IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 October 2022, accepted 14 November 2022, date of publication 18 November 2022,
date of current version 23 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223362

SURVEY

Security of Multi-Agent Cyber-Physical Systems: A Survey

RICHARD OWOPUTI, (Student Member, IEEE), AND SANDIP RAY, (Senior Member, IEEE)
Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

Corresponding author: Richard Owoputi (rowoputi@ufl.edu)

This work was supported by the National Science Foundation under Grant CNS-1908549.

ABSTRACT Multi-agent systems are becoming increasingly popular due to their successful implementation in several sectors. However, there are a variety of threats that might undermine the agent's security and imperil system security. As a result, security concerns should be addressed during the design of multi-agent systems. This survey reviews different models for securing multi-agent systems, which were developed based on the concepts regarding the agent's role and communications. This paper presents and categorizes the most common attacks on MASs. Then, we study and analyze numerous security strategies in the literature, classifying them as prevention, detection, and resiliency approaches based on reputation and trust. Finally, we recommend which security approach is the best countermeasure for specific types of attacks based on recent advances in the research field.

INDEX TERMS Cyber physical system, multi agent systems, resiliency, security.

I. INTRODUCTION

The term *multi-agent system* (MAS) is used to refer to a cyber-physical system composed of a collection of autonomous entities (or *agents*) that collaborate to solve a task. The term ''agent'' here denotes anything that perceives its environment and takes actions autonomously without direct or continuous supervision from any centralized control. The use of multiple agents (rather than a single, centralized entity) induces additional flexibility resulting from the possibility for different agents to perceive different aspects of the environment and make independent judgments while facilitating coordination and knowledge sharing [1], [2], [3]. Furthermore, using multiple agents induces redundancy and heterogeneity that fosters robustness against failure of individual components [4].

With the increasing advancement of Artificial Intelligence, connectivity, sensors, and robotics, the use of MASs is emerging into many critical applications, including military, space, manufacturing, electronic business, supply chain management, and many others. However, these agents move in uncertain and adversarial environments, making their activities unreliable and unsafe. It is challenging to detect malicious agents that transmit harmful communication messages. These malicious agents can lead to undesirable effects like the leakage of sensitive information or the destruction of the agents in a mission-critical scenario. It is imperative to provide security mechanisms to guarantee the tenets of security such as the confidentiality, integrity, availability, accountability, and non-repudiation of the various autonomous agents and systems that might face [5].

Over the last decade, there has been significant research on various facets of MAS security. This includes exploration and identification of various attack mechanisms on the one hand and novel techniques to mitigate such attacks on the other. Nevertheless, — and despite its great need — we did not find a comprehensive survey of the area that provides a holistic view of the challenges, progress made in the research, and limitations of the current state of the art.

The goal of this paper is to provide a comprehensive, systematic overview of the research in the security of MASs. We develop a systematic categorization of research advances in various aspects of both attacks and defenses, point out the constraints and tradeoffs within which they must operate, and state-of-the-art limitations.

The associate editor coordinating the review of this manuscript and approving it for publication was Engang Tian.

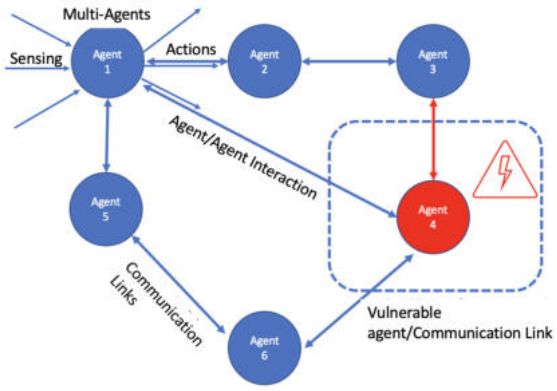

FIGURE 1. An example MAS containing a vulnerable agent.

The rest of the paper is organized as follows. Section II gives a tutorial overview of MASs and various categorizations developed from the perspective of security. We review the attacks on MASs in Section III. We divide the discussion of security solutions or ''defenses'' into three categories: *prevention* (Section IV-A), *detection* (Section IV-B) and *resiliency* (Section IV-C). We conclude in Section V.

II. MULTI-AGENT SYSTEMS

Russell and Norvig [6] defines an agent as ''a flexible, autonomous entity capable of perceiving the environment through the sensors connected to it.'' The agent senses different parameters that are used to make a decision based on the goal of the entity. This definition of agents is based on several keywords, e.g., ''entities,'' ''environment,'' and ''parameters.'' Each *agent* aims to complete its assigned work while adhering to certain additional constraints, such as a deadline. To achieve this objective, the agent first senses *parameters* from the *environment*. Vested with this data, the agent can accumulate knowledge about the environment. An agent might also use the knowledge of its neighbors. This knowledge, along with the record of the previous actions taken and the goal, is fed to an extrapolation engine that decides on the agent's appropriate action.

Dorri et al. [7] discuss the features that enabled agents to solve complex tasks. These features include sociability, autonomy, and proactivity. *Sociability* is the ability of the agents to socialize with other agents. *Autonomy* allows the agents to make autonomous decisions. *Proactivity* provides agents with the ability to predict future actions using their history of past occurrences, information from other agents, and sensed parameters. In general, MASs attempt to solve complicated issues through agent collaboration. Furthermore, many MASs are deployed without centralized control, and their data is often decentralized. To strengthen MAS security, it is necessary to address the security of individual agents and the security of the interaction between the various agents. To fully comprehend agents, [8], [9] represented MASs as a graph, with the vertices of the network representing

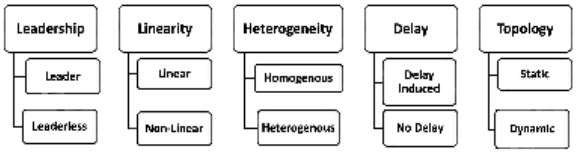

FIGURE 2. Characteristics of multi agent systems.

autonomous agents and the edges of the graph representing communication links between the agents.

Different previous papers have developed different ways to classify MASs in terms of security features. The different classifications expose different facets of MAS designs and applications in the context of security. Here we briefly review some of these different categorizations to give a flavor of the different contexts. Nevertheless, in our own description of attacks (Section III), we eschew the classifications based on MAS features and find it convenient to develop a taxonomy of the attacks themselves.

Dorri et al. [7] consider the *architectural characteristics* of MASs and their role in security. They characterize MASs based on the following features.

- *Leadership.* A MAS with a leader enables ''centralized decision-making'' in the sense that actuation or conflict resolution is controlled by one central agent. In contrast, a leaderless MAS would operate through consensus.
- *Linearity.* In *linear MAS*, an agent's behavior is based on environmental characteristics that are perceived by the system. An example of this is when a robot in a multi-robot system detects a direct barrier and pauses. The agent behaviors in a nonlinear system are not exclusively reliant on the environmental characteristics that are perceived.
- *Heterogeneity.* A system is *homogeneous* if each agent has the same features and functionalities, heterogeneous otherwise. *Heterogeneous* agents have a variety of characteristics, actions, and rules. For instance, a sensor network where protocols only depend on how many sensors send any given signal is an homogeneous system
- *Delay.* The agents in the MAS might take delay into consideration when performing their tasks. In a scenario without delay, the MAS does not consider processing and communication time.
- *Topology.* The topology of a MAS can be *static* or *dynamic*. In a static topology, the position and relations of an agent remain unchanged over the lifetime of the agent. In dynamic topology, the agent's location and connections change while the agents form new communication and move from one place to another.

On the other hand, there have been efforts to create taxonomy of MASs based on security vulnerabilties. Trcek [10] defined the role of cybersecurity to be the minimization of vulnerabilities of assets and resources. The term ''vulnerabilities'' is generally attributed to a weakness in the system that, if discovered by an adversary, might be exploited, causing loss and damage to the system [5].

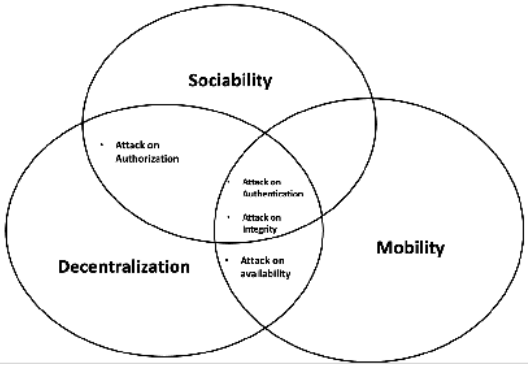

FIGURE 3. Vulnerabilities inducing features of multi-agent systems.

Vulnerabilities can also be defined as a secret passageway enables an adversary control in a system to perform malicious actions. Pfleeger et al. [11] define threats as the internal and external factors and circumstances that have the potential to cause damage to the system. An attack has been orchestrated by exploiting a vulnerability in the design. Research has shown that apparently innocuous errors or vulnerabilities on an agent can degrade the performance and even paralyze the whole MAS [12]. The following classification define MAS features that give rise to orthogonal (and complementary) vulnerabilities.

- *Sociability.* The ability of agents in the system to communicate with other agents and use this information gotten from these agents for decision-making makes the system vulnerable to malicious entities that can share falsified information or subvert the communication agents, e.g., by corrupting messages in transit.
- *Decentralization.* The absence of a centralized controller to verify the identity of the individual agents or verifying the legitimacy of the messages in transit leaves the system vulnerable to masquerading attacks by rogue entities.
- *Mobility.* A mobile agent that has been compromised by an adversary can attack other agents indirectly by sharing compromised messages to other agents in encounters thereby disrupting the system. It can also directly attack the agent it encounters.

Finally, we can also characterize MASs based on the type of subversions. The following catetorizations adapt traditional security classifications to MASs. Fig. 2 depicts the relationship between the vulnerabilities inducing characteristics of MAS and various attacks on the system's security requirement. We need to survey the possible attacks that have been recognized and implemented against multi-agent systems.

