A Framework for Characterising the Value of Information in Communication Systems



Zijing Wang

Balliol College

Department of Engineering Science, University of Oxford

Supervised by Professor Justin P. Coon

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy in Engineering Science

Trinity 2023

Abstract

Thesis Title: A Framework for Characterising the Value of Information in Communication Systems
Name: Zijing Wang
College: Balliol College
Degree: Doctor of Philosophy
Term and Year of Submission: Trinity 2023

With the rapid development of sensor-based Internet of Things (IoT) technology, there are increasingly many real-time control systems, in which the timeliness of information is becoming increasingly essential for decision-making and control. In this thesis, a mutual information-based value of information (VoI) framework is formulated to characterise how useful and valuable the data is for latent variable models. The VoI is explored in the context of various hidden Markov models, in which data samples generated at the source are assumed to be transmitted over a noisy channel to the receiver. The signal-to-noise ratio (SNR) indicates the level of noise which is used to represent the quality of transmission in such communication systems. First, in hidden Gauss-Markov models, closed-form expressions and bounds of the VoI are derived. The effects of sampling and the number of observations on VoI are investigated in different SNR regimes. The VoI-optimal sampling policy is obtained under a sampling rate constraint. Next, the VoI is further analysed in an important hidden Gauss-Markov model, the noisy Ornstein-Uhlenbeck process, in more depth. The distribution of the VoI is obtained. Functional relationships between the VoI and other traditional metrics are explored. Finally, the VoI is applied and optimised in an IoT network with multiple static sensors and one mobile central entity. The mobility of the centre and the scheduling strategy of sensors are jointly optimised for VoI maximisation. A reinforcement learning-based algorithm is proposed to solve this problem efficiently. Numerical results are presented by Monte Carlo simulations. The proposed VoI framework is able to capture the data freshness, the correlation of the underlying process, and the SNR through the transmission channel. The results illustrated in this thesis can be used in the design of real-time communication systems.

Acknowledgements

First and foremost, I would like to express my sincerest thanks to my supervisor Prof. Justin P. Coon for his continuous guidance, encouragement, support, and patience over the past four years. I cannot thank him enough for spending a lot of time providing insightful comments on my papers and thesis. I truly enjoy working with him, and his attitude towards scientific research really inspires and motivates me.

Meanwhile, I especially thank my collaborator Dr. Mihai-Alin Badiu for inspiring and helping me with my research problems. He is always patient to answer my questions and is always encouraging in my research. He provides a lot of helpful suggestions on our publications and points out promising research directions for our joint work.

I would like to thank Tianxiong Wang, Dr. Ferhat Yarkin, Dr. Sunghwan Cho, Tanmayee Deshprabhu, Martin Wafula, Amirmohammad Farzaneh, Dr. Jinchuan Tang as well as other colleagues in the group. I would also like to thank Dr. Gaojie Chen from the University of Surrey. They are open to sharing experiences and discussions. I really appreciate the chance to learn from them and broaden my horizons through this journey.

I would like to say thank you to my Transfer and Confirmation of Status assessors Prof. Ekaterina Shamonina and Prof. Dominic O'Brien, my internal and external examiners Dr. Konstantinos Gatsis and Prof. Deniz Gunduz from Imperial College London, and all the reviewers of my publications, for their constructive comments to improve the quality of my work.

In addition, I would like to express my gratitude to the Clarendon Scholarship Fund and Jason Hu Scholarship for awarding me a full scholarship for my DPhil study at Oxford. I would also like to thank the Department of Engineering Science and Balliol College for choosing me. It is always a privilege for me to be a member here.

Most importantly, I would like to thank my beloved parents, grandparents as well as other family members, for everything they have done for me. I would not be who I am today without their unconditional love, support and belief.

List of Publications

Journal Papers

- **Z. Wang**, M.-A. Badiu, and J. P, Coon, "Optimal mobility and communication strategy to maximize the value of information in IoT networks," ***IEEE Transactions on Wireless Communications***, under review.
- **Z. Wang**, M.-A. Badiu, and J. P, Coon, "A framework for characterizing the value of information in hidden Markov models," ***IEEE Transactions on Information Theory***, vol. 68, no. 8, pp. 5203-5216, 2022.
- **Z. Wang**, M.-A. Badiu, and J. P. Coon, "On the value of information and mean squared error for noisy Gaussian models," ***IEEE Communications Letters***, vol. 26, no. 9, pp. 2023-2026, 2022.
- **Z. Wang**, M.-A. Badiu, and J. P. Coon, "Relationship between age and value of information for a noisy Ornstein–Uhlenbeck process," ***Entropy***, 23(8), 940, 2021.

Conference Papers

- **Z. Wang**, M.-A. Badiu, and J. P. Coon, "Trading off SNR and the number of observations to improve the value of information in IoT networks," ***IEEE Vehicular Technology Conference (VTC)***, Fall 2022, pp. 1-6.
- **Z. Wang**, M.-A. Badiu, and J. P. Coon, "A value of information framework for latent variable models," ***IEEE Global Communications Conference (GLOBECOM)***, 2020, pp. 1-6.

Acronyms

AHP	Analytic hierarchy process
AI	Artificial intelligence
AoI	Age of information
AWGN	Additive white Gaussian noise
CDF	Cumulative distribution function
FCFS	First come first serve
i.i.d.	Independent and identically distributed
IoT	Internet of Things
LCFS	Last come first serve
LoS	Line of sight
MDP	Markov decision process
MSE	Mean-squared error
MTC	Machine-type communications
MGF	Moment-generating function
MEC	Mobile edge computing
OU process	Ornstein-Uhlenbeck process
PDF	Probability density function
PPP	Poisson point process
SDE	Stochastic differential equation
SNR	Signal-to-noise ratio
SWIPT	Simultaneous wireless information and power transfer
TDMA	Time division multiple access
UAV	Unmanned aerial vehicle
VoI	Value of information
WSN	Wireless sensor network

Mathematical Notations

$\mathrm{P}(\cdot)$	Probability operator
$f_X(x)$	Probability density function of random variable X
$F_X(x)$	Cumulative distribution of random variable X
$\mathrm{E}[\cdot]$	Expectation operator
$\mathrm{Var}[\cdot]$	Variance operator
$R_{XX}(t_1, t_2)$	Autocorrelation function between X_{t_1} and X_{t_2}
$K_{XX}(t_1, t_2)$	Autocovariance function between X_{t_1} and X_{t_2}
$\xi_{XX}(t_1, t_2)$	Autocorrelation coefficient between X_{t_1} and X_{t_2}
$\overline{(\cdot)}^*$	Complex conjugation operator
$\log(\cdot)$	Natural logarithm
$\mathrm{H}(X)$	Entropy of the random variable X
$\mathrm{H}(X\|Y)$	Conditional entropy
$\mathrm{h}(X)$	Differential entropy of the random variable X
$\mathrm{h}(X\|Y)$	Conditional differential entropy
$\mathrm{H}(\mathcal{X})$	Entropy rate of the random process X
$\mathrm{I}(X;Y)$	Mutual information
$\det[\cdot]$	Determinant operator
$[\cdot]^{\mathrm{T}}$	Transpose operator
$[\cdot]^{-1}$	Inverse operator

Contents

1 Introduction **1**

1.1 Internet of Things 1

1.2 Motivation 4

1.3 Thesis Outline and Contributions 6

2 Background **9**

2.1 Status Update System Model 9

2.2 Data Freshness Metrics 12

2.2.1 Age of Information and its Variants 12

2.2.2 Non-linear Age of Information 14

2.2.3 Information-theoretic Metrics 16

2.2.3.1 Correlation Function 16

2.2.3.2 Estimation Error 17

2.2.3.3 Entropy 19

2.2.4 Value of Information 22

2.2.5 Other Metrics 22

2.3 Applications of Data Freshness Metrics 24

2.3.1 Age of Information in Queueing Models 24

2.3.1.1 Queueing Theory 24

2.3.1.2 AoI in Queues 26

2.3.2 Age of Information in Internet of Things 27

2.4 Summary 29

3 Value of Information Formulation **31**

3.1 VoI for Latent Variable Models 32

3.2 VoI for Hidden Markov Models 35

3.3 VoI for Hidden Gauss-Markov Models 37

3.3.1 Noisy Gauss-Markov Model 37

3.3.2 Value of Information . . . 39
3.3.3 Connection to Mean-squared Error . . . 41
3.4 Numerical Results . . . 44
3.5 Summary . . . 47
3.6 Appendix . . . 48
3.6.1 Proof of Proposition 1 . . . 48
3.6.2 Proof of Corollary 1 . . . 49

4 Analysis of Hidden Gauss-Markov Process Models 50
4.1 Effective Sampling Policies for Value of Information . . . 51
4.1.1 Arbitrary Fixed Sampling Intervals . . . 51
4.1.1.1 High SNR Regime . . . 53
4.1.1.2 Low SNR Regime . . . 54
4.1.2 Uniform Sampling Intervals . . . 56
4.2 Trading off SNR and the Number of Observations to Improve the Value of Information . . . 58
4.2.1 High SNR Regime . . . 60
4.2.2 Low SNR Regime . . . 61
4.3 Optimal Sampling Policy to Maximise the Value of Information . . . 62
4.3.1 Problem Formulation . . . 62
4.3.2 A Threshold-based Sampling Policy . . . 65
4.4 Numerical Results . . . 66
4.5 Summary . . . 71
4.6 Appendix . . . 72
4.6.1 Proof of (4.24) . . . 72
4.6.2 Proof of Proposition 3 . . . 75
4.6.3 Proof of Corollary 9 . . . 76
4.6.4 Proof of Corollary 11 . . . 77

5 Analysis of the Ornstein-Uhlenbeck Process Model 78
5.1 Ornstein–Uhlenbeck Process . . . 79
5.2 VoI for Noisy OU Model . . . 80
5.2.1 Value of Information . . . 80
5.2.2 Connection to Age of Information . . . 81
5.3 Analysis for M/M/1 Queueing Model . . . 83
5.3.1 M/M/1 Queueing Model . . . 83
5.3.2 Stationary Distribution of VoI . . . 85

5.3.2.1 Distribution of the VoI . . . 86
5.3.2.2 Moments and Bounds . . . 88
5.4 Numerical Results . . . 90
5.5 Summary . . . 98
5.6 Appendix . . . 99
5.6.1 Proof of Monotonicity in μ . . . 99
5.6.2 Proof of the Existence of the Optimal λ . . . 100
5.6.3 Proof of Proposition 7 . . . 100
5.6.4 Proof of Corollary 17 . . . 101
5.6.5 Proof of Proposition 8 . . . 102

6 Optimal Mobility and Communication Strategy in IoT Networks 104
6.1 System Model and Problem Formulation . . . 105
6.1.1 Mobility Model . . . 107
6.1.2 Communication Model . . . 108
6.1.3 Value of Information Model . . . 108
6.1.4 Problem Formulation . . . 110
6.2 A Reinforcement Learning-based Algorithm . . . 111
6.2.1 Markov Decision Process . . . 112
6.2.2 A Q-learning based Algorithm . . . 115
6.3 Numerical Results . . . 117
6.4 Summary . . . 124

7 Conclusions 125
7.1 Thesis Summary . . . 125
7.2 Future Works . . . 127
7.3 Conclusions . . . 128

Bibliography 130

List of Figures

2.1 Model of a status update system. 10
2.2 Time evolution of the AoI. 13
2.3 Model of a single queue. 25
2.4 Model of a birth-death process. 25

3.1 Sampling of a target random process. 33
3.2 Temporal evolution of the Markov model. 34
3.3 Temporal evolution of the hidden Markov model. 35
3.4 AWGN channel model. 38
3.5 Time evolution of VoI in Markov and hidden Markov process models; correlation parameter $\rho = 0.8$, $\sigma_x^2 = 5$, noise parameter $\sigma_n^2 = 1$, and sampling interval $\Delta t = 2$. 45
3.6 VoI in the hidden Markov model versus the number of observations m for $\sigma_n^2 \in \{0.1, 2, 5, 10\}$ at $t = 100$; correlation parameter $\rho = 0.9$, $\sigma_x^2 = 10$, and sampling interval $\Delta t = 2$. 46
3.7 VoI versus the correlation parameter for $m \in \{1, 10\}$ and $\gamma \in \{1, 10\}$; sampling rate $\lambda = 0.5$. 47

4.1 Time evolution of the VoI. 63
4.2 High SNR regime: Comparison of the exact VoI and the approximated VoI with arbitrary sampling times for $\rho \in \{0.7, 0.9\}$ at $t = 100$; sampling rate $\lambda = 0.5$ and the number of observations $m = 5$. 67
4.3 Low SNR regime: Comparison of the exact VoI and the approximated VoI with arbitrary sampling times for $\rho \in \{0.4, 0.6\}$ at $t = 100$; sampling rate $\lambda = 0.5$ and the number of observations $m = 5$. 67
4.4 The normalised value of information with different number of observations $m \in \{1, 2, 3, 4, 5, 6\}$ and different SNR $\gamma \in (0, 8)$. The correlation parameter $\rho = 0.7$. 68

4.5 The improvement of the value of information by increasing the number of observations $m \in \{1, 2, 3, \cdots, 8, 9, 10\}$ under different SNR conditions $\gamma \in (0, 10)$. The correlation parameter $\rho = 0.7$. 68
4.6 Average VoI in the hidden Markov process versus the transmission rate parameter μ for different sampling policies without sampling rate constraints; correlation parameter $\rho = 0.9$ and the SNR $\gamma = 50$. . . . 70
4.7 Average VoI in the hidden Markov process versus $1/f_{\max}$ for different sampling policies; transmission rate parameter $\mu = 0.5$, correlation parameter $\rho = 0.9$ and the SNR $\gamma = 50$. 70

5.1 Two examples of the service time in M/M/1 Queueing Model. 84
5.2 Low SNR regime: Comparison of the exact VoI and the exponential VoI versus κ for $\sigma_n^2 \in \{10, 30\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 91
5.3 High SNR regime: Comparison of the exact VoI and the logarithmic VoI versus κ for $\sigma_n^2 \in \{1, 5\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 92
5.4 Intermediate SNR regime: Comparison of the exact VoI and the linear VoI versus κ for $\sigma_n^2 \in \{10, 30\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 92
5.5 The density function of the VoI; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. . . 93
5.6 The cumulative distribution function of the VoI versus v for $\sigma_n^2 \in \{0.5, 1\}$ and $\kappa \in \{0.05, 0.1, 0.2\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 94
5.7 The VoI outage probability versus λ for $\kappa \in \{0.05, 0.1, 0.2\}$; threshold $v = 0.4$, noise parameter $\sigma_n^2 = 0.5$ and service rate $\mu = 1$. 94
5.8 The VoI outage probability versus μ for $\kappa \in \{0.05, 0.1, 0.2\}$; threshold $v = 0.4$, noise parameter $\sigma_n^2 = 0.5$ and sampling rate $\lambda = 0.2$. 95
5.9 The average VoI and its bounds versus the sampling rate λ; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$ and service rate $\mu = 1$. . 96
5.10 The average VoI and its bounds versus the service rate μ; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$ and sampling rate $\lambda = 0.2$. 96
5.11 The average VoI and the lower bound versus κ for $\sigma_n^2 \in \{1, 5\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 97

5.12 The average VoI and the upper bound versus κ for $\sigma_n^2 \in \{0.1, 0.5, 1\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$. 97

6.1 System model of an industrial IoT network. 106
6.2 A case study. The number of nodes $N = 3$, the correlation parameter of each node $\rho = [0.7, 0.6, 0.5]$, initial age of each node $A = [2, 4, 4]$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 10$. 119
6.3 The average value of information evolution among all nodes in the network given in Fig. 6.2. 120
6.4 The time average minimum value of information among all nodes versus the transmission power. The number of nodes $N = 4$, the correlation parameter is generated uniformly from $[0, 1]$, and the total number of time slots $T_{\max} = 30$. 121
6.5 Time average minimum value of information among all nodes versus different correlation parameters. The number of nodes $N = 4$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 30$. 122
6.6 Time average minimum value of information among all nodes versus the number of nodes. The correlation parameter is generated uniformly from $[0, 1]$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 30$. 123

Chapter 1

Introduction

1.1 Internet of Things

The term Internet of Things (IoT) was coined by Kevin Ashton in 1999 to describe a system where objects are connected to the Internet [1]. Nowadays, the IoT is a network of physical devices which are equipped with sensors, software, processing and communication abilities, allowing them to collect and exchange data with other devices as well as humans over the network. With the exponential growth of connected devices, the IoT has become one of the most important technological developments. The scale of IoT is immense and the potential for innovation and transformation is enormous. There is an explosion in the number of IoT devices all over the world and it is expected to continuously increase in the following years. The number of connected IoT devices in use worldwide was estimated to be 7 billion in 2019, and this number is estimated to be over 75 billion by 2025 and 80 billion by 2030 [2, 3].

Driven by the rapid advance in intelligent sensor-based IoT networks, there are increasingly many real-time control applications, such as environmental surveillance, smart homes, autonomous driving, industrial automation, e-health and so on. These applications can be modelled as a generic status update system, in which sensing

devices are deployed to continuously monitor the ambient environment and sample data. Sensor-based IoT devices range from simple things like watches and phones to complex things like vehicles, industrial machines and medical equipment. They are used to monitor various physical phenomena, such as temperature, light, motion, blood pressure and so on. The sampled data represents the status information of the target physical process, and it is required to be transmitted to a remote monitor as timely as possible with the aim of supporting real-time management, providing insights for decision-making, enabling precise control of entities in the real world and thus improving the quality of our life.

IoT technology is expected to create new opportunities and potential to change the world. It also presents new engineering challenges. The first challenge of IoT technology is energy efficiency. Since most IoT devices rely on batteries which are difficult to be replaced or recharged, reducing energy consumption is crucial to extending the lifetime of IoT networks. Fortunately, there are several ways to improve the energy efficiency of IoT networks. For example, multihop networks can be used to reduce the transmission distance and balance the traffic load. Low-power processors can also be utilised to decrease energy consumption. Moreover, energy harvesting technology [4] is able to provide additional energy to IoT devices. This technology utilises energy sources such as solar power and wind power to recharge IoT devices' batteries, reducing the need for frequent battery replacements. Simultaneous wireless information and power transfer (SWIPT) is a new technology which enables the transfer of wireless power and data between devices at the same time [5]. This technology has the potential to revolutionise power delivery by enabling devices to work without the need for batteries or other external energy sources.

The second challenge is efficient data processing. The growth of IoT technology has resulted in an increasing number of connected devices, leading to a massive volume of data generated by these devices. However, traditional computing and storage

infrastructure is not sufficient enough to handle a large amount of data. One approach to address this challenge is cache and mobile edge computing (MEC) technologies [6]. The cache is able to bring extra storage resources to IoT devices, allowing them to store and process data more efficiently. By utilising MEC technology, computing and storage infrastructures can be deployed closer to the edge of the network (i.e., closer to the place where IoT devices are located). This allows data to be analysed and processed locally, reducing the amount of data that needs to be transmitted over the core network and improving overall network efficiency. The other approach is the use of artificial intelligence (AI) [7]. AI can be used to analyse data and remove redundant or unnecessary information, optimising the storage and processing spaces of data. For example, machine learning algorithms can be used to identify patterns in raw data and make predictions based on that data.

The third challenge that arises in IoT is the issue of security and data privacy. With the increasing number of IoT devices, there is an increasing risk of cyber-attacks on these devices and this means that the privacy of the data may be unsecured. To address this challenge, advances in blockchain and physical layer security technologies can be utilised. Blockchain is a decentralised and distributed technology which enables secure and transparent record keeping and prevents unauthorised access to data [8]. Physical layer security technology is another method which comprises advanced encryption techniques and signal processing algorithms to secure the wireless communication between IoT devices. This makes it harder for attackers to intercept the data, thereby improving the overall security of the IoT network.

The timeliness of data is also significant in IoT networks. In many applications, it is not enough to simply receive data, but the data packets are required to be transmitted in a timely manner to support real-time management and precise control. Therefore, fresh and useful information has become increasingly important in such communication systems to guarantee efficient network operation. The question of how

to measure the timeliness of information has become an interesting topic of research in real-time communication systems. Although some existing measurements have enabled IoT to become a reality, there still remain challenges in this area to address to further reach its full potential. This thesis aims to propose a framework which can be used to characterise the value and usefulness of information in IoT networks.

1.2 Motivation

Existing performance metrics, such as delay and throughput, successfully meet the demand of traditional wireless sensor networks (WSN), in which the data packet is treated individually and the value of each sampled data packet does not change over time. However, for real-time communication systems, data packets are temporally correlated with each other, thus a newly generated packet should be more important than an old packet.

Against this background, the age of information (AoI) is proposed in [9] as a new performance metric for data freshness from the receiver's perspective. It is defined as the time duration since the generation time of the latest received data sample. The AoI has received considerable attention due to its utility in characterising the timeliness of information, and it has been widely studied as a concept, a metric, and a tool in a variety of communication systems [10, 11]. Many works focus on AoI and its variants (e.g., peak AoI and average AoI) at the fundamental level, studying its distribution, moments and bounds [9,12–16]. In addition to the fundamental research, AoI-based sampling, scheduling and optimisation problems are extensively explored in various practical contexts [17, 18], including single-hop networks [19], multiple-hop networks [20], energy harvesting networks [21], unmanned aerial vehicle (UAV) networks [22, 23], and so on.

However, it is recognised that AoI is not a perfect metric of utility in all applica-

tions. The AoI grows linearly with time until a new status update is received, which means that the AoI is independent of the statistical variations inherent in the underlying source data and is unable to capture the non-linear performance degradation in information quality. For example, some data sources (e.g., node mobility) change frequently over time, thus even fresh samples may hold little useful information. While other data sources (e.g., temperature) change more slowly, old samples may still be valuable for further actuation. In practice, old information with large age may still have value while new information with small age may have less value. Therefore, the AoI may not be suitable to be applied in various communication systems.

In this regard, other approaches are further investigated to quantify the information value. A general way is to utilise non-linear AoI functions [24]. The "age penalty" is introduced in [25], which maps the AoI to a non-linear and non-decreasing penalty function to evaluate the level of "dissatisfaction" related to the outdated information. Despite these contributions, it seems that the non-linear functions in the existing work have been chosen arbitrarily without any particular principled basis. Information-theoretic AoI research is also widely discussed to provide a theoretical interpretation of non-linear age functions. The mean-squared error (MSE) in remote estimation [26–29], conditional entropy [30] and mutual information [31] in information theory can remove the linearity and has been widely utilised to measure the information value.

Moreover, in practical applications, noise, interference, errors and other features can lead to severe performance degradation. This means that status updates generated at the source can be negatively affected, and may be hidden from observation when they are delivered to the receiver. However, existing works mostly treat Markov models, in which data samples are assumed to be directly observable at the receiver and information value only relates to the most recently received measurements. In practice, all the past observations should hold valuable information about the under-

lying random process, and the sample received via a poor wireless channel contains less value than the sample transmitted under good channel conditions.

Therefore, a more appropriate metric is required to evaluate the information value for latent variable models. We are motivated to develop a value of information (VoI) framework in the context of information theory which can be used to measure the usefulness of data for real-time communication systems.

1.3 Thesis Outline and Contributions

In this thesis, we focus on the effective monitoring process of real-time communication systems and propose a mutual information-based value of information framework. The VoI and its applications are investigated in the context of various latent variable models. This thesis includes four research chapters (Chapters 3 to 6), preceded by a research overview and followed by a conclusion. The details of this thesis are given as follows.

- In Chapter 2, we provide the relevant research overview and literature review for the thesis, aiming at giving preliminary knowledge and a full picture of this work for readers' better understanding. First, we provide the system model of the IoT-assisted status update system. Next, we introduce the concepts of existing performance metrics for data freshness, including the AoI and its variants. We also briefly introduce the basic knowledge of information theory and information-theoretic measures. Finally, the AoI-based applications in various networks are summarised.

- In Chapter 3, we first propose a mutual information-based value of information framework to characterise the usefulness of data for general latent variable models. The VoI definition gives the standard interpretation of the reduction in uncertainty in the current status of a latent process given a sequence of noisy

observations. Next, we analyse the VoI for one of the most important latent variable models: hidden Markov models, deriving the bound of the VoI. Finally, we further explore the VoI for hidden Gauss-Markov models, deriving its closed-form expression and studying the connection between the VoI and MSE which illuminates useful relationships between the information theory and signal processing theory. The research in this chapter is published in *IEEE Globecom 2020* [32] and *IEEE Communications Letters* [33].

- In Chapter 4, we focus on general hidden Gauss-Markov models to explore the impact of sampling, the number of observations and the signal-to-noise ratio (SNR) on the VoI. First, we investigate the VoI with different sampling intervals. The simplified closed-form expressions of the VoI are derived in high and low SNR regimes with arbitrary and uniform sampling intervals. We show that the VoI can be improved by increasing the SNR regardless of how many samples are used. Increasing the number of observations is effective to improve the VoI only in the low SNR regime, while it yields a zero rate of change of the VoI in the high SNR regime. Next, we study the trade-off between the SNR and the observations amount, and scale back SNR to achieve the same VoI improvement by adding one extra observation. Finally, the optimal sampling policy to maximise the average VoI under the sampling rate constraint is obtained. The research in this chapter is published in *IEEE Transactions on Information Theory* [34] and *IEEE VTC Fall 2022* [35].

- In Chapter 5, we look at a specific hidden Gauss-Markov model: the noisy Ornstein-Uhlenbeck process, to explore VoI and its applications in more depth. First, the functional relationship between the VoI and AoI is studied. We explore the effects of correlation and noise on their relationship, and find logarithmic, exponential and linear dependencies between the two in three different

SNR regimes. This gives the formal justification for the selection of non-linear AoI functions previously reported in other works. Next, by utilising the distribution of the AoI, we further study the distribution of the VoI. Distribution functions, moments, and lower and upper bounds of the VoI are also derived. The research in this chapter is published in *Entropy* [36].

- In Chapter 6, we apply the VoI framework in an Internet of Things network and study the VoI-based optimisation problem. We consider an industrial IoT network with various types of sensing devices (transmitters) and a mobile central entity (receiver). We aim to maximise the VoI by jointly optimising the moving path of the central entity and user scheduling policy under the mobility, communication and VoI constraints. This problem is formulated as a Markov decision process (MDP), and a reinforcement learning-based algorithm is proposed to solve the problem efficiently. We show that the proposed scheme is able to measure the data value in temporal and spatial domains. Compared with other existing methods, the proposed method achieves better performance when the wireless resource is limited by exploiting the correlation properties of sensing devices. The research in this chapter is submitted to *IEEE Transactions on Wireless Communications*.

- In Chapter 7, we provide the conclusions of this thesis and discuss potential areas of future work.

Chapter 2

Background

In this chapter, the relevant research overview is provided. We aim at giving the fundamentals and preliminary knowledge of the timeliness of information. First, we introduce the IoT-assisted status update system. Next, we begin with the definition of the age of information and proceed with its variants. Finally, we review the applications of the data freshness metrics, including the AoI and its statistical properties in different queueing models and the AoI in various wireless communication systems.

2.1 Status Update System Model

Nowadays, there are many IoT-assisted real-time control applications, such as smart transport, industrial control, environment surveillance, e-health, and so on, which interconnect cyberspace with the physical world by integrating wireless sensor networks, massive machine-type communications (MTC), and computational engines [37, 38]. Specifically, sensing devices are deployed to monitor a variety of physical phenomena, and continuously sample data of the target physical process. The sampled data represents the status of the underlying random process and will be transmitted through the wireless network to a remote monitor or computing device with the aim of supporting real-time monitoring, decision-making and control of the entities in the physical

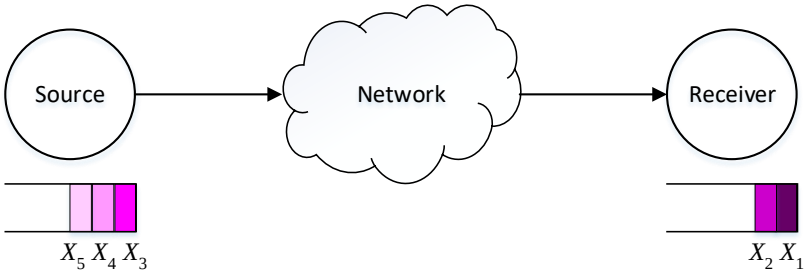


Figure 2.1: Model of a status update system.

world.

These real-time monitoring applications have a common need and can be modelled as a generic term: the status update system. Fig. 2.1 gives a simple model of the status update system. The physical phenomenon under observation is modelled by a random process $\{X_t\}$. The source monitors this target physical process and samples data to get the timely status of $\{X_t\}$. X_i represents the value of the i-th data sample. In such systems, the remote monitor relies on the real-time status of the data source for further analysis, thus the sampled data is required to be updated and transmitted as timely as possible so that the monitor can hold the latest information on the target source. In practice, the information kept on the monitor is staler than the real-time status of the source. There are two reasons. On the one hand, some data sources change very quickly, such as the stock and the position of a moving car. If the source does not update its status in a timely manner, the data on the monitor can be useless. On the other hand, due to the congestion in the queue buffer and transmission channel, the network may experience random delay and the sample cannot be received by the monitor as soon as it is generated at the source. Therefore, the timeliness of information is of critical importance in such IoT-assisted status update systems.

There are two major differences between the novel status update systems and traditional sensor networks. First, the target random process at the data source has

Markovian property, in which the future is independent of the past states, given the present state. Specifically, we have

$$\mathrm{P}(X_{n+1}|X_n, \cdots, X_1) = \mathrm{P}(X_{n+1}|X_n). \tag{2.1}$$

This means that the new data sample is able to fully replace the old generated data sample. Second, the timeliness of a status update system is not equivalent to the transmission time of a traditional WSN. One of the challenges in the design of a status update system is to find a suitable metric to quantify the system performance in terms of data freshness. Throughput, delay and estimation error are widely utilised as the metric for a traditional communication system, but they may not be perfect for a novel status update system [39]. The reasons are given as follows.

- Throughput measures the transmission rate over a communication channel, which reflects how many bits of information can be transmitted successfully to the receiver during a given time period. Unfortunately, it is not able to reflect how many bits of fresh and timely information can be delivered. For example, the maximum throughput can be achieved when the queueing buffer is kept full and the transmission channel is kept busy. In this case, the network may experience a large delay due to the congestion and thus data samples can be stale.

- Delay measures the time duration from the generation of a data sample at the source to its received time at the receiver, which reflects the freshness of a single data packet. However, the data samples generated in a status update system are correlated, and this means that the newly received data sample should be more useful than the previous one. Delay treats each status update sample independently and the value of each packet does not change over time, which is not appropriate to measure the timeliness of information in a status update

system with correlated data samples.

- Estimation error measures the difference between the current status and the estimation. The measured observations at the receiver can be utilised to make estimates of the current state of a system, which is referred to as a filtering problem. The quality of the estimation can be quantified by using a distortion measure, the most common of which for unbiased estimators is the mean-squared error. The freshness of measured data can only be quantified when the estimator exists. But how to quantify when it is not possible or there is no need to treat the measurements in a filtering framework is still a question.

Against this background, how to measure data freshness is an interesting and challenging topic in the IoT-assisted status update system.

2.2 Data Freshness Metrics

2.2.1 Age of Information and its Variants

In the past few years, the concept of the age of information has been introduced in [9, 40] as a new performance metric for data freshness from the receiver's perspective. It is defined as the time elapsed since the generation time of the most recently received data sample. The AoI at time t can be expressed as[1]

$$A(t) = t - \delta(t), \tag{2.2}$$

where $\delta(t)$ is a function of time, representing the generation time of the latest received sample during the time period $(0, t)$.

