Doctoral theses at NTNU, 2022:325

Victor Wattin Håkansson

Transmission Schemes for Resource-constrained Wireless Sensor Networks

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Norwegian University of
Science and Technology

Victor Wattin Håkansson

Transmission Schemes for Resource-constrained Wireless Sensor Networks

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

ISBN 978-82-326-6262-3 (printed ver.)
ISBN 978-82-326-5221-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2022:325

Printed by NTNU Grafisk senter

Till mormor Ulla

Never go to sleep
without a request
to your subconscious.

THOMAS A. EDISON

Abstract

Wireless sensor networks (WSN) provide a versatile monitoring infrastructure to track physical processes for autonomous and manual decision-making. In WSN and wireless networked control systems, sensors observe physical processes and transmit measurements to remote estimators or a fusion center that tracks process parameters. The information update rate from sensors is limited by the number of communication channels and the sensor nodes' energy storage capabilities. Commonly, sensors share a limited number of communication channels, and if the number of transmitting sensors exceeds the number of channels, interference occurs. Furthermore, sensor nodes in WSN are usually not supported by power grids, and instead, they rely on small batteries, from which radio communication consumes a significant amount of energy. Therefore, to prolong the lifetime of sensor nodes, there is a need for approaches that reduce the number of data transmissions from the sensors without compromising the data accuracy. Also, given the limited resources for communication channels, the data transmission should be coordinated to maximize system utility.

This thesis develops transmission schemes that determine when a sensor should transmit an observation to achieve a high level of state-estimation in channel- and energy-constrained WSN for remote estimation. For a channel-constrained WSN, we exploit spatio-temporal dependencies among sensors to improve the overall estimation accuracy for remote estimators tracking different processes. We derive an optimal scheduling policy that minimizes the time average mean squared error (MSE) by modeling the scheduling problem as a Markov decision process. We also consider event-triggered transmission schemes, where a sensor transmits a measurement if it exceeds a predefined threshold. As an extension of the dual prediction scheme framework, a cost-aware dual prediction scheme is presented to further reduce data transmission in a WSN where sensors observe non-stationary processes. Finally, we consider a system of multiple sensors implementing threshold-based transmission over a limited number of shared channels, resulting in collisions. From statistical parameters, we derive optimal thresholds minimizing the MSE.

Acknowledgment

Similar to many people's experience, doing a PhD has been an emotionally bumpy ride. One may feel like a sailor, single-handedly crossing an ocean with a clear, straight, almost-visible navigation point ahead, but in whims of weather conditions and unexpected events, constantly forcing to re-route and practice the art of persistence. Fortunately, I was never alone, but guided and supported by many lodestars to help me fulfill my lifelong dream of becoming a scientist.

I want to express my gratitude to my main supervisor Stefan Werner for giving me the opportunity to be a PhD candidate, for his academic guidance and for being always available when I needed help. I want to thank Naveen Venkategowda for his friendship and scientific collaboration on all my academic work included in this thesis. I am thankful to Prof. Pramod Varshney for his valuable suggestions and feedback on my research. Also, I wish to thank my co-supervisors Frank Alexander Kræmer and Ingelin Steinsland for their support and mentorship during my PhD studies. Special thanks to all the members of the assessment committee: Assoc. Prof. Nikolaos Pappas and Assoc. Prof. Donatella Darsena, together with Kimmo Kansanen and Pierluigi Salvo Rossi.

I gratefully acknowledge the support from NTNU Digital for enabling all of this. Special thanks to all at the Department of Electronic Systems (IES) at NTNU in Trondheim for supporting me.

I am grateful to my family and friends for cheering for me on every adventure I embark. Also, to all the new friendships built during my PhD, which led me explore Trondheim and colored it with joyful memories. Especially, my fantastic roommate Míska, with whom I have spent numerous hours reflecting on schola vitae. Finally, I express my love for my dog Melody for her calmness, warmth and personal growth. Particularly, for taking me out on long walks in the Gibhlian forests of Trondheim, where I was able to relax, clear my head and shed light on ideas for my research.

Contents

Abstract **v**

Acknowledgment **vii**

List of Publications **xi**

List of Figures **xiii**

List of Tables **xv**

1 Introduction **1**

1.1 Scope and Objective 3

1.2 Contributions 4

1.3 Thesis Outline 6

2 Background **7**

2.1 The General System Model 7

2.2 Technical Concepts 9

2.3 Mathematical Tools 11

3 AoI-based Scheduling of Spatio-temporally Dependent Measurements **17**

3.1 Prior work 18

3.2 System Model and Problem Formulation 19
3.3 Optimal Scheduling of Multiple Sensors 21
3.3.1 Theoretical Results 22
3.3.2 Numerical Results 28
3.4 Scheduling of Two Sensors with a Transmission Constraint 29
3.4.1 Theoretical Results 30
3.4.2 Numerical Results 34
3.5 Discussion 35

4 Optimized Threshold-triggered Transmission Schemes 37
4.1 Cost-Aware Dual Prediction Scheme 38
4.1.1 Prior Work 38
4.1.2 System Model 39
4.1.3 Bootstrap-based Cost-Aware Dual Prediction Scheme . . . 41
4.1.4 Numerical Results 42
4.2 Optimal Threshold and Channel Allocation 45
4.2.1 Prior Work 45
4.2.2 System Model 47
4.2.3 Theoretical Results 48
4.2.4 Numerical Results 53
4.3 Discussion 54

5 Conclusions and Future Work 57

Appendix A Publications on AoI-based Scheduling of Spatio-temporally Dependent Measurements 71

Appendix B Publications on Optimized Threshold-triggered Transmission Schemes 109

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their roman numerals.

I V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal scheduling policy for spatio-temporally dependent observations using age-of-information," *IEEE 23rd International Conference on Information Fusion (FUSION)*, pp. 1154–1160, 2020, [Runner up - Best Paper Award]

II V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal scheduling of multiple spatio-temporally dependent observations using age-of-information," *54th Asilomar conference on signals, systems, and computers*, pp. 1188–1193, 2020

III V. W. Håkansson, N. K. D. Venkategowda, S. Werner and P. K. Varshney, "Optimal scheduling of multiple spatio-temporally dependent observations for remote estimation using age-of-information," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20308-20321, Oct., 2022

IV V. W. Håkansson, N. K. D. Venkategowda, S. Werner and P. K. Varshney, "Optimal transmission-constrained scheduling of spatio-temporally dependent observations using age-of-information," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15596-15606, Aug., 2022

V V. W. Håkansson, N. K. D. Venkategowda, F. A. Kræmer and S. Werner, "Cost-aware dual prediction scheme for reducing transmissions at IoT sensor nodes," *27th European Signal Processing Conference (EUSIPCO)*, pp. 2166–2171, 2019

VI V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal transmission threshold and channel allocation strategies for heterogeneous sensor data," *55th Asilomar Conference on Signals, Systems, and Computers*, pp. 757–761, 2021

List of Figures

1.1 WSN topology considered for this thesis. 3

2.1 Schematic of a generalized WSN system model considered in this thesis. 8

3.1 Schematic of WSN scheduling problem with $D = 2$. 19

3.2 Asymptotic average cost, $\lim_{T\to\infty} J(\gamma, T)$, vs λ_t with $N = 5$, $D = 2$, $\sigma_i = 1$, $\forall i = 1, 2, ..., N$, $\xi = 0.5$, and $r_0 = 0.5$. 29

3.3 Asymptotic average cost, $\lim_{T\to\infty} J(\gamma, T)$, vs r_0 with $N = 5$, $D = 2$, $\sigma_i = 1$, $\forall i = 1, 2, ..., N$, $\xi = 0.5$, and $\lambda_t = 0.8$. 29

3.4 Average cost $J(\gamma)$ versus n_1^γ for system parameters $\sigma_1 = 2, \sigma_2 = 1, \rho_{12} = -0.5$, $\bar{n}_i \geq T, i = 1, 2$, and $\xi = 0.5$ and $T = 100$. Solid lines show results derived from theory and markers show simulation results. Red asterix show optimal performance at n_1^*. . 35

3.5 Optimal number of scheduling instances of Sensor 1, n_1^*, versus spatial correlation ρ_{12} for system parameters $\sigma_1 = 2, \sigma_2 = 1, \rho_{12} = -0.5$, $\bar{n}_i \geq T, i = 1, 2$, and $\xi = 0.5$. 36

4.1 Percentage of data transmitted versus accuracy β for CM and CA-DPS with sliding window length $\eta = 10$, forecast horizon $T = 10$ and $L = 50$ trajectory simulations. 42

4.2 Percentage of data transmitted versus accuracy β for Mote ID 30 for CA-DPS with sliding window length $\eta = 30$, forecast horizon $T = 20$, and $L = 40$ trajectory simulations, LMS [56] with filter length $N = 4$, and step-size $\mu = 10^{-5}$ and AM-DR [33] with fixed window size $\eta_f = 4$, slow window size $\eta_s = 8$, and learning rate $\alpha = 10^{-7}$. 43

4.3 Percentage of data transmitted with accuracy set to $\beta = 0.1$ for all temperature sensors in the Intel office data set, between March 6 and 9, for CM and CA-DPS with sliding window length $\eta = 30$, forecast horizon $T = 20$ and $L = 40$ trajectory simulations. Numbers represent Mote IDs. 44

4.4 Schematic of WSN channel allocation problem. 47

4.5 Optimal MSE J_1^* and $\tilde{J}_1$ for $P = 1$, $N_1 = 200$ sensors with $\sigma_1 = 1$ and $\bar{p} = 1$. 53

4.6 Optimal MSE vs total number of channels D for $P = 3$, $N_i = 20$, $i = 1, 2, 3$, with $\sigma_1 = 1$, $\sigma_2 = 1.2$, $\sigma_3 = 1.4$ and $\bar{p} = 0.5$. 54

List of Tables

4.1 Transmission strategies t and cost $M_k(t)$ for current time k. 40

Abbrevations and Symbols

Abbreviations

AoI	Age of information
CA-DPS	Cost-aware dual prediction scheme
DPS	Dual prediction scheme
FC	Fusion center
FHM	Finite horizon minimization
IoT	Internet-of-things
LSTM	Long short-term memory
LTI	Linear time-invariant
MDP	Markov decision process
MMSE	Minimum mean square error
MSE	Mean squared error
NCS	Network control system
ORR	Optimized round robin
RR	Round robin
WNCS	Wireless network control system
WSN	Wireless sensor network

Symbols

$x[k]$	Sensor measurement at time instant k
$w_i[k]$	Stochastic measurement noise of the ith process at time instant k
$x^{FC}[k]$	Registered measurement at the fusion center at time instant k
$x_i[k]$	Measurement of the ith sensor at time instant k
$x_{i,j}$	Measurement of the ith sensor in the jth cluster
$y_i[k]$	Most recently broadcast measurement from the ith sensor
n_i^γ	Number of scheduling instances for the ith sensor given policy γ
$\bar{n}_i$	Transmission constraint for the ith sensor
n_i^*	Optimal number of scheduling instances for the ith sensor
$\hat{x}_i[k]$	Estimated measurement of the ith sensor at time instant k
p_i	Transmission rate, equivalent to the transmission probability, at the ith cluster
s_i	Variable indicating either a single- or double-bounded threshold strategy at the ith cluster
D	Number of communication channels
D_i	Number of communication channels for the ith cluster
N	Number of sensors
N_i	Number of sensors in the ith cluster
γ	Policy in a Markov decision process
σ_i	Standard deviation of the ith process
ξ_i	Standard deviation of measurement noise $w_i[k]$
$\Delta_i[k]$	AoI of the ith sensor at time instant k
$\Delta_{ij}[k]$	AoI differences between the ith and the jth process at time instant k
ρ_{ij}	Spatial correlation between the ith and the jth process

$\theta_i[k]$	Value of the ith process at time instant k
$\tilde{\Delta}_i[k]$	Truncated AoI of the ith sensor at time instant k
$\tilde{\Delta}_{ij}[k]$	Truncated AoI differences between the ith and the jth process at time instant k
$\bar{\Delta}$	The lowest maximum AoI (minmax) for all sensors
$\Delta_i^{\gamma}[k]$	AoI of the ith sensor at instant k given policy γ
$\hat{\theta}_i[k]$	Estimated value of the ith process at time instant k
β	Accuracy threshold in a DPS
α_i	Lower transmission threshold at the ith cluster
β_i	Upper transmission threshold at the ith cluster
$\boldsymbol{a}[k]$	Action at instant k
$\boldsymbol{s}[k]$	State at instant k
$\boldsymbol{y}[k]$	Vector of most recently broadcasted measurements from each sensor
$\boldsymbol{\Delta}[k]$	AoI vector at time instant k
$\boldsymbol{\pi}[k]$	Scheduling variable denoting an index set of sensors to be scheduled at time instant k
$\boldsymbol{\theta}[k]$	Vector of process values at time instant k
$\tilde{\boldsymbol{\Delta}}[k]$	Truncated AoI vector at time instant k
$\hat{\boldsymbol{\theta}}[k]$	Vector of estimated process values at time instant k
$\boldsymbol{\kappa}[k]$	Vector of prediction model parameters in a DPS at time instant k
$\mathbf{C}_{\theta y}[k]$	Cross-covariance matrix of processes $\boldsymbol{\theta}[k]$ and $\mathbf{y}[k]$
$\mathbf{C}_{yy}[k]$	Covariance matrix of process $\mathbf{y}[k]$
$\mathcal{M}$	Markov decision process tuple
$\mathcal{A}$	Set of all possible actions in $\mathcal{M}$
$\mathcal{S}$	Set of all possible states in $\mathcal{M}$

$\tilde{\mathcal{M}}$	Markov decision process tuple based on the truncated AoI
$\tilde{\mathcal{S}}$	Set of all possible states in $\tilde{\mathcal{M}}$
$\mathcal{I}_i$	Output from the ith collision-channel to the ith estimator
$\mathcal{Y}_i$	Set of indices of the transmitting sensors from the ith cluster
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of nonnegative real numbers
$\mathbb{R}_{++}$	Set of positive real numbers
$\mathbb{N}$	Set of natural numbers
$\mathbb{N}_+$	Set of nonnegative natural numbers
$\mathbb{N}_{++}$	Set of positive natural numbers
$\delta(\cdot)$	Dirac delta function
$\mathbb{1}(\cdot)$	Indicator function having value 1 if the condition in the argument is true and 0 otherwise
$\Phi(\cdot)$	Cumulative distribution function for a Gaussian distribution $x \sim \mathcal{N}(0,1)$
tr$(\cdot)$	Trace of its argument matrix
$\varphi(\cdot)$	Temporal correlation function
$\lceil\cdot\rceil$	Ceil operator
$\lfloor\cdot\rfloor$	Floor operator
$[\,\cdot\,]_+^m$	Operator truncating value to m
$\bar{\Delta}_i(\cdot)$	Function giving the lowest maximum AoI (minmax) for the ith sensor
$\gamma_k(\cdot)$	Strategy at instant k based on state $\boldsymbol{s}[k]$
$\tilde{\gamma}_k(\cdot)$	Scheduling strategy at instant k based on truncated AoI $\tilde{\boldsymbol{\Delta}}[k-1]$
$\tilde{b}(\cdot)$	Mapping from the AoI $\boldsymbol{\Delta}[k]$ to the truncated AoI $\tilde{\boldsymbol{\Delta}}[k]$

Chapter 1

Introduction

Wireless sensor networks (WSN) enhance our ability to monitor physical processes and is an essential technology in the internet-of-things (IoT) [1]. In WSN and wireless networked control systems (WNCS), sensors observe physical processes and transmit their measurements to remote estimators or a centralized fusion center (FC) that track process parameters to form decisions to maximize the utility of the system application. This data collection infrastructure provides the backbone of IoT technologies, such as autonomous driving cars, smart cities [2], industry 4.0 [3], smart healthcare [4], and modern-day distributed WNCS [5], that will impact our daily lives and a wide range of industries. Recent technological advancements in component development and wireless communications have led to sensors becoming cheaper, smaller, increasing their communication capabilities, and even making them mobile, e.g., sensors attached to unmanned aerial vehicles (UAVs) [6]. This trend will enable larger and more versatile WSNs, allowing better monitoring capabilities to support future IoT applications. Another critical factor in the evolution of WSN and IoT is the widespread accessibility of the internet. This allows for applications to be supported by a cloud-based fusion center, providing vast computational resources to process large amounts of data quickly. Hence, in the era of data-driven decision-making and big data, any additional sensor measurement registered at a central unit could be of use in improving the utility of the IoT application. However, in WSN, the information update rate is constrained by the sensors limited energy resources and the number of communication channels over which they communicate.

Firstly, sensor nodes in WSN are usually not supported by power grids and rely on small batteries with limited energy storage capabilities. In some WSNs, sensor nodes can re-charge their batteries from energy harvesting, e.g., solar energy [7].

Unfortunately, such renewable energy sources are unpredictable and fluctuating, complicating energy resource planning [8]. In general, radio communication consumes significantly more energy than other sensor tasks, e.g., sensing and data processing [9]. Empirical data shows that roughly 80 percent of the node energy is consumed when transmitting and receiving data [10]. Excessive measurement transmission significantly reduces sensor lifetime, which can be critical for applications that depend on long-lasting and durable WSN. For many WSN applications, due to the placement of the sensor node, it can be a demanding, complicated and costly procedure to manually change batteries, e.g., in rural areas [11], in-body [12], or built into concrete [13]. Furthermore, applications made to prevent accidents require stable and reliable WSN, e.g., avalanche detection [14] or healthcare emergency event detection [15].

Secondly, the information update rate is also limited by the number of available communication channels in a WSN. Commonly, sensors share communication channels, and if the number of transmitted data packages exceeds the number of available communication channels, a packet collision occurs, and information is lost due to interference. Over the years, communication protocols have been developed, e.g., ALOHA [16], or hybrid automatic repeat request (HARQ) [17], that facilitate that packet losses are acknowledged by the sender and re-transmitted. However, the primary purpose of these protocols is to enable the transmitted packets to be delivered, and they are not designed to schedule or prioritize measurements based on the information that the transmitted packets carry. Consequently, simply queuing or re-transmitting failed measurement packages until they finally arrive can result in the staleness of information. Hence, the utility of a channel-constrained WSN can be further enhanced by coordinating the sensor transmissions based on the contributing information of their measurements.

Today, there is a need for approaches that can maximize the system utility and prolong the lifetime of sensors with consideration for limited energy and communication resources. One method is to design transmission schemes that reduce the number of data transmissions from the sensors without compromising the data accuracy. Alternatively, to focus on reducing the size of data packets transmitted from sensors. Since larger data packets consume more energy for radio transmission and require more channel capacity for communication. Another approach would be to coordinate the sensor transmissions by selecting sensors that contribute the most information to enhance the overall estimation accuracy while considering the available resources. One way to achieve these approaches is to exploit redundancy in sensor measurement data.

Sensors often observe processes, e.g., temperature, or humidity, that follow trends, seasonalities, or periodicities [18]. As a result, sensor measurement data tend to

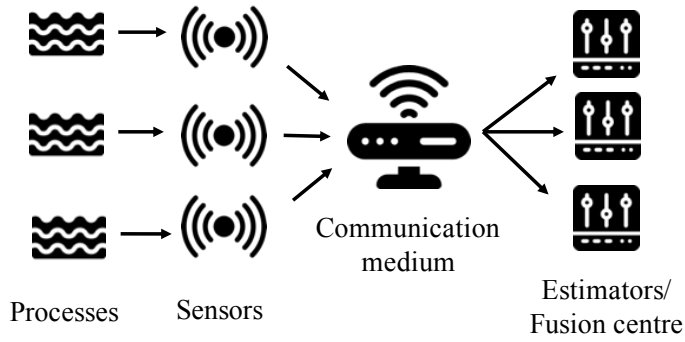

Figure 1.1: WSN topology considered for this thesis.

be redundant, demonstrate pattern-like behavior or even spatio-temporal correlation across sensors [19], [20]. These properties can be exploited to develop data reduction approaches in channel- and energy-constrained WSNs. Instead of transmitting all the raw sensor data directly to remote estimators or a fusion center, it can be locally aggregated [21] or compressed [22] at a cluster-head node appointed to a cluster of neighboring sensors, which reduces the data package size before transmitting the data. Other approaches are based on adjusting the sample rate with respect to statistical properties of the measurement distribution, e.g., adaptive sensing [23], [24], or compressed sensing [25], which gives the lowest possible sample rate such that the complete measurement time series can be reconstructed.

Although the methods above can be beneficial in reducing the transmitted data packet size or even the transmission rate, i.e., the average number of transmitted measurements per time instant, they have implications for the privacy and latency of the system. For example, if a cluster-head collects all measurements before performing compression, it can be targeted by an eavesdropper or malicious attacker that wants to sabotage the system [26]. Furthermore, it is essential to satisfy real-time monitoring requirements with a given information update rate in delay-sensitive and safety-critical applications, where compressing or aggregating data is not an option. Another vital aspect to acknowledge is that compressing or aggregating data requires specific topology solutions of routing or communication among sensors, which can not be generalized to sensors transmitting directly to the end receiver.

1.1 Scope and Objective

This thesis aims to design transmission schemes that determine when an individual sensor should transmit an observation to facilitate high data quality in resource-

constrained wireless sensor networks. We focus on WSN for remote estimation, where the system utility depends on the estimation accuracy at the remote estimators or a centralized FC. The two primary goals are achieving the highest possible estimation accuracy and prolonging the lifetime of sensor nodes.

We consider the two following approaches when designing transmission schemes to achieve our goals:

- Reducing the number of transmissions or transmitted data packet sizes from sensors
- Coordinating the measurement transmission from all sensors to maximize the estimation accuracy

We aim to design transmission schemes suitable to an arbitrary WSN topology and do not require distributed calculations among sensors, where sensors need to share measurements with each other. We consider a WSN topology, as depicted in Figure 1.1, where sensors communicate their measurements directly to the remote estimators or a fusion center via a communication medium, not via other sensors.

1.2 Contributions

The contributions of this thesis apply to two different approaches of transmission schemes, defined by their transmission triggering mechanisms; time-triggered, i.e., scheduled time slots, or event-triggered, e.g., a measurement breaching a threshold. The main contributions of this thesis are summarized as follows.