- *Authentication*: This guarantees that each agent is what it claims to be.
- *Authorization*: This provides assurance that the agent has legitimate right to have access to what it requires
- *Integrity*: This provides an assurance that the messages in transit have not been mutated since it was generated.
- *Availability*: The services and resources including the messages transmitted are available to the authorized agents in the system.
- *Confidentiality*: This ensures that only permitted agents have access to the resources and services, that is, only authorized agents can read the particular data.

III. ATTACKS ON MULTI-AGENT SYSTEMS

Fig. 4 shows a taxonomy of various attacks on MASs. We classify an attack according to its impact on the victim, effect on communication (which we call "Attack Operation"), origin, victim, and frequency. Our taxonomy can be viewed as a complete black-box classification of MAS attacks. It is *black-box* in the sense that we deliberately eschew the *mechanism* of the attack and focus on categorizing attacks based on the features of its manifestation that can be used by an external observer to describe the attack. It is *complete* in the sense that the external effect on any attack can be described with the features specified in the taxonomy. Furthermore, combining these five characteristics ensures encapsulation and precision in the resulting attack space required for a comprehensive evaluation. We analyze the usefulness of our resilient technique by listing numerous example attacks as a combination of the five identifying criteria. In the description below, we use this taxonomy to systematically explore research on MAS attacks.

A. ATTACK OPERATION

1) DISCLOSURE ATTACKS

A disclosure attack is an attack on confidentiality, i.e., the attacker accesses confidential information from agents. An obvious way to achieve this is by intercepting sensitive messages in transit. An adversary can also gain authorized access to the agent's data, such as the state and the internal code of the agents. Provenance attacks such as disclosing a mobile agent's itinerary information to an adversary are other attacks that the agents should be cautious about. Other, more sophisticated disclosure attacks are probing attacks, where the adversary probes the victim's private database of confidential information. One specific type of probing attack is the ontology attack, which works as follows. An ontology is generally defined as "a formal, explicit specification of a shared conceptualization" [13]. Ontologies are widely used by agents in MASs for sending queries; this is achieved via a semi-open ontology-based system. Obviously, all agents must have the same understanding of the ideas communicated in the messages to interpret them consistently. The adversary can subvert this process and actively probe the agents by accessing the victim agent's private local knowledge (e.g., decision rules and policies) and injecting facts into the agent's knowledge base, asking queries, and evaluating results. Other probing attacks have also been explored in recent research [14], [15], [16]. In particular, Bijani et al. [14] demonstrated four probing attacks on MASs controlled by electronic institutions: explicit query, implicit query, injection, and indirect query attacks. These

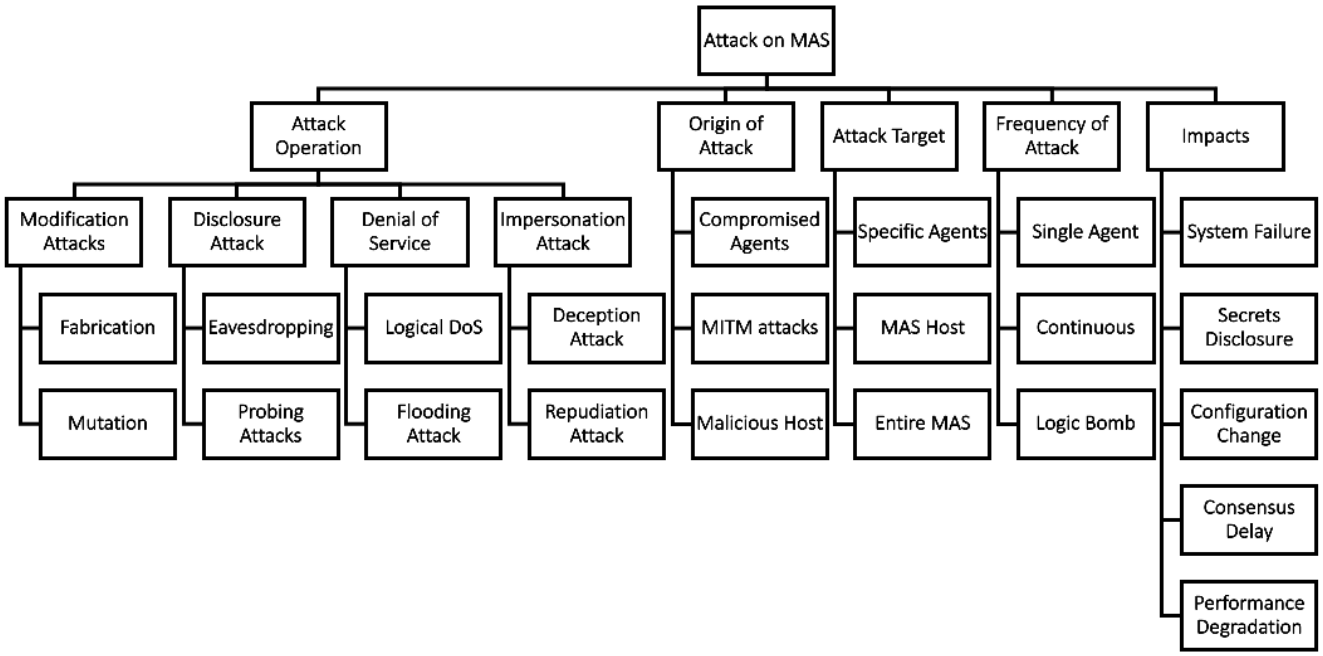

FIGURE 4. A taxonomy of MAS attacks.

attacks utilize Lightweight Coordination Calculus interaction models. They developed a secrecy analysis framework for the interaction models to detect probing attacks.

2) MUTATION ATTACKS

A key expectation from an integrated system is that information is not tampered with [17]. A *mutation attack* on a MAS occurs when this expectation is undermined, possibly (though not exclusively) by a malevolent agent. Mutation attacks are classified into *altering* and *injection* attacks. In an altering attack, the adversary modifies the agents' interaction, resulting in an alteration of the information transfer process. For instance, Rehak et al. [18] show an altering attack where a malicious agent can exploit vulnerabilities in the system by performing a buffer overflow. The adversary can also alter the agent code, data and configuration, and the event logging of the multi-agent system. Yue et al. [19] present a type of mutation attack in which two or more malicious agents surround a sender agent and collude to mutate the sent messages. Bijani et al. [20] present another type of mutation attack in which the adversary infuses forged information into the system to cause a malicious outcome or infer confidential information. This attack could be implemented via *message injection* attack or *knowledge injection*. In message injection, malicious messages are introduced to control the victim agent's interaction with other agents; knowledge injection entails the introduction of false information into the victim agent's knowledge base to affect its decisions.

3) DENIAL OF SERVICE

In a denial-of-service (DoS) attack on MASs, the attacker attempts to prevent the system from providing the intended services to its legitimate users. DoS attacks may target wasting other agents' resources [21], delaying the service [22], making real users forsake the system [23], or ruining the system's reputation [24]. A malicious agent may attack just one agent or a group of agents. A *distributed DoS* (DDoS) attack involves collaboration by more than one attacking entity. Persis et al. [25] established a general DoS attack model that imposes limits on the frequency and duration of attacks. Traynor et al. [26] divided Denial of Service (DoS) attacks into two classes: *flooding* and *logical* attacks. Flooding occurs when a malicious sends a high number of messages to one or more agents in a bid to overwhelm the agent or connections between agents, The adversary can do this by consuming the agent or network resources such as the communication bandwidth. Logical DoS uses more sophisticated methods to exploit the system. For instance, the adversary can create falsified interaction models to make the agents in the multi-agent system perform useless tasks or remain in infinite loops.

4) IMPERSONATION

In an impersonation attack, an adversary produces a large number of anonymous agents, possibly with the goal of misusing the resources and changing the system configuration of the victim agent. Impersonation attacks are generally categorized as *deception* or *repudiation*. In a deception attack, the adversary can send deceptive messages to other agents [27]. Ma et al. [28] proposed a deception attack where the attack signals are injected by the adversary into the measurement data during the process of information transmitted via the communication network. In repudiation attacks, the adversary can create fake agents (together with

fake interaction) for impersonation [29]. Once done, the adversary can use these fake agents to increase their trust value, e.g., by making the fake agents falsely certify their veracity. Eventually, the increased trust value would cause the benign agents to believe in the reliability of a malicious agent. The malicious agents then perform some interactions and then repudiate the existence of those interactions, thereby defaming the system.

B. ORIGIN OF ATTACK

The attack origin is the vulnerability point through which the adversaries can gain unauthorized access to the MAS application. Unauthorized access can be accomplished by compromising one or more agents or implementing a man-in-the-middle attack. Mustafa and Modares [30] demonstrated that an attack on a compromised agent could have a negative impact on intact agents that are reachable from it. They presented a mathematical framework for the analysis of this attack. Their study emphasizes the importance of developing unique resilient control mechanisms to offset the impacts of the attacker and establish an invulnerable consensus. Yang et al. [31] also described various attacks based on MITM to disrupt the formation control process of multi-agent systems. They analyzed potential formation control security issues on a multi-robot system. The experimental findings of the paper revealed that the MITM-based attacks may readily disrupt the formation motions of a multi-robot system and that several designed MITM attacks can even cause irreparable loss. There has also been work on MAS compromises through malicious or compromised hosts [32]. For instance, the host might include local access control mechanisms to prevent agents from reading and manipulating information belonging to other agents. It would also incorporate resource management, which would provide equitable resource allocation to agents running on the host, preventing a single agent (or a collection of agents) from using too many resources and preventing other agents from functioning optimally.

C. TARGET, FREQUENCY, AND IMPACT

The adversaries can target a single agent as seen in [30] or the entire fleet of agents in the MAS. The attackers can also target the system's host or the middleware controlling the MAS. The attack can be continuous within a time limit or orchestrated at intervals during the MAS operation. An external event, such as time, location, or the arrival of a certain agent, triggers an event-triggered attack, also known as a logic bomb [33]. Finally, the impact of the attack can include degradation of the performance [34], leakage of sensitive information (e.g., through eavesdropping or probing attack [13], [20]), a change in the configuration of the data, agent or the entire MAS system leading to the system not behaving in the intended manner [18]. The attack can lead to an increase in the time it would take for the agents to reach a consensus which would lead to undesirable effects, including a complete system failure [35].