We denote the generation time of the i-th status update sample as t_i and its

[1] $A(t)$ can be either continuous-time or discrete-time. Here, we take the continuous-time AoI as an example.

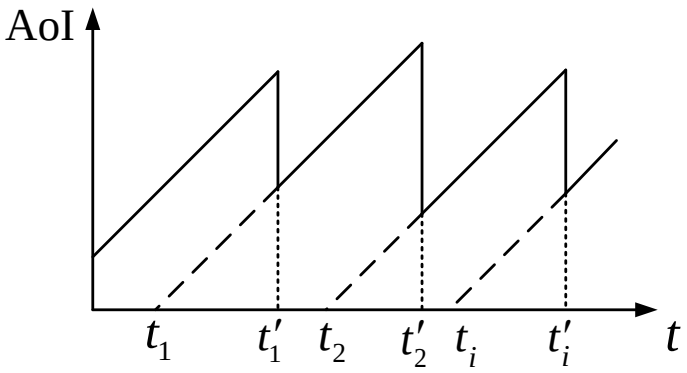


Figure 2.2: Time evolution of the AoI.

received time at the receiver is denoted as t_i'. Fig. 2.2 shows how the AoI varies with time. In this model, the function $\delta(t) = t_i$ given that $t_i' \leq t < t_{i+1}'$ for any positive integer i. It is easy to see that the AoI involves two parts. One is the transmission delay of a single data sample which refers to $t_i' - t_i$. The other is the waiting time for a new data packet at the receiver which refers to $t - t_i'$. The AoI increases linearly with the unit slope with time t and decreases sharply to $t_i' - t_i$ when the i-th data sample is delivered. This behaviour illustrates that the delivery of a new data sample can lead to AoI reduction and refresh the receiver's knowledge about the underlying process from the time domain.

The AoI has received much attention due to its novelty and simplicity. In addition to the instantaneous AoI, the average AoI and the peak AoI are the other two main data freshness performance metrics which are more tractable to evaluate the system-level performance in terms of data freshness. The average AoI over time T is given as

$$A(T) = \frac{1}{T} \int_0^T A(t) \, \mathrm{d}t. \tag{2.3}$$

As $T \to \infty$, the time average AoI will converge to its corresponding stochastic average. Therefore, the average AoI can be obtained as

$$\overline{A} = \lim_{T \to \infty} A(T) = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(t) \, \mathrm{d}t. \tag{2.4}$$

The peak AoI and average peak AoI are investigated in [41, 42] as a new metric to evaluate the worst-case data freshness performance. In Fig. 2.2, we can observe that the AoI will reach a local maximum immediately before a new update is transmitted to the receiver. The peak AoI of the i-th status update is given as

$$A_P(i) = t'_{i+1} - t_i. \tag{2.5}$$

As the number of updates grows large, the average peak AoI becomes

$$A_P = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} A_P(i). \tag{2.6}$$

Compared with the instantaneous AoI and the average AoI, the peak AoI is easy to calculate. It provides information about the timeliness of information in the worst case, and it is believed to be of interest in applications with a threshold constraint on the AoI.

2.2.2 Non-linear Age of Information

The information age given in (2.2) increases linearly with time at a unit slope until a new status update is received, which means that the concept of AoI is independent of the statistical variations inherent in the underlying source data. However, in some practical cases, old information may still have value while new information may have less value. In the example of an autonomous driving system, the mobility of a car evolves very frequently over time, thus even fresh samples may hold little valuable information. While the engine temperature evolves slowly, old samples may be sufficient enough to be used for further analysis and actuation. This means that the AoI cannot fully capture the non-linear performance degradation in information quality caused by the time lapse between status updates or the relevant properties the underlying random process might exhibit, such as how correlated it is. In this

regard, the AoI may not be a perfect metric. Therefore, other performance metrics should be further investigated to quantify the information value.

A common approach to characterise the non-linearity is to utilise the non-linear AoI-related functions to quantify the timeliness of information [24, 25]. The concept of the "age penalty" is introduced in [25], which maps the AoI to a non-decreasing penalty function $p(A(t))$ to evaluate the level of "dissatisfaction" related to outdated information. The function $p(\cdot)$ is assumed to be non-negative and non-decreasing, and it can be either continuous or discrete. Similarly, the concept of "age utility" is also introduced which maps the AoI to a non-increasing utility function $u(A(t))$ to evaluate the level of "satisfaction" related to data refreshment. A simple method is to set $u(A(t)) = -p(A(t))$. The selection of the monotonic age penalty function can be application-specific.

The exponential and logarithmic AoI penalty functions are investigated in [43,44]. The authors suggest that, for highly correlated samples, the logarithmic function is utilised and exponential functions can be chosen in low-correlation scenarios. Closed-form expressions of the exponential penalty function under different queueing models are derived in [45] for energy harvesting networks. Binary penalty functions are used to evaluate the freshness for web crawling [46, 47] and cache updating systems [48]. This function takes the value 0 when the data is up-to-date (i.e., AoI is below a threshold); otherwise, it takes the value 1. This metric is appropriate for the application with a threshold constraint on the information age.

Despite these contributions, it seems that the non-linear functions reported in existing works are chosen arbitrarily. Theoretical interpretation or formal justification for how to choose non-linear functions based on the structure of the data source and how they relate to the correlation of the underlying process are not provided.

2.2.3 Information-theoretic Metrics

To provide the theoretical basis of non-linear data freshness performance, probability theory, signal processing theory and information theory are widely employed to characterise the importance of fresh information.

2.2.3.1 Correlation Function

In probability theory, autocorrelation, autocovariance, and autocoefficient functions can be used to represent the correlation properties of a random process.

The autocorrelation of a random process is the Pearson correlation between values of the process at different time instances. It is a function of the two time instances, and represents the correlation of the data with a delayed copy of itself as a function of delay. The autocorrelation between two variables at times t_1 and t_2 is defined as

$$R_{XX}(t_1, t_2) = \mathrm{E}[X_{t_1}\overline{X_{t_2}}^*] \tag{2.7}$$

where $\overline{X}^*$ is the complex conjugation of X. The autocovariance function is defined by subtracting the mean of the autocorrelation function which is given as

$$K_{XX}(t_1, t_2) = \mathrm{E}[(X_{t_1} - \mu_1)\overline{(X_{t_2} - \mu_2)}^*] \tag{2.8}$$

where the mean of X_{t_1} and X_{t_2} are denoted as μ_1 and μ_2, respectively. In some practical cases, one may be interested in the normalised correlation function. The normalised autocovariance function is called the autocorrelation coefficient which is defined as

$$\xi_{XX}(t_1, t_2) = \frac{\mathrm{E}[(X_{t_1} - \mu_1)\overline{(X_{t_2} - \mu_2)}^*]}{\sigma_1\sigma_2} \tag{2.9}$$

where the variance of X_{t_1} and X_{t_2} are denoted as σ_1^2 and σ_2^2, respectively. The value of the autocorrelation coefficient ranges from -1 to 1. Particularly, $\xi_{XX}(t_1, t_2) = -1$

represents the perfect negative correlation and 1 represents perfect positive correlation. If the random process is stationary, its statistical properties do not change over time. In this case, the mean and variance do not change over time, and they can be denoted as μ and σ^2, respectively. Moreover, these correlation functions only relate to the time difference between two variables which can be given as

$$R_{XX}(\tau) = \mathrm{E}[X_{t+\tau}\overline{X_t}^*] \tag{2.10}$$

$$K_{XX}(\tau) = \mathrm{E}[X_{t+\tau}\overline{X_t}^*] - \mu\overline{\mu}^* \tag{2.11}$$

and

$$\xi_{XX}(\tau) = \frac{\mathrm{E}[(X_{t+\tau} - \mu)\overline{(X_t - \mu)}^*]}{\sigma^2}. \tag{2.12}$$

The autocorrelation function $K_{XX}(t, t-A(t))$ is used in [44,49] to measure the data freshness. When the underlying random process is stationary, the autocorrelation function can be presented by a function of the AoI, i.e., $K_{XX}(A(t))$. The authors in [49] investigate the age of the channel which is characterised by the autocorrelation function of the fading channel coefficients. The timely updating strategy (the optimal waiting time) for two correlated information sources is investigated in [50].

2.2.3.2 Estimation Error

The mean squared error is also extensively utilised to measure the information value [26–29, 51–61]. In an estimation system, the estimator aims to give the best guess $\hat{X}_t$ of the real but unobserved status X_t of a random process $\{X_t\}$ at time t based on the knowledge of past observations $X^k = \{X_1, \cdots, X_k\}$. The MSE measures the quality of the estimator. It is defined as the average of the squared difference between the estimator and the real value which can be written as follows

$$\mathrm{mse} = \mathrm{E}[(X_t - \hat{X}_t)^2]. \tag{2.13}$$

The relationship between the MSE and the AoI is investigated for various Markov processes, such as the Wiener process [28,51] and the Ornstein–Uhlenbeck process [29, 52–56]. The estimator is taken by minimising the MSE [62] and the authors show that the MSE can be presented by a non-linear AoI function for these two Markov processes. Various sampling policies are studied in [28, 29, 51, 52] to minimise the MSE. The authors prove that the MSE-optimal strategy is a threshold-based policy but not equivalent to the AoI-optimal strategy. In the context of the OU process model, the authors in [53, 54] further take the noise channel into consideration, in which the samples are not directly observable, to optimise the MSE. The impact of the sampling, quantisation and coding on the role of the AoI in remote estimation systems is explored in [55, 56].

The concept of the age of incorrect information (AoII) is introduced in [57], which incorporates the AoI and the traditional error-based indicator metrics together to measure the freshness of informative data. The AoII-based sampling problems are formulated as a Markov decision process and the optimal policies are obtained for a binary Markov source [26] and a semantics-based communication system [58], respectively.

A context-aware metric named the urgency of information (UoI) is proposed in [59–61], which can be utilised to evaluate the non-linear time-varying importance and the non-uniform context dependence of the correlated information in a status update system. Specifically, it is defined as the product of the weight parameter and the estimation error and it generalises the AoI and MSE under different conditions. If the weight is not time-varying, the UoI is equivalent to MSE; if the estimation error increases linearly with time and the status is not updated, the UoI is regarded as a non-linear AoI function.

2.2.3.3 Entropy

Information theory was introduced by Claude E. Shannon in 1948 [63], which is one of the most important frameworks for the study of communication. The major measure in information theory is *entropy*, which is widely utilised as a tool to quantify the amount of uncertainty contained in a random variable or a random process.

For a discrete random variable X, its entropy is defined as

$$\mathrm{H}(X) = -\mathrm{E}[\log p(x)] = -\sum_{x} p(x) \log p(x) \tag{2.14}$$

where $p(x)$ represents the probability function $\mathrm{P}(X = x)$. The units of the entropy can be "nats" when the natural logarithm is used and "bits" when the base of the logarithm is 2. For two discrete random variables X and Y, their joint entropy is defined as

$$\mathrm{H}(X, Y) = -\sum_{x} \sum_{y} p(x, y) \log p(x, y) \tag{2.15}$$

where $p(x, y)$ represents the joint probability function $\mathrm{P}(X = x, Y = y)$. Conditional entropy is defined as

$$\begin{aligned} \mathrm{H}(X|Y) &= -\sum_{y} \sum_{x} p(x, y) \log p(x|y) \\ &= -\sum_{y} p(y) \sum_{x} p(x|y) \log p(x|y). \end{aligned} \tag{2.16}$$

It is also presented by

$$\mathrm{H}(X|Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y). \tag{2.17}$$

The conditional entropy represents the uncertainty involved in a variable X given another variable Y. If X and Y are independent of each other, $\mathrm{H}(X|Y) = \mathrm{H}(X)$; if $X = Y$, we have $\mathrm{H}(X|Y) = 0$.

The entropy of a continuous random variable is called *differential entropy*. De-

note X as a continuous random variable and denote $f_X(x)$ as its probability density function. The differential entropy is defined as

$$\mathrm{h}(X) = -\mathrm{E}[\log f_X(x)] = -\int f_X(x) \log f_X(x)\, \mathrm{d}x. \tag{2.18}$$

Similarly, the joint and conditional differential entropy are defined as

$$\mathrm{h}(X, Y) = -\int\int f(x, y) \log f(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{2.19}$$

$$\mathrm{h}(X|Y) = -\int\int f(x, y) \log f(x|y)\, \mathrm{d}x\, \mathrm{d}y \tag{2.20}$$

where $f(x, y)$ is the joint density function of variables X and Y.

The entropy of a stochastic process is captured by the *entropy rate*, which represents the average level of uncertainty involved in one random variable in a random process. The entropy rate of a random process $\{X_n, n \in \mathbb{N}\}$ with countable index is defined as

$$\mathrm{H}(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}(X_n, \cdots, X_1). \tag{2.21}$$

The random process $\{X_n, n \in \mathbb{N}\}$ is said to be stationary when

$$\mathrm{P}(X_n|X_{n-1}) = \mathrm{P}(X_2|X_1). \tag{2.22}$$

In this case, the entropy rate of a stationary process can be simplified as

$$\mathrm{H}(\mathcal{X}) = \lim_{n\to\infty} \mathrm{H}(X_n|X_{n-1}, \cdots, X_1). \tag{2.23}$$

The random process $\{X_n, n \in \mathbb{N}\}$ is said to be Markov when

$$\mathrm{P}(X_n|X_{n-1}, \cdots, X_1) = \mathrm{P}(X_n|X_{n-1}). \tag{2.24}$$

Therefore, the entropy rate of a stationary Markov chain can be further simplified as

$$\mathrm{H}(\mathcal{X}) = \mathrm{H}(X_2|X_1). \tag{2.25}$$

Another key measure in information theory is *mutual information*, which quantifies the reduction in the uncertainty of a random variable given another random variable. The mutual information between variables X and Y is defined as

$$\mathrm{I}(X;Y) = \sum_{y}\sum_{x} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \tag{2.26}$$

The relationship between mutual information and entropy is given as follows

$$\begin{aligned}\mathrm{I}(X;Y) &= \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)\\ &= \mathrm{H}(X) - \mathrm{H}(X|Y)\\ &= \mathrm{H}(Y) - \mathrm{H}(Y|X).\end{aligned} \tag{2.27}$$

Entropy-based metrics are employed to measure the timeliness of information in status update systems. Since transferring a new version of information decreases the uncertainty in the knowledge of the monitor, conditional entropy is used as a measure of timely information in [64]. Mutual information is also used to measure how valuable the information is [31,65,66]. The authors in [31] consider a stationary Markov chain and prove that the mutual information can be presented by a non-negative and non-increasing function of the AoI. The mutual information based-metric is applied in satellite-integrated IoT networks and the optimal channel slot allocation scheme is obtained in [65]. Besides the time domain, the space domain is further taken into consideration for data freshness. A mutual information-based spatio-temporal model is proposed in [66], which decays with space and time, to quantify the usefulness of information and guide effective decision-making.

2.2.4 Value of Information

The notion of "value of information" is also used in the existing literature. Although the name of the concept is the same as the concept we propose in this thesis, their definitions are different.

In [30], conditional entropy is used to represent the timeliness of information and the value of information is defined as the maximum transmission cost that the observer would like to pay for decision-making. Moreover, the optimal information updating and scheduling problem is investigated in [67] with the aim of minimising the estimation error. In this work, the value of information is defined as a function which decreases with the AoI, and increases with the precision of the source. The authors show that the optimal policy is to schedule the packet with the largest current value of information. The analytic hierarchy process (AHP) is exploited in [68] to define the value of information in vehicular networks. The information value relates to the AoI and transmission distance which can be used to balance dissemination of the critical and non-critical data by altering the weights in AHP.

2.2.5 Other Metrics

A discrete concept of the age of version (AoV) is introduced in [69–71] for a cache updating system, in which each status update at the source is referred to as a "version" and the age at the receiver is the number of versions older than the most recent update from the source. The AoV measures the time difference between the current time and the time when the local copy is desynchronised with the remote source. An alternative metric named the age of synchronization (AoS) is proposed in [72]. It is defined as the time duration since the earliest time that the last refresh of the local copy is updated.

Two "effective age" metrics are proposed in [27, 73] to capture the impact of the sampling strategy on remote estimation for a Markov source. One is the concept of the sampling age which is defined as $A_{\text{samp}}(t) = s(t) - g(t)$. Given the time

instant t, $g(t)$ is the sampling time of the packet that yields the minimum estimation error, representing the most recent ideal sampling time. $s(t)$ is the sampling time of the first packet generated after $g(t)$. If no packet is generated, $s(t)$ is set to t. This metric measures the timeliness of the sampling at the data source. The other effective age metric is the concept of cumulative marginal error. It is defined as $A_{\mathrm{CME}}(t) = \int_{r(t)}^{t} h(\tau)\,\mathrm{d}\tau$, where $r(t)$ is the receiving time of the most recently received sample and $h(\cdot)$ is an AoI penalty function. This metric is appropriate to measure the effect of infrequent sampling on the AoI.

The works mentioned above all consider the "hard update" of the data packet, in which the information age (or the non-linear age) decreases sharply when a new status update is received. On the contrary, the "soft update" is proposed in [74,75], in which the information age decreases gradually when a new status update is received. The authors consider two soft update models. One is the exponential decay model in which the rate of decrease in the AoI follows $\frac{\mathrm{d}A(t)}{\mathrm{d}t} = -\alpha A(t)$ where α is a constant parameter. This model is suitable for the data source that changes rapidly with time. The other is the linear decay model in which the rate of change in the AoI is a constant, i.e., $\frac{\mathrm{d}A(t)}{\mathrm{d}t} = -\alpha$.

Besides the time-varying correlation properties, the spatial correlation properties can also affect the freshness of data. Therefore, the concept of spatial-temporal model is also widely investigated [76–78]. By modelling the location of the source and destination as a Poisson point process (PPP), the spatial distribution of the average AoI is observed in [77]. The effects of spatial configurations (caused by the path loss) and temporal dynamics on the AoI are investigated in large-scale wireless networks [78].

Despite these contributions, existing works largely focus on Markov models in which variables are assumed to be direct observable at the receiver, and the timeliness of the system only relates to the most recently received data sample. However, the

status updates generated at the source can be negatively affected by the noise during the transmission. This means they may be invisible from observation when they are delivered to the destination. Moreover, from the receiver's perspective, the observed process is not Markov anymore, thus the timeliness of data not only relates to the most recent single observation, but a sequence of past observations. Therefore, how to measure the data freshness with latent variables is a challenge in IoT-assisted status update systems.

2.3 Applications of Data Freshness Metrics

2.3.1 Age of Information in Queueing Models

After the AoI framework was formalised, a lot of early works focused on exploring AoI in different queueing systems and its statistical properties, aiming to provide theoretical contributions at the fundamental level.

2.3.1.1 Queueing Theory

Queuing theory focuses on exploring the dynamics of waiting lines or queues and how a system operates more efficiently through a queue. The behaviour of a queue is modelled as a birth-death process which describes the arrivals and departures to and from the queue. The model of a queue with a single server is shown in Fig. 2.3 with arrival rate λ and departure rate μ. Denote the number of jobs in a queue as k which includes the waiting jobs as well as the jobs being served. As we illustrated in Fig. 2.4, the system transits between different values of k by "births" and "deaths". Queueing systems are widely used to model the packet transmission process in the field of telecommunications.

There are four typical service disciplines.

- First come first serve (FCFS) discipline, in which all packets are served in the

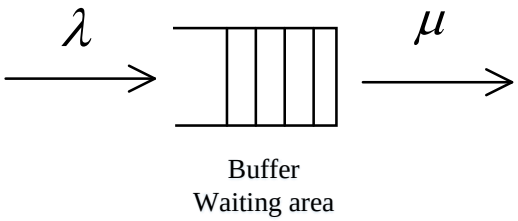


Figure 2.3: Model of a single queue.

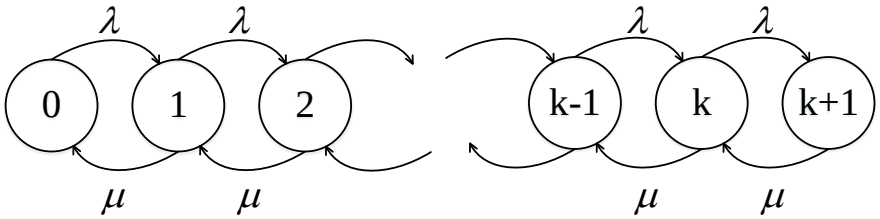


Figure 2.4: Model of a birth-death process.

sequence of their arrival.

- Preemptive last come first serve (LCFS), in which the newest packet has the highest priority. The newest packet starts to be served as soon as it arrives and any ongoing services will be interrupted immediately.

- Non-preemptive LCFS with discarding, in which the newest packet will be served until the ongoing service is completed. The other waiting packets in the buffer will be discarded.

- Non-preemptive LCFS without discarding. The only difference with the non-preemptive LCFS with discarding is that the other waiting packets will not be discarded, but remain in the buffer and wait to be served eventually.

For each discipline, there are various queue models which are described according to the probability distribution. A general queue model can be described by $X/Y/Z/A$. Here, X and Y are the distributions of the packet arrival interval and service time of each packet, respectively. Z is the number of servers. A is the buffer capacity which

is defaulted by infinity. Specifically, "M" represents the exponential distribution, "D" represents the fixed time, "Gamma" represents the Gamma distribution, "Ek" represents the k-th order Erlang distribution, "G" represents a general distribution, and "GI" represents a general distribution with independent and identically distributed (i.i.d.) random variables. The M/M/1 model is the most elementary of queueing model [79] in which the arrival process is a Poisson process (i.e., the arrival interval is exponentially distributed) and the service time is an exponential process. The closed-form expressions for many metrics based on this model have been obtained.

2.3.1.2 AoI in Queues

The average AoI is widely investigated with the FCFS discipline. It is shown that the AoI in M/D/1 and D/M/1 queue models outperform its counterpart in the M/M/1 model [9]. This means that fixed arrival or service time is able to reduce the average queue length thus further reducing the AoI. Fixing the service rate, the optimal arrival rate is also investigated for age minimisation. The M/M/1 queue model with packet replacement is treated in [41, 42, 80]. The performance of the average AoI and the peak AoI is studied in the context of the FCFS M/M/1 model with and without discarding. The update scheme is explored in the M/G/1/1 queue model [81], in which the authors show that discarding the new one policy is better than preempting the ongoing update in this case. The network with multiple servers is further considered in [82]. The effect of different message transmission routes on the AoI is studied for the M/M/2 and M/M/∞ models.

The LCFS discipline also receives much attention due to the fact that it can greatly reduce the information age. The preemptive and non-preemptive LCFS M/M/1 queue models are studied in [25]. The AoI in preemptive LCFS M/Gamma/1 and non-preemptive LCFS M/Ek/1 systems is explored in [83]. Furthermore, there are many works focusing on the performance analysis for the mean AoI in the queueing systems

with extra features, such as the queues with the packet deadline [84], link capacity constraint [85], constrained buffer size [86], packet priority [87], multiple class [88] and the presence of error [89].

Besides the average or the worst-case performance, the statistical distribution of the AoI is also widely explored by using queueing theory. A general formula for the stationary distribution of the AoI is first derived in [13, 90] which holds for various queueing systems with a single server. The distribution of the AoI is also looked at in the FCFS D/G/1 system [91], and the tail of the AoI distribution function is minimised with respect to the sampling rate constraint. The closed-form expression and the bounds of the AoI are derived in the G/G/1/1 [92], GI/GI/1/1 and GI/GI/1/2 queueing systems [93]. The closed-form expressions of the first and second moments of the AoI are obtained in [94, 95] with an energy harvester.

2.3.2 Age of Information in Internet of Things

In addition to the fundamental-level research works, the role of the AoI in the efficient design of freshness-aware applications also received much attention in various IoT-assisted status update systems [37]. The scheduling and optimisation problems are the most important problems in a wireless network. In the past, the relevant problems have been thoroughly studied with the aim of optimising the common traditional metrics (delay minimisation or throughput optimisation). The concept of AoI opens a new research direction.

In single-hop networks with a single transmitter–receiver pair, the optimal sampling problems at the data source are widely studied with different types of random processes [28, 29, 31] with and without sampling rate constraints. The authors show that the optimal sampling policy is a threshold-based policy in which the next sample is generated when the current AoI or the non-linear AoI exceeds a threshold, rather than the minimum wait policy. In single-hop networks with multiple transmitter–

receiver pairs, the user or link scheduling problems are studied to minimise the peak AoI of multiple data sources in [19, 96, 97]. A few works extend the AoI concept to multi-hop networks with multiple users. Age-optimal link scheduling problem is researched in [98] by fixing the link activation status according to a stationary probability distribution. In [20], a multi-hop cyber-physical system with multiple data flows is considered, and age minimisation is achieved by jointly optimising data sampling at the data sources and link scheduling along the path of each data flow. The scheduling problem falls in the form of the mixed integer non-linear programming which is difficult to be solved. Therefore, a heuristic algorithm with a polynomial time complexity is proposed to solve the problem efficiently. Moreover, the routing selection is first taken into consideration in [99] to study the optimal AoI and throughput tradeoff in multi-hop networks.

The above works focus on age-oriented data transmission in static networks, in which the transmission path for each data flow is either prefixed or not time-varying. However, in dynamic or mobile networks, the transmission link of each data flow changes over time. Against this background, there is a rising trend in applying the concept of AoI in various dynamic IoT-assisted applications, such us energy harvesting networks, UAV networks, vehicular networks, and so on. The timely status update and transmission are constrained by the changeable availability of energy at the transmitter in communication systems. How to reduce energy consumption to deliver status updates in a timely manner becomes an important issue. The minimisation of the average age under such considerations is studied in energy harvesting networks [18, 100]. The effect of the battery size of an energy harvester is further investigated in [21]. The AoI is also investigated and optimised in vehicular networks [101–103]. A joint resource allocation and sampling rate optimisation problem is formulated in [101] to maximise the link capacity under the AoI outage probability. A joint data collection and routing selection problem is formulated in [102]

to minimise the AoI, and an efficient algorithm based on reinforcement learning is proposed to solve the problem. It is shown that the reinforcement learning method is able to accommodate the dynamic network by exploring and interacting with the environment, which is helpful for decision-making [104, 105].

Due to the manoeuvrability, flexibility and low cost of UAVs [106], the AoI-oriented mobility and communication design problem receives much interest in UAV-enabled networks [23, 107–114]. The authors in [107] firstly explore two UAV path planning problems with the aim of minimising the maximum and average AoI, respectively. Dynamic programming and genetic algorithm methods are utilised to solve the proposed two problems. This work is further extended to a network where the sensors are densely deployed [108]. The sensors are divided into clusters with the aim of minimising the uploading time, and UAV hovers to collect the sensed data at the collection point of each cluster. The maximum AoI is minimised by jointly optimising ground user association and UAV trajectory planning policy. The authors in [109] consider a grid-based trajectory to optimise the AoI. This problem is formulated as a Markov decision process which is solved by deep reinforcement learning. Energy consumption has been further taken into consideration in the trajectory design, and AoI-energy-aware data collection schemes are proposed in [23, 110, 114]. The authors in [111] jointly optimise the AoI, moving time and user association for data acquisition in UAV-assisted sensor networks. Furthermore, the role of a UAV as a mobile relay node is investigated in [22] to minimise the average peak AoI for a source-destination pair by convex optimisation.

2.4 Summary

In this chapter, we provided the technical background and relevant literature review of this thesis. First, we showed the basic concept of IoT-assisted status update systems,

and explained why the timeliness of information is important in such networks. Next, we introduced a few performance metrics which can be used to measure the freshness of data, including the age of information and its variants. The advantages and disadvantages of each metric were discussed. The motivation to look for a more systematic metric was provided. Finally, the applications of the existing data freshness metrics were given. We reviewed the fundamental research on AoI in the context of queueing models. We also reviewed the application-based research on AoI in various emerging IoT-assisted networks.

Based on the overview and current research achievements, in the following chapters, we will propose a more appropriate measurement which can be used to quantify the usefulness of data. We will further analyse its performance, investigate its relationships with other existing metrics, and explore its application and optimisation problems.

Chapter 3

Value of Information Formulation

In this chapter, we formalise the value of information notion that forms the central study in the thesis and explore it in various latent variable models. The contributions of this chapter are given as follows:

- A VoI framework is formalised for the general latent variable model. The VoI is defined as the mutual information between the current status and a dynamic sequence of past observations, which gives a theoretical interpretation of the reduction in uncertainty in the current (unobserved) status of a latent process given that we have noisy measurements.
- The VoI is explored in the context of the most important latent variable model, the hidden Markov process, in which the underlying random process is Markov. The bound of the VoI is obtained.
- The VoI is further explored in the hidden Gauss-Markov process in which the underlying random process is Gauss-Markov. The closed-form expressions of the VoI are obtained. The connection between the VoI and estimation error is explored, which illuminates a useful relationship between information theory and signal processing theory and provides a possible way to infer the information value by utilising the MSE.

- Numerical results are provided to verify the theoretical analysis. The performance of VoI for Markov and hidden Markov models are presented. The effects of noise, number of observations, and correlation on VoI are discussed.

3.1 VoI for Latent Variable Models

We consider a general status update system with a pair of source and destination nodes. The source node continuously monitors a random process and generates a sequence of time-stamped messages representing updates on the status of the targeted process. These time-stamped data samples are required to be transmitted as timely as possible so that the destination node is able to hold the timely information of the underlying physical process. Although ideally, the destination node would receive a status update at the moment it is generated at the source, it is assumed that the communication system has limited resources, such that the message reaches the destination after some time.

As shown in Fig. 3.1, we denote $\{X_t\}$ as the random process under observation at the source node which is used to model the physical process in the practical world. Here, the time variable t can be either continuous or discrete. Denote (t_i, X_{t_i}) as the message which is generated at arbitrary time t_i, and contains the corresponding value X_{t_i} of the underlying random process. During a time period, the source node generates a sequence of data packets at times $t_1, t_2, \cdots, t_i, \cdots$, and i represents the packet index. From the receiver's perspective, the status update is received at time t_i'. We assume the transmission delay exists, thus we have $t_i' > t_i$. The observations at the receiver are recorded in the corresponding observed random process $\{Y_t\}$. $(t_i', Y_{t_i'})$ denotes the observation which is captured at time t_i' and contains the value $Y_{t_i'}$. We consider a practical communication system in which the data packets may be corrupted by the transmission noise, thus the exact value of the observation at the

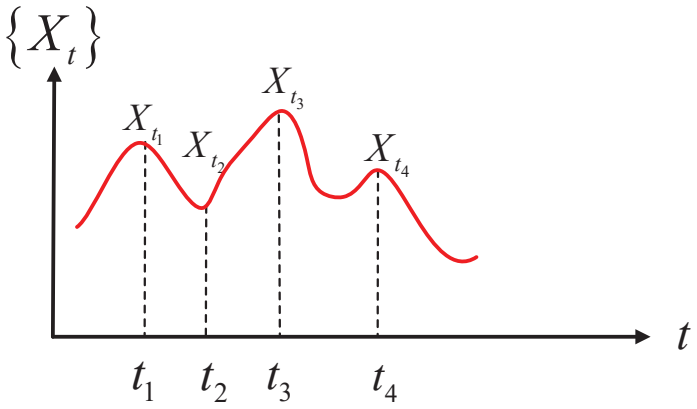


Figure 3.1: Sampling of a target random process.

receiver may be different from the exact value of random variable X_{t_i}, i.e., $Y_{t'_i} \neq X_{t_i}$ and $\{Y_t\} \neq \{X_t\}$. For the given time period $(0, t)$, we denote n as the index of the most recent data packet which is received at time t'_n with $t'_n < t \leq t'_{n+1}$.

