Titre: Title:	Secure and Privacy-Preserving Cyber-Physical Systems
Auteur: Author:	Kwassi Holali Degue
Date:	2021
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Degue, K. H. (2021). Secure and Privacy-Preserving Cyber-Physical Systems [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5589/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/5589/
Directeurs de recherche: Advisors:	Jérôme Le Ny
Programme: Program:	Génie électrique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Secure and privacy-preserving cyber-physical systems

KWASSI HOLALI DEGUE
Département de génie électrique

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie électrique

JANVIER 2021

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Secure and privacy-preserving cyber-physical systems

présentée par **Kwassi Holali DEGUE**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Lahcen SAYDY, président
Jérôme LE NY, membre et directeur de recherche
Roland MALHAMÉ, membre
George PAPPAS, membre externe

DEDICATION

To my family

ACKNOWLEDGEMENTS

I was very fortunate to have an exceptional advisor, Prof. Jérôme LE NY, during my Ph.D. program. I thank him for giving me the opportunity to perform my Ph.D. research under his supervision, and for his patience and insightful advice in research. His understanding and expertise added considerably to my academic experience. I thank Prof. Lahcen SAYDY for his invaluable advice when I came to Canada. I thank Dr. Denis EFIMOV for his willingness to answer my questions and to always provide me with assistance when needed.

I am very grateful to have remarkable supervisors during the research visits that I have done during this thesis. I thank Prof. Eric FERON, Prof. Magnus EGERSTEDT and Prof. Matthieu BLOCH for their invaluable advice during my visit at Georgia Institute of Technology. I am grateful to Prof. Sandra HIRCHE for giving me the opportunity to work in her team during my visit at the Technical University of Munich (TUM). I thank Prof. Hamsa BALAKRISHNAN for being an exceptional supervisor during my visit at Massachusetts Institute of Technology (MIT).

I thank Fondation Arbour for their financial support that helps me conclude my Ph.D. program. Especially, I am grateful to Dr. Marine HADENGUE and Mrs. Diane DE CHAMPLAIN, whose support was invaluable.

I thank the members of my comitee Prof. Lahcen SAYDY, Prof. Roland MALHAMÉ, Prof. George PAPPAS, and Prof. Michaël KUMMERT for their time and effort to read this thesis.

I was very fortunate to have great professors and colleagues at École Polytechnique de Montréal and GERAD. I thank them for their support, especially my labmates Saad ABOBAKR, Kaijun YANG, Sahar SEDAGHATI, Rabih SALHAB, Wassim RAFRAFI and Feng LI. I am grateful also to the staff of Automation and systems section and the Department of Electrical Engineering of Polytechnique, in particular Mrs. Suzanne LE BEL and Mrs. Nathalie LEVESQUE for making my life at Polytechnique much easier.

I am deeply grateful to my mother, father, sisters, the rest of the family and my family-in-law for their endless support and love. Finally, I dedicate this thesis to my wife, Alexandrine, whose unconditional support, love and patience have always motivated me. I am also grateful to my daughter Ariana, for giving us unlimited happiness and pleasure.

RÉSUMÉ

Dans cette thèse de doctorat, nous étudions le problème de conception d'estimateur et de commande préservant la confidentialité de données dans un système multi-algent composé de systèmes individuels linéaires incertains ainsi que le problème de conception d'attaques furtives et d'estimateurs résilients aux attaques dans les système cyber-physiques. Les systèmes de surveillance et de commande à grande échelle permettant une infrastructure de plus en plus intelligente s'appuient de plus en plus sur des données sensibles obtenues auprès d'agents privés. Par exemple, ces systèmes collectent des données de localisation d'utilisateurs d'un système de transport intelligent ou des données médicales de patients pour une détection intelligente d'épidémie. Cependant, les considérations de confidentialité peuvent rendre les agents réticents à partager les informations nécessaires pour améliorer les performances d'une infrastructure intelligente. Dans le but d'encourager la participation de ces agents, il s'avère important de concevoir des algorithmes qui traitent les données d'une manière qui preserve leur confidentialité.

Durant la première partie de cette thèse, nous considérons des scénarios dans lesquels les systèmes individuels sont indépendants et sont des systèmes linéaires gaussiens. Nous revisitons les problèmes de filtrage de Kalman et de commande linéaire quadratique gaussienne (LQG), sous contraintes de preservation de la confidentialité. Nous aimerions garantir la confidentialité differentielle, une définition formelle et à la pointe de la technologie concernant la confidentialité, et qui garantit que la sortie d'un algorithme ne soit pas trop sensible aux données collectées auprès d'un seul agent. Nous proposons une architecture en deux étapes, qui agrège et combine d'abord les signaux des agents individuels avant d'ajouter du bruit préservant la confidentialité et post-filtrer le résultat à publier. Nous montrons qu'une amélioration significative des performances est offerte par cette architecture par rapport aux architectures standards de perturbations d'entrée à mesure que le nombre de signaux d'entrée augmente. Nous prouvons qu'un pré-filtre optimal d'agrégation statique peut être conçu en résolvant un programme semi-défini. L'architecture en deux étapes, que nous développons d'abord pour le filtrage de Kalman, est ensuite adaptée au problème de commande LQG en exploitant le principe de séparation. A travers des simulations numériques, nous illustrons les améliorations de performance de notre architecture par rapport aux algorithmes de confidentialité différentielle qui n'utilisent pas d'agrégation de signal.

Durant la seconde partie de cette thèse, nous considérons le problème de la conception d'estimateurs préservant la confidentialité pour un système multi-agents composé de différents systèmes linéaires invariants dans le temps affectés par des incertitudes dont les propriétés statistiques ne sont pas connues. Seules des bornes sur ces incertitudes sont connues à priori. Nous proposons une architecture d'estimateur d'intervalle préservant la confidentialité, qui publie des estimations des bornes inférieure et supérieure d'une agrégation des états des systèmes individuels. En particulier, nous ajoutons un bruit borné préservant la confidentialité aux données de chaque agent avant de l'envoyer à l'estimateur. Les estimations publiées par l'observateur garantissent une confidentialité différentielle pour les données des agents. Nous évaluons le comportement de l'architecture proposée lors d'une simulation numérique. Par ailleurs, nous illustrons que les performances de l'architecture proposée peuvent être améliorées en agrégant convenablement les données avant d'ajouter du bruit borné préservant la confidentialité.

Durant la troisième partie de cette thèse, nous étudions le problème de sécurité dans les systèmes cyber-physiques. Les systèmes industriels de commande ont été fréquemment la cible de cyber-attaques au cours de la dernière décennie. Les adversaires peuvent entraver le fonctionnement de ces systèmes en altérant leurs capteurs et actionneurs tout en s'assurant que les systèmes de surveillance ne soient pas en mesure de détecter de telles attaques rapidement. Durant cette partie de la thèse, nous présentons des techniques pour concevoir et faire face aux attaques furtives sur des systèmes de commande linéaires, qui estiment des bornes de leur état à l'aide d'un observateur d'intervalle, en présence de bruit et perturbations inconnus mais bornés. Nous analysons des scénarios dans lesquels un agent malveillant compromet les capteurs et/ou les actionneurs du système avec des signaux d'attaque additifs pour diriger l'estimation d'état en dehors des bornes fournies par l'observateur d'intervalle. D'abord, nous montrons que les séquences d'attaque optimales perturbatrices qui ne sont pas détectées par un moniteur linéaire peuvent être calculées de manière récursive via programmation linéaire. Nous concevons ensuite un observateur d'intervalle résilient aux attaques pour l'état du système. Nous identifions les conditions suffisantes sur les données du capteur pour que la conception d'un tel observateur soit réalisable. Nous proposons une méthode de calcul pour déterminer le gain optimal pour l'observateur en utilisant la programmation semidéfinie et nous construisons aussi des bornes pour le signal d'attaque inconnu. A travers des simulations numériques, nous illustrons la capacité de ces observateurs d'intervalle à toujours fournir des estimations précises en cas d'attaque.

ABSTRACT

This thesis studies the problem of privacy-preserving estimator and control design in a multi-agent system composed of uncertain individual linear systems and the problem of design of undetectable attacks and attack-resilient estimators for cyber-physical systems. Large-scale monitoring and control systems enabling a more intelligent infrastructure increasingly rely on sensitive data obtained from private agents, e.g., location traces collected from the users of an intelligent transportation system or medical records collected from patients for intelligent health monitoring. Nevertheless, privacy considerations can make agents reluctant to share the information necessary to improve the performance of an intelligent infrastructure. In order to encourage the participation of these agents, it becomes then critical to design algorithms that process information in a privacy-preserving way. The first part of this thesis consider scenarios in which the individual agent systems are linear Gaussian systems and are independent. We revisit the Kalman filtering and Linear Quadratic Gaussian (LQG) control problems, subject to privacy constraints. We aim to enforce differential privacy, a formal, state-of-the-art definition of privacy ensuring that the output of an algorithm is not too sensitive to the data collected from any single participating agent. We propose a two-stage architecture, which first aggregates and combines the individual agent signals before adding privacy-preserving noise and post-filtering the result to be published. We show a significant performance improvement offered by this architecture over input perturbation schemes as the number of input signals increases and that an optimal static aggregation stage can be computed by solving a semidefinite program. The two-stage architecture, which we develop first for Kalman filtering, is then adapted to the LQG control problem by leveraging the separation principle. We provide numerical simulations that illustrate the performance improvements over differentially private algorithms without first-stage signal aggregation.

The second part of this thesis considers the problem of privacy-preserving estimator design for a multi-agent system composed of individual linear time-invariant systems affected by uncertainties whose statistical properties are not available. Only bounds are given a priori for these uncertainties. We propose a privacy-preserving interval estimator architecture, which releases publicly estimates of lower and upper bounds for an aggregate of the states of the individual systems. Particularly, we add a bounded privacy-preserving noise to each participant's data before sending it to the estimator. The estimates published by the observer guarantee differential privacy for the agents' data. We provide a numerical simulation that illustrates the behavior of the proposed architecture. Furthermore, we illustrate that the performance of the proposed architecture can be improved by suitably combining the data before adding a bounded privacy-preserving noise.

The third part of this thesis considers the problem of security in cyber-physical systems. Industrial control systems have been frequent targets of cyber attacks during the last decade. Adversaries can hinder the safe operation of these systems by tampering with their sensors and actuators while ensuring that the monitoring systems are not able to detect such attacks in time. This part of the thesis presents methods to design and overcome stealthy attacks on linear time-invariant control systems that estimate lower and upper bounds for their state using an interval observer, in the presence of unknown but bounded noise and perturbations. We analyze scenarios in which a malicious agent compromises the sensors and/or the actuators of the system with additive attack signals to steer the state estimate outside of the bounds provided by the interval observer. We first show that maximally disruptive attack sequences that remain undetected by a linear monitor can be computed recursively via linear programming. We then design an attack-resilient interval observer for the system's state, identifying sufficient conditions on the sensor data for such an observer to be realizable. We propose a computational method to determine optimal observer gains using semi-definite programming and compute bounds for the unknown attack signal as well. In numerical simulations, we illustrate the ability of such interval observers to still provide accurate estimates when under attack.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND NOTATIONS

LQG	Linear Quadratic Gaussian
LTI	Linear Time-Invariant
SDP	Semi-Definite Program
GAS	Globally Asymptotically Stability
SCADA	Supervisory Control And Data Acquisition
ISS	Input-to-State Stability
$X \sim \mathcal{N}(\mu, \Sigma)$	X is a Gaussian random vector with mean vector μ and covariance matrix Σ
iid	Independent and identically distributed
$\lvert x\rvert_p$, for $p \in [1, \infty)$	$(\sum_{i=1}^{k} \lvert x_i\rvert^p)^{1/p}$
$\lVert A\rVert_2$	For a matrix A, the induced 2-norm (maximum singular value of A)
$\lVert A\rVert_F$	The Frobenius norm $\sqrt{\mathrm{Tr}(A^{\mathrm{T}} A)}$
$A \succeq B$	$A - B$ is positive semi-definite
$A \succ B$	$A - B$ is positive definite
$\mathrm{diag}(A_1, \dots, A_n)$	A block-diagonal matrix with the matrices A_i on the diagonal
$\mathbf{1}_n$	The column vector of size n with all components equal to 1
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integer numbers
$\mathbb{R}_+$	$\{\tau \in \mathbb{R} : \tau \geq 0\}$
$\mathbb{R}^n_{>0}$	The cones of vectors of dimension n with positive components
$\mathbb{R}^n_+$	The cones of vectors of dimension n with nonnegative components
$\lvert x\rvert$	The Euclidean norm for a vector $x \in \mathbb{R}^n$
I_n	The $n \times n$ identity matrix
$\lVert u\rVert_{[t_0,t_1]}$	$\sup_{t\in[t_0,t_1]} \lvert u_t\rvert$
$\lVert u\rVert$	$\lVert u\rVert_{[t_0,+\infty]}$
$\mathcal{L}^n_\infty$	The set of all $\mathbb{R}^n$-valued signals u with the property $\lVert u\rVert < \infty$
$\underline{u}$	*Component-wise* lower bounds for the vector or signal u
$\overline{u}$	*Component-wise* upper bounds for the vector or signal u
$s_{t_1:t_2}$, for any $t_1 \leq t_2$	$s_{t_1:t_2} = \begin{bmatrix} s_{t_1}^{\mathrm{T}} & s_{t_1+1}^{\mathrm{T}} & \dots & s_{t_2}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$
$x_1 \leq x_2$	For two vectors $x_1, x_2 \in \mathbb{R}^n$, it is understood componentwise
$A_1 \leq A_2$	For two matrices $A_1, A_2 \in \mathbb{R}^{n\times n}$, it is understood componentwise
$\mathbb{D}^n_{>0}$	The set of diagonal matrices of dimension $n \times n$ with diagonal positive elements.

A^+	$\max\{0, A\}$ applied elementwise
A^-	$A^+ - A$
$\lvert A \rvert$	Given a matrix $A \in \mathbb{R}^{m \times n}$, $\lvert A \rvert = A^+ + A^-$
$*$	In a symmetric block matrix, the symbol $*$ denotes a term that can be deduced by symmetry

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

To monitor and control intelligent infrastructure systems such as smart grids, smart buildings or smart cities, data needs to be continuously collected from the people interacting with these systems, either through sensors installed in the environment such as cameras and smart meters, or through personal devices such as smartphones. Hence, in contrast to more traditional control systems, the measured signals for such systems often contain highly privacy-sensitive information, e.g., related to the real-time location or health of a person. For example, the accuracy of crowd-sourced traffic maps and congestion-aware routing applications is increased by using data provided by smartphones and connected vehicles [1]. However, individual location data turns out to be very difficult to properly anonymize because individuals have highly unique mobility patterns [2, 3], and in fact individual trajectories can be reconstructed even from just aggregate location data [4, 5]. Similarly, fine-grained measurements of a house's electric power consumption collected by a smart meter can enable demand-response schemes, but can also be used to infer the activities of the occupants, by identifying the usage of individual appliances [6–9]. Therefore, it is necessary to implement privacy-preserving mechanisms when sensitive data must be shared to improve a system's performance. The first part of this thesis focuses on the problems of designing *Kalman filters and LQG controllers under privacy constraints on the measured signals*. These problems arise when a data collector measures private signals originating from a population of agents, whose dynamics can be modeled as linear Gaussian systems, in order to publish in real-time either an estimate of an aggregate state of the agent population, or a control signal shared with the agents and aimed at regulating such an aggregate state. As a motivating example, one can consider the problem of controlling the distribution of vehicles on a road network by means of traffic messages broadcasted to all cars, with the current density estimated from location data obtained from the smartphones of individual drivers. The second part of this thesis considers the problem of designing an *interval observer under privacy constraints on the measured signals*. This problem arises when a data aggregator collects privacy-sensitive signals originating from a population of agents, whose dynamics can be modeled as uncertain linear time-invariant (LTI) systems with unknown but bounded inputs.

Moreover, an increasing number of safety-critical systems that once required physical access for interaction can now be remotely monitored and controlled, and are sometimes even connected to untrusted networks. While this can lead to cost savings and increased operational efficiency for Cyber-Physical Systems (CPS), i.e., systems that tightly integrate computing and communication resources to control physical processes, it also creates new vulnerabilities, allowing for new types of cyberattacks that can lead to disastrous physical damage. Examples of CPS are industrial control systems, sensor networks, and critical infrastructures such as power generation and distribution networks, transportation networks, water and gas distribution networks, and advanced manufacturing systems. According to [10], there were 675,186 Supervisory control and data acquisition (SCADA) attacks targeting operational capabilities within power plants, factories, and refineries worldwide in 2014, which represents an increase of 636 percent between 2012 and 2014. Some particularly harmful recent examples of attacks on CPS include the StuxNet malware [11] and the Maroochy sewage control incident [12]. Commercial drones [13] and military vehicles [14] have also been targeted. Adversaries can hinder the safe operation of these systems by attacking sensors and actuators embedded in them, if the monitoring system is not able to detect such attacks in time. For the StuxNet attack, several control systems of an Iranian nuclear-enrichment plant were infected by a computer worm designed for this purpose. The centrifuges' measurements were corrupted in order to alter the centrifuges' actuation signals to compel them to spin out of control while remaining undetected by the monitor by indicating a normal operation. The StuxNet example illustrates vulnerabilities of CPS and motivates the need to design efficient attack-resilient monitors. The third part of this thesis considers the problem of *designing and overcoming stealthy attacks on LTI systems subject to unknown but bounded uncertainties.*

1.1 Notation

Let us introduce some notation used throughout this thesis. We fix a generic probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ a probability measure defined on $\mathcal{F}$. The notation $X \sim \mathcal{N}(\mu, \Sigma)$ means that X is a Gaussian random vector with mean vector μ and covariance matrix Σ. "Independent and identically distributed" is abbreviated iid. We denote the p-norm of a vector $x \in \mathbb{R}^k$ by $|x|_p := (\sum_{i=1}^k |x_i|^p)^{1/p}$, for $p \in [1, \infty)$. For a matrix A, the induced 2-norm (maximum singular value of A) is denoted $\|A\|_2$ and the Frobenius norm $\|A\|_F := \sqrt{\mathrm{Tr}(A^{\mathrm{T}} A)}$. If A and B are symmetric matrices, $A \succeq B$ (resp. $A \succ B$) means that $A - B$ is positive semi-definite (resp. positive definite). We use the notation $\mathrm{diag}(A_1, \ldots, A_n)$ to represent a block-diagonal matrix with the matrices A_i on the diagonal. The column vector of size n with all components equal to 1 is denoted $\mathbf{1}_n$. The minimum and maximum eigenvalues of a symmetric matrix A are $\lambda_{min}(A)$, $\lambda_{\max}(A)$.

We denote the real and integer numbers by $\mathbb{R}$ and $\mathbb{Z}$ respectively, and let $\mathbb{R}_+ = \{\tau \in \mathbb{R} : \tau \geq 0\}$ and $\mathbb{Z}_+ = \mathbb{Z} \cap \mathbb{R}_+$. The cones of vectors of dimension n with positive and nonnegative components are denoted $\mathbb{R}^n_{>0}$ and $\mathbb{R}^n_+$ respectively. The Euclidean norm for a vector $x \in \mathbb{R}^n$ is written $|x|$. The symbol I_n denotes the $n \times n$ identity matrix. For a bounded vector-valued signal $u : \mathbb{Z}_+ \to \mathbb{R}^n$, we denote its ℓ_∞-norm $\|u\|_{[t_0,t_1]} = \sup_{t\in[t_0,t_1]} |u_t|$, and if $t_1 = +\infty$ then we simply write $\|u\|$. We denote by $\mathcal{L}_\infty^n$ the set of all $\mathbb{R}^n$-valued signals u with the property $\|u\| < \infty$. The symbols $\underline{u}$ and $\overline{u}$ denote *component-wise* lower and upper bounds for the vector or signal u. For any signal s, we assume throughout that $s_t = 0$ if $t < 0$ and we define the notation $s_{t_1:t_2} = \begin{bmatrix} s_{t_1}^{\mathrm{T}} & s_{t_1+1}^{\mathrm{T}} & \dots & s_{t_2}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, for any $t_1 \leq t_2$. For two vectors $x_1, x_2 \in \mathbb{R}^n$ or matrices $A_1, A_2 \in \mathbb{R}^{n\times n}$, the relations $x_1 \leq x_2$ and $A_1 \leq A_2$ are understood componentwise. A matrix $A \in \mathbb{R}^{n\times n}$ is called Schur stable if all its eigenvalues have absolute value less than one. It is called nonnegative if all its elements are nonnegative, i.e., if $A \geq 0$. The set of diagonal matrices of dimension $n \times n$ with diagonal positive elements is denoted as $\mathbb{D}_{>0}^n$. Given a matrix $A \in \mathbb{R}^{m\times n}$, we define $A^+ = \max\{0, A\}$ applied elementwise, $A^- = A^+ - A$ and denote the matrix of absolute values of all elements by $|A| = A^+ + A^-$. In a symmetric block matrix, the symbol $*$ denotes a term that can be deduced by symmetry.

1.2 Research Objectives

The first objective of this thesis consists in designing novel privacy-preserving algorithms for CPS, in which data streams are shared by individuals to optimize monitoring and control performance. Consider for example a scenario in which fine-grained measurements of individual houses electric power consumption are collected by a smart meter to enable demand-response schemes. By using Kirchhoff's voltage law and Kirchhoff's current law, one can remark in [15] that dynamics of currents and voltages of individual distributed energy resources such as solar cells and photovoltaics can be expressed as uncertain linear systems (linear Gaussian systems). Consider another scenario in which Public Health Services (PHS) must publish for a population infected by a disease the number of infectious people, i.e., those who have the disease and are able to infect others. PHS use privacy-sensitive data collected from hospitals, with each hospital i recording the number of infectious people in its area, as well as the number of recovered people, i.e, those who were infected by the disease and are now immune. For each area i, it has been proved in [16] that these numbers follow a linear Gaussian SEIR epidemiological model for the specific disease.

The above examples motivate the purpose of the first part of this thesis, which consists in developing privacy-preserving observers for CPS, whose dynamics can be modeled as uncertain linear systems. The first problem of privacy-preserving observer design that we consider (Chapter 2) involves a multi-agent system composed of n linear Gaussian individual systems whose uncertainties have known statistical properties. First, we assume that a data aggregator aims to release publicly an estimate of a linear combination of the individual states, computed from the individual measured signals. For privacy reasons, the individual measured signals are not released by the data aggregator. Moreover the publicly released estimate should also formally preserve the privacy of these measured signals. Second, motivated by the aforementioned example of control of the distribution of vehicles on a road network, we consider a situation where the data aggregator uses the individual measured signals to compute and broadcast a (causal) control signal that minimizes a given quadratic cost for the n agents. Only the control signal is published (in particular, it is available to the n agents), and releasing it must still guarantee the privacy of the measured signals. A related problem of privacy-preserving observer design is formulated in chapter 3. Again, we aim to publicly release an estimate of a linear combination of n individual linear system states, however the statistical properties of the uncertainties arising in the system models are now unknown, and only bounds are given on these uncertainties. For example, epidemiological models for which only bounds of uncertainties are known are discussed in [17–19].

For these problems, this thesis aims to answer the following questions:

- What architecture can be constructed to preserve the privacy of measured signals when publishing an aggregate estimate of a combination of individual system states?
- What architecture can be built to preserve privacy for measured signals when publishing a control signal that minimizes a given quadratic cost function of the state of a linear Gaussian system?

In Chapter 4, a problem of secure observer design is formulated by considering an uncertain LTI system subject to unknown but bounded disturbances and noise, which moreover can be attacked by an adversary capable of adding malicious signals to the actuator and sensor signals. The adversary aims to remain undetected by a specifically designed monitor. For this problem, our work aims to answer the following questions:

- How can an adversary design attack signals that are undetected by the monitor?
- How can we design attack-resilient estimators of the state of a system under attack, assuming some information about the location of such attacks?
- How can a monitor evaluate the vulnerability of a system to undetected attacks?

1.3 Methodology and Research Related to Privacy-Preserving Observer Design

1.3.1 The Right to Privacy

Despite the fact that the "*right to privacy*" is now recognized as fundamental [20, 21], its introduction dates back only of the end of the 19th century [22]. The "*right to privacy*" is part of the right to life [20], one of the four fundamental legal rights recognized by Canada's Constitution. Article 12 of the Universal Declaration of Human Rights (1948) is also devoted to privacy. Key moments that have shaped privacy laws are presented in [23]. Title 13 Chapter 1.1 of the U.S. Code states that no individual should be re-identifiable in any data that is publicly released. Further information on privacy laws can be found in [20, 24, 25].

Nowadays, protecting the privacy of individual users has been recognized as a central issue for emerging large-scale infrastructure systems such as as smart grids or intelligent transportation systems. Indeed, these systems require individual users to continuously send privacy-sensitive information to external data aggregators performing monitoring or control tasks. For example, real-time traffic maps are useful for drivers to plan optimal routes and for city transportation administrators to plan for future capacity or detect and respond to accidents. However, they require real-time location data from individuals, which turns out to be very difficult to properly anonymize, because each person tends to have a highly unique mobility pattern. In general, privacy considerations can make people reluctant to share the information necessary to improve the performance of these systems, thereby reducing their impact.

1.3.2 Privacy Models

Dalenius was one of the first to propose in [26] a formal definition of privacy. According to him, by releasing statistics computed from a private database we should not contribute to increasing the knowledge about any specific individual's confidential information. Dwork [27] mentions that this point of view on privacy is too strict, namely, Dalenius' goal is not achievable when background information is present. Privacy criteria that are used in real life offer only partial disclosure limitation guarantees. Note that data privacy is different from data security. Data breaches that result from leaks, hacks and stolen data are considered as security failures and not as data privacy failures. Privacy-preserving data analysis is concerned with situations where one aims to purposefully release some aggregate information about a dataset, while controlling certain aspects of this data release that relate to personal information.

An Example of Privacy Failure in Data Release

There have been some examples of failures to protect privacy in recent times, such as the case of the Massachusetts Health Records in the 1990s [28]. In the state of Massachusetts, USA, state employees' health insurance is purchased by the Group Insurance Commission (GIC). In the mid 1990s, the GIC had gathered the medical records of $135,000$ employees and their families. It was decided to share part of this data with researchers. The governor of Massachusetts William Weld claimed that the privacy of individuals was protected because personally identifiable information such as names or addresses were removed from the database [29]. Nevertheless, Latanya Sweeney, then a graduate student at MIT, showed that those claims are not true. For 20 dollars, she purchased Cambridge's voter registration list. Then, she crossed this information with three attributes also provided by the GIC dataset, namely, ZIP code, gender and date of birth, and was in fact able to re-identify the medical record of the governor William Weld himself. She also showed that 87 per cent of the US population can be uniquely identified from these three pieces of information [30].

Preserving individual privacy implies some loss on the utility of the data that is protected, in comparison with the original data. Consequently, the guarantees that are offered by each privacy technique have to be limited in order to keep the data useful for further analysis. One can classify privacy methods in two broad categories: *non perturbative* methods and *perturbative* methods. Next, we provide examples for each class of methods.

Non-perturbative methods [31]

Non-perturbative methods do not change values of the data, they limit the amount of data that is released publicly. Some primary non-perturbative methods are *cell suppression, recoding, sampling, query restriction.* Cell suppression consists of suppressing individual attributes or data points' combination of attributes to preserve privacy. Recoding consists of merging groups of data points together. Sampling consists of releasing publicly a randomly chosen subset of the dataset instead of the entire dataset. In scenarios in which a database is sent to a data recipient via a querying mechanism, query restriction consists of suppressing some queries whose answers are suspected to breach a given privacy protection requirement or placing a limit on the number of queries that are answered.

Perturbative methods [31]

Perturbative methods change values of data points' attributes. Some perturbative methods are *noise addition, rounding, re-sampling, the Post-Randomization Method (PRAM) and swapping.* Noise addition technique can be applied generally only to continuous numerical data such as ages, salaries, etc., although discrete distributions have been proposed to add noise to discrete numerical data such as integers. Re-sampling consists in generating a synthetic dataset, sampled from a distribution generated from the original database. The PRAM is a technique that has been inspired by the randomized response technique [32]. The difference is that one can apply randomized response during the interviewing step of a survey, whereas PRAM is used after a survey has been completed and the dataset is formed. After applying PRAM on a dataset, the scores on certain variables for some records in the original database are modified according to a given probability mechanism. Data swapping consists in swapping individual values in a database.

We now review some privacy models that have been proposed in the last decades. A more exhaustive survey of privacy models can be found in [33].

Ad Hoc Anonymization

To anonymize a dataset, personally identifiable information such as addresses, names, social security numbers and telephone numbers are removed. It has been shown that using these basic anonymization techniques alone leaves systems vulnerable to *linkage* attacks [28] and other third party attacks [34]. An illustration of how to re-identify anonymized data is the approach used by Sweeney to deduce the medical record of governor W. Weld in the GIC dataset. Another vulnerability of anonymized datasets has been shown by Sweeney in [35], where she tested the vulnerability of Washington State's hospital data to re-identification. She obtained a correct matching of 43 percent (35 of 81 individuals identified) in Washington State's inpatient data by combining the anonymized discharge data released by the states local newspaper stories with anonymized hospital visits.

$k-$anonymity

Following her re-identification of the GIC database, Sweeney proposed a novel suppressive privacy model: *$k-$anonymity.* $k-$anonymity [28, 36, 37] is based on the anonymity principle and aims at hiding individual data within groups of at least $(k-1)$ other indistinguishable records. This privacy model seems sound: it would not have been possible to perform the attack that Sweeney performed on the GIC database if $k-$anonymity had been used. However, this privacy model is also vulnerable to some types of attacks. A trivial example is a scenario in which a group of k records with the same quasi-identifiers [1] share the same sensitive attribute value: re-identification can be performed. Recently in 2018, Sweeney re-identified individual records in a $k-$anonymous dataset [38]. There are other privacy models that mix both anonymity and secrecy. For example $l-$diversity [39] and $t-$closeness [40], similarly to $k-$anonymity, aim to hide each individual among a group of individuals, but, unlike $k-$anonymity, also require the confidential information of the individuals in the group to be sufficiently diverse to improve secrecy. Additional background on these privacy models

[1] a quasi-identifier is a set of attributes that can be combined with external information to identify a specific individual

can be found in [41].

Various other definitions of privacy have been proposed that are amenable to formal analysis. While a survey of such definitions is out of the scope of this subsection, we can mention some recent work focusing on signal processing and control problems. Privacy is measured by a *lower bound* on the mutual information between published and private signals in [42], on the Fisher information in [43], or on the error covariance of the estimator of a sensitive signal in [44,45]. The concept of k-anonymity and its extensions has been applied to the publication of location traces in [46]. But much of the recent research on privacy-preserving data analysis relies on the notion of *differential privacy* [47–49]. Next, we present this privacy model.

Differential Privacy

In the standard set-up for differential privacy, which is also the situation considered in this thesis, a data holder aims to release the results of computations based on private data. Differential privacy is enforced by adding an appropriate amount of noise to the published results, in such a way that the probability distribution over the outputs does not depend too much on the data of any single individual. As a result, the ability of a third party observing the outputs to make new inferences about a given person is roughly the same, whether or not that person chooses to contribute their data. This guarantee can then be used to weigh the risks of information disclosure against the benefits of publishing more accurate analyses. Differential privacy provides a disclosure limitation guarantee that is similar to that of Dalenius (see Section 1.3.2). The difference between them is that while Dalenius compared the knowledge before and after accessing the released data, differential privacy compares the knowledge before and after a single individual contributes his or her data. Instead of limiting the knowledge provided by the dataset, differential privacy limits the knowledge provided by each individual in the dataset [21].

A large number of techniques have been developed to compute differentially private versions of various statistics, see [49] for an overview. Nevertheless, the differentially private analysis of *streaming* data remains relatively less explored [42,50,51], despite its importance for signal processing and control applications. Some previous work has focused on the design of differentially private dynamic estimators [52–55], controllers [56,57], consensus algorithms [58,59], or anomaly detectors [60–62]. In particular, [52] discusses the Kalman filtering problem under a differential privacy constraint and compares schemes introducing noise either directly on the measured signals (input perturbation mechanisms) or on the published estimate (output perturbation mechanisms). Output perturbation provides better performance as the number of input signals increases, but has the drawback of leaving unfiltered noise on the output.

In [56], the authors consider a multi-agent linear quadratic tracking problem where the trajectory tracked by each agent should remain private, while [57] considers an LQG control problem where each agent wishes to keep its individual state private. In both cases, noise is added directly on the individual measurements, a form of input perturbation.

Differential privacy is a property satisfied by certain randomized algorithms (also called mechanisms), which in an abstract setting compute outputs in a space R based on sensitive data in a space D. To be differentially private, the algorithm must ensure that the probability distribution of its randomized output is not very sensitive to certain variations in the input data, which are specified as part of the privacy requirement. More concretely, we equip the input space D with a symmetric binary relation called adjacency and denoted Adj, which captures the variations in the input datasets that we want to make hard to detect by observing the outputs. A mechanism is a stochastic system producing an output based on its input y. For any two inputs that satisfy the adjacency relation, the following definition characterizes the deviation that is allowed for a differentially private mechanism's output distribution.

Definition 1 *Let D be a space equipped with a symmetric binary relation denoted Adj and let $(\mathsf{R}, \mathcal{R})$ be a measurable space, where $\mathcal{R}$ is a given σ-algebra over R. Let $\epsilon \geq 0$, $1 \geq \delta \geq 0$. A randomized mechanism M from D to R is (ϵ, δ)-differentially private (for Adj) if for all $d, d' \in \mathsf{D}$ such that Adj(d, d'),*

$$\mathbb{P}(M(d) \in S) \leq e^{\epsilon}\, \mathbb{P}(M(d') \in S) + \delta, \; \forall S \in \mathcal{R}. \tag{1.1}$$

In Definition 1, smaller values of ϵ and δ correspond to stronger privacy guarantees, i.e., distributions for $M(d)$ and $M(d')$ that are closer in (1.1). (ϵ, δ)-differential privacy ensures that for every pair of adjacent databases d, d', retroactively, the observed value $M(d)$ will be much more or much less likely to be produced when the database is d than when it is d' [49]. Nevertheless, given an output $\Xi \sim M(d)$, there may exist a database d' such that Ξ is much more likely to be generated on d' than it is when the database is d. Consequently, the mass of Ξ in the distribution $M(d')$ can be larger than its mass in the distribution $M(d)$. In contrast, $(\epsilon, 0)$-differential privacy says that, for every run of the mechanism $M(d)$, the observed output is (almost) equally likely to be observed on every adjacent database, simultaneously. Moreover, define the *privacy loss* induced by observing Ξ as follows

$$P_{\mathrm{Adj}(d,d')}(\Xi) = \ln\left(\frac{\mathbb{P}(M(d) = \Xi)}{\mathbb{P}(M(d') = \Xi)}\right).$$

(ϵ, δ)-differential privacy guarantees that for all adjacent d, d', the absolute value of $P_{\mathrm{Adj}(d,d')}(\Xi)$

is upper bounded by ϵ with a probability of at least $1-\delta$ [49, Lemma 3.17].

A fundamental property of differential privacy is that manipulating an output that is already differentially private does not imply any additional privacy loss if the original dataset is not re-accessed (resilience to postprocessing property [49]). For the next theorem, define two measurable spaces $(\mathsf{R}_1, \mathcal{R}_1)$ and $(\mathsf{R}_2, \mathcal{R}_2)$.