FIGURE 5. MAS defense taxonomy.

IV. MAS DEFENSE STRATEGIES

Security assurance for MASs has been a very active research topic integrating a variety of architecture, design, and validation techniques. In this paper, we categorize defenses into three basic classes as shown in Fig. 5. By *prevention strategies* we mean the use of design solutions to eliminate the possibility specific classes of attacks, while *resiliency strategies* entail monitoring for specific attack classes after deployment and performing mitigations.

A. PREVENTION STRATEGIES

A number of prevention mechanisms have been proposed to protect multi-agents against adversaries. In the description below, we categorize the strategies based on the key security feature used in the prevention, e.g., cryuptography, authentication, or acess control restriction.

1) ENCRYPTION AND CRYPTOGRAPHY

Cryptography is one of the most popular and widely used security mechanisms, with a history dating back to the history of written language itself. The approach is to develop methods for *encrypting* information into *ciphers* to protect it from illegal users. Modern cryptography includes mathematical methods for protecting digital information, systems, and distributed computing from hostile actions. Encryption of sensitive data and authentication is the first and most effective step toward countering MITM attacks, unauthorized access to agents, provenance attacks, and agent log modification.

Among cryptographic protections in MASs, Kapitonov et al. [36] describes how to organize a communication system between agents in a peer-to-peer network using the decentralized Ethereum Blockchain technology and smart contracts. Each agent of the network owns a shared distributed database, which consists of a chain of information blocks in which each current block relates to

the previous block using a unique cryptographic identifier. Adding a new block to the database is impossible without the agreement between all the agents. The disadvantages of this method, as the authors indicate, are due to the limitations of blockchain, which include disruptions and discredits generated by the principle of the blockchain network, the legal status of this network has not been completely worked out as the technology has to be regulated according to the law of the various states. Blockchain technology is also in its crude state and needs much more work to be done. Costa et al. [37] introduces a message-oriented middleware that improves communication efficiency and adds a new component called the certification authority service to the typical multi-agent system architecture. This component is responsible for generating certificates that agents in the platform can use to ensure their identity and communicate messages safely in the FIPA (Foundation for Intelligent Physical Agents architecture) [38] for multi-agent systems. The author describes the limitations of the approaches. The approach has three potential points of failure that need to be addressed by future research. The failure points include the RabbitMQ broker failure, Master node failure, and CA certification service failure. The proposed approach does not include a form of protection for the agent's decision process regarding the choice to trust the sender of a given message.

Horvat et al. [39] proposes an extension to the existing TFTP that is required for use in embedded systems of low computation power such as multi-agent system called STFTP. STFTP is a more secure version in the form of added authentication and established confidentiality for use in the multi-agent system platform. The added security comes from Digest Access Authentication combined with the SHA-1 hash function, as compared to the previous MD5 scheme, which was proved cryptographically unsafe.

Voronova and Zhilenkov [40] proposed a method for increasing the cryptographic stability of communication channels between agents in a multi-agent system to ensure information security. This is accomplished using an encryption process based on dynamic chaos systems, as well as synchronization. The author proposed that the essential security and privacy of communication channels between agents can be enforced using chaotic circuits capable of producing chaotic fluctuations from audio frequencies to the optical range and used as sources of chaotic carriers in a variety of applications such as broadband communications, signal masking, chaotic modulation, spectrum expansion, radars, and cryptography for high-entropy information sources.

A downside common to the above approaches is that they do not shield the agents from the influence of programs operating on network nodes. Because of the presence of a large number of malicious programs that can illegally alter the operation of agents and modify sensitive information that agents operate with. This problem remains largely unaddressed. This prompted a rise in resilience study, which would be described in Section IV-C. Another issue with employing symmetric encryption techniques in an open multi-agent system is the necessity for a distinct secret key for each pair (or group) of agents exchanging secret keys, which might cause scalability issues for numerous agents. For example, in a MAS set up with a high number of agents, it appears impossible to allocate and keep a distinct encryption key for each pair of agents. According to Bijani et al. [14], This issue can be solved by combining symmetric encryption with public-key cryptography schemes, which as used in some other papers. Wong & Sycara [41] proposed a security infrastructure to address the RETSINA framework's security and trust [42] which is a reusable multi-agent infrastructure and provides solutions for secure communication, the integrity of system-level services (such as naming and matchmaking services), and accountability. To achieve agent communication security, they coupled unique agent IDs and the Secure Socket Layer (SSL) protocol beneath their agent communication layer.

Other cryptographic techniques for preventing attacks against MAS have been explored [43], [44], [45], [46], [47]. Besides encryption, certificate-based security approaches such as SSL [48], digital signature [49], and integrity checking [50] have also been used in preventing interaction modification attacks.

2) AUTHENTICATION

Authentication is the process of recognizing an agent's identity in the MAS by associating an incoming request with a set of identifying credentials. Since a request may originate on a remote host and may traverse several machines and network channels that are secured in different ways (and are not equally trusted) implies that it is non-trivial to authenticate the original source of communication in a distributed system.

Sabir et al. [51] developed an authentication scheme that uses JSON Web Token (JWT) to secure the communication between the agents by ensuring the integrity of the exchanged messages. JWT is a compact, URL-safe means of representing claims to be transferred between two parties. An authentication agent is responsible for the generation of JWT for the different agents who want to communicate with other agents.

Chatterjee et al. [52] proposed a timestamp-based mutual authentication protocol to enhance security during communication in monitoring systems like multi-agent systems. The scheme is based on Elliptic Curve Diffie–Hellman (ECDH) in wireless sensor network, which gives authentication and confidentiality. The proposed protocol is suitable for multi-agent systems as it requires less bandwidth and has a low storage requirement for the user side hence low computation costs. There are other papers that proposed the use of the Elliptic Curve Cryptosystem (ECC) for authentication. [53], [54], [55], [56].

In vehicular communications, Zhang et al. [57] proposed RAISE, a unique RSU-assisted message authentication scheme in which Roadside Units are in charge of evaluating the authenticity of messages transmitted from cars and

informing them of the results. The proposed RAISE system has numerous advantages due to its lower processing and computational overhead. RAISE also uses the k-anonymity technique to safeguard the privacy of the cars. Hao et al. [58] also developed a way for authenticating VANETs using roadside units in which roadside unit (RSU) acts as the key distributor for the group of vehicles. This leads to another problem in which the RSUs can be compromised. The authors created a secure key distribution strategy that can prevent us from acting irrationally. The protocol ensures that compromised RSUs and hostile vehicles can be tracked. Aside from these, there has been research on other authentication schemes suitable for MASs [59], [60], [61].

3) ACCESS CONTROL MECHANISMS

Access control is the way of safeguarding the MAS by restraining what resources can be accessed or the entity accessing the resources. Role-based access control supports centralized security management, simplifies authorization management of MAS, and protects the integrity of information. Access control using security policy is common in MASs. Agents usually stipulate their policies for revealing information and information collection processes from other agents. If an agent's policy for disclosing information matches another agent's policy for information collection, the information is transmitted between the agents. Various middleware systems have been developed over the years to ensure access control policies in MASs. These architectures include JADE-S [62] which is a decentralized and inflexible access control scheme, SeMoA [63] which is centralized but also inflexible, and AgentScape [64] which uses role-based access control methods and can be customized.

Bijani et al. [14] identified the significant challenges of access control-based security methods: they are usually low-level middleware and only control access to low-level objects. This is challenging as this method finds it difficult to detect high-level violations such as impersonation and probing attacks. To solve the issue of end-to-end security interoperability, Tan and Poslad [65] presented a policy-based architecture for dynamic security reconfiguration in open and heterogeneous systems. This architecture detects and investigates policy disputes and the need for security reconfiguration, then resolves them at the meta-level without modifying the implementation of essential techniques. Paruchuri et al. [66] introduced a randomized policy to reduce the predictability of an agent's action using a decision-theoretic model based on the Multi-agent Constrained Markov Decision Problem (MCMDP). They also developed an algorithm called the Rolling Down Optimization that effectively generates new access control policies through linear programming.

B. DETECTION STRATEGIES

Preventing all types of assaults is not always practicable or viable in a MAS platform. As a result, as the second line of defense, the design and implementation of effective detection and reaction systems against potential assaults are critical. The earlier an assault is identified, the less influence it has on the victim agents. Detection strategies can be subdivided into two classes according to the method used: *behavior-based* and *knowledge-based*. Behavior-based detection methods ''learn'' normal behavior and communication patterns among agents and then detect intrusions by analyzing differences from expected or normal traffic. On the other hand, knowledge-based detection strategies use the knowledge pertaining to specific attacks.

1) ANOMALY DETECTION

Anomaly detection refers to the problem of discovering patterns in data that do not conform to expected behavior. In various application fields, these nonconforming patterns are referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, oddities, or contaminants [2]. Anomaly detection is used in a broad range of applications, including credit card [67], insurance [68], and health-care fraud detection [69], intrusion detection for cybersecurity [70], defect detection in safety-critical systems, and military surveillance for enemy operations [71].

Mateos and García [72] created an architecture for anomaly detection in MASs. This architecture is made up of several agents, each of which is modeled as a virtual digital shell of each asset in the manufacturing line. These agents gather information generated by the asset. There is also a central agent, such as a middle-ware. This central agent consists of a learning algorithm that detects abnormal behavior among other agents. Anomaly detection then constitutes the following tasks.

- Collecting, analyzing, and processing vast amounts of data in real-time under the supervision of intelligent agents; and
- Applying learning algorithms such as machine learning models (predictive models) in the multi-agent system to predict the normal state of the MAS.

Anomaly detection techniques in open MASs use a diversity of methods, including classification, clustering, or statistical analysis techniques. The anomaly detection algorithm assists the MAS system in differentiating aberrant activity from normal behaviors. Following are three main aspects of the predictive models used in detecting abnormality.