In this scenario, the traditional age of information at a given time t can be written as the following based on (2.2)

$$A(t) = t - t_n, \quad t > t'_n \tag{3.1}$$

which is conditioned on the fixed times $\{t_i\}$ and $\{t'_i\}$. If there is no other new data sample observed by the receiver (i.e., $t < t'_{n+1}$), the AoI increases linearly with time t. Otherwise, the AoI will be updated accordingly.

The value of information is formulated in the context of the information theory. We define the VoI as the mutual information between the current status of the underlying random process at the source and a dynamic sequence of past observations captured by the receiver. For the given time instants, the general definition of VoI is given as

$$v(t) = \mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}), \quad t > t'_n, \tag{3.2}$$

which is conditioned on fixed times $\{t_i\}$ and $\{t'_i\}$. Here, n is the total number of

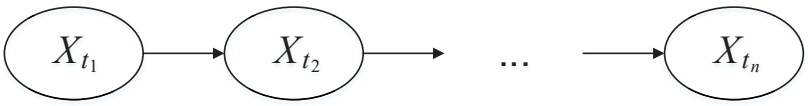


Figure 3.2: Temporal evolution of the Markov model.

recorded observations during the time period $(0, t)$ and it is also the index of the most recently received data sample. We look back in time and use a dynamic time window containing the most recent m of n samples $(1 \leq m \leq n)$ to measure the information value of the underlying hidden process. m is the number of observations we used which can be chosen based on the property of the random process model. For example, m can take 1 in the context of the Markov model. This definition is an average definition which is not conditioned on the exact value that samples and observations take. The VoI can be interpreted as the reduction in uncertainty in the current and unobserved status of a hidden process given that we have a sequence of noisy measurements.

Based on the chain rule for information [115], the VoI defined in (3.2) can also be written as

$$v(t) = \sum_{i=1}^{m} \mathrm{I}(X_t; Y_{t'_{n-m+i}} | Y_{t'_{n-m+i-1}}, \cdots, Y_{t'_{n-m+1}}). \tag{3.3}$$

The VoI for general latent variable models can be calculated by using the joint and marginal density functions of $\{X_{t_i}\}$ and $\{Y_{t'_i}\}$. In the following sections, we will explore the VoI in the context of two important latent variable models. If $\{X_t\}$ is Markov, the VoI is further manipulated in sec. 3.2. If $\{X_t\}$ is Gaussian and Markov, the VoI is manipulated in sec. 3.3.

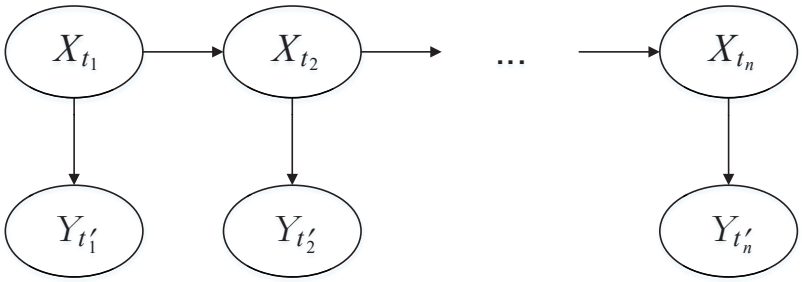


Figure 3.3: Temporal evolution of the hidden Markov model.

3.2 VoI for Hidden Markov Models

In the Markov model (Fig. 3.2), the random process $\{X_t\}$ is Markov and directly visible, and the observations are also Markov, i.e., $Y_{t'_i} = X_{t_i}$, for all $1 \leq i \leq n$. In this case, the VoI in the Markov model can be simplified to

$$v_{\mathrm{MM}}(t) = \mathrm{I}(X_t; X_{t_n}), \quad t > t'_n. \tag{3.4}$$

This result is obtained directly by

$$\begin{aligned} v(t) &= \mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \\ &= \mathrm{I}(X_t; X_{t_n}, \cdots, X_{t_{n-m+1}}) \\ &= \mathrm{h}(X_t) - \mathrm{h}(X_t | X_{t_n}, \cdots, X_{t_{n-m+1}}) \\ &= \mathrm{h}(X_t) - \mathrm{h}(X_t | X_{t_n}) \\ &= \mathrm{I}(X_t; X_{t_n}) \end{aligned} \tag{3.5}$$

It is easy to see that the VoI in the Markov model is independent of the number of observations m and only depends on the most recent single status update.

In the hidden Markov model (Fig. 3.3), the underlying random process $\{X_t\}$ is Markov and the observations at the receiver may be different from the exact value of

the initial random variable, i.e., $Y_{t'_i} \neq X_{t_i}$, but where

$$\mathrm{P}[Y_{t'_i} \in A | X_{t_1}, \ldots, X_{t_i}] = \mathrm{P}[Y_{t'_i} \in A | X_{t_i}] \tag{3.6}$$

for all admissible A. Hence, the initial samples $\{X_{t_i}\}$ are invisible at the receiver. In this case, we have

$$\mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \geq \mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+2}}), \tag{3.7}$$

and

$$v(t) \leq \mathrm{I}(X_t; X_{t_n}) \tag{3.8}$$

for $2 \leq m \leq n$. This result is obtained by

$$\begin{aligned} v(t) &= \mathrm{h}(X_t) - \mathrm{h}(X_t | Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \\ &\leq \mathrm{h}(X_t) - \mathrm{h}(X_t | Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}, X_{t_n}) \\ &= \mathrm{h}(X_t) - \mathrm{h}(X_t | X_{t_n}) \\ &= \mathrm{I}(X_t; X_{t_n}) \end{aligned} \tag{3.9}$$

Equation (3.7) shows that the VoI increases with the number of observations m. Moreover, (3.8) shows that the VoI in the Markov model can be regarded as the upper bound of the VoI in the hidden Markov model which illustrates that the lack of a direct route to observe $\{X_t\}$ reduces the information value.

The difference between the VoI in the Markov model and its counterpart in the

hidden Markov model can be expressed as

$$\begin{aligned}
&\mathrm{I}(X_t; X_{t_n}) - \mathrm{I}(X_t; Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \\
&= \mathrm{h}(X_t | Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) - \mathrm{h}(X_t | X_{t_n}) \\
&= \mathrm{h}(X_t | Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) - \mathrm{h}(X_t | X_{t_n}, Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}) \\
&= \mathrm{I}(X_t; X_{t_n} | Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}).
\end{aligned} \tag{3.10}$$

This reduction can be interpreted as the "correction" which captures the VoI gap due to the indirect observation in the hidden Markov model. In other words, we can think of the VoI for the hidden Markov model as the VoI for the Markov model minus the correction. The "correction" part can be quantified by the mutual information between the current status X_t and the most recent (unobserved) status update X_{t_n} conditioned on the knowledge of a sequence of past observations $\{Y_{t'_n}, \cdots, Y_{t'_{n-m+1}}\}$.

3.3 VoI for Hidden Gauss-Markov Models

The above part investigates the VoI in the general hidden Markov model. Many physical processes can be modelled by various Gauss-Markov processes, such as the discrete autoregressive process [116] and the continuous Ornstein–Uhlenbeck process [117]. In this part, we further look at the VoI in the context of the hidden Gauss-Markov process.

3.3.1 Noisy Gauss-Markov Model

The underlying hidden process $\{X_t\}$ is assumed to be a general Gaussian, Markov and stationary process. We assume that samples $\{X_{t_i}\}$ are Gaussian variables with constant mean $\theta = 0$ and constant variance σ_x^2. In this model, the correlation of two random variables only depends on their time difference due to the stationary property. Moreover, the autocorrelation coefficient which is given in (2.12) can be written as an

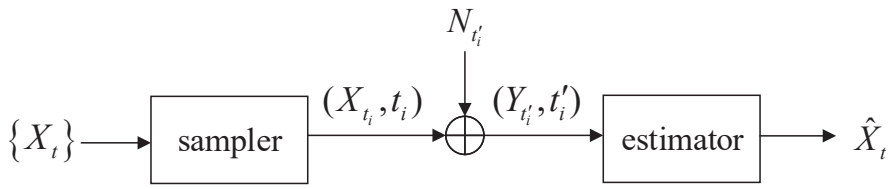


Figure 3.4: AWGN channel model.

exponential function. Therefore, the autocorrelation coefficient of two variables X_{t_i} and X_{t_j} can be written as ρ^τ where $\tau = |t_j - t_i|$ and ρ is a constant with $0 < \rho < 1$ (Theorem 6.8.2, [118]), i.e.,

$$\xi_{XX}(t_i, t_j) = \frac{\mathrm{E}[(X_{t_i} - \theta)(X_{t_j} - \theta)]}{\sigma_x^2} = \frac{\mathrm{E}[X_{t_i} X_{t_j}]}{\sigma_x^2} = \rho^\tau. \tag{3.11}$$

If X_{t_i} and X_{t_j} are less correlated, ρ^τ approaches 0; otherwise, it approaches 1. The correlation can be affected by either the inherent properties of the underlying data source (which depends on the correlation parameter ρ) or the time duration between the generation time of two samples (which depends on the time difference τ). Let vector $\boldsymbol{X} = [X_{t_n}, \cdots, X_{t_{n-m+1}}]^{\mathrm{T}}$ be the sequence of status updates of the underlying random process and let $\mathbf{K}_{\boldsymbol{X}}$ be its normalised autocovariance matrix, i.e.,

$$\mathbf{K}_{\boldsymbol{X}} = \begin{bmatrix} 1 & \rho^{t_n - t_{n-1}} & \cdots & \rho^{t_n - t_{n-m+1}} \\ \rho^{t_n - t_{n-1}} & 1 & \cdots & \rho^{t_{n-1} - t_{n-m+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{t_n - t_{n-m+1}} & \rho^{t_{n-1} - t_{n-m+1}} & \cdots & 1 \end{bmatrix}. \tag{3.12}$$

We assume that this random process is observed through an additive white Gaussian noise (AWGN) channel, which is shown in Fig. 3.4. We denote $N_{t_i'}$ as the noise sample induced by the receiver at t_i'. Then, the corresponding noisy measurement is defined as

$$Y_{t_i'} = X_{t_i} + N_{t_i'}. \tag{3.13}$$

The samples $\{N_{t_i'}\}$ are assumed to be i.i.d. Gaussian variables with zero mean and constant variance σ_n^2. Let γ be

$$\gamma = \frac{\sigma_x^2}{\sigma_n^2}. \tag{3.14}$$

The parameter γ gives a comparison between the randomness in the underlying process and the noise process in the communication channel, thus it can be compared to the concept of the signal-to-noise ratio in communication systems. Similarly, let vectors $\boldsymbol{Y} = [Y_{t_n'}, \cdots, Y_{t_{n-m+1}'}]^{\mathrm{T}}$ and associated $\boldsymbol{N} = [N_{t_n'}, \cdots, N_{t_{n-m+1}'}]^{\mathrm{T}}$ be the noisy observations and noise samples, respectively. Let $\boldsymbol{\Sigma_Y}$ be the autocovariance matrix of $\boldsymbol{Y}$. Thus, the observations and their autocovariance matrix can be collectively represented by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}, \quad \boldsymbol{\Sigma_Y} = \sigma_x^2\left(\mathbf{K_X} + \frac{1}{\gamma}\mathbf{I}\right) \tag{3.15}$$

where $\mathbf{I}$ is the identity matrix.

3.3.2 Value of Information

Let the m-dimensional vector $\boldsymbol{u}$ be given by

$$\boldsymbol{u} = \left[\rho^{t-t_n}, \rho^{t-t_{n-1}}, \cdots, \rho^{t-t_{n-m+1}}\right]^{\mathrm{T}}. \tag{3.16}$$

Based on the model described above, we can state the following main result.

Proposition 1. *The VoI for the noisy Gaussian Markov model is given as*

$$v(t) = -\frac{1}{2}\log\left[1 - \boldsymbol{u}^{\mathrm{T}}\left(\mathbf{K_X} + \frac{1}{\gamma}\mathbf{I}\right)^{-1}\boldsymbol{u}\right]. \tag{3.17}$$

Proof. See the Appendix in sec. 3.6.1. □

It is easy to see that the VoI relates to the following parts: the SNR in the transmission environment, which is captured by γ; the correlation of the underlying

data source, which is captured by ρ in matrix $\mathbf{K}_{\boldsymbol{X}}$; the age of information $A(t)$ in vector $\boldsymbol{u}$, which equals the time difference $t-t_n$; the number of observations m, which is the dimension of vector $\boldsymbol{u}$ and matrix $\mathbf{K}_{\boldsymbol{X}}$; the sampling and transmission time of each data packet.

Proposition 1 gives the general VoI expression for noisy Gauss-Markov models which is also the key result in this chapter. To explore this result further, we consider two special cases. First, we look at the non-noisy Gauss-Markov model to see how the VoI operates.

Corollary 1. *The VoI for the non-noisy Gaussian Markov model is given as*

$$v_{MM}(t) = -\frac{1}{2}\log\left[1-\rho^{2(t-t_n)}\right], \quad t > t'_n. \tag{3.18}$$

Proof. See the Appendix in sec. 3.6.2. □

In this case, the data samples are assumed to be directly observed without any corruption. The difference between the VoI in Proposition 1 and its Markov counterpart in Corollary 1 quantifies a "correction" to the VoI of the hidden process that arises due to the indirect observation of the process through the noisy channel, and it evaluates the result of (3.10) in the example of the Gauss-Markov model.

Next, we consider another special case where one may also wish to know how much value the most recently received observation (at time t'_n) contains about the status of a process at time t. In this case, the VoI can be calculated by replacing the m-dimensional vector $\boldsymbol{Y}$ with the single variable $Y_{t'_n}$ in (3.2), which leads to the following corollary.

Corollary 2. *The VoI for the noisy Gauss-Markov process with a single observation is given by*

$$v(t) = -\frac{1}{2}\log\left[1-\frac{\gamma}{1+\gamma}\rho^{2(t-t_n)}\right]. \tag{3.19}$$

Proof. This result follows directly from Proposition 1 where $m = 1$. □

This corollary shows that for fixed t_n, as time t increases, the VoI will decrease and the newly received update can cause a corresponding reset of $v(t)$. This is somewhat similar to the concept of AoI, which is equal to $t'_n - t_n$ at the moment the n-th update arrives and then increases with the unit slope until the next update comes. However, the VoI will decrease like $O(\rho^{2t})$ until a new status update is received. The parameter ρ can be used to represent how correlated the updates are. Compared with the traditional AoI metric, the proposed VoI framework not only reflects the time evolution of a random process, but also captures the correlation properties of the underlying data source and the noise in the transmission channel.

3.3.3 Connection to Mean-squared Error

The relationship between the AoI and the estimation error has been widely investigated in other works. In this section, we focus on general hidden Gauss-Markov models to investigate the relationship between the proposed VoI and the MSE. In [119], an important relationship between MSE and the input-output mutual information was established in the absence of transmission delay. In that work, data samples are treated individually and the data value of each sample does not change with time. In our work, we extend this result to status update systems in the presence of transmission delay.

As we show in Fig. 3.4, the estimator for remote estimation is defined by the conditional expectation given some noisy measurements, i.e., $\hat{X}_t = \mathrm{E}[X_t|\boldsymbol{Y} = \boldsymbol{y}]$, which is given by (Theorem 11.1, [62])

$$\hat{X}_t = \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{y}. \tag{3.20}$$

Based on (3.20), we can state the following result.

Lemma 1. *The MSE with multiple observations for the noisy Gaussian model is given by*

$$\mathrm{mse}(t) = \sigma_x^2 \left[1 - \boldsymbol{u}^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \boldsymbol{u} \right]. \tag{3.21}$$

Proof.

$$\begin{aligned} \mathrm{mse}(t) =& \mathrm{E}[(X_t - \hat{X}_t)^2] \\ =& \mathrm{E}[(X_t - \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{y})^2] \\ =& \mathrm{E}[X_t^2] + \sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}] \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} \\ & - 2\sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \mathrm{E}[X_t \boldsymbol{y}] \\ =& \sigma_x^2 + \sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{\Sigma}_{\mathbf{Y}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} - 2\sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} \\ =& \sigma_x^2 (1 - \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u}). \end{aligned} \tag{3.22}$$

□

The MSE in Lemma 1 shows similar behaviour to the VoI in Proposition 1. It also relates the SNR γ, the normalised autocovariance matrix $\mathbf{K}_{\boldsymbol{X}}$, and the vector $\boldsymbol{u}$. For the single observation case, it is easy to see that the MSE is a non-decreasing function of the AoI for Gauss-Markov models. This result holds for specific Gaussian processes (e.g., Wiener process [28] and OU process [29]).

Based on this lemma and Proposition 1, we can state the following main result regarding the VoI–MSE dependency.

Proposition 2. *For a fixed time instant t, the relationship between the VoI and the MSE for noisy Gaussian models satisfies*

$$v(t) = \frac{1}{2} \log \frac{\sigma_x^2}{\mathrm{mse}(t)}. \tag{3.23}$$

This proposition links the VoI to a logarithmic function of the MSE. In practice, if we think about a communication system and try to determine how valuable the status updates are, it may not be easy to estimate the mutual information but it is

easy to estimate the MSE. Therefore, by estimating the MSE and using this link, it is possible to infer what the value of information is. The VoI is defined based on mutual information, and it has the potential to be used as the objective function for effective data collection in robotic networks [120].

Proposition 2 illustrates the relationship between the VoI and the MSE with multiple observations. In the following, we consider the single observation case to explore this result further. The VoI and MSE can be calculated by replacing the m-dimensional vector $\boldsymbol{u}$ and matrix $\mathbf{K}_{\boldsymbol{X}}$ with the single variable ρ^{t-t_n} and 1, respectively. Then, we have the following corollaries.

Corollary 3. *If the latest sample X_{t_n} and the current status X_t of the underlying random process are highly correlated (i.e., $\rho^{t-t_n} \to 1$), the relationship between the VoI and MSE can be written as*

$$\frac{\mathrm{d}v}{\mathrm{d}\gamma} = \frac{\mathrm{mse}}{2\sigma_x^2}. \tag{3.24}$$

If the latest sample X_{t_n} and the current status X_t of the underlying random process are uncorrelated (i.e., $\rho^{t-t_n} \to 0$), we have

$$\mathrm{mse} \to \sigma_x^2, \quad \frac{\mathrm{d}v}{\mathrm{d}\gamma} \to 0. \tag{3.25}$$

In this corollary, (3.24) links the derivative of the VoI with respect to the SNR to the MSE for highly correlated data samples. This relationship is also given in [119], which links the derivative of the input-output mutual information to the MSE in the absence of transmission delay and holds for both Gaussian and non-Gaussian signals. Motivated by that work, we study their relationship in Gaussian models in the presence of transmission delay. This corollary shows that a high correlation condition means that either the AoI is very small (i.e., $t - t_n \to 0$) or the underlying data source does not change quickly with time (e.g., the random process $\{X_t\}$ is a constant). For fewer correlated data samples, (3.25) shows that the MSE is independent of the SNR.

Moreover, a low correlation condition yields a zero rate of change of the VoI with increasing SNR.

Corollary 4. *The derivative of the VoI with respect to the AoI can be written as*

$$\frac{\partial v}{\partial A} = -\frac{1}{2\,\mathrm{mse}}\frac{\partial\,\mathrm{mse}}{\partial A}. \tag{3.26}$$

This corollary shows that the rate of change of the VoI is not equivalent to the rate of change of MSE when increasing the AoI. When the estimation error is large, the rate of change of the VoI is 0, and it does not relate to the AoI.

3.4 Numerical Results

In this section, numerical results are provided. We compare the VoI performance in both Markov and hidden Markov models. The results also illustrate the effects of the number of observations, the correlation and the noise on the VoI performance. The performance of the VoI is verified by Monte Carlo simulations. In the simulation, we consider the FCFS queueing and transmission model, and the service time of each status update is generated randomly according to a rate $\mu = 1$ exponential distribution. For uniform sampling times, the fixed sampling interval is set to $\Delta t = 2$. For arbitrary sampling times, the sampling process is chosen to be a rate $\lambda = 0.5$ Poisson process. These instants are generated for 10^5 times and we take their average as the service time and sampling intervals when plotting the figures.

Fig. 3.5 shows the VoI in Markov and hidden Markov models for different observation numbers m. In this figure, all the received observations are used for the result labelled "$m = n$" which is obtained by Proposition 1. Only the most recently received single observation is used for the result labelled "$m = 1$" which is obtained by Corollary 2. The black curve represents the VoI in the underlying Markov model, which represents the result given in Corollary 1. The VoI decreases with time until a

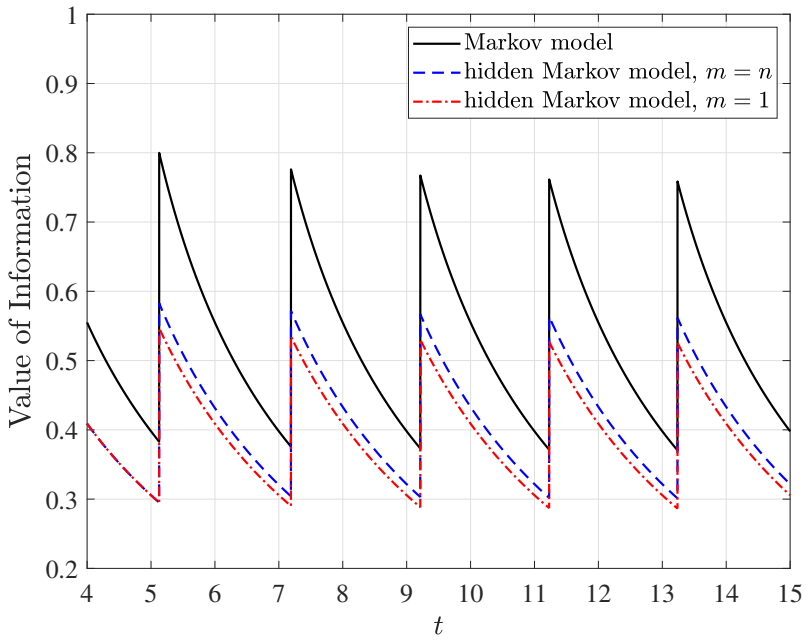


Figure 3.5: Time evolution of VoI in Markov and hidden Markov process models; correlation parameter $\rho = 0.8$, $\sigma_x^2 = 5$, noise parameter $\sigma_n^2 = 1$, and sampling interval $\Delta t = 2$.

new update is transmitted, which shows a behaviour that is similar to the traditional AoI evolution. The gap between the result in the Markov model and its counterpart in the hidden Markov model represents the "VoI correction" due to the indirect observation. The number of observations can affect the VoI in the hidden Markov model. However, the VoI in the Markov model does not depend on m.

Fig. 3.6 further shows how the VoI varies with the number of observations m for different values of noise σ_n^2. The horizontal axis represents the number of observations we used to predict the value of the current status of the random process. The vertical axis is the normalised VoI, which represents the ratio of $v(t)$ to $v_{\mathrm{MM}}(t)$, where $v_{\mathrm{MM}}(t)$ is the VoI in the underlying Markov process. This result shows that the VoI in the noisy Gauss-Markov process increases with the number of observations, and converges as more past observations are used. This can be explained as more past observations can give more information about the current status of the latent random process.

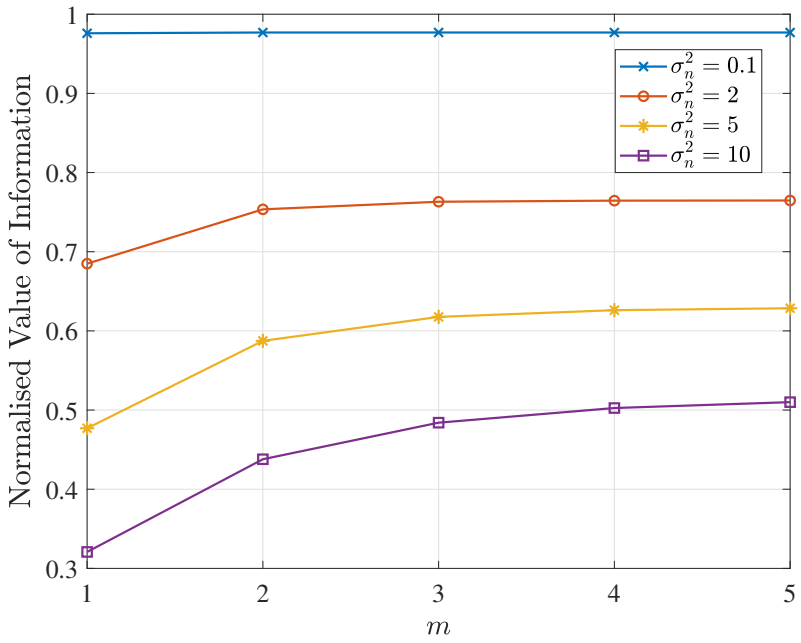


Figure 3.6: VoI in the hidden Markov model versus the number of observations m for $\sigma_n^2 \in \{0.1, 2, 5, 10\}$ at $t = 100$; correlation parameter $\rho = 0.9$, $\sigma_x^2 = 10$, and sampling interval $\Delta t = 2$.

Moreover, the normalised VoI approaches 1 for small σ_n^2, which means that the VoI in the Markov model can be regarded as the upper bound of its counterpart in the hidden Markov model, verifying the result given in (3.8).

Fig. 3.7 shows how the VoI varies with the autocorrelation parameter ρ. Large ρ represents the highly-correlated data source and small ρ means weak correlation. Regardless of the SNR and the number of observations, the VoI always increases with the correlation parameter. As ρ rises, the past data samples will be more dependent and contain more valuable information about the underlying random process, which further yields higher VoI. This figure illustrates that the value of highly correlated samples is larger, compared with the less correlated samples, which shows that "old" samples from the highly correlated source may still offer value.

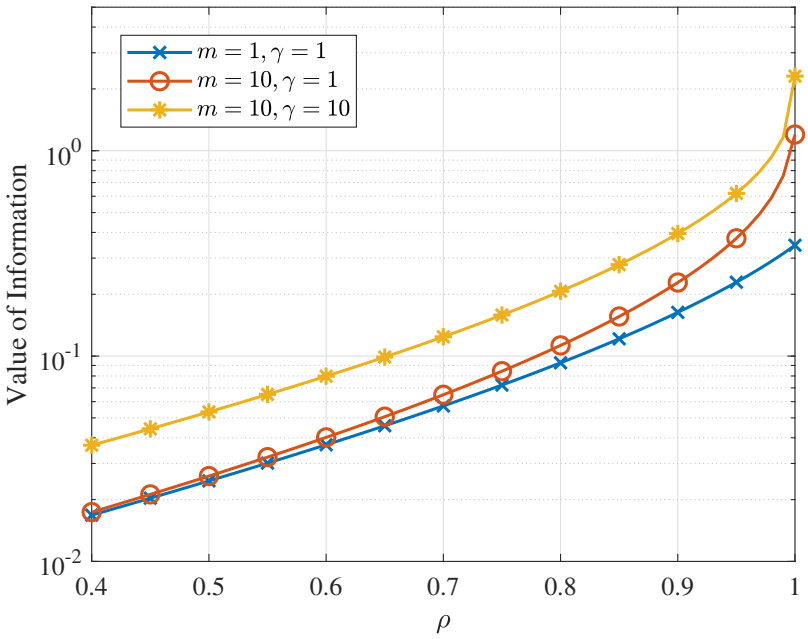


Figure 3.7: VoI versus the correlation parameter for $m \in \{1, 10\}$ and $\gamma \in \{1, 10\}$; sampling rate $\lambda = 0.5$.

3.5 Summary

In this chapter, we proposed a mutual information-based value of information framework which can be used to characterise the usefulness of data for general latent variable models. Next, we looked at the VoI in the context of hidden Markov models and hidden Gauss-Markov models, deriving its closed-form expression and the bound. Then, the functional relationship between the VoI and MSE was investigated. We also studied the rate of change of the two with respect to the SNR and the AoI. The results illuminated interesting relationships between the timeliness of information and remote estimation in an information-theoretic context, which also provided a possible way to infer the information value by utilising the estimation error. Finally, the effects of the noise, the number of observations and the correlation on the VoI were investigated through simulations. In the next chapter, we will focus on general hidden Gauss-Markov models to further explore the effects of sampling and the number of

observations on the VoI.

3.6 Appendix

3.6.1 Proof of Proposition 1

The VoI definition given in (3.2) follows from the relation $\mathrm{I}(X_t;\boldsymbol{Y}) = \mathrm{h}(X_t) + \mathrm{h}(\boldsymbol{Y}) - \mathrm{h}(X_t,\boldsymbol{Y})$. Since $(X_t,\boldsymbol{Y}^{\mathrm{T}})$ is multivariate Gaussian and stationary distributed, the VoI can be further written as (Example 8.5.1, [115])

$$v(t) = \frac{1}{2}\log\frac{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}})}{\det(\boldsymbol{\Sigma}_{X_t,\mathbf{Y}})} \tag{3.27}$$

where $\boldsymbol{\Sigma}_{X_t,\mathbf{Y}}$ is the covariance matrix of $(X_t,\boldsymbol{Y}^{\mathrm{T}})^{\mathrm{T}}$. The covariance between X_t and $Y_{t_i'}$ is given as

$$\mathrm{E}[X_tY_{t_i'}] = \mathrm{E}[X_t(X_{t_i}+N_{t_i'})] = \mathrm{E}[X_tX_{t_i}] = \sigma_x^2\rho^{t-t_i}. \tag{3.28}$$

Therefore, $\boldsymbol{\Sigma}_{X_t,\mathbf{Y}}$ can be presented by $\boldsymbol{\Sigma}_{\mathbf{Y}}$, i.e.,

$$\boldsymbol{\Sigma}_{X_t,\mathbf{Y}} = \begin{bmatrix} \sigma_x^2 & \sigma_x^2\boldsymbol{u}^{\mathrm{T}} \\ \sigma_x^2\boldsymbol{u} & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{bmatrix}. \tag{3.29}$$

Substituting (3.29) into (3.27), the VoI is given as

$$\begin{aligned} v(t) &\overset{(a)}{=} \frac{1}{2}\log\frac{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}})}{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}}-\sigma_x^2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})} \\ &\overset{(b)}{=} \frac{1}{2}\log\frac{1}{1-\sigma_x^2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{u}} \\ &= -\frac{1}{2}\log\left(1-\sigma_x^2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{u}\right) \end{aligned} \tag{3.30}$$

where (a) is obtained by the block matrix determinant and (b) is obtained by the determinant lemma.

3.6.2 Proof of Corollary 1

The VoI for the non-noisy Markov model in (3.4) follows from the relation $\mathrm{I}(X_t; X_{t_n}) = \mathrm{h}(X_t) + \mathrm{h}(X_{t_n}) - \mathrm{h}(X_t, X_{t_n})$. Similar to Proposition 1, the VoI can be written as

$$\begin{aligned} v(t) &= \frac{1}{2}\log\frac{\sigma_x^4}{\det(\mathbf{\Sigma}_{X_t,X_{t_n}})} \\ &= \frac{1}{2}\log\frac{\sigma_x^4}{\det\begin{pmatrix} \sigma_x^2 & \sigma_x^2\rho^{t-t_n} \\ \sigma_x^2\rho^{t-t_n} & \sigma_x^2 \end{pmatrix}} \\ &= -\frac{1}{2}\log\Big[1-\rho^{2(t-t_n)}\Big]. \end{aligned} \tag{3.31}$$

Chapter 4

Analysis of Hidden Gauss-Markov Process Models

In Chapter 3, we derived the closed-form VoI expression in general hidden Gauss-Markov models in Proposition 1. In this chapter, we will utilise this key result to give a detailed analysis of the VoI in the context of hidden Gauss-Markov models. Specifically, we will go into more depth with regard to different sampling policies and the effect of the number of observations. The contributions of this chapter are given as follows:

- The VoI with different sampling policies is investigated. Simplified VoI expressions are derived in both large and small SNR regimes for arbitrary and uniform sampling times.
- We study the rate of change of the VoI when generating more data samples and increasing the SNR. We also find a trade-off between the SNR and the number of observations, and scale back SNR to achieve the same VoI improvement by adding one extra observation.
- The optimal sampling problem is formulated to maximise the average VoI under

the sampling rate constraints. A threshold-based VoI-optimal sampling policy is obtained.

