ABSTRACT

Title of dissertation: CONTENT MODIFICATION ATTACKS ON NETWORKED ROBOTIC SYSTEMS

Yimeng Dong
Doctor of Philosophy, 2018

Dissertation directed by: Associate Professor Nikhil Chopra
Department of Mechanical Engineering

With the advent of communication networks in robotic systems, distributed networked robotic systems can be deployed to perform certain tasks collaboratively. However, this makes the networked robotic systems vulnerable to cyber attacks. Thus, the rigorous study of the impact of cyber attacks and the development of corresponding defense mechanisms are necessary.

In this dissertation, the cyber-physical security issue of networked robotic systems is studied under a specific type of cyber attack called content modification attack, which can modify the data content transmitted in the communication networks among the robots. Specifically, algorithms for attack design and detection for content modification attacks are studied. The physics of the robotic system is utilized to design and detect the cyber attacks for networked robotic systems.

Content modification attacks are studied for the synchronization problem in networked robotic systems. The considered systems include multi-robot systems, bilateral teleoperation systems and bilateral tele-driving systems. To demonstrate

the potential severity of the attack, a constructive methodology for attack design is also developed. Specifically, a destabilizing content modification attack referred to as a malignant content modification attack (MCoMA) is designed based on the system storage function, which can lead to system instability and even physical system damage. To protect the system, a physics-based attack detection scheme with an encoding-decoding structure is proposed for general content modification attacks. As part of the tele-driving system study, a novel passivity-based adaptive bilateral tele-driving control scheme is also proposed in the presence of network delays and dynamics parametric uncertainties. Simulations and experiments have also been conducted to validate the proposed algorithms. This study demonstrates the potential of utilizing the physics of the robotic system to better understand and strengthen the security of the networked robotic systems.

CONTENT MODIFICATION ATTACKS ON NETWORKED ROBOTIC SYSTEMS

by

Yimeng Dong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Associate Professor Nikhil Chopra, Chair/Advisor
Professor Balakumar Balachandran
Associate Professor Jin-Oh Hahn
Professor Derek Paley, Dean's Representative
Professor Miao Yu

Acknowledgments

I would like to acknowledge all the people who helped me in the process of finishing this dissertation.

First and foremost, I would like to thank my advisor, Dr. Nikhil Chopra for the guidance and encouragement during all this time. This dissertation will not be possible without your constant support. Special thanks to my labmate and coauthor, Nirupam Gupta for all the discussions and collaborations during these five years, and I really learn a lot. I would like to thank all my other labmates, Prasad Chanekar, Gurtajbir Herr, Bryan Hays, Tianchen Liu and Jeffrey Twigg for all the help and discussions towards this dissertation. I would also like to thank Dr. Balakumar Balachandran, Dr. Jin-Oh Hahn, Dr. Derek Paley and Dr. Miao Yu for being my dissertation committee members and providing the valuable advices for this dissertation. Thanks to all the friends I met at the University of Maryland, it is all of you that make this journey unique. Finally, I would especially like to thank my parents for their unwavering support and thank my wife, Xin Xu for her unconditional acceptance, support and love.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CPS	Cyber-physical System
BTOS	Bilateral Teleoperation System
BTDS	Bilateral Tele-driving System
DOF	Degree-of-freedom
MAS	Multi-agent System
MCoMA	Malignant Content Modification Attack
OMCoMA	Optimal Malignant Content Modification Attack

Chapter 1: Introduction

1.1 Motivation and Objective

In this dissertation, the cyber-physical security issue of networked robotic systems is investigated. The networked robotic system refers to a robotic system connected to a wired or wireless communication network. Cyber-physical security is an emerging research field for cyber-physical systems (CPS) like networked robotic systems [1–3]. Recent publicly known examples of cyber attacks on CPS include Stuxnet malware sabotaging Iran's nuclear infrastructure [4] and the power transmission network attack [5].

The traditional cyber security methods for information systems like cryptography are not adequate for cyber-physical security and the necessity of cyber-physical security study can be expounded as follows:

- Most of the CPS such as networked robotic systems are designed without security considerations, and cyber attacks can directly impact or even damage the critical physical system, so adequate safety nets are required.

- The CPS has real-time requirements and may additionally have energy consumption/processing power constraints. Several modern cryptographic algo-

rithms cannot be implemented in constrained devices and cannot satisfy the real-time requirement as the algorithms were designed for desktop/server environments [6].

- The underlying physics of CPS can potentially be exploited to better understand the attack impact and provide a different layer of security over existing cyber security methods.

Based on these observations, the research objectives of this dissertation can be summarized as follows:

A fundamental problem of ***synchronization*** is considered for the networked robotic systems including bilateral teleoperation systems (BTOS), multi-robot systems, and bilateral tele-driving systems (BTDS). By utilizing the physics of robotic systems, content modification attacks on synchronization of these systems are studied from two different perspectives:

- To better understand the severity of attack impact, the content modification attack design is investigated from the attacker's perspective.

- To effectively safeguard the system from content modification attack, the attack detection scheme is investigated from the defender's perspective.

The content modification attack is a cyber attack that can modify the data content transmitted in the networks among the robots, and a formal definition is given in Chapter 2.

In the rest of this chapter, three types of networked robotic systems are introduced and compared. Additionally, the related works in cyber-physical security

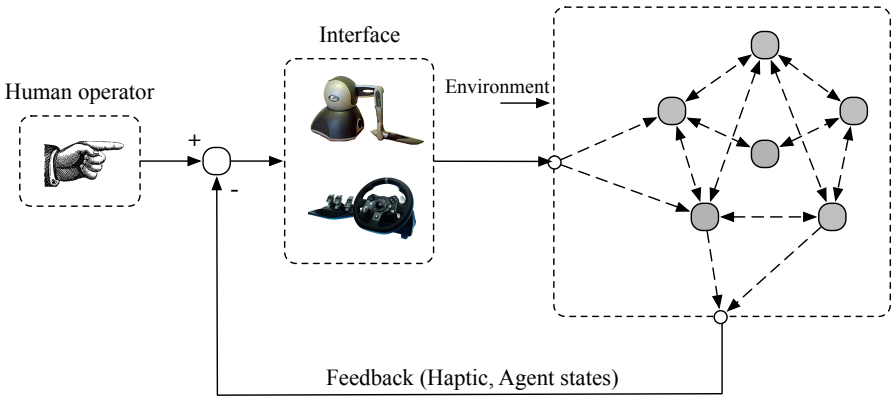


Figure 1.1: General framework of a networked robotic system.

are discussed and compared with the proposed work, the research results in the dissertation are summarized, and finally the technical notations are introduced.

1.2 Background

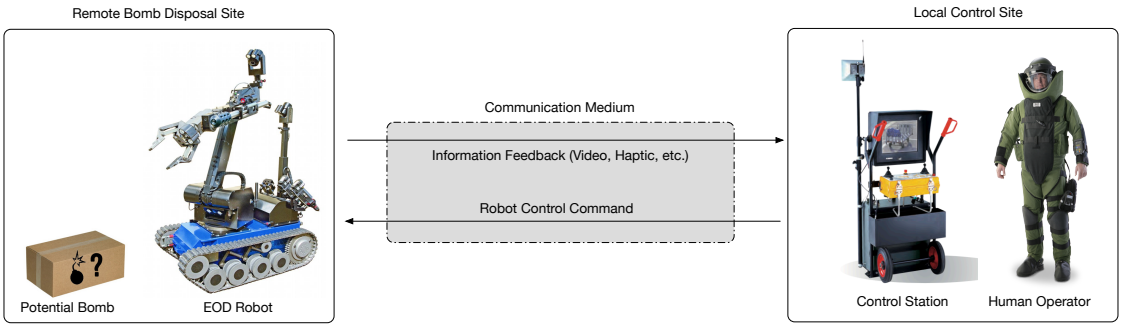
Networked robotic systems can be divided into two types: (1) teleoperated, where a human operator can send commands and receive feedback via the communication network; (2) autonomous, where robots exchange data via the communication network allowing them to communicate with one another over long distances for coordination without human's involvement. A general framework of networked robotic systems is shown in Fig. 1.1, which covers all three types of networked robotic systems in this dissertation. Considering the human in the loop scenario, the BTOS and BTDS are introduced in Section 1.2.1 and 1.2.3. To address the case of only autonomous robots, the multi-robot system is introduced in Section 1.2.2.

1.2.1 Bilateral Teleoperation System

BTOS can extend human capability to manipulate objects in remote environments with the help of robotic manipulators and communication networks. Such systems can be used for various tasks, like handling hazardous materials [8], space and underwater exploration [9,10], and telesurgery [11], etc. Two application examples of BTOS including an explosive ordnance disposal tele-robot and a tele-surgical robot are shown in Fig. 1.2.

Typically, a BTOS consists of two robot manipulators termed the master and slave robot. A human operator can teleoperate the remote slave robot to perform certain tasks by manipulating the local master robot when two robots are coupled and synchronized through a communication network. In Chapter 3 of this dissertation, a BTOS setup as shown in Fig. 1.3 is considered. On being manipulated by a human operator, the state of the master robot is transmitted to the slave robot through a communication channel (wired or wireless). The slave robot is coupled to the master robot state through an appropriate controller which then guides the slave robot to complete a desired task in the remote environment. Simultaneously, the state of the slave robot is communicated to the master robot, and through the master controller a bilateral coupling is established between the two robots.

With the deployment of a communication network, the BTOS performance suffers from the network effects including delays, noises, quantization errors and packet drops. Numerous works have been conducted to overcome these issues and they have yielded fairly successful results. The reader is referred to [12] and [13] for



(a) The explosive ordnance disposal tele-robot by Cobham.

(b) The da Vinci telesurgery system by Intuitive Surgical. [7]

Figure 1.2: Application examples of BTOS.

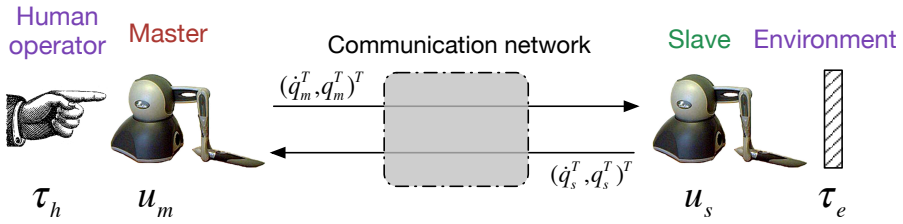


Figure 1.3: A BTOS setup.

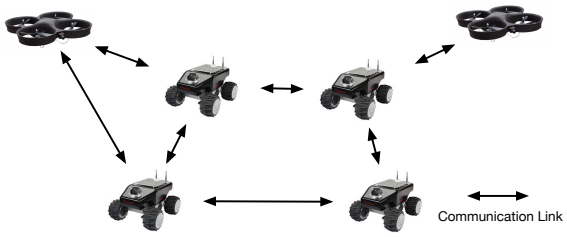


Figure 1.4: Multi-robot system.

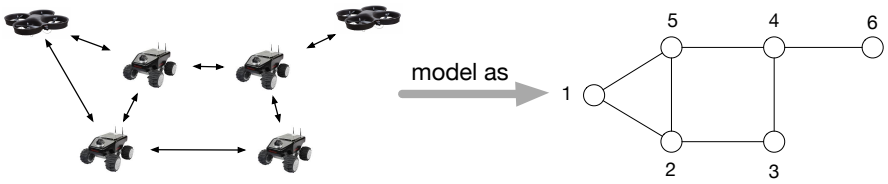


Figure 1.5: Multi-robot system can be modeled as a graph.

an overview of BTOS control algorithms.

1.2.2 Multi-Robot System

The multi-robot system in Fig. 1.4 can be deployed in various tasks such as surveillance, searching and mapping. Each robot can share the information with its neighbors through communication links. One of the critical problems in distributed cooperative control of multi-robot systems is called synchronization, which means that multiple robots reach an agreement on a common value by interacting with neighbors. The synchronization protocol can be applied in cases such as formation control [14] and attitude alignment [15] for multi-robot systems.

As shown in Fig. 1.5, the multi-robot system can be abstracted as a multi-

agent system (MAS) and modeled as a graph with agents as nodes and communication links as edges. The coupling of master and slave robots in BTOS can be considered as a special case of MAS synchronization as two nodes in BTOS are robot manipulators with nonlinear dynamics.

The agreement, consensus and synchronization problems for networked MAS have been extensively studied by [16–30] amongst others, wherein different scenarios such as directed/undirected graphs, fixed/switching topologies, network effects, deterministic/stochastic topologies, uncertainties/disturbances, and linear/nonlinear dynamics are considered. In practice, a broad class of systems can be modeled by a double integrators dynamics model, for example, certain vehicle dynamics can be feedback linearized into double integrators. Thus, considerable research attention has been paid to the consensus-related problem for MAS with double-integrator dynamics [31–35]. In Chapter 4 of this dissertation, the synchronization problem of a multi-robot system modeled as a MAS with double-integrator dynamics is considered.

The data transmission among agents also suffers from various network effects. Over the past few decades, significant research effort has been accomplished to address these issues [18, 25, 27, 29] for consensus seeking MAS.

1.2.3 Bilateral Tele-Driving System

BTDS can allow a human driver to remotely operate a vehicle through wireless communication networks with information (video, haptic) feedbacks. In practice,

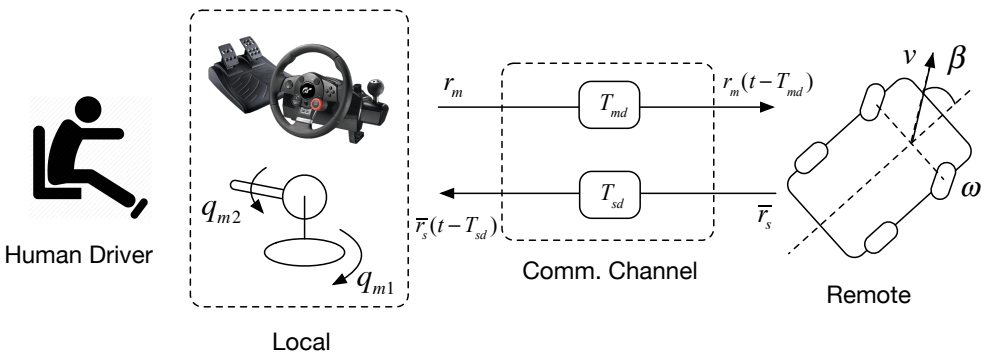


Figure 1.6: The sketch for tele-driving scheme: q_{m1}, q_{m2} are two joint variables of joystick, β is the steering angle, ω is the angular velocity of front wheel, v is the linear velocity of front wheel, T_{md}, T_{sd} are constant communication delays and $r_m, \bar{r}_s$ is the transmitted data defined later in Chapter 5.

as discussed in [36], remote driving (tele-driving a real car) has many potential applications such as transportation of goods, return of unoccupied vehicles, supervision and fail-safe operation of passenger or autonomous cars. The advantage of tele-driving over fully-autonomous driving is to keep the human in the control loop of the unmanned vehicle so as to handle the driving tasks in complex situations. In Chapter 5 of this dissertation, a BTDS for tele-driving a car-like mobile robot is considered. Compared with the traditional bilateral teleoperation between two robotic manipulators, one major difference for tele-driving a car-like mobile robot is the kinematic dissimilarity between the local and remote robot. Fig. 1.6 displays a BTDS, in which the human driver can use local two degree-of-freedom (DOF) local robot (joystick) to tele-drive a car-like remote robot through a communication network. This two DOF system can be analogously considered as steering wheel and gas pedal inputs.

Considering previous work in control of bilateral tele-driving systems, the notion of feedback r-passivity was proposed and utilized in [37] and [38] to study the bilateral tele-driving scheme for a two-wheeled mobile robot. An impedance control framework was proposed for bilateral teleoperation of a car-like rover in [39]. In [40], a wave-variable method was applied on a kinematic model called extended virtual-mass model for the car-like mobile robot teleoperation. In the previously described algorithms, one DOF of the local device was used to control the remote car's linear velocity. In [39] and [40], the haptic feedback on the environmental force was achieved for obstacle avoidance by defining a virtual environmental force based on the relative distance and speeds between rover and obstacle. To better emulate normal car driving, we utilize an additional DOF of the local robot as a gas/breaking pedal for controlling the remote car's acceleration, and to generate haptic feedback when the remote car is in hard contact with an obstacle in the environment.

Thus in Chapter 5 of this dissertation, a new tele-driving scheme is proposed, which achieves (q_{m1}, β)-coordination and $(q_{m2}, \dot{\omega})$-coordination and provides haptic feedback that can improve situation awareness of the local driver. Here $(.,.)$ implies that these signals track each other asymptotically. Specifically, (q_{m1}, β)-coordination and $(q_{m2}, \dot{\omega})$-coordination imply that in the proposed control scheme, a local robot's link variables q_{m1}, q_{m2} are used to control the steering angle β and angular acceleration $\dot{\omega}$, respectively. It is worth noting that $(q_{m2}, \dot{\omega})$-coordination is equivalent to $(q_{m2}, \dot{v})$-coordination as $v = r_F\omega$, where r_F is the radius of the front wheel.

1.2.4 Comparison of Three Systems

In this dissertation, the algorithms for content modification attack design and detection are studied for three kinds of networked robotic systems, respectively. A comparison of the three systems is given in the following table:

Table 1.1: Comparison of three networked robotic systems.

System	Dynamics	Topology	Network Delay and Parametric Uncertainty	Human Operator
BTOS	Nonlinear	Two identical nodes	No	Yes
MAS	Linear	Multiple identical nodes	No	No
BTDS	Nonlinear	Two different nodes	Yes	Yes

BTOS and BTDS have two nodes in the system with nonlinear robotic dynamics, and human operator is involved in the system. Here MAS is considered to have multiple identical nodes in the system with a linear double-integrator dynamics, and no human operator is involved in MAS. In BTOS, two nodes have the identical dynamics, but two nodes in BTDS have different dynamics. Also, for BTDS, the network delay and dynamic parametric uncertainty are considered in the study.

1.2.5 Dynamics and Properties of Euler-Lagrange System

In this dissertation, the robotic system can be modeled as an Euler-Lagrange system with the following dynamics equation [64]:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = \tau \tag{1.1}$$

where $q \in \mathbb{R}^n$ represents the generalized coordinates, $M(q)$ is the inertial matrix, $C(q, \dot{q})$ is the centrifugal and Coriolis matrix, $g(q)$ is the gravitational torque and τ is the generalized force acting on the system.

Due to the dynamics structure, (1.1) have the following properties discussed in [64], which are utilized later in this dissertation:

(P1.1) The inertia matrix $M(q)$ is symmetric positive definite matrix, which is lower and upper bounded by

$$\lambda_m I_n \leq M(q) \leq \lambda_M I_n, \tag{1.2}$$

where λ_m, λ_M are positive minimum and maximum eigenvalues of $M(q)$ for all configurations q.

(P1.2) Under an appropriate definition of $C(q, \dot{q})$, the matrix $\dot{M}(q) - 2C(q, \dot{q})$ is skew symmetric.

(P1.3) The centrifugal and Coriolis term $C(q, \dot{q})\dot{q}$ satisfies

$$|C(q, \dot{q})\dot{q}| \leq k_0 |\dot{q}|^2 \tag{1.3}$$

for some $k_0 \in \mathbb{R}^+$.

(P1.4) The dynamics are linearly parametrizable in the sense that

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q) = Y(q, \dot{q}, \ddot{q})\phi \tag{1.4}$$

where $Y(q, \dot{q}, \ddot{q}) \in \mathbb{R}^{n \times p}$ is called regressor which is a matrix of known functions of generalized coordinates and their derivatives, ϕ is called parameters vector which is a constant p-dimensional vector of the inertia parameters (such as mass, moment of inertia, etc.).

1.3 Related Works in Cyber-Physical Security

In addition to the network effects discussed in the previous section that can negatively affect the system performance, severe performance and safety degradation can only occur due to cyber attacks. In this part, an introduction to the related works in cyber-physical security study is given.

1.3.1 General Networked Control System Security

The study of cyber-physical security in networked control systems has been conducted from two perspectives: (i) attack model and design; (ii) defense mechanisms. For attack model and design, [41] defined a 3D attack space by the adversary's system knowledge, disclosure, and disruption resources. In [42], an optimal control problem was studied under the denial-of-service attack where the attacker denies data availability in the system. A replay attack was considered in [43], where normal sensor readings can be recorded and repeated while implementing the attack. Attack design and detection were studied for the LQG control problem. For a similar LQG problem, a data injection attack was designed in [44]. In [45] an undetected covert attack using full system knowledge was designed for both linear and nonlinear networked control system. For defense mechanisms, [46] adopted a quantitative risk management approach. A game theoretical method was studied in [47], where a zero-sum differential game for robust control was coupled with a zero-sum stochastic game for security policy. [48] designed a physical authentication scheme by watermarking the control input. For a linear descriptor system, [49]

defined attack detectability and identifiability, and constructed an observer-based attack detection and identification monitor. Unlike the observer-based method, [50] proposed an energy-based attack detection for passive networked CPS based on the fact that the attack can affect the energy balance of the system.

1.3.2 Multi-Agent System Security

With the fact that MAS can be modeled as a graph with agents as nodes and communication links as edges, the cyber attacks on MAS can be divided into two categories: node attacks and edge attacks. In node attacks, the attacker can convert normal nodes into malicious nodes, whereas in edge attacks, the attacker can break the edge connections or modify the data in edges such as content modification attacks. Some works in the literature have focused on node attacks and resilient MAS design. In [51] and [52], the number of tolerable malicious nodes in the network was characterized by the notion of network connectivity and network robustness, respectively. [53] studied the robustness of consensus tracking under an edge attack called connectivity-broken attack. Other works studied the attack detection and identification methods. An observer-based malicious node detection and identification scheme for first order MAS consensus was proposed in [54]. An unknown input observer-based distributed faulty node and edge detection and isolation scheme for second order MAS consensus was studied in [55].

1.3.3 Bilateral Teleoperation System Security

The security of BTOS has attracted recent attention from the research community since the attacker has the potential to cause harm to the robots, humans and the environment involved, which is highly undesired in any bilateral teleoperation application. The security threats in surgical telerobotics were identified by [56], and an experimental analysis of security threats on the RAVEN surgical robot was presented in [57]. However, no rigorous theoretical analysis of the attack strategies and the prevention/mitigation solutions were discussed. Also it is worth noting that [57] studied a so-called "Surgeon's Intent Modification" attack, which only modifies the data packet from surgeon to surgical robot. The content modification attacks discussed in this dissertation is one kind of "man-in-the-middle-attack" known as "Message modification and spoofing attack" discussed in [56]. This kind of attack can first block the communication between master and slave, and connect independently with both sides, then forward the modified malicious message to both sides. Other works include secure communication protocol design for telesurgery [58] [59], however a theoretical analysis of the attacks has not been considered in this study.

1.4 Contributions of the Dissertation

Compared with the aforementioned works, this dissertation has the following contributions:

For networked robotic systems including BTOS, MAS and BTDS

- In order to better demonstrate the severe attack impact, different from the works [53, 55] considering edge attacks, a systematic framework for designing destabilizing malignant content modification attacks (termed MCoMA) is developed based on the notion of physical storage function.

- To safeguard the system, a physics-based attack detection scheme with an encoding-decoding structure is proposed for general content modification attacks, which can be arbitrary modifications on data content. Compared with the detection schemes in the aforementioned works, **(i)** [49, 50, 54, 55] consider linear system dynamics, whereas the tele-robotic systems with nonlinear robotic dynamics are considered in Chapter 3 and 5, thus their detection schemes on linear systems cannot be directly applied; **(ii)** The observer-based detection schemes [49, 54, 55] and the energy-based detection scheme [50] require the system model and are prone to high computation and communication costs. In contrast, the proposed attack detection scheme does not require the detailed knowledge of system parameters and is solely based on the inherent physical relation within the transmitted data: once the data content is corrupted, the inherent physical relation is violated, allowing the attack to be detected, thus it is fast-response, distributed, computationally light and can detect general content modification attacks (including MCoMA as a special case).

- As part of the research results on BTDS in Chapter 5, a passivity-based adaptive bilateral tele-driving control scheme is proposed in the presence of com-

munication delays and dynamic parametric uncertainties. **(i)** Different from previous works [37–40], the proposed scheme can achieve a new control mode for tele-driving, which is the position-acceleration $(q_{m2}, \dot{\omega})$ coordination as described in Section 1.2.3; **(ii)** Inspired by the formulation in [60], the proposed algorithm avoids the typically assumed passivity assumption on the human and the environment. It should be noted that the control algorithms in [13, 61, 62] can be analogously modified to to make them more broadly applicable.

1.5 Organization

The main results of the dissertation are organized as follows:

- **Overview of Attack Design and Detection Algorithms (Chapter 2):** In this chapter, an overview of attack design and detection algorithms is given. A formal modeling of the attacker is first presented. The design idea of the MCoMA is then described. Based on a simple physics-based detection condition, an attack detection algorithm called physics-based attack detection scheme with an encoding-decoding structure is proposed to detect any general content modification attack, which can be arbitrary modification on the original data content.

- **Content Modification Attack on BTOS (Chapter 3):** Based on the algorithms proposed in Chapter 2, content modification attacks on BTOS are studied, where the attacker can modify the states being exchanged between the master and the slave robot. To demonstrate the damaging attack impact,

a static MCoMA is designed, where the modified state is in the static state feedback form of the original state. The simple physics-based detection condition is utilized for detecting MCoMA on BTOS. To protect the system, the physics-based attack detection scheme with an encoding-decoding structure is applied for detecting any general content modification attack. A BTOS experiment platform consisting of two PHANToM Omni robots is developed. The efficacy of the proposed attacks design and detection algorithms is studied through experiments.

- **Content Modification Attack on MAS (Chapter 4):** Following a similar approach for BTOS, to demonstrate the severe attack impact on second order MAS synchronization, a static MCoMA is first designed. Since multiple links among MAS can be attacked, an optimal MCoMA that is distributed and compromises the least number of links is also designed. The same physics-based attack detection condition is applied for detecting static MCoMA. The physics-based attack detection scheme with an encoding-decoding structure is also applied to protect against general content modification attacks. Additionally, a velocity observer-based attack mitigation scheme is also discussed for MCoMA. The efficacy of the proposed results is illustrated through numerical simulations.

- **Content Modification Attack on BTDS (Chapter 5):** A passivity-based adaptive bilateral tele-driving control scheme is first proposed that enables a human operator to tele-drive a car-like mobile robot with haptic feedback in

the presence of communication delays and dynamic parametric uncertainties. The static MCoMA is designed for the proposed tele-driving system. The physics-based attack detection scheme with an encoding-decoding structure is adopted for detecting general content modification attacks on the tele-driving system. MCoMA design and attack detection scheme are verified through various simulations. An initial BTDS experimental platform is also developed.

1.6 Notations

Throughout the dissertation, the symbols $\mathbb{Z}$, $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, $\mathbb{R}_0^+$ and $\mathbb{R}^+$ denote the sets of positive integers, n-dimensional real-valued vectors, n by m matrices with real-valued elements, sets of nonnegative and positive real numbers, respectively. $|\cdot|$ denotes the Euclidean norm for a vector and denotes the cardinality of a set. $||\cdot||_0$ denotes the l_0 norm that is a total number of non-zero elements in a vector. I_n, $\mathbf{0}_n$, 0_n and 1_n denotes n by n identity matrix, n by n zero matrix, n-dimensional column vector of zeros and n-dimensional column vector of ones, respectively. For any matrix $A \in \mathbb{R}^{n\times n}$, $A \succ 0$ denotes it's positive definite, $A \succeq 0$ denotes it's positive semidefinite, A^T denotes its transpose, A^- denotes generalized inverse, if A is invertible then A^{-1} denotes its inverse and A_i represents its i-th column. $diag$ denotes the diagonal matrix. $\otimes$ denotes Kronecker product. $\mathcal{N}(\cdot)$ denotes the null space of a matrix. $dim(\cdot)$ denotes the dimension of matrix or space. For any function $f : \mathbb{R}_0^+ \to \mathbb{R}^n$, the $\mathcal{L}_\infty$-norm is defined as $\|f\|_\infty = sup_{t\geq 0}|f(t)|$, and $\mathcal{L}_2$-norm is defined as $\|f\|_2 = (\int_0^\infty |f(t)|^2 dt)^{1/2}$. The $\mathcal{L}_\infty$ and $\mathcal{L}_2$ spaces are defined as

the set $\{f : \|f\|_\infty < \infty\}$ and $\{f : \|f\|_2 < \infty\}$, respectively.

Chapter 2: Overview of Attack Design and Detection Algorithms

In this chapter, an overview of attack design and detection algorithms is given. A formal modeling of the attacker is first presented in Section 2.1. Then the design idea of the MCoMA is described in Section 2.2. In Section 2.3, based on a simple physics-based detection condition, an attack detection algorithm called physics-based attack detection scheme with an encoding-decoding structure is proposed to detect any general content modification attack, which can be an arbitrary modification on the original data content.

2.1 Attack Modeling

As shown in Fig. 2.1, two types of cyber attacks can be defined based on the graph structure for networked robotic systems:

- Node attack: an attacker can convert a normal node to a malicious node.
- Edge attack: an attacker can break edge connections or modify information in edges.