C1 The first contribution of this thesis applies to time-triggered transmission schemes, for which we study the scheduling of dependent sensor measurements in a channel-constrained WSN. We design an optimal scheduling policy for a system where a network manager is responsible for scheduling multiple sensors observing spatio-temporally correlated observations. At each time instant, the scheduling decision is based on the age-of-information (AoI, introduced in Chapter 2), representing the time elapsed since the last received measurement. Although AoI-based scheduling policies have been previously studied for remote estimation, only a few works deal with dependent observations. Our work extends the previous literature on AoI-based scheduling of dependent processes, and our results demonstrate that the performance can be improved by exploiting spatio-temporal dependency in a channel-constrained system. The main contributions of Publications **III** and **IV** can be summarized as follows:

- We prove the existence and derive an optimal scheduling policy for a system of multiple spatio-temporally dependent observations based on age-of-information. An optimal policy minimizes the average mean squared estimation error over an infinite time horizon.
- We show that a policy can be derived by formalizing the problem as a finite-state MDP. The finite-state MDP is possible by exploiting the property that increasing time-distance between consecutive transmissions from a single source decreases spatio-temporal correlation to observations from other sources.
- We show that an optimal policy yields a periodic scheduling pattern, which has earlier been demonstrated for single-sensor scheduling [27]–[29]. This property simplifies the practical implementations and saves data storage at the network manager.
- We show that the finite state space implies that any deterministic policy results in a periodic structure. The performance of any periodic scheduling policy can be easily calculated using the theoretical framework given in the paper.
- For the particular case of two sensors and one communication channel, we include a transmission constraint and present a low-complexity method of deriving an optimal scheduling policy for a given time horizon.

C2 The second contribution of this thesis applies to event-triggered transmission schemes, for which we address the implications of sensors observing non-stationary measurement processes and the importance of implementing updating protocols in a dual prediction scheme (DPS, introduced in Chapter 2) to avoid excessive transmission. We develop a cost-aware dual prediction scheme (CA-DPS) for deciding when to update the prediction model or transmit a set of measurements from the sensor to the fusion center (FC) to achieve minimal data transmission in a DPS. In a CA-DPS, the sensors choose among a set of transmission strategies and decide the one that achieves minimum expected transmission cost within a given forecast horizon. In a practical setting, statistical information of the measurements might be limited, and to derive a robust estimate of the future transmission cost, we apply model-based bootstrapping. The main contributions of Publication **V** can be summarized as follows.

- We present a CA-DPS where a sensor evaluates several strategies at each transmission instance, and select the best transmission strategy to reduce future transmission costs.

- We introduce model-based bootstrapping in a DPS framework to estimate future transmission costs.

C3 The third contribution of this thesis also applies to event-triggered transmission schemes, where a sensor transmits a measurement to a remote estimator if it exceeds a pre-defined threshold. All sensors share a limited number of communication channels. The sensors are divided into spatially distributed clusters, each assigned a set of communication channels. The measurement distributions among the different clusters are assumed heterogeneous. We develop a method to find optimal transmission thresholds and channel allocation for each cluster. Furthermore, we numerically compare the performance of assigning collision channels for each cluster to having all clusters sharing collision channels. The main contributions of Publication **VI** can be summarized as follows.

- We present a method to find optimal transmission thresholds and channel allocation for sensors transmitting observations to remote estimators over multiple collision channels.
- We show that by allowing for both single- and double-sided thresholds, our decentralized scheme can outperform the centralized scheme presented in [30].
- We also derive an optimal performance upper boundary for any given number of channels.

1.3 Thesis Outline

This thesis begins by introducing the background and motivation for developing transmission schemes to enhance the utility and prolong sensor lifetime in resource-constrained WSN. Chapter 2 presents an overview of the general system model, together with some important technical concepts and mathematical tools used to develop our transmission schemes. Chapter 3, considers time-triggered transmission schemes and presents the results from Publications **I**-**IV**, where we derive an optimal AoI-based scheduling policy for sensors observing spatio-temporally dependent measurements. After that, in Chapter 4, we consider event-triggered transmission schemes. In the first part of Chapter 4, we summarize the main results from Publication **V**, where we present the CA-DPS. In the second part of Chapter 4, we present the results from Publication **VI**, where we develop a method to find optimal transmission thresholds and channel allocation for sensors transmitting observations to remote estimators over multiple collision channels. Finally, in Chapter 5, we discuss our main contribution, implications of our work, and future extensions.

Chapter 2

Background

This chapter introduces parts from the theoretical framework, specifically our system model, technical concepts from relevant literature, and necessary mathematical tools to derive our protocols. We begin this chapter by presenting a generalized system model. Later on, we present two important technical concepts from the related literature; the age-of-information and the dual prediction scheme. Finally, we present two mathematical frameworks; Markov decision processes and bootstrapping, used to design our transmission schemes.

2.1 The General System Model

In this thesis, we consider a WSN topology where sensors observe individual processes and transmit their measurements via a communication medium to remote estimators, or a centralized fusion center, as depicted in Figure 2.1. Sensor i observes the stochastic process $\theta_i[k] \in \mathbb{R}$, $i = 1, ..., N$, and at time instant $k \in \mathbb{N}$ acquires measurement $x_i[k] \in \mathbb{R}$, which is modeled as

$$x_i[k] = \theta_i[k] + w_i[k], \quad k \in \mathbb{N},\ i = 1, 2, ..., N, \tag{2.1}$$

where $w_i[k] \in \mathbb{R}$ denotes independent identically distributed (iid) stochastic measurement noise, following a Gaussian distribution $w_i[k] \sim \mathcal{N}(0, \xi^2)$. For each process $\theta_i[k]$, there is a remote estimator or a fusion center that tracks the process and forms an estimate $\hat{\theta}_i[k]$ based on received sensor measurements. The number of remote estimators P can be less than or equal to the number of sensors, i.e., $P \leq N$.

Remark 2.1. In Chapter 4, to simplify the mathematical calculations, we do not assume that the remote estimators have knowledge of the distribution of the measurement noise $w_i[k]$ in (3.2). Instead, the estimator forms an estimate of the meas-

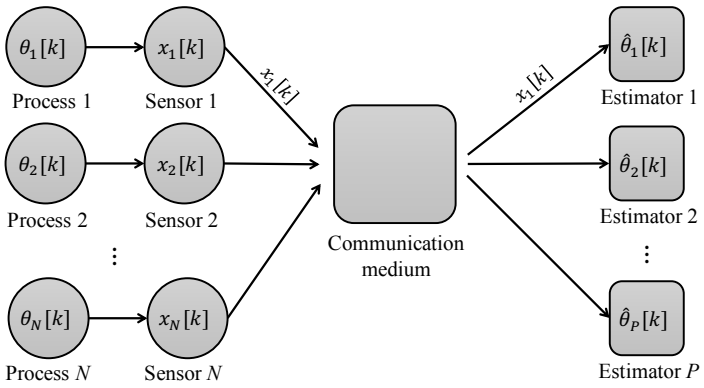

Figure 2.1: Schematic of a generalized WSN system model considered in this thesis.

urement $\hat{x}_i[k]$. Thus, in Chapter 4, the estimate at time instant k is denoted as $\hat{x}_i[k]$ and not as $\hat{\theta}_i[k]$.

Transmission Triggering Mechanisms

For all transmission schemes, a sensor must know when it should transmit a measurement. In this thesis, we consider the following two types of triggering mechanisms that determine when a transmission instance occur:

- **Time-triggered**, i.e., allocating time slots for each sensor transmission.
- **Event-triggered**, i.e., transmitting an observation in case of an event, e.g., a measurement exceeds a predefined threshold.

Both these approaches have their advantages when designing transmission schemes. In channel-constrained WSN, time-triggered transmission can result in collision-free communication [29], [31]. It also makes energy resource-planning for measurement transmission easier than event-triggered transmission, which involves stochastic transmission.

Basically, in event-triggered transmission, a sensor only transmits an observation when it provides sufficiently valuable information. If a triggering event rarely occurs, event-triggered transmission schemes can result in a low transmission rate and, thereby, save energy due to reduced communication. Another benefit of event-triggered transmission is that it can result in a fast detection time, compared to scheduling sensors, if a crucial or unexpected event occurs. Event-triggered transmission schemes for remote-estimation WSN are usually designed such that

a sensor should only transmit if an observation deviates beyond a pre-defined threshold from the estimated value at the current time instant [30]–[34]. As will be discussed in Chapter 4, the transmission rate depends on how well the model for estimation fits the physical process.

2.2 Technical Concepts

This section introduces two technical concepts from related literature. We begin by presenting age-of-information (AoI), which we exploit to derive time-triggered transmission schemes in Chapter 3. Later on, we present the dual prediction scheme (DPS) framework, which we present an extended version of in Chapter 4.

Age-of-information

The AoI refers to the freshness of information [35], i.e., the time elapsed since the information received was generated at the source. Recently, the AoI has gained increased attention as a system performance metric to optimize with respect to system settings. Commonly, a WSN is modeled as a queuing system, defined by a queuing model, e.g., first-come, first-served [36], where the AoI is influenced by stochastic arrival processes and service times of transmitted packets. In general, due to factors such as collisions, queuing, packet dropouts, and random arrival and service times, the AoI is modeled as a stochastic process. Hence, it can become complicated to derive the relationship between the AoI and other system parameters of a WSN. As seen in [35], for various communication system set-ups, simply maximizing the transmission rate does not result in a minimum average AoI. Over time, several performance metrics of the AoI have been introduced, e.g., peak and average, which have been studied under different system settings [35], [37]–[43].

In many WSN applications, the AoI is an important factor that directly or indirectly impacts the main utility of the system. For this reason, the AoI has been utilized as a state variable to assist in designing and evaluating scheduling policies for a variety of tasks. For example, in [44], the AoI is used to determine an optimal scheduling policy for updating model parameters for federated learning in a mobile edge network. Whereas, in [45], the AoI is used to decide content to multicast in a cache-enabled content-centric wireless network. Terms such as the value-of-information, or the cost-of-delay [46], [47], have been used to quantify the utility cost of the AoI, represented by a non-linear function of the AoI. In some cases, the utility cost of the AoI is not static over time, but depends on the current state of the monitored process. For example, in [48], a sensor monitors and transmits status updates of a two-state stochastic process, where each state is associated with a different cost. In remote estimation, the object is to maximize the estimation

accuracy at the estimators, where usually the performance metric of the system utility is the overall mean squared error (MSE) of the estimation error. Although an increasing AoI results in reduced estimation accuracy, minimizing the average AoI across sensors does not always result in a minimum overall MSE.

Scheduling with AoI in remote estimation and networked control has been considered in several previous works, e.g., [5], [47], [49]–[54], wherein the AoI has been used as a state variable to assist in designing scheduling policies. In [5], [47], [51], the authors investigate scheduling for state estimation of discrete-time linear time-invariant (LTI) systems in which multiple sensors transmit their local state estimates to a remote estimator over a limited number of time-varying wireless communication channels resulting in occasional packet drops. The task is to determine the system conditions in terms of LTI system parameters and channel statistics that guarantee stability conditions for remote estimation applications or WNCS, i.e., the existence of a scheduling policy that results in a bounded process variance or average estimation MSE. Most works regarding AoI-based scheduling for remote estimation consider systems where multiple sensors share a single communication channel, and only one sensor can be scheduled at each time instant. Adding more communication channels improves the overall real-time accuracy but also adds complexity to finding optimal scheduling policies as the number of possible scheduling decisions increases. Recently, there has been an increasing amount of work regarding scheduling multiple devices, including sensors and controllers, that share multiple communication channels [5], [51].

The majority of work that regards AoI-based scheduling for remote estimation deal with independent sensor observations [5], [47], [49]–[54]. As mentioned earlier, sensor observations tend to be dependent that can be exploited to improve the remote estimators' overall accuracy. In Chapter 3, we extend the current literature on AoI-based scheduling by designing an optimal scheduling policy for a WSN where sensors observe spatio-temporally correlated measurements.

Dual Prediction Scheme

A dual prediction scheme (DPS) is an event-triggered transmission scheme based on the idea of reducing the number of transmissions from sensors to an FC, or a remote estimator, by replacing missing transmissions with predictions. In a DPS, both the sensor and FC store an identical prediction model of future values of the sensor measurement process. Hence, the term *dual prediction*. Each time a sensor observes a measurement and compares it with the predicted measurement. If the measurement lies within a pre-defined error-tolerance from the predicted measurement, the sensor does not transmit the measurement, and the FC registers the predicted measurement as the most recent measurement. Suppose the sensor

measurement deviates beyond the pre-defined error-tolerance level. In that case, the sensor transmits the measurement, and the FC registers the received measurement. For a DPS to work, the sensor and the FC must store and apply the same prediction model and input data to create identical predicted measurements.

The benefit of using a DPS is that it can result in a bounded measurement error with low latency. Thus, it provides a suitable data-reduction approach in applications where compressing or aggregating data is not an option. A DPS can be generalized to any prediction model as long as the following criterion is satisfied; that the sensor and the FC can recreate the same predicted and registered measurements. In the literature, the prediction models can vary from simple linear models, e.g., the form of time series models [55], adaptive filters [33], [56], a weighted average [18], [57], to more complex non-linear models, e.g., deep learning long short-term memory (LSTM)-model [58], or a hybrid of two prediction models [59].

In Chapter 4, we discuss the implications of applying a DPS for non-stationary processes, where the prediction model parameters need to be re-estimated and updated each time the measurement distribution changes to avoid excessive measurement transmissions due to threshold breaches. We present a cost-aware DPS, where the sensor evaluates transmission strategies each time a threshold breach occurs, which includes the alternative of transmitting re-estimated model parameters or not.

2.3 Mathematical Tools

This section presents two mathematical frameworks, bootstrapping and Markov decision processes, relevant to the proposed methods presented in Chapters 3 and 4.

Bootstrapping

Bootstrapping is a numerical method to estimate a statistic of a stochastic variable by random resampling with replacement from observed data [60]. For this purpose, we first generate synthetic samples of the statistic using the observed data by sampling from the empirical probability distribution. Thereafter, we estimate the statistic as the mean of the generated samples. Following is a mathematical representation of bootstrapping.

Let $\boldsymbol{X} \in \mathbb{R}^n$, $n \in \mathbb{N}_{++}$, and $Y \in \mathbb{R}$, be two stochastic variables, where Y is a function of $\boldsymbol{X}$, i.e., $Y = g(\boldsymbol{X})$. The expected value of Y is denoted as μ_Y, where $\mu_Y = \mathbb{E}[Y] = \mathbb{E}[g(\boldsymbol{X})]$. Now, let $\mathcal{X}$ be a set of m realized data points of variable $\boldsymbol{X}$, i.e., $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m\}$, $\boldsymbol{x}_i \in \mathbb{R}^n$. With the bootstrapping method, we estimate μ_Y as $\hat{\mu}_Y$, by generating a set of L values of Y by re-sampling L values

of $\mathcal{X}$ as $\mathcal{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_L\}$, $\boldsymbol{z}_i \in \mathcal{X}$, $i = 1, 2, ..., L$, and taking the average, i.e.,

$$\hat{\mu}_Y = \frac{1}{L} \sum_{\boldsymbol{z}_i \in \mathcal{Z}} g(\boldsymbol{z}_i).$$

Two practical benefits of using bootstrapping are, firstly, it provides a straightforward numerical method to estimate any statistic when a closed-form expression of an estimator is not available. This can occur if a closed-form expression of the statistic is intractable to derive analytically. Secondly, it provides a robust estimator where one does not have to rely on assumptions of parametric distributions of the observed data [61]. In practice, model-fitting is time-consuming and becomes arbitrary when the observed data is sparse, and the empirical distribution has a non-smooth shape.

Bootstrapping can also be applied to generate stochastic processes that are temporally dependent. A common way to model stochastic processes $x[k] \in \mathbb{R}$ is to assume that the process value at a given time instant, k, is a deterministic function, h, of $p + 1$ previous values with an additive stochastic term, i.e.,

$$x[k] = h(\tilde{\boldsymbol{x}}[k-1]) + e[k], \quad e[k] \sim \mathcal{F}_e \tag{2.2}$$

where $\tilde{\boldsymbol{x}}[k-1] = \big[x[k-(p+1)], ..., x[k-1]\big]^{\mathrm{T}}$ and $e[k]$ is the model residual. To generate a trajectory of the process $x[k]$ in (2.2), one can apply *model-based bootstrapping* [62], in which only the stochastic component of the process, $e[k]$, is re-sampled and added to the deterministic part, given by $h(\tilde{\boldsymbol{x}}[k])$. These steps are then repeated to generate a complete sequence. Following is a mathematical representation of model-based bootstrapping.

Given a set of previous observations $\mathcal{X} = \{x[0], x[1], ..., x[k]\}$, we can generate L future trajectories of $x[k]$ of length $T \in \mathbb{N}_{++}$ as $\big[x^{(l)}[k+1], x^{(l)}[k+2], ..., x^{(l)}[k+T]\big]$, $l = 1, ..., L$. We begin by first extracting a set of historical model residuals $\mathcal{E}[k] = \{e_{p+2}, e_{p+3}, ..., e_k\}$, using (2.2) and $\mathcal{X}$ as

$$e_i = x[i] - h(\tilde{\boldsymbol{x}}[i-1]), \quad i = p+2, p+3, ..., k.$$

To generate process value $x^{(l)}[k+1]$ for trajectory l, we first randomly draw $e^{(l)}[k+1]$ from $\mathcal{E}[k]$, i.e., $e[k+1] \in \mathcal{E}[k]$, and then calculate

$$x^{(l)}[k+1] = h(\tilde{\boldsymbol{x}}^{(l)}[k]) + e^{(l)}[k+1],$$

and finally update $\tilde{\boldsymbol{x}}^{(l)}[k+1] = \big[x^{(l)}[k-p], ..., , x^{(l)}[k]\big]^{\mathrm{T}}$, where $x^{(l)}[t] = x[t]$, $\forall t \leq k$, $l = 1, 2, ..., L$. The same process is then repeated to generate $\big[\tilde{\boldsymbol{x}}^{(l)}[k+2], \tilde{\boldsymbol{x}}^{(l)}[k+3], ..., \tilde{\boldsymbol{x}}^{(l)}[k+T]\big]$.

Another bootstrapping approach to simulate temporally correlated stochastic processes is to re-sample from the observed data in a restricted order, for example, by following the maximum entropy bootstrap algorithm (with scale adjustment) described in [63], [64]. In Chapter 4, we use a combination of model-based bootstrapping and the maximum entropy bootstrap algorithm (with scale adjustment) to generate simulated trajectories of sensor measurement data.

A real-world application of bootstrapping is in quantitative finance when the task is to price financial derivatives, i.e., financial instruments where the monetary pay-out is a function of an underlying asset, e.g., a stock price [65], [66]. Take, for example, an Asian option, where the payout is the positive difference between time average stock price and a fixed value. For many derivatives, deriving closed-form functions of the expected future pay-out is prohibitive, and bootstrapping provides a straightforward method by simply re-sampling historical prices. Also, financial markets tend to be highly dynamic and non-stationary. Therefore, statistical models are based on short historical windows to capture current market states. For this reason, empirical distributions are based on small-sized data sets that demonstrate properties of non-smoothness and thick tails of the probability distribution, which can be difficult to fit to parametric distributions. Thus, re-sampling and bootstrapping provide a practical method to overcome these issues.

Markov Decision Process and Dynamic Programming

In Chapter 3, we consider a WSN where sensors observe spatio-temporally dependent measurements that are transmitted to remote estimators that track the processes. We derive an optimal AoI-based scheduling policy that minimizes the overall time average MSE. For this scenario, as in most cases regarding scheduling, finding an optimal scheduling policy involves solving a sequential decision-making problem. To derive an optimal policy, we first model the problem as a Markov decision process (MDP) [67], being a discrete-time stochastic control process. We then use dynamic programming to solve the optimization problem numerically.

A Markov decision process is a mathematical framework that can be applied to study and solve recursive optimization problems, i.e., path-dependent optimization problems. An MDP is an extension of a Markov chain process, which is defined by a set of *states* and *transition probabilities* between states. The fundamental property of a Markov chain is that the probability of transitioning from a particular state to another is independent of previous states. In contrast to a Markov chain, an MDP involves a *decision-maker* who, for each current state, takes an *action* that determines the transition probabilities to the next possible state, which then

results in a *reward*. When solving a sequential decision-making problem, the goal is to determine a state-action *policy* for the decision-maker that maximizes the accumulated reward over time.

Reinforcement learning [68] is a machine learning approach based on the MDP framework to derive a state-action policy when there does not exist full knowledge of the transition probabilities and the reward function beforehand. In reinforcement learning, the transition probabilities and the reward function are continuously accessed after each time instant a state-action occurs. For this reason, each time the transition probabilities and the reward function are updated, the state-action policy is re-calculated. In the literature of WSN [69], reinforcement learning has gained an increasing amount of interest, motivated by the fact that sensors can learn complex tasks, where transition probabilities and the reward functions are difficult to model beforehand. However, training a reinforcement learning model often requires an extensive amount of training data [68], which is not always feasible in practical settings.

A given MDP, $\mathcal{M}$, is formally defined by a 4-tuple, $\{\mathcal{S}, \mathcal{A}, \mathbb{P}(\cdot|\cdot), r(\cdot)\}$, which represents the following components.

- **State space** $\mathcal{S}$: The set of all possible states in $\mathcal{M}$, where at time instant k the state is given by $\boldsymbol{s}[k] \in \mathcal{S}$.
- **Action space** $\mathcal{A}$: The set of all possible actions in $\mathcal{M}$, where at time instant k the state is given by $\boldsymbol{a}[k] \in \mathcal{S}$.
- **Transition probabilities** $\mathbb{P}(\cdot|\cdot)$: The probability distributions of transitioning from a particular state $\boldsymbol{s}[k]$ to $\boldsymbol{s}[k+1]$ given action $\boldsymbol{a}[k]$, i.e., $P(\boldsymbol{s}[k+1]|\boldsymbol{s}[k], \boldsymbol{a}[k]) \in [0, 1]$.
- **Reward function** $r(\boldsymbol{s}[k], \boldsymbol{a}[k]) \in \mathbb{R}$: The function that determines the reward at instant k given state $\boldsymbol{s}[k]$ and action $\boldsymbol{a}[k]$.

For a system to be modeled as an MDP, it should satisfy the Markovian property that at instant k, the next state $\boldsymbol{s}[k+1]$ only depends on the current state $\boldsymbol{s}[k]$ and the action $\boldsymbol{a}[k]$.

