (multi-agent system, MAS) 在开放的分布式环境中被开发, 并被广泛应用于与日常生活及工业生产相关的领域[2], 例如: 智能机器人 (协同式机器人[3]、水下机器人[4]), 交通控制 (智能城市交通[5]、路径规划[6]), 柔性制

**收稿日期:** 2018−12−13.
**基金项目:** 国家重点研发计划项目 (2018YFB0904900).
**通信作者:** 丁俐夫. E-mail: dinglifu@zju.edu.cn.

造，网络自动化 (银行管理系统、医疗保健系统[7]) 等。因此，在不久的将来，MAS 将进一步渗透到生产生活的各个领域，成为促进社会进步的关键技术。

在传统的集中式人工智能领域，单个的智能体已经拥有较为健全的防御机制。但是在 DAI 领域中，MAS 的多个智能体的逻辑和物理位置呈现分布状态，所以每个智能体需要通过预先定制的通信协议，与其他智能体通过网络进行数据传输，以此协调彼此的任务，并协同实现总体目标。在此过程中，开放的环境带来了安全隐患[8]。一方面，智能体由于自身的自治性和灵活性，容易受到攻击并成为恶意智能体。另一方面，MAS 在通信过程中容易产生恶意控制或信息泄露的问题。这些问题都可能影响 MAS 在生产生活中应用的推广。例如，针对医疗保健系统的攻击可能导致严重的隐私泄露问题。此外，针对自动驾驶车辆和城市交通控制系统的攻击可能产生恶意控制的事件，对交通安全造成威胁。

近年来，有关 MAS 安全问题的研究越来越多，许多研究人员着眼于可能的 MAS 攻击手段，并提出相对应的防御机制。针对开放网络中的智能体，各种安全模型和安全服务被提出，用于提高单个智能体通信能力的同时保证其安全性能。除此之外，一些研究人员关注于 MAS 的通信过程和访问控制，在多个智能体之间的交互问题上建立可靠的防御机制，保证 MAS 的完整性、保密性、可用性、可控性及不可否认性。

## 1　多智能体系统及其安全性问题

### 1.1　多智能体系统及其特点

随着人工智能技术的发展和普及，人工智能在各个方面都发挥着不可替代的作用。作为人工智能的前沿学科和研究热点之一，MAS 旨在通过智能体间通信与协调建立大型和复杂的软硬件系统。

智能体和 MAS 的定义并不统一，对于智能体，本文采用 Jennings 等[9] 的定义，即智能体是一个软件实体，能够在某些环境中灵活、自主地行动，以实现其设计目标。智能体拥有自治性、社交性和灵活性的特点，这不仅意味着智能体可以自主地控制其状态与行为，在没有外界介入的情况下操作与运行，还表示多个智能体可以在目标驱动下进行交流与沟通，并采取合适的行动。

MAS 由多个可交互的智能体组成，每个智能体在逻辑或物理位置上呈现分布状态，因此 MAS 没有全局控制，各智能体的数据通常是分散的。MAS 有一个全局目标，通过预先定制的协议，各智能体相互协作，共同解决全局问题。MAS 对于解决复杂问题意义重大，由于时间和资源的动态约束，系统需要解决的关键问题包括任务调度、资源分配和冲突调解。MAS 的出现克服了单一智能体的局限性，并具有并行性、分布性、开放性和容错性的优点。

### 1.2　多智能体系统面临的安全性威胁

当智能体在开放环境中，MAS 通过智能体的分布式协作降低了复杂问题的计算量并提高了系统性能，但同时 MAS 的机制对协作控制、访问控制和通信机制提出了更高的要求。因此，这些方面的安全漏洞可能对整个 MAS 造成损害。尤其是对于一些安全敏感的领域，如银行业和医疗服务领域，许多研究人员已经认识到 MAS 的安全漏洞并发现了可能的攻击。

早在 1998 年，Greenberg 等[1] 就已经归纳了几种可能针对开放环境中移动智能体的攻击方法。Wang 等[10] 则提出了针对移动智能体的几种攻击路径和后果。首先，MAS 中各个智能体拥有自己的职能、行为模式、偏好、授权和隐私，因此，恶意代码可以访问智能体授权，修改其首选项，窃取私有数据，甚至更改其行为模式和角色，使其成为影响整个 MAS 的攻击性智能体。其次，MAS 是一个开放的分布式系统，因此有大量的分布式节点相互独立但又相互依存，通过攻击最弱的节点，攻击者可以强制整个 MAS 停止服务。此外，主机可能容易受到伪装、DoS 或未经授权的访问的攻击。现有的 MAS 缺乏严格的认证机制，恶意代码可以假装智能体并试图连接到 MAS。这种情况可能导致非法资源占用，发送虚假信息以及窃取隐私数据等安全威胁。

对于智能体的攻击数据传输和通信协议过程，可能会发生两种类型的安全威胁：一种是探索性攻击 (exploratory attack)，即攻击者不会干扰通信，但会尝试提取可用信息，在这种情况下，传输的敏感信息可能泄露；另一种是诱发性攻击 (causative attack)，此时攻击者可能会尝试拦截、修改、删除甚至替换数据包。

## 2　多智能体系统的安全框架

许多研究已经指出了 MAS 的安全漏洞，并且针对 MAS 的攻击进行了尝试。为了进一步完善 MAS 的安全防御机制，本文首先针对 MAS 的安全需求搭建了框架，为 MAS 提供了通用的安全