In this dissertation, we are specifically interested in one kind of edge attack termed content modification attack that can modify the data content transmitted

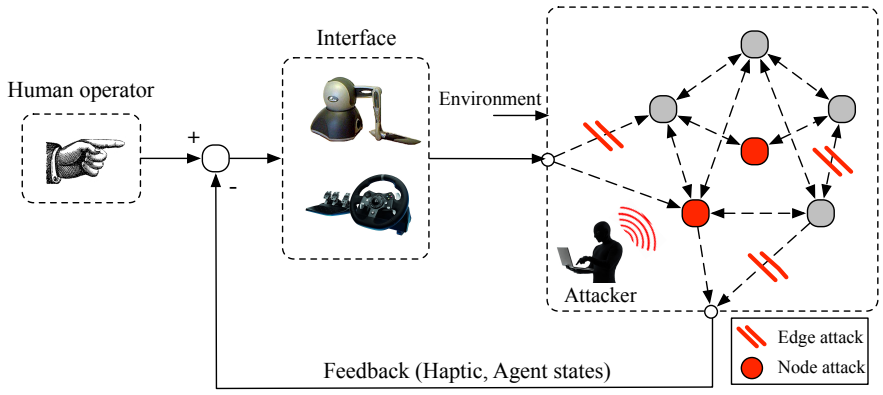


Figure 2.1: General framework of a networked robotic system under cyber attacks.

in communication links.

In this dissertation, the assumptions hold for the attacker.:

(A2.1) Attacker has the knowledge of system dynamics structure, controller structure and controller gains except a set of state initial values (made clear in the sequel).

(A2.2) Attacker has the ability to receive, interpret, manipulate and forward the data between any pair of robots.

In addition, a formal model for the attacker is provided next.

Definition 2.1.1. The attacker is a malicious external entity with assumptions **(A2.1)-(A2.2)** which can launch an edge attack called content modification attack by modifying original data $d \in \mathbb{R}^p$ to modified data $\tilde{d} \in \mathbb{R}^p$ as the following mapping g:

$$g : \mathbb{R}^p \to \mathbb{R}^p, \ d \mapsto \tilde{d}. \tag{2.1}$$

In fact, $\tilde{d} \in \mathbb{R}^p$ can always be expressed as

$$\tilde{d} = d + \hat{d}, \tag{2.2}$$

where $\hat{d} \in \mathbb{R}^p$ is referred to as injection data in this dissertation. In this dissertation, the attack as described by (2.1) and (2.2) is termed as a general content modification attack.

2.2 Attack Design

To demonstrate the severe impact of content modification attack, from the attacker's perspective, a special content modification attack called malignant content modification attack (MCoMA) is designed, which results in an unbounded growth of a physical storage function thus causing system instability and even physical system damage. The formal definition of MCoMA can be given as:

Definition 2.2.1. A content modification attack that modifies the data content transmitted in communication links and ensures that a positive semidefinite system storage function V satisfies $\lim_{t\to\infty} V = \infty$ with $\dot{V} \geq 0$, $\forall t \geq t_a \geq 0$ (t_a is the attack launch time) is called the malignant content modification attack (MCoMA) with respect to storage function V.

Typically, the physical storage function V can be interpreted as the synchronization metric for the system. When V keeps decreasing, the synchronization can be achieved, however, if the content modification attack is designed such that V keeps increasing, the system states can become unbounded and thus cause physical

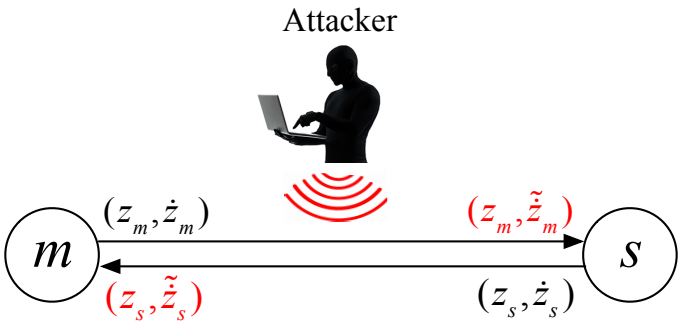


Figure 2.2: An edge with nodes (m, s) in networked robotic system under attack.

damage to the system. Hence, in this dissertation, following the above definition, we will design a static MCoMA for three kinds of networked robotic systems, respectively, where the injection data of attacker is in a static feedback form of the original data. The design procedure for each system is detailed in the following chapters.

2.3 Attack Detection

To establish a general framework for the proposed attack detection algorithms, let us consider an edge in the networked robotic system as shown in Fig. 2.1, in which two nodes (m, s) exchange their states information through the network. As shown in Fig. 2.2, two nodes transmit their states $(z_i, \dot{z}_i),\ i = m, s$ in the network, and the attacker defined in **Definition 2.1.1** can modify the data into $(\tilde{z}_i, \tilde{\dot{z}}_i),\ i = m, s$.

Utilizing the physics of the robotic system, a simple physics-based detection condition can be used to detect the content modification attack. This condition checks if the physical relationship between the received data $\tilde{\dot{z}}_i$ and $\tilde{z}_i$ is violated under the attack. To be specific, the received data $(\tilde{z}_i, \tilde{\dot{z}}_i)$ is only accepted when the

following detection condition is satisfied:

$$\tilde{\dot{z}}_i(t) = \frac{d\tilde{z}_i(t)}{dt} \tag{2.3}$$

However, the simple detection condition (2.3) is not always adequate for protecting the system, because it is easy for the attacker to design the injection data $(\hat{z}_i, \hat{\dot{z}}_i)$ for $(z_i, \dot{z}_i)$ such that the detection condition (2.3) is satisfied, which means the detection is bypassed.

Thus, an attack detection algorithm called physics-based attack detection scheme with an encoding-decoding structure is proposed for any general content modification attack defined as (2.1) and (2.2) (including MCoMA as a special case). The main motivation for designing this encoding-decoding structure is to guarantee that once the encoded data is modified by attacker, the physical relationship between the decoded data must be violated, and hence the attack is detected.

In practice, the scheme is implemented in discrete time. For $i = m, s$, let T_i denote the sufficiently small sampling period for two nodes. On each side, the data is transmitted at the time instance $t = \{k_i T_i | k_i = 1, 2, 3...\}$, where $k_i \in \mathbb{Z}$ is the transmission sequence number.

In the rest of the dissertation, for the notation such as ${}^{i}\lambda_{ia}^{k_i}$, the superscript k_i denotes the sequence number k_i, and the upper left i denotes the transmission direction. In the proposed encoding-decoding scheme, for $i = m$ or s and $j = s$ or m, a set of encoding factors ${}^{i}\lambda_{ia}, {}^{i}\lambda_{ib} \in \mathbb{R}$ and decoding factors ${}^{i}\lambda_{ja}, {}^{i}\lambda_{jb} \in \mathbb{R}$ are utilized. For each transmission sequence k_i, the encoding and decoding factors are

updated with the same dynamics f_a, f_b as

$$ {}^i\lambda_{ia}^{k_i} = {}^i\lambda_{ia}^{k_i-1} + T_i f_a({}^i\lambda_{ia}^{k_i-1}, k_i), \;\; {}^i\lambda_{ib}^{k_i} = {}^i\lambda_{ib}^{k_i-1} + T_i f_b({}^i\lambda_{ib}^{k_i-1}, k_i) \tag{2.4} $$

and

$$ {}^i\lambda_{ja}^{k_i} = {}^i\lambda_{ja}^{k_i-1} + T_i f_a({}^i\lambda_{ja}^{k_i-1}, k_i), \;\; {}^i\lambda_{jb}^{k_i} = {}^i\lambda_{jb}^{k_i-1} + T_i f_b({}^i\lambda_{jb}^{k_i-1}, k_i) \tag{2.5} $$

where functions $f_a, f_b : \mathbb{R} \times \mathbb{Z} \to \mathbb{R}$ are known to both sides and f_a, f_b need to be designed such that ${}^i\lambda_{ja}^{k_i} \neq {}^i\lambda_{jb}^{k_i}$.

In a discrete time setting, the detection condition (2.3) can be checked as

$$ \frac{z_i^{k_i} - z_i^{k_i-1}}{T_i} = \dot{z}_i^{k_i}. \tag{2.6} $$

If the initial values for $({}^i\lambda_{ia}, {}^i\lambda_{ib})$ and $({}^i\lambda_{ja}, {}^i\lambda_{jb})$ can set to be equal as $({}^i\lambda_{ia}^0, {}^i\lambda_{ib}^0) = ({}^i\lambda_{ja}^0, {}^i\lambda_{jb}^0) = ({}^i\lambda_a^0, {}^i\lambda_b^0)$, then the encoding and decoding factors remain equal as ${}^i\lambda_{ia}^{k_i} = {}^i\lambda_{ja}^{k_i} = {}^i\lambda_a^{k_i}$ and ${}^i\lambda_{ib}^{k_i} = {}^i\lambda_{jb}^{k_i} = {}^i\lambda_b^{k_i}$. Recall that in assumption **(A1.1)**, it is assumed that the attacker does have the knowledge of a set of state initial values. To be specific here, we assume the initial values ${}^i\lambda_a^0, {}^i\lambda_b^0, z_i^0$ are unknown to the attacker. In other words, for the proposed detection scheme to succeed, we assume:

(A2.3) For $i = m$ or s, the initial values for the encoding/decoding factors $({}^i\lambda_a^0, {}^i\lambda_b^0)$ and state z_i^0 can be shared securely between both nodes before the operation starts.[1]

If ${}^i\lambda_a^0, {}^i\lambda_b^0$ are unknown to attacker, then the attacker does not know the value of ${}^i\lambda_a^{k_i}, {}^i\lambda_b^{k_i}$ during the operation. If z_i^0 is unknown to attacker, then the detection

[1]It can be argued that the attacker may deduce the initial values using an observer-based approach. Initial investigation to mitigate this possibility has been accomplished in [63].

condition (2.6) can be checked when both sides start to sending the data packet from $k_i = 1$.

Next, the implementation of the proposed attach detection scheme with an encoding-decoding structure is outlined.

For each transmission sequence $k_i = 1, 2, 3, ...$

1. On the sending end $i = m$ or s, first update the encoding factors ${}^i\lambda_{ia}, {}^i\lambda_{ib}$ as (2.4). Then instead of sending $(\dot{z}_i^{k_i}, z_i^{k_i})$, send the encoded data $(r_{ia}^{k_i}, r_{ib}^{k_i})$ with $(r_{ia}^{k_i}, r_{ib}^{k_i})$ given as

$$r_{ia}^{k_i} = \dot{z}_i^{k_i} + {}^i\lambda_{ia}^{k_i} z_i^{k_i}, \ r_{ib}^{k_i} = \dot{z}_i^{k_i} + {}^i\lambda_{ib}^{k_i} z_i^{k_i} \tag{2.7}$$

2. On the receiving end $j = s$ or m, after receiving the data, first update the decoding factors ${}^i\lambda_{ja}, {}^i\lambda_{jb} \in \mathbb{R}$ as (2.5). Then recover $\dot{z}_i^{k_i}, z_i^{k_i}$ by decoding $(r_{ia}^{k_i}, r_{ib}^{k_i})$ as

$$z_i^{k_i} = {}^ia^{k_i}(r_{ia}^{k_i} - r_{ib}^{k_i}), \quad \dot{z}_i^{k_i} = {}^ib^{k_i} r_{ia}^{k_i} - {}^ic^{k_i} r_{ib}^{k_i}, \tag{2.8}$$

where ${}^ia^{k_i} = 1/({}^i\lambda_{ja}^{k_i} - {}^i\lambda_{jb}^{k_i}), {}^ib^{k_i} = {}^i\lambda_{jb}^{k_i}/({}^i\lambda_{jb}^{k_i} - {}^i\lambda_{ja}^{k_i}), {}^ic^{k_i} = {}^i\lambda_{ja}^{k_i}/({}^i\lambda_{jb}^{k_i} - {}^i\lambda_{ja}^{k_i})$. Note ${}^ia^{k_i}, {}^ib^{k_i}, {}^ic^{k_i}$ are only valid when ${}^i\lambda_{ja}^{k_i} \neq {}^i\lambda_{jb}^{k_i}$.

3. After decoding, check the physical relationship between $z_i^{k_i}$ and $\dot{z}_i^{k_i}$ using (2.6).

4. If the detection condition (2.6) is not violated, utilize $\dot{z}_i^{k_i}, z_i^{k_i}$ on the receiving end. Otherwise, the attack is detected, and reject the received data to protect the system.

The following theorem gives a necessary condition that a general content modification attack can avoid the detection.

Theorem 2.3.1. With the proposed attack detection scheme, suppose a general content modification attack is launched at transmission sequence k_i^a and for $k_i \geq k_i^a$, the attacker can modify $(r_{ia}^{k_i}, r_{ib}^{k_i})$ to $(\tilde{r}_{ia}^{k_i}, \tilde{r}_{ib}^{k_i})$ as

$$\tilde{r}_{ia}^{k_i} = r_{ia}^{k_i} + \hat{r}_{ia}^{k_i}, \quad \tilde{r}_{ib}^{k_i} = r_{ib}^{k_i} + \hat{r}_{ib}^{k_i}, \tag{2.9}$$

where $\hat{r}_{ia}^{k_i}, \hat{r}_{ib}^{k_i}$ are arbitrary injection data. Then the above content modification attack can avoid the proposed detection scheme only if

$$({}^{i}a^{k_i} - T_i\, {}^{i}b^{k_i})\hat{r}_{ia}^{k_i} - ({}^{i}a^{k_i} - T_i\, {}^{i}c^{k_i})\hat{r}_{ib}^{k_i} = {}^{i}a^{k_i-1}(\hat{r}_{ia}^{k_i-1} - \hat{r}_{ib}^{k_i-1})$$

where T_i is the sufficiently small sampling period for $i = m, s$.

Proof:

Consider the content modification attack (2.9) is launched. Now by the decoding equation (2.8),

$$\tilde{z}_i^{k_i} = {}^{i}a^{k_i}(\tilde{r}_{ia}^{k_i} - \tilde{r}_{ib}^{k_i}), \quad \tilde{\dot{z}}_i^{k_i} = {}^{i}b^{k_i}\tilde{r}_{ia}^{k_i} - {}^{i}c^{k_i}\tilde{r}_{ib}^{k_i}, \tag{2.10}$$

Executing the detection condition (2.6) for $\tilde{z}_i^{k_i}$ and $\tilde{\dot{z}}_i^{k_i}$ from (2.10), using (2.8), and substituting $\tilde{z}_i^{k_i}$ from (2.10) into the left hand side of (2.6) gives

$$\frac{\tilde{z}_i^{k_i} - \tilde{z}_i^{k_i-1}}{T_i} = \frac{z_i^{k_i} - z_i^{k_i-1}}{T_i} + \frac{{}^{i}a^{k_i}(\hat{r}_{ia}^{k_i} - \hat{r}_{ib}^{k_i})}{T_i} - \frac{{}^{i}a^{k_i-1}(\hat{r}_{ia}^{k_i-1} - \hat{r}_{ib}^{k_i-1})}{T_i}. \tag{2.11}$$

Substituting $\tilde{\dot{z}}_i^{k_i}$ from (2.10) into the right hand side of (2.6), we get

$$\tilde{\dot{z}}_i^{k_i} = \dot{z}_i^{k_i} + {}^{i}b^{k_i}\hat{r}_{ia}^{k_i} - {}^{i}c^{k_i}\hat{r}_{ib}^{k_i}. \tag{2.12}$$

Assume the attacker can avoid the detection condition (2.6), which means $(\tilde{z}_i^{k_i} - \tilde{z}_i^{k_i-1})/T_i = \tilde{\dot{z}}_i^{k_i}$, thus it requires

$$\frac{{}^{i}a^{k_i}(\hat{r}_{ia}^{k_i} - \hat{r}_{ib}^{k_i})}{T_i} - \frac{{}^{i}a^{k_i-1}(\hat{r}_{ia}^{k_i-1} - \hat{r}_{ib}^{k_i-1})}{T_i} = {}^{i}b^{k_i}\hat{r}_{ia}^{k_i} - {}^{i}c^{k_i}\hat{r}_{ib}^{k_i} \tag{2.13}$$

which is equivalent to

$$({}^{i}a^{k_i} - T_i\,{}^{i}b^{k_i})\hat{r}_{ia}^{k_i} - ({}^{i}a^{k_i} - T_i\,{}^{i}c^{k_i})\hat{r}_{ib}^{k_i} = {}^{i}a^{k_i-1}(\hat{r}_{ia}^{k_i-1} - \hat{r}_{ib}^{k_i-1}) \tag{2.14}$$

Hence, the content modification attack (2.9) can avoid the detection only if (2.14) is satisfied.

■

Condition (2.14) represents a complex relation between two consecutive injection data. When ${}^{i}\lambda_a^{k_i}$ and ${}^{i}\lambda_b^{k_i}$ are unknown to the attacker during the operation, the condition (2.14) cannot be satisfied, and thus the attack can be detected. The following corollary provides a simpler sufficient condition for the attack detection.

Corollary 2.3.1. With the proposed attack detection scheme, suppose the attack (2.9) is launched at transmission sequence k_i^a. The general content modification attack is detected if

$$\frac{\hat{r}_{ia}^{k_i^a}}{\hat{r}_{ib}^{k_i^a}} \neq \frac{1 + T_i\,{}^{i}\lambda_{ja}^{k_i^a}}{1 + T_i\,{}^{i}\lambda_{jb}^{k_i^b}}$$

where ${}^{i}\lambda_a^{k_i}, {}^{i}\lambda_b^{k_i}$ are the encoding and decoding factors.

Proof:

The content modification attack (2.9) is launched at $k_i = k_i^a \geq 1$, then the data content before $k_i = k_i^a$ is not compromised. Thus, $\hat{r}_{i1}^{k_i-1}, \hat{r}_{i2}^{k_i-1}$ equal 0_n in condition (2.14) at $k_i = k_i^a$.

Hence at $k_i = k_i^a$, the condition (2.14) is reduced to

$$\frac{\hat{r}_{ia}^{k_i^a}}{\hat{r}_{ib}^{k_i^a}} = \frac{{}^{i}a^{k_i^a} - T_i\,{}^{i}c^{k_i^a}}{{}^{i}a^{k_i^a} - T_i\,{}^{i}b^{k_i^a}} = \frac{1 + T_i\,{}^{i}\lambda_{ja}^{k_i^a}}{1 + T_i\,{}^{i}\lambda_{jb}^{k_i^a}}. \tag{2.15}$$

Since ${}^i\lambda_{ja}^{k_i}$ and ${}^i\lambda_{jb}^{k_i}$ are unknown to the attacker during the operation, the content modification attack (2.9) is detected once the attack is launched at $k_i = k_i^a$ if (2.15) is not satisfied.

■

Remark 2.3.1. As described in the detection scheme, the scheme is solely based on the data between two consecutive sequences k_i and $k_i - 1$. Here the detection scheme is presented in a discrete time setting, which is more realistic in implementation, and the encoding and decoding factors are updated asynchronously once the packet is sent or received as (2.4) and (2.5). Consequently, current scheme is independent of the network delay. Due to other network factors such as noises and quantization errors, the checking condition in (2.6) can be relaxed as

$$e_i^{k_i} := |\frac{z_i^{k_i} - z_i^{k_i-1}}{T_i} - \dot{z}_i^{k_i}| < \epsilon_i, \tag{2.16}$$

where e_i is termed checking error and the constant threshold $\epsilon_i > 0$ can be determined empirically. Since there exists a sudden jump in e_i when an attack is launched, there is a sharp spike in the trajectory of e_i's derivative at the attack time points. Thus, we can have another attack detection condition based on e_i's derivative, which is defined as e_{id}. Then we can say if

$$|e_{id}| := |\frac{e_i^{k_i} - e_i^{k_i-1}}{T_i}| > \sigma_i, \tag{2.17}$$

where σ_i is a large enough positive number, then the attack is detected. The detection condition (2.17) can be used as an additional indication on the attack.

Remark 2.3.2. Unlike the model-based attack detection schemes in [49, 50, 54, 55], the proposed attack detection scheme is solely based on the inherent physical relation within the transmitted data and does not require the knowledge of the system parameters. The detection condition is easy to compute and only relies on two consecutive data samples. Thus, the proposed detection scheme has the following advantages: it is fast, distributed, computationally light, and does not require knowledge of system parameters.

2.4 Summary

In this chapter, a formal attack modeling is first given for this dissertation. Then the design idea of MCoMA is presented, and the specific design procedure will be explained in the following chapters. Considering two nodes exchanging states data in an edge of networked robotic system, a physics-based attack detection scheme with an encoding-decoding structure is proposed for general content modification attacks, which will be applied respectively for the three kinds of networked robotic systems in the rest of the dissertation. The proposed attack detection algorithm can also potentially be applied for detecting general content modification attacks in a variety of networked control/robotic systems.

Chapter 3: Content Modification Attacks on Bilateral Teleoperation System

In this chapter, the content modification attack on BTOS synchronization shown in Fig. 3.1 is studied. The master and slave robotic manipulators exchange their states (joint position and velocity) through a communication network to achieve the BTOS synchronization. As discussed in Chapter 1, the BTOS synchronization can be viewed as a special case of MAS synchronization, only in this case **(i)** two nodes with nonlinear manipulator dynamics are considered and **(ii)** interactions with human/environment are involved in the system. The BTOS suffers from the attacker defined in **Definition** 2.1.1, who can launch content modification attacks to modify the content of the transmitted data. The attack on BTOS has potentials to cause harm to the robots, the human and the environment involved, which is highly undesired in any bilateral teleoperation application.

The results of this chapter can be summarized as follows: Based on the algorithms proposed in Chapter 2, content modification attacks on BTOS are studied, where the attacker can modify the states being exchanged between the master and the slave robot. To demonstrate the damaging attack impact, a static MCoMA is designed, where the modified state is in the form of static state feedback of the original

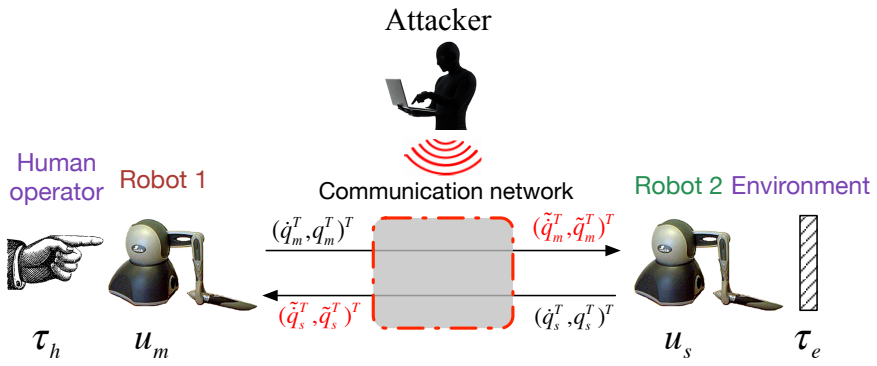


Figure 3.1: The attacker can externally compromise the data content in communication network.

state. The simple physics-based detection condition is utilized for detecting MCoMA on BTOS. To protect the system, the physics-based attack detection scheme with an encoding-decoding structure is applied for detecting any general content modification attack. A BTOS experiment platform consisting of two PHANToM Omni robots is developed. The efficacy of the proposed attacks design and detection algorithms is studied through experiments.

The rest of this chapter is organized as follows: In Section 3.1, the attack design is discussed for a PD-like BTOS control scheme. In Section 3.1.1, a PD-like BTOS control scheme is first introduced as a simple case for the attack design. Then in Section 3.1.2, the MCoMA design is described in the context of BTOS. In Section 3.2, the attack detection scheme for content modification attacks is discussed. Finally, the experiment results are presented in Section 3.3.

3.1 Attack Design for BTOS

In this section, to demonstrate the potential of a severe attack, a special content modification attack called MCoMA is designed for a simple PD-like BTOS control scheme based on Section 2.2. The MCoMA can result in an unbounded growth of a physical storage function, thereby causing system instability and even physical system damage.

3.1.1 PD-like control scheme for BTOS

Consider a BTOS consisting of a pair of nonlinear revolute robotic manipulators coupled via a communication network, as shown in Fig. 3.1. Ignoring external disturbances and friction, given Section 1.2.5, the dynamics of n-link master and slave robots are given as

$$\begin{aligned} M_m(q_m)\ddot{q}_m + C_m(q_m,\dot{q}_m)\dot{q}_m + G_m(q_m) &= \tau_h + \tau_m, \\ M_s(q_s)\ddot{q}_s + C_s(q_s,\dot{q}_s)\dot{q}_s + G_s(q_s) &= \tau_s - \tau_e, \end{aligned} \tag{3.1}$$

where subscript $i = m,\ s$ denotes the master and slave robot, respectively. Henceforth, subscript i will represent both master and slave robots. Here, $\ddot{q}_i, \dot{q}_i, q_i \in \mathbb{R}^n$ are the angular acceleration, velocity and position, respectively, $M_i(q_i) \in \mathbb{R}^{n\times n}$ is the inertia matrix, $C_i(q_i,\dot{q}_i) \in \mathbb{R}^{n\times n}$ is the centrifugal and Coriolis matrix, $G_i(q_i) \in \mathbb{R}^n$ is the gravitational torque, $\tau_i \in \mathbb{R}^n$ is the robot control input, and $\tau_h, \tau_e \in \mathbb{R}^n$ are torques exerted by the human operator and the environment, respectively.

In this section, to simplify the theoretical analysis of the attack design, the

following assumptions are made for the BTOS:

(A3.1) The network condition is perfect, which means the network effects such as time delays, data losses are not considered in the attack design.

(A3.2) The human operator and environment can be modeled as passive systems,

$$-\int_0^t \dot{q}_m(s)^T \tau_h(s) ds \geq -\beta_h, \quad \int_0^t \dot{q}_s(s)^T \tau_e(s) ds \geq -\beta_e \tag{3.2}$$

for $t > 0$ and some $\beta_h, \beta_e \in \mathbb{R}_0^+$.

(A3.3) The gravitational forces are pre-compensated such that,

$$\tau_m = u_m + G_m(q_m), \quad \tau_s = u_s + G_s(q_s) \tag{3.3}$$

in the given dynamics (3.1). This reduces the overall dynamics of the BTOS as

$$\begin{aligned} M_m(q_m)\ddot{q}_m + C_m(q_m, \dot{q}_m)\dot{q}_m &= \tau_h + u_m, \\ M_s(q_s)\ddot{q}_s + C_s(q_s, \dot{q}_s)\dot{q}_s &= u_s - \tau_e. \end{aligned} \tag{3.4}$$

The control structure is shown in Fig. 3.1, the data consisting of velocity and position as $(\dot{q}_i^T, q_i^T)^T$ is transmitted between the two robots, the received data is denoted by $(\tilde{\dot{q}}_i^T, \tilde{q}_i^T)^T$.

Proposition 3.1.1. Assume **(A3.1)-(A3.3)** hold and the data-transmission through the communication link is perfect ($\tilde{\dot{q}}_i = \dot{q}_i, \tilde{q}_i = q_i$). Consider the BTOS dynamics given in (3.4) controlled by a PD-like controller:

$$\begin{aligned} u_m &= -K_d(\dot{q}_m - \tilde{\dot{q}}_s) - K_p(q_m - \tilde{q}_s) - K_{dm}\dot{q}_m, \\ u_s &= -K_d(\dot{q}_s - \tilde{\dot{q}}_m) - K_p(q_s - \tilde{q}_m) - K_{dm}\dot{q}_s, \end{aligned} \tag{3.5}$$

where $K_d, K_p, K_{dm} \in \mathbb{R}^+$. Then,

(a) The position error $e_q := q_s - q_m$ is bounded, and velocity error $\dot{e}_q := \dot{q}_s - \dot{q}_m$ asymptotically converges to zero, $\lim_{t\to\infty} \dot{e}_q = 0_n$.

(b) In the free motion case ($\tau_h = \tau_e = 0_n$), the states get synchronized, $\lim_{t\to\infty} e_q = \lim_{t\to\infty} \dot{e}_q = 0_n$.

(c) If $\dot{q}_i = \ddot{q}_i = 0_n$, the environment contact force is accurately transmitted back to the human operator, $\tau_h = \tau_e$.

It should be noted that **Proposition 3.1.1** does not constitute the main results of this chapter, but here a proof is given just for completeness. The proof idea is similar to proof of Theorem 3.5 in [13] and Proposition 2 in [65]. Before the proof is given, we also need to state the following Barbalat's lemma [66]:

Lemma 3.1.1. The Barbalat's lemma can be stated in following two equivalent forms:

(a) If function $f(t) \in \mathcal{L}_2$ and $\dot{f}$ is bounded, then $f \to 0$ as $t \to \infty$.

(b) If function $f(t)$ has a finite limit as $t \to \infty$ and if $\dot{f}$ is uniformly continuous ($\ddot{f} \in \mathcal{L}_\infty$), then $\dot{f} \to 0$ as $t \to \infty$.

