Cyber-Physical Attack Detection for Networked Control Systems

Seyed Eman Mousavinejad

B.Eng., M.Sc., MPhil.

School of Engineering and Built Environment
Griffith Sciences
Griffith University

Submitted in fulfillment of the requirements of the degree of
Doctor of Philosophy

January 2020
Statement of Originality

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part by another person for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text. This dissertation contains fewer than 45,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 55 figures.

Signed: 

Seyed Eman Mousavinejad
January 2020
This work is dedicated

to

My beautiful wife SAMIRA

&

My lovely parents MONIR & VALLY.
Until 1960s, control systems consisted mainly of mechanical or analog electronic devices exchanging information among system components, i.e., sensors, controllers, and actuators, via wired communication. However, recent advancement in computer and communication industries have led to the growing use of Internet, embedded systems, wireless and digital communication technologies in many industrial control systems and transformed them into Networked Control Systems (NCSs).

A defining feature of an NCS is that it consists of a number of devices implemented distributively so that system information is exchanged through a shared communication network. In light of many distinct advantages of NCSs including flexible architectures and less installation and maintenance costs, the development and application of NCSs have been recently boosted in a wide range of practical areas and critical infrastructures including transportation systems, electrical power systems and smart grids, remote surgery, industrial and manufacturing systems.

Owing to heterogeneous IT components and open network connections among controllers, sensors, actuators, and other networked components, the Confidentiality, Integrity, and Availability (CIA) of exchanged data in an NCS may suffer from vulnerability to malicious cyber attacks. Undoubtedly, this kind of threat is mainly launched by an adversary in either the physical world or the cyber-space with the aim of substantial economic benefits or disrupting human life. Therefore, it is imperative to properly address security issues of NCSs so as to ensure their reliable and safe performance.

In securing NCSs, reliable attack detection is of utmost importance. Generally speaking, when cyber attacks are detected and located in a timely fashion, the damage to overall systems can be controlled within a tolerable limit. Motivated by security concerns of NCSs, the first major contribution of this thesis is the development of a novel centralized detection method based on set-membership filtering technique so as to detect cyber attacks in an NCS subject to Unknown-But-Bounded (UBB) process noise and UBB measurement noise. In response to it, a set-membership filter is designed so as to construct two ellipsoidal sets: 1) a prediction set and 2) an estimation set. The estimation ellipsoidal set is calculated through updating the prediction ellipsoidal set with the current sensor measurement data. Whether the filter can detect the occurrence of such an attack is determined by the existence of intersection between these two sets.
The developed centralized detection method may not be straightforwardly applicable for a large-scale NCS because it requires full knowledge of the entire network information. Furthermore, the computational overhead for this detection method is quite high and hence, it may make the use of the detection system unrealistic. Therefore, the second major contribution of this thesis is the development of a distributed attack detection method for a vehicular platoon system, which is one of the large-scale NCSs from real engineering world. Moreover, two recovery mechanisms are developed to mitigate the adversarial impacts of attacks on the performance of the vehicle platooning system. With these two recovery mechanisms, the system can be brought back to the normal condition after detection of the attacks.

In some practical situations, it is quite common for a crafty adversary to launch assorted attacks of different models and strategies for comprehensively compromising the sensor measurements and control signals. It has been well acknowledged that different attack strategies are generally stealthy to any detection method. Motivated by this observation, the property of system’s resiliency is of utmost significance. In this study, the focus lies on resilient remote tracking control through a shared communication network. Thus, the third major contribution of this thesis is the analysis of the joint problem of resilient tracking control and resilient estimation in NCSs subject to the presence of various cyber attacks that are modeled in a unified framework which leads the NCS to be operated and controlled via some digital and unprotected communication networks.
I would like to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other during this incredible journey and without any of them, this research work would not have been possible.

First and foremost, I wish to place on records my sincere gratitude to my principal supervisor Prof. Emer. Ljubo Vlacic for his continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His valuable suggestions, comments and guidance helped me in all the stages of candidature including but not limited to design of the research projects, research methods, analyzing the results, and writing of this thesis. I would like to tell Ljubo “You have not been only a supervisor but also a kind and supportive friend whose help cannot be mentioned by only thankful words. I could not have imagined having a better supervisor and mentor throughout my Ph.D study.”

I am extremely grateful to my joint supervisors, A. Prof. Fuwen Yang and Prof. Qing-Long Han for their contributions of time and ideas to make my work productive and stimulating. I would like to thank them for encouraging me during my research and for supporting me to publish my research work on top-rank journals. Their important suggestions and advice are priceless. Thank you for your valuable time, co-operation, and generosity which set this work possible as it is till end. Your support has been the most profitable experience for me.

My heartfelt thanks is reserved for Dr. Xiaohua Ge for his invaluable insights and suggestions. I really appreciate his willingness to help me whenever I needed some clarification. I am amazed that even during his busy schedule, he was always willing to help me with my publications. His useful comments widen my knowledge in various fields of the subject throughout my Ph.D research.

Words cannot express the feelings I have for my parents for their constant unconditional support in my study. Mom and dad, you have been a constant source of strength and inspiration to me. Your prayer for me was what sustained me this far and put me through from many difficulties that I am facing for all these years.

Above all I would like to thank my better half, Samira who is not only my dearest and passionate wife but also my very supportive and close friend. I would like to say “Thank you, my love, for your moral support, constant love, and being so patient with me during this journey. There are not enough words in the world to express my feelings for you, so I will just say: I love You.”
Journal publications during candidature


Conference publications during candidature


It is worth mentioning that

- Impact factor of IEEE Transactions on Cybernetics is 10.3.
- Impact factor of IEEE Transactions on Intelligent Transportation Systems is 5.7.
- Paper [1] has attracted 21 citations although it was published just 16 months ago.
- Paper [2] has attracted 1 citation although it was published just 3 months ago.
1 Introduction and outline
  1.1 What is an NCS? ............................................. 1
  1.2 Real-world cyber incidents .............................. 2
  1.3 Security properties in NCSs ............................ 3
  1.4 Cyber attacks in NCSs ................................... 4
    1.4.1 DoS attacks ......................................... 4
    1.4.2 Deception attacks .................................. 4
  1.5 Attack detection strategies for NCSs ................. 5
  1.6 Resilient control and estimation for NCSs ........... 6
  1.7 Aim of the research .................................... 7
  1.8 Thesis outline ......................................... 9
  1.9 Original contributions ................................ 10

2 A novel cyber attack detection method .................. 12
  2.1 Abstract ................................................. 13
  2.2 Introduction ........................................... 13
  2.3 Problem statement and formulation .................... 17
    2.3.1 Prediction step .................................... 20
    2.3.2 Measurement update step .......................... 21
  2.4 Attack detection using set-membership filtering .... 22
    2.4.1 Design of the prediction ellipsoidal set .......... 22
    2.4.2 Update on prediction ellipsoidal set with current measurement 24
    2.4.3 Recursive algorithm for attack diagnosis .......... 27
  2.5 Case studies .............................................. 29
    2.5.1 Test system ........................................ 29
    2.5.2 Cyber attack model ................................ 31
### 3 Distributed cyber attack detection

3.1 Abstract ......................................................... 40
3.2 Introduction ................................................... 40
3.3 Problem formulation ............................................. 45
   3.3.1 CACC law ................................................. 45
   3.3.2 Cyber attack model ....................................... 49
   3.3.3 Set-membership state estimation problem ................. 50
   3.3.4 Cyber attack detection criteria .......................... 51
3.4 Distributed attack detection and recovery mechanism design .... 53
   3.4.1 Design criteria .......................................... 53
   3.4.2 Recursive convex optimization and cyber attack detection algo-
        rithm with recovery mechanism ............................ 54
3.5 Simulation results and case studies ................................ 58
   3.5.1 Simulation setup ......................................... 58
   3.5.2 DoS attack on communication network .................... 59
   3.5.3 Replay attack on sensor data ............................. 61
   3.5.4 Message falsification attack on sensor data ............... 64
   3.5.5 String stability analysis under different attack locations .. 65
3.6 Conclusion .................................................... 67

### 4 Resilient control and estimation

4.1 Abstract ......................................................... 70
4.2 Introduction ................................................... 70
4.3 Problem formulation ............................................. 73
   4.3.1 The system model ......................................... 74
   4.3.2 Data quantization ......................................... 75
   4.3.3 Cyber attacks models ..................................... 77
   4.3.4 The resilient observer-based tracking control protocol ..... 80
   4.3.5 The resilient set-membership tracking control problem ..... 81
4.4 Main results .................................................... 84
   4.4.1 Design criteria .......................................... 84
List of Figures

1.1 Different data properties compromised by an attack on communication channels of an NCS. ............................................. 3

2.1 Some possible attack points to a networked control system. .......... 18

2.2 Vehicle tracking system with GPS. ........................................ 30

2.3 Intersection between the prediction ellipsoidal set $\chi_{k+1|k}$ and the updated estimation ellipsoidal set $\chi_{k+1}$ at (a) $x_1 - x_2$ phase-plane (b) $x_3 - x_4$ phase-plane ("1" indicates there exists intersection, "0" indicates there is no intersection). .......................................................... 33

2.4 Intersection between the prediction ellipsoidal set $\chi_{k+1|k}$ and the updated estimation ellipsoidal set $\chi_{k+1}$ at $x_1 - x_2$ phase-plane. ..................................................... 34

2.5 The prediction ellipsoid $\chi_{k+1|k}$ (green, dotted line), the updated estimation ellipsoid $\chi_{k+1}$ (blue, solid line) (a) step 29 (b) step 30 (c) step 36 (the updated estimation ellipsoid at step 30 and 36 are magnified with the ratio of $10^7$ and 10, respectively). .......................................................... 35

2.6 Intersection between the prediction ellipsoidal set $\chi_{k+1|k}$ and the previous updated estimation ellipsoid set $\chi_k$ at $x_3 - x_4$ phase-plane. ..................................................... 36

2.7 The prediction ellipsoid $\chi_{k+1|k}$ (green, dotted line), the updated estimation ellipsoid $\chi_k$ (blue, solid line) (a) step 19 (b) step 20 (c) step 36. ............................................................. 37

3.1 A platoon-based vehicular control system with some possible attack points. ................................................................. 45

3.2 Schematic of distributed attack detection with recovery mechanisms for vehicle $i$ under set-membership filtering technique. ............................................ 57

3.3 (a) Intersection between $\chi_{k+1|k}^2$ and $\chi_k^2$ ("1" indicates there exists intersection, "0" indicates there is no intersection); (b) $\chi_{111|110}^2$ (green, dotted line), $\chi_{110}^2$ (blue, solid line), true state (red, star). ........................................... 59

3.4 (a) and (b) Vehicle’s acceleration without and with recovery mechanisms; (c) and (d) Inter-vehicle distance without and with recovery mechanisms; (e) String stability analysis. .................................................. 61
3.5 (a) Intersection between $\chi_{k+1|k}^2$ and $\chi_{k+1}^2$; (b) $\chi_{105|104}^2$ (green, dotted line), $\chi_{105}^2$ (blue, solid line, magnified with the ratio of $10^2$), true state (red, star). .................................................. 62

3.6 (a) and (b) Vehicle’s velocity without and with recovery mechanisms; (c) and (d) Inter-vehicle distance without and with recovery mechanisms; (e) String stability analysis. ............................................................. 63

3.7 (a) Intersection between $\chi_{k+1|k}^3$ and $\chi_{k+1}^3$; (b) $\chi_{80|79}^3$ (green, dotted line), $\chi_{80}^2$ (blue, solid line, magnified with the ratio of 10), true state (red, star). 64

3.8 (a) and (b) Relative velocity without and with recovery mechanisms; (c) and (d) Inter-vehicle distance without and with recovery mechanisms; (e) String stability analysis. ............................................................. 66

3.9 Stability analysis under DoS attack on different communication channels. 67

4.1 Schematic of resilient observer-based tracking control over shared communication networks subject to cyber attacks. ................................. 74

4.2 Schematic of the ITTS with the developed resilient observer-based tracking control strategy. ................................................................. 89

4.3 Non-resilient tracking control and estimation performance on liquid levels (a) $h_1$; (b) $h_2$; and (c) $h_3$. .................................................. 91

4.4 (a) Non-resilient tracking error. (b) Non-resilient estimation error. 92

4.5 Resilient tracking control and estimation performance on liquid levels (a) $h_1$; (b) $h_2$; and (c) $h_3$. .................................................. 92

4.6 (a) Resilient tracking error. (b) Resilient estimation error. 93
Abbreviations

**ACC** Adaptive Cruise Control.

**ACSC** Australian Cyber Security Center.

**BD** Bidirectional.

**BDL** Bidirectional Leader.

**CACC** Cooperative Adaptive Cruise Control.

**CIA** Confidentiality, Integrity and Availability.

**CP** Controller-to-Plant.

**CTH** Constant Time Headway.

**DoS** Denial of Service.

**FDI** False Data Injection.

**GPS** Global Positioning System.

**IFT** Information Flow Topology.

**ITS** Intelligent Transportation Systems.

**ITTS** Internet-based Three-Tank System.

**KL** Kullback–Leibler.

**LQG** Linear Quadratic Gaussian.

**NCS** Networked Control System.

**NIST** National Institute of Standards and Technology.

**OP** Optimization Problem.

**PD** Proportional-Derivative.
Abbreviations

**PF** Predecessor Following.

**PFL** Predecessor-Following Leader.

**PLC** Programmable Logic Controller.

**RLMI** Recursive Linear Matrix Inequality.

**SC** Sensor-to-Controller.

**TPF** Two Predecessor-Following.

**TPFL** Two Predecessor-Following Leader.

**TTS** Three-Tank System.

**UBB** Unknown-But-Bounded.

**VANET** Vehicular Ad-hoc Network.

**WLS** Weighted Least Square.
Nomenclature

A  System state matrix.
\( \hat{A} \)  Prediction filter gain.
\( a_i \)  Absolute acceleration of vehicle \( i \).
B  Input matrix.
\( \hat{B} \)  Estimation filter gain.
\( B_{c,i} \)  Input matrix from vehicle \( i - 1 \).
\( B_{s,i} \)  Input matrix of vehicle \( i \).
\( B_w \)  Process noise matrix.
C  Output matrix.
D  Measurement noise matrix.
\( d_i \)  Inter-vehicle distance.
\( h_d \)  Headway-time for vehicle.
i  Index of vehicle in platoon.
k  Time instant.
M  Controller gain.
\( P_k \)  Shape matrix of estimation ellipsoid.
\( P_{k+1|k} \)  Shape matrix of prediction ellipsoid.
\( q_i \)  Absolute position of vehicle \( i \).
R  Shape matrix of reference ellipsoid.
u  System input.
u_{fb}  Measurement-based feedback component.
u_{ff}  Communication-based forward component.
Nomenclature

$u^r$ Reference acceleration profile.

$V$ Shape matrix of measurement noise ellipsoid.

$v_i$ Absolute velocity of vehicle $i$.

$w$ Process noise.

$W$ Shape matrix of process noise ellipsoid.

$x$ System state.

$\hat{x}_k$ Estimation of the state.

$\hat{x}_{k+1|k}$ Prediction of the state.

$y$ Measurement output.

$\mathcal{T}_{k+1}$ Reference ellipsoidal set.

$\mathcal{X}_k$ Estimation ellipsoidal set.

$\mathcal{X}_{k+1|k}$ Prediction ellipsoidal set.

$\xi$ Internal actuator dynamics parameter of vehicle.

$\mu_{t,0}$ Quantization scaling constant.

$\nu$ Measurement noise.

$\omega_c$ Bandwidth of controller.

$\rho_{\ell}$ Quantization density.

$\tau_s$ Sampling period.
This thesis consists of five chapters: a general introduction in Chapter 1, three results chapters including the published journal papers in Chapter 2, Chapter 3, and Chapter 4, and the thesis summary and conclusions in Chapter 5. The conference publications have not been completely included in this thesis due to their minor contributions, however, the summary of each conference paper has been discussed at the conclusion section of its relevant chapter.

The results chapters are in the form of manuscripts formatted to meet the requirement of the peer-reviewed academic journals where they were published. The thesis has been done in accordance with Griffith University policy (Appendix C).

My contribution to each co-authored paper is outlined at the front of the relevant chapter. The corresponding literature review has been conducted in each relevant chapter. However, references are collected at the end of this thesis in order to avoid repetition of the identical references.

The bibliographic details of the co-authored papers, including all authors, are:

- **Chapter 2:**
  - Mousavinejad, E., F. Yang, Q.-L. Han, and L. Vlacic, “Cyber-physical attacks detection in networked control systems with limited communication bandwidth”, *ANZCC 2017*, pp. 53–58. DOI: 10.1109/ANZCC.2017.8298484.

- **Chapter 3:**

- **Chapter 4:**


According to the IEEE copyright policy, the IEEE does not require individuals working on a thesis to obtain a formal reuse license. For further information, refer to Appendix D.
CHAPTER 1

Introduction and outline

Until 1960s, control systems consisted mainly of mechanical or analog electronic devices exchanging information among system components, i.e., sensors, controllers, and actuators, via wired communication. However, recent advancement in computer and communication industries have led to the growing use of Internet, embedded systems, wireless and digital communication technologies in many industrial control systems and transformed them into NCSs.

1.1 What is an NCS?

A defining feature of an NCS is that it consists of a number of devices implemented distributively so that system information is exchanged through a shared communication network. In light of many distinct advantages of NCSs including flexible architectures and less installation and maintenance costs, the development and application of NCSs have been recently boosted in a wide range of practical areas and critical infrastructures including transportation systems, electrical power systems and smart grids, remote surgery, industrial and manufacturing systems, and so on [1].

However, owing to heterogeneous IT components and open network connections among system components, NCSs become insecure and vulnerable to cyber attacks which can maliciously destroy the information transmitted through common communication setting. Undoubtedly, this kind of threat is mainly launched by an adversary in either the physical world or the cyber-space with the aim of substantial economic benefits or disrupting human life. Therefore, over the past few decades, considerable effort has been devoted to enhance the security of NCSs against cyber-physical attacks. For example, the Australian Cyber Security Center (ACSC) [2] and the National Institute of Standards and Technology (NIST) in the U.S. [3] have issued even a guide to industrial control system security.
1.2 Real-world cyber incidents

Security enhancement of NCSs relies on reviewing recent cyber incidents in real-world infrastructures. Investigation of cyber incidents helps to understand the vulnerable points of the real-world systems and how they might be subject to cyber attacks. Therefore, it provides us with a general idea of how the attacks have been launched recently so that detection algorithms and protection methods can be implemented to avoid future attacks. The following presents some representative examples of cyber attacks in real-world NCSs.

1. **Maroochy water breach** [4]: In 2000, an ex-employee used his laptop computer and a radio transmitter so as to take control of 150 sewage pumping stations in Maroochy Shire, Queensland, Australia. During his attack for a period of three months, one million liters of untreated sewage were released into a storm-water drain through which local waterways were contaminated.

2. **Slammer worm crashed Ohio nuke plant network** [5]: In 2003, an attacker exposed a private computer network at Ohio’s Davis-Besse power plant to a Slammer worm and disabled a safety monitoring system for about five hours although the network was protected by a firewall. The Slammer worm spread from the network to the systems controlling the power plant. It was reported that this attack caused serious trouble to system operators as the human-machine interface and the plant process computers had crashed for hours.

3. **Electricity grid in U.S. penetrated by spies** [6]: In 2009, cyber spies entered into the U.S. electric power grid and spread a software program that could be used to disrupt the system. Therefore, it has raised a big concern about the security of electric power grids since disrupting power systems might cause substantial damage on economic losses and even human life.

4. **Stuxnet virus** [7]: Stuxnet is a computer worm that was initially written to target Iranian nuclear centrifuges in 2010. The final goal of this worm is to hack industrial control systems and modify the programs implemented on PLCs to change their responses to that the attacker intended and to hide those changes from system operators. Stuxnet was applied to a computer network through a removable drive infected with the worm.
1.3 Security properties in NCSs

There exists a vast literature on computer security mentioning three fundamental properties on data and IT services, namely confidentiality, integrity, and availability [9]. Confidentiality relates to protecting the information from disclosure to unauthorized parties. Integrity concerns the trustworthiness of data, referring to protection of information from being modified by unauthorized parties. Availability considers that authorized parties are able to access the information when needed. Figure 1.1 demonstrates how these properties are compromised by an adversary. In three following cases, a signal $y_k$ containing the measurement data of a local physical plant is transmitted to a remote controller via a shared communication network. Since the transmitted data are private information, the sender and receiver must only know their contents. In Figure 1.1a, an adversary is able to hack the communication channels so as to illegally access the private information. Therefore, data confidentiality is compromised. In Figure 1.1b, an adversary succeeds to modify the content of the signal through sending a false signal into the communication channels and adding some bogus information to the original signal. Hence, data integrity is violated. Finally, in Figure 1.1c, an
adversary has this ability to completely block the communication channels so that the information does not reach the receiver. As a result, data availability is compromised.

1.4 Cyber attacks in NCSs

Depending on which one of the above-mentioned data properties is compromised, the existing attacks triggered through unreliable communication networks can be arguably categorized into two main types: 1) Denial-of-Service (DoS) attacks and 2) deception attacks.

1.4.1 DoS attacks

DoS attacks are aimed at deteriorating the availability of the system information, usually either control commands or sensor data, by jamming the communication channels and therefore, preventing information exchange between components of NCSs. From a technological perspective, an adversary can perpetrate DoS attacks through disrupting the radio frequencies on wireless communication channels, which leads to congestion of those channels. Since an adversary requires little prior knowledge about the system and does not need any special devices to launch DoS attacks, this type of attack can seriously threaten an NCS’s performance.

1.4.2 Deception attacks

Deception attacks represent a kind of attempt to violate the confidentiality and integrity of the system information. Through the deception attacks, an adversary tries to manipulate the system into following his desired behavior via accessing disclosure resources and injecting deception information into either sensor data or control commands. Two typical types of deception attacks on NCSs are False Data Injection (FDI) attacks and replay attacks.

In FDI attacks, an adversary starts collecting the information transmitted through the wireless medium and simultaneously modifies the content meaningfully such that he can severely damage the stability of NCSs by rebroadcasting the bogus information. Replay attacks, however, do not require prior knowledge of the system components. It can be easily implemented by an adversary in two phases. In the first phase, which is known as disclosure phase, the adversary overhears and stores the packets transmitted over communication channels without injecting any malicious input to the system. In
the second phase, the adversary manipulates the system by replaying the recorded data as if they are new packets received from those tampered channels. Even though the content of the data packets is not falsified by the replay attacks, any outdated information may mislead the controller, having a great potential to destroy stability of NCSs that may lead to hazardous effects on the performance of real-world critical infrastructures.

1.5 Attack detection strategies for NCSs

Notice that the purpose of an adversary is to carefully design an attack through which he can severely destroy the desired performance of safety-critical NCSs. However, when attacks are detected and located in a timely fashion, the damage to overall systems can be controlled in a tolerable limit. From such a perspective, reliable attack detection is of utmost importance in securing NCSs and maintaining the healthy performance of such systems.

Up to date, considerable effort has been devoted to attack detection by means of different approaches, which can be arguably divided into three categories: 1) Bayesian detection with binary hypothesis; 2) Weighted Least Square (WLS) approaches; and 3) $\chi^2$-detector based on Kalman filters.

1. **Bayesian detection with binary hypothesis**: In Bayesian detection, attack detection is usually formulated as a binary hypothesis test with prior probability. With the help of such a test, the performance limit of collaborative spectrum sensing is investigated in [10]. The closed form expression is proposed in [11] for an optimal attacking strategy of the Byzantines. Similar results can be found in [12] and [13], and the references therein. Recently, intrusion detection issues have been addressed by developing a novel Bayesian game-theoretic framework [14], [15]. Such a method is useful when players do not have complete information on other players.

2. **WLS approach**: For steady measurement data, WLS approach is a reliable and efficient scheme for the defense of attacks and therefore it is widely employed to obtain the gain of various detectors, such as generalized likelihood ratio detectors [16], detectors based on the nuclear norm minimization and the low rank matrix factorization [17], and a heuristic approach for detecting abrupt changes in the system outputs in light of the singular value decomposition [18].
The WLS approach is a residual-based framework where the measurement residual is calculated by some weighted least-square observers so that it can be compared with a predetermined threshold in order to determine whether there exists a bad measurement.

3. \(\chi^2\)-detector based on Kalman filters: Different from the WLS approach, a \(\chi^2\)-detector employs a Kalman filter to generate a residual signal. This detection approach is suitable to be used in the presence of both bad and false data (such as DoS attacks, short-term/long-term random attacks) and therefore, it has been extensively studied recently. For instance, in [19], a method to detect a replay attack on sensors of a control system is introduced. The proposed method requires the usage of an LQG controller and a Kalman filter with \(\chi^2\)-detector. In the proposed method, the control signal is redesigned by adding a zero-mean authentication signal in order to increase the sensitivity of \(\chi^2\)-detector to a cyber attack. In [20], a Kalman filtering technique is utilized to estimate the state of the system based on the report from sensor measurement and the past state values. Also, an additional detector based on Euclidean distance along with \(\chi^2\)-detector is proposed to detect a false data injection attack on smart grid systems. A method in which KL divergence, a Kalman filter, and an LQG controller are adopted to design an optimal Neyman–Pearson detector for replay and covert attacks is proposed in [21]. In [22], a cosine similarity matching-based approach to design a detector is presented. The detector incorporate the Kalman filter estimation to measure any deviation between actual measurements and estimated values.

1.6 Resilient control and estimation for NCSs

Motivated by security concerns of NCSs, the property of system’s resiliency, either in state estimation or various control tasks, is of utmost significance. Generally, resiliency refers to the ability of restoration of a system after being corrupted by unexpected adversarial attacks.

In the context of better understanding system dynamical behaviors and developing some specific control strategies, state estimation plays a crucial role specially when full information of the system’s state measured by existing sensors is not implicitly trusted in the presence of attacks. Therefore, resilient estimation has been deemed
as an effective means to guarantee a reliable estimation of the system’s state in the simultaneous presence of noises and attacks, and thus has been intensively studied in the field of NCSs. For some latest results on the problem of resilient estimation in NCSs, readers are referred to the survey paper [23] and many references therein.

In addition to the resilient state estimation, the threat from stealthy attackers needs to be mitigated via various resilient control strategies. During the past two decades, the problem of guaranteeing resilient NCSs has attracted considerable interest from various perspectives and there are some recent results available in [24]–[32]. However, in the context of resilient tracking control, there exist only a few studies in the literature. For example, in [33], the design of a specific deception attack, FDI attack, violating both feedback and forward channels is considered in the sense that the attack can heavily destroy the tracking performance without being detected. The tracking problem of heterogeneous dynamical networks in the presence of malicious cyber attacks is investigated in [34]. The attacker therein launches asynchronous attacks on both controllers’ and observers’ communication channels. Their proposed algorithm gives the sufficient conditions on the length of attacking intervals in order to achieve the tracking objective. The distributed tracking control problem of multi-agent systems under a strategic DoS attack whose dynamics are modeled by a random Markov process is addressed in [35]. Based on a hybrid stochastic secure control framework, they developed a distributed resilient control law to achieve exponential consensus tracking in mean square sense. In [36], the tracking control of mobile robots in the presence of DoS attacks is studied. To ensure both the tracking convergence and efficient usage of communication resources, a general hybrid model consisting of an event-triggering strategy and DoS attacks is established.

1.7 Aim of the research

Up until now, most attack detection approaches by means of the state estimation in the presence of noises can be arguably divided into two categories depending on the noise model: 1) stochastic noises (e.g., Gaussian noise) and 2) bounded non-stochastic noises (e.g., \(l_2\)-norm energy-bounded noise).

It should be pointed out that modeling the noise in a stochastic framework usually requires accurate statistical properties of the noise such as known mean and covariance to describe the state distributions modeled as random variables, which is conservative for some practical applications when the noise is unknown. On the other hand, in an
energy-bounded noise model, e.g., in the $H_\infty$ sense, it is assumed that the noise has norm-bounded energy so as to satisfy the aim of finding the worst-case solution to the estimation problem. However, in many real-world applications, such as target tracking and attack, system guidance and navigation, 100% confidence in the state estimation is of paramount importance.

The objective of a conventional estimation problem, e.g., the Kalman filtering technique, is to compute a single vector estimation with regard to the system’s state such that the estimation error asymptotically converges zero. However, such an estimation approach only provides a pointwise estimation of the system’s state. Therefore, there is no guarantee that the distribution of the system’s state can be achieved within a confidence region where all true states of the system may reside in, specially when there exist unpredictable environmental changes due to the influences of cyber attacks and noises. Therefore, it is much more appropriate to model the state distributions in certain sets considering UBB noises. As a result, an alternative method called set-membership estimation is developed in the literature. The core idea of this method is to calculate a bounding ellipsoidal set in state-space, which always encloses the true state of the system by assuming UBB noise signals. Indeed, an assumption of UBB noise eliminates the requirement of prior knowledge of the accurate statistical characteristics of noise because only the knowledge of a bound on the realization is needed.

Motivated by the discussion above, the first outcome of this research is the development of a novel centralized method based on set-membership filtering so as to detect cyber attacks in an NCS subject to UBB process noise and UBB measurement noise. The designed set-membership filter constructs two ellipsoidal sets: 1) a prediction set and 2) an estimation set. The estimation ellipsoidal set is calculated through updating the prediction ellipsoidal set with the current sensor measurement data. Whether the filter can detect the occurrence of such an attack is determined by the existence of intersection between these two sets.

The developed centralized detection method is not straightforwardly applicable for a large-scale NCS because it requires full knowledge of the entire network information. Furthermore, the computational overhead for this detection method is quite high and hence it may make the use of the detection system unrealistic. Therefore, the second outcome of this research is the development of a distributed attack detection method for a vehicular platoon system, which can represent a large-scale NCS. Moreover, two recovery mechanisms are developed to mitigate the adversarial impacts of attacks on
the performance of the vehicle platooning system. With these two recovery mechanisms, the system can be brought back to the normal condition after detection of the attacks. In other words, the recovery mechanisms make the state estimation secure against attacks and then the controller uses the secure estimation to generate the control signal.