- *Feature selection* allows the machine learning model makes sense of the data provided.
- *Learning parameters* are parameters that algorithms employ to manage various learning aspects such as model size, the maximum number of iterations in the learning data, and regularization type.
- *Probability* of the presence of an anomaly is calculated and compared against a predefined threshold to infer an anomaly.

Servin and Kudenko [73] shows how a group of agents coordinate their actions to reach the common goal of anomaly detection. Decision agents learn how to understand action

TABLE 1. Defence strategies.

	Prevention
Encryption and Certificate	[14], [36]–[50]
Authentication	[51]–[61]
Access Control	[14], [62]–[66]
	Detection
Behavior-based (Anomaly Detection)	[2], [67]–[79]
Knowledge-based (Point-based Approaches)	[80]–[87]
	Resiliency
Trust and Reputation	[88]–[100]
Game-Theoretical Approach	[101]–[116]
Stochastic Approach	[22], [85], [87], [117]–[133]

signals supplied by sensor agents without any previously assigned interpretations or logic throughout this process. These action signals integrate the partial information gathered by sensor agents and are utilized by decision agents to reconstruct the global state of the environment. The technique was then used to recognize abnormal activities when the multi-agent system was introduced to DoS attacks. The main advantage of this learning approach is that the machine learning model does not need to be trained with prior information from the DoS data. However, the limitation of this approach is that it requires a large amount of data and high computational resources. Replay attack detection in a multi-agent system using stability analysis and loss effective watermarking

Anomaly detection has also been successfully applied to multi-robot systems (MRS), which constitute an essential class of MASs. One such approach is an online data-driven anomaly detection approach (ODDAD) proposed by Khalastchi et al. [74], in which the anomaly is detected in real-time using the current input data. The occurrence of a fault in the system was detected in three steps. The input was initially filtered to reduce noise. Dimension reduction was applied by splitting the data into sets of correlated attributes. Finally, the Mahalanobis Distance calculation was applied to each set-in order to return the probability of a data instance being an outlier. If the likelihood falls above a calculated threshold, an anomaly is detected.

Goh et al. [75] proposed using an LSTM-RNN to forecast a data sequence for anomaly detection in cyber-physical systems. Because anomalies or cyber-attacks usually happen over time, correlating time-series data information over time provides a way to recognize anomalies. The LSTM-RNN is used as a predictor to model normal behavior. The Cumulative Sum technique then identifies abnormal behavior with a meager false positive rate. This strategy is essential in real-world CPS applications where abnormal behavior is uncommon. The limitation of this paper is that only a small amount of data gotten from one aspect of the system was used for training or validating the data set due to limited resources and infrastructures.

Khazraei et al. [76] proposed a method for detecting replay attacks in a MAS by analyzing the stability of the system and using loss-effective watermarking techniques. Each agent is assigned a local estimator and an anomaly detector. An adversary detection method is then proposed based on a watermarked control strategy, in which watermarking signal is shared among the agents within the network. However, sharing watermarked signals through the network reduces the performance.

Boddupalli et al. [77], [78] proposed an anomaly technique based on machine learning that can be used in conjunction with several cooperative autonomous vehicle applications to identify and mitigate communication attacks. The primary concept is to create models that can learn normal behavior associated with benign V2X communication and detect unusual behavior to detect possibly harmful communication. This approach was applied to applications such as Platooning, Cooperative adaptive cruise control, Smart Intersection detection, and Dynamic cooperative route management. However, the framework only considers attacks from a single communication channel.

Additional information on machine learning-based anomaly detection in MAS was presented by Kim et al. [79]. Their survey analyzes attack-detection methods based on ML approaches and threats that harm the CPS. The physical system, the network, and the application layer were abstracted from the complicated CPS structure in their hierarchical CPS model. They also provided examples of attack implementations and cyber-physical attacks for each CPS tier. Additionally, they offered a variety of ML-based approaches for detecting cyber-physical attacks, such as anomaly detection, which took into account the hierarchical

CPS model to identify and handle attacks that target different layers.

2) KNOWLEDGE-BASED APPROACHES

Knowledge-based approaches for intrusion detection, — also sometimes referred to as *misuse detection*, — look for run-time indicators that match a given pattern of misbehavior. The advantage of this approach over behavioral-based techniques is low false positive rates. However, the drawback of this approach is the need for an up-to-date database of each attack vector. Consequently, the approach is ineffective against zero-day attacks.

Huang et al. [80] proposed a real-time detection scheme against false data injection attacks in smart grid networks. The authors built an analytical model to configure the detection system for performance assurance based on the fundamental detection criteria using the adaptive CUMSUM algorithm. The approach could detect false data attacks when the post-change probability density function is unknown.

Varshovi et al. [81] proposed a fuzzy IDS based on a DoS attack to address the uncertainty problem in distinguishing between normal and malicious network traffic. In more than 5 million test sessions, the system had a detection rate of 99.9 ZZXCC. 1%, with only roughly 1600 false alarms against DoS flooding attacks that used a limited set of features. Lima et al. [82] developed a defense approach for supervisory control systems that detects intrusions and prevents damage induced by man-in-the-middle cyber-attacks in sensor and control communication channels. They developed a deterministic model of systems subjected to the sensor and actuator channel attacks, as well as a defense method for detecting intrusions and protecting the system from damage caused by man-in-the-middle attacks on communication network channels in CPS. Vuong et al. [83] developed a method based on decision trees for constructing simple detection criteria, which they tested against DoS and command injection attacks. They discovered that adding physical input elements to cyber-physical systems can significantly reduce false positive rates and improve overall intrusion detection accuracy. Besides these, there have been several other approaches to knowledge-based intrusion detection system [84], [134], [86], [87].

C. RESILIENCY STRATEGIES

A significant requirement of multi-agent systems is resiliency, i.e., the capacity to supply services consistently despite bugs, failures, attacks, and subversions. Incorporating resilience into MASs is an active field of research with a large volume of work. Our focus in this section is confined to resiliency against security attacks.

1) REPUTATION AND TRUST

There has been significant research on models of trust and confidence for building resilient MAS against adversarial attacks. Lee and See [88] define trustworthiness as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability. Since conventional network security techniques such as encryption, firewall, and access control cannot predict agent behavior from a trust standpoint, trust concerns have grown in popularity. In various applications, the term ''trustworthiness of an agent'' has been used to signify a variety of notions, including (but not limited to the following: (1) the agent precisely executes the coordinator's directions; (2) the agent accurately communicates its status (e.g., position, velocity, etc.); or (3) the agent is not malicious.

Das and Islam [89] defined a reputation trust model as a model that collects, distributes, and aggregates feedback about participants' past behavior. These models assist agents in deciding who to trust, encourage trustworthy conduct, and discourage involvement by dishonest agents. The reputation trust model was divided into two categories based on how the evaluator agents evaluate the information process.

- Direct experience model/Local Trust Models in which the reputation and trust values are based on direct encounters and observations (firsthand value)
- Global reputation/Trust models in which the agents combine information about the reputation of the target agent from all other agents interacting with the target agent.

Although global reputation models converge to a better decision than the direct experience model, they are more complex and more challenging to manage than the natural experience models. The global reputation model also fails when the adversaries change their behavior strategically to benefit them. The high workload demand of the global reputation problem is also an issue that needs to be tackled.

Das and Islam proposed a technique called Secured Trust. This dynamic trust computation technique can identify unexpected strategic changes in malicious behavior and has the added function of balancing workload among service providers in the multi-agent system. This approach employed a unique policy of using an exponential averaging process to decrease storage costs in computing agent trust. However, this is counter-intuitive as it adds additional workload to the system. The limitation of the method was that it did not consider how to transmit trust data securely. However, this can be addressed by combining the proposed approach with other security issues like cryptography.

Zhang et al. [90] proposed a method called SFtrust, which uses two trust metrics, one for service trust and one for feedback trust. Even with variable feedback, this technique can fully use all the agents' service capabilities. One limitation of this approach is the lengthy computational time. Another limitation is the inability of the static weighted average of the local feedback trust to incorporate the evaluating agent's cumulative experience appropriately.

Hu et al. [91] proposed a robust feedback credibility global trust model called FCtrust that distinguishes between offering feedback and delivering services. This approach assesses the reliability of any recommender who provides feedback using transaction density and similarity measures. However, this

approach still must deal with the limitation of the inability to find a way to transmit trust data securely. Coordination may be utilized for autonomous and mobile MASs, such as those found in ground transportation systems or unmanned aerial vehicles, to provide higher safety over human-operated agents, to increase system efficiency, or both. When the MAS comprises both trustworthy and untrusted agents, the MAS must provide safe and efficient coordination.

A critical application domain of multi-agent cooperation is multi-vehicle platooning, which employs artificial intelligence to examine platoon members and assess their level of trustworthiness to avoid attacks that might result in accidents. Hu et al. [92] proposed a technique called REPLACE for creating a trust-based Platoon service. In REPLACE, users may assess the conduct of lead drivers, and the data can then be used to advise others on whether to join a platoon with that leader. An iterative filtering technique is used to cope with false feedback from user automobiles. One limitation of this approach is that the attack primarily aims to provide unjustified trust in the leader vehicle, but other attacks apply to the platooning application. These methodologies, however, examine individual vehicles in isolation and do not consider vehicle communication and cooperation. Another limitation is that the proposed methodologies do not consider scenarios in which the Platooning scheme is attacked by multiple adversaries simultaneously.

Cheng et al. [93] addressed these issues by building a universal framework based on a logical characterization of trust that enables systematic quantification of trust in individual agents. To assess an agent's trustworthiness, the proposed includes short-term and long-term behavioral histories. The framework helps select coordinating rules that accomplish the required trade-off between safety and efficiency based on quantifiable trust values. The proposed framework was successfully applied in three multi-agent platforms, such as the Cooperative adaptive cruise control (CACC), Autonomous intersection management (AIM), and the Reinforcement Learning Traffic light control system, to demonstrate its feasibility and applicability. The limitation of this approach, particularly in the CACC application, is that it considers a centralized authority that maintains the trust distribution and reputation ratings. The centralized system presents a single point of failure on the network.