- Numerical results are provided to illustrate how the sampling policy, number of observations and SNR affect the performance of the VoI.

4.1 Effective Sampling Policies for Value of Information

The closed-form expression of the VoI given in Proposition 1 involves a large dimensional vector and metric which is hard to manipulate in the practical network. Meanwhile, the VoI with multiple observations is closely related to the sampling times of the status updates and the SNR. Therefore, it is important to take a view of different sampling policies and study how the sampling time affects the VoI. In this section, we consider two sampling cases and aim to simplify the VoI under different SNR conditions. We start with an arbitrary but fixed sampling policy, and then we move to a uniform sampling policy.

4.1.1 Arbitrary Fixed Sampling Intervals

We assume that status updates are generated at arbitrary but fixed times $\{t_i\}$, and denote T_i as the sampling interval of two packets, i.e.,

$$T_i = t_{n+1-i} - t_{n-i}, \quad 1 \leq i \leq m-1. \tag{4.1}$$

In this case, the general VoI expression with multiple observations is given by Proposition 1:

$$v(t) = -\frac{1}{2}\log\left[1 - \boldsymbol{u}^{\mathrm{T}}\left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma}\mathbf{I}\right)^{-1}\boldsymbol{u}\right] \tag{4.2}$$

where the vector $\boldsymbol{u}$ is

$$\boldsymbol{u} = \rho^{t-t_n}\left[1, \rho^{T_1}, \cdots, \rho^{\sum_{i=1}^{m-1} T_i}\right]^{\mathrm{T}}. \tag{4.3}$$

$\mathbf{K}_{\boldsymbol{X}}$ is the normalised autocovariance matrix of $\boldsymbol{X}$ which can be written as

$$\mathbf{K}_{\boldsymbol{X}} = \begin{bmatrix} 1 & \rho^{T_1} & \cdots & \rho^{\sum_{i=1}^{m-1} T_i} \\ \rho^{T_1} & 1 & \cdots & \rho^{\sum_{i=2}^{m-1} T_i} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{\sum_{i=1}^{m-1} T_i} & \rho^{\sum_{i=2}^{m-1} T_i} & \cdots & 1 \end{bmatrix}. \tag{4.4}$$

The inverse of the normalised autocovariance matrix of $\boldsymbol{X}$ is tridiagonal, which can be written as [121, 122]

$$\mathbf{K}_{\boldsymbol{X}}^{-1} = \begin{bmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & a_{m-1} & b_{m-1} \\ & & & b_{m-1} & a_m \end{bmatrix}. \tag{4.5}$$

Here,

$$a_i = \begin{cases} \frac{1}{1-\rho^{2T_1}} & i = 1 \\ \frac{1}{1-\rho^{2T_{m-1}}} & i = m \\ \frac{1}{1-\rho^{2T_i-1}} + \frac{1}{1-\rho^{2T_i}} - 1 & \text{others} \end{cases} \tag{4.6}$$

and

$$b_i = -\frac{\rho^{T_i}}{1-\rho^{2T_i}}, \quad 1 \le i \le m-1. \tag{4.7}$$

Thus, we have a clear and general presentation of the matrix $\mathbf{K}_{\boldsymbol{X}}$ for arbitrary sampling intervals. The SNR parameter γ can significantly affect the VoI in the hidden Markov model. If γ is large, the underlying latent process is dominant; otherwise, the

noise process is dominant. Therefore, it is interesting to explore the VoI expressions with different levels of noise.

4.1.1.1 High SNR Regime

First, we consider the high SNR regime in which the small variance of noise leads to large γ. This case is suitable to model the communication system which is less likely to be affected by noise. The series expansion is employed at $\frac{1}{\gamma} \to 0$ to study the VoI in the high SNR regime and we can state the following result.

Corollary 5. *For the hidden Markov process with arbitrary but fixed sampling times, the VoI in the high SNR regime can be written as*

$$v(t) = -\frac{1}{2}\log\left[1-\rho^{2(t-t_n)}\left(1-\frac{1}{\gamma}+\frac{1}{1-\rho^{2(t_n-t_{n-1})}}\frac{1}{\gamma^2}\right)\right]+O(\frac{1}{\gamma^3}). \tag{4.8}$$

Proof. Let function $g(x)$ be

$$g(x) = \boldsymbol{u}^{\mathrm{T}}\Big(\mathbf{K}_{\boldsymbol{X}}+x\mathbf{I}\Big)^{-1}\boldsymbol{u}. \tag{4.9}$$

Since we have

$$\frac{\partial\Big(\mathbf{K}_{\boldsymbol{X}}+x\mathbf{I}\Big)^{-1}}{\partial x} = -\left[\Big(\mathbf{K}_{\boldsymbol{X}}+x\mathbf{I}\Big)^{-1}\right]^2, \tag{4.10}$$

the k-th order derivative of $g(x)$ with respect to x can be given as

$$\frac{\partial^k g}{\partial x^k} = (-1)^k\prod_{i=1}^{k} i\boldsymbol{u}^{\mathrm{T}}\left[\Big(\mathbf{K}_{\boldsymbol{X}}+x\mathbf{I}\Big)^{-1}\right]^{k+1}\boldsymbol{u}. \tag{4.11}$$

Since

$$\boldsymbol{u}^{\mathrm{T}}\mathbf{K}_{\boldsymbol{X}}^{-1} = \rho^{t-t_n}[1,0,\cdots,0] \tag{4.12}$$

the function $g(x)$ and its derivatives at $x = 0$ can be given as

$$g(0) = \rho^{2(t-t_n)} \tag{4.13}$$

$$g'(0) = -\rho^{2(t-t_n)} \tag{4.14}$$

and

$$g''(0) = 2\frac{\rho^{2(t-t_n)}}{1-\rho^{2T_1}}. \tag{4.15}$$

Therefore, the function g can be presented by the following series expansion

$$g(x) = \rho^{2(t-t_n)} - \rho^{2(t-t_n)}x + \frac{\rho^{2(t-t_n)}}{1-\rho^{2T_1}}x^2 + O(x^3). \tag{4.16}$$

The result of this corollary is obtained by substituting (4.16) into (4.2). □

The expression of $v(t)$ is conditioned on the past time instants t_n and t_{n-1}, and the VoI does not depend on m (the number of samples). This is because when γ is large, the Markov randomness is dominant, and the noisy channel is not expected to play a vital role in the calculation of VoI. Therefore, the VoI in the high SNR regime approaches its Markov counterpart, which is not related to m.

Moreover, if the VoI given in (4.8) is truncated to the second-order term $\frac{1}{\gamma^2}$, the approximated VoI will first decrease and then increase with $\frac{1}{\gamma}$. The turning point is at $\frac{1}{\gamma} = \frac{1}{2}(1-\rho^{2(t_n-t_{n-1})})$. Therefore, the valid region of the approximated VoI is

$$\gamma \geq \frac{2}{1-\rho^{2(t_n-t_{n-1})}}. \tag{4.17}$$

4.1.1.2 Low SNR Regime

Next, we study the VoI expression in the low SNR regime (i.e., $\gamma \to 0$). This case can be used to model the communication system which experiences poor channel conditions. We can state the following result.

Corollary 6. *For the hidden Markov process with arbitrary but fixed sampling times, the VoI in the low SNR regime can be written as*

$$v(t) = -\frac{1}{2}\log\left[1 - \rho^{2(t-t_n)}\left(1 + \sum_{j=1}^{m-1} \rho^{2\sum_{i=1}^{j} T_i}\right)\gamma\right] + O(\gamma^2). \tag{4.18}$$

Proof. In this case, we let function $g(x)$ be

$$g(x) = x\boldsymbol{u}^{\mathrm{T}}\left(x\mathbf{K}_{\boldsymbol{X}} + \mathbf{I}\right)^{-1}\boldsymbol{u} \tag{4.19}$$

in which we have

$$g(0) = 0. \tag{4.20}$$

Its first-order derivative is given as

$$g'(0) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = \rho^{2(t-t_n)}\left(1 + \sum_{j=1}^{m-1} \rho^{2\sum_{i=1}^{j} T_i}\right). \tag{4.21}$$

In the low SNR regime, the function g is presented by the following series expansion

$$g(x) = -\rho^{2(t-t_n)}\left(1 + \sum_{j=1}^{m-1} \rho^{2\sum_{i=1}^{j} T_i}\right)x + O(x^2). \tag{4.22}$$

The result of this corollary is obtained by substituting (4.22) into (4.2). □

The VoI in the low SNR regime is conditioned on all past time instants. The randomness of the noise dominates in the low SNR regime, thus the VoI relates to the number of observations m and increases as m grows larger. Compared with Corollary 5, the VoI in the low SNR regime can be improved by using more past observations, while the number of observations m is not able to help to improve the VoI in the high SNR regime.

4.1.2 Uniform Sampling Intervals

Corollaries 5 and 6 give general results for the VoI with arbitrary sampling times in high and low SNR regimes, respectively. In some cases, status updates are generated at a fixed rate. It may still be of interest to have a clear representation of VoI with uniform sampling intervals. In this subsection, we consider this specific case in which status updates are sampled at regular times.

Let the sampling time $t_i = i\Delta t$ where the constant Δt ($\Delta t > 0$) denotes the fixed sampling interval for all $1 \leq i \leq n$. In this case, we are able to derive the closed-form expression of the VoI given in Proposition 1 when sampling times are regular. Let the m-dimensional matrix $\mathbf{A} = \frac{1}{\gamma}\mathbf{K}_{\boldsymbol{X}}^{-1} + \mathbf{I}$, and denote $\mathbf{A}_{ij}$ as the $(m-1) \times (m-1)$ matrix constructed by removing the i-th row and the j-th column of the matrix $\mathbf{A}$. The VoI in Proposition 1 can be rewritten as

$$v(t) = -\frac{1}{2}\log\left[1 - \rho^{2(t-t_n)}\left(1 - \frac{\det(\mathbf{A}_{11})}{\gamma\det(\mathbf{A})}\right)\right]. \tag{4.23}$$

The closed-form expression of the determinant ratio can be given as

$$\frac{\det(\mathbf{A}_{11})}{\gamma\det(\mathbf{A})} = \frac{1-\rho^2}{\rho} \cdot \frac{ac(\lambda_1^{m-2} - \lambda_2^{m-2}) + (ab+bc)(\lambda_1^{m-3} - \lambda_2^{m-3}) + b^2(\lambda_1^{m-4} - \lambda_2^{m-4})}{a^2(\lambda_1^{m-1} - \lambda_2^{m-1}) + 2ab(\lambda_1^{m-2} - \lambda_2^{m-2}) + b^2(\lambda_1^{m-3} - \lambda_2^{m-3})} \tag{4.24}$$

where

$$\lambda_1 = \frac{-c + \sqrt{c^2 - 4b^2}}{2b}, \quad \lambda_2 = \frac{-c - \sqrt{c^2 - 4b^2}}{2b} \tag{4.25}$$

and

$$a = \frac{1}{\gamma(1-\rho^2)} + 1, \quad b = \frac{-\rho}{\gamma(1-\rho^2)}, \quad c = \frac{1+\rho^2}{\gamma(1-\rho^2)} + 1. \tag{4.26}$$

The proof of this result is given in the appendix in sec. 4.6.1. This means that we can obtain the general closed-form expression of the VoI given in Proposition 1 when sampling times are regular.

In addition to the closed-form expression, we are also able to obtain approximated VoI expressions in the high and low SNR regimes by substituting $T_i = \Delta t$ in Corollaries 5 and 6. We have the following corollaries.

Corollary 7. *For the hidden Markov process with uniform sampling intervals, the VoI in the high SNR regime can be written as*

$$v(t) = -\frac{1}{2}\log\left[1 - \rho^{2(t-t_n)}\left(1 - \frac{1}{\gamma} + \frac{1}{1-\rho^{2\Delta t}}\frac{1}{\gamma^2}\right)\right] + O(\frac{1}{\gamma^3}). \tag{4.27}$$

The valid region of the approximated VoI is

$$\gamma \geq \frac{2}{1-\rho^2}. \tag{4.28}$$

Corollary 8. *For the hidden Markov process with uniform sampling intervals, the VoI in the low SNR regime can be written as*

$$v(t) = -\frac{1}{2}\log\left[1 - \rho^{2(t-t_n)}\frac{1-\rho^{2m\Delta t}}{1-\rho^{2\Delta t}}\gamma\right] + O(\gamma^2). \tag{4.29}$$

For uniform sampling intervals, these corollaries further illustrate that the number of observations has a very small effect on the VoI in the high SNR regime. While in the low SNR regime, m has a larger effect. As $0 < \rho < 1$, the VoI converges to

$$v(t) \to -\frac{1}{2}\log\left[1 - \frac{\rho^{2(t-t_n)}}{1-\rho^{2\Delta t}}\gamma\right] \tag{4.30}$$

when m grows to infinity in the low SNR regime.

4.2 Trading off SNR and the Number of Observations to Improve the Value of Information

The above section shows that using more past observations and increasing the SNR can both improve the VoI. In this section, we will further explore their trade-off on VoI performance.

The VoI is largely affected by the number of past observations m and the SNR γ. Fixing the time instant t, the VoI can be presented by a function of m and γ and it is rewritten as

$$v(m,\gamma) = -\frac{1}{2}\log\left[1 - \boldsymbol{u}_m^{\mathrm{T}}\left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma}\mathbf{I}_m\right)^{-1}\boldsymbol{u}_m\right] \tag{4.31}$$

where vector $\boldsymbol{u}_m$ and matrices $\mathbf{K}_{\boldsymbol{X}_m}, \mathbf{I}_m$ have the same format as they are in Proposition 1 and m represents their dimension. For simplicity, we denote η as the absolute value of the exponential VoI difference, i.e.,

$$\eta = \left|e^{-2v(m,\gamma)} - e^{-2v(m',\gamma')}\right|. \tag{4.32}$$

η is a non-negative function of m, m', γ and γ'. It can be used to represent the growth of the amount of valuable information by using more past observations or increasing the SNR. We consider the following two cases to further explore the VoI:

- For a fixed SNR γ, the change of the VoI when increasing m by one, which is denoted as η_m;
- For a fixed number of observations m, the change of the VoI when increasing a certain value of the SNR, which is denoted as η_γ.

Then, we can state the following two propositions.

Proposition 3. *For a fixed SNR γ, the exponential VoI difference over the number of observations ($m' = m + 1$) can be given by*

$$\eta_m = \delta \left[\boldsymbol{u}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma} \mathbf{I}_m \right)^{-1} \boldsymbol{w}_m - \rho^{t - t_{n-m}} \right]^2. \tag{4.33}$$

where vector $\boldsymbol{w}_m$ is given by

$$\boldsymbol{w}_m = \left[\rho^{t_n - t_{n-m}}, \rho^{t_{n-1} - t_{n-m}}, \cdots, \rho^{t_{n-m+1} - t_{n-m}} \right]^{\mathrm{T}}. \tag{4.34}$$

δ is given by

$$\delta = \left[1 + \frac{1}{\gamma} - \boldsymbol{w}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma} \mathbf{I}_m \right)^{-1} \boldsymbol{w}_m \right]^{-1}. \tag{4.35}$$

Proof. See the Appendix in sec. 4.6.2. □

This proposition shows the improvement of the VoI when increasing the number of observations by 1 for a fixed SNR, i.e., $\eta_m = e^{-2v(m,\gamma)} - e^{-2v(m+1,\gamma)}$. $\rho^{t - t_{n-m}}$ captures the correlation between the current status X_t and the $(m+1)$-th measurement $Y_{t'_{n-m}}$.

Proposition 4. *For a fixed m, the exponential VoI difference over the SNR can be given by*

$$\eta_\gamma = \boldsymbol{u}_m^{\mathrm{T}} \left[\left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma'} \mathbf{I}_m \right)^{-1} - \left(\mathbf{K}_{\boldsymbol{X}_m} + \frac{1}{\gamma} \mathbf{I}_m \right)^{-1} \right] \boldsymbol{u}_m. \tag{4.36}$$

Proposition 4 shows the improvement of the VoI when increasing the SNR from γ to γ' for a fixed m, i.e., $\eta_\gamma = e^{-2v(m,\gamma)} - e^{-2v(m,\gamma')}$ where $\gamma' > \gamma$.

Propositions 3 and 4 give a clear mathematical representation of how the VoI changes with the number of observations and the SNR. It is not easy to give insight directly from the above two propositions. Therefore, we will further investigate different SNR conditions to see the VoI performance. The following corollaries can be derived from Propositions 3 and 4 under different SNR conditions.

4.2.1 High SNR Regime

Corollary 9. *In the high SNR regime, for a fixed SNR γ, the exponential VoI difference over the number of observations satisfies*

$$\eta_m = O(\gamma^{-m}). \tag{4.37}$$

Moreover, $\eta_m = 0$ holds in the absence of noise.

Proof. See the Appendix in sec. 4.6.3. □

Corollary 9 shows how the VoI varies when increasing m by 1 for a fixed and high SNR. The difference between the VoI with $m+1$ and m observations is presented by a big O term in which γ^{-m} approaches 0 exponentially quickly as m grows. This means that increasing the length of the observation window does not help to improve the VoI in this case. Regardless of the correlation condition, the high SNR regime yields a nearly zero rate of change of the VoI when increasing the number of observations.

Corollary 10. *In the high SNR regime, for a fixed m, the exponential VoI difference over the SNR satisfies*

$$\eta_\gamma = \left(\frac{1}{\gamma} - \frac{1}{\gamma'}\right)\rho^{2(t-t_n)} + O\left(\frac{1}{\gamma^2}\right) + O\left(\frac{1}{\gamma'^2}\right). \tag{4.38}$$

Proof. In this case, we assume that $\frac{1}{\gamma} \to 0$ and $\frac{1}{\gamma'} \to 0$. This result is obtained directly by $(\mathbf{K}_{\boldsymbol{X}_m} + \epsilon \mathbf{I}_m)^{-1} = \mathbf{K}_{\boldsymbol{X}_m}^{-1} - \epsilon(\mathbf{K}_{\boldsymbol{X}_m}^{-1})^2 + O(\epsilon^2)$ when $\epsilon \to 0$ in Proposition 3. □

Corollary 10 shows the change of the VoI when increasing the SNR for a fixed m. The increase of the VoI with different SNR does not relate to m, thus the SNR is the dominant factor affecting the VoI in the high SNR regime. When $m = 1$, the difference of the VoI in Corollary 9 can be presented as the same order in Corollary 10. This means that, only for very small m, we may scale back SNR as we increase m by one to achieve the same VoI.

4.2.2 Low SNR Regime

Corollary 11. *In the low SNR regime, for a fixed SNR γ, the exponential VoI difference over the number of observations satisfies*

$$\eta_m = \rho^{2(t-t_n)} \rho^{2\sum_{i=1}^{m} T_i} \left[\gamma - \left(1 + 2m \right) \gamma^2 \right] + O(\gamma^3). \tag{4.39}$$

Proof. See the Appendix in sec. 4.6.4. □

In the low SNR regime, Corollary 11 shows that increasing the number of observations is a possible way to improve the VoI. However, η_m decreases with increasing m. This means that infinite observations cannot increase the VoI infinitely. Moreover, the low correlation condition ($\rho \to 0$) also yields a zero rate of change of the VoI when increasing m which is the same as the high SNR regime. A highly correlated data source ($\rho \to 1$) yields a positive VoI difference.

Corollary 12. *In the low SNR regime, for a fixed m, the exponential VoI difference over the SNR satisfies*

$$\eta_\gamma = (\gamma' - \gamma) \left(1 + \sum_{j=1}^{m-1} \rho^{2\sum_{i=1}^{j} T_i} \right) \rho^{2(t-t_n)} + O(\gamma^2) + O(\gamma'^2). \tag{4.40}$$

Proof. This result is obtained by series expansion at $\gamma = 0$ and $\gamma' = 0$. □

Corollary 12 shows that increasing the SNR is another possible way to improve the VoI in the low SNR regime. There is a trade-off existing in this case, i.e., either increasing SNR or increasing m by one can achieve the same target value. Compared with Corollaries 11 and 12, we have the following statement.

Increasing the SNR by a certain amount $\Delta\gamma$ has the same effect on the VoI as increasing m by 1 in the low SNR regime. Specifically, for fixed m and γ, we have

$v(m+1,\gamma) = v(m,\gamma+\Delta\gamma)$ when

$$\Delta\gamma = \frac{\gamma}{\sum_{j=1}^{m} \rho^{-2\sum_{i=j}^{m} T_i}}. \tag{4.41}$$

4.3 Optimal Sampling Policy to Maximise the Value of Information

In the AoI literature, the optimal sampling policy has been studied in [17, 25, 28, 29, 123, 124]. This is done by formulating a constrained optimisation problem to minimise the average AoI under the sampling rate constraint. In this section, we formulate and analyse this problem in the context of the VoI case. Our objective is to maximise the average VoI by optimising the sampling times at the data source under the maximum sampling rate constraint.

4.3.1 Problem Formulation

In this subsection, we focus on the VoI with a single observation which is given in (3.19) of Corollary 2, i.e.,

$$v(t) = -\frac{1}{2}\log\left[1 - \frac{\gamma}{1+\gamma}\rho^{2(t-t_n)}\right].$$

This VoI expression can be further presented by a function of the AoI which is given as

$$v_A(a) = -\frac{1}{2}\log\left(1 - \frac{\gamma}{1+\gamma}\rho^{2a}\right). \tag{4.42}$$

By using the functional relationship between the AoI and the VoI, the time average

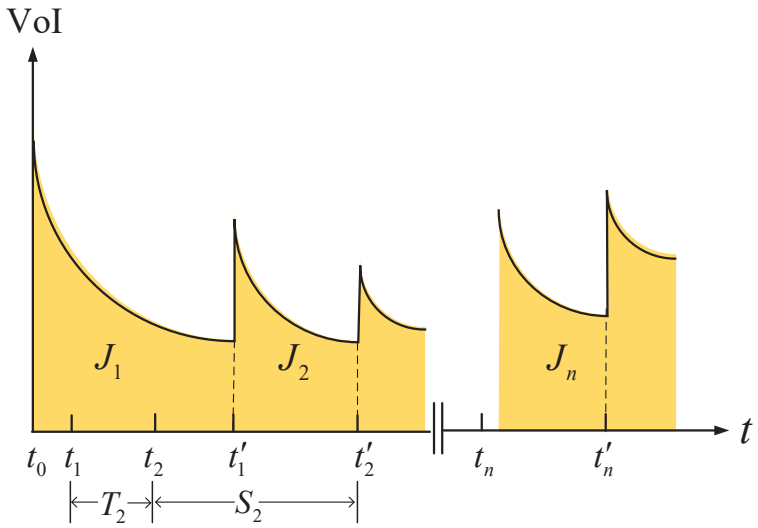


Figure 4.1: Time evolution of the VoI.

of the VoI can be presented by

$$\begin{aligned}
\mathrm{E}[V] &= \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^{\tau} v(t)\,\mathrm{d}t \\
&= \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1}\left[\int_{t_i'}^{t_{i+1}'} v(t)\,\mathrm{d}t\right]}{\sum_{i=0}^{n-1}(t_{i+1}'-t_i')} \\
&= \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1}\left[\int_{t_i'-t_i}^{t_{i+1}'-t_i} v_A(a)\,\mathrm{d}a\right]}{\sum_{i=0}^{n-1}(t_{i+1}'-t_i')}. \\
&= \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1}\int_{S_i}^{T_{i+1}+S_{i+1}} v_A(a)\,\mathrm{d}a}{\sum_{i=1}^{n} T_i + S_n}
\end{aligned} \tag{4.43}$$

where $S_i = t_i' - t_i$ is the transmission delay and $T_i = t_i - t_{i-1}$ is the sampling interval. Fig. 4.1 shows the time evolution of the VoI. Its time average is calculated by the integral which is also the summation of each yellow area in this figure. Transmission times of each data sample are i.i.d. and the distribution is assumed to be known. We

assume the initial value $S_0 = 0$ and $T_0 = t_1$. The transmission process is assumed to be stationary and ergodic. Therefore, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_{S_i}^{T_{i+1}+S_{i+1}} v_A(a)\,\mathrm{d}a = \mathrm{E}\left[\int_{S}^{T+S'} v_A(a)\,\mathrm{d}a\right] \tag{4.44}$$

$$\lim_{n\to\infty} \frac{1}{n}\left(\sum_{i=1}^{n} T_i + S_n\right) = \mathrm{E}[T]. \tag{4.45}$$

Then the average VoI can be further given as

$$\mathrm{E}[V] = -\frac{\mathrm{E}[J(S,T,S')]}{\mathrm{E}[T]} \tag{4.46}$$

where S' has the same distribution of S and the function $J(\cdot)$ is denoted as

$$J(S,T,S') = \int_{S}^{T+S'} \frac{1}{2}\log\left(1 - \frac{\gamma}{1+\gamma}\rho^{2a}\right)\mathrm{d}a. \tag{4.47}$$

Denote $f_{\max}$ as the maximum sampling rate at the data source. We aim to find the optimal sampling intervals T_i ($T_i \geq S_i$) to maximise the average VoI with the sampling rate constraint. This optimisation problem can be formulated as

$$\begin{aligned} \min_{T} \quad & \frac{\mathrm{E}[J(S,T,S')]}{\mathrm{E}[T]} \\ \text{s.t.} \quad & \mathrm{E}[T] \geq \frac{1}{f_{\max}}. \end{aligned} \tag{4.48}$$

For any integer i and sampling interval T_i, past time variables $\{S_{i-1}, \cdots, S_1\}$ and $\{T_{i-1}, \cdots, T_1\}$ are known. Therefore, the objective can be presented by the conditional expectation and the optimisation problem can be further formulated as

$$\begin{aligned} \min_{\tau} \quad & \frac{\mathrm{E}[J(S,\tau,S')|S=s]}{\tau} \\ \text{s.t.} \quad & \tau \geq \frac{1}{f_{\max}}. \end{aligned} \tag{4.49}$$

4.3.2 A Threshold-based Sampling Policy

Our goal is to find the optimal sampling interval τ given $S_i = s$ to maximise the average VoI under the sampling rate constraint. A similar problem was investigated in [125] with the aim of minimising the AoI, and the authors showed that the optimal sampling policy is a threshold-based policy. This result also holds for non-linear AoI optimisation problems. Therefore, the VoI-optimal sampling policy is given as the following.

Theorem 1 (Theorem 5.5, [125]). *The optimal sampling interval to maximise the average VoI is given by*

$$\tau = \inf\left\{\tau \geq s : \mathrm{E}\left[-\frac{1}{2}\log\left(1 - \frac{\gamma}{1+\gamma}\rho^{2(\tau+S')}\right)\right] \leq \beta\right\}. \tag{4.50}$$

If $\tau \geq \frac{1}{f_{\max}}$, *then* β *is the root of*

$$\beta = \frac{\mathrm{E}\left[\int_s^{\tau+S'} -\frac{1}{2}\log(1 - \frac{\gamma}{1+\gamma}\rho^{2a})\,\mathrm{d}a\right]}{\tau}; \tag{4.51}$$

otherwise, β *is determined by solving*

$$\tau = \frac{1}{f_{\max}}. \tag{4.52}$$

Here, we consider the non-preemptive FCFS queue. β is the threshold which can be solved by the bisection search algorithm. The sampling interval τ is conditioned on the transmission time s. The random variable S' represents the transmission delay and its distribution is assumed to be known. The optimal policy for VoI has the same form as the AoI-optimal policy, i.e., they are both threshold-based policies. Given all the past time instants, the next data packet is generated when the current time average VoI is below a threshold β, rather than generating as soon as possible when

the transmission channel is not busy, i.e., $\tau = s$.

4.4 Numerical Results

Numerical results obtained by Monte Carlo simulations are provided in this section. For arbitrary sampling times, the sampling process is chosen to be a rate λ Poisson process. For uniform sampling times, the sampling interval Δt is chosen to be fixed. The service time of each sample is also generated randomly according to an exponential distribution with rate μ. We set $\lambda = 0.5$, $\Delta t = 2$ and $\mu = 1$, and evaluate the VoI at a fixed time instant $t = 100$. Simulation results verify the validity of the simplified VoI in the high and low SNR regimes and illustrate the effects brought about by altering the number of observations and the SNR.

Figs. 4.2 and 4.3 show the numerical validation of the exact general VoI given in Proposition 1 and the approximated VoI with arbitrary sampling intervals in different SNR regimes which are discussed in Corollaries 5 and 6. In Fig. 4.2, we compare the exact VoI given in (4.2) with the approximated VoI in the high SNR regime which is given in (4.8). It is not surprising that the exact VoI increases as γ increases, and the large correlation parameter ρ leads to higher VoI. As expected, the approximated VoI is very close to the actual VoI when γ is large, while the gap increases when the system experiences larger noise. Fig. 4.3 considers the low SNR regime, and compares the exact VoI with the approximated VoI given in (4.18). Compared to the high SNR regime, we observe the opposite behaviour, i.e., the approximated VoI is approaching the exact VoI when γ is small. Therefore, these simulation results verify the analysis in Corollaries 5 and 6.

Fig. 4.4 shows the normalised VoI with the different number of observations m and different SNR γ, and illustrates the trade-off between m and γ. The normalised VoI is the ratio of the VoI and its non-noisy Markov counterpart which is given in (3.18).

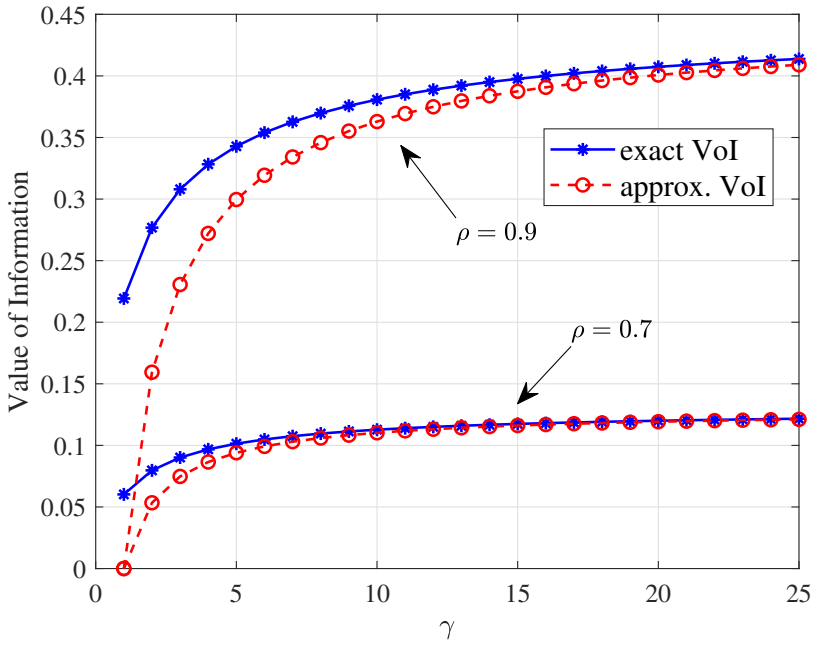


Figure 4.2: High SNR regime: Comparison of the exact VoI and the approximated VoI with arbitrary sampling times for $\rho \in \{0.7, 0.9\}$ at $t = 100$; sampling rate $\lambda = 0.5$ and the number of observations $m = 5$.