Furthermore, it should be pointed out that the existing literature on resilient tracking control and estimation methods concentrates primarily on one specific type of attack, and relatively few results deal with the tracking control and estimation problems in the presence of various attacks on communication channels. In some practical situations, it is quite common for a crafty adversary to launch assorted attacks of different models and strategies so that he could heavily disrupt the tracking and estimation performance. Thus, the third outcome of this research is to address the joint problem of resilient tracking control and resilient estimation in an NCS which is operated and controlled via some digital and unprotected communication networks whose channels are subject to various cyber attacks that are modeled in a unified framework.

1.8 Thesis outline

The following briefly outlines the content of each chapter.

Chapter 2 provides a novel cyber attack detection method in a centralized framework for an NCS. The developed detection method includes two steps: 1) a prediction step and 2) a measurement update step. In these two steps, the goal is to find an estimation set through updating the prediction set with the one yielding from the current measurement. In the measurement update step, the unconstrained state estimate is projected onto the linear measurement (output) constrained surface to obtain the updated estimation set by using Finsler’s lemma. A cyber attack on sensors or a network exchanging data between sensors and controllers can be detected if there is no intersection between the estimation set updated at the current time instant and the prediction set, and a cyber attack on actuators or a network transmitting data between controllers and actuators can be detected if the prediction set has no intersection with the estimation set updated at the previous time instant. The proposed method is able to distinguish attacks on control signals from those on measurement outputs.

Chapter 3 presents a distributed cyber attack detection method for a vehicular platoon system. The platoon of vehicles can be regarded as a multi-agent NCS in a large-scale framework where vehicles represent mobile agents connected together via a
Vehicular Ad-hoc Network (VANET). First, a set-membership filtering technique in a distributed framework has been developed such that each vehicle can determine a group of confidence ellipsoids and each vehicle’s true state always resides in bounding ellipsoidal sets regardless of UBB process noise and measurement noise, providing the vehicle is free of any attack. From such a perspective, a refined attack detection method has been proposed to discern when the occurrence of an attack can be detected and alarmed. Second, two recovery mechanisms have been introduced into the proposed detection algorithm so that each ellipsoidal set will adopt the compensated state prediction and/or state estimation to increase the resilience of estimation system. Based on the proposed recovery mechanisms, the controllers in the platoon are able to alleviate the adversarial effect of an attack through accessing the secure state estimation.

Chapter 4 investigates the problem of resilient tracking control in an NCS subject to various cyber attacks and data quantization. First, a resilient set-membership tracking control method has been developed to ensure that the true state of the system is guaranteed to be always included in a bounding ellipsoidal set of the reference state in the simultaneous presence of UBB process and measurement noises, and two various attacks including DoS attacks and deception attacks launched on Controller-to-Plant (CP) and Sensor-to-Controller (SC) communication channels, respectively. Second, in the case that full information of the system’s state is not implicitly trusted in the presence of attacks, a resilient set-membership estimation strategy has been provided to secure the state estimates against those attacks.

Finally, conclusions and directions for future research are outlined in Chapter 5.

1.9 Original contributions

- The first major contribution of this research is the development of a novel attack detection method based on a set-membership filtering technique. To achieve this, in Chapter 2, an ellipsoidal set-membership state estimate algorithm has been designed, which recursively computes the state of a general linear time-varying system in two steps: 1) a prediction step and 2) a measurement update step. The aim of the two-step state estimate algorithm is to find a group of confidence prediction ellipsoidal sets and estimation ellipsoidal sets so that the true state of the system is always enclosed by these sets regardless of UBB process noise and measurement noise, provided that the system is free of any attack. Therefore,
these two ellipsoidal sets will always have intersection as they both contain the
true state of the attack-free system. However, if there is an attack on the system,
one of the two sets may not include the true state since the center of that set
is affected by the attack. Thus, one can conclude that there must be an attack
compromising the system performance if there is no intersection between the two
ellipsoidal sets.

• The second major contribution of this research is the implementation of the
developed novel attack detection in a fully distributed framework into a large-scale
vehicular platoon system. Since the developed attack detection method presented
in Chapter 2 follows a centralized framework, it is not straightforwardly applicable
for a large-scale vehicular platoon system due to its requirement of full knowledge
of the entire platoon information which leads to the high computational overhead.
To overcome this limitation, in Chapter 3, a novel attack detection method has
been developed in a fully distributed manner in the sense that each vehicle is
equipped with its own detection system with no need of having full knowledge of
the entire platoon. Furthermore, two recovery mechanisms have been added into
the distributed attack detection algorithm so as to mitigate the malicious effects
of attacks. The developed recovery mechanisms make the state estimation secure
against attacks and then the controller uses the secure estimation to generate
the control signal so that the string stability of the platoon is satisfied while the
attack is present.

• The third major contribution of this research is the development of a resilient
tracking control method for an NCS subject to UBB process noise and measure-
ment noise, limited communication capacity, and two various attacks including
DoS attacks and deception attacks launched on CP and SC communication
channels, respectively. In Chapter 4, first, a resilient set-membership tracking
control protocol has been developed, through which the system’s true state is
guaranteed to reside in a bounding ellipsoidal set of the reference state regardless
of the existence of attacks and UBB noises. Second, in the case that full infor-
mation of the system’s state is not implicitly trusted in the presence of attacks,
a resilient set-membership estimation strategy has been provided to secure the
state estimates against attacks.
A novel cyber attack detection method

In this chapter, a novel cyber attack detection in a centralized framework has been designed for an NCS. The core of the developed detection strategy lies in a set-membership ellipsoidal estimation method which consists of two steps: 1) a prediction step where a prediction ellipsoidal set is calculated and 2) a measurement update step where an estimation ellipsoidal set is obtained through updating the prediction ellipsoidal set with the current sensor measurement data. Whether the occurrence of a cyber attack can be detected is determined by the existence of intersection between these two sets.

This chapter includes the full published version of a co-authored journal paper and a summary of a co-authored conference paper. The bibliographic details of the co-authored papers, including all authors, are:


Mousavinejad, E., F. Yang, Q.-L. Han, and L. Vlacic, “Cyber-physical attacks detection in networked control systems with limited communication bandwidth”, *ANZCC 2017*, pp. 53–58. DOI: [10.1109/ANZCC.2017.8298484](https://doi.org/10.1109/ANZCC.2017.8298484).

My contribution to the papers involved: literature review, problem formulation, design, model simulation, model testing, analysis of simulation results, drafting of concluding remarks, manuscript writing and editing.

Signed:  
PhD Candidate: Seyed Eman Mousavinejad (Principal author)  
Date: 03/01/2020

Countersigned:  
Principal Supervisor: Prof. Emer. Ljubo Vlacic  
Date: 03/01/2020

Countersigned:  
Principal Supervisor: A. Prof. Fuwen Yang  
Date: 03/01/2020

Countersigned:  
Associate Supervisor: Prof. Qing Long Han (Corresponding author)  
Date: 03/01/2020
2.1 Abstract

This paper is concerned with cyber attack detection in an NCS. A novel cyber attack detection method, which consists of two steps: 1) a prediction step and 2) a measurement update step, is developed. An estimation ellipsoidal set is calculated through updating the prediction ellipsoidal set with the current sensor measurement data. Based on the intersection between these two ellipsoidal sets, two criteria are provided to detect cyber attacks injecting malicious signals into physical components (i.e., sensors and actuators) or into a communication network through which information among physical components is transmitted. There exists a cyber attack on sensors or a network exchanging data between sensors and controllers if there is no intersection between the prediction set and the estimation set updated at the current time instant. Actuators or network transmitting data between controllers and actuators are under a cyber attack if the prediction set has no intersection with the estimation set updated at the previous time instant. Recursive algorithms for the calculation of the two ellipsoidal sets and for the attack detection on physical components and the communication network are proposed. Simulation results for two types of cyber attacks, namely a replay attack and a bias injection attack, are provided to demonstrate the effectiveness of the proposed method.

2.2 Introduction

Nowadays, NCSs have been playing a crucial role in many critical infrastructures including power grids, transportation systems, robotic platforms, target tracking, and system guidance and navigation [1], [37]–[42]. Due to physical constraints or technological limitations, exchanged data among sensors, actuators and other networked components may be subject to maliciously destroy in common communication setting or wireless communication one [23], [43]–[45]. This kind of phenomena is usually implemented by cyber-attackers with the aim of the enormous economy benefits or the disturbing social order. For real-world NCSs, representative examples of cyber attacks include an attack on Maroochy Shire Council’s sewage control system in Queensland, Australia (2000) [4], Slammer worm on Davis–Besse power plant in Ohio, U.S. (2003) [5], and recent Stuxnet worm targeted many industrial control systems [7] (see [8] for more real-world cyber incidents). Obviously, NCSs are becoming more and more vulnerable to cyber attacks on both physical infrastructures and communication networks. As
a result, security issues of NCSs need to be properly addressed to ensure that the systems are operating in a safe manner.

In securing NCSs, reliable attack detection is of utmost importance (see [46]–[50]). Generally speaking, when cyber attacks are detected and located in a timely fashion, the damage to overall systems can be controlled within a tolerable limit. For instance, bad data detectors in power systems or sensor networks are generally equipped to detect the deviation of the estimated states and provide an alarm operation [51]. Recently, considerable effort has been devoted to attack detection by means of different approaches, which can be arguably divided into three types: 1) Bayesian detection approaches [10], [11]; 2) WLS approaches [16]–[18], [52]; and 3) $\chi^2$-detector strategies based on Kalman filters [19]–[22]. In Bayesian detection, attack detection is usually formulated as a binary hypothesis test with prior probability. With the help of such a test, the performance limit of collaborative spectrum sensing is investigated in [10]. The closed form expression is proposed in [11] for an optimal attacking strategy of the Byzantines. For steady data, a WLS approach is widely employed to obtain the gain of various detectors, such as generalized likelihood ratio detectors [16], detectors based on the nuclear norm minimization and the low rank matrix factorization [17], and a heuristic approach for detecting abrupt changes in the system outputs in light of the singular value decomposition [18].

The next phase in enhancement of the NCSs’ security is designing a resilient control system that provides NCSs with the ability to tolerate adversaries and recover from cyber attacks [35]. Investigating the problem of resilient control system design is beyond the scope of this paper, and therefore, the propagation of the output of the attack detection system throughout the control system is not considered in this paper. Instead, it has just been considered as an alarm signal for the time being.

When taking the dynamics of plants into account, a generally used state estimation method (i.e., Kalman filter) and attack detector (i.e., performance index test, also known as $\chi^2$-detector) have been extensively studied recently. For instance, in [19], a method to detect a replay attack on sensors of a control system is introduced. The proposed method requires the usage of an LQG controller and a Kalman filter with $\chi^2$-detector. In the method, the control signal is redesigned by adding a zero-mean authentication signal in order to increase the sensitivity of $\chi^2$-detector to a cyber attack. In [20], a Kalman filtering technique is utilized to estimate the state of the system based on the report from sensor measurement and the past state values. Also, an additional detector based on Euclidean distance along with $\chi^2$-detector is proposed
to detect a false data injection attack on smart grid systems. A method in which KL divergence, a Kalman filter, and an LQG controller are adopted to design an optimal Neyman–Pearson detector for replay and covert attacks is proposed in [21]. In [22], a cosine similarity matching-based approach to design a detector is presented. The detector incorporates the Kalman filter estimation to measure any deviation between actual measurements and estimated values.

Up until now, most of attack detection approaches by means of the state estimation require systems noises in a stochastic framework, which provides a probabilistic state estimation. As is well-known, estimation with the nature of a probabilistic approach, particularly the Kalman filtering technique, requires the use of mean and variance to describe the state distributions modeled as random variables (usually white and Gaussian perturbations). However, in many real-word applications, such as target tracking and attack, system guidance and navigation, 100% confidence in the state estimation is of paramount importance. Therefore, it is much more appropriate to model the state distributions in certain sets considering UBB noises. In addition, a widely used attack detection scheme, i.e., the performance index test ($\chi^2$-detector) relies on a robust residual signal to capture discrepancies between estimated behavior and that predicted by a model. In the Kalman filtering technique, the estimated and predicted states are single vectors and therefore, they cannot guarantee that state is included in some region. Besides, the resulting UBB noises are suboptimal for Kalman-type filtering, and thus could mostly reduce the reliability of attack detection. The requirement of set-valued estimation stimulates the development of ellipsoidal state estimation technique [53]. The idea of this technique is to provide a set of state estimates in state space which always contains the true state of the system [54], [55]. In other words, the actual estimate is a set in state space rather than a single vector. Consequently, this kind of technique is known as a set-membership or set-valued state estimation (filtering). Recently, the set-membership approach has been extensively studied in filtering problems (see [56]–[63]). On the other hand, an optimal ellipsoid with minimal size in set-membership estimation can be determined by using convex optimization approaches, which provides a pathway to improve the state estimation performance or detection performance.

Motivated by the discussion above and inspired by the Kalman and set-membership filtering technique, it is the objective of this paper to design a set-membership filter to detect cyber attacks in an NCS. We first propose a recursive convex optimization algorithm to compute the state estimate ellipsoid that guarantees to contain the true
state for an attack-free system. The state estimate algorithm consists of a prediction step and a measurement update step. We then add two sub-algorithms into our algorithm to detect cyber attacks by introducing two criteria based on the intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated with the measurement output. According to these two developed criteria combined with the ellipsoids calculated in prediction steps and measurement update steps, we determine the following.

1. The control signal is violated by a cyber attack if there is no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated with the previous (one-step-behind) measurement output.

2. The sensor signal is targeted by a cyber attack if there is no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time instant.

The main contribution of this paper is to design a novel cyber attack detection method which includes two steps: 1) a prediction step and 2) a measurement update step. In these two steps, our aim is to find an estimation set through updating the prediction set with the one yielding from the current measurement. In the measurement update step, the unconstrained state estimate is projected onto the linear measurement (output) constrained surface to obtain the updated estimation set by using Finsler’s lemma. A cyber attack on sensors or a network exchanging data between sensors and controllers can be detected if there is no intersection between the estimation set updated at the current time instant and the prediction set, and a cyber attack on actuators or a network transmitting data between controllers and actuators can be detected if the prediction set has no intersection with the estimation set updated at the previous time instant. To the best of authors’ knowledge, it is the very first time that not only has the set-membership filtering approach been utilized for the purpose of the attack detection problem but it has also been studied especially in distinguishing attacks on control signals from attacks on measurement outputs.

The rest of this paper is organized as follows. Section 2.3 presents the framework of the system of interest and the proposed attack detection strategy and formulates the set-membership filtering problem with measurement update for NCSs. Section 2.4 provides the design of an attack detection procedure based on the prediction and updated sets and then the associated algorithms are presented in this section. Section 2.5 presents the mathematical model of the test system and also the two cyber attacks
models used to demonstrate the effectiveness of the proposed method. The conclusion is drawn in Section 2.6.

Notation: The notations used in this paper are fairly standard except where otherwise stated. The notation $X > 0$ means that $X$ is “positive definite”. The superscript $T$ stands for matrix transposition. The notation $\text{Tr}(P)$ denotes the trace of $P$.

Ellipsoid: An ellipsoidal set is denoted as $\mathcal{X} \triangleq \{ \zeta : \zeta = c + Ez, \|z\| \leq 1 \}$, where $c \in \mathbb{R}^n$ is the center and $E \in \mathbb{R}^{n \times m}$ with $\text{rank}(E) = m \leq n$ is the shape matrix of the ellipsoid. This representation is widely used to deal with all bounded ellipsoids including “flat” ellipsoids such as points or intervals. Let $E$ be a lower triangular matrix whose every diagonal element is positive. By a Cholesky factorization, one can conclude that $P = EE^T > 0$ and $z^T z = (\zeta - c)^T P^{-1} (\zeta - c) \leq 1$. Hence, an alternative representation of the ellipsoidal set is $\mathcal{X} \triangleq \{ \zeta : (\zeta - c)^T P^{-1} (\zeta - c) \leq 1 \}$. The size of the ellipsoid is a function of the squared shape matrix $P$ which can be measured by means of $\text{Tr}(P)$, i.e. the sum of squared semi-axes lengths [64].

2.3 Problem statement and formulation

Consider the framework of an NCS as shown in Figure 2.1. In this framework, the controller receives the measurement data from the sensor and sends the control signal data to the actuator via a network which is often equipped with wireless communication [65], [66]. As the attack detector located at the remote site, the measurement data received from the wireless communication and the output of the actuator are sent to the attack detector via a communication network.

In practical situations, an attack on an NCS can be considered in two ways: 1) attacks on physical components, i.e., sensors and actuators and 2) attacks on the communication network exchanging data among the physical components [21]. For instance, as shown in Figure 2.1:

1. an attack on a sensor can either be considered as an attack on the sensor itself, attack point $A_4$, by forcing it to transmit an incorrect signal, $\bar{y}_k$, or it can also be considered as an attack on the communication network between the sensor and the controller, attack point $A_1$, to force the communication network to transmit an incorrect signal, $\bar{y}_k$;
Figure 2.1 Some possible attack points to a networked control system.

2. an attack on an actuator can either be interpreted as an attack directly on the actuator, attack point $A_3$, to send an incorrect data, $\bar{u}_k$, to the plant or on the communication network, attack point $A_2$, to transmit an incorrect data, $\tilde{u}_k$, from the controller to the actuator.

In this paper, it will be considered that only the communication network exchanging data among the physical components is under cyber attacks (attacks points $A_1$ and $A_2$). Therefore, the measurement output and the control signal at the plant side can be considered as $\bar{y}_k = y_k$ and $\bar{u}_k = \tilde{u}_k$, respectively, and at the controller side can be denoted as $\tilde{y}_k$ and $u_k$, respectively. If there is no attack on this communication network, then $\bar{u}_k = u_k$ and $\tilde{y}_k = y_k$. Also, it will be assumed that there is an ideal communication network sending data to the attack detector. Considering attacks on the communication network between the system’s components and the proposed attack detection system is considerable part of the ongoing research but is currently beyond the scope of this paper. The assumption of an ideal communication network between the system’s components and the proposed attack detector in Figure 2.1 is due to just focusing on the detection of the attacks targeting the communication channels between the system’s components, i.e. sensor, controller, and actuator.

In the following, a physical plant considered in the proposed framework, Figure 2.1, is introduced in a mathematical way. The physical plant under consideration is a
discrete time-varying system in the form of

\[ x_{k+1} = A_k x_k + B_k u_k + B_{w,k} w_k \quad (2.1) \]
\[ y_k = C_k x_k + D_k \nu_k \quad (2.2) \]

where \( x \in \mathbb{R}^{n_x} \) is the system state; \( u_k \in \mathbb{R}^{n_u} \) is the known deterministic input; \( y_k \in \mathbb{R}^{n_y} \) is the measurement output; \( A_k, B_k, C_k, D_k, \) and \( B_{w,k} \) are known time-varying matrices with appropriate dimensions; \( w_k \in \mathbb{R}^{n_w} \) and \( \nu_k \in \mathbb{R}^{n_v} \) are the process and measurement noises, respectively, and satisfy the following assumption.

**Assumption 2.1:** The process noise \( w_k \) and measurement noise \( \nu_k \) are UBB, which are assumed to belong to the following specified ellipsoidal sets

\[ \mathcal{W}_k \triangleq \{ w_k : w_k^T W_k^{-1} w_k \leq 1 \} \]
\[ \mathcal{V}_k \triangleq \{ \nu_k : \nu_k^T V_k^{-1} \nu_k \leq 1 \} \]

where \( W_k = W_k^T > 0 \) and \( V_k = V_k^T > 0 \) are known matrices with compatible dimensions.

**Remark 2.1:** There are some practical situations where noises on system inputs and measurement outputs are unknown but can be bounded. For instance, for a vehicle tracking system, the maximum acceleration of the vehicle is always known due to the engine dynamics and construction of the vehicle; however, it is not exactly known how much the acceleration is when the vehicle is running. Therefore, it is quite reasonable to assume that noises on the vehicle acceleration are bounded, and thus they can be considered as UBB process noises. Furthermore, bounded noises on measurement outputs can be viewed as some bound due to quantization errors and measurement errors. Therefore, an assumption of UBB noise does not require a prior knowledge of the actual pattern of the noise in the sense that only the knowledge of a bound on the realization is necessary.

**Assumption 2.2:** The initial state \( x_0 \) is assumed to belong to a given ellipsoid:

\[ \mathcal{X}_0 \triangleq \{ x_0 : (x_0 - \hat{x}_0)^T P_0^{-1} (x_0 - \hat{x}_0) \leq 1 \} \]

where \( \hat{x}_0 \) is the given estimate of \( x_0 \), and \( P_0 = P_0^T > 0 \) is a known matrix.

**Remark 2.2:** In practical situations, due to some unknown errors and noises in measuring the initial state of the system, there might be some uncertainties in the
measurement of the initial state. These uncertainties can be viewed as UBB noises as discussed in Remark 2.1. Therefore, Assumption 2.2 can be considered on the initial state in the deterministic framework.

In light of the set-membership filtering approach, our strategy to detect the cyber attack can be summarized as below.

1. The detection on cyber attacks for sensor measurement data:
   - If the updated ellipsoidal set $X_{k+1}$ and the prediction ellipsoidal set $X_{k+1|k}$ have the intersection, i.e., $X_{k+1} \cap X_{k+1|k} \neq \emptyset$, it can be concluded that there is no attack.
   - Otherwise, if the updated ellipsoidal set $X_{k+1}$ has no intersection with the prediction ellipsoidal set $X_{k+1|k}$, i.e., $X_{k+1} \cap X_{k+1|k} = \emptyset$, it indicates that the sensor measurement data is affected by the attack.

2. The detection on cyber attacks for control signal data:
   - If the prediction ellipsoidal set $X_{k+1|k}$ and the previous updated ellipsoidal set $X_k$ have the intersection, i.e., $X_k \cap X_{k+1|k} \neq \emptyset$, it indicates that there is no attack.
   - Otherwise, if the prediction ellipsoidal set $X_{k+1|k}$ has no intersection with the previous updated ellipsoidal set $X_k$, i.e., $X_k \cap X_{k+1|k} = \emptyset$, it indicates that the control signal data is affected by the attack.

Clearly, for the given ellipsoid $X_k$ on previous state estimation, two key steps for attack detection are to determine the corresponding prediction ellipsoid $X_{k+1|k}$ (i.e. the prediction step) and the updated ellipsoid $X_{k+1}$ (i.e. the measurement update step) via the current sensor measurement data, respectively. In what follows, some preliminaries will be provided for these two steps.

### 2.3.1 Prediction step

First, the prediction filter is considered in the form of

$$x_{k+1|k} = \hat{A}_k \hat{x}_k + B_k u_k,$$  \hfill (2.6)

where $\hat{x}_k$ is the estimation of the state $x_k$ and $\hat{A}_k$ is the filter parameter to be determined.
For the given state estimation ellipsoid set $X_k$ with the centre $\hat{x}_k$ and the shape matrix $E_k$, the real state $x_k$ can be described by
\[ x_k = \hat{x}_k + E_k \vartheta_1. \] (2.7)
where $\vartheta_1$ is a vector satisfying $\| \vartheta_1 \| \leq 1$.

Then, our goal is to obtain the prediction ellipsoid set:
\[ X_{k+1|k} \triangleq \{ x_{k+1} : (x_{k+1} - \hat{x}_{k+1|k})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1 \}. \] (2.8)
It should be pointed out that such an ellipsoidal set contains the state $x_{k+1}$ for any value of the system noises belong to their specified sets.

### 2.3.2 Measurement update step

The update based on the current measurement is considered for the system (2.1)-(2.2), which is in the form of
\[ \hat{x}_{k+1} = \hat{x}_{k+1|k} + \hat{B}_{k+1}(y_{k+1} - \hat{y}_{k+1|k}), \] (2.9)
where $\hat{B}_{k+1}$ is the filter parameter to be determined.

According to the prediction ellipsoidal set $X_{k+1|k}$ given by (2.8), the state $x_{k+1}$ can be written as
\[ x_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k} \vartheta_2. \] (2.10)
where $\vartheta_2$ is a vector satisfying $\| \vartheta_2 \| \leq 1$.

Our objective is to update this prediction set with the one yielding from the current measurement $y_{k+1}$. In other words, we look for an updated ellipsoidal set $X_{k+1}$ with the center $\hat{x}_{k+1}$ and the shape matrix $E_{k+1}$ for the state $x_{k+1}$, given by the current measurement information at the time instant $k + 1$. Thus, the updated ellipsoidal set should satisfy the condition
\[ (x_{k+1} - \hat{x}_{k+1})^T P_{k+1}^{-1} (x_{k+1} - \hat{x}_{k+1}) \leq 1, \] (2.11)
whenever the equality (output constraint)
\[ y_{k+1} = C_{k+1} \hat{x}_{k+1|k} + C_{k+1} E_{k+1|k} \vartheta_2 + D_{k+1} \nu_{k+1} \] (2.12)
holds for some $\| \vartheta \| \leq 1$ and $\nu_{k+1} \in \mathcal{V}_{k+1}$.

2.4 Attack detection using set-membership filtering

In this section, a set-membership filter will be designed to solve the proposed cyber attack detection problem. First, in Section 2.4.1, a prediction ellipsoidal set is designed for state $x_{k+1}$. Then, Section 2.4.2 utilizes the method of projecting the unconstrained state estimate onto the linear output constrained surface in order to update the prediction ellipsoidal set for state $x_{k+1}$ with the current measurement. Finally two convex optimization problems (OPs) and one algorithm are provided to expose the cyber attack diagnosis scheme.

2.4.1 Design of the prediction ellipsoidal set

From the system model (2.1), and the filter (2.6) and (2.7), the prediction error $x_{k+1} - \hat{x}_{k+1|k}$ can be written as

$$x_{k+1} - \hat{x}_{k+1|k} = (A_k - \hat{A}_k)\hat{x}_k + A_kE_k\vartheta_1 + B_{w,k}w_k.$$  

(2.13)

Denoting

$$\eta_{1,k} = \begin{bmatrix} 1 \\ \vartheta_1 \\ w_k \end{bmatrix},$$

(2.13) can be written in a compact form

$$x_{k+1} - \hat{x}_{k+1|k} = \Pi_{1,k}\eta_{1,k},$$

(2.14)

where

$$\Pi_{1,k} = \begin{bmatrix} (A_k - \hat{A}_k)\hat{x}_k & A_kE_k & B_{w,k} \end{bmatrix}.$$  

Thus, the condition in (2.8)

$$(x_{k+1} - \hat{x}_{k+1|k})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1$$

can be written as

$$\eta_{1,k}^T [\Pi_{1,k}^T P_{k+1|k}^{-1} \Pi_{1,k} - \text{diag}\{1, 0, 0\}] \eta_{1,k} \leq 0.$$  

(2.15)
From (2.3) and (2.7), the unknown variables \( w_k \) and \( \vartheta_1 \) satisfy the following constraints

\[
\begin{align*}
\begin{cases}
    w_k^T W_k^{-1} w_k \leq 1, \\
    \|\vartheta_1\| \leq 1,
\end{cases}
\end{align*}
\]

which can be expressed in \( \eta_{1,k} \) as

\[
\begin{align*}
\begin{cases}
    \eta_{1,k}^T \text{diag}\{-1, 0, W_k^{-1}\} \eta_{1,k} \leq 0, \\
    \eta_{1,k}^T \text{diag}\{-1, I, 0\} \eta_{1,k} \leq 0.
\end{cases}
\end{align*}
\]

(2.16)

Applying \( S \)-procedure [67] to (2.15) and (2.16), one has that the inequality (2.15) holds if there exist nonnegative scalars \( \tau_{1,k} \) and \( \tau_{2,k} \) such that

\[
\Pi_{1,k}^T P_{k+1|k}^{-1} \Pi_{1,k} - \text{diag}\{1, 0, 0\} - \tau_{1,k} \text{diag}\{-1, I, 0\} - \tau_{2,k} \text{diag}\{-1, 0, W_k^{-1}\} \leq 0. \quad (2.17)
\]

In addition, inequality (2.17) can be rewritten in the following compact form:

\[
\Pi_{1,k}^T P_{k+1|k}^{-1} \Pi_{1,k} - \text{diag}\{1 - \tau_{1,k} - \tau_{2,k}, \tau_{1,k} I, \tau_{2,k} W_k^{-1}\} \leq 0. \quad (2.18)
\]

Finally, denoting

\[ \Theta_{1,k} = -\text{diag}\{1 - \tau_{1,k} - \tau_{2,k}, \tau_{1,k} I, \tau_{2,k} W_k^{-1}\}, \]

the above inequality can be further written as

\[
\Pi_{1,k}^T P_{k+1|k}^{-1} \Pi_{1,k} + \Theta_{1,k} \leq 0. \quad (2.19)
\]

By using the Schur complements, (2.19) is equivalent to

\[
\begin{bmatrix}
    -P_{k+1|k} & \Pi_{1,k} \\
    \Pi_{1,k}^T & \Theta_{1,k}
\end{bmatrix} \leq 0. \quad (2.20)
\]

Based on the discussion above, we have the following result.

**Theorem 2.1:** For the system (2.1)-(2.2), suppose that the state \( x_k \) belongs to its state estimation ellipsoid \((x_k - \hat{x}_k)^T P_k^{-1} (x_k - \hat{x}_k) \leq 1\). Then the one-step ahead state \( x_{k+1} \) resides in its state prediction ellipsoid \((x_{k+1} - \hat{x}_{k+1})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1}) \leq 1\), if there exist \( P_{k+1|k} > 0, \hat{A}_k, \tau_{1,k} \geq 0, \) and \( \tau_{2,k} \geq 0 \) such that (2.20) holds. Moreover the center of the state prediction ellipsoid is determined by (2.6).
Theorem 2.1 outlines the principle of determining the state prediction ellipsoid containing $x_{k+1}$. In order to determine an optimal ellipsoid and reduce the conservativeness, the convex optimization is performed in (2.21), in which the trace of $P_{k+1|k}$ is optimized at each time step in an effort to find the prediction ellipsoid set with minimal size.

$$
\text{minimize} \quad \text{Tr}(P_{k+1|k})
$$

subject to (2.20)

$$
(2.21)
$$

2.4.2 Update on prediction ellipsoidal set with current measurement

At this section, our purpose is to develop a scheme to determine the shape matrix $E_{k+1}$ and the filter gain $\hat{B}_{k+1}$ with the output constraint (2.12).

From the system (2.1)-(2.2), the prediction ellipsoid set (2.10), and the filter based on the current measurement (2.9), the current estimation error $x_{k+1} - \hat{x}_{k+1}$ can be written as

$$
x_{k+1} - \hat{x}_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k}\vartheta_2 - \hat{x}_{k+1|k} - \hat{B}_{k+1}C_{k+1}\hat{x}_{k+1|k} - \hat{B}_{k+1}D_{k+1}\nu_{k+1}
$$

$$
= (I - \hat{B}_{k+1}C_{k+1})E_{k+1|k}\vartheta_2 - \hat{B}_{k+1}D_{k+1}\nu_{k+1}.
$$

(2.22)

As the unknown variables are $\vartheta_2$ and $\nu_{k+1}$, we can define

$$
\eta_{2,k+1} = \begin{bmatrix} 1 \\ \vartheta_2 \\ \nu_{k+1} \end{bmatrix}.
$$

Thus, the above estimation error dynamics can be written in a compact form

$$
x_{k+1} - \hat{x}_{k+1} = \Pi_{2,k+1}\eta_{2,k+1},
$$

(2.23)

where

$$
\Pi_{2,k+1} = \begin{bmatrix} 0 & (I - \hat{B}_{k+1}C_{k+1})E_{k+1|k} & -\hat{B}_{k+1}D_{k+1} \end{bmatrix}.
$$

Taking (2.23) into account, the condition (2.11) in Section 2.3.2 can be described as

$$
\eta_{2,k+1}^T[\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1} - \text{diag}(1, 0, 0)]\eta_{2,k+1} \leq 0.
$$

(2.24)
On the other hand, from (2.4) and (2.10), the unknown variables $\nu_{k+1}$, and $\vartheta_2$ satisfy the following constraints

$$\begin{cases}
\nu_{k+1}^T V_{k+1}^{-1} \nu_{k+1} \leq 1,
\|\vartheta_2\| \leq 1,
\end{cases}$$

which can be expressed in $\eta_{2,k+1}$ as

$$\begin{cases}
\eta_{2,k+1}^T \text{diag}\{-1,0,V_{k+1}^{-1}\} \eta_{2,k+1} \leq 0,
\eta_{2,k+1}^T \text{diag}\{-1,I,0\} \eta_{2,k+1} \leq 0.
\end{cases}$$

By applying $S$-procedure to (2.24) and (2.25), one has that the inequality (2.24) is true if there exist nonnegative scalars $\tau_{3,k}$ and $\tau_{4,k}$ such that

$\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1} - \text{diag}\{1,0,0\} - \tau_{3,k} \text{diag}\{-1,I,0\} - \tau_{4,k} \text{diag}\{-1,0,V_{k+1}^{-1}\} \leq 0,$

which can be written in the following compact form

$$\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1} - \Theta_{2,k} \leq 0.$$ (2.27)

Denote

$$\Theta_{2,k} = \text{diag}\{1 - \tau_{3,k} - \tau_{4,k}, \tau_{3,k} I, \tau_{4,k} V_{k+1}^{-1}\}.$$ (2.28)

It follows from (2.27) that

$$\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1} - \Theta_{2,k} \leq 0.$$ (2.29)

Now, we deal with the output constraint (2.12) in Section 2.3.2. First, it can be described by

$$\Pi_{y,k+1} \eta_{2,k+1} = 0,$$ (2.30)

where

$$\Pi_{y,k+1} = \begin{bmatrix}
C_{k+1} \hat{x}_{k+1|k} - y_{k+1} & C_{k+1} E_{k+1|k} & D_{k+1}
\end{bmatrix}.$$ (2.31)

By virtue of Finsler’s lemma [67], the inequality (2.24) [i.e. (2.11)] under constraint (2.30) [i.e. (2.12)] is true if there exists an $N_{k+1}$ such that

$$\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1} - \Theta_{2,k} + N_{k+1}^T \Pi_{y,k+1} + \Pi_{y,k+1}^T N_{k+1} \leq 0.$$ (2.32)
For the purpose of simplicity, denote
\[ \Theta_{3,k} = -\Theta_{2,k} + N_{k+1}^T \Pi_{g,k+1} + \Pi_{y,k+1}^T N_{k+1}. \] (2.33)

Then, by using Schur complements, (2.32) is equivalent to
\[ \begin{bmatrix} -P_{k+1} & \Pi_{2,k+1} \\ \Pi_{2,k+1}^T & \Theta_{3,k} \end{bmatrix} \leq 0. \] (2.34)

Then we arrive at the following theorem.

**Theorem 2.2:** For the system (2.1)-(2.2), if the state \( x_{k+1} \) belongs to its state prediction ellipsoid \( (x_{k+1} - \hat{x}_{k+1|k})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1 \), then such a state also resides in its updated state estimation ellipsoid \( (x_{k+1} - \hat{x}_{k+1})^T P_{k+1}^{-1} (x_{k+1} - \hat{x}_{k+1}) \leq 1 \) with the center determined by (2.9), where \( P_{k+1} > 0 \) satisfies matrix inequality (2.34) with other decision variables \( \hat{B}_{k+1}, \tau_{3,k} \geq 0, \tau_{4,k} \geq 0 \), and \( N_{k+1} \).

Now the convex optimization approach is applied to determine an optimal ellipsoid with the minimal size. \( P_{k+1} \) is obtained by solving the following OP:
\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(P_{k+1}) \\
\text{subject to} & \quad (2.34)
\end{align*}
\] (2.35)

**Remark 2.3:** From Theorems 2.1 and 2.2, the OPs (2.21) and (2.35) are based on Recursive Linear Matrix Inequalities (RLMIs) (2.20) and (2.34) which are linear to \( P_{k+1|k}, \hat{A}_{k}, \hat{B}_{k+1}, N_{k+1}, \) and \( \tau_{m,k}, m = 1, 2, \cdots, 4 \). Hence, these OPs can be solved by some existing semi-definite programming via an interior-point algorithm at each time step. The interior-point algorithm usually has a polynomial-time complexity \( \mathcal{O}(\ell \mathcal{M}^3) \), where \( \ell \) is the total row size of the main RLMIs, \( \mathcal{M} \) is the total number of scalar decision variables of the main RLMIs (2.20) and (2.34). Since \( \ell \) and \( \mathcal{M} \) are dependent on \( n_x, n_u, n_y, n_w, \) and \( n_v \), where \( n \) denotes the number of its subscript, the computational complexity of the developed recursive algorithm depends polynomially on the dimensions of system’s parameter variables \([59]\). As discussed in \([68]\), in practical situations, interior-point methods for semi-definite programs are competitive with other methods for small programs and substantially faster for medium and large-scale problems.
### 2.4.3 Recursive algorithm for attack diagnosis

The recursive algorithm based on the set-membership filtering to compute the state ellipsoids so that a cyber attack can be detected is summarized below.

**Algorithm 2.1**: Recursive convex optimization algorithm / Attack detection

1. **Initialization**
   - Given an initial ellipsoid $\mathcal{X}_0(\hat{x}_0, E_0)$, the current value of input $u_0$, recursive times $N$, and set $k = 0$. Let $\hat{x} \leftarrow \hat{x}_0$, $E \leftarrow E_0$, $u \leftarrow u_0$.

2. **Prediction**
   1. Calculate $P_{k+1|k}$ and $\hat{A}_k$ by solving the OP (2.21). Obtain the matrix $E_{k+1|k}$.
   2. Calculate the center of the prediction ellipsoid $\hat{x}_{k+1|k}$ by (2.6).

3. **Sub-algorithm 2.1.a: Control signal data cyber attack diagnosis**
   1. If $\mathcal{X}_k \cap \mathcal{X}_{k+1|k} \neq \emptyset$, there is no attack;
   2. If $\mathcal{X}_k \cap \mathcal{X}_{k+1|k} = \emptyset$, data is subject to attack and then $\mathcal{X}_{k+1|k} \leftarrow \mathcal{X}_k$

4. **Measurement update**
   1. Calculate $P_{k+1}$ and $\hat{B}_{k+1}$ by solving the OP (2.35). Obtain the new $E_{k+1}$.
   2. Calculate the center of the updated estimation ellipsoid $\hat{x}_{k+1}$ by (2.9).

5. **Sub-algorithm 2.1.b: Sensor measurement data cyber attack diagnosis**
   1. If $\mathcal{X}_{k+1} \cap \mathcal{X}_{k+1|k} \neq \emptyset$, there is no attack;
   2. If $\mathcal{X}_{k+1} \cap \mathcal{X}_{k+1|k} = \emptyset$, data is subject to attack and then $\mathcal{X}_{k+1} \leftarrow \mathcal{X}_{k+1|k}$

6. **Loop**
   - If $k == N$ then Exit, Else $k \leftarrow k + 1$ and Goto Prediction step.

Algorithm 2.1 recursively computes the prediction ellipsoid $\mathcal{X}_{k+1|k}$ and its update $\mathcal{X}_{k+1}$ with the current measurement $y_{k+1}$. The two sub-algorithms are proposed to detect cyber attacks that affect control signals and sensor measurements.

**Remark 2.4**: Theorems 2.1 and 2.2 show that the proposed ellipsoidal set-membership filtering problem can be converted into the feasibility problem of a set of RLMIs (2.20) and (2.34) to determine two optimal ellipsoidal sets. Ideally, in an attack-free system, the two sets will always have intersection since they both contain the true state $x_{k+1}$. However, if there is an attack on the system, one of the two sets does not contain...
the true state since the center of that set is affected by the attack. Therefore, one can conclude that there must be no intersection between the ellipsoidal sets as a results of the attack’s impact on their centres. However, this effect may result in infeasibility of Theorems 2.1 and 2.2. As one can see in subalgorithms 2.1.a and 2.1.b, in order to overcome this situation, if there is an attack compromising the control signal then $\mathcal{X}_{k+1|k} \leftarrow \mathcal{X}_k$, and if there is an attack targeting the measurement outputs then $\mathcal{X}_{k+1} \leftarrow \mathcal{X}_{k+1|k}$. These modifications in the ellipsoidal sets recover them so that they become free of the attack for their succeeding steps in Algorithm 2.1. Consequently, the feasibility of the proposed RLMII problems can be kept at each time step.

**Remark 2.5:** From subalgorithm 2.1.a, if there is an attack that violates the control signal commencing at time $k$, the input of the system $u_k$ is affected by the attack. Therefore, the prediction ellipsoid set $\mathcal{X}_{k+1|k}$ is affected by the attack. However, the updated estimation ellipsoid $\mathcal{X}_k$, which is based on the prediction ellipsoidal set at time $k - 1$, i.e., $\mathcal{X}_{k|k-1}$, is not affected by the attack. Thus, one can conclude that if there is no intersection between the prediction ellipsoid set $\mathcal{X}_{k+1|k}$ and the updated estimation ellipsoidal set $\mathcal{X}_k$, the control signal is compromised by an attack.

**Remark 2.6:** From subalgorithm 2.1.b, if the sensor measurement is violated by an attack at time $k + 1$, the output of the sensor $y_{k+1}$ is affected by the attack. In this case, the estimation ellipsoidal set $\mathcal{X}_{k+1}$ updated with the current measurement $y_{k+1}$ is affected by the attack. However, the prediction ellipsoidal set $\mathcal{X}_{k+1|k}$, which is based on the measurement at time $k$, i.e. $y_k$, is not affected by the attack. Thus, it can be indicated that if there is no intersection between the prediction ellipsoidal set and the updated estimation ellipsoidal set, the sensor measurement is compromised by an attack.

**Remark 2.7:** As the existence of noises and parameter uncertainties in many practical situations is unavoidable, the proposed method is aimed at considering the noises as an inherent property of the system so that only abrupt changes deliberately brought into the system by an attacker can be detected in a timely fashion. Considering noises directly into solving the two RLMIs in (2.20) and (2.34) can improve the performance of the proposed detection algorithm through focusing just on attacks. Therefore, the detection algorithm can tolerate the changes caused by the system noises and therefore, it does not recognize them as an attack. In other words, the proposed attack detection system is able to distinguish the system’s changes due to process and measurement
noises from the attacker’s deliberate changes into the system, and as a result, the proposed attack detection method can tolerate the system changes due to noises and just report the abrupt changes brought to the system by the attacker.

Remark 2.8: It should be pointed out that if the abrupt change caused by an attacker is fairly small at its time of occurrence, the proposed detection algorithm cannot recognize it, i.e. there might be an intersection between the two ellipsoidal sets. This situation continues till the change into the control signals and/or measurement outputs can reach the certain level (threshold) where there is no intersection between the two ellipsoid sets. As is well-known, most of attack detection approaches require a threshold and if the abrupt changes are below of this threshold, they cannot be detected. However, it is noteworthy that the size of these ellipsoid sets are minimized through OPs (2.21) and (2.35), which implies that the threshold for the proposed attack detection algorithm reaches its minimum value by minimizing the size of these sets. Consequently, the optimization approach in our proposed attack detection algorithm is aimed at minimizing the attacks damages which may be tolerated by a resilient control algorithm.

2.5 Case studies

2.5.1 Test system

In this section, a vehicle tracking system is considered as the physical plant including a Global Positioning System (GPS) receiver. The schematic of a vehicle tracking system is illustrated in Figure 2.2. The main objective of a tracking system is to estimate the state trajectories of a vehicle. As a result, accuracy in determination of a vehicle position with respect to a fixed reference frame (absolute positioning) is of practical importance to navigate autonomous vehicles. If the navigation system’s components, i.e., sensors, actuators, and controllers are connected through a wireless communication system as shown in Figure 2.1, this system becomes more vulnerable to cyber attacks. Consequently, it is our objective to evaluate the performance of the proposed attack detection method for a vehicle tracking system.
The vehicle dynamics model with a known commanded acceleration is

\[
\dot{x}(t) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ \sin \theta \\ \cos \theta \end{bmatrix} u(t)
\]

(2.36)

where the first two components of the state vector \(x(t)\) are latitude and longitude positions, and the last two components are latitude and longitude velocities, respectively. \(u(t)\) is the commanded acceleration and \(\theta\) is the road orientation angle measured counterclockwise from due east.

Considering finite difference approximation of derivatives as

\[
\dot{x}(t) \approx \frac{x_{t+1} - x_t}{\tau_s}
\]

through discretization process with a sampling period \(\tau_s\) and changing the subscript from \(t\) to \(k \in \mathbb{Z}^+\), the discrete-time vehicle dynamics model is described by [69]

\[
x_{k+1} = \begin{bmatrix} 1 & 0 & \tau_s & 0 \\ 0 & 1 & 0 & \tau_s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 0 \\ \tau_s \sin \theta \\ \tau_s \cos \theta \end{bmatrix} u_k + \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} w_k
\]

(2.37)

where \(w_k\) represents process noises due to potholes which belongs to a specified ellipsoidal set.
The GPS measurement equation can be written as

\[
y_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x_k + \frac{1}{2} \nu_k
\]  

(2.38)

where \( y_k \) is the GPS measurement; and \( \nu_k \) is the GPS measurement noise.

The sampling period \( \tau_s \) is chosen as 1.1 s. The road orientation angle is chosen as a constant, \( \theta = 60^\circ \). The control signal is considered as a known commanded acceleration \( u_k \) with a value which is set to \( \pm 0.1 \text{m/s}^2 \), as if the vehicle alternately accelerating and decelerating in traffic. \( w_k \) and \( \nu_k \) are assumed as \( 0.5 \sin(2k) \) and \( 0.5 \sin(30k) \), respectively. \( W_k = 1 \) and \( V_k = 1 \). The initial state is set as

\[
x_0 = \begin{bmatrix} 5 & 5\sqrt{3} & 2 & 2\sqrt{3} \end{bmatrix}^T
\]

which belongs to the ellipsoid

\[
X_0 \triangleq \{ x_0 : (x_0 - \hat{x}_0)^T P_0^{-1} (x_0 - \hat{x}_0) \leq 1 \}
\]

where \( \hat{x}_0 \) and \( P_0 \) are set to \( \hat{x}_0 = \begin{bmatrix} 0 & 7 & 0 & 3 \end{bmatrix}^T \), and \( P_0 = \text{diag}\{100, 100, 10, 10\} \).

### 2.5.2 Cyber attack model

In the following sections, we consider two types of the deception attacks, namely replay attacks and bias injection attacks, that affect the sensor measurement data and control signal data transmitted via a communication network, respectively.

In this paper, there are some assumptions about the ability of an attacker to perform a successful attack. Considering these assumptions, the attacker is able to modify the true values of \( y_k \) and \( u_k \) to arbitrary ones. The required assumptions will be explained in the following sections.

1) **Replay attacks on sensor measurement data:** A successful replay attack does not need \textit{a priori} knowledge of the system components. As discussed in [70], a successful replay attack can be separated into two phases. In the first phase, which is known as disclosure attacks, an attacker must have this ability to gather sequences of data from the sensors’ measurements through violating the disclosure resources, i.e., set of accessible sensor channels. Having this ability, the attacker starts to record sequences of data from sensors’ communication channels without inserting any input to the system. Then, in the second phase the attacker attacks the system by replaying the recorded data to the system via tampering those channels from which data have been previously recorded.
It is assumed that the attacker can record sensor’s measurement data from $k_o$ till $k_r$ with the window size $\zeta = k_r - k_o$ in the first phase. Then, in the second phase, the attacker replays the recorded data to the system from $k = k_r + d$ till the end on the attack at $k = k_f$, where $d$ is the delay between the recording time and replaying time. This attack can be modeled as

$$a_k^y = y_{k-\zeta} - y_k$$

Thus, the sensor’s data affected by the attack is

$$\tilde{y}_k = y_k + a_k^y$$

2) **Bias injection attacks on control signal data:** In a bias injection attack, the attacker injects a constant bias into the system [70]. A successful bias injection attack requires the knowledge of the system model. Although the attacker must be able to compromise the integrity of control signal data, there is no need for the attacker to have a prior knowledge of those channels exchanging control signal data. The bias injection attack on the control signal can be modeled as

$$a_k^u = \delta_u$$

where $\delta_u$ is a constant value injected by the attacker. Therefore, the control signal’s data affected by the attack is

$$\tilde{u}_k = u_k + a_k^u$$

To illustrate the effectiveness of the proposed method, the following three cases are considered. The simulation results are obtained under Matlab 8.6 with YALMIP and the solver Sdpt 3 during 50 sampling steps.

### 2.5.3 Attack-free system

Consider that the system (2.37)-(2.38) is an attack-free system, i.e., $\tilde{u}_k = u_k$ and $\tilde{y}_k = y_k$. Since in the attack-free system the prediction ellipsoidal set and the updated estimation ellipsoidal set both contain the true state, there must always exist the intersection between these two ellipsoidal sets. Figure 2.3 confirms the existence of the intersection between the two sets which are both projected onto the first two components of the state ($x_1 - x_2$ phase-plane) and the last two components of the state ($x_3 - x_4$ phase-plane) shown in Figure 2.3a and Figure 2.3b, respectively.
2.5 Case studies

2.5.4 Replay attack on sensor data

In this case, it is considered that the attacker implements a replay attack on the sensor measurement data via targeting the wireless communication channel between the sensor and the controller, i.e., the attack point $A_1$ shown in Figure 2.1. In the simulation, it is assumed that the attacker records the sensor’s data from step $k = 14$ till $k = 19$ and then replaces the sensor’s data with the recorded data from step $k = 30$ to $k = 35$. From (2.38), it can be observed that the sensors measure the data of the first two components of the state, i.e. $x_1$ and $x_2$. Therefore, the attack can be detected by projecting the prediction and updated ellipsoid sets onto the $x_1 - x_2$ phase-plane.

Figure 2.4 demonstrates the sequence of the intersection between the prediction and the updated estimation ellipsoidal sets during the simulation time. In particular, as the replay attack starts at $k = 30$, the prediction ellipsoidal set is calculated from the sensor measurement data obtained at $k = 29$ when there is no attack. However, the estimation ellipsoidal set is updated with the current sensor measurement data at $k = 30$. Therefore, it is expected from the subalgorithm 2.1.b that there must be no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time instant when the attack starts at $k = 30$, i.e. $\chi_{30|29} \cap \chi_{30} = \emptyset$, as depicted in Figure 2.5b.

From Remark 2.4, since $X_{k+1} \leftarrow X_{k+1|k}$ from $k = 30$ to $k = 35$, one can conclude that the prediction ellipsoidal set remains free of the attack when the system is under the
replay attack. Therefore, once the attack finishes at $k = 36$, the prediction ellipsoidal set and the estimation ellipsoidal set updated with the current sensor measurement data at $k = 36$ are both free of attack. As a result, from the subalgorithm 2.1.b, there must be intersection between these two sets, i.e., $\chi_{36|35} \cap \chi_{36} \neq \emptyset$, as shown in Figure 2.5c.

At $k = 29$ the system is free of the attack, so from the subalgorithm 2.1.b it can be concluded that there exists intersection between the two sets as illustrated in Figure 2.5a.

### 2.5.5 Bias injection attack on control signal data

In this case, it is considered that the attacker carries out a bias injection attack on the control signal via targeting the wireless communication channel between the actuator and the controller, i.e., the attack point $A_2$ shown in Figure 2.1. In the simulation, the attack vector is modeled as $a_k^u = 4$ from step $k = 20$ to $k = 35$, and so

$$\tilde{u}_k = u_k + 4$$

From (2.37), it can be concluded that this attack is directly applied to the last two components of the state, i.e., $x_3$ and $x_4$. Therefore, the attack on the control signal can be detected by projecting the prediction ellipsoidal set and the estimation ellipsoidal set updated with the previous time instant onto the $x_3 - x_4$ phase-plane.
2.5 Case studies

Figure 2.5 The prediction ellipsoid $\chi_{k+1|k}$ (green, dotted line), the updated estimation ellipsoid $\chi_{k+1}$ (blue, solid line) (a) step 29 (b) step 30 (c) step 36 (the updated estimation ellipsoid at step 30 and 36 are magnified with the ratio of $10^7$ and 10, respectively).

Figure 2.6 shows the sequence of the intersection between these two sets during the simulation time. As the bias injection attack starts at $k = 20$, the center of the prediction ellipsoidal set (2.6) is affected by the attack; however the attack does not have any effect on the previous updated estimation ellipsoidal set since it is based on the prediction set at step $k = 19$. Therefore, it is expected from the subalgorithm 2.1.a that there must be no intersection between the prediction ellipsoidal set and the previous updated estimation ellipsoidal set when the attack starts at $k = 20$, i.e., $\chi_{21|20} \cap \chi_{20} = \emptyset$, as depicted in Figure 2.7b.
A novel cyber attack detection method based on the ellipsoidal set-membership filtering approach in NCSs is developed. Two ellipsoidal sets are calculated for the system state in two steps: the prediction step and the measurement update step. The attack detection method relies on the intersection of the two sets resulting in two detection criteria. From subalgorithm 2.1.a, the cyber attack on control signals can be detected if the prediction ellipsoidal set does not have intersection with the estimation set updated with the measurement data received at the previous time instant. The subalgorithm.

**Figure 2.6** Intersection between the prediction ellipsoid set $\chi_{k+1|k}$ and the previous updated estimation ellipsoid set $\chi_k$ at $x_3 - x_4$ phase-plane.

From Remark 2.4, since there is no intersection between two sets from $k = 20$ till $k = 35$, then $\chi_{k+1|k} \leftarrow \chi_k$ from $k = 20$ to $k = 35$ and as a result, the estimation ellipsoidal set remains same as its updated set at $k = 20$. Therefore, when the attack finishes at $k = 36$, the center of the prediction ellipsoid set is based on the control signal at $k = 36$ which is free of the attack, and the estimation ellipsoidal set is same as its updated set at $k = 20$ which is calculated from the prediction ellipsoid set obtained at $k = 19$. Since the center of the prediction ellipsoidal set at $k = 19$ is not affected by the attack, from the subalgorithm 2.1.a, there exists intersection between these two sets at $k = 36$, i.e., $\chi_{37|36} \cap \chi_{36} \neq \emptyset$, as shown in Figure 2.7c.

At $k = 19$ the system is free of the attack, so from the subalgorithm 2.1.a it can be concluded that there exists intersection between the two sets as illustrated in Figure 2.7a.

## 2.6 Conclusion

A novel cyber attack detection method based on the ellipsoidal set-membership filtering approach in NCSs is developed. Two ellipsoidal sets are calculated for the system state in two steps: the prediction step and the measurement update step. The attack detection method relies on the intersection of the two sets resulting in two detection criteria. From subalgorithm 2.1.a, the cyber attack on control signals can be detected if the prediction ellipsoidal set does not have intersection with the estimation set updated with the measurement data received at the previous time instant. The subalgorithm.
2.6 Conclusion

Figure 2.7 The prediction ellipsoid $\chi_{k+1|k}$ (green, dotted line), the updated estimation ellipsoid $\chi_k$ (blue, solid line) (a) step 19 (b) step 20 (c) step 36.

2.1.b indicates that the cyber attack on sensors measurements can be detected if there is no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated with the measurement data obtained at the current time instant. A practical system is considered to show the effectiveness of the proposed attack detection method. In practical situations, when the system is subjected to the UBB noises, certain attack detectors such as the widely used $\chi^2$-detector cannot be effective. Therefore, it was of practical importance to provide a novel method that is able to detect cyber attacks on NCSs in which noises are not in the stochastic framework.

The co-authored conference paper entitled “Cyber-physical attacks detection in networked control systems with limited communication bandwidth” has presented a
cyber-physical attack detection algorithm for NCSs subject to limited communication bandwidth. This constraint arises when an attack detection system is located at a remote site and so the required signals, measurement output and control signals, need to be transmitted over a digital communication channel. Therefore, data before being sent to the remote site must be encoded and converted from analog signals to digital signals by using quantizer. A quantizer maps the amount of information from a continuous space to a finite set which is compatible with the limited communication bandwidth. The algorithm consists of a prediction ellipsoid set and an estimation ellipsoid set updated with the quantized measurement output. The so-called “zoomin/zoom-out” uniform quantized strategy is employed to model the quantization process. In this paper, it has been investigated that the uniform quantizer in the system makes the attack detection system sensitive to the quantization levels and zoom-in/zoom-out scaling parameters.
CHAPTER 3

Distributed cyber attack detection

The developed centralized detection method in Chapter 2 may not be straightforwardly applicable for a large-scale NCS since it needs full knowledge of the entire network information, which leads to a high computational overhead for the centralized detection method and hence, it may make the use of this detection method unrealistic. Therefore, in this chapter, a fully distributed detection method is developed for a vehicular platoon system, which represents a large-scale NCS. Furthermore, two recovery mechanisms are introduced to mitigate the adversarial impacts of attacks so that the system can be brought back to the normal condition after detection of the attacks.

This chapter includes the full published version of a co-authored journal paper and summaries of two co-authored conference papers. The bibliographic details of the co-authored papers, including all authors, are:


My contribution to the papers involved: literature review, problem formulation, design, model simulation, model testing, analysis of simulation results, drafting of concluding remarks, manuscript writing and editing.

Signed: ________________________________ Date: 03/01/2020
PhD Candidate: Seyed Eman Mousavinejad (Principal author)

Countersigned: ________________________________ Date: 03/01/2020
Principal Supervisor: Prof. Emer. Ljubo Vlacic

Countersigned: ________________________________ Date: 03/01/2020
Principal Supervisor: A. Prof. Fuwen Yang

Countersigned: ________________________________ Date: 03/01/2020
Associate Supervisor: Prof. Qing Long Han (Corresponding author)

Countersigned: ________________________________ Date: 03/01/2020
Co-author: Dr. Xiaohua Ge
3.1 Abstract

This paper is concerned with the distributed attack detection and recovery in a vehicle platooning control system, wherein inter-vehicle information is propagated via a wireless communication network. An active adversary may launch malicious cyber attacks to compromise both sensor measurements and control command data due to the openness of the wireless communication. First, a distributed attack detection algorithm is developed to identify any of those attacks. The core of the algorithm lies in that each designed filter can provide two ellipsoidal sets: a state prediction set and a state estimation set. Whether a filter can detect the occurrence of such an attack is determined by the existence of intersection between these two sets. Second, two recovery mechanisms are put forward, through which the adversarial effects of cyber attacks can be mitigated in a timely manner. The recovery mechanisms depend on reliable modifications of the attacked signals required for the computation of the two ellipsoidal sets. Finally, simulation is provided to validate the effectiveness of the proposed method in both detection and recovery phases.

3.2 Introduction

With the growing demand for mobility and development of urbanization, the number of vehicles has been significantly increased in recent years. As a result, there has been mounting concern about modern transportation systems due to traffic congestion, traffic accidents, energy waste, and pollution. To deal with these issues, developing Intelligent Transportation Systems (ITS) technologies for the driving pattern change from individual driving to platoon-based driving is of utmost importance. As an advanced automated technology, vehicle platooning is aimed at steering a team of vehicles into a platoon on a road where all vehicles’ speed can be automatically adjusted such that the inter-vehicle distance is reduced while not compromising safety.

The Cooperative Adaptive Cruise Control (CACC), which extends the conventional Adaptive Cruise Control (ACC) technology [71], is deemed as one of the most developed ITS technologies employed in vehicle platooning systems. The objective of CACC is to guarantee that all vehicles in the platoon achieve a prescribed leader vehicle’s velocity/acceleration and meanwhile preserve a desirable inter-vehicle distance. To fulfil this, vehicles normally are equipped with suitable on-board sensors, such as radar, camera, and lidar, enabling each of them to monitor its predecessor’s states. Moreover,
vehicles exchange their inter-vehicle data used by CACC among their neighbors via short-range to medium-range wireless communication channels so as to coordinate their behaviors. From this perspective, a platoon-based vehicular control system can be regarded as a multi-agent system where vehicles represent mobile agents connected together via a VANET.

Recently in ITS industry, considerable research interest has been paid to the impact of network-induced phenomena in vehicle platooning and the improvement of string stability performance, see [72]–[85]. Besides, with the ever-increasing utilization of communication networks in ITS industry, a multitude of studies have investigated the impact of cyber attacks on VANET security because of the openness of the communication networks. Readers are referred to the survey papers [86]–[90] and many references therein for some latest results. Generally, there exist mainly two types of cyber attacks which can deteriorate the performance of VANET, namely DoS attacks and deception attacks [45], [91]. DoS attacks are aimed at jamming the communication channels in order to prevent information exchange between vehicles. From a technological perspective, an adversary can perpetrate DoS attacks through disrupting the radio frequencies on wireless communication channels, which leads to congestion of those channels. Platoon behavior under DoS attacks is investigated in [92], where a drone flying above the platoon, namely the adversary, disrupts the platoon by using its limited power. It is shown that the best possible location to carry out an attack with a huge adverse impact is above the second vehicle and the impact decreases when the adversary moves down in the string. A real-time detection scheme for DoS attacks is presented in [93]. The effects of the DoS attacks are modeled by time-delayed data transmission via a communication network and the attack is diagnosed through tracking the delay in information processing by a set of observers. Deception attacks, on the other hand, represent a kind of attempt to violate the integrity of sensor and/or control data, thus manipulating the vehicle toward the desired behavior of the adversary. To launch a successful deception attack, the adversary usually has sufficient knowledge of the information transmitted through VANET in real time. For example, an adversary may be a trusted insider, i.e., an authenticated member of the platoon, so that the adversary can discard the pseudonymous certificates and digital signatures employed to protect the content of beacon messages. Two typical types of deception attacks on VANET are message falsification attacks [94]–[96] and replay attacks [86], [87]. In message falsification attacks, an adversary starts collecting the information transmitted through the wireless medium and simultaneously modifies the content
meaningfully such that the adversary can severely damage the string stability by rebroadcasting bogus information. The impacts of cyber attacks and sensor tampering are described in [94] through launching a message falsification attack by an external vehicle into the platoon. It is shown that degrading the CACC strategy to ACC can be a potential countermeasure to this attack. In [95], a two-component controller consisting of a radar-based Proportional-Derivative (PD) feedback component and a communication-based feedforward component is proposed to design an attack detection strategy. The proposed attack detection algorithm is based on the estimate of the expected behavior of the front vehicle and the switching policy to ACC if any abnormal behavior exists. String stability of a vehicular platoon under message falsification attacks is investigated in [96], where an attacker is able to force the platoon to oscillate at a resonant frequency by modifying the control signal through which he may cause a serious accident. Replay attacks, however, do not require prior knowledge of the system components. It can be easily implemented by an adversary in two phases. In the first phase, which is known as disclosure phase, the adversary overhears and stores the packets transmitted over communication channels without injecting any malicious input to the system. In the second phase, the adversary manipulates the system by replaying the recorded data as if they are new packets received from those tampered channels. Even though the content of the vehicle data packets is not falsified by the replay attacks, any outdated information may mislead the platoon members, having a great potential to destroy the platoon stability. For example, the attacker may modify a following vehicle’s control command, which contains the desired acceleration received from its predecessor, by some outdated information with the aim of speeding up the following vehicle and causing a collision.

The lack of attention to detection of cyber attacks and recovery of the vehicular platoon system has led the vehicle industry to raise concern about the risks of cyber attacks in this system, which constitutes one of the most challenging issues when a vehicle platoon control system is commercialized in highways [97]. Therefore, the ongoing research on vehicle platooning includes significant efforts to develop effective attacks detection and protection strategies. This motivates us to develop an attack detection strategy and its associated recovery mechanism so as to not only detect attacks that maliciously disrupt the performance of VANET but also recuperate the system from the detected attacks in a timely fashion.

Note that most of attack detection approaches [10], [11], [16]–[22], [52], [98], [99], particularly the $\chi^2$-detector based on the celebrated Kalman filters, require system
noises in a stochastic framework. Whereas, a stochastic noise model requires particular statistical properties of the noise such as known mean and covariance, which may be unrealistic in some practical situations in vehicular systems. For example, acceleration of a vehicle is always deterministically bounded because of the engine dynamics and construction of the vehicle, but becomes uncertain when the vehicle is running. In this case, noise on control data and sensor measurements of a vehicle can be assumed to be UBB. Indeed, an assumption of UBB noise eliminates prior knowledge of the accurate statistical characteristics of noise because only the knowledge of a bound on the realization is needed. Therefore, an alternative filtering method called set-membership filtering is developed in the literature [53]. This filtering method aims to calculate a bounding ellipsoidal state estimation set in state space, which always encloses the true state of the system [54], [55]. During the past decade, such a filtering method has been intensively studied for different problem formulations of various system models (see [56]–[58], [61], [63]).

Motivated by the discussion above, the main objective of this paper is twofold. 1) A distributed attack detection issue for vehicle platooning will be addressed. In our preliminary work [100], centralized cyber attack detection method based on ellipsoidal filtering has been developed for general linear time-varying systems. However, this method is not straightforwardly applicable for a large-scale vehicular platoon because it requires full knowledge of the entire platoon information. Furthermore, the computational overhead for this detection method is quite high, thereby rendering the detection system inadvisable. To overcome these limitations, a distributed attack detection method will be developed in this paper in the sense that each vehicle is equipped with its own detection system. More specifically, each vehicle has its own filter to estimate the state of the vehicle without having the full knowledge from other agents states, which helps apply such a system to a large-scale platoon and decrease the computational overhead. 2) It is our intention to introduce two recovery mechanisms to mitigate the adversarial impacts of the attacks on the performance of the vehicle platooning system, which represents a distinct difference between this paper and our preliminary work [100]. With these two recovery mechanisms, the system can be brought back to the normal condition after detection of the attacks. In other words, the recovery mechanisms will make the state estimation secure against the attacks and then the controller will use the secure estimation to generate the desired control command.
The major contribution is threefold. 1) A set-membership filtering technique in a distributed framework is developed such that each vehicle can determine a group of confidence ellipsoids and each vehicle’s true state always resides in bounding ellipsoidal sets regardless of UBB process noise and measurement noise, providing the vehicle is free of any attack. In particular, each vehicle is equipped with a set-membership filter constructing two ellipsoidal sets: a prediction set and an estimation set. 2) A refined attack detection method is proposed to discern when the occurrence of an attack can be detected and alarmed. The alarm is triggered once there exists no intersection between the two ellipsoidal sets. When the system is free of attacks, the two sets must both enclose the true state, thus guaranteeing the existence of the intersection. When the system is subject to attacks, however, the center of at least one of the two sets is biased by the attacks, which excludes the intersection between the ellipsoidal sets. 3) Two recovery mechanisms are introduced into the proposed algorithm so that each ellipsoidal set will adopt the compensated state prediction and/or state estimation to increase the resilience of estimation system. In other words, the recovery mechanisms promise reliable modifications of the attacked signals required for the computation of the prediction and estimation ellipsoidal sets. Based on the proposed recovery mechanisms, the controllers in CACC system are able to alleviate the adversarial effect of an attack through accessing the secure state estimation.

The rest of this paper is organized as follows. Section 3.3 formulates the longitudinal vehicle dynamics and the control strategy, which together form the CACC-equipped vehicle that will be employed in this paper. Besides, the models of some potential cyber attacks that will be studied in this paper are constructed in this section. The attack detection criteria under a two-step set-membership filtering method are also provided in this section. Section 3.4 presents the design of the prediction and updated estimation ellipsoidal sets, and then the associated attack detection algorithm and recovery mechanisms in a distributed manner are proposed at the end of this section. The simulation results are demonstrated in Section 3.5 to illustrate the effectiveness of the proposed method. Finally, Section 3.6 concludes the paper.

**Notation:** The notation $X > 0$ means that $X$ is “positive definite”. The superscript $T$ stands for matrix transposition. The notation $\text{Tr}(P)$ denotes the trace of $P$. Other notations in this paper are quite standard.

**Ellipsoid:** An ellipsoidal set is denoted as $\mathcal{X} \triangleq \{\zeta \in \mathbb{R}^n : (\zeta - c)^T P^{-1} (\zeta - c) \leq 1\}$, where $c \in \mathbb{R}^n$ is the center and $P = P^T > 0$ is the shape matrix of the ellipsoid. Let $E \in \mathbb{R}^{n \times m}$ with $\text{rank}(E) = m \leq n$ be a lower triangular matrix whose every diagonal
3.3 Problem formulation

3.3.1 CACC law

A platoon-based vehicular control system with some possible attack points is shown in Figure 3.1. To model the dynamics of each vehicle in the string, the linear longitudinal dynamic model presented in [75] is adopted for each vehicle in the string as

$$q_i = v_i, \quad \dot{v}_i = a_i, \quad \dot{a}_i = -\xi_i^{-1}a_i + \xi_i^{-1}u_i, \quad i \in \mathbb{N}_{[1,n]}$$  (3.1)

where $q_i$, $v_i$, $a_i$ denote the absolute position, velocity, and acceleration of vehicle $i$, respectively; $\xi_i$ is the internal actuator dynamics parameter; and $u_i$ represents the commanded acceleration for vehicle $i$. Let us consider vehicle $i$’s state variables as $x_i = [d_i, v_i, a_i, v_{i-1} - v_i, a_{i-1} - a_i]^T \in \mathbb{R}^n$, where $d_i$, $v_{i-1} - v_i$, and $a_{i-1} - a_i$ represent the inter-vehicle distance between two consecutive vehicles, their relative velocity, and acceleration, respectively.

Considering process and measurement noises in a UBB sequence for each vehicle in the string, the discrete-time dynamics model of the $i$-th CACC-equipped vehicle at sampling instants $t_k = k\tau_s, k \in \mathbb{Z}^+$ can be expressed as

$$x_{k+1}^i = A_ix_k^i + B_{s,i}u_k^i + B_{c,i}v_k^{i-1} + B_{w,i}w_k^i$$  (3.2)
where \( A_i \) is the state matrix of vehicle \( i \); \( B_{s,i} \) denotes the input matrix of vehicle \( i \); and \( B_{c,i} \) represents the input matrix from vehicle \( i - 1 \) through the wireless communication network. These matrices can be obtained by

\[
A_i = e^{\bar{A}_i \tau_s}, \quad B_{s,i} = \int_0^{\tau_s} e^{\bar{A}_s s} d\bar{B}_{s,i}, \quad B_{c,i} = \int_0^{\tau_s} e^{\bar{A}_s s} d\bar{B}_{c,i}
\]  

(3.3)

with

\[
\bar{A}_i = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\xi_i^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\xi_i^{-1} & 0 \end{bmatrix}, \quad \bar{B}_{s,i} = \begin{bmatrix} 0 \\ 0 \\ \xi_i^{-1} \\ 0 \\ 0 \end{bmatrix}, \quad \bar{B}_{c,i} = \begin{bmatrix} 0 \\ 0 \\ -\xi_i^{-1} \\ 0 \\ \xi_i^{-1} \end{bmatrix}.
\]

The measurement output obtained from each individual distributed sensor on each vehicle is given by

\[
y^i_k = C_i x^i_k + D_i \nu^i_k
\]

(3.4)

where \( y^i_k \in \mathbb{R}^{n_y} \) is the measurement on sensor \( i \) which is based on the inter-vehicle distance, the vehicle absolute velocity, and the relative velocity; \( B_{w,i}, C_i \) and \( D_i \) are known matrices with appropriate dimensions; \( w^i_k \in \mathbb{R}^{n_w} \) and \( \nu^i_k \in \mathbb{R}^{n_v} \) are the process noise and the measurement noise of vehicle \( i \), respectively, which satisfy the following assumption.

**Assumption 3.1:** The noises \( w^i_k \) and \( \nu^i_k \) are UBB and confined to the following specified ellipsoids

\[
W^i_k \triangleq \{ w^i_k : w^i_k W^i_k^{-1} w^i_k \leq 1 \}
\]

(3.5)

\[
V^i_k \triangleq \{ \nu^i_k : \nu^i_k V^i_k^{-1} \nu^i_k \leq 1 \}
\]

(3.6)

where \( W^i_k = W^i_k^T > 0 \) and \( V^i_k = V^i_k^T > 0 \) are known matrices with compatible dimensions.

**Remark 3.1:** A reference vehicle (denoted by index \( i = 0 \) and with state \( x_0^0 \)) must be introduced as a trajectory generator in the lead vehicle in case there are no preceding vehicles. The discrete-time dynamics of the reference vehicle can be described by

\[
x^0_{k+1} = A_0 x^0_k + B_{s,0} u^0_k
\]

(3.7)
where \( x_0^k = [v_0^k \ a_0^k]^T \), \( u^r_k = u_r(t_k) \) is the sampled reference acceleration profile at the sampling instants \( t_k = k\tau_s \), and \( A_0 \) and \( B_{s,0} \) are calculated in the same method as \( A_i \) and \( B_{s,i} \) in (3.3) when \( i = 0 \), with

\[
\tilde{A}_0 = \begin{bmatrix} 0 & 1 \\ 0 & -\xi_0^{-1} \end{bmatrix}, \quad \tilde{B}_{s,0} = \begin{bmatrix} 0 \\ \xi_0^{-1} \end{bmatrix}.
\]

In practice, the reference vehicle normally serves as a command generator, thus not being affected by the followers. For this purpose, it is assumed that the reference vehicle’s dynamics are not subject to UBB process noise but deterministic to the leader. In this sense, the leader can be informed by the reference vehicle. It should be also pointed out that the lead vehicle (denoted by index \( i = 1 \) and with state \( x_1^k \)) in the string requires special consideration.

Assumption 3.2: The reference acceleration profile \( u_r^k \) is locally available to the lead vehicle without any network-induced imperfection or any cyber attack impact.

It is noteworthy that Assumption 3.2 is mild since \( u_r^k \) is generated locally by the lead vehicle. Therefore, the dynamics of the lead vehicle can be represented by (3.2) with \( i = 1 \) and \( u_{i-1}^k = u_0^k = u_r^k \).

Assumption 3.3: The platoon is composed of homogenous vehicles whose longitudinal dynamic properties are identical.

In actual traffic with different types of vehicles, Assumption 3.3 can be achieved through adequately designed pre-compensators, i.e., low-level acceleration controllers. Hence, for a homogenous platoon, the index \( i \in \mathbb{N}_{[1,n]} \) can be omitted from the system’s matrices \( A_i, B_{s,i}, B_{c,i}, C_i, B_{w,i}, D_i \) and the vehicle parameter \( \xi_i \) in the rest of the paper.

In this study, the control strategy is similar to the one that is presented in [75], which is compatible with the structure shown in Figure 3.1. This control strategy requires a combination of a communication-based forward component \( u_{ff,k} \) and a measurement-based feedback component \( u_{fb,k} \). Thus, the total control signal can be written as

\[
u_i^k = u_{ff,k} + u_{fb,k} \quad (3.8)\]

with the discretized feedforward control signal in the form of

\[
u_{ff,k+1}^i = (1 - \tau_d h_{d}^{-1}) u_{ff,k}^i + \tau_d h_{d}^{-1} u_{i-1}^k \quad (3.9)\]
where $h_d$ is the headway-time for vehicle $i$ to arrive at the same position as its predecessor under a Constant Time Headway (CTH) spacing policy [75].

Different from [75], which utilizes the state measured by sensors directly into calculation of the control signal’s feedback component, we are interested in employing the state estimation. This is because in real-world applications, the vehicle system’s state may not be fully available or not be implicitly trusted due to the restricted sensor perceiving and computing capabilities or faulty sensors. However, full availability of state information is vital for detection of an attack and its integrity is crucial in recovery of the violated system in a timely manner. Motivated by this discussion, the feedback signal can be expressed in an observer-based state feedback protocol as

$$u_{fb,k}^i = K\Omega\hat{x}_k^i$$

where $K = [k_p \ k_d]$ with $k_d = \omega_c$ and $k_p = \omega_c^2$, $\omega_c$ is the bandwidth of the controller and is chosen such that $\omega_c \ll 1/\xi$, $\Omega = \begin{bmatrix} 1 & -h_d & 0 & 0 \\ 0 & 0 & -h_d & 1 \end{bmatrix}$, and $\hat{x}_k^i$ is the estimation of state $x_k^i$.

Note that the control signal is sampled at the sampling instants $t_k = k\tau_s$ and then transmitted over the wireless network. Hence, the sampled and transmitted data may typically experience some network-induced constraints, such as data communication delay, data packet dropouts/disorders, channel fading, and quantization effects [1]. In this paper, however, we pose the following assumption on the wireless communication channels, which allows us to focus on the cyber attack detection and protection issue.

**Assumption 3.4:** The wireless communication network is free of network-induced constraints but suffers from cyber threats.

**Remark 3.2:** The control objective is to regulate the spacing error $e_i = d_i - d_{r,i}$ such that $e_i$ goes to zero asymptotically, where the desired spacing between two consecutive vehicles under CTH policy is represented as

$$d_{r,i} = h_d v_i.$$ 

The spacing error is produced by the leader’s acceleration and deceleration. **String stability** requires spacing error attenuation as vehicles move upstream in the string. This phenomenon is well investigated in [75] and tested experimentally for the adopted control strategy (3.8). The time domain definition of string stability [92] is expressed...
3.3 Problem formulation

as

\[
\max_{t_k} \| e_n(t_k) \| < \max_{t_k} \| e_{n-1}(t_k) \| < \ldots < \max_{t_k} \| e_1(t_k) \|. \tag{3.11}
\]

Since string stability may be degraded in the presence of cyber attacks, a detection method will be introduced such that the attack can be detected in a timely manner and then, two recovery mechanisms will be considered so as to provide the control unit with the secure state prediction and estimation required to calculate the control signal. In this manner, the string maintains its normal performance and the string stability is satisfied in the existence of cyber attacks, which will be further investigated in Section 3.5.

3.3.2 Cyber attack model

In this paper, we investigate the scenario that there exists an adversary whose aim is to manipulate the data transmitted through the communication network, namely, \( \Upsilon_k^i \) which represents either the inter-vehicle signal or the sensor measurement output of vehicle \( i \). The under-attack signal \( \tilde{\Upsilon}_k^i \in \mathbb{R}^q \) on vehicle \( i \) can be expressed as

\[
\tilde{\Upsilon}_k^i = \Upsilon_k^i + \Lambda_k^i \sigma_k^i \tag{3.12}
\]

where \( \sigma_k^i \in \mathbb{R}^q \) is an attack signal carefully designed by the adversary and \( \Lambda_k^i \in \mathbb{R}^{q \times q} \) represents the physical constraints imposed on the attack signals and is assumed to be of the following diagonal structure

\[
\begin{aligned}
\Lambda_k^i &= \text{diag}\{\lambda_{1,k}^i, \lambda_{2,k}^i, \ldots, \lambda_{q,k}^i\} \\
\Lambda_{j,k}^i &\leq \lambda_{j,k}^i \leq \Lambda_{j,k}^i, \quad j = 1, 2, \ldots, q
\end{aligned} \tag{3.13}
\]

where \( \Lambda_{j,k}^i, \Lambda_{j,k}^i \in [0, 1] \) are the scalars representing lower- and upper-bounds on \( \lambda_{j,k}^i \).

**Remark 3.3:** In engineering practice, a realistic adversary may suffer from some physical constraints, such as device saturations, finite battery, and limited interference capacities. This means that the adversary may not be able to launch the attacks at all times. For this reason, a diagonal matrix \( \Lambda_k^i \) is exposed on the attack signal, as shown in (3.12). Note that the scalar parameters \( \lambda_{j,k}^i \) can also be interpreted as the healthy status of the \( j \)-th component of the transmitted data under attacks. Specifically, if \( \lambda_{j,k}^i = 0 \), there is no attack on the \( j \)-th component of the transmitted data at time \( k \) and therefore, this component is securely transmitted. When \( 0 < \lambda_{j,k}^i < 1 \), the \( j \)-th
component is partially contaminated by the attack signal. The case \( \lambda^i_{j,k} = 1 \) represents the worst case scenario that the \( j \)-th component is fully corrupted by the attack signal.

**Remark 3.4:** The under-attack signal model in (3.12) includes different attack strategies as its special cases. For example, if the attack signal \( \sigma^i_k = -\Upsilon^i_k \), there exist DoS attacks compromising the signal at time \( k \). For replay attacks, it is assumed that the attacker can record sequences of data from a signal \( \Upsilon^i_k \) from \( k_o \) till \( k_r \) with the window size \( \varsigma = k_r - k_o \) in the first phase. In the second phase, the attacker replays the recorded data to the system from \( k = k_r + d \) till the end of the attack at \( k = k_f \), where \( d \) is the delay between the recording time and the replaying time. Hence, the replay attacks can be implemented if \( \sigma^i_k = \Upsilon^i_{k-\varsigma} - \Upsilon^i_k \). The case \( \sigma^i_k = \delta^i_k \), where \( \delta^i_k \) is any arbitrary signal chosen by the attacker, represents message falsification attacks.

### 3.3.3 Set-membership state estimation problem

Before proceeding further, let us provide some definitions about set-membership state estimation which is the foundation of the cyber attack detection algorithm. In what follows, a state estimation problem is formulated for each vehicle such that its state prediction and estimation ellipsoids can be calculated to guarantee the enclosing of the true vehicle state in an attack-free scenario.

The state estimation problem is formulated at two steps described as follows.

**Prediction Step:** We are interested in constructing a set-membership filter which runs the following prediction

\[
\hat{x}^i_{k+1|k} = \hat{A}^i_k \hat{x}^i_k + B_s u^i_k + B_c u^{-1}_k
\]

(3.14)

where \( \hat{A}^i_k \) is the filter gain matrix sequence to be determined. The system described by (3.2) and (3.4) is said to achieve set-membership state estimation at the prediction step if there exists \( \hat{A}^i_k \) such that each vehicle’s state \( x^i_{k+1} \) resides in a prediction ellipsoid \( \mathcal{X}^i_{k+1|k} \), which always contains the true state of vehicle \( i \), where

\[
\mathcal{X}^i_{k+1|k} \triangleq \{ x^i_{k+1} : (x^i_{k+1} - \hat{x}^i_{k+1|k})^T P^{i-1}_{k+1|k} (x^i_{k+1} - \hat{x}^i_{k+1|k}) \leq 1 \}
\]

(3.15)

for any values of the process noise \( w^i_k \in \mathcal{W}^i_k \) and the measurement noise \( v^i_k \in \mathcal{V}^i_k \).

**Measurement Update Step:** The set-membership filter updates its state based on the current sensor measurement of vehicle \( i \). More specifically, the state update of the
filter is given in the form of
\[
\hat{x}_{k+1}^i = \hat{x}_{k+1|k}^i + \hat{B}_{k+1}^i (y_{k+1}^i - \hat{y}_{k+1|k}^i)
\]  
(3.16)
where $\hat{B}_{k+1}^i$ is the filter gain matrix sequence to be determined and $\hat{y}_{k+1|k}^i = C\hat{x}_{k+1|k}^i$.

The system described by (3.2) and (3.4) is said to achieve set-membership state estimation at the measurement update step if there exists $\hat{B}_{k+1}^i$ such that each vehicle’s state $x_{k+1}^i$ resides in an estimation ellipsoid $X_{k+1}^i$, which always contains the true state of vehicle $i$, where
\[
X_{k+1}^i \triangleq \{ x_{k+1}^i : (x_{k+1}^i - \hat{x}_{k+1}^i)^T P_{k+1}^{-1} (x_{k+1}^i - \hat{x}_{k+1}^i) \leq 1 \} 
\]  
(3.17)
whenever the equality
\[
y_{k+1}^i = C\hat{x}_{k+1|k}^i + CE_{k+1|k}^i \vartheta_2 + D\nu_{k+1}^i
\]  
(3.18)
holds for some $\| \vartheta_2 \| \leq 1$ and any values of the process noise $w_i^k \in W_i^k$ and the measurement noise $\nu_i^k \in V_i^k$.

The initial state $x_0^i$ satisfies the following assumption.

**Assumption 3.5:** The initial state of vehicle $i$ belongs to the given ellipsoid
\[
X_0^i \triangleq \{ x_0^i : (x_0^i - \hat{x}_0^i)^T P_0^{-1} (x_0^i - \hat{x}_0^i) \leq 1 \}.
\]  
(3.19)

### 3.3.4 Cyber attack detection criteria

To achieve successful attack detection, the following detection criteria are developed to identify attacks violating the data measured by each vehicle’s sensors in the platoon and the inter-vehicle data transmitted through the communication network.

1. If the prediction ellipsoidal set $X_{k+1|k}^i$ has no intersection with the previous updated ellipsoidal set $X_k^i$, i.e., $X_k^i \cap X_{k+1|k}^i = \emptyset$, it indicates that the communication network channel transmitting inter-vehicle data is affected by the attack.

2. If the updated ellipsoidal set $X_{k+1}^i$ has no intersection with the prediction ellipsoidal set $X_{k+1|k}^i$, i.e., $X_{k+1}^i \cap X_{k+1|k}^i = \emptyset$, it indicates that the sensor measurement of vehicle $i$ is affected by the attack.
Remark 3.5: In the case that an attack is present, there exists one among the two sets not enclosing the true state because the center of that set is biased by the attack. From (3.14), if the attack violates the network channel that transmits the inter-vehicle signal $u_{i-1}^k$ sent from vehicle $i-1$ to vehicle $i$, at time $k$, the center of the $i$-th prediction ellipsoidal set $X_{i|k}^{i+1}$ is affected by the attack. Hence, the condition (3.15) may not be satisfied, which leads the prediction ellipsoidal set not to include the true state of the system. However, the previously updated estimation ellipsoidal set $X_k^i$ is based on its prediction ellipsoidal set at time $k-1$ and thus not affected by the attack. Thus, one can conclude that if there is no intersection between $X_{i|k}^{i+1}$ and $X_k^i$, the transmitted signal $u_{i-1}^k$ and its corresponding network channel are compromised by an adversary.

Remark 3.6: If the sensor of vehicle $i$ is manipulated by an adversary at time $k+1$, its current measurement output $y_{i,k+1}$ is violated. Hence, from (3.16), the center of the $i$-th estimation ellipsoidal set $X_{k+1}^i$, which is updated with the current measurement, is affected. This renders the condition (3.17) unsatisfied and further leads the estimation ellipsoidal set not to confine the true state of the system. However, the $i$-th prediction ellipsoidal set $X_{i|k}^{i+1}$, which is based on the measurement at time $k$, is not affected by the attack. Thus, it can be indicated that if there is no intersection between $X_{i|k}^{i+1}$ and $X_{i+1}^i$, the sensor measurement of vehicle $i$ is compromised by an attack.

Remark 3.7: Note that attacks can be carefully designed to be undetectable and stealthy. For example, if the attacks have slow dynamics, such attacks are difficult to be distinguished from model uncertainty and noise. Therefore, most of attack detection approaches presented in the literature involve a threshold and if the abrupt changes are below of this threshold, they cannot be detected and thus, remain undetected [101]. Also, if malicious falsification delivered by an attacker is fairly small, the proposed detection strategy may not be able to recognize the attack in a timely fashion, i.e., there might be an intersection between the two ellipsoidal sets. Thus, the size of the ellipsoidal sets must be minimized, which necessitates the usage of an optimization approach aimed at minimizing the threshold. This concept will be outlined in Section 3.4.2. In other words, the optimization approach in our proposed attack detection algorithm is aimed at softening the adversarial impact of attacks which may be restored by a well-designed resilient control strategy. The main trust of this paper is in its attack detection strategy while the design of a resilient control method constitutes one of our future research work.
3.4 Distributed attack detection and recovery mechanism design

3.4.1 Design criteria

We first present the following theorems, which establish sufficient conditions on the existence of the prediction and estimation ellipsoidal sets that guarantee to contain the true state of the vehicle.

**Theorem 3.1:** For vehicle $i$ described by (3.2) and (3.4) subject to UBB noises $w^i_k \in W^i_k$ and $\nu^i_k \in V^i_k$, suppose that at time $k$ the vehicle’s state $x^i_k$ belongs to its state estimation ellipsoid $(x^i_k - \hat{x}^i_k)^T P^{-1}_i (x^i_k - \hat{x}^i_k) \leq 1$. Then the one-step ahead state $x^i_{k+1}$ resides in its state prediction ellipsoid $(x^i_{k+1} - \hat{x}^i_{k+1})^T P^{-1}_i (x^i_{k+1} - \hat{x}^i_{k+1}) \leq 1$, if there exist matrix sequences $P^i_{k+1|k} > 0$, $\hat{A}^i_k$, and scalar sequences $\tau^i_{m,k} > 0$, $m = 1, 2$ such that

$$\begin{bmatrix}
-P^i_{k+1|k} & \Theta^i_1,k & A E^i_k & B w \\
* & \Theta^i_2,k & 0 & 0 \\
* & * & -\tau^i_{2,k} I & 0 \\
* & * & * & \Theta^i_3,k
\end{bmatrix} \leq 0$$

(3.20)

where

$$\Theta^i_1,k = (A - \hat{A}^i_k) \hat{x}^i_k, \quad \Theta^i_2,k = -1 + \tau^i_{1,k} + \tau^i_{2,k}, \quad \Theta^i_3,k = -\tau^i_{1,k} W^{-1}_i.$$

**Proof:** See the Appendix A.1. □

**Theorem 3.2:** For vehicle $i$ described by (3.2) and (3.4) subject to UBB noises $w^i_k \in W^i_k$ and $\nu^i_k \in V^i_k$, suppose that the one-step ahead state $x^i_{k+1}$ belongs to its state prediction ellipsoid $(x^i_{k+1} - \hat{x}^i_{k+1})^T P^{-1}_{k+1|k} (x^i_{k+1} - \hat{x}^i_{k+1}) \leq 1$. Then such a state resides in its state estimation ellipsoid $(x^i_{k+1} - \hat{x}^i_{k+1})^T P^{-1}_{k+1|k} (x^i_{k+1} - \hat{x}^i_{k+1}) \leq 1$, which is updated with the measurement output $y^i_{k+1}$, if there exist matrix sequences $P^i_{k+1} > 0$, $\hat{B}^i_{k+1}$, $N^i_{k+1}$, and scalar sequences $\tau^i_{m,k} > 0$, $m = 3, 4$ such that

$$\begin{bmatrix}
-P^i_{k+1} & 0 & \Theta^i_1,k & -\hat{B}^i_{k+1} D \\
* & \Theta^i_2,k & \Theta^i_3,k & \Theta^i_4,k \\
* & * & \Theta^i_5,k & \Theta^i_6,k \\
* & * & * & \Theta^i_7,k
\end{bmatrix} \leq 0$$

(3.21)
Distributed cyber attack detection

where

\[ \Theta_{1,k} = (I - \hat{B}_{k+1} C) E_{k+1|k}, \]
\[ \Theta_{2,k} = -1 + \tau_{4,k} + \tau_{4,k} + F_{k}[1, 1], \]
\[ \Theta_{3,k} = F_{k}[\alpha_{1}, \beta_{1}] \alpha_{1}=1, \beta_{2} \in \mathbb{N}[2, 7], \]
\[ \Theta_{4,k} = F_{k}[1, 8], \]
\[ \Theta_{5,k} = -\tau_{4,k} I + F_{k}[\alpha_{2}, \beta_{2}] \alpha_{2} \in \mathbb{N}[2, 7], \beta_{2} \in \mathbb{N}[2, 7], \]
\[ \Theta_{6,k} = F_{k}[\alpha_{3}, \beta_{3}] \alpha_{3} \in \mathbb{N}[2, 7], \beta_{3} = 8, \]
\[ \Theta_{7,k} = -\tau_{3,k} V_{k}^{-1}; \]
\[ F_{k} = N_{k+1} \Pi_{y,k+1} + \Pi_{y,k+1} N_{k+1}; \]
\[ \Pi_{y,k+1} = \left[ C_{k+1|k} \hat{x}_{k+1|k} - y_{k+1} \quad CE_{k+1|k} \quad D \right]. \]

Proof: See the Appendix A.2.

Remark 3.8: It is clearly shown in Theorems 3.1 and 3.2 that the proposed ellipsoidal set-membership filtering problem can be cast into the feasibility problem of a set of RLMIs (3.20) and (3.21). Thus, Theorems 3.1 and 3.2 provide criteria for designing two ellipsoidal sets which always have intersection since they both guarantee to contain the true state \( x_{k+1} \) in an attack-free system.

3.4.2 Recursive convex optimization and cyber attack detection algorithm with recovery mechanism

In light of Theorems 3.1 and 3.2, for the \( i \)-th attack-free vehicle, the one-step ahead state \( x_{k+1} \) always resides in both the state prediction ellipsoidal set \( \mathcal{X}_{k+1|k} \) and its updated state estimation ellipsoidal set \( \mathcal{Y}_{k+1|k} \) if (3.20) and (3.21) hold. Therefore, there exist two vectors \( \vartheta_{k}, \kappa = 1, 2 \), satisfying \( \| \vartheta_{k} \| \leq 1 \) such that \( x_{k+1} = \hat{x}_{k+1} + E_{k+1|k} \vartheta_{1} \), and \( x_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k} \vartheta_{2} \), respectively. Furthermore, the center of the state prediction ellipsoid \( \hat{x}_{k+1} \) and the center of the state estimation ellipsoid \( \hat{x}_{k+1} \) are determined by (3.14) and (3.16), respectively. Thus, Theorems 3.1 and 3.2 outline the principles of determining the state prediction ellipsoidal set and the state estimation ellipsoidal set calculated through updating the prediction set via the current measurement data. However, Theorems 3.1 and 3.2 do not provide optimal state prediction and state estimation ellipsoids. Therefore, a convex optimization approach is applied in (3.22) and (3.23), in which the traces of \( P_{k+1|k} \) and \( P_{k+1} \) are optimized at each time step in an effort to find the prediction ellipsoidal set and the estimation ellipsoidal set with
minimal size.

\[
\begin{aligned}
\text{minimize} & \quad P_{k+1|k}^i > 0, A^i_k, \tau^i_{1,k} > 0, \tau^i_{2,k} > 0, \\
\text{subject to} & \quad (3.20), \\
\end{aligned}
\]

\[
\begin{aligned}
\text{minimize} & \quad \text{Tr}(P_{k+1}^i) \\
\text{subject to} & \quad (3.21).
\end{aligned}
\]

**Remark 3.9:** In Theorems 3.1 and 3.2, one can observe that RLMIs (3.20) and (3.21) are linear to \(P_{k+1|k}^i, A^i_k, \hat{A}^i_k, N^i_{k+1}, \tau^i_{m,k}, m = 1, 2, \ldots, 4\). As a result, the OPs (3.22) and (3.23) can be solved by some existing semi-definite programming via an interior-point algorithm at each time step. In practice, interior-point methods for semi-definite programs are competitive with other methods for small programs and substantially faster for medium and large-scale problems [68].

Based on OPs (3.22) and (3.23), we next present an algorithm that recursively computes the state ellipsoidal sets for each vehicle in the platoon such that cyber attacks can be detected by the existence of intersection between these sets. Besides, two recovery mechanisms are provided to secure each vehicle in the platoon against cyber attacks violating its data measured by sensors and the inter-vehicle data transmitted through the communication network.

**Algorithm 3.1:** Recursive convex optimization algorithm / Attack detection & Recovery mechanism

**Step 1. Initialization**

Given initial conditions \(x_0^i, \hat{x}_0^i, \text{ and } x_0^0\). Choose suitable \(W_k^i, V_k^i, \text{ and } P_0^i\) such that (3.5), (3.6), and (3.19) hold. Calculate \(E_0^i\) according to \(E_0^i = E_0^i E_0^{iT}\).

Set \(k = 0, x_k^i = x_0^i, \hat{x}_k^i = \hat{x}_0^i, x_0^i = x_0^0, \text{ and } E_k^i = E_0^i\). Define the reference acceleration profile \(u_k^i\). Let the simulation run recursively for \(N\) steps.

**Step 2. Prediction**

1. Solve the OP (3.22) to calculate \(P_{k+1|k}^i\) and \(A_k^i\). Obtain \(E_{k+1|k}^i\) such that \(P_{k+1|k}^i = E_{k+1|k}^i E_{k+1|k}^{iT}\).

2. Compute the center of the state prediction ellipsoid \(\hat{x}_{k+1|k}^i\) by (3.14).

**Step 3. Diagnosis of Cyber Attack on Inter-Vehicle Signal**

1. If \(X_{k+1|k}^i \neq \emptyset\), report "No Attack" and go to Step 5.

2. If \(X_{k+1|k}^i = \emptyset\), report "Attack" and go to Step 4.
Step 4. Recovery Mechanism
1. Set \( u_{k-1}^i \leftarrow u_{k-1}^i \).
2. Recompute \( u_{ff,k}^i \) and \( u_k^i \) by (3.9) and (3.8), respectively.
3. Recalculate the center of the state prediction ellipsoid by (3.14) with the new \( u_{k-1}^i \) and \( u_k^i \).
4. Set \( E_{k+1|k}^i \leftarrow E_{k|k-1}^i \).

Step 5. Measurement Update
1. Solve the OP (3.23) to calculate \( P_{k+1}^i \) and \( \hat{B}_{k+1}^i \). Obtain the new \( E_{k+1}^i \) such that \( P_{k+1}^i = E_{k+1}^i E_{k+1}^{TT} \).
2. Compute the center of the updated state estimation ellipsoid \( \hat{x}_{k+1}^i \) by (3.16).

Step 6. Diagnosis of Cyber Attack on Sensor Signal
1. If \( \mathcal{X}_{k+1|k}^i \bigcap \mathcal{X}_{k+1|k}^i \neq \emptyset \), report “No Attack” and go to Step 8.
2. If \( \mathcal{X}_{k+1|k}^i \bigcap \mathcal{X}_{k+1|k}^i = \emptyset \), report “Attack” and go to Step 7.

Step 7. Recovery Mechanism
Set \( x_{k+1}^i \leftarrow x_{k+1|k}^i \) and \( E_{k+1}^i \leftarrow E_{k+1|k}^i \).

Step 8. Loop
If \( k = N \) then Exit, Else \( k \leftarrow k + 1 \) and go to Step 2.

Remark 3.10: The impacts of an attack on the centers of the prediction and estimation ellipsoidal sets result in infeasibility of the conditions (3.15) and (3.17), i.e., they may not contain the true state of the system. Therefore, it is of utmost importance to recover these sets once the attack is detected, which not only ensures the feasibility of those conditions but also preserves the system’s string stability while mitigating the adversarial effects of the attack. To address this, two recovery mechanisms are proposed in the recursive algorithm. 1) In step 4, if \( \mathcal{X}_k^i \bigcap \mathcal{X}_{k+1|k}^i = \emptyset \), the prediction ellipsoidal set must be reobtained through modifying the transmitted signal \( u_{k-1}^i \) with its one-step-behind value. This modification requires a command signal received by the \( i \)-th controller from which the controller recalculates the feedforward control signal \( u_{ff,k}^i \) and sends a new control signal \( u_k^i \) to vehicle \( i \). 2) In step 7, if \( \mathcal{X}_{k+1}^i \bigcap \mathcal{X}_{k+1|k}^i = \emptyset \), the estimation ellipsoidal set must be replaced with its prediction ellipsoidal set, which is still based on the previous measurement output. Therefore, a command signal is sent to the measurement update step so that the replacement occurs in this step. From
3.4 Distributed attack detection and recovery mechanism design

Figure 3.2 Schematic of distributed attack detection with recovery mechanisms for vehicle $i$ under set-membership filtering technique.

Figure 3.2, one can conclude that the $i$-th controller always uses the free-of-attack state prediction and state estimation so as to generate the control signal for vehicle $i$ through which the string stability can be satisfied in the entire platoon.

**Remark 3.11:** The Information Flow Topology (IFT) defines how information is exchanged between connected vehicles. So far, in this paper, the proposed attack detection algorithm and recovery mechanisms have been constructed over a typical IFT, which is the Predecessor Following (PF) topology shown in Figure 3.1. In addition to the PF topology, representative approaches for IFT in the literature [102] can be arguably categorized as the Bi-Directional (BD), Predecessor-Following Leader (PFL), Bi-Directional Leader (BDL), Two Predecessor-Following (TPF), and Two Predecessor-Following Leader (TPFL) topologies. Considering different topologies is currently beyond the scope of this paper. However, the proposed attack detection algorithm and recovery mechanisms have a great potential to be extended for the specific communication topologies aforementioned and therefore, it constitutes one of our future research work.

**Remark 3.12:** Note that faults in the system’s components, i.e., sensors and actuators, commonly exist in many practical engineering systems. For example, consider $f_{1,k}$ and $f_{2,k}$ as a UBB actuator’s fault and a UBB sensor’s fault added into (3.2) and (3.4), respectively. The proposed two-step filter in Section 3.3.3 can now be refined to provide a state prediction ellipsoidal set and a state estimation ellipsoidal set which always contain the true state of the system in the presence of the faults. If the system is affected by an attack with a different pattern of the fault signal, this attack may lead
to empty set intersection and thus can be successfully identified through the proposed
detection method and its impact on string stability can be eliminated by the proposed
recovery mechanisms. However, if the attack hides in the fault (e.g., follows a similar
UBB pattern of the fault signal), the intersection of the two sets may not always be
empty, the proposed detection method fails to distinguish the fault-like attack from
the fault. However, the refined filter is able to tolerate this type of attack in the same
way as it does with respect to the fault. From this perspective, the proposed detection
method is only sensitive to those attacks whose patterns are different from those of
fault signals. As a matter of fact, one key merit of this method is that attack must
occur when it is detected by the designed filter.

3.5 Simulation results and case studies

In this section, the developed attack detection algorithm and the two recovery mecha-
nisms will be applied to a platooning system of five vehicles to validate their effectiveness
under various cases of attacks.

3.5.1 Simulation setup

The platoon consists of one leader \((i = 1)\) and four followers \((i = 2, 3, 4, 5)\) whose
desired acceleration informed by the reference vehicle in the platoon \((i = 0)\) is in the form of
\[
 u_k^r = \begin{cases} 
 1 \text{ m/s}^2, & \text{if } 50 < k < 150 \\
 0, & \text{otherwise},
\end{cases}
\]
and its initial velocity and acceleration are 15 m/s and 0 m/s\(^2\), respectively. The
parameters for each vehicle in the platoon are assumed to be identical, i.e., \(\xi = 0.1,\)
\(h_d = 0.7 \text{ s}\), which indicates a homogeneous vehicle platoon. As discussed in [103],
the bandwidth of the ACC controller must be taken as \(\omega_c = 0.05\xi^{-1}\) to satisfy the
internal stability of the vehicle dynamics. Hence, the PD controller gains are chosen
as \(k_p = 0.25\) and \(k_d = 0.5\). The continuous-time model is discretized with a sampling
period \(\tau_s = 0.1 \text{ s}\).

The initial velocity of each vehicle is taken as 15 m/s with the initial inter-vehicle
distance 10.5 m and zero initial acceleration, which means zero initial spacing er-
ror from the definition of the spacing error in Remark 3.2. Therefore, the initial
condition for all vehicles in the platoon is \(x_0^i = [10.5, 15, 0, 0, 0]^T, i \in \mathbb{N}_{[1,5]}\). For
3.5 Simulation results and case studies

![Graph showing simulation results](image)

Figure 3.3 (a) Intersection between $\chi^{2}_{k+1|k}$ and $\chi^{2}_{k}$ (“1” indicates there exists intersection, “0” indicates there is no intersection); (b) $\chi^{2}_{111|110}$ (green, dotted line), $\chi^{2}_{110}$ (blue, solid line), true state (red, star).

each vehicle, suppose the UBB noises are $w_{k}^{i} = 0.1 \sin(2k)$ and $\nu_{k}^{i} = 0.2 \cos(5k)$ and set $W_{k}^{i} = 2$ and $V_{k}^{i} = 4$. Then it can be easily checked that $w_{k}^{i}$ and $\nu_{k}^{i}$ belong to the ellipsoidal sets defined in (3.5) and (3.6), respectively. Furthermore, let $B_{w} = [0.2, 0.2, 0.1, 0.2, 0.1]^{T}$ and $D = [1, 1, 1]^{T}$. The initial conditions of estimators are chosen as $\hat{x}_{0}^{1} = [10.42, 14.98, 0, 0.02, 0]^{T}$, $\hat{x}_{0}^{2} = [10.431, 14.972, 0, 0.018, 0]^{T}$, $\hat{x}_{0}^{3} = [10.458, 14.96, 0, 0.012, 0]^{T}$, $\hat{x}_{0}^{4} = [10.474, 14.942, 0, 0.018, 0]^{T}$, $\hat{x}_{0}^{5} = [10.498, 14.911, 0, 0.031, 0]^{T}$, which define the centers of initial ellipsoidal sets (3.19) for each vehicle with the shape matrix taken as $P_{0}^{i} = 50I_{5 \times 5}$. The simulation results are obtained during 250 sampling instants.

3.5.2 DoS attack on communication network

In this case, the adversary launches a DoS attack on the control signal $u_{k}^{i}$, which is sent by vehicle $i = 1$ and received by vehicle $i = 2$, through jamming the communication network channel between these two vehicles from $k = 110$ to $k = 130$. In the simulation, the lower- and upper-bounds on $\Lambda_{k}^{i}$ in (3.13) are chosen as $0.8 \leq \Lambda_{k}^{i} \leq 1$. Hence, the actual control signal received by vehicle $i = 2$ is

$$
\begin{align*}
\hat{u}_{k}^{i} &= u_{k}^{i} - \Lambda_{k}^{i} u_{k}^{i}, & \text{if } 110 \leq k \leq 130 \\
\hat{u}_{k}^{i} &= u_{k}^{i}, & \text{otherwise},
\end{align*}
$$

where $\Lambda_{k}^{i} = 1$ indicates that the signal $u_{k}^{i}$ is completely lost during its transmission; $0.8 \leq \Lambda_{k}^{i} < 1$ represents that vehicle $i = 2$ receives up to 20% of the signal $u_{k}^{i}$. 

\[\begin{bmatrix}
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}\]
Figure 3.3a shows the sequences of intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the previous time for vehicle $i = 2$ during the simulation time. Since the DoS attack starts at $k = 110$, the prediction ellipsoidal set for vehicle $i = 2$, i.e., $X^2_{111|110}$, is affected by the attack as its center is obtained from the signal data received from vehicle $i = 1$ at this time. Therefore, from Remark 3.5, it can be concluded that the prediction set $X^2_{111|110}$ does not enclose the true state as shown in Figure 3.3b. However, the attack does not have any influence on the previous updated estimation ellipsoidal set, i.e., $X^2_{110}$, since it is based on its prediction set at time $k = 109$. Thus, from step 3 in the proposed algorithm, it is expected that there must be no intersection between these two sets in the existence of the attack at time $k = 110$, i.e., $X^2_{110} \cap X^2_{111|110} = \emptyset$.

Once the communication channel between vehicles $i = 1$ and $i = 2$ is jammed, vehicle $i = 2$ in the string does not completely receive its desired acceleration from the leader ($i = 1$) and therefore, all its following vehicles’ acceleration in the string are affected by this attack as shown in Figure 3.4a. In consequence, as depicted in Figure 3.4c, the inter-vehicle distance between all followers cannot reach the desired distance defined under CTH policy, which opposes the objective of increasing the traffic throughput. Therefore, from Figure 3.4e, the string stability (3.11) cannot be achieved as $\max_t \|e^3(t_k)\|$ and $\max_t \|e^2(t_k)\|$ are greater than $\max_t \|e^1(t_k)\|$, which indicates that the spacing error does not decrease upstream the string.

The next step after the detection of the attack is to recover the predicted state $\hat{x}^2_{111|110}$ that is used to calculate the estimated state $\hat{x}^2_{111}$. From step 4, the control signal $u^1_{110}$ must be replaced with its one-step-behind value from which the feedforward control signal $u^2_{ff,110}$ is recalculated. Then, the predicted state is reobtained from these modified signals and it will be sent to step 5, the measurement update step, in order to generate the secure estimated state. These modifications are required for every instant until the attack finishes at $k = 130$. With the recovery mechanism, the platoon obtains this capability to maintain its string stability in the existence of the DoS attack since the spacing error is attenuated toward the tail of the platoon as shown in Figure 3.4e. Thus, all vehicles in the string follow the leader’s desired acceleration and reach the safety inter-vehicle distance in accordance with CTH policy, as shown in Figure 3.4b and Figure 3.4d, respectively.
3.5 Simulation results and case studies

Figure 3.4 (a) and (b) Vehicle’s acceleration without and with recovery mechanisms; (c) and (d) Inter-vehicle distance without and with recovery mechanisms; (e) String stability analysis.

3.5.3 Replay attack on sensor data

To manipulate the measurement output with a replay attack, it is considered that the attacker uses his ability to perform a disclosure attack so that he can store the data measured through on-board sensors. In the simulation, it is assumed that he obtains access to vehicle $i = 2$’s sensor through which its velocity is measured and records the signal data from $k = 65$ till $k = 75$. In the second phase of the attack, the attacker modifies the current data of the signal with his recorded data from $k = 105$ to $k = 115$. The lower- and upper-bounds on $\Lambda_k^2$ are chosen as $\text{diag\{0,0.8,0\}} \leq \Lambda_k^2 \leq \text{diag\{0,1,0\}}$, 
which indicates that there is only a replay attack on the second component of the measurement output $y_2^k$, i.e., the measured velocity of vehicle $i = 2$. Therefore, the actual measurement output is

$$\begin{cases} \tilde{y}_2^k = y_2^k + \Lambda_k^2(y_2^k - 40 - y_2^k), & \text{if } 105 \leq k \leq 115 \\ \tilde{y}_2^k = y_2^k, & \text{otherwise} \end{cases}$$

Figure 3.5a demonstrates the sequences of intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time for vehicle $i = 2$ during the simulation time. In particular, as the replay attack starts at $k = 105$, the prediction ellipsoidal set $\mathcal{X}_{105|104}^2$ is calculated from the measurement data obtained at $k = 104$ when there is no attack. However, the estimation ellipsoidal set $\mathcal{X}_{105}^2$ is updated with the current sensor measurement output at $k = 105$, which is compromised by the attack. Therefore, one can conclude from Remark 3.6 that the estimation set $\mathcal{X}_{105}^2$ does not include the true state as depicted in Figure 3.5b. Hence, it is expected from step 6 in the proposed algorithm that there must be no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time when the attack starts at $k = 105$, i.e., $\mathcal{X}_{105}^2 \cap \mathcal{X}_{105|104}^2 = \emptyset$.

Replacing the second vehicle’s velocity from $k = 105$ till $k = 115$ with the recorded value from $k = 65$ till $k = 75$ causes the controller to consider that the driver reduces the velocity. Thus, the controller suddenly increases the acceleration of the vehicle which results in an abrupt increase in the real-time velocities of the vehicle and its followers that may exceed the limitations of vehicle dynamics causing the vehicles to lose their internal stability, as demonstrated in Figure 3.6a. It can be easily seen from Figure 3.6c that the controller fails to mitigate the effect of the attack which causes
Figure 3.6 (a) and (b) Vehicle’s velocity without and with recovery mechanisms; (c) and (d) Inter-vehicle distance without and with recovery mechanisms; (e) String stability analysis.

vehicle $i = 2$ not to be able to properly adjust its inter-vehicle distance based on CTH policy and hence, it crashes into its predecessor. Furthermore, the string stability (3.11) is violated by the replay attack since $\max_{t_k} \|e^i(t_k)\|$, $i = 2, 3, \ldots, 5$ is greater than $\max_{t_k} \|e^1(t_k)\|$ as shown in Figure 3.6e, which contradicts the definition of string stability.

Once the attack is detected, the estimated state $\hat{x}_{105}^2$, which is utilized to generate the control command, must be recovered from the attack. To achieve this, the center and the shape matrix of the estimation ellipsoidal set are replaced with the same properties of its prediction ellipsoidal set as concluded from step 7. In doing so,
the controller uses the state prediction instead of its estimation since the predicted state is secure from the attack. This replacement must be made till the end of the attack at $k = 115$. With the recovery mechanism, the controller receives the secure predicted state through which it can compensate the malicious effect of the attack. As a consequence of the proposed recovery mechanism, the spacing error does not propagate upstream the string as shown in Figure 3.6e, which satisfies the definition of string stability in (3.11) and hence, all vehicles in the string follow the velocity pattern of the leader and reach the desired inter-vehicle distance, as depicted in Figure 3.6b and Figure 3.6d, respectively.

### 3.5.4 Message falsification attack on sensor data

To perform a message falsification attack, it is assumed that the attacker is able to collect the information from vehicle $i = 3$’s sensor through which its distance from vehicle $i = 2$ is measured. Simultaneously, the attacker manipulates the platoon toward his desired performance through adding the signal $\delta^3_k = 0.2 + 0.05 \sin(k)$ into the content of the measurement output from $k = 80$ to $k = 95$. The lower- and upper-bounds on $\Lambda^3_k$ in (3.13) are taken as $\text{diag}\{0.6, 0, 0\} \preceq \Lambda^3_k \preceq \text{diag}\{1, 0, 0\}$, which represents that there only exists a message falsification attack on the first component of the measurement output $y^3_k$, which is the measured distance between vehicles $i = 2$ and $i = 3$. Thus, the actual measurement output is

$$
\begin{cases}
\tilde{y}^3_k = y^3_k + \Lambda^3_k \delta^3_k, & \text{if } 80 \leq k \leq 95 \\
\tilde{y}^3_k = y^3_k, & \text{otherwise},
\end{cases}
$$
Figure 3.7a illustrates the sequences intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time for vehicle $i = 3$ over the simulation time. When the attack starts at $k = 80$, the prediction ellipsoidal set $\mathcal{X}_{80|79}^3$ is calculated from the measurement data received at $k = 79$ when there is no attack. However, the estimation ellipsoidal set $\mathcal{X}_{80}^3$ is updated with the current measurement output at $k = 80$, which is affected by the attack. Therefore, as discussed in Remark 3.6, the estimation set $\mathcal{X}_{80}^3$ does not include the true state as shown in Figure 3.7b. Thus, from step 6 in the proposed algorithm, one can conclude that there must be no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated at the current time when the attack starts at $k = 80$, i.e., $\mathcal{X}_{80}^3 \cap \mathcal{X}_{80|79}^3 = \emptyset$.

Tampering the content of the measured inter-vehicle distance between vehicles $i = 2$ and $i = 3$ with the bogus signal $\delta_k^3$ deceives the controller into considering that there exists an increase in the measured distance. Therefore, the controller generates its command so as to reduce the relative velocity between vehicles $i = 2$ and $i = 3$ as shown in Figure 3.8a, which results in a sudden decrease in the inter-vehicle distance between these vehicles. Hence, vehicle $i = 3$ fails to follow its required inter-vehicle distance and crashes into vehicle $i = 2$ as shown in Figure 3.8c. These abrupt modifications degrade the string stability of the platoon since $\max_{t_k} \| \mathbf{e}(t_k) \|, i = 3, 4, 5$ is greater than $\max_{t_k} \| \mathbf{e}^2(t_k) \|$ as shown in Figure 3.8e.

To tackle this, the estimated state $\hat{x}_{80}^3$ must be recovered from the attack so as to generate the control command by using the secure state estimation. Hence, from step 7 in the proposed algorithm, the center and the shape matrix of the estimation ellipsoidal set $\mathcal{X}_{80}^3$ are replaced with the same properties of its prediction ellipsoidal set $\mathcal{X}_{80|79}^3$ so that the controller employs the predicted state which is secure from the attack. This recovery strategy must be implemented for every instant till the end of the attack at $k = 95$. As a result of the recovery mechanism, the string stability (3.11) is achieved since the spacing error decreases toward the end of the platoon as shown in Figure 3.8e. Moreover, all vehicles in the platoon follow the relative velocity pattern introduced by the leader and reach the desired inter-vehicle distance, as demonstrated in Figure 3.8b and Figure 3.8d, respectively.

### 3.5.5 String stability analysis under different attack locations

In this section, we investigate the performance of the proposed algorithm on string stability while the attacker launches an attack on different vehicles in the string. For
the brevity of presentation, only the DoS attack with the same properties as discussed in Section 3.5.2 is considered over the communication network through which the desired control signal of each vehicle is received by its predecessor in the string. The discussion about the results for a replay attack and a message falsification attack on sensor data of different vehicles in the string is similar to that of the DoS attack and therefore, it is omitted here.

As depicted in Figure 3.9, once the attacker moves toward the tail of the platoon, although the spacing error propagates in the string from vehicle $i$ to vehicle $i+1$, the maximum magnitude of $\max_t \|e^i(t_k)\|$ decreases. This is due to the fact that the
more spacing error is attenuated by CACC controllers as the attacker goes far away from the lead vehicle. Thus, the further attacker moves away from the lead vehicle, the less he is able to destabilize the platoon. However, as shown in Figure 3.9, no matter where the attack is launched, the platoon can achieve its string stability with the proposed recovery mechanisms once the attack is detected at its time of occurrence by the proposed distributed detection method.

3.6 Conclusion

A novel distributed attack detection method has been developed to address the problem of the detection of cyber attacks compromising the shared communication network and on-board sensors employed in a vehicle platooning system. To build a foundation of the attack detection method, a recursive state estimation algorithm characterized by ellipsoidal set-membership filters in a distributed framework is developed. The distributed set-membership filtering algorithm allows each vehicle in the platoon to estimate their states with no need of having full knowledge of the entire platoon, which reduces the computational overhead for this model-based detection method and makes the method suitable for a large-scale platoon. To mitigate the malicious effects of cyber attacks, two recovery mechanisms are introduced into the proposed algorithm. The idea of these two recovery mechanisms is based on reliable modifications of the attacked signals required for the computation of the prediction ellipsoidal set and estimation ellipsoidal set so that each ellipsoidal set will adopt the compensated state prediction and/or state estimation to increase security of the estimation system. The simulation
results demonstrated that with the proposed recovery methods, the controllers in CACC system are able to compensate the adversarial effect of an attack through accessing the secure state estimation so that the string stability of the platoon is satisfied while the attack is present. In the coexistence of the faults in the system, the proposed set-membership state estimation method can be refined as a fault tolerant filter by considering the UBB fault signals in the system dynamics. In doing so, the refined filter can tolerate the impact of the fault-like attack in the same way as it does with respect to the fault.

The co-authored conference paper entitled “Cyber attack detection in platoon-based vehicular networked control systems” has presented a centralized cyber attack detection algorithm for vehicle platooning systems in which the information among vehicles is transmitted through a shared wireless communication network and also each vehicle has access to its own information measured by local sensors. As concluded from this paper, the developed centralized attack detection strategy is not applicable to a large-scale platoon since it requires full knowledge of the entire network information.

The co-authored conference paper entitled “Detection of cyber attacks on leader-following multi-agent systems” has addressed an attack detection problem for a networked leader-following multi-agent system subject to UBB system noises and logarithmic quantization effects, where an adversary launches malicious cyber attacks on agents’ measurement outputs aiming to distrust the leader-following consensus. An effective distributed attack detection algorithm has been developed for each follower such that the attack can be identified at the time of its occurrence. The core of the algorithm lies in a set-membership filtering approach as discussed in this chapter. Furthermore, a convex optimization algorithm has been established to solve out anticipated consensus protocol and two-step set-membership filter by resorting to some RLMIs.
CHAPTER 4

Resilient control and estimation

It is quite common for a crafty adversary to launch assorted attacks of different models and strategies for comprehensively compromising performance of an NCS. It has been well acknowledged that different attack strategies are generally stealthy to any detection method including those developed in Chapters 2 and 3. Motivated by this observation, in this chapter, the focus lies on resilient remote tracking control through a shared communication network. Thus, this chapter investigates the joint problem of resilient tracking control and resilient estimation in NCSs subject to the presence of various cyber attacks that are modeled in a unified framework. This chapter includes the full published version of a co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:


My contribution to the paper involved: literature review, problem formulation, design, model simulation, model testing, analysis of simulation results, drafting of concluding remarks, manuscript writing and editing.

Signed: ___________________________ Date: 03/01/2020

PhD Candidate: Seyed Eman Mousavinejad (Principal author)

Countersigned: ___________________________ Date: 03/01/2020

Principal Supervisor: Prof. Emer. Ljubo Vlacic

Countersigned: ___________________________ Date: 03/01/2020

Principal Supervisor: A. Prof. Fuwen Yang

Countersigned: ___________________________ Date: 03/01/2020

Associate Supervisor: Prof. Qing Long Han (Corresponding author)

Countersigned: ___________________________ Date: 03/01/2020

Co-author: Dr. Xiaohua Ge
4.1 Abstract

This article is concerned with the resilient tracking control of a networked control system under cyber attacks. The attacker is an active adversary whose aim is to severely degrade the tracking performance of the system by launching deception attacks on the Sensor-to-Controller (SC) communication channels and denial-of-service attacks on the Controller-to-Plant (CP) channels, respectively. First, a concept of resilient set-membership tracking control is presented, through which the system’s true state is guaranteed to reside in a bounding ellipsoidal set of the reference state regardless of the existence of attacks and UBB noises. Second, in the case that full information of the system’s state is not implicitly trusted in the presence of attacks, a resilient set-membership estimation strategy is provided to secure the state estimates against the deception attacks. Furthermore, based on a recursive computation of a reference state ellipsoid and confidence state estimation ellipsoids, a convex optimization algorithm in terms of recursive linear matrix inequalities is proposed to obtain the gain parameters for both the desired resilient state estimator and the tracking controller. Finally, the effectiveness of the proposed method is illustrated through an Internet-based three-tank system.

4.2 Introduction

Advancement and use of Internet, embedded systems, wireless communication technologies and novel control strategies over the past few decades have significantly elevated the development and application of NCSs in a wide range of practical fields, such as transportation systems, electrical power systems and smart grids, remote surgery, industrial and manufacturing systems, and so on [1]. Due to heterogeneous IT components and open network connections among controllers, sensors, actuators, and other networked components, the confidentiality, integrity, and availability of exchanged data in an NCS may suffer from vulnerability to malicious cyber attacks. Undoubtedly, this kind of threat is mainly launched by an adversary in either the physical world or the cyberspace with the aim of substantial economic benefits or disrupting human life. Therefore, it is imperative to properly address security issues of NCSs so as to ensure their reliable and safe performance [33], [50].

Note that NCSs have several distinct advantages, including flexible architectures and less installation and maintenance costs. The remote tracking control through a
shared communication network has received an ever-increasing interest from researchers due to its extensive practical applications in industry and military, such as motor control [104], [105], robotic trajectory tracking control [106]–[108], and flight control [109]–[111]. Generally, the ultimate objective of tracking control is to provide the system with a proper control command such that the state (or output) of the system tracks a state (or output) trajectory predefined by a given reference model. In many real-world applications due to physical constraints, technological limitations, or even human safety aspects, it is an irrational approach to locate the physical plant, actuators, sensors, and controllers at the same place. Hence, in modern industrial systems, the system components are connected remotely via some wireless network medium.

Motivated by security concerns of NCSs, the property of the system’s resiliency is of utmost significance. Generally speaking, resiliency refers to the ability of restoration of a system after being corrupted by unexpected adversarial attacks. Hence, a resilient tracking controller must be able, in the presence of a cyber attack, to recover its computational performance as well as to resume the system’s operation which was in place just before the cyber attack. Up to date, cyber attacks triggered through unreliable communication networks can be arguably categorized into two main types: 1) DoS attacks; and 2) deception attacks. DoS attacks are aimed at deteriorating the availability of the system information, usually either control commands or sensor data, by jamming the communication channels and therefore, preventing information exchange between components of NCSs. Technologically, DoS attacks can be implemented by disrupting the radio frequencies on wireless communication channels, which results in channel congestion. Since an adversary requires little prior knowledge about the system and does not need any special devices to launch a DoS attack, this type of attack can seriously threaten the NCS’s performance. On the other hand, deception attacks represent a kind of attempt to violate the confidentiality and integrity of system information. Through the deception attacks, an adversary tries to manipulate the system into following his or her desired behavior via accessing disclosure resources and injecting deception information into either sensor data or control commands. For example, radars in a target tracking system can be deceived by deception attacks if the adversary successfully injects the deception signals into frequency-shifted copies of the radar’s signals and hence, degrades the tracking performance [112].

During the past two decades, the problem of guaranteeing resilient NCSs has attracted considerable interest from different perspectives. (See [24]–[32] and the references therein.) However, in the context of resilient tracking control, there exist only
Resilient control and estimation

a few studies in the literature. For example, in [33], the design of a specific deception attack, FDI attack, violating both feedback and forward channels, is considered in the sense that the attack can heavily destroy the tracking performance without being detected. The tracking problem of heterogeneous dynamical networks in the presence of malicious cyber attacks is investigated in [34]. The attacker therein launches asynchronous attacks on both controllers’ and observers’ communication channels. The proposed algorithm provides sufficient conditions on the length of attacking intervals in order to achieve the tracking objective. The distributed tracking control problem of multiagent systems under a strategic DoS attack, whose dynamics are modeled by a random Markov process, is addressed in [35]. Based on a hybrid stochastic secure control framework, a distributed resilient control law is developed to achieve exponential consensus tracking in a mean square sense. In [36], the tracking control of mobile robots in the presence of DoS attacks is studied. To ensure both the tracking convergence and efficient usage of communication recourses, a general hybrid model consisting of an event-triggering strategy and DoS attacks is established. Nevertheless, it should be pointed out that the existing literature concentrates primarily on one specific type of attack, and relatively few results deal with the tracking control problem in the presence of different attacks on communication channels. In some practical situations, it is quite common for a crafty adversary to launch assorted attacks of different models and strategies so that he or she could heavily disrupt the tracking performance. To the best of our knowledge, the resilient tracking control problem in NCSs subject to the presence of different cyber attacks that are modeled in a unified framework has not been adequately addressed and the problem remains challenging when the NCS is operated and controlled via some digital and unprotected communication networks, which motivates the current study.

In this article, we aim to achieve resilient tracking control for an NCS operating over a digital and unsafe communication network in the presence of simultaneous data quantization and two malicious attacks, including DoS attacks and deception attacks on the communication channels. Furthermore, we consider that the NCS also suffers from both UBB process noise and measurement noise. More specifically, two logarithmic quantizers will be considered over each of controller-to-plant CP and SC channels so as to adequately account for the quantization effect on the resilient tracking and estimation performance. The DoS attack will be launched by an adversary on the CP communication channel so as to prevent the physical plant from receiving the tracking control command. On the other hand, the deception attack will be
4.3 Problem formulation

The purpose of this article is to design a resilient tracking control algorithm in the case that there exist DoS attacks on quantized control signals and deception attacks considered on the SC communication channel so that the attacker can severely damage the tracking performance by modifying the measurement outputs meaningfully. By fully exploring two different attack models and strategies, the adversary aims to disrupt the tracking control performance of the system in a subtle way. The main contributions are summarized as follows.

- A resilient set-membership tracking control method is developed such that the true state of the system is guaranteed to be included in a bounding ellipsoidal set of the reference state in the simultaneous presence of UBB process and measurement noises as well as DoS attacks and deception attacks.

- A resilient set-membership estimation technique is delicately proposed on account of the scenario that the full states of the system may not be trustworthy and/or available at the controller side due to the existence of an adversary. The estimation procedure includes two steps: a) a predicted state estimate step and b) an updated state estimate step. The two-step estimation determines a group of confidence sets which contain the true state of the system while the attack may be present.

- A recursive convex optimization algorithm, which is based on a recursive computation of the reference state ellipsoid and confidence state estimation ellipsoids, is presented to determine the desired gain parameters for both the resilient tracking controller and estimator in an NCS whose CP and SC communication channels are under the presence of DoS and deception attacks, respectively.

The reminder of this article is organized as follows. In Section 4.3, models of DoS and deception attacks are presented and the system dynamics subject to these attacks and UBB noises are formulated. The observer-based tracking control protocol and the two-step state estimator are also constructed in this section. The main results are presented in Section 4.4 where resilient tracking control and estimation design criteria are derived. Section 4.4 also includes a recursive convex optimization algorithm through which the design parameters of interest are calculated. In Section 4.5, an Internet-based Three-Tank System (ITTS) is employed to illustrate the effectiveness of the proposed method. Finally, Section 4.6 draws a conclusion.

4.3 Problem formulation

The purpose of this article is to design a resilient tracking control algorithm in the case that there exist DoS attacks on quantized control signals and deception attacks
Figure 4.1 Schematic of resilient observer-based tracking control over shared communication networks subject to cyber attacks.

on quantized measurement outputs. The design criteria are presented in Section 4.4.1. To provide a solution for NCSs from a practical engineering perspective, it is anticipated that sensors, controllers, and actuators are at the distance from each other, i.e., mutually interconnected through a digital and unprotected communication network as shown in Figure 4.1. In such a communication setting, it is expected that the transmitted signals are subject to quantization effects and are also exposed to cyber attacks. Details of data quantization and cyber attacks models are discussed in Sections 4.3.2 and 4.3.3, respectively.

4.3.1 The system model

The physical plant is described by a discrete time-varying linear system of the following form:

\[
\begin{align*}
\dot{x}_{k+1} &= A_k x_k + B_k u_k + B_{w,k} w_k \\
y_k &= C_k x_k + D_k \nu_k
\end{align*}
\] (4.1)

where \( x_k \in \mathbb{R}^n \), \( u_k \in \mathbb{R}^n \), and \( y_k \in \mathbb{R}^n \) are, respectively, the state vector, the control input, and the measurement output; \( A_k, B_k, B_{w,k}, C_k, \) and \( D_k \) are time-varying matrices with appropriate dimensions; \( x(0) = x_0 \) is a given initial state of the system; \( w_k \) and \( \nu_k \) are the process noise and measurement noise, respectively, which are UBB. Before proceeding further, the following definitions and assumption are introduced.

**Definition 4.1:** [113] A nonlinearity \( \varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^n \) is said to satisfy a sector condition if

\[
(\varphi(\varepsilon) - \Omega_1 \varepsilon)^T (\varphi(\varepsilon) - \Omega_2 \varepsilon) \leq 0, \forall \varepsilon \in \mathbb{R}^n
\] (4.2)
for some real matrices $\Omega_1, \Omega_2 \in \mathbb{R}^{n \times n}$, where $\Omega = \Omega_2 - \Omega_1$ is a positive-definite symmetric matrix. In this case, the vector-valued nonlinearity $\varphi(\cdot)$ belongs to the sector $[\Omega_1, \Omega_2]$.

**Definition 4.2**: [64] A bounded ellipsoid is denoted as $(c, P, n) \triangleq \{\zeta \in \mathbb{R}^n : (\zeta - c)^TP^{-1}(\zeta - c) \leq 1\}$, where $c \in \mathbb{R}^n$ is the center and $P > 0$ is the shape matrix of the ellipsoid. Let $E \in \mathbb{R}^{n \times m}$ with rank$(E) = m \leq n$ be a lower triangular matrix whose every diagonal element is positive. By a Cholesky factorization, we have $P = EE^T > 0$. Hence, an alternative representation of the ellipsoid is $\mathcal{E} \triangleq \{\zeta : \zeta = c + Ez, \|z\| \leq 1\}$. The size of the ellipsoid is a function of the squared shape matrix $P$ which can be measured by means of trace$(P)$, that is, the sum of squared semiaxes lengths.

**Assumption 4.1**: The system noises $w_k$ and $\nu_k$ are confined to the following specified ellipsoidal sets:

$$\mathcal{W}_k \triangleq \{w_k : w_k^TW_k^{-1}w_k \leq 1\} \quad (4.3)$$

$$\mathcal{V}_k \triangleq \{\nu_k : \nu_k^TV_k^{-1}\nu_k \leq 1\} \quad (4.4)$$

where $W_k = W_k^T > 0$ and $V_k = V_k^T > 0$ are the time-varying matrices of compatible dimensions.

**Remark 4.1**: The UBB assumption of process and measurement noises is fully justified in the literature [58], [63] and therefore, it is omitted here.

### 4.3.2 Data quantization

In a typical NCS, system components, such as sensors, controllers, and actuators, are remotely connected, and thus data transmissions among these components need to be carried out through some digital communication network medium. In this sense, the quantization effects and cyber attack impacts on both control signal and measurement output data are inevitable. It is our aim to investigate the problem of resilient observer-based tracking control for the system (4.1) via a digital and unprotected communication network. In particular, we consider that the signal is quantized by a logarithmic quantizer prior to enter the communication network, and then it may be compromised by a cyber attacker while being transmitted via communication network channels.
For the logarithmic quantizer, let an \( m \)-dimensional vector signal as \( \epsilon_k = [\epsilon_{1,k}, \epsilon_{2,k}, \ldots, \epsilon_{m,k}]^T \in \mathbb{R}^m \). Then quantizer \( q(\cdot) \) is expressed as

\[
\epsilon_k^q = q(\epsilon_k) = [q_1(\epsilon_{1,k}), q_2(\epsilon_{2,k}), \ldots, q_m(\epsilon_{m,k})]^T
\]

where \( \epsilon_k^q \in \mathbb{R}^m \) is the quantized signal. For each \( q_\ell(\cdot) \) \( 1 \leq \ell \leq m \), suppose that the set of quantization levels is described by

\[
\mathcal{U}_\ell := \{ \pm \mu_{\ell,i} : \mu_{\ell,i} = \rho_\ell \mu_{\ell,0}, \quad i = 0, \pm 1, \pm 2, \ldots \} \cup \{ 0 \}
\]

\[
0 < \rho_\ell < 1, \quad \mu_{\ell,0} > 0 \tag{4.5}
\]

where \( \rho_\ell \) is the quantization density and \( \mu_{\ell,0} \) is a scaling constant. Then from [114], the logarithmic quantizer \( q_\ell(\cdot) \) is defined as

\[
q_\ell(\epsilon_{\ell,k}) = \begin{cases} 
\mu_{\ell,i}, & 1 + \frac{\rho_\ell}{2} \mu_{\ell,i} \leq \epsilon_{\ell,k} \leq 1 + \frac{\rho_\ell}{2} \mu_{\ell,i} \\
0, & \epsilon_{\ell,k} = 0 \\
-q_\ell(-\epsilon_{\ell,k}), & \epsilon_{\ell,k} < 0.
\end{cases} \tag{4.6}
\]

A small \( \rho_\ell \) implies coarse quantization and a large \( \rho_\ell \) means dense quantization. Denoting \( H_1 = \text{diag}_m \{2\rho_\ell/(1 + \rho_\ell)\} \) and \( H_2 = \text{diag}_m \{2/(1 + \rho_\ell)\} \), it can be easily checked from the above definition (4.6) that the quantized signal \( \epsilon_k^q \) satisfies the following inequality:

\[
(\epsilon_k^q - H_1 \epsilon_k)(\epsilon_k^q - H_2 \epsilon_k) \leq 0. \tag{4.7}
\]

Since \( 0 < \rho_\ell < 1 \), one can check that \( 0 \leq H_1 < I \leq H_2 \). Then \( \epsilon_k^q \) can be decomposed as

\[
\epsilon_k^q = H_1 \epsilon_k + \sigma(\epsilon_k) \tag{4.8}
\]

where \( \sigma(\epsilon_k) \) is a nonlinear vector-valued function which, from Definition 4.1, satisfies

\[
\sigma(\epsilon_k)^T(\sigma(\epsilon_k) - H \epsilon_k) \leq 0 \tag{4.9}
\]

with \( H = H_2 - H_1 \).

**Remark 4.2:** The use of network in tracking control not only raises concern over security issues but also renders the assumption of data transmission with infinite precision no longer valid due to signal quantization or limited communication capacity. It is therefore of great significance to consider the quantization effects during the
network transmission. A rich body of literature in quantization of sensor measurement and/or control input signals is available, among which two quantization approaches are widely employed: 1) linear (or uniform) quantization and 2) logarithmic quantization. As discussed in [114]–[116], the logarithmic quantization, whose step-size increases exponentially with respect to the increase in the input, outperforms the linear quantization in control problems. This is due to the fact that, in the logarithmic quantization, the relative quantization error is guaranteed to be roughly constant, which serves as an important feature in many applications.

Remark 4.3: Since the quantization effects are considered on both the control signal \( u_k \) and the measurement output \( y_k \), the corresponding matrices, scalars, and nonlinear vector-valued function, that is, \( H_1, H_2, H, \rho, \mu, \rho_0, \) and \( \sigma(\cdot) \) will be denoted with superscripts \( u \) and \( y \), respectively, in the reminder of this article.

4.3.3 Cyber attacks models

Consider the scenario that there exists an attacker who deteriorates the remote estimation performance by injecting certain deception signals into the true value of the measurement output \( y_k \) and also degrades the remote tracking performance through jamming the wireless communication channels between the controller and the physical plant, as shown in Figure 4.1.

Before introducing the cyber attacks models, there are some assumptions on the system knowledge being available to the attacker for launching a successful attack. Regarding the DoS attack on control signals, the attacker is an active adversary in the sense that the control signal will be dropped once the wireless channels are successfully jammed. After the occurrence of the DoS attack, one of the following cases occurs: 1) the control signal will be successfully received by the physical plant if the attacker completely fails to jam the corresponding communication channels. This case might occur if the attacker has to desist from jamming certain channels due to a limited energy budget; 2) the control signal will be partially received if the communication channels are not heavily jammed; and 3) the control signal will be completely lost if the communication channels are under a severe DoS attack. For the deception attack on measurement outputs, in the first phase, the attacker must have this ability to know the accurate value of the measurement output in real time, and in the second phase, the attacker must be able to modify the true value of the measurement output to his or her arbitrary one.
Motivated by the above observation, the actual control input, which is under a DoS attack and received by the physical plant, is proposed as

\[
\tilde{u}_k = \Gamma_k u_k
\]  

(4.10)

where \( \Gamma_k \in \mathbb{R}^{n_u \times n_u} \) represents an attack model parameter matrix prescribed by the attacker and it has the following diagonal structure:

\[
\begin{cases} \\
\Gamma_k = \text{diag}\{\gamma_{1,k}, \gamma_{2,k}, \ldots, \gamma_{n_u,k}\} \\
\tilde{\gamma}_{j,k} \leq \gamma_{j,k} \leq \overline{\gamma}_{j,k}, \quad j = 1, 2, \ldots, n_u
\end{cases}
\]  

(4.11)

where \( \gamma_{j,k}, \overline{\gamma}_{j,k} \in [0, 1] \) are known scalars representing the lower and upper bounds of the unknown parameter \( \gamma_{j,k} \).

**Remark 4.4:** The unknown parameter \( \gamma_{j,k} \) represents the transmission status of the \( j \)-th control signal’s component so as to reflect the jamming status of the communication channel under attack. If \( \gamma_{j,k} = 1 \), there is no DoS attack on the \( j \)-th control signal’s component at time \( k \) and therefore, this component can be successfully delivered to the physical plant. When \( 0 < \gamma_{j,k} < 1 \), it represents the case of partial transmission of the \( j \)-th component. The case \( \gamma_{j,k} = 0 \) characterizes the worst-case scenario on transmission of the \( j \)-th control signal’s component, which means the \( j \)-th component is completely lost while being transmitted at time \( k \).

To facilitate subsequent analysis, we define the following matrices:

\[
\hat{\Gamma}_k = \text{diag}\{\overline{\gamma}_{1,k} + \overline{\gamma}_{1,k}, \overline{\gamma}_{2,k} + \overline{\gamma}_{2,k}, \ldots, \overline{\gamma}_{n_u,k} + \overline{\gamma}_{n_u,k}\}
\]

\[
\check{\Gamma}_k = \text{diag}\{\overline{\gamma}_{1,k} - \overline{\gamma}_{1,k}, \overline{\gamma}_{2,k} - \overline{\gamma}_{2,k}, \ldots, \overline{\gamma}_{n_u,k} - \overline{\gamma}_{n_u,k}\}
\]

(4.12)

Then, one has

\[
\Gamma_k = \hat{\Gamma}_k + \check{\Gamma}_k
\]  

(4.13)

where \( \hat{\Gamma}_k = \text{diag}\{\hat{\gamma}_{1,k}, \hat{\gamma}_{2,k}, \ldots, \hat{\gamma}_{n_u,k}\} \) and \( |\hat{\gamma}_{j,k}| \leq (\overline{\gamma}_{j,k} - \underline{\gamma}_{j,k})/2 \). Thus,

\[
||\hat{\Gamma}_k|| \leq \hat{\Gamma}_k
\]  

(4.14)
Furthermore, the signal generated by the attacker for the deception attack is proposed as
\[ y_k^a = -y_k + \delta_k \]  
where \( \delta_k \) is a UBB signal belonging to the following ellipsoid:
\[ S_k \triangleq \{ \delta_k : \delta_k^T S_k^{-1} \delta_k \leq 1 \} \]
with \( S_k = S_k^T > 0 \).

**Remark 4.5:** Note that the signal \( \delta_k \) used by the attacker to launch the deception attack has similar form with the system noises \( w_k \) and \( \nu_k \). As discussed in [43], considering the signal \( \delta_k \) to be unknown but confined to a specified ellipsoidal set not only makes the deception attack difficult to be distinguished by the detectors but also facilitates the attacker to successfully bypass the most widely employed false data detector, namely, the \( \chi^2 \) detector since this detector is only effective when system noises comply with Gaussian distribution.

The actual measurement output, which is under a deception attack and sent to the estimator, can be represented by
\[ \tilde{y}_k = y_k + \Lambda_k y_k^a \]
where the matrix \( \Lambda_k \) represents the physical constraints imposed on the attack signal due to some hardware constraints including device saturations, bandwidth limitations, and channel fading. This matrix is assumed to have the following diagonal structure:
\[ \Lambda_k = \text{diag}\{\lambda_{1,k}, \lambda_{2,k}, \ldots, \lambda_{n_y,k}\} \]
\[ \Lambda_{s,k} \leq \lambda_{s,k} \leq \overline{\lambda}_{s,k}, \quad s = 1, 2, \ldots, n_y \]
where \( 0 \leq \lambda_{s,k} < 1, \overline{\lambda}_{s,k} \geq 1 \) are known scalars describing the lower and upper bounds on \( \lambda_{s,k} \).

**Remark 4.6:** The parameter \( \lambda_{s,k} \) describes how the physical constraints affect the behavior of the deception attack. If \( \lambda_{s,k} = 1 \), there exists a complete deception attack on the \( s \)-th component of the measurement output as the attacker plans. When \( 0 \leq \lambda_{s,k} < 1 \) or \( \lambda_{s,k} > 1 \), the deception signal \( y_k^a \) might be degraded or amplified, respectively, from the attacker’s point of view. As a result, taking physical constraints
into consideration offers a comprehensive yet realistic deception attack model which reflects the impacts of hardware constraints that the attacker would be confronted with.

By denoting \( \Lambda_k = \text{diag}\{\lambda_{1,k}, \lambda_{2,k}, \ldots, \lambda_{n_y,k}\} \) and \( \overline{\Lambda}_k = \text{diag}\{\overline{\lambda}_{1,k}, \overline{\lambda}_{2,k}, \ldots, \overline{\lambda}_{n_y,k}\} \), the lower and upper bounds on \( \lambda_{p,k} \), (4.18), can be written in a compact form

\[
\Delta_k \leq \Lambda_k \leq \overline{\Lambda}_k. \tag{4.19}
\]

To simplify the subsequent derivation, the deception attack signal \( \Lambda_k y^a_k \) can be split into two terms as

\[
\Lambda_k y^a_k = \Delta_k y^a_k + \psi(y^a_k) \tag{4.20}
\]

where \( \psi(y^a_k) \) is a nonlinear vector-valued function which from Definition 4.1 satisfies a sector condition with \( \Omega_1 = 0 \) and \( \Omega_2 = \tilde{\Lambda}_k \), in which \( \tilde{\Lambda}_k = \overline{\Lambda}_k - \Lambda_k > 0 \), that is, \( \psi(y^a_k) \) satisfies the following inequality:

\[
\psi^T(y^a_k)(\psi(y^a_k) - \tilde{\Lambda}_k y^a_k) \leq 1. \tag{4.21}
\]

### 4.3.4 The resilient observer-based tracking control protocol

On account of DoS and deception attacks, and quantization effects discussed above, the original physical plant (4.1) can be reformulated as

\[
\begin{align*}
\dot{x}_{k+1} &= A_k x_k + B_k \tilde{u}^q_k + B_{w,k} w_k, \\
y_k &= C_k x_k + D_k \nu_k, \\
\tilde{u}^q_k &= \Gamma_k u^q_k, \\
\tilde{y}^q_k &= y^q_k + \Lambda_k y^a_k, \\
y^q_k &= H^q_1 y_k + \sigma^q(y_k), \\
y^a_k &= -y^q_k + \delta_k.
\end{align*} \tag{4.22}
\]

In engineering practice, sensing and computing capabilities of a sensor are often restricted to a small voltage source, and thus its performance can be degraded by some potential attacks and external disturbances. Thus, the assumption of full-state information availability and/or integrity may not be realistic. In what follows, a
two-step estimator is developed to monitor and observe the system’s true state as

\[
\begin{align*}
\hat{x}_{k+1|k} &= \hat{A}_k \hat{x}_k, \\
\hat{x}_{k+1} &= \hat{x}_{k+1|k} + \hat{B}_{k+1}(\tilde{y}_{k+1} - C_{k+1} \hat{x}_{k+1|k})
\end{align*}
\] (4.23)

where \(\hat{x}_k\) is the estimation of the state \(x_k\); \(\hat{A}_k\) and \(\hat{B}_{k+1}\) are time-varying parameter matrices of the estimator to be designed, and the initial condition of the estimator is \(\hat{x}_0\).

Hence, the following observer-based tracking control protocol, which utilizes the estimated state instead of the full-state information measured by sensors, is proposed

\[
u_k = M_k (\hat{x}_k - x^r_k)
\] (4.24)

where \(M_k\) is the controller gain to be designed; and \(x^r_k\) is the state of the specified reference model whose dynamics are given by

\[
x^r_{k+1} = A^r_k x^r_k, \quad x^r(0) = x^r_0.
\] (4.25)

### 4.3.5 The resilient set-membership tracking control problem

We next develop a suitable set-membership tracking control framework through which the true state of the system will be always enclosed in some confidence region of the reference state and also, simultaneously, some confidence region of the system’s state estimation despite the existence of UBB noises and cyber attacks. To proceed with, the definition of the two-step set-membership state estimation is firstly presented.

**Definition 4.3:** System (4.22) is said to achieve set-membership state estimation at the predicted (a priori) state estimate step under the tracking protocol (4.24) if there exist gain sequences \(\hat{A}_k\) and \(M_k\) such that the system’s state \(x_{k+1}\) resides in a state prediction ellipsoid \(\mathcal{X}_{k+1|k}\), which always contains the true state of the system, where

\[
\mathcal{X}_{k+1|k} \triangleq \{x_{k+1} : (x_{k+1} - \hat{x}_{k+1|k})^T P^{-1}_{k+1|k} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1\}
\] (4.26)

for all \(k \in \mathbb{N}\); UBB process noise \(w_k \in \mathcal{W}_k\) and UBB measurement noise \(\nu_k \in \mathcal{V}_k\) with \(P_{k+1|k} = P^T_{k+1|k} > 0\) representing a time-varying matrix.
**Definition 4.4:** System (4.22) is said to achieve set-membership state estimation at the updated (a posteriori) state estimate step under the tracking protocol (4.24) if there exists a gain sequence $\hat{B}_{k+1}$ such that the system’s state $x_{k+1}$ resides in a state estimation ellipsoid $\mathcal{X}_{k+1}$, which always contains the true state of the system, where

$$
\mathcal{X}_{k+1} \triangleq \{ x_{k+1} : (x_{k+1} - \hat{x}_{k+1})^T P_{k+1}^{-1} (x_{k+1} - \hat{x}_{k+1}) \leq 1 \} \tag{4.27}
$$

for all $k \in \mathbb{N}$; UBB process noise $w_k \in \mathcal{W}_k$ and UBB measurement noise $\nu_k \in \mathcal{V}_k$ with $P_{k+1} = P_{k+1}^T > 0$ representing a time-varying matrix.

**Remark 4.7:** Resilient estimation has been deemed as an effective means of guaranteeing a reliable estimation of the system’s state in the simultaneous presence of noises and attacks, and thus has been intensively studied in the field of NCSs. Arguably, prior work on resilient estimation in the presence of noises and attacks can be divided into two categories depending on the noise model [117] 1) stochastic noises (e.g., Gaussian noise) [118]–[121] and 2) bounded non-stochastic noises (e.g., $l_2$-norm energy-bounded noise) [122]–[125]. Fruitful results have been available on resilient estimation, it should be pointed out that modeling the noise in a stochastic framework usually requires accurate statistical properties of the noise, such as known mean and covariance, which is conservative for some practical applications when the noise is unknown. In an energy-bounded noise model, e.g., in the $H_\infty$ sense, it is assumed that the noise has norm-bounded energy so as to satisfy the aim of finding the worst-case solution to the estimation problem. Furthermore, the objective of a conventional estimation problem is to compute a single-vector estimation $\hat{x}_k$ with regard to the system’s state such that $\lim_{k \to \infty} \| x_k - \hat{x}_k \| = 0$ in the asymptotic convergence case. However, such an estimation approach only provides a point-wise estimation of the system’s state [63]. Therefore, there is no guarantee that the distribution of the system’s state can be achieved within a confidence region where all true states of the system may reside in, specially when there exist unpredictable environmental changes due to the influences of cyber attacks and noises. As a result, an alternative method called set-membership estimation is developed in [53]. The core idea of this method is to calculate a bounding ellipsoidal set in state-space, which always encloses the true state of the system by assuming UBB noise signals [54], [55]. Indeed, an assumption of UBB noise eliminates the requirement of prior knowledge of the accurate statistical characteristics of noise because only the knowledge of a bound on the realization is needed. During the past two decades, the set-membership method has been intensively studied for different problem formulations.
of various system models [56]–[58], [61], [63]. However, to the best of our knowledge, there are rare results available that consider the conjunct problem of resilient tracking control and set-membership estimation over a shared communication network subject to two different cyber attacks independently corrupting the sensor data and control signals, and UBB process and measurement noises.

We then present the definition of the proposed resilient set-membership tracking control.

**Definition 4.5:** System (4.22) subject to DoS and deception attacks on control signals and measurement outputs is said to achieve resilient set-membership tracking control with respect to the reference model (4.25) under the tracking protocol (4.24) if there exist gain sequences \( \hat{A}_k \) and \( M_k \) such that the system’s state \( x_{k+1} \) resides in a reference state ellipsoid \( T_{k+1} \), which always contains the true state of the system, where

\[
T_{k+1} \triangleq \left\{ x_{k+1} : (x_{k+1} - x_{k+1}^r)^T R_{k+1}^{-1} (x_{k+1} - x_{k+1}^r) \right\} \leq 1
\]  (4.28)

for all \( k \in \mathbb{N} \); UBB process noise \( w_k \in \mathcal{W}_k \) and UBB measurement noise \( \nu_k \in \mathcal{V}_k \) with \( R_{k+1} = R_{k+1}^T > 0 \) representing a time-varying matrix.

The initial state \( x_0 \) satisfies the following assumption.

**Assumption 4.2:** The initial state of the system belongs to given ellipsoids

\[
T_0 \triangleq \left\{ x_0 : (x_0 - x_0^r)^T R_0^{-1} (x_0 - x_0^r) \leq 1 \right\} \quad (4.29)
\]

\[
\mathcal{X}_0 \triangleq \left\{ x_0 : (x_0 - \hat{x}_0)^T P_0^{-1} (x_0 - \hat{x}_0) \leq 1 \right\}. \quad (4.30)
\]

From Definitions 4.3–4.5, resilient set-membership tracking control and estimation problem to be addressed is now stated as: For prescribed time-varying matrices \( \Gamma_k \), \( \Lambda_k \), scalars \( \rho_u^r \), \( \rho_y^r \), \( \mu_u^0 \), \( \mu_y^0 \) satisfying (4.5), UBB deception attack \( \delta_k \in \mathcal{S}_k \), UBB noises \( w_k \in \mathcal{W}_k \) and \( \nu_k \in \mathcal{V}_k \), the objective is to design a desired tracking control protocol in the form of (4.24) and to find appropriate matrix sequences \( R_{k+1} \), \( P_{k+1|k} \), and \( \hat{P}_{k+1} \) such that the system’s one-step ahead state \( x_{k+1} \) always resides in the ellipsoids \( \mathcal{X}_{k+1|k} \), \( \mathcal{X}_{k+1} \), and \( T_{k+1} \), simultaneously.
4.4 Main results

4.4.1 Design criteria

In the following text, we first present two theorems, which establish sufficient conditions on the existence of a two-step set-membership estimator (4.23) and a desired tracking control protocol (4.24) such that (4.26), (4.27), and (4.28) hold.

**Theorem 4.1:** For the system (4.22) subject to UBB noises \( w_k \in \mathcal{W}_k \) and \( v_k \in \mathcal{V}_k \), and DoS attacks and deception attacks on control signals and measurement outputs, respectively, suppose that at time \( k \) the system’s state \( x_k \) belongs to its state estimation ellipsoid \( (x_k - \hat{x}_k)^T P_k^{-1} (x_k - \hat{x}_k) \leq 1 \) and its reference state ellipsoid \( (x_k - x_r^k)^T R_k^{-1} (x_k - x_r^k) \leq 1 \). Then, for any prescribed time-varying matrices \( \Gamma_k \), \( \Lambda_k \), scalars \( \rho_u, \rho_y, \mu_u, \mu_y \) satisfying (4.5), and UBB deception attack \( \delta_k \in S_k \), the system’s one-step ahead state \( x_{k+1} \) always resides in the state prediction ellipsoid \( X_{k+1|k} \) and the reference state ellipsoid \( T_{k+1|k} \), if there exist matrix sequences \( P_{k+1|k} > 0, R_{k+1} > 0, \hat{A}_k, M_k \), and scalar sequences \( \tau_{m,k} > 0, m = 1, 2, \ldots, 4 \), such that

\[
\begin{bmatrix}
- I & 0 & \Phi_k \\
* & - P_{k+1|k} & \Pi_k^p \\
* & * & \Theta_k^p
\end{bmatrix} \leq 0, \\
\begin{bmatrix}
- I & 0 & \Phi_k \\
* & - R_{k+1} & \Pi_k^c \\
* & * & \Theta_k^c
\end{bmatrix} \leq 0
\] (4.31)

for all \( k \in \mathbb{N} \), where

\[
\Pi_k^p = \begin{bmatrix} \Pi_{11}^p & A_k E_k & B_{w,k} & B_k & B_k \hat{\Gamma}_k \end{bmatrix}, \\
\Pi_{11}^c = (A_k - \hat{A}_k + B_k \hat{\Gamma}_k H_1^u M_k) \hat{x}_k - B_k \hat{\Gamma}_k H_1^u M_k x_r^k, \\
\Pi_k^c = \begin{bmatrix} \Pi_{11}^c & A_k L_k & B_{w,k} & B_k & B_k \hat{\Gamma}_k \end{bmatrix}, \\
\Pi_{11}^c = (A_k - \hat{A}_k + B_k \hat{\Gamma}_k H_1^u M_k) x_r^k + B_k \hat{\Gamma}_k H_1^u M_k \hat{x}_k, \\
\Phi_k = \begin{bmatrix} \hat{\Gamma}_k H_1^u M_k (\hat{x}_k - x_r^k) & 0 & 0 & 0 \end{bmatrix},
\]

and \( \Theta_k^p = [\Theta_{o,\beta}^c(k)]_{5 \times 5} \) and \( \Theta_k^c = [\Theta_{o,\beta}^c(k)]_{5 \times 5} \) are time-varying sparse symmetric block matrices with their nonzero entries given by

\[
\Theta_{1,1}^p(k) = -1 + \tau_{3,k} + \tau_{4,k}, \quad \Theta_{1,1}^c(k) = -1 + \tau_{1,k} + \tau_{2,k}, \\
\Theta_{5,5}^p(k) = \Theta_{5,5}^c(k), \quad \Theta_{5,5}^c(k) = (1/2) (\hat{x}_k - x_r^k)^T M_k^T H^u,
\]
\[\Theta^p_{2,2}(k) = -\tau_{3,k} I, \quad \Theta^p_{5,2}(k) = -\tau_{1,k} I, \]
\[\Theta^p_{3,3}(k) = -\tau_{4,k} W_k^{-1}, \quad \Theta^p_{3,3}(k) = -\tau_{2,k} W_k^{-1}, \]
\[\Theta^p_{4,4}(k) = \Theta^p_{4,4}(k) = -I, \quad \Theta^p_{5,5}(k) = \Theta^p_{5,5}(k) = 2\Gamma_k^2 - I.\]

**Proof:** See the Appendix B.1.

**Theorem 4.2:** For the system (4.22) subject to UBB noises \( w_k \in \mathcal{W}_k \) and \( \nu_k \in \mathcal{V}_k \), and DoS attacks and deception attacks on control signals and measurement outputs, respectively, suppose that the one-step ahead state \( x_{k+1} \) belongs to its state prediction ellipsoid \( (x_{k+1} - \hat{x}_{k+1})^T P_{k+1}^{-1} (x_{k+1} - \hat{x}_{k+1}) \leq 1 \). Then, for any prescribed time-varying matrices \( \Gamma_k, \Lambda_k \), scalars \( \rho_k^p, \rho_k^m, \mu_k^w, \mu_k^\ell \) satisfying (4.5), and UBB deception attack \( \delta_k \in \mathcal{S}_k \), such a state belongs to its state estimation ellipsoid \( \mathcal{E}_{k+1} \), if there exist matrix sequences \( P_{k+1} > 0, \hat{B}_{k+1}, \hat{S}_{k+1} > 0 \), and scalar sequences \( \tau_{m,k} > 0 \), \( m = 5, 6, \ldots, 8 \), such that
\[
\begin{bmatrix}
-P_{k+1} & \Pi_k^\ell \\
* & \Theta_k^\ell
\end{bmatrix} \leq 0
\]
for all \( k \in \mathbb{N} \), where
\[
\Pi_k^\ell = \begin{bmatrix}
\Pi_{11}^\ell & \Pi_{12}^\ell & \Pi_{13}^\ell & \Pi_{14}^\ell & \Pi_{15}^\ell & -\hat{B}_{k+1}
\end{bmatrix},
\]
\[
\Pi_{11}^\ell = -\hat{B}_{k+1}((I - \Delta_{k+1})\Gamma_k - I)C_{k+1} \hat{x}_{k+1},
\]
\[
\Pi_{12}^\ell = (I - \hat{B}_{k+1}(I - \Delta_{k+1})\Gamma_k C_{k+1})E_{k+1},
\]
\[
\Pi_{13}^\ell = -\hat{B}_{k+1}(I - \Delta_{k+1})\Gamma_k D_{k+1},
\]
\[
\Pi_{14}^\ell = -\hat{B}_{k+1} \Delta_{k+1}, \quad \Pi_{15}^\ell = -\hat{B}_{k+1}(I - \Delta_{k+1}),
\]
and \( \Theta_k^\ell = [\Theta_{\alpha,\beta}(k)]_{6 \times 6} \) is a time-varying sparse symmetric block matrix with its nonzero entries given by
\[
\Theta_{1,1}(k) = -1 + \tau_{5,k} + \tau_{6,k} + \tau_{7,k}, \quad \Theta_{1,5}(k) = (1/2)\tau_{8,k} \hat{x}_{k+1}^T C_{k+1}^T H_k^y, \]
\[
\Theta_{1,6}(k) = -(1/2)\tau_{8,k} \hat{x}_{k+1}^T C_{k+1}^T H_k^y \Delta_{k+1}, \]
\[
\Theta_{2,2}(k) = -\tau_{5,k} I, \quad \Theta_{2,5}(k) = (1/2)\tau_{8,k} E_{k+1}^T C_{k+1} H_k^y, \]
\[
\Theta_{2,6}(k) = -(1/2)\tau_{8,k} E_{k+1}^T C_{k+1} H_k^y \Delta_{k+1}, \]
\[
\Theta_{3,3}(k) = -\tau_{6,k} V_{k+1}^{-1}, \quad \Theta_{3,5}(k) = (1/2)\tau_{8,k} D_{k+1}^T H_k^y, \]
\[
\Theta_{3,6}(k) = -(1/2)\tau_{8,k} D_{k+1}^T H_k^y \Delta_{k+1}, \]
\[
\Theta_{4,4}(k) = -\hat{S}_{k+1}, \quad \Theta_{4,6}(k) = (1/2)\tau_{8,k} \Delta_{k+1},
\]
\[ \Theta_{5,5}(k) = \Theta_{6,6}(k) = -\tau_{8,k} I, \quad \Theta_{5,6}(k) = -\Theta_{4,6}(k) \]

where the parameter \( S_{k+1} \) can be calculated by \( S_{k+1} = \tau_{7,k} \tilde{S}_{k+1}^{-1} \).

\[ \text{Proof:} \quad \text{See the Appendix B.2.} \]

**Remark 4.8:** By Theorems 4.1 and 4.2, the proposed resilient set-membership tracking control and two-step set-membership estimation problem can be transferred into the feasibility problem of a set of RLMIs (4.31) and (4.32). Thus, Theorems 4.1 and 4.2 provide criteria for designing tracking control and two-step estimation gain matrix sequences \( M_k, \hat{A}_k, \) and \( \hat{B}_{k+1} \) such that the true state of the system is always included in the ellipsoids \( \mathcal{T}_{k+1}, \mathcal{X}_{k+1|k}, \) and \( \mathcal{X}_{k+1} \) despite the existence of UBB process noise, UBB measurement noise, DoS attacks on quantized control signals, and UBB deception attacks on quantized measurement outputs.

**Remark 4.9:** Notice that the developed resilient tracking control framework assumes only DoS attacks on the CP communication channels and deception attacks on the SC communication channels, respectively. The reasons are twofold. First, we consider that the attacker is a cunning adversary who launches cyber attacks under two different attack models and strategies for comprehensively compromising the sensor measurements and control signals. It has been well acknowledged that different attack strategies are generally stealthy to any defender. Second, in practice, some transmission layers between system components such as the CP communication channel may gain more protection than the others such as the SC communication channel. This is due to the fact that more security requires heavy weight computations and large memory capacity. However, the widely used cryptographic techniques to encrypt data in the SC communication channel are not sufficient due to the constraints of memory and weak processing capability of sensors [126]. Thus, from the attacker’s perspective, it is much easier to launch a DoS attack on the CP channel because a DoS attack requires less system model knowledge and disruption resources to cause destruction.

**4.4.2 Recursive convex optimization algorithm**

In light of Theorems 4.1 and 4.2, the system’s one-step ahead state is always confined in the reference state ellipsoid \( \mathcal{T}_{k+1} \), its state prediction ellipsoid \( \mathcal{X}_{k+1|k} \), and its state estimation ellipsoid \( \mathcal{X}_{k+1} \) if (4.31) and (4.32) hold. Thus, there exist three vectors \( \vartheta_\kappa \), \( \kappa = 1, 2, 3 \), satisfying \( \| \vartheta_\kappa \| \leq 1 \) such that \( x_{k+1} = x^r_k + L_{k+1} \vartheta_1, \ x_{k+1} = \hat{x}_{k+1} + E_{k+1} \vartheta_2, \) and \( x_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3 \), respectively. Furthermore, the centers of ellipsoids
\(X_{k+1/k}, X_{k+1}, \) and \(T_{k+1}\) are determined by (4.23) and (4.25), respectively. Note that even though Theorems 4.1 and 4.2 outline the principles of determining reference state ellipsoid, prediction state and estimation state ellipsoids, they do not provide an optimal reference state ellipsoid or optimal prediction state and estimation state ellipsoids. Hence, by applying a convex optimization approach, the proposed resilient set-membership tracking control and two-step set-membership estimation problems are cast into the following OPs so as to find some optimal ellipsoids:

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(\Psi_{k+1}) \\
\text{subject to} & \quad (4.31),
\end{align*}
\]

where \(\Psi_{k+1} = \text{diag}\{R_{k+1}, P_{k+1}\}\), and

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(P_{k+1}) \\
\text{subject to} & \quad (4.32).
\end{align*}
\]

**Remark 4.10:** Notice that (4.31) and (4.32) are linear to \(R_{k+1}, P_{k+1|k}, P_{k+1}, \tilde{S}_{k+1}, \hat{A}_k, \hat{B}_{k+1}, M_k, \tau_{m,k}, \tau_{m,k}, \tau_{m,k}, \tau_{m,k}, \tau_{m,k}, \tau_{m,k}, \)

Based on the OPs (4.33) and (4.34), we are in a position to present a recursive algorithm, i.e., Algorithm 4.1, which computes the gain matrix sequences for the observer-based tracking protocol (4.24) and also solves out the gain matrix sequences for the two-step set-membership estimator (4.23) as well as determining matrix sequences \(R_{k+1}, P_{k+1|k}, P_{k+1}\) for optimal ellipsoids.
Algorithm 4.1: Recursive convex optimization algorithm

**Step 1.** Given initial conditions $x_0$, $\hat{x}_0$, $x^r_0$, known lower- and upper-bounds on $\gamma_{j,k}$ and $\lambda_{s,k}$ such that $\gamma_{j,k}, \gamma_{j,k} \in [0, 1]$, $\lambda_{s,k} \in [0, 1)$, $\lambda_{s,k} \geq 1$, and given quantization parameters $\rho^u_\ell$, $\rho^y_\ell$, $\mu^u_0$, and $\mu^y_0$ satisfying (4.5). Choose suitable $W_k$, $V_k$, $R_0$ and $P_0$ such that (4.3), (4.4), (4.29) and (4.30) hold. Calculate $E_0$ and $L_0$ according to $P_0 = E_0E_0^T$ and $R_0 = L_0L_0^T$. Set $k = 0$, $x_k = x_0$, $\hat{x}_k = \hat{x}_0$, $x^r_k = x^r_0$, $E_k = E_0$, $L_k = L_0$. Let the simulation run recursively for $T_n$ seconds.

**Step 2.** Solve the OP (4.33) to obtain $R_{k+1}$, $P_{k+1|k}$, $\hat{A}_k$, and $M_k$. Compute $L_{k+1}$ and $E_{k+1|k}$ such that $R_{k+1} = L_{k+1}L_{k+1}^T$ and $P_{k+1|k} = E_{k+1|k}E_{k+1|k}^T$.

**Step 3.** Calculate the state prediction $\hat{x}_{k+1|k}$ by (4.23).

**Step 4.** Solve the OP (4.34) to determine $P_{k+1}$, and $\hat{B}_{k+1}$. Obtain $E_{k+1}$ such that $P_{k+1} = E_{k+1}E_{k+1}^T$.

**Step 5.** Compute the reference state $x^r_{k+1}$ by (4.25) and the state estimation $\hat{x}_{k+1}$ by (4.23).

**Step 6.** If $k = T_n$ go to Step 7; otherwise, set $k = k + 1$, go to Step 2.

**Step 7.** Output matrix sequences $\{R_{k+1}\}$, $\{P_{k+1|k}\}$, $\{P_{k+1}\}$, $\{\hat{A}_k\}$, $\{\hat{B}_{k+1}\}$, $\{M_k\}$, the reference state ellipsoidal set $T_{k+1}$, the state prediction ellipsoidal set $X_{k+1|k}$, and the state estimation ellipsoidal set $\hat{X}_{k+1}$. Exit.

4.5 Illustrative example

In this section, the developed resilient observer-based tracking control method will be applied to an ITTS which has been adopted from [127] as a benchmark system in order to investigate the resiliency of such a system subject to DoS and UBB deception attacks.

We first introduce the ITTS in which the physical plant, that is, a Three-Tank System (TTS), is linked with the remote observer-based tracking controller and its two-step set-membership estimator through a shared communication network, as shown in Figure 4.2. The three tanks in this system have been labeled as $T_1$, $T_3$, and $T_2$, from left to right, that is, the central tank is the tank tagged as 3. There exist two sensors measuring the liquid level, $h$, in $T_1$ and $T_2$ with the assumption of
4.5 Illustrative example

Figure 4.2 Schematic of the ITTS with the developed resilient observer-based tracking control strategy.

$h_1 > h_3 > h_2$. Moreover, there are two pumps as actuators to supply the tanks $T_1$ and $T_2$ with water. Since the sensor measurement output $y_k$ is transferred into the remote estimator/controller and also the control signal $u_k$ is sent to the local TTS through the network, not only must these signals be quantized before transmission but they may also suffer from DoS attacks and UBB deception attacks compromised CP and SC communication channels, respectively.

As discussed in [127], the dynamics of TTS can be linearized around its equilibrium point $x^* = [0.3182, 0.1517, 0.2314]^T$, which represents the steady liquid levels in meter, and then discretized with a sampling period $\tau_s = 1$ s. Thus, the system parameters in (4.22) are given as

$$A_k = \begin{bmatrix} 0.9889 & 0.0001 & 0.0110 \\ 0.0001 & 0.9774 & 0.0119 \\ 0.0110 & 0.0119 & 0.9770 \end{bmatrix}, \quad B_k = \begin{bmatrix} 64.5993 & 0.0015 \\ 0.0015 & 64.2236 \\ 0.3604 & 0.3910 \end{bmatrix}.$$

The other system matrices in (4.22) and (4.25) are chosen as

$$A_r^k = \begin{bmatrix} -0.1093 & 0.0001 & -0.2474 \\ 0.0001 & -0.1080 & -0.2707 \\ 0.0049 & 0.0053 & 0.9738 \end{bmatrix}, \quad B_{w,k} = \begin{bmatrix} 0.3 \\ 0.4 \\ 0.1 \end{bmatrix}, \quad C_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad D_k = \begin{bmatrix} 1 \end{bmatrix}.$$
The three state variables in $x_k$ are the liquid levels $h_1$, $h_2$, and $h_3$, respectively. In the simulation, the initial conditions of (4.22), (4.23), and (4.25) are taken as $x_0 = [0.28, 0.08, 0.18]^T$, $\dot{x}_0 = [0.271, 0.074, 0.17]^T$, and $x_0^* = [0.3, 0.1, 0.2]^T$. Suppose that the UBB noises are $w_k = 0.02\sin(20k)$ and $\nu_k = 0.01\cos(2k)$ and set $W_k = V_k = 0.001$. Then it can be easily checked that $u_k$ and $\nu_k$ belong to the ellipsoidal sets defined in (4.3) and (4.4), respectively. Set $R_0 = \text{diag}\{40, 40, 40\}$ and $P_0 = \{30, 30, 30\}$. Furthermore, the parameters of the logarithmic quantizers are taken as $\mu_{t,0}^a = \mu_{t,0}^b = 3$, $\rho_{t}^a = 0.5$, and $\rho_{t}^b = 0.7$. Then, it yields that $H_1^a = 0.67$, $H_2^a = 1.33$, $H_1^b = 0.8235$, and $H_2^b = 1.1765$. Due to the fluctuation of liquid surface caused by the inflow from the pumps, the liquid level may be biased by up to $0.03$ m around the equilibrium point in each tank [127]. Therefore, it is expected from the developed resilient tracking controller to reach a tracking error less than the bias value.

For a DoS attack on the CP communication channels, it is assumed that the lower and upper bounds of the unknown model parameter matrix $\Gamma_k = \text{diag}\{\gamma_{1,k}, \gamma_{2,k}\}$ in (4.10) are as $\text{diag}\{0, 0.1\} \leq \Gamma_k \leq \text{diag}\{0.2, 0.4\}$. One can see from the DoS attack model parameter matrix that the jamming status on the channel transmitting the first component of the control signal $u_1^k$ is more severe than the one compromising the channel of the second component $u_2^k$ since the control signal $u_1^k$ may be completely lost due to successful attack on its channel. In the simulation, it is considered that the attacker performs the DoS attack from $k = 1$ till $k = 15$.

In the case of a UBB deception attack on the SC communication channels, suppose that the physical constraint matrix $\Lambda_k = \text{diag}\{\lambda_{1,k}, \lambda_{2,k}\}$ is characterized by $\text{diag}\{0.2, 0.6\} \leq \Lambda_k \leq \text{diag}\{1.3, 1.3\}$. From the lower and upper bounds on the physical constraints affecting the UBB deception attack signal, it can be seen that the attack on the tank $T_2$ sensor’s measurement output is less degraded by the imposed physical constraints than the one violating the tank $T_1$ sensor’s measurement output. In the simulation, the UBB deception attack is launched by the attacker from $k = 10$ till $k = 50$. The UBB deception attack signal is assumed to be $\delta_k = [0.07\sin(k), 0.08\cos(k)]^T$ whose entries represent the attacks on the SC channels transmitting the measurement outputs of tanks $T_1$ and $T_2$, respectively.

Figure 4.3 depicts the liquid level in each tank when the tracking controller and estimator are not resilient against the attacks, that is, tracking control and estimation strategies are designed without any prior consideration of the existence of attacks. Since the attacker perpetrates the DoS attack on the CP communication channels so as to prevent the tracking control commands $u_1$ and $u_2$ from transferring to tanks $T_1$
and \( T_2 \), respectively, the liquid levels in the three tanks of ITTS fail to track their reference signals. Furthermore, due to the existence of the deception attack on the SC communication channels, the estimator receives the corrupted measurement outputs. The simulated deception attack affects the amplitude and frequency of the signal measured by each sensor and hence, large fluctuations occur on the state estimation at each tank causing the estimation deviates from the true state significantly. As a result of the attacks on both CP and SC communication channels, tracking error, \( x_k - x^*_k \), increases dramatically with the maximum of 0.3 m in tank \( T_1 \) and 0.1 m in tanks \( T_2 \) and \( T_3 \) as shown in Figure 4.4a. Additionally, as demonstrated in Figure 4.4b, the attack makes the estimation error, \( x_k - \hat{x}_k \) rise by the maximum of 0.28 m, 0.15 m, and 0.05 m in tanks 1, 2, and 3, respectively, which are considerably above the predetermined value of the bias, that is, 0.03 m.

Let us carry out the same attacks on CP and SC communication channels while considering the developed resilient set-membership tracking control and estimation methods. As illustrated in Figure 4.5, the liquid level in each tank can perfectly track the reference signal although the availability of the control commands \( u_1 \) and \( u_2 \) and the integrity of the measurement outputs \( y_1 \) and \( y_2 \) are compromised by the DoS attack.
Resilient control and estimation

Figure 4.4 (a) Non-resilient tracking error. (b) Non-resilient estimation error.

Figure 4.5 Resilient tracking control and estimation performance on liquid levels (a) \( h_1 \); (b) \( h_2 \); and (c) \( h_3 \).

on the CP channels and the deception attack on the SC channels, respectively. More specifically, when these attacks are present, the two-step set-membership estimator adjusts the gains \( \hat{A}_k \) and \( \hat{B}_{k+1} \) in (4.23) such that deviation of the state estimation from the true state is minimized and thus, the prediction and estimation errors are constrained by the predefined bounding ellipsoidal sets in (4.26) and (4.27). The secure state estimation is employed by the observer-based tracking control protocol (4.24) in order to generate the tracking control command. However, the control signal is subject to be corrupted by the DoS attack on CP communication channels. Therefore,
the tracking controller modifies the gain $M_k$ in the proposed protocol (4.24) at every instant of time so that the tracking error is constrained by the predefined bounding ellipsoidal set in (4.28). Therefore, the proposed resilient set-membership tracking control and estimation methods can lead to a significant reduction in the tracking and estimation errors compared with the non-resilient case. As shown in Figure 4.6a, the maximum tracking error drops approximately into 0.025 m in tank $T_1$ and 0.02 m in tanks $T_2$, and $T_3$, respectively, which are less than the predetermined value of the bias, that is, 0.03 m. Furthermore, from Figure 4.6b, the maximum estimation error decreases to about 0.01 m in each tank.

### 4.6 Conclusion

The problem of resilient tracking control in an NCS subject to two different attacks, UBB process noise and measurement noise, and limited communication capacity has been addressed. More specifically, a DoS attack has been considered on the communication channels between a remote controller and a physical plant, where the adversary violates the availability of the tracking control command. In addition, the adversary compromises the integrity of the communication channels between a sensor and a remote controller by launching a deception attack on those channels. Furthermore, due to limited communication capacity in real-world application of wireless networks, it has been considered that the signals are quantized by a logarithmic quantizer prior to enter the communication networks. The concept of resilient set-membership tracking control technique has been introduced. Also, a resilient two-step set-membership estimator is provided to secure the state estimation against those attacks. To solve out the parameters of the resilient tracking controller and the resilient estimator, a convex
optimization algorithm has been proposed by recursively computing of confidence reference state ellipsoid and state estimation ellipsoids, which contain the true state of the system. The effectiveness and applicability of the derived results have been verified via an ITTS. It should be mentioned that when modeling cyber attacks on the CP and SC communication channels, one may take into account a combination of deception attacks and DoS attacks to form a sophisticated and stealthy attack model and therefore, it constitutes one of our future research work. Furthermore, this article can be further developed by considering the resource-efficient issue in NCSs subject to the presence of cyber attacks and limited communication resources. The exiting event-triggered control techniques provide some useful insights for the design of the desired resilient event-triggered tracking control methods for this class of NCSs.
Summary and Conclusions

In this chapter, the contributions of this thesis are summarized and its principal research outcomes are highlighted. This chapter concludes with some suggestions for the further research and explorations in the relevant field of study.

5.1 Research accomplishments

(i) The first outcome of this research is the development of a novel attack detection method based on a set-membership filtering technique. To yield this, an ellipsoidal set-membership state estimate algorithm has been constructed, which recursively computes the state of a general linear time-varying system in two steps: 1) a prediction step and 2) a measurement update step. The aim of the two-step state estimate algorithm is to find a group of confidence prediction ellipsoidal sets and estimation ellipsoidal sets which always guarantee to enclose the true state of the system regardless of UBB process noise and measurement noise, providing the system is free of any attack. In the measurement update step, an estimation ellipsoidal set is calculated through updating the prediction ellipsoidal set with the current measurement output. The developed novel attack detection method relies on the existence of intersection between the two ellipsoidal sets. Ideally, in an attack-free system, the two ellipsoidal sets will always have intersection as they both contain the true state of the system. However, if there is an attack on the system, one of the two sets may not include the true state since the center of that set is affected by the attack. Therefore, one can conclude that there must be an attack compromising the system performance if there is no intersection between the ellipsoidal sets. As illustrated in Chapter 2, attacks on control signals can be detected if the prediction ellipsoidal set does not have intersection with the estimation set updated with the measurement data received at the previous time instant. Furthermore, attacks on sensors measurements can be detected if there is no intersection between the prediction ellipsoidal set and the estimation ellipsoidal set updated with the measurement data obtained at the current time instant.
(ii) The second outcome of this research is the implementation of the developed novel attack detection in a fully distributed framework into a vehicular platoon system, which represents a large-scale NCS. The centralized attack detection method presented in Chapter 2 is not straightforwardly applicable for a large-scale vehicular platoon system due to the fact that this method requires full knowledge of the entire platoon information. Hence, the computational overhead for this centralized detection method is quite high and it may make the use of the detection system unrealistic. To address the problem of the detection of attacks compromising the shared communication network and on-board sensors employed in a vehicle platooning system, as illustrated in Chapter 3, a novel attack detection method has been developed in a fully distributed manner in the sense that each vehicle is equipped with its own detection system. The distributed set-membership filtering algorithm allows each vehicle in the platoon to estimate their states with no need of having full knowledge of the entire platoon, which reduces the computational overhead for this model-based detection method and makes the method suitable for a large-scale platoon. Furthermore, in Chapter 3, two recovery mechanisms have been introduced into the proposed algorithm so as to mitigate the malicious effects of attacks. The developed recovery mechanisms make the state estimation secure against attacks and then the controller uses the secure estimation to generate the control signal so that the string stability of the platoon is satisfied while the attack is present.

(iii) The final outcome of this research is the development of a resilient tracking control method for an NCS subject to two various attacks, UBB process noise and measurement noise, and limited communication capacity. More specifically, a DoS attack has been considered on the communication channels between a remote controller and a physical plant, where the adversary violates the availability of the tracking control command. In addition, the adversary compromises the integrity of the communication channels between a sensor and a remote controller by launching a UBB deception attack on those channels. Furthermore, due to limited communication capacity in real-world application of wireless networks, it has been considered that the signals are quantized by a logarithmic quantizer prior to enter the communication networks. The main motivation of this study is based on the fact that it is quite common for a crafty adversary to launch assorted attacks of different models and strategies so that he could heavily disrupt the tracking performance. Furthermore, it has been well acknowledged that different
attack strategies are generally stealthy to any defender. Therefore, in Chapter 4, first, a concept of resilient set-membership tracking control has been presented, through which the system’s true state is guaranteed to reside in a bounding ellipsoidal set of the reference state regardless of the existence of attacks and UBB noises. Second, in the case that full information of the system’s state is not implicitly trusted in the presence of attacks, a resilient set-membership estimation strategy has been provided to secure the state estimates against attacks. To solve out the parameters of the resilient tracking controller and the resilient estimator, a convex optimization algorithm has been proposed by recursively computing of confidence reference state ellipsoid and state estimation ellipsoids, which contain the true state of the system.

5.2 Recommendations for future work

Based on the literature review performed on each chapter and research outcomes presented in this thesis, the following gaps have been identified. We strongly believe that scientific community would benefit from research that completely or partially addresses the following key issues.

(i) In many practical engineering systems, faults in the system’s components, i.e., sensors and actuators, commonly exist. In the coexistence of the faults in the system, the proposed set-membership state estimation method in Chapter 2 and Chapter 3 can be refined as a fault tolerant filter by considering the UBB fault signals in the system dynamics. In doing so, the refined filter can tolerate the impact of the fault-like attack in a same way as it does with respect to the fault. Such a fault-tolerant state estimation method would be advantageous in the enhancement of our proposed attack detection system in the sense that attacks can be distinguished from the faults whose patterns are different from those of attacks.

(ii) As discussed in [102], there are many approaches representing the different IFT strategies among vehicles in a platoon. Our distributed attack detection technique and recovery mechanisms presented in Chapter 3 have this flexibility to be modified for a specific type of IFT. Therefore, it would be beneficial to researchers to develop a refined attack detection method which can be employed for various types of IFT.
(iii) In Chapter 4, the developed resilient tracking control framework assumes only DoS attacks on the CP communication channels and deception attacks on the SC communication channels, respectively. It is worthwhile to a sophisticated and stealthy attack model by considering a combination of deception attack and DoS attacks. Interested researchers may follow similar analysis and design procedures for the manipulated control signals presented in Chapter 4 to derive the relevant results.

(iv) This study can be further developed by considering the resource-efficient issue in NCSs subject to the presence of cyber attacks and limited communication resources. The exiting event-triggered control techniques provide some useful insights for the design of the desired resilient event-triggered tracking control methods for this class of NCSs.
References


References


Appendix A

A.1 Proof of Theorem 3.1

At time $k$, if $(x^i_k - \hat{x}^i_k)^T P_k^{-1} (x^i_k - \hat{x}^i_k) \leq 1$, by Schur complement [67], the inequality is equivalent to $(x^i_k - \hat{x}^i_k)(x^i_k - \hat{x}^i_k)^T \leq P_k^i$. From Cholesky factorization, one has $P_k^i = E_k^i E_k^{iT}$. Let $\vartheta_1 = E_k^{-1}(x^i_k - \hat{x}^i_k)$, then we have

$$\vartheta_1^T \vartheta_1 = (x^i_k - \hat{x}^i_k)^T P_k^{-1} (x^i_k - \hat{x}^i_k) \leq 1 \tag{A.1}$$

which means that there exists a vector $\vartheta_1$ satisfying $\|\vartheta_1\| \leq 1$ such that $x^i_k = \hat{x}^i_k + E_k^i \vartheta_1$.

The prediction tracking error can be calculated as

$$x^i_{k+1} - \hat{x}^i_{k+1|k} = \Pi^i_{1,k} \vartheta^i_{1,k} \tag{A.2}$$

where $\Pi^i_{1,k} = [\Theta^i_{1,k}, AE_k^i, B_w]$ and $\vartheta^i_{1,k} = [1, \vartheta_1^T, w^i_k]^T$. Thus, the condition in (3.15) can be rewritten as

$$\eta^T_{1,k} (-\text{diag} \{1, 0, 0\} + \Pi^T_{1,k} P_{k+1|k}^{-1} \Pi^i_{1,k}) \eta^i_{1,k} \leq 0. \tag{A.3}$$

From (3.5) and (A.1), the unknown variables $w^i_k$ and $\vartheta_1$ satisfy

$$\begin{cases} w^i_k \Pi^T_{k+1|k} w^i_k \leq 1 \\ \|\vartheta_1\| \leq 1 \end{cases}$$

which can be expressed in $\eta^i_{1,k}$ as

$$\begin{cases} \eta^T_{1,k} \Xi_{1,k} \eta^i_{1,k} \geq 0 \\ \eta^T_{1,k} \Xi_{2,k} \eta^i_{1,k} \geq 0 \end{cases} \tag{A.4}$$

where $\Xi_{1,k} = \text{diag} \{1, 0, -W^i_k\}$ and $\Xi_{2,k} = \text{diag} \{1, -I, 0\}$.

By virtue of $S$-procedure [67], one obtains from (A.3) and (A.4) that the inequality (A.3) is satisfied if there exist scalar sequences $\tau^i_{m,k} > 0$, $m = 1, 2$ such that

$$-\text{diag} \{1, 0, 0\} + \Pi^T_{1,k} P_{k+1|k}^{-1} \Pi^i_{1,k} + \tau^i_{1,k} \Xi_{1,k} + \tau^i_{2,k} \Xi_{2,k} \leq 0. \tag{A.5}$$
Using Schur complement to (A.5) straightforwardly gets (3.20). This completes the proof. ■

A.2 Proof of Theorem 3.2

From Theorem 3.1, if \(x_{k+1}^i\) belongs to the ellipsoid \((x_{k+1}^i - \hat{x}_{k+1|k}^i)^T P_{k+1|k}^{-1} (x_{k+1}^i - \hat{x}_{k+1|k}^i) \leq 1\), then there exists a vector \(\vartheta_2\) satisfying

\[
\vartheta_2^T \vartheta_2 = (x_{k+1}^i - \hat{x}_{k+1|k}^i)^T P_{k+1|k}^{-1} (x_{k+1}^i - \hat{x}_{k+1|k}^i) \leq 1
\]  

(A.6)

such that \(x_{k+1}^i = \hat{x}_{k+1|k}^i + E_{k+1|k}^i \vartheta_2\), where \(E_{k+1|k}^i\) is a factorization of \(P_{k+1|k}^i = E_{k+1|k}^i E_{k+1|k}^T\). The one-step ahead state estimation error can be obtained as

\[
x_{k+1}^i - \hat{x}_{k+1|k}^i = \Pi_{2,k+1}^i \eta_{2,k+1}^i
\]  

(A.7)

where \(\Pi_{2,k+1}^i = [0, \hat{\Theta}_{1,k}^i - \hat{B}_{k+1}^i D]\) and \(\eta_{2,k+1}^i = [1, \vartheta_2^T, \nu_{k+1}^T, \nu_{k+1}^T]^T\). Hence, the condition in (3.17) can be rewritten as

\[
\eta_{2,k+1}^T (\text{diag} \{1, 0, 0\} + \Pi_{2,k+1}^i P_{k+1|k}^{-1} \Pi_{2,k+1}^i) \eta_{2,k+1}^i \leq 0.
\]  

(A.8)

From (3.6) and (A.6), it is clear that the unknown variables \(\nu_{k+1}^i\) and \(\vartheta_2\) satisfy

\[
\begin{align*}
\nu_{k+1}^T V_{k+1}^{-1} \nu_{k+1}^i & \leq 1 \\
\|\vartheta_2\| & \leq 1
\end{align*}
\]

which can be rewritten as

\[
\begin{align*}
\eta_{2,k+1}^T \Xi_{3,k+1} \eta_{2,k+1}^i & \geq 0 \\
\eta_{2,k+1}^T \Xi_{4,k+1} \eta_{2,k+1}^i & \geq 0
\end{align*}
\]  

(A.9)

where \(\Xi_{3,k+1} = \text{diag} \{1, 0, -V_{k+1}^{-1}\}\) and \(\Xi_{4,k+1} = \Xi_{2,k}\).

Applying \(S\)-procedure, (A.8) is satisfied if there exist scalar sequences \(\tau_{m,k}^i > 0, m = 3, 4\) such that

\[
-\text{diag} \{1, 0, 0\} + \Pi_{2,k+1}^i P_{k+1|k}^{-1} \Pi_{2,k+1}^i + \tau_{3,k}^i \Xi_{3,k+1} + \tau_{4,k}^i \Xi_{4,k+1} \leq 0.
\]  

(A.10)
The output constraint (3.18) can be expressed in terms of $\tilde{\psi}_{k+1}^i$ as

$$\Pi_{y,k+1}^i \eta_{2,k+1}^i = 0$$  \hspace{1cm} (A.11)

with $\Pi_{y,k+1}^i$ defined in (3.21). Applying Finsler’s lemma [67], (A.10) under constraint (A.11) (i.e. (3.18)) holds if there exists an $N_{k+1}^i$ such that

$$\Pi_{2,k+1}^T P_{k+1}^{-1} \Pi_{2,k+1}^i + \tau_{3,k}^i \Xi_{3,k+1} + \tau_{4,k}^i \Xi_{4,k+1} + N_{k+1}^i \Pi_{y,k+1}^i + \Pi_{y,k+1}^T N_{k+1}^i \leq 0.$$  \hspace{1cm} (A.12)

Using Schur complement to (A.12) yields (3.21). This completes the proof. ■
Appendix B

B.1 Proof of Theorem 4.1

The proof is twofold.

1) Proof of resilient set-membership tracking control.

At time $k$, if $(x_k - x^r_k)^T R_k^{-1} (x_k - x^r_k) \leq 1$, by Schur complement, the inequality is equivalent to $(x_k - x^r_k)(x_k - x^r_k)^T \leq R_k$. From Cholesky factorization, one has $R_k = L_k L_k^T$. Let $\vartheta_1 = L_k^{-1} (x_k - x^r_k)$, then we have

$$\vartheta_1^T \vartheta_1 = (x_k - x^r_k)^T R_k^{-1} (x_k - x^r_k) \leq 1 \quad (B.1)$$

which means that there is a vector $\vartheta_1$ satisfying $\|\vartheta_1\| \leq 1$ such that $x_k = x^r_k + L_k \vartheta_1$.

For brevity, we denote $\sigma^u_k = \sigma^u(u_k)$ and $\bar{u}_k = \tilde{\Gamma}_k u^w_k$. Then, the one-step ahead state tracking error can be calculated as

$$x_{k+1} - x^r_{k+1} = \Pi^c_k \eta^c_k \quad (B.2)$$

where $\eta^c_k = [1, \vartheta_1^T, w_k^T, \bar{u}_k^T, \sigma^u_k]^T$ and $\Pi^c_k$ is defined in (4.31). Thus, the condition in (4.28) can be rewritten as

$$\eta^c_k^T (\Delta^c + \Pi^c_k^T R_{k+1}^{-1} \Pi^c_k) \eta^c_k \leq 0 \quad (B.3)$$

where $\Delta^c = -\text{diag}\{1, 0, 0, 0, 0\}$.

By virtue of the elementary inequality $2a^T b \leq a^T a + b^T b$, (4.14), and (4.22), one can conclude that

$$\bar{u}_k^T \bar{u}_k \leq 2 (w_k^T H_1 u^w_k + \sigma^u_k^T \tilde{\Gamma}_k^2 \sigma^u_k) \quad (B.4)$$

where $\tilde{\Gamma}_k^2 = \tilde{\Gamma}_k^T \tilde{\Gamma}_k$ since $\tilde{\Gamma}_k$ is a diagonal matrix. From (4.3) and (B.1), it is clear that $w_k$ and $\vartheta_1$ satisfy

$$\begin{cases}
    w_k^T W_k^{-1} w_k \leq 1 \\
    \|\vartheta_1\| \leq 1
\end{cases} \iff \begin{cases}
    \eta^c_k^T \Xi^c_k \eta^c_k \leq 0 \\
    \eta^c_k^T \Xi^c_k \eta^c_k \leq 0
\end{cases} \quad (B.5)$$

where $\Xi^c_k = \text{diag}\{-1, 0, W_k^{-1}, 0, 0\}$ and $\Xi^c_k = \text{diag}\{-1, I, 0, 0, 0\}$. 
On the other hand, the constraint in (B.4) can be written in $\eta^c_k$ as

$$\eta^T_k \Omega^c_k \eta_k \leq 0$$  \hspace{1cm} (B.6)

where $\Omega^c_k = \text{diag}\{-2(\hat{x}_k - x^r_k)^T M_k^T H_1^u T \hat{\Gamma}^2_k H_1^u M_k(\hat{x}_k - x^r_k), 0, 0, I, -2\Gamma^2_k\}$. Furthermore, substituting $u_k$ from (4.24) into (4.9) yields

$$\sigma_k^T (\sigma_k^u - H_k^u M_k(\hat{x}_k - x^r_k)) \leq 0$$

which can be expressed in $\eta^c_k$ as

$$\eta^T_k \Omega^c_k \eta_k \leq 0$$  \hspace{1cm} (B.7)

where $\Omega^c_k = [\Omega^c_{\alpha,\beta}(k)]_{5 \times 5}$ is a time-varying sparse symmetric block matrix whose nonzero entries given by $\Omega^c_{\alpha,\beta}(k) = -\Theta^c_{\alpha,\beta}(k)$, $\Omega^c_{\beta,\beta}(k) = -\Theta^c_{\beta,\beta}(k) + 2\Gamma^2_k$.

Applying S-procedure, (B.3) holds if there exist positive scalar sequences $\tau_{m,k}$, $m = 1, 2$, such that

$$\Pi_k^T R_{k+1}^{-1} \Pi_k + \Delta^c - \Omega^c_k - \tau_{1,k} \Xi^c_k - \tau_{2,k} \Xi^c_k \leq 0.$$  \hspace{1cm} (B.8)

Applying Schur complement to (B.8) straightforwardly gets the right matrix inequality in (4.31). This completes the proof. \hfill \blacksquare

2) Proof of resilient set-membership state estimation at the predicted state estimate step.

The proof is similar to that of Theorem 4.1, part 1, and is therefore omitted due to the page limit.

\section*{B.2 Proof of Theorem 4.2}

From Theorem 4.1, if $x_{k+1}$ belongs to $(x_{k+1} - \hat{x}_{k+1|k})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1$, then there exists a vector $\vartheta_3$ satisfying

$$\vartheta_3^T \vartheta_3 = (x_{k+1} - \hat{x}_{k+1|k})^T P_{k+1|k}^{-1} (x_{k+1} - \hat{x}_{k+1|k}) \leq 1$$  \hspace{1cm} (B.9)

such that $x_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3$, where $E_{k+1|k}$ is a factorization of $P_{k+1|k} = E_{k+1|k} E_{k+1|k}^T$. By denoting $\sigma^y(y_k) = \sigma^y_k$ and $\psi(y^q_k) = \psi_k$, the one-step ahead state
estimation error can be obtained as

\[ x_{k+1} - \hat{x}_{k+1} = \Pi_k^e \eta_k^e \tag{B.10} \]

where \( \eta_k^e = [1 \ \theta_3^T \ \nu_{k+1}^T \ \delta_{k+1}^T \ \sigma_{k+1}^y \ \psi_{k+1}^T]^T \) and \( \Pi_k^e \) is defined in (4.32). Hence, the condition in (4.27) can be rewritten as

\[ \eta_k^e^T (\Delta^e + \Pi_k^e \gamma_{k+1} \Pi_k^e) \eta_k^e \leq 0 \tag{B.11} \]

where \( \Delta^e = -\text{diag}\{1, 0, 0, 0, 0\} \).

From (4.4), (4.16), and (B.9), we have the following constraints

\[
\begin{cases} 
\nu_{k+1}^T V_{k+1}^{-1} \nu_{k+1} \leq 1 \\
\delta_{k+1}^T S_{k+1}^{-1} \delta_{k+1} \leq 1 \\
\left\| \vartheta_3 \right\| \leq 1
\end{cases} \iff \begin{cases} 
\eta_k^e \Xi_{k,1} \eta_k^e \leq 0 \\
\eta_k^e \Xi_{k,2} \eta_k^e \leq 0 \\
\eta_k^e \Xi_{k,3} \eta_k^e \leq 0
\end{cases} \tag{B.12} \]

where \( \Xi_{k,1} = \text{diag}\{-1, 0, V_{k+1}^{-1}, 0, 0\} \), \( \Xi_{k,2} = \text{diag}\{-1, 0, S_{k+1}^{-1}, 0, 0\} \), and \( \Xi_{k,3} = \text{diag}\{-1, I, 0, 0, 0\} \).

From (4.22) and noting that \( x_{k+1} = \hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3 \), it can be easily obtained that

\[
y_{k+1} = C_{k+1}(\hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3) + D_{k+1} v_{k+1} \\
y_{k+1}^q = -H_1^y C_{k+1}(\hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3) - H_1^y D_{k+1} v_{k+1} - \sigma_{k+1}^y + \delta_{k+1}. \tag{B.13} \]

Substituting (B.13) into (4.9) and also inserting (B.14) into (4.20), respectively, yield

\[
\begin{cases} 
\sigma_{k+1}^y \left( \sigma_{k+1}^y - H^y C_{k+1}(\hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3) - H^y D_{k+1} v_{k+1} \right) \leq 0 \\
\psi_{k+1}^T \left( \psi_{k+1} - \tilde{\Lambda}_{k+1}( -H_1^y C_{k+1}(\hat{x}_{k+1|k} + E_{k+1|k} \vartheta_3) - H_1^y D_{k+1} v_{k+1} - \sigma_{k+1}^y + \delta_{k+1} \right) \leq 0 
\end{cases} \tag{B.15} \]

which can be written in \( \eta_k^e \) as

\[ \eta_k^e \Omega \eta_k^e \leq 0 \tag{B.16} \]
where $\Omega^e_k = [\Omega^e_{\alpha,\beta}(k)]_{6 \times 6}$ is a time-varying sparse symmetric block matrix with its nonzero entries given by

\begin{align*}
\Omega^e_{1,5}(k) &= -(1/2)\hat{x}^T_{k+1|k}C^T_{k+1}H^y, \\
\Omega^e_{1,6}(k) &= (1/2)\hat{x}^T_{k+1|k}C^T_{k+1}H^y\tilde{\Lambda}_{k+1}, \\
\Omega^e_{2,5}(k) &= -(1/2)E^T_{k+1|k}C^T_{k+1|k}H^y, \\
\Omega^e_{2,6}(k) &= (1/2)E^T_{k+1|k}C^T_{k+1|k}H^y\tilde{\Lambda}_{k+1}, \\
\Omega^e_{3,5}(k) &= -(1/2)D^T_{k+1}H^y, \\
\Omega^e_{3,6}(k) &= (1/2)D^T_{k+1}H^y\tilde{\Lambda}_{k+1}, \\
\Omega^e_{4,6}(k) &= -(1/2)\Lambda_{k+1}, \\
\Omega^e_{5,6}(k) &= -\Omega^e_{4,6}(k), \\
\Omega^e_{5,5}(k) &= \Omega^e_{6,6}(k) = I.
\end{align*}

By virtue of $\mathcal{S}$-procedure, (B.11) holds if there exist positive scalar sequences $\tau_{m,k}$, $m = 5, 6, \ldots, 8$, such that

\begin{equation}
\Pi_k^T P_{k+1}^{-1} \Pi_k^e + \Delta^e - \tau_{5,k} \Xi^e_{k} - \tau_{6,k} \Xi^e_{k} - \tau_{7,k} \Xi^e_{k} - \tau_{8,k} \Omega^e_k \leq 0. \tag{B.17}
\end{equation}

By applying the Schur complement and noting that $S_{k+1} = \tau_{7,k} \tilde{S}_{k+1}$, (4.32) can be derived from (B.17). This completes the proof. $\blacksquare$
Appendix C

Inclusion of papers within the thesis guidelines and policy

Overview

HDR candidates may include one or more papers within the body of their thesis where such papers have been produced under supervision and during the period of candidature; and where the quality of such papers is appropriate to Doctoral or Masters (Research) level research. A thesis prepared in this way is a different thesis format, it is not a different degree. There are several advantages to organising a thesis in this way:

- Preparing papers for publication saves time when preparing the thesis for examination as papers may make up one, or several, chapters within the thesis.

- It is to your advantage to publish work from your thesis as a means of disseminating your research, and developing your writing skills.

- It may improve the quality of your thesis as part of your thesis has already been subjected to peer review.

- Examiners may have more confidence in your thesis if they can see that you have already published your research. In addition, you will have already met one of the criteria of examination, with the thesis suitable for publication.

As a candidature requirement, all doctoral candidates are expected to have at least one peer reviewed output accepted for publication during candidature. Whilst not compulsory, candidates are encouraged to include this publication in the body of the thesis due to the advantages as outlined above.

Requirements of inclusion of papers within the thesis

Higher degree by research is a program of independent supervised study that produces significant and original research outcomes, culminating in a thesis, exegesis or equivalent. Inclusion of papers within a thesis is not a suitable thesis format for all research projects,
for example: collaborative projects where there may be several co-authors for each paper which may make it difficult for the examiner to establish the independence of the candidates work; where primary data is not collected, or results obtained, until late in the candidature; or where the research will not produce a logical sequence of papers that are able to be presented as an integrated whole. Candidates should also take into account whether this thesis format is an accepted practice within their discipline and likely to be received well by the thesis examiners. Candidates are required to consult with their supervisor(s) early in their candidature to determine if this thesis format is appropriate. It is expected that candidates will identify as part of the confirmation of candidature milestone if their thesis is to be prepared in this format. Candidates should consult their Group specific guidelines in addition to the requirements detailed below. Candidates are also encouraged to attend the workshop: “Inclusion of papers within a thesis” offered by the Griffith Graduate Research School.

**Status of papers:** A thesis may include papers that have been submitted, accepted for publication, or published. Some disciplines may specify a variation to the status of papers requirement, refer to your Group specific guidelines.

**Type of papers:** For the purpose of this requirement, papers are defined as a journal article, conference publication, book or book chapter. Papers which have been rejected by a publisher must not be included unless they have been substantially rewritten to address the reviewers’ comments, or have since been accepted for publication. Some disciplines may specify a variation to the type of papers requirement, refer to your Group specific guidelines.

**Number of papers:** A thesis may be entirely or partly comprised of papers. A paper may be included as a single chapter if the paper contributes to the argument of the thesis, or several papers may form the core chapters of the theses where they present a cohesive argument. Where a thesis is entirely comprised of papers, there is no minimum requirement for the number of papers that must be included (except as noted below) and is a matter of professional judgment for the supervisor and the candidate. Overall, the material presented for examination needs to reflect the research thesis standard required for the award of the degree. For example, PhD candidates, on the basis of a program of independent supervised study, must produce a thesis that makes a significant and original contribution to knowledge and understanding in the relevant field of study. This remains a matter of professional judgment for the supervisor and the candidate.
Where a thesis is entirely comprised of papers, some disciplines may specify a minimum number of papers to be included, refer to your Group specific guidelines.

**Authorship:** The candidate should normally be principal author (that is, responsible for the intellectual content and the majority of writing of the text) of any work included in the body of the thesis. Where a paper has been co-authored, the candidate is required to have made a substantial contribution to the intellectual content and writing of the text, co-authored work in which the candidate was a minor author can only be used and referenced in the way common to any other research publication cited in the thesis. A signature from the corresponding author is required in order to include co-authored material in the body of the thesis.

For co-authored papers, the attribution of authorship must be in accordance with the Griffith University Code for the Responsible Conduct of Research, which specifies that authorship must be based on substantial contributions in one or more of:

- conception and design of the research project
- analysis and interpretation of research data
- drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output.

Some disciplines may specify a variation to the authorship requirement, refer to your Group specific guidelines.

**Quality of papers:** Candidates should endeavour to publish their research in high quality peer reviewed publications. Papers to be included in the body of the thesis should be published (or submitted for publication) in reputable outlets that are held in higher regard in the relevant field of research. Candidates should consult their supervisor(s) for advice on suitable publications specific to their research discipline. Some disciplines may specify quality standards that must be met for papers to be included, refer to your Group specific guidelines.

**Copyright:** As copyright in an article is normally assigned to a publisher, the publisher must give permission to reproduce the work in the thesis and put a digital copy on the institutional repository. Information on how to seek permission is available at: Copyright and Articles in thesis. If permission cannot be obtained, students may still include the publication in the body of the thesis, however following examination the relevant chapter(s) will be redacted from the digital copy to be held by the Griffith University Library so that the copyright material is not made publicly available in the institutional repository.
Format of thesis

**General:** Consult the thesis preparation and formatting guidelines for general information about the requirements for formatting the thesis. Some disciplines may specify a variation to the thesis format requirements below, refer to your Group specific guidelines.

**Structure of Thesis and linking Chapters:** The structure of the thesis will vary depending on whether the thesis is partly or entirely comprised of papers. Whatever the format, the thesis must present as a coherent and integrated body of work in which the research objectives, relationship to other scholarly work, methodology and strategies employed, and the results obtained are identified, analysed and evaluated.

In general every thesis should include a general introduction and general discussion to frame the internal chapters. The introduction should outline the scope of the research covered by the thesis and include an explanation of the organisation and structure of the thesis. The general discussion should draw together the main findings of the thesis and establish the significance of the work as a whole, and should not just restate the discussion points of each paper.

It is important that candidates explicitly argue the coherence of the work and establish links between the various papers/chapters throughout the thesis. Linking text should be added to introduce each new paper or chapter, with a foreword which introduces the research and establishes its links to previous papers/chapters.

**Format of papers:** The papers may be rewritten for the thesis according to the general formatting guidelines; or they can be inserted in their published format, subject to copyright approval as detailed above. Pagination Candidates may repaginate the papers to be consistent with the thesis. However, this is at the discretion of the candidate.

**Declarations:** All theses that include papers must include declarations which specify the publication status of the paper(s), your contribution to the paper(s), and the copyright status of the paper(s). The declarations must be signed by the corresponding author (where applicable). If you are the sole author, this still needs to be specified. The declaration will need to be inserted at the beginning of the thesis, and for any co-authored papers, additional declarations will need to be inserted at the beginning of each relevant chapter. You may wish to consult the declaration requirements for inclusion of papers diagram to ensure that you insert the correct declaration(s) within the thesis. Please note that completion of the declaration(s)
does not negate the need to comply with any other University requirement relating to co-authored works as outlined in the Griffith University Code for the Responsible Conduct of Research.

Examination requirements

Assessment by Examiners: Theses that include papers are subject to the same examination criteria as theses submitted in the traditional format. It should also be noted that the inclusion of published papers within the thesis does not prevent an examiner from requesting amendments to that material.

Candidates should discuss the suitability of this thesis format for examination with their supervisor(s).

Nomination of examiners: It is the responsibility of the principal supervisor to nominate thesis examiners, and the process dictates that the principal supervisor must approach all nominees to determine their willingness to examine. Where a candidate’s thesis is formatted to include papers, the principal supervisor must also ensure that the examiners are familiar with and/or accepting of, this thesis format.

Upon dispatch of a candidate’s thesis to an examiner, the examiner will be reminded that the thesis has been formatted to include papers. The examiner will also be provided with the relevant information and regulations regarding this thesis format.
Appendix D

The IEEE policy for thesis / dissertation reuse of published material

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, the following statement can be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1. In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

2. In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

3. If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author’s approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1. The following IEEE copyright/credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

2. Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

3. In placing the thesis on the author’s university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity’s name goes here]’s products or
services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.