Ensuring secure and trusted communications within Vehicular Adhoc Networks (VANETs) is another exciting application area. Several works have described the trust solutions in VANETS, a subset of multi-agent systems. Dotzer et al. [94] proposed a VANET reputation framework called VARS, which is an entirely distributed approach based on reputation. Peers can develop opinions about a message based on the aggregated opinions of other nodes and direct interactions with the sender. What sets this approach from other methods is the fact that the approach gives more importance to the views coming from the closest agents to the reported events. The limitation of the approach is the added overhead in the message transmitted by aggregating other nodes' opinions.

Tajeddine et al. [95] proposed a trust-based privacy-preserving model for VANETS. The methodology is unique in maintaining accurate reputation-based trust while protecting privacy. The vehicles are arranged into groups, each with a reputation value. Each group's reputation improves if the average of its members' opinions is consistent with the road condition. The limitation of this approach is that the vehicles are hardly resilient against colluding vehicles in the same group. Many trust and reputation approaches also suffer badly from their total resource consumption. To combat this problem, a subjective logic-based technique is presented for modifying reliability information in data exchanged by the MAS [37]. Subjective logic is a branch of probabilistic logic that explicitly considers cognitive uncertainty and source trust. The paper's method provides a trade-off between security, generality, and resource usage. The technique is shown to cope with heterogeneous agents, isolate faulty agents, and show little resource management.Other research work that discussed the role of reputation and trust in securing MAS includes [96], [97], [98], [99], and [100].

2) GAME THEORY

Game theory is a mathematical model for understanding conflict and cooperation among rational, intelligent decision-makers [101]. The idea is to model interactions between different parties as games, such that a ''win'' in the game corresponds to achieving some specific goal for the cooperative artifacts being modeled. The game can either be static or dynamic. Various MAS security can be naturally modeled through game theory [102], [103], [104], [105], [106], [107], [108], [109], [110].

Farhadi et al. [111] explored a dynamic network with strategic agents that discreetly monitor their security state and are exclusively concerned with increasing their utility. The author posed a mechanism design issue for a network manager whose goal is to dynamically allocate his limited security resources across the network to maximize overall network security over time. The authors took advantage of the dynamic correlation between the security states of the agents to generate a set of inference signals for all agents over time. They outlined a dynamic incentive mechanism that maintains the agents' incentive compatibility and individual rationality and delivers a socially efficient conclusion using the proposed inference signals.

Farhadi et al. [112] proposed an incomplete information two-player game platform that models the interaction between the service providers and various clients, which could be agents in our case. The game is analyzed using perfect Bayesian Nash equilibrium (PBNE) under different situations. This helps prevent malicious activities of the attacker clients leading to the provision of quality services to the benign client.

La et al. [113] studied the application of game-theoretic concepts to the idea of honeypots in IoT and

cyber-physical systems. They considered a game-theoretic model of deception involving an attacker and a defender. The game model tries to answer fundamental problems such as how the defender should react to various observations made by the attacker and if deception benefits the attacker and the defender. The authors devised a Bayesian game of incomplete information to reflect the defender's poor awareness of impending attacks.

In the presence of replay attacks, Miao et al. [114] created a zero-sum, finite horizon, non-stationary stochastic game to minimize the worst-case control and detection costs. The game also obtains an optimal control policy for switching between control-cost optimal (but non-secure) and secure (but cost-suboptimal) controllers. The authors demonstrated that the system's optimal strategy exists and presented a sub-optimal algorithm for calculating the system's strategy using a combination of robust game approaches and a stationary stochastic game algorithm with a finite horizon. This approach can also be extended to other cyber-physical systems to find the optimal policy.

In the military settings, Anwar et al. [115] introduced a novel method for cyber-deception that protects critical nodes while also trapping the adversary. This is done by formulating a stochastic game to study the interactions between the administrator of the system and the other clients or agents. In this game, the defender's goal is to stop the attacker early in the cyber-attack chain and prevent more dangerous scenarios from materializing. The limitation of this approach is the cost, as the game is computationally expensive for each player, and this cost grows with the network size.

Çeker et al. [116] provided a game theoretic approach in preventing DoS attacks.They presented a deception-based protection system that uses game theory to describe the defender-attacker interaction. They created a new defense architecture that proactively uses deception to aid in the development of effective responses to adversary attacks that are unconventional, coordinated, and complicated. They also used a new quantification method for the cost variables to generate Bayesian equilibrium solutions for this model and assess the complementary strategies of the participants. The limitation of this paper is that it only considered a one-period game between the game players due to simplicity, as this would not be the case in a real-life scenario.

3) STOCHASTIC RESILIENCY

MASs operating in a real-world environment are subject to noise, interference, and obstructions that can severely affect the agents' communication quality. There is a need to develop various model-based resilient control algorithms that enable the team of autonomous agents to accomplish their formation tasks even in the presence of different adversaries. The *stochastic resiliency* technique, also called *probabilistic resiliency*, deals with adversarial attacks and probabilistic communication constraints.

A fundamental problem for stochastic resiliency is *consensus*, which entails information exchange so that the group of agents can agree on a specific quantity of interest [87]. Shang [117] explored the consensus problem for MASs over directed networks with state constraints. The authors developed the robust state constrained consensus in the presence of rogue agents who may have been compromised.

Yual et al. [118] proposed a method of establishing robust consensus in multi-hop communication in a scenario where some nodes are malicious and try to prevent consensus by sending false information. Using the proposed Mean Subsequence Reduced (MW-MSR) algorithm, they investigated robust asymptotic (approximate) consensus under the malicious scenario. Normal nodes will filter out the extreme values created by malicious multi-hop neighbors. The MW-MSR method not only improves the robustness of the network but also speeds up the convergence in consensus forming even in adversarial environments.

There has been significant research on the resiliency of MASs against DoS attacks. Lu and Yang [119] proposed such a resiliency scheme and demonstrated that despite the DoS assault, the proposed distributed state-feedback and observer-based controller achieves consensus. The primary concept is to ignore the information gained from the attacked channels to reach a safe consensus. The limitation of this approach is that the consensus approach may be triggered even without the presence of a DoS attack. Cetinkaya et al. [120] proposed a stochastic communication strategy for multi-agent consensus under Jamming attacks. In this protocol, agents seek to communicate information with their neighbors at uniformly distributed random time instants. Agent communication attempt timings are randomized in this protocol, and the attacker is unaware of them until after the agents have made their attempts. They show that when the suggested communication protocol is paired with a stochastic ternary control rule, agents may attain consensus regardless of the number of attacks using probabilistic analysis. The limitation of this paper is that the MAS system utilizing the approach might take a longer time to reach consensus compared to the base approach, but that is a realistic trade-off for resiliency. Other papers that explored how the multi-agent system can still reach consensus even when the system is attacked include [121], [122], [123], and [124].

A common control strategy against DoS attacks is event-triggered control. It determines when and how often data samplings, transmissions, and security control mechanisms should be done depending on well-defined events rather than predetermined times. Cheng et al. [22] proposed a distributed event-triggered consensus of a generally linear multi-agent system subjected to periodic denial-of-service (DoS) jamming attacks. In this paper, DoS attacks are carried out by sending out pulse-width modulated (PWM) jamming signals. A distributed event-triggered mechanism was developed to mitigate the attack, and a resilient event-triggered coordination protocol was constructed, enabling the MAS to reach exponential consensus. Persis et al. [125] investigated event-resilient control techniques for linear systems operating

in a DoS environment. The suggested control strategy's resiliency stems from its capacity to adjust the sample rate in response to the state of the process and the existence of DoS attacks.

Ma et al. [126] The authors proposed a resilient wireless cyber-physical management system that is resilient against both physical and wireless interference. Their solution incorporates a holistic controller that generates actuation signals to physical plants and reconfigure the Wireless sensor-actuator networks to retain the desired control performance while conserving wireless resources. Under ideal network settings, the controller considers the worst-case evolution of the plant's Lyapunov function. The authors simulated the process using real-world data, and the simulation's outcome demonstrated that their holistic controller could maintain secure physical operation despite considerable wireless interference and sensor disturbances

Fang et al. [127] described how stochastic systems could use two-way coding to protect linear time-invariant (LTI) feedback control systems against injection attacks. The two-way transformation known as two-way coding, which operates in a feedback loop, takes the signals in the forward path and the feedback path and outputs a new signal to the forward path and a second new signal that continues in the feedback path. The two-way coding can distort the attacker's viewpoint of the control system; it has been shown that this distorted vision on the attacker's side makes it easier to detect attacks or limit the attacker's options.

Pole-dynamics attacks are typical covert cyber-physical attacks based on system theory that tampers with control-related information in communication networks. To counter PDA, Kim et al. [128] suggested a real-time resilient CPS framework. The proposed architecture is a holistic approach needed to detect, isolate, and recover from the PDA in real-time. We incorporate the PDA detection technique on SDN switches, and the proposed detection algorithm distinguishes the PDA from other anomalies, such as disruptions and transient behaviors. The PDA attacker should be removed from the network to limit further damage. A unique data transmission line is needed for sensor measurement because of the attacker's isolation. After restoring the transmission line, the computing system receives the accurate sensor reading and restores the physical system to its initial condition.

Another extensible innovative cybersecurity architecture for ICS is proposed by Paridari et al. [129]. The framework comprises two components: an attack detection module that uses data analytics to identify threats and a resilient control policy that keeps the physical system secure during and after an attack. The architecture ensures that attacks such as man-in-the-middle and DoS attacks cannot destabilize the system. This resiliency technique is accomplished by safeguarding a few carefully chosen sensors. To do this, the supervisory controller generates a correction vector signal, sent to the local controllers to correct the attacked signals. As a result, the controller is reconfigured.

Other control-oriented resiliency techniques against attacks against various cyber-attacks were described in a survey by Kim et al. [130]. The authors thoroughly analyzed how cyber-physical attacks affect CPS, such as Multi-agent systems, due to the disruption of the controlled networks and how to construct CPS that is resilient to such attacks. The survey characterized CPS as a hierarchical networked control system with physical, network, and application layers. One limitation of the survey was that it only focuses on attacks that can disrupt the physical dynamics of the Multi-agent system.