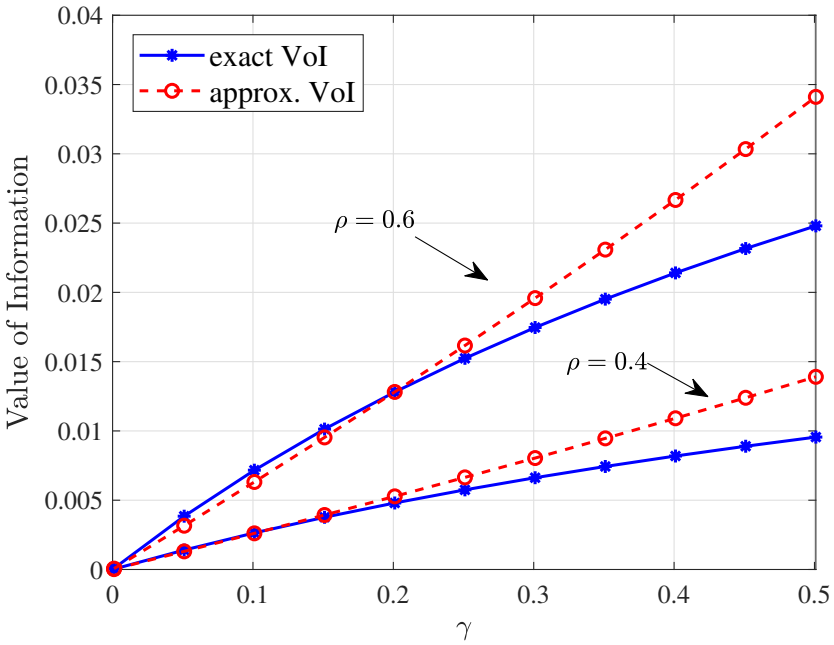


Figure 4.3: Low SNR regime: Comparison of the exact VoI and the approximated VoI with arbitrary sampling times for $\rho \in \{0.4, 0.6\}$ at $t = 100$; sampling rate $\lambda = 0.5$ and the number of observations $m = 5$.

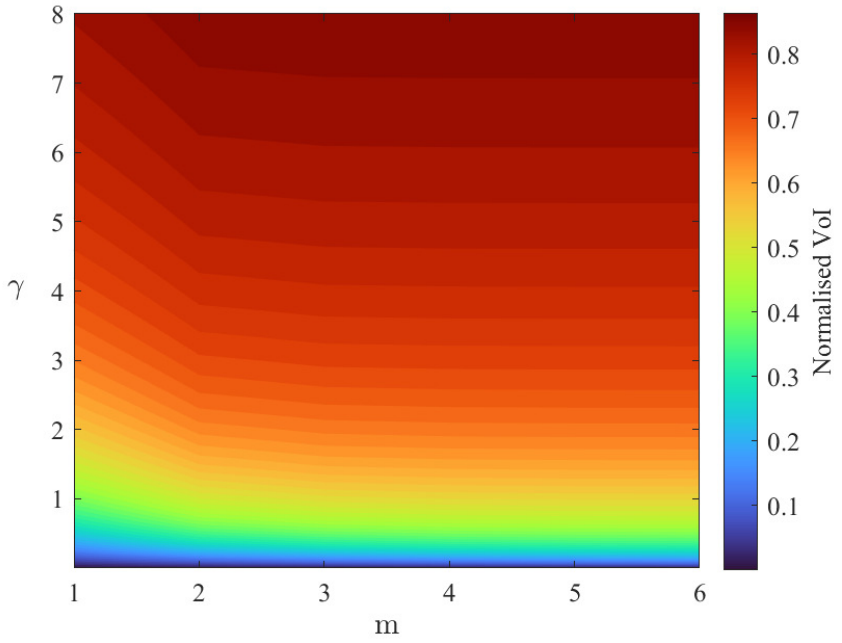


Figure 4.4: The normalised value of information with different number of observations $m \in \{1, 2, 3, 4, 5, 6\}$ and different SNR $\gamma \in (0, 8)$. The correlation parameter $\rho = 0.7$.

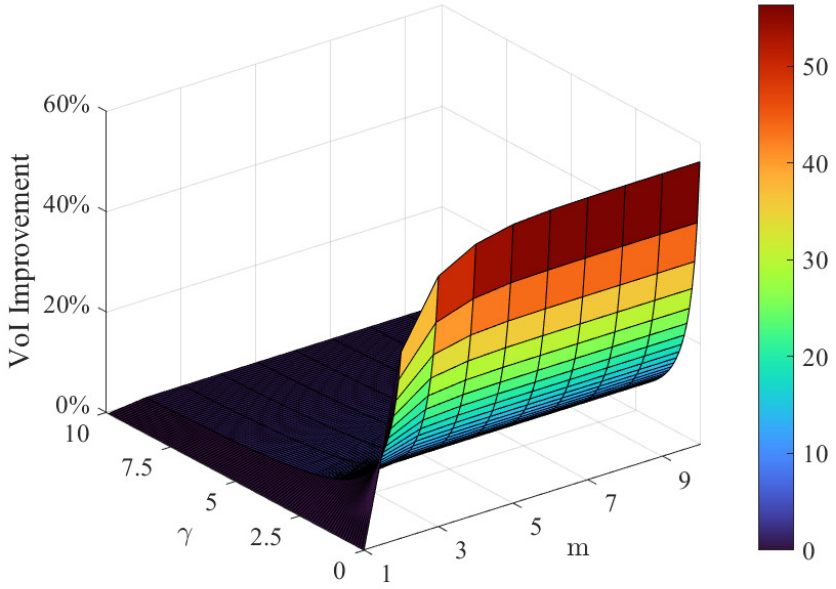


Figure 4.5: The improvement of the value of information by increasing the number of observations $m \in \{1, 2, 3, \cdots, 8, 9, 10\}$ under different SNR conditions $\gamma \in (0, 10)$. The correlation parameter $\rho = 0.7$.

It is shown that the VoI increases and then converges when increasing the SNR or using more data samples. Fixing the SNR, the VoI increases with m when the SNR is small, while the change is not obvious when the SNR is large. This result verifies the statements given in Corollaries 9 and 11, i.e., only in the low SNR regime, the VoI can be increased by increasing m. Fixing m, the VoI increases with the SNR regardless of the number of observations. This means that increasing the SNR is helpful to increase the VoI and verifies the results given in Corollaries 10 and 12.

Fig. 4.5 further shows the performance of the VoI when m rises from 1 to 10 for different SNR. The vertical axis represents the percentage increase in the VoI which is calculated by $\frac{v(m,\gamma)-v(1,\gamma)}{v(1,\gamma)} \times 100\%$. When the SNR is low, the VoI can be improved up to 60% by using more past observations. When the SNR is high, large m is not helpful to improve the VoI and the SNR is the dominant term affecting the VoI.

Figs. 4.6 and 4.7 are presented to illustrate the performance of the VoI under different sampling policies. In Theorem 1, we show that the optimal sampling policy to maximise the average VoI is a threshold-based policy. In addition to this VoI-optimal sampling policy, the following three policies are considered to evaluate the network-level information value.

- AoI-optimal policy [25]: the optimal threshold-based solution to minimise the average AoI.

- Minimum wait policy: a new sample is generated once the previous sample is delivered under the sampling constraint, i.e., $\tau = \max(s, \frac{1}{f_{\max}})$ where τ represents the sampling interval and s represents the transmission time. Without the sampling constraint, we have $\tau = s$ which is called the zero wait policy. In this case, the data packet is generated as soon as the previous one has been successfully delivered.

- Uniform sampling policy: the sampling intervals are constant $\tau = \Delta t$.

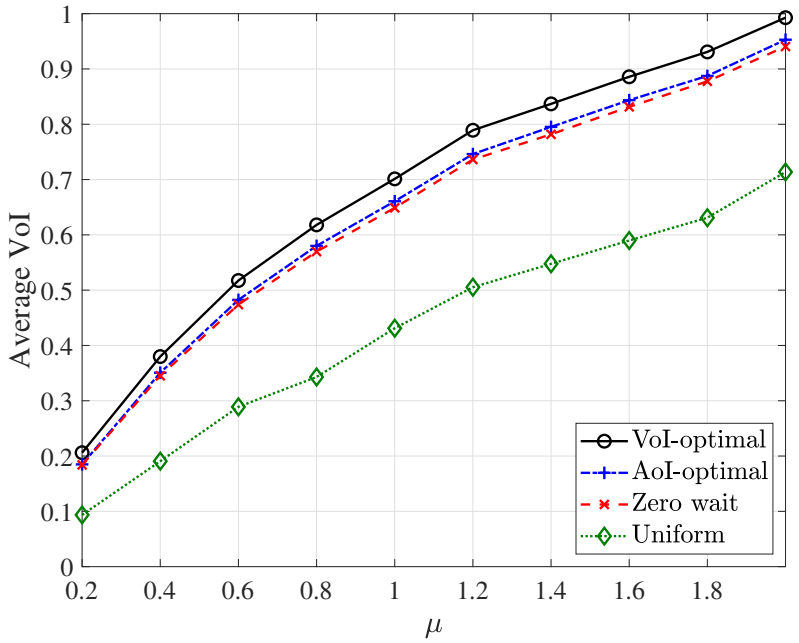


Figure 4.6: Average VoI in the hidden Markov process versus the transmission rate parameter μ for different sampling policies without sampling rate constraints; correlation parameter $\rho = 0.9$ and the SNR $\gamma = 50$.

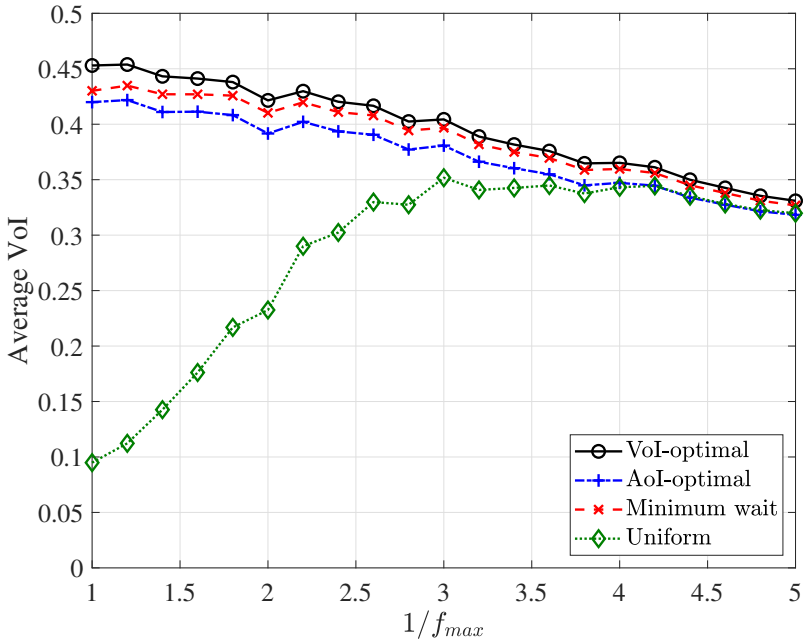


Figure 4.7: Average VoI in the hidden Markov process versus $1/f_{\max}$ for different sampling policies; transmission rate parameter $\mu = 0.5$, correlation parameter $\rho = 0.9$ and the SNR $\gamma = 50$.

Fig. 4.6 shows the average VoI versus the transmission rate μ for different sampling policies without the sampling rate constraint. Here, the uniform sampling interval $\Delta t = \frac{1}{\mu}$. The average VoI increases with the transmission rate μ and the threshold-based sampling policy achieves the largest average VoI. Fig. 4.7 shows the average VoI versus the minimum sampling interval $\frac{1}{f_{\max}}$ for different sampling policies. Here, the uniform sampling interval $\Delta t = \frac{1}{f_{\max}}$. As $\frac{1}{f_{\max}}$ goes larger, the constraint given in (4.48) will be an equality. This means that the sampling rate constraint is active, and the VoI-optimal policy will turn to the minimum wait sampling policy. These numerical results further illustrate that the threshold-based sampling policy achieves the highest information value, and the VoI-optimal strategy is not equivalent to the AoI-optimal strategy.

4.5 Summary

In this chapter, we focused on the noisy Gauss-Markov process models and explored the VoI performance with different sampling intervals in different SNR conditions. The approximated VoI expressions with arbitrary and uniform sampling times were derived in high and low SNR regimes, respectively. We also investigated the effects of the number of observations and the SNR on the VoI. It is interesting to find that, in the high SNR regime, increasing the SNR can largely increase the VoI regardless of how many observations are captured. However, using more past data samples is not helpful on the VoI in this case. In the low SNR regime, we showed that increasing the SNR by a certain value can achieve the equivalent VoI performance as increasing the number of observations by 1. Moreover, the optimal sampling policy at the source is obtained under the sampling rate constraints for VoI maximisation. The results given in this chapter are general. In the following chapter, we will apply the VoI in a specific hidden Gauss-Markov process, the noisy Ornstein–Uhlenbeck process. In

this model, we are going to study the statistical property of the VoI, and provide mathematical interpretations to show how the proposed VoI metric relates to other traditional performance metrics.

4.6 Appendix

4.6.1 Proof of (4.24)

For simplicity, we use ν to denote $\rho^{\Delta t}$. When the sampling intervals are constant, the inverse normalised autocovariance matrix of $\boldsymbol{X}$ can be further written as

$$\mathbf{K}_{\boldsymbol{X}}^{-1} = \frac{1}{(1-\nu^2)} \begin{bmatrix} 1 & -\nu & & & & \\ -\nu & 1+\nu^2 & -\nu & & & \\ & -\nu & \ddots & \ddots & & \\ & & \ddots & 1+\nu^2 & -\nu & \\ & & & -\nu & 1 & \end{bmatrix}. \tag{4.53}$$

The matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \frac{1}{\gamma}\mathbf{K}_{\boldsymbol{X}}^{-1} + \mathbf{I} = \begin{bmatrix} a & b & & & \\ b & c & b & & \\ & b & \ddots & \ddots & \\ & & \ddots & c & b \\ & & & b & a \end{bmatrix}, \tag{4.54}$$

where

$$a = \frac{1}{\gamma(1-\nu^2)} + 1, \quad b = \frac{-\nu}{\gamma(1-\nu^2)}, \quad c = \frac{1+\nu^2}{\gamma(1-\nu^2)} + 1. \tag{4.55}$$

Since the matrix $\mathbf{A}$ is tridiagonal, we are able to calculate its determinant [126].

Let the m-dimensional circulant matrix $\boldsymbol{\eta}$ where

$$(\boldsymbol{\eta})_{i,j} = \begin{cases} 1 & i = m, j = 1 \\ 1 & j = i + 1 \\ 0 & \text{others} \end{cases} \tag{4.56}$$

$$\det(\boldsymbol{\eta}) = (-1)^{m-1}. \tag{4.57}$$

The product of matrix $\mathbf{A}$ and $\boldsymbol{\eta}$ can be partitioned into four blocks

$$\mathbf{A}\boldsymbol{\eta} = \left[\begin{array}{c|ccccc} 0 & a & b & 0 & \cdots & 0 \\ \hline 0 & b & c & b & & \\ \vdots & & b & c & \ddots & \\ 0 & & & \ddots & \ddots & b \\ b & & & & b & c \\ a & & & & & b \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{\eta_{11}} & \boldsymbol{\eta_{12}} \\ \hline \boldsymbol{\eta_{21}} & \boldsymbol{\eta_{22}} \end{array}\right]. \tag{4.58}$$

Taking the determinant of both side, then we have

$$\det(\mathbf{A}) = \frac{\det(\boldsymbol{\eta_{22}}) \det(\boldsymbol{\eta_{11}} - \boldsymbol{\eta_{12}}\boldsymbol{\eta_{22}}^{-1}\boldsymbol{\eta_{21}})}{\det(\boldsymbol{\eta})}. \tag{4.59}$$

Here,

$$\det(\boldsymbol{\eta_{22}}) = b^{m-1}. \tag{4.60}$$

As $\boldsymbol{\eta_{22}}$ is a tri-band Toeplitz matrix, the inverse matrix can be expressed by [127]

$$\boldsymbol{\eta_{22}}^{-1} = \begin{bmatrix} J_1 & J_2 & \cdots & J_{m-1} \\ & J_1 & \ddots & \vdots \\ & & \ddots & J_2 \\ & & & J_1 \end{bmatrix} \tag{4.61}$$

where J_i falls in the form of the following recurrence relation

$$J_i = -\frac{c}{b}J_{i-1} - J_{i-2} \tag{4.62}$$

with $J_1 = \frac{1}{b}$ and $J_2 = -\frac{c}{b^2}$. Substituting (4.58), (4.60) and (4.61) into (4.59), we have

$$\det(\mathbf{A}) = (-1)^m b^{m-1}(a^2 J_{m-1} + 2abJ_{m-2} + b^2 J_{m-3}). \tag{4.63}$$

The recurrence relation can be solved by the roots of the characteristic polynomial. The characteristic equation is given by

$$\lambda^2 + \frac{c}{b}\lambda + 1 = 0, \tag{4.64}$$

and the eigenvalues are

$$\lambda_1 = \frac{-c + \sqrt{c^2 - 4b^2}}{2b}, \quad \lambda_2 = \frac{-c - \sqrt{c^2 - 4b^2}}{2b}. \tag{4.65}$$

Thus, J_i can be written as

$$J_i = \frac{1}{\sqrt{c^2 - 4b^2}}(\lambda_1^i - \lambda_2^i). \tag{4.66}$$

Substituting J_i into (4.63), we obtain the following result.

$$\det(\mathbf{A}) = \frac{(-1)^m b^{m-1}}{\sqrt{c^2 - 4b^2}}\left[a^2(\lambda_1^{m-1} - \lambda_2^{m-1}) + 2ab(\lambda_1^{m-2} - \lambda_2^{m-2}) + b^2(\lambda_1^{m-3} - \lambda_2^{m-3})\right], \tag{4.67}$$

where

$$\lambda_1 = \frac{-c + \sqrt{c^2 - 4b^2}}{2b}, \quad \lambda_2 = \frac{-c - \sqrt{c^2 - 4b^2}}{2b}. \tag{4.68}$$

Similarly, we also have

$$\det(\mathbf{A}_{11}) = \frac{(-1)^{m-1}b^{m-2}}{\sqrt{c^2-4b^2}}\left[ac(\lambda_1^{m-2}-\lambda_2^{m-2})+(ab+bc)(\lambda_1^{m-3}-\lambda_2^{m-3})+b^2(\lambda_1^{m-4}-\lambda_2^{m-4})\right]. \tag{4.69}$$

4.6.2 Proof of Proposition 3

The normalised autocovariance matrix of $\boldsymbol{X}_{m+1}$ can be written as

$$\mathbf{K}_{\boldsymbol{X}_{m+1}} = \begin{bmatrix} \mathbf{K}_{\boldsymbol{X}_m} & \boldsymbol{w}_m \\ \boldsymbol{w}_m^{\mathrm{T}} & 1 \end{bmatrix}. \tag{4.70}$$

For simplicity, we denote

$$\boldsymbol{\Sigma}_{m+1} = \mathbf{K}_{\boldsymbol{X}_{m+1}} + \frac{1}{\gamma}\mathbf{I}_{m+1} = \begin{bmatrix} \boldsymbol{\Sigma}_m & \boldsymbol{w}_m \\ \boldsymbol{w}_m^{\mathrm{T}} & 1+\frac{1}{\gamma} \end{bmatrix}. \tag{4.71}$$

The correlation between the current status X_t and the past $m+1$ observations can be captured by the following vector

$$\boldsymbol{u}_{m+1} = [\boldsymbol{u}_m^{\mathrm{T}}, \rho^{t-t_{n-m}}]^{\mathrm{T}}. \tag{4.72}$$

Based on (4.31), the VoI with $m+1$ observations is given by

$$v(m+1,\gamma) = -\frac{1}{2}\log\left[1-\boldsymbol{u}_{m+1}^{\mathrm{T}}\boldsymbol{\Sigma}_{m+1}^{-1}\boldsymbol{u}_{m+1}\right]. \tag{4.73}$$

Based on the block matrix inversion, we have

$$\boldsymbol{\Sigma}_{m+1}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{-1} + \delta \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} & -\delta \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m \\ -\delta \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} & \delta \end{bmatrix} \tag{4.74}$$

where

$$\delta = \left(1 + \frac{1}{\gamma} - \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m\right)^{-1} = \frac{\gamma}{1 + \gamma - \gamma \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m}. \tag{4.75}$$

Therefore, the VoI difference can be presented by

$$\boldsymbol{u}_{m+1}^{\mathrm{T}} \boldsymbol{\Sigma}_{m+1}^{-1} \boldsymbol{u}_{m+1} = \boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{u}_m + \delta \left(\boldsymbol{u}_m^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{w}_m - \rho^{t - t_{n-m}}\right)^2. \tag{4.76}$$

4.6.3 Proof of Corollary 9

In the high SNR regime, series expansion is used at $\frac{1}{\gamma} = 0$ to study how the VoI varies with the length of the time window m. For simplicity, we denote $\frac{1}{\gamma}$ as z and let $\eta(z) = \frac{g^2(z)}{h(z)}$ where

$$\begin{aligned} g(z) &= \boldsymbol{u}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m\right)^{-1} \boldsymbol{w}_m - \rho^{t - t_{n-m}} \\ h(z) &= 1 + z - \boldsymbol{w}_m^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m\right)^{-1} \boldsymbol{w}_m. \end{aligned} \tag{4.77}$$

It is easy to see that $g(0) = 0$ and $g'(0) = 0$, thus we have $\eta(0) = 0$ and $\eta'(0) = 0$.

For any positive integer $k > 1$, k-th order derivatives of $g(z)$ and $h(z)$ are given as

$$\frac{\partial^k g}{\partial z^k} = (-1)^k \boldsymbol{u}_m^{\mathrm{T}} \left[\left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m\right)^{-1}\right]^{k+1} \boldsymbol{w}_m \prod_{i=1}^{k} i, \tag{4.78}$$

$$\frac{\partial^k h}{\partial z^k} = (-1)^{k+1} \boldsymbol{w}_m^{\mathrm{T}} \left[\left(\mathbf{K}_{\boldsymbol{X}_m} + z\mathbf{I}_m\right)^{-1}\right]^{k+1} \boldsymbol{w}_m \prod_{i=1}^{k} i. \tag{4.79}$$

Since

$$\boldsymbol{u}_m^{\mathrm{T}} \mathbf{K}_{\boldsymbol{X}_m}^{-1} = \rho^{t - t_n} [1, 0, \cdots, 0] \tag{4.80}$$

$$\mathbf{K}_{\boldsymbol{X}_m}^{-1}\boldsymbol{w}_m = \rho^{t_{n-m+1}-t_{n-m}}[0,\cdots,0,1]^{\mathrm{T}}, \tag{4.81}$$

we have $\frac{\partial^k g}{\partial z^k} = 0$ holds for any $1 \leq k \leq m-1$ when $z = 0$. Therefore, we can state that $\frac{\partial^k \eta}{\partial z^k} = 0$ at $z = 0$ for any $1 \leq k \leq m-1$. If $\frac{1}{\gamma} = 0$ we have $\boldsymbol{u}_m^{\mathrm{T}}\boldsymbol{\Sigma}_m^{-1}\boldsymbol{w}_m = \rho^{t-t_{n-m}}$ in Proposition 3, thus the VoI with $m+1$ observations equals its counterpart with m observations when there is no noise.

4.6.4 Proof of Corollary 11

Similar to the analysis given in the high SNR regime, we let $\eta(\gamma) = \gamma\frac{g^2(\gamma)}{h(\gamma)}$ where

$$\begin{aligned} g(\gamma) &= \gamma\boldsymbol{u}_m^{\mathrm{T}}\left(\gamma\mathbf{K}_{\boldsymbol{X}_m} + \mathbf{I}_m\right)^{-1}\boldsymbol{w}_m - \rho^{t-t_{n-m}} \\ h(\gamma) &= 1 + \gamma - \gamma^2\boldsymbol{w}_m^{\mathrm{T}}\left(\mathbf{K}_{\boldsymbol{X}_m} + \mathbf{I}_m\right)^{-1}\boldsymbol{w}_m. \end{aligned} \tag{4.82}$$

It is easy to see that $g(0) = -\rho^{t-t_{n-m}}$ and $h(0) = 1$. The derivatives of $g(\gamma)$ and $h(\gamma)$ are given as

$$g'(0) = \boldsymbol{u}_m^{\mathrm{T}}\boldsymbol{w}_m, \quad g''(0) = -2\boldsymbol{u}_m^{\mathrm{T}}\mathbf{K}_{\boldsymbol{X}_m}^{-1}\boldsymbol{w}_m \tag{4.83}$$

$$h'(0) = 1, \quad h''(0) = 2\boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{w}_m. \tag{4.84}$$

Therefore, derivatives of $\eta(\gamma)$ are given as

$$\begin{aligned} \eta'(0) &= \rho^{2(t-t_{n-m})}, \\ \eta''(0) &= -2\rho^{2(t-t_{n-m})} - 4m\rho^{t-t_{n-m}}\rho^{t-t_n}\rho^{\sum_{i=1}^{m} T_i}. \end{aligned} \tag{4.85}$$

Chapter 5

Analysis of the Ornstein-Uhlenbeck Process Model

In Chapter 3, we formalised the basic notion of the value of information in general latent variable models. In Chapter 4, we commenced the investigation of VoI in the context of general hidden Gauss-Markov models. In this chapter, the VoI is further explored in the example of the Ornstein–Uhlenbeck process model, which is a typical and important Guass-Markov process model. In this scenario, we investigate the relationship between the proposed VoI and traditional AoI metric and study the stationary distribution of the VoI. The contributions of this chapter are given as follows:

- The VoI is applied in the context of one of the most important hidden Markov models: the noisy Ornstein-Uhlenbeck process, and its closed-form expression is derived.
- The relationship between the VoI and traditional AoI metric is investigated. We give the mathematical explanation for how the correlation and the noise affect their relationship, and obtain the functional dependency between them. A clear statistical framework linking the VoI to the AoI and a formal justification for

the selection of non-linear age functions are provided.

- Distribution of the VoI is explored in the M/M/1 queue model. We derive its probability density function (PDF), cumulative distribution function (CDF), and moment-generating function (MGF). We also derive the average VoI and its upper and lower bounds, which are tractable and useful for the design and optimisation of freshness-aware applications.

- Numerical results are provided to verify the theoretical analysis. The effects of SNR, sampling rate and correlation on its distribution and its relationship to AoI are discussed.

5.1 Ornstein–Uhlenbeck Process

In this section, we assume the underlying random process is an Ornstein–Uhlenbeck process which is an important continuous, stationary and Gauss-Markov process and is able to represent various practical applications [117]. For example, it can be used to model the mobility of a drone that hovers at a fixed point but experiences positional disturbances in UAV networks.

The OU process $\{X_t\}$ satisfies the following stochastic differential equation (SDE)

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{5.1}$$

where $\{W_t\}$ is standard Brownian motion, κ is the rate of mean reversion, θ is the long-term mean, and σ is the volatility of the random fluctuation. For any t, the variable X_t is normally distributed with mean and variance:

$$\mathrm{E}[X_t] = \theta, \quad \mathrm{Var}[X_t] = \frac{\sigma^2}{2\kappa}. \tag{5.2}$$

We assume the initial value X_0 is also normally distributed, specifically, $\mathcal{N}(\theta, \frac{\sigma^2}{2\kappa})$. Furthermore, X_t conditioned on X_s is also Gaussian with mean and variance:

$$\begin{aligned} \mathrm{E}[X_t|X_s] &= \theta + (X_s - \theta)e^{-\kappa(t-s)}, \\ \mathrm{Var}[X_t|X_s] &= \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right). \end{aligned} \tag{5.3}$$

In the OU model, the autocorrelation coefficient given in (3.11) can be written as $\rho = e^{-\kappa}$. We assume that the mean $\theta = 0$. Based on (3.12), the autocorrelation matrix is given as

$$\mathbf{K}_{\boldsymbol{X}} = \begin{bmatrix} 1 & e^{-\kappa(t_n - t_{n-1})} & \cdots & e^{-\kappa(t_n - t_{n-m+1})} \\ e^{-\kappa(t_n - t_{n-1})} & 1 & \cdots & e^{-\kappa(t_{n-1} - t_{n-m+1})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\kappa(t_n - t_{n-m+1})} & e^{-\kappa(t_{n-1} - t_{n-m+1})} & \cdots & 1 \end{bmatrix}. \tag{5.4}$$

5.2 VoI for Noisy OU Model

5.2.1 Value of Information

Corollary 2 in Chapter 3 gives the expression of the VoI with a single observation in general hidden Gauss-Markov models. According to Corollary 2, we can state the following main result in the noisy OU process model.

Corollary 13. *The VoI for the noisy OU process with a single observation is given by*

$$v(t) = -\frac{1}{2}\log\left[1 - \frac{\gamma}{1+\gamma}e^{-2\kappa(t-t_n)}\right], \quad t > t'_n \tag{5.5}$$

where the SNR γ is given by

$$\gamma = \frac{\sigma^2}{2\kappa\sigma_n^2}. \tag{5.6}$$

This corollary shows that the VoI ranges from 0 to $\frac{1}{2}\log(1+\gamma)$.

5.2.2 Connection to Age of Information

In the definition of the AoI $A(t) = t - t_n$ in (3.1), we consider that the time instant t_n is fixed which means that we view the AoI as deterministic. What we do here is to create a relationship between the VoI and the conditional AoI (i.e., the AoI conditioned on the most recent sample time). Therefore, the VoI is further written as a function of the AoI. Let $a = A(t)$; then, the VoI can be written as follows

$$v_A(a) = -\frac{1}{2}\log\left[1 - \frac{\gamma}{1+\gamma}e^{-2\kappa a}\right]. \tag{5.7}$$

The VoI with a single observation in (5.7) and its relationship to the AoI can be largely affected by system parameters. Fixing the random fluctuation parameter σ^2 of the OU process, the SNR γ relates to two parameters, κ and σ_n^2. κ can be used to represent the correlation property of the underlying OU process. If κ is small, the status updates are highly correlated; as κ increases, they become less correlated. σ_n^2 represents the noise level in the transmission environment. If σ_n^2 is small, the underlying hidden Markov process is dominant, and the VoI approaches its Markov counterpart in the OU model; otherwise, the noise process is dominant.

In the following part, we consider different SNR conditions to further investigate the relationship between the VoI and AoI. We can state the following corollaries regarding their connection.

Corollary 14. *In the low SNR regime, the VoI can be approximated as an exponential function of the AoI, which is given by*

$$v_A(a) \approx \frac{\gamma}{2(1+\gamma)}e^{-2\kappa a}. \tag{5.8}$$

Proof. In the low SNR regime (small γ), large κ and $\sigma_n^2 > 0$ (or large σ_n^2 and $\kappa > 0$) can lead to small SNR in (5.6). When γ approaches 0, the term $\frac{\gamma}{1+\gamma}e^{-2\kappa a}$ in (5.7) is

small. For small x, we have $\log(1+x) \approx x$, thus the result in (5.8) is obtained. □

In the low SNR regime, the dependency between the VoI and AoI is exponential. Less correlated samples or large noise can negatively affect the VoI at the receiver, thus the approximated VoI decreases faster as the AoI increases. For a less correlated data source, even fresh updates may contain little valuable information about the underlying OU process. For a high level of noise, status updates are corrupted, due to the indirect observation.

Corollary 15. *In the high SNR regime resulting from high correlation, the VoI can be approximated as a logarithmic function of the AoI, which is given by*

$$v_A(a) \approx -\frac{1}{2}\log(2\kappa\gamma a + 1) + \frac{1}{2}\log(1+\gamma). \tag{5.9}$$

Proof. For small x, we have $e^x \approx 1 + x$. Therefore, when $\kappa \to 0$ in (5.7), $e^{-2\kappa a} \approx 1 - 2\kappa a$. □

For highly correlated status updates, the VoI is expressed as a logarithmic function, and this means that the VoI decreases slower as the AoI increases. In this case, correlated updates can be transmitted under good channel conditions, thus old samples may still hold enough valuable information.

Corollary 16. *In the intermediate SNR regime where $\kappa \to 0$, $\sigma_n^2 \to \infty$ with $\kappa\sigma_n^2$ being constant, the VoI can be approximated as a linear function of the AoI, which is given by*

$$v_A(a) \approx -\kappa\gamma a + \frac{1}{2}\log(1+\gamma). \tag{5.10}$$

In the intermediate SNR regime where $\kappa \to \infty$, $\sigma_n^2 \to 0$ with $\kappa\sigma_n^2$ being constant, the VoI can be approximated as an exponential function of the AoI, which is given by

$$v_A(a) \approx \frac{\gamma}{2(1+\gamma)} e^{-2\kappa a}. \tag{5.11}$$

Proof. The result in (5.10) can be derived from Corollary 15 directly. When $\sigma_n^2 \to \infty$, the term $2\kappa\gamma a$ in (5.9) is small. Therefore, we have $\log(2\kappa\gamma a + 1) \approx 2\kappa\gamma a$. The result in (5.11) matches Corollary 14. When $\kappa \to \infty$, the term $e^{-2\kappa a}$ in (5.7) is small. For small x, we have $\log(1 + x) \approx x$, thus the result in (5.11) is obtained. □

The three corollaries stated above provide compelling insight into the adoption of non-linear AoI functions. In some existing works, exponential and logarithmic non-linear age functions are widely utilised to measure the information value, but they do not give a formal justification for why these functions are selected. Corollaries 14 to 16 provide a theoretic interpretation and explain how the correlation, noise and SNR affect the VoI and its relationship to the AoI in the noisy OU process. Generally, low SNR and high SNR conditions yield exponential and logarithmic relationships. The intermediate SNR regime yields an exponential or linear relationship, which depends on the value of noise and correlation. Therefore, the proposed VoI framework is more appropriate to measure the timeliness of information in different SNR regimes.

5.3 Analysis for M/M/1 Queueing Model

Corollaries 14 to 16 show general relationships between the VoI and the AoI in the noisy OU process. In this section, we relax the "fixed time instants" restriction and view the AoI as a random variable to study the distribution of the VoI. We explore the VoI in a specific FCFS M/M/1 queue system and derive its stationary distribution (including the PDF, CDF, expectation value and MGF).

5.3.1 M/M/1 Queueing Model

The M/M/1 system is the most elementary queue model in which there is one server in the system. The arrival process is a Poisson process and the service time follows

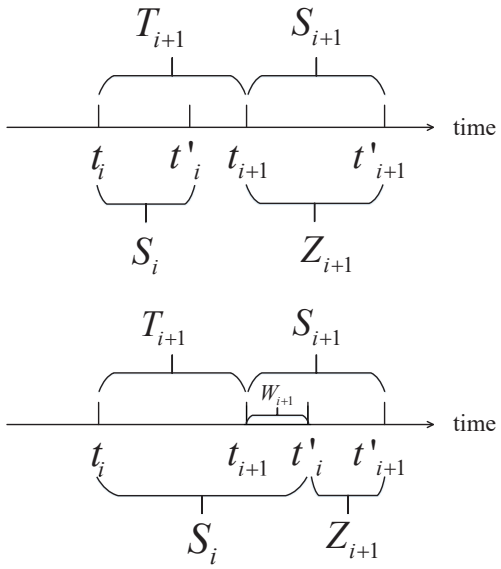


Figure 5.1: Two examples of the service time in M/M/1 Queueing Model.

the exponential distribution. This model has been widely used to model the packet generation and transmission processes in the area of telecommunications.

We denote T_i as the sampling interval of two packets which is given by

$$T_i = t_i - t_{i-1}, \quad 2 \leq i \leq n. \tag{5.12}$$

Denote S_i as the system time of the data packet i which is the total time from the generation at the source and observation at the receiver, i.e.,

$$S_i = t'_i - t_i, \quad 1 \leq i \leq n. \tag{5.13}$$

Denote Z_i as the service time of each data packet i which is the transmission delay (not including the queueing delay at the source). Fig. 5.1 illustrates two cases regarding the service time of the $(i+1)$-th packet. If $t_{i+1} > t'_i$, the service time of the $(i+1)$-th packet equals its system time. Otherwise, the service time equals the system time minus the waiting time, in which the waiting time of each packet is denoted by

W_i. Specifically, we can write

$$Z_{i+1} = \begin{cases} t'_{i+1} - t_{i+1}, & t_{i+1} \geq t'_i \\ t'_{i+1} - t'_i, & t_{i+1} < t'_i \end{cases} \tag{5.14}$$

$$W_{i+1} = (t'_i - t_{i+1})^+. \tag{5.15}$$

In the M/M/1 model, status updates are assumed to be sampled as a rate λ Poisson process, and this means that $\{T_i\}$ are i.i.d. exponentially distributed random variables with

$$\mathrm{E}[T_i] = \frac{1}{\lambda}, \quad \mathrm{Var}[T_i] = \frac{1}{\lambda^2}. \tag{5.16}$$

The service process is assumed to be a rate μ exponential process, and this means that $\{Z_i\}$ are i.i.d. exponential random variables with

$$\mathrm{E}[Z_i] = \frac{1}{\mu}, \quad \mathrm{Var}[Z_i] = \frac{1}{\mu^2}. \tag{5.17}$$

The arrival process and service process are independent of each other, but the service rate μ is assumed to be larger than the sampling rate λ. This assumption guarantees that all the packets can be served eventually and the system can reach a steady status. When the system reaches the steady state, the system times are also i.i.d. exponential random variables with [9, 128]

$$\mathrm{E}[S_i] = \frac{1}{\mu - \lambda}, \quad \mathrm{Var}[S_i] = \frac{1}{(\mu - \lambda)^2}. \tag{5.18}$$

5.3.2 Stationary Distribution of VoI

In this section, the VoI is applied in the FCFS M/M/1 queueing system to further explore the distribution. First, we investigate the stationary distribution of the VoI by utilising its functional relationship to AoI. This operating point is of interest in

applications with an outage restriction on the information value. Second, we further explore the moments and bounds of the VoI. This investigation is able to provide insight into potential applications of the proposed VoI framework.

5.3.2.1 Distribution of the VoI

The functional relationship between the VoI and AoI is given in (5.7). In this section, the AoI is regarded as a random variable which is utilised to study the stationary distribution of the VoI.

Lemma 2. *In the M/M/1 queueing model, the PDF and CDF of the AoI are given as [13]*

$$f_A(a) = \mu\left[\frac{\mu-\lambda}{\mu}e^{-(\mu-\lambda)a} - \left(\frac{\mu}{\mu-\lambda} + \lambda a - \frac{\lambda}{\mu}\right)e^{-\mu a} + \frac{\lambda}{\mu-\lambda}e^{-\lambda a}\right], \tag{5.19}$$

and

$$F_A(a) = 1 - e^{-(\mu-\lambda)a} + \left(\frac{\mu}{\mu-\lambda} + \lambda a\right)e^{-\mu a} - \frac{\mu}{\mu-\lambda}e^{-\lambda a}. \tag{5.20}$$

It can be seen that the distribution of the AoI only relates to the queue discipline, which means that it is independent of the inherent statistical characterisations of the underlying random process. As for the distribution of the VoI of a latent OU process with a single observation, we can state the following propositions.

Proposition 5. *In the M/M/1 queue model, the PDF of the VoI for the noisy OU process is given by*

$$f_V(v) = \frac{\mu e^{-2v}}{\kappa(1-e^{-2v})}\left[\frac{\mu-\lambda}{\mu}r(v)^{\frac{\mu-\lambda}{2\kappa}} - \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu} - \frac{\lambda}{2\kappa}\log r(v)\right)r(v)^{\frac{\mu}{2\kappa}} + \frac{\lambda}{\mu-\lambda}r(v)^{\frac{\lambda}{2\kappa}}\right], \tag{5.21}$$

where $r(v)$ is denoted as follows:

$$r(v) = \frac{(1+\gamma)(1-e^{-2v})}{\gamma}. \tag{5.22}$$

Proof. Since (5.7) is a monotonically decreasing function, the PDF of the VoI can be calculated by

$$f_V(v) = f_A(v_A^{-1}(v)) \left| \frac{\mathrm{d}}{\mathrm{d}v}\left(v_A^{-1}(v)\right) \right|. \tag{5.23}$$

Here, v_A^{-1} denotes the inverse function of the VoI given in (5.7), which can be written as

$$v_A^{-1}(v) = -\frac{1}{2\kappa} \log\left(\frac{(1+\gamma)(1-e^{-2v})}{\gamma}\right), \tag{5.24}$$

and we have

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(v_A^{-1}(v)\right) = -\frac{e^{-2v}}{\kappa(1-e^{-2v})}. \tag{5.25}$$

Therefore, the PDF of the VoI given in (5.21) is obtained by substituting (5.24), (5.19) and (5.25) into (5.23). □

Proposition 6. *In the M/M/1 queue model, the CDF of the VoI for the noisy OU process is given as*

$$F_V(v) = r(v)^{\frac{\mu-\lambda}{2\kappa}} - \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{2\kappa}\log r(v)\right) r(v)^{\frac{\mu}{2\kappa}} + \frac{\mu}{\mu-\lambda} r(v)^{\frac{\lambda}{2\kappa}}. \tag{5.26}$$

Proof. The CDF is obtained directly by the integral of the PDF, i.e., $F_V(v) = \mathrm{P}(V \leq v) = \int_0^v f_V(x)\,\mathrm{d}x$. □

Propositions 5 and 6 show that the distribution of the VoI relates to the sampling rate λ, service rate μ, correlation parameter κ and noise parameter σ_n^2, while the AoI distribution only relates to parameters λ and μ for the M/M/1 queue system.

The CDF of the VoI given in Proposition 6 can be interpreted as the "VoI outage probability", i.e., the probability that the VoI right before a new sample arrives is

below a threshold v, which can play a vital role in the system design. It is interesting to note that Proposition 6 implies that the VoI outage probability is a monotonically decreasing function of the service rate μ, and it converges to $r(v)^{\frac{\lambda}{2\kappa}}$ as μ goes to infinity. The proof of this is given in the Appendix of sec. 5.6.1. The reason for this decreasing nature of the VoI with μ is predictable because one would expect the information value to increase if the service time in the queue reduces.

Proposition 6 also implies that the VoI outage probability first decreases and then increases as the sampling rate λ increases. The optimal sampling rate λ^* satisfies $\frac{\partial \mathrm{P}(V \leq v)}{\partial \lambda}|_{\lambda=\lambda^*} = 0$. The proof of this is provided in the Appendix of sec. 5.6.2. It is not surprising that the small sampling rate λ can lead to a high outage, due to the lack of fresh updates at the source. It is interesting to find that the large sampling rate can also lead to high outage probability, due to the traffic congestion in the queue.

5.3.2.2 Moments and Bounds

In this subsection, we derive the expectation and two bounds of the VoI with a single observation, and calculate the moment-generating function of the VoI. We can state the following two propositions.

Proposition 7. *In the M/M/1 queue model, the average VoI for the noisy OU process is given as*

$$\begin{aligned}\mathrm{E}[V] = \frac{1}{2}\Bigg[&\log(1+\gamma) - g_1\left(\frac{\gamma}{1+\gamma}, \frac{\mu-\lambda}{2\kappa}\right) - \frac{\mu}{\mu-\lambda} g_1\left(\frac{\gamma}{1+\gamma}, \frac{\lambda}{2\kappa}\right) \\ &+ \left(\frac{\mu}{\mu-\lambda} + \frac{\lambda}{2\kappa}\log\frac{\gamma}{1+\gamma}\right) g_1\left(\frac{\gamma}{1+\gamma}, \frac{\mu}{2\kappa}\right) - \frac{\lambda}{2\kappa} g_2\left(\frac{\gamma}{1+\gamma}, \frac{\mu}{2\kappa}\right)\Bigg], \end{aligned} \tag{5.27}$$

where two functions $g_1(x,y)$ and $g_2(x,y)$ are defined for $x > 0$ and $y > 0$ with

$$g_1(x,y) = \frac{1}{x^y}\int_0^x \frac{z^y}{1-z}\,\mathrm{d}z, \tag{5.28}$$

$$g_2(x,y) = \frac{1}{x^y}\int_0^x \frac{z^y \log z}{1-z}\,\mathrm{d}z. \tag{5.29}$$

Proof. See the Appendix in sec. 5.6.3. □

Moreover, we have the following two bounds of the average VoI.

Corollary 17. *In the M/M/1 queue model, the average VoI is lower bounded by*

$$\mathrm{E}[V] \geq -\frac{1}{2}\log\left[1 - \frac{\gamma}{1+\gamma}\left(\frac{\frac{\mu-\lambda}{2\kappa}}{\frac{\mu-\lambda}{2\kappa}+1} - \frac{\frac{\mu-\lambda}{2\kappa}(\frac{\mu+\lambda}{2\kappa}+1)}{(\frac{\mu}{2\kappa}+1)^2(\frac{\lambda}{2\kappa}+1)}\right)\right], \tag{5.30}$$

and it is upper bounded by

$$\mathrm{E}[V] \leq \frac{1}{2}\left[H\left(\frac{\mu-\lambda}{2\kappa}\right) + \frac{\mu}{\mu-\lambda}H\left(\frac{\lambda}{2\kappa}\right) - \frac{\mu}{\mu-\lambda}H\left(\frac{\mu}{2\kappa}\right) + \frac{\lambda}{2\kappa}\psi^{(1)}\left(1+\frac{\mu}{2\kappa}\right)\right]. \tag{5.31}$$

Here, $H(\cdot)$ represents the harmonic number and the integral representation is given by $H(x) = \int_0^1 \frac{1-z^x}{1-z}\,\mathrm{d}z$ [129]. $\psi^{(1)}(\cdot)$ represents the first order polygamma function which is given by $\psi^{(1)}(x) = -\int_0^1 \frac{z^{x-1}\log z}{1-z}\,\mathrm{d}z$ [130].

Proof. See the Appendix in sec. 5.6.4. □

This corollary gives two bounds of the average VoI in the noisy OU process. The lower bound is based on Jensen's inequality. The equality holds if the VoI is a linear function of the Laplace transform of the AoI ($\mathrm{E}[e^{-2\kappa a}]$). In Corollary 14, we show that the dependence between the VoI and $\mathrm{E}[e^{-2\kappa a}]$ is approximately linear under the low SNR condition. Therefore, the average VoI approaches this lower bound in the low SNR regime. Moreover, the upper bound is based on the average VoI in the Markov model which is stated in (3.8). Hence, in the high SNR regime, the upper bound is tight. Compared with the general average VoI, the bounds are more tractable and may be useful for network design and optimisation.

Proposition 8. *In the M/M/1 queue, the MGF of the VoI for the noisy OU process is given as follows:*

$$M_v(t) = {}_2F_1\left(\frac{\mu-\lambda}{2\kappa}, \frac{t}{2}; \frac{\mu-\lambda}{2\kappa}+1; \frac{\gamma}{1+\gamma}\right) + \frac{\mu}{\mu-\lambda}{}_2F_1\left(\frac{\lambda}{2\kappa}, \frac{t}{2}; \frac{\lambda}{2\kappa}+1; \frac{\gamma}{1+\gamma}\right) - \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu}\right){}_2F_1\left(\frac{\mu}{2\kappa}, \frac{t}{2}; \frac{\mu}{2\kappa}+1; \frac{\gamma}{1+\gamma}\right) - \frac{\lambda}{\mu}{}_3F_2\left(\frac{\mu}{2\kappa}, \frac{\mu}{2\kappa}, \frac{t}{2}; \frac{\mu}{2\kappa}+1, \frac{\mu}{2\kappa}+1; \frac{\gamma}{1+\gamma}\right). \tag{5.32}$$

Here, ${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z)$ represents the generalised hypergeometric function which is given by the following series:

$${}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{z^n}{n!}. \tag{5.33}$$

where $(\cdot)_n$ represents the Pochhammer symbol, which is given as follows:

$$(x)_n = \begin{cases} 1 & n = 0 \\ \prod_{i=0}^{n-1} (x-i) & n \geq 1 \end{cases}. \tag{5.34}$$

Proof. See the Appendix in sec. 5.6.5. □

Moments of the VoI can be obtained by derivatives of the MGF at $t = 0$. The average VoI given in Proposition 7 is the first-order moment and can be derived from the MGF directly. Using this MGF, higher-order moments can also be used for the system design and optimisation instead of just utilising the average value.

5.4 Numerical Results

In this section, the relationship between the VoI and AoI and the distribution of the VoI are investigated through Monte Carlo simulations. In the simulation, the volatility parameter σ^2 of the OU model is fixed and set as 1. The sampling times

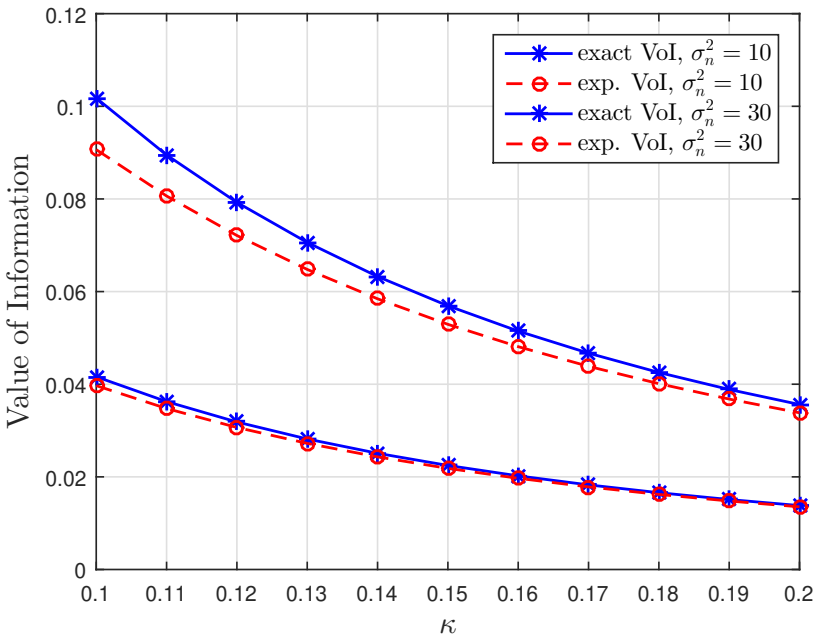


Figure 5.2: Low SNR regime: Comparison of the exact VoI and the exponential VoI versus κ for $\sigma_n^2 \in \{10, 30\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

$\{t_i\}$ are randomly generated by the rate λ Poisson process. The service times of each sample are randomly generated by the rate μ exponential process. We set time $t = 100$. For each running round, we record the sampling time of the most recently received update as t_n, and the VoI is calculated by (5.7).

Figs. 5.2–5.4 show the non-linear relationships between the VoI and the AoI under low, high and intermediate SNR conditions, respectively. Figs. 5.5–5.8 illustrate the distribution of the VoI, including the PDF, CDF and the outage probability. Figs. 5.9–5.12 provide the numerical results about the VoI expectation and bounds.

Figs. 5.2–5.4 compare the exact VoI in (3.19) and the approximated VoI for different SNR regimes which are given in (5.8) to (5.10). Fig. 5.2 shows that the exponential approximation is suitable when updates are less correlated and the noise is large. Figure 5.3 shows the opposite behaviour. The logarithmic approximation is more accurate when κ and σ_n^2 are small. Fig. 5.4 shows that the linear approximation is accurate when κ is small but σ_n^2 is large. These results verify the functional depen-

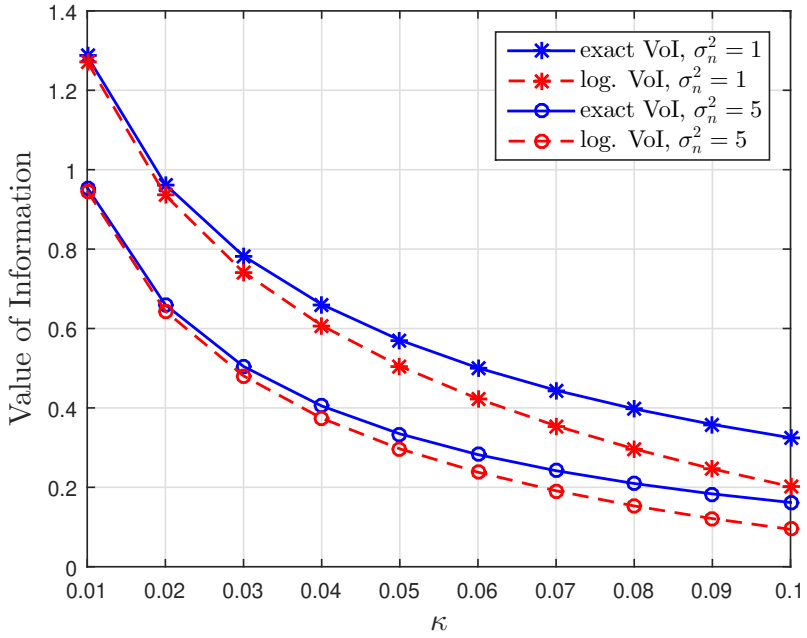


Figure 5.3: High SNR regime: Comparison of the exact VoI and the logarithmic VoI versus κ for $\sigma_n^2 \in \{1, 5\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

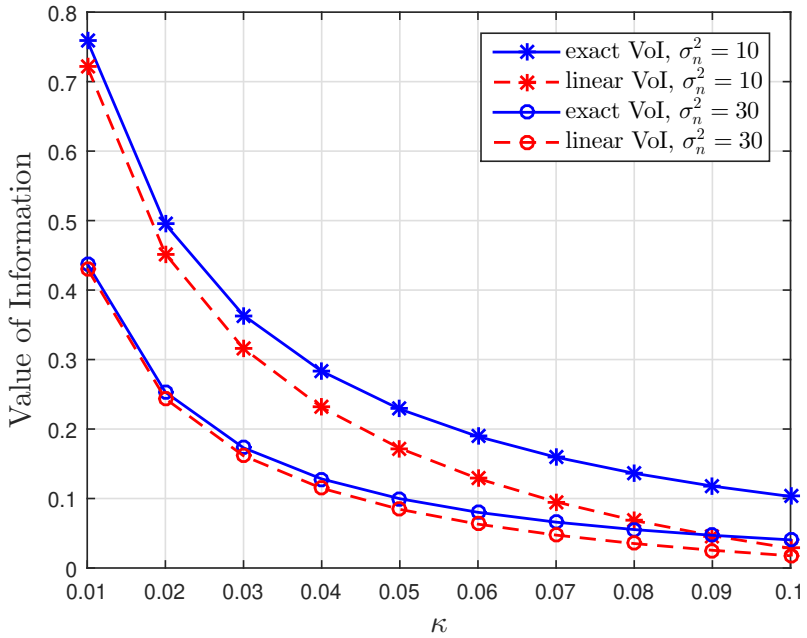


Figure 5.4: Intermediate SNR regime: Comparison of the exact VoI and the linear VoI versus κ for $\sigma_n^2 \in \{10, 30\}$ at $t = 100$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

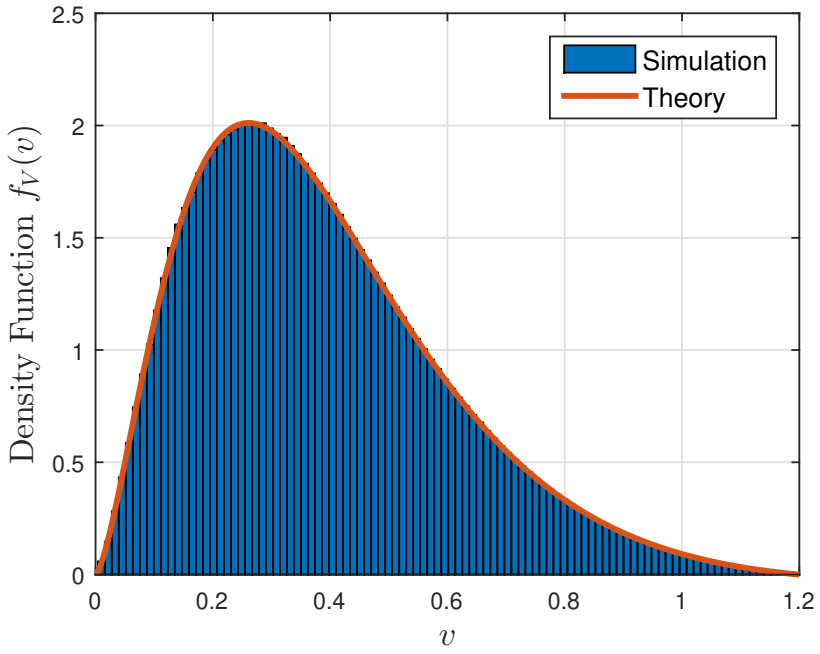


Figure 5.5: The density function of the VoI; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$, sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

dencies between VoI and the AoI, which are discussed in Corollaries 14, 15 and 16, illustrating that the low, high and intermediate SNR conditions yield exponential, logarithmic and linear relationships.

Fig. 5.5 gives the numerical validation of the theoretical PDF given in Proposition 5 and the density of the discrete path of the VoI obtained from the Monte Carlo simulations. Figs. 5.6–5.8 show the VoI outage probability given in Proposition 6 for different system parameters. In Fig. 5.6, the VoI outage probability is high when the status updates are less correlated or when the system experiences large noise. For a particular threshold v, Fig. 5.7 shows that either a too-small or too-large sampling rate can lead to a large VoI outage probability. Fixing μ, small λ means that we do not have sufficient newly generated status updates about the underlying OU process for prediction. Large λ means that enough newly generated updates have been sampled at the source, but they have to wait for a longer time, due to the packet congestion in the FCFS queue. Fig. 5.8 shows that VoI outage probability decreases

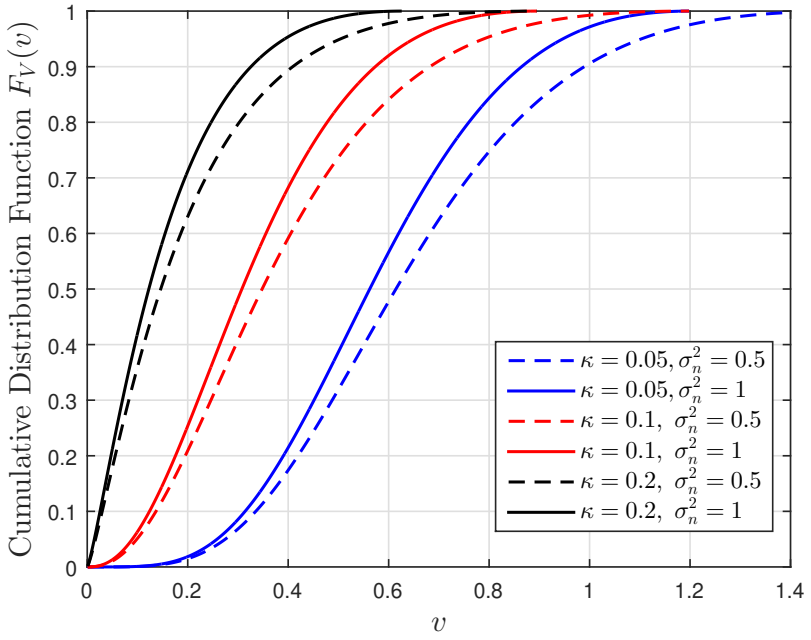


Figure 5.6: The cumulative distribution function of the VoI versus v for $\sigma_n^2 \in \{0.5, 1\}$ and $\kappa \in \{0.05, 0.1, 0.2\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

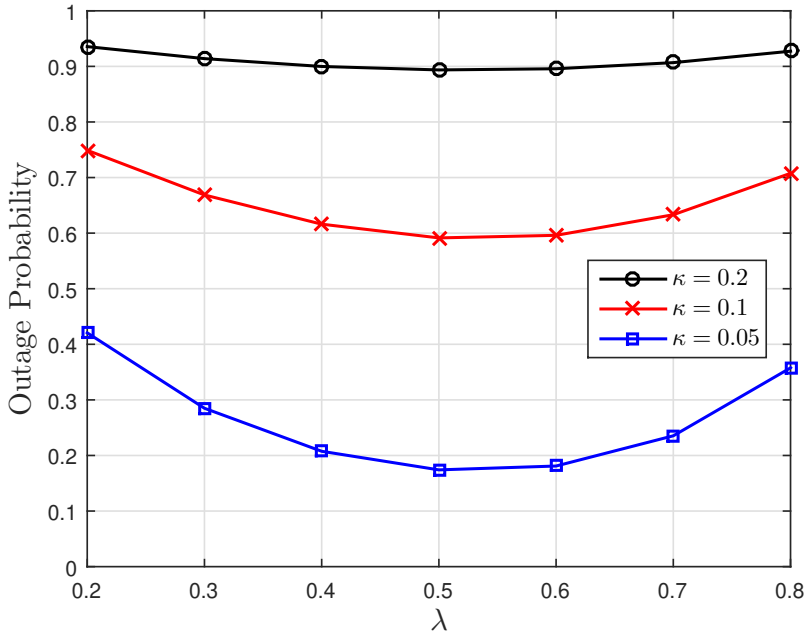


Figure 5.7: The VoI outage probability versus λ for $\kappa \in \{0.05, 0.1, 0.2\}$; threshold $v = 0.4$, noise parameter $\sigma_n^2 = 0.5$ and service rate $\mu = 1$.

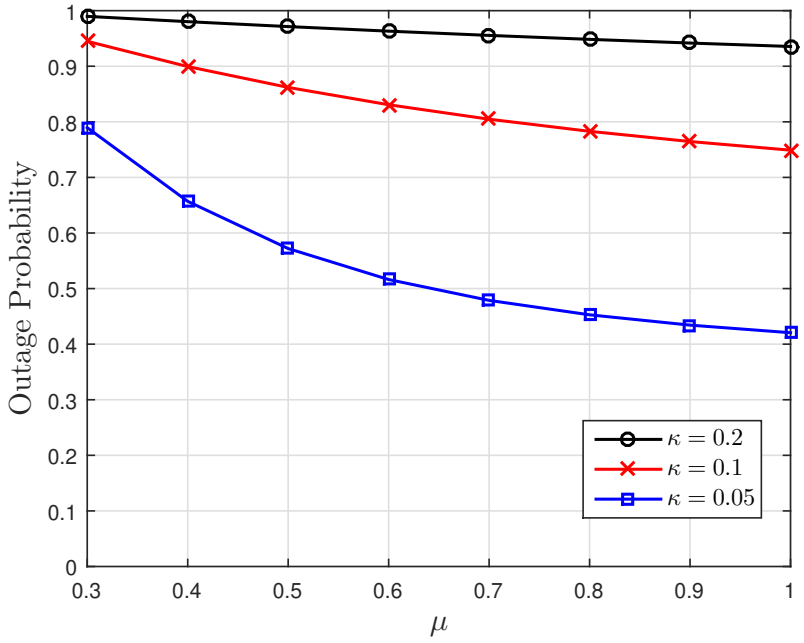


Figure 5.8: The VoI outage probability versus μ for $\kappa \in \{0.05, 0.1, 0.2\}$; threshold $v = 0.4$, noise parameter $\sigma_n^2 = 0.5$ and sampling rate $\lambda = 0.2$.

as the service rate μ increases. In the M/M/1 model, λ is smaller than μ. Fixing λ, large μ means that status updates can be served and transmitted more quickly, thus the receiver can hold more valuable information about the underlying process. These two figures verify the discussion given in Proposition 6.