A state-action policy $\gamma = (\gamma_0, \gamma_1, , ..., \gamma_T)$, where $\gamma_k : \mathcal{S} \rightarrow \mathcal{A}$, provides a mapping between a state and an action at instant k over time-period $k = 0, 1, ..., T$, i.e.,

$$\boldsymbol{a}[k] = \gamma_k(\boldsymbol{s}[k]).$$

In an MDP, the objective is to find an optimal policy γ^* that maximizes the cumulatitive sum of rewards over a finite or infinite time-horizon. Commonly, an optimal policy γ^* is defined such that it satisfies

$$\max_{\gamma \in \Gamma} g_\gamma(\boldsymbol{s}[k]).$$

where g_γ is a function defined by

$$g_\gamma(\boldsymbol{s}[k]) = \lim_{T \to \infty} \sum_{i=k}^{T} \lambda^{(i-k)} \mathbb{E}\Big[r(\boldsymbol{s}[i], \boldsymbol{a}[i]) \Big| \gamma, \boldsymbol{s}[k] \Big], \tag{2.3}$$

and $\lambda \in [0, 1]$ is a discount factor. The function g_γ in (2.3), is often referred to as the value function, since it refers to the current value of discounted future rewards. For a given policy γ, the value function is equivalent to

$$g_\gamma(\boldsymbol{s}[k]) = r(\boldsymbol{s}[k], \boldsymbol{a}[k]) + \lambda \sum_{\boldsymbol{s}[k+1] \in \mathcal{S}} P(\boldsymbol{s}[k+1] | \boldsymbol{s}[k], \boldsymbol{a}[k]) g_\gamma(\boldsymbol{s}[k+1]). \tag{2.4}$$

Based on (2.4), an optimal policy γ^* should satisfy maximizing the value function at the current state as

$$\max_{\gamma \in \Gamma} \{ r(\boldsymbol{s}[k], \boldsymbol{a}[k]) + \lambda \sum_{\boldsymbol{s}[k+1] \in \mathcal{S}} P(\boldsymbol{s}[k+1] | \boldsymbol{s}[k], \boldsymbol{a}[k]) g_\gamma(\boldsymbol{s}[k+1]) \},$$

where Γ is the set of all feasible policies.

An optimal policy γ^* can be derived numerically using dynamic programming, e.g., policy iteration or value iteration [67].

Remark 2.2. For simplicity, in this presentation of the MDP framework, we assume that the reward function is deterministic and that the state-transitions, reward function and set of actions are independent of the time instant k. Also, that the sequential optimization problem concerns an infinite time horizon, i.e., $k = 1, ..., T$, where $T \to \infty$. For a more extensive introduction to MDPs, see [67].

Chapter 3

AoI-based Scheduling of Spatio-temporally Dependent Measurements

This chapter presents our contributions in AoI-based scheduling for remote estimation. In practice, sensor measurements tend to be spatio-temporally dependent, which has been exploited in resource-constrained WSN to achieve energy-efficient routing [70], [71], optimal sensor location selection [72], missing data inference in environmental mobile crowdsensing [73], and reducing traffic load [74]. Although there exists work considering the scheduling of dependent measurement for remote estimation, this body of work is small compared to the amount that has been studying the scheduling of independent sensors observations. In Publications **I**-**IV**, we demonstrate how spatio-temporal dependencies can be further exploited to improve the performance in sensor scheduling for remote estimation in a channel-constrained WSN.

This chapter summarizes Publications **III** and **IV**, which are, respectively, the extended journal work of the conference papers **I** and **II**. We consider a system of sensors observing spatio-temporally dependent measurements that are broadcasted to remote estimators, each tracking one process, over a resource-limited broadcast channel. At each time instant, a measurement-blind network scheduler broadcasts a limited number of measurements to remote estimators. The network scheduler is responsible for scheduling the sensor observations to broadcast. The scheduler cannot observe the measurements and decides the scheduling policy based on the age-of-information. In our work, we model the scheduling problem as a Markov

decision process (MDP), presented in Section 2.3, to derive an optimal scheduling policy that minimizes the time average MSE across the remote estimators over an infinite time horizon. We also present a low-complexity numerical method for deriving an optimal scheduling policy for a system of two sensors sharing a single communication channel, where the number of scheduling instances for each sensor is constrained due to limited energy resources.

3.1 Prior work

Optimal scheduling schemes for remote estimation where sensors observe independent processes have been studied under a variety of different system scenarios and various resource constraints, e.g., limited packet sizes for different source processes [75], limited battery [76], single or multiple communication channels [29], [77], and in the presence of eavesdroppers [78]. As of today, only a few work focused on scheduling for remote estimation have exploited dependence among sensor measurements to enhance the overall estimation accuracy in resource-constrained WSN.

In [79], an optimal scheduling policy is presented for a system of two sensors observing dependent processes where a network manager, responsible for the scheduling, can observe measurements before scheduling. Optimal observation-driven scheduling of sensors observing independent measurements have earlier been derived with consideration for energy-harvesting [76], temporally-correlated measurements [80], and multiple communication channels [81]. A WSN where a scheduler can observe measurement before scheduling provides more information to base every scheduling decision and can benefit in improving the overall estimation accuracy, compared to not observing. However, such a system setup has implications on both the system privacy and latency.

Instead of an observation-driven scheduling policy, one approach is to apply AoI-based scheduling, which has been studied for scheduling of independent measurements [5], [47], [49]–[54]. Recently, a number of works have studied AoI-based scheduling of dependent observation. In [19], [82], the authors consider a system of energy-constrained sensors observing a spatio-temporally correlated random field and determine the transmission rate that maximizes the battery life-time for a given estimation accuracy. The same spatio-temporal model was considered in [83], which focuses on deriving the jointly optimal transmission rate and the spatial density among sensors to minimize the overall MSE. In [84], [85], the average AoI is minimized for a system where individual sensors observe multiple overlapping sources, whereas in [85], sensors only observe partial information from each source, and multiple status packet updates are required at the receiver for proper reconstruction. The system scenarios considered in [84], [85], are more applic-

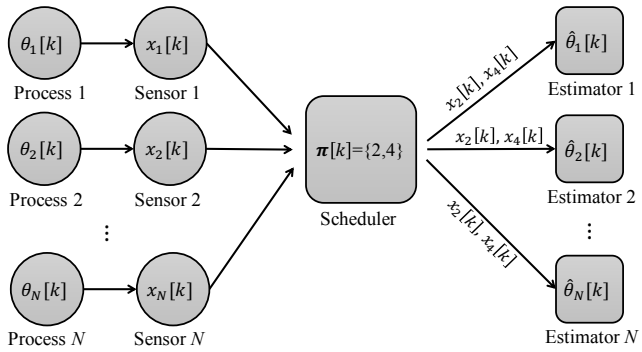

Figure 3.1: Schematic of WSN scheduling problem with $D = 2$.

able to a video-surveillance IoT security application than for remote estimation, in which the system utility depends on the estimation accuracy of the controllers and estimators.

Given the benefits of exploiting dependency among sensors to improve the overall estimation accuracy in remote estimation, we intend to extend the body of work concerning AoI-based scheduling of dependent observations. In the rest of this chapter, we present a summary of the works Publications **III** and **IV**, in which we derive an optimal AoI-based scheduling policy for a system of multiple sensors observing spatio-temporally dependent observations. Our work resembles that of [19], [82], [83], assuming a similar model for spatio-temporal dependency. The difference to the related works [19], [82], [83], can be summarized as follows.

- We allow for multiple sensors to be scheduled over multiple communication channels at each time instant.
- We do not assume periodic measurement scheduling. Therefore, in contrast to [19], [82], [83], we do not focus on finding update rates for individual sensors or learning the average transmission frequency for all sensors.
- We do not assume homogenous distributions among the processes.

The main contributions from Publications **III** and **IV** can be found in Contribution **C1** in Section 1.2.

3.2 System Model and Problem Formulation

We consider a WSN as presented in Section 2.1 that consist of N sensors, one scheduler, and N remote estimators as depicted in Fig. 3.1. The sensor observa-

tions are communicated to the remote estimators via a network scheduler. At time instant k, due to limited channel capacity, the scheduler broadcasts $D \leq N$, sensor observations to the remote estimators over D orthogonal channels. In our system model, we consider error-free parallel communication channels and assume that packet losses are addressed by retransmission through higher layers of the communication protocol.

Source Processes

Each process $\theta_i[k]$ follows a Gaussian distribution $\theta_i[k] \sim \mathcal{N}(0, \sigma_i^2)$. The processes $\{\theta_i[k]\}_{i=1}^N$ are correlated over space and time with the cross-covariance given by a positive-definite function [86], [87]

$$\mathbb{E}[\theta_i[k]\theta_j[l]] = \sigma_i\sigma_j\rho_{ij}\varphi(|k-l|), \quad i,j \in \{1,...,N\}, \tag{3.1}$$

where $\rho_{ij} \in [-1,1]$ represents the spatial correlation and $\varphi : \mathbb{R}_+ \to (0,1]$ is the temporal correlation, which is a strictly decreasing function with $\varphi(0) = 1$ and $\lim_{n\to\infty} \varphi(n) = 0$.

At time instant k, the ith sensor acquires measurement $x_i[k] \in \mathbb{R}$, i.e.,

$$x_i[k] = \theta_i[k] + w_i[k], \quad k \in \mathbb{N},\ i = 1,2,...,N, \tag{3.2}$$

where $w_i[k] \in \mathbb{R}$ denotes independent identically distributed (iid) measurement noise with distribution $w_i[k] \sim \mathcal{N}(0, \xi^2)$.

Scheduler

Let $\boldsymbol{\pi}[k] \in \{1,...,N\}^D$ be a scheduling variable denoting an index set of sensors to be scheduled at time k. The AoI of the ith sensor is denoted by $\Delta_i[k] \in \mathbb{N}_+$, $i = 1,...,N$, and defined as the time elapsed between two measurement transmissions, i.e.,

$$\Delta_i[k] = \begin{cases} 0, & \text{if } i \in \boldsymbol{\pi}[k], \\ \Delta_i[k-1]+1, & \text{if } i \notin \boldsymbol{\pi}[k]. \end{cases} \tag{3.3}$$

The scheduler is not allowed to observe the measurements, $\boldsymbol{x}[k] = [x_1[k], x_2[k], ..., x_N[k]]^{\mathrm{T}}$, but can keep track of the previous AoI at each sensor through vector $\boldsymbol{\Delta}[k-1]$, where $\boldsymbol{\Delta}[k-1] = [\Delta_1[k-1], \Delta_2[k-1], ..., \Delta_N[k-1]]^{\mathrm{T}}$. Let $\gamma_k : \boldsymbol{\Delta}[k-1] \to \{1,...,N\}^D$ denote the *scheduling strategy* at time k, i.e.,

$$\boldsymbol{\pi}[k] = \gamma_k(\boldsymbol{\Delta}[k-1]), \tag{3.4}$$

which provides a mapping from $\boldsymbol{\Delta}[k-1]$ to the scheduling decision at instant k.

Remote Estimators

The data available at the ith remote estimator at time instant k consists of $\boldsymbol{\Delta}[k]$ and $\boldsymbol{y}[k] = [y_1[k], y_2[k], ..., y_N[k]]^{\mathrm{T}}$, where $y_i[k]$ is the most recently broadcast measurement from Sensor i, i.e.,

$$y_i[k] = x_i[k - \Delta_i[k]], \quad i = 1, ..., N. \tag{3.5}$$

The estimate $\hat{\boldsymbol{\theta}}[k] = [\hat{\theta}_1[k], \hat{\theta}_2[k], ..., \hat{\theta}_N[k]]^{\mathrm{T}}$ is the linear minimum mean square error (MMSE) estimate [88] given as a function of $\boldsymbol{\Delta}[k]$ and $\boldsymbol{y}[k]$ as follows,

$$\hat{\boldsymbol{\theta}}[k] = \mathbb{E}[\boldsymbol{\theta}[k]|\boldsymbol{\Delta}[k], \boldsymbol{y}[k]] = \mathbf{C}_{\theta y}[k]\mathbf{C}_{yy}^{-1}[k]\boldsymbol{y}[k], \tag{3.6}$$

where the elements of the cross-covariance and covariance matrices are given by

$$\begin{aligned}
[\mathbf{C}_{\theta y}[k]]_{i,j} &= \sigma_i \sigma_j \rho_{ij} \varphi(\Delta_j[k]), \ i = 1, ..., N, j = 1, ..., N, \\
[\mathbf{C}_{yy}[k]]_{i,j} &= \sigma_i \sigma_j \rho_{ij} \varphi(\Delta_{ij}[k]) + \xi^2 \delta(i - j),
\end{aligned} \tag{3.7}$$

with $\Delta_{ij}[k] = |\Delta_i[k] - \Delta_j[k]| \in \mathbb{N}_+$ being the AoI differences between the two processes, and $\delta(\cdot)$ the Dirac delta function.

Scheduling Policy

The *scheduling policy* γ is defined as the collection of scheduling strategies from time instant $k = 1$ to $k = T$, i.e., $\gamma = (\gamma_1, \gamma_2, \dots \gamma_T)$. As performance measure (cost), we adopt the total mean squared error (MSE) of the estimate (3.6) over T time slots, given by

$$J(\gamma, T) = \frac{1}{TN} \sum_{k=1}^{T} \sum_{i=1}^{N} \mathbb{E}\Big[(\theta_i[k] - \hat{\theta}_i[k])^2 \Big| \gamma, \boldsymbol{\Delta}[k-1]\Big], \tag{3.8}$$

where $\boldsymbol{\Delta}[0]$ is known at the scheduler.

Problem Statement

Our objective is to find an *optimal scheduling policy* γ^* that minimizes the average cost in (3.8) over an infinite time horizon

$$\min_{\gamma \in \Gamma} \lim_{T \to \infty} J(\gamma, T), \tag{3.9}$$

where Γ is the set of all feasible policies.

3.3 Optimal Scheduling of Multiple Sensors

In this section, we show how (3.9) can be solved by modeling the problem as a Markov decision process (MDP). A mathematical introduction to MDPs can be found in Section 2.3.

3.3.1 Theoretical Results

To solve (3.9), we must be able to calculate the cost in (3.8), which depends on the process $\boldsymbol{\Delta}[k]$ during interval $k \in [1, T]$. The MSE at instant k can be expressed as a function $f : \boldsymbol{\Delta}[k] \to \mathbb{R}_+$, i.e.,

$$\begin{aligned} f(\boldsymbol{\Delta}[k]) &= \sum_{i=1}^{N} \mathbb{E}\Big[(\theta_i[k] - \hat{\theta}_i[k])^2 \Big| \boldsymbol{\Delta}[k]\Big] \\ &= \mathrm{tr}\Big(\mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta y}[k]\mathbf{C}_{yy}^{-1}[k]\mathbf{C}_{\theta y}^{\mathrm{T}}[k]\Big), \end{aligned} \tag{3.10}$$

where $\mathbf{C}_{\theta\theta}$ is the covariance matrix of $\boldsymbol{\theta}[k]$ and $\mathrm{tr}(\cdot)$ denotes the trace of its argument matrix.

From (3.3) and (3.10) we see that the MSE at instant k depends on $\boldsymbol{\Delta}[k]$, which in turn depends on $\boldsymbol{\Delta}[k-1]$ and $\boldsymbol{\pi}[k]$. For this reason, the problem in (3.9) can be modeled as the Markov decision process, where at instant k, the state is $\boldsymbol{\Delta}[k-1]$, the action is $\boldsymbol{\pi}[k]$ and the reward is $-N^{-1}f(\boldsymbol{\Delta}[k])$. The mathematical model fulfills Markovian properties (see Section 2.3) such that the state transition and reward at time k only depend on the previous state and the current action.

Markov Decision Process Formulation

Before defining the MDP, we present the set of all feasible values of $\boldsymbol{\Delta}[k]$ that will represent the state-space. If a round-robin scheduling policy [77] is applied, there would be a maximum AoI across all sensors, denoted as $\bar{\Delta} \in \mathbb{N}_+$, i.e.,

$$\begin{aligned} \bar{\Delta} &= \min_{\gamma \in \Gamma} \lim_{k\to\infty} \sup\, \mathbb{E}\big[\Delta_i[k]\big|\gamma\big], \quad \forall i = 1, 2, ..., N, \\ &= \begin{cases} N/D - 1, & \text{if } 0 = N \mod D, \\ \lfloor N/D \rfloor, & \text{else}, \end{cases} \end{aligned} \tag{3.11}$$

where $\lfloor \cdot \rfloor$ is the floor operator. For a round-robin scheduling policy, at each time instant k, there will be $\bar{N}$ sensors with an AoI equal to $\bar{\Delta}$. The value $\bar{N}$ is given by

$$\bar{N} = \begin{cases} D, & \text{if } 0 = N \mod D, \\ N \mod D, & \text{else.} \end{cases} \tag{3.12}$$

Let $c : \mathbb{N}_+^N \times \mathbb{N}_+ \to \mathbb{N}_+$ be an operator counting the number of elements in a vector $\boldsymbol{u} = [u_1, u_2, ..., u_N]^{\mathrm{T}}$ that equal to $l \in \mathbb{N}_+$, i.e.,

$$c(\boldsymbol{u}, l) = \sum_{i=1}^{N} \mathbb{1}(u_i = l),$$

where $\mathbb{1}(\cdot)$ is an indicator function having value 1 if the condition in the argument is true and 0 otherwise.

Given N, D, (3.11) and (3.12), the set of possible AoI values $\mathcal{S}$, $\boldsymbol{\Delta}[k] \in \mathcal{S}$, $k \in \mathbb{N}_+$ generated by any policy $\gamma \in \Gamma$ becomes

$$\begin{aligned}\mathcal{S} = \Big\{ \boldsymbol{u} \in \mathbb{N}_+^N \mid & c(\boldsymbol{u}, 0) = D, \\ & c(\boldsymbol{u}, l) \leq D, \quad l \in \mathbb{N}_{++}, \\ & c(\boldsymbol{u}, \bar{\Delta}) \geq \bar{N} \Big\}.\end{aligned} \tag{3.13}$$

Assumption 3.1. *We assume $\boldsymbol{\Delta}[k] \in \mathcal{S}$, for $k \in \mathbb{N}_+$*

Definition 3.1. *We define the MDP $\mathcal{M}$ in the following way;*

- ***State** at instant k is $\boldsymbol{\Delta}[k-1]$ and state space $\mathcal{S}$.*
- ***Action** at instant k is $\boldsymbol{\pi}[k]$ and the action space $\mathcal{A} = \{1, ..., N\}^D$.*
- ***Transition probabilities** $P(\boldsymbol{\Delta}[k] \mid \boldsymbol{\Delta}[k-1], \boldsymbol{\pi}[k])$, given state and action at instant k, can be derived using (3.3).*
- ***Reward** $r(\boldsymbol{\Delta}[k-1], \boldsymbol{\pi}[k]) \in \mathbb{R}_-$ at instant k equals $-N^{-1} f(\boldsymbol{\Delta}[k])$ in (3.10) and is given by the reward function $r : \{\mathcal{S}, \mathcal{A}\} \to \mathbb{R}_-$.*

Comparing (3.3) and (3.9) with Definition 3.1, we can see that an optimal scheduling policy γ^* also satisfies maximizing the time average reward in $\mathcal{M}$, i.e.,

$$\min_{\gamma \in \Gamma} \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}\Big[r(\boldsymbol{\Delta}[k-1], \boldsymbol{\pi}[k]) \Big| \gamma, \boldsymbol{\Delta}[0] \Big]. \tag{3.14}$$

One way to find γ^* is to use dynamic programming to derive an average reward optimal policy in $\mathcal{M}$ that satisfies (3.14) for all possible initial states, i.e., $\forall \boldsymbol{\Delta}[0] \in \mathcal{S}$. However, as shown in (3.13), the MDP $\mathcal{M}$ has an infinite countable state-space $\mathcal{S}$, for which an average reward optimal policy γ^* may not exist or is prohibitively complex to derive [67]. Therefore, we shall introduce another state-variable for the MDP that corresponds to a finite state-space.

Finite-state MDP

Following are two important properties of the function $f(\boldsymbol{\Delta}[k])$ in (3.10) that enables the finite-state MDP model. Firstly, due to the declining temporal correlation

in (3.1), the MSE increases with respect to the AoI, i.e.,

$$f([\Delta_1[k], ..., \Delta_i[k], ..., \Delta_N[k]]^{\mathrm{T}}) \leq \\ f([\Delta_1[k], ..., \Delta_i[k]+1, ..., \Delta_N[k]]^{\mathrm{T}}), \quad i = 1, ..., N. \tag{3.15}$$

Secondly, the function is upper bounded by the sum of the marginal variances, i.e.,

$$f(\boldsymbol{\Delta}[k]) \leq \mathrm{tr}\Big(\mathbf{C}_{\theta\theta}\Big) = \sum_{i=1}^{N} \sigma_i^2. \tag{3.16}$$

From (3.10), (3.15) and (3.16), we see that as the AoI grows, the temporal correlation becomes negligible, and the MSE does not increase with respect to the marginal AoI, i.e.,

$$\lim_{\Delta_i[k]\to\infty} |f([\Delta_1[k], ..., \Delta_i[k]+1, ..., \Delta_N[k]]^{\mathrm{T}}) \\ - f([\Delta_1[k], ..., \Delta_i[k], ..., \Delta_N[k]]^{\mathrm{T}})| = 0, \quad i = 1, ..., N.$$

This implies that the growth of MSE decreases with each instant the same sensors are consecutively scheduled, and, secondly, that many AoI states in $\mathcal{S}$ correspond to approximately the same MSE values. Following, we will demonstrate that some states in $\mathcal{M}$ can be merged to model the scheduling problem as a finite state-space MDP. Since φ in (3.1) is continuous, we restrict the set of possible correlation functions φ as stated in Assumption 3.2.

Assumption 3.2. *The temporal correlation function $\varphi : \mathbb{R}_+ \to [0, 1]$ in (3.1), satisfies $\varphi(x) = 0$, for all $x \geq m$, $m \in \mathbb{N}_{++}$.*

Assumption 3.2, together with (3.7), gives that the information at Estimator j, $y_j[k]$, whose AoI exceeds m, i.e., $\Delta_j[k] \geq m$, is uncorrelated with all processes at time k, i.e., $\mathrm{E}\big[\theta_i[k] y_j[[k]\big] = 0, \forall i = 1, ..., N$. As a consequence, the infinite state space $\mathcal{S}$ maps to a finite-set of MSE values, i.e., $f : \mathcal{S} \to \mathcal{Y}$ with $|\mathcal{Y}| < \infty$. This gives that any of the elements $\tilde{\Delta}_i[k], \tilde{\Delta}_{ij}[k] \in \{0, 1, ..., m\}$, $\forall i, j = 1, .., N$, belonging to the AoI vector $\boldsymbol{\Delta}[k]$ can be truncated to m while still corresponding to the same MSE value in (3.10).