Cetinkaya et al. [85] describes another event-controlled stochastic resiliency strategy in which the authors explored control of linear dynamical systems in networks subject to random packet losses and malicious attacks. The authors demonstrated that the proposed probabilistic characterization could handle independent and dependent loss instances by using a tail probability inequality for the sum of processes representing random packet losses and malicious attacks. Other event-controlled resilient mechanisms have explored resiliency against jamming attacks and random packet losses, which also explores resiliency for the system against denial of service attacks [131], [132], [133].

V. SUMMARY AND CONCLUSION

We presented a comprehensive high-level discussion on various methodologies needed to secure the MAS. We comprehensively reviewed recent developments in the physical safety and cyber-security of MASs. We discussed various MAS characteristics (e.g., sociability, mobility, and decentralization) and their influence on vulnerabilities. We provided an extensive survey of the spectrum of attacks on MASs and various prevention, detection, and resiliency strategies. Despite significant research on the security of open MAS, many security vulnerabilities remain. A viable solution must account for various challenges, including limitations in computational resources and real-time requirements for the solutions. We believe that the overview of potential threats and security techniques in MASs will pave the way for a comprehensive understanding of the state of the art in the research area and facilitate future research.

REFERENCES

[1] R. Benaboud and T. Marir, "Flexibility measurement model of multi-agent systems," *Multiagent Grid Syst.*, vol. 16, no. 3, pp. 309–341, Oct. 2020.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[3] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination," *Comput. Netw.*, vol. 101, pp. 158–168, Jun. 2016.

[4] M. J. Wooldridge, *An Introduction to MultiAgent Systems*. Hoboken, NJ, USA: Wiley, 2012.

[5] R. C. Cavalcante, I. I. Bittencourt, A. P. da Silva, M. Silva, E. Costa, and R. Santos, "A survey of security in multi-agent systems," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 4835–4846, 2012.

[6] S. Russell and P. Norvig, "A modern, agent-oriented approach to introductory artificial intelligence," *ACM SIGART Bull.*, vol. 6, no. 2, pp. 24–26, Apr. 1995.

[7] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-agent systems: A survey," *IEEE Access*, vol. 6, pp. 28573–28593, 2018.

[8] R. Merris, "Laplacian matrices of graphs: A survey," *Linear Algebra Appl.*, vols. 197–198, pp. 143–176, Jan./Feb. 1994.

[9] M. Rigo, *Advanced Graph Theory and Combinatorics*. Hoboken, NJ, USA: Wiley, 2016.

[10] D. Trcek, *Managing Information Systems Security and Privacy*. Heidelberg, Germany: Springer-Verlag, 2006.

[11] C. P. Pfleeger, S. L. Pfleeger, and J. Margulies, *Security in Computing*. New Delhi, India: Pearson, 2018.

[12] D. Zhang, G. Feng, Y. Shi, and D. Srinivasan, "Physical safety and cyber security analysis of multi-agent systems: A survey of recent advances," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 2, pp. 319–333, Feb. 2021.

[13] A. M. Talib, R. Atan, R. Abdullah, and M. A. A. Murad, "Security framework of cloud data storage based on multi agent system architecture—A pilot study," in *Proc. Int. Conf. Inf. Retr. Knowl. Manag.*, Mar. 2012, pp. 54–59.

[14] S. Bijani, D. Robertson, and D. Aspinall, "Probing attacks on multi-agent systems using electronic institutions," in *Declarative Agent Languages and Technologies IX*. Berlin, Germany: Springer, 2012, pp. 33–50.

[15] V. Graveto, L. Rosa, T. Cruz, and P. Simões, "A stealth monitoring mechanism for cyber-physical systems," *Int. J. Crit. Infrastruct. Protection*, vol. 24, pp. 126–143, Mar. 2019.

[16] G. Na and Y. Eun, "A multiplicative coordinated stealthy attack and its detection for cyber physical systems," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Aug. 2018, pp. 1698–1703.

[17] C. Mitchell, "Security for mobility," *Electron. Commun. Eng. J.*, vol. 14, no. 5, p. 178, 2002.

[18] M. Rehák, M. Pechoucek, M. Grill, J. Stiborek, K. Bartos, and P. Celeda, "Adaptive multiagent system for network traffic monitoring," *IEEE Intell. Syst.*, vol. 24, no. 3, pp. 16–25, May/Jun. 2009.

[19] X. Yue, X. Qiu, Y. Ji, and C. Zhang, "P2P attack taxonomy and relationship analysis," in *Proc. 11th Int. Conf. Adv. Commun. Technol.*, 2009, pp. 1207–1210.

[20] S. Bijani and D. Robertson, "A review of attacks and security approaches in open multi-agent systems," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 607–636, Dec. 2014.

[21] H. L. Nguyen and U. T. Nguyen, "Study of different types of attacks on multicast in mobile ad hoc networks," in *Proc. Int. Conf. Netw., Int. Conf. Syst. Int. Conf. Mobile Commun. Learn. Technol. (ICNICONSMCL06)*, 2006, p. 149.

[22] Z. Cheng, D. Yue, S. Hu, H. Ge, and L. Chen, "Distributed event-triggered consensus of multi-agent systems under periodic DoS jamming attacks," *Neurocomputing*, vol. 400, pp. 458–466, Aug. 2020.

[23] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, "An integrated trust and reputation model for open multi-agent systems," *Auton. Agents Multi-Agent Syst.*, vol. 13, no. 2, pp. 119–154, 2006.

[24] E. M. Amullen, S. Shetty, and L. H. Keel, "Model-based resilient control for a multi-agent system against denial of service attacks," in *Proc. World Autom. Congr. (WAC)*, Jul. 2016, pp. 1–6.

[25] C. D. Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2930–2944, Nov. 2015.

[26] P. Traynor, P. McDaniel, and T. L. Porta, "Future directions and challenges," in *Security for Telecommunications Networks*. Berlin, Germany: Springer, 2008, pp. 157–162.

[27] D. Wang, Z. Wang, B. Shen, F. E. Alsaadi, and T. Hayat, "Recent advances on filtering and control for cyber-physical systems under security and resource constraints," *J. Franklin Inst.*, vol. 353, no. 11, pp. 2451–2466, Jul. 2016.

[28] L. Ma, Z. Wang, and Y. Yuan, "Consensus control for nonlinear multi-agent systems subject to deception attacks," in *Proc. 22nd Int. Conf. Autom. Comput. (ICAC)*, Sep. 2016, pp. 21–26.

[29] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proc. 1st Int. Conf. High Confidence Netw. Syst. (HiCoNS)*, 2012, pp. 55–64.

[30] A. Mustafa and H. Modares, "Attack analysis for discrete-time distributed multi-agent systems," in *Proc. 57th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2019, pp. 230–237.

[31] Y. Yang, Y. Xiao, and T. Li, "Attacks on formation control for multiagent systems," *IEEE Trans. Cybern.*, early access, Jul. 14, 2021, doi: 10.1109/TCYB.2021.3089375.

[32] G. A. Sujitha, "A software agent framework to overcome malicious host threats and uncontrolled agent clones," *Int. J. Adv. Inf. Technol.*, vol. 2, no. 2, pp. 13–27, Apr. 2012.

[33] M. S. Greenberg, J. C. Byington, and D. G. Harper, "Mobile agents and security," *IEEE Commun. Mag.*, vol. 36, no. 7, pp. 76–85, Jul. 1998.

[34] F. Sharifi, Y. Zhang, and A. G. Aghdam, "A distributed deployment strategy for multi-agent systems subject to health degradation and communication delays," *J. Intell. Robot. Syst.*, vol. 73, nos. 1–4, pp. 623–633, Jan. 2014.

[35] Y. Wenwu, W. Ren, M. Cao, and G. Chen, "Delay-induced consensus and quasi-consensus in multi-agent systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, pp. 214–228, 2016.

[36] A. Kapitonov, S. Lonshakov, A. Krupenkin, and I. Berman, "Blockchain-based protocol of autonomous business activity for multi-agent systems consisting of UAVs," in *Proc. Workshop Res., Educ. Develop. Unmanned Aerial Syst. (RED-UAS)*, Oct. 2017, pp. 84–89.

[37] D. Costa, D. Garrido, and D. Silva, "Efficient secure communication for distributed multi-agent systems," in *Proc. 13th Int. Conf. Agents Artif. Intell.*, 2021, pp. 543–552.

[38] F. Bellifemine, A. Poggi, and G. Rimassa, "Developing multi-agent systems with a FIPA-compliant agent framework," *Softw., Pract. Exp.*, vol. 31, no. 2, pp. 103–128, Feb. 2001.

[39] G. Horvat, D. Zagar, and G. Martinovic, "STFTP: Secure TFTP protocol for embedded multi-agent systems communication," *Adv. Electr. Comput. Eng.*, vol. 13, no. 2, pp. 23–32, 2013.

[40] A. V. Voronova and A. A. Zhilenkov, "Cryptographic strength of encryption in a multi-agent system," in *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (ElConRus)*, Jan. 2021, pp. 739–741.

[41] H. C. Wong and K. Sycara, "Adding security and trust to multiagent systems," *Appl. Artif. Intell.*, vol. 14, no. 9, pp. 927–941, Oct. 2000.

[42] K. Sycara, A. Joseph Giampapa, B. Langley, and M. Paolucci, "The retsina MAS, a case study," in *Software Engineering for Large-Scale Multi-Agent Systems*. Berlin, Germany: Springer, 2003, pp. 232–250.

[43] D. D. Khudhur and M. S. Croock, "Developed security and privacy algorithms for cyber physical system," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, p. 5379, Dec. 2021.

[44] M. Kirkpatrick, E. Bertino, and F. T. Sheldon, "Restricted authentication and encryption for cyber-physical systems," in *Proc. DHS CPS Workshop Restricted Authentication Encryption Cyber-Phys. Syst.*, 2009, pp. 1–4.

[45] K. Muniasamy, S. Srinivasan, J. Vain, and M. Sethumadhavan, "Formal methods based security for cloud-based manufacturing cyber physical system," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1198–1203, 2019.

[46] Y. Kim, V. Kolesnikov, and M. Thottan, "Resilient end-to-end message protection for cyber-physical system communications," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2478–2487, Jul. 2018.

[47] Y.-J. Kim, V. Kolesnikov, and M. Thottan, "Resilient end-to-end message protection for large-scale cyber-physical system communications," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2012, pp. 193–198.

[48] J. Davies, *Implementing SSL/TLS Using Cryptography and PKI*. Hoboken, NJ, USA: Wiley, 2011.

[49] R. Sulaiman, D. Sharma, W. Ma, and D. Tran, "A multi-agent security architecture," in *Proc. 3rd Int. Conf. Netw. Syst. Secur.*, 2009, pp. 184–191.

[50] K. Pietak, A. Wos, A. Byrski, and M. Kisiel-Dorohinicki, "Functional integrity of multi-agent computational system supported by component-based implementation," in *Holonic and Multi-Agent Systems for Manufacturing*. Berlin, Germany: Springer, 2009, pp. 82–91.

[51] B. E. Sabir, M. Youssfi, O. Bouattane, and H. Allali, "Authentication and load balancing scheme based on JSON token for multi-agent systems," *Proc. Comput. Sci.*, vol. 148, pp. 562–570, Jan. 2019.

[52] K. Chatterjee, A. De, and D. Gupta, "A secure and efficient authentication protocol in wireless sensor network," *Wireless Pers. Commun.*, vol. 81, no. 1, pp. 17–37, Mar. 2015.

[53] A. Zakerolhosseini and M. Nikooghadam, "Secure transmission of mobile agent in dynamic distributed environments," *Wireless Pers. Commun.*, vol. 70, no. 2, pp. 641–656, May 2013.

[54] M. Nikooghadam, F. Safaei, and A. Zakerolhosseini, "An efficient key management scheme for mobile agents in distributed networks," in *Proc. 1st Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Oct. 2010, pp. 32–37.

[55] H. Hasan, T. Salah, D. Shehada, M. J. Zemerly, C. Y. Yeun, M. Al-Qutayri, and Y. Al-Hammadi, "Secure lightweight ECC-based protocol for multi-agent IoT systems," in *Proc. IEEE 13th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2017, pp. 1–8.
[56] Y. Berguig, J. Laassiri, and S. Hanaoui, "Anonymous and lightweight secure authentication protocol for mobile agent system," *J. Inf. Secur. Appl.*, vol. 63, Dec. 2021, Art. no. 103007.
[57] C. Zhang, X. Lin, R. Lu, P. H. Ho, and X. Shen, "An efficient message authentication scheme for vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3357–3368, Nov. 2008.
[58] Y. Hao, Y. Cheng, C. Zhou, and W. Song, "A distributed key management framework with cooperative message authentication in VANETs," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 616–629, Mar. 2011.
[59] X. Liu, Z. Fang, and L. Shi, "Securing vehicular ad hoc networks," in *Proc. 2nd Int. Conf. Pervasive Comput. Appl.*, Jul. 2007, pp. 424–429.
[60] X. Lin, C. Zhang, X. Sun, P.-H. Ho, and X. S. Shen, "Performance enhancement for secure vehicular communications," in *Proc. IEEE Global Telecommun. Conf. (IEEE GLOBECOM)*, Nov. 2007, pp. 480–485.
[61] W. Shen, L. Liu, X. Cao, Y. Hao, and Y. Cheng, "Cooperative message authentication in vehicular cyber-physical systems," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 84–97, Jun. 2013.
[62] S. Vitabile, V. Conti, C. Militello, and F. Sorbello, "An extended JADE-S based framework for developing secure multi-agent systems," *Comput. Standards Interfaces*, vol. 31, no. 5, pp. 913–930, Sep. 2009.
[63] Y. Jung, M. Kim, A. Masoumzadeh, and J. B. D. Joshi, "A survey of security issue in multi-agent systems," *Artif. Intell. Rev.*, vol. 37, no. 3, pp. 239–260, Mar. 2012.
[64] N. J. E. Wijngaards, B. J. Overeinder, M. van Steen, and F. M. T. Brazier, "Supporting internet-scale multi-agent systems," *Data Knowl. Eng.*, vol. 41, nos. 2–3, pp. 229–245, Jun. 2002.
[65] J. Tan and S. Poslad, "Dynamic security reconfiguration for the semantic web," *Eng. Appl. Artif. Intell.*, vol. 17, no. 7, pp. 783–797, Oct. 2004.
[66] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordóñez, and S. Kraus, "Coordinating randomized policies for increasing security of agent systems," *Inf. Technol. Manag.*, vol. 10, no. 1, pp. 67–79, Mar. 2009.
[67] M. S. Gupta, S. Patel, S. Kumar, and G. Chauhan, "Anomaly detection in credit card transactions using machine learning," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 8, no. 3, pp. 1–5, May 2020.
[68] B. Moharram, "Multi-dimensional approaches to anomaly detection: A study of insurance claims," in *Rutgers Studies in Accounting Analytics: Audit Analytics in the Financial Industry*. Bingley, U.K.: Emerald Publishing Limited, 2019, pp. 111–144.
[69] A. Gangopadhyay and S. Chen, "Health care fraud detection with community detection algorithms," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2016, pp. 1–5.
[70] B. Morel, "Anomaly based intrusion detection and artificial intelligence," in *Intrusion Detection Systems*. London, U.K.: IntechOpen, 2011.
[71] T. Brotherton and T. Johnson, "Anomaly detection for advanced military aircraft using neural networks," in *Proc. IEEE Aerosp. Conf.*, vol. 6, Mar. 2001, pp. 3113–3123.
[72] N. M. García, "Multi-agent system for anomaly detection in industry 4.0 using machine learning techniques," *Adv. Distrib. Comput. Artif. Intell. J.*, vol. 8, no. 4, pp. 33–40, Sep. 2019.
[73] A. Servin and D. Kudenko, "Multi-agent reinforcement learning for intrusion detection: A case study and evaluation," in *Proc. German Conf. Multiagent Syst. Technol.* Berlin, Germany: Springer-Verlag, 2008, pp. 159–170.
[74] E. Khalastchi, M. Kalech, and L. Rokach, "A hybrid approach for fault detection in autonomous physical agents," Dept. Inf. Syst. Eng., Ben-Gurion Univ. Negev, Beersheba, Israel, Tech. Rep., 2014.
[75] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *Proc. IEEE 18th Int. Symp. High Assurance Syst. Eng. (HASE)*, Jan. 2017, pp. 140–145.
[76] A. Khazraei, H. Kebriaei, and F. R. Salmasi, "Replay attack detection in a multi agent system using stability analysis and loss effective watermarking," in *Proc. Amer. Control Conf. (ACC)*, May 2017, pp. 4778–4783.
[77] S. Boddupalli and S. Ray, "REDEM: Real-time detection and mitigation of communication attacks in connected autonomous vehicle applications," in *Proc. IFIP Int. Internet Things Conf.* Manhattan, NY, USA: Springer, 2019, pp. 105–122.
[78] S. Boddupalli, A. S. Rao, and S. Ray, "Resilient cooperative adaptive cruise control for autonomous vehicles using machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15655–15672, Sep. 2022.
[79] S. Kim and K.-J. Park, "A survey on machine-learning based security design for cyber-physical systems," *Appl. Sci.*, vol. 11, no. 12, p. 5458, Jun. 2021.
[80] Y. Huang, J. Tang, Y. Cheng, H. Li, K. A. Campbell, and Z. Han, "Real-time detection of false data injection in smart grid networks: An adaptive CUSUM method and analysis," *IEEE Syst. J.*, vol. 10, no. 2, pp. 532–543, Jun. 2016.
[81] A. Varshovi, M. Rostamipour, and B. Sadeghiyan, "A fuzzy intrusion detection system based on categorization of attacks," in *Proc. 6th Conf. Inf. Knowl. Technol. (IKT)*, May 2014, pp. 50–55.
[82] P. M. Lima, M. V. S. Alves, L. K. Carvalho, and M. V. Moreira, "Security against network attacks in supervisory control systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12333–12338, 2017.
[83] T. P. Vuong, G. Loukas, D. Gan, and A. Bezemskij, "Decision tree-based detection of denial of service and command injection attacks on robotic vehicles," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Nov. 2015, pp. 1–6.
[84] H. Sedjelmaci and S. M. Senouci, "A new intrusion detection framework for vehicular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 538–543.
[85] A. Cetinkaya, H. Ishii, and T. Hayakawa, "Networked control under random and malicious packet losses," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2434–2449, May 2017.
[86] T. Li, B. Chen, L. Yu, and W.-A. Zhang, "Active security control approach against DoS attacks in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 66, no. 9, pp. 4303–4310, Sep. 2021.
[87] H. Qu, L. Lei, X. Tang, and P. Wang, "A lightweight intrusion detection method based on fuzzy clustering algorithm for wireless sensor networks," *Adv. Fuzzy Syst.*, vol. 2018, pp. 1–12, Jun. 2018.
[88] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 46, no. 1, pp. 50–80, 2004.
[89] A. Das and M. M. Islam, "SecuredTrust: A dynamic trust computation model for secured communication in multiagent systems," *IEEE Trans. Depend. Sec. Comput.*, vol. 9, no. 2, pp. 261–274, Mar./Apr. 2012.
[90] Y. Zhang, S. Chen, and G. Yang, "SFTrust: A double trust metric based trust model in unstructured P2P system," in *Proc. IEEE Int. Symp. Parallel Distrib. Process.*, May 2009, pp. 1–7.
[91] J. Hu, Q. Wu, and B. Zhou, "FCTrust: A robust and efficient feedback credibility-based distributed P2P trust model," in *Proc. 9th Int. Conf. Young Comput. Sci.*, Nov. 2008, pp. 1963–1968.
[92] H. Hu, R. Lu, Z. Zhang, and J. Shao, "REPLACE: A reliable trust-based platoon service recommendation scheme in VANET," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1786–1797, Feb. 2017.
[93] M. Cheng, C. Yin, J. Zhang, S. Nazarian, J. Deshmukh, and P. Bogdan, "A general trust framework for multi-agent systems," in *Proc. 20th Int. Conf. Auto. Agents MultiAgent Syst.*, 2021, pp. 332–340.
[94] F. Dotzer, L. Fischer, and P. Magiera, "VARS: A vehicle ad-hoc network reputation system," in *Proc. 6th IEEE Int. Symp. World Wireless Mobile Multimedia Netw.*, Jun. 2005, pp. 454–456.
[95] A. Tajeddine, A. Kayssi, and A. Chehab, "A privacy-preserving trust model for VANETs," in *Proc. 10th IEEE Int. Conf. Comput. Inf. Technol.*, Jun. 2010, pp. 832–837.
[96] D. Calvaresi, V. Mattioli, A. Dubovitskaya, A. F. Dragoni, and M. Schumacher, "Reputation management in multi-agent systems using permissioned blockchain technology," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 719–725.
[97] R. Khalid, O. Samuel, N. Javaid, A. Aldegheishem, M. Shafiq, and N. Alrajeh, "A secure trust method for multi-agent system in smart grids using blockchain," *IEEE Access*, vol. 9, pp. 59848–59859, 2021.
[98] J. Jost and B. Mättig, "Trust and reputation in heterogeneous multi-agent system for production processes based on the market economy," in *Proc. 25th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2020, pp. 1321–1324.
[99] T. D. Huynh, "Trust and reputation in open multi-agent systems," Ph.D. thesis, Dept. Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., 2006.
[100] M. Sievers, A. M. Madni, P. Pouya, and R. Minnichelli, "Trust and reputation in multi-agent resilient systems," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 741–747.
[101] S. R. Etesami and T. Ba ar, "Dynamic games in cyber-physical security: An overview," *Dyn. Games Appl.*, vol. 9, no. 4, pp. 884–913, Dec. 2019.
[102] A. Agah and S. K. Das, "Preventing DoS attacks in wireless sensor networks: A repeated game theory approach," *Int. J. Netw. Secur.*, vol. 5, no. 2, pp. 145–153, Sep. 2007.

[103] K. Durkota, V. Lisỳ, B. Bošanskỳ, and C. Kiekintveld, "Optimal network security hardening using attack graph games," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.
[104] T. E. Carroll and D. Grosu, "A game theoretic investigation of deception in network security," *Secur. Commun. Netw.*, vol. 4, no. 10, pp. 1162–1172, 2011.
[105] C. A. Kamhoua, "Game theoretic modeling of cyber deception in the internet of battlefield things," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, p. 862.
[106] M. Rasouli, E. Miehling, and D. Teneketzis, "A supervisory control approach to dynamic cyber-security," in *Proc. Int. Conf. Decis. Game Theory Secur.* Manhattan, NY, USA: Springer, 2014, pp. 99–117.
[107] C. A. Kamhoua, N. Pissinou, and K. Makki, "Game theoretic modeling and evolution of trust in autonomous multi-hop networks: Application to network security and privacy," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–6.
[108] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attack on cyber-physical systems: A game-theoretic approach," in *Proc. IEEE Int. Conf. Cyber Technol. Autom., Control Intell. Syst.*, May 2013, pp. 252–257.
[109] M. E. Mkiramweni, C. Yang, J. Li, and Z. Han, "Game-theoretic approaches for wireless communications with unmanned aerial vehicles," *IEEE Wireless Commun.*, vol. 25, no. 6, pp. 104–112, Dec. 2018.
[110] J. Sun, W. Wang, Q. Da, L. Kou, G. Zhao, L. Zhang, and Q. Han, "An intrusion detection based on Bayesian game theory for UAV network," in *Proc. 11th EAI Int. Conf. Mobile Multimedia Commun.*, 2018, p. 56.
[111] F. Farhadi, H. Tavafoghi, D. Teneketzis, and J. Golestani, "A dynamic incentive mechanism for security in networks of interdependent agents," in *Proc. Int. Conf. Game Theory Netw.* Manhattan, NY, USA: Springer, 2017, pp. 86–96.
[112] S. Farhang, M. H. Manshaei, M. N. Esfahani, and Q. Zhu, "A dynamic Bayesian security game framework for strategic defense mechanism design," in *Proc. Int. Conf. Decis. Game Theory Secur.* Manhattan, NY, USA: Springer, 2014, pp. 319–328.
[113] Q. D. La, T. Q. S. Quek, and J. Lee, "A game theoretic model for enabling honeypots in IoT networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[114] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 1854–1859.
[115] A. H. Anwar, C. Kamhoua, and N. Leslie, "A game-theoretic framework for dynamic cyber deception in internet of battlefield things," in *Proc. 16th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, Nov. 2019, pp. 522–526.
[116] H. Çeker, J. Zhuang, S. Upadhyaya, Q. D. La, and B.-H. Soong, "Deception-based game theoretical approach to mitigate DoS attacks," in *Proc. Int. Conf. Decis. Game Theory Secur.* Manhattan, NY, USA: Springer, 2016, pp. 18–38.
[117] Y. Shang, "Resilient consensus in multi-agent systems with state constraints," *Automatica*, vol. 122, Dec. 2020, Art. no. 109288.
[118] L. Yuan and H. Ishii, "Resilient consensus with multi-hop communication," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 2696–2701.
[119] A.-Y. Lu and G.-H. Yang, "Distributed consensus control for multi-agent systems under denial-of-service," *Inf. Sci.*, vols. 439–440, pp. 95–107, May 2018.
[120] A. Cetinkaya, K. Kikuchi, T. Hayakawa, and H. Ishii, "Randomized transmission protocols for protection against jamming attacks in multi-agent consensus," *Automatica*, vol. 117, Jul. 2020, Art. no. 108960.
[121] J. Usevitch and D. Panagou, "Resilient finite-time consensus: A discontinuous systems perspective," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2020, pp. 3285–3290.
[122] J. Usevitch and D. Panagou, "Resilient leader-follower consensus to arbitrary reference values in time-varying graphs," *IEEE Trans. Autom. Control*, vol. 65, no. 4, pp. 1755–1762, Apr. 2020.
[123] H. J. LeBlanc, H. Zhang, S. Sundaram, and X. Koutsoukos, "Resilient continuous-time consensus in fractional robust networks," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 1237–1242.
[124] Y. Shang, "Consensus of hybrid multi-agent systems with malicious nodes," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 4, pp. 685–689, Apr. 2020.
[125] C. De Persis and P. Tesi, "Resilient control under Denial-of-Service," *IFAC Proc. Volumes*, vol. 47, no. 3, pp. 134–139, 2014.
[126] Y. Ma, D. Gunatilaka, B. Li, H. Gonzalez, and C. Lu, "Holistic cyber-physical management for dependable wireless control systems," *ACM Trans. Cyber-Phys. Syst.*, vol. 3, no. 1, pp. 1–25, 2018.
[127] S. Fang, K. H. Johansson, M. Skoglund, H. Sandberg, and H. Ishii, "Two-way coding in control systems under injection attacks: From attack detection to attack correction," in *Proc. 10th ACM/IEEE Int. Conf. Cyber-Phys. Syst.*, Apr. 2019, pp. 141–150.
[128] S. Kim, Y. Eun, and K.-J. Park, "Stealthy sensor attack detection and real-time performance recovery for resilient CPS," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7412–7422, Nov. 2021.
[129] K. Paridari, N. O'Mahony, A. E.-D. Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration," *Proc. IEEE*, vol. 106, no. 1, pp. 113–128, Jan. 2018.
[130] S. Kim, K.-J. Park, and C. Lu, "A survey on network security for cyber–physical systems: From threats to resilient design," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1534–1573, 3rd Quart., 2022.
[131] H. S. Foroush and S. Martínez, "On triggering control of single-input linear systems under pulse-width modulated DoS signals," *SIAM J. Control Optim.*, vol. 54, no. 6, pp. 3084–3105, Jan. 2016.
[132] A. Cetinkaya, H. Ishii, and T. Hayakawa, "Event-triggered output feedback control resilient against jamming attacks and random packet losses," *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 270–275, 2015.
[133] S. Liu, S. Li, and B. Xu, "Event-triggered resilient control for cyber-physical system under denial-of-service attacks," *Int. J. Control*, vol. 93, no. 8, pp. 1907–1919, Aug. 2020.
[134] V. S. Dolk, P. Tesi, C. De Persis, and W. P. M. H. Heemels, "Event-triggered control systems under denial-of-service attacks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 93–105, Mar. 2017.

RICHARD OWOPUTI (Student Member, IEEE) received the B.S. degree in electrical engineering from the Federal University of Technology, Akure, Nigeria, in 2017, and the M.S. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 2021, where he is currently pursuing the Ph.D. degree in electrical engineering. He is also a Research Assistant with the Rising Laboratory, University of Florida. He works in the domain of connected autonomous vehicles, including fleet management. His research interests include the investigation of the security vulnerabilities in multi-agent coordination and developing resiliency solutions to defend the coordination scheme against a spectrum of adversaries.

SANDIP RAY (Senior Member, IEEE) received the Ph.D. degree from The University of Texas at Austin. He is currently an Endowed IoT Term Professor at the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. His research involves developing correct, dependable, secure, and trustworthy computing through cooperation of specification, synthesis, architecture and validation technologies. He focuses on next generation computing applications, including the Internet of Things applications, autonomous automotive systems, smart homes, and intelligent implants. Before joining the University of Florida, he was a Senior Principal Engineer at NXP Semiconductors, and prior to that, a Research Scientist at Intel Strategic CAD Laboratories. In addition to Intel and NXP, his research found applications in AMD, IBM, Galois, Microsoft, and Rockwell Collins.