性模型。

### 2.1　多智能体系统安全性问题的分类

多智能体系统安全性问题尚未形成完善的分类体系。Jung 等[11] 曾提出 MAS 安全问题分类的不同角度。本文总结为如图 1 所示的分类体系。从 MAS 系统环节的角度, MAS 的安全性应分为分布式调控中的安全问题与单个智能体中的安全问题。从 MAS 安全性需求考虑, 其安全性问题可以分为信息来源安全 (information origin security)、授权机制安全 (authorization mechanism security)、通信安全 (communication security)、主机保护安全 (host protection security)、智能体保护安全 (agent protection security) 以及共享资源安全 (resources sharing security)。另外, 根据攻击造成的 MAS 安全损害分类, MAS 安全问题可分为完整性安全 (integrity security)、可用性安全 (availability security) 和私密性安全 (privacy security)。

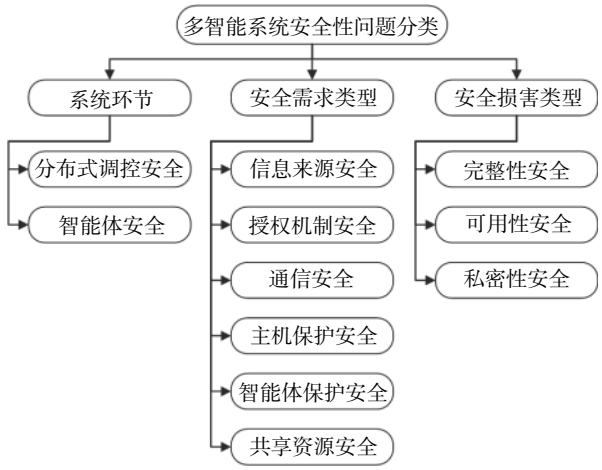


**图 1　多智能体系统的安全问题分类体系**

**Fig. 1　Taxonomy of the security issues of MAS**

### 2.2　多智能体系统的安全需求

为了实现 MAS 安全问题的标准化, 智能物理 Agent 基金会 (Foundation Intelligent Physical Agents, FIPA) 很早就着手于将 MAS 安全要求纳入其规范中。从 2002 年起, FIPA 已制定了若干结构规范条例, 对智能体系统的访问控制、数据传输以及智能体管理等方面[12-14] 安全要求进行了定义。

但是, 迄今为止, 依旧只有少数 MAS 设计者会将合理的安全防范纳入系统开发阶段[15]。这种瓶颈来自于两个方面[16]。一方面, 很多情况下安全要求会与 MAS 功能要求相冲突, 权衡安全性与功能性之间平衡的过程十分耗时。另一方面, 很多开发人员缺乏安全软件开发的专业知识。尤其是在 MAS 中, 每个智能体会分配不同的职能, 如果某个智能体出现过载, 将无法承担其安全要求的部分, 因此开发人员必须能够合理地定义系统设计, 平均分配子系统的负载。因此, 降低 MAS 的复杂性与单个智能体关键性成为了 MAS 的安全要求之一。

此外, 从系统设计者的角度来说, 针对 MAS 的特点考虑其安全需求是很有必要的。首先, MAS 的每个智能体逻辑或物理上分布的位置特征, 将成为身份验证的一个关键问题。其次, 由于智能体的自治性可能带来的安全问题, 需要每个智能体拥有识别并预防未授权访问的能力, 以此保护 MAS 免受开放网络中恶意智能体的入侵。对于智能体的社交性, MAS 的设计者需要保证沟通过程中的安全性。

### 2.3　多智能体系统的通用安全模型

针对 MAS 系统性能的评估, 很多研究者已经逐渐将系统安全性纳入考虑范围。因此, 对于现在的 MAS 设计人员而言, 有很多通用的安全模型可供参考。

2002 年 Poslad 等[17] 提出了 MAS 资产安全模型, 他们以 FIPA 安全条例为标准, 将 MAS 的数据信息视为一组有价物品, 并通过类比建立了安全措施与数据保护系统的通用模型。这一模型也成为 MAS 通用安全模型的基础, 图 2 给出了这种通用模型的概念图。

此后, 有关 MAS 安全模型的理论不断完善。Tropos 方法[18] 作为智能体软件系统开发的常用方法, 在安全方面被不断完善[19-20], 通过添加安全约束, 成为了 MAS 的软件开发过程中安全建模的重要方法。Beydoun 等[21] 提出了一种 MAS 建模的元模型 (metamodel), Moradian[22] 在元模型基础上基于 Gaia 方法增加了认证模块、搜索模块和检查模块以增强安全性。2015 年 Basheer 等[23] 通过智能体的可信度建立了 AgentOpCo(agent opinion confidence) 模型, 这种通用模型通过检测 MAS 中智能体的置信度来保证系统安全。2006 年 Huynh 等[9] 提出了有关 MAS 中智能体的信任与声誉认证模型, 为子系统验证提供了框架。2016 年, Zikratov 等[24] 设计了动态网络下的信任模型, 由于加入了时间驱动的信任级别认证, 提升了 MAS 在开放网络下的安全性。Cheribi 等[25] 提出了一个用于保护复杂的应用程序 MAS 安全模型。该安全模型规范了智能体个体及智能体群体内的统一行为。Majd 等[26] 则通过设置传递性、符合性、相似性与可靠性 4 个信任组件, 提出了名为 TtSSR 的信任模型。

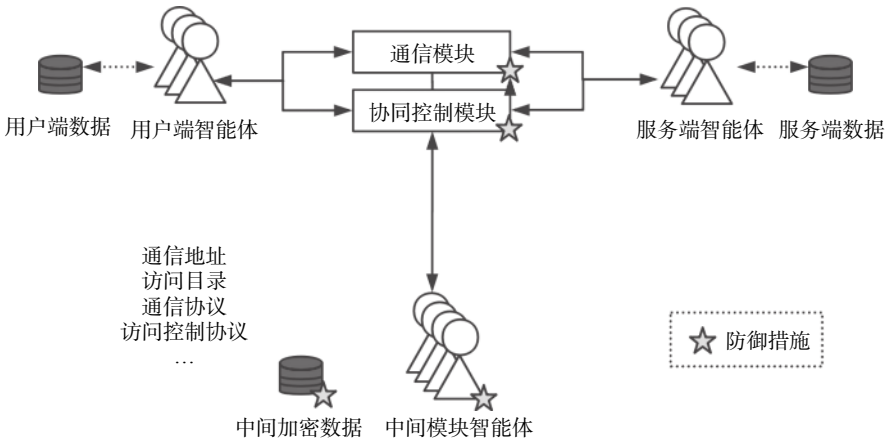


**图 2　多智能体系统较为通用的安全模型**

**Fig. 2　A generic security model of MAS**

除此以外，一些 MAS 的设计平台和开放环境，也在原有的基础上添加了安全性模块。其中最为常用的 JADE 和 SeMoA 平台，已有研究论证其安全性能，并有研究者进行测试以证明这两个平台可以阻止对其智能体的未授权访问和攻击[27-28]。在此基础上，研究人员对 MAS 开放平台进行进一步扩展，使其在拓展应用中保持安全性。2009 年 Vitabile 等[29] 设计了基于 JADE-S 的扩展框架，进一步完善了多智能体系统开放环境的安全性。

## 3　多智能体系统安全防御技术

如前文所述，如今各种研究已经确定了多智能体系统的安全框架。接下来，本文将关注具体的安全解决方案。MAS 的防御技术是根据其系统环节和安全威胁制定的。从图 3 中的系统框图可以看出，安全防御技术主要有针对智能体的防御技术与针对分布式调控的防御技术，分布式调控的防御技术又包括通信保护、访问控制与数据加密等防御手段。

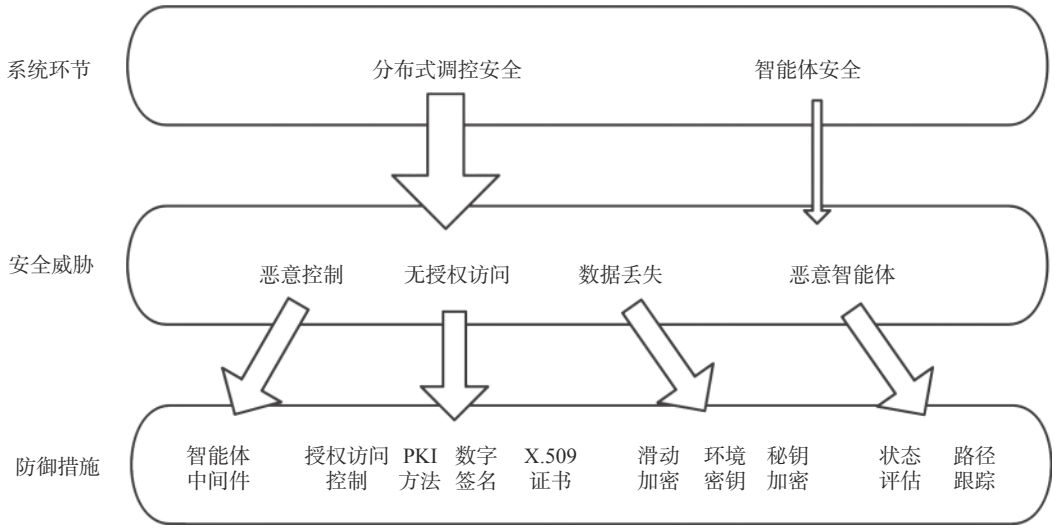


**图 3　多智能体系统的安全需求与防御措施**

**Fig. 3　Security requirements and defense measures of MAS**

### 3.1　针对智能体的防御技术

单个的智能体可以看作一个小型的集中式人工智能系统，除了传统的集中式人工智能保护机制外，防御者可以使用智能体的加密与历史表单数据来防止恶意智能体。

Jansen[30] 的文章很好地概括了智能体的防御技术。沙箱法，又称软件的故障隔离法，将应用程序模块隔离，转换不受信任子系统的可解释代码模块，以便所有存储器访问都限制在其故障域内的代码和数据段，由此确定子系统的安全性。数字签名技术通过智能体的创建者、所有者或可信任的第三方执行数字签名加密，以确保智能体的真实性。路径跟踪技术[31] 使智能体记录它们之前访问过的主机，从而防止恶意主机控制的智

能体破坏 MAS。状态评估法[32] 旨在确定智能体的状态是否已被修改或其信息是否已被破坏, 状态评估依赖于评估函数, 根据条件因素和状态不变量来判断智能体的状态。携带证明口令[33] 的方法使 MAS 能够自动运行由智能体提供的程序代码, 无需解释或运行时检查, 证明口令是系统开发者为子系统预置的安全凭据。

策略随机化是近几年作用于单一智能体的一个新方法。Paruchuri 等[34] 运用滚动随机化 (rolling down randomization,RDR) 算法,在单个智能体中实现策略随机化, 以避免恶意攻击对智能体的可预测性。

### 3.2　针对分布式调控的防御技术

MAS 的分布式调控中, 除了关注每个子系统的安全防御问题, 对于全局的安全控制, 系统设计者也需要合适的防御手段。本文主要研究了协同控制、数据交换、访问控制以及群体身份验证等几个阶段的防御技术。

#### 3.2.1　协同控制安全防御

为了保证各自独立的智能体共同协作实现目标, MAS 的设计者需要考虑对智能体进行协同控制。因此, 一致性算法成为 MAS 协同控制问题的核心内容。多智能体系统中, 一致性算法保证了网络有向信息流的作用, 对链路与节点故障引起的网络拓扑变化的鲁棒性以及通信时延下的系统新能提供了支持。在此过程中, 恶意攻击可能影响复杂网络的频率和结构特性之间的直接联系以及一致性算法的信息扩散速度, 对具有非局部信息流的网络系统的性能造成危害[35]。因此, 近年来 MAS 一致性的安全成为该领域一个新的热点。

很多研究者关注 MAS 一致性算法和结构的设计。He 等[36] 提出了一种分布式脉冲控制器, 研究了在数据欺骗攻击下, 错误数据注入引起的有界协同控制问题。Amullen 等[37] 提出了一种基于模型的安全编队控制算法, 该算法对一组智能体进行编队并记录其状态, 通过状态恢复的方法使得 MAS 不受 DoS 攻击的影响。Torre 等[38] 开发了一个新的分布式自适应控制架构, 当存在行为不当的智能体的情况下, 算法会启用本地状态仿真器对其进行控制。

其他研究从通信延时、有限时间、量化通信等几个条件[39] 优化 MAS 一致性算法, 使得 MAS 的性能可以不受特定攻击的影响。针对通信延时的情况, 2011 年 Liu 等[40] 提出了时延和噪声干扰下的多智能体系统的安全一致性算法。Wu 等[41] 分析了通信延时下的非线性 MAS 的安全一致性问题, 本文根据相邻节点的延迟信息设计了延迟鲁棒安全一致性 (delay robust secure consensus, DRSC) 算法。对于有限时间问题, Meng 等[42] 通过智能体协作和对抗, 使用邻近规则构造了两个一致性协议。结果表明, 该协议可以保证所有智能体在有限时间内达成一致性。作者在随后的研究中, 提出了另一种分布式非线性协议[43], 实现了 MAS 在固定时间内的安全一致性。Liu 等[44] 设计了一个集中式切换的一致性协议, 实现了快速的有限时间一致性问题。迭代学习控制 (iterative learning control, ILC) 被应用于 MAS 一致性研究后[45], 成为智能体安全一致性控制的新方法。为了应对恶意攻击, 随后的研究[46] 基于 ILC 实现了有限时间内的智能体一致性控制。对于量化通信下的安全一致性, 2016 年的一种基于量化数据的安全控制算法[47], 通过确定攻击节点在邻域中的有界性, 在有恶意智能体的情况下达到安全一致性。此外, Feng 等[48] 在连接维持和连接断开两类攻击下, 研究了一种安全一致性追踪算法。作者通过定向拓扑交换对问题进行了建模和探究, 并在之后的研究中, 用此方法研究了 DoS 攻击下的 MAS 安全一致性问题[49-50]。

#### 3.2.2　数据交换安全防御

针对数据交换的保护首先需要设计合理的通信结构和通信协议。

Pitt 等[51] 定义了 MAS 通信的通用语义框架, 并提出了智能体通信语言 (agent communation language, ACL)。ACL 被 FIPA 纳入 MAS 通信框架标准之后, FIPA-ACL 也成为 MAS 数据交换和通信协议的规范准则[13]。在此基础上, 之后的研究添加了安全性内容。

Yokoo 等[52] 设计了动态加密的 MAS 通信协议。通过使用这种方法, 多个智能体可以在解决组合优化问题的同时, 避免泄露隐私信息。Zhu 等[53] 基于传统代理通信语言, 提出了一种多智能体安全通信协议和 3 种多智能体组重新密钥协议, 并提出了一种更加完善的 ACL。Abdennadher 等[54] 在 2010 年提出了 MAS 动态通信协议, 它使 MAS 能够在不完美的通信环境中灵活, 快速且顺畅地进行通信。2017 年提出的一种基于广播的移动智能体协议 BROSMAP(broadcast based secure mobile agent protocol)[55], 使用对称和非对称混合加密的方法保证了分布式应用安全性。Elshaafi 等[56] 实现了 JADE 平台通用协议的实例化, 并为 JADE 平台添加了 JAVA 安全库, 使得 JADE 平台

支持加密和身份验证功能。

除了可靠的通信协议之外，许多研究还关注数字签名和 PKI 方法[57-58]，以确认通信地址的真实性与准确性。

3.2.3　访问控制

访问控制策略分为自主访问控制 (discretionary access control, DAC) 和强制访问控制 (mandatory access control, MAC)。MAC 基于系统确定的强制性规则。在 DAC 中，授权由子系统自行决定，子系统是某些资源的控制者或所有者。MAS 的访问控制中，需要 DAC 和 MAC 相结合的访问控制策略。