Theorem 1 [52, Theorem 1] *Let $\epsilon, \delta \geq 0$. Consider an (ϵ, δ)-differentially private mechanism M_1 from D to $(\mathsf{R}_1, \mathcal{R}_1)$. Consider another mechanism M_2 from D to $(\mathsf{R}_2, \mathcal{R}_2)$, such that there exists a probability kernel k from $\mathsf{R}_1 \times \mathcal{R}_2$ to $[0,1]$ that satisfies, for all $S \in \mathcal{R}_2$ and $d \in \mathsf{D}$*

$$\mathbb{P}(M_2(d) \in S | (M_1(d)) = k(M_1(d), S), \text{ a.s.} \tag{1.2}$$

Then M_2 is (ϵ, δ)-differentially private.

The equality (1.2) means that once $M_1(d)$ is known, the distribution of $M_2(d)$ is not a function of d anymore. In other words, Theorem 1 is equivalent to saying that a mechanism M_2 that has access to database only via the output of a differentially private mechanism M_1 is not able to weaken the privacy guarantee.

Next, we need tools that can be used to enforce the property of Definition 1. The following definition is useful for mechanisms such as ours that produce outputs in vector spaces.

Definition 2 *Let D be a space equipped with an adjacency relation Adj. Let R be a vector space equipped with a norm $\| \cdot \|_{\mathsf{R}}$. The sensitivity of a mapping $q : \mathsf{D} \mapsto \mathsf{R}$ is defined as*

$$\triangle q := \sup_{\{d,d' : Adj(d,d')\}} \|q(d) - q(d')\|_{\mathsf{R}}.$$

For R equipped with the ℓ_2-norm, this defines the ℓ_2-sensitivity of q, denoted $\triangle_2 q$.

The *Gaussian mechanism* [63] consists in adding Gaussian noise proportional to the ℓ_2-sensitivity of a mapping to enforce (ϵ, δ)-differential privacy. A fairly tight upper bound on the proportionality constant is provided in [52]. Recall first the definition of the $\mathcal{Q}$-function $\mathcal{Q}(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2}) du$, which is monotonically decreasing from $(-\infty, \infty)$ to $(0, 1)$. Then, for $1 > \delta > 0$, define $\kappa_{\delta,\epsilon} := \frac{1}{2\epsilon}\left(Q^{-1}(\delta) + \sqrt{(Q^{-1}(\delta))^2 + 2\epsilon}\right)$. The following theorem can be found in [52].

Theorem 2 *Let $\epsilon > 0$, $1 > \delta > 0$. Let G be a dynamic system with p inputs and q outputs. Then the mechanism $M(d) = Gd + \nu$, where ν is a white Gaussian noise (sequence of iid zero-mean Gaussian vectors) with $\nu_t \sim \mathcal{N}(0, \kappa^2_{\delta,\epsilon}(\Delta_2 G)^2 I_q)$, is (ϵ, δ)-differentially private.*

In other words, Theorem 2 says that we can produce a differentially private signal by adding white Gaussian noise at the output of a system G processing the sensitive signal d, with covariance matrix $\sigma^2 I_q$, and σ proportional to the ℓ_2-sensitivity of G. A first solution is the *input perturbation mechanism*: it consists in perturbing each individual signal directly to release differentially private versions of these signals. By using the resilience to post-processing property of differential privacy, the data aggregator can apply further computations to such an output that is differentially private without degrading the differential privacy guarantee, as long as the original dataset is not re-accessed for these computations. Nevertheless, we show in Chapter 2 that input perturbation typically leads to a high level of noise and hence performance degradation, which motivates the search for better mechanisms.

Another solution, called the *Laplace Mechanism* [47], consists in adding iid zero-mean noise whose distributed as a Laplace distribution. Denote by $\text{Lap}(b)$, the Laplace distribution with scale parameter b and mean zero. Its variance is $2b^2$ and its density is $p(x; b) = \frac{1}{2b} e^{-\frac{|x|}{b}}$.

Theorem 3 [47] *Let $\epsilon > 0$. Consider a query $q : \mathsf{D} \to \mathbb{R}$. Then the mechanism $M(d) = q(d) + \nu$, with $\nu \sim \text{Lap}(b)$ and $b \geq \frac{\Delta_1 q}{\epsilon}$, is ϵ-differentially private, where $\Delta_1 q$ is the ℓ_1-sensitivity of q.*

Kalman filter

Consider the following linear system

$$\begin{aligned} x_{t+1} &= F_t\, x_{i,t} + B_t\, u_t + w_t, \\ y_t &= H_t\, x_t + v_t, \end{aligned} \tag{1.3}$$

where $x_t \in \mathbb{R}^n$ is the state, $y_t \in \mathbb{R}^p$ is the output, $w_t \sim \mathcal{N}(0, W_t)$ and $v_t \sim \mathcal{N}(0, V_t)$ represent independent sequences of iid zero-mean Gaussian random vectors with covariance matrices $W_t \succ 0, V_t \succ 0$. The sequence u with $u_t \in \mathbb{R}^h$ is the control input. The initial condition x_0 is an independent Gaussian random vector that is also independent of the noise processes w and v, with mean $\overline{x}_0$ and covariance matrix $\Sigma_0^- \succ 0$.

To estimate the value of the state x_t at each period $t \geq 0$ from the observed outputs $y_{t-1}, y_{t-2}, \ldots, y_0$, a Kalman filter is optimal in the minimum mean square error (MMSE) sense. Denote by $\hat{x}_t^- = \mathbb{E}[x_t | y_{0:t-1}]$ and $\hat{x}_t = \mathbb{E}[x_t | y_{0:t}]$, the state estimates provided by the

Kalman filter after the prediction step and the measurement update step respectively [64]. Denote by $\bar{\Sigma}_t = \mathbb{E}[(x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)^T | y_{0:t-1}]$ and $\Sigma_t = \mathbb{E}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T | y_{0:t}]$, the corresponding error covariance matrices. In addition, let $\bar{\Sigma}_0$ be the covariance matrix for the initial state x_0. Equations of the Kalman filter can be written as follows. Given the dynamics (1.3) and the measurement equation, we get for $t \geq 0$ and starting from $\hat{x}_0^- := \bar{x}_0$

$$\begin{aligned} \hat{x}_t &= \hat{x}_t^- + K_t(y_t - H_t \hat{x}_t^-), \\ \hat{x}_{t+1}^- &= F_t \hat{x}_t + B_t u_t, \qquad \text{with} \\ K_t &= \bar{\Sigma}_t F_t^T (F_t \bar{\Sigma}_t F_t^T + V_t)^{-1}. \end{aligned} \tag{1.4}$$

Dynamics of the error covariance matrices evolve for $t \geq 0$ as

$$\begin{aligned} \Sigma_t^{-1} &= \bar{\Sigma}_t^{-1} + H_t^T V_t^{-1} H_t, \\ \bar{\Sigma}_{t+1} &= F_t \Sigma_t F_t^T + W_t. \end{aligned}$$

The Kalman filter has sundry applications in technology. It is commonly used in guidance, navigation, and control of vehicles [65], prediction and estimation of outbreaks of infectious disease [16], state estimation and control in smart grids [66]. This motivates us to consider the problem of Kalman filtering under privacy constraints in Chapter 2.

Furthermore, over the last decade, the notion of differential privacy has been extended to dynamical systems and has been applied to signal filtering [52, 67]. However, these papers assume that the statistical properties of the disturbances in the signal models are available. Furthermore, a differentially private mechanism has been proposed by using *output perturbation* for positive linear systems without disturbance in [68], and for nonlinear systems in [69], both assuming that point-wise estimation (via a Luenberger-type observer design) is possible. Nevertheless, state disturbances for certain systems can be modeled as bounded uncertain signals. Instead of point-wise observers, *interval estimators* [19, 70] that provide estimates of lower and upper bounds of the state can address state estimation problems for such systems. This motivates the design of differentially private observers that handle the presence of disturbances or uncertain parameters whose values are only known to belong to given intervals or polytopes in Chapter 3.

1.4 Methodology and Research Related to Secure Observer Design

1.4.1 Some examples of cyber-attacks in CPS

During the last decades, numerous cyber-attacks have been noticed on safety-critical infrastructures including many power blackouts in Brazil [71] and the SQL Slammer worm attack on the Davis-Besse nuclear plant [72], which illustrates the vulnerabilities of CPS. When analyzing these vulnerabilities, a common approach consists in considering particular attacks against specific systems. For example, [73] defines *denial of service* and *deception* attacks against a networked control system. In addition, [73] proposed SDP-based approaches to counteract *denial of service* attacks, which compromise the availability of certain resources, for example by jamming communication channels. Deception attacks, instead, compromise measurements and control packets by modifying the behavior of actuators and sensors. Furthermore, *stealthy deception attacks* against SCADA systems are discussed in [74]. Stealthy attacks against legacy systems and schemes that can used to counteract them are also considered in [75–77]. *Replay attacks* against control systems are studied in [78, 79]. Adversaries design replay attacks by hijacking the sensors, recording the measurements over an interval of time, and repeating such measurements while adding an exogenous signal into the system. Moreover, *covert attacks* against control systems are discussed in [80]. Specifically, a covert agent uses a decoupling structure to modify the physical plant's behavior while remaining undetected by the original controller. Nevertheless, the aforementioned attacks have been designed only by considering point-wise models.

1.4.2 Attack-resilient estimation and control in CPS

An attack-resilient control problem is discussed in [81]: an adversary corrupts control packets transmitted over a network in order to cause a harm to the system. A stabilizing receding-horizon Stackelberg control law is designed in the presence of the attack. More recently, [82] designed an estimator for the state of a linear system whose measurements are corrupted. They characterize the maximum number of tolerable faulty sensors, and propose a decoding algorithm to detect corrupted measurements. However, only point-wise estimation has been considered when discussing these problems in the aforementioned articles.

Standard fault detection algorithms, while generally useful, are unsuccessful in some cases against the attacks of a smart adversary [83]. Classical bad data detection strategies, such as the largest residue test [84], have been applied extensively to static linear models with Gaussian noise, e.g., in the context of state estimation for power systems. Notwithstanding, an adversary who knows the configuration of a power grid for example is able to carry out a false-data injection attack [85], i.e., inject a stealthy signal into the measurements to compromise the state estimator of the power grid while leaving the residue unchanged [86].

To carry out false-data injection attacks on a dynamical system, an attacker must select attack strategies that are consistent not only with static observations but also with the state dynamics at all times [85]. This type of attack is discussed in [82] for noiseless models. Dynamic false-data injection attacks are studied in [74], by assuming that the statistical properties of the disturbances are available, and [87] designs optimal *stealthy* attack strategies against CPS by making the same assumption. However, these disturbances are commonly modeled as bounded signals for control design [88]. Robust failure detection algorithms are designed so that they can handle disturbance and measurement noise that are *a priori* bounded [89]. This allows an intelligent adversary to hide an attack within these bounds, remaining undetected while causing serious harm. State estimators have been developed for dynamical systems under sensor attacks with bounded noise in [90], under the assumption however that there is no uncertainty on the initial value of the state, which is a drawback in applications where only bounds on this value are known.

An analysis of stealthy attacks for CPS with bounded parameters is performed in [83] using a set-membership estimation approach, which computes the set of states consistent with the model and the measurements [91, 92]. However, this technique may be difficult to apply in practice to design Fault Detection and Isolation (FDI) systems, where simpler observers with tunable gains are more common. On the other hand, *interval observers* [93, 94], which are a subclass of set-membership estimators, need less computational power and have become one of the most common approaches for FDI during the last decade.

1.4.3 Interval estimation

Next, we provide a formal definition of *interval observers*. To this end, we review some basic lemmas from interval estimation theory.

Lemma 1 [95] *Consider the following linear time-invariant (LTI) system*

$$\begin{aligned} x_{t+1} &= Ax_t + B\omega_t,\, \omega : \mathbb{Z}_+ \to \mathbb{R}^q_+, \\ y_t &= Cx_t + D\omega_t, \end{aligned} \tag{1.5}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$ and the matrix $A \in \mathbb{R}^{n\times n}_+$. Any solution of the LTI system (1.5) is elementwise nonnegative for all $t \geq 0$ provided that $x_0 \geq 0$ and $B \in \mathbb{R}^{n\times q}_+$. In addition, the output solution y_t of such a system is nonnegative if $C \in \mathbb{R}^{p\times n}_+$ and $D \in \mathbb{R}^{p\times q}_+$.

A dynamical system that satisfies all these restrictions is called cooperative (monotone) or nonnegative [95].

Lemma 2 [96] *Let $x \in \mathbb{R}^n$ be a vector variable, $\underline{x} \leq x \leq \overline{x}$ for some $\underline{x}, \overline{x} \in \mathbb{R}^n$. If $A \in \mathbb{R}^{m \times n}$ is a constant matrix, then*

$$A^+\underline{x} - A^-\overline{x} \leq Ax \leq A^+\overline{x} - A^-\underline{x}. \tag{1.6}$$

Proof. By definition of A^+ and A^-, we get $Ax = (A^+ - A^-)x$. The inequality $\underline{x} \leq x \leq \overline{x}$ leads to the result (1.6). □

Consider a LTI system

$$\begin{aligned} x_{t+1} &= Ax_t + w_t, \\ y_t &= Cx_t + v_t, \end{aligned} \tag{1.7}$$

where $x_t \in \mathbb{R}^n$ represents the state vector, $y_t \in \mathbb{R}^p$ is the output vector, $w : \mathbb{Z}_+ \to \mathbb{R}^n$ stands for an *unknown* input in $\mathcal{L}_\infty^n$, $v : \mathbb{Z}_+ \to \mathbb{R}^p$ is an *unknown* measurement noise in $\mathcal{L}_\infty^p$, and $A \in \mathbb{R}^{n \times n}, C \in \mathbb{R}^{p \times n}$ are known constant matrices. Assume that the initial condition x_0 is *unknown* but satisfy the bounds $\underline{x}_0 \leq x_0 \leq \overline{x}_0$, where $\underline{x}_0, \overline{x}_0 \in \mathbb{R}^n$ are given. Assume also that two functions $\underline{w}$, $\overline{w} : \mathbb{Z}_+ \to \mathcal{L}_\infty^n$ and two functions $\underline{v}$, $\overline{v} : \mathbb{Z}_+ \to \mathcal{L}_\infty^p$ are given such that $\underline{w}_t \leq w_t \leq \overline{w}_t$ and $\underline{v}_t \leq v_t \leq \overline{v}_t$ for all $t \geq 0$.

Equations of an interval observer can be written as follows

$$\begin{aligned} \underline{x}_{t+1} =& (A - LC)\underline{x}_t + Ly_t + \underline{w}_t - L^+\overline{v}_t + L^-\underline{v}_t, \\ \overline{x}_{t+1} =& (A - LC)\overline{x}_t + Ly_t + \overline{w}_t - L^+\underline{v}_t + L^-\overline{v}_t, \end{aligned} \tag{1.8}$$

where $\underline{x}_t \in \mathbb{R}^n$ and $\overline{x}_t \in \mathbb{R}^n$ stand for the lower and the upper interval estimates of the system state x_t. Denote the estimation errors $\underline{e} = x - \underline{x}$ and $\overline{x} - x$.

Theorem 4 [70] *Consider a matrix $L \in \mathbb{R}^{n \times p}$ such that the matrix $A - LC$ is Schur stable and nonnegative. Then, we get for (1.7)*

$$\underline{x}_t \leq x_t \leq \overline{x}_t, \forall t \geq 0. \tag{1.9}$$

Furthermore, the estimation errors $\underline{e}, \overline{e} \in \mathcal{L}_\infty^n$.

Proof. The errors' dynamics can be written as follows

$$\begin{aligned}\underline{e}_{t+1} =&(A-LC)\underline{e}_t + \sum_{i=1}^{i=2} \underline{g}_i,\\ \overline{e}_{t+1} =&(A-LC)\overline{e}_t + \sum_{i=1}^{i=2} \overline{g}_i,\end{aligned} \tag{1.10}$$

where

$$\begin{aligned}\underline{g}_1 =&w_t - \underline{w}_t,\\ \overline{g}_1 =&\overline{w}_t - w_t,\\ \underline{g}_2 =&L^+\overline{v}_t - L^-\underline{v}_t - Lv_t,\\ \overline{g}_2 =&Lv_t - (L^+\underline{v}_t - L^-\overline{v}_t).\end{aligned}$$

We deduce by using Lemma 2 that the signals $\{\underline{g}_i, \overline{g}_i, 1 \leq i \leq 2\}$ are nonnegative. Therefore, when $(A-LC)$ is nonnegative, by applying Lemma 1, we get $\underline{e}_t \geq 0, \overline{e}_t \geq 0$ since the system (1.10) is cooperative ($\underline{e}_0 \geq 0$ and $\overline{e}_0 \geq 0$). Hence, the order relation $\underline{x}_t \leq x_t \leq \overline{x}_t$ holds for all $t \geq 0$. Since the system (1.10) is linear, the GAS property of the system (3.22) for $\{\underline{g}_i\}_{i=1}^{i=2} \equiv 0, \{\overline{g}_i\}_{i=1}^{i=2} \equiv 0$ implies its ISS [97]. It can be inferred that $\underline{e}, \overline{e} \in \mathcal{L}_\infty^n$. □

The third part of this thesis focuses on control-theoretic approaches to CPS security. Namely, we consider security issues for uncertain linear systems for which only lower and upper bounds are known for uncertainties.

1.5 Main contributions and structure of the thesis

We divide the main contributions of this thesis into three groups. The first one is related to the problem of privacy-preserving observer design in a multi-agent system composed of independent linear Gaussian individual systems whose uncertainties have known statistical properties. The second group of contributions is related to the design of a privacy-preserving observer when the uncertainties affecting the individual systems are not available and when these individual systems are interconnected. The third group of contributions of this thesis is related to the problem of designing undetectable attacks and attack-resilient observers for CPS with unknown and bounded uncertainties using interval estimation approaches.

1.5.1 First Group of Contributions

The first group of contributions of this thesis concerns the design of a two-stage architecture for differentially private Kalman filtering, where the privacy-preserving noise is added only after an input stage appropriately combining the independent measured signals of the individual agents, while an output stage filters out this noise. Inspired by [98], such two-stage architectures were discussed in [52] but have not yet been applied to the Kalman filtering problem. We show that the optimal input stage can be computed by solving a semidefinite program (SDP), hence, a tractable convex optimization problem. The fact that the input stage design problem admits an SDP formulation is reminiscent of other Kalman filtering problems subject to resource or communication constraints, see, e.g., [99–101], but the SDP capturing specifically the differential privacy constraint is new. The design procedure is then adapted in Section 2.3 of Chapter 2 to the LQG control problem. By exploiting the classical properties of the optimal LQG controller (linearity and separation principle), we can view the control problem as the problem of estimating a certain linear combination of the agent states as in Section 2.2 of Chapter 2, but for a specific cost on the estimation error.

1.5.2 Second Group of Contributions

The second group of contributions of this thesis concerns the design of privacy-preserving interval estimators for multi-agents systems in which the signals of individual participants are interconnected and are modeled using uncertain linear time-invariant systems with bounded disturbances. We consider uncertain initial conditions as well as uncertain time-varying inputs and outputs. Extending the differentially private mechanism with bounded noise of [102] for the publication of a single scalar value to the publication of vectors and signals, we obtain in Chapter 3 an input perturbation mechanism where privacy-preserving noise is added to each individual's data before sending it to an interval observer. Moreover, our estimator handles multi-agent systems in which the dynamics of the agents are coupled, in contrast to [52] that has considered only independent dynamics.

1.5.3 Third Group of Contributions

Chapter 4 of this thesis focuses on stealthy attacks in CPS with unknown but bounded disturbances. We design stealthy attacks on CPS in the presence of unknown but bounded uncertainty on initial conditions, on the dynamics and on the measurements. We use interval estimation methods and show that these attacks can be computed by repeatedly solving linear programs. Furthermore, we construct interval observers that are resilient to stealthy attacks. Using semi-definite programming, we compute efficiently the observer gains that minimize the estimation errors. The construction is based on the method proposed in [103], extended in this thesis to LTI discrete-time systems with bounded measurement noise and disturbances. The required detectability condition is easier to satisfy for our interval observer than the condition in [104]. In addition, we compute bounds on stealthy attacks, which can be useful in practice to assess the vulnerability of a system.

CHAPTER 2 DIFFERENTIALLY PRIVATE KALMAN FILTERING AND LQG CONTROL WITH SIGNAL AGGREGATION

In this chapter, we discuss the problem of privacy-preserving observer design and control in a multi-agent system composed of linear Gaussian individual systems. We formulate the problem of privacy-preserving state estimation and LQG Control for a population of dynamic agents in Section 2.1. Then, we argue in Section 2.2.1 via a simple example that significant performance improvements can be expected compared to input perturbation mechanisms (see Subsection 1.3.2). We compute the optimal input stage by solving an SDP, hence, a tractable convex optimization problem. Then, we adapt the design procedure in Section 2.3 to the LQG control problem. By exploiting the classical properties of the optimal LQG controller (linearity and separation principle), we prove that we can view the control problem as the problem of estimating a certain linear combination of the agent states as in Section 2.2, but for a specific cost on the estimation error.

2.1 Problem Statement

2.1.1 Privacy-Preserving State Estimation and LQG Control for a Population of Dynamic Agents

Consider a set of n privacy-sensitive signals $\{y_{i,t}\}_{0 \leq t \leq T}$, $i = 1, \dots, n$, with $y_{i,t} \in \mathbb{R}^{p_i}$, collected by a data aggregator, and which could originate from n distinct agents. Let $p = \sum_{i=1}^{n} p_i$. We assume that a mathematical model capturing known dynamic and statistical properties of these signals is publicly available, consisting of a linear system with n independent (vector-valued) states associated to the n measured signals

$$\begin{aligned} x_{i,t+1} &= A_{i,t}\, x_{i,t} + B_{i,t}\, u_t + w_{i,t}, \quad 0 \leq t \leq T-1, \\ y_{i,t} &= C_{i,t}\, x_{i,t} + v_{i,t}, \quad 0 \leq t \leq T, \end{aligned} \tag{2.1}$$

for $i = 1, \dots, n$, where $x_{i,t}, w_{i,t} \in \mathbb{R}^{m_i}$. Here $w_{i,t} \sim \mathcal{N}(0, W_{i,t})$ and $v_{i,t} \sim \mathcal{N}(0, V_i)$ are independent sequences of iid zero-mean Gaussian random vectors with covariance matrices $W_{i,t} \succ 0, V_i \succ 0$, for $i = 1, \dots, n$. In particular, assuming that the matrices $W_{i,t}$ are invertible is necessary in the following to be able to use the "information filter" form of the Kalman filter equations [64]. The sequence u with $u_t \in \mathbb{R}^h$ represents a control input that is shared by the n individuals. This is motivated by scenarios in which a common signal is broadcast to drive the aggregate state of a population, while individual signals can still be subject to privacy constraints. The initial conditions $x_{i,0}$ are independent Gaussian random vectors that are also independent of the noise processes w and v, with mean $\overline{x}_{i,0}$ and covariance matrices $\Sigma^-_{i,0} \succ 0$. Let $x_t := [x^T_{1,t}, \dots, x^T_{n,t}]^T$, $y_t := [y^T_{1,t}, \dots, y^T_{n,t}]^T$, $w_t = [w^T_{1,t}, \dots, w^T_{n,t}]^T$ and $v_t = [v^T_{1,t}, \dots, v^T_{n,t}]^T$ denote the global state, measurement and noise signals of (2.1). Define $A_t := \text{diag}(A_{1,t}, \dots, A_{n,t})$, $B_t = [B^T_{1,t}, \dots, B^T_{n,t}]^T$, $C_t = \text{diag}(C_{1,t}, \dots, C_{n,t})$, $W_t = \text{diag}(W_{1,t}, \dots, W_{n,t})$, $V = \text{diag}(V_1, \dots, V_n)$. Then the system (2.1) can be rewritten more compactly as

$$x_{t+1} = A_t\, x_t + B_t\, u_t + w_t, \quad 0 \leq t \leq T-1, \tag{2.2}$$

$$y_t = C_t\, x_t + v_t, \quad 0 \leq t \leq T, \tag{2.3}$$

with $w_t \sim \mathcal{N}(0, W_t)$ and $v_t \sim \mathcal{N}(0, V)$. Throughout the chapter the model parameters $\overline{x}_{i,0}$, $\Sigma^-_{i,0}$, $A_{i,t}$, $B_{i,t}$, $C_{i,t}$, $W_{i,t}$, V_i, are assumed to be publicly known information.

In Section 2.2, we first consider a filtering problem (with u a known signal) where the data aggregator aims to publish at each period t a causal estimate $\hat{z}_t$ of a linear combination $z_t = L_t x_t = \sum_{i=1}^n L_{i,t} x_{i,t}$ of the individual states, computed from the signals y_i, with $L_t := \left[L_{i,t}, \dots, L_{n,t}\right]$ some given (publicly known) matrices. This estimator should minimize the Mean Square Error (MSE) performance measure

$$E_T := \frac{1}{T+1} \sum_{t=0}^{T} \mathbb{E}\left[\|z_t - \hat{z}_t\|_2^2\right]. \tag{2.4}$$

For privacy reasons, the signals y_i are not released by the data aggregator and moreover the publicly released estimate $\hat{z}$ should also guarantee the differential privacy of the input signals y_i (see Definition 1). For example, the signals y_i could represent position measurements of n individuals, each state x_i could consist of the position and velocity of individual i, and the goal might be to publish only a real-time estimate of the average velocity of all individuals. Note that *in the absence of privacy constraint*, the optimal estimator is $\hat{z}_t = \sum_{i=1}^n L_{i,t} \hat{x}_{i,t}$, with $\hat{x}_{i,t}$ provided by the (time-varying) Kalman filter estimating the state x_i of subsystem i from the signal y_i [64], and in particular the estimation problem then decouples for the n subsystems.

Next, in Section 2.3 we build on the results obtained for the filtering problem to study the following privacy-constrained LQG regulation problem. The data aggregator uses the measured signals $y_i, 1 \leq i \leq n$, to compute and broadcast a (causal) control signal u that minimizes the following quadratic cost for the n agents

$$J_T = \frac{1}{T+1}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_t^T Q_t x_t + u_t^T R_t u_t\right) + x_T^T Q_T x_T\right], \tag{2.5}$$

where $Q_t \succeq 0$ for $0 \leq t \leq T$ and $R_t \succ 0$ for $0 \leq t \leq T-1$ are publicly known weight matrices. Again, only the signal u is published (in particular, it is available to the n agents), and releasing u must guarantee the differential privacy of the measured signals y_i. It is worth noting that the cost function (2.5) can be used to drive an aggregate value of the global population state toward 0 rather than the individual agent states, since trying to do the latter might be in direct conflict with the privacy requirement (which, essentially, aims to hide the value of the individual signals y_i, and hence indirectly of the individual states x_i). For example, we can have $x_t Q_t x_t = \left(\frac{1}{n}\sum_{i=1}^n x_{i,t}\right)^2$ if $Q_t = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n^T$, in order to regulate the average population state. Figure 2.1 represents the estimation and control problem setups, including a basic scheme to enforce differential privacy by using the input perturbation mechanism (described in Subsection 1.3.2). The thicker lines correspond to published signals (estimate $\hat{z}$ or broadcasted control input u) and dashed lines represent the injection of privacy-preserving noise by the input perturbation mechanism.

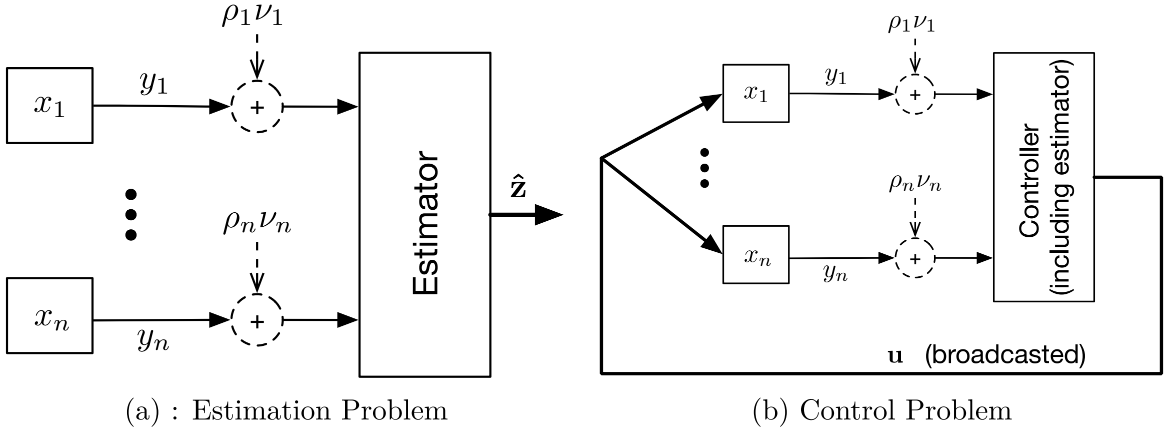
Note that the steady-state versions of both the filtering and LQG control problems are also considered, with the model (2.2)-(2.3) in this case assumed time-invariant and the performance measures defined as

$$E_\infty := \lim_{T\to\infty} E_T, \quad J_\infty := \lim_{T\to\infty} J_T. \tag{2.6}$$

To finish stating the above problems formally, we define next the differential privacy constraint imposed on the signal $\hat{z}$ for the filtering problem and on u for the LQG problem. In this chapter, the input space D (see Definition 1) is the vector space $\mathbb{R}^{p(T+1)}$ of global measurement signals y, and a mechanism is a causal stochastic system producing an output signal ($\hat{z}$ or u in the previous section) based on its input y. We consider two measured signals to be adjacent if the following condition holds

$$\begin{aligned}\text{Adj}(y,y') \text{ iff for some } 1 \leq i \leq n, \; \|y_i - y_i'\|_2 \leq \rho_i,\\ \text{and } y_j = y_j' \text{ for all } j \neq i,\end{aligned} \tag{2.7}$$

with $\{\rho_i\}_{i=1}^n \in \mathbb{R}_+^n$ a given set of positive numbers and with the definition of the ℓ_2-norm $\|v\|_2 := \left(\sum_{t=0}^T |v_t|_2^2\right)^{1/2}$ for a vector-valued signal v. In other words, two adjacent measurement signals can differ by the values of a single participant, with only ℓ_2-bounded signal



(a) : Estimation Problem (b) Control Problem

Figure 2.1 : Estimation and control problem setup. Signals $\{y_i\}_{1\leq i\leq n}$ produced by n independent agents with states $\{x_i\}_{1\leq i\leq n}$ are collected for monitoring or control purposes

deviations (with predefined bounds) allowed for each individual.

2.1.2 A First Solution: Input Perturbation Mechanism

As explained in Subsection 1.3.2, we can use the resilience to post-processing property to design an input perturbation mechanism, which consists in perturbing each measured signal y_i directly to release differentially private versions of these signals.

First, note that the memoryless system defined by $(Gy)_t = My_t$, where M is the diagonal matrix $M = \text{diag}(I_{p_1}/\rho_1, \ldots, I_{p_n}/\rho_n)$, has the sensibility bound

$$\Delta G := \sup_{\text{Adj(y,y')}} \|My - My'\|_2 \leq 1$$

for the adjacency relation (2.7). Hence, by Theorem 2, releasing the signals $\{y_i/\rho_i + \nu_i\}_{1\leq i\leq n}$, where each signal ν_i is a white Gaussian noise with covariance matrix $\kappa^2_{\delta,\epsilon} I_{p_i}$, is (ϵ,δ)-differentially private. Equivalently, using the resilience to post-processing to multiply this output by M^{-1}, we see that releasing the signals $\{\tilde{y}_i := y_i + \rho_i\nu_i\}_{1\leq i\leq n}$ is (ϵ,δ)-differentially private. Once these signals are released, applying further processing on them does not impact the differential privacy guarantee. Moreover, these signals are of the same form as the outputs of system (2.1), except for a higher level of (still Gaussian) noise due to the addition of the artificial privacy-preserving noise. One can therefore produce an (ϵ,δ)-differentially private estimate $\hat{z}$ or control signal u discussed in Section 2.1.1 by applying standard Kalman filtering and LQG design techniques to the signals $\tilde{y}_i$, see Figure 2.1. An advantage of input perturbation mechanisms is that each agent can release directly the differentially private signal $\tilde{y}_i$, and hence does not need to trust the data aggregator to enforce the differential privacy property. Moreover, this mechanism has the potentially useful feature of publishing the individual signals $\tilde{y}_i$, which could be used for other purposes than the original estimation or control problem. Nonetheless, as we discuss in the following sections, input perturbation typically leads to a high level of noise and hence performance degradation, which motivates the search for better mechanisms.

2.2 Differentially Private Kalman Filtering

Input perturbation for the differentially private Kalman filtering problem, as discussed in Section 2.1.2 and represented on Figure 2.1a, was considered in [105]. Here, we show first in Section 2.2.1 via a simple example that the performance of this mechanism can be significantly improved by combining the individual input signals before adding the privacy preserving noise. This leads to the two-stage mechanism of Figure 2.2, whose systematic design is discussed in Section 2.2.2.

2.2.1 A Scalar Example

Consider a scalar homogeneous version of model (2.1) with $A_{i,t} = a \in \mathbb{R}$, $B_{i,t} = b$, $C_{i,t} = c$, $W_{i,t} = \sigma_w^2$, $V_i = \sigma_v^2$ and $L_{i,t} = 1$ (so $z_t = \sum_{i=1}^n x_{i,t}$), and assume $\rho_i = \rho$ for all i in (2.7). In other words, $A_t = aI_n$, $B_t = b\mathbf{1}_n$, $C_t = cI_n$, $W_t = \sigma_w^2 I_n$, $V = \sigma_v^2 I_n$ in (2.2). We also let $T \to \infty$ in the problem statement and consider the steady-state MSE E_∞, see (2.6), as performance measure for a given estimator $\hat{z}$ of z. Moreover in this section we consider for simplicity the minimum mean square error (MMSE) estimate $\hat{z}_t$ of x_t given the measurements up to time $t-1$ only, since the corresponding MSE is directly obtained by solving an algebraic Riccati equation (ARE).

Let $\alpha := \kappa_{\delta,\epsilon}\,\rho$. Since $\sup_{y,y':\mathrm{Adj}(y,y')} \|y - y'\|_2 = \rho$, by Theorem 2 and as explained in Section 2.1.2, releasing the signal $r_t = y_t + \nu_t$, with white noise ν such that $\nu_t \sim \mathcal{N}(0, \alpha^2 I_n)$, is (ϵ, δ)-differentially private for the adjacency relation (2.7). The steady-state MSE of a Kalman filter estimating z for the system with dynamics as in (2.2) and measurements r is obtained by solving a scalar ARE, which leads to the following expression

$$E_\infty^1 = \frac{n}{2c^2}\left(-\beta + \sqrt{\beta^2 + 4(\alpha^2 + \sigma_v^2)\sigma_w^2 c^2}\right), \tag{2.8}$$

where $\beta = (1 - a^2)(\sigma_v^2 + \alpha^2) - c^2\sigma_w^2$.

Instead of input perturbation, we can use the architecture shown on Figure 2.2 with $D = \mathbf{1}_n^T$,

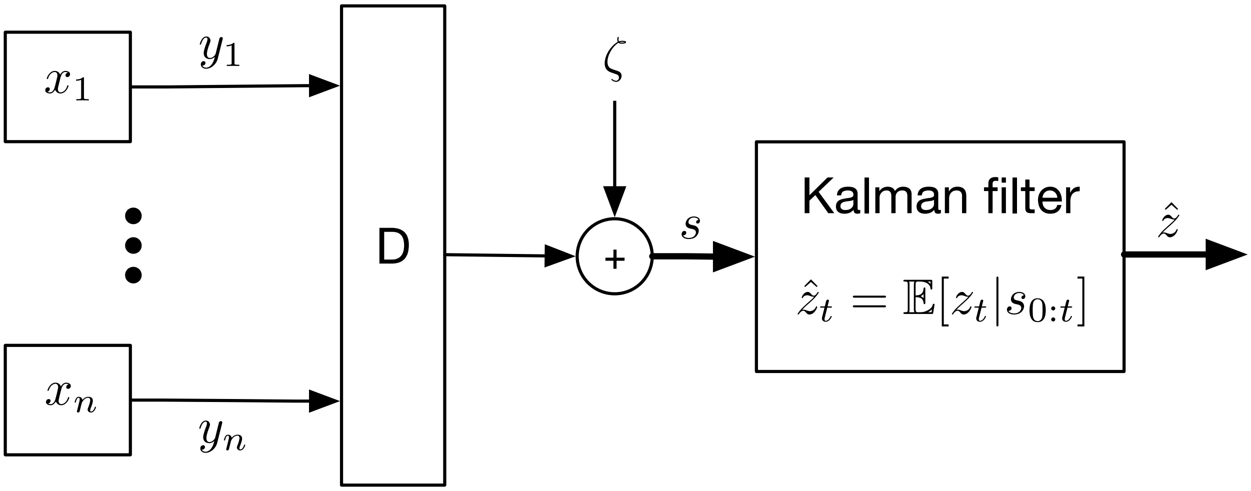


Figure 2.2 : Differentially private Kalman filtering architecture with first-stage aggregation

a $1 \times n$ row vector of ones. Consider the same adjacency relation (2.7) and denote $\eta_t = D y_t = \sum_{i=1}^n y_{i,t}$, $\theta_t = \sum_{i=1}^n w_{i,t}$ and $\lambda_t = \sum_{i=1}^n v_{i,t}$. We have

$$\begin{aligned} z_{t+1} &= a z_t + n b u_t + \theta_t, \\ \eta_t &= c z_t + \lambda_t. \end{aligned}$$

Since again $\sup_{y,y':\mathrm{Adj}(y,y')} \|Dy - Dy'\|_2 = \rho$, releasing the scalar signal $s_t = \eta_t + \zeta_t$, with $\zeta_t \sim \mathcal{N}(0, \alpha^2)$, is (ϵ, δ)-differentially private for the adjacency relation (2.7). The MSE of a Kalman filter estimating z from this signal s, with the dynamics of the model (2.2), can again be obtained by solving an ARE, which leads to the following expression

$$E_\infty^2 = \frac{n}{2c^2} \left(-\beta_{(n)} + \sqrt{\beta_{(n)}^2 + 4 \left(\frac{\alpha^2}{n} + \sigma_v^2 \right) \sigma_w^2 c^2} \right), \tag{2.9}$$

where $\beta_{(n)} = (1 - a^2) \left(\sigma_v^2 + \frac{\alpha^2}{n} \right) - c^2 \sigma_w^2$.

Comparing (2.8) and (2.9), we see that the only difference is the vanishing influence of the privacy preserving noise on E_∞^2 as n increases, with the term α^2/n replacing α^2 in E_∞^1. For example, if $n = 100$, $a = 1$, $c = 1$, $\rho = 50$, $\epsilon = \ln(3)$, $\delta = 0.05$, $\sigma_w^2 = 0.5$, $\sigma_v^2 = 0.9$, we obtain $E_\infty^1 \approx 6235$ and $E_\infty^2 \approx 650$. It is indeed desirable that as the number of agents n increases, differential privacy becomes easier to enforce and the impact of the privacy requirement on achievable performance decreases, a feature that the architecture of Figure 2.2 has the potential to achieve. The design of this architecture is discussed in the next section for the general filtering problem of Section 2.1.

2.2.2 Design of the Two-Stage Mechanism

Following Figure 2.2, we construct a differentially private estimate $\hat{z}_t$ of z_t by first multiplying the global signal y with a constant matrix

$$D = \begin{bmatrix} D_1 & D_2 & \dots & D_n \end{bmatrix}, \tag{2.10}$$

with the matrices $D_i \in \mathbb{R}^{q \times p_i}$, $1 \leq i \leq n$, to be designed and q to be determined. Then, we add white Gaussian noise ζ according to the Gaussian mechanism, in order to make the signal s differentially private, with

$$s_t = Dy_t + \zeta_t = DC_t x_t + Dv_t + \zeta_t, \; 0 \leq t \leq T. \tag{2.11}$$

Therefore, the role of the matrix D is to combine the individual signals appropriately before adding the privacy-preserving noise, in order to decrease the overall sensitivity (see Definition 2), while preserving enough information for z to be estimated with sufficient accuracy. Finally, we construct a causal MMSE estimator $\hat{z}$ of z from s, a task for which it is optimal to use a Kalman filter, since the system model producing s with the state dynamics of (2.2) is still linear and Gaussian. This Kalman filter produces a state estimate $\hat{x}$ of x and then $\hat{z}_t = L_t \hat{x}_t$ for all t.

Given D, for measurement signals y and y' adjacent according to (2.7) and differing at index i, we have

$$\|Dy - Dy'\|_2 = \|D_i y_i - D_i y_i'\|_2 \leq \rho_i \|D_i\|_2,$$

where $\|D_i\|_2$ denotes the maximum singular value of the matrix D_i, and there are adjacent signals y_i, y_i' achieving the bound. Hence, we can bound the sensitivity of the memoryless system $y \mapsto Dy$ as follows

$$\begin{aligned} \triangle_2 D &:= \sup_{y,y' : \text{Adj}(y,y')} \|Dy - Dy'\|_2 \\ &= \max_{1 \leq i \leq n} \{\rho_i \|D_i\|_2\}. \end{aligned} \tag{2.12}$$

Therefore, from Theorem 2, for any matrix D, releasing $s_t = Dy_t + \zeta_t$, with

$$\zeta_t \sim \mathcal{N}(0, (\kappa_{\delta,\epsilon} \triangle_2 D)^2 I_q),$$

is (ϵ, δ)-differentially private for the adjacency relation (2.7). The estimate $\hat{z}$ is then also

(ϵ, δ)-differentially private, since it is obtained by post-processing s, without re-accessing the sensitive signal y.

Input Transformation Optimization

We can now consider the problem of optimizing the choice of matrix D. Let $\hat{x}_t^- = \mathbb{E}[x_t | s_{0:t-1}]$ and $\hat{x}_t = \mathbb{E}[x_t | s_{0:t}]$ be the state estimates produced by the Kalman filter of Figure 2.2 after the prediction step and the measurement update step respectively [64]. Let $\bar{\Sigma}_t = \mathbb{E}[(x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)^T | s_{0:t-1}]$ and $\Sigma_t = \mathbb{E}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T | s_{0:t}]$ be the corresponding error covariance matrices. We also denote by $\bar{\Sigma}_0 = \text{diag}(\Sigma_{1,0}^-, \ldots \Sigma_{n,0}^-)$ the covariance matrix for the initial state x_0. For completeness, we recall here the equations of the Kalman filter. Given the dynamics (2.2) and the measurement equation (2.11), with ζ a Gaussian noise with covariance matrix $(\kappa_{\delta,\epsilon} \triangle_2 D)^2 I_q$, we have for $t \geq 0$ and starting from $\hat{x}_0^- := \bar{x}_0$

$$\begin{aligned} \hat{x}_t &= \hat{x}_t^- + K_t^f (s_t - DC_t \hat{x}_t^-), \\ \hat{x}_{t+1}^- &= A_t \hat{x}_t + B_t u_t, \qquad \text{with} \\ K_t^f &= \bar{\Sigma}_t C_t^T D^T (D(C_t \bar{\Sigma}_t C_t^T + V) D^T + \kappa_{\delta,\epsilon}^2 \Delta_2 D^2 I_q)^{-1}. \end{aligned} \tag{2.13}$$

The error covariance matrices evolve for $t \geq 0$ as

$$\Sigma_t^{-1} = \bar{\Sigma}_t^{-1} + C_t^T \Pi C_t, \;\; \bar{\Sigma}_{t+1} = A_t \Sigma_t A_t^T + W_t,$$

where $\Pi = D^T (DVD^T + (\kappa_{\delta,\epsilon} \triangle_2 D)^2 I_q)^{-1} D$. With $z = L_t x_t$ and its estimator $\hat{z}_t = L_t \hat{x}_t$, we can rewrite the MSE E_T in (2.4) as

$$E_T = \frac{1}{T+1} \sum_{t=0}^{T} \text{Tr}(L_t \Sigma_t L_t^T).$$

As a result, a matrix D minimizing the MSE can be found by solving the following optimization problem

$$\min_{q \in \mathbb{N}, D \in \mathbb{R}^{q \times p}} \quad \frac{1}{T+1} \sum_{t=0}^{T} \text{Tr}\left(L_t \Sigma_t L_t^T \right) \tag{2.14a}$$

$$\text{s.t. } \Sigma_0^{-1} = \; \bar{\Sigma}_0^{-1} + C_0^T \Pi C_0, \tag{2.14b}$$

$$\Sigma_{t+1}^{-1} = (A_t \Sigma_t A_t^T + W_t)^{-1} + C_{t+1}^T \Pi C_{t+1}, \;\; 0 \leq t \leq T-1, \tag{2.14c}$$

$$\Pi = D^T (DVD^T + \kappa_{\delta,\epsilon}^2 (\Delta_2 D)^2 I_q)^{-1} D. \tag{2.14d}$$

Note that the error Σ_t is finite and the Kalman filter converges when the system (2.2) is observable from the measurements z. In the minimization (2.14a), we have emphasized that finding the first dimension q of the matrix D is part of the optimization problem. Note also that we can write the optimization problem above equivalently as a minimization over the variables D, Π, and $\{\Sigma_t\}_{0 \leq t \leq T}$, but the variables other than D can be immediately eliminated using the equality constraints (2.14b)-(2.14d).

With our assumption $V \succ 0$, we obtain an equivalent form for (2.14d) by using the matrix inversion lemma

$$\Pi = V^{-1} - V^{-1} \left(V^{-1} + \frac{D^T D}{(\kappa_{\delta,\epsilon} \Delta_2 D)^2} \right)^{-1} V^{-1}, \tag{2.15}$$

or, alternatively,

$$\kappa_{\delta,\epsilon}^2 \left[(V - V \Pi V)^{-1} - V^{-1} \right] = \frac{D^T D}{\Delta_2 D^2} \quad . \tag{2.16}$$

Semidefinite Programming-based Synthesis

In this section, we show that the optimization problem (2.14a)-(2.14d) can be recast as a semidefinite program (SDP) and hence solved efficiently [106], if we impose the following additional constraints on D

$$\Delta_2 D = 1 = \rho_1 \|D_1\|_2 = \ldots = \rho_n \|D_n\|_2. \tag{2.17}$$

First, the following Lemma shows that in fact no loss of performance occurs by adding the constraint (2.17) to (2.14a)-(2.14d), i.e., that this constraint is satisfied automatically by some matrix D^* that is optimal for (2.14a)-(2.14d).

Lemma 3 *Adding the constraint* (2.17) *to the problem* (2.14a)-(2.14d) *does not change the value of the minimum nor the existence of a minimizer.*

Proof. Consider a matrix D and a corresponding sequence $\{\Sigma_t\}_{0 \leq t \leq T}$ defined by the iterations (2.14b)-(2.14d). First, rescaling D to λD for any $\lambda \neq 0$ does not impact the constraint (2.14d) (note that $\Delta_2(\lambda D) = \lambda \Delta_2 D$), and so we can add the constraint $\Delta_2 D = 1$ without changing the solution of the optimization problem (2.14a)-(2.14d).

Next, if the other constraints of (2.17) are not satisfied by D, construct the $p \times p$ matrix $M = D^T D + \operatorname{diag}\left(\{\eta_i I_{p_i}\}_{1 \leq i \leq n} \right)$, with $\eta_i = (\Delta_2 D / \rho_i)^2 - \|D_i\|_2^2$. Since $\eta_i \geq 0$ by (2.12), M is positive semi-definite. The i^{th} diagonal block of M is $M_{ii} = D_i^T D_i + [(\Delta_2 D/\rho_i)^2 - \|D_i\|_2^2] I_{p_i}$, which has maximum eigenvalue $(\Delta_2 D/\rho_i)^2$. Define some matrix $\tilde{D}$ such that $\tilde{D}^T \tilde{D} = M$ and group the columns of $\tilde{D}$ as $\tilde{D} = \begin{bmatrix} \tilde{D}_1 & \dots \tilde{D}_n \end{bmatrix}$ as for D, so that $\tilde{D}_i$ consists of p_i columns. In particular $M_{ii} = \tilde{D}_i^T \tilde{D}_i$, so $\tilde{D}_i^T \tilde{D}_i$ has maximum eigenvalue $(\Delta_2 D/\rho_i)^2$ and hence $\tilde{D}_i$ has maximum singular value $(\Delta_2 D/\rho_i)$. In other words, $\tilde{D}$ satisfies (2.17) with a sensitivity $\Delta_2 D = \Delta_2 \tilde{D}$ that is unchanged, and moreover $\tilde{D}^T \tilde{D} = M \succeq D^T D$.

Therefore, when we replace D by $\tilde{D}$, $\Delta_2 D$ in the denominator of (2.15) remains unchanged, and moreover

$$\left(V^{-1} + \frac{\tilde{D}^T \tilde{D}}{(\kappa_{\delta,\epsilon} \Delta_2 \tilde{D})^2} \right)^{-1} \preceq \left(V^{-1} + \frac{D^T D}{(\kappa_{\delta,\epsilon} \Delta_2 D)^2} \right)^{-1}$$

hence $\tilde{\Pi} \succeq \Pi$, where $\tilde{\Pi}$ and Π are defined according to (2.15) or equivalently (2.14d) for $\tilde{D}$ and D respectively. Let $K := \tilde{\Pi} - \Pi \succeq 0$. Replacing Π by $\tilde{\Pi}$ in (2.14b), we obtain a matrix $\tilde{\Sigma}_0$ satisfying $\tilde{\Sigma}_0^{-1} = \Sigma_0^{-1} + C_0^T K C_0 \succeq \Sigma_0^{-1}$, so $\tilde{\Sigma}_0 \preceq \Sigma_0$. Now if we have two matrices $\tilde{\Sigma}_t \preceq \Sigma_t$, and we use these two matrices together with $\tilde{\Pi}$ and Π to define $\tilde{\Sigma}_{t+1}, \Sigma_{t+1}$ according to (2.14c), then immediately

$$\begin{aligned} \tilde{\Sigma}_{t+1}^{-1} &= (A_t \tilde{\Sigma}_t A_t^T + W_t)^{-1} + C_{t+1}^T \tilde{\Pi} C_{t+1} \\ &\succeq (A_t \Sigma_t A_t^T + W_t)^{-1} + C_{t+1}^T \tilde{\Pi} C_{t+1} \\ &= \Sigma_{t+1}^{-1} + C_{t+1}^T K C_{t+1} \succeq \Sigma_{t+1}^{-1}. \end{aligned}$$

Therefore, $\tilde{\Sigma}_{t+1} \preceq \Sigma_{t+1}$. Hence, by induction, starting from $\tilde{D}$ we obtain a sequence $\{\tilde{\Sigma}_t\}_t$ such that $\tilde{\Sigma}_t \preceq \Sigma_t$ for all $t \geq 0$. This gives a smaller or equal cost

$$\frac{1}{T+1} \sum_{t=0}^{T+1} \operatorname{Tr}(L_t \tilde{\Sigma}_t L_t^T) \leq \frac{1}{T+1} \sum_{t=0}^{T+1} \operatorname{Tr}(L_t \Sigma_t L_t^T),$$

and so the lemma is proved. □

By Lemma 3, we can add without loss of optimality the constraints (2.17) to (2.14a)-(2.14d), which allows us in the following to recast the problem as an SDP. Let $\alpha_i = \kappa_{\delta,\epsilon} \rho_i$, for all $1 \leq i \leq n$. Denote $E_i = \begin{bmatrix} 0 & \dots & I_{p_i} & \dots & 0 \end{bmatrix}^T$ the $p \times p_i$ matrix whose elements are zero except for an identity matrix in its i^{th} block. Define the following symmetric matrices, for $1 \leq i \leq n$

$$\Phi_i(\Pi) = \frac{I_{p_i}}{\alpha_i^2} - E_i^T \left[(V - V \Pi V)^{-1} - V^{-1} \right] E_i.$$

Lemma 4 *If* Π, D *satisfy the constraints* (2.16)-(2.17)*, then* Π *satisfies the constraints* $V -$

$V\Pi V \succeq 0$ and $\lambda_{\min}(\Phi_i(\Pi)) = 0$ for all $1 \leq i \leq n$.

Conversely, if Π satisfies these constraints, then there exists a matrix D such that Π, D satisfy (2.16)-(2.17)*. One such D can be obtained by the factorization of*

$$\kappa_{\delta,\epsilon}^2 \left[(V - V\Pi V)^{-1} - V^{-1} \right] = D^T D \tag{2.18}$$

(e.g., via singular value decomposition (SVD)) and will then satisfy $\Delta_2 D = 1$.

Proof. If Π, D satisfy (2.16)-(2.17), then $V - V\Pi V \succeq 0$ is immediate from (2.15), since it is equal to $\left(V^{-1} + \frac{D^T D}{(\kappa_{\delta,\epsilon}\Delta_2 D)^2}\right)^{-1}$. For $1 \leq i \leq n$, since $\|D_i\|_2 = 1/\rho_i$ by (2.17), the i^{th} diagonal block of the matrix on the left hand side of (2.16), which is equal to $D_i^T D_i$, has maximum eigenvalue equal to $1/\rho_i^2$. This is equivalent to saying that $\lambda_{\min}(\Phi_i(\Pi)) = 0$.

For the converse, since $V \succeq V - V\Pi V \succeq 0$, we have $(V - V\Pi V)^{-1} \succeq V^{-1}$ and hence the matrix on the left hand side of (2.16) or (2.18) is positive semidefinite, so a matrix factor D satisfying (2.18) can be found. As above, the condition $\lambda_{\min}(\Phi_i(\Pi)) = 0$ is then equivalent to saying that

$$\lambda_{\max}\left(E_i^T \left[(V - V\Pi V)^{-1} - V^{-1} \right] E_i \right) = \frac{1}{\alpha_i^2},$$

and hence $\lambda_{\max}(D_i^T D_i) = 1/\rho_i^2$, or $\|D_i\|_2 = 1/\rho_i$. □

Note that if we relax the constraints of Lemma 4 to $\lambda_{\min}(\Phi_i(\Pi)) \geq 0$, i.e., $\Phi_i(\Pi) \succeq 0$, we can rewrite these constraints as the LMIs

$$\begin{bmatrix} I_{p_i}/\alpha_i^2 + V_i^{-1} & E_i^T \\ E_i & V - V\Pi V \end{bmatrix} \succeq 0, \quad 1 \leq i \leq n, \tag{2.19}$$

by noting that $E_i^T V^{-1} E_i = V_i^{-1}$ and taking a Schur complement.

Next, define the information matrices $\Omega_t = \Sigma_t^{-1}$, for $0 \leq t \leq T$. If the matrices W_t are invertible, denoting $\Xi_t = W_t^{-1}$ and using the matrix inversion lemma in (2.14c), one gets

$$\begin{aligned} & C_{t+1}^T \Pi C_{t+1} - \Omega_{t+1} + \Xi_t \\ & \quad - \Xi_t A_t (\Omega_t + A_t^T \Xi_t A_t)^{-1} A_t^T \Xi_t = 0. \end{aligned} \tag{2.20}$$

Replacing the equality in (2.20) by $\succeq 0$ and taking a Schur complement, together with the inequalities (2.19), leads to the following SDP with variables $\Pi \succeq 0, \{X_t \succeq 0, \Omega_t \succ 0\}_{0 \leq t \leq T}$

$$\min_{\Pi \succeq 0, \{X_t, \Omega_t\}_{0 \leq t \leq T}} \quad \frac{1}{T+1} \sum_{t=0}^{T} \mathrm{Tr}(X_t) \quad \text{s.t.} \tag{2.21a}$$

$$\begin{bmatrix} X_t & L_t \\ L_t^T & \Omega_t \end{bmatrix} \succeq 0, \quad 0 \leq t \leq T, \tag{2.21b}$$

$$\Omega_0 = \bar{\Sigma}_0^{-1} + C_0^T \Pi C_0, \tag{2.21c}$$

$$\begin{bmatrix} C_{t+1}^T \Pi C_{t+1} - \Omega_{t+1} + \Xi_t & \Xi_t A_t \\ A_t^T \Xi_t & \Omega_t + A_t^T \Xi_t A_t \end{bmatrix} \succeq 0,$$

$$0 \leq t \leq T-1, \tag{2.21d}$$

$$\begin{bmatrix} I_{p_i}/\alpha_i^2 + V_i^{-1} & E_i^T \\ E_i & V - V \Pi V \end{bmatrix} \succeq 0, \quad 1 \leq i \leq n. \tag{2.21e}$$

Here the minimization of the cost (2.14a) has been replaced by the minimization of (2.21a), after introducing the slack variable X_t satisfying (2.21b), or equivalently $X_t \succeq L_t \Omega_t^{-1} L_t^T$ by taking a Schur complement. Since we replaced the equality in (2.20) by an inequality and relaxed the constraints of the first part of Lemma 4 to (2.21e), the SDP above is a relaxation of the original problem (2.14a)-(2.14d). The purpose of the next theorem is to show that this relaxation is tight. Once an optimal solution for this SDP is obtained, we recover an optimal matrix D from Π by the factorization (2.18).

Theorem 5 *Let* $\Pi^* \succeq 0$,$\{X_t^* \succeq 0, \Omega_t^* \succ 0\}_{0 \leq t \leq T}$ *be an optimal solution for* (2.21a)-(2.21e). *Suppose that for some* $0 \leq t \leq T$, *we have* $L_t (\Omega_t^*)^{-1} C_t^T \neq 0$. *Let* D^* *be a matrix obtained from* Π^* *by the factorization* (2.18). *Then* D^* *is an optimal solution for* (2.14a)-(2.14d), *which moreover satisfies* $\|D_i^*\|_2 = 1/\rho_i$ *for* $1 \leq i \leq n$, *with the decomposition* (2.10). *The corresponding optimal covariance matrices* $\{\Sigma_t^*\}_{0 \leq t \leq T}$ *for the Kalman filter can be computed using the equations* (2.14b)-(2.14d). *Finally, the optimal costs of* (2.14a)-(2.14d) *and* (2.21a)-(2.21e) *are equal, i.e., the SDP relaxation is tight.*

Remark 1 *Even though the condition* $L_t (\Omega_t^*)^{-1} C_t^T \neq 0$ *introduced to guarantee the possibility of constructing the matrix D in the proof is not an explicit condition expressed directly in term of the problem parameters, it appears to be a weak requirement in practice.*

Proof. Consider Π^*, $\{X_t^*, \Omega_t^*\}_{0 \leq t \leq T}$ an optimal solution of the SDP (2.21a)-(2.21e). As explained below Lemma 4, the constraint (2.21e) is equivalent to $\lambda_{\min}(\Phi_i(\Pi^*)) \geq 0$. We show that we cannot have $\lambda_{\min}(\Phi_i(\Pi^*)) > 0$. Indeed, otherwise there exists $\eta > 0$ such that the matrix $\tilde{\Pi} = \Pi^* + \eta I_p$ still satisfies (2.21e). Using this matrix $\tilde{\Pi}$ in (2.21c), we obtain a matrix $\tilde{\Omega}_0 = \Omega_0^* + \eta C_0^T C_0$ feasible for (2.21c). Now define $\tilde{\Omega}_1 = \Omega_1^* + \eta C_1^T C_1$. One can immediately check that $\tilde{\Pi}, \tilde{\Omega}_0$ and $\tilde{\Omega}_1$ satisfy (2.21d) for $t = 0$, using the fact that $\Omega_0^*, \Omega_1^*, \Pi^*$ are feasible and that $C_0^T C_0 \succeq 0$. Similarly the matrices $\tilde{\Omega}_t = \Omega_t^* + \eta C_t^T C_t$ are feasible in (2.21d) for all $0 \leq t \leq T$. Now, taking a Schur complement in (2.21b), we obtain that the matrices $\tilde{X}_t = L_t \tilde{\Omega}_t^{-1} L_t^T$ are feasible. By the matrix inversion lemma we can write

$$\tilde{X}_t = X_t^* - L_t (\Omega_t^*)^{-1} C_t^T K_t C_t (\Omega_t^*)^{-1} L_t^T.$$

for some matrices $K_t \succ 0$. These matrices $\tilde{X}_t$ give a cost $\frac{1}{T+1} \sum_{t=0}^T \mathrm{Tr}(X_t^*) - \| L_t (\Omega_t^*)^{-1} C_t^T K_t^{1/2} \|_F$, which is a strict improvement over the assumed optimal solution, as soon as one matrix $L_t (\Omega_t^*)^{-1} C_t^T$ is not zero (since the K_t's are invertible). Hence, we have a contradiction and so we must have $\lambda_{\min}(\Phi_i(\Pi^*)) = 0$. We can then apply Lemma 4 and construct a matrix D^* from Π^* as in (2.18), so that the pair Π^*, D^* satisfies (2.16)-(2.17).

Let $\mathcal{V}^*$ be the optimum value of (2.21a)-(2.21e), and V^* that of (2.14a)-(2.14d). First, $\mathcal{V}^* \leq V^*$ since the constraints of the original problem have been relaxed to obtain the SDP. We now show how to construct a sequence $\{\Sigma_t^*\}_{0 \leq t \leq T}$, which together with Π^* satisfy the constraints of (2.14a)-(2.14d) and achieve the cost $\mathcal{V}^*$, thereby proving the remaining claims of the theorem. Note that since $\Omega_t^* + A_t^T \Xi_t A_t \succ 0$, (2.21d) is equivalent to $\mathcal{R}_t(\Omega_t^*, \Omega_{t+1}^*) \succeq 0$, where

$$\begin{aligned} \mathcal{R}_t(\Omega_t, \Omega_{t+1}) := & C_{t+1}^T \Pi^* C_{t+1} - \Omega_{t+1} + \Xi_t \\ & - \Xi_t A_t (\Omega_t + A_t^T \Xi_t A_t)^{-1} A_t^T \Xi_t. \end{aligned}$$

First, we take $\Sigma_0^* = (\Omega_0^*)^{-1}$. If $\mathcal{R}_t(\Omega_t^*, \Omega_{t+1}^*) = 0$ for all $0 \leq t \leq T-1$, then the matrices Ω_t^* satisfy (2.20) and we can take $\Sigma_t^* = (\Omega_t^*)^{-1}$ for all t, since these matrices satisfy the equivalent condition (2.14c). Otherwise, let $\tilde{t}$ be the first time index such that $\mathcal{R}_{\tilde{t}}(\Omega_{\tilde{t}}^*, \Omega_{\tilde{t}+1}^*)$ is not zero. For $t \leq \tilde{t}$, we take $\Sigma_t^* = (\Omega_t^*)^{-1}$ and so in particular we have $\mathcal{R}_{\tilde{t}}((\Sigma_{\tilde{t}}^*)^{-1}, \Omega_{\tilde{t}+1}^*) \succeq 0$ and not zero. Consider the matrix $\tilde{\Omega}_{\tilde{t}+1} = \Omega_{\tilde{t}+1}^* + \mathcal{R}_{\tilde{t}}(\Omega_{\tilde{t}}^*, \Omega_{\tilde{t}+1}^*)$, which then satisfies $\mathcal{R}_{\tilde{t}}(\Omega_{\tilde{t}}^*, \tilde{\Omega}_{\tilde{t}+1}) = 0$ by definition. We set $\Sigma_{\tilde{t}+1}^* = \tilde{\Omega}_{\tilde{t}+1}^{-1}$.

Now note that we again have $\mathcal{R}_{\tilde{t}+1}((\Sigma_{\tilde{t}+1}^*)^{-1}, \tilde{\Omega}_{\tilde{t}+2}) \succeq 0$, by verifying that (2.21d) is satisfied at $t+1$, using the fact that $(\Sigma_{\tilde{t}+1}^*)^{-1} = \tilde{\Omega}_{\tilde{t}+1} \succeq \Omega_{\tilde{t}+1}^*$. From here, we can proceed by induction, assuming that $\Sigma_0^*, \ldots, \Sigma_t^*$ are set and taking

$$\Sigma_{t+1}^* = (\tilde{\Omega}_{t+1})^{-1} := (\Omega_{t+1}^* + \mathcal{R}_t((\Sigma_t^*)^{-1}, \Omega_{t+1}^*))^{-1}, \tag{2.22}$$

which reduces to $(\Omega^*_{t+1})^{-1}$ if $\mathcal{R}_t((\Sigma^*_t)^{-1}, \Omega^*_{t+1}) = 0$.

The procedure above provides matrices Π^*, $\{\Sigma^*_t\}_{0 \leq t \leq T}$ satisfying the constraints of the original program (2.14a)-(2.14d). By construction, we have $(\Sigma^*_t)^{-1} \succeq \Omega^*_t$ and the matrices $(\Sigma^*_t)^{-1}$ also satisfy (2.21d). Therefore, replacing, for each $0 \leq t \leq T$, Ω^*_t by $(\Sigma^*_t)^{-1}$ and X^*_t by $L_t \Sigma_t L_t^T$ in the solution of (2.21a)-(2.21e) that we started with gives a cost $\mathcal{V} \leq \mathcal{V}^*$ for the SDP, hence $\mathcal{V} = \mathcal{V}^*$ by optimality of $\mathcal{V}^*$. But this cost $\mathcal{V}$ is also equal to the cost $\frac{1}{T+1} \sum_{t=0}^{T} \mathrm{Tr}(L_t \Sigma^*_t L_t^T)$ of (2.14a). Hence, we have shown that (2.14a)-(2.14d) and (2.21a)-(2.21e) have the same minimum value, and constructed an optimal solution D^*, Π^*, $\{\Sigma^*_t\}_{0 \leq t \leq T}$ to (2.14a)-(2.14d) achieving this value. □

2.2.3 Stationary problem

In the stationary case with $T \to \infty$ and the model (2.2)-(2.3) now assumed time-invariant and detectable, we wish to find a signal aggregation matrix D followed by a time-invariant Kalman filter to minimize the steady-state MSE E_∞. This can be done by solving the following SDP with variables $\Pi \succeq 0$, $X \succeq 0$, $\Omega \succ 0$

$$\min_{\Pi \succeq 0, X, \Omega} \quad \mathrm{Tr}(X) \quad \text{s.t.} \tag{2.23a}$$

$$\begin{bmatrix} X & L \\ L^T & \Omega \end{bmatrix} \succeq 0, \tag{2.23b}$$

$$\begin{bmatrix} C^T \Pi C - \Omega + \Xi & \Xi A \\ A^T \Xi & \Omega + A^T \Xi A \end{bmatrix} \succeq 0, \tag{2.23c}$$

$$\begin{bmatrix} I_{p_i}/\alpha_i^2 + V_i^{-1} & E_i^T \\ E_i & V - V \Pi V \end{bmatrix} \succeq 0, \quad 1 \leq i \leq n. \tag{2.23d}$$

Compared to (2.21a)-(2.21e), this SDP is of much smaller size, due to the fact that the transient behavior is neglected in the performance measure. The proof of the following theorem is similar to that of Theorem 5.

Theorem 6 *Let* $\Pi^* \succeq 0$*,* $X^* \succeq 0$*,* $\Omega^* \succ 0$ *be an optimal solution for* (2.23a)-(2.23d)*. Suppose that we have* $L(\Omega^*)^{-1}C^T \neq 0$*. Let* D^* *be a matrix obtained from* Π^* *by the factorization* (2.18)*. Then* D^* *minimizes the steady-state MSE* E_∞ *among all possible matrices* D *introduced as in Figure 2.2, and the corresponding value of* E_∞ *is equal to the optimal value of the SDP.*

Note that given the optimum matrix D^* and corresponding Π^*, an alternative way of computing E_∞ is by solving an ARE to obtain the steady-state prediction error covariance matrix

$\bar{\Sigma}_\infty$ for $\hat{x}^-$, then compute the steady-state error covariance matrix $\Sigma_\infty = (\bar{\Sigma}_\infty^{-1} + C^T \Pi^* C)^{-1}$ for the estimator $\hat{x}$, and finally $E_\infty = \mathrm{Tr}(L \Sigma_\infty L^T)$.

2.2.4 Syndromic Surveillance Example

To illustrate the differentially private filtering methodology, including issues related to the choice of model and adjacency relation (2.7), we discuss in this section an example motivated by the analysis of epidemiological data. Consider a scenario in which Public Health Services (PHS) must publish for a population infected by a disease the number I_t of infectious people, i.e., those who have the disease and are able to infect others. PHS use privacy-sensitive data collected from $n = 12$ hospitals, with each hospital i recording the number $I_{i,t}$ of infectious people in its area, as well as the number $R_{i,t}$ of recovered people, i.e, those who were infected by the disease and are now immune. For each area i, these numbers are assumed to follow a discrete-time SEIR epidemiological model for the specific disease [19, 107], written here first without process noise, obtained by discretizing a classical continuous-time model [108] using a forward Euler discretization [109]

$$\begin{aligned} S_{i,t+1} &= S_{i,t} - \beta_i S_{i,t} I_{i,t} / N_i, \\ E_{i,t+1} &= (1 - \tau_i) E_{i,t} + \beta_i S_{i,t} I_{i,t} / N_i, \\ I_{i,t+1} &= (1 - \vartheta_i) I_{i,t} + \tau_i E_{i,t}, \\ R_{i,t+1} &= R_{i,t} + \vartheta_i I_{i,t}, \end{aligned} \tag{2.24}$$

where S_i represents the number of susceptible people, i.e., those who are not infected but could become infected, E_i the number of exposed people, i.e., those infected but not yet able to infect others, and N_i the total number of people. The parameters τ_i, β_i and ϑ_i represent the transition rates from one disease stage to the next.

For each hospital i's area, N_i in (2.24) is assumed constant for the time interval of interest. Moreover, let us make an approximation that this period is short enough or the disease at an early-enough stage so that $S_{i,t}$ can also be assumed approximately constant, equal to $S_{i,0}$. Then, the remaining states $\eta_{i,t} = [E_{i,t}, I_{i,t}, R_{i,t}]^T \in \mathbb{R}^3$ evolve as a linear system of the form

$$\eta_{i,t+1} = \mathcal{A}_i \, \eta_{i,t} + \varphi_{i,t}, \quad 0 \leq t \leq T - 1, \tag{2.25}$$

where

$$\mathcal{A}_i = \begin{bmatrix} 1-\tau_i & \beta_i S_{i,0}/N_i & 0 \\ \tau_i & 1-\vartheta_i & 0 \\ 0 & \vartheta_i & 1 \end{bmatrix}, \tag{2.26}$$

and we introduced the white Gaussian process noise $\varphi_{i,t} \sim \mathcal{N}(0, \Phi_i)$ in the model, with covariance matrices $\Phi_i \succ 0$, for $1 \leq i \leq 12$. Now, consider the problem of choosing the level ρ_i in the adjacency relation (2.7) to provide a meaningful privacy guarantee to the patients. If we assume that the measurement for hospital i in our model is $y_{i,t} = [I_{i,t}, R_{i,t}]^T$, then once a person becomes sick, he or she is counted at each period either in the signal I_i or, after recovery, in the signal R_i. As a result, the impact of a single individual on the measurement y_i could be quite large, proportional to the time horizon T, requiring in turn a large value of ρ_i to provide strong privacy guarantees, and hence a high level of noise.

A practical solution to this issue is to not continuously record the same individuals, but simply count at each time period t the number of newly infectious individuals, i.e., $y^{(1)}_{i,t} = I_{i,t} - I_{i,t-1}$, as well as the number of newly recovered individuals, $y^{(2)}_{i,t} = R_{i,t} - R_{i,t-1}$. Such a measurement model is much more beneficial from a privacy point of view (as well as closer to an actual counting process that could be used in practice). Indeed, assuming that a given individual can only become sick once, he or she can affect $y^{(1)}_{i,t}$ by ±1 for at most 2 periods t (when the person becomes sick and recovers), and $y^{(2)}_{i,t}$ for at most one period by 1. As a result, one can take $\rho_i = \sqrt{3}$ in (2.7) to provide a strong privacy guarantee, i.e., insensitivity of the published output to the complete record of a single individual. The measurement noise $v_{i,t}$ in the model represents counting errors, e.g., due to people not being diagnosed by the hospital.

To obtain a dynamic model compatible with the measurements $y_{i,t} = [y^{(1)}_{i,t}, y^{(2)}_{i,t}]^T$, define for each hospital i the 4-dimensional state $x_{i,t} = [I_{i,t-1}, R_{i,t} - R_{i,t-1}, E_{i,t}, I_{i,t}]^T$. These states $x_{i,t}$ evolve as (2.1) with

$$A_i = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \vartheta_i \\ 0 & 0 & 1-\tau_i & \beta_i S_{i,0}/N_i \\ 0 & 0 & \tau_i & 1-\vartheta_i \end{bmatrix}, \; B_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.27}$$

and the noise $w_{i,t} = [\varpi_{i,t}, \varphi_{1,i,t}, \varphi_{2,i,t}, \varphi_{3,i,t}]^T$ includes the components of φ_i introduced in (2.25) as well as a small discrete independent Gaussian noise ϖ_i with small variance σ^2, added so that the covariance matrices of $W_i = \text{diag}(\sigma^2, \Phi_i)$ are invertible as required by our algorithms (ideally $\varpi_{i,t}$ would be 0 since the first line of A_i corresponds to the delay in the model). The measurement matrices are immediately

$$C_i = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \ 1 \leq i \leq 12,$$

and one can then verify that the model is observable. Let us assume the following values for the parameters in (2.27)

$$\begin{aligned} &\tau_i = 0.2, \ \beta_i S_{i,0}/N_i = 0.5, \ \vartheta_i = 0.1, \ \text{for} \ \ 1 \leq i \leq 3 \\ &\tau_i = 0.3, \ \beta_i S_{i,0}/N_i = 0.3, \ \vartheta_i = 0.5, \ \text{for} \ \ 4 \leq i \leq 6 \\ &\tau_i = 0.5, \ \beta_i S_{i,0}/N_i = 0.7, \ \vartheta_i = 0.15, \ \text{for} \ \ 7 \leq i \leq 9 \\ &\tau_i = 0.7, \ \beta_i S_{i,0}/N_i = 0.6, \ \vartheta_i = 0.3, \ \text{for} \ \ 10 \leq i \leq 12. \end{aligned}$$

Moreover, assume for all $1 \leq i \leq 12$

$$V_i = 0.4\, I_2, \Phi_i = \begin{bmatrix} 0.3 & -0.15 & 0 \\ -0.15 & 0.3 & -0.15 \\ 0 & -0.15 & 0.3 \end{bmatrix}.$$

The goal of the surveillance system is to continuously release, at each period t, an estimate of the quantity

$$z_t = \sum_{i=1}^{n} \left(\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \times x_{i,t} \right).$$

Let us set the privacy parameters to $\delta = 0.02$ and $\epsilon = \ln(3)$ for example, and ρ_i to $\sqrt{3}$ as discussed above. We design the two-stage architecture of Figure 2.2 by first solving the stationary optimization problem (2.23a)-(2.23d), which provides an optimal matrix Π^*. Recall that the matrix D can be then obtained from the factorization (2.18). The number of rows q of D is then equal to the rank of the matrix $M^* := \kappa_{\delta,\epsilon}^2 \left[(V - V\Pi^* V)^{-1} - V^{-1} \right]$. Depending on the numerical tolerance chosen, this rank can be close to maximal the numerical linear algebra routines used here returned a matrix of rank 22 for example, close to the maximum 24. However, we plot on Figure 2.3 the ratios $\sigma_i/\sigma_{\max}$ of the singular values of M^*, with $\sigma_{\max}$ the maximum singular value. We see that if we select for example only the singular values $\sigma_i \geq 10^{-4}\sigma_{\max}$ in the SVD of M^* and set the smaller ones manually to 0, we obtain a matrix of rank 14, hence a matrix D with 14 rows instead of 24. Reducing the number of rows of D is also beneficial for example in terms of processing complexity of the Kalman filter, which now has fewer inputs. We then verify (by solving an algebraic Riccati

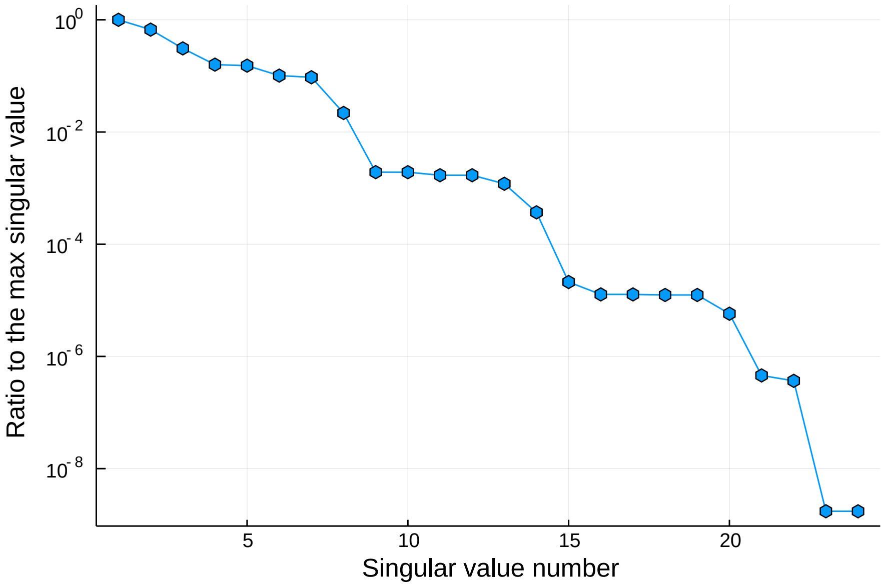


Figure 2.3 : Ratio $\sigma_i/\sigma_{\max}$ for the singular values of M^* (on a logarithmic scale)

equation) that the performance of the steady-state Kalman filter is left virtually unchanged by this truncation, with a steady-state MSE of about 182, i.e., a root mean square error (RMSE) of 13.5 for the estimate of the number I_t of infectious people. In contrast, the input perturbation mechanism (i.e., taking $D = I$) gives a steady-state MSE of 941 or RMSE on I_t of 30.6. For comparison purposes, the non private Kalman filter using all the measurements, without aggregation or privacy-preserving noise, has an RMSE of 5.36. For a two-stage algorithm where we select naively the D matrix as $[I_2, \ldots, I_2]$, we get an RMSE of 23.4.

For $\delta = 0.01$, we compare on Figure 2.4 the steady-state RMSE of the two-stage mechanism and the input perturbation architecture for different values of the privacy parameter ϵ. One can see that by aggregating the input signals, we obtain a much better performance, especially in the high-privacy regime (when ϵ is small). Other measures of performance could also be of interest for the final filtering architecture, such as the convergence time of the estimates. For illustration purposes, Figure 2.5 shows sample paths of the error trajectories of differentially private estimates both for the two-stage mechanism and for the input perturbation mechanism when $\delta = 0.01$ and $\epsilon = \ln(3)$, with the steady-state version of the filters.

2.3 Differentially Private LQG Control

We now turn to the LQG control problem introduced at the end of Section 2.1.1. For concreteness, we assume here that the control input u_t at time t can depend on the measurements

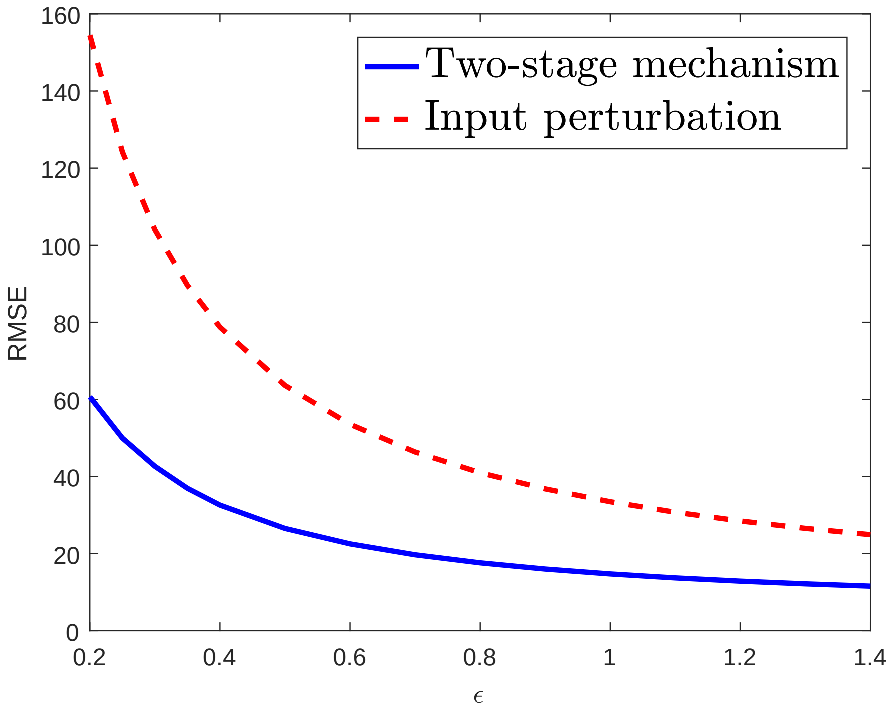


Figure 2.4 : Steady-state RMSE of the total infectious population estimate for the two-stage and input perturbation mechanisms, as a function of the privacy parameter ϵ (here $\delta = 0.01$)

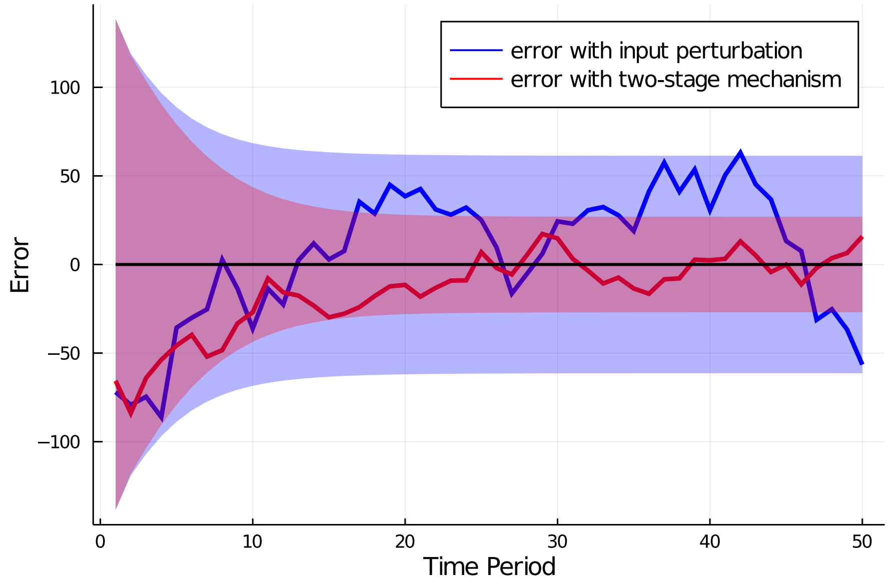


Figure 2.5 : Sample paths and 2σ bands for the error trajectories of differentially private estimates of the total infectious population, for the two-stage and input perturbation mechanisms

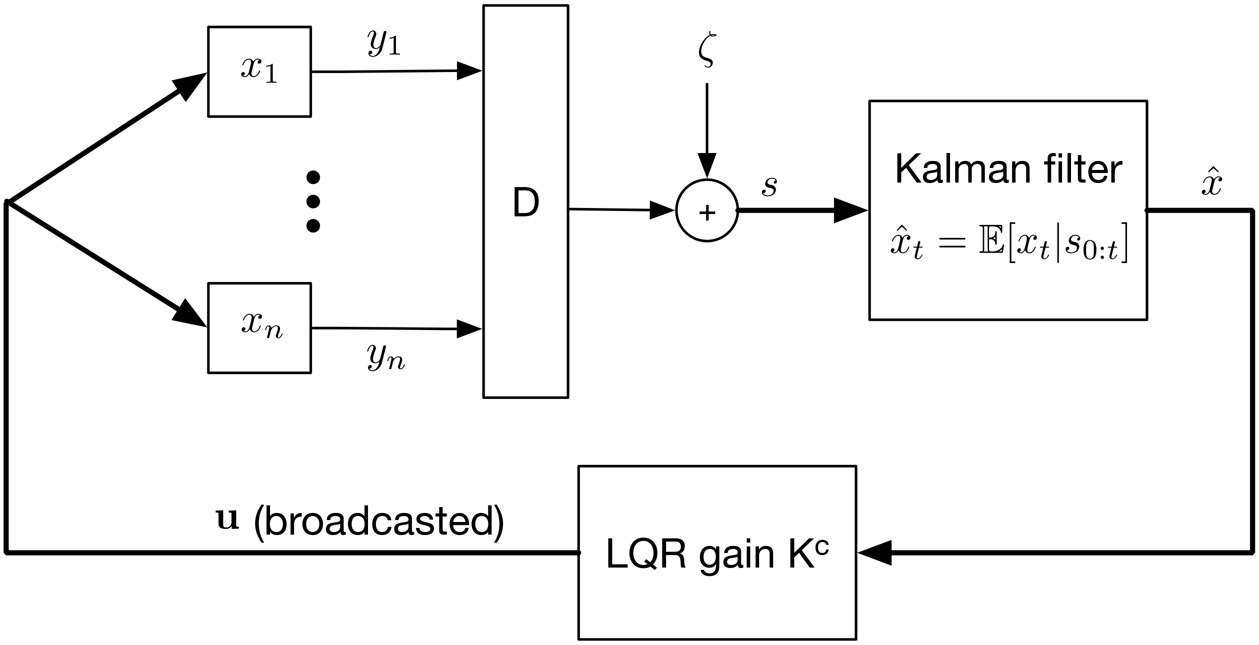


Figure 2.6 : Differentially private LQG control architecture with first-stage signal aggregation

$y_{0:t}$ up to time t. It is straightforward to adapt the discussion to the case where only $y_{0:t-1}$ are available to compute u_t. By the separation principle [110, Chapter 8] for the standard LQG control problem (i.e., with no privacy constraint), the optimal control law for the system (2.2)-(2.3) and quadratic cost (2.5) is of the form $u_t = K_t^c \hat{x}_t$, where: i) $\hat{x}_t = \mathbb{E}[x_t|y_{0:t}]$ is the MMSE estimator, computed by the Kalman filter (2.13) independently of the design of the optimal control law; and ii) K_t^c is the optimal gain for the deterministic linear quadratic regulator (LQR) problem, i.e., assuming that $w = 0$ in (2.2) and $C = I_n$, $v = 0$ in (2.3). In particular, since the sequence of control gains K_t^c can be precomputed, the LQG control problem is similar to the filtering problem considered in the previous section, with the desired published output $\hat{z}_t = L_t \hat{x}_t$ simply replaced by $u_t = K_t^c \hat{x}_t$. This motivates the architecture proposed on Figure 2.6 for differentially private LQG control, which, compared to Figure 2.1b, aggregates the measured signals y_i before adding the privacy-preserving noise. Essentially, the only difference with the Kalman filtering problem is that the performance is measured by (2.5) instead of the MSE (2.4), so that the cost function in the optimization problem for the matrix D needs to be changed. The following theorem summarizes the discussion above and the classical results (see for example [110, Chapter 8]) that allow us to formulate in the following an efficiently solvable optimization problem for the choice of aggregation matrix D on Figure 2.6.

Theorem 7 *Given a choice of matrix D for the differentially private LQG control architec-*

ture of Figure 2.6, the control law $u_t(y_{0:t})$, $t \geq 0$, *minimizing the cost function* (2.5) *takes the form*

$$u_t = K_t^c \hat{x}_t,$$

where $\hat{x}_t$ *is computed by the Kalman filter* (2.13) *and the gains* K_t^c *are precomputed independently of the filtering problem as*

$$K_t^c := -(R_t + B_t^T P_{t+1} B_t)^{-1} B_t^T P_{t+1} A_t$$

with the matrices $P_t \succeq 0$ *given by* $P_T = Q_T$ *and the backward Riccati difference equation*

$$\begin{aligned} P_t = & A_t^T P_{t+1} A_t + Q_t \\ & - A_t P_{t+1} B_t (R_t + B_t^T P_{t+1} B_t)^{-1} B_t^T P_{t+1} A_t. \end{aligned} \tag{2.28}$$

Moreover, the optimal objective (2.5) *corresponding to this control law can be written*

$$J_T = J_T^c + J_T^f(D) \tag{2.29}$$

where

$$J_T^c = \frac{1}{T+1} \left(\bar{x}_0^T P_0 \bar{x}_0 + \mathrm{Tr}(P_0 \bar{\Sigma}_0) + \sum_{t=1}^{T} \mathrm{Tr}(P_t W_t) \right)$$

is a term independent of D *and*

$$J_T^f(D) = \frac{1}{T+1} \sum_{t=0}^{T-1} \mathrm{Tr}(N_t \Sigma_t), \quad \textit{with} \tag{2.30}$$

$$N_t = Q_t + A_t^T P_{t+1} A_t - P_t, \;\; 0 \leq t \leq T-1, \tag{2.31}$$

and the matrices Σ_t *for* $0 \leq t \leq T-1$ *defined by* (2.14b)-(2.14d).

Note that the dependence on D in (2.30) is due to the fact that the error covariance matrices Σ_t depend on D via (2.14b)-(2.14d). Moreover, from (2.28) we see that N_t defined in (2.31) is positive semidefinite. Hence, we can define for all $0 \leq t \leq T-1$ matrices L_t such that $N_t = L_t^T L_t$, and $L_T = 0$, to rewrite the cost (2.30) as $\frac{1}{T+1} \sum_{t=0}^{T} \mathrm{Tr}(L_t \Sigma_t L_t^T)$. Minimizing this cost over the matrices D with the relations (2.14b)-(2.14d) leads to an optimal aggregation matrix D for the architecture of Figure 2.6. The reformulation of this optimization problem as an SDP then follows exactly from the same argument as in Section 2.2.2, which led to Theorem 5. In other words, we have the following result.

Proposition 1 *Let L_t, $0 \leq t \leq T-1$, be any matrices obtained from the factorization*

$$L_t^T L_t = N_t, 0 \leq t \leq T-1,$$

with N_t defined by (2.31), *and let $L_T = 0$. Let $\Pi^* \succeq 0$,$\{X_t^* \succeq 0, \Omega_t^* \succ 0\}_{0 \leq t \leq T}$ be an optimal solution for* (2.21a)-(2.21e) *with this choice of matrices L_t. Suppose that for some* $0 \leq t \leq T$, *we have $L_t(\Omega_t^*)^{-1}C_t^T \neq 0$. Let D^* be a matrix obtained from Π^* by the factorization* (2.18). *Then D^* minimizes the LQG cost* (2.5) *among all the aggregation matrices D of Figure 2.6. This cost is equal to $J_T^c + \frac{1}{T+1}\sum_{t=0}^T \mathrm{Tr}(X_t^*)$, with J_T^c defined in* (2.29).

2.3.1 Stationary Problem

As in Section 2.2.3 for the filtering problem, we can consider the steady-state LQG problem by letting $T \to \infty$ and assuming the model (2.2)-(2.3) and the weight matrices Q and R in the cost (2.5) to be time-invariant. We assume the model to be detectable and stabilizable and the pair $(A, Q^{1/2})$ to be detectable, in order to be able to implement a stabilizing LQG controller. We can take the optimal gains K^c and K^f of the controller and the Kalman filter respectively to be also independent of time. Following Theorem 6 and Proposition 1, we then immediately have the following result for the design of the optimal D matrix.

Proposition 2 *Let P be the positive semidefinite solution of the following algebraic Riccati equation*

$$P = A^T PA + Q - A^T PB(R + B^T PB)^{-1} B^T PA.$$

Let L be any matrix obtained from the factorization

$$L^T L = A^T PA + Q - P.$$

Let $\Pi^* \succeq 0$, $X^* \succeq 0$, $\Omega^* \succ 0$ be an optimal solution for (2.23a)-(2.23d), *for this choice of matrix L. Suppose that we have $L(\Omega^*)^{-1}C^T \neq 0$. Let D^* be a matrix obtained from Π^* by the factorization* (2.18). *Then D^* minimizes the steady-state LQG cost J_∞ among all possible matrices D introduced as in Figure 2.6, the corresponding value of the cost is $J_\infty = \mathrm{Tr}(PW) + \mathrm{Tr}(X^*)$.*

2.3.2 Numerical Simulations

We illustrate the above results numerically for $n = 10$ independent scalar systems, with states $x_{i,t}$ evolving as first order systems with time-invariant dynamics (2.1), where

$$\begin{aligned} &A_1 = 1.1, A_2 = 0.85, A_3 = 0.84, A_4 = 0.7, A_5 = 0.75, \\ &A_6 = 0.9, A_7 = 0.8, A_8 = 1.05, A_9 = 0.99, A_{10} = 1, \end{aligned}$$

$C_i = 1$, $W_i = 0.02$ and $V_i = 0.1$ for all $1 \leq i \leq 10$, and B is a 10×3 matrix with $B_{ij} = 0$ except for

$$\begin{aligned} &B_{3,1} = B_{6,1} = B_{9,1} = 1 \\ &B_{1,2} = B_{4,2} = B_{7,2} = B_{10,2} = 1 \\ &B_{2,3} = B_{5,3} = B_{8,3} = 1. \end{aligned}$$

In other words, the published control signal is 3-dimensional, with control input u_1 simultaneously actuating systems $3, 6, 9$, u_2 actuating systems $1, 4, 7, 10$ and u_3 actuating systems $2, 5$ and 8. We wish to regulate the sum of the states $\sum_{i=1}^{10} x_i$ to 0, hence we take Q to be the 10×10 all-ones matrix and $R = I_3$ in (2.5). We set the privacy parameters to $\delta = 0.05$ and $\epsilon = \ln(3)$, and $\rho_i = 1$ for $1 \leq i \leq 10$.

To design the differentially private LQG controller with signal aggregation for the stationary problem, we compute the matrix L of Proposition 2 and solve the optimization problem (2.23a)-(2.23d). Following the methodology discussed at the end of Section 2.2.4, we find that one can take the matrix D to be a 4×10 matrix at the matrix factorization stage (2.18). The corresponding steady-state cost J_∞ is found to be 1.37, whereas it is 2.17 for the input perturbation mechanism (i.e., with $D = I_{10}$). Hence, signal aggregation results in a significant improvement. Figure 2.7 shows a comparison of the cost J_∞ for this problem, with the two architectures, for different values of ϵ. Finally, Figure 2.8 illustrates the sample paths obtained under closed-loop control with the differentially private controllers. We see in particular on Figure 2.8b that the two-stage architecture provides a much better transient behavior for the regulated average trajectory (or sum of trajectories) compared to the input perturbation architecture, in addition to a better steady-state performance.

2.4 Conclusion

In this chapter, we consider the Kalman filtering and LQG optimal control problems under a differential privacy constraint. We propose an architecture combining an input stage aggre-

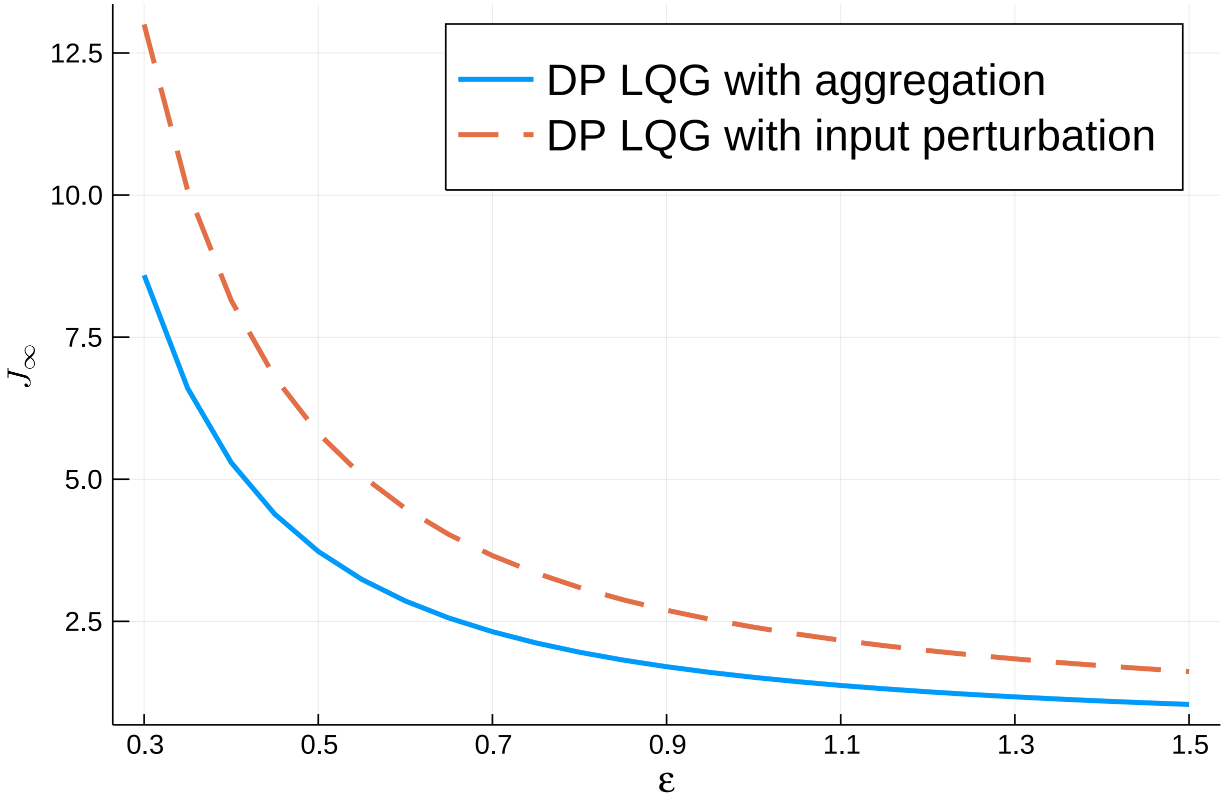
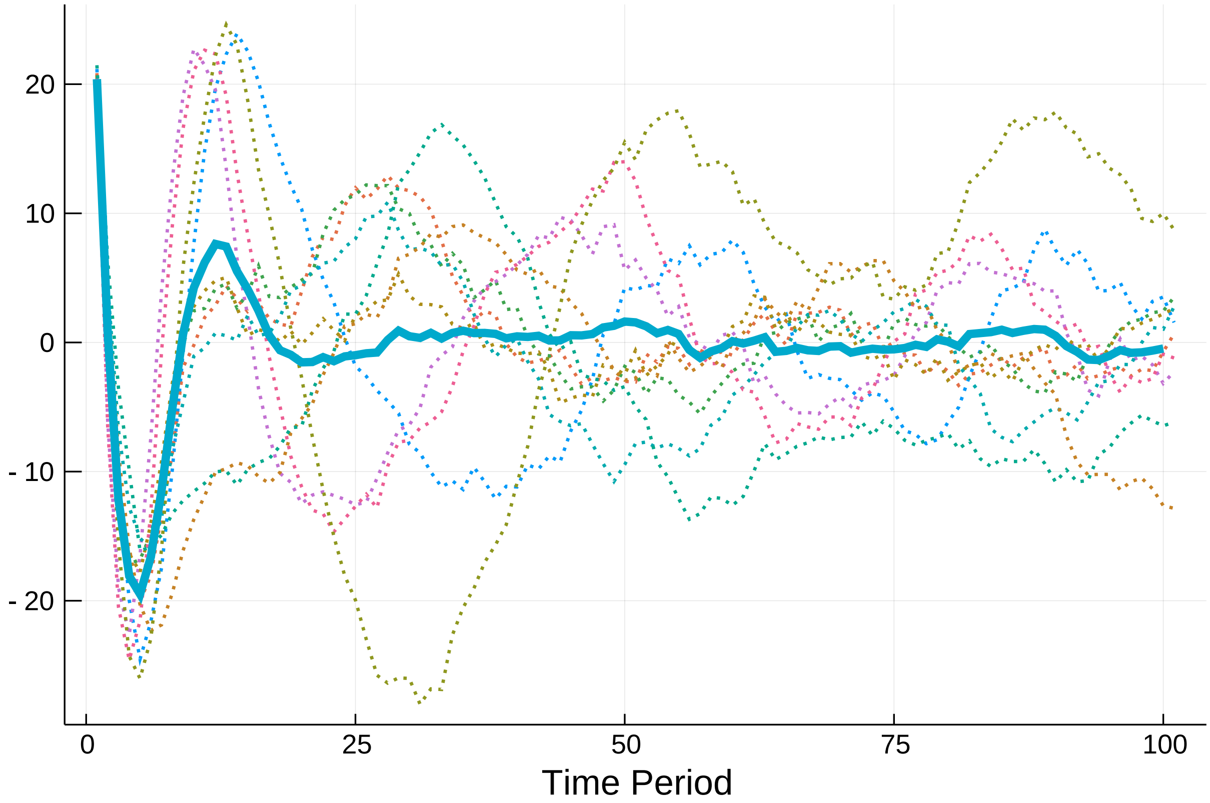
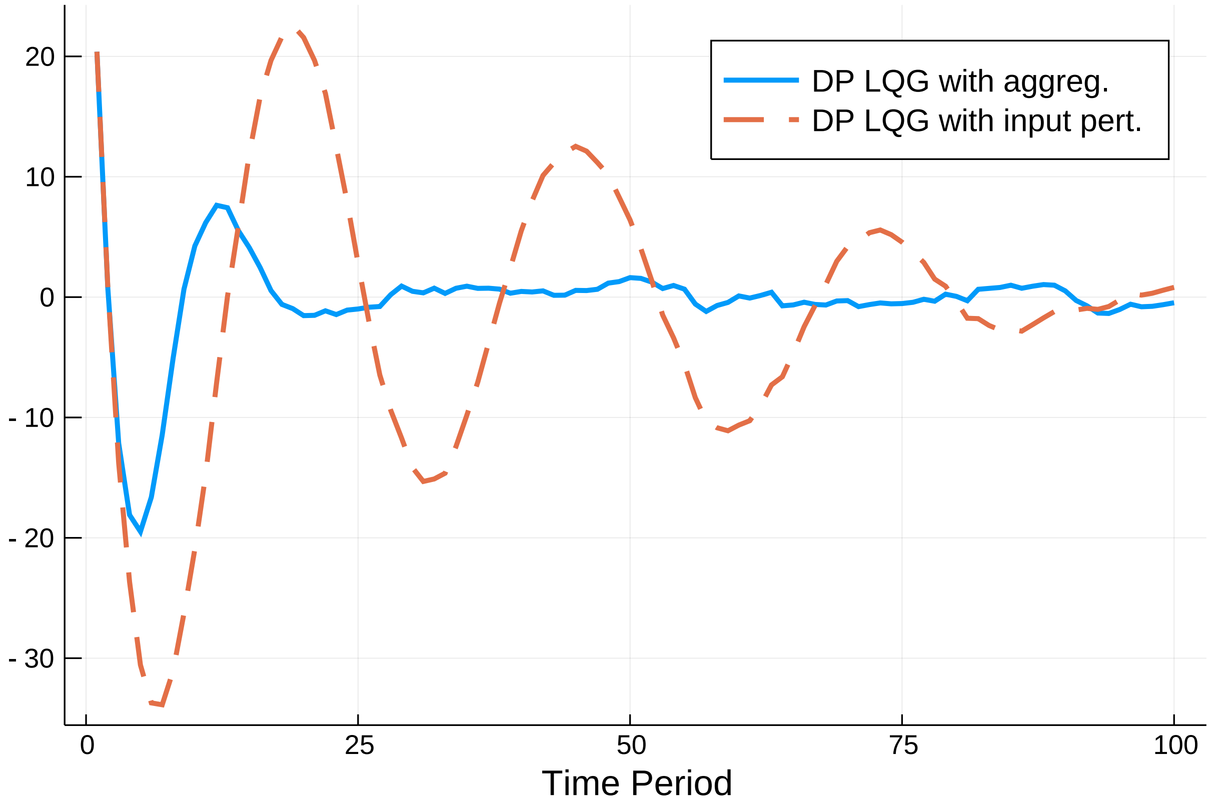


Figure 2.7 : Optimal steady-state quadratic cost J_∞ as a function of the privacy parameter ϵ, for the architecture with signal aggregation and for the input perturbation mechanism. Here $\delta = 0.05$



(a) : Sample paths for the individual trajectories and the regulated average trajectory ($\frac{1}{10}\sum_{i=1}^{10} x_{i,t}$, thick solid line) using the LQG controller with signal aggregation



(b) : Sample paths of the regulated average trajectories for the LQG controllers with input perturbation and signal aggregation gating the individual signals appropriately, the Gaussian mechanism to enforce differential privacy and a Kalman filter to reconstruct the desired estimate. Optimizing the parameters of this architecture can be recast as an SDP. Examples illustrate the performance improvements compared to the input perturbation mechanism, which adds noise directly on the individual signals. The methodology is then adapted to propose a similar two-stage architecture for an LQG control problem, where the goal is to compute a shared control broadcasted to the agent population. In the next chapter, we consider scenarios in which statistical properties of the disturbances $w_{i,t}$ and $v_{i,t}$ are unknown.

CHAPTER 3 DIFFERENTIALLY PRIVATE INTERVAL OBSERVER DESIGN WITH BOUNDED PERTURBATION

In this chapter, we formulate in Section 3.1 a problem of designing privacy-preserving interval observers for multi-agents systems in which the signals of individual participants are modeled using uncertain linear time-invariant systems with bounded disturbances. In Section 3.2, we design a bounded privacy-preserving noise that is added to each agent's data following an input perturbation mechanism (described in Subsection 1.3.2). The differentially private data is sent to an observer, which releases publicly lower and upper bounds for an aggregation of the individual agent systems' states.

3.1 Background

3.1.1 System Model

Let $\left\{y_t^{(i)}, 0 \leq t \leq T\right\}$, $1 \leq i \leq n$ be a set of n measured and privacy-sensitive scalar signals, i.e., $y_t^{(i)} \in \mathbb{R}$, originating from n distinct agents. The case $T = \infty$ is also of interest. A mathematical model for this data is publicly known and consists of a set of linear systems with n individual (vector-valued) states that are coupled and correspond to the n measured signals

$$\begin{aligned} x_{t+1}^{(i)} &= A^{(i)} x_t^{(i)} + \sum_{j \neq i} A^{(i,j)} x_t^{(j)} + w_t^{(i)}, 0 \leq t \leq T-1, \\ y_t^{(i)} &= C^{(i)} x_t^{(i)} + v_t^{(i)}, \quad 0 \leq t \leq T, \end{aligned} \tag{3.1}$$

for $i = 1, ..., n$, where $x_t^{(i)} \in \mathbb{R}^{p_i}$ represents the state vector of the agent i, $w^{(i)} : \mathbb{Z}_+ \to \mathbb{R}^{p_i}$ stands for an *unknown* input in $\mathcal{L}_\infty^{p_i}$, $v^{(i)} : \mathbb{Z}_+ \to \mathbb{R}$ is an *unknown* measurement noise in $\mathcal{L}_\infty$, and $A^{(i)}, A^{(i,j)} \in \mathbb{R}^{p_i \times p_j}, C^{(i)} \in \mathbb{R}^{1 \times p_i}$ are known constant matrices. The matrices $A^{(i,j)}$ represent coupling matrices that capture the influence of the other agents on the agent i. One can express the dynamics of the global system formed by the n agents as follows

$$\begin{aligned} x_{t+1} &= A x_t + w_t, \quad t = 0, 1, ..., T-1, \\ y_t &= C x_t + v_t, \quad t = 0, 1, ..., T, \end{aligned} \tag{3.2}$$

with

$$\begin{aligned}
x_t &= \begin{bmatrix} x_t^{(1)\mathrm{T}} & \dots & x_t^{(n)\mathrm{T}} \end{bmatrix}^{\mathrm{T}},\ y_t = \begin{bmatrix} y_t^{(1)} & \dots & y_t^{(n)} \end{bmatrix}^{\mathrm{T}},\\
w_t &= \begin{bmatrix} w_t^{(1)\mathrm{T}} & \dots & w_t^{(n)\mathrm{T}} \end{bmatrix}^{\mathrm{T}},\ v_t = \begin{bmatrix} v_t^{(1)} & \dots & v_t^{(n)} \end{bmatrix}^{\mathrm{T}},\\
C &= \mathrm{diag}(C^{(1)}, \dots, C^{(n)}),\\
A &= \begin{bmatrix} A^{(1)} & A^{(1,2)} & \dots & A^{(1,n)} \\ A^{(2,1)} & A^{(2)} & \dots & A^{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n,1)} & A^{(n,2)} & \dots & A^{(n)} \end{bmatrix},
\end{aligned}$$

where diag() denotes a block-diagonal matrix. Denote $p = \sum_{i=1}^{n} p_i$. Throughout this chapter, we assume that the initial condition x_0 is *unknown* but satisfy the bounds $\underline{x}_0 \leq x_0 \leq \overline{x}_0$, where $\underline{x}_0, \overline{x}_0 \in \mathbb{R}^p$ are given. A data aggregator aims at releasing an estimate $\hat{z}_t$ of a linear combination $z_t = \Phi x_t = \sum_{i=1}^{n} \Phi_i x_{i,t}$ of the individual states, where the matrices $\Phi_i \geq 0$ are given, by using the data y. To reach this goal, we make the following assumption, stating that the noise signals are bounded.

Assumption 1 *Two functions $\underline{w}$, $\overline{w}$: $\mathbb{Z}_+ \to \mathcal{L}_\infty^p$ and two functions $\underline{v}$, $\overline{v}$: $\mathbb{Z}_+ \to \mathcal{L}_\infty^n$ are given such that*

$$\underline{w}_t \leq w_t \leq \overline{w}_t \text{ and } \underline{v}_t \leq v_t \leq \overline{v}_t,\ \forall t \geq 0.$$

3.1.2 Interval Observer Design and Problem Statement

The following assumption, which is common in the interval observer literature, is needed.

Assumption 2 *There exists a matrix $L \in \mathbb{R}^{m \times n}$ such that the matrix $A - LC$ is Schur stable and nonnegative.*

The equations of an interval observer take the form

$$\begin{aligned}
\underline{x}_{t+1} &= (A - LC)\underline{x}_t + Ly_t + \underline{w}_t - L^+\overline{v}_t + L^-\underline{v}_t,\\
\overline{x}_{t+1} &= (A - LC)\overline{x}_t + Ly_t + \overline{w}_t - L^+\underline{v}_t + L^-\overline{v}_t,
\end{aligned} \tag{3.3}$$

where $\underline{x}_t \in \mathbb{R}^m$ and $\overline{x}_t \in \mathbb{R}^m$ stand for the lower and the upper interval estimates of the system state x_t.

Theorem 8 [70] *Let Assumptions 1 and 2 be satisfied. Then, we get for (3.2)*

$$\underline{x}_t \leq x_t \leq \overline{x}_t, \forall t \geq 0. \tag{3.4}$$

The data $\underline{x}_0, \overline{x}_0, \underline{w}, \overline{w}, \underline{v}, \overline{v}, A, C, \Phi_i$ is assumed to be public information. By using the publicly released estimates $\Phi\underline{x}_t$ and $\Phi\overline{x}_t$, it might be possible to deduce new information about the data $\{y_t\}_{t\geq 0}$ for example by using *linkage attacks*, where someone can combine the newly published information with other available data to make new inferences about specific individuals [37]. Hence, we aim to ensure that the publicly released estimates $\underline{z}_t$ and $\overline{z}_t$, which are lower and upper bounds on z_t, also guarantee differential privacy for each agent's data, as defined next.

Recall that a differentially private version of the interval observer (3.3) must provide estimates that are not too sensitive to some variations in the participating agents' signals y. Let $\mathcal{D}$ denote the space of measured signals $t \mapsto y_t$. We consider here the following adjacency relation

$$\text{Adj}(y, \tilde{y}) \text{ iff } \|y - \tilde{y}\|_1 \leq \rho, \tag{3.5}$$

with $\rho \in \mathbb{R}_+$ given. Such an interpretation of adjacent datasets implies that a single participant contributes additively to possibly each $y_t^{(i)}$ in a way that its overall impact on the dataset y is bounded in ℓ_1-norm by ρ. In the special case $T = 0$, we have a single vector $y_0 \in \mathbb{R}^n$, in which case $\mathcal{D} = \mathbb{R}^n$ and the norm in the adjacency relation (3.5) reduces to the 1-norm $|\cdot|_1$. Here we aim to publish estimates $\underline{z}$ and $\overline{z}$ of $z = \Phi x$ that are accurate and respect Definition 1 for given values of ϵ and δ. Smaller values of ϵ and δ give stronger privacy guarantees. We consider an input perturbation architecture [52], where each individual participant perturbs their signal $y^{(i)}$ by adding privacy-preserving white noise in order to render these signals differentially private, before sending them to the data aggregator implementing an observer. In this case, because the signals received by the aggregator are already differentially private, so are the results of the observer's computations, since differential privacy guarantees are preserved by post-processing [52, Theorem 1].

3.2 Design of the Differentially Private Bounded Noise

Normally, to produce a differentially private vector or signal by using additive white noise, the distribution of each noise sample is taken to be Laplace or Gaussian [48], hence has unbounded support. However, to design an interval observer we need to know lower and upper bounds for the noise signals. Therefore, a scheme adding bounded noise is required here.

A bounded Laplace mechanism is proposed in [111] but requires to know a priori lower and upper bounds on the signals $y^{(i)}$. Here we do not assume knowledge of such bounds. Instead we build on a noise distribution considered in [102] for the scalar case in which one publishes a single real number in a differentially private way. Here we consider the vector- and signal-valued cases.

Let the privacy parameters be $\epsilon > 0$ and $0 < \delta < \frac{1}{2}$. Define for any integer m the probability density function of a truncated Laplace distribution as follows

$$p(x) = \begin{cases} \phi_m \, e^{\frac{-|x|}{\lambda}} & \text{if } x \in [-a_m \ a_m], \\ 0 & \text{otherwise}, \end{cases} \tag{3.6}$$

with

$$\lambda = \frac{\rho}{\epsilon}, \; a_m = \frac{\rho}{\epsilon} \ln \left(1 + e^{\epsilon} \frac{m \left(1 - e^{-\epsilon/m}\right)}{2\delta} \right), \tag{3.7}$$
$$\phi_m = \frac{1}{2\lambda \left(1 - e^{-\frac{a_m}{\lambda}}\right)}.$$

For $m = 1$, we recover the distribution of [102]. Moreover, since $m\left(1 - e^{-\epsilon/m}\right) \leq \epsilon$, and also

$$\lim_{m \to \infty} m \left(1 - e^{-\epsilon/m}\right) = \epsilon,$$

we define

$$a_\infty = \frac{\rho}{\epsilon} \ln \left(1 + \frac{\epsilon \, e^{\epsilon}}{2\delta} \right) \tag{3.8}$$

and the corresponding distribution with support $[-a_\infty, a_\infty]$ through (3.6).

Theorem 9 *Let $\epsilon > 0$, $\frac{1}{2} > \delta > 0$. Publishing the sequence of n-dimensional vectors $\hat{y}_t = y_t + \zeta_t$, $0 \leq t \leq T$, where the coordinates of the noise vectors ζ_t are iid with probability distribution* (3.6) *for $m = n(T + 1)$, and the successive samples of ζ are also iid, is (ϵ, δ)-differentially private for the adjacency relation* (3.5)*. If $T = \infty$, we take $m = \infty$ and the support for the distribution of each noise component is defined by* (3.8)*.*

Proof. We prove the result for a single time period ($T = 0$), i.e., for the problem of publishing an n-dimensional vector $\hat{y} = y + \zeta$ so that $\hat{y} \in \mathbb{R}^n$ is differentially private. Hence, we suppress the time index in the rest of the proof. The extension to the publication of signals with $T > 0$ or even $T = \infty$ (i.e., "infinitely long vectors") follows from this result and [52, Lemma 2].

For each measurable set S in $\mathbb{R}^n$,

$$\mathbb{P}(\zeta \in S) = \int_{\mathbb{R}^n} \mathbf{1}_S(x) \prod_{i=1}^{n} p(x_i)\, dx_1 \dots dx_n,$$

where p is the pdf of any component of ζ, of the form (3.6), and $\mathbf{1}_{\{\cdot\}}$ represents the indicator function. The differential privacy property (1.1) is equivalent to

$$\sup_{S \in \mathcal{B}(\mathbb{R}^n)} \{\mathbb{P}(y + \zeta \in S) - e^{\epsilon}\mathbb{P}(\tilde{y} + \zeta \in S)\} \leq \delta, \tag{3.9}$$

for any adjacent vectors y and $\tilde{y}$, i.e., such that $|y - \tilde{y}|_1 \leq \rho$, where $\mathcal{B}(\mathbb{R}^n)$ represents the Borel σ-algebra. Now, for any set S and vector v, define the notation for the shifted set

$$S - v = \{x | x + v \in S\}.$$

We can rewrite the left-hand side of (3.9) as $\mathbb{P}(\zeta \in S - y) - e^{\epsilon}\mathbb{P}(\zeta \in S - \tilde{y})$, and then, renaming $S_1 := S - \tilde{y}$, $\mathbb{P}(\zeta \in S_1 - d) - e^{\epsilon}\mathbb{P}(\zeta \in S_1)$, with $d = y - \tilde{y}$. Since (3.9) must hold for all Borel sets S, the sets S_1 also consist of all Borel sets, and (3.9) can be rewritten equivalently (we renamed S_1 to S to simplify the notation)

$$\sup_{S \in \mathcal{B}(\mathbb{R}^n)} \{\mathbb{P}(\zeta \in S - d) - e^{\epsilon}\mathbb{P}(\zeta \in S)\} \leq \delta, \tag{3.10}$$

for any vector d such that $|d|_1 \leq \rho$. This places a condition on the distribution of ζ.

Let C_a be the hypercube $[-a, a]^n$ (here a denotes the support of the distribution (3.6), which is determined below, so we omit the subscript m from the notation for now). Since the support of the distribution of ζ is contained in C_a, the expression to be upper bounded by δ in (3.10) for all S and admissible d can be written

$$\begin{aligned} &\mathbb{P}(\zeta \in S - d) - e^{\epsilon}\mathbb{P}(\zeta \in S) \\ &= \mathbb{P}(\zeta \in (S - d) \cap C_a) - e^{\epsilon}\mathbb{P}(\zeta \in S \cap C_a). \end{aligned}$$

We now exploit the form of the Laplace distribution to work with a more convenient upper bound. We have

$$\begin{aligned}
&\mathbb{P}(\zeta \in (S-d) \cap C_a) \\
&= \phi^n \int_{\mathbb{R}^n} e^{-\frac{|x|_1}{\lambda}} \mathbf{1}_{S-d}(x)\, \mathbf{1}_{C_a}(x) dx \\
&= \phi^n \int_{\mathbb{R}^n} e^{-\frac{|x|_1}{\lambda}} \mathbf{1}_{S}(x+d)\, \mathbf{1}_{C_a+d}(x+d) dx \\
&= \phi^n \int_{\mathbb{R}^n} e^{-\frac{|z-d|_1}{\lambda}} \mathbf{1}_{S}(z)\, \mathbf{1}_{C_a+d}(z) dz,
\end{aligned}$$

using the fact that $\mathbf{1}_A(x) = \mathbf{1}_{A+d}(x+d)$ and the change of variable $z = x + d$. Since $|z-d|_1 \geq |z|_1 - |d|_1$, if $|d|_1 \leq \rho$, we get

$$\mathbb{P}(\zeta \in (S-d) \cap C_a) \leq e^{\frac{\rho}{\lambda}} \phi^n \int_{\mathbb{R}^n} e^{-\frac{|x|_1}{\lambda}} \mathbf{1}_{S \cap (C_a+d)}(x) dx.$$

If moreover we take $\frac{\rho}{\lambda} \leq \epsilon$, we get the upper bound

$$\begin{aligned}
&\mathbb{P}(\zeta \in S-d) - e^{\epsilon}\mathbb{P}(\zeta \in S) \\
&\leq e^{\epsilon} \phi^n \int_{\mathbb{R}^n} e^{-\frac{|x|_1}{\lambda}} \left\{ \mathbf{1}_{S \cap (C_a+d)}(x) - \mathbf{1}_{S \cap C_a}(x) \right\} dx.
\end{aligned}$$

From here on, we fix $\lambda = \rho/\epsilon$. Since the only points x that contribute something positive to the integral are those that are simultaneously in S, in $C_a + d$ and not in C_a, the integral is maximized for $S = (C_a + d) \setminus C_a$, and we get

$$\mathbb{P}(\zeta \in S-d) - e^{\epsilon}\mathbb{P}(\zeta \in S) \leq e^{\epsilon} F(d), \tag{3.11}$$

with

$$F(d) := \phi^n \int_{\mathbb{R}^n} e^{-\frac{|x|_1}{\lambda}} \mathbf{1}_{(C_a+d) \setminus C_a}(x)\, dx.$$

We now maximize the upper bound (3.11) over admissible d (i.e., such that $|d|_1 \leq \rho$) and obtain a condition under which this upper bound is less than δ.

By symmetry (see Figure 3.1), it is sufficient to consider the case $d = [d_1, \ldots, d_n]$ with $d_i \geq 0$ for all i and $\sum_i d_i \leq \rho$. Next, note from Figure 3.1 that $(C_a + d) \setminus C_a \subset \cup_{i=1}^n R_i$, where R_i is the hyperrectangle

$$\begin{aligned}
R_i = \{x \in \mathbb{R}^n \,|\, a \leq x_i \leq a + d_i, &\text{and} \\
&- a + d_j \leq x_j \leq a + d_j, \forall j \neq i\}.
\end{aligned}$$

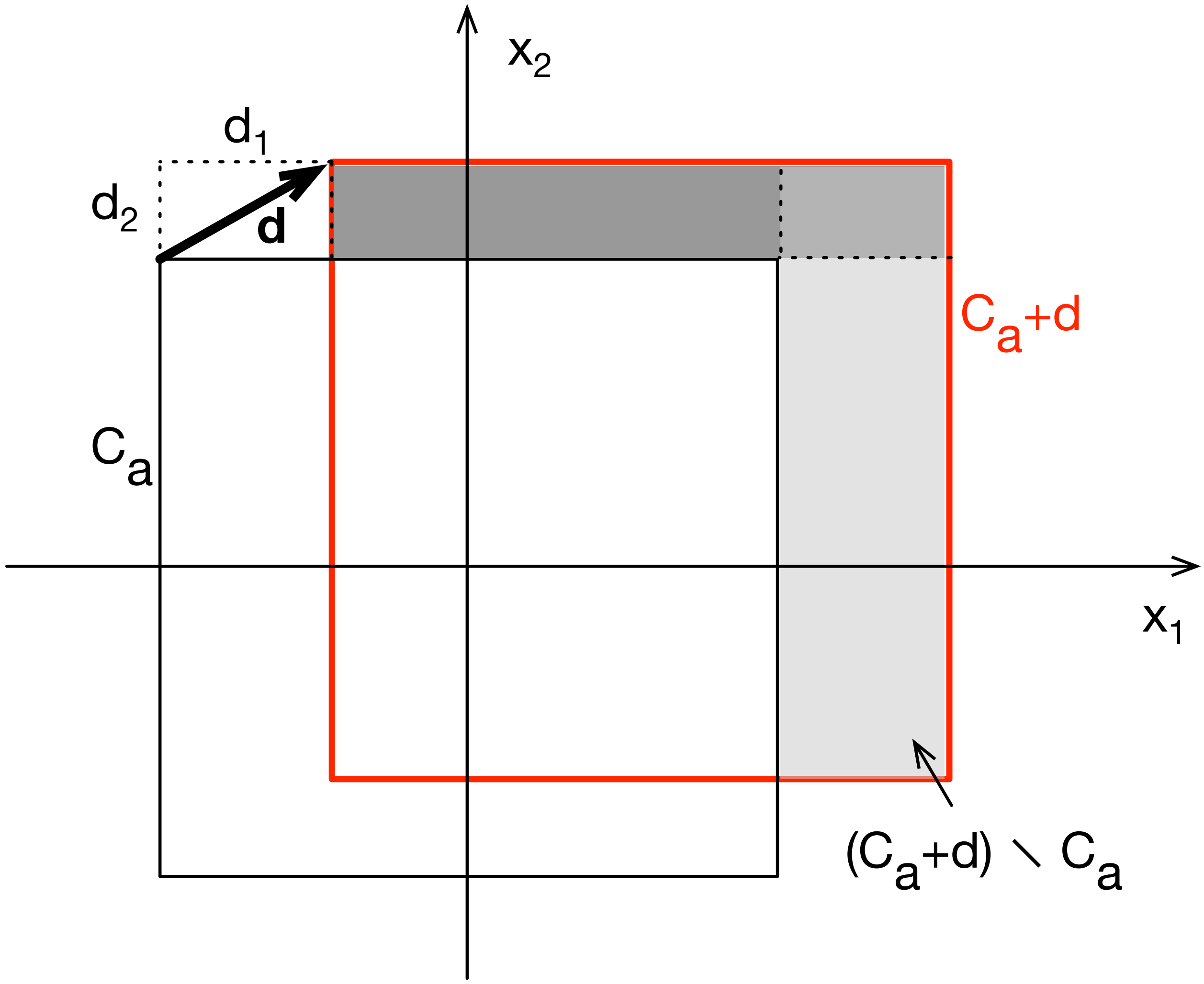


Figure 3.1 : Geometry for the (ϵ, δ)-differential privacy argument. The set $(C_a + d) \setminus C_a$ has been subdivided into n (here $n = 2$) rectangles, which intersect in the "top-right corner" of $C_a + d$

Therefore, we get the bound

$$F(d) \leq \sum_{i=1}^{n} \phi^n \int_{R_i} e^{-|x|_1/\lambda} dx.$$

Now

$$\phi^n \int_{R_i} e^{-|x|_1/\lambda} dx = \phi \int_{a}^{a+d_i} e^{-x_i/\lambda} dx_i \prod_{j \neq i} T_j,$$

where

$$T_j = \phi \int_{-a+d_j}^{a+d_j} e^{-|x_j|/\lambda} dx_j \leq \phi \int_{-a}^{a} e^{-|x_j|/\lambda} dx_j = 1,$$

because of the shape of the function $e^{-|x_j|/\lambda}$. Hence, we have

$$\begin{aligned} F(d) &\leq \sum_{i=1}^{n} \phi \int_{a}^{a+d_i} e^{-x_i/\lambda} dx_i \\ &= \phi \, \lambda \, e^{-a/\lambda} \sum_{i=1}^{n} \left(1 - e^{-d_i/\lambda}\right). \end{aligned}$$

Consider then the maximization problem

$$\begin{aligned} &\max_{d \in \mathbb{R}^n} \sum_{i=1}^{n} \left(1 - e^{-d_i/\lambda}\right) \\ &\text{s.t. } \; d_i \geq 0, 1 \leq i \leq n, \; \text{ and } \; \sum_{i=1}^{n} d_i \leq \rho. \end{aligned}$$

Since the objective is strictly concave in d and the constraints define a polytope, this problem has a unique maximizer. A straightfoward analysis, e.g., using the KKT conditions, shows that this maximizer is given by $d_i = \rho/n$ for all $1 \leq i \leq n$. Hence, we finally get

$$F(d) \leq \phi \, \lambda \, e^{-a/\lambda} \, n \left(1 - e^{-\frac{\rho}{n\lambda}}\right).$$

We now choose a to ensure that the upper bound on $e^{\epsilon} F(d)$ is always less than δ. We get the condition

$$\begin{aligned} e^{\epsilon} \frac{e^{-\frac{a}{\lambda}} n \left(1 - e^{-\rho/n\lambda}\right)}{1 - e^{-\frac{a}{\lambda}}} &\leq 2\delta \\ \text{hence, } e^{a/\lambda} &\geq 1 + e^{\epsilon} \frac{n \left(1 - e^{-\rho/n\lambda}\right)}{2\delta} \\ a &\geq \frac{\rho}{\epsilon} \ln \left(1 + e^{\epsilon} \frac{n \left(1 - e^{-\epsilon/n}\right)}{2\delta}\right) \end{aligned}$$

where we used $\lambda = \frac{\rho}{\epsilon}$ on the last line. This gives the distribution (3.6) with $m = n$, as stated in the Theorem for the case $T = 0$. □

When considering the design of a differentially private bounded noise, another possibility consists in designing an uniform noise mechanism, i.e., designing a bounded noise with the following probability density function

$$p(x) = \begin{cases} \frac{\delta}{\rho} & \text{if}\, x \in [-\frac{\rho}{2\delta}\ \frac{\rho}{2\delta}], \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

Corollary 1 *Let $\frac{1}{2} > \delta > 0$. Publishing the sequence of n-dimensional vectors $\hat{y}_t = y_t + \zeta_t$, $0 \le t \le T$, where the coordinates of the noise vectors ζ_t are iid with probability distribution* (3.12)*, and the successive samples of ζ are also iid, is $(0, \delta)$-differentially private for the adjacency relation* (3.5)*.*

Proof. The proof uses the exact argumentation of the proof of Theorem 9. Consider a noise with probability density function (3.6) with $m = \infty$. We prove that its limit is a noise with probability density function (3.12) when $\epsilon \to 0$. Denote $f(\epsilon) = \ln\left(1 + \frac{\epsilon\, e^{\epsilon}}{2\delta}\right), \forall \epsilon \ge 0$. We get

$$\begin{aligned} \lim_{\epsilon \to 0} a_\infty &= \rho \lim_{\epsilon \to 0} \frac{\ln\left(1 + \frac{\epsilon\, e^{\epsilon}}{2\delta}\right)}{\epsilon}, \\ &= \rho \lim_{\epsilon \to 0} \frac{f(\epsilon) - f(0)}{\epsilon - 0}, \\ &= \rho f'(0) = \frac{\rho}{2\delta}. \end{aligned} \tag{3.13}$$

Furthermore, denote $g(\epsilon) = e^{\frac{-\epsilon}{2\delta}}, \forall \epsilon \ge 0$. We get

$$\lim_{\epsilon \to 0} \phi_\infty = \frac{1}{2\rho} \lim_{\epsilon \to 0} \frac{\epsilon}{\left(1 - e^{-\frac{\epsilon a_\infty}{\rho}}\right)}.$$

We deduce from (3.13) that

$$\begin{aligned}\lim_{\epsilon\to 0}\phi_\infty &= \frac{1}{2\rho}\lim_{\epsilon\to 0}\frac{\epsilon}{\left(1-e^{-\frac{\epsilon}{2\delta}}\right)} = \frac{-1}{2\rho}\times\frac{1}{\lim_{\epsilon\to 0}\frac{1-e^{-\frac{\epsilon}{2\delta}}}{-\epsilon}}\\ &= \frac{-1}{2\rho}\times\frac{1}{\lim_{\epsilon\to 0}\frac{g(0)-g(\epsilon)}{0-\epsilon}} = \frac{-1}{2\rho}\times\frac{1}{g'(0)}\\ &= \frac{\delta}{\rho}.\end{aligned}$$

The result of Corollary 1 follows directly. □

When $m = +\infty$, Corollary 1 implies that the limit of the variance of the privacy-preserving noise ζ_t with probability distribution (3.6) as $\epsilon \to 0$ is

$$\begin{aligned}\lim_{\epsilon\to 0}\int_{-a_m}^{a_m} x^2 p(x)\,dx &= \int_{-\frac{\rho}{2\delta}}^{\frac{\rho}{2\delta}} x^2\frac{\delta}{\rho}\,dx\\ &= \frac{\rho^2}{12\delta^2}.\end{aligned} \tag{3.14}$$

Proposition 3 *Consider the privacy-preserving noise ζ_t with probability distribution* (3.6)*. As the privacy parameter $\delta \to 0$, $\zeta_t \sim \mathrm{Lap}(\frac{\rho}{\epsilon})$.*

Proof. Consider the probability distribution (3.6). We get

$$\begin{aligned}\lim_{\delta\to 0} a_m &= \lim_{\delta\to 0}\frac{\rho}{\epsilon}\ln\left(1+e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right)\\ &= +\infty,\end{aligned} \tag{3.15}$$

and

$$\begin{aligned}\lim_{\delta\to 0}\phi_m &= \lim_{\delta\to 0}\frac{1}{2\frac{\rho}{\epsilon}\left(1-e^{-\frac{a_m}{\frac{\rho}{\epsilon}}}\right)}\\ &= \frac{1}{2\frac{\rho}{\epsilon}}.\end{aligned} \tag{3.16}$$

This leads to the result of Proposition 3. □

Proposition 3 implies that as $\delta \to 0$, the bounded mechanism proposed in Theorem 9 is equivalent to the Laplace mechanism of Theorem 3.

Next, we calculate the variance of ζ_t in the general case. It can be computed as follows

$$\begin{aligned}
\mathrm{Var}(\zeta_t) &= \int_{-a_m}^{a_m} x^2 p(x)\,dx \\
&= \int_{-a_m}^{a_m} x^2 \phi_m e^{-\frac{|x|}{\lambda}}\,dx \\
&= 2\int_{0}^{a_m} x^2 \phi_m e^{-\frac{|x|}{\lambda}}\,dx \\
&= 2\lambda\phi_m\left(-a_m^2 e^{-\frac{a_m}{\lambda}} + 2\int_0^{a_m} x e^{-\frac{-x}{\lambda}}\,dx\right) \\
&= 2\lambda\phi_m\left(-a_m^2 e^{-\frac{a_m}{\lambda}} + 2\lambda(-a_m e^{-\frac{a_m}{\lambda}} + \lambda - \lambda e^{-\frac{a_m}{\lambda}})\right) \\
&= \frac{1}{1-e^{-\frac{a_m}{\lambda}}}\left(-a_m^2 e^{-\frac{a_m}{\lambda}} - 2\lambda a_m e^{-\frac{a_m}{\lambda}} + 2\lambda^2(1-e^{-\frac{a_m}{\lambda}})\right) \\
&= 2\lambda^2 - \frac{a_m^2 e^{-\frac{a_m}{\lambda}} + 2\lambda a_m e^{-\frac{a_m}{\lambda}}}{1-e^{-\frac{a_m}{\lambda}}} \\
&= 2\lambda^2 - \frac{a_m^2 + 2\lambda a_m}{e^{\frac{a_m}{\lambda}} - 1}. \qquad (3.17)
\end{aligned}$$

By replacing λ and a_m by their expressions of (3.7), we get

$$\begin{aligned}
\mathrm{Var}(\zeta_t) &= 2(\frac{\rho}{\epsilon})^2 - \frac{(\frac{\rho}{\epsilon})^2 \ln^2\left(1+e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right) + 2(\frac{\rho}{\epsilon})^2 \ln\left(1+e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right)}{e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}} \\
&= 2(\frac{\rho}{\epsilon})^2\left(1 - \frac{\frac{1}{2}\ln^2\left(1+e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right) + \ln\left(1+e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right)}{e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}}\right). \qquad (3.18)
\end{aligned}$$

3.3 Design of the Differentially Private Interval Observer

The privacy-preserving noise introduced in Theorem 9 is bounded and we get

$$\underline{\zeta}^{(i)} \le \zeta_t^{(i)} \le \overline{\zeta}^{(i)}, \quad \forall t \ge 0, \qquad (3.19)$$

with $\overline{\zeta}^{(i)} = -\underline{\zeta}^{(i)} = a_m$ for $i = 1, ..., n$, $m = n(T+1)$. In the sequel, denote $\overline{\zeta} = -\underline{\zeta} = \begin{bmatrix} a_m & \dots & a_m \end{bmatrix}^{\mathrm{T}}$ and $\zeta_t = \begin{bmatrix} \zeta_t^{(1)} & \dots & \zeta_t^{(n)} \end{bmatrix}^{\mathrm{T}}$. When the data aggregator receives the differentially private signal $\hat{y}$, it can design an interval observer and publish lower and upper estimates $\underline{\hat{x}}$ and $\hat{\overline{x}}$ for the state x. Releasing $\Phi\underline{\hat{x}}$ and $\Phi\hat{\overline{x}}$ preserves (ϵ, δ)-differentially privacy for the data y by the resilience to post-processing property.

The equations of the interval estimator can be written as follows

$$\begin{aligned}
\underline{\hat{x}}_{t+1} &= (A - LC)\underline{\hat{x}}_t + L\hat{y}_t + \underline{w}_t - L^+(\overline{v}_t + \overline{\zeta}) + L^-(\underline{v}_t + \underline{\zeta}), \quad \underline{\hat{x}}_0 = \underline{x}_0, \\
\hat{\overline{x}}_{t+1} &= (A - LC)\hat{\overline{x}}_t + L\hat{y}_t + \overline{w}_t - L^+(\underline{v}_t + \underline{\zeta}) + L^-(\overline{v}_t + \overline{\zeta}), \quad \hat{\overline{x}}_0 = \overline{x}_0. \qquad (3.20)
\end{aligned}$$

Theorem 10 *Let Assumptions 1 and 2 be satisfied. Then, we get for* (3.2)

$$\underline{\hat{x}}_t \leq x_t \leq \hat{\overline{x}}_t, \;\; \forall t \geq 0. \qquad (3.21)$$

Furthermore, we get $\underline{\hat{e}}, \hat{\overline{e}} \in \mathcal{L}_\infty^p$ *for the error signals* $\underline{\hat{e}}_t := x_t - \underline{\hat{x}}_t$, $\hat{\overline{e}}_t := \hat{\overline{x}}_t - x_t$.

Note that the differentially private signals $\underline{\hat{x}}, \hat{\overline{x}}$, while random, still maintain the order relation (3.21) of an interval observer for each trajectory.

Proof. The errors' dynamics satisfy the equations

$$\begin{aligned}
\underline{\hat{e}}_{t+1} =& (A - LC)\underline{\hat{e}}_t + \sum_{i=1}^{i=2} \underline{g}_i, \\
\hat{\overline{e}}_{t+1} =& (A - LC)\hat{\overline{e}}_t + \sum_{i=1}^{i=2} \overline{g}_i, \qquad (3.22)
\end{aligned}$$

where

$$\begin{aligned}
\underline{g}_1 =& w_t - \underline{w}_t, \; \overline{g}_1 = \overline{w}_t - w_t, \\
\underline{g}_2 =& L^+(\overline{v}_t + \overline{\zeta}) - L^-(\underline{v}_t + \underline{\zeta}) - L(v_t + \zeta), \\
\overline{g}_2 =& L(v_t + \zeta) - (L^+(\underline{v}_t + \underline{\zeta}) - L^-(\overline{v}_t + \overline{\zeta})).
\end{aligned}$$

When Assumption 1 is satisfied, we deduce by using Lemma 2 that the signals $\{\underline{g}_i, \overline{g}_i, 1 \leq i \leq 2\}$ are nonnegative. Consequently, when Assumption 2 holds, we deduce by applying Lemma 1 that $\underline{\hat{e}}_t \geq 0, \hat{\overline{e}}_t \geq 0$ since $\underline{\hat{e}}_0 \geq 0$ and $\hat{\overline{e}}_0 \geq 0$ (the system (3.22) is cooperative). Consequently, the order relation $\underline{\hat{x}}_t \leq x_t \leq \hat{\overline{x}}_t$ is satisfied for all $t \geq 0$. Since the system (3.22) is linear, we deduce that the global asymptotic stability (GAS) property of system (3.22) for $\{\underline{g}_i\}_{i=1}^{i=2} \equiv 0, \{\overline{g}_i\}_{i=1}^{i=2} \equiv 0$ implies its input-to-state stability (ISS) [97]. We conclude that $\underline{\hat{e}}, \hat{\overline{e}} \in \mathcal{L}_\infty^p$. □

The differentially private interval estimation developed in this section is summarized in Figure 3.2.

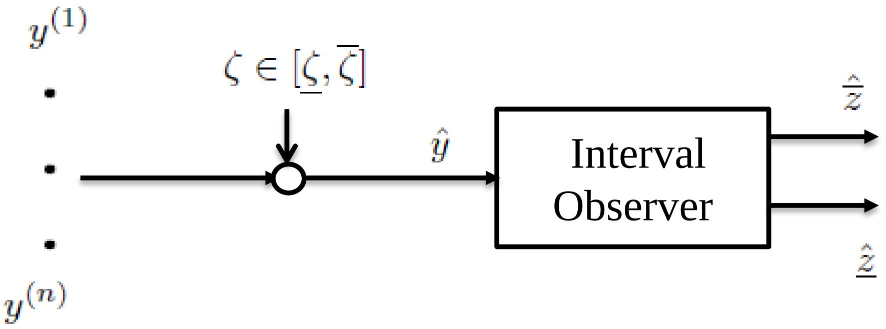


Figure 3.2 : Differentially private interval observer

Assumption 2 can be restrictive since it requires the matrix $(A - LC)$ to be not only Schur stable but also nonnegative. It can be relaxed by using a change of coordinates if the pair (A, C) is observable [Lemma 1] [96]. The following assumption is satisfied when the pair (A, C) is observable.

Assumption 3 *There exists a matrix $P \in \mathbb{R}^{p\times p}$ such that $P(A - LC)P^{-1}$ is Schur stable and nonnegative.*

Using a change of coordinates $s = Px$, equations of the interval estimator can be written as follows

$$\begin{aligned}
\hat{\underline{s}}_{t+1} &= P(A-LC)P^{-1}\hat{\underline{s}}_t + PL\hat{y}_t + P^+\underline{w}_t - P^-\overline{w}_t - P^+[L^+(\overline{v}_t+\overline{\zeta}) - L^-(\underline{v}_t+\underline{\zeta})]+ \\
&\quad P^-[L^+(\underline{v}_t+\underline{\zeta}) - L^-(\overline{v}_t+\overline{\zeta})], \quad \hat{\underline{s}}_0 = P^+\underline{x}_0 - P^-\overline{x}_0, \\
\hat{\overline{s}}_{t+1} &= P(A-LC)P^{-1}\hat{\overline{s}}_t + PL\hat{y}_t + P^+\overline{w}_t - P^-\underline{w}_t - P^+[L^+(\underline{v}_t+\underline{\zeta}) - L^-(\overline{v}_t+\overline{\zeta})]+ \\
&\quad P^-[L^+(\overline{v}_t+\overline{\zeta}) - L^-(\underline{v}_t+\underline{\zeta})], \quad \hat{\overline{s}}_0 = P^+\overline{x}_0 - P^-\underline{x}_0, \qquad (3.23) \\
\hat{\underline{x}}_t &= (P^{-1})^+\hat{\underline{s}}_t - (P^{-1})^-\hat{\overline{s}}_t, \\
\hat{\overline{x}}_t &= (P^{-1})^+\hat{\overline{s}}_t - (P^{-1})^-\hat{\underline{s}}_t.
\end{aligned}$$

Theorem 11 *Let Assumptions 1 and 3 be satisfied. Then, we get for* (3.2) *and* (3.23)

$$\hat{\underline{x}}_t \leq x_t \leq \hat{\overline{x}}_t, \quad \forall t \geq 0. \qquad (3.24)$$

Moreover, the error signals $\hat{\underline{e}}_t := x_t - \hat{\underline{x}}_t$, $\hat{\overline{e}}_t := \hat{\overline{x}}_t - x_t$ satisfy $\hat{\underline{e}}, \hat{\overline{e}} \in \mathcal{L}_\infty^p$.

Proof. Denote the estimation errors $\underline{\hat{\eta}}_t := s_t - \underline{\hat{s}}_t$ and $\hat{\overline{\eta}}_t := \hat{\overline{s}}_t - s_t$. Their dynamics satisfy the equations

$$\begin{aligned}\underline{\hat{\eta}}_{t+1} =& P(A - LC)P^{-1}\underline{\hat{\eta}}_t + \sum_{i=1}^{i=2} \underline{g}_i,\\ \hat{\overline{e}}_{t+1} =& (A - LC)\hat{\overline{e}}_t + \sum_{i=1}^{i=2} \overline{g}_i,\end{aligned} \tag{3.25}$$

with

$$\begin{aligned}\underline{g}_1 =& Pw_t - (P^+\underline{w}_t - P^-\overline{w}_t,\\ \overline{g}_1 =& P^+\overline{w}_t - P^-\underline{w}_t - Pw_t,\\ \underline{g}_2 =& P^+[L^+(\overline{v}_t + \overline{\zeta}) - L^-(\underline{v}_t + \underline{\zeta})] - P^-[L^+(\underline{v}_t + \underline{\zeta}) - L^-(\overline{v}_t + \overline{\zeta})] - PL(v_t + \zeta),\\ \overline{g}_2 =& PL(v_t + \zeta) - P^+[(L^+(\underline{v}_t + \underline{\zeta}) - L^-(\overline{v}_t + \overline{\zeta}))] + P^-[(L^+(\overline{v}_t + \overline{\zeta}) - L^-(\underline{v}_t + \underline{\zeta}))].\end{aligned}$$

One can use the exact argumentation used for the proof of Theorem 10 to deduce that $\underline{\hat{\eta}}_t \geq 0, \hat{\overline{\eta}}_t \geq 0$ and hence, the order relation $\underline{\hat{s}}_t \leq s_t \leq \hat{\overline{s}}_t$ is satisfied for all $t \geq 0$. Since $x = P^{-1}s$ by definition, we get (3.24) by applying Lemma 2. Again, the proof argument of Theorem 10 can be used to deduce that $\underline{\hat{\eta}}, \hat{\overline{\eta}} \in \mathcal{L}_\infty^p$ and then, we get $\underline{\hat{e}}, \hat{\overline{e}} \in \mathcal{L}_\infty^p$. □

3.4 Simulations

Consider an abstract scenario involving $n = 7$ subsystems, whose dynamics are governed by (3.1) with

$$\begin{aligned}
A^{(i)} &= \begin{bmatrix} -0.85 & 0 \\ 0 & 0.5 \end{bmatrix}, \forall\, 1 \leq i \leq 5, \\
A^{(i)} &= \begin{bmatrix} 0.3 & 0 \\ 0 & 0.9 \end{bmatrix}, \forall\, 6 \leq i \leq 7, \\
A^{(i,i+1)} &= \begin{bmatrix} 0.5 & 0 \\ -0.2 & 0 \end{bmatrix}, \forall\, 1 \leq i \leq 5, \\
A^{(i,i+1)} &= \begin{bmatrix} 0.4 & 0.1 \\ 0 & -0.7 \end{bmatrix}, \forall\, 6 \leq i \leq 7, \\
A^{(i,j)} &= 0_{2\times 2}, \forall\, j < i,\ i = 1, \ldots, 7,\ j = 1, \ldots, 7, \\
A^{(i,j)} &= 0_{2\times 2}, \forall\, i + 1 < j,\ i = 1, \ldots, 7,\ j = 1, \ldots, 7, \\
C^{(i)} &= \begin{bmatrix} 1 & 1 \end{bmatrix}, \forall\, 1 \leq i \leq 7.
\end{aligned}$$

The process noise $w_t^{(i)}$ and the measurement noise $v_t^{(i)}$ of each firm i are iid uniform random variables in the interval $[0, W]$ and $[0, V]$ respectively, with $W = \mathbf{1}_{14}^{\mathrm{T}}$ and $V = \mathbf{1}_{7}^{\mathrm{T}}$. Therefore, we have $\underline{w}_t^{(i)} = 0$, $\underline{v}_t^{(i)} = 0$, $\overline{w}_t^{(i)} = W$ and $\overline{v}_t^{(i)} = V$. The initial conditions of the state of each subsystem i are $x_0^{(i)} = 200 \times \mathbf{1}_{14}^{\mathrm{T}}$, and for the design of the observer we assume known the bounds on initial conditions $\underline{x}_0^{(i)} = (200 - \sigma) \times \mathbf{1}_{14}^{\mathrm{T}}$, $\overline{x}_0^{(i)} = (200 + \sigma) \times \mathbf{1}_{14}^{\mathrm{T}}$, with $\sigma = 15$. We take $T = \infty$.

The matrix $\Phi = \mathbf{1}_7$ for the data aggregator. We consider the case in which the data aggregator needs to provide privacy guarantees for the subsystems (each data $y^{(i)}$ is highly sensitive). Consider the adjacency relation (3.5) with $\rho = 1$. We set the privacy parameters to $\epsilon = 0.3$

and $\delta = 0.1$. We select the interval observer gain as follows

$$L = \begin{bmatrix} -0.8560 & 0.4830 & -0.0150 & -0.0150 & -0.0150 & 0 & 0 \\ -0.0028 & -0.2021 & -0.0036 & -0.0036 & -0.0036 & 0 & 0 \\ -0.0034 & -0.8470 & -0.0008 & -0.0008 & -0.0008 & 0 & 0 \\ -0.0016 & -0.0003 & -0.0013 & -0.0013 & -0.0013 & 0 & 0 \\ -0.0113 & -0.0067 & -0.8579 & -0.0132 & -0.0132 & 0 & 0 \\ -0.0037 & -0.0023 & -0.0032 & -0.0044 & -0.0044 & 0 & 0 \\ -0.0113 & -0.0067 & -0.0132 & -0.8579 & -0.0132 & 0 & 0 \\ -0.0037 & -0.0023 & -0.0044 & -0.0032 & -0.0044 & 0 & 0 \\ -0.0113 & -0.0067 & -0.0132 & -0.0132 & -0.8579 & 0 & 0 \\ -0.0037 & -0.0023 & -0.0044 & -0.0044 & -0.0032 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0811 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2115 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0811 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2115 \end{bmatrix},$$

to satisfy Assumption 2. To provide differential privacy guarantees for each firm's data, we compute differentially private interval estimates by applying Theorem 10. Figure 3.3 shows the difference between the differentially private observed bounds $\hat{\underline{z}}, \hat{\overline{z}}$ and bounds provided by standard non private interval observers obtained from Theorem 8.

3.5 Discussion About A Two-stage Architecture For Differentially Private Interval Estimation

In this section, we aim to illustrate that the performance of the architecture presented in Section 3.2 can be significantly improved by suitably aggregating the measurements data $\{y_t^{(i)}\}_{i=1}^{i=n}$ before adding the privacy preserving noise. Such a two-stage architecture has been developed in Chapter 2 for Kalman filtering and LQG control, but here we consider an interval estimation framework.

Consider the scalar case of the system (3.1) with $A^{(i)} = a$ and $C^{(i)} = c$ and assume that the individual states are independent, i.e., $A^{(i,j)} = 0$ for $i \neq j$. A data aggregator aims to release an estimate of the sum of the individual states, i.e., $\Phi_i = 1$ and $z_t = \sum_{i=1}^{i=n} x_t^{(i)}$. In the sequel, we fix the adjacency relation (3.5) parameter $\rho^{(i)} = \rho$ for $i = 1, ..., n$.

First, let us use the input perturbation architecture of Figure 3.2. Denote the estimation errors $\underline{\eta}_t = z_t - \hat{\underline{z}}_t$ and $\overline{\eta}_t = \hat{\overline{z}}_t - z_t$, where $\hat{\underline{z}}_t$ and $\hat{\overline{z}}_t$ are the released differentially private

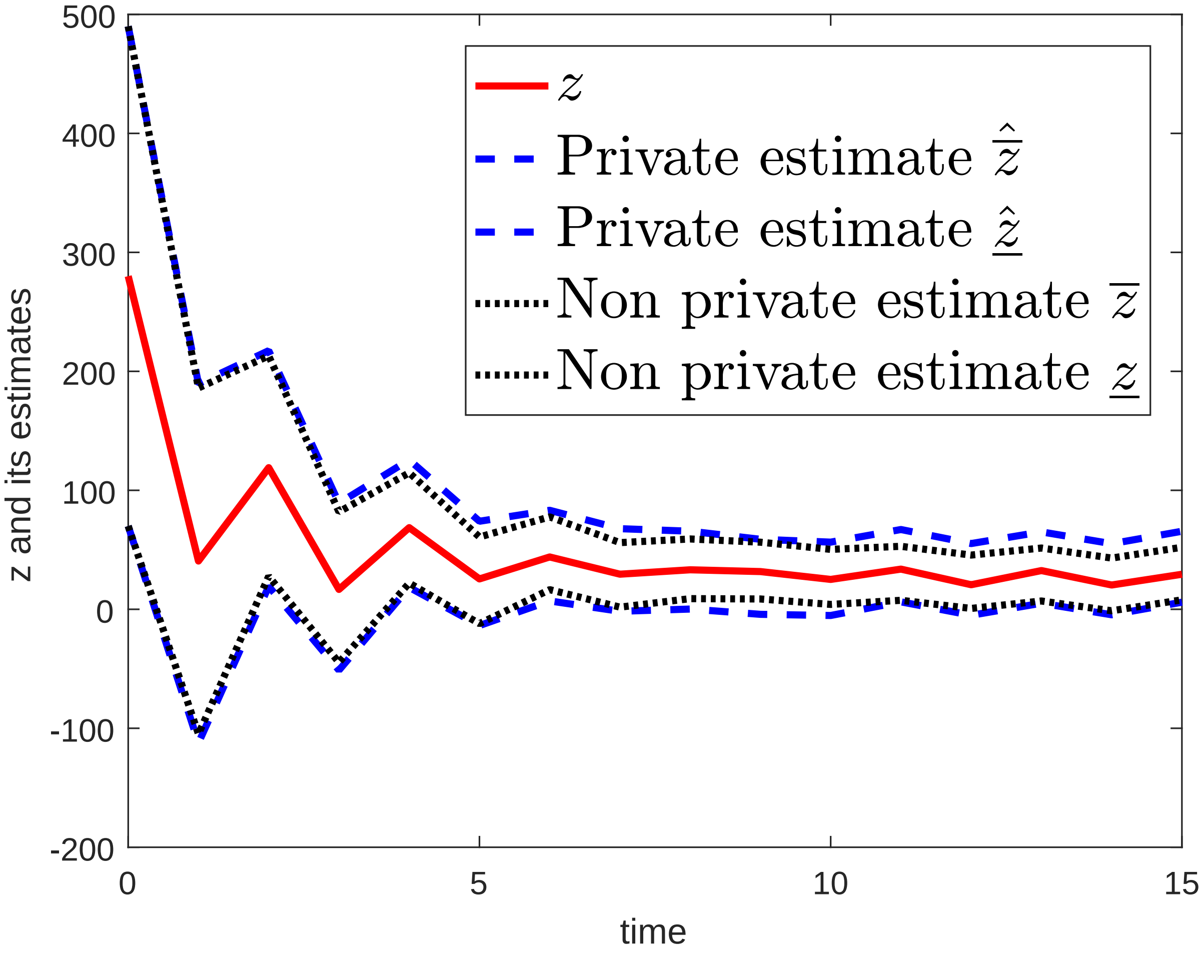


Figure 3.3 : Evolution of $z = \sum_{i=1}^{7} \Phi_i x_i$, the differentially private observer bounds $\sum_{i=1}^{7} \Phi_i \hat{\underline{x}}_i$ and $\sum_{i=1}^{7} \Phi_i \hat{\overline{x}}_i$ and the non private standard observer bounds

estimates of z_t in case of input perturbation. Denote $\hat{y}_t^{(i)} = y_t^{(i)} + \zeta_t^{(i)}$, the differentially private signal released after applying Theorem 9. Assume that there exists an observer gain $l \geq 0$ such that Assumption 2 is satisfied. It can be inferred from Theorem 10 that $\underline{\hat{z}}_t \leq z_t \leq \overline{\hat{z}}_t$, $\forall t \geq 0$ with

$$\begin{aligned}
\underline{\hat{z}}_{t+1} =&(a-lc)\underline{\hat{z}}_t + l\sum_{i=1}^{i=n}\hat{y}_t^{(i)} + \sum_{i=1}^{i=n}\underline{w}_t^{(i)} - l(\sum_{i=1}^{i=n}\overline{v}_t^{(i)} + \sum_{i=1}^{i=n}\overline{\zeta}^{(i)}),\\
\overline{\hat{z}}_{t+1} =&(a-lc)\overline{\hat{z}}_t + l\sum_{i=1}^{i=n}\hat{y}_t^{(i)} + \sum_{i=1}^{i=n}\overline{w}_t^{(i)} - l(\sum_{i=1}^{i=n}\underline{v}_t^{(i)} + \sum_{i=1}^{i=n}\underline{\zeta}^{(i)}),
\end{aligned}$$

$\underline{\hat{z}}_0 = \sum_{i=1}^{i=n}\underline{x}_0^{(i)} \leq z_0 \leq \overline{\hat{z}}_0 = \sum_{i=1}^{i=n}\overline{x}_0^{(i)}$ and

$$\overline{\zeta}^{(i)} = -\underline{\zeta}^{(i)} = \frac{\rho}{\epsilon}\ln\left(1 + e^{\epsilon}\frac{m\left(1-e^{-\epsilon/m}\right)}{2\delta}\right). \tag{3.26}$$

Furthermore, we get

$$\begin{aligned}
\underline{\eta}_{t+1} =&(a-lc)\underline{\eta}_t + \sum_{i=1}^{i=n}(w_t^{(i)} - \underline{w}_t^{(i)}) + l\sum_{i=1}^{i=n}(\overline{v}_t^{(i)} + \overline{\zeta}^{(i)}) - l(\sum_{i=1}^{i=n}v_t^{(i)} + \sum_{i=1}^{i=n}\zeta_t^{(i)}), \qquad (3.27)\\
\overline{\eta}_{t+1} =&(a-lc)\overline{\eta}_t + \sum_{i=1}^{i=n}(\overline{w}_t^{(i)} - w_t^{(i)}) + l(\sum_{i=1}^{i=n}v_t^{(i)} + \sum_{i=1}^{i=n}\zeta_t^{(i)}) - l(\sum_{i=1}^{i=n}\underline{v}_t^{(i)} + \sum_{i=1}^{i=n}\underline{\zeta}^{(i)}).
\end{aligned}$$

On the other hand, one can use the architecture of Figure 3.4 instead of input perturbation. Let the matrix $F = \mathbf{1}_n^{\mathrm{T}}$ and denote $h_t = Fy_t$. In such a case, one can bound the sensitivity

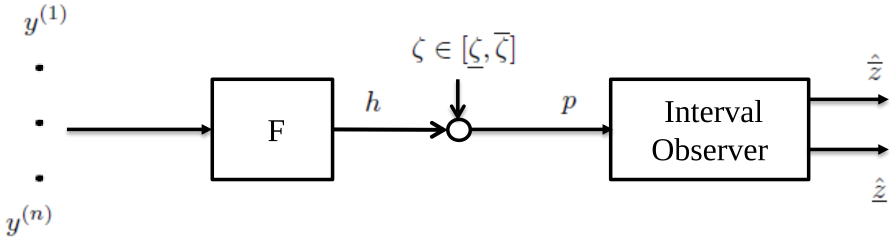


Figure 3.4 : Two-stage architecture for differentially private interval estimation

as follows

$$\begin{aligned}\Delta F =& \sup_{y,\tilde{y}:\mathrm{Adj}(y,\tilde{y})} |Fy - F\tilde{y}|_1\\ =& \sup_{y,\tilde{y}:\mathrm{Adj}(y,\tilde{y})} |\sum_{i=1}^{i=n} y^{(i)} - \sum_{i=1}^{i=n} \tilde{y}^{(i)}|_1\\ =&\rho.\end{aligned}$$

By applying Theorem 9, we deduce that releasing $p_t = h_t + \zeta_t$, where the probability distribution of ζ is defined in (3.6) is (ϵ, δ)-differentially private for the adjacency relation (3.5). We have $\underline{\zeta} \leq \zeta_t \leq \overline{\zeta}$ for all $t \geq 0$ with $\underline{\zeta} = \underline{\zeta}^{(i)}$ and $\overline{\zeta} = \overline{\zeta}^{(i)}$, where $\underline{\zeta}^{(i)}$ and $\overline{\zeta}^{(i)}$ are defined in (3.26). By using such an two-stage architecture, we can write equations of an interval estimator as follows

$$\begin{aligned}\hat{\underline{z}}_{t+1} =&(a-lc)\hat{\underline{z}}_t + lp_t + \sum_{i=1}^{i=n} \underline{w}_t^{(i)} - l(\sum_{i=1}^{i=n} \overline{v}_t^{(i)} + \overline{\zeta}^{(i)}),\\ \hat{\overline{z}}_{t+1} =&(a-lc)\hat{\overline{z}}_t + lp_t + \sum_{i=1}^{i=n} \overline{w}_t^{(i)} - l(\sum_{i=1}^{i=n} \underline{v}_t^{(i)} + \underline{\zeta}^{(i)},\end{aligned} \tag{3.28}$$

with $\hat{\underline{z}}_0 = \sum_{i=1}^{i=n} \underline{x}_0^{(i)} \leq z_0 \leq \hat{\overline{z}}_0 = \sum_{i=1}^{i=n} \overline{x}_0^{(i)}$. Denote the estimation errors $\underline{\beta}_t = z_t - \hat{\underline{z}}_t$ and $\overline{\beta}_t = \hat{\overline{z}}_t - z_t$, where $\hat{\underline{z}}_t$ and $\hat{\overline{z}}_t$ are the released differentially private estimates of z_t when we use the two-stage architecture of Figure 3.4.

Theorem 12 *Let Assumptions 1 and 2 be satisfied. Then, we get for* (3.28)

$$\hat{\underline{z}}_t \leq z_t \leq \hat{\overline{z}}_t, \forall t \geq 0, \tag{3.29}$$

provided that $\hat{\underline{z}}_0 \leq z_0 \leq \hat{\overline{z}}_0$, and moreover $\underline{\beta}, \overline{\beta} \in \mathcal{L}_\infty^1$.

Proof. The proof uses the exact argumentation used for the proof of Theorem 10. □

We can write the estimation errors' dynamics of the two-stage architecture as follows

$$\begin{aligned}\underline{\beta}_{t+1} =&(a-lc)\underline{\beta}_t + \sum_{i=1}^{i=n} w_t^{(i)} - \sum_{i=1}^{i=n} \underline{w}_t^{(i)} + l^+(\sum_{i=1}^{i=n} \overline{v}_t^{(i)} + \overline{\zeta}^{(i)}) - l(\sum_{i=1}^{i=n} v_t^{(i)} + \zeta_t),\\ \overline{\beta}_{t+1} =&(a-lc)\overline{\beta}_t + \sum_{i=1}^{i=n} \overline{w}_t^{(i)} - \sum_{i=1}^{i=n} w_t^{(i)} + l(\sum_{i=1}^{i=n} v_t^{(i)} + \zeta_t) - l(\sum_{i=1}^{i=n} \underline{v}_t^{(i)} + \underline{\zeta}^{(i)}).\end{aligned} \tag{3.30}$$

Let us compare now the estimation errors $\underline{\eta}_t$ and $\overline{\eta}_t$, obtained when one use the input perturbation architecture of Figure 3.2, to the estimation errors $\underline{\beta}_t$ and $\overline{\beta}_t$, obtained when we use the two-stage architecture of Figure 3.4. One can notice from (3.27) and (3.30) that the difference between the two-stage architecture and the input perturbation is the impact of the privacy-preserving noise and its lower and upper bounds on $\underline{\beta}_t$ and $\overline{\beta}_t$ as the number of participants n increases, through the expressions $\underline{\zeta}^{(i)}$, $\overline{\zeta}^{(i)}$ and ζ_t that replace the terms $\sum_{i=1}^{i=n}\underline{\zeta}^{(i)}$, $\sum_{i=1}^{i=n}\overline{\zeta}^{(i)}$ and $\sum_{i=1}^{i=n}\zeta_t^{(i)}$ in the expressions of $\underline{\eta}_t$ and $\overline{\eta}_t$. Denote $\underline{\kappa}_t = \underline{\eta}_t - \underline{\beta}_t$ and $\overline{\kappa}_t = \overline{\eta}_t - \overline{\beta}_t$. Since the noise ζ_t of the two-stage architecture and the noises $\zeta_t^{(i)}$ of the input perturbation mechanism are both distributed according to the distribution (3.6), we select $\zeta_t = \zeta_t^{(n)}$ for the next proposition.

Proposition 4 *Let Assumption 2 be satisfied. We get*

$$\begin{aligned}
\underline{\kappa}_{t+1} =& (a-lc)\underline{\kappa}_t + l\left((n-1)\overline{\zeta}^{(i)} - \sum_{i=1}^{i=n-1}\zeta_t^{(i)}\right), \\
\overline{\kappa}_{t+1} =& (a-lc)\overline{\kappa}_t + l\left(\sum_{i=1}^{i=n-1}\zeta_t^{(i)} - (n-1)\underline{\zeta}^{(i)}\right).
\end{aligned} \tag{3.31}$$

Furthermore, $\underline{\kappa}_t \geq 0$, $\overline{\kappa}_t \geq 0, \forall t \geq 0$.

Proof. It can be inferred from (3.27) and (3.30) that

$$\begin{aligned}
\underline{\kappa}_{t+1} =& (a-lc)\underline{\kappa}_t + l(\sum_{i=1}^{i=n}\overline{\zeta}^{(i)} - \overline{\zeta}^{(i)}) - l(\sum_{i=1}^{i=n}\zeta_t^{(i)} - \zeta_t^{(n)}), \\
\overline{\kappa}_{t+1} =& (a-lc)\overline{\kappa}_t + l(\sum_{i=1}^{i=n}\zeta_t^{(i)} - \zeta_t^{(n)}) - l(\sum_{i=1}^{i=n}\underline{\zeta}^{(i)} - \underline{\zeta}^{(i)}),
\end{aligned}$$

which leads to the dynamics (3.31). The system (3.31) is cooperative when Assumption 2 is satisfied. Then, by applying Lemma 1, we get $\underline{\kappa}_t \geq 0, \overline{\kappa}_t \geq 0$ since $\underline{\kappa}_0 = 0$ and $\overline{\kappa}_0 = 0$. □

It can be inferred from Proposition 4 that the two-stage architecture of Figure 3.4 has better performance than the input perturbation architecture of Figure 3.2 and the difference of performance increases with the number of participants n.

3.6 Conclusion

In this chapter, we consider the problem of interval estimation under a differential privacy constraint. We design an input perturbation architecture for differentially private interval estimation. The performance of our private interval estimator is illustrated through numerical simulations. Moreover, we show that a two-stage architecture can significantly improve the performance of the differentially private interval estimator.

CHAPTER 4 STEALTHY ATTACKS AND ATTACK-RESILIENT INTERVAL OBSERVERS

In this chapter, first, we present the monitor that is considered throughout this part of the thesis and the problem statement in Section 4.1. In Section 4.2, we design attack signals that are undetected by the considered monitor. Then, we construct interval state observers that are resilient to such attacks in Section 4.3. Furthermore, we present computational methods to obtain interval observer parameters that minimize the $\mathcal{H}_\infty$ norms of the estimation error dynamics in Sections 4.4. Next, we design bounds for the attack signals in Section 4.5.

4.1 Problem statement

4.1.1 System model and Monitor

Consider the following discrete-time linear time-invariant (LTI) system for $t \in \mathbb{Z}_+$

$$\begin{aligned} x_{t+1} &= Ax_t + Mw_t + Ea_t, \\ y_t &= Cx_t + Nw_t + Da_t, \end{aligned} \tag{4.1}$$

where $x_t \in \mathbb{R}^n$ is the state vector, $w : \mathbb{Z}_+ \to \mathbb{R}^q$ is an *unknown* input in $\mathcal{L}_\infty^q$ representing process and measurement noises and disturbances, $y_t \in \mathbb{R}^p$ is a measured output signal, $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, $M \in \mathbb{R}^{n\times q}$, $N \in \mathbb{R}^{p\times q}$ are known constant matrices. The signal $a \in \mathcal{L}_\infty^m$ denotes an “attack signal” applied by an adversary, which influences the dynamics and sensor measurements through the matrices $E \in \mathbb{R}^{n\times m}$ and $D \in \mathbb{R}^{p\times m}$. We do not impose any statistical restrictions or a priori bounds on the signal a. The exact value of the initial condition x_0 is unknown but satisfies $\underline{x}_0 \le x_0 \le \overline{x}_0$, where $\underline{x}_0, \overline{x}_0 \in \mathbb{R}^n$ are given vectors. Furthermore, the exact values of the input w_t are unknown, but two bounded signals $\underline{w}_t$, $\overline{w}_t \in \mathcal{L}_\infty^q$ are given such that $\underline{w}_t \le w_t \le \overline{w}_t, \quad \forall t \ge 0$.

A monitor aims to detect attacks, i.e., to decide if a non-zero signal a is present in (4.1). We assume its design follows the parity equation approach [112, Chapter 10], extended here to discrete-time LTI systems with bounded uncertainties, as follows. Given $d \in \mathbb{Z}_+$, it can be inferred from (4.1) that

$$y_{t-d:t} = \mathcal{O}^{(d)} x_{t-d} + \mathcal{Q}_w^{(d)} w_{t-d:t} + \mathcal{Q}_a^{(d)} a_{t-d:t}, \tag{4.2}$$

with $\mathcal{O}^{(d)} = \begin{bmatrix} C^T & (CA)^T & (CA^2)^T & \dots & (CA^d)^T \end{bmatrix}^T$ and

$$\mathcal{Q}_w^{(d)} = \begin{bmatrix} N & 0 & 0 & 0 & \dots & 0 \\ CM & N & 0 & 0 & \dots & 0 \\ CAM & CM & N & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots \\ CA^{(d-1)}M & CA^{(d-2)}M & \dots & \dots & CM & N \end{bmatrix},$$

$$\mathcal{Q}_a^{(d)} = \begin{bmatrix} D & 0 & 0 & 0 & \dots & 0 \\ CE & D & 0 & 0 & \dots & 0 \\ CAE & CE & D & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots \\ CA^{(d-1)}E & CA^{(d-2)}E & \dots & \dots & CE & D \end{bmatrix}.$$

We can generate residuals that are independent of the unknown state x_{t-d} by pre-multiplying (4.2) by a matrix R such that $R\mathcal{O}^{(d)} = 0$. Such a matrix always exist, if we take d sufficiently large. Indeed, the rank of the matrix $\mathcal{O}^{(d)}$ can be at most n, the dimension of the state x. Consequently, we can take the columns of R^T to form a basis of the null space of $(\mathcal{O}^{(d)})^{\mathrm{T}}$, when $(d+1)\,p > \mathrm{rank}(\mathcal{O}^{(d)})$. In the following, we assume the value of d used by the monitor known and fixed.

Ideally, we would like to select R such that it also satisfies $R\mathcal{Q}_w^{(d)} = 0$, in order to obtain residuals that are also insensitive to the disturbances. However, this is not always possible, since for example N and hence $\mathcal{Q}_w^{(d)}$ could be full row rank. Consequently, in general, we get

$$Ry_{t-d:t} = R\mathcal{Q}_w^{(d)} w_{t-d:t} + R\mathcal{Q}_a^{(d)} a_{t-d:t}. \tag{4.3}$$

Once R is selected, the following result follows from Lemma 2.

Proposition 5 *Consider the system* (4.1)*. Then, if it is not under attack, i.e., if $a_t = 0$ for all t, we have*

$$\underline{\Theta}_t \leq Ry_{t-d:t} \leq \overline{\Theta}_t, \forall t \geq 0, \ \textit{with} \tag{4.4}$$

$$\begin{aligned} \underline{\Theta}_t &= \left(R\mathcal{Q}_w^{(d)}\right)^+ \underline{w}_{t-d:t} - \left(R\mathcal{Q}_w^{(d)}\right)^- \overline{w}_{t-d:t}, \\ \overline{\Theta}_t &= \left(R\mathcal{Q}_w^{(d)}\right)^+ \overline{w}_{t-d:t} - \left(R\mathcal{Q}_w^{(d)}\right)^- \underline{w}_{t-d:t}. \end{aligned} \tag{4.5}$$

The monitor can then raise an alarm if the measurements $y_{t-d:t}$ do not satisfy the bounds

(4.4), since this implies that an attack signal is present. Provided a measurement signal y does not raise an alarm, it is then used by an interval observer to provide lower and upper bounds on the state of the system (4.1). For this, we make the following assumption, which simplifies the interval estimation approach. Although restrictive, this assumption can be relaxed by using a change of coordinates if the pair (A, C) is observable [94, Lemma 1].

Assumption 4 *There exists a matrix $L \in \mathbb{R}^{n \times p}$ such that the matrix $(A - LC)$ is Schur stable and nonnegative.*

Given a matrix L satisfying Assumption 4, if we do not take into account the possible presence of an attack signal, we can design the following interval observer for the state x

$$\begin{aligned}\underline{x}_{t+1} &= (A-LC)\underline{x}_t + Ly_t + M^+\underline{w}_t - M^-\overline{w}_t - (LN)^+\overline{w}_t + (LN)^-\underline{w}_t,\\ \overline{x}_{t+1} &= (A-LC)\overline{x}_t + Ly_t + M^+\overline{w}_t - M^-\underline{w}_t - (LN)^+\underline{w}_t + (LN)^-\overline{w}_t,\end{aligned} \tag{4.6}$$

where $\underline{x}_t \in \mathbb{R}^n$ and $\overline{x}_t \in \mathbb{R}^n$ represent lower and upper interval estimates for the state x_t. We obtain the following result by applying Lemma 1, see for example [113, 114].

Theorem 13 *Let L in* (4.6) *satisfy the conditions of Assumption 4 and suppose that $a \equiv 0$ in* (4.1) *and $\underline{x}_0 \leq x_0 \leq \overline{x}_0$. Then, we have*

$$\underline{x}_t \leq x_t \leq \overline{x}_t, \quad \forall t \geq 0. \tag{4.7}$$

Note also that if $\Delta_t = \overline{x}_t - \underline{x}_t$, then $\Delta_0 \geq 0$ and

$$\Delta_{t+1} = (A - LC)\Delta_t + (|LN| + |M|)(\overline{w}_t - \underline{w}_t). \tag{4.8}$$

Consequently, the size Δ_t of the interval estimate does not depend on the measurement signal y_t but only on the noise parameters M, N and on the gain matrix L.

4.1.2 Adversary Model

We assume that an adversary with detailed knowledge of the system designs an attack signal a entering (4.1), with the goal of perturbing the interval estimates (4.7), e.g., rendering the bounds invalid, while trying to remain undetected by the monitor.

Definition 3 *An attack signal a in* (4.1) *is called* stealthy *for the monitor's bounds $\underline{\Theta}$, $\overline{\Theta}$ given by* (4.5) *if it produces a plausible output signal for the monitor, i.e., if the corresponding output signal y satisfies* (4.4)*.*

Remark 2 *The notion of stealthiness in Definition 3 depends on the monitor used, given by* (4.4)*. This is similar to the definition of stealthy attacks used in [115] for example, although there the authors consider an observer-based monitor and stochastic noise. Other models consider attacks to be stealthy or undetectable if they only excite the zero dynamics of* (4.1)*, in other words, if the attack signal has no influence on y [115, 116]. Definition 3 gives more freedom to the adversary than zero dynamics attacks, allowing him/her to take advantage of the noise to hide an attack signal. A small value of the delay d used by the monitor could also be exploited, even in a noise free scenario.*

We consider a powerful adversary who knows the model of the system (4.1), i.e., the matrices A, C, M, N, the bounds $\underline{x}_0, \overline{x}_0$ and $\underline{w}, \overline{w}$, as well as the matrix R, the parameter d and therefore the bounds (4.5) used by the monitor. Before selecting a_t at period t, the adversary has also access to the sensor measurements

$$\tilde{y}_t = Cx_t + Nw_t. \tag{4.9}$$

and to

$$\tilde{x}_{t+1} = Ax_t + Mw_t. \tag{4.10}$$

As in [116], this represents a worst case scenario where the adversary knows the full state at all times. Finally, as in [117], the adversary also knows the observer gain L. He/she can select an attack signal a that depends on all of this information, in order to degrade the state estimates. As (4.8) shows however, the width of the interval estimate and order of the bounds cannot be influenced. The monitor, on the other hand, does not have access to the state of (4.1) or to $\tilde{y}_t$, but only to the attacked measurements $y_t = \tilde{y}_t + Da_t$.

In the rest of the chapter, for a given system (4.1) under attack, first we describe a methodology for the adversary to design and optimize stealthy attacks. Next, we design an attack-resilient interval observer for the monitor, i.e., we design lower and upper bounds for the state x that remain valid even under attack. Finally, we compute interval observer gains that provide tight bounds on x, minimizing the $\mathcal{H}_\infty$-norm from disturbances to the estimation errors.

4.2 Design of Stealthy Attacks

To understand the impact of attack signals on the interval observer of Theorem 13, in this section we develop a strategy to design dynamic stealthy attacks that aim to invalidate the bounds (4.7). Consider the error signals $\underline{e}_t = x_t - \underline{x}_t$ and $\overline{e}_t = \overline{x}_t - x_t$, whose dynamics read

$$\begin{aligned} \underline{e}_{t+1} &= \Lambda \underline{e}_t + \Psi w_t + \Pi a_t + \underline{\nu}_t \\ \overline{e}_{t+1} &= \Lambda \overline{e}_t - \Psi w_t - \Pi a_t + \overline{\nu}_t. \end{aligned} \tag{4.11}$$

with

$$\Lambda = A - LC, \Psi = M - LN, \Pi = E - LD, \underline{\nu}_t = -(M^+ \underline{w}_t - M^- \overline{w}_t) + (LN)^+ \overline{w}_t - (LN)^- \underline{w}_t,$$
$$\overline{\nu}_t = M^+ \overline{w}_t - M^- \underline{w}_t - \Big((LN)^+ \underline{w}_t - (LN)^- \overline{w}_t\Big).$$

Define the notation, for $h \in \mathbb{Z}_+$ and matrices P, Q,

$$\mathcal{H}^h(P, Q) := \begin{bmatrix} Q & 0 & 0 & 0 & \dots & 0 \\ PQ & Q & 0 & 0 & \dots & 0 \\ P^2 Q & PQ & Q & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots \\ P^h Q & P^{h-1} Q & \dots & \dots & PQ & Q \end{bmatrix},$$

as well as the matrices

$$\mathcal{J}_1^h = \begin{bmatrix} \Lambda^{\mathrm{T}} & \dots & (\Lambda^{h+1})^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \ \mathcal{J}_2^h = \mathcal{H}^h(\Lambda, \Psi), \ \mathcal{J}_3^h = \mathcal{H}^h(\Lambda, \Pi), \ \mathcal{J}_4^h = \mathcal{H}^h(\Lambda, I).$$

The we have from (4.11)

$$\begin{aligned} \underline{e}_{t+1:t+h+1} &= \mathcal{J}_1^h \, \underline{e}_t + \mathcal{J}_2^h w_{t:t+h} + \mathcal{J}_3^h a_{t:t+h} + \mathcal{J}_4^h \underline{\nu}_{t:t+h} \\ \overline{e}_{t+1:t+h+1} &= \mathcal{J}_1^h \, \overline{e}_t - \mathcal{J}_2^h w_{t:t+h} - \mathcal{J}_3^h a_{t:t+h} + \mathcal{J}_4^h \overline{\nu}_{t:t+h}. \end{aligned}$$

In view of these relations, the proposed strategy consists in computing at time t an attack vector $a_{t:t+h}$ that minimizes the objective

$$\min_{a_{t:t+h}} \ (c_1 - c_2)^{\mathrm{T}} \mathcal{J}_3^h \, a_{t:t+h} \tag{4.12}$$

and hence attempts to minimize $c_1^{\mathrm{T}} \underline{e}_{t+1:t+h+1} + c_2^{\mathrm{T}} \overline{e}_{t+1:t+h+1}$. The horizon $h \in \mathbb{Z}_+$ and vectors $c_1, c_2 \in \mathbb{R}^{n(h+1)}$ are design parameters. For example, choosing $c_1^{\mathrm{T}} = [0, \dots, \mathbf{1}_n^T]$ and $c_2 = 0$

aims to make $\underline{e}_{t+h+1}$ small and hopefully *negative* at time $t+h+1$, thereby violating the interval observer bound. In the following, we describe additional constraints imposed on $a_{t:t+h}$ to enforce stealthiness of the attack. Moreover, as in model-predictive control, once $a_{t:t+h}$ is computed we only apply the first input a_t and then let the system evolves. As new information becomes available to the attacker, at period $t+1$ a new vector $a_{t+1:t+1+h}$ is computed, a_{t+1} is applied, and so on.

To generate stealthy attacks, the adversary must ensure that $Ry_{t-d:t}$ remains within the bounds (4.4) for all t. At time t, when selecting a_t, the adversary knows $\tilde{y}_t$ and $\tilde{x}_{t+1}$, see (4.9) and (4.10). Moreover, the past measurements y_{t-1}, y_{t-2}, etc., have already been realized. To select a_t, we compute $a_{t:t+h}$ to make sure that y passes the test of the monitor at time t *and also at future times.* First, let us write $R = \begin{bmatrix} R_d & \dots & R_0 \end{bmatrix}$, where each R_i has p columns, and define correspondingly

$$RQ_w^{(d)} = \begin{bmatrix} \Xi_d & \dots \Xi_1 & \Xi_0 \end{bmatrix}, \; RQ_a^{(d)} = \begin{bmatrix} \Phi_d & \dots \Phi_1 & \Phi_0 \end{bmatrix}.$$

At time t, the attack signal a_t needs to satisfy

$$\underline{\Theta}_t \leq \sum_{i=1}^{d} R_i y_{t-i} + R_0 \tilde{y}_t + R_0 D a_t \leq \overline{\Theta}_t,$$
$$\text{that is, } \; \underline{\Theta}_t \leq \sum_{i=1}^{d} R_i y_{t-i} + R_0 \tilde{y}_t + \Phi_0 a_t \leq \overline{\Theta}_t.$$

However, a_t also impacts future states and measurements. At time $t+1$, we must have

$$\underline{\Theta}_{t+1} \leq \sum_{i=1}^{d-1} R_{i+1} y_{t-i} + R_1(\tilde{y}_t + D a_t) + R_0 y_{t+1} \leq \underline{\Theta}_{t+1},$$

where y_{t+1} is yet unrealized but can be written

$$y_{t+1} = C\tilde{x}_{t+1} + CEa_t + Nw_{t+1} + Da_{t+1}.$$

Hence, when choosing a_t we must also ensure that there exists a_{t+1} such that

$$\underline{\Theta}_{t+1} \leq \sum_{i=1}^{d-1} R_{i+1} y_{t-i} + R_1 \tilde{y}_t + R_0 C \tilde{x}_{t+1} + R_0 N w_{t+1} + (R_1 D + R_0 CE) a_t + R_0 D a_{t+1} \leq \overline{\Theta}_{t+1}.$$

Since w_{t+1} is still unknown, we assume a worst case scenario and aim to ensure that these constraints remain feasible even for $R_0 N w_{t+1} = \Xi_0 w_{t+1}$ replaced by its upper and lower bounds from Lemma 2. Given the definitions of $\underline{\Theta}_{t+1}$ and $\overline{\Theta}_{t+1}$ however, the terms involving the bounds on w_{t+1} simplify and we obtain the following constraints involving a_t, a_{t+1}

$$\begin{aligned}
&\sum_{i=1}^{d}(\Xi_i^+ \underline{w}_{t+1-i} - \Xi_i^- \overline{w}_{t+1-i}) \leq \\
&\sum_{i=1}^{d-1} R_{i+1} y_{t-i} + R_1 \tilde{y}_t + R_0 C \tilde{x}_{t+1} + \Phi_1 a_t + \Phi_0 a_{t+1} \\
&\leq \sum_{i=1}^{d}(\Xi_i^+ \overline{w}_{t+1-i} - \Xi_i^- \underline{w}_{t+1-i}).
\end{aligned}$$

Now, because of the delayed samples present in the monitor's test, the vector y_t and hence a_t appear directly in this test up to time $t+d$. This gives a total of $d+1$ constraints involving a_t directly through the measurements, which, by a straightforward calculation following the argument above, can be written, for $0 \leq k \leq d$,

$$\begin{aligned}
&\sum_{i=k}^{d}(\Xi_i^+ \underline{w}_{t+k-i} - \Xi_i^- \overline{w}_{t+k-i}) \leq \\
&\sum_{i=1}^{d-k} R_{i+k} y_{t-i} + R_k \tilde{y}_t + \Upsilon_k \tilde{x}_{t+1} + \sum_{i=0}^{k} \Phi_i \, a_{t+k-i} \\
&\leq \sum_{i=k}^{d}(\Xi_i^+ \overline{w}_{t+k-i} - \Xi_i^- \underline{w}_{t+k-i}),
\end{aligned} \tag{4.13}$$

with $\Upsilon_0 = 0$ and $\Upsilon_k = \sum_{i=0}^{k-1} R_i C A^{k-1-i}$ for $1 \leq k \leq d$.

For $k > d$, a_t only indirectly affects the monitor's residual signals. The reasoning above then just leads to the equality constraints $RQ_a^{(d)} \, a_{t+i:t+i+d} = 0$ for $i \geq 1$, i.e.,

$$\sum_{j=0}^{d} \Phi_j \, a_{t+d+i-j} = 0, \quad \forall i \geq 1, \tag{4.14}$$

which can be seen to guarantee that future attack vectors do not impact the residuals in (4.3).

The constraints (4.13) and (4.14) leave some freedom to choose an attack sequence $a_{t:t+h}$ optimizing (4.12). One possible strategy is to fix $a_{t+i} = \alpha$ for all $i > h_0 \geq 0$, with h_0 possibly strictly larger than h and α any element such that

$$\Phi_d \, \alpha = \ldots = \Phi_0 \, \alpha = 0, \tag{4.15}$$

where $\alpha = 0$ is always a possible choice. Then, (4.14) reduces to a *finite* number of linear equality constraints on the remaining finite number of variables $a_{t:t+h_0}$. Finding a vector

$a_{t:t+h_0}$ satisfying these constraints together with (4.13) while minimizing (4.12) is a *linear programming problem*, which is always feasible and can be solved efficiently. This gives a design strategy for non-trivial attacks that are guaranteed to remain stealthy for the monitor (4.4).

Remark 3 *If the system* (A, E, C, D) *is not input observable [118, Corollary 5], then there exists* $\alpha \neq 0$ *such that* $D\alpha = CE\alpha = CAE\alpha = \ldots = CA^{n-1}E\alpha = 0$, *in which case the solutions of* (4.15) *form a linear space of dimension at least* 1.

Remark 4 *Imposing all the constraints* (4.13) *and* (4.14) *provides a guarantee that an attack signal will remain stealthy as time evolves, but more aggressive attack signals could be obtained by dropping some of these constraints, for example* (4.14)*, without necessarily being detected by the monitor. Indeed, this will depend on the actual realization of the disturbance signal* w*, which attack vectors at future times can still take advantage of.*

When the system (4.1) is under attack, the guarantee (4.7) of the observer (4.6) is not necessarily satisfied, see Figure 4.1 for example. In the next section, we design an interval observer that is resilient to stealthy attacks, i.e., provides bounds $\underline{\chi}_t \leq x_t \leq \overline{\chi}_t$ for all $t \geq 0$ that are guaranteed to hold despite the presence of the signal a.

4.3 Attack-Resilient Interval Observer

Herein, we extend the methodology proposed in [103] to LTI discrete-time systems with measurement noise and bounded disturbances. We can assume without loss of generality that $\text{rank}\left(\begin{bmatrix} E \\ D \end{bmatrix}\right) = m$, otherwise some columns of this matrix can simply be removed. Moreover, we make here the following assumption, which constitutes a limit on the capability of the attacker.

Assumption 5 *There exist integers* $0 \leq n_a \leq n - 1$ *and* $1 \leq n_f \leq p$ *such that rank*$(E) = n_a < n$ *and rank*$(D) = p - n_f < p$.

As a result of Assumption 5, there exist two nonzero and full row rank matrices $T_1 \in \mathbb{R}^{(n-n_a)\times n}$ and $F \in \mathbb{R}^{n_f \times p}$ such that $T_1 E = 0$ and $FD = 0$. Note also that in case of an attack on actuators only, i.e., when $D = 0$, the matrix F can be selected as $F = I_p$.

Example 1 *Consider the system*

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} -0.5 & 1 \\ 0 & 0.5 \end{bmatrix} x_t + w_{1,t} + \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \\ y_t &= x_t + w_{2,t} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}. \end{aligned}$$

Then we can simply take $T_1 = \begin{bmatrix} 1 & -\frac{1}{2} \end{bmatrix}$ *and* $F = \begin{bmatrix} 1 & -1 \end{bmatrix}$.

Let us complete T_1 to a non-singular matrix $T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$, with $T_2 \in \mathbb{R}^{n_a \times n}$. With the change of coordinates $x = T^{-1}z$, the system (4.1) can be rewritten as follows

$$\begin{aligned} z_{t+1} &= \bar{A} z_t + \bar{M} w_t + \bar{E} a_t, \\ y_t &= \bar{C} z_t + N w_t + D a_t, \end{aligned} \tag{4.16}$$

with

$$\begin{aligned} z_t &= \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix}, \ \bar{A} = TAT^{-1} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix}, \\ \bar{M} &= TM, \ \bar{E} = TE = \begin{bmatrix} 0 \\ T_2 E \end{bmatrix}, \ \bar{C} = CT^{-1}, \end{aligned}$$

where $z_t^1 \in \mathbb{R}^{n-n_a}$ and $z_t^2 \in \mathbb{R}^{n_a}$. In (4.16), notice that the difference equation corresponding to the state component z^2 is the only one affected by the attack a.

Using $FD = 0$, we get the following descriptor system

$$\begin{aligned} \begin{bmatrix} I_{n-n_a} & 0 \end{bmatrix} z_{t+1} &= \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \end{bmatrix} z_t + \bar{M}_1 w_t, \\ \mathcal{Y}_t &= F y_t = FCT^{-1} z_t + FN w_t \\ &= FCB z_t^1 + FCG z_t^2 + FN w_t, \end{aligned} \tag{4.17}$$

where we have denoted $\bar{M}_1 = \begin{bmatrix} I_{n-n_a} & 0 \end{bmatrix} \bar{M}$ and $T^{-1} = \begin{bmatrix} B & G \end{bmatrix}$, with $B \in \mathbb{R}^{n \times (n-n_a)}$ and $G \in \mathbb{R}^{n \times n_a}$. In the following, using an idea from [103], we express the state component z^2 as a function of $\mathcal{Y}$ and z^1, in order to transform (4.17) into a standard linear system. For this, we need the following assumption.

Assumption 6 *The matrix FCG is full column rank, equal to n_a.*

Assumption 6 means that the virtual measurements $\mathcal{Y}$ still contain enough information about the components of the state under attack. Note that it requires that $n_f \geq n_a$. For our Example 1 above, we can take $T_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and compute $FCG = -1/2$, so Assumption 6 is satisfied.

When Assumption 6 holds, there exists a non-singular matrix $V = \begin{bmatrix} FCG & Q \end{bmatrix}$, with $Q \in \mathbb{R}^{n_f \times (n_f - n_a)}$. Define

$$V^{-1} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \; V_1 \in \mathbb{R}^{n_a \times n_f}, \; V_2 \in \mathbb{R}^{(n_f - n_a) \times n_f}.$$

Notice that when $n_f = n_a$, we have $V^{-1} = V_1$ and the matrices Q and V_2 do not exist. Hence, for the Example 1 above, $V^{-1} = V_1 = -2$. In general, we have

$$V^{-1}V = \begin{bmatrix} V_1FCG & V_1Q \\ V_2FCG & V_2Q \end{bmatrix} = \begin{bmatrix} I_{n_a} & 0 \\ 0 & I_{n_f - n_a} \end{bmatrix}. \tag{4.18}$$

By multiplying both sides of the equation of $\mathcal{Y}_t$ in (4.17) by V^{-1} and by using (4.18), we get

$$\begin{aligned} V_1\mathcal{Y}_t &= V_1FCBz_t^1 + z_t^2 + V_1FNw_t, \\ V_2\mathcal{Y}_t &= V_2FCBz_t^1 + V_2FNw_t, \end{aligned} \tag{4.19}$$

which in turn implies that

$$z_t^2 = V_1\mathcal{Y}_t - V_1FCBz_t^1 - V_1FNw_t. \tag{4.20}$$

By combining (4.17) with (4.19)-(4.20), we get

$$\begin{aligned} z_{t+1}^1 &= \bar{A}_{11}z_t^1 + \bar{A}_{12}z_t^2 + \bar{M}_1w_t, \\ &= (\bar{A}_{11} - \bar{A}_{12}V_1FCB)z_t^1 + \bar{A}_{12}V_1\mathcal{Y}_t + (-\bar{A}_{12}V_1FN + \bar{M}_1)w_t, \\ \hat{y}_t &= V_2\mathcal{Y}_t = V_2FCBz_t^1 + V_2FNw_t. \end{aligned} \tag{4.21}$$

Denote $\mathcal{A} = \bar{A}_{11} - \bar{A}_{12}V_1FCB$ and $\mathcal{C} = V_2FCB$. We need the following assumption to design an attack-resilient interval observer for the state of (4.21).

Assumption 7 *There exists $\mathcal{L} \in \mathbb{R}^{(n-n_a)\times(n_f-n_a)}$ such that the matrix $(\mathcal{A} - \mathcal{L}\mathcal{C})$ is Schur stable and nonnegative.*

As for Assumption 4 , Assumption 7 can be relaxed by using a change of coordinates if the pair $(\mathcal{A}, \mathcal{C})$ is observable [94, Lemma 1]. Note also that if $n_f = n_a$, the matrices $\mathcal{C}$, $\mathcal{L}$ do not exist and the assumption means that $\mathcal{A}$ itself should be stable (as in our Example 1, where $\mathcal{A} = 1/2$). In such a case, an open-loop interval observer can be constructed for z^1, which does not use measurements $\hat{y}_t$.

An interval observer for z^1 takes the form

$$\begin{aligned}\underline{z}^1_{t+1} =&(\mathcal{A} - \mathcal{L}\mathcal{C})\underline{z}^1_t + \mathcal{L}\hat{y}_t + \underline{\hat{w}}_t - \mathcal{L}^+\overline{\hat{v}}_t + \mathcal{L}^-\underline{\hat{v}}_t + \bar{A}_{12}V_1\mathcal{Y}_t, \ \forall t \geq 0,\\ \overline{z}^1_{t+1} =&(\mathcal{A} - \mathcal{L}\mathcal{C})\overline{z}^1_t + \mathcal{L}\hat{y}_t + \overline{\hat{w}}_t - \mathcal{L}^+\underline{\hat{v}}_t + \mathcal{L}^-\overline{\hat{v}}_t + \bar{A}_{12}V_1\mathcal{Y}_t, \ \forall t \geq 0,\end{aligned} \tag{4.22}$$

where $\underline{z}^1_0 = T_1^+\underline{x}_0 - T_1^-\overline{x}_0$, $\overline{z}^1_0 = T_1^+\overline{x}_0 - T_1^-\underline{x}_0$, the gain matrix $\mathcal{L}$ verifying the condition of Assumption 7 needs to be determined and

$$\begin{aligned}&\underline{\hat{w}}_t = W^+\underline{w}_t - W^-\overline{w}_t, \ \ \overline{\hat{w}}_t = W^+\overline{w}_t - W^-\underline{w}_t,\\ &\underline{\hat{v}}_t = (V_2FN)^+\underline{w}_t - (V_2FN)^-\overline{w}_t, \ \ \overline{\hat{v}}_t = (V_2FN)^+\overline{w}_t - (V_2FN)^-\underline{w}_t,\end{aligned}$$

with $W = -\bar{A}_{12}V_1FN + \bar{M}_1$.

Define the estimation errors $\underline{\eta}^1_t = z^1_t - \underline{z}^1_t$, $\overline{\eta}^1_t = \overline{z}^1_t - z^1_t$ and let $\hat{w}_t = Ww_t$, $\hat{v}_t = V_2FNw_t$. The error dynamics read

$$\begin{aligned}\underline{\eta}^1_{t+1} &= (\mathcal{A} - \mathcal{L}\mathcal{C})\underline{\eta}^1_t + \underline{g}_{1,t} + \underline{g}_{2,t},\\ \overline{\eta}^1_{t+1} &= (\mathcal{A} - \mathcal{L}\mathcal{C})\overline{\eta}^1_t + \overline{g}_{1,t} + \overline{g}_{2,t},\end{aligned} \tag{4.23}$$

with

$$\underline{g}_{1,t} = \hat{w}_t - \underline{\hat{w}}_t, \ \ \underline{g}_{2,t} = \mathcal{L}^+\overline{\hat{v}}_t - \mathcal{L}^-\underline{\hat{v}}_t - \mathcal{L}\hat{v}_t, \ \overline{g}_{1,t} = \overline{\hat{w}}_t - \hat{w}_t, \ \ \overline{g}_{2,t} = \mathcal{L}\hat{v}_t - (\mathcal{L}^+\underline{\hat{v}}_t - \mathcal{L}^-\overline{\hat{v}}_t).$$

Theorem 14 *Let $\mathcal{L}$ be such that the properties of Assumption 7 are satisfied. Then we get, for* (4.21) *and* (4.22)*,*

$$\underline{z}^1_t \leq z^1_t \leq \overline{z}^1_t, \quad \forall t \geq 0. \tag{4.24}$$

and moreover $\overline{\eta}^1_t$, $\underline{\eta}^1_t$ are in $\mathcal{L}_\infty^{(n-n_a)}$.

Proof. From Lemma 2, we get $\underline{z}^1_0 \leq z^1_0 \leq \overline{z}^1_0$. We also get that all terms $\underline{g}_{i,t}$, $\overline{g}_{i,t}$ in (4.23) are nonnegative. Since $\mathcal{A} - \mathcal{L}\mathcal{C}$ is nonnegative, by applying Lemma 1, we obtain that $\underline{\eta}^1_t$ and $\overline{\eta}^1_t$ are nonnegative, and (4.24) follows. Furthermore, since the inputs $\underline{g}_{i,t}$ and $\overline{g}_{i,t}$ are bounded for all $i = 1, 2$ and $\mathcal{A} - \mathcal{L}C$ is stable, we get that $\overline{\eta}^1_t$, $\underline{\eta}^1_t$ are in $\mathcal{L}_\infty^{(n-n_a)}$. □

From Theorem 14, Lemma 2 and (4.20), we get the following bounds on z^2.

Corollary 2 *For the signals $\underline{z}^1$, $\overline{z}_t^1$ defined by (4.22), we have $\underline{z}_t^2 \leq z_t^2 \leq \overline{z}_t^2$ for all $t \geq 0$, with*

$$\begin{aligned}\underline{z}_t^2 =& V_1\mathcal{Y}_t + (-V_1FCB)^+\underline{z}_t^1 - (-V_1FCB)^-\overline{z}_t^1 + (-V_1FN)^+\underline{w}_t - (-V_1FN)^-\overline{w}_t,\\ \overline{z}_t^2 =& V_1\mathcal{Y}_t + (-V_1FCB)^+\overline{z}_t^1 - (-V_1FCB)^-\underline{z}_t^1 + (-V_1FN)^+\overline{w}_t - (-V_1FN)^-\underline{w}_t.\end{aligned} \tag{4.25}$$

Since $x = T^{-1}z$, attack-resilient estimates for the original state x of (4.1) can now be computed as follows

$$\begin{aligned}\underline{\chi}_t =& (T^{-1})^+\underline{z}_t - (T^{-1})^-\overline{z}_t,\\ \overline{\chi}_t =& (T^{-1})^+\overline{z}_t - (T^{-1})^-\underline{z}_t,\end{aligned} \tag{4.26}$$

with $\underline{z}_t := \begin{bmatrix}(\underline{z}_t^1)^T & (\underline{z}_t^2)^T\end{bmatrix}^T$, $\overline{z}_t := \begin{bmatrix}(\overline{z}_t^1)^T & (\overline{z}_t^2)^T\end{bmatrix}^T$. We summarize these results in the following theorem.

Theorem 15 *Let Assumptions 5, 6 and 7 be satisfied. We then have*

$$\underline{\chi}_t \leq x_t \leq \overline{\chi}_t, \quad \forall t \geq 0, \tag{4.27}$$

provided that $\underline{x}_0 \leq x_0 \leq \overline{x}_0$. Moreover, the error signals $(x - \underline{\chi})$ and $(\overline{\chi} - x)$ are in $\mathcal{L}_\infty^n$.

Remark 5 *In case of an attack on sensors only, i.e., when $E = 0$, we take $T = B = I_n$, $z = z^1 = x$, $A = \bar{A} = \bar{A}_{11}$, $V = V_2 = I_{n_f}$ and z^2 and V_1 do not exist. Since the dynamics (4.1) now read $x_{t+1} = Ax_t + Mw_t$ subject to the measurements $\hat{y}_t = Fy_t = FCx_t + FNw_t$, an interval observer for x simply reads, for all $t \geq 0$,*

$$\begin{aligned}\underline{\chi}_{t+1} =& (A - \mathcal{L}FC)\underline{\chi}_t + \mathcal{L}\hat{y}_t + \underline{\hat{w}}_t - \mathcal{L}^+\overline{\hat{v}}_t + \mathcal{L}^-\underline{\hat{v}}_t,\\ \overline{\chi}_{t+1} =& (A - \mathcal{L}FC)\overline{\chi}_t + \mathcal{L}\hat{y}_t + \overline{\hat{w}}_t - \mathcal{L}^+\underline{\hat{v}}_t + \mathcal{L}^-\overline{\hat{v}}_t,\end{aligned}$$

with $\underline{\chi}_0 = \underline{x}_0$, $\overline{\chi}_0 = \overline{x}_0$, and

$$\begin{aligned}&\underline{\hat{w}}_t = M^+\underline{w}_t - M^-\overline{w}_t,\ \underline{\hat{v}}_t = (FN)^+\underline{w}_t - (FN)^-\overline{w}_t,\\ &\overline{\hat{w}}_t = M^+\overline{w}_t - M^-\underline{w}_t,\ \overline{\hat{v}}_t = (FN)^+\overline{w}_t - (FN)^-\underline{w}_t.\end{aligned} \tag{4.28}$$

4.4 Optimization of the interval observer gains

In this section, we discuss for the case $n_f > n_a$ a computational method to select an observer gain matrix $\mathcal{L}$ that provides tight bounds $\underline{z}^1$, $\overline{z}^1$ in (4.22), and hence tight bounds $\underline{\chi}$, $\overline{\chi}$ on the state in (4.26). The matrix $\mathcal{L}$ solves an optimization problem minimizing the ℓ_2-gain from disturbances to the estimation errors $\overline{z}^1 - z^1$ and $z^1 - \underline{z}^1$. A similar optimization problem is proposed for uncertain LTI systems with bounded uncertainties in [119]. In contrast to the approach of [119], we need to take into account the nonnegativity constraints on the dynamics of the estimation errors. In addition, the interval estimation technique based on zonotopes considered in [119] does not include the gains $\mathcal{L}^+$ and $\mathcal{L}^-$ of the observer in the error dynamics and so we need to solve a different optimization problem here. A similar problem is also addressed in [120], whose solution requires solving a semidefinite program (SDP) and then computing a matrix pseudo-inverse. Our method simply solves an SDP.

Define the following inputs

$$\underline{f}_t = \begin{bmatrix} \hat{w}_t - \underline{\hat{w}}_t \\ \overline{\hat{v}}_t - \hat{v}_t \\ \hat{v}_t - \underline{\hat{v}}_t \end{bmatrix}, \ \overline{f}_t = \begin{bmatrix} \overline{\hat{w}}_t - \hat{w}_t \\ \hat{v}_t - \underline{\hat{v}}_t \\ \overline{\hat{v}}_t - \hat{v}_t \end{bmatrix}.$$

The estimation errors' dynamics (4.23) can be rewritten as follows

$$\underline{\eta}^1_{t+1} =(\mathcal{A} - \mathcal{L}\mathcal{C})\underline{\eta}^1_t + H\underline{f}_t, \tag{4.29}$$

$$\overline{\eta}^1_{t+1} =(\mathcal{A} - \mathcal{L}\mathcal{C})\overline{\eta}^1_t + H\overline{f}_t, \tag{4.30}$$

where $H = \begin{bmatrix} I_{n-n_a} & \mathcal{L}^+ & \mathcal{L}^- \end{bmatrix} \in \mathbb{R}^{(n-n_a)\times n'}$ with $n' = n + 2n_f - 3n_a$. Our objective is then to select a matrix $\mathcal{L}$ minimizing the $\mathcal{H}_\infty$ norm from $\underline{f}$ to $\underline{\eta}^1$ in (4.29) and from $\overline{f}$ to $\overline{\eta}^1$ in (4.30). First, we recall the following *Bounded Real Lemma* for nonnegative systems [121, Theorem 10, Remark 12].

Lemma 5 *Consider the following* nonnegative *LTI system*

$$\begin{aligned} x_{t+1} &= A_z x_t + B_z u_t, \\ y_t &= C_z x_t + D_z u_t, \end{aligned} \tag{4.31}$$

with $A_z \in \mathbb{R}_+^{n\times n}$*,* $B_z \in \mathbb{R}_+^{n\times q}$*,* $C_z \in \mathbb{R}_+^{p\times n}$ *and* $D_z \in \mathbb{R}_+^{p\times q}$*. For a scalar* $\gamma > 0$*, the following statements are equivalent:*

(a) The system (4.31) is stable and the $\mathcal{H}_\infty$ *norm of its transfer function from* u *to* y *is strictly less than* γ*.*

(b) There exists a diagonal *matrix* $P \succ 0$ *such that the following matrix inequality is*

feasible:

$$\begin{bmatrix} A_z^T P A_z - P + C_z^T C_z & A_z^T P B_z + C_z^T D_z \\ * & Z \end{bmatrix} \prec 0, \tag{4.32}$$
$$Z = B_z^T P B_z + D_z^T D_z - \gamma^2 I_q.$$

We can now use this result to select the matrix $\mathcal{L}$ minimizing the $\mathcal{H}_\infty$ norm of the estimation error dynamics.

Theorem 16 *Let $n_f > n_a$. Consider the following SDP with variables $\Omega_1, \Omega_2 \in \mathbb{R}_+^{(n-n_a)\times(n_f-n_a)}$, $\Gamma \in \mathbb{R}_+$ and $P \in \mathbb{D}_{>0}^{(n-n_a)}$*

$$\inf_{\Gamma>0,\, P\in\mathbb{D}_{>0},\, \Omega_1\geq 0,\, \Omega_2\geq 0} \quad \Gamma, \tag{4.33}$$

$$s.t. \quad \begin{bmatrix} -P & \Theta_1 & \Theta_2 \\ * & (I_{n-n_a} - P) & 0_{(n-n_a)\times n'} \\ * & * & -\Gamma\, I_{n'} \end{bmatrix} \prec 0, \tag{4.34}$$

$$\Theta_1 \geq 0, \tag{4.35}$$

where

$$\Theta_1 = P\mathcal{A} + (\Omega_2 - \Omega_1)\mathcal{C},\ \Theta_2 = \begin{bmatrix} P & \Omega_1 & \Omega_2 \end{bmatrix}. \tag{4.36}$$

Suppose (4.33)–(4.35) *has an optimal solution $\Omega_1^\star$, $\Omega_2^\star$, $P^\star$, $\Gamma^\star$. Then, the gain matrix*

$$\mathcal{L}^\star = (P^\star)^{-1}(\Omega_1^\star - \Omega_2^\star),$$

satisfies Assumption 7 and the $\mathcal{H}_\infty$-norm of the estimation error dynamics (4.29)-(4.30) *is bounded by $\sqrt{\Gamma^*}$. In particular, for this $\mathcal{L}^\star$ we have $\underline{\eta}^1, \overline{\eta}^1 \in \mathcal{L}_\infty^{n-n_a}$.*

Proof. The $\mathcal{H}_\infty$-norms from $\underline{f}$ to $\underline{\eta}^1$ in (4.29) and from $\overline{f}$ to $\overline{\eta}^1$ in (4.30) are the same and equal to $\|\mathcal{F}(z)\|_\infty$ with $\mathcal{F}(z) = (zI - (\mathcal{A} - \mathcal{L}\mathcal{C}))^{-1}H$. Using a Schur complement, the inequality (4.32) is equivalent to

$$\begin{bmatrix} -P & PA_z & PB_z \\ * & C_z^{\mathrm{T}} C_z - P & C_z^{\mathrm{T}} D_z \\ * & * & D_z^{\mathrm{T}} D_z - \Gamma I_q \end{bmatrix} \prec 0, \tag{4.37}$$

where $\Gamma = \gamma^2$. Consequently, by applying Lemma 5, we deduce that if there exists a diagonal matrix $P \succ 0$ such that the following LMI is feasible:

$$\begin{bmatrix} -P & P(\mathcal{A}-\mathcal{L}\mathcal{C}) & PH \\ * & I_{n-n_a}-P & 0_{(n-n_a)\times n'} \\ * & * & -\Gamma I_{n'} \end{bmatrix} \prec 0, \tag{4.38}$$

we get $\|\mathcal{F}(z)\|_\infty < \Gamma$.

Recall that by definition we have $\mathcal{L} = \mathcal{L}^+ - \mathcal{L}^-$. Define the new variables $\Omega_1 = P\mathcal{L}^+$, $\Omega_2 = P\mathcal{L}^-$. Then $P(\mathcal{A}-\mathcal{L}\mathcal{C}) = \Theta_1$ and $PH = \Theta_2$, with Θ_1, Θ_2 defined in (4.36). Imposing $\mathcal{A}-\mathcal{L}\mathcal{C} \geq 0$ is equivalent to imposing $\Theta_1 \geq 0$ because $P \in \mathbb{D}_{>0}^{n-n_a}$. An optimal solution is in particular feasible, so it satisfies the linear matrix inequality (4.38) and hence by Lemma 5(a) the systems (4.29), (4.30) are also stable. Since the inputs $\underline{f}_t$ and $\overline{f}_t$ are bounded, we get $\underline{\eta}^1, \overline{\eta}^1 \in \mathcal{L}_\infty^{n-n_a}$. □

4.5 Estimation of attack signals

In this section we develop lower and upper bounds for the attack signals in (4.1), which we can rewrite as follows

$$\begin{bmatrix} x_{t+1} - Ax_t - Mw_t \\ y_t - Cx_t - Nw_t \end{bmatrix} = \begin{bmatrix} E \\ D \end{bmatrix} a_t, \quad \forall t \in \mathbb{Z}_+.$$

Assuming as in Section 4.3 that $\begin{bmatrix} E^{\mathrm{T}} & D^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ is full column rank, there exists $\mathcal{G} \in \mathbb{R}^{m\times(n+p)}$ such that $\mathcal{G}\begin{bmatrix} E^{\mathrm{T}} & D^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} = I_m$. Consequently, we get

$$\mathcal{G}\begin{bmatrix} x_{t+1} - Ax_t - Mw_t \\ y_t - Cx_t - Nw_t \end{bmatrix} = a_t. \tag{4.39}$$

Let $\mathcal{G} = \begin{bmatrix} \mathcal{G}_1 & \mathcal{G}_2 \end{bmatrix}$ with $\mathcal{G}_1 \in \mathbb{R}^{m\times n}$. The following proposition provides bounds on the attack signal in (4.1).

Proposition 6 *Assume that $\begin{bmatrix} E^T & D^T \end{bmatrix}^T$ is full column rank. Suppose that we have bounds $\underline{\chi}_t \leq x_t \leq \overline{\chi}_t$ for all $t \in \mathbb{Z}_+$, as provided for example by* (4.26) *under the conditions of Theorem 15. Then we have the bounds*

$$\underline{a}_t \leq a_t \leq \overline{a}_t, \forall t \in \mathbb{Z}_+, \tag{4.40}$$

with

$$\begin{aligned}\underline{a}_t &= \mathcal{G}_1^+\underline{\chi}_{t+1} - \mathcal{G}_1^-\overline{\chi}_{t+1} + (-\mathcal{G}_1 A - \mathcal{G}_2 C)^+\underline{\chi}_t - (-\mathcal{G}_1 A - \mathcal{G}_2 C)^-\overline{\chi}_t + (-\mathcal{G}_1 M - \mathcal{G}_2 N)^+\underline{w}_t \\ &\quad - (-\mathcal{G}_1 M - \mathcal{G}_2 N)^-\overline{w}_t + \mathcal{G}_2 y_t, \\ \overline{a}_t &= \mathcal{G}_1^+\overline{\chi}_{t+1} - \mathcal{G}_1^-\underline{\chi}_{t+1} + (-\mathcal{G}_1 A - \mathcal{G}_2 C)^+\overline{\chi}_t - (-\mathcal{G}_1 A - \mathcal{G}_2 C)^-\underline{\chi}_t + (-\mathcal{G}_1 M - \mathcal{G}_2 N)^+\overline{w}_t \\ &\quad - (-\mathcal{G}_1 M - \mathcal{G}_2 N)^-\underline{w}_t + \mathcal{G}_2 y_t.\end{aligned} \tag{4.41}$$

Proof. This results follows from Lemma 2 by rewriting (4.39) as

$$a_t = \mathcal{G}_1 x_{t+1} + (-\mathcal{G}_1 A - \mathcal{G}_2 C)x_t + (-\mathcal{G}_1 M - \mathcal{G}_2 N)w_t + \mathcal{G}_2 y_t.$$

□

Remark 6 *In case of a sensor attack, i.e., when $D \neq 0$ and $E = 0$, one can select $\mathcal{G}_1 = 0$ in* (4.41) *and $\mathcal{G}_2$ such that $\mathcal{G}_2 D = I_m$.*

4.6 Numerical simulations

This section illustrates the results of this chapter via simulations on an abstract model. The matrices A, C, M, N, D and E of the system (4.1) are defined as follows

$$A = \begin{bmatrix} 0.9 & 0.3 & 0.9 & 0.2 \\ 0 & 0.5 & 0.03 & 0.36 \\ 0 & 0.2 & 0.1 & 0.67 \\ 0 & 0.32 & 0 & 0.5 \end{bmatrix},\ C = I_4,\ M = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix},$$

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},\ E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},\ D = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The unknown input signal $w_t^{\mathrm{T}} = \begin{bmatrix} -1 & -1 \end{bmatrix}$ for all $t \geq 0$, with given bounds $\underline{w}_t^{\mathrm{T}} = \begin{bmatrix} -1 & -1 \end{bmatrix}$ and $\overline{w}_t^{\mathrm{T}} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ for all $t \geq 0$. The unknown initial condition x_0 and its given bounds are selected as follows

$$x_0^{\mathrm{T}} = \begin{bmatrix} 2 & 2 & 2 & 2 \end{bmatrix},\ \underline{x}_0^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix},\ \overline{x}_0^{\mathrm{T}} = \begin{bmatrix} 3 & 3 & 3 & 3 \end{bmatrix}.$$

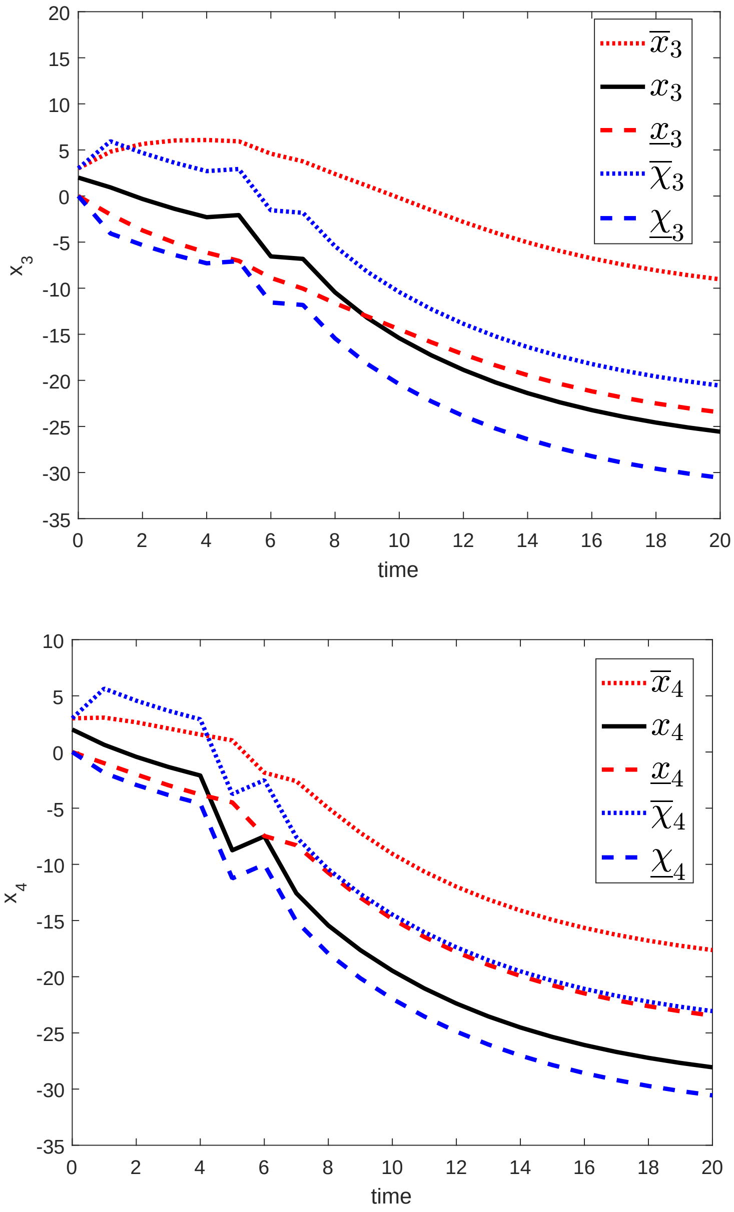


Figure 4.1 : Trajectory estimates of x_3 and x_4 under attack, with the standard observer estimates $\underline{x}_i$, $\overline{x}_i$ and the attack-resilient observer estimates $\underline{\chi}_i$, $\overline{\chi}_i$

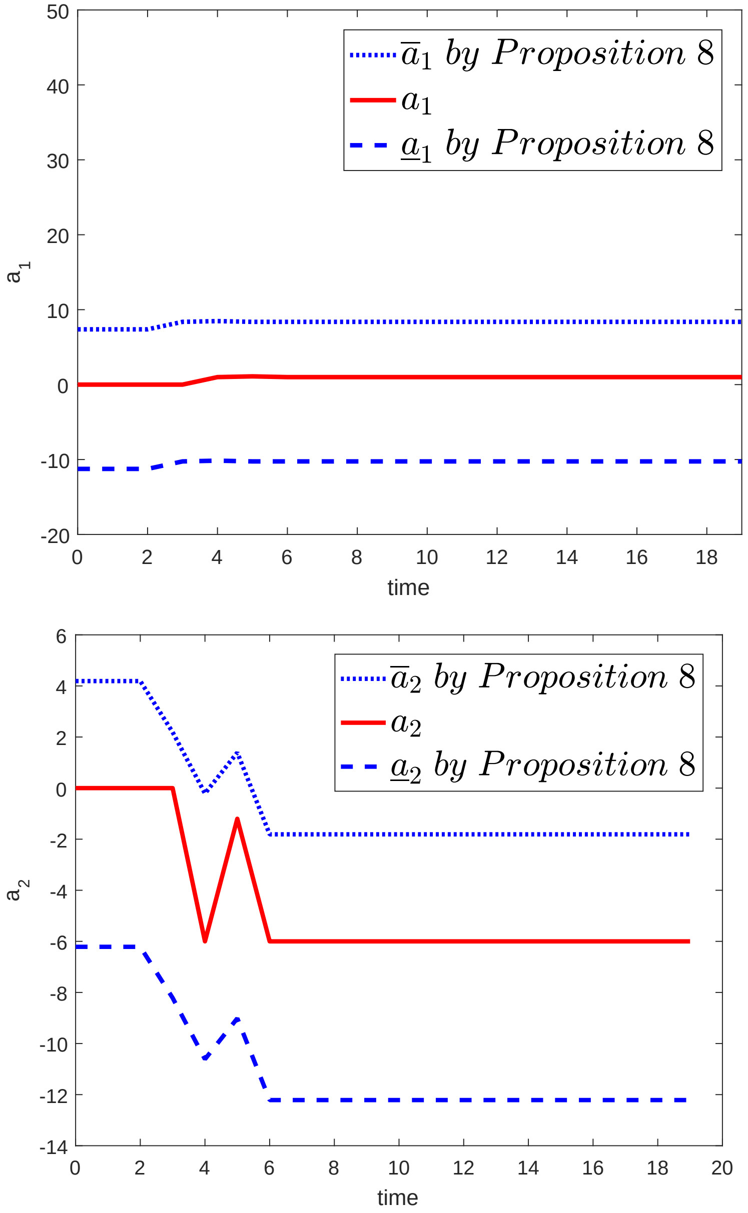


Figure 4.2 : Optimized stealthy attack

The following matrix L is selected such that Assumption 4 is satisfied:

$$L = \begin{bmatrix} 0.5 & 0.3 & 0.8 & -0.5 \\ -0.1 & 0.2 & 0 & 0.3 \\ 0 & 0.2 & -0.5 & 0.4 \\ 0 & 0.1 & 0 & 0.3 \end{bmatrix}.$$

We design a stealthy sensor attack by using the methodology of Section 4.2, selecting $d = 5$ for the monitor of Proposition 5. The effects of the optimally disruptive attack signal can be seen in Figure 4.1: the inclusion relation (4.7) is violated for x_3 and x_4 when using standard interval observers. Furthermore, we design the attack-resilient interval observer (4.26) by solving the SDP (4.33)–(4.35) using the SeduMi solver [122] together with the YALMIP toolbox [123] as the interface. We select the matrices F, T and V as follows

$$F = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \end{bmatrix},\ T = I_4,\ V = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The matrices $\mathcal{A}$ and $\mathcal{C}$ are as follows

$$\mathcal{A} = \begin{bmatrix} 1.1 & 1.2 \\ 0.36 & 0.53 \end{bmatrix},\ \mathcal{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

As optimal gain for the interval observer (4.22), we obtain $\mathcal{L}^\star = \begin{bmatrix} 1.1 & 0.36 \end{bmatrix}^T$, together with an optimal $\mathcal{H}_\infty$-norm $\gamma^\star = \sqrt{\Gamma^*} = 4$. Notice in Figure 4.1 that unlike the bounds provided by a standard interval observer (4.6), which are clearly violated for x_3 and x_4, the bounds provided by the attack-resilient interval observer (4.26) always include the actual value of the state x. Moreover, we use (4.41) to compute bounds for the attack signals shown in Figure 4.2.

4.7 Conclusion

This chapter discusses the problem of state estimation for discrete LTI systems with unknown but bounded uncertainties under attacks. We construct stealthy attack signals and we design interval observers for the state of the system under sensor and actuator stealthy attacks. Moreover, we compute the corresponding stealthy attack-resilient interval observer gains that minimize the $\mathcal{H}_\infty$ norm of the estimation error dynamics as solutions of semi-definite programs. We illustrate the threat of these cyber-attacks through numerical simulations.

CHAPTER 5 CONCLUSION

5.1 Summary

In this thesis, we study first the problem of privacy-preserving observer and controller design in a multi-agent system composed of uncertain linear systems. The system is composed of independent linear Gaussian individual systems and we solve the problem of differentially private state estimation (Kalman filtering) by proposing a two-stage architecture in Chapter 2. The two-stage architecture first aggregates and combines the individual agent signals before adding privacy-preserving noise and post-filtering the result to be published. We prove that the optimal input aggregation stage can be computed by solving an SDP and we show a significant performance improvement offered by our architecture over input perturbation schemes as the number of input signals increases. Then, by using the separation principle, we adapt this architecture to solve the LQG control problem, by estimating a certain linear combination of the agent states as for the case of differentially private Kalman filtering, but for a specific cost on the estimation error.

In Chapter 3, we consider a multi-agent system composed of interconnected linear individual systems with unknown but bounded uncertainties. We design a privacy-preserving interval observer architecture, by adding a bounded privacy-preserving noise to each participant's data, which is subsequently taken into account by the observer. We provide characteristics for the bounded privacy-preserving noise and we illustrate that the performance of the input perturbation mechanism can be improved by suitably aggregating the measurements data before adding the bounded privacy preserving noise. The estimates published by the observer guarantee differential privacy for the agents' data by the post-processing property.

In Chapter 4, we focus on the problems of designing undetectable attacks and attack-resilient observers for CPS. We design attack signals against LTI systems with unknown uncertainties that are undetected by a monitor based on parity equations. Only bounds on the uncertainties are given. Furthermore, we design attack-resilient interval observers for such systems and we compute observer gains that minimize the $\mathcal{H}_\infty$-norms of the estimation error dynamics, by solving a semi-definite program (SDP). In addition, we evaluate the vulnerability of the system by computing bounds for the admissible stealthy attack signals.

5.2 Future Directions

For the problem of Kalman Filtering and LQG control that we solve in Chapter 3, future research could consider the extension of these ideas to nonlinear systems, improving on the input and output perturbation mechanisms of [54]. Indeed, we have seen that two-stage architectures offer significant performance improvements over these mechanisms, yet there is currently no example of such an architecture for a nonlinear control system. Challenges include computing tight bounds on sensitivity for nonlinear systems, as well as characterizing overall performance in order to optimize the architecture.

In addition, since the size of the SDP increases rapidly with the number of agents (and the time horizon in the non-stationary case), it would be useful to develop numerical methods and a problem-specific solver that can take advantage of the sparsity of the matrices involved in the constraints, as in [124] for example. Furthermore, it is of interest for future work to consider dynamic pre-filters. Such a pre-filter can lead to a better result than the static pre-filter that we have designed in this thesis. Indeed, in the initial work of [52] for stationary signals, the optimal pre-filters are found to be dynamic. It is also of interest for future work to consider a multi-agent system with individual linear Gaussian systems that are *dependent*. Such a situation occurs when for example one considers the problem of traffic control based on location data collected from vehicles on a road, with the speed and the position of each vehicle being impacted by the speed and the position of the other vehicles on the road. A suitable adjacency relation needs to be defined and the structure of the optimal static pre-filter needs to be discussed in such a case.

A future direction for the problem of control of a multi-agent system under differential privacy constraints is to consider the same problem for individual systems whose models are not known a priori, but for which only input-output data is available. Machine learning algorithms [125, 126] have been proposed to control networks of model-free dynamic processes, specifically by learning the models from data before designing a controller. Nevertheless, situations in which the privacy of individuals' data needs to be preserved have not been discussed. To solve the problem, first, we can add a privacy-preserving noise to each individual signal by using the Gaussian mechanism. Next, we can use machine learning techniques based on ideas from [127] for example to estimate each individual model. Then, we can design a model predictive controller based on the learned prediction model.

For the problem of differentially private interval observer that is considered in Chapter 3, a future direction is to design a two-stage architecture that first aggregates and combines the individual agent signals before adding privacy-preserving noise and sending the result to an interval observer (only a scalar example has been given to introduce such an architecture in Chapter 3). A future direction is also to design an output perturbation architecture to solve the problem. Such an architecture consists in sending directly the individual agent signals to an interval observer, and then add a privacy-preserving noise to the estimates provided by the observer. A challenge in this case is to preserve the ordering of the interval bounds. A terminal filter can be added to smooth out the privacy-preserving noise when using this output perturbation architecture.

Finally, for the problem of undetected attack signals and attack-resilient interval observer design that is considered in Chapter 4, it is of interest for future work to consider uncertain linear time-varying systems, with the state matrix A_t being unknown but subject to known bounds. This additional uncertainty offers more opportunities to hide an attack for the adversary and the design of an attack-resilient interval observer may be more challenging in this case. It is also of interest to consider uncertain nonlinear systems for this problem.

REFERENCES

[1] J. C. Herrera *et al.*, “Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568 – 583, 2010.

[2] R. Shokri *et al.*, “Quantifying location privacy,” in *Proceedings of the IEEE Symposium on Security and Privacy*, Oakland, California, May 2011, pp. 247–262.

[3] Y.-A. de Montjoye *et al.*, “Unique in the crowd: The privacy bounds of human mobility,” *Scientific Reports*, vol. 3, 2013.

[4] F. Xu *et al.*, “Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data,” in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1241–1250.

[5] A. Pyrgelis, C. Troncoso, and E. D. Cristofaro, “What does the crowd say about you? evaluating aggregation-based location privacy,” *Proceedings on Privacy Enhancing Technologies*, vol. 4, pp. 156–176, 2017.

[6] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, December 1992.

[7] G. Bauer, K. Stockinger, and P. Lukowicz, “Recognizing the use-mode of kitchen appliances from their current consumption,” *EuroSSC*, pp. 163–176, 2009.

[8] M. A. Lisovich, D. K. Mulligan, and S. B. Wicker, “Inferring personal information from demand-response systems,” *IEEE Security and Privacy*, vol. 8, no. 1, pp. 11–20, 2010.

[9] A. Molina-Markham *et al.*, “Private memoirs of a smart meter,” in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, New York, NY, USA, 2010, pp. 61–66.

[10] Dell, “2015 Dell security annual threat report,” Tech. Rep., 2015.

[11] A. A. Cárdenas *et al.*, “Attacks against process control systems: Risk assessment, detection, and response,” in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, Hong Kong, China, 2011, pp. 355–366.

[12] A. A. Cárdenas, S. Amin, and S. Sastry, “Research challenges for the security of control systems,” in *Proceedings of the 3rd Conference on Hot Topics in Security*, San Jose, CA, USA, 2008.

[13] D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, “Drone hack,” *GPS World*, vol. 23, no. 8, pp. 30–33, 2012.

[14] S. Peterson and P. Faramarzi, “Iran hijacked US drone, says iranian engineer,” *The Christian Science Monitor*, vol. 15, Dec. 2011.

[15] M. M. Rana and L. Li, “An overview of distributed microgrid state estimation and control for smart grids,” *Sensors (Basel)*, vol. 15, no. 2, pp. 4302–4325, 2015.

[16] P. J. Costa, J. P. Dunyak, and M. Mohtashemi, “Models, prediction, and estimation of outbreaks of infectious disease,” in *Proceedings of IEEE SoutheastCon*, Ft. Lauderdale, FL, USA, April 2005.

[17] P.-A. Bliman and B. D’Avila Barros, “Interval observer for SIR epidemic model subject to uncertain seasonality,” in *Proc. of the 5th International Symposium on Positive Systems Theory and Applications (POSTA 2016)*, Roma, Italy, September 2016.

[18] K. H. Degue and J. Le Ny, “An interval observer for discrete-time SEIR epidemic models,” in *Proceedings of American Control Conference (ACC 2018)*, Milwaukee, Wisconsin, USA, Jun. 2018.

[19] ——, “Estimation and outbreak detection with interval observers for uncertain discrete-time SEIR epidemic models,” *International Journal of Control*, pp. 1–12, 2019.

[20] J. Terstegge, “Privacy in the law,” in *Security, Privacy, and Trust in Modern Data Management*, M. Petkovic and W. Jonker, Eds. Springer, 2007, pp. 11–20.

[21] J. S. Comas, “Improving data utility in differential privacy and k-anonymity,” Ph.D. dissertation, Universitat Rovira i Virgili, Tarragona, July 2013.

[22] S. D. Warren and L. D. Brandeis, “The right to privacy,” in *Harvard Law Review*, 1890, vol. IV, pp. 193–220.

[23] “Timeline: Privacy and the law,” *NPR*, 2009.

[24] “Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.” *Official Journal of the European Communities*, pp. 31–50, October 1995.

[25] “Standard for privacy of individually identifiable health information,” *Federal Register, Special Edition*, pp. 768–769, October 2007.

[26] T. Dalenius, “Towards a methodology for statistical disclosure control,” *Statistik Tidskrift*, vol. 15, pp. 429–444, 1977.

[27] C. Dwork, “Differential privacy,” in *Automata, Languages and Programming*, ser. Lecture Notes in Computer Science, B. Bugliesi, M. and. Preneel, V. Sassone, and I. Wegener, Eds. Berlin / Heidelberg: Springer, 2006, vol. 4052, pp. 1–12.

[28] L. Sweeney, “k-anonymity: a model for protecting privacy,” *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, pp. 557–570, 2002.

[29] H. T. Greely, “The uneasy ethical and legal underpinnings of large-scale genomic biobanks,” *Annual Review of Genomics and Human Genetics*, vol. 8, no. 1, pp. 343–364, 2007.

[30] L. Sweeney, “Uniqueness of simple demographics in the us population,” *Technical report, Carnegie Mellon University*, 2000.

[31] L. Willenborg and T. De Waal, “Elements of statistical disclosure control,” *Springer Science & Business Media*, vol. 155, 2012.

[32] S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[33] B. C. M. Fung *et al.*, “Privacy-preserving data publishing: A survey of recent developments,” *ACM Computing Surveys*, vol. 42, no. 4, 2010.

[34] L. Backstrom, C. Dwork, and J. Kleinberg, “Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography,” in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW ’07. New York, NY, USA: Association for Computing Machinery, 2007, p. 181–190.

[35] L. Sweeney, “Only you, your doctor, and many others may know,” *Technology Science*, September 2015.

[36] W. Jiang and C. Clifton, *Privacy-Preserving Distributed k-Anonymity.* Berlin, Heidelberg: Springer, 2005, pp. 166–177. [Online]. Available: http://dx.doi.org/10.1007/11535706_13

[37] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, pp. 571–588, 2002.

[38] L. Sweeney, M. Von Loewenfeldt, and M. Perry, "Saying it's anonymous doesn't make it so: Re-identifications of "anonymized" law school data," *Technology Science*, November 2018.

[39] A. Machanavajjhala *et al.*, "l-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, March 2007.

[40] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *ICDE*, R. Chirkova *et al.*, Eds. IEEE, 2007, pp. 106–115.

[41] S. Venkatasubramanian, "Measures of anonymity," in *Privacy-Preserving Data Mining: Models and Algorithms*, ser. Advances in Database Systems, C. C. Aggarwal and P. S. Yu, Eds. US: Springer, 2008, vol. 34, pp. 81–103.

[42] L. Sankar, S. Raj Rajagopalan, and V. H. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, June 2013.

[43] F. Farokhi and H. Sandberg, "Fisher information privacy with application to smart meter privacy using HVAC units," in *Privacy in Dynamical Systems*, F. Farokhi, Ed. Springer, 2020, pp. 3–17.

[44] Y. Mo and R. M. Murray, "Privacy preserving average consensus," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 753–765, 2016.

[45] Y. Song, C. X. Wang, and W. P. Tay, "Compressive privacy for a linear dynamical system," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 895 – 910, 2020.

[46] F. Fei *et al.*, "A k-anonymity based schema for location privacy preservation," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 2, pp. 156–167, Apr. 2019.

[47] C. Dwork *et al.*, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Third Theory of Cryptography Conference*, 2006, pp. 265–284.

[48] C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, ser. Lecture Notes in Computer Science, vol. 4052. Springer-Verlag, 2006.

[49] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, August 2014.

[50] C. Dwork *et al.*, "Differential privacy under continual observations," in *Proceedings of the ACM Symposium on the Theory of Computing (STOC)*, Cambridge, MA, June 2010.

[51] L. Fan and L. Xiong, "An adaptive approach to real-time aggregate monitoring with differential privacy," *IEEE Transactions on knowledge and data engineering*, vol. 26, no. 9, pp. 2094–2106, 2014.

[52] J. Le Ny and G. J. Pappas, "Differential private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, February 2014.

[53] J. Le Ny and M. Mohammady, "Differentially private MIMO filtering for event streams," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, 01 2018.

[54] J. Le Ny, "Differentially private nonlinear observer design using contraction analysis," *International Journal of Robust and Nonlinear Control*, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rnc.4392

[55] A. McGlinchey and O. Mason, "Bounding the $\ell 2$ sensitivity for positive linear observers," in *European Control Conference (ECC)*, 2018, pp. 1214–1219.

[56] Y. Wang *et al.*, "Differential privacy in linear distributed control systems: Entropy minimizing mechanisms and performance tradeoffs," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 118–130, Mar. 2017.

[57] M. T. Hale, A. Jones, and K. Leahy, "Privacy in feedback: The differentially private LQG," in *Proceedings of the American Control Conference (ACC)*, Milwaukee, WI, USA, Jun. 2018.

[58] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, 2012, pp. 81—-90.

[59] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design," *Automatica*, vol. 81, pp. 221–231, 2017.

[60] R. Cummings *et al.*, “Differentially private change-point detection,” in *Advances in Neural Information Processing Systems*, 2018.

[61] K. H. Degue and J. Le Ny, “On differentially private Gaussian hypothesis testing,” in *Proceedings of the 56th Annual Allerton Conference on Communication, Control, and Computing*, Allerton Park and Retreat Center, Monticello, Illinois, USA, Oct. 2018.

[62] V. Rostampour *et al.*, “Differentially-private distributed fault diagnosis for large-scale nonlinear uncertain systems,” in *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, 2018.

[63] C. Dwork *et al.*, “Our data, ourselves: Privacy via distributed noise generation,” *Advances in Cryptology-EUROCRYPT*, vol. 4004, pp. 486–503, 2006.

[64] B. D. O. Anderson and J. B. Moore, *Optimal filtering.* New York, NY, USA: Dover: Dover, 2005.

[65] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach.* American Institute of Aeronautics and Astronautics, 2000.

[66] M. Rana, L. Li, and S. Su, “Kalman filter based microgrid state estimation and control using the iot with 5g networks,” in *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2015, pp. 1–5.

[67] K. H. Degue and J. Le Ny, “On differentially private Kalman filtering,” in *Proceedings of the 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, Canada, Nov. 2017.

[68] A. McGlinchey and O. Mason, “Differential privacy and the l_1 sensitivity of positive linear observers,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3111–3116, Jul. 2017, 20th IFAC World Congress.

[69] J. Le Ny, “Differentially private nonlinear observer design using contraction analysis,” *International Journal of Robust and Nonlinear Control*, pp. 1–19, Nov. 2018.

[70] F. Mazenc, T. N. Dinh, and S. I. Niculescu, “Robust interval observers and stabilization design for discrete-time systems with input and output,” *Automatica*, vol. 49, pp. 3490–3497, Sep. 2013.

[71] J. P. Conti, “The day the samba stopped,” *Engineering Technology*, vol. 5, no. 4, pp. 46–47, 2010.

[72] S. Kuvshinkova, “SQL slammer worm lessons learned for consideration by the electricity sector,” North American Electric Reliability Council, Atlanta, GA, Tech. Rep., 2003.

[73] S. Amin, A. A. Cárdenas, and S. Sastry, “Safe and secure networked control systems under denial-of-service attacks,” in *Proceedings of Hybrid Systems: Computation and Control*, San Francisco, CA, USA, April 2009, pp. 31–45.

[74] A. Teixeira *et al.*, “A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator,” *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 11 271–11 277, 2011.

[75] ——, “Revealing stealthy attacks in control systems,” in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, Allerton Park and Retreat Center, Monticello, IL, USA, Oct. 2012.

[76] S. D. Bopardikar and A. Speranzon, “On analysis and design of stealth-resilient control systems,” in *2013 6th International Symposium on Resilient Control Systems (ISRCS)*, 2013, pp. 48–53.

[77] J. Y. Keller and D. Sauter, “Monitoring of stealthy attack in networked control systems,” in *2013 Conference on Control and Fault-Tolerant Systems (SysTol)*, 2013, pp. 462–467.

[78] Y. Mo and B. Sinopoli, “False data injection attacks in control systems,” in *First workshop on secure control systems*, Stockholm, Sweden, 2010.

[79] Y. Mo *et al.*, “Cyber–physical security of a smart grid infrastructure,” *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[80] R. Smith, “A decoupled feedback structure for covertly appropriating network control systems,” in *IFAC World Congress*, 2011, pp. 90–95.

[81] M. Zhu and S. Martínez, “Stackelberg-game analysis of correlated attacks in cyber-physical systems,” in *Proceedings of the 2011 American Control Conference*, 2011, pp. 4063–4068.

[82] H. Fawzi, P. Tabuada, and S. Diggavi, “Secure estimation and control for cyber-physical systems under adversarial attacks,” *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[83] A. Dutta and C. Langbort, “Stealthy output injection attacks on control systems with bounded variables,” *International Journal of Control*, vol. 90, no. 7, pp. 1389–1402, 2017.

[84] A. Abur and A. G. Exposito, *Power system state estimation:Theory and implementation.* Boca Raton, FL: CRC Press, Mar. 2004.

[85] Y. Mo, S. Weerakkody, and B. Sinopoli, “Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs,” *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, Feb. 2015.

[86] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Chicago, Illinois, USA, 2009, pp. 21–32.

[87] Y. Chen, S. Kar, and J. M. F. Moura, “Optimal attack strategies subject to detection constraints against cyber-physical systems,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1157–1168, Sep. 2018.

[88] V. Puig *et al.*, “Robust fault detection for LPV systems using a consistency-based state estimation approach and zonotopes,” in *Proc. of the 10th European Control Conference (ECC)*, Budapest, Hungary, Aug. 2009, pp. 3184–3189.

[89] Q. Zhu and T. Basar, “Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems,” *IEEE Control Systems*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[90] M. Pajic *et al.*, “Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators,” *IEEE Control Systems*, vol. 37, no. 2, pp. 66–81, Apr. 2017.

[91] M. Milanese and C. Novara, “Unified set membership theory for identification, prediction and filtering of nonlinear systems,” *Automatica*, vol. 47, no. 10, pp. 2141 – 2151, Oct. 2011.

[92] M. Pourasghar, V. Puig, and C. Ocampo-Martinez, “Comparison of set-membership and interval observer approaches for state estimation of uncertain systems,” in *Proceedings of the 15th European Control Conference (ECC)*, Jun. 2016, pp. 1111–1116.

[93] F. Mazenc and O. Bernard, “Interval observers for linear time-invariant systems with disturbances,” *Automatica*, vol. 47, no. 1, pp. 140–147, 2011.

[94] T. Raïssi, D. Efimov, and A. Zolghadri, “Interval state estimation for a class of nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 260–265, 2012.

[95] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications.* New York: Wiley, 2000.

[96] D. Efimov *et al.*, “Interval estimation for LPV systems applying high order sliding mode techniques,” *Automatica*, vol. 48, pp. 2365–2371, 2012.

[97] E. D. Sontag, “On the Input-to-State Stability Property,” *European Journal of Control*, vol. 1, no. 1, pp. 24–36, 1995.

[98] C. Li *et al.*, “The matrix mechanism: optimizing linear counting queries under differential privacy,” *The VLDB Journal*, vol. 24, pp. 757 – 781, 2015.

[99] J. Le Ny, E. Feron, and M. Dahleh, “Scheduling continuous-time Kalman filters,” *IEEE Transactions on Automatic Control*, vol. 56, no. 6, June 2011.

[100] A. I. Mourikis and S. I. Roumeliotis, “Optimal sensor scheduling for resource-constrained localization of mobile robot formations,” *IEEE Transactions on Robotics*, vol. 22, no. 5, pp. 917–931, 2006.

[101] T. Tanaka *et al.*, “Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs,” *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1896–1910, April 2017.

[102] Q. Geng *et al.*, “Privacy and utility tradeoff in approximate differential privacy,” *ArXiv e-prints*, Feb. 2019.

[103] M. Hou and P. C. Muller, “Design of observers for linear systems with unknown inputs,” *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 871–875, Jun. 1992.

[104] W. Zhang *et al.*, “A state augmentation approach to interval fault estimation for descriptor systems,” *European Journal of Control*, vol. 51, pp. 19 – 29, Jan. 2020.

[105] J. Le Ny and G. J. Pappas, “Differentially private Kalman filtering,” in *Proceedings of the 50th Annual Allerton Conference*, Allerton House, UIUC, Illinois, USA, October 2012.

[106] S. Boyd *et al.*, *Linear matrix Inequalities in system and control theory.* SIAM, 1994, vol. 15.

[107] V. Dukic, H. F. Lopes, and N. G. Polson, "Tracking epidemics with Google flu trends data and a state-space SEIR model," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1410–1426, 2012.

[108] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.

[109] Z. Hu, Z. Teng, and H. Jiang, "Stability analysis in a class of discrete SIRS epidemic models," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 5, pp. 2017–2033, 2012.

[110] K. J. Åström, *Introduction to stochastic control theory.* Academic Press, 1970.

[111] N. Holohan *et al.*, "The Bounded Laplace Mechanism in Differential Privacy," *ArXiv e-prints*, Aug. 2018.

[112] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance.* Springer-Verlag Berlin Heidelberg, 2006, no. 1.

[113] T. Raïssi, G. Videau, and A. Zolghadri, "Interval observers design for consistency checks of nonlinear continuous-time systems," *Automatica*, vol. 46, no. 3, pp. 518–527, 2010.

[114] D. Efimov, T. Raïssi, and A. Zolghadri, "Control of nonlinear and LPV systems: interval observer-based framework," *IEEE Trans. Automatic Control*, vol. 58, no. 3, pp. 773–782, 2013.

[115] A. Teixeira *et al.*, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, Beijing, China, Apr. 2012, pp. 55–64.

[116] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[117] S. Mishra *et al.*, "Secure state estimation and control using multiple (insecure) observers," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, California, USA, Dec. 2014.

[118] M. Sain and J. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Transactions on Automatic Control*, vol. 14, no. 2, pp. 141–149, 1969.

[119] W. Tang *et al.*, "Interval estimation methods for discrete-time linear time-invariant systems," *IEEE Transactions on Automatic Control*, 2019.

[120] Z. Wang, C.-C. Lim, and Y. Shen, "Interval observer design for uncertain discrete-time linear systems," *Systems & Control Letters*, vol. 116, pp. 41 – 46, Jun. 2018.

[121] C. Wang, "Bounded real lemma for positive discrete systems," *IET Control Theory Applications*, vol. 7, no. 4, pp. 502–507, March 2013.

[122] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[123] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in matlab," *Automatic Control Laboratory,ETHZ*, 2004.

[124] S. Benson, Y. Ye, and X. Zhang, "Solving large-scale sparse semidefinite programs for combinatorial optimization," *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 443–461, 2000.

[125] H. Hasanbeig *et al.*, "Reinforcement learning for temporal logic control synthesis with probabilistic satisfaction guarantees," in *Proceedings of the 58th IEEE Conference on Decision and Control*, Nice, France, Dec. 2019.

[126] M. Eisen *et al.*, "Learning in wireless control systems over non-stationary channels," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1123–1137, Mar. 2019.

[127] J. M. Manzano *et al.*, "Robust learning-based MPC for nonlinear constrained systems," *Automatica*, vol. 117, p. 108948, Jul. 2020.

APPENDIX A PUBLICATIONS

Publications Related to Chapters 2, 3 and 4

1. Kwassi H. Degue, Jerome Le Ny and Denis Efimov. “Stealthy Attacks and Attack-Resilient Interval Observers”, Submitted to Automatica, January 2021.

2. Kwassi H. Degue and Jerome Le Ny. “Differentially Private Kalman Filtering and LQG Control with Signal Aggregation”, Submitted to IEEE Transactions on Automatic Control, March 2020.

3. Kwassi H. Degue and Jerome Le Ny. “Differentially private interval observer design with input perturbation”, In Proceedings of the 2020 American Control Conference (ACC 2020), Denver, Colorado, USA, July 2020.

4. Kwassi H. Degue, Denis Efimov, Jerome Le Ny and Eric Feron. “Interval Observers for Secure Estimation in Cyber-Physical Systems”, In Proceedings of the 57th IEEE Conference on Decision and Control (CDC 2018), Miami Beach, Florida, USA, December 2018.

5. Kwassi H. Degue and Jerome Le Ny. “On Differentially Private Kalman Filtering”, In Proceedings of the 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP 2017), November 2017, Montreal, Canada.

Other Papers published as a PhD Student

6. Kwassi H. Degue, Karthik Gopalakrishnan, Max Z. Li, Hamsa Balakrishnan and Jerome Le Ny. “Differentially Private Outlier Detection in Multivariate Gaussian Signals”, Accepted for presentation at the 2021 American Control Conference (ACC 2021).

7. Kwassi H. Degue, Denis Efimov and Jerome Le Ny. “An interval observer-based feedback control for rehabilitation in Tremor”, In Proceedings of the 2020 European Control Conference (ECC 2020), Saint Petersburg, Russia, May 2020.

8. Kwassi H. Degue, Denis Efimov and Abderrahman Iggidr. “Interval Observer Design for Sequestered Erythrocytes Concentration Estimation in Severe Malaria Patients”, European Journal of Control, Elsevier, September 2020.

9. Kwassi H. Degue and Jerome Le Ny. “Estimation and outbreak detection with interval observers for uncertain discrete-time SEIR epidemic models”, International Journal of Control, July 2019.

10. Kwassi H. Degue, Denis Efimov and Jean-Pierre Richard. “Stabilization of linear impulsive systems under dwell-time constraints: Interval observer-based framework”, European Journal of Control, Elsevier, vol. 42, pp. 1-14, July 2018.

11. Kwassi H. Degue and Jerome Le Ny. “On Differentially Private Gaussian Hypothesis Testing”, In Proceedings of the 56th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, USA, October 2018.

12. Kwassi H. Degue and Jerome Le Ny. “An Interval Observer for Discrete-Time SEIR Epidemic Models”, In Proceedings of the 2018 American Control Conference (ACC 2018), Milwaukee, Wisconsin, USA, June 2018.

13. Kwassi H. Degue, Denis Efimov and Jerome Le Ny. “Interval Observer approach to Output Stabilization of Linear Impulsive Systems”, In Proceedings of the 20th IFAC World Congress, July 2017, Toulouse, France.

14. Kwassi H. Degue, Denis Efimov and Jean-Pierre Richard. “Interval observers for linear impulsive systems”, In Proceedings of the 10th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2016), August 2016, Monterey, California, USA.

15. Kwassi H. Degue, Denis Efimov and Abderrahman Iggidr. “Interval estimation of sequestered infected erythrocytes in malaria patients”, In Proceedings of the 15th IEEE European Control Conference (ECC16), June 2016, Aalborg, Denmark.

Book Chapter submitted as a PhD Student

16. Riccardo Ferrari, Kwassi H. Degue and Jerome Le Ny. “Differentially Private Anomaly Detection for Interconnected Systems”, Submitted, May 2020.