Based on the former mentioned properties, we introduce a variable that pertains to all possible MSE values and belongs to a finite set. Let $\tilde{\boldsymbol{\Delta}}[k] \in \{0, 1, ..., m\}^{N^2}$ contain the elements $\tilde{\Delta}_i[k], \tilde{\Delta}_{ij}[k] \in \{0, 1, ..., m\}$, $\forall i, j = 1, .., N$, i.e.,

$$\tilde{\Delta}_i[k] = \big[\Delta_i[k]\big]_+^m, \quad i = 1, ..., N, \\ \tilde{\Delta}_{ij}[k] = \big[|\Delta_i[k] - \Delta_j[k]|\big]_+^m = \big[\Delta_{ij}[k]\big]_+^m, \quad i, j = 1, ..., N, \tag{3.17}$$

where $m \in \mathbb{N}_+$, $[\,\cdot\,]_+^m$ is defined as the truncation operator $[x]_+^m \triangleq \min\{x, m\}$, $x \in \mathbb{R}_+$ and $\tilde{\boldsymbol{\Delta}}[k]$ denotes the truncated AoI [89].

Let $\tilde{b} : \mathbb{N}_+^N \to \{0, 1, ..., m\}^{N^2}$ be a mapping from $\boldsymbol{\Delta}[k]$ to $\tilde{\boldsymbol{\Delta}}[k]$, i.e., $\tilde{\boldsymbol{\Delta}}[k] = \tilde{b}(\boldsymbol{\Delta}[k])$. Applying $\tilde{b}$ on the set of possible AoI values $\mathcal{S}$ in (3.13), gives the finite set of possible truncated AoI values

$$\tilde{\mathcal{S}} = \{\tilde{b}(\boldsymbol{\Delta}) \mid \boldsymbol{\Delta} \in \mathcal{S}\}. \tag{3.18}$$

We can express the MSE as a function of $\tilde{\boldsymbol{\Delta}}[k]$ as follows

$$\begin{aligned} \tilde{f}(\tilde{\boldsymbol{\Delta}}[k]) &= \sum_{i=1}^{N} \mathbb{E}\Big[(\theta_i[k] - \hat{\theta}_i[k])^2 \Big| \tilde{\boldsymbol{\Delta}}[k]\Big] \\ &= \mathrm{tr}\Big(\mathbf{C}_{\theta\theta} - \tilde{\mathbf{C}}_{\theta y}[k](\tilde{\mathbf{C}}_{yy})^{-1}[k](\tilde{\mathbf{C}}_{\theta y}[k])^{\mathrm{T}}\Big), \end{aligned} \tag{3.19}$$

with $\tilde{\mathbf{C}}_{yy}[k]$ and $\tilde{\mathbf{C}}_{\theta y}[k]$ calculated using $\tilde{\boldsymbol{\Delta}}[k]$ as

$$\begin{aligned} [\tilde{\mathbf{C}}_{yy}[k]]_{i,j} &= \sigma_i \sigma_j \rho_{ij} \varphi(\tilde{\Delta}_{ij}[k]) + \xi^2 \delta(i-j), \\ [\tilde{\mathbf{C}}_{\theta y}[k]]_{i,j} &= \sigma_i \sigma_j \rho_{ij} \varphi(\tilde{\Delta}_{j}[k]), \quad i, j \in \{1, ..., N\}. \end{aligned} \tag{3.20}$$

Proposition 3.1. *Under Assumption 3.2, the following relationship holds*

$$f(\boldsymbol{\Delta}[k]) = \tilde{f}(\tilde{\boldsymbol{\Delta}}[k]), \quad \forall \boldsymbol{\Delta}[k] \in \mathbb{N}_+^N. \tag{3.21}$$

Proof. The proof is given in Publication **III**. □

In the following proposition, we show that the truncated AoI, $\tilde{\boldsymbol{\Delta}}[k]$, satisfies the Markovian property that it can be expressed as a function of the previous value $\tilde{\boldsymbol{\Delta}}[k-1]$ and scheduling variable $\boldsymbol{\pi}[k]$.

Proposition 3.2. *The truncated AoI $\tilde{\boldsymbol{\Delta}}[k]$ can be expressed as a function of $\tilde{\boldsymbol{\Delta}}[k-1]$ and $\boldsymbol{\pi}[k]$ as*

$$\tilde{\Delta}_i[k] = \begin{cases} 0, & \text{if } i \in \boldsymbol{\pi}[k], \\ [\tilde{\Delta}_i[k-1] + 1]_+^m, & \text{if } i \notin \boldsymbol{\pi}[k], \end{cases} \tag{3.22}$$

$$\tilde{\Delta}_{ij}[k] = \begin{cases} 0, & \text{if } i, j \in \boldsymbol{\pi}[k], \\ [\tilde{\Delta}_{ij}[k-1]]_+^m, & \text{if } i, j \notin \boldsymbol{\pi}[k], \\ [\tilde{\Delta}_i[k-1] + 1]_+^m, & \text{if } i \notin \boldsymbol{\pi}[k],\ j \in \boldsymbol{\pi}[k], \\ [\tilde{\Delta}_j[k-1] + 1]_+^m, & \text{if } i \in \boldsymbol{\pi}[k],\ j \notin \boldsymbol{\pi}[k]. \end{cases} \tag{3.23}$$

Proof. The proof is given in Publication **III**. □

Proposition 3.1 and Proposition 3.2 demonstrate that $\tilde{\mathbf{\Delta}}[k]$ and $\mathbf{\Delta}[k]$ corresponds to the same MSE and, if either $\mathbf{\Delta}[k]$, or $\tilde{\mathbf{\Delta}}[k]$, is known, any given scheduling sequence that follows after k will result in the same sequence of MSE values. Thus, if Assumption 3.2 holds, we can model the scheduling problem as finite state space using the truncated AoI.

Definition 3.2. *We define the finite state-space MDP, $\tilde{\mathcal{M}}$, as follows;*

- ***Action** at instant k is the scheduling decision $\boldsymbol{\pi}[k]$ belonging to action-space $\mathcal{A} = \{1, ..., N\}^D$.*
- ***State** at instant k is the truncated AoI $\tilde{\mathbf{\Delta}}[k-1]$ belonging to state-space $\tilde{\mathcal{S}}$ in (3.18).*
- ***Transition probabilities** $P(\tilde{\mathbf{\Delta}}[k] \mid \tilde{\mathbf{\Delta}}[k-1], \boldsymbol{\pi}[k]) \in \{0, 1\}$ are binary and given by (3.22) and (3.23) in Proposition 3.2.*
- ***Reward** $\tilde{r}(\tilde{\mathbf{\Delta}}[k-1], \boldsymbol{\pi}[k]) \in \mathbb{R}$ at instant k equals $\tilde{r}[k] = -N^{-1}\tilde{f}(\tilde{\mathbf{\Delta}}[k])$ in (3.19) given by the reward function $\tilde{r} : \{\tilde{\mathcal{S}}, \mathcal{A}\} \to \mathbb{R}_-$.*

Similar to (3.4), let $\tilde{\gamma}_k : \tilde{\mathcal{S}} \to \mathcal{A}$ be a scheduling strategy based on $\tilde{\mathbf{\Delta}}[k]$ as

$$\boldsymbol{\pi}[k] = \tilde{\gamma}_k(\tilde{\mathbf{\Delta}}[k-1]),$$

where $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, ..., \tilde{\gamma}_T)$ is a scheduling policy $\tilde{\gamma} \in \tilde{\Gamma}$. We define an *optimal truncated scheduling policy* $\tilde{\gamma}^*$ as a policy that maximizes the time average reward in $\tilde{\mathcal{M}}$ for all possible initial states $\tilde{\mathbf{\Delta}}[0]$, i.e,

$$\max_{\tilde{\gamma} \in \tilde{\Gamma}} \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}\Big[\tilde{r}(\tilde{\mathbf{\Delta}}[k-1], \boldsymbol{\pi}[k]) \Big| \tilde{\gamma}, \tilde{\mathbf{\Delta}}[0]\Big], \quad \forall \tilde{\mathbf{\Delta}}[0] \in \tilde{\mathcal{S}}.$$

Since the state-space $\tilde{\mathcal{S}}$ is finite, an optimal truncated scheduling policy $\tilde{\gamma}^*$ possible to derive numerically using dynamic programming. In the following theorem we present the relationship between $\tilde{\gamma}^*$ and an optimal scheduling policy γ^*.

Theorem 3.1. *Under Assumption 3.1 and Assumption 3.2, there exists a stationary optimal scheduling policy $\gamma^* = (\gamma_0^*, ..., \gamma_0^*)$, where $\gamma_0^* = \tilde{\gamma}_0^* \circ \tilde{b}$ and $\tilde{\gamma}^* = (\tilde{\gamma}_0^*, \tilde{\gamma}_0^*, ..., \tilde{\gamma}_0^*)$, which can be derived in a finite number of iterations using policy iteration. The policy results in a periodic scheduling sequence.*

Algorithm 1 Finding scheduling policy γ^*

1: Define $\tilde{\mathcal{M}} = \{\mathcal{A}, \tilde{\mathcal{S}}, \tilde{r}, P(\cdot \mid \cdot)\}$, given N, D, $\mathbf{C}_{\theta\theta}$, ξ, φ and m
2: Set $n = 0$ and select arbitrary policy $\tilde{\gamma}^n \in \tilde{\Gamma}^S$
3: Obtain $g \in \mathbb{R}$ and $\boldsymbol{h} \in \mathbb{R}^{|\tilde{\mathcal{S}}|}$, $[\boldsymbol{h}]_i = h(\tilde{\boldsymbol{\Delta}}_i)$, $\forall \tilde{\boldsymbol{\Delta}}_i \in \tilde{\mathcal{S}}$, by solving below equation

$$\tilde{\boldsymbol{r}}_{\tilde{\gamma}^n} + [-\mathbf{1} \mid (\mathbf{P}_{\tilde{\gamma}^n} - \mathbf{I})] \begin{bmatrix} g \\ \boldsymbol{h} \end{bmatrix} = \mathbf{0},$$

where $\tilde{\boldsymbol{r}}_{\tilde{\gamma}^n} \in \mathbb{R}^{|\tilde{\mathcal{S}}|}$, $[\tilde{\boldsymbol{r}}]_i = r(\tilde{\boldsymbol{\Delta}}_i, \tilde{\gamma}^n(\tilde{\boldsymbol{\Delta}}_i))$, $\forall \tilde{\boldsymbol{\Delta}}_i \in \tilde{\mathcal{S}}$, is a reward vector, $\mathbf{1} = (1, 1, ..., 1)^{\mathrm{T}}$, $\mathbf{1} \in \mathbb{R}^{|\tilde{\mathcal{S}}|}$ is a vector of ones, $\mathbf{P}_{\tilde{\gamma}^n} \in \mathbb{R}^{|\tilde{\mathcal{S}}| \times |\tilde{\mathcal{S}}|}$, $[\mathbf{P}_{\tilde{\gamma}^n}]_{i,j} = P(\tilde{\boldsymbol{\Delta}}_j | \tilde{\boldsymbol{\Delta}}_i, \tilde{\gamma}^n(\tilde{\boldsymbol{\Delta}}_i))$ is the transition matrix..
4: Get policy $\tilde{\gamma}^{n+1} = (\tilde{\gamma}_0^{n+1}, \tilde{\gamma}_0^{n+1}, ..., \tilde{\gamma}_0^{n+1})$, $\forall \tilde{\boldsymbol{\Delta}} \in \mathcal{S}$, by solving

$$\tilde{\gamma}_0^{n+1}(\tilde{\boldsymbol{\Delta}}) = \arg\max_{\boldsymbol{\pi} \in \mathcal{A}} \{ r(\tilde{\boldsymbol{\Delta}}, \boldsymbol{a}) + \sum_{\tilde{\boldsymbol{\Delta}}' \in \tilde{\mathcal{S}}} P(\tilde{\boldsymbol{\Delta}}' | \tilde{\boldsymbol{\Delta}}, \boldsymbol{a}) h(\tilde{\boldsymbol{\Delta}}') \}$$

5: **if** $\tilde{\gamma}^{n+1} = \tilde{\gamma}^n$ **then**
6: Stop and set $\tilde{\gamma}^* = \tilde{\gamma}^{n+1}$
7: **else**
8: Return to Step 2 using $\tilde{\gamma}^{n+1}$
9: **end if**
10: Obtain $\gamma^* = (\gamma_0^*, ..., \gamma_0^*)$ as $\gamma_0^* := \tilde{\gamma}_0^* \circ b$

Proof. The proof is given in Publication **III**. □

Theorem 3.1 states that γ^* results in a periodic scheduling sequence. In Publication **III**, due to the finite-state space, we prove that applying a periodic scheduling sequence results in the same periodic sequence of truncated AoI states, regardless of the initial truncated AoI. This result can be utilized to save data storage at the scheduler, as it only needs to store the periodic scheduling pattern and not the entire policy, mapping every truncated AoI value to a scheduling decision. Furthermore, as shown in Publication **III**, it becomes straightforward to calculate the performance of any periodic scheduling policy.

Although the state-space is finite, the dimensionality grows rapidly with the number of sensors N and the truncation value m. As N and m grows, it can quickly become prohibitively complex to derive an optimal policy γ^*. In the next section, we compare the performance of an optimal scheduling policy to periodic sub-optimal policies that can be derived with less computational effort.

3.3.2 Numerical Results

We assume a system where N sensors observe dependent processes with equal marginal variances, i.e., $\sigma_i = 1$, $i = 1, ..., N$, measurement noise $\xi = 0.5$ and the spatio-temporal dependency components in (3.1) are given by [86], [87]

$$\rho_{ij} = e^{-r_0|i-j|}, \quad \varphi(x) = e^{-\lambda_t x}\mathbb{1}(e^{-\lambda_t x} \geq 0.1),\ x \in \mathbb{R}_+,$$

where $\lambda_t \in \mathbb{R}_+$, is the temporal correlation decay factor, and $r_0 \in \mathbb{R}_+$, is the spatial correlation decay factor.

We compare the results of an optimal policy to a set of sub-optimal policies, presented in Definition 3.3 below. A more detailed mathematical presentation for each sub-optimal policy can be found in Publication **III**.

Definition 3.3. *Sub-optimal policies;*

- ***Random (RANDOM)*** *- Randomly selecting D sensors at each time instant.*
- ***Round robin (RR)*** *- Scheduling the D sensors with the highest AoI in a round-robin fashion.*
- ***Optimized round robin (ORR)*** *- Applying a round-robin periodic scheduling sequence that is organized in an optimal way to minimize the time average MSE.*
- ***Finite horizon minimization (FHM-L)*** *- The scheduler minimizes the time average MSE over L time steps given the current truncated AoI.*

Notably, all the sub-optimal policies in Definition 3.3, except for RANDOM, result in a periodic scheduling sequence that can be efficiently executed at the network scheduler.

Figure 3.2 shows the asymptotic average cost versus λ_t for different scheduling policies, where $N = 5$, $D = 2$, and $r_0 = 0.5$. Solid lines show theoretical results, whereas dots represent numerical simulations, calculated by averaging 200 ensembles for a time horizon of $T = 100$. We see that as the correlation increases, i.e., $\lambda_t \to 0$, all policies results in a decreasing cost. We see that for a higher temporal correlation, both FHM policies and ORR, performs close to optimal. Most interestingly, for $\lambda_t > 0.5$, a RANDOM policy outperforms RR, whereas ORR consistently outperforms RANDOM. This indicates the performance improvement of selecting an optimal scheduling order for a RR-policy that takes into account the spatial correlations between the sensors.

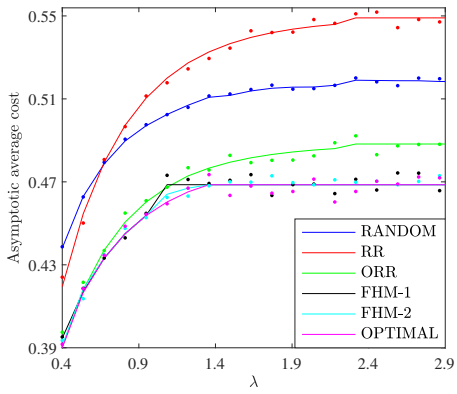

Figure 3.2: Asymptotic average cost, $\lim_{T\to\infty} J(\gamma, T)$, vs λ_t with $N = 5$, $D = 2$, $\sigma_i = 1$, $\forall i = 1, 2, ..., N$, $\xi = 0.5$, and $r_0 = 0.5$.

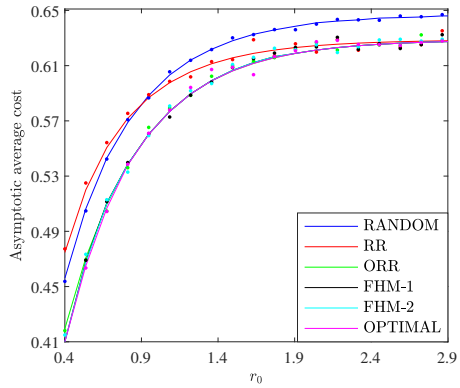

Figure 3.3: Asymptotic average cost, $\lim_{T\to\infty} J(\gamma, T)$, vs r_0 with $N = 5$, $D = 2$, $\sigma_i = 1$, $\forall i = 1, 2, ..., N$, $\xi = 0.5$, and $\lambda_t = 0.8$.

Figure 3.3 shows the asymptotic average cost versus the distance between neighboring sensors for $N = 5$, $D = 2$, and $T = 0.8$. Similar to Figure 3.2, we see that as the spatial correlation increases, i.e., $r_0 \to 0$, all policies results in a decreasing cost. Both FHM policies and ORR performs near-optimal, and consistently outperforms RANDOM and RR. At $r_0 < 0.8$, RR is outperformed by RANDOM, since the temporal correlation is surpassed by the spatial correlation, which an RR does not exploit efficiently.

3.4 Scheduling of Two Sensors with a Transmission Constraint

In this section, we study a particular case of the scheduling problem presented in Section 3.2. We consider a system scenario of two sensors, i.e., $N = 2$, scheduled over a finite time horizon, i.e., $T < \infty$ in (3.9), where the number of scheduling instances for each sensor is limited by an individual transmission constraint. Commonly, sensor have limited energy resources and so, the transmission rate of observations and estimators are not only limited by the number of communication channels but also by their energy supplies.

In general, adding transmission constraints and assuming a finite time horizon adds complexity to the scheduling problem. However, for non-stationary processes or stochastic energy supply, e.g., energy harvesting, optimal scheduling policies for finite horizons can offer better performance. Fortunately, as we will demonstrate for the $N = 2$ and $D = 1$ system scenario, we are able to derive an optimal policy

for a given horizon using a low-complexity numerical method presented in this section.

To include a transmission constraint in the scheduling problem in (3.9), we must first introduce some additional notations. Assume that $\mathbf{\Delta}[0] = [1, 0]^{\mathrm{T}}$ when initializing the system and let $\mathbf{\Delta}^{\gamma}[k] = [\Delta_1^{\gamma}[k], \Delta_2^{\gamma}[k]]^{\mathrm{T}}$ denote the AoI at time k generated by policy γ. Let n_1^{γ} and n_2^{γ}, respectively, denote the number of instants Sensor 1 and Sensor 2 are scheduled using policy γ. The values of n_1^{γ} and n_2^{γ} can be computed as

$$n_1^{\gamma} = \sum_{i=1}^{T} \mathbb{1}(\Delta_1^{\gamma}[i] = 0), \quad n_2^{\gamma} = \sum_{i=1}^{T} \mathbb{1}(\Delta_2^{\gamma}[i] = 0), \tag{3.24}$$

where $n_1^{\gamma}, n_2^{\gamma} \in \{0, 1, ..., T\}$, $n_1^{\gamma} + n_2^{\gamma} = T$.

Without loss of generality, let us assume each transmission costs unit energy. Also, let $\bar{n}_1, \bar{n}_2 \in \mathbb{N}_{++}$, be the energy resources available for transmitting data, i.e., Sensor i can transmit $\bar{n}_i$ measurements. We assume that the total energy available at the sensors satisfies $\bar{n}_1 + \bar{n}_2 \geq T$.

Our objective is to find an *optimal scheduling policy* γ^* that minimizes the average cost in (3.8) over any time horizon

$$\begin{aligned} \min_{\gamma \in \Gamma} \quad & J(\gamma, T), \\ \text{s. t.} \quad & n_i^{\gamma} \leq \bar{n}_i, \quad \bar{n}_i \in \mathbb{N}_{++},\, i = 1, 2, \\ & n_1^{\gamma} + n_2^{\gamma} = T, \end{aligned} \tag{3.25}$$

where Γ is the set of all feasible policies.

Remark 3.1. In Section 3.4, for the temporal correlation φ in (3.1), we do not restrict the set of feasible functions to Assumption 3.2. Instead, φ can either be continuous, or satisfy Assumption 3.2.

3.4.1 Theoretical Results

In this section, we present a summary of the theoretical results presented in Publication **IV** and demonstrate how to derive γ^* that solves (3.25). We begin by deriving an optimal scheduling policy for the specific case; having no transmission constraints, i.e., $\bar{n}_i > T$, $i = 1, 2$, and the number of transmission instances for the ith sensor must equal $n_i \in \mathbb{N}_+$, $i = 1, 2$. This results in the following optimization problem

$$\begin{aligned} \min_{\gamma \in \Gamma} \quad & J(\gamma, T), \\ \text{s. t.} \quad & n_i^{\gamma} = n_i, \quad i = 1, 2, \quad n_1 + n_2 = T. \end{aligned} \tag{3.26}$$

Later on, we show how to derive an optimal number of scheduling instances for each sensor, n_i^*, $i = 1, 2$. To simplify the calculations, we present a numerical method to reduce the feasible set of n_i^*, $i = 1, 2$, to a set of two pairs to evaluate. In the following theorem, we present a solution to (3.26). Before doing so, we need to introduce some additional notations.

Similar to (3.11), let $\bar{\Delta}_i : \mathbb{N}_+ \to \{0, 1, ..., T+1\}$, represent the lowest maximum AoI (minmax) for Sensor i, given that Sensor i is scheduled l instances during interval $k \in [1, T]$, and be defined as

$$\bar{\Delta}_i(l) = \min_{\gamma \in \Gamma} \sup_{k=1,...,T} \left\{ \Delta_i^\gamma[k] \middle| n_i^\gamma = l \right\}, \quad i = 1, 2.$$

The value $\bar{\Delta}_i(l)$ can be expressed as the following function

$$\begin{aligned} \bar{\Delta}_1(l) &= \left\lceil \frac{T+1-l}{l+1} \right\rceil, \ \bar{\Delta}_2(l) = \left\lceil \frac{T-l}{l+1} \right\rceil, \ l = 0, ..., T-1 \\ \bar{\Delta}_1(T) &= \bar{\Delta}_2(T) = 0, \end{aligned} \tag{3.27}$$

where $\lceil \cdot \rceil$ is the ceil operator.