Tekbacak 等[59] 提出了基于 XACML 的 MAS 平台访问控制方法，通过制定 XACML 策略文档，检查智能体或代码的访问控制是否符合要求。Vitabile 等[60] 使用策略和所有者访问控制，通过信誉机制来阻止可信代理与恶意代理交互。Wang 等[61] 则由基于的信任访问控制 (TACA) 和角色访问控制 (RACA) 组成的安全机制，通过子系统和用户的信任与角色分配建立授权。Xiao 等[62] 结合了角色访问控制 (RACA) 的安全管理方法和面向代理的软件工程 (AOSE) 中的代理角色扮演行为的需求。捕获功能和非功能需求的模型由业务专家在高级抽象中持续维护。最终，代理商从最新模型中解释其约束行为。

3.2.4　群体身份验证方法

为了防止恶意代码伪装控制的威胁，除了对单个智能体的身份验证，还需要建立智能体之间广泛和开放的信任关系。

Robles 等[63] 提出了用于管理移动通信网络资源的多智能体系统的通用信任架构，结合 MAS 的市场应用，从公平性和可靠性的角度出发，为市场提供 MAS 的基本信任框架。Poggi 等[64] 提出了智能体系统中间件的信任机制问题。Noordende 等[65] 提出 Mansion 多层中间件系统，旨在为大规模智能体系统提供身份验证服务。Chae 等[66] 将认证证书和通信密钥组合在一起，提出一种智能体相互认证方法。在之后的研究中，研究者提出了信任社区[67] 的概念，将 MAS 开发平台的信任机制嵌入智能体本身内，在智能体之间建立信任评估体系。Fagiolini 等[68] 设计了分散式入侵检测方法，应对 MAS 独立任务的智能体之间的安全监视与身份验证。解决方案中每个智能体仅使用本地可用信息运行本地监视器，从而保证状态监测与身份验证的独立性。Such 等[69] 设计了面向群组安全的多智能体平台，作者提出的 Magentix MAP 的安全基础结构支持智能体组的身份验证并可以保护子系统的身份隐私。

## 4　多智能体系统安全的未来发展

尽管 MAS 的安全性问题已经逐渐受到系统开发人员的重视，但是仍然有尚未解决的问题。此外，随着 MAS 的不断发展，也会有新的安全问题不断涌现出来。

首先，智能体的自治性很难掌握。如果系统开发人员为 MAS 预设定访问授权协议、通信协议等，智能体有自主选择不遵循的权力。强制性的系统控制会很大程度影响到智能体即 MAS 的性能。

其次，系统的分布性、全局的信任机制在实现上有一定的困难。这是因为在智能体各司其职的情况下，每个智能体不会与其他所有智能体相关联或协作，因此很难获得其他智能体的完整信息，这种安全信息的缺失和 MAS 分布性的平衡很难处理。

此外，在 MAS 的不同应用上，MAS 安全问题需要考虑的信任机制是不同的。如何在通用的安全模型的基础上，建立通用的评估体系，是很有必要的。

因此，在未来的相关研究中，需要解决多智能体系统全局信任机制的问题并设计通用的评估体系。随着分布式安全这一概念[70] 的提出，如何利用分布式优化合理地进行智能体资源分配与智能体社区协作，将为 MAS 的安全协作提供很好的思路。也就是说，如何利用分布式方法，解决分布式系统的安全问题，将成为多智能体系统安全性问题未来的研究重点。

## 5　结束语

随着大数据和云计算的快速发展，分布式人工智能将为生产生活中的各个领域提供重要方法。越来越多的 MAS 已经成为不同领域应用开发的重要方法。因此，确保基于智能体的系统环境安全性的问题非常关键，MAS 的安全问题和防御机制也受到了学术界高度重视。本文对 MAS 安全问题的现有相关工作进行了综述和分类。在现有的安全问题中，访问控制和智能体信任机制问题仍是研究的重点，并且随着新技术的发展不断出现新的挑战。通过对 MAS 安全问题发展过程的调查和分析，本文对当前 MAS 安全问题及其防御技术研究做出如下总结：

1) 分布式人工智能仍面临诸多风险, 新的 MAS 安全威胁可能涌现。由于 MAS 的工作机制, 智能体需要在开放环境中协作, 一方面, 智能体的自治性给系统安全带来了不确定性, 另一方面, MAS 在很大程度上依赖于通信, 并且访问控制易受影响。因此, 潜在的安全威胁可能造成非法资源占用、窃取隐私数据、恶意控制甚至停机等现象。

2) MAS 安全评估依旧是难点所在。由于 MAS 的系统复杂性, MAS 设计人员将需要重视其系统安全性能评估。通用的安全评估标准, 需要进行统一的量化, 以便 MAS 防御体系的应用推广。

3) 在安全的 MAS 设计过程中, 应合理平衡算法的安全性和实用性, 在尽可能少地牺牲 MAS 系统性能的情况下建立安全机制, 以更好地满足实际应用需求。

4) 为 MAS 建立合理的信任机制是如今 MAS 研究的热点之一。利用智能体的社交性, 建立智能体社区信任评价体系, 在保证系统性能的同时, 减小安全隐患。

5) 充分考虑分布式方法在 MAS 安全性问题中的应用, 通过建立分布式安全模型, 增强 MAS 安全性与自我管理功能的同时, 实现采用 MAS 方法保护分布式系统的目标。

## 参考文献:


[1] GREENBERG M S, BYINGTON J C, HARPER D G. Mobile agents and security[J]. IEEE communications magazine, 1998, 36(7): 76–85.

[2] CENTENO R, FAGUNDES M, BILLHARDT H, et al. Supporting medical emergencies by MAS[C]//Proceedings of the 3rd Kes International Symposium on Agent and Multi-Agent Systems: Technologies and Applications. Uppsala, Sweden, 2009: 823–833.

[3] DIOUBATE M, TAN Guanzheng, MOHAMED L T. An artificial immune system based multi-agent model and its application to robot cooperation problem[C]//Proceedings of the 2008 7th World Congress on Intelligent Control and Automation. Chongqing, China, 2008: 3033–3039.

[4] SZYMAK P, PRACZYK T. Control systems of underwater vehicles in multi-agent system of underwater inspection[C]//Proceedings of the 11th WSEAS International Conference on Automatic Control, Modelling and Simulation. Istanbul, Turkey, 2009: 153–156.

[5] HUANG Xiao, ZHANG Qi, WANG Yu. Research on multi-agent traffic signal control system based on VANET information[C]//Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems. Yokohama, Japan, 2018: 1–6.

[6] 姚尧, 卢淑娟, 徐德民. IMMAS 算法在复杂环境下路径规划中的应用 [J]. 计算机仿真, 2007, 24(12): 148–151.
YAO Yao, LU Shujuan, XU Demin. IMMAS for robot path planning in complex environment[J]. Computer simulation, 2007, 24(12): 148–151.

[7] IQBAL S, ALTAF W, ASLAM M, et al. Application of intelligent agents in health-care: review[J]. Artificial intelligence review, 2016, 46(1): 83–112.

[8] BIJANI S, ROBERTSON D. A review of attacks and security approaches in open multi-agent systems[J]. Artificial intelligence review, 2014, 42(4): 607–636.

[9] HUYNH T D, JENNINGS N R, SHADBOLT N R. An integrated trust and reputation model for open multi-agent systems[J]. Autonomous agents and multi-agent systems, 2006, 13(2): 119–154.

[10] WANG S, HU J, LIU A, et al. Security frame and evaluation in mobile agent system[C]. International Conference on Mobile Technology Applications and Systems. Guangzhou, China, 2005: 1–6.

[11] JUNG Y, KIM M, MASOUMZADEH A, et al. A survey of security issue in multi-agent systems[J]. Artificial intelligence review, 2012, 37(3): 239–260.

[12] Foundation for Intelligent Physical Agents (FIPA). FIPA Abstract Architecture Specification 2002 [EB/OL]. (2002-02-25)[2018-12-01]http://www.fipa.org/specs/fipaoo001/SC00001L.pdf.

[13] Foundation for Intelligent Physical Agents (FIPA). FIPA ACL Message Structure Specification [EB/OL]. (2000-01-08)[2018-12-01]http//www.fipa.org/specs/fipa00061/SC00061G.html.

[14] Foundation for Intelligent Physical Agents (FIPA). FIPA agent management specification [EB/OL]. (2004-03-18)[2018-12-01]http://www.fipa.org/specs/fipaO0023/SCO0023K.pdf.

[15] MOURATIDIS H, GIORGINI P, MANSON G. Modelling secure multiagent systems[C]//Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems. Melbourne, Australia, 2003: 859–866.

[16] BRESCIANI P, GIORGINI P, MOURATIDIS H, et al. Multi-agent systems and security requirements analysis[M]// LUCENA C, GARCIA A, ROMANOVSKY A, et al. Software Engineering for Multi-Agent Systems II. Berlin, Germany: Springer, 2004: 35–48.

[17] POSLAD S, CHARLTON P, CALISTI M. Specifying standard security mechanisms in multi-agent systems[M]//

FALCONE R, BARBER S, KORBA L, et al. Trust, Reputation, and Security: Theories and Practice. Berlin, Germany: Springer, 2003: 163–176.

[18] BRESCIANI P, PERINI A, GIORGINI P, et al. Tropos: an agent-oriented software development methodology[J]. Autonomous agents and multi-agent systems, 2004, 8(3): 203–236.

[19] MOURATIDIS H, GIORGINI P. Secure tropos: a security-oriented extension of the tropos methodology[J]. International journal of software engineering and knowledge engineering, 2007, 17(2): 285–309.

[20] MOURATIDIS H, GIORGINI P. Enhancing secure tropos to effectively deal with security requirements in the development of multiagent systems[M]//BARLEY M, MOURATIDIS H, UNRUH A, et al. Safety and Security in Multiagent Systems. Berlin: Springer, 2009: 8–26.

[21] BEYDOUN G, LOW G, MOURATIDIS H, et al. A security-aware metamodel for multi-agent systems (MAS)[J]. Information and software technology, 2009, 51(5): 832–845.

[22] MORADIAN E. Security of E-commerce software systems[M]//HAKANSSON A, HARTUNG R. Agent and Multi-Agent Systems in Distributed Systems-Digital Economy and E-Commerce. Berlin: Springer, 2013: 95–103.

[23] BASHEER G S, AHMAD M S, TANG A Y C, et al. Certainty, trust and evidence: towards an integrative model of confidence in multi-agent systems[J]. Computers in human behavior, 2015, 45: 307–315.

[24] ZIKRATOV I, MASLENNIKOV O, LEBEDEV I, et al. Dynamic trust management framework for robotic multi-agent systems[C]//Proceedings of the 16th International Conference on Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Petersburg, Russia, 2016.

[25] CHERIBI H, KHOLLADI M K. A security model for complex applications based on normative multi-agents system[C]//Proceedings of the 2015 2nd International Conference on Information Security and Cyber Forensics. Cape Town, South Africa, 2015:41–46.

[26] MAJD E, BALAKRISHNAN V. A trust model for recommender agent systems[J]. Soft computing — a fusion of foundations, methodologies and applications, 2017, 21(2): 417–433.

[27] BÜRKLE A, HERTEL A, MÜLLER W, et al. Evaluating the security of mobile agent platforms[J]. Autonomous agents and multi-agent systems, 2009, 18(2): 295–311.

[28] VILA X, SCHUSTER A, RIERA A. Security for a multi-agent system based on JADE[J]. Computers & security, 2007, 26(5): 391–400.

[29] VITABILE S, CONTI V, MILITELLO C, et al. An extended JADE-S based framework for developing secure Multi-Agent Systems[J]. Computer standards & interfaces, 2009, 31(5): 913–930.

[30] JANSEN W A. Countermeasures for mobile agent security[J]. Computer communications, 2000, 23(17): 1667–1676.

[31] ROTH V. Secure recording of itineraries through co-operating agents[C]//Proceedings of 1998 European Conference on Object-Oriented Programming. Belgium, Germany, 1998: 297–298.

[32] FARMER W M, GUTTMAN J D, SWARUP V. Security for mobile agents: authentication and state appraisal[C]//Proceedings of the 4th European Symposium on Research in Computer Security. Rome, Italy, 1996: 118–130.

[33] NECULA G C, LEE P. Safe, untrusted agents using proof-carrying code[M]//VIGNA G. Mobile Agents and Security. Berlin: Springer, 1998: 61–91.

[34] PARUCHURI P, TAMBE M, ORDÓÑEZ F, et al. Security in multiagent systems by policy randomization[C]//Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems. Hakodate, Japan, 2006.

[35] OLFATI-SABER R, FAX J A, MURRAY R M. Consensus and cooperation in networked multi-agent systems[J]. Proceedings of the IEEE, 2007, 95(1): 215–233.

[36] HE Wangli, GAO Xiaoyang, ZHONG Weimin, et al. Secure impulsive synchronization control of multi-agent systems under deception attacks[J]. Information sciences, 2018, 459: 354–368.

[37] AMULLEN E M, SHETTY S, KEEL L H. Secured formation control for multi-agent systems under DoS attacks[C]//Proceedings of 2016 IEEE Symposium on Technologies for Homeland Security. Waltham, USA, 2016.

[38] DE LA TORRE G, YUCELEN T, PETERSON J D. Resilient networked multiagent systems: a distributed adaptive control approachy[C]//Proceedings of the 53rd IEEE Conference on Decision and Control. Los Angeles, USA, 2014.

[39] 伍益明. 恶意攻击下的多智能体系统安全一致性问题研究 [D]. 杭州: 浙江工业大学, 2016.
WU Yiming. Research on secure consensus for multi-agent systems under malicious attacks[D]. Hangzhou: Zhejiang University of Technology, 2016.

[40] LIU S, XIE L, ZHANG H, et al. Distributed consensus for multi-agent systems with delays and noises in transmission channels[J]. Automatica, 2011, 47(5): 920–934.

[41] WU Yiming, HE Xiongxiong. Secure consensus control for multiagent systems with attacks and communication delays[J]. IEEE/CAA journal of automatica sinica, 2017, 4(1): 136–142.

[42] MENG Deyuan, JIA Yingmin, DU Junping. Finite-time consensus for multiagent systems with cooperative and antagonistic interactions[J]. IEEE transactions on neural networks and learning systems, 2016, 27(4): 762–770.

[43] MENG Deyuan, ZUO Zongyu. Signed-average consensus for networks of agents: a nonlinear fixed-time convergence protocol[J]. Nonlinear dynamics, 2016, 85(1): 155–165.

[44] LIU Xiaoyang, LAM J, YU Wenwu, et al. Finite-time consensus of multiagent systems with a switching protocol[J]. IEEE transactions on neural networks and learning systems, 2016, 27(4): 853–862.

[45] LI Jinsha, LI Junmin. Adaptive iterative learning control for coordination of second-order multi-agent systems[J]. International journal of robust and nonlinear control, 2014, 24(18): 3282–3299.

[46] WU Yiming, XU Ming, ZHENG Ning, et al. Attack tolerant finite-time consensus for multi-agent networks[C]// Proceedings of the 2017 13th IEEE International Conference on Control & Automation. Ohrid, Macedonia, 2017.

[47] WU Yiming, HE Xiongxiong, LIU Shuai. Resilient consensus for multi-agent systems with quantized communication[C]//Proceedings of 2016 American Control Conference. Boston, USA, 2016.

[48] FENG Zhi, HU Guoqiang, WEN Guanghui. Distributed consensus tracking for multi-agent systems under two types of attacks[J]. International journal of robust and nonlinear control, 2016, 26(5): 896–918.

[49] FENG Zhi, WEN Guanghui, HU Guoqiang. Distributed secure coordinated control for multiagent systems under strategic attacks[J]. IEEE transactions on cybernetics, 2017, 47(5): 1273–1284.

[50] FENG Zhi, HU Guoqiang. Distributed secure leader-following consensus of multi-agent systems under DoS attacks and directed topology[C]//Proceedings of 2017 IEEE International Conference on Information and Automation. Macau, China, 2017.

[51] PITT J, MAMDANI A. Communication protocols in multi-agent systems: a development method and reference architecture[M]//DIGNUM F, GREAVES M. Issues in Agent Communication. Berlin: Springer, 2000: 160–177.

[52] YOKOO M, SUZUKI K. Secure multi-agent dynamic programming based on homomorphic encryption and its application to combinatorial auctions[C]//Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems. Bologna, Italy, 2002.

[53] ZHU Liehuang, CAO Yuanda, LIAO Lejian, et al. SKQML: a secure multi-agent communication language [C]//Proceedings of the 5th WSEAS International Conference on Applied Computer Science. Hangzhou, China, 2006.

[54] ABDENNADHER S, ABDALLAH M, MUELLER H J. DCP-MASR: a dynamic communication protocol for robot-based multi-agent systems[C]//Proceedings of 2010 5th International Conference on Systems. Menuires, France, 2010: 12–17.

[55] SHEHADA D, YEUN C Y, ZEMERLY M J, et al. BROSMAP: a novel broadcast based secure mobile agent protocol for distributed service applications[J]. Security and communication networks, 2017, 2017: 3606424.

[56] ELSHAAFI H, VINYALS M, GRIMALDI I, et al. Secure automated home energy management in multi-agent smart grid architecture[J]. Technology and economics of smart grids and sustainable energy, 2018, 3(1): 4.

[57] HU Y J, TANG Chaowei. Agent-oriented public key infrastructure for multi-agent e-service[C]//Proceedings of the 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems. Oxford, UK, 2003: 1215–1221.

[58] LU Feng, HUANG Mei. Research and design of security in Multi-agent System[C]//Proceedings of 2016 IET International Conference on Wireless, Mobile and Multimedia Networks. Hangzhou, China, 2006: 1–4.

[59] TEKBACAK F, TUGLULAR T, DIKENELLI O. An architecture for verification of access control policies with multi agent system ontologies[C]//Proceedings of the 2009 33rd Annual IEEE International Computer Software and Applications Conference. Seattle, USA, 2009: 52–55.

[60] VITABILE S, MILICI G, SCOLARO S, et al. A MAS security framework implementing reputation based policies and owners access control[C]//Proceedings of the 20th International Conference on Advanced Information NETWORKING and Applications. Vienna, Austria, 2006: 746–752.

[61] WANG Xuan, YAN Jinglong. Research on security mechanism for Multi-Agent system[C]//Proceedings of the 2011 IEEE 3rd International Conference on Communication Software and Networks. Xi'an, China, 2011: 345–348.

[62] XIAO Liang, HU Bo. Towards adaptive and secure multiagent systems[C]//Proceedings of the 2007 2nd International Conference on Pervasive Computing and Applications. Birmingham, UK, 2007: 56–61.

[63] ROBLES S, BORRELL J, BIGHAM J, et al. Design of a trust model for a secure multi-agent marketplace[C]//Proceedings of the 5th International Conference on Autonomous agents. Montreal, Canada, 2001: 77–78.

[64] POGGI A, TOMAIUOLO M, VITAGLIONE G. Security and trust in agent-oriented middleware[C]//Proceedings of 2003 on the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops. Sicily, Italy, 2003: 989–1003.

[65] VAN 'T NOORDENDE G J, BRAZIER F M T, TANENBAUM A S. Security in a mobile agent system[C]//Proceedings of the IEEE 1st Symposium on Multi-Agent Security and Survivability. Drexel, USA, 2004: 35–45.

[66] CHAE C J, CHOI K N, CHOI K. Information interoperability system using multi-agent with security[J]. Wireless personal communications, 2016, 89(3): 819–832.

[67] JONES K I. A trust based approach to mobile multi-agent systems[D]. Leicester: De Montfort University, 2010.

[68] FAGIOLINI A, VALENTI G, PALLOTTINO L, et al. Decentralized intrusion detection for secure cooperative multi-agent systems[C]//Proceedings of the 2007 46th IEEE Conference on Decision and Control. New Orleans, USA, 2007.

[69] SUCH J M, ALBEROLA J M, ESPINOSA A, et al. A group-oriented secure multiagent platform[J]. Software: practice and experience, 2011, 41(11): 1289–1302.

[70] RASHVAND H F, SALAH K, CALERO J M A, et al. Distributed security for multi-agent systems-review and applications[J]. IET information security, 2010, 4(4): 188–201.

**作者简介：**

丁俐夫，博士研究生，主要研究方向为多智能体系统、分布式优化算法。

颜钢锋，教授，博士生导师，中国系统工程学会理事，主要研究方向为多智能体系统。发表学术论文 100 余篇。

# CAAI 获评 2020 年“全国科技工作者日”全国学会十佳优秀组织单位

为响应中国科学技术协会的号召，庆祝第四个“全国科技工作者日”，中国人工智能学会在疫情特殊时期，积极探索线上学术交流新模式，凭借一系列精心策划并组织的高质量活动，获评中国科协颁发的 2020 年“全国科技工作者日”全国学会十佳优秀组织单位荣誉称号。在为全国科技工作者日增色添彩的同时，也有力弘扬了新时代科学家精神，展现了中国科技工作者的良好精神风貌。

在此期间，学会为会员和 AI 从业者开设 CAAI 云课堂、CAAI 云论坛系列活动，惠及百万人；为响应“让科学家精神光耀时代，让科技创造新的价值”的号召，开展中国人工智能学会优秀科技工作者风采及成果评选工作，经评审，分别有 13 位科技工作者、14 项科技成果获得荣誉；为体现“科技为民，奋斗有我”的决心，特别推出限时一周免费入会活动，收获满满赞誉。

中国人工智能学会

2020 年 08 月 30 日