Now the proof of **Proposition** 3.1.1 can be given:

Proof: Consider the following Lyapunov function candidate:

$$V = \frac{1}{2}\dot{q}_s^T M_s \dot{q}_s + \frac{1}{2}\dot{q}_m^T M_m \dot{q}_m + \frac{1}{2}(q_s - q_m)^T K_p (q_s - q_m) - \int_0^t \dot{q}_m^T \tau_h + \beta_h + \int_0^t \dot{q}_s^T \tau_e + \beta_e \tag{3.6}$$

The property **(P1.1)** and assumption **(A3.2)** can guarantee the positive semidefiniteness of V. The derivative of V along the system trajectories described by (3.4)

and (3.5) with the property **(P1.2)** is given by

$$\dot{V} = -\dot{q}_s^T K_{dm} \dot{q}_s - \dot{q}_m^T K_{dm} \dot{q}_m - \dot{q}_s^T (K_d(\dot{q}_s - \bar{\dot{q}}_m) + K_p(q_s - \tilde{q}_m)) - \dot{q}_m^T (K_d(\dot{q}_m - \bar{\dot{q}}_s) \tag{3.7}$$

$$+ K_p(q_m - \tilde{q}_s)) + (\dot{q}_s - \dot{q}_m)^T K_p(q_s - q_m)$$

If transmission is perfect ($\bar{\dot{q}}_i = \dot{q}_i, \tilde{q}_i = q_i$), then

$$\begin{aligned} \dot{V} &= -\dot{q}_s^T K_{dm} \dot{q}_s - \dot{q}_m^T K_{dm} \dot{q}_m - (\dot{q}_s - \dot{q}_m)^T K_d (\dot{q}_s - \dot{q}_m) \\ &= -K_{dm} |\dot{q}_m|^2 - K_{dm} |\dot{q}_s|^2 - K_d |\dot{e}_q|^2. \end{aligned} \tag{3.8}$$

Since $\dot{V}$ is negative semi-definite, $\lim_{t\to\infty} V$ is finite, thus $\dot{q}_s, \dot{q}_m, q_s - q_m$ are bounded. Integrating $\dot{V}$ from 0 to t, we get

$$V(t) - V(0) = -K_{dm} ||\dot{q}_m||_{L_2}^2 - K_{dm} ||\dot{q}_s||_{L_2}^2 - K_d ||\dot{e}_q||_{L_2}^2 \tag{3.9}$$

Thus, $\dot{q}_m, \dot{q}_s, \dot{e}_q \in \mathcal{L}_2$.

From (3.4), we have

$$\begin{aligned} \ddot{q}_m &= M_m(q_m)^{-1}(-C(q_m, \dot{q}_m)\dot{q}_m + \tau_h + u_m) \\ \ddot{q}_s &= M_s(q_s)^{-1}(-C(q_s, \dot{q}_s)\dot{q}_s + \tau_e + u_s) \end{aligned} \tag{3.10}$$

We can find $\ddot{q}_m, \ddot{q}_s$ are also bounded. Invoking Barbalat's lemma (a), we can see $\dot{q}_m, \dot{q}_s, \dot{e}_q \to 0$ as $t \to \infty$. This completes the proof of part (a).

Consider the free motion case, differentiating $\ddot{q}_m$ and using property **(P1.1)** and **(P1.3)**, one can show $\dddot{q}_m \in \mathcal{L}_\infty$, thus $\ddot{q}_m$ is uniformly continuous. Invoking Barbalat's lemma (b), since $\dot{q}_m$ has a finite limit as $t \to \infty$ and $\ddot{q}_m$ is uniformly continuous, then $\ddot{q}_m \to 0$ as $t \to \infty$.

Now consider the dynamic (3.10) again, then we can find $e_q = q_s - q_m \to 0$ as $t \to \infty$. This completes the proof of part (b).

Additionally, in the static case ($\dot{q}_i = \ddot{q}_i = 0$), from dynamic (3.4), we can see $\tau_h = K_p(q_m - q_s) = \tau_e$, which completes the proof of part (c).

■

Remark 3.1.2. Since the essential goal of Section 3.1 is to demonstrate a constructive methodology for attack design, assumptions **(A3.1)**-**(A3.3)** are required so that a PD-like BTOS control scheme can be considered for the theoretical attack design. Attack design for other advanced BTOS control architectures can be similarly accomplished and assumptions specific to the chosen architecture would then replace **(A3.1)**-**(A3.3)**. The network assumption **(A3.1)** is used in the current section to simplify the theoretical attack design. It is not assumed in the experiments of Section 3.3, and it is also not explicitly required in the attack detection scheme in Section 3.2. The consequences of realistic network effects for implementing the proposed detection scheme are discussed in ***Remark*** *2.3.1*.

3.1.2 MCoMA Design

In this subsection, the MCoMA design for BTOS is discussed following the design idea in Section 2.2. Based on **Definition** 2.1.1, for the BTOS with a PD-like control scheme in Fig. 3.1, the attacker can launch a content modification attack by modifying the data content $(\dot{q}_i^T, q_i^T)^T$ being exchanged by the master and slave robots and intelligently replace them with $(\tilde{\dot{q}}_i^T, \tilde{q}_i^T)^T$ for $i = m, s$.

Besides the assumptions made in Section 2.1, an additional assumption is made for the attacker:

(A3.4) The attack starts at $t = t_a \geq 0$, assume $q(t_a) \notin \boldsymbol{E}$, where

$$\boldsymbol{E} = \{q := (\dot{q}_s^T, q_s^T, \dot{q}_m^T, q_m^T)^T | \dot{q}_s = \dot{q}_m = 0_n, q_s = q_m\}. \tag{3.11}$$

is the equilibrium set for BTOS.

In this subsection, based on **Definition** 2.2.1, the MCoMA with respect to storage function (3.6) is studied for BTOS. The storage function (3.6) can be used as a Lyapunov-like function for the system (3.4) to demonstrate master slave synchronization in the system. In other words, (3.6) can be viewed as a metric for BTOS synchronization. Clearly, MCoMA can not only prevent the BTOS synchronization but also lead to violent instability in the overall system. V is positive semidefinite due to the property **(P1.1)** and assumption **(A3.2)**. An increasing V implies that controllers can eventually drive the motors of the revolute robotic manipulators to their maximum speed limits for an indefinite amount of time. This can put stress on the mechanical system of the robots, thereby making the BTOS unsafe for human users and the environment. Hence, it is essential to study the impact of such content modification attacks. To be noted, it is not always necessary for the attacker to launch MCoMA if the attacker's goal is just to disrupt the system, but here the MCoMA can be considered as the worst case design from the system's perspective in terms of the physical storage function and can be used to justify the potentially damaging impact of the content modification attack. The detection scheme for general content modification attacks is discussed in Section 3.2.

Specifically, a type of MCoMA called static MCoMA is proposed for the BTOS, where the injection data is in static feedback form of the original data.

Theorem 3.1.1. Consider BTOS dynamics (3.4) controlled by (3.5). A content modification attack, which modifies the states being exchanged between the master and the slave robot as

$$\tilde{q} = q + \hat{q} = q + Kq, \tag{3.12}$$

with original data $q = (\dot{q}_s^T, q_s^T, \dot{q}_m^T, q_m^T)^T$, modified data $\tilde{q} = (\tilde{\dot{q}}_s^T, \tilde{q}_s^T, \tilde{\dot{q}}_m^T, \tilde{q}_m^T)^T$, injection data $\hat{q} = (\hat{\dot{q}}_s^T, \hat{q}_s^T, \hat{\dot{q}}_m^T, \hat{q}_m^T)^T$ and 'attack gain' constant $K \in \mathbb{R}^{4n\times 4n}$, is a MCoMA with respect to storage function V in (3.6) if and only if

(a) $P := A + BK$ is positive semidefinite;

(b) the set $\boldsymbol{S}\backslash\boldsymbol{E}$ is not positively invariant, where $\boldsymbol{S} := \{q \in \mathbb{R}^{4n} | q^T P q = 0\}$,

$$A = \begin{bmatrix} -(K_{dm}+K_d)I_n & \mathbf{0}_n & K_d I_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ K_d I_n & \mathbf{0}_n & -(K_{dm}+K_d)I_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}, \quad B = \begin{bmatrix} \mathbf{0}_n & \mathbf{0}_n & K_d I_n & K_p I_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ K_d I_n & K_p I_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}.$$

Proof:

Consider the storage function given in (3.6) for the BTOS. Clearly, $V > 0, \forall q \in \mathbb{R}^{4n}\backslash\boldsymbol{E}$.

Using property **(P1.1)**, the derivative of V along the system trajectories de-

scribed by (3.4) and (3.5) is given by

$$\dot{V} = \frac{1}{2}\dot{q}_s^T \dot{M}_s \dot{q}_s + \dot{q}_s^T M_s \ddot{q}_s + \frac{1}{2}\dot{q}_m^T \dot{M}_m \dot{q}_m + \dot{q}_m^T M_m \ddot{q}_m + (\dot{q}_s - \dot{q}_m)^T K_p(q_s - q_m) - \dot{q}_m^T \tau_h + \dot{q}_s^T \tau_e$$
$$= \frac{1}{2}\dot{q}_s^T \dot{M}_s \dot{q}_s + \dot{q}_s^T(-C_s\dot{q}_s + u_s) + \frac{1}{2}\dot{q}_m^T \dot{M}_m \dot{q}_m + \dot{q}_m^T(-C_m\dot{q}_m + u_m) + (\dot{q}_s - \dot{q}_m)^T K_p(q_s - q_m). \tag{3.13}$$

Using property **(P1.2)**,

$$\dot{V} = \dot{q}_s^T u_s + \dot{q}_m^T u_m + (\dot{q}_s - \dot{q}_m)^T K_p(q_s - q_m) = -\dot{q}_s^T K_{dm}\dot{q}_s - \dot{q}_m^T K_{dm}\dot{q}_m - \dot{q}_s^T(K_d(\dot{q}_s - \tilde{\dot{q}}_m)$$
$$+ K_p(q_s - \tilde{q}_m)) - \dot{q}_m^T(K_d(\dot{q}_m - \tilde{\dot{q}}_s) + K_p(q_m - \tilde{q}_s)) + (\dot{q}_s - \dot{q}_m)^T K_p(q_s - q_m). \tag{3.14}$$

Substituting $\tilde{q}$ from (3.12) into (3.14) gives

$$\dot{V} = q^T(A + BK)q = q^T P q. \tag{3.15}$$

Sufficiency:

From **(A3.4)**, $q(t_a) \notin \boldsymbol{E}$, which means $V(t_a)$ is positive. Using (3.15), the following conclusions can be made:

(a) if P is positive semidefinite, then V always stays positive $\forall t \geq t_a$ and q never enters $\boldsymbol{E}$.

(b) if the set $\boldsymbol{S}\backslash\boldsymbol{E}$ is not positively invariant, then $\dot{V}$ cannot always stay in 0, thus the V ultimately keeps increasing.

Hence from **Definition** 2.2.1, a content modification attack given by (3.12) is MCoMA with respect to storage function V in (3.6).

Necessity:

By **Definition** 2.2.1, if the content modification attack is MCoMA with respect to storage function V in (3.6), then $\dot{V} \geq 0$ in (3.15), which implies P is

positive semidefinite. Further, $V \to \infty$ as $t \to \infty$, implies the set $\mathbf{S}\backslash\mathbf{E}$ should not be positively invariant. Hence, both **(a)** and **(b)** are necessary.

■

As the 'attack gain' K is a constant matrix, the content modification attack (3.12) is referred to as static MCoMA.

Remark 3.1.3. Among different choices for the attacker to modify the content, the proposed MCoMA (3.12) guarantees the increase of the storage function V (given by (3.6)) and also gives a provision for controlling the degree of instability in the BTOS by adjusting $\dot{V}$ (given by (3.15)), and this point will be further demonstrated in the sequel.

The following proposition discusses a simple MCoMA design method with the help of **Theorem** 3.1.1.

Proposition 3.1.4. Consider BTOS dynamics (3.4) controlled by (3.5). A content modification attack that modifies the states being exchanged between the master and slave robots in the form given by (3.12) is a MCoMA with respect to storage function V in (3.6), if the 'attack gain'

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13}+\Lambda_3 & K_{14} \\ K_{21} & K_{22} & K_{23}+\Lambda_4 & K_{24} \\ K_{31}+\Lambda_1 & K_{32} & K_{33} & K_{34} \\ K_{41}+\Lambda_2 & K_{42} & K_{43} & K_{44} \end{bmatrix},$$

where $K_{ij} \in \mathbb{R}^{n\times n}$ are constant coefficient matrices satisfying

$$\begin{aligned}
-(K_d + K_{dm})I_n + K_d K_{31} + K_p K_{41} &= \mathbf{0}_n \\
-(K_d + K_{dm})I_n + K_d K_{13} + K_p K_{23} &= \mathbf{0}_n \\
K_d I_n + K_d K_{11} + K_p K_{21} &= \mathbf{0}_n \\
K_d I_n + K_d K_{33} + K_p K_{43} &= \mathbf{0}_n \\
K_d K_{32} + K_p K_{42} = \mathbf{0}_n,\ \ K_d K_{12} + K_p K_{22} &= \mathbf{0}_n \\
K_d K_{34} + K_p K_{44} = \mathbf{0}_n,\ \ K_d K_{14} + K_p K_{24} &= \mathbf{0}_n
\end{aligned} \tag{3.16}$$

and $\{\Lambda_j\} \in \mathbb{R}^{n\times n}$ are positive definite matrices.

Proof:

Consider the storage function V given by (3.6) and from (3.15), $\dot{V} = q^T(A + BK)q$.

Let $K = K^* + \hat{K}$, where

$$K^* = \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{33} & K_{34} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{bmatrix}, \quad \hat{K} = \begin{bmatrix} \mathbf{0}_n & \mathbf{0}_n & \Lambda_3 & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \Lambda_4 & \mathbf{0}_n \\ \Lambda_1 & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ \Lambda_2 & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}.$$

Substituting K into $\dot{V}$ gives

$$\dot{V} = q^T(A + BK^*)q + q^T B\hat{K}q. \tag{3.17}$$

Using (3.16) in (3.17) gives

$$\begin{aligned}\dot{V} &= q^T B \hat{K} q \\ &= \dot{q}_s^T K_d \Lambda_1 \dot{q}_s + \dot{q}_s^T K_p \Lambda_2 \dot{q}_s + \dot{q}_m^T K_d \Lambda_3 \dot{q}_m + \dot{q}_m^T K_p \Lambda_4 \dot{q}_m. \end{aligned} \tag{3.18}$$

It is obvious that $\dot{V}$ is positive semidefinite, which implies the condition **(a)** in **Theorem** 3.1.1 is satisfied.

In this case, $\boldsymbol{S} = \{q \in \mathbb{R}^{4n} | \dot{q}_s = \dot{q}_m = 0_n\}$ (refer to condition **(b)** in **Theorem** 3.1.1), thus $\boldsymbol{S} \backslash \boldsymbol{E} = \{q \in \mathbb{R}^{4n} | \dot{q}_s = \dot{q}_m = 0_n, q_s \neq q_m\}$.

With $\tilde{q} = q + Kq$, the controller (3.5) becomes

$$\begin{aligned} u_m &= (K_d \Lambda_3 + K_p \Lambda_4) \dot{q}_m - K_p (q_m - q_s), \\ u_s &= (K_d \Lambda_1 + K_p \Lambda_2) \dot{q}_s - K_p (q_s - q_m). \end{aligned} \tag{3.19}$$

When $\dot{q}_s$ and $\dot{q}_m$ are 0_n, the nonzero $|q_m - q_s|$ term in control can drive the state away from zero thereby rendering the velocities nonzero. Thus, $\boldsymbol{S} \backslash \boldsymbol{E}$ is not positively invariant set, which satisfies condition **(b)** of **Theorem** 3.1.1.

Hence, the content modification attack given in the proposition is MCoMA with respect to storage function V in (3.6).

■

As it is mentioned in **Remark** 3.1.3, $\dot{V}$ can be adjusted by appropriately selecting $\{\Lambda_j\}$.

3.2 Attack Detection for BTOS

The detection algorithms proposed in Section 2.3 can be applied here to detect the content modification attacks on BTOS by replacing the notation $(\dot{z}_i, z_i)$ in Section 2.3 with $(\dot{q}_i, q_i)$ here. In this section, the simple physics-based detection condition discussed in Section 2.3 is first applied to detect the static MCoMA proposed in Section 3.1.2. Then the physics-based detection scheme with an encoding-decoding structure proposed in Section 2.3 is adopted for general content modification attacks on BTOS.

3.2.1 Attack Detection on Static MCoMA

For detecting the static MCoMA proposed in Section 3.1.2, the simple physics-based detection condition (2.3) can be executed on both receiving ends, where the inherent physical relation between constituents of the transmitted data is checked. To be specific in this case, the received data $(\tilde{q}_i, \tilde{\dot{q}}_i)$ is only accepted when the following detection condition is satisfied:

$$\tilde{\dot{q}}_i(t) = \frac{d\tilde{q}_i(t)}{dt}, \quad \forall t \geq 0. \tag{3.20}$$

The following proposition shows the above simple condition can detect the proposed static MCoMA:

Proposition 3.2.1. Consider BTOS dynamics (3.4) controlled by (3.5). The proposed static MCoMA in **Theorem** 3.1.1 can be detected if the detection condition (3.20) is checked for the received data on both master and slave side.

Proof:

Let the 'attack gain' be

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{33} & K_{34} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{bmatrix}, \tag{3.21}$$

where $K_{ij} \in \mathbb{R}^{n\times n}, \quad i,j \in \{1,2,3,4\}$ are constant matrices. Thus, the modified state $\tilde{q}$ is given by (3.12) as

$$\begin{aligned} \tilde{\dot{q}}_s &= (K_{11} + I_n)\dot{q}_s + K_{12}q_s + K_{13}\dot{q}_m + K_{14}q_m \\ \tilde{q}_s &= K_{21}\dot{q}_s + (K_{22} + I_n)q_s + K_{23}\dot{q}_m + K_{24}q_m \\ \tilde{\dot{q}}_m &= K_{31}\dot{q}_s + K_{32}q_s + (K_{33} + I_n)\dot{q}_m + K_{34}q_m \\ \tilde{q}_m &= K_{41}\dot{q}_s + K_{42}q_s + K_{43}\dot{q}_m + (K_{44} + I_n)q_m \end{aligned} \tag{3.22}$$

Using the dynamics equations (3.4) in (3.22) gives

$$\begin{aligned} \dot{\tilde{q}}_s &= K_{21}M_s^{-1}(q_s)(u_s - C_s(\dot{q}_s, q_s)\dot{q}_s) + (K_{22} + I_n)\dot{q}_s + K_{23}M_m^{-1}(q_m)(u_m \\ &\quad - C_m(\dot{q}_m, q_s)\dot{q}_m) + K_{24}\dot{q}_m, \\ \dot{\tilde{q}}_m &= K_{41}M_s^{-1}(q_s)(u_s - C_s(\dot{q}_s, q_s)\dot{q}_s) + K_{42}\dot{q}_s + K_{43}M_m^{-1}(q_m)(u_m - C_m(\dot{q}_m, q_s)\dot{q}_m) \\ &\quad + (K_{44} + I_n)\dot{q}_m. \end{aligned} \tag{3.23}$$

Assume the static MCoMA (3.22) can bypass detection condition (3.20), thus both $\dot{\tilde{q}}_s$ and $\dot{\tilde{q}}_m$ given by (3.23) should be linear in $\dot{q}_s$ and $\dot{q}_m$. Equivalently,

$$K_{21} = K_{23} = K_{41} = K_{43} = \mathbf{0}_n. \tag{3.24}$$

Substituting (3.24) in (3.23) gives,

$$\begin{aligned}\dot{\tilde{q}}_s &= (K_{22} + I_n)\dot{q}_s + K_{24}\dot{q}_m,\\ \dot{\tilde{q}}_m &= K_{42}\dot{q}_s + (K_{44} + I_n)\dot{q}_m.\end{aligned} \tag{3.25}$$

Using (3.25) and (3.22), it is clear that (3.20) holds if and only if

$$\begin{aligned}&K_{12} = K_{14} = K_{32} = K_{34} = \mathbf{0}_n,\\ &K_{11} = K_{22} = G_1,\ K_{13} = K_{24} = G_2,\\ &K_{31} = K_{42} = G_3,\ K_{33} = K_{44} = G_4.\end{aligned} \tag{3.26}$$

Thus, the static MCoMA given by (3.12) satisfies the detection condition (3.20) if and only if

$$K = \begin{bmatrix} G_1 & \mathbf{0}_n & G_2 & \mathbf{0}_n \\ \mathbf{0}_n & G_1 & \mathbf{0}_n & G_2 \\ G_3 & \mathbf{0}_n & G_4 & \mathbf{0}_n \\ \mathbf{0}_n & G_3 & \mathbf{0}_n & G_4 \end{bmatrix}, \tag{3.27}$$

where $G_j \in \mathbb{R}^{n\times n}$ are constant matrices. In **Theorem** 3.1.1, it is shown that with $\tilde{q} = q + Kq$, $\dot{V} = q^T P q$, where $P \coloneqq A + BK$ is positive semidefinite and given by

$$P = \begin{bmatrix} -(K_d + K_{dm})I_n + K_d G_3 & K_p G_3 & K_d(I_n + G_4) & K_p G_4 \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ K_d(I_n + G_1) & K_p G_1 & -(K_d + K_{dm})I_n + K_d G_2 & K_p G_2 \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}. \tag{3.28}$$

which implies,

$$\dot{V}(0) = q(0)^T P q(0) \geq 0, \quad \forall q(0) \in \mathbb{R}^{4n}\backslash \boldsymbol{E}. \tag{3.29}$$

Inequality (3.29) should hold for

$$q(0) = (\dot{q}_s^T, q_s^T, 0_n^T, 0_n^T)^T, \quad \dot{q}_s, q_s \in \mathbb{R}^n \backslash \{0_n\}.$$

Substituting this $q(0)$ in (3.29) implies,

$$P_1 = \begin{bmatrix} -(K_d + K_{dm})I_n + K_d G_3 & K_p G_3 \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} \tag{3.30}$$

should be a positive semidefinite matrix.

Matrix P_1 given in (3.30) is positive semidefinite if and only if

$$P_1 + P_1^T = \begin{bmatrix} -2(K_d + K_{dm})I_n + K_d(G_3 + G_3^T) & K_p G_3 \\ K_p G_3^T & \mathbf{0}_n \end{bmatrix}$$

is symmetric positive semidefinite. From generalized Schur's complement condition [67], $P_1 + P_1^T$ is a symmetric positive semidefinite if and only if

$$P_1^{'} = -2(K_d + K_{dm})I_n + K_d(G_3 + G_3^T)$$

$$P_1^{''} = -K_p^2 G_3^T \left(-2(K_d + K_{dm})I_n + K_d(G_3 + G_3^T)\right)^{-} G_3$$

are both symmetric positive semidefinite matrices. However, matrices $P_1^{'}$ and $P_1^{''}$ can not be symmetric positive semidefinite simultaneously. Thus, by contradiction, the static MCoMA in **Theorem** 3.1.1 can be detected by the detection condition (3.20), which completes the proof.

■

As discussed in Section 2.3, the simple detection condition (3.20) is not sufficient for protecting the system, because it is easy for the attacker to design the injection data $(\hat{q}_i, \hat{\dot{q}}_i)$ for $(q_i, \dot{q}_i)$ such that the detection condition (3.20) is satisfied, which

means the detection is bypassed. Thus, the physics-based attack detection scheme with an encoding-decoding structure proposed in Section 2.3 should be adopted for detecting general content modification attack, which includes MCoMA as a special case.

3.2.2 Attack Detection on General Content Modification Attack

As shown in Section 3.1.2, MCoMA is only a special content modification attack. In the case that the attacker just launches a general content modification attack defined as (2.1) and (2.2) to disrupt the system, we can adopt the physics-based attack detection scheme with an encoding-decoding structure proposed in Section 2.3 for BTOS as shown in Fig. 3.2. For simplicity, only one transmission direction from the master robot to the slave robot is shown in Fig. 3.2, and the implementation in the opposite transmission direction is the same.

3.3 Experiment

In this section, the theoretical results are tested on a BTOS experimental platform which consists of two PHANToM Omni robots. The PHANToM Omni robot, developed by SensAble Technology, is a 6 degree-of-freedom (DOF) haptic device which can apply a force feedback on the user's hand, and allow the user to feel virtual objects and interact with the virtual environment.

The Omni robot consists of 6 revolute joints where the last 3 joints are not actuated. The device is shown in Fig. 4.11, where $J1 - J6$ denote 6 revolute joints

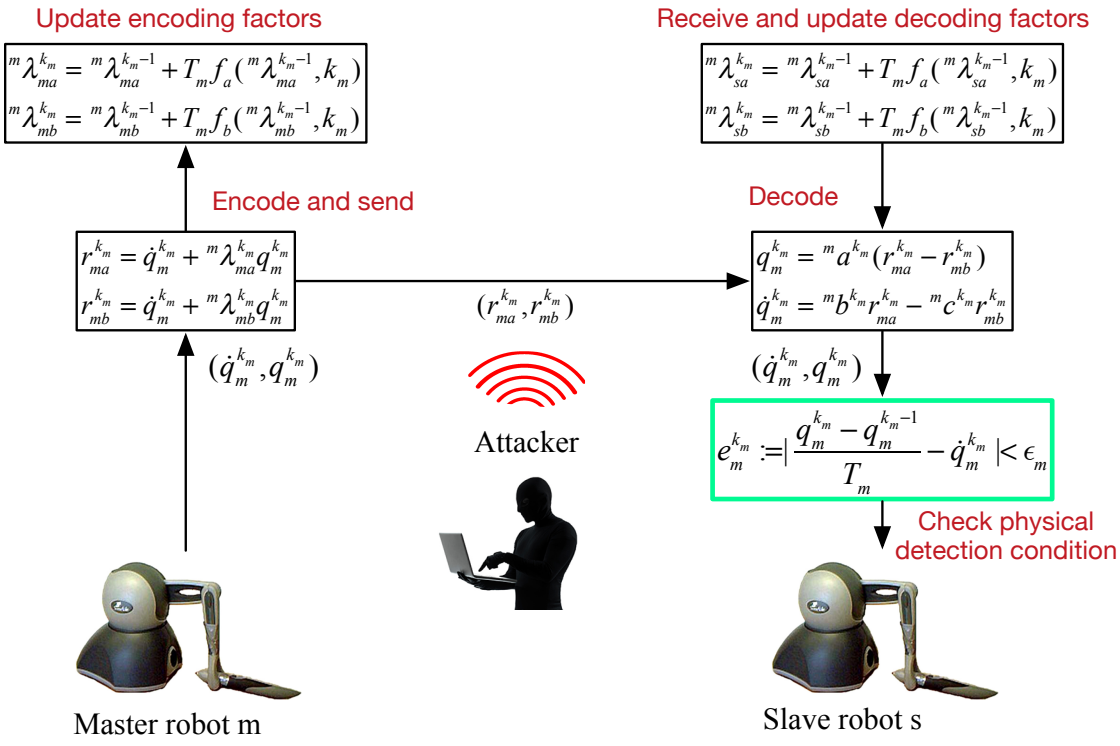


Figure 3.2: Physics-based attack detection scheme with an encoding-decoding structure for BTOS.

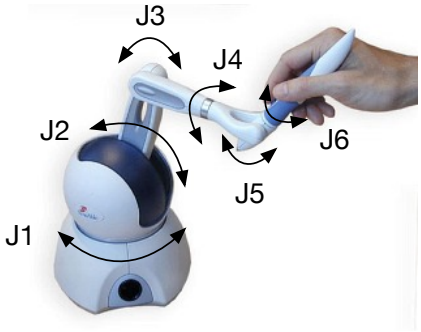


Figure 3.3: The PHANToM Omni haptic device.

of the device. Since the first 3 joints are actuated, it can be used as a low-cost 3 DOF manipulator in the research of robotics control.

To set up a simple BTOS, two PHANToM Omni haptic robots are used. Each robot is interfaced to a desktop powered with Intel Core2Quad processor, 16GB RAM and operating system Windows 7, using FireWire 400 (IEEE 1394) cable. The program to control each robot was written in C++ using the OpenHaptics API (v 3.1), which is developed by SensAble Technologies for PHANToM devices. For nominal BTOS operations, the considered PD-like control with empirically fine tuned gains is implemented. The complete experiment setup is shown in Fig. 4.12. For networking, Windows socket API is utilized and packets are transmitted in UDP format through the Ethernet. The payload of each packet consists of the robot's joint angles and rate of change of respective joint angles. There is no packet fragmentation required.

The application layer of the program consists of three supplementary threads along with the primary - main thread, which initializes the network sockets, and

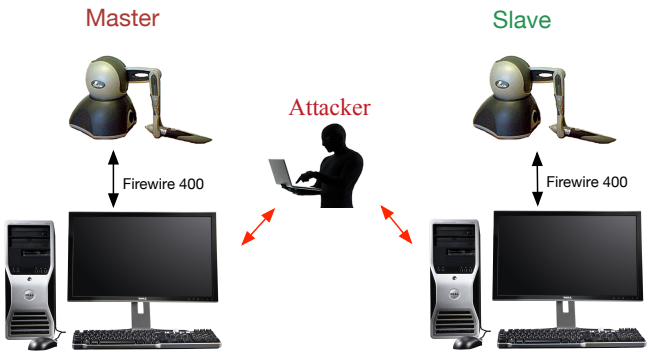


Figure 3.4: The BTOS experiment setup.

starts and stops the supplementary threads. The first thread is responsible for interacting with the robot, synchronizing with encoders and joint motors, computing the rate of change of joint angles and the control action for each motor. The second thread is responsible for handling the inbound packets, reading the UDP sockets, parsing the packet payload, and extracting the values of joint angles and their rate of change. The third thread is responsible for handling the outbound packets, creating the packet payload, and queuing it at the UDP socket for transmission. The transport layer and layers beneath it are implemented by the Windows kernel.

To emulate the 'man-in-the-middle' attacker, equivalent changes to the received packet payload are made in the reception thread of the robots' application programs (refer to Fig. 3.5). The 'man-in-the-middle' can also be easily emulated by a third computer relaying between the two robots.

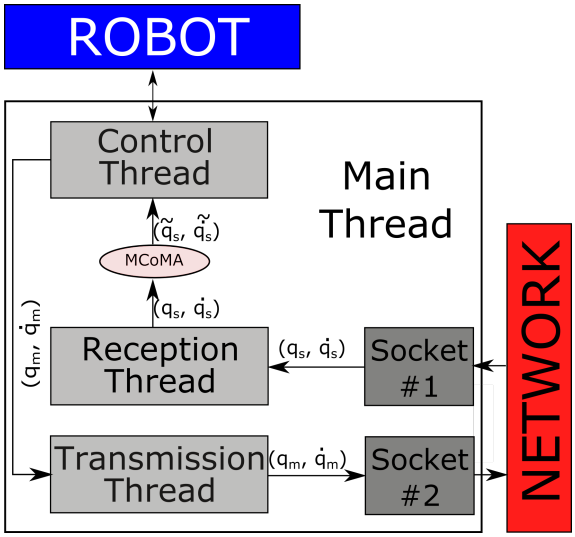


Figure 3.5: The application architecture of the master robot for the MCoMA experiment. The same architecture is used for the slave robot's application program.

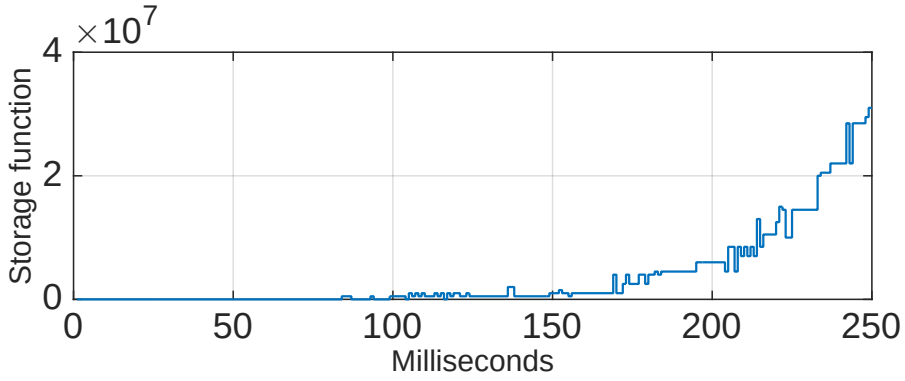


Figure 3.6: The storage function V under static MCoMA.

3.3.1 MCoMA Design

Assuming no detection scheme in BTOS, a simple static MCoMA is designed using **Proposition** 3.1.4 to attack the system. The following control gains are selected:

$$K_p = 600, \quad K_d = 5 \text{ and } K_{dm} = 5$$

The 'attack gain' $K \in \mathbb{R}^{12\times 12}$ is designed as

$$K = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ K_{21} & \mathbf{0}_3 & K_{23} + \Lambda_4 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ K_{41} + \Lambda_2 & \mathbf{0}_3 & K_{43} & \mathbf{0}_3 \end{bmatrix}.$$

From (3.16), $K_{41} = K_{23} = diag([s_1, s_1, s_1])$, $K_{21} = K_{43} = diag([s_2, s_2, s_2])$, with $s_1 = (K_d + K_{dm})/K_p, s_2 = -K_d/K_p$. Λ_4 and Λ_2 are chosen as $diag([1.5, 1.5, 1.5])$.

In order to protect the system from real damage, an inherent maximum joint speed limit is added on the robots. Additionally, the system storage function value is only observed and recorded during a short time interval after launching the attack.

Under this attack, the storage function V of BTOS is shown in Fig. 3.6. The storage function is chattering while it is increasing, which is due to the inherent physical limit of the robot joint space.

3.3.2 Attack Detection on General Content Modification Attack

The proposed detection scheme in Section 3.2.2 is tested under general content modification attacks. Here the sampling period $T_i = 4ms$. In the detection scheme,

for $i = m, s$, the initial values of $({}^i\lambda_{ia}, {}^i\lambda_{ib})$ are set to be $({}^i\lambda_a^0, {}^i\lambda_b^0) = (5, -2)$ and the update laws f_a, f_b in (2.4) and (2.5) are designed as

$$f_a = 20\cos(10k_iT_i),\ f_b = -20\sin(10k_iT_i),\ \ k_i = 1, 2, 3, \ldots$$

It can be verified that ${}^i\lambda_{ia}^{k_i} \neq {}^i\lambda_{ib}^{k_i}, k_i = 1, 2, 3, \ldots$.

In the **Case 1**, the proposed detection scheme is tested without additional artificial network delay. Then in the **Case 2**, in order to verify that the detection scheme is independent of network delay as discussed in ***Remark*** *2.3.1*, a $400ms$ one-way network delay is added artificially using a software called clumsy[1].

Case 1:

In order to determine the threshold ϵ_i in (2.16), we can run the BTOS with the detection scheme implemented under a normal operation condition (no attack) for a certain period of time, save all the position and velocity data, then compute the checking error e_i in (2.16) for all the saved data points, and set the value of ϵ_i larger than all the computed e_i. In this experiment, to determine ϵ_m, we use the master robot to teleoperate the slave robot under a normal operation condition for a certain period of time, and the computed checking error on the slave side e_m is shown in Fig. 3.7. Based on this plot, we select the threshold as $\epsilon_m = 0.0015$. It is also worth noting that the teleoperation performance remains unaffected during the operation in the absence of any content modification attack.

In the experiment, the attacker launches a general content modification attack on $(r_{ma}^{k_m}, r_{mb}^{k_m})$ from $k_m = 3000$. As (2.9), the injection data for $(r_{ma}^{k_m}, r_{mb}^{k_m})$ is designed

[1]https://jagt.github.io/clumsy/index.html

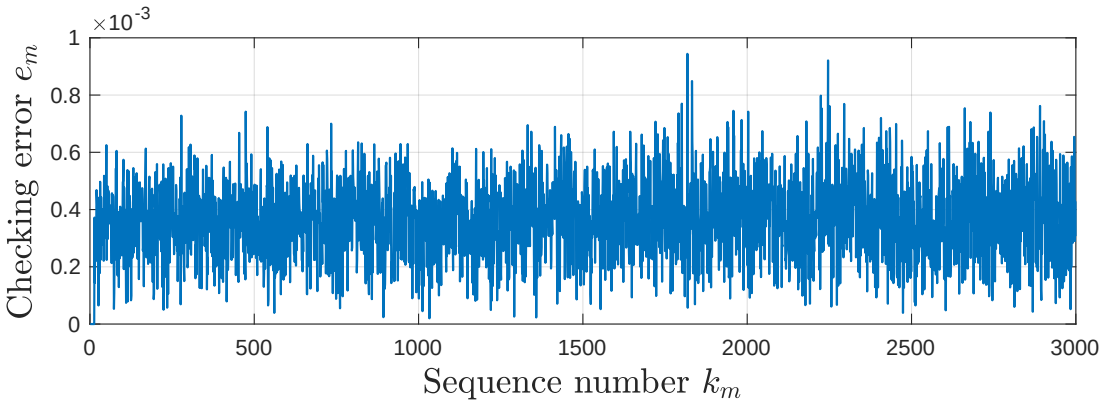


Figure 3.7: The checking error e_m on the slave side under a normal operation condition in **Case 1**.

as $(\hat{r}_{ma}^{k_m}, \hat{r}_{mb}^{k_m}) = (-0.01\sin(0.2t), 0.01\cos(0.2t))$. The plot of the checking error e_m on the slave side is shown in Fig. 3.8. The attack is successfully detected once the attack is launched at $k_m = 3000$ since the checking error e_m instantly exceeds the threshold $\epsilon_m = 0.0015$ (red line in Fig. 3.8). In this experiment, when checking condition in (2.16) is violated, the robots just drop the received data and stay still to protect the system.

Based on (2.17), checking error rate e_{md} can be computed, and the its plot is shown in Fig. 3.9. We can find there is sharp spike in the trajectory when the attack appears, which can be used as an additional indication on the attack.

Case 2:

In this case, a $400ms$ one-way delay is added in the network. The same approach in **Case 1** is adopted to determine the threshold ϵ_i. The computed error e_m is shown in Fig. 3.10, and ϵ_m is also set as 0.0015.

The attacker also launched the same general content modification attack on

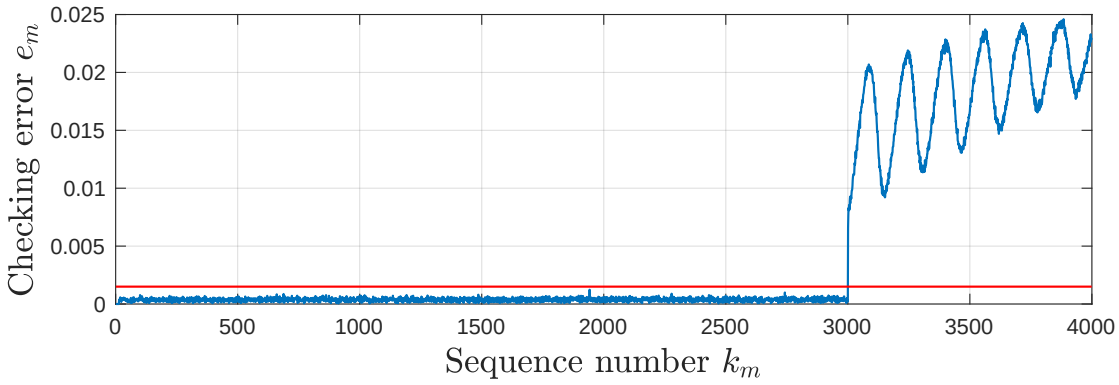


Figure 3.8: The checking error e_m (blue line) on the slave side under a general content modification attack in **Case 1**. The red line is the determined threshold $\epsilon = 0.0015$.

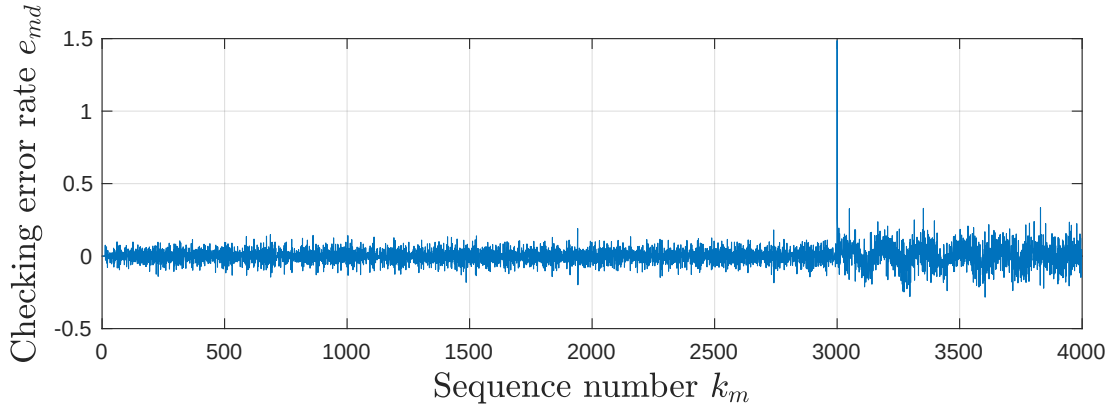


Figure 3.9: The checking error rate e_{md} on the slave side under a general content modification attack in **Case 1**.

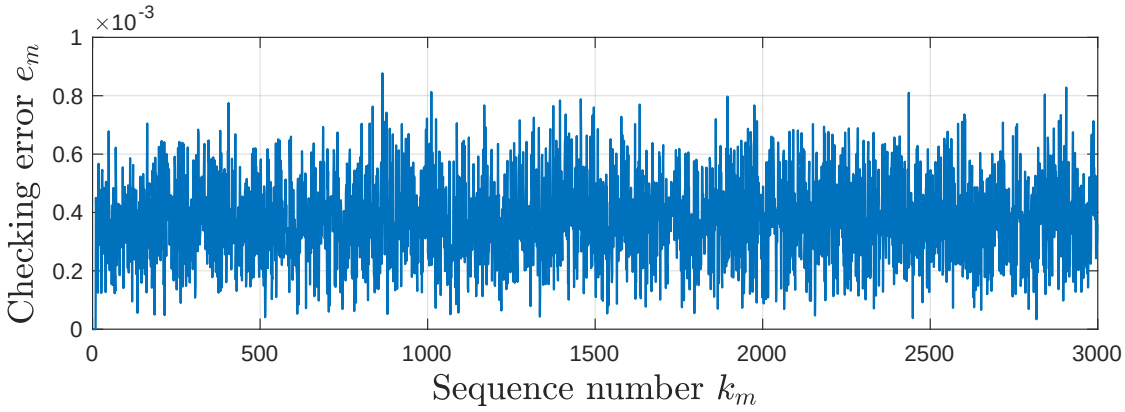


Figure 3.10: The checking error e_m on the slave side under a normal operation condition in **Case 2**.

$(r_{ma}^{k_m}, r_{mb}^{k_m})$ from $k_m = 3000$ as **Case 1**. The plot of the checking error e_m on the slave side is shown in Fig. 3.11. The attack is successfully detected after the attack is launched at $k_m = 3000$, as the checking error e_m exceeds the threshold $\epsilon_m = 0.0015$ (red line in Fig. 3.11).

Based on (2.17), checking error rate e_{md} can also be computed for this case, and the its plot is shown in Fig. 3.12. We can find there is sharp spike in the trajectory when the attack appears, which can be used as an additional indication on the attack.

3.4 Summary

In this chapter, content modification attacks are studied for a nonlinear bilateral teleoperation system where an attacker can change the data contents being communicated between the master and slave robot. The main motivation for the

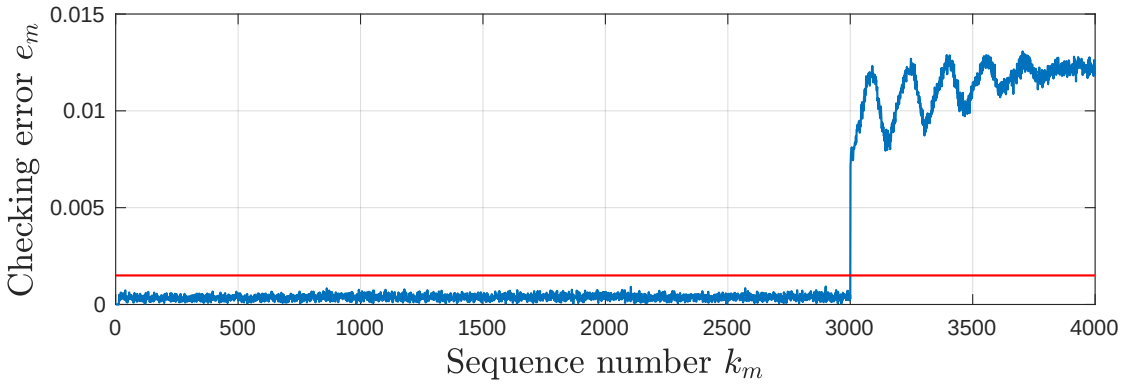


Figure 3.11: The checking error e_m (blue line) on the slave side under a general content modification attack with an artificial network delay in **Case 2**. The red line is the determined threshold $\epsilon = 0.0015$.

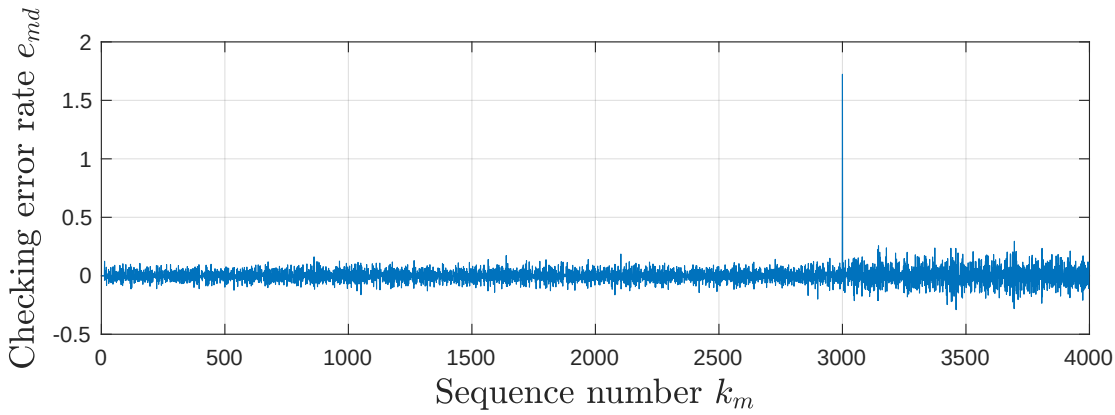


Figure 3.12: The checking error rate e_{md} on the slave side under a general content modification attack in **Case 2**.

work is to utilize the underlying physics of robotic systems to better understand the attack impact and provide a different perspective on the system security from existing cryptographic tools. Consequently, the static MCoMA is designed to demonstrate the severe attack impact, and the physics-based attack detection scheme with an encoding-decoding structure is adopted to detect general content modification attacks. The effectiveness of the attack design and the detection scheme is verified through experiments.

Chapter 4: Content Modification Attack on Multi-Robot System

In this chapter, a multi-robot system modeled as a MAS with double-integrator dynamics is considered. Specifically, we consider the synchronization problem for a second order MAS with undirected topology under the content modification attack, which can externally intercept the communication links and corrupt the data content.

The results of this chapter can be summarized as follows: Following a similar approach for BTOS, to demonstrate the severe attack impact on second order MAS synchronization, a static MCoMA is first designed. Since multiple links among MAS can be attacked, an optimal MCoMA that is distributed and compromises the least number of links is also designed. The physics-based attack detection condition (4.29) is applied for detecting static MCoMA. The physics-based attack detection scheme with an encoding-decoding structure proposed in Section 2.3 is also applied to protect against general content modification attacks. Additionally, a velocity observer-based attack mitigation scheme is also discussed for MCoMA. The efficacy of the proposed results is illustrated through numerical simulations.

The rest of the chapter is organized as follows: Section 4.1 provides a brief review on graph theory and existing synchronization algorithm. The content mod-

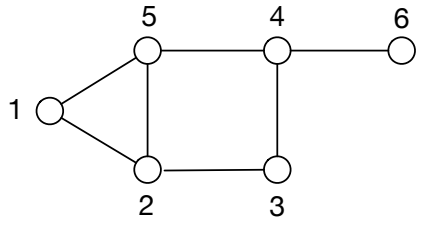


Figure 4.1: Multi-agent system modeled by an undirected graph with agents as the nodes and communication links as edges.

ification attack design is presented in Section 4.2. In Section 4.3, the attack detection/mitigation schemes are discussed for MAS. Finally, the simulation results are shown in Section 4.4.

4.1 Preliminary

4.1.1 Graph theory

The MAS can be modeled as a graph (Fig. 4.1). This section reviews some concepts and facts in the graph theory that are used in this chapter.

Consider a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ consists of a set of nodes $\mathcal{V} = \{1, 2, \ldots, n\}$ and a set of edges $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} | i, j \text{ adjacent}\}$. Nodes i, j are adjacent means there exists an edge (i, j) between two nodes.

The graph $\mathcal{G}$ is called undirected if $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$. The adjacency matrix is a square matrix $A \in \mathbb{R}^{n \times n}$ with element $a_{ij} = 1$ if i, j are adjacent and $a_{ij} = 0$ otherwise. The diagonal elements a_{ii} are zero since the self-loop case will not be considered. The degree matrix is a diagonal matrix $D \in \mathbb{R}^{n \times n}$ with element

d_i equals to the cardinality of the node i's neighbor set $\mathcal{N}_i = \{j \in \mathcal{V} | (i,j) \in \mathcal{E}\}$. The Laplacian matrix $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$, which means its elements are

$$l_{ii} = \sum_{j=1, j \neq i}^{n} a_{ij}, \; l_{ij} = -a_{ij}, \; i \neq j. \tag{4.1}$$

For an undirected graph $\mathcal{G}$, L is a symmetric and positive semidefinite. Observing the fact that the row sum of L is zero, the vector $1_n = [1, 1, \ldots, 1]^T \in \mathbb{R}^n$ is a right eigenvector of L associated with the eigenvalue $\lambda = 0$, i.e., $L1_n = 0$.

A path from node i to j is a sequence of distinct nodes from i to j, such that each pair of consecutive nodes are adjacent. If there is a path from i to j, then i, j are called connected. If all pairs of nodes in $\mathcal{G}$ are connected, then $\mathcal{G}$ is called connected. For connected graphs, L has exactly one zero eigenvalue. The eigenvalues of L can be listed in an increasing order as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$. The second smallest eigenvalue λ_2 is called algebraic connectivity of a graph, which is a measure of performance/speed of synchronization algorithm [18].

4.1.2 Existing synchronization protocol

This section reviews the existing graph-Laplacian based synchronization protocol for MAS with double-integrator dynamics presented in [33–35].

The MAS is modeled by a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with agents as the nodes and communication links as edges. The following assumptions hold throughout the chapter:

(A4.1) Graph $\mathcal{G}$ is undirected and connected.

(A4.2) The communication condition is perfect (network effects are neglected),

which can isolate the effect of cyber attack on MAS in later sections.

Dynamics of each agent in MAS is identical and is given as:

$$\begin{bmatrix} \dot{q}_i \\ \\ \ddot{q}_i \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \\ 0 & 0 \end{bmatrix} \begin{bmatrix} q_i \\ \\ \dot{q}_i \end{bmatrix} + \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix} u_i, \quad i \in \mathcal{V} = \{1, 2, \ldots, n\}, \tag{4.2}$$

where $q_i \in \mathbb{R}, i \in \mathcal{V}$ and $\dot{q}_i \in \mathbb{R}, i \in \mathcal{V}$ are generalized coordinates and their rate of change, respectively. Henceforth, for simplicity we shall just refer to q_i as position and $\dot{q}_i$ as velocity of any agent $i \in \mathcal{V}$. The state $q_i, \dot{q}_i$ are treated as scalar here, but the results can be extended to higher dimensional case using the concepts of Kronecker product.

Consider the distributed synchronization protocol proposed in [34], where

$$u_i = -\sum_{j \in \mathcal{N}_i} \left[(q_i - q_j) + \beta(\dot{q}_i - \dot{q}_j) \right], \quad i \in \mathcal{V} \tag{4.3}$$

where $\beta \in \mathbb{R}^+$ is referred to as coupling gain of relative velocities.

Definition 4.1.1. Consensus is achieved by the MAS described by (4.2) - (4.3), if for any $(i, j) \in \mathcal{V} \times \mathcal{V}$

$$\lim_{t \to \infty} (q_i(t) - q_j(t)) = 0 \quad \text{and} \quad \lim_{t \to \infty} (\dot{q}_i(t) - \dot{q}_j(t)) = 0 \tag{4.4}$$

Let $q = [q_1, q_2, \ldots, q_n]^T$, $\dot{q} = [\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n]^T$ and $x = \left[q^T, \dot{q}^T\right]^T$. By applying synchronization protocol (4.3), (4.2) can be written in compact matrix form as

$$\dot{x} = \Gamma x, \; \Gamma = \begin{bmatrix} \mathbf{0}_n & I_n \\ \\ -L & -\beta L \end{bmatrix} \tag{4.5}$$

Following from these notational substitutions, (4.4) is equivalent to $x(t) \to \mathbf{E}$ as $t \to \infty$, where

$$\mathbf{E} := \{x \in \mathbb{R}^{2n} | q \in \text{span}(1_n),\ \dot{q} \in \text{span}(1_n)\} \tag{4.6}$$

The convergence analysis of synchronization protocol (4.3) is done by Lyapunov function analysis similar to Theorem 1 of [35].

Theorem 4.1.1. The synchronization is achieved by the MAS (4.2)-(4.3) for any $\beta \in \mathbb{R}^+$.

Proof:

From **(A4.1)**, both L and L^2 are positive semidefinite with exactly one zero valued eigenvalue and

$$\mathcal{N}(L) = \mathcal{N}(L^2) = \text{span}(1_n). \tag{4.7}$$

Let $q = \alpha 1_n + \delta_q$ and $\dot{q} = \gamma 1_n + \delta_{\dot{q}}$, where $\delta_q \in \text{span}\{1_n\}^\perp$, $\delta_{\dot{q}} \in \text{span}\{1_n\}^\perp$ are referred to as position and velocity disagreement vectors [68]. Real-valued scalars $\alpha = \frac{1}{n} q^T 1_n$, $\gamma = \frac{1}{n} \dot{q}^T 1_n$ are simply the components of q and $\dot{q}$ along 1_n, respectively.

Substituting $q = \alpha 1_n + \delta_q$ and $\dot{q} = \gamma 1_n + \delta_{\dot{q}}$ in (4.5) gives us the following dynamics of the disagreement vectors,

$$\dot{\delta} = \Gamma \delta, \quad \Gamma = \begin{bmatrix} \mathbf{0}_n & I_n \\ -L & -\beta L \end{bmatrix} \tag{4.8}$$

where $\delta = \left[\delta_q^T,\ \delta_{\dot{q}}^T\right]^T \in D_\delta$ and $\dot{\alpha} = \gamma$, $\dot{\gamma} = 0$. $D_\delta = \text{span}\left\{\left[1_n^T,\ 0_n^T\right]^T,\ \left[0_n^T,\ 1_n^T\right]^T\right\}^\perp$ is a vector space of dimension $2n - 2$. From (4.7), (4.8) has a single equilibrium point at the origin in D_δ.

Consider the following storage function of (4.5),

$$V = \frac{1}{2}x^T P x = \frac{1}{2}x^T \begin{bmatrix} L^2 & \mathbf{0}_n \\ \mathbf{0}_n & L \end{bmatrix} x = \frac{1}{2}q^T L^2 q + \frac{1}{2}\dot{q}^T L\dot{q} \tag{4.9}$$

Since $L \succeq 0$, we have $L^2 \succeq 0$, which implies $P \succeq 0$. Thus $V \geq 0, \forall x \in \mathbb{R}^{2n}$.

From (4.9) and (4.5), we get

$$\dot{V} = \frac{1}{2}x^T \left(\Gamma^T P + P\Gamma\right) x = -\beta \dot{q}^T L^2 \dot{q} \leq 0, \tag{4.10}$$

Substituting $q = \alpha 1_n + \delta_q$ and $\dot{q} = \gamma 1_n + \delta_{\dot{q}}$ in (4.9) gives,

$$V = \frac{1}{2}\delta_q^T L^2 \delta_q + \frac{1}{2}\delta_{\dot{q}}^T L \delta_{\dot{q}} \tag{4.11}$$

From (4.7), V defined in (4.11) is positive definite and radially unbounded in D_δ. Thus, V is a Lyapunov candidate function for the system (4.8).

Similarly, substituting $\dot{q} = \gamma 1_n + \delta_{\dot{q}}$ in (4.10) gives

$$\dot{V} = -\beta \delta_{\dot{q}}^T L^2 \delta_{\dot{q}} \leq 0 \tag{4.12}$$

Since V in (4.11) is radially unbounded and $\dot{V} \leq 0$, $\forall \delta \in D_\delta$, the set $\Omega_c = \{\delta \in D_\delta \,|\, V \leq c,\, c > 0\}$ is a compact, positively invariant set. Let

$$\mathbf{S}_\delta = \{\delta \in D_\delta \,|\, \dot{V} = 0\} \tag{4.13}$$

From (4.7), $\mathbf{S}_\delta = \{\delta \in D_\delta \,|\, \delta_{\dot{q}} = 0_n\}$. It is easy to verify that $\mathbf{E}_\delta = \{\delta \in D_\delta \,|\, \delta_q = 0_n,\, \delta_{\dot{q}} = 0_n\}$ is the largest invariant set in $\mathbf{S}_\delta$. Hence, from LaSalle's invariance principle [66], the origin of (4.8) is globally asymptotically stable.

This implies, $x(t) \to \mathbf{E}$ as $t \to \infty$.

■

4.2 Attack Design for MAS

In this section, the content modification attack design is discussed in the context of MAS. In subsection 4.2.1, to better understand the vulnerability of MAS and severity of attack issue, a static MCoMA is designed from the attacker's perspective, where the attacker's goal is to ensure unbounded growth of the storage function (4.9). In Section 4.2.2, an optimal MCoMA (OMCoMA) is further discussed based on the static MCoMA.

Based on **Definition** 2.1.1, the attacker can launch a content modification attack on MAS by compromising an edge set $\bar{\mathcal{E}} \subseteq \mathcal{E}$ and therefore modify the information being exchanged in the edge set $\bar{\mathcal{E}}$ at will.

Given the assumptions made in Section 2.1, in the context of MAS, we assume

(A3.3) The attacker is able to completely hack the communication links represented by the edges in the set $\bar{\mathcal{E}}$. Here, completely hack refers to the attacker's capability of breaking existing cryptographic tools like encryption or message authentication code (MAC).

(A3.4) The attack is launched at $t = t_a \geq 0$, and MAS state x does not lie in the synchronization set $\mathbf{E}$ defined in (4.6) at $t = t_a$, i.e., $x(t_a) \notin \mathbf{E}$.

(A3.5) The attacker knows the complete graph topology and synchronization protocol (4.3).

4.2.1 MCoMA Design

In this subsection, based on **Definition** 2.2.1, the MCoMA with respect to storage function V in (4.9) is studied.

Associate a basis vector e_i with every node $i \in \mathcal{V}$, where e_i is the i-th column of I_n. Let $\bar{A}$ and $\bar{D}$ be the adjacency matrix and degree matrix of subgraph $\bar{G} := \{\mathcal{V}, \bar{\mathcal{E}}\}$, then the information of node i is available to the attacker if and only if $||\bar{A}e_i||_0 = 1_n^T \bar{A} e_i \neq 0$.

As $\bar{A} = \bar{A}^T$ (from **(A3.3)**), thus the total number of completely hacked edges associated with node i is given as $1_n^T \bar{A} e_i$. Define $\mathcal{V}_a = \{i \in \mathcal{V} | 1_n^T \bar{A} e_i \neq 0\} \subseteq V$ as the set of compromised nodes, hence the attack space which is defined as $\mathcal{S}_a^n = \text{span}\{e_i,\, i \in \{1, \ldots, n\} |\, i \in \mathcal{V}_a\}$ has dimension $|\mathcal{V}_a|$. For simplifying the attack analysis, let us define an attack operator

$$O_a = \bar{D} = diag(\bar{A} 1_n) \tag{4.14}$$

which basically maps $\mathbb{R}^n$ to $\mathcal{S}_a^n$ linearly.

For each node $i \in \mathcal{V}_a$ and edge $(i,\, j) \in \bar{\mathcal{E}}$, the attacker can modify the information $(q_i,\, \dot{q}_i)$ of node i to $(\tilde{q}_i,\, \tilde{\dot{q}}_i)$ being received by node j as

$$\begin{aligned} \tilde{q}_i &= q_i + \bar{q}_{ij},\, \bar{q}_{ij} = K_{ij}^p \left[(O_a q)^T,\, (O_a \dot{q})^T \right]^T \\ \tilde{\dot{q}}_i &= \dot{q}_i + \bar{\dot{q}}_{ij},\, \bar{\dot{q}}_{ij} = K_{ij}^v \left[(O_a q)^T,\, (O_a \dot{q})^T \right]^T \end{aligned} \tag{4.15}$$

where $K_{ij}^p \in \mathbb{R}^{1 \times 2n}$ and $K_{ij}^v \in \mathbb{R}^{1 \times 2n}$ are termed as the position attack gain and velocity attack gain associated with edge $(i,\, j) \in \bar{\mathcal{E}}$, respectively.

Hence, the effective control input of any node $j \in \mathcal{V}_a$ (refer to (4.3)) is given

as

$$u_j = -\sum_{i\in\mathcal{N}_j}[(q_j - q_i) + \beta(\dot{q}_j - \dot{q}_i)] + \sum_{i\in\bar{\mathcal{N}}_j}[\bar{q}_{ij} + \beta\bar{\dot{q}}_{ij}]$$

$$u_j = [L_j^T,\, \beta L_j^T]x + \bar{A}_j^T(K_j^p + \beta K_j^v)(I_2 \otimes O_a)x \tag{4.16}$$

where $\bar{\mathcal{N}}_j = \{i \in \mathcal{V} | (i,j) \in \bar{\mathcal{E}}\}$ and $x = [q^T,\, \dot{q}^T]^T$.

For uniformity, we associate position attack gains and velocity attack gains with every node $j \in \mathcal{V}$ (not necessarily in $\mathcal{V}_a$) as

$$K_j^p = [K_j^{pp} |\, K_j^{pv}] = \begin{bmatrix} K_{1j}^p \\ \vdots \\ K_{nj}^p \end{bmatrix}, \quad K_j^v = [K_j^{vp} |\, K_j^{vv}] = \begin{bmatrix} K_{1j}^v \\ \vdots \\ K_{nj}^v \end{bmatrix} \tag{4.17}$$

where, $K_j^p, K_j^v \in \mathbb{R}^{n\times 2n}$ and $K_j^{pp}, K_j^{pv}, K_j^{vp}, K_j^{vv} \in \mathbb{R}^{n\times n}$ for all $j \in \mathcal{V}$. Ideally, if $(i,\, j) \notin \bar{\mathcal{E}}$ then $K_{ij}^p = K_{ij}^v = 0_{2n}^T$. But, even in case these gains are not identically zero, the adjacency vector $\bar{A}_j$ associated with $j \in \mathcal{V}$ and the attack operator O_a will limit the effect of the attack only to the compromised nodes (nodes that belong to $\mathcal{V}_a$).

Substituting (4.16) in MAS dynamics (4.5) gives

$$\dot{x} = \Gamma x + \Delta x,\ \Delta = \begin{bmatrix} \mathbf{0}_n & \mathbf{0}_n \\ \Delta_p & \Delta_v \end{bmatrix} \tag{4.18}$$

where,

$$\Delta_p = \begin{bmatrix} \bar{A}_1\,(K_1^{pp} + \beta K_1^{vp}) \\ \vdots \\ \bar{A}_n\,(K_n^{pp} + \beta K_n^{vp}) \end{bmatrix} O_a, \quad \Delta_v = \begin{bmatrix} \bar{A}_1\,(K_1^{pv} + \beta K_1^{vv}) \\ \vdots \\ \bar{A}_n\,(K_n^{pv} + \beta K_n^{vv}) \end{bmatrix} O_a \tag{4.19}$$

are referred to as attack gains.

Next, we give sufficient and necessary conditions for the attack (4.15) to be MCoMA.

Theorem 4.2.1. Consider the dynamics of MAS (4.5) under content modification attack (4.15), i.e. (4.18). The content modification attack (4.15) is a MCoMA with respect to storage function V in (4.9) if and only if

(a) attack gains Δ_p and Δ_v in (4.18) satisfies $\Delta_p = \mathbf{0}_n$ and $R := -2\beta L^2 + L\Delta_v + \Delta_v^T L \succeq 0$,

(b) set $\mathbf{F}\backslash\mathbf{E}$ is not positively invariant, where $\mathbf{F} := \{x \in \mathbb{R}^{2n} | \dot{q}^T R \dot{q} = 0\}$.

Proof:

Consider the storage function $V = \frac{1}{2}x^T P x$ in (4.9). Clearly, $V = 0,\ \forall x \in \mathbf{E}$ and $V > 0,\ \forall x \in \mathbb{R}^{2n}\backslash\mathbf{E}$.

Computing the time derivative of V along the trajectory of (4.18), we have

$$\dot{V} = \frac{1}{2}x^T(\Gamma^T P + P\Gamma + \Delta^T P + P\Delta)x = \frac{1}{2}x^T \begin{bmatrix} \mathbf{0}_n & \Delta_p^T L \\ L\Delta_p & R \end{bmatrix} x := \frac{1}{2}x^T M x \quad (4.20)$$

Sufficiency:

From **(A3.4)**, $x(t_a) \notin \mathbf{E}$, thus $V(t_a)$ is positive. Now

If **(a)** $\Delta_p = \mathbf{0}_n$ and $R \succeq 0$, then $\dot{V} \geq 0$. Thus, $V > 0$ and $x \notin \mathbf{E}, \forall t \geq t_a$.

For condition **(b)**, we need first show $\mathbf{E} \subset \mathbf{F}$. For $x \in \mathbf{E}$, we have $\dot{q} \in \text{span}(1_n)$. We can easily verify that $\dot{q}^T R \dot{q} = 0$ when $\dot{q} \in \text{span}(1_n)$. Thus, $x \in \mathbf{F}$, which implies $\mathbf{E} \subset \mathbf{F}$.

If **(b)** the set $\mathbf{F}\backslash\mathbf{E}$ is not positively invariant, $\dot{V}$ will not always stay in 0, thus

$V \to \infty$ as $t \to \infty$.

Hence, by **Definition** 2.2.1, the content modification attack is MCoMA with respect to storage function V in (4.9).

Necessity:

By **Definition** 2.2.1, if the content modification attack is MCoMA, then $\dot{V} \geq 0$ in (4.20), which implies $M \succeq 0$. By generalized Schur's complement condition [67], for $M \succeq 0$, we should have $R \succeq 0$ and $-\Delta_p^T L R^- L \Delta_p \succeq 0$. This can only be satisfied when $\Delta_p = \mathbf{0}_n$. Further, $V \to \infty$ as $t \to \infty$, implies the set $\mathbf{F} \backslash \mathbf{E}$ should not be positively invariant. Hence, both **(a)** and **(b)** are necessary.

■

Since the attack gains (4.19) are constant matrices, the proposed MCoMA is also referred to as static MCoMA.

Remark 4.2.1. Amongst all possible choices for the attacker to modify the data content in the compromised link, the proposed MCoMA guarantees nonnegative increase rate of the storage function (4.9) and ensures that $\lim_{t\to\infty} V = \infty$. This could be catastrophic for the individual agent in the MAS and in some cases, can even cause physical damage to the agents.

Remark 4.2.2. Given O_a and $\bar{A}$, the attacker can compute the position attack gains and velocity attack gains by choosing an appropriate $Q = Q^T \succ 0$ such that the following feasibility problem has a nonempty solution set. This problem is a linear programming problem as the constraints are all linear matrix inequalities (LMI),

which can be solved efficiently.

$$\begin{aligned} &\text{minimize } 1 \\ &\text{subject to:} \\ &\quad \textbf{(C1)}\ \Delta_p = \mathbf{0}_n \\ &\quad \textbf{(C2)}\ L\Delta_v + \Delta_v^T L = Q + 2\beta L^2 \end{aligned} \tag{4.21}$$

If there exists $Q = Q^T \succ 0$ such that the solution set of (4.21) is nonempty then clearly $R = Q \succ 0$ and condition **(a)** of **Theorem** 4.2.1 holds. $R \succ 0$ implies $\mathbf{F}\backslash\mathbf{E} = \{x \in \mathbb{R}^{2n} | \, q \notin \text{span}(1_n), \, \dot{q} = 0\}$ and from (4.18) we get, $\ddot{q} = -Lq \neq 0$ in $\mathbf{F}\backslash\mathbf{E}$ which makes $\mathbf{F}\backslash\mathbf{E}$ a non-invariant set. Hence, the content modification attack with position attack gains and velocity attack gains belonging to the solution set of (4.21) is MCoMA.

4.2.2 Optimal MCoMA Design

In order to maximize the damage and minimize the cost of attack at the same time, it is desirable for MCoMA to be distributed and compromise least number of edges possible in any given graph.

A distributed MCoMA means for any $(i, j) \in \bar{\mathcal{E}}$, only the i-th and j-th elements of the associated K_{ij}^p and K_{ij}^v are non-zero. In words, this means that a content modification of states being exchanged on any completely hacked link is not dependent in any way on the content modification of states being exchanged on any other completely hacked links. Due to this reason distributed MCoMA is much easier to execute and has higher scalability and lower communication cost.

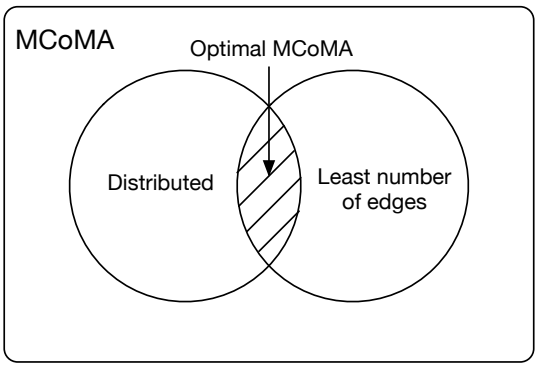


Figure 4.2: The relation between OMCoMA and MCoMA.

The relation between OMCoMA and MCoMA can be illustrated by Fig. 4.2.

Definition 4.2.1. The optimal MCoMA or OMCoMA is referred to as a MCoMA which is distributed and compromises least number of edges for any given graph.

Next proposition gives a distributed MCoMA design:

Proposition 4.2.3. Consider the dynamics of MAS (4.5) under content modification attack (4.15), i.e. (4.18). If

$$\Delta_p = \mathbf{0}_n, \quad \Delta_v = \Delta_v^T = \theta I_n \tag{4.22}$$

where θ satisfies $\theta \geq \beta\lambda_n$, where λ_n is the maximum eigenvalue of Laplacian matrix L, then the content modification attack (4.15) is a MCoMA with respect to storage function V in (4.9).

Before the proof is provided, a lemma which will be used in the proof is stated as follows:

Lemma 4.2.1. (Theorem 3 in [69]) Let A and B be two real positive semidefinite matrices, then $AB \succeq 0$ if and only if AB is normal.

Now the proof of **Proposition** 4.2.3 is given:

Proof:

Choosing $\Delta_v^T = \Delta_v = \theta I_n$, we have

$$R = -2\beta L^2 + 2L\Delta_v = 2L(\Delta_v - \beta L) = 2L(\theta I_n - \beta L). \tag{4.23}$$

Since $\theta \geq \beta\lambda_n$, we have $\theta I_n - \beta L \succeq 0$. Now from **Lemma** 4.2.1, $R = 2L(\theta I_n - \beta L) \succeq 0$ since R is normal and thus condition (a) in **Theorem** 4.2.1 is satisfied.

As $L = T\Lambda T^{-1}$ by a similarity transformation and from (4.7), we get

$$R = 2L(\theta I_n - \beta L) = T(2\theta\Lambda - 2\beta\Lambda^2)T^{-1}, \tag{4.24}$$

where $2\theta\Lambda - 2\beta\Lambda^2$ also has exactly one zero eigenvalue, which means $dim(\mathcal{N}(R)) = 1$ and $\mathcal{N}(R) = \text{span}(1_n)$.

This implies

$$\mathbf{F} = \{x \in \mathbb{R}^{2n} | \dot{q}^T R \dot{q} = 0\} = \{x \in \mathbb{R}^{2n} | \dot{q} \in \text{span}(1_n)\}. \tag{4.25}$$

Hence, $\mathbf{F} \backslash \mathbf{E} = \{x \in \mathbb{R}^{2n} | \dot{q} \in \text{span}(1_n), q \notin \text{span}(1_n)\}$.

With Δ_p, Δ_v defined in (4.22), the velocity dynamics be written as

$$\ddot{q} = -Lq - \beta L\dot{q} + \theta I_n \dot{q}. \tag{4.26}$$

Multiplying by L on both sides of equation, we get

$$L\ddot{q} = -L^2 q - \beta L^2 \dot{q} + \theta L\dot{q}. \tag{4.27}$$

When the system state is in $\mathbf{F}\backslash\mathbf{E}$, $L\dot{q} = 0, L^2\dot{q} = 0$ and $L^2 q \neq 0$, thus $L\ddot{q} \neq 0$, which implies $\ddot{q} \notin \text{span}(1_n)$. Thus, $\mathbf{F}\backslash\mathbf{E}$ is not positively invariant set, which satisfies condition (b) in **Theorem** 4.2.1.

Hence, by **Theorem** 4.2.1, the content modification attack given in the proposition is MCoMA with respect to storage function V in (4.9).

■

Observing the MCoMA injection term $\theta I_n \dot{q}$ in (4.26), it is evident that the proposed static MCoMA in **Proposition** 4.2.3 is in fact distributed, i.e. there is no information exchange required between the attacks targeting different edges.

The result proposed in the following corollary of **Theorem** 4.2.1 is critical for the design of MCoMA with minimum $|\bar{\mathcal{E}}|$.

Corollary 4.2.1. For a graph $G = \{\mathcal{V}, \mathcal{E}\}$, suppose the attack space of MCoMA is denoted by $\mathcal{S}_a^n$ as defined in Section 4.2.1 and $dim(\mathcal{S}_a^n) = |\mathcal{V}_a|$, then we always have $|\mathcal{V}_a| = |\mathcal{V}|$.

Proof: Clearly, for any content modification attack, we have $|\mathcal{V}_a| \leq |\mathcal{V}|$.

(Proof by Contradiction.) Assume $|\mathcal{V}_a| < |\mathcal{V}|$, which means some of the nodes are not compromised by the MCoMA. This implies, there exists non-zero number of zero rows in matrix Δ_v, and symmetrically, there exist zero columns in Δ_v Consequently, corresponding diagonal element(s) of $L\Delta_v + \Delta_v^T L$ are zero as well. Hence, the matrix R as defined in (a) of **Theorem** 4.2.1 has negative diagonal element(s) and hence is not positive semidefinite. However, a necessary condition for a content modification attack to be MCoMA is $R \succeq 0$ (refer to **Theorem** 4.2.1). This

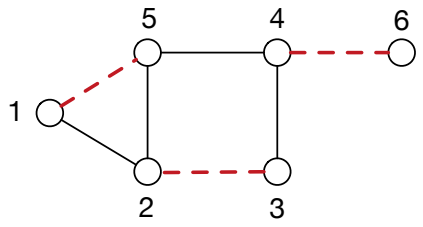


Figure 4.3: An instance of minimum edge cover for the graph in Fig. 4.1.

contradicts the given fact that the content modification is MCoMA.

Hence, by contradiction $|\mathcal{V}_a| < |\mathcal{V}|$ cannot hold, thus $|\mathcal{V}_a| = |\mathcal{V}|$.

■

The implications of **Corollary** 4.2.1 are

- None of the diagonal entries of the attack operator O_a is zero for MCoMA.
- The OMCoMA design problem is now reduced to the minimum edge cover problem (given in (4.28)) : finding the least number of compromised edges such that every node of the graph is incident to at least one edge in the set $\bar{\mathcal{E}}$,

$$\begin{aligned} &\underset{(i,j)\in\mathcal{E}}{\text{minimize}} \ |\bar{\mathcal{E}}| \\ &\text{subject to:} \ \bigcup_{(i,j)\in\bar{\mathcal{E}}} \{i, j\} = \mathcal{V} \end{aligned} \tag{4.28}$$

Note: The minimum edge cover problem (4.28) is also known as maximum matching problem. [70] introduces an algorithm in Section 10.5 to solve this problem for a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ in $O(|\mathcal{V}|^4)$ time.

To summarize, OMCoMA defined in **Definition** 4.2.1 can be designed as following:

- Firstly, solve the minimum edge cover problem (4.28) to obtain optimal $\bar{\mathcal{E}}$.
- Secondly, implement the distributed attack based on **Proposition** 4.2.3.

4.3 Attack Detection/Mitigation for MAS

In this section, the attack detection/mitigation scheme on static MCoMA for MAS is first studied, and then the attack detection scheme for general content modification attacks is discussed.

4.3.1 Attack Detection/Mitigation on Static MCoMA

In this subsection, the attack detection and mitigation scheme for MAS synchronization against the static MCoMA in **Theorem** 4.2.1 is proposed.

4.3.1.1 Attack Detection

In this part, the detection condition in (2.3) can be applied for detecting static MCoMA for MAS. To further interpret this, let us consider one compromised edge $(i,j) \in \bar{\mathcal{E}}$ as Fig. 4.4. Suppose the attacker modifies $(q_i, \dot{q}_i)$ to $(\tilde{q}_i, \dot{\tilde{q}}_i)$, the data is only accepted by agent j when the following detection condition is satisfied:

$$\dot{\tilde{q}}_i(t) = \dot{\tilde{q}}_i(t), \quad \forall t \geq 0. \tag{4.29}$$

The following proposition demonstrates that the proposed static MCoMA can be detected with simple detection condition (4.29).

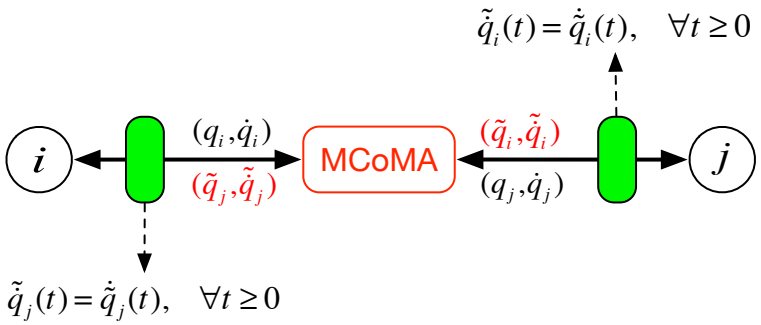


Figure 4.4: Attack detection scheme for static MCoMA on edge (i, j).

Proposition 4.3.1. Consider the dynamics of MAS (4.5) under content modification attack (4.15), i.e. (4.18). The static MCoMA in **Theorem** 4.2.1 can be detected if the detection condition (4.29) is checked for the received data at all the agents in MAS.

Proof:

For each node $i \in \mathcal{V}_a$ and edge $(i,\ j) \in \bar{\mathcal{E}}$, consider the attack (4.15),

$$\begin{aligned} \tilde{q}_i &= q_i + \bar{q}_i = q_i + K_{ij}^p (I_2 \otimes O_a) x, \\ \tilde{\dot{q}}_i &= \dot{q}_i + \bar{\dot{q}}_i = \dot{q}_i + K_{ij}^v (I_2 \otimes O_a) x, \end{aligned} \tag{4.30}$$

where $K_{ij}^p = [K_{ij}^{pp} | \, K_{ij}^{pv}]$ and $K_{ij}^v = [K_{ij}^{vp} | \, K_{ij}^{vv}]$ with $K_{ij}^{pp}, K_{ij}^{pv}, K_{ij}^{vp}, K_{ij}^{vv} \in \mathbb{R}^{1\times n}$.

Because for a MCoMA, Δ_p and Δ_v in (4.18) satisfies $\Delta_p = \mathbf{0}_n$ and $R = -2\beta L^2 + L\Delta_v + \Delta_v^T L \succeq 0$ as shown in **Theorem** 4.2.1, which implies $K_{ij}^p = [0_n^T | \, K_{ij}^{pv}]$ and $K_{ij}^v = [0_n^T | \, K_{ij}^{vv}]$.

(Proof by Contradiction.) Now assume the MCoMA avoids the detection

condition (4.29) imposed on (4.30), which means the attacker can make

$$K_{ij}^{v}(I_2 \otimes O_a)x = K_{ij}^{p}(I_2 \otimes O_a)\dot{x} \tag{4.31}$$

Substituting (4.18) into above equation gives

$$K_{ij}^{v}(I_2 \otimes O_a)x = K_{ij}^{p}(I_2 \otimes O_a)(\Gamma + \Delta)x \tag{4.32}$$

The left hand side of (4.32) is

$$K_{ij}^{v}(I_2 \otimes O_a)x = [0_n^T |\, K_{ij}^{vv}] \begin{bmatrix} O_a & \mathbf{0}_n \\ \mathbf{0}_a & O_a \end{bmatrix} x = [0_n^T |\, K_{ij}^{vv} O_a]x \tag{4.33}$$

The right hand side of (4.32) is

$$\begin{aligned} & K_{ij}^{p}(I_2 \otimes O_a)(\Gamma + \Delta)x \\ &= [0_n^T |\, K_{ij}^{pv}] \begin{bmatrix} O_a & \mathbf{0}_n \\ \mathbf{0}_n & O_a \end{bmatrix} \begin{bmatrix} \mathbf{0}_n & I_n \\ -L & -\beta L + \Delta_v \end{bmatrix} x \\ &= [0_n^T |\, K_{ij}^{pv}] \begin{bmatrix} \mathbf{0}_n & O_a \\ -O_a L & -\beta O_a L + O_a \Delta_v \end{bmatrix} x \\ &= [-K_{ij}^{pv} O_a L |\, K_{ij}^{pv}(-\beta O_a L + O_a \Delta_v)]x \end{aligned} \tag{4.34}$$

So (4.32) is equivalent to

$$[0_n^T |\, K_{ij}^{vv} O_a]x = [-K_{ij}^{pv} O_a L |\, K_{ij}^{pv}(-\beta O_a L + O_a \Delta_v)]x \tag{4.35}$$

To make the above equality hold, the attacker has to choose $K_{ij}^{pv} = K_{ij}^{vv} = 0_n^T$, which further implies $\Delta_v = \mathbf{0}_n$ and thereby $R = -2\beta L^2$ is negative semidefinite.

However, the MCoMA requires $R \succeq 0$. Hence, by contradiction, the static MCoMA in **Theorem** 4.2.1 can be detected by the detection condition (4.29), which completes the proof.

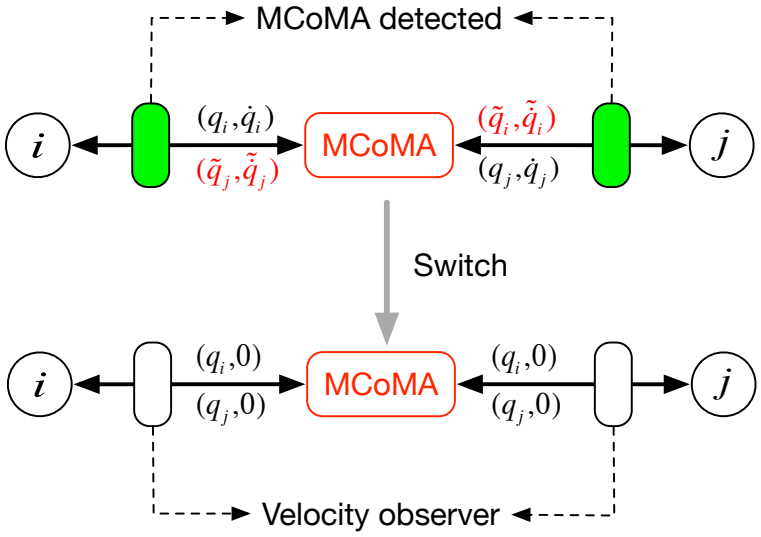


Figure 4.5: Attack mitigation scheme for static MCoMA on edge (i, j).

■

4.3.1.2 Attack Mitigation

In this part, a velocity observer from [34] is used to mitigate the proposed static MCoMA.

Proposition 4.3.2. Consider the dynamics of MAS (4.5) under content modification attack (4.15), i.e. (4.18). As shown in Fig. 4.5, once the proposed static MCoMA is detected, each agent $i \in \mathcal{V}$ transmits $(q_i, 0)$ instead of $(q_i, \dot{q}_i)$ and switches to the following local observer based synchronization algorithm [34] for MAS dy-

namics (4.2):

$$u_i = -\sum_{j\in\mathcal{N}_i}(q_i - q_j) - p\dot{\hat{z}}_i, \tag{4.36}$$

$$\dot{\hat{z}}_i = -\tau\hat{z}_i + \sum_{j\in\mathcal{N}_i}(q_i - q_j), \quad i \in \mathcal{V} \tag{4.37}$$

where $p, \tau \in \mathbb{R}^+$ and $\hat{z}_i \in \mathbb{R}$ is the observer state.

With above attack mitigation scheme, the synchronization of MAS defined by **Definition** 4.1.1 is still achieved under the static MCoMA.

Proof:

By (4.30) and (4.33), the injection signals are

$$\begin{aligned}\bar{q}_i &= K_{ij}^p(I_2 \otimes O_a)x = [0_n^T|\, K_{ij}^{pv}O_a]x = K_{ij}^{pv}O_a\dot{q},\\ \bar{\dot{q}}_i &= K_{ij}^v(I_2 \otimes O_a)x = [0_n^T|\, K_{ij}^{vv}O_a]x = K_{ij}^{vv}O_a\dot{q},\end{aligned} \tag{4.38}$$

which are in static feedback form of the velocities.

Since each agent $i \in \mathcal{V}$ transmits $(q_i, 0)$ instead of $(q_i, \dot{q}_i)$ once the MCoMA is detected, thus from (4.38), the injection signals are zero.

Thus, for each agent $i \in \mathcal{V}$, the received position information $q_j, j \in \mathcal{N}_i$ for synchronization algorithm (4.36) and (4.37) is not compromised.

For the convergence analysis for local observer based synchronization algorithm (4.36) and (4.37), the reader is referred to Theorem 4.1 in [34] and is omitted here.

■

4.3.2 Attack Detection for General Content Modification Attack

In the case that the attacker just launches a general content modification attack defined as (2.1) and (2.2) to disrupt the MAS synchronization, we can implement

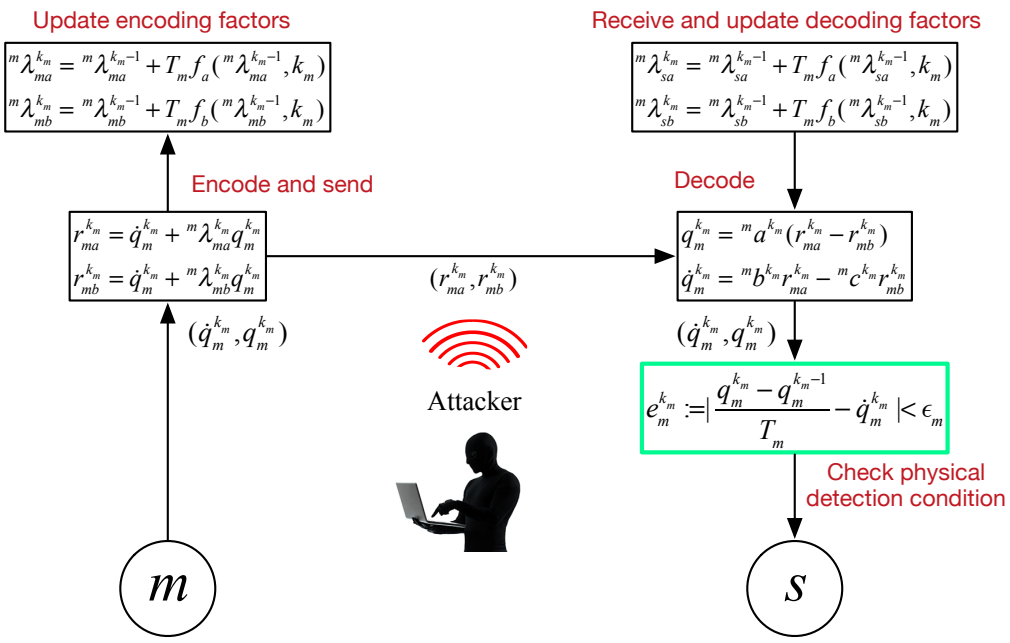


Figure 4.6: Attack detection scheme for general content modification attack on edge (m, s) for MAS.

the physics-based attack detection scheme with an encoding-decoding structure proposed in Section 2.3 for each pair of nodes (m, s) in every edge of MAS as shown in Fig. 4.6.

4.4 Simulation

Consider a MAS modeled by a graph in Fig. 4.1 satisfying assumption **(A4.1)** and **(A4.2)**. The Laplacian matrix L of this graph is

$$L = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

Firstly, suppose there is no attack and the synchronization algorithm (4.3) with $\beta = 2$ is implemented for each agent. The simulation results are shown in Fig. 4.7 and Fig. 4.8. We can find the synchronization is reached after 6s, and the storage function V in (4.9) and function rate $\dot{V}$ converge to zero rapidly. This verifies the efficacy of existing synchronization algorithm as analyzed in **Theorem** 4.1.1.

Secondly, suppose OMCoMA is designed as **Proposition** 4.2.3 with $\theta = \beta\lambda_n$ and launched as Fig. 4.3, wherein the edge (1,5), (2,3), (4,6) are compromised. The simulation results are shown in Fig. 4.9 and Fig. 4.10. It is evident that the position and velocity states diverge rapidly, the storage function V in (4.9) keeps increasing rapidly, and function rate $\dot{V}$ is nonnegative. This simply verifies the effect of proposed OMCoMA as in **Definition** 4.2.1.

Next, the attack detection scheme in **Proposition** 4.3.1 and mitigation scheme

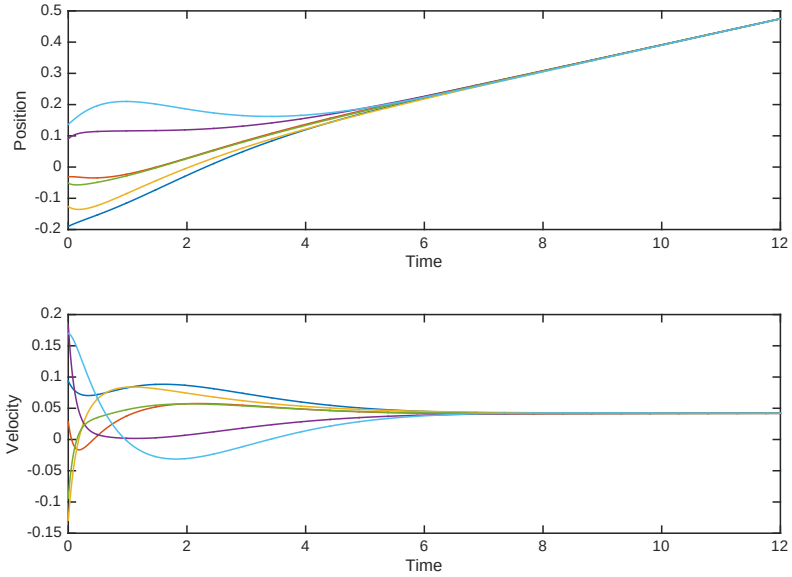


Figure 4.7: The position and velocity states of MAS under synchronization algorithm (4.3).

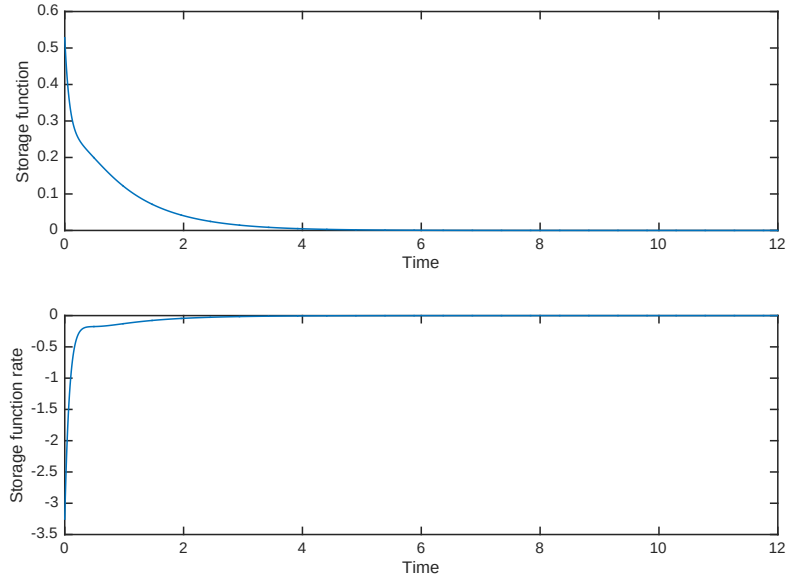


Figure 4.8: The storage function V and storage function rate $\dot{V}$ of MAS under synchronization algorithm (4.3).

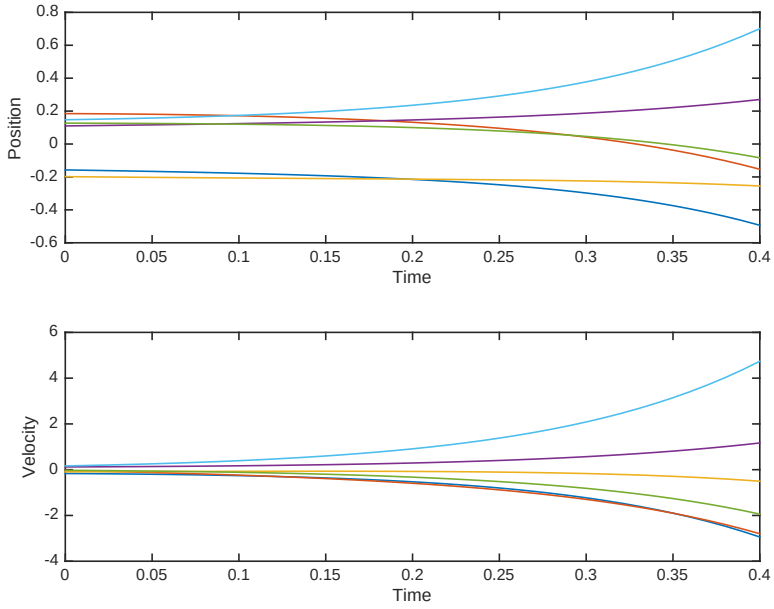


Figure 4.9: The position and velocity states of MAS under OMCoMA.

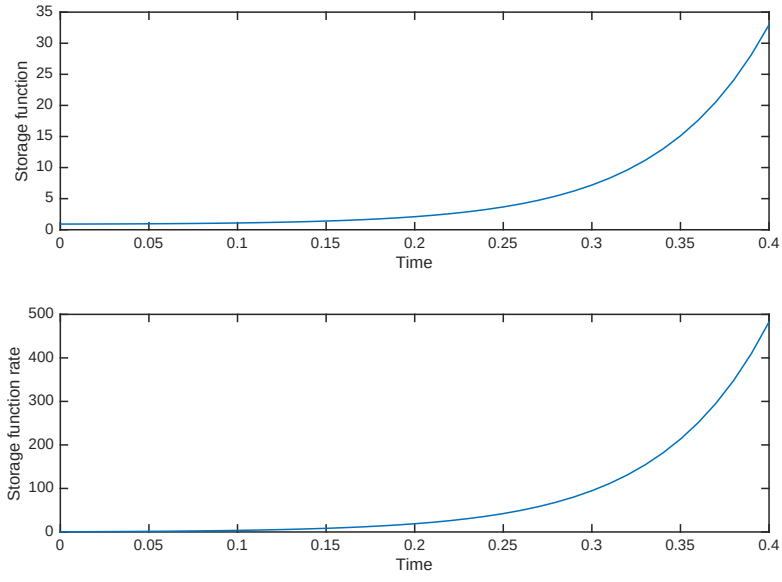


Figure 4.10: The storage function V and the storage function rate $\dot{V}$ of MAS under OMCoMA.

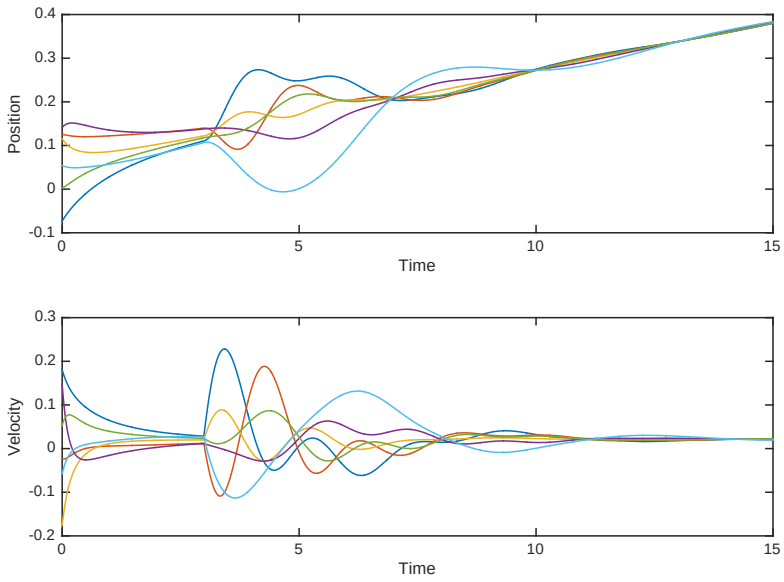


Figure 4.11: The position and velocity states of MAS under OMCoMA with attack detection scheme in **Proposition** 4.3.1 and mitigation scheme in **Proposition** 4.3.2.

in **Proposition** 4.3.2 are implemented in the MAS. The above OMCoMA is launched at t_a=3s. The parameters for the observer based synchronization algorithm (4.36) and (4.37) are selected as $p = \tau = 2$ and the initial values $\hat{z}_i, \forall i \in \mathcal{V}$ are randomly chosen from $[-0.2, 0.2]$. The simulation results are shown in Fig. 4.11 and Fig. 4.12. The position and velocity states converge before 3s. When OMCoMA is launched at 3s, it is detected by (4.29). MAS switches the synchronization algorithm from (4.3) to (4.36) and (4.37). The synchronization is finally reached. Additionally, the storage function V in (4.9) and the function rate $\dot{V}$ finally converge to zero. This demonstrates that the proposed attack detection and mitigation method secures the synchronization under the proposed MCoMA.

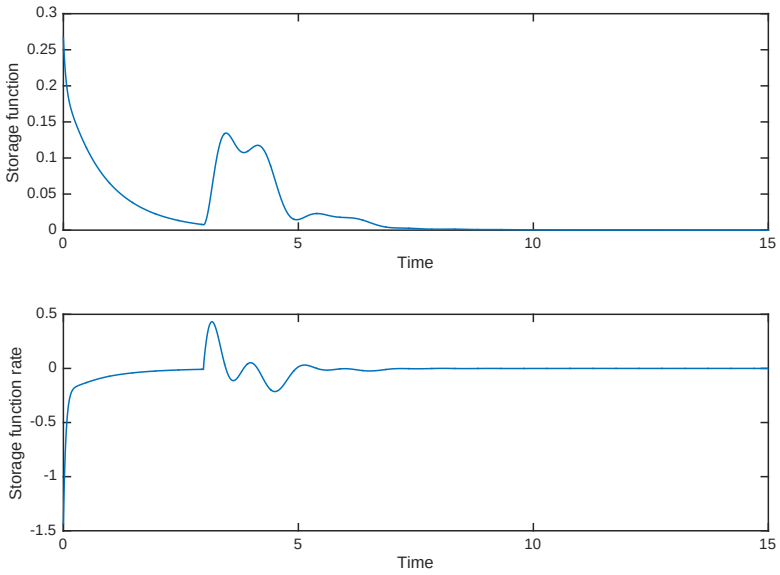


Figure 4.12: The storage function V and rate $\dot{V}$ of MAS under OM-CoMA with attack detection scheme in **Proposition** 4.3.1 and mitigation scheme in **Proposition** 4.3.2.

4.5 Summary

In this chapter, the vulnerability of MAS synchronization has been demonstrated by designing and analyzing MCoMA. An optimal MCoMA is then designed. The attack detection/mitigation scheme for MCoMA and the attack detection scheme for general content modification attacks are also discussed.

Chapter 5: Content Modification Attacks on Bilateral Tele-Driving System

In this chapter, the cyber-physical security issue is studied for a BTDS under content modification attacks, which can modify the data transmitted between the human operator and the tele-driven vehicle.

The results of this chapter can be summarized as follows: A passivity-based adaptive bilateral tele-driving control scheme is first proposed that enables a human operator to tele-drive a car-like mobile robot with haptic feedback in the presence of communication delays and dynamic parametric uncertainties. The static MCoMA is designed for the proposed tele-driving system. The physics-based attack detection scheme with an encoding-decoding structure is adopted for detecting general content modification attacks on the tele-driving system. MCoMA design and attack detection scheme are verified through various simulations. An initial BTDS experimental platform is also developed.

The rest of the chapter is organized as follows: In Section 5.1, the proposed tele-driving control scheme is first presented. Then, the attack design is discussed in Section 5.2. The detection scheme is discussed in Section 5.3. Finally, all the proposed schemes are simulated in Section 5.4.

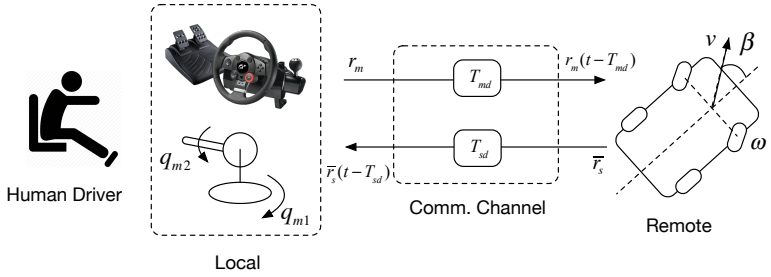


Figure 5.1: The sketch for tele-driving scheme: q_{m1}, q_{m2} are two joint variables of the local joystick, β is the steering angle, ω is the angular velocity of the front wheel, v is the linear velocity of the front wheel, T_{md}, T_{sd} are constant communication delays and $r_m, \bar{r}_s$ are the transmitted data defined later in Section 5.1.

5.1 Bilateral Tele-driving Control Scheme

As shown in Fig. 5.1, in BTDS, a human driver can a use 2 DOF local robot (joystick) to tele-drive a car-like remote robot through communication networks. This two DOF system can be analogously considered as steering wheel and gas/breaking pedal inputs. As introduced in Section 1.2.3, a tele-driving scheme with a new control mode is proposed in this work, where (q_{m1}, β)-coordination and $(q_{m2}, \dot{\omega})$-coordination (see Fig. 5.1) is achieved. Here $(.,.)$ implies that these signals track each other asymptotically. Specifically, (q_{m1}, β)-coordination and $(q_{m2}, \dot{\omega})$-coordination imply that the local robot's link variables q_{m1}, q_{m2} are used to control the steering angle β and the angular acceleration $\dot{\omega}$, respectively. In the proposed setup, the $(q_{m2}, \dot{\omega})$-coordination is equivalent to a $(q_{m2}, \dot{v})$-coordination as $v = r_F\omega$, where r_F is the radius of the front wheel.

It should be noted that many existing passivity-based methods cannot be

directly applied in the $(q_{m2}, \dot{\omega})$-coordination since the pair $(q_{m2}, \dot{\omega})$ implies position-acceleration coordination. In this work, the proposed scheme for the $(q_{m2}, \dot{\omega})$-coordination is inspired by the control algorithm for the (q_{m2}, v)-coordination in [37]. A new variable $r_{m2} = \dot{q}_{m2} + \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds, \lambda_1, \lambda_2 > 0$ is defined and transmitted instead of transmitting q_{m2} as was accomplished in [37]. Then $(q_{m2}, \dot{\omega})$-coordination can be approximately achieved by coupling r_{m2} with ω when the magnitude of $\dot{q}_{m2}$ and $\ddot{q}_{m2}$ are relatively small.

On the other hand, the (q_{m1}, β)-coordination requires position-position coordination. A PD-based control was applied in [37], but its controller gain is time delay dependent as shown in equation (11) in [37]. In the current work, a control scheme similar to the state synchronization scheme in Section 4.3.2 of [13] is proposed for the (q_{m1}, β)-coordination and the (r_{m2}, ω)-coordination, thereby avoiding delay dependent control gains.

Additionally, an adaptive control approach is utilized to address the uncertainties in the system dynamics. Furthermore, the synchronization algorithms discussed in [13, 61, 62] were developed under passivity assumptions on the human operator, which not always be the case in practice as discussed in [71]. In the proposed work, inspired by the scheme in [60], the passivity assumption is replaced with a boundedness condition on the human and environment input. The proposed control framework is different from [60] in two main respects: **(i)** The formulation is different. The scheme in [60] achieved the position tracking while the velocities are driven to zero and it cannot be applied here to achieve the new control mode for tele-driving. **(ii)** The system is different. [60] studied a traditional teleoperation

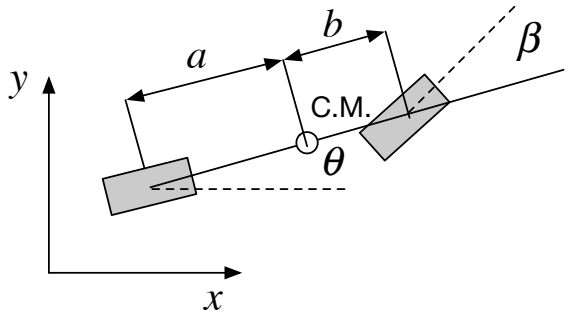


Figure 5.2: The single track model of the car-like robot: β is the steering angle, θ is the angel between the world reference frame and the longitudinal axis of the robot, a, b are the distance from center of mass to each wheel.

system, while a tele-driving system is considered here.

As part of the BTDS study, a new tele-driving control framework is proposed in this section. Compared with other related works, the contributions of the proposed control scheme are summarized in Section 1.4. Next, the formulation of the proposed tele-driving control scheme is described.

Based on Section 1.2.5, in the proposed tele-driving scheme in Fig. 5.1, the dynamics for the local manipulator are given as

$$M_m(q_m)\ddot{q}_m + C_m(q_m, \dot{q}_m)\dot{q}_m + g_m(q_m) = \tau_m + f_h \tag{5.1}$$

where $q_m = [q_{m1}, q_{m2}]^T \in \mathbb{R}^2$ represents the angular positions of the local robot, M_m is the inertial matrix, C_m is the centrifugal and Coriolis matrix, g_m is the gravitational torque, τ_m is the local robot control input and f_h is the human input.

The dynamics of the remote car-like mobile robot are approximated by a single-

track model shown in Fig. 5.2. Following [72], the dynamics are given as

$$M_s(q_s)\dot{\eta} + C_s(q_s, \dot{q}_s)\eta = \tau_s - f_e \tag{5.2}$$

where $\eta = [\dot{\beta}, \omega]^T \in \mathbb{R}^2$, $q_s = [\beta, \dot{\beta}]^T \in \mathbb{R}^2$, M_s is the inertial matrix, C_s is the centrifugal and Coriolis matrix, τ_s is the remote robot control input and f_e is the environment input.

To simplify the car-like robot model in [72], the following assumptions are made in this work:

- The robot moves on a horizontal plane without slip.
- Each wheel is modeled as a rigid wheel and the tire mechanics are not considered.
- Frictions is neglected in the model.

Following [72], the mathematical expression for each matrix of (5.2) is given as

$$M_s = \begin{bmatrix} J_{F,v} & r_F J_{F,v} \sin\beta/l \\ r_F J_{F,v} \sin\beta/l & m_{22} \end{bmatrix},$$
$$C_s = \begin{bmatrix} 0 & r_F J_{F,v} \cos\beta\dot{\beta}/l \\ 0 & c_{22} \end{bmatrix} \tag{5.3}$$

where

$$m_{22} = r_F^2 J_B \sin^2\beta/l^2 + J_{F,h} + r_F^2[m_F + \cos^2\beta(m_R + J_{R,h}/r_R^2)] + [(a^2 + l^2 + b(a+l)\cos 2\beta)m_B + 2\sin^2\beta(J_{F,v} + J_{R,v})]/l^2$$

$$c_{22} = \sin 2\beta[(J_B + J_{R,v} + J_{F,v})r_R^2 - (J_{R,h} + m_R r_R^2)l^2 - (a+l)bm_B r_R^2]r_F^2\dot{\beta}/2l^2 r_R^2$$

In the above equations, besides the notations introduced in Fig. 5.2, $J_B, J_{R,v}$, $J_{R,h}, J_{F,v}, J_{F,h}$ are the moments of inertia of body, and the moments of inertia of rear wheel and front wheel along vertical and horizontal wheel axis, respectively; r_F, r_R are the radius of front and rear wheel, respectively; $l := a + b$ is the distance between front and rear wheel; m_B, m_F, m_R are the mass of body, front wheel and rear wheel, respectively.

Given the structure of M_s, C_s in (5.3) and the fact that the local joystick can be modeled as a manipulator with two revolute joints as (5.1), it can be verified that the properties **(P1.2)** and **(P1.4)** of Lagrangian dynamics also hold for the dynamics of the local and remote robot.

The proposed adaptive coordination control framework consists of two steps:

(i) An adaptive control law is first used to render the local and remote robot dynamics passive with respect to the new defined outputs;

(ii) Then a passive coordination control is applied to achieve the desired coordination between the local and remote robot.

As shown in Fig. 5.1, define the two coordination signals r_m and $\bar{r}_s$ as

$$r_m = \begin{bmatrix} r_{m1} \\ \\ r_{m2} \end{bmatrix} = \begin{bmatrix} \dot{q}_{m1} + \lambda q_{m1} \\ \\ \dot{q}_{m2} + \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds \end{bmatrix} \tag{5.4}$$

$$\bar{r}_s = \begin{bmatrix} \bar{r}_{s1} \\ \\ \bar{r}_{s2} \end{bmatrix} = Kr_s = K \begin{bmatrix} \dot{\beta} + \lambda\beta \\ \\ \omega \end{bmatrix} \tag{5.5}$$

where β is the steering angle, ω is the angular velocity of the front wheel as shown in Fig. 5.1, and $\lambda, \lambda_1, \lambda_2 > 0$ are constant coefficients. It is natural to consider a scaling factor between r_m and r_s due to the kinematic dissimilarity between the local and remote robot. Hence, a constant diagonal positive definite scaling factor matrix K is considered here. Consequently, in the ideal no delay case, the coordination control would guarantee that Kr_s tracks r_m, where $K = diag([k_1, k_2]), k_1, k_2 > 0$.

Assume there exists constant time delays T_{md} and T_{sd} between the local and remote robot in the network as shown in Fig. 5.1. Then, the coordination error signals can be defined as

$$e_{rm} = r_m - \bar{r}_s(t - T_{sd}), \quad e_{rs} = \bar{r}_s - r_m(t - T_{md}) \tag{5.6}$$

Definition 5.1.1. In this chapter, the coordination is said to be achieved by the proposed tele-driving system if

$$\begin{aligned}
&\lim_{t\to\infty}(q_{m1}(t - T_{md}) - k_1\beta) = \lim_{t\to\infty}(q_{m1} - k_1\beta(t - T_{sd})) = 0 \\
&\lim_{t\to\infty}(\dot{q}_{m1}(t - T_{md}) - k_1\dot{\beta}) = \lim_{t\to\infty}(\dot{q}_{m1} - k_1\dot{\beta}(t - T_{sd})) = 0 \\
&\lim_{t\to\infty}(r_{m2}(t - T_{md}) - k_2\omega) = \lim_{t\to\infty}(r_{m2} - k_2\omega(t - T_{sd})) = 0
\end{aligned} \tag{5.7}$$

Assuming that the dynamics parameters of the local and remote robots are uncertain, their control inputs τ_m and τ_s are then designed as

$$\tau_m = u_m - \hat{M}_m(\begin{bmatrix} \lambda\dot{q}_{m1} \\ \lambda_1\dot{q}_{m2} + \lambda_2 q_{m2} \end{bmatrix} - \dot{\bar{r}}_s(t - T_{sd})) - \hat{C}_m(\begin{bmatrix} \lambda q_{m1} \\ \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds \end{bmatrix}$$

$$- \bar{r}_s(t - T_{sd})) + \hat{g}_m \tag{5.8}$$

$$\tau_s = u_s - \hat{M}_s(\begin{bmatrix} \lambda\dot{\beta}, \\ 0 \end{bmatrix} - K^{-1}\dot{r}_m(t - T_{md})) - \hat{C}_s(\begin{bmatrix} \lambda\beta \\ 0 \end{bmatrix} - K^{-1}r_m(t - T_{md})) \tag{5.9}$$

where $\hat{M}_i$, $\hat{C}_i$ $(i = m, s)$ and $\hat{g}_m$ are the estimates of the model matrices, and u_m, u_s are the coordination control inputs to be designed.

Using property **(P1.4)**, the above control inputs can be rewritten as

$$\begin{aligned} \tau_m &= u_m + Y_m(q_m, \dot{q}_m, \bar{r}_s(t - T_{sd}), \dot{\bar{r}}_s(t - T_{sd}))\hat{\phi}_m \\ &= u_m + Y_m\phi_m + Y_m\tilde{\phi}_m, \end{aligned} \tag{5.10}$$

$$\begin{aligned} \tau_s &= u_s + Y_s(\beta, \dot{\beta}, r_m(t - T_{md}), \dot{r}_m(t - T_{md}))\hat{\phi}_s \\ &= u_s + Y_s\phi_s + Y_s\tilde{\phi}_s \end{aligned} \tag{5.11}$$

where $\hat{\phi}_m, \hat{\phi}_s$ are the time-varying estimates of the robots model parameters given by ϕ_m, ϕ_s respectively and $\tilde{\phi}_m := \hat{\phi}_m - \phi_m, \tilde{\phi}_s := \hat{\phi}_s - \phi_s$ are the parameters estimation errors.

Then substituting (5.10) and (5.11) into (5.1) and (5.2) gives

$$M_m\dot{e}_{rm} + C_m e_{rm} = u_m + Y_m\tilde{\phi}_m + f_h \tag{5.12}$$

$$\bar{M}_s\dot{e}_{rs} + \bar{C}_s e_{rs} = u_s + Y_s\tilde{\phi}_s - f_e \tag{5.13}$$

where $\bar{M}_s = M_s K^{-1}, \bar{C}_s = C_s K^{-1}$.

The update laws for the uncertain parameters estimates are given as

$$\dot{\hat{\phi}}_m = -\Gamma_m^{-1}Y_m^T e_{rm}, \quad \dot{\hat{\phi}}_s = -\Gamma_s^{-1}Y_s^T e_{rs} \tag{5.14}$$

where Γ_m and Γ_s are constant positive definite matrices.

We next establish passivity properties of the local and remote systems, in the absence of external inputs f_h and f_e.

Lemma 5.1.1. Consider $f_h = 0$ in (5.12), system (5.12) is passive with (u_m, e_{rm}) as the input-output pair with respect to the storage function $S_m = \frac{1}{2}e_{rm}^T M_m e_{rm} + \frac{1}{2}\tilde{\phi}_m^T \Gamma_m \tilde{\phi}_m$.

Proof:

Using the property **(P1.2)**, the derivative of S_m along the trajectory of (5.12) (5.14) is computed as

$$\dot{S}_m = e_{rm}^T(-C_m e_{rm} + u_m + Y_m\tilde{\phi}_m) + \frac{1}{2}e_{rm}^T\dot{M}_m e_{rm} - \tilde{\phi}_m^T Y_m^T e_{rm} = e_{rm}^T u_m$$

Following the definition of passivity from [13], the system (5.12) is passive with (u_m, e_{rm}) as the input-output pair with respect to the storage function S_m.

■

Similarly, when $f_e = 0$, system (5.13) is passive with (u_s, e_{rs}) as the input-output pair with respect to the storage function $S_s = \frac{1}{2}e_{rs}^T \bar{M}_s e_{rs} + \frac{1}{2}\tilde{\phi}_s^T \Gamma_s \tilde{\phi}_s$.

In this scheme, the coordination control u_m and u_s are designed as

$$u_m = -K_u e_{rm}, \quad u_s = -K_u e_{rs} \tag{5.15}$$

where K_u is a constant diagonal positive definite control gain matrix.

Theorem 5.1.1. Consider the tele-driving system described by (5.1)-(5.15). Then,

(a) if $f_h, f_e \in \mathcal{L}_2 \cap \mathcal{L}_\infty$, then the signal $e_{rm}, e_{rs}, \tilde{\phi}_m, \tilde{\phi}_s$ are bounded and the state coordination is achieved for the tele-driving system in the sense of **Definition 5.1.1**;

(b) if the tele-driven car is in hard contact with the environment assuming steady state $(\dot{q}_m, \ddot{q}_m, \dot{\beta}, \ddot{\beta}, \omega, \dot{\omega} \to 0)$, then the driver can perceive the environment contact force as $f_h \to f_e + K_u \begin{bmatrix} 0 \\ \lambda_2 \int_{t-T_{md}}^{t} q_{m2}(s)ds \end{bmatrix} - Y_m\tilde{\phi}_m - Y_s\tilde{\phi}_s + (M_m + \bar{M}_s) \begin{bmatrix} 0 \\ \lambda_2 q_{m2} \end{bmatrix}$.

Proof:

(a) A positive semidefinite storage functional V is considered for the tele-driving system as

$$V = \frac{1}{2}e_{rm}^T M_m e_{rm} + \frac{1}{2}\tilde{\phi}_m^T \Gamma_m \tilde{\phi}_m + \frac{1}{2}e_{rs}^T \bar{M}_s e_{rs} + \frac{1}{2}\tilde{\phi}_s^T \Gamma_s \tilde{\phi}_s \tag{5.16}$$

With the help of the property **(P1.2)**, the derivative of V along the trajectories of system (5.12), (5.13) and (5.14) can be computed as

$$\begin{aligned}
\dot{V} &= e_{rm}^T(-C_m e_{rm} + u_m + Y_m\tilde{\phi}_m + f_h) + \frac{1}{2}e_{rm}^T\dot{M}_m e_{rm} - \tilde{\phi}_m^T Y_m^T e_{rm} \\
&+ e_{rs}^T(-\bar{C}_s e_{rs} + u_s + Y_s\tilde{\phi}_s - f_e) + \frac{1}{2}e_{rs}\dot{\bar{M}}_s e_{rs} - \tilde{\phi}_s^T Y_s^T e_{rs} \\
&= e_{rm}^T u_m + e_{rs}^T u_s + e_{rm}^T f_h - e_{rs}^T f_e
\end{aligned}$$

Substituting u_m, u_s in (5.15) into above equation gives

$$\begin{aligned}
\dot{V} &= -e_{rm}^T K_u e_{rm} - e_{rs}^T K_u e_{rs} + e_{rm}^T f_h - e_{rs}^T f_e \\
&\leq -\lambda_K |e_{rm}|^2 - \lambda_K |e_{rs}|^2 + |e_{rm}||f_h| + |e_{rs}||f_e|
\end{aligned}$$

where $\lambda_K = \lambda_m(K_u) > 0$.

Using Young's inequality, we have

$$|e_{rm}||f_h| \leq \frac{|f_h|^2}{2\lambda_K} + \frac{\lambda_K |e_{rm}|^2}{2}, \ |e_{rs}||f_e| \leq \frac{|f_e|^2}{2\lambda_K} + \frac{\lambda_K |e_{rs}|^2}{2},$$

Thus,

$$\dot{V} \leq -\frac{\lambda_K}{2}|e_{rm}|^2 - \frac{\lambda_K}{2}|e_{rs}|^2 + \frac{1}{2\lambda_K}|f_h|^2 + \frac{1}{2\lambda_K}|f_e|^2$$

Integrating $\dot{V}$ from 0 to t gives

$$V(t) + \frac{\lambda_K}{2}\int_0^t |e_{rm}|^2 ds + \frac{\lambda_K}{2}\int_0^t |e_{rs}|^2 ds \leq \frac{1}{2\lambda_K}\int_0^t |f_h|^2 ds + \frac{1}{2\lambda_K}\int_0^t |f_e|^2 ds + V(0)$$

Letting $t \to \infty$, and using the assumption that $f_h, f_e \in \mathcal{L}_2$, we have $e_{rm}, e_{rs} \in \mathcal{L}_2$ and V is bounded. Hence, from (5.16) the signals $e_{rm}, e_{rs}, \tilde{\phi}_m, \tilde{\phi}_s$ are bounded.

From (5.12) and (5.13), and using the assumption that $f_h, f_e \in \mathcal{L}_\infty$, $\dot{e}_{rm}, \dot{e}_{rs}$ are also bounded. Using Barbalat's Lemma [66] gives $\lim_{t\to\infty} e_{rm}(t) = \lim_{t\to\infty} e_{rs}(t) = 0$, which means

$$\lim_{t\to\infty}(\bar{r}_{s1}(t - T_{sd}) - r_{m1}) = \lim_{t\to\infty}(r_{m1}(t - T_{md}) - \bar{r}_{s1}) = 0 \tag{5.17}$$

$$\lim_{t\to\infty}(\bar{r}_{s2}(t - T_{sd}) - r_{m2}) = \lim_{t\to\infty}(r_{m2}(t - T_{md}) - \bar{r}_{s2}) = 0 \tag{5.18}$$

Using (5.5), (5.18) is equivalent to

$$\lim_{t\to\infty}(r_{m2}(t - T_{md}) - k_2\omega) = \lim_{t\to\infty}(r_{m2} - k_2\omega(t - T_{sd})) = 0 \tag{5.19}$$

The signal $\bar{r}_{s1}(t - T_{sd}) - r_{m1}$ can be rewritten as

$$\bar{r}_{s1}(t - T_{sd}) - r_{m1} = \dot{e}_\beta + \lambda e_\beta \tag{5.20}$$

where $e_\beta := k_1\beta(t - T_{sd}) - q_{m1}$.

As (5.20) is an exponentially stable linear system with input $\bar{r}_{s1}(t-T_{sd}) - r_{m1}$ and state e_β, by (5.17) and Theorem A.4 in [13], if $\bar{r}_{s1}(t - T_{sd}) - r_{m1} \in \mathcal{L}_2$ and asymptotically converges to zero, then

$$\lim_{t\to\infty} e_\beta = \lim_{t\to\infty} \dot{e}_\beta = 0 \tag{5.21}$$

Similarly, if $r_{m1}(t - T_{md}) - \bar{r}_{s1}$ is rewritten as

$$r_{m1}(t - T_{md}) - \bar{r}_{s1} = \dot{e}_{q_{m1}} + \lambda e_{q_{m1}} \tag{5.22}$$

where $e_{q_{m1}} := q_{m1}(t - T_{md}) - k_1\beta$, we can also show

$$\lim_{t\to\infty} e_{q_{m1}} = \lim_{t\to\infty} \dot{e}_{q_{m1}} = 0 \tag{5.23}$$

Hence, the proof of the part (a) is complete.

(b) Now assuming $\dot{q}_m, \ddot{q}_m, \dot{\beta}, \ddot{\beta}, \omega, \dot{\omega} \to 0$, then we have $C_m(q_m, \dot{q}_m) \to 0$ as $\dot{q}_m \to 0$ and $\dot{e}_{rm} \to [0, \lambda_2 q_{m2}]^T$ in (5.12), thus

$$\begin{aligned} f_h &\to -u_m - Y_m\tilde{\phi}_m + M_m \begin{bmatrix} 0 \\ \\ \lambda_2 q_{m2} \end{bmatrix} \\ &= K_u \begin{bmatrix} -k_1\lambda\beta(t - T_{sd}) + \lambda q_{m1} \\ \\ \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds \end{bmatrix} - Y_m\tilde{\phi}_m + M_m \begin{bmatrix} 0 \\ \\ \lambda_2 q_{m2} \end{bmatrix} \end{aligned} \tag{5.24}$$

In (5.13), $\dot{\bar{r}}_s \to 0$ and $\bar{C}_s(q_s, \dot{q}_s) \to 0$ due to the structure of C_s in (5.3). Hence,

$$\begin{aligned} f_e &\to u_s + Y_s\tilde{\phi}_s - \bar{M}_s \begin{bmatrix} 0 \\ \\ \lambda_2 q_{m2}(t - T_{md}) \end{bmatrix} \\ &= K_u \begin{bmatrix} -k_1\lambda\beta + \lambda q_{m1}(t - T_{md}) \\ \\ \lambda_1 q_{m2}(t - T_{md}) + \lambda_2 \int_0^{t-T_{md}} q_{m2}(s)ds \end{bmatrix} + Y_s\tilde{\phi}_s - \bar{M}_s \begin{bmatrix} 0 \\ \\ \lambda_2 q_{m2}(t - T_{md}) \end{bmatrix} \end{aligned}$$

For a signal $x(t)$ and $i = m, s$, $x(t) \to x(t - T_{id})$ as $\lim_{t\to\infty} \dot{x}(t) = 0$, thus when $\dot{q}_m, \ddot{q}_m, \dot{\beta}, \ddot{\beta}, \omega, \dot{\omega} \to 0$, $f_h \to$

$$K_u \begin{bmatrix} -k_1\lambda\beta(t) + \lambda q_{m1} \\ \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds \end{bmatrix} - Y_m\tilde{\phi}_m + M_m \begin{bmatrix} 0 \\ \lambda_2 q_{m2} \end{bmatrix}$$
$$= f_e + K_u \begin{bmatrix} 0 \\ \lambda_2 \int_{t-T_{md}}^t q_{m2}(s)ds \end{bmatrix} - Y_m\tilde{\phi}_m - Y_s\tilde{\phi}_s + (M_m + \bar{M}_s) \begin{bmatrix} 0 \\ \lambda_2 q_{m2} \end{bmatrix} \tag{5.25}$$

which completes the proof of the part (b).

■

Remark 5.1.1. Compared with the state synchronization control scheme for the traditional bilateral teleoperation system in [13,61,62], the proposed scheme formulates a passive system as (5.12) and (5.13) with the new defined outputs e_{rm} and e_{rs} as (5.6). Additionally, the passivity assumption on the human and environment in [13, 61, 62] can be avoided as has been accomplished in part (a) of ***Theorem 5.1.1***. It should be noted that the bilateral teleoperation algorithms and results developed in [13, 61, 62] can be made less conservative by avoiding the passivity assumption on the human operator and environment, as has been done in the proposed work.

Remark 5.1.2. From part (b) of **Theorem 5.1.1**, the force reflection from the environment can be perceived by the human driver when the car is in hard contact with the environment and in steady state. For example, when the car's tire hits an obstacle like the curb on the road, the environment force information can be

provided to the local driver by (5.25), and hence the driver's situational awareness can be improved. In (5.25), the first and second row reflect the force feedback on the steering direction and the forward driving direction, respectively. Observing the final expression of (5.25), the effect of time delay on the force reflection is represented by the second term, the effect of model parameters estimation errors on the force reflection is represented by the third and fourth term, and the coupling between two directions is represented by the last term.

5.2 Attack Design for Tele-driving System

The tele-driving system described in the previous section is also vulnerable to content modification attacks discussed in Chapter 2. Given **Definition** 2.1.1, the attacker is a malicious external entity which can launch content modification attack by modifying the data content being exchanged by local and remote robot. Next the MCoMA design for proposed tele-driving system based on **Definition** 2.2.1 is presented. Here we consider two cases of MCoMA: **(i)** MCoMA on one side (remote or local side); **(ii)** MCoMA on both sides. In the following designs, for simplicity, we assume the there is no interaction between robot system and human/environment, i.e., $f_h = f_e = 0$.

5.2.1 MCoMA on One Side

MCoMA on the remote side can be implemented as Fig. 5.3. The attacker is co-located with the remote robot, and it can intercept the communication channel

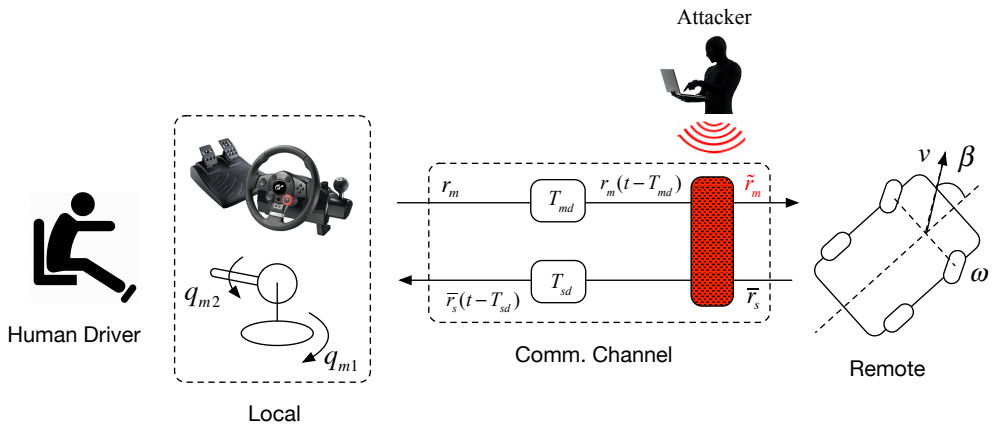


Figure 5.3: MCoMA on remote robot side.

and get access to $r_m(t - T_{md})$ and $\bar{r}_s$.

Proposition 5.2.1. Consider the proposed tele-driving system and the attacker is co-located with the remote robot, a content modification attack that modifies the received data for remote robot as

$$\tilde{r}_m = r_m(t - T_{md}) + \delta_{s1} r_m(t - T_{md}) + \delta_{s2} \bar{r}_s, \tag{5.26}$$

where $\delta_{s1}, \delta_{s2} \in \mathbb{R}$ are constant attack gain, is MCoMA with respect to storage function

$$V_s = \frac{1}{2} e_{rs}^T \bar{M}_s e_{rs} + \frac{1}{2} \tilde{\phi}_s^T \Gamma_s \tilde{\phi}_s \tag{5.27}$$

if (a) $\delta_{s1} = -k - 1, \delta_{s2} = 1 + k$, where $k > 0$ is a constant; (b) $V_s \neq 0$ at $t = t_a$.

Proof:

For remote robot, we consider the storage function given in (5.27).

Under the attack (5.26), the control input (5.15) for remote robot becomes

$$\begin{aligned} u_s &= -K_u(\bar{r}_s - \tilde{r}_m) \\ &= -K_u((1-\delta_{s2})\bar{r}_s - (1+\delta_{s1})r_m(t-T_{md})) \end{aligned} \tag{5.28}$$

Thus the derivative of V_s along the trajectory of (5.13) is

$$\dot{V}_s = -e_{rs}^T K_u(\bar{r}_s - \tilde{r}_m) = -e_{rs}^T K_u((1-\delta_{s2})\bar{r}_s - (1+\delta_{s1})r_m(t-T_{md})) \tag{5.29}$$

Now (a) $\delta_{s1} = -k-1, \delta_{s2} = 1+k, k>0$ leads to

$$\dot{V}_s = k e_{rs}^T K_u e_{rs} > 0, \ \forall e_{rs} \neq 0, \ t \geq t_a \tag{5.30}$$

Since (b) $V_s \neq 0$ at $t = t_a$ and $\dot{V}_s \geq 0$ for $t \geq t_a$, $\tilde{\phi}_s \neq 0$ when $e_{rs} = 0$. Then given (5.13), when $e_{rs} = 0$, the nonzero $\tilde{\phi}_s$ can render $\dot{e}_{rs} \neq 0$, which can drive e_{rs} away from zero. Thus, $\dot{V}_s$ stays positive and $\lim_{t\to\infty} V_s = \infty$ for $t \geq t_a$, which proves the proposed attack (5.26) is MCoMA for the remote robot with respect to V_s.

■

Following the similar approach, MCoMA attack can be designed on the local side:

Proposition 5.2.2. Consider the proposed tele-driving system and the attacker is co-located with the local robot, a content modification attack that modifies the received data for local robot as

$$\tilde{r}_s = \bar{r}_s(t - T_{sd}) + \delta_{m1}\bar{r}_s(t - T_{sd}) + \delta_{m2} r_m, \tag{5.31}$$

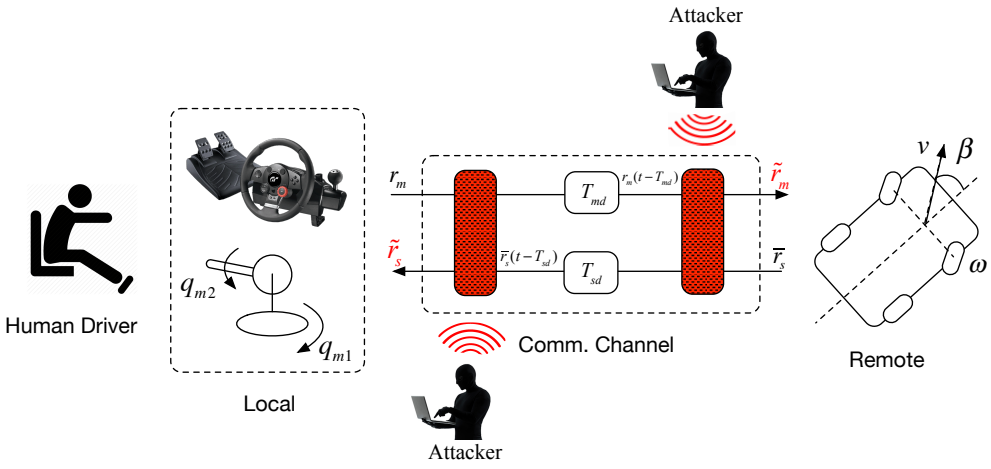


Figure 5.4: Distributed MCoMA on both sides.

where $\delta_{m1}, \delta_{m2} \in \mathbb{R}$ are constant attack gain, is MCoMA with respect to storage function

$$V_m = \frac{1}{2} e_{rm}^T M_m e_{rm} + \frac{1}{2} \tilde{\phi}_m^T \Gamma_m \tilde{\phi}_m \tag{5.32}$$

if (a) $\delta_{m1} = -k - 1, \delta_{m2} = 1 + k$, where $k > 0$ is a constant; (b) $V_m \neq 0$ at $t = t_a$.

The proof is similar to proof of **Proposition** 5.2.1, thus it is omitted here.

5.2.2 MCoMA on Both Sides

MCoMA can be designed on both sides of the tele-driving system with respect to the storage function V given in (5.16). As shown in Figure 5.4, there are two non-colluding distributed attackers which reside on the local and remote side, respectively.

Proposition 5.2.3. Consider the proposed tele-driving system and two non-colluding distributed attackers which reside on each side, content modification attack that

modifies the received data for local and remote robot as

$$\tilde{r}_m = r_m(t - T_{md}) + \delta_{s1} r_m(t - T_{md}) + \delta_{s2} \bar{r}_s \tag{5.33}$$

$$\tilde{r}_s = \bar{r}_s(t - T_{sd}) + \delta_{m1} \bar{r}_s(t - T_{sd}) + \delta_{m2} r_m,$$

where $\delta_{s1}, \delta_{s2}, \delta_{m1}, \delta_{m2} \in \mathbb{R}$ are constant attack gain, is MCoMA with respect to storage function given in (5.16) if (a) $\delta_{m1} = \delta_{s1} = -k - 1, \delta_{m2} = \delta_{s2} = 1 + k$, where $k > 0$ is a constant; (b) $V \neq 0$ at $t = t_a$.

The proof is also similar to proof of **Proposition** 5.2.1, thus it is omitted here.

5.3 Attack Detection for Tele-driving System

In the previous section, the MCoMA design is discussed, which is a special content modification attack. In more general cases, the attacker can stay anywhere between the local and remote side, and can just launch a general content modification attack defined as (2.1) and (2.2) to arbitrarily modify the data content. Here the attack detection algorithm proposed in Section 2.3 is adopted to detect the general content modification attack on the proposed tele-driving system.

Before proceeding, some notations need to be defined. The r_m in (5.4) received

by the remote vehicle can be written as

$$r_m = \begin{bmatrix} \dot{q}_{m1} + \lambda q_{m1} \\ \dot{q}_{m2} + \lambda_1 q_{m2} + \lambda_2 \int_0^t q_{m2}(s)ds \end{bmatrix} \tag{5.34}$$
$$= \begin{bmatrix} \dot{q}_{m1} + \lambda q_{m1} \\ \dot{q}_{m2} + \alpha q_{m2} + \lambda(q_{m2} + \alpha \int_0^t q_{m2}(s)ds) \end{bmatrix},$$

where α is a constant coefficient satisfying $\alpha + \lambda = \lambda_1$ and $\lambda\alpha = \lambda_2$. Let us define $S := q_{m2} + \alpha \int_0^t q_{m2}(s)ds$, we have

$$r_m = \begin{bmatrix} \dot{q}_{m1} + \lambda q_{m1} \\ \dot{S} + \lambda S \end{bmatrix} = \dot{p}_m + \lambda p_m, \quad p_m := [q_{m1}, S]^T \tag{5.35}$$

Also, p_s is defined as $p_s := [\beta, \omega]$. As shown in Fig. 5.5, the physics-based attack detection scheme with an encoding-decoding structure proposed in Section 2.3 can be implemented for BTDS by replacing the notation $(\dot{z}_i, z_i)$ in Section 2.3 with $(\dot{p}_i, p_i)$ here.

5.4 Simulation

In this section, the proposed tele-driving scheme, attack design and attack detection scheme are simulated in Matlab.

5.4.1 Tele-driving Control Scheme

To simplify the presentation, it is assumed that the local robot dynamics parameters are known and the remote robot dynamics parameters are uncertain. The local robot consisted of two decoupled individual single link planar manipulators,

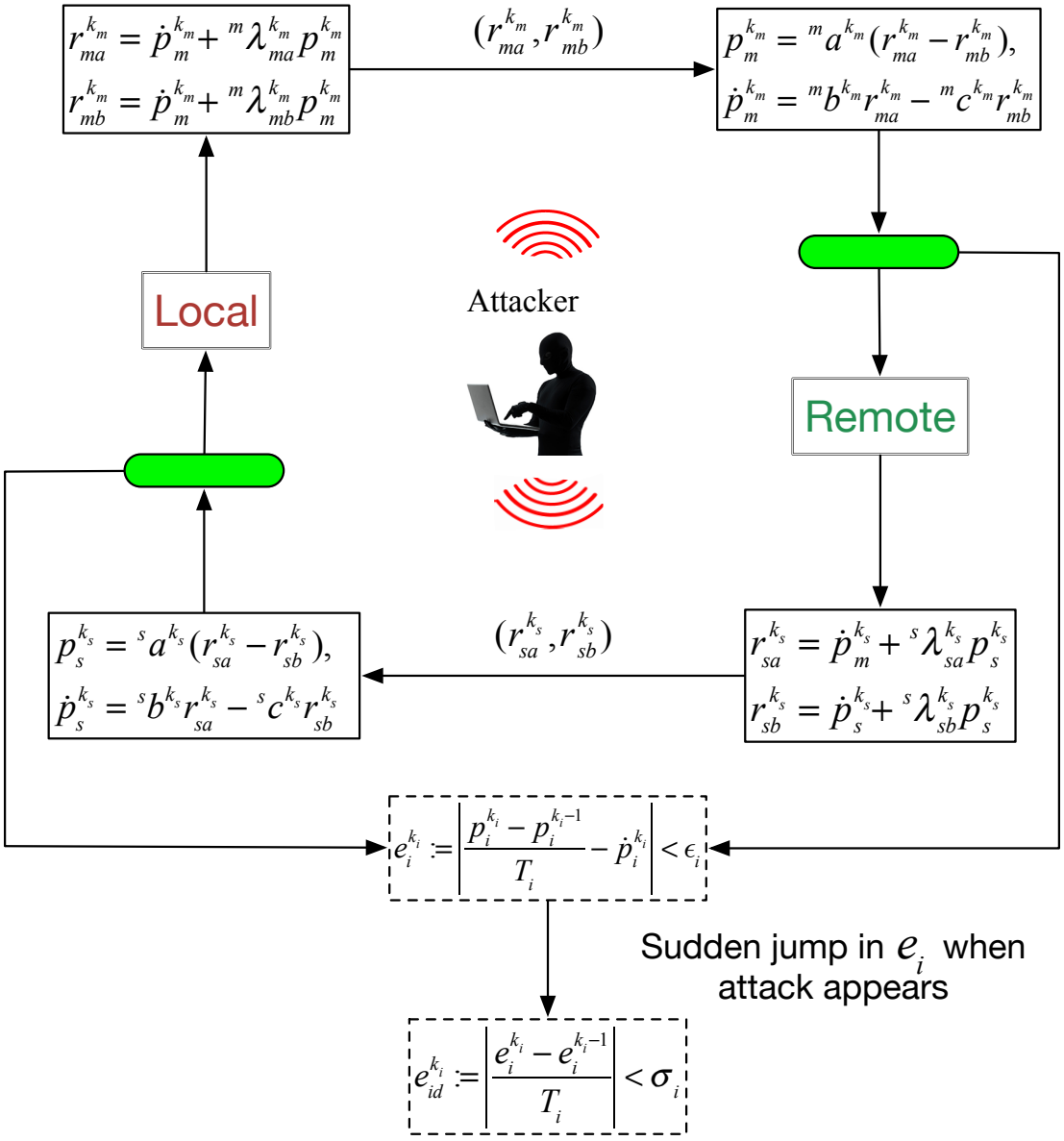


Figure 5.5: Physics-based attack detection scheme with an encoding-decoding structure for BTDS.

and hence the dynamics can be written as

$$M_m\ddot{q}_m = \tau_m + f_h, \quad q_m \in \mathbb{R}^2 \tag{5.36}$$

where $M_m = diag([J_{m1}, J_{m2}])$.

For the remote robot, following (5.2) (5.3), it can be shown that

$$-M_s(\begin{bmatrix} \lambda\dot{\beta}, \\ 0 \end{bmatrix} - K^{-1}\dot{r}_m(t - T_{md})) - C_s(\begin{bmatrix} \lambda\beta \\ 0 \end{bmatrix} - K^{-1}r_m(t - T_{md})) = Y_s\phi_s, \tag{5.37}$$

and it can be verified that the parameters vector $\phi_s \in \mathbb{R}^7$ and the regressor matrix $Y_s(\beta, \dot{\beta}, r_m(t - T_{md}), \dot{r}_m(t - T_{md})) \in \mathbb{R}^{2\times 7}$.

The parameters in the control scheme are taken as following: $\lambda = 3, \lambda_1 = 2, \lambda_2 = 2, k_1 = 2, k_2 = 3$ and the control gain matrix $K_u = diag([4, 10])$. The simulation parameters are taken as the following: simulation time $t = 12s$, time step $T_m = T_s = 0.005s$, time delays $T_{sd} = 0.1s, T_{md} = 0.15s$, the initial condition for system states $(q_{m1}, q_{m2}, \dot{q}_{m1}, \dot{q}_{m2}, \beta, \dot{\beta}, \omega)$ is $(-0.1, 0, 0, 0, 0.2, 0, 0)$, the nominal value of ϕ_s in (5.37) is chosen as $(1, 0.2, 4.04, 8.12, 1.24, 5.04, -0.5808)$. The human input on the local robot $f_h = [f_{h1}, f_{h2}]$ is shown in the Fig. 5.6. In (5.8) and (5.9), the derivative of $\bar{r}_s(t - T_{sd})$ and $r_m(t - T_{md})$ are computed and used, and in the simulation, a low-pass filter is applied to $\bar{r}_s(t - T_{sd})$ and $r_m(t - T_{md})$ before the derivatives are computed to get rid of the noise issue in the derivative computation. In the simulation, the nominal value of ϕ_s is unknown, and a perturbation is added on the nominal value of ϕ_s to generate an initial condition of $\hat{\phi}_s$ in the adaptation law (5.14) for the remote robot.

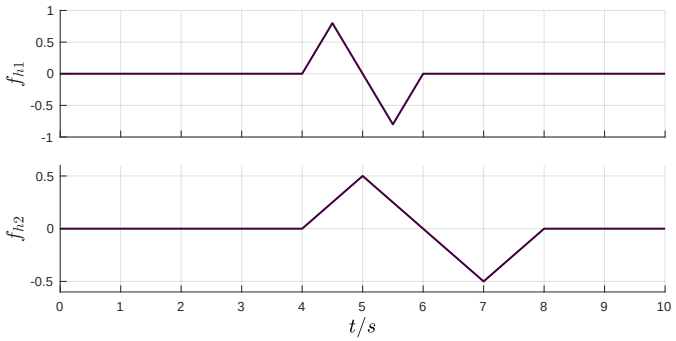


Figure 5.6: The human input $f_h = [f_{h1}, f_{h2}]$ in the simulation.

The coordination performance of the proposed control scheme is shown in Fig. 5.7, and a zoomed-in plot of the simulation results for $t \in [3s, 9s]$ is shown in Fig. 5.8. A good coordination performance under the human input is achieved with communication delays and dynamic parametric uncertainties. In the fourth subplot of Fig. 5.8, it is also verified that the new control mode (position-acceleration $(q_{m2}, \dot{\omega})$ coordination) is achieved by the proposed control scheme. Fig. 5.9 displays the trajectories of the parameter vector $\hat{\phi}_s$, and as demonstrated in part (a) of ***Theorem 5.1.1***, the parameter estimation errors remain bounded.

5.4.2 MCoMA Design

In this part, a MCoMA is designed on the remote robot side as **Proposition 5.2.1**. The simulation time $t = 2s$, k in **Proposition 5.2.1** is set as 0.5, the attack gains $\delta_{s1} = -1 - k = -1.5, \delta_{s2} = 1 + k = 1.5$ and the rest of the parameters are the same. The simulation result is shown in Fig. 5.10, and the storage function V_s keeps increasing with non-negative increasing rate, which verifies the impact of MCoMA.

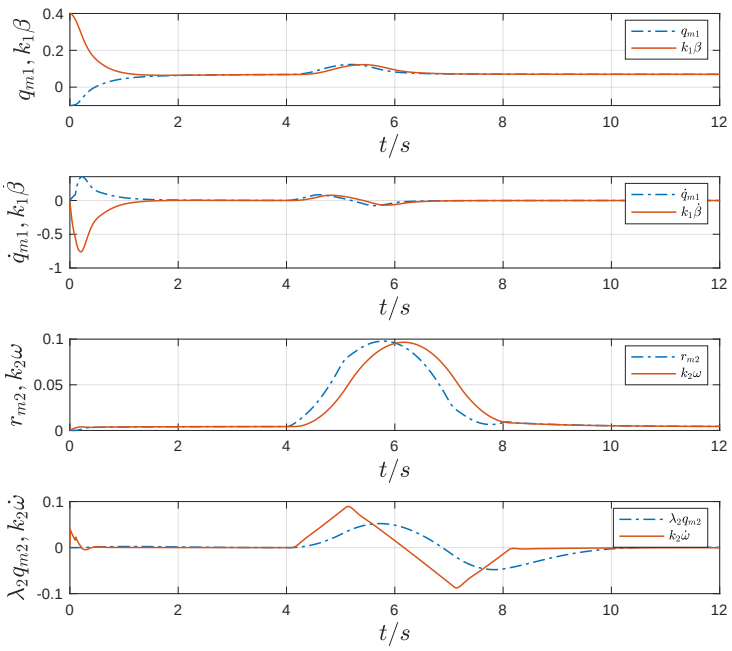


Figure 5.7: The coordination performance of the proposed control scheme.

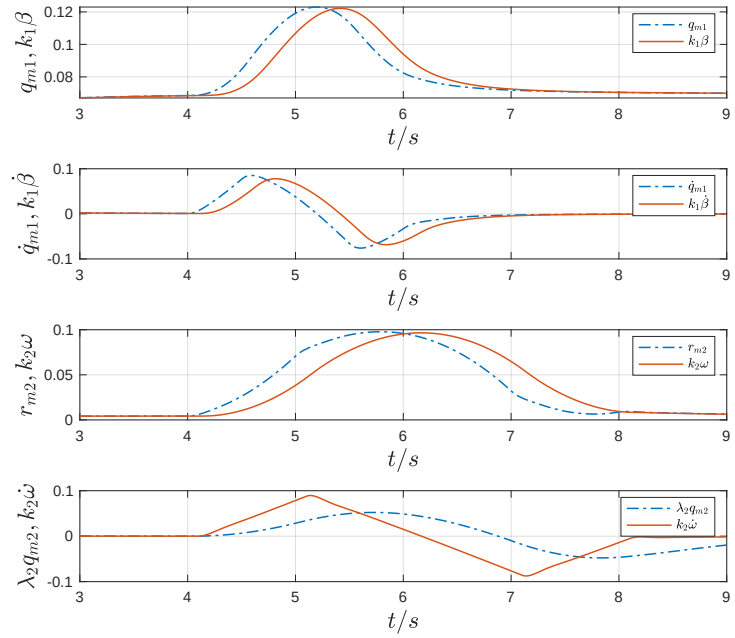


Figure 5.8: The coordination performance for $t \in [3s, 9s]$.

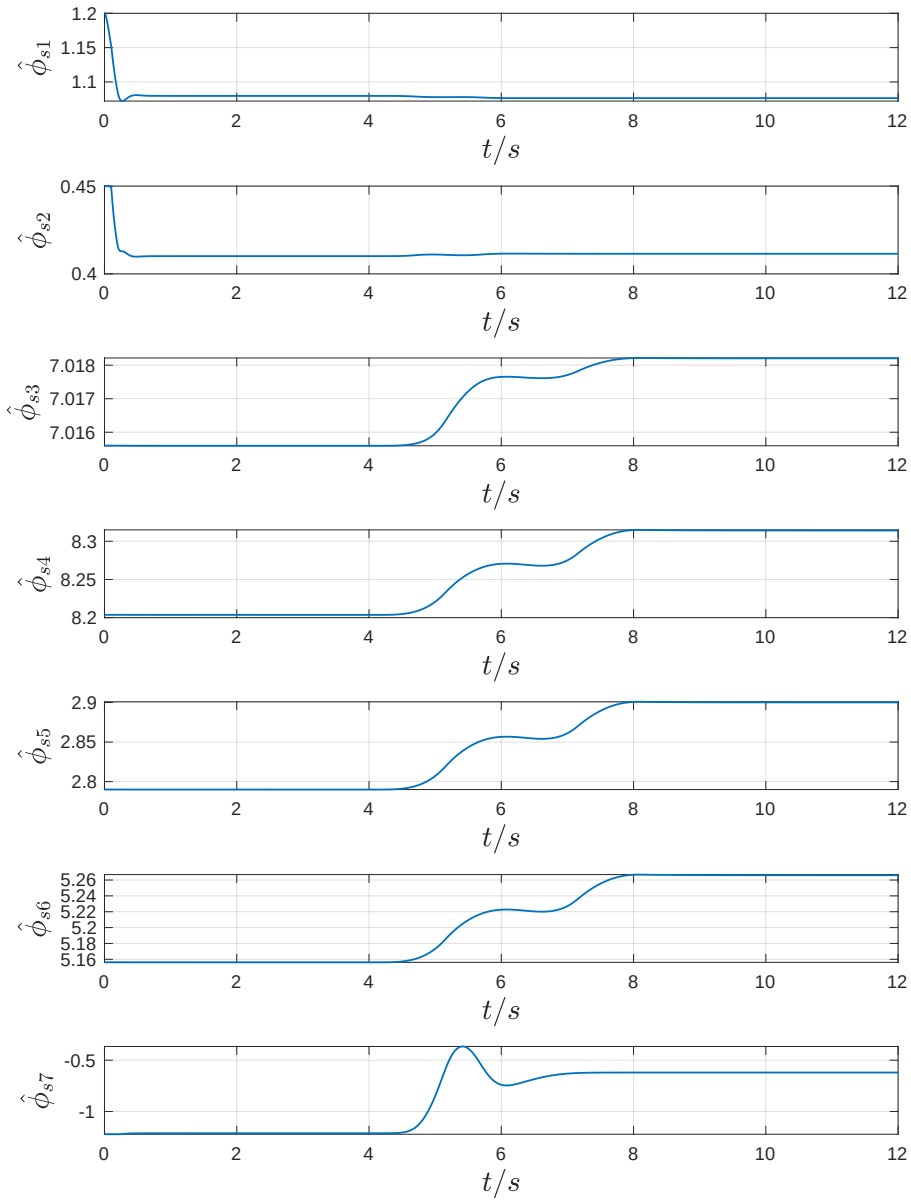


Figure 5.9: The trajectories of the parameters vector $\hat{\phi}_s$ for the remote robot.

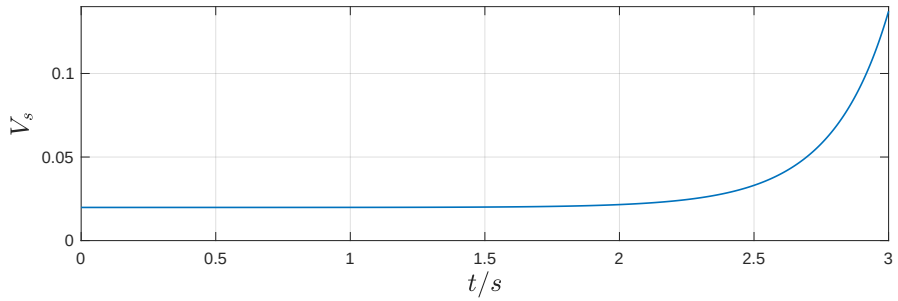


Figure 5.10: V_s under MCoMA on remote robot side.

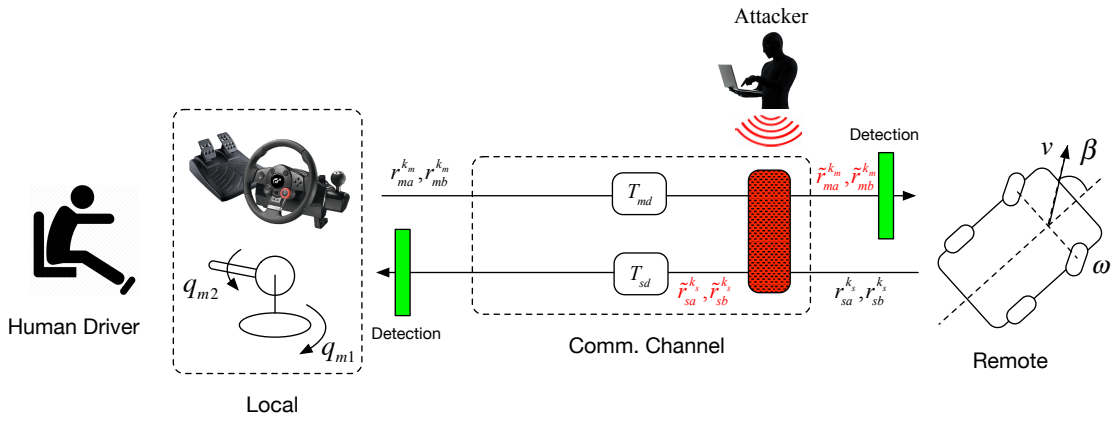


Figure 5.11: The attacker resides on the remote side.

5.4.3 Attack Detection on General Content Modification Attack

The detection scheme described in Section 5.3 is simulated for a general content modification attack. The network time delays are set as $T_{md} = T_{sd} = 0.2s$ in this case. In the encoding/decoding scheme, $\alpha = 1$, $\lambda_a = 2$ and $\lambda_b = 3$ are chosen. Here λ_a, λ_b are taken as constant, which is a special case for (2.4) and (2.5).

The attack is launched from $t = 6s$. As in Fig. 5.11, the attacker resides on the remote side, and the injection data on $r_{ma}^{k_m}, r_{mb}^{k_m}, r_{sa}^{k_s}, r_{sb}^{k_s}$ are designed as $\hat{r}_{sa}^{k_s} =$

$0.2[\cos(5t); \sin(5t)]$, $\hat{r}_{sb}^{k_s} = 0.2[\sin(5t); \cos(5t)]$, $\hat{r}_{ma}^{k_m} = 0.2[\sin(4t); \cos(4t)]$ and $\hat{r}_{mb}^{k_m} = 0.2[\sin(4t); \cos(4t)]$ for $t \geq 6s$.

The checking error e_i and its changing rate e_{id} are shown in Fig. 5.12. The threshold $\epsilon_i = 0.1$ is shown as dashed red line in the figure. In the first two subplots, the checking error exceeds the threshold ϵ_i after the attack is launched, which shows the attack is detected. In the second subplot, we can observe that the sudden jump of checking error is $0.2s$ delayed compared with first subplot, which reflects the network delay. In the last two subplots, the sharp spikes at the attack time point can be clearly observed, hence, the e_{id} can actually be used as an additional attack indication as discussed in **Remark 2.3.1**.

5.5 Tele-driving Experimental Platform Development

As part of the results, an initial tele-driving experimental platform is developed as shown in Fig. 5.13. In this system, a human driver uses a local Logitech G920 racing wheel with pedals to tele-drive a remote car-like robot called Elre-Rover developed by Elre Robotics. The human driver can send the control commands to remote car and get the video feedback in real time, where the data is transmitted through Wi-Fi using socket programming. The steering wheel SDK provided by Logitech is used to query the wheel/pedals state, and the control commands for the remote robot is implemented within ROS. This experimental platform can be used for future research on tele-driving.

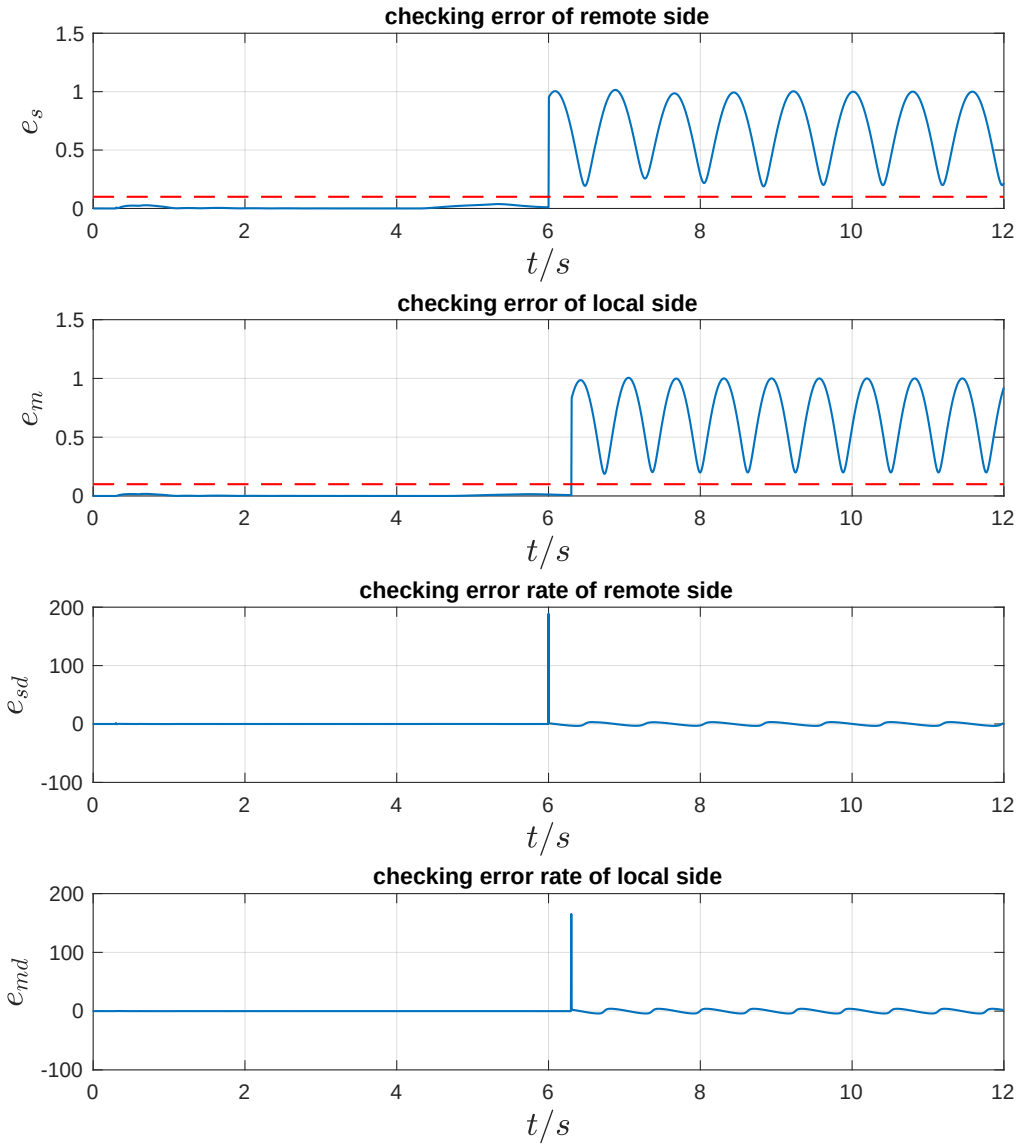


Figure 5.12: Checking error e_i and changing rate e_{id} under a general content modification attack with attack detection.

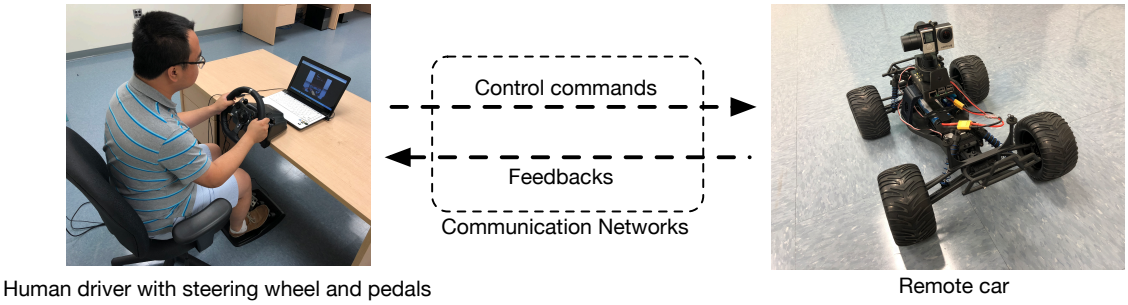


Figure 5.13: The developed tele-driving experimental platform.

5.6 Summary

In this chapter, a novel passivity-based adaptive coordination control scheme is proposed for a new bilateral tele-driving system in the presence of communication delays and dynamic parametric uncertainties. The proposed tele-driving scheme can achieve a new control mode and haptic feedback to better emulate normal car driving compared with the existing schemes in the mobile robot teleoperation such as [37–40]. Unlike the synchronization control scheme in [13, 26, 62] for traditional bilateral teleoperation system, the passivity assumption on the human and environment is replaced with an appropriately defined boundedness condition. For the proposed tele-driving system, the content modification attack design and detection are studied. All the proposed schemes are verified through simulations. An initial tele-driving experimental platform is also developed.

Chapter 6: Conclusion

The results of this dissertation can be briefly summarized as follows: By utilizing the physics of the robotic system, the content modification attacks are studied on a fundamental problem called synchronization for networked robotic systems including bilateral teleoperation system, multi-robot system and bilateral tele-driving system.

- To demonstrate the potential severity of the attack, a systematic methodology for designing destabilizing content modification attacks, termed as MCoMA, is proposed. The proposed design framework utilizes the system's physical storage function for constructing the attack.

- To defend the system, a physics-based attack detection scheme with an encoding-decoding structure is proposed to detect general content modification attacks. The proposed scheme is fast-response, distributed, computationally light and does not require the parameters of the system model. Additionally, the proposed scheme is independent of network delays because the encoding and decoding factors are updated asynchronously, and it can work under the normal network effects such as noises, packet drops and quantization errors with the relaxation on the detection condition such as (2.16).

- As part of the tele-driving system study, a novel passivity-based adaptive bilateral tele-driving control scheme is proposed in the presence of network delays and dynamic parametric uncertainties.

The proposed attack design and detection schemes for BTOS are verified on an experimental platform, and the proposed schemes for MAS and BTDS are verified through simulations.

This study explores the possibilities of utilizing the physics of the robotic system to better understand and strengthen the security of the networked robotic systems, which provides a different perspective for the cyber-physical security problem of CPS. Although the study is a meaningful exploration, the results do have some limitations, for instance, **(i)** the determination of threshold constant ϵ_i in (2.16) depends on the network conditions, which means the threshold might need to be recalibrated under different network conditions; **(ii)** the initial values of encoding-decoding factors in the proposed detection scheme have to be shared securely before the operation, which imposes a constraint on the scheme.

The cyber-physical security of CPS is anticipated to becoming increasingly important. The proposed work can be extended in several directions: **(i)** with only attack detection schemes discussed in this work, the attack mitigation methods for content modification attacks can be further explored, which can maintain the functioning of the networked robotic systems under attacks; **(ii)** the impact of other edge attacks (such as intentional delay attack and packet dropping attack) or node attacks on networked robotic systems can also be studied, and an attacker

with a mixed attack strategies is also worth investigations; **(iii)** other physics-based schemes are worth being further explored for CPS, which might provide a different and even more effective way for securing CPS.

Bibliography

[1] Alvaro A Cardenas, Saurabh Amin, and Shankar Sastry. Secure control: Towards survivable cyber-physical systems. In *Distributed Computing Systems Workshops, 2008. ICDCS'08. 28th International Conference on*, pages 495–500. IEEE, 2008.

[2] Clifford Neuman. Challenges in security for cyber-physical systems. In *DHS Workshop on Future Directions in Cyber-Physical Systems Security*, pages 22–24, 2009.

[3] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, and Shankar Sastry. Challenges for securing cyber physical systems. In *Workshop on future directions in cyber-physical systems security*, volume 5, 2009.

[4] Nicolas Falliere, Liam O Murchu, and Eric Chien. W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*, 5, 2011.

[5] Siobhan Gorman. Electricity grid in us penetrated by spies. *The Wall Street Journal*, 8, 2009.

[6] Kerry A McKay, Larry Bassham, Meltem Sönmez Turan, and Nicky Mouha. Report on lightweight cryptography. *NIST DRAFT NISTIR*, 8114, 2016.

[7] http://www.davincisurgery.com/da-vinci-surgery/da-vinci-surgical-system.

[8] Wang Wei and Yuan Kui. Teleoperated manipulator for leak detection of sealed radioactive sources. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 2, pages 1682–1687. IEEE, 2004.

[9] Woo-Keun Yoon, Toshihiko Goshozono, Hiroshi Kawabe, Masahiro Kinami, Yuichi Tsumaki, Masaru Uchiyama, Mitsushige Oda, et al. Model-based space robot teleoperation of ets-vii manipulator. *Robotics and Automation, IEEE Transactions on*, 20(3):602–612, 2004.

[10] Janez Funda and Richard P Paul. A symbolic teleoperator interface for time-delayed underwater robot manipulation. In *OCEANS'91. Ocean Technologies and Opportunities in the Pacific for the 90's. Proceedings.*, pages 1526–1533. IEEE, 1991.

[11] Sajeesh Kumar and Jacques Marescaux. *Telesurgery*. Springer Science & Business Media, 2008.

[12] Peter F Hokayem and Mark W Spong. Bilateral teleoperation: An historical survey. *Automatica*, 42(12):2035–2057, 2006.

[13] Takeshi Hatanaka, Nikhil Chopra, Masayuki Fujita, and Mark W Spong. Passivity-based control and estimation in networked robotics. *Communications and Control Engineering Series, Springer-Verlag*, 2015.

[14] Wei Ren and Randal W Beard. Decentralized scheme for spacecraft formation flying via the virtual structure approach. *Journal of Guidance, Control, and Dynamics*, 27(1):73–82, 2004.

[15] Dario Bauso, Laura Giarré, and Raffaele Pesenti. Attitude alignment of a team of uavs under decentralized information structure. In *Control Applications, 2003. CCA 2003. Proceedings of 2003 IEEE Conference on*, volume 1, pages 486–491. IEEE, 2003.

[16] Reza Olfati-Saber and Richard M Murray. Consensus protocols for networks of dynamic agents. In *Proceedings of the 2003 American Controls Conference*, 2003.

[17] J Alexander Fax and Richard M Murray. Information flow and cooperative control of vehicle formations. *Automatic Control, IEEE Transactions on*, 49(9):1465–1476, 2004.

[18] Reza Olfati-Saber and Richard M Murray. Consensus problems in networks of agents with switching topology and time-delays. *Automatic Control, IEEE Transactions on*, 49(9):1520–1533, 2004.

[19] Luc Moreau. Stability of multiagent systems with time-dependent communication links. *Automatic Control, IEEE Transactions on*, 50(2):169–182, 2005.

[20] Wei Ren, Randal W Beard, et al. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on automatic control*, 50(5):655–661, 2005.

[21] Yuko Hatano and Mehran Mesbahi. Agreement over random networks. *IEEE Transactions on Automatic Control*, 50(11):1867–1872, 2005.

[22] Nikhil Chopra and Mark W Spong. Passivity-based control of multi-agent systems. In *Advances in robot control*, pages 107–134. Springer, 2006.

[23] Wei Wang and Jean-Jacques E Slotine. Contraction analysis of time-delayed communications and group cooperation. *Automatic Control, IEEE Transactions on*, 51(4):712–717, 2006.

[24] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2508–2530, 2006.

[25] Akshay Kashyap, Tamer Başar, and Ramakrishnan Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.

[26] Nikhil Chopra and Mark W Spong. Output synchronization of nonlinear systems with relative degree one. In *Recent advances in learning and control*, pages 51–64. Springer, 2008.

[27] Pedram Hovareshti, John S Baras, and Vijay Gupta. Average consensus over small world networks: A probabilistic framework. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 375–380. IEEE, 2008.

[28] Ya Zhang and Yu-Ping Tian. Consentability and protocol design of multi-agent systems with stochastic switching topology. *Automatica*, 45(5):1195–1201, 2009.

[29] Tao Li and Ji-Feng Zhang. Mean square average-consensus under measurement noises and fixed topologies: Necessary and sufficient conditions. *Automatica*, 45(8):1929–1936, 2009.

[30] Hongkeun Kim, Hyungbo Shim, and Jin Heon Seo. Output consensus of heterogeneous uncertain linear multi-agent systems. *IEEE Transactions on Automatic Control*, 56(1):200–206, 2011.

[31] Herbert G Tanner, Ali Jadbabaie, and George J Pappas. Stable flocking of mobile agents, part i: Fixed topology. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 2, pages 2010–2015. IEEE, 2003.

[32] Reza Olfati Saber and Richard M Murray. Flocking with obstacle avoidance: cooperation with limited communication in mobile networks. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 2, pages 2022–2028. IEEE, 2003.

[33] Wei Ren and Ella Atkins. Distributed multi-vehicle coordinated control via local information exchange. *International Journal of Robust and Nonlinear Control*, 17(10-11):1002–1033, 2007.

[34] Wei Ren and Randal W Beard. Consensus algorithms for double-integrator dynamics. *Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications*, pages 77–104, 2008.

[35] Jiahu Qin, Wei Xing Zheng, and Huijun Gao. Coordination of multiple agents with double-integrator dynamics under generalized interaction topologies. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(1):44–57, 2012.

[36] Ruilin Liu, Kostas Bekris, Ahmed Elgammal, Vinod Ganapathy, Mario Gerla, Liviu Iftode, Melchi Michel, and Jingang Yi. Remote driving: A ready-to-go approach to autonomous car?-opportunities and challenges.

[37] Dongjun Lee, Oscar Martinez-Palafox, and Mark W Spong. Bilateral teleoperation of a wheeled mobile robot over delayed communication network. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3298–3303. IEEE, 2006.

[38] Dongjun Lee and Daye Xu. Feedback r-passivity of lagrangian systems for mobile robot teleoperation. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2118–2123. IEEE, 2011.

[39] Lei Ma, Zhihao Xu, and Klaus Schilling. Robust bilateral teleoperation of a car-like rover with communication delay. In *Control Conference (ECC), 2009 European*, pages 2337–2342. IEEE, 2009.

[40] Zhihao Xu, Lei Ma, and Klaus Schilling. Passive bilateral teleoperation of a car-like mobile robot. In *Control and Automation, 2009. MED'09. 17th Mediterranean Conference on*, pages 790–796. IEEE, 2009.

[41] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. Attack models and scenarios for networked control systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, pages 55–64. ACM, 2012.

[42] Saurabh Amin, Alvaro A Cárdenas, and S Shankar Sastry. Safe and secure networked control systems under denial-of-service attacks. In *International Workshop on Hybrid Systems: Computation and Control*, pages 31–45. Springer, 2009.

[43] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 911–918. IEEE, 2009.

[44] Yilin Mo and Bruno Sinopoli. False data injection attacks in control systems. In *Preprints of the 1st workshop on Secure Control Systems*, pages 1–6, 2010.

[45] Roy S Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *Control Systems, IEEE*, 35(1):82–92, 2015.

[46] Antonio Teixeira, Kin Cheong Sou, Henrik Sandberg, and Karl H Johansson. Secure control systems: A quantitative risk management approach. *Control Systems, IEEE*, 35(1):24–45, 2015.

[47] Quanyan Zhu and Tamer Basar. Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems. *Control Systems, IEEE*, 35(1):46–65, 2015.

[48] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *Control Systems, IEEE*, 35(1):93–109, 2015.

[49] Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *Control Systems, IEEE*, 35(1):110–127, 2015.

[50] Emeka Eyisi and Xenofon Koutsoukos. Energy-based attack detection in networked control systems. In *Proceedings of the 3rd international conference on High confidence networked systems*, pages 115–124. ACM, 2014.

[51] Shreyas Sundaram and Christoforos N Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, 56(7):1495–1508, 2011.

[52] Heath J LeBlanc, Haotian Zhang, Xenofon Koutsoukos, and Shreyas Sundaram. Resilient asymptotic consensus in robust networks. *IEEE Journal on Selected Areas in Communications*, 31(4):766–781, 2013.

[53] Zhi Feng and Guoqiang Hu. Distributed tracking control for multi-agent systems under two types of attacks. *IFAC Proceedings Volumes*, 47(3):5790–5795, 2014.

[54] Fabio Pasqualetti, Antonio Bicchi, and Francesco Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 57(1):90–104, 2012.

[55] André Teixeira, Iman Shames, Henrik Sandberg, and Karl H Johansson. Distributed fault detection and isolation resilient to network model uncertainties. *Cybernetics, IEEE Transactions on*, 44(11):2024–2037, 2014.

[56] Tamara Bonaci and Howard Jay Chizeck. Surgical telerobotics meets information security. In *21st Usenix Security Symposium*, 2012.

[57] Tamara Bonaci, Jeffrey Herron, Tariq Yusuf, Junjie Yan, Tadayoshi Kohno, and Howard Jay Chizeck. To make a robot secure: An experimental analysis of cyber security threats against teleoperated surgical robots. *arXiv preprint arXiv:1504.04339*, 2015.

[58] Gregory S Lee and Bhavani Thuraisingham. Cyberphysical systems security applied to telesurgical robotics. *Computer Standards & Interfaces*, 34(1):225–229, 2012.

[59] M Engin Tozal, Yongge Wang, Ehab Al-Shaer, Kamil Sarac, Bhavani Thuraisingham, and Bei-Tseng Chu. On secure and resilient telesurgery communications over unreliable networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pages 714–719. IEEE, 2011.

[60] Ioannis Sarras, Emmanuel Nuño, and Luis Basañez. An adaptive controller for nonlinear teleoperators with variable time-delays. *Journal of the Franklin Institute*, 351(10):4817–4837, 2014.

[61] Nikhil Chopra, Mark W Spong, and Rogelio Lozano. Synchronization of bilateral teleoperators with time delay. *Automatica*, 44(8):2142–2148, 2008.

[62] Hsin-Chen Hu and Yen-Chen Liu. Passivity-based control framework for task-space bilateral teleoperation with parametric uncertainty over unreliable networks. *ISA transactions*, 70:187–199, 2017.

[63] Yimeng Dong and Nikhil Chopra. Observability-based secure state encryption design for cyberphysical systems. In *Indian Control Conference (ICC). 2018*, pages 24–29. IEEE, 2018.

[64] Mark W Spong, Seth Hutchinson, and Mathukumalli Vidyasagar. *Robot modeling and control*, volume 3. Wiley New York, 2006.

[65] Emmanuel Nuño, Luis Basañez, Romeo Ortega, and Mark W Spong. Position tracking for non-linear teleoperators with variable time delay. *The International Journal of Robotics Research*, 28(7):895–910, 2009.

[66] Hassan K Khalil. *Nonlinear control*. Pearson Higher Ed, 2014.

[67] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.

[68] Reza Olfati-Saber, Alex Fax, and Richard M Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[69] AR Meenakshi and C Rajian. On a product of positive semidefinite matrices. *Linear algebra and its applications*, 295(1):3–6, 1999.

[70] Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1982.

[71] S Farokh Atashzar, Ilia G Polushin, and Rajni V Patel. Networked teleoperation with non-passive environment: Application to tele-rehabilitation. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 5125–5130. IEEE, 2012.

[72] Roongrote Wangkiet. Modeling of the dynamics for outdoor rovers. Master's thesis, Bayerische Julius- Maximilians-Universitat Wurzburg, 2008.