Finally, let $g_i^\gamma : \mathbb{N}_+ \to \mathbb{N}_+, i = 1, 2$, represent the number of instances the ith sensor reaches AoI $l \in \mathbb{N}_+$ given policy $\gamma \in \Gamma$, and be defined as

$$g_i^\gamma(l) = \sum_{k=1}^{T} \mathbb{1}(\Delta_i^\gamma[k] = l). \tag{3.28}$$

Theorem 3.2. *i) A policy $\gamma \in \Gamma$ is a solution to (3.26), if it minimizes the maximum AoI as*

$$\Delta_i^\gamma[k] \leq \bar{\Delta}_i(n_i), \quad k = 1, ..., T, \quad i = 1, 2, \tag{3.29}$$

and the number of instances the AoI equals $\bar{\Delta}_i(n_i)$ and $\bar{\Delta}_i(n_i) - 1$, $i =, 1, 2$, given in (3.27), can be computed as; if $n_i = T$ and $n_j = 0$, $i \neq j$, then $\bar{\Delta}_i(n_i) = 0$, and

$$g_i^\gamma(0) = n_i, \quad g_j^\gamma\left(\bar{\Delta}_j(0)\right) = 1, \tag{3.30}$$

else, if $1 \leq n_j \leq n_i$, then $\bar{\Delta}_i(n_i) = 1$, and

$$\begin{aligned} g_i^\gamma(1) = n_j, \quad g_i^\gamma(0) = n_i \\ g_j^\gamma\left(\bar{\Delta}_j(n_j)\right) = n_i - (n_j + 1)\left(\bar{\Delta}_j(n_j) - 1\right) + \mathbb{1}(j = 1), \\ g_j^\gamma\left(\bar{\Delta}_j(n_j) - 1\right) = n_j + 1 - \mathbb{1}(j = 1)\mathbb{1}\left(\bar{\Delta}_j(n_j) = 2\right) \\ - \mathbb{1}\left(\bar{\Delta}_j(n_j) = 1\right). \end{aligned} \tag{3.31}$$

ii) For problem (3.25), the optimal number of scheduling instances satisfies

$$\begin{aligned} n_1^* &= \bar{n}_1, && \text{if} \quad \bar{n}_1 \leq T/2 \\ n_1^* &\geq T/2, && \text{if} \quad \bar{n}_1 > T/2, \end{aligned} \tag{3.32}$$

and $n_2^* = T - n_1^*$.

Proof. The proof is given in Publication **IV**. □

Theorem 3.2 implies that if the number of scheduling instances n_1^* and n_2^* are known, we can use expressions (3.29) -(3.31) to derive an optimal policy γ^*. The criteria for an optimal policy is that; for Sensor i, $i = 1, 2$, the maximum AoI must be minimized, i.e., being equal to $\bar{\Delta}_i(n_i^*)$ in (3.29), and the number of instants the AoI reaches $\bar{\Delta}_i(n_i^*)$ and $\bar{\Delta}_i(n_i^*) - 1$ during time interval $[1, T]$ must satisfy (3.30) and (3.31). Hence, we conclude, that there can be one or more optimal policies as long as they satisfy (3.29) to (3.31).

From (3.32), we know that if $\bar{n}_1 \leq T/2$, then $n_1^* = \bar{n}_1 = T - n_2^*$ and we can again use (3.29)–(3.31), to obtain γ^*.

We will now show how to derive optimal number of transmission instances for each sensor, n_1^* and n_2^*, for the case $\bar{n}_1 > T/2$ in (3.32), to later obtain γ^*. For $\bar{n}_1 > T/2$, let $\mathcal{L}$ be the set of feasible values of n_2^*, $n_2^* \in \mathcal{L}$, given by

$$\mathcal{L} = \{n \in \mathbb{N}_+ | \max\{T - \bar{n}_1, 0\} \leq n \leq \min\{\bar{n}_2, \lfloor T/2 \rfloor\}\}.$$

Lemma 3.1. *For $\bar{n}_1 > T/2$, the optimal number of transmission instances for Sensor 2, n_2^*, should be equal to the value n, $n \in \mathcal{L}$, satisfying*

$$\min_{n \in \mathcal{L}} p(n) = J(\gamma^*, T), \tag{3.33}$$

with $p : \mathcal{L} \to \mathbb{R}_+$ defined as

$$p(n) = \begin{cases} \bar{f}(2), & \text{if } n = T/2, \\ \nu(n)\bar{f}(\bar{m}) + \omega(n)\bar{f}(\bar{m} - 1) + w_0, & \text{else}, \end{cases} \tag{3.34}$$

where $\bar{f}(\bar{m})$, $\nu(n), \omega(n)$ and $\bar{m}$ are given as

$$\begin{aligned} \bar{m} &= \left\lceil \frac{T+1}{n+1} \right\rceil, \quad w_0 = -\frac{f([1,0]^{\mathrm{T}})}{T}, \\ \nu(n) &= \frac{\bar{m}}{T}(T + 1 - (n+1)(\bar{m} - 1)), \\ \omega(n) &= \frac{\bar{m} - 1}{T}\left(n + 1 - \nu(n)\frac{T}{\bar{m}}\right), \\ \bar{f}(\bar{m}) &= \frac{\sum_{i=1}^{\bar{m}-1} f([0,i]^{\mathrm{T}}) + f([1,0]^{\mathrm{T}})}{\bar{m}}, \end{aligned}$$

with $\nu(n)+\omega(n) = (T+1)/T$ for all $n \in \mathcal{L}$, and $f(\cdot)$ is the function for the MSE given in (3.10).

Proof. The proof of Lemma 3.1 follows from the mathematical derivations presented in Publication **IV** in Section III.A. Note that some considered notations in this thesis and Publication **IV** differ. □

Lemma 3.1, shows that n_2^* can be derived by evaluating all possible values of $\mathcal{L}$ to see which minimizes (3.33). However, depending on the system parameters, the number of possible values of $\mathcal{L}$ can be substantially large. In the following theorem, we present a formula to reduce the set of feasible values of n_2^* to only two values.

Theorem 3.3. *For $\bar{n}_1 > T/2$, the optimal number of scheduling instances for Sensor 2, n_2^*, belongs to set $n_2^* \in \{n^-, n^+\}$, where*

$$n^- = \sup\left\{ n \in \mathcal{L} \,\middle|\, \left\lceil \frac{T+1}{n+1} \right\rceil > \hat{m} \right\},$$
$$n^+ = \inf\left\{ n \in \mathcal{L} \,\middle|\, \left\lceil \frac{T+1}{n+1} \right\rceil \leq \hat{m} \right\}, \tag{3.35}$$

and where the value $\hat{m}$ is $\hat{m} = \infty$ if

$$\lim_{\Delta_2[k]\to\infty} \sum_{i=1}^{\infty} \left(f([0, \Delta_2[k]]^{\mathrm{T}}) - f([0,i]^{\mathrm{T}}) \right) \leq f([1,0]^{\mathrm{T}}) - f([0, \Delta_2[k]]^{\mathrm{T}}), \tag{3.36}$$

else

$$\hat{m} = \inf\left\{ \bar{m} \geq 2 \,\middle|\, \frac{\sum_{i=1}^{\bar{m}-1} f([0,i]^{\mathrm{T}}) + f([1,0]^{\mathrm{T}})}{\bar{m}} \leq f([0,\bar{m}]^{\mathrm{T}}) \right\}. \tag{3.37}$$

Proof. The proof can be found in Publication **IV**. Note that some considered notations in this thesis and Publication **IV** differ. □

Using Theorem 3.3, we now have low-complexity numerical method to derive γ^*. For $\bar{n}_1 > T/2$, Theorem 3.3 implies that γ^* can be derived by, firstly, calculating $\hat{m}$ from expressions (3.36) and (3.37). This can be done in a straightforward way using the recursive method presented in Algorithm 2. Secondly, by deriving n^- and n^+ in (3.35) and evaluating the value that minimizes (3.33) as

$$n_2^* = \underset{n \in \{n^-, n^+\}}{\arg\min}\ p(n). \tag{3.38}$$

Finally, deriving γ^* by applying n_2^* to Theorem 3.2.

Algorithm 2 Finding $\hat{m}$

if Inequality (3.36) is satisfied **then**
  set $\hat{m} = \infty$
else
  Initialize $\bar{m} = 2$
  while $\bar{f}(\bar{m}) > f([0, \bar{m}]^{\mathrm{T}})$ **do**
    $\bar{f}(\bar{m}+1) = \dfrac{\bar{m}\bar{f}(\bar{m}) + f([0,\bar{m}]^{\mathrm{T}})}{\bar{m}+1}$
    $\bar{m} = \bar{m} + 1$
  end while
  set $\hat{m} = \bar{m}$
end if

3.4.2 Numerical Results

We assume a system with statistical parameters $\sigma_1 = 2$, $\sigma_2 = 1$, $\rho_{12} = -0.5$, and $\xi = 0.5$. For the temporal correlation φ in (3.1), we use $\varphi(x) = e^{-\lambda_t x}$, $x \in \mathbb{N}_+$, where $\lambda_t \in \mathbb{R}_+$ [86].

Fig. 3.4 shows the average cost $J(\gamma)$ versus n_1^γ for $\lambda_t = (0.1, 0.2, 0.4)$ with time horizon $T = 100$ and $\bar{n}_i = T$, $i = 1, 2$. Given the value n_1^γ, an optimal policy is derived from the results in Theorems 3.2, referred to as OPTIMAL. Solid lines depict theoretical values and markers show Monte Carlo simulations of 1000 sequences with $T = 100$, which matches the theory. The red markers show the performance given the optimal number of transmission instances $n_1^* = (76, 84, 100)$ given by Theorem 3.3 together with (3.38). The results show that for $\lambda_t \leq 0.2$, an optimal performance is achieved for $n_1^* < \bar{n}_1 = T$.

In Fig. 3.4, the optimal performance is compared to a policy where Sensor 1 is scheduled n_1^γ instances in a random order during interval $[1, T]$, referred to as RANDOM, similar as in Definition 3.3. The performance of RANDOM is calculated as the average MSE after simulating 1000 scheduling sequences for each value n_1^γ. We see that an optimal scheduling order outperforms a random scheduling policy for every value of n_1^γ. Furthermore, we see that the optimal number of scheduling instances for Sensor 1 for policies OPTIMAL and RANDOM lies close to each other.

Fig. 3.5 shows the optimal number of scheduling instances for Sensor 1, n_1^*, versus the spatial correlation ρ_{12} for $\lambda_t = (0.01, 0.05, 0.1, 0.14)$. The results show that n_1^* is lower bounded at $\rho_{12} = 0$, with values respectively $n_1^* = (75, 75, 88)$, and then increases with ρ_{12}. We see that as the temporal correlation increases, i.e.,

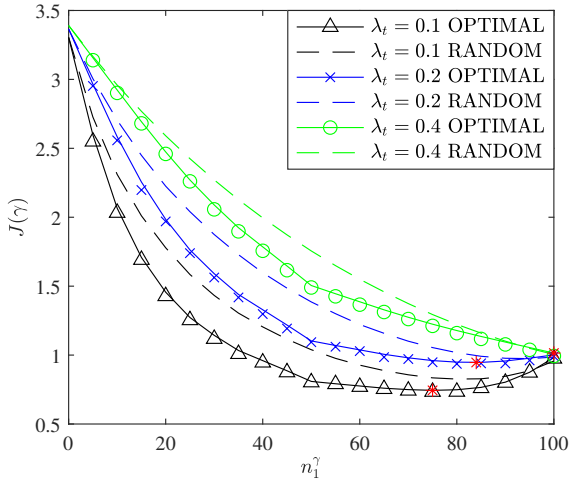

Figure 3.4: Average cost $J(\gamma)$ versus n_1^γ for system parameters $\sigma_1 = 2, \sigma_2 = 1, \rho_{12} = -0.5$, $\bar{n}_i \geq T, i = 1, 2$, and $\xi = 0.5$ and $T = 100$. Solid lines show results derived from theory and markers show simulation results. Red asterix show optimal performance at n_1^*.

$\lambda_t \to 0$, n_1^* decreases. This implies that optimal performance can be achieved at transmission constraint $\bar{n}_2 \leq 25$, for all $\rho_{12} \in [0, 1]$.

3.5 Discussion

In this chapter, we considered a scheduling problem in a WSN where sensors observe spatio-temporally dependent processes and transmit their measurements over a limited number of commutation channels to corresponding remote estimators. The system involves a measurement-blind network scheduler responsible for scheduling the sensors that broadcast every measurement to all remote estimators to improve the overall estimation accuracy. The task was to derive an optimal AoI-based scheduling policy that minimized the time average MSE over an infinite time horizon. By modeling the scheduling problem as a finite-state MDP, we proved the existence of an optimal policy resulting in a periodic scheduling sequence. We also presented a numerical method for deriving an optimal policy.

Our work extends the minority of work regarding AoI-based scheduling of dependent observation for remote estimation in WSN and WNCS. Numerical results show that spatio-temporal dependencies can be exploited in AoI-based scheduling for both channel and transmission constrained WSN to improve the overall estimation accuracy. Most previous work concerning the AoI has been focused on designing scheduling schemes or optimizing system parameters to minimize the AoI [35], [90]. In our work, numerical results showed that when designing policies that minimize the AoI, the estimation accuracy varied from worse than a

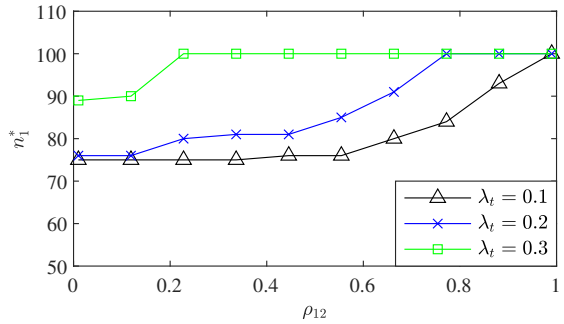

Figure 3.5: Optimal number of scheduling instances of Sensor 1, n_1^*, versus spatial correlation ρ_{12} for system parameters $\sigma_1 = 2, \sigma_2 = 1, \rho_{12} = -0.5$, $\bar{n}_i \geq T, i = 1, 2$, and $\xi = 0.5$.

randomized scheduling approach to near-optimal. To maximize performance, the intrinsic order of a round-robin scheduling policy should be designed with consideration for the marginal variances and the spatial dependencies.

By modeling the scheduling problem as finite-state MDP, we were able to show that any stationary scheduling policy results in a periodic scheduling pattern. Furthermore, regardless of the initial AoI, executing the periodic scheduling pattern would achieve the same time average MSE. A practical benefit of these results is that it can save data storage at the scheduler that only needs to execute the periodic scheduling sequence instead of keeping track of the current AoI while storing a lookup table for every possible AoI state.

We also studied the particular case of two sensors sharing a single communication channel where the number of transmission is limited by individual transmission constraints to account for limited energy resources. A low-complexity numerical method was presented for deriving an optimal scheduling policy for a finite time horizon. Although including a transmission constraint and defining a finite time horizon for the scheduling problem adds additional complexity, our method does not rely on dynamic programming, and the computational complexity does not become prohibitive beyond a given time horizon. A benefit of the presented numerical method is that it allows to quickly derive scheduling policies that minimize the estimation error over shorter time horizons when considering non-stationary processes or stochastic energy resources, e.g., energy harvesting. Our theoretical result can also be used to derive necessary energy resources to achieve optimal performance in similar systems.

Chapter 4

Optimized Threshold-triggered Transmission Schemes

This chapter presents our work regarding event-triggered transmission schemes, specifically threshold-based transmission schemes, where a sensor transmits a measurement if it surpasses a pre-defined threshold. In the first part of this chapter, we present our work regarding dual prediction schemes (DPS), introduced in 2.2. First, we discuss the implications of non-stationarity in measurement processes. Each time a change in measurement distribution occurs, we must update the DPS prediction model to avoid excessive transmission. Then, we present the results from Publication **V**, a cost-aware dual prediction scheme (CA-DPS), which extends the DPS framework to further reduce the data transmission energy cost. The CA-DPS framework is based on the idea that a sensor is given a set of options, referred to as transmission strategies, that determine whether or not an update should be transmitted.

In contrast to time-triggered scheduling schemes, event-triggered schemes can result in collisions due to a lack of centralized coordination preventing the number of simultaneous transmissions from exceeding the number of communication channels. In the second part of this chapter, we present the results from Publication **VI**, which considers a system of multiple sensors implementing event-triggered transmission over a limited number of shared channels. The sensors are divided into spatially distributed clusters assigned to a set of channels. The sensor measurement distributions are assumed heterogeneous among the clusters. The task is to derive optimal transmission threshold and channel allocation strategies for each cluster.

4.1 Cost-Aware Dual Prediction Scheme

In a DPS framework, the number of transmission instances depends on the prediction accuracy of future sensor measurements, which depends on the prediction model and the input data used for the predictions. Over time, the prediction accuracy may decline if; the measurement distribution changes or the input data set chiefly consists of previously predicted values. In practice, sensors often observe processes that are non-stationary and dynamic. Even physical processes that demonstrate clear periodicities over more extended periods, e.g., daily or weekly cycles, tend to show properties of non-stationarity at a more granular level.

To maintain an acceptable prediction accuracy in a DPS when the measurement distribution changes, the model parameters need to be re-estimated and updated at both the sensor and the FC. Otherwise, it will result in extensive threshold breaches and measurement transmission that drains the sensor battery. Furthermore, the predicted input variables need to be replaced with true measurements over time.

4.1.1 Prior Work

Previous works have incorporated similar protocols to update the prediction model and reset the input data in a DPS. In [33], [91], an update is triggered each time a measurement breaches a pre-defined error-tolerance level over a fixed number of instances. The works [18], [33], [57] consider an adaptive filter in a DPS, where the predicted measurement is solely based on previously registered measurements, including both predicted and true measurements, and do not depend on estimated model parameters. Specifically, in [18], [57], the sensor transmits a measure at every threshold breach but never refreshes all the input data with true measurements. Whereas in [33], once an error threshold is breached, measurements are transmitted at every time instant until the prediction error is below the error tolerance a fixed number of times. A linear prediction model is applied in [92], wherein the sensor re-estimates and transmits updated model parameters together with a recent measurement every time the prediction error is breached a consecutive number of time instances. In [91], the DPS is based on an ARIMA model where the model parameters are re-estimated and transmitted together with a true measurement each time the prediction error is breached.

Although previously mentioned works include protocols for updating the prediction model and transmitting sensor measurements, they do not evaluate whether such an update reduces the number of future transmission instances. Neither do they consider the additional energy cost of transmitting new model parameters. If the mathematical structure of the prediction model provides a poor fit for the measurement distribution, transmitting re-estimated model parameters might have

a marginal effect on reducing future transmission instances. Therefore, energy efficiency for communication can be enhanced by protocols that are selective in the transmission of measurement data and model parameters.

In Publication **V**, we introduce a cost-aware dual prediction scheme (CA-DPS) to reduce the data transmissions in the sensor-to-FC communication link in a DPS. Before each transmission instance, the sensor evaluates several transmission strategies and decides on the one that minimizes the expected future transmission cost. To estimate the expected future transmission cost for each strategy, we generate measurement trajectories by bootstrapping model residuals and estimate the future transmission costs. As discussed in Section 2.3, the bootstrap method provides a robust estimator based on resampling from the empirical distribution [93], [94], which resolves to rely on assumptions of the measurement distribution. To preserve critical stochastic properties of the empirical distribution, e.g., temporal correlation, when doing the resampling, we apply model-based bootstrap and resample the model residuals using the maximum entropy bootstrap algorithm [63], [64]. Our approach to estimating the transmission cost is inspired by pricing financial derivatives in quantitative finance, where the future pay-out is a path-dependent function of an underlying financial asset [65], [66].

The main contributions from Publication **V** can be found in Contribution **C2** in Section 1.2.

4.1.2 System Model

We consider a system of a sensor and an FC, where a DPS coordinates the sensor measurement transmission. At each time-instant, $k \in \mathbb{N}$, the sensor observes the measurement process $x[k]$, which has been predicted as $\hat{x}[k]$ using a model stored at both the sensor and the FC. The sensor only transmits a measurement if the prediction error $|x[k] - \hat{x}[k]|$ exceeds pre-defined the accuracy threshold β. The measurement registered at the FC can be expressed as

$$x^{FC}[k] = \begin{cases} \hat{x}[k], & \text{if } |x[k] - \hat{x}[k]| \leq \beta, \\ x[k], & \text{if } |x[k] - \hat{x}[k]| > \beta. \end{cases} \tag{4.1}$$

The predicted value $\hat{x}[k]$ at the FC and sensor is obtained using the prediction model given by

$$\hat{x}[k] = h(\boldsymbol{\kappa}[k], \boldsymbol{x}^{FC}[k-1]), \tag{4.2}$$

where the function $h : \mathbb{R}^q \times \mathbb{R}^p \rightarrow \mathbb{R}$ defines the model, $\boldsymbol{\kappa}[k] = [\kappa_1[k], \kappa_2[k], \ldots, \kappa_q[k]]^T \in \mathbb{R}^q$ denotes the prediction model parameter of dimension q and $\boldsymbol{x}^{FC}[k] = [x^{FC}[j_1], x^{FC}[j_2], \ldots, x^{FC}[j_p]]^T \in \mathbb{R}^p$ is the input data available at the FC for predictions, where $j_i \in \{1, 2, \ldots, k-1\}$ and p is the

Table 4.1: Transmission strategies t and cost $M_k(t)$ for current time k.

Strategy t	Transmit	Current Transmission Cost $M_k(t)$
0	-	0
1	$x[k]$	1
2	$\tilde{\boldsymbol{x}}[k]$	s
3	$\hat{\boldsymbol{\kappa}}[k]$	q
4	$\tilde{\boldsymbol{x}}[k]$ and $\hat{\boldsymbol{\kappa}}[k]$	$s+q$

number of registered measurements. Note that $\boldsymbol{x}^{FC}[k]$ can consist of both measurements and predictions according to (4.1).

From (4.1) and (4.2), we see that a measurement $x[k]$ is transmitted due to: volatility of the process $x[k]$, outdated model parameter $\boldsymbol{\kappa}[k]$, or, inaccurate predictor input $\boldsymbol{x}^{FC}[k]$. Therefore, the model parameters $\boldsymbol{\kappa}[k]$ and input variables $\boldsymbol{x}^{FC}[k-1]$ must occasionally be updated to reduce the number of future transmissions. The model parameters are continuously re-estimated at the sensor node and can be transmitted to the FC to replace the model parameters currently used by the predictor. The input variables can be updated by having the sensor transmit a set $\tilde{\boldsymbol{x}}[k]$ of s, $s \leq p$, of missing measurements, such that all the input variables for the upcoming prediction are true measurements.

Let $d[k] \in \mathcal{T}$ denote the decision by sensor at time k and $\mathcal{T} = \{0, 1, 2, 3, 4\}$ denote the set of transmission strategies available at the sensor. If the prediction error magnitude is below the accuracy tolerance level β, no transmission takes place and $d[k] = 0$. Otherwise, the sensor either decides to transmit the measurement $x[k]$, i.e., $d[k] = 1$, or to improve the prediction accuracy by transmitting a set of missing measurements $\tilde{\boldsymbol{x}}[k]$, i.e., $d[k] = 2$, or re-estimated model parameters $\hat{\boldsymbol{\kappa}}[k]$ to the FC, i.e., $d[k] = 3$ or transmit both $\tilde{\boldsymbol{x}}[k]$ and $\hat{\boldsymbol{\kappa}}[k]$, which is represented by $d[k] = 4$. The transmission strategies available at the sensor is summarized in Table 4.1.

After a sensor transmission is triggered according to (4.1), the sensor-to-FC communication protocol decides on the strategy that leads to the lowest future transmission cost. The transmission cost for the current time $M_k(t)$ that is associated with each strategy $t \in \mathcal{T}$ is summarized in Table 4.1.

4.1.3 Bootstrap-based Cost-Aware Dual Prediction Scheme

Given that a transmission is triggered at time k according to (4.1), the expected number of transmission instances in future $N_k(t)$ during time period $(k, k+T]$ is given by

$$N_k(t) = \sum_{k=k+1}^{k+T} \mathbb{E}[\mathbb{1}(|x[k] - \hat{x}[k]| > \beta) \mid d[k] = t]. \tag{4.3}$$

The expected total transmission cost for time period $[k, k+T]$ in future due to choosing strategy $d[k] = t$ at time k is given by

$$C_k(t) = M_k(t) + N_k(t),$$

where $M_k(t)$ is the current transmission cost at time k as described in Table 4.1. In a CA-DPS, the threshold strategy chosen at triggering time k, $d[k] = t^*$, is the one that minimizes the expected total transmission cost

$$t^* = \arg\min_{t \in \mathcal{T}} \; M_k(t) + N_k(t).$$

Since $M_k(t)$ is known *a priori*, we only need to determine $N_k(t)$ from (4.3).

However, evaluating (4.3) is not straightforward. Firstly, it requires knowledge of the distribution of $x[k]$ and secondly, it can be intractable to derive a close-form function given the prediction model and distribution of $x[k]$. To overcome this, we estimate $\hat{N}_k(t)$ of the number of future transmissions $N_k(t)$ using the bootstrap paradigm [60]. An introduction to bootstraping can be found in Section 2.3. At instant k, the estimate $N_k(t)$ is used to determine the decision criterion as $t^* = \arg\min_{t \in \mathcal{T}} M_k(t) + \hat{N}_k(t)$.

To estimate $N_k(t)$, we first generate L future trajectories of the measurements $[x^{(l)}[k+1], x^{(l)}[k+2], \ldots, x^{(l)}[k+T]]^{\mathrm{T}}$, $l = 1, 2, \ldots, L$, in the time interval $[k+1, k+T]$ by drawing resamples from the empirical distribution of measurement process $x[k]$. Thereafter, for each trajectory l, we simulate a DPS for the given model parameters and input variables corresponding to each strategy $t \in \mathcal{T}$ to compute the number of transmissions, which is denoted by $N_k^{(l)}(t)$. We then estimate $N_k(t)$ by taking the average of $N_k^{(l)}(t)$, i.e.,

$$\hat{N}_k(t) = \frac{1}{L} \sum_{l=1}^{L} N_k^{(l)}(t), \quad \forall t \in \mathcal{T}.$$

We use model-based bootstrap [62] to draw the resamples of $x[k]$. The future trajectories of $x[k]$ are obtained by using the estimated prediction model in (4.2),

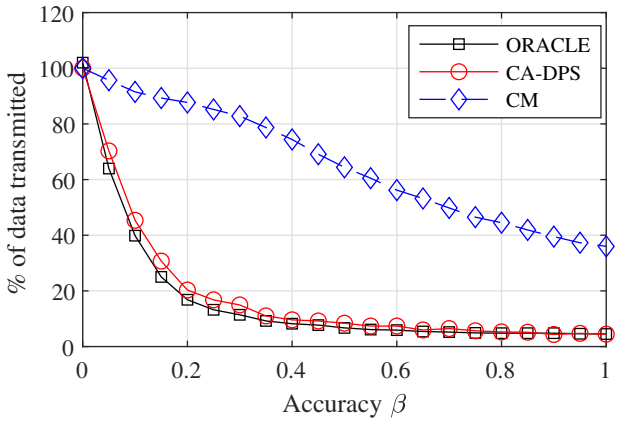

Figure 4.1: Percentage of data transmitted versus accuracy β for CM and CA-DPS with sliding window length $\eta = 10$, forecast horizon $T = 10$ and $L = 50$ trajectory simulations.

with re-estimated model parameters $\hat{\boldsymbol{\kappa}}[k]$, and adding resampled residuals from the empirical distribution of the estimated model residuals. To achieve this, we define the model residual e_i as

$$e_i = x_i - h(\hat{\boldsymbol{\kappa}}[k], \boldsymbol{x}_{i-1}),$$

where $\boldsymbol{x}_i$ is a set $\boldsymbol{x}_i = [x[j_1], x[j_2], \ldots, x[j_p]]^T \in \mathbb{R}^p$ and $j_i \in \{1, 2, \ldots, k-1\}$. Starting from measurement $x[k]$, we generate the next time step $x^{(l)}[k+1]$ of the lth simulated trajectory as

$$x^{(l)}[k+1] = h(\hat{\boldsymbol{\kappa}}[k], \boldsymbol{x}[k]) + e^{(l)}[k+1],$$

where $e^{(l)}[k+1]$ is resampled from the empirical distribution of model residuals. The model residuals are resampled using the maximum entropy bootstrap algorithm (with scale adjustment) described in [63], [64], which preserves critical statistical properties from the empirical distribution, such as temporal correlation, variance and mean. We repeat the same procedure to create a full trajectory $[x^{(l)}[k+1], x^{(l)}[k+2], \ldots, x^{(l)}[k+T]]^{\mathrm{T}}$.

4.1.4 Numerical Results

Synthetic Data – Gaussian Random Walk with Drift

The synthetic data set follows a Gaussian random walk with drift, $x[k+1] = x[k] + \kappa[k] + z[k]$, $\kappa[k] \in \mathbb{R}$, where $z[k] \sim N(0, \sigma^2)$ and $\sigma = 0.1$. The drift parameter $\kappa[k]$ gives the linear trend of the process and was simulated as constant

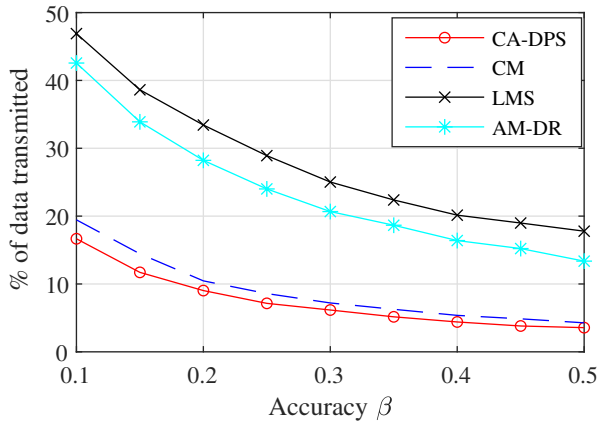

Figure 4.2: Percentage of data transmitted versus accuracy β for Mote ID 30 for CA-DPS with sliding window length $\eta = 30$, forecast horizon $T = 20$, and $L = 40$ trajectory simulations, LMS [56] with filter length $N = 4$, and step-size $\mu = 10^{-5}$ and AM-DR [33] with fixed window size $\eta_f = 4$, slow window size $\eta_s = 8$, and learning rate $\alpha = 10^{-7}$.

until it made random normal distributed jumps $\sigma_\kappa \sim N(0, 1)$ with a probability of 0.02 at each time instant k. The assigned predictor h has the form $\hat{x}_{k+n} = x[k] + n\hat{\kappa}^{FC}[k]$, $n \in \mathbb{N}_+$, where $\hat{\kappa}^{FC}[k]$ is the estimate of $\kappa[k]$ available at FC. The maximum likelihood estimate of $\kappa[k]$, computed at the sensor whenever a transmission is triggered, is given by [95]

$$\hat{\kappa}[k] = \frac{x[k] - x[k - \eta + 1]}{\eta - 1}$$

where η is the window size. Consequently, we have $\hat{\kappa}^{FC}[k] = \hat{\kappa}[k]$ for transmission strategies $t = 3$ and $t = 4$ whenever the error bound is violated, and $\hat{\kappa}^{FC}[k] = \hat{\kappa}^{FC}[k - 1]$ otherwise.

We compare the method to an oracle solution (ORACLE), where the sensor knows the distribution of $x[k]$ and transmits the model parameters every time the distribution changes. We also compare the proposed method to a constant prediction model (CM), where the prediction is the most recently transmitted measurement for which the registered measurement at the FC is [32]

$$x^{FC}[k] = \begin{cases} x^{FC}[k-1], & \text{if } |x[k] - x^{FC}[k-1]| \leq \beta, \\ x[k], & \text{if } |x[k] - x^{FC}[k-1]| > \beta. \end{cases}$$

Figure 4.1 shows the percentage of data transmitted versus accuracy β for ORACLE, CM, and CA-DPS with window size $\eta = 10$ to estimate $\kappa[k]$, $T = 10$

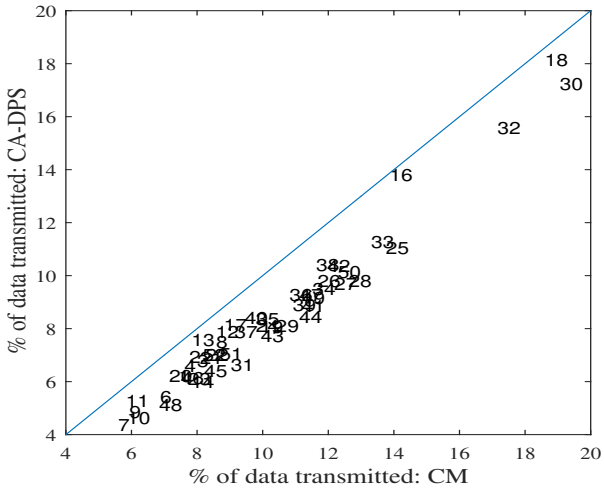

Figure 4.3: Percentage of data transmitted with accuracy set to $\beta = 0.1$ for all temperature sensors in the Intel office data set, between March 6 and 9, for CM and CA-DPS with sliding window length $\eta = 30$, forecast horizon $T = 20$ and $L = 40$ trajectory simulations. Numbers represent Mote IDs.

time steps, and $L = 50$ bootstrapped trajectories. We see that CM performs poorly since the model does not capture the drift component $\kappa[k]$ in the measurement process $x[k]$. For CA-DPS, the FC model parameter is updated in line with the ORACLE method.

Real World Data and Benchmarks

We also test CA-DPS on a real-world data set containing the readings from 54 Mica2Dot sensors in the Intel Berkeley Research lab between March 6 and 9, 2004[1]. Each sensor is identified by a Mote ID number and has humidity, temperature, light, and voltage readings every 31 seconds. This data set has repeatedly been used as test data in the literature on prediction-based data reduction [33], [56], [92], [96].

Figure 4.2 shows the percentage of data transmitted for CA-DPS and benchmark methods when applied to the temperature data of Mote ID 30, as was also done in [33], [56]. We see that CA-DPS achieves the lowest transmission rate for β between 0.1 and 0.5. At $\beta = 0.1$ it achieves a transmission rate of 17% with a mean absolute error (MAE) of 0.05 between the true and predicted readings $|x^{FC}[k] - x[k]|$.

[1] Available at http://db.lcs.mit.edu/labdata/labdata.html

Figure 4.3, shows the transmission rate for CA-DPS and CM for all 54 mote IDs, at $\beta = 0.1$. We see that CA-DPS achieves a lower transmission rate than CM for all 54 sensors, since all sensors are below the 45° line. CA-DPS was able to achieve an average transmission rate of 8%, at threshold $\beta = 0.1$, for all sensors using the settings $(\eta = 30, T = 20, L = 40)$.

4.2 Optimal Threshold and Channel Allocation

In the second part of this chapter, we consider a system of multiple sensors following event-triggered transmission schemes while sharing a limited number of collision channels. If the number of transmitting sensors exceeds the number of available communication channels, a package collision occurs, resulting in interference. Thus, we cannot guarantee a bounded error for such a system scenario. A statistic measure, the MSE, will represent the overall estimation accuracy performance metric.

Intuitively, and as shown in Figures 4.1 and 4.2, for a single-sensor system, reducing the accuracy threshold β in (4.1) will result in a lower bounded error, i.e., a higher estimation accuracy. However, it will come at an energy expanse by transmitting more frequently. For a multi-sensor system with a limited number of communication channels, the relationship between the accuracy threshold and the estimation accuracy is not straightforward. A smaller threshold across sensors leads to an increased transmission rate and a higher probability of a packet collision. Thus, minimizing the MSE with respect to the transmission thresholds becomes a function of the measurement statistics and the number of communication channels and sensors.

Following, we present the results from Publication **VI**, where we derive a method for finding optimal transmission thresholds and channel allocation for multiple sensors observing heterogeneous processes and sharing a limited number of collision channels.

4.2.1 Prior Work

Previous work in event-triggered transmission schemes has focused on analyzing how different system performance metrics depend on the accuracy threshold [32], [33], [92]. As mentioned earlier, the accuracy threshold determines transmission probability, which determines the transmission rate and energy consumption for communication. In channels constrained systems, it also influences the likelihood of a collision. Finding optimal transmission thresholds that maximize the overall estimation accuracy has been studied with consideration for energy-harvesting [76], [97], presence of an eavesdropper [98] and multiple sensors sharing a limited number of communication channels [30], [76], [99].

The threshold values are often determined by a threshold strategy, which can be either deterministic [30] or stochastic [100]–[102]. Commonly, the transmission thresholds are two-folded and deterministic, with threshold values symmetrically centered around the estimated value at the current time instant [30]–[34]. In the works [76], [79], [99], authors study jointly optimal threshold values and estimator models to minimize the average MSE across sensors.

In a *centralized* event-triggered transmission scheme [30], [76], [79], an agent observes measurements across sensors before deciding which measurements to transmit to the estimators. The alternative approach is to apply a *decentralized* transmission scheme, in which sensors cannot communicate with each other before transmitting their measurements. A centralized scheme can improve the performance compared to a decentralized scheme since it allows for defining transmission regions of multiple sensor measurements. For example, as shown in [80], for a WSN of multiple sensors observing equally distributed independent Gauss-Markov processes, an optimal strategy is always transmitting the sensor measurement furthest from the current estimate of the corresponding remote estimator. Hence, this transmission sequence would not be possible to execute in a decentralized scheme. An optimal centralized transmission policy of a similar structure holds in [30], [103], where multiple sensors observe independent measurements following symmetric measurement distributions. For the system considered in [30], an optimal centralized scheme always outperforms a decentralized optimal symmetric two-sided threshold strategy.

A centralized scheme where a cluster-head observes measurements before transmitting has implications for privacy and latency. Furthermore, it requires a specific system topology. To support the claim in [30], there is a need to find optimal transmission policies for decentralized schemes. We also see a limited number of works studying optimal threshold strategies for sensors observing heterogeneous processes.

In Publication **VI**, we consider a multi-sensor system that follows event-triggered transmission schemes and shared collision channels. The sensors are divided into spatially distributed clusters, each assigned a set of collision channels, where the measurement distributions among the different clusters are assumed heterogeneous. Due to limited energy resources, the sensors' transmission rates are restricted by a transmission constraint. The goal is to allocate communication channels and determine the optimal transmission thresholds for each cluster to maximize the estimation accuracy.

The main contributions from Publication **VI** can be fund in Contribution **C3** in Section 1.2.

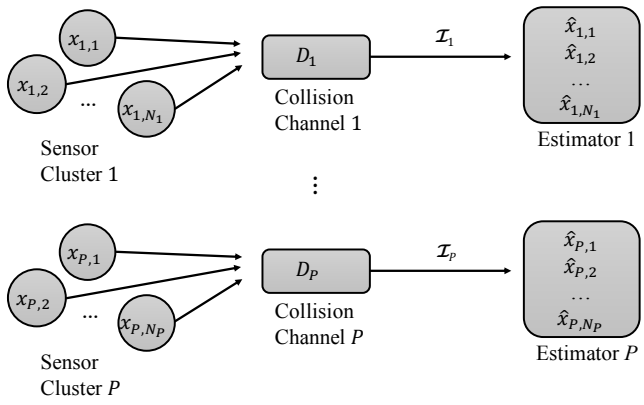

Figure 4.4: Schematic of WSN channel allocation problem.

4.2.2 System Model

We consider an IoT system comprising $N \in \mathbb{N}_+$ sensors, spatially distributed into $P \in \mathbb{N}_+$ clusters, with $N_i \in \mathbb{N}_+$ sensors in cluster i and $N = \sum_{i=1}^{P} N_i$, as shown in Figure 4.4. Sensors in cluster i, $i = 1, ..., N_i$ observe independent Gaussian processes $x_{i,j} \sim \mathcal{N}(0, \sigma_i^2)$, where $i = 1, ..., P$ and $j = 1, ..., N_i$. The measurement distributions among the different clusters are assumed heterogeneous, i.e., $\sigma_i^2 \neq \sigma_j^2$, $\forall i, j = 1, ..., P$.

The sensors in cluster i, $i = 1, ..., P$ transmit their measurements via an assigned communication channel i, with $D_i \in \mathbb{N}_+$ available channels, to a corresponding remote estimator i. An event-triggered transmission scheme is set up such that a sensor only transmits if the measurement value falls outside a predefined interval $x_{i,j} \notin (\alpha_i, \beta_i)$, where $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}$ are the transmission thresholds.

The threshold values α_i and β_i depend on the *threshold strategy* for cluster i, defined by the two parameters; $p_i \in [0, 1]$, representing the transmission rate, which is equivalent to the transmission probability $p_i = P(x_{i,j} \notin (\alpha_i, \beta_i))$, and $s_i \in \{1, 2\}$, indicating either a single- or double-bounded threshold strategy, i.e.,

$$(\alpha_i, \beta_i) = \begin{cases} (-\infty,\, Q_i\,(p_i))\,, & \text{if } s_i = 1, \\ \left(-Q_i(\dfrac{p_i}{2}),\, Q_i(\dfrac{p_i}{2})\right) & \text{if } s_i = 2, \end{cases} \tag{4.4}$$

where $Q_i(p_i) = \inf\{x \in \mathbb{R} | 1 - p_i \leq \Phi(x/\sigma_i)\}$ represents the $(1 - p_i)$-percentile for Gaussian distribution $x \sim \mathcal{N}(0, 1)$, with cumulative distribution function, Φ.

Similar to the system in [30], we assume that if more than $D_i \in \mathbb{N}_{++}$ sensors transmit through the channel allocated to cluster i, a packet collision occurs. In that

case, the ith estimator does not receive the measurements but is able to decode the sensor indices of the transmitting sensors [30]. Let $\mathcal{Y}_i$ be the set of indices of the transmitting sensors from cluster i, i.e., $\mathcal{Y}_i = \{j \in \{1, ..., D_i\} | \, x_{i,j} \notin (\alpha_i, \beta_i)\}$. The output $\mathcal{I}_i$ from collision-channel i to estimator i is

$$\mathcal{I}_i = \begin{cases} \varnothing, & \text{if } |\mathcal{Y}_i| = 0, \\ \{j, x_{i,j} | j \in \mathcal{Y}_i\}, & \text{if } 1 \leq |\mathcal{Y}_i| \leq D_i, \\ \{\mathfrak{C}, \mathcal{Y}_i\}, & \text{if } |\mathcal{Y}_i| \geq D_i, \end{cases} \tag{4.5}$$

where $\mathfrak{C}$ is a collision indicator symbol. From (4.5), the estimate at the remote estimators $\hat{x}_{i,j} = \mathbb{E}[x_{i,j}|\mathcal{I}_i]$ equals

$$\hat{x}_{i,j} = \begin{cases} \mathbb{E}[x_{i,j} | x_{i,j} \in (\alpha_i, \beta_i)], & \text{if } j \notin \mathcal{Y}_i, \\ x_{i,j}, & \text{if } j \in \mathcal{Y}_i, \, |\mathcal{Y}_i| \leq D_i, \\ \mathbb{E}[x_{i,j} | x_{i,j} \notin (\alpha_i, \beta_i)], & \text{if } j \in \mathcal{Y}_i, \, |\mathcal{Y}_i| > D_i. \end{cases} \tag{4.6}$$

The mean squared error (MSE) at the ith estimator, as shown in (4.6), depends on the threshold strategy s_i, the transmission rate p_i and the channel allocation D_i, is given by

$$J_i(s_i, p_i, D_i) = \sum_{j=1}^{N_i} \mathbb{E}\Big[(x_{i,j} - \hat{x}_{i,j})^2\Big]. \tag{4.7}$$

Our objective is to find an optimal joint transmission and channel allocation strategy $\boldsymbol{s} = [s_1, s_2, ..., .s_P]^{\mathrm{T}}$, $\boldsymbol{p} = [p_1, p_2, ..., .p_P]^{\mathrm{T}}$ and $\boldsymbol{D} = [D_1, D_2, ..., .D_P]^{\mathrm{T}}$ that minimizes the total MSE across all sensors, i.e.,

$$\begin{aligned} \min_{\boldsymbol{s},\boldsymbol{p},\boldsymbol{D}} \; & \sum_{i=1}^{P} J_i(s_i, p_i, D_i) \\ \text{s.t.} & \sum_{i=1}^{P} D_i = D, \quad D_i \geq 1, \\ & 0 \leq p_i \leq \bar{p}, \quad i = 1, ..., P, \end{aligned} \tag{4.8}$$

where D is the total number of available channels for the system and $\bar{p} \in \mathbb{R}_+$, $0 < \bar{p} \leq 1$, is a transmission probability constraint to account for the sensors limited energy resources.

4.2.3 Theoretical Results

To solve (4.8), we need to calculate the cost in (4.7) based on s_i, p_i and D_i. We begin this section by presenting a closed-form expression of the cost function

$J_i(s_i, p_i, D_i)$. We then present theoretical results of the optimal performance for each individual cluster i, $J_i(s_i^*, p_i^*, D_i^*)$. Finally, we present an algorithm of how to step-wise derive $\{\boldsymbol{s}^*, \boldsymbol{p}^*, \boldsymbol{N}^*\}$.