Figs. 5.9 and 5.10 show the effect of the sampling rate and service rate on the average VoI and its bounds given in Proposition 7 and Corollary 17. The average VoI and its bounds first increase and then decrease as λ increases, and they increase as μ increases. This behaviour is similar to the VoI outage and can be explained similarly to Figs. 5.7 and 5.8. Moreover, it can be seen that the theoretical average VoI is consistent with the result obtained from the Monte Carlo simulations.

Figs. 5.11 and 5.12 plot the theoretical average VoI in (5.27) and the lower and upper bounds in (5.30) and (5.31) for different κ and σ_n^2. Small noise σ_n^2 and small κ can lead to large average VoI. In Fig. 5.11, the gap between the exact value and the lower bound is small for large σ_n^2, and it decreases as κ increases. The gap between the

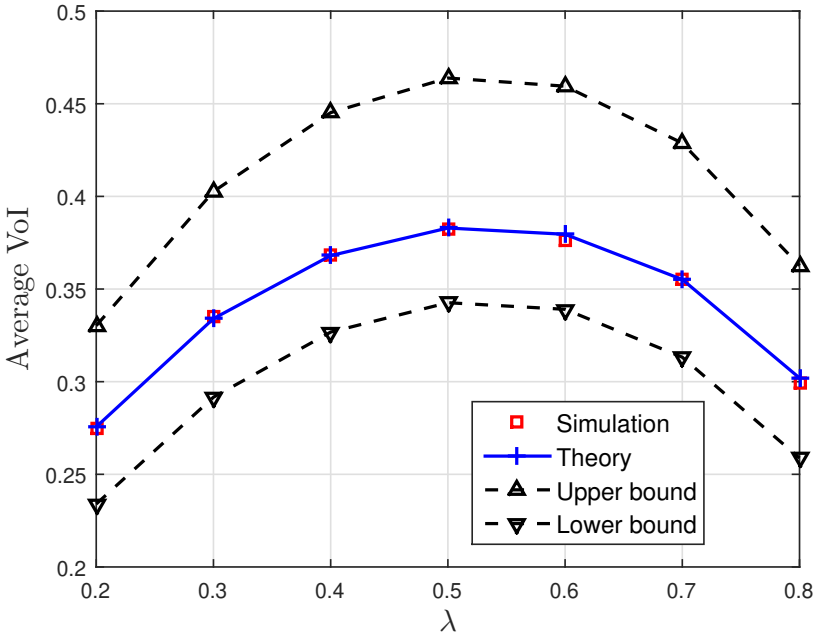


Figure 5.9: The average VoI and its bounds versus the sampling rate λ; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$ and service rate $\mu = 1$.

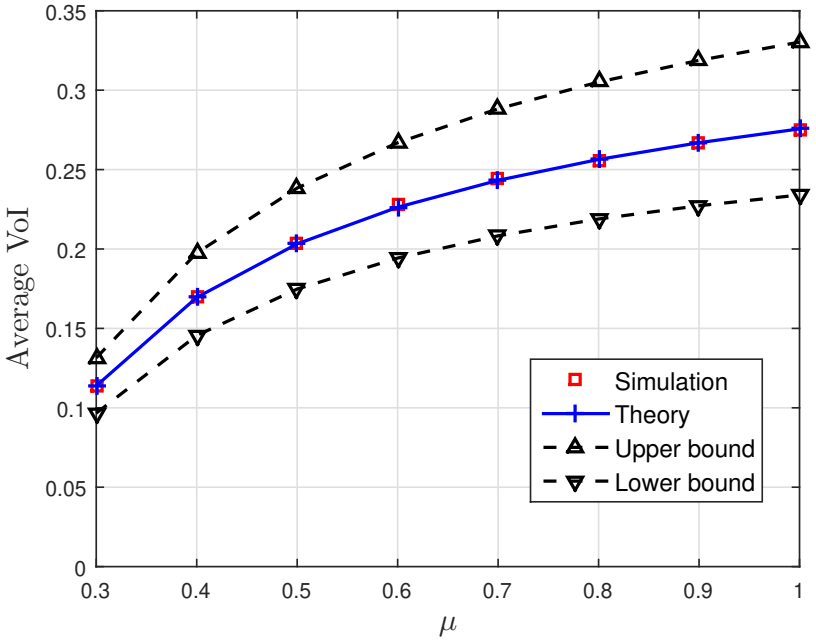


Figure 5.10: The average VoI and its bounds versus the service rate μ; correlation parameter $\kappa = 0.1$, noise parameter $\sigma_n^2 = 0.5$ and sampling rate $\lambda = 0.2$.

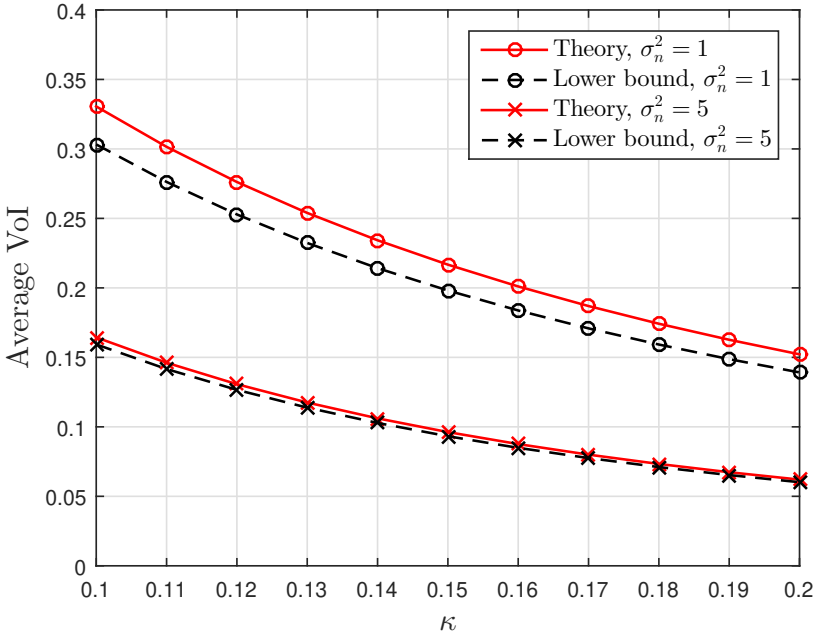


Figure 5.11: The average VoI and the lower bound versus κ for $\sigma_n^2 \in \{1, 5\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

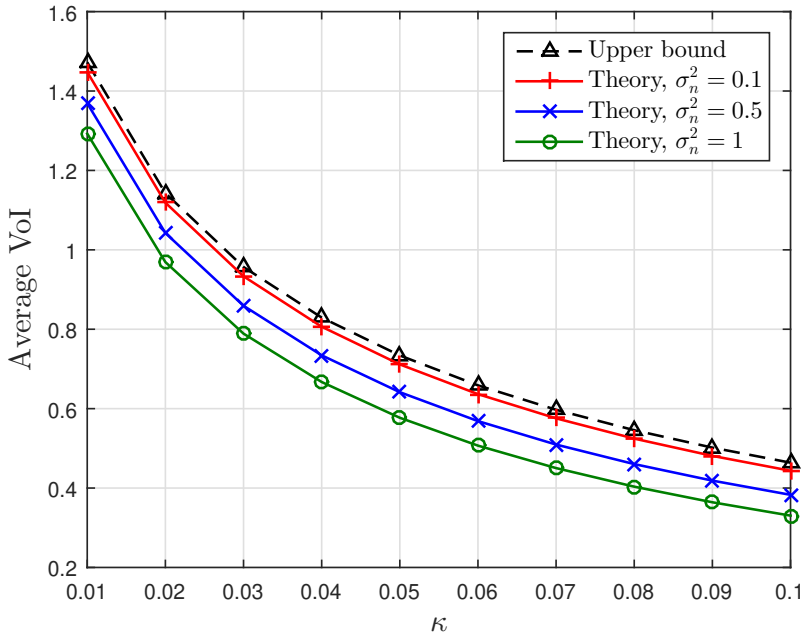


Figure 5.12: The average VoI and the upper bound versus κ for $\sigma_n^2 \in \{0.1, 0.5, 1\}$; sampling rate $\lambda = 0.5$ and service rate $\mu = 1$.

exact value and the upper bound in Fig. 5.12 shows the opposite behaviour; the gap narrows as σ_n^2 decreases. These two figures verify the discussion given in Corollary 17, illustrating that the average VoI approaches lower and upper bounds in low and high SNR regimes, respectively.

5.5 Summary

In this chapter, we analysed the performance of the VoI in the context of the noisy OU model. The dependency between the proposed VoI and the AoI was investigated. Functional relationships were obtained in different SNR conditions. We showed that the low, high and intermediate SNR conditions yield exponential, logarithmic and linear relationships between the two. Our work gave the formal justification for the selection of non-linear AoI functions previously reported in other works. Moreover, the distribution, moments and bounds of the VoI were investigated in the example of the M/M/1 queue model, which can be used as an optimisation tool in the design of real-time applications in the real world. Finally, we performed Monte Carlo simulations to obtain numerical validation of the theoretical analysis. In the following chapter, we will apply the concept of VoI in a practical Internet of Things network and study its optimisation problem to improve network efficiency.

5.6 Appendix

5.6.1 Proof of Monotonicity in μ

First, we prove the monotonicity in μ. For any particular VoI threshold v, the derivative of the VoI outage is given as

$$\frac{\partial \mathrm{P}(V \leq v)}{\partial \mu} = \frac{\log r(v)}{2\kappa}\left[r(v)^{\frac{\mu-\lambda}{2\kappa}} - \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{2\kappa}\log r(v)\right) r(v)^{\frac{\mu}{2\kappa}}\right] + \frac{\lambda}{(\mu-\lambda)^2}\left(r(v)^{\frac{\mu}{2\kappa}} - r(v)^{\frac{\lambda}{2\kappa}}\right). \quad (5.35)$$

For simplicity, let $x_1 = \frac{\lambda}{2\kappa}\log r(v)$ and $x_2 = \frac{\mu}{2\kappa}\log r(v)$. Then, (5.35) can be written as

$$\begin{aligned}\frac{\partial \mathrm{P}(V \leq v)}{\partial \mu} &= \frac{\log r(v)}{2\kappa}\left[e^{x_2-x_1} + \left(x_1 - \frac{x_2}{x_2-x_1}\right)e^{x_2}\right] + \frac{\log r(v)}{2\kappa}\frac{x_1}{(x_2-x_1)^2}\left(e^{x_2} - e^{x_1}\right) \\ &= \frac{\log r(v)}{2\kappa}e^{x_2}\left[e^{-x_1} + x_1 - \frac{x_2}{x_2-x_1} + \frac{x_1(1-e^{x_1-x_2})}{(x_2-x_1)^2}\right].\end{aligned} \quad (5.36)$$

Since $\lambda < \mu$ and $0 < r(v) < 1$, thus $x_2 < x_1 < 0$ and $\log r(v) < 0$. Moreover, for any x, we have $e^x \geq 1 + x$. Therefore, (5.36) can be further given as

$$\begin{aligned}\frac{\partial \mathrm{P}(V \leq v)}{\partial \mu} &\leq \frac{\log r(v)}{2\kappa}e^{x_2}\left[1 - \frac{x_2}{x_2-x_1} + \frac{x_1(x_2-x_1)}{(x_2-x_1)^2}\right] \\ &= 0.\end{aligned} \quad (5.37)$$

As the derivative is non-positive, the VoI outage is a monotonic function of μ.

5.6.2 Proof of the Existence of the Optimal λ

Next, we prove that the optimal sampling rate exists. The derivative of the VoI outage with respect to λ is given as

$$\begin{aligned}\frac{\partial \mathrm{P}(V \le v)}{\partial \lambda} = -\frac{\log r(v)}{2\kappa}\left(r(v)^{\frac{\mu-\lambda}{2\kappa}} - r(v)^{\frac{\mu}{2\kappa}} - \frac{\mu}{\mu-\lambda} r(v)^{\frac{\lambda}{2\kappa}}\right) \\ - \frac{\mu}{(\mu-\lambda)^2}\left(r(v)^{\frac{\mu}{2\kappa}} - r(v)^{\frac{\lambda}{2\kappa}}\right) \\ = -\frac{\log r(v)}{2\kappa}\left[e^{x_2-x_1} - e^{x_2} - \frac{x_2 e^{x_1}}{x_2-x_1} + \frac{x_2(e^{x_2}-e^{x_1})}{(x_2-x_1)^2}\right]. \quad (5.38)\end{aligned}$$

When λ approaches 0, we can write

$$\lim_{x_1 \to 0} \frac{\partial \mathrm{P}(V \le v)}{\partial \lambda} = -\frac{\log r(v)}{2\kappa}\left(\frac{e^{x_2}-1}{x_2} - 1\right) \le 0. \quad (5.39)$$

When λ approaches μ, we have

$$\lim_{x_1 \to x_2} \frac{\partial \mathrm{P}(V \le v)}{\partial \lambda} = -\frac{\log r(v)}{2\kappa}\left[1 - e^{x_2}\left(1 - \frac{x_2}{2}\right)\right] \ge -\frac{\log r(v)}{2\kappa}\left(1 - e^{\frac{x_2}{2}}\right) \ge 0. \quad (5.40)$$

We show that the VoI outage probability decreases with λ when λ is small, and increases when λ is large. Therefore, there exists the optimal sampling rate λ^*, and the minimum outage probability is achieved when the derivative in (5.38) is 0.

5.6.3 Proof of Proposition 7

The average VoI can be obtained directly by

$$\begin{aligned}\mathrm{E}[V] = \int_0^{\frac{1}{2}\log(1+\gamma)} v f_V(v)\,\mathrm{d}v = -\frac{\mu}{4\kappa}\int_0^1 \log\left(1 - \frac{\gamma}{1+\gamma} r\right)\left[\frac{\mu-\lambda}{\mu} r^{\frac{\mu-\lambda}{2\kappa}-1}\right. \\ \left. + \frac{\lambda}{\mu-\lambda} r^{\frac{\lambda}{2\kappa}-1} - \left(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu} - \frac{\lambda}{2\kappa}\log r\right) r^{\frac{\mu}{2\kappa}-1}\right] \mathrm{d}r. \quad (5.41)\end{aligned}$$

Here, for the given x and y, we have

$$\begin{aligned}\int_0^1 \log(1-xr)\cdot r^{y-1}\,\mathrm{d}r &= \frac{1}{x^y}\int_0^x \log(1-z)\cdot z^{y-1}\,\mathrm{d}z \\ &= \frac{1}{x^y}\frac{z^y}{y}\log(1-z)\Big|_{z=0}^{z=x} + \frac{1}{yx^y}\int_0^x \frac{z^y}{1-z}\,\mathrm{d}z \\ &= \frac{1}{y}\Big[\log(1-x) + g_1(x,y)\Big],\end{aligned} \tag{5.42}$$

and

$$\begin{aligned}&\int_0^1 \log r\log(1-xr)\cdot r^{y-1}\,\mathrm{d}r \\ &\quad= \frac{1}{x^y}\bigg(\int_0^x \log z\cdot\log(1-z)\cdot z^{y-1}\,\mathrm{d}z - \log x\int_0^x \log(1-z)\cdot z^{y-1}\,\mathrm{d}z\bigg) \\ &\quad= \frac{1}{x^y}\frac{z^y}{y}\log(1-z)\cdot\log z\Big|_{z=0}^{z=x} - \frac{\log x}{x^y}\int_0^x \log(1-z)\cdot z^{y-1}d\,\mathrm{d}z \\ &\qquad - \frac{1}{yx^y}\int_0^x z^y\bigg(\frac{\log(1-z)}{z} - \frac{\log z}{1-z}\bigg)\,\mathrm{d}z \\ &= \frac{\log(1-x)\cdot\log x}{y} - \bigg(\frac{1}{yx^y} + \frac{\log x}{x^y}\bigg)\int_0^x \log(1-z)\cdot z^{y-1}\,\mathrm{d}z + \frac{1}{yx^y}\int_0^x \frac{z^y\log z}{1-z}\,\mathrm{d}z \\ &\quad= -\frac{\log(1-x)}{y^2} - \bigg(\frac{1}{y^2} + \frac{\log x}{y}\bigg)g_1(x,y) + \frac{g_2(x,y)}{y}.\end{aligned} \tag{5.43}$$

Therefore, the average VoI is derived by substituting (5.42) and (5.43) into (5.41).

5.6.4 Proof of Corollary 17

The lower bound in (5.30) is obtained by applying Jensen's inequality, i.e.,

$$\begin{aligned}\mathrm{E}[V] &= \mathrm{E}\bigg[-\frac{1}{2}\log\bigg(1-\frac{\gamma}{1+\gamma}e^{-2\kappa a}\bigg)\bigg] \\ &\geq -\frac{1}{2}\log\bigg(1-\frac{\gamma}{1+\gamma}\,\mathrm{E}[e^{-2\kappa a}]\bigg),\end{aligned} \tag{5.44}$$

where

$$\mathrm{E}[e^{-2\kappa a}] = \int_0^{+\infty} e^{-2\kappa a} f_A(a)\,\mathrm{d}a = \frac{\frac{\mu-\lambda}{2\kappa}}{\frac{\mu-\lambda}{2\kappa}+1} - \frac{\frac{\mu-\lambda}{2\kappa}(\frac{\mu+\lambda}{2\kappa}+1)}{(\frac{\mu}{2\kappa}+1)^2(\frac{\lambda}{2\kappa}+1)}. \tag{5.45}$$

As we prove in (3.8), the upper bound in (5.31) is the average VoI in the Markov OU process which is denoted by V_{MM}. In the hidden Markov model, we can write

$$\begin{aligned}
&\mathrm{E}[V] \le \mathrm{E}[V_{\mathrm{MM}}] = \mathrm{E}\bigg[-\frac{1}{2}\log\Big(1-e^{-2\kappa a}\Big)\bigg] \\
&= -\frac{\mu}{4\kappa}\int_0^1 \log(1-r)\bigg[\frac{\mu-\lambda}{\mu} r^{\frac{\mu-\lambda}{2\kappa}-1} + \frac{\lambda}{\mu-\lambda} r^{\frac{\lambda}{2\kappa}-1} - \Big(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu} - \frac{\lambda}{2\kappa}\log r\Big) r^{\frac{\mu}{2\kappa}-1}\bigg]\,\mathrm{d}r.
\end{aligned} \tag{5.46}$$

Similar to the calculation given in (5.42) and (5.43), for the given y, we have [129,130]

$$\begin{aligned}
\int_0^1 \log(1-r)\cdot r^{y-1}\,\mathrm{d}r &= -\frac{1}{y}H(y), \\
\int_0^1 \log r\cdot\log(1-r)\cdot r^{y-1}\,\mathrm{d}r &= \frac{1}{y^2}H(y) - \frac{1}{y}\varphi^{(1)}(y).
\end{aligned} \tag{5.47}$$

Therefore, the upper bound of the average VoI is derived by substituting (5.47) into (5.46).

5.6.5 Proof of Proposition 8

The MGF of the VoI is obtained directly by

$$\begin{aligned}
M_v(t) = \int_0^{\frac{1}{2}\log(1+\gamma)} e^{tv} f_V(v)\,\mathrm{d}v = \frac{\mu}{2\kappa}\int_0^1 \Big(1-\frac{\gamma}{1+\gamma}r\Big)^{-\frac{t}{2}} \bigg[\frac{\mu-\lambda}{\mu} r^{\frac{\mu-\lambda}{2\kappa}-1} \\
+ \frac{\lambda}{\mu-\lambda} r^{\frac{\lambda}{2\kappa}-1} - \Big(\frac{\mu}{\mu-\lambda} - \frac{\lambda}{\mu} - \frac{\lambda}{2\kappa}\log r\Big) r^{\frac{\mu}{2\kappa}-1}\bigg]\,\mathrm{d}r.
\end{aligned} \tag{5.48}$$

Here, for the given x, y and t, we have the following [130]:

$$\int_0^1 (1-xr)^{-\frac{t}{2}} \cdot r^{y-1}\,\mathrm{d}r = \frac{1}{y}{}_2F_1\left(y, \frac{t}{2}; y+1; x\right), \tag{5.49}$$

and

$$\int_0^1 \log r \cdot (1-xr)^{-\frac{t}{2}} \cdot r^{y-1}\,\mathrm{d}r = -\frac{1}{y^2}{}_3F_2\left(y, y, \frac{t}{2}; y+1, y+1; x\right). \tag{5.50}$$

Therefore, the MGF of VoI is derived by substituting (5.49) and (5.50) into (5.48).

Chapter 6

Optimal Mobility and Communication Strategy in IoT Networks

In previous chapters, we have defined a clear VoI framework and analysed its performance at the fundamental level in various hidden Markov models. In this chapter, we apply VoI in practical IoT networks to study a VoI-oriented network optimisation problem. We consider an industrial IoT network with different types of sensing devices (data sources) and a mobile central entity (receiver). We aim to find the optimal mobility of the central entity and the data transmission strategy to maximise the minimum VoI at the central entity. The results given in this chapter can be further utilised in the design of timeliness-critical IoT applications. The contributions of this chapter are given as follows:

- An optimisation problem is formulated for data collection in IoT networks with different types of sensing devices. We jointly consider the path planning of a central entity and the communication scheme to maximise the value of information. This problem is formulated as a mixed integer non-linear programming

problem with mobility, scheduling, and VoI constraints.

- As the proposed problem is intractable, we reformulate it as an MDP and propose a competitive off-line algorithm based on reinforcement learning to solve it efficiently. We obtain the optimal moving direction, time slot allocation and user scheduling scheme so that the network-level VoI at the entity is maximised.

- Numerical results are presented to evaluate the performance of the proposed algorithm. Compared with other existing methods, the proposed method achieves better performance when the wireless resource is limited by exploiting the correlation properties of sensed data.

6.1 System Model and Problem Formulation

As shown in Fig. 6.1, we consider a wireless network with a set of fixed sensing nodes and a mobile central entity. In a practical industrial IoT network, the sensing nodes represent various IoT devices, and the central entity can be used to represent the intelligent robot or the smart vehicle. Various sensing nodes are deployed to monitor the dynamics of heterogeneous physical processes, such as temperature, emergency traffic, light intensity and so on. They are required to upload the status update of the target random process to the central entity in a timely manner and then the entity offloads the data to the server for further analysis to achieve real-time surveillance and precise control. Due to the limited wireless resources, it is difficult to transmit the data packet to the centre via long distance. Therefore, the central entity needs to move to the best position to fulfil the communication demand by controlling its moving path. To achieve real-time industrial automation, the usefulness of the data at the centre plays an important role. How to plan the path and schedule the nodes is a key problem in such networks.

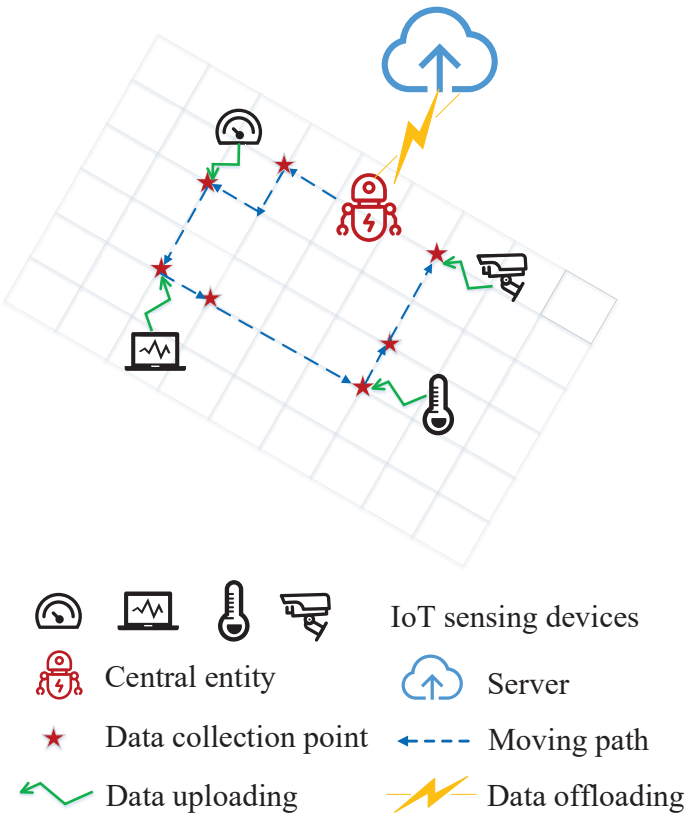


Figure 6.1: System model of an industrial IoT network.

6.1.1 Mobility Model

We consider a network with N fixed sensing devices and one mobile central entity. This network is assumed to be time-slotted and we denote T as the total number of time slots which means that the time is discrete. For simplicity, the duration of each slot is assumed to be 1. The location of each sensor is denoted by (x_i, y_i) with $1 \leq i \leq N$, which is assumed to be fixed and known. The location of the mobile central entity within time slot t is denoted by $(x(t), y(t))$. The moving path can be obtained by connecting the location of each separate line segment. The distance between the node i and the central entity at time slot t is denoted as $d_i(t)$ which is given as

$$d_i(t) = \sqrt{\Big(x_i - x(t)\Big)^2 + \Big(y_i - y(t)\Big)^2}, \quad (1 \leq i \leq N, 1 \leq t \leq T). \tag{6.1}$$

For the central entity's mobility, we assume it moves with a fixed speed and we denote z as the moving distance within a unit time slot. $m(t)$ represents the moving direction of the central entity at time slot t, and there are five options which can be chosen by the central entity. $\{\theta_N, \theta_S, \theta_E, \theta_W\}$ represents north, south, east and west directions, respectively. If $m(t) = \theta_0$, the central entity will retain its current position. Specifically, the location of the central entity is presented by

$$\Big(x(t+1), y(t+1)\Big) = \begin{cases} \Big(x(t), y(t) + z\Big), & m(t) = \theta_N \\ \Big(x(t), y(t) - z\Big), & m(t) = \theta_S \\ \Big(x(t) + z, y(t)\Big), & m(t) = \theta_E \\ \Big(x(t) - z, y(t)\Big), & m(t) = \theta_W \\ \Big(x(t), y(t)\Big), & m(t) = \theta_0 \end{cases} \qquad (1 \leq t < T). \tag{6.2}$$

6.1.2 Communication Model

Compared with the urban outdoor network, the quality of transmission in an industrial IoT is less likely to be affected by channel fading caused by blocks from buildings, and mountains [131]. Therefore, we assume that sensor devices are able to communicate with the central entity via the line of sight (LoS) link with additive white Gaussian noise, thus the channel condition largely depends on the transmission distance. Based on the free-space path loss model, the SNR between the node i and the central entity at time slot t can be given as

$$\gamma_i(t) = \alpha_i(t)\frac{P_i g_0 d_i^{-2}(t)}{BN_0}, \quad (1 \leq i \leq N, 1 \leq t \leq T) \tag{6.3}$$

where P_i is the transmission power of node i, g_0 is the channel power gain at the reference distance $d_0 = 1\text{m}$, and B is the available bandwidth. $\alpha_i(t)$ is a binary variable indicating the scheduling policy. If the node i transmits its data successfully at time slot t, then $\alpha_i(t) = 1$, and vice versa. In this work, we consider the time-division multiple access (TDMA)-based user scheduling scheme and this means that at most one node can be served for a given time slot. Therefore, the time slot allocation constraints can be given as

$$\sum_{i=1}^{N} \alpha_i(t) \leq 1, \quad (1 \leq t \leq T). \tag{6.4}$$

6.1.3 Value of Information Model

The transmission model of the individual sensing node is the same as the model given in Fig. 3.4, in which we consider the AWGN channel. For each sensing node, we denote $\{X_t^{(i)}\}$ as the random process under observation at node i with $1 \leq i \leq N$. We model the physical process $\{X_t^{(i)}\}$ as a stationary Gauss-Markov random process which generalises various random processes [117], such as the Ornstein–Uhlenbeck

process. Denote $X_{t_j}^{(i)}$ as the data packet which is sampled by the sensor i at time t_j. It represents the status of the target physical process $\{X_t^{(i)}\}$ at time t_j. We assume the sampled packet is transmitted through an AWGN channel and received by the central entity at time t'_j. We denote $Y_{t'_j}^{(i)}$ as the corresponding noisy observation, and denote $\{Y_t^{(i)}\}$ as the observing random process at the central entity.

In this model, we focus on the VoI with a single observation. For a given time instant t and a given device i, the VoI is defined as

$$v_i(t) = \mathrm{I}(X_t^{(i)}; Y_{t'_n}^{(i)}), \quad t > t'_n. \tag{6.5}$$

We denote ρ_i $(0 < \rho_i < 1)$ as the autocorrelation parameter of node i. According to (3.19), the VoI (in bits/s) of the node i at time t can be written as

$$v_i(t) = -B \log \left(1 - \frac{P_i g_0 \rho_i^{A_i(t)}}{B N_0 q_i^2(t) + P_i g_0}\right), \quad (1 \le i \le N, 1 \le t \le T). \tag{6.6}$$

Here, $A_i(t)$ represents the AoI of node i at time t. $q_i(t)$ represents the "effective distance" between the node i and the central entity at time slot t. We consider the "generate-at-will" sampling policy in which the AoI drops to 1 when a node is served. In the following part, we will consider two cases to show how $q_i(t)$ relates to $d_i(t)$ and the AoI evolution.

- Case 1: $\gamma_i(t) \ge \gamma_{th}$. If the SNR is larger than a given threshold γ_{th}, we assume that the transmission is successful. In this case, we have $q_i(t) = d_i(t)$ and $A_i(t) = 1$. If the transmission at time t is successful, the effective distance is equivalent to the Euclidean distance. The AoI will be updated to 1 when a new packet is received by the central entity, and the VoI can be updated accordingly.

- Case 2: $\gamma_i(t) < \gamma_{th}$. If the SNR is smaller than the given threshold, we assume that the transmission is unsuccessful and we need to discard the packet. In the

case with network outage, we have $q_i(t) = q_i(t-1)$ and $A_i(t) = A_i(t-1) + 1$. This is because the central entity does not receive any effective information regarding the status of the underlying random process, and it can only hold the information at the previous time slot. Therefore, the unsuccessful status transmission will not cause an update but a decrease in the VoI.

Thus, the AoI and the effective distance constraints are given as

$$A_i(t) = \begin{cases} 1, & \gamma_i(t) \geq \gamma_{th} \\ A_i(t-1)+1, & \gamma_i(t) < \gamma_{th} \end{cases}, \quad (1 \leq i \leq N, 1 \leq t < T). \tag{6.7}$$

and

$$q_i(t) = \begin{cases} d_i(t), & \gamma_i(t) \geq \gamma_{th} \\ q_i(t-1), & \gamma_i(t) < \gamma_{th} \end{cases}, \quad (1 \leq i \leq N, 1 \leq t \leq T). \tag{6.8}$$

In (6.6), the AoI of each device $A_i(t)$ measures the freshness of data from the time domain. The effective transmission distance $q_i(t)$ measures the timeliness of data from the space domain. The correlation parameter ρ_i reflects the physical properties of the underlying random process. Therefore, the proposed VoI notion is able to capture the usefulness of data in noisy spatial-temporal networks.

6.1.4 Problem Formulation

Our objective is to find the optimal central entity's path planning and the node scheduling scheme to maximise the minimum VoI among all sensor nodes. This max-

min optimisation problem is formulated as follows:

$$
\begin{aligned}
\max_{\mathbf{m},\alpha} \quad & V \\
\text{s.t.} \quad & v_i(t) \geq V, \quad (1 \leq i \leq N, 1 \leq t \leq T) \\
& \text{Mobility constraints: (6.1), (6.2)} \\
& \text{Communication constraints: (6.3), (6.4)} \\
& \text{VoI constraints: (3.19), (6.7), (6.8)} \\
& \alpha_i(t) \in \{0, 1\}, \quad (1 \leq i \leq N, 1 \leq t \leq T) \\
& m(t) \in \{\theta_N, \theta_S, \theta_E, \theta_W, \theta_0\}, \quad (1 \leq t \leq T)
\end{aligned}
$$

In this problem formulation, $v_i(t)$, $q_i(t)$, $d_i(t)$, and $\gamma_i(t)$ are continuous variables. $A_i(t)$, $m(t)$, $x(t)$ and $y(t)$ are discrete variables. $\alpha_i(t)$ is a binary variable. T, N, P_i, γ_{th}, B, N_0, ρ_i, g_0 and z are constants. Our objective is to find the optimal path planning of the central entity ($m(t)$) and the time slot allocation scheme of each sensor node ($\alpha_i(t)$) at each time slot to maximise the minimum VoI over all the nodes in the network.

6.2 A Reinforcement Learning-based Algorithm

This VoI optimisation problem is formulated as a mixed integer non-linear programming problem. Due to the high dynamics and uncertainty of the network, it is not practical to find the optimal solution using traditional methods. Therefore, we reformulate the problem as a Markov decision process first and then propose a reinforcement learning-based offline method to solve this problem efficiently.

6.2.1 Markov Decision Process

The MDP is widely used to model the decision-making process of an agent. A general MDP includes four parts, i.e., $\{\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}\}$, where $\mathcal{S}$ is the state set, $\mathcal{A}$ is the action set, $\mathcal{R}$ is the reward when moving from one state to another state due to an action, and $\mathcal{P}$ is the corresponding state transition probability.

In our work, the central entity is regarded as the agent, and the corresponding MDP is described as follows.

- State: For each time slot t, the state includes the central entity's position, the AoI and the effective distance between the sensor and the central entity, which is given as

$$s_t = \Big\{ x(t), y(t), \boldsymbol{A}(t), \boldsymbol{q}(t) \Big\} \tag{6.9}$$

where

$$\begin{aligned} \boldsymbol{A}(t) &= \Big[A_1(t), A_2(t), \cdots, A_N(t) \Big]^{\mathrm{T}}, \\ \boldsymbol{q}(t) &= \Big[q_1(t), q_2(t), \cdots, q_N(t) \Big]^{\mathrm{T}}. \end{aligned} \tag{6.10}$$

Here, $x(t)$ and $y(t)$ jointly represent the position of the central entity, vector $\boldsymbol{A}(t)$ represents the AoI of all nodes, and vector $\boldsymbol{q}(t)$ represents the effective distance of all nodes.

- Action: For each time slot t, the action includes the moving direction and the user scheduling scheme, which is given as

$$a_t = \Big\{ m(t), \boldsymbol{\alpha}(t) \Big\} \tag{6.11}$$

where

$$\begin{aligned} m(t) &\in \Big\{ \theta_N, \theta_S, \theta_E, \theta_W, \theta_0 \Big\}, \\ \boldsymbol{\alpha}(t) &= \Big[\alpha_1(t), \alpha_2(t), \cdots, \alpha_N(t) \Big]^{\mathrm{T}}. \end{aligned} \tag{6.12}$$

Here, $m(t)$ represents the moving direction of the agent, and $\boldsymbol{\alpha}(t)$ represents the scheduling policy of all nodes at time slot t with $\alpha_i(t) \in \{0, 1\}$ for all $1 \leq i \leq N$.

- Reward: For each time slot t, the reward relates to the state and action at the previous time slot, which is defined as the minimum VoI among all nodes, i.e.,

$$r_t(s_{t-1}, a_{t-1}) = \min\{v_1(t), \cdots, v_N(t)\} - f_{penalty}. \tag{6.13}$$

  Here, $f_{penalty}$ is a non-negative and constant parameter which is used to represent the penalty. Specifically, if no node is visited for a given time slot, we take $f_{penalty} = 100$; otherwise, $f_{penalty} = 0$. Due to the existence of this penalty parameter, the learning process can be more efficient by stimulating the central agent to collect data as timely as possible and avoiding it stopping at a position with no information being collected. The VoI can be calculated by $\boldsymbol{A}(t)$ and $\boldsymbol{q}(t)$. Since the total number of time slots T is assumed to be finite, the AoI of each node is finite. Therefore, the VoI of each node in (6.6) is also finite.

- Transition probability: The state transition probability is defined as the probability of the state in the next time slot $(t+1)$ given the state and action in the time slot t, which is denoted as $\mathrm{P}(s_{t+1}|s_t, a_t)$.

For any given time slot, the MDP is in a state s_t and the central entity needs to choose a feasible action a_t from the available set $\mathcal{A}_{s_t}$. For the following time slot, the MDP moves to the next state s_{t+1} and the central entity receives the related reward $r_{t+1}(s_t, a_t)$. The objective of the MDP is to find a policy that can be used by the central entity to choose how to move and serve sensor nodes for each time slot.

We denote $\pi(a|s)$ as the stationary policy which is a distribution of actions given states, i.e.,

$$\pi(a|s) = \mathrm{P}(a_t = a|s_t = s). \tag{6.14}$$

The optimal policy can be obtained by maximising the state-value function. Particularly, the state-value function is defined as the expected long-term discounted reward function which is given as [132]

$$\mathbb{V}_{\pi}(s) = \mathrm{E}_{\pi}\left[\sum_{\tau=t}^{\infty} \omega^{\tau} r_{\tau+1}(s_{\tau}, a_{\tau}) \mid s_t = s\right]. \tag{6.15}$$

$\mathrm{E}_{\pi}[\cdot]$ represents the expectation taken over different decision-making given different states. Here, ω is the discount factor ($0 \leq \omega \leq 1$) which measures the level of the importance of future states given the current state. Based on Bellman's expectation equation [132], the state-value function can be presented by the immediate reward plus the discounted value of the future state, i.e.,

$$\begin{aligned} \mathbb{V}_{\pi}(s) &= \mathrm{E}_{\pi}\left[r_{t+1}(s_t, a_t) + \omega \mathbb{V}_{\pi}(s_{t+1}) \mid s_t = s\right] \\ &= \mathrm{E}_{\pi}[r_{t+1}(s_t, a_t)] + \omega \sum_{s' \in \mathcal{S}} \mathrm{P}(s'|s, a) \mathbb{V}_{\pi}(s'). \end{aligned} \tag{6.16}$$

Here, s and a represent the current state and action. s' represents the next state when action a is taken.

We denote $\pi^*(a|s)$ as the optimal policy which can lead to the optimal state-value function, i.e.,

$$\begin{aligned} \mathbb{V}_{\pi^*}(s) &= \max_{\pi} \mathbb{V}_{\pi}(s) \\ &= \max_{a \in \mathcal{A}_s} \left[\mathrm{E}_{\pi}[r_{t+1}(s_t, a_t)] + \omega \sum_{s' \in \mathcal{S}} \mathrm{P}(s'|s, a) \mathbb{V}_{\pi}(s')\right]. \end{aligned} \tag{6.17}$$

We introduce a state-action-value function $\mathbb{Q}_{\pi}(s, a)$ as right hand side of (6.17) and it is defined as

$$\mathbb{Q}_{\pi}(s, a) = \mathrm{E}_{\pi}[r_{t+1}(s_t, a_t)] + \omega \sum_{s' \in \mathcal{S}} \mathrm{P}(s'|s, a) \mathbb{V}_{\pi}(s'). \tag{6.18}$$

Thus, we can write

$$\mathbb{V}_{\pi^*}(s) = \max_{a \in \mathcal{A}_s} \mathbb{Q}_{\pi^*}(s, a). \tag{6.19}$$

The optimal state-value function $\mathbb{V}$ can be obtained by the state-action-value function $\mathbb{Q}$ which is given as

$$\mathbb{Q}_{\pi^*}(s, a) = \mathrm{E}_{\pi}[r_{t+1}(s_t, a_t)] + \omega \sum_{s' \in \mathcal{S}} \mathrm{P}(s'|s, a) \max_{a' \in \mathcal{A}_{s'}} \mathbb{Q}_{\pi^*}(s', a'). \tag{6.20}$$

An algorithm is proposed in the following section to find the solution π^* of our formulated problem.

6.2.2 A Q-learning based Algorithm

Generally, dynamic programming or integer programming can be used to solve the MDP problem based on the given state transition probability. Considering the dynamic nature and uncertainty of the mobile IoT network, the transition probabilities can be unknown. Therefore, reinforcement learning is used to solve the MDP problem in the context of implicit transition probabilities.

Q-leaning is an important and lightweight RL algorithm which can learn the reward of action in a given state with random transition probability based on past experience. The aim of a Q-learning task is to find the optimal action a_t based on the current state s_t. Specifically, the central entity has a Q-table to store the Q-value which relates to the current position and the information value. The central entity will choose an available action based on its Q-value, and the Q-table will be updated after moving to the next state. Through update iterations, the optimal Q-value can be derived eventually in a recursive way.

By implementing the Q-learning algorithm, the optimal Q-value can be achieved

by an iterative update based on the following recursion based on (6.20)

$$Q_{t+1}(s_t, a_t) \leftarrow (1-\phi)Q_t(s_t, a_t) + \phi r_{t+1}(s_t, a_t) + \phi\omega \max_{a \in \mathcal{A}_{s_{t+1}}} Q_t(s_{t+1}, a) \tag{6.21}$$

where ϕ is the learning rate ($0 < \phi \leq 1$) and ω is the discounted factor. The updated Q-value relates to three parts. The first term $(1-\phi)Q_t(s_t, a_t)$ represents the current Q-value. The second term $\phi r_{t+1}(s_t, a_t)$ reflects the reward when action a_t is taken in state s_t. The third term $\phi\omega \max_{a \in \mathcal{A}_{s_{t+1}}} Q_t(s_{t+1}, a)$ measures the optimal future reward which can be obtained in state s_{t+1}. Based on (6.19), the optimal action is chosen that maximises the Q-value.

Algorithm 1 A Q-learning based Algorithm

Initialisation: learning rate ϕ, discounted factor ω, Q-table, running rounds $R_{\max}$, total time slots $T_{\max}$, parameters ϵ and ϵ_0

for $i = 1:1:R_{\max}$ **do**
  set initial state s_1
  for $t = 1:1:T_{\max}$ **do**
    $U_{\max} = -1$
    Generate a random number x ranging from $[0, 1]$
    if $x < \epsilon$ **then**
      Randomly select an action $a_t \in \mathcal{A}_{s_t}$
    else
      for each action $a \in \mathcal{A}_{s_t}$ **do**
        if $Q(s_t, a) > U_{\max}$ **then**
          $U_{\max} = Q(s_t, a)$
          $a_t = a$
        end if
      end for
    end if
    Adopt action a_t
    Record the next state s_{t+1} and calculate the reward $r(s_t, a_t)$ based on (6.13)
    Update the Q-table based on (6.21)
  end for
  $\epsilon \leftarrow \epsilon - \epsilon_0$
end for

The details of the proposed Q-learning-based method are presented in Algorithm 1. First, we need to initialise the following constant parameters: the learning rate

ϕ, the discounted factor ω, the total number of running rounds $R_{\max}$ and the total number of time slots $T_{\max}$. Q-value in the Q-table is set to an arbitrary but fixed value before running the algorithm. ϵ and ϵ_0 represent the exploration and exploitation probabilities. The central entity can exploit the learning and estimate the Q-value based on the recursive method to maximise the VoI. However, at the beginning of the algorithm, the central entity does not have the chance to visit all states, so it has no knowledge of whether the estimated value is optimal. Therefore, it is important to explore the next available state-action pair for a given state. In the proposed algorithm, an ϵ-greedy method is utilised to balance exploration and exploitation. The exploitation phase means that the entity chooses the available action that maximises the Q-value. The exploration phase means that the entity chooses a random action from the available action set which is independent of the Q-value. For a given state s, the action is selected by the following

$$a = \begin{cases} \arg\max\limits_{a \in \mathcal{A}_s} Q(s,a), & \text{with probability } 1-\epsilon \\ \text{a random action in } \mathcal{A}_s, & \text{with probability } \epsilon. \end{cases} \tag{6.22}$$

The central entity adopts the selected action, records the next state, calculates the corresponding reward based on (6.13), and updates the Q-table based on (6.21). For each running round, the probability ϵ is deceased by ϵ_0 at a fixed rate. As the learning goes on, it will decrease to 0 and this means that the agent is able to exploit the learning to take optimal action instead of exploration.

6.3 Numerical Results

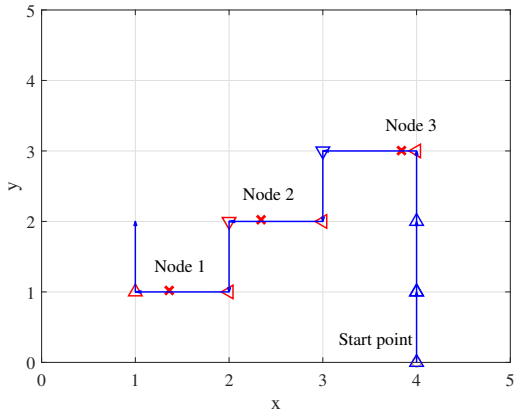
In this section, numerical results are provided to show the efficiency of the proposed algorithm. We compare the proposed method with the other two solutions to evaluate their performance. They are described as followed.

- "VoI-optimal" scheme denotes the proposed method in this chapter.
- "AoI-optimal" scheme in which the objective function is to minimise the maximum age of information among all nodes. Similarly to the proposed VoI-optimal scheme, we also employ reinforcement learning to solve the problem. The difference is that the minimum VoI is replaced by the negative maximum AoI in the design of the reward function.
- "Shortest path" scheme in which the objective is to minimise the total length of the mobility path. This scheme can be used as the benchmark to compare the performance between different methods.

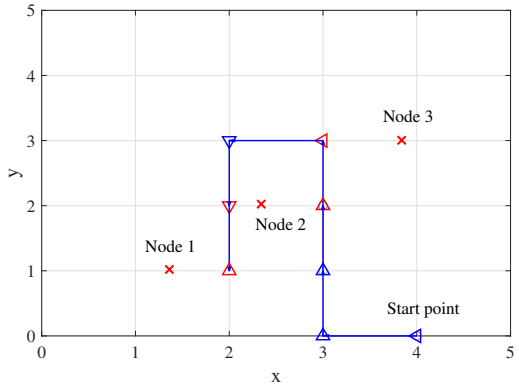
In the simulation, the fixed sensor nodes are deployed randomly in the 5×5 area. The starting position of the mobile central entity is also selected randomly. The simulation parameters are given as followed. The channel power gain g_0 at the reference distance $d_0 = 1$m is set to -50dB. The available bandwidth B is set to 2×10^6Hz. The noise power spectral density N_0 is set to -110dBm/Hz. The moving distance per time slot is set to $z = 1$. We assume that all nodes have the same transmission power. In this case, the SNR threshold γ_{th} only depends on the transmission distance. We use d_{th} to represent the distance threshold of the successful transmission and set d_{th} as 1. The learning rate β is set to 0.75, and the discount factor γ is set to 0.9. The total running round $R_{\max}$ is set to 10^6, the initial exploration probability ϵ is set to 1 and exploitation probability ϵ_0 is set to 10^{-6}.

Fig. 6.2 gives a case study to illustrate how the proposed algorithm and the benchmark method work in the same network. Fig. 6.3 further evaluates and compares the network performance in this given network scenario. Figs. 6.4 to 6.6 illustrate the average network performance in terms of the VoI, in which 40 network instances are generated randomly for each network setting.

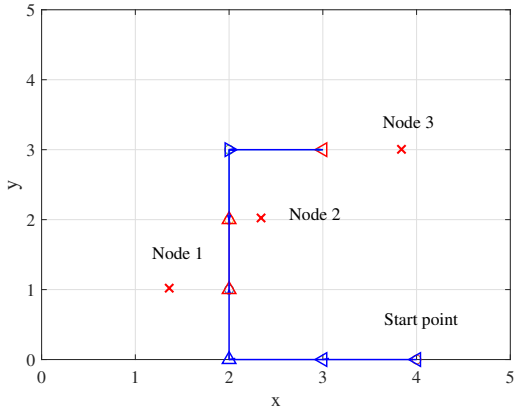
Fig. 6.2 gives three examples to show the mobility of the central entity and the



(a) The VoI-optimal scenario.



(b) The AoI-optimal scenario.



(c) The shortest path scenario.

Figure 6.2: A case study. The number of nodes $N = 3$, the correlation parameter of each node $\rho = [0.7, 0.6, 0.5]$, initial age of each node $A = [2, 4, 4]$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 10$.

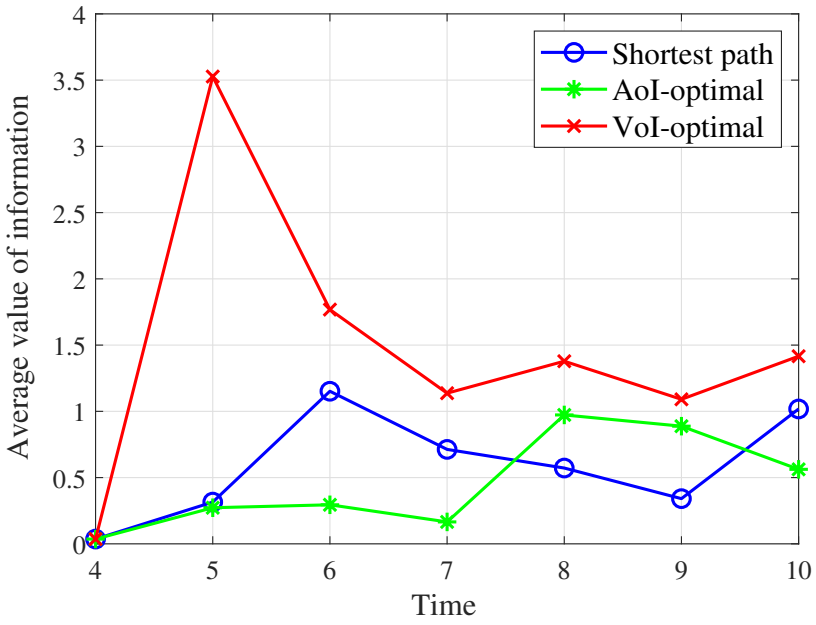


Figure 6.3: The average value of information evolution among all nodes in the network given in Fig. 6.2.

communication scheme between the sensing nodes and the central entity by three different solutions. The sensing devices are fixed and marked by the "red cross". The position of the central entity at each time slot is marked by the "triangle" and the trajectory is marked by the "blue line". The direction of the triangle represents the moving direction and the colour represents the communication indicator. "Red triangle" means that a node i is scheduled to communicate with the central entity at this position (i.e., $\alpha_i(t) = 1$); otherwise, the triangle is blue. For example, in Fig. 6.2(a), the central entity moves to $(4, 3)$ at time slot 4. In this position, node 3 transmits its data to the central entity and the entity will move to $(3, 3)$ at the next slot.

The solution of the proposed algorithm is visualised in Fig. 6.2(a), and the AoI-optimal and shortest path schemes are shown in Figs. 6.2(b) and 6.2(c), respectively. Fig. 6.3 shows how the average VoI evolves with time from the central entity's perspective, illustrating that the proposed algorithm achieves better performance. This

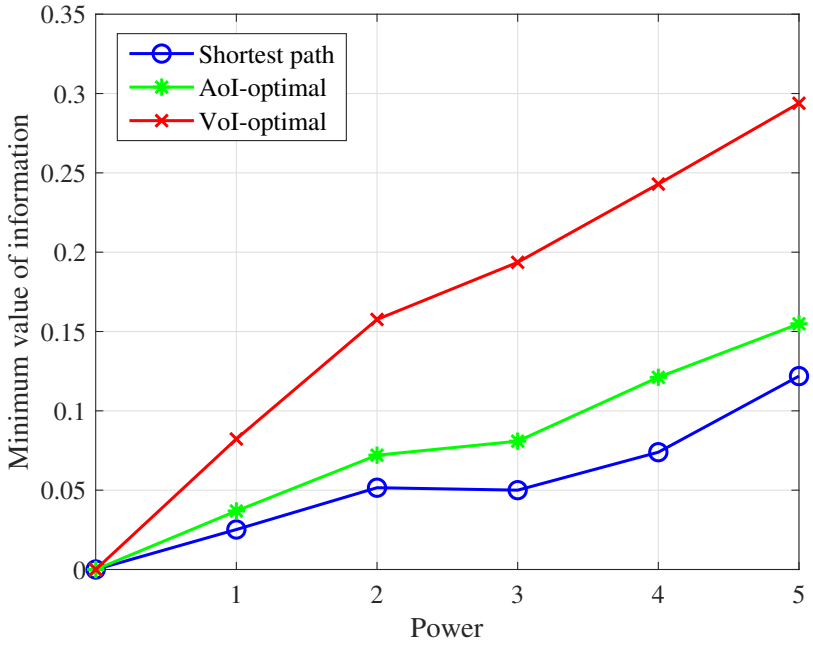


Figure 6.4: The time average minimum value of information among all nodes versus the transmission power. The number of nodes $N = 4$, the correlation parameter is generated uniformly from $[0, 1]$, and the total number of time slots $T_{\max} = 30$.

is because the VoI metric is able to capture the timeliness of data and the correlation properties of data. In this network setting, node 1 is highly correlated, node 2 is intermediately correlated and node 3 is less-correlated. Therefore, node 3 varies more quickly with time thus it should be firstly served with priority. However, the shortest path solution only considers the moving distance, and the AoI-optimal solution does not take the correlation into consideration. Moreover, in Fig. 6.2(a), the data collection position of node 3 is $(4, 3)$ while this position is $(3, 3)$ in the other two schemes. This means that the proposed VoI-optimal solution is able to capture the performance degradation caused by the different transmission distances. Therefore, this case study illustrates that the VoI metric reflects the correlation properties, and the timeliness of data in both time and space domains.

Fig. 6.4 presents how the minimum value of information varies with the transmission power. This figure shows that the VoI increases with the transmission power,

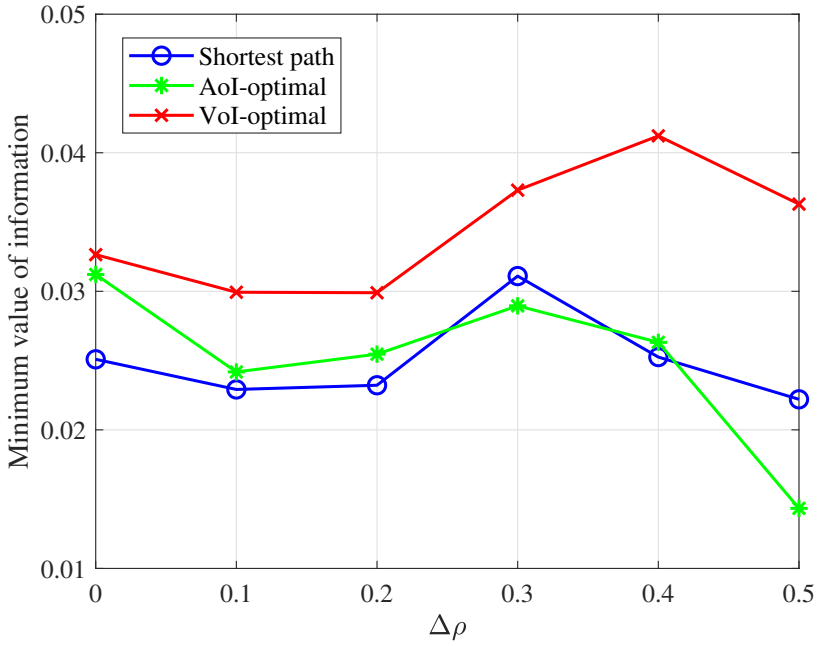


Figure 6.5: Time average minimum value of information among all nodes versus different correlation parameters. The number of nodes $N = 4$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 30$.

and this means that the VoI is able to capture the SNR in the transmission environment and verifies that the packet received via a good channel is more valuable than the packet received via a poor channel. However, the performance of the traditional AoI-optimal or shortest-path scheme is regardless of the transmission distance, thus they are unable to capture the usefulness of data from the spatial dimension.

Fig. 6.5 explores the performance of the VoI in the context of a network with heterogeneous data sources. The correlation parameter of each node is generated uniformly from $(0.5-\Delta\rho, 0.5+\Delta\rho)$. The horizontal axis represents how heterogeneous the data sources are. Large $\Delta\rho$ represents high heterogeneity; small $\Delta\rho$ represents less heterogeneity. The vertical axis is the time average minimum VoI among all nodes. When $\Delta\rho = 0$ (the data sources are homogeneous), the VoI can be written as a monotonic function of the AoI, thus the VoI-optimal solution achieves similar performance as its AoI counterpart. However, their performance gap increases with

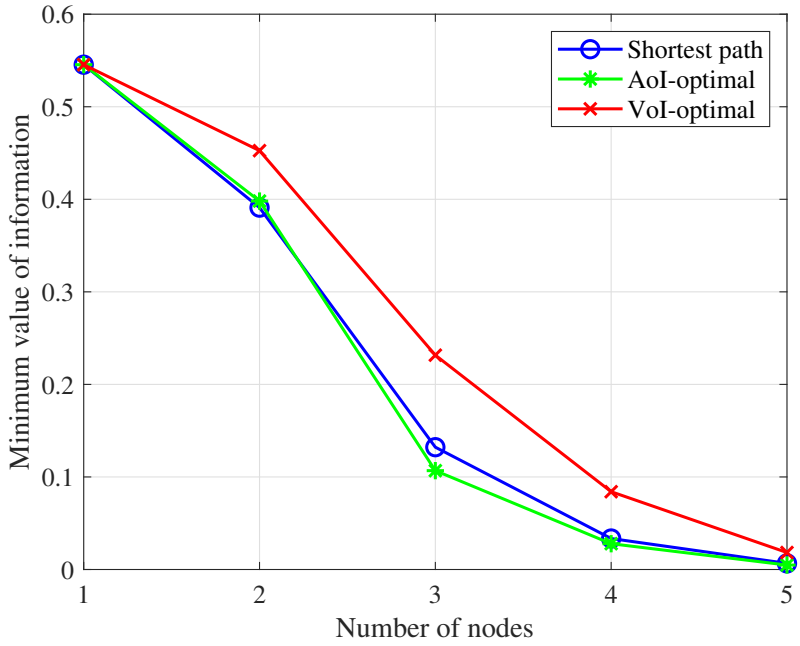


Figure 6.6: Time average minimum value of information among all nodes versus the number of nodes. The correlation parameter is generated uniformly from $[0, 1]$, the transmission power of each node $P = 1$W, and the total number of time slots $T_{\max} = 30$.

$\Delta\rho$. This is because the proposed method is more likely to serve the less-correlated node to balance the VoI. But the other two methods do not consider the correlation property and they are not able to help improve the VoI. This figure further shows that our proposed method is more appropriate to be applied in the network with different types of data sources.

Fig. 6.6 shows the performance of the minimum value of information versus the number of nodes. When the number of nodes is small, all three methods have similar performance. In this case, the wireless resource and the time slots are enough to serve the node as timely as possible. When increasing the number of nodes, the advantage of the proposed method becomes more significant. This is because the VoI is able to explore the correlation properties of the data source, thus less-correlated sources can be served with higher priority. Compared to the shortest path or the AoI-optimal scheme, the proposed method is complex in the algorithm design, but it is more

suitable to be applied in networks with limited wireless resources by considering the correlation of the underlying data source.

6.4 Summary

In this chapter, we studied an optimisation problem to show how VoI can be applied in practical IoT networks. We considered the network with multiple heterogeneous data sources and a mobile central entity. We studied the optimal path planning of the central entity and the communication strategy between the source and the central entity to maximise the minimum VoI in the network. To solve this problem efficiently, this problem was reformulated as an MDP, and a reinforcement learning-based algorithm was proposed. Compared with other benchmark methods, we showed that the proposed method achieves higher VoI especially when the wireless resource is limited by exploiting the correlation properties of the underlying data sources.

Chapter 7

Conclusions

7.1 Thesis Summary

In this thesis, we formulated a general value of information framework in the context of information theory, which can be used to measure the usefulness of data in real-time communication systems. Compared with the traditional age of information metric, the proposed VoI measurement is able to capture the timeliness of information, the non-linear performance degradation caused by the correlation properties of the underlying data source, and the noise in the transmission environment. We analysed the VoI performance in various latent variable models, including the general hidden Markov models, hidden Gauss-Markov models and the noisy Ornstein-Uhlenbeck process model. Moreover, we studied an optimisation problem to show how VoI can be applied in practical IoT networks.

Specifically, in Chapter 3, the basic concept of VoI was defined in general latent variable models based on mutual information. The VoI was interpreted as the reduction in the uncertainty of the current status given some past observations, which quantified how valuable the data is. Next, we focused on general hidden Markov models and derived the bound of the VoI. Finally, the VoI was investigated in general

hidden Gauss-Markov models. The closed-form expression of the VoI was obtained and its functional relationship to the estimation error was also obtained. We showed that the VoI relates to the number of observations, the correlation of the underlying random process, the AoI, and the SNR in the network.

In Chapter 4, we focused on general hidden Gauss-Markov models, and went into more depth with regard to different sampling policies and the effect of the number of observations on VoI. We obtained the simplified closed-form VoI expressions with arbitrary and uniform sampling intervals under high and low SNR conditions. Through these results, the effects of the number of observations and the SNR on the VoI were given. We found that the VoI can be improved by using more observations in the low SNR regime, while it does not relate to the number of observations in the high SNR regime. Meanwhile, increasing the SNR was able to improve the VoI. The trade-off between increasing SNR and adding one extra observation was obtained. Moreover, the optimal sampling problem at the data source was formulated and solved to maximise the VoI under the sampling rate constraint.

In Chapter 5, we looked at a specific hidden Gauss-Markov model, the noisy Ornstein-Uhlenbeck process, to investigate the VoI and its application in detail. First, we studied how the correlation and SNR affect the VoI and its relationship to AoI. We derived three non-linear functional relationships between the two in different SNR conditions, which gave the mathematical justification for how to choose non-linear AoI functions. Next, we investigated the distribution of the VoI in the example of the M/M/1 queueing system, deriving its closed-form PDF, CDF, MGF and bounds.

In Chapter 6, the VoI was optimised in IoT networks with multiple various transmitters and one mobile receiver. The minimum VoI was maximised by jointly optimising the mobility and user scheduling. This problem was reformulated as an MDP, and a reinforcement learning-based method was introduced to solve this problem more efficiently.

In this thesis, we formalised a novel and general framework to characterise the value of information in communication systems. The analysis and results in this thesis provided useful insights, which can be further used in the design of timeliness-critical applications.

7.2 Future Works

There are a lot of interesting directions in terms of the timeliness of information. This thesis reveals a small part of them, and there are still many problems that remain to be studied. In this section, we give some new directions which are worth future investigation.

In this thesis, the VoI is explored in the context of Gauss-Markov models, in which the underlying random process is a stationary and Gauss-Markov process. It would be interesting to look at other random processes to model the physical process more precisely in the real world, such as the higher-order Markov processes. In that case, the trade-off between the number of observations and the order of the random process can be further explored, which may provide many useful insights into VoI performance. Moreover, it would also be interesting to utilise real practical data to model the property of the underlying data source. By employing the method of machine learning, it is possible to use various distribution functions to fit the physical phenomenon. This direction provides the potential to apply VoI in the real world.

The VoI could be used in more complex communication systems which are closer to actual systems. In this thesis, the data packