Simultaneous fault detection and consensus for multi-agent systems under false data injection attacks

Meilin Li
School of Automation Engineering
University of Electronic Science and Technology of China
Chengdu, China
meilinli0126@163.com

***Abstract*—In the realm of multiagent systems (MAS), maintaining system integrity and achieving consensus in the presence of malicious activities is a critical challenge. This paper addresses the problem of simultaneous fault detection and consensus under false data injection (FDI) attacks. FDI attacks, where adversaries strategically alter the data exchanged among agents, can lead to severe disruptions in the system's functionality, potentially causing erroneous consensus or total system failure. We propose a robust detection and consensus mechanism that integrates advanced fault detection techniques with resilient consensus algorithms specifically designed to withstand FDI attacks. By leveraging the inherent structure of multiagent systems and employing a decentralized approach, our method ensures that agents can not only detect and isolate compromised nodes but also maintain accurate consensus across the network. Simulation results validate the effectiveness of the proposed strategy, demonstrating its capability to mitigate the impact of FDI attacks and sustain system performance.**

***Index Terms*—Multi-agent systems, fault detection, consensus, false data injection attacks.**

I. Introduction

Multiagent systems (MAS) represent a sophisticated class of distributed systems where multiple autonomous agents collaborate to achieve common objectives. Each agent in a MAS is an intelligent entity capable of sensing, processing, and acting within its environment. These systems have found widespread applications in various domains, such as autonomous robotics, distributed sensing networks, smart grids, and cooperative vehicle systems. The ability of agents to work together and reach a common decision or statełtermed consensusłis vital for the success of such systems.

The consensus problem in MAS is one of the most extensively studied topics in control theory and distributed computing. Consensus algorithms aim to ensure that all agents in the system, despite starting from different initial states and having access to only partial information, can eventually agree on a single state. This consensus process is essential for tasks such as coordinated motion in robotics, distributed optimization in sensor networks, and load balancing in power systems.

This work is supported in part by the National Natural Science Foundation of China under Grants 51939001, 62273072, 62203088, the Natural Science Foundation of Sichuan Province under Grant 2022NSFSC0903.

However, the decentralized nature of MAS, while offering advantages like scalability, robustness, and flexibility, also introduces significant challenges. Unlike centralized systems where a single point of control oversees the entire operation, MAS operate without a central coordinator. Each agent relies on local information and interactions with neighboring agents, which can make the system vulnerable to various types of faults and malicious attacks. Among these, false data injection (FDI) attacks have emerged as a particularly insidious threat.

FDI attacks involve an adversary injecting incorrect information into the system, with the intent to disrupt normal operations. These attacks can be particularly damaging in MAS because they exploit the decentralized nature of the system. By corrupting the data that agents use to make decisions, an FDI attack can cause the system to reach incorrect consensus or even lead to system-wide failures. The severity of such attacks is amplified in scenarios where the system relies on consensus for critical operations, such as in autonomous vehicles or power grid management.

The challenge posed by FDI attacks is not merely one of detection but also of maintaining reliable operation in their presence. Traditional consensus algorithms often assume that all agents are cooperative and that communication channels are secure. These assumptions do not hold in environments where adversaries can tamper with the information being exchanged between agents. As a result, there is a growing need for new strategies that can detect and mitigate the effects of FDI attacks while still ensuring that the system reaches a correct consensus.

This paper addresses the dual challenge of fault detection and consensus in MAS under FDI attacks. We propose a novel approach that integrates advanced fault detection mechanisms with robust consensus algorithms. Our approach is designed to operate effectively in environments where communication between agents is not only directed but also potentially compromised by adversarial actions.

Consensus in MAS involves all agents agreeing on a common value or state, despite differences in their initial conditions. This agreement is crucial for the coordinated operation of the system. For example, in a fleet of autonomous

vehicles, consensus might involve all vehicles agreeing on a common speed or direction. In a sensor network, consensus might mean all sensors agreeing on a common estimate of an environmental variable.

The consensus problem is typically modeled using graph theory, where each agent is represented as a node in a graph, and communication links between agents are represented as edges. The graph can be directed or undirected, depending on whether communication is bidirectional or unidirectional. The consensus problem can be mathematically formulated as ensuring that the state of each agent converges to a common value over time, given certain conditions on the communication graph and the dynamics of the agents.

Traditional consensus algorithms assume that all agents are honest and that the communication links are reliable. However, these assumptions may not hold in practice, particularly in environments where security is a concern. Adversaries can target the system by injecting false data into the communication links, leading to incorrect consensus or preventing the system from reaching consensus altogether.