The N_i observations are i.i.d., and the distribution for the number of transmissions $|\mathcal{Y}_i|$ for cluster i follows a binomial distribution, i.e., $|\mathcal{Y}_i| \sim Bin(N_i, p_i)$. From the law of total expectation, we can express (4.7) as the probability weighted MSE given the number of transmitted sensors, i.e.,

$$
\begin{aligned}
&J_i(s_i, p_i, D_i) = \\
&\sum_{y=0}^{N_i} P(|\mathcal{Y}_i| = y) \Bigg((N_i - y)\mathbb{E}\Big[(x_{i,1} - \hat{x}_{i,1})^2 | x_{i,1} \in (\alpha_i, \beta_i)\Big] \\
&+ y\mathbb{1}(y > D_i)\mathbb{E}\Big[(x_{i,1} - \hat{x}_{i,1})^2 | x_{i,1} \notin (\alpha_i, \beta_i)\Big] \Bigg) \\
&= \underbrace{\mathbb{E}\Big[(x_{i,1} - \hat{x}_{i,1})^2 | x_{i,1} \in (\alpha_i, \beta_i)\Big]}_{=A_i(s_i,p_i)} \underbrace{\Big(N_i - \mathbb{E}[|y|]\Big)}_{=\omega_i(p_i)} + \\
&\underbrace{\mathbb{E}\Big[(x_{i,1} - \hat{x}_{i,1})^2 | x_{i,1} \notin (\alpha_i, \beta_i)\Big]}_{=B_i(s_i,p_i)} \underbrace{\mathbb{E}\Big[|y|\mathbb{1}(|y| > D_i)\Big]}_{=\tilde{\omega}_i(p_i,D_i)},
\end{aligned}
\tag{4.9}
$$

where $\mathbb{E}[|y|] = N_i p_i$ and $\mathbb{1}(\cdot)$ is an indicator function having value 1 if the condition in the argument is true and 0 otherwise. Given that all sensors in Cluster i observe i.i.d. distributions, the variable $x_{i,1}$ in (4.9), represents the measurement process from an arbitrary sensor in Cluster i.

As seen in (4.9), the cost can be expressed in the form

$$
J_i(s_i, p_i, D_i) = A_i(s_i, p_i)\omega_i(p_i) + B_i(s_i, p_i)\tilde{\omega}_i(p_i, D_i),
$$

where the functions $A_i(s_i, p_i)$ and $B_i(s_i, p_i)$, respectively, represent the conditional MSE for a non-transmitted measurement, and a collided measurement, and the functions $\omega_i(p_i)$ and $\tilde{\omega}_i(p_i, D_i)$, respectively, represent the expected number of non-transmitted measurements and collided measurements from cluster i. The

values of $A_i(s_i, p_i)$ and $B_i(s_i, p_i)$ can be calculated [104], [105] as

$$A_i(1,p_i) = \sigma_i^2 \left(1 - \frac{\beta_i}{\sigma_i}\frac{\phi(\beta_i/\sigma_i)}{(1-p_i)} - \left(\frac{\phi(\beta_i/\sigma_i)}{1-p_i}\right)^2\right),$$
$$B_i(1,p_i) = \sigma_i^2 \left(1 + \frac{\beta_i}{\sigma_i}\frac{\phi(\beta_i/\sigma_i)}{p_i} - \left(\frac{\phi(\beta_i/\sigma_i)}{p_i}\right)^2\right),$$
$$A_i(2,p_i) = \sigma_i^2 \left(1 - \frac{2\beta_i\phi(\beta_i/\sigma_i)}{\sigma_i(1-p_i)}\right),$$
$$B_i(2,p_i) = \sigma_i^2 \left(1 + \frac{2\beta_i\phi(\beta_i/\sigma_i)}{\sigma_i p_i}\right),$$

where ϕ is the pdf of Gaussian $\mathcal{N}(0,1)$.

An important property to acknowledge, is how the conditional MSE of collided measurements, $B_i(s_i, p_i)$, depends on the threshold strategy s_i. As seen in (4.4)-(4.6), for $s_1 = 1$, the estimator always has information of the region where a measurement lies, i.e., if $x_{i,j} \leq \alpha_i$ or $x_{i,j} \geq \beta_i$, regardless if a collision has occurred or not. From (4.6), for $s_i = 1$, the estimate of a collided measurement, $\hat{x}_{i,j} = \mathbb{E}[x_{i,j}|x_{i,j} \notin (\alpha_i, \beta_i)]$, depends on the transmission rate p_i, while for $s_i = 2$, the estimate is independent of p_i and always equals the distribution mean, i.e., $\hat{x}_{i,j} = 0$, $\forall p \in [0,1]$. For this reason, the conditional MSE for a collided measurement in (4.9) is smaller for $s_i = 1$ than for $s_i = 2$, and results in the inequality

$$B_i(1,p_i) \leq \sigma_i^2 \leq B_i(2,p_i), \quad 0 \leq p_i \leq 1.$$

Another property to acknowledge, is that the cost decreases with respect to D_i, i.e,

$$J_i(s_i, p_i, D_i + \epsilon) \leq J_i(s_i, p_i, D_i), \quad \epsilon \in \mathbb{N}_+, \tag{4.10}$$

which allows us to state the following lemma.

Lemma 4.1. *For $\bar{p} = 1$, if the following inequality holds*

$$\min_{p_i \in [0,1]} J_i(1, p_i, 0) \leq \min_{p_i \in [0,1]} J_i(2, p_i, \bar{D}),$$

then a single-sided strategy, $s_i = 1$, is optimal for any number of channels being less than $\bar{D}$, i.e.,

$$\min_{p_i \in [0,1]} J_i(1, p_i, D_i) \leq \min_{p_i \in [0,1]} J_i(2, p_i, \bar{D}), \quad \forall D_i \leq \bar{D}. \tag{4.11}$$

Proof. Lemma 4.1 follows from property (4.10). □

Theoretically, Lemma 4.1 implies that for a system with no transmission constraint, i.e., $\bar{p} = 1$, if a one-sided strategy using zero channels outperform a two-sided strategy using $\bar{D}_i$ available channels, $D_i = 0$, a one-sided strategy is then consistently optimal for all number of channels D_i equal or less than $\bar{D}_i$.

In the following theorem, we present an upper bound of the optimal performance for any given number of channels $D_i \leq D$. Before presenting the theorem, we introduce $J_i^*(D_i)$, $i = 1, ..., P$ as the minimum cost for Cluster i, if assigned D_i channels, being defined as

$$J_i^*(D_i) = \min_{s_i \in \{1,2\}, p_i \leq \bar{p}} J_i(s_i, p_i, D_i). \tag{4.12}$$

Theorem 4.1. *For $\bar{p} = 1$, the optimal performance for any threshold strategy using D_i channels is upper bounded by*

$$J_i^*(D_i) \leq N_i A_i(1, 0.5), \tag{4.13}$$

and, if the following inequality holds,

$$N_i A_i(1, 0.5) \leq \min_{p_i \in [0,1]} J_i(2, p_i, \bar{D}), \tag{4.14}$$

then (4.11) holds, and a single-sided strategy, $s_i = 1$, is optimal for all $D_i \leq \bar{D}$.

Proof. The proof can be found in Publication **VI**. □

Theorem 4.1 presents an upper bound of the optimal performance for any given number of channels $D_i \leq D$. An additional lower performance boundary for an optimal double-sided strategy $s_i = 2$ can be found in [30], where a cluster-head observes all measurements from Cluster i before transmitting the D_i largest measurements. Let $\tilde{J}_i(D_i)$ denote the MSE for the centralized transmission scheme, being defined as

$$\tilde{J}_i(D_i) = \sum_{n=D_i+1}^{N_i} \mathbb{E}\big[x_{i,(n)}^2\big], \tag{4.18}$$

where $x_{i,(n)}$ is the nth largest value in cluster i, i.e., $\{|x_{i,j}|\}_{j=1}^{N_i}$. Hence, for $s_i = 2$, the cost is lower-bounded by

$$\tilde{J}_i(D_i) \leq \min_{p_i \in [0,1]} J_i(2, p_i, D_i). \tag{4.19}$$

Remark 4.1. For the centralized scheme presented in [30], the output $\mathcal{I}_i$ from the cluster-head to the remote estimator only includes the D_i largest measurement values in cluster i, i.e., $\mathcal{I}_i = \{|x_{i,j}|\}_{j=1}^{D_i}$, and does not include information of measurements breaching any threshold values α_i and β_i.

Algorithm 3 Deriving $\{\boldsymbol{s}^*, \boldsymbol{p}^*, \boldsymbol{D}^*\}$ step-wise

1: For each cluster i, $i = 1, ..., P$, and for all possible number of channels, $\forall D_i \leq D$, solve the two problems

$$\min_{p_i \leq \bar{p}} J_i(1, p_i, D_i), \text{ and } \min_{p_i \leq \bar{p}} J_i(2, p_i, D_i). \tag{4.15}$$

2: From step 1, calculate $J_i^*(D_i)$ in (4.12), for all i, $i = 1, ..., P$, and D_i, $D_i \leq D$.

3: From step 2, find $\boldsymbol{D}^* = (D_1^*, D_2^*, ..., D_P^*)$

$$\begin{aligned} &\min_{\boldsymbol{D}} \sum_{i=1}^{P} J_i^*(D_i), \\ &\text{s.t.} \sum_{i=1}^{P} D_i = D, \end{aligned} \tag{4.16}$$

4: Compute the corresponding parameters $\{\boldsymbol{s}^*, \boldsymbol{p}^*\}$ as

$$(s_i^*, p_i^*) = \arg\min_{s_i, p_i \leq \bar{p}} J_i(s_i, p_i, D_i^*). \tag{4.17}$$

The function $J_i(s_i, p_i, D_i)$ is a continuous and well-behaving function, which prove the existence of an optimal set $\{\boldsymbol{s}^*, \boldsymbol{p}^*, \boldsymbol{D}^*\}$ that can be derived by following the steps presented in Algorithm 3.

In the next section, we compare the performance to have all clusters sharing collision channels. First, we present a method for deriving an optimal parameter set for such a system scenario. Let $\mathcal{Y} = (|\mathcal{Y}_1|, |\mathcal{Y}_2|, ..., |\mathcal{Y}_P|)$, $\mathcal{Y} \in \mathbb{N}_+^P$, be a stochastic variable representing the number of transmitted measurements from all P clusters. Since, $|\mathcal{Y}_i| \sim Bin(N_i, p_i)$, the variable $\mathcal{Y}$ follows a multinomial distribution. From the law of total expectation, let $f(\mathcal{Y}, \boldsymbol{s}, \boldsymbol{p})$ represent the total MSE

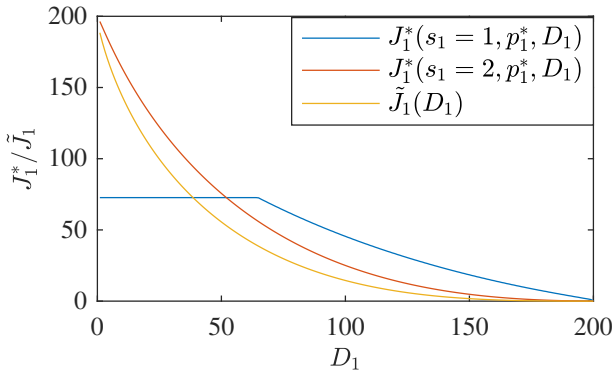

Figure 4.5: Optimal MSE J_1^* and $\tilde{J}_1$ for $P = 1$, $N_1 = 200$ sensors with $\sigma_1 = 1$ and $\bar{p} = 1$.

given outcome $\mathcal{Y}$ and system setting $\{\boldsymbol{s}, \boldsymbol{p}\}$, $\boldsymbol{s} \in \{1, 2\}^P$, $\boldsymbol{p} \in [0, \bar{p}]^P$,

$$\begin{aligned}\hat{J}_i(\boldsymbol{s}, \boldsymbol{p}) &= \sum_{j=1}^{N_i} \mathbb{E}\Big[(x_{i,j} - \hat{x}_{i,j})^2\Big] \\ &= \sum_{\mathcal{Y}' \in \boldsymbol{\mathcal{Y}}} P(\mathcal{Y} = \mathcal{Y}')\Big((N_i - |\mathcal{Y}_i'|)A_i(s_i, p_i) \\ &+ \mathbb{1}(\sum_{j=1}^{P} |\mathcal{Y}_j'| > D) \sum_{i=0}^{P} |\mathcal{Y}_i'| B_i(s_i, p_i)\Big),\end{aligned}$$

where $\boldsymbol{\mathcal{Y}}$ is the set of all possible outcomes of $\mathcal{Y}$.

An optimal pair $\{\boldsymbol{s}^*, \boldsymbol{p}^*\}$, is obtained by solving

$$\begin{aligned}&\min_{\boldsymbol{s},\boldsymbol{p}} \sum_{i=1}^{P} \hat{J}_i(\boldsymbol{s}, \boldsymbol{p}) \\ &\text{s.t. } 0 \leq p_i \leq \bar{p}, \quad i = 1, ..., P.\end{aligned} \tag{4.20}$$

We numerically derive approximated values of $\{\boldsymbol{s}^*, \boldsymbol{p}^*\}$ using a greedy approach, where we first discretize the range of $\boldsymbol{p}$, $\boldsymbol{p} \in [0, \bar{p}]$, into $M \in \mathbb{N}_+$ discrete values, i.e., $\boldsymbol{\varrho} = \{0, \varrho, 2\varrho, ..., M\varrho\}$, $\varrho = \bar{p}/M$, and then compare the performance for all possible combinations of $\boldsymbol{s}, \boldsymbol{p} \in \{1, 2\}^P \times \boldsymbol{\varrho}$, to find which pair minimizes (4.20).

4.2.4 Numerical Results

We begin by considering a single-cluster system, i.e., $P = 1$, with size $N_1 = 200$, variance $\sigma_1 = 1$ and without any transmission probability constrain $\bar{p} = 1$. Figure

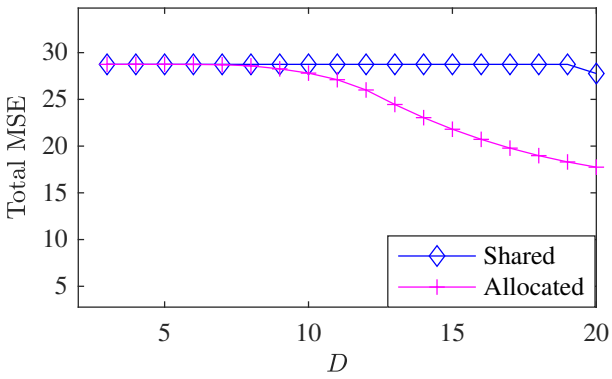

Figure 4.6: Optimal MSE vs total number of channels D for $P = 3$, $N_i = 20$, $i = 1, 2, 3$, with $\sigma_1 = 1$, $\sigma_2 = 1.2$, $\sigma_3 = 1.4$ and $\bar{p} = 0.5$.

4.5 shows $J_i(1, p_i, D_i)$ in (4.7) minimized with respect to $s_1 = 1$ and $s_1 = 2$, denoted as J_1^*, and $\tilde{J}_1^*$ from (4.18) vs D_1. As stated in (4.19), a centralized strategy outperforms a decentralized double-sided threshold strategy $s_1 = 2$, for all $D_1 \leq 200$. However, a single-sided strategy $s_1 = 1$ outperforms a centralized strategy for all $D_i \leq 38$. As presented in Proposition 4.1 and Theorem 4.1 a single-sided threshold strategy $s_1 = 1$ is consistently optimal for $D_1 \leq 52$.

We now consider a system of $P = 3$, with respective cluster sizes of $N_i = 20$, $i = 1, 2, 3$, with $\sigma_1 = 1$, $\sigma_2 = 1.2$, $\sigma_3 = 1.4$ and $\bar{p} = 0.5$. Figure 4.6 displays an optimal performance for the two approaches; allocating channels across cluster as in (4.8) or having all clusters sharing all channels as in (4.20), vs total number of channels D. For the case of sharing channels, the values $\{\boldsymbol{s}^*, \boldsymbol{p}^*\}$ are obtained using $M = 50$ discretization points. As can be seen, allocating channels across clusters outperforms having all clusters sharing channels.

4.3 Discussion

This chapter has studied threshold-based event-triggered transmission schemes for both a single- and a multi-sensor system. We discussed some implications of sensors observing non-stationary measurement distributions. Also, if the number of sensors exceeded the number of communication channels. We developed methods to address these issues and derived optimal system settings.

Firstly, to maintain a high prediction accuracy in a DPS, the sensors must transmit re-estimated model parameters to account for changes in the measurement distribution and occasionally transmit historical measurements to re-fresh the input data for future predictions. In previous literature, updating protocols are ad hoc,

where an update is triggered at every or a given number of threshold breaches. We presented our work in Publication **V**, in which we introduced a CA-DPS, which is an extension of the DPS framework, where the sensor evaluated the transmission cost among a set of transmission strategies. Using the empirical measurement distribution, we applied model-based bootstrapping to assess the different transmission strategies to avoid deriving closed-form estimators of the future transmission cost based on parametric model assumptions. Our numerical results showed that by applying a CA-DPS on both synthetic and real-world data, the amount of data transmitted between IoT sensor nodes and a fusion center could be further reduced. An essential aspect of our work is that we exploited the available computational resources at the sensor to evaluate the transmission strategies. We allowed for the decisions making of the transmission strategies to take place at the sensor and not rely on a fusion center.

Later on in the chapter, we presented our work in Publication **VI**, in which we presented a method to derive optimal transmission thresholds for multiple sensors sharing a limited number of communication channels. Each time a collision occurs, an estimator always has knowledge of the region a measurement lies in for a single-sided strategy, which results in a lower MSE than if a collision occurred using a double two-sided strategy. Analytical results showed that a single-sided threshold strategy could outperform a double-sided strategy depending on the system parameters; the number of channels, sensors, and statistical parameters. A system engineer can use our theoretical results to set necessary energy resources and the number of channels for a required MSE. Numerical results validated the analytical results and showed that a single-sided strategy could outperform a centralized transmission scheme, where a cluster-head can observe all measurements before deciding to transmit the ones that deviated the largest from the current estimate [30].

Chapter 5

Conclusions and Future Work

In a WSN or WNCS, sensors transmit measurements to remote estimators or a centralized FC that tracks physical processes to support decision-making and maximize the application utility. The information update rate from sensors is limited by the sensor node's energy resources and the number of available communication channels. This thesis has developed transmission schemes to support the data collection infrastructure in resource-constrained WSN and pro-long sensor nodes' battery lifetime. The two considered approaches have been to reduce the number of transmissions and the amount of data from sensors and, secondly, to schedule sensors in an optimal order to maximize estimation accuracy. The design of the transmission schemes has been based on exploiting statistical patterns in the measurement data, e.g., temporal and spatial correlation. We considered a simple WSN topology where sensors transmit measurements directly to a remote estimator or a fusion center via a communication medium to generalize our results to other WSN topologies. The theoretical results from this thesis can be of value for a system architect designing a WSN, e.g., choice of transmission scheme and determining technical requirements.

Chapter 3 presented our work in AoI-based scheduling of dependent measurements. We considered a channel-constrained WSN where sensors observed spatio-temporally dependent measurements that were broadcasted to remote estimators via a network manager. At each instant, the number of broadcasted measurements was limited by the number of available channels. Therefore, we derived an optimal AoI-based scheduling policy that minimized the overall time average MSE by modeling the scheduling problem as a finite-state MDP. We showed that an optimal policy resulted in a periodic scheduling sequence. This result has practical benefits as it could be easily executed and save data storage at the network sched-

uler. Later on, in Chapter 3, we considered the particular case for a system of two sensors, where the sensor's individual energy resources constrained the number of measurement transmissions. We presented a low-complexity numerical method to derive an optimal AoI-based scheduling policy for a finite horizon.

The AoI has previously been studied with consideration for queuing [35] and stochastic arrival processes, e.g., due to packet losses [5], [51]. In our system model, we consider parallel communication channels that solve the issue of packet collisions and assume that packet losses are addressed by retransmission through higher layers of the communication protocol, e.g., an ARQ protocol. As an extension of our work, one could design AoI-based scheduling policies for multiple-access channels instead of parallel access channels. However, this type of study requires incorporating queuing processes and random arrival times in the system model. Consequently, the mathematical modeling would need to be re-designed concerning the particular queuing system and the probability distribution of the arrival time.

In our work, we have chosen a simple estimation model in the sense that the estimate is only based on the most recently scheduled measurements from each sensor. We adopt this model since it: i) renders the problem mathematically tractable, ii) offers insights into the properties of an optimal scheduling policy for spatio-temporally correlated processes; iii) helps in determining the necessary conditions that an optimal scheduling policy must satisfy, and; iv) allows the derivation of an optimal policy numerically through a manageable dimensionality of the MDP finite state-space. The estimation model can be extended by including previous measurements from the sensors to improve the estimation accuracy. However, there is a trade-off between the simplicity of the estimation model and the computational complexity of deriving an optimal policy. If more recent measurements are included, the state variable must include the AoI for all the used measurements and the absolute AoI differences between them. For this reason, the finite state-space increases substantially for each additional measurement stored at the estimators. Even for a small network, storing only the two most recent measurements can make the numerical calculations prohibitive. For future work, a sequential estimator incorporating past measurements can be considered to derive the optimal scheduling policies for estimating spatio-temporal processes. We cannot conclude that the proposed mathematical framework and the ensuing results would hold for this scenario.

Optimal scheduling policies for multiple sensors that observe independent discrete-time linear time-invariant (LTI) processes and share a single communication channel have earlier been derived in [28], [29]. To the best of our knowledge, optimal scheduling policies of multiple sensors observing multiple spatio-

temporally dependent Gauss-Markov processes have not earlier been presented. We believe that this still is an open problem to derive necessary conditions for the existence of an optimal policy for processes that follows a spatio-temporally correlated Gauss-Markov model.

Chapter 4 focused on event-triggered transmission schemes, specifically, threshold-based transmission schemes. In the first part of this chapter, we considered a system of one sensor and an FC, where a DPS was implemented to save energy at the sensor by reducing the number of transmitted measurements to the FC. In practice, sensors often observe non-stationary and dynamic processes, which requires the prediction model in a DPS to be updated occasionally to maintain the prediction accuracy. A re-estimated prediction model and new input variables must be transmitted from the sensor to the FC when the measurement distribution changes. We discussed the implications of using an ad-hoc event-triggered updating protocol and introduced a CA-DPS. In a CA-DPS, the sensor evaluates a set of transmission strategies at each transmission instance and decides on the strategy that results in the lowest expected future transmission cost. The transmission cost for each strategy was estimated using model-based bootstrapping. By exploiting the computational resources at the sensor, we showed that energy efficiency for communication could be further improved by protocols that are selective in the data transmitted.

The CA-DPS framework could be extended by allowing the sensor to evaluate several prediction models among a set of prediction models to transmit. e.g., changing from an AR(2)-model to an AR(3)-model. Another extension could be to have the sensor decide on the transmission strategy that results in the largest expected number of transmissions until a threshold breach occurs, i.e., the stopping time [106]. Similarly, the expected stopping time can be estimated using model-based bootstrapping from the empirical measurement distribution.

In the second part of Chapter 4, we stated the fact that event-triggered transmission can result in packet collision if the number of simultaneous transmissions exceeds the number of communication channels. We considered a system of multiple sensors following event-triggered transmission schemes, sharing a limited number of collision-channels. The sensors were placed into spatial clusters, observing heterogeneous measurement distributions. We presented a method to derive optimal transmission threshold strategies and channel allocation for each cluster. We considered a single-sided threshold strategy and demonstrated that it could outperform the centralized scheme presented in [30], where a cluster-head observes all measurements before deciding which to transmit. We suggested exploiting spatio-temporal dependencies for future work to improve the estimation accuracy of collided measurements that the remote estimators do not observe. Further extensions

would be to incorporate measurement noise in the system model. However, this would require a different estimator model and additional mathematical derivations.

In summary, we have studied both time-triggered, i.e., scheduling sensors, and event-triggered transmission schemes, which both have their advantages. By scheduling sensors, we can coordinate the transmission of the sensor to avoid collisions that result in packet drops. Scheduling also makes it easy to plan how long available resources for communication will last. Suppose an estimation model of the physical process provides a good fit in a threshold-based event-triggered transmission scheme. In that case, it can result in a low transmission rate, saving energy for communication. For future work, a third option would be studying hybrid transmission schemes, where sensors are scheduled time-slots to avoid collisions but are not obliged to transmit if the measurement lies within some region.

Bibliography

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications", *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] A. Kirimtat, O. Krejcar, A. Kertesz and M. F. Tasgetiren, "Future trends and current state of smart city concepts: A survey", *IEEE Access*, vol. 8, pp. 86 448–86 467, 2020.

[3] B. Bajic, A. Rikalovic, N. Suzic and V. Piuri, "Industry 4.0 implementation challenges and opportunities: A managerial perspective", *IEEE Systems Journal*, vol. 15, no. 1, pp. 546–559, 2021.

[4] F. A. Kraemer, A. E. Braten, N. Tamkittikhun and D. Palma, "Fog computing in healthcare–a review and discussion", *IEEE Access*, vol. 5, pp. 9206–9222, 2017.

[5] W. Liu, K. Huang, D. E. Quevedo, B. Vucetic and Y. Li, *Deep reinforcement learning for wireless scheduling in distributed networked control*, ArXiv Preprint, 2021. arXiv: `2109.12562 [eess.SY]`. [Online]. Available: `https://arxiv.org/abs/2109.12562`.

[6] F. Wu, H. Zhang, J. Wu, Z. Han, H. Vincent Poor and L. Song, "UAV-to-device underlay communications: Age of information minimization by multi-agent deep reinforcement learning", *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4461–4475, 2021.

[7] F. A. Kraemer, D. Palma, A. E. Braten and D. Ammar, "Operationalizing Solar Energy Predictions for Sustainable, Autonomous IoT Device Management", *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11 803–11 814, 2020.

[8] A. Murad, K. Bach, F. A. Kraemer and G. Taylor, “Iot sensor gym: Training autonomous iot devices with deep reinforcement learning”, in *Proceedings of the 9th International Conference on the Internet of Things*, 2019.

[9] M. Srbinovska, V. Dimcev and C. Gavrovski, “Energy consumption estimation of wireless sensor networks in greenhouse crop production”, in *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, 2017, pp. 870–875.

[10] M. A. Alsheikh, S. Lin, D. Niyato and H. P. Tan, “Machine learning in wireless sensor networks: Algorithms, strategies, and applications”, *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.

[11] N. Chae, H. Yang, B. Y. Lee and C. Lee, “Measurement of environmental parameters in polar regions based on a ubiquitous sensor network”, *Cold Regions Science and Technology*, vol. 123, pp. 22–31, 2016.

[12] A. K. Teshome, B. Kibret and D. T. H. Lai, “A review of implant communication technology in wban: Progress and challenges”, *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 88–99, 2019.

[13] X. Ma, Z. Dong and Y. Dong, “Toward asphalt pavement health monitoring with built-in sensors: A novel application to real-time modulus evaluation”, *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.

[14] G. V. Iordache, G. Suciu, S. Segarceanu, G. Petrescu, R.-I. Vatasoiu, L. Rodriguez, J. M. Strout, P. Vergara, S.-E. Calescu, E. Nygard, C. P. Hijón and G. Langaas, “Mews - an iot and cloud-based avalanche detection and prediction platform”, in *2021 20th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, 2021, pp. 1–5.

[15] S. Misra, P. K. Bishoyi and S. Sarkar, “I-mac: In-body sensor mac in wireless body area networks for healthcare iot”, *IEEE Systems Journal*, vol. 15, no. 3, pp. 4413–4420, 2021.

[16] R. D. Yates and S. K. Kaul, “Status updates over unreliable multiaccess channels”, in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 331–335.

[17] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem and E. Hossain, “Hybrid automatic repeat request (harq) in wireless communications systems and standards: A contemporary survey”, *IEEE Communications Surveys Tutorials*, vol. 23, no. 4, pp. 2711–2752, 2021.

[18] F. A. Aderohunmu, G. Paci, D. Brunelli, J. D. Deng and L. Benini, "Prolonging the lifetime of wireless sensor networks using light-weight forecasting algorithms", *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 461–466, 2013.

[19] J. Hribar, M. Costa, N. Kaminski and L. A. Dasilva, "Using correlated information to extend device lifetime", *IEEE Internet Things*, vol. 6, no. 2, pp. 2439–2448, 2019.

[20] Z. Guo, J. Peng, W. Xu, W. Liang, W. Wu, Z. Xu, B. Guo and Y. Ivan Wu, "Minimizing redundant sensing data transmissions in energy-harvesting sensor networks via exploring spatial data correlations", *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 512–527, 2021.

[21] H. Harb, A. Makhoul, S. Tawbi and R. Couturier, "Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks", *IEEE Access*, vol. 5, pp. 4250–4263, 2017.

[22] M. Abu Alsheikh, S. Lin, D. Niyato and H. P. Tan, "Rate-distortion balanced data compression for wireless sensor networks", *IEEE Sensors Journal*, vol. 16, no. 12, pp. 5072–5083, 2016.

[23] Z. Chen, J. Ranieri, R. Zhang and M. Vetterli, "DASS: Distributed adaptive sparse sensing", *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2571–2583, 2015.

[24] M. Mathew and N. Weng, "Quality of information and energy efficiency optimization for sensor networks via adaptive sensing and transmitting", *IEEE Sensors Journal*, vol. 14, no. 2, pp. 341–348, 2014.

[25] D. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[26] Y. Zhou, Y. Fang and Y. Zhang, "Securing wireless sensor networks: A survey", *IEEE Communications Surveys Tutorials*, vol. 10, no. 3, pp. 6–28, 2008.

[27] M. P. Vitus, W. Zhang, A. Abate, J. Hu and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems", *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[28] L. Shi and H. Zhang, "Scheduling two gauss-markov systems: An optimal solution for remote state estimation under bandwidth constraint", *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2038–2042, 2012.

[29] D. Han, J. Wu, H. Zhang and L. Shi, "Optimal sensor scheduling for multiple linear dynamical systems", *Automatica*, vol. 75, pp. 260 –270, 2017.

[30] X. Zhang, M. M. Vasconcelos, W. Cui and U. Mitra, "Distributed remote estimation over the collision channel with and without local communication", *IEEE Trans. Control Netw. Syst.*, vol. 5870, no. c, pp. 1–1, 2021.

[31] M. Xia, V. Gupta and P. J. Antsaklis, "Networked state estimation over a shared communication medium", *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1729–1741, 2017.

[32] S. Werner and J. Lundén, "Smart load tracking and reporting for real-time metering in electric power grids", *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1723–1731, 2016.

[33] Y. Fathy and P. Barnaghi, "Quality-based and energy-efficient data communication for the internet of things networks", *IEEE Internet of Things J.*, vol. 6, no. 6, pp. 10 318–10 331, 2019.

[34] L. He, J. Chen and Y. Qi, "Event-based state estimation: Optimal algorithm with generalized closed skew normal distribution", *IEEE Trans. Autom. Control*, vol. 64, pp. 321–328, 2019.

[35] A. Kosta, N. Pappas and V. Angelakis, "Age of information: A new concept, metric, and tool", *Foundations and Trends® in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[36] S. Kaul, R. Yates and M. Gruteser, "Real-time status: How often should one update?", in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2731–2735.

[37] M. A. Abd-Elmagid, N. Pappas and H. S. Dhillon, "On the role of age of information in the internet of things", *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, 2019.

[38] H. Tang, J. Wang, L. Song and J. Song, "Minimizing age of information with power constraints: Multi-user opportunistic scheduling in multi-state time-varying channels", *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 854–868, 2020.

[39] T. Park, W. Saad and B. Zhou, "Centralized and distributed age of information minimization with nonlinear aging functions in the internet of things", *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8437–8455, 2021.

[40] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things", *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7468–7482, 2019.

[41] L. Guo, Z. Chen, D. Zhang, K. Liu and J. Pan, "AoI inspired collaborative information collection for auv assisted internet of underwater things", *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3851–3863, 2021.

[42] M. Emara, H. Elsawy and G. Bauch, "A spatiotemporal model for peak AoI in uplink IoT networks: Time versus event-triggered traffic", *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6762–6777, 2020.

[43] L. Hu, Z. Chen, Y. Dong, Y. Jia, L. Liang and M. Wang, "Status update in IoT networks: Age of information violation probability and optimal update rate", *IEEE Internet of Things Journal*, no. early acces, 2021. DOI: 10.1109/JIOT.2021.3051722.

[44] H. H. Yang, A. Arafa, T. Q. S. Quek and H. Vincent Poor, "Age-based scheduling policy for federated learning in mobile edge networks", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8743–8747, 2020.

[45] B. Zhou, Y. Cui and M. Tao, "Optimal dynamic multicast scheduling for cache-enabled content-centric wireless networks", *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2956–2970, 2017.

[46] A. Kosta, N. Pappas, A. Ephremides and V. Angelakis, "The cost of delay in status updates and their value: non-linear ageing", *IEEE Transactions on Communications*, vol. 68, pp. 1–1, 2020.

[47] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems", *Proceedings 10th ACM/IEEE International Conference on Cyber-Physical Systems*, pp. 109–117, 2019.

[48] G. Stamatakis, N. Pappas and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms", in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[49] Y. Sun, Y. Polyanskiy and E. Uysal, "Sampling of the Wiener Process for Remote Estimation over a Channel with Random Delay", *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, 2020. arXiv: 1707.02531.

[50] M. Klügel, M. H. Mamduhi, S. Hirche and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss", in *IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 189–196.

[51] W. Liu, D. E. Quevedo, K. H. Johansson, B. Vucetic and Y. Li, *Remote state estimation of multiple systems over multiple markov fading channels*, ArXiv Preprint, 2021. arXiv: 2104.04181 [eess.SY]. [Online]. Available: https://arxiv.org/abs/2104.04181.

[52] K. Huang, W. Liu, Y. Li, A. Savkin and B. Vucetic, "Wireless feedback control with variable packet length for industrial IoT", *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1586–1590, 2020.

[53] K. Huang, W. Liu, M. Shirvanimoghaddam, Y. Li and B. Vucetic, "Real-time remote estimation with hybrid ARQ in wireless networked control", *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3490–3504, 2020.

[54] K. Huang, W. Liu, Y. Li, B. Vucetic and A. Savkin, "Optimal downlink–uplink scheduling of wireless networked control for industrial IoT", *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1756–1772, 2020.

[55] G. Li and Y. Wang, "Automatic ARIMA modeling-based data aggregation scheme in wireless sensor networks", *Eurasip Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–13, 2013.

[56] S. Santini and K. Römer, "An adaptive strategy for quality-based data reduction in wireless sensor networks", in *Proc. International Conference on Networked Sensing Systems (INSS)*, 2006, pp. 29–36.

[57] F. A. Aderohunmu, G. Paci, D. Brunelli, J. D. Deng, L. Benini and M. Purvis, "An application-specific forecasting algorithm for extending wsn lifetime", *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 374–381, 2013.

[58] A. Fathalla, K. Li, A. Salah and M. F. Mohamed, "An lstm-based distributed scheme for data transmission reduction of iot systems", *Neurocomputing*, 2021.

[59] R. Askari Moghadam and M. Keshmirpour, "Hybrid ARIMA and neural network model for measurement estimation in energy-efficient wireless sensor networks", in *Informatics Engineering and Information Science*, Springer Berlin Heidelberg, 2011, pp. 35–48.

[60] B. Efron, "Bootstrap methods: Another look at the jackknife", *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[61] M. Golz, V. Koivunen and A. Zoubir, "Nonparametric detection using empirical distributions and bootstrapping", *25th European Signal Processing Conference, EUSIPCO 2017*, pp. 1450–1454, 2017.

[62] S. N. Lahiri, "Resampling methods for dependent data", eng, in. Springer-Verlag New York, 2003, ISBN: 1-4757-3803-X.

[63] J. Lopez-de Lacalle and H. Vinod, "Maximum entropy bootstrap for time series: The meboot R package", *Journal of Statistical Software*, vol. 29, no. 05, 2009.

[64] H. D. Vinod, "Maximum Entropy Bootstrap Algorithm Enhancements", *SSRN Electronic Journal*, pp. 1–15, 2013.

[65] P. Rebentrost, B. Gupt and T. R. Bromley, "Quantum computational finance: Monte Carlo pricing of financial derivatives", *Physical Review A*, vol. 98, no. 2, pp. 1–15, 2018.

[66] R. W. Shonkwiler, "Pricing exotic options", in *Finance with Monte Carlo*, Springer New York, 2013.

[67] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[68] T. Hubert, J. Schrittwieser, L. Baker, F. Hui, D. Hassabis, I. Antonoglou, D. Silver, A. Bolton, K. Simonyan, L. Sifre, T. Lillicrap, T. Graepel, G. van den Driessche, Y. Chen, M. Lai, A. Huang and A. Guez, "Mastering the game of Go without human knowledge", *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[69] W. Chen, X. Qiu, T. Cai, H.-N. Dai, Z. Zheng and Y. Zhang, "Deep reinforcement learning for internet of things: A comprehensive survey", *IEEE Communications Surveys Tutorials*, vol. 23, no. 3, pp. 1659–1692, 2021.

[70] L. A. Villas, A. Boukerche, H. A. De Oliveira, R. B. De Araujo and A. A. Loureiro, "A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks", *Ad Hoc Networks*, vol. 12, no. 1, pp. 69–85, 2014.

[71] S. Hu, G. Li and G. Huang, "Dynamic spatial-correlation-aware topology control of wireless sensor networks using game theory", *IEEE Sensors J.*, vol. 21, no. 5, pp. 7093–7102, 2021.

[72] V. Roy, A. Simonetto and G. Leus, "Spatio-temporal sensor management for environmental field estimation", *Signal Processing*, vol. 128, 2016.

[73] N. Marchang and R. Tripathi, "KNN-ST: Exploiting spatio-temporal correlation for missing data inference in environmental crowd sensing", *IEEE Sensors J.*, vol. 21, no. 3, pp. 3429–3436, 2021.

[74] H. Yetgin, K. T. K. Cheung, M. El-Hajjar and L. Hanzo, "A Survey of Network Lifetime Maximization Techniques in Wireless Sensor Networks", *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, pp. 828–854, 2017.

[75] S. Wu, X. Ren, S. Dey and L. Shi, "Optimal scheduling of multiple sensors over shared channels with packet transmission constraint", *Automatica*, vol. 96, pp. 22 –31, 2018.

[76] M. M. Vasconcelos, M. Gagrani, A. Nayyar and U. Mitra, "Optimal scheduling strategy for networked estimation with energy harvesting", *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 4, pp. 1723–1735, 2020.

[77] A. Leong, A. Ramaswamy, D. Quevedo, H. Karl and L. Shi, "Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems", *Automatica*, vol. 113, pp. 1–8, 2020.

[78] A. S. Leong, D. E. Quevedo, D. Dolz and S. Dey, "Transmission scheduling for remote state estimation over packet dropping links in the presence of an eavesdropper", *IEEE Trans. Automat. Control*, vol. 64, no. 9, pp. 3732–3739, 2019.

[79] M. M. Vasconcelos and U. Mitra, "Observation-driven scheduling for remote estimation of two gaussian random variables", *IEEE Trans. Control-Netw. Syst.*, vol. 7, no. 1, pp. 232–244, 2020.

[80] M. Gagrani, M. M. Vasconcelos and A. Nayyar, "Scheduling and estimation strategies in a sequential networked estimation problem", *2018 56th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2018*, pp. 871–878, 2019.

[81] M. M. Vasconcelos, A. Nayyar and U. Mitra, "Optimal sensor scheduling strategies in networked estimation", pp. 5378–5384, 2017.

[82] J. Hribar, A. Marinescu, G. A. Ropokis and L. A. DaSilva, "Using deep Q-learning to prolong the lifetime of correlated internet of things devices", in *IEEE In. Conf. Commun. Workshops*, 2019, pp. 1–6.

[83] Z. Jiang and S. Zhou, "Status from a random field: How densely should one update?", *IEEE International Symposium on Information Theory - Proceedings*, vol. 2019-July, pp. 1037–1041, 2019.

[84] A. E. Kalor and P. Popovski, "Minimizing the age of information from sensors with common observations", *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1390–1393, 2019.

[85] B. Zhou and W. Saad, "On the Age of Information in Internet of Things Systems with Correlated Devices", 2020. arXiv: 2001.11162.

[86] N. Cressie and C. Wikle, *Statistics for Spatio-Temporal Data*. Wiley, 2011.

[87] R. Furrer, M. G. Genton and D. Nychka, "Covariance tapering for interpolation of large spatial datasets", *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 502–523, 2006.

[88] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[89] Y. Hsu, E. Modiano and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms", *IEEE International Symposium on Information Theory (ISIT)*, pp. 561–565, 2017.

[90] H. Tang, J. Wang, L. Song and J. Song, "Minimizing age of information with power constraints: Multi-user opportunistic scheduling in multi-state time-varying channels", *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 854–868, 2020.

[91] G. Li and Y. Wang, "Automatic arima modeling-based data aggregation scheme in wireless sensor networks", *EURASIP Journal on Wireless Communications and Networking*, vol. 85, no. 1, 2013.

[92] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas and G. P. Picco, "Practical data prediction for real-world wireless sensor networks", *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2231–2244, 2015.

[93] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment", *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

[94] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing", *IEEE Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, 1998.

[95] "Maximum-likelihood estimators and random walks in long memory models", *Statistics*, vol. 45, no. 4, pp. 361–374, 2011.

[96] "Data reduction in wireless sensor networks: a hierarchical LMS prediction approach", *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1708–1715, 2016.

[97] J. Huang, D. Shi and T. Chen, "Event-triggered state estimation with an energy harvesting sensor", *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4768–4775, 2017.

[98] A. S. Leong, D. E. Quevedo, D. Dolz and S. Dey, "Transmission scheduling for remote state estimation over packet dropping links in the presence of an eavesdropper", *IEEE Trans. Autom. Control*, vol. 64, 2019.

[99] M. M. Vasconcelos and N. C. Martins, "Optimal remote estimation of discrete random variables over the collision channel", *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1519–1534, 2019.

[100] H. Yu, J. Shang and T. Chen, "On stochastic and deterministic event-based state estimation", *Automatica*, vol. 123, p. 109 314, 2021.

[101] A. Mohammadi and K. N. Plataniotis, "Event-based estimation with information-based triggering and adaptive update", *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4924–4939, 2017.

[102] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han and L. Shi, "Multi-sensor scheduling for state estimation with event-based, stochastic triggers", *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2695–2701, 2016.

[103] M. M. Vasconcelos and N. C. Martins, "Optimal estimation over the collision channel", *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 321–336, 2017.

[104] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous univariate distributions*. Wiley, 1994, vol. 1.

[105] D. R. Barr and E. T. Sherrill, "Mean and variance of truncated normal distributions", *American Statistician*, vol. 53, no. 4, pp. 357–361, 1999.

[106] I. V. Lotov, "On some boundary crossing problems for Gaussian random walks", *The Annals of Probability*, vol. 24, no. 4, pp. 2154–2171, 1996.

Appendix A

Publications on AoI-based Scheduling of Spatio-temporally Dependent Measurements

I V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal scheduling policy for spatio-temporally dependent observations using age-of-information," *IEEE 23rd International Conference on Information Fusion (FUSION)*, pp. 1154–1160, 2020, [Runner up - Best Paper Award]

II V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal scheduling of multiple spatio-temporally dependent observations using age-of-information," *54th Asilomar conference on signals, systems, and computers*, pp. 1188–1193, 2020

III V. W. Håkansson, N. K. D. Venkategowda, S. Werner and P. K. Varshney, "Optimal scheduling of multiple spatio-temporally dependent observations for remote estimation using age-of-information," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20308-20321, Oct., 2022

IV V. W. Håkansson, N. K. D. Venkategowda, S. Werner and P. K. Varshney, "Optimal transmission-constrained scheduling of spatio-temporally dependent observations using age-of-information," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15596-15606, Aug., 2022

Appendix B

Publications on Optimized Threshold-triggered Transmission Schemes

V V. W. Håkansson, N. K. D. Venkategowda, F. A. Kræmer and S. Werner, "Cost-aware dual prediction scheme for reducing transmissions at IoT sensor nodes," *27th European Signal Processing Conference (EUSIPCO)*, pp. 2166–2171, 2019

VI V. W. Håkansson, N. K. D. Venkategowda and S. Werner, "Optimal transmission threshold and channel allocation strategies for heterogeneous sensor data," *55th Asilomar Conference on Signals, Systems, and Computers*, pp. 757–761, 2021

These papers are not included due to IEEE copyright

ISBN 978-82-326-6262-3 (printed ver.)
ISBN 978-82-326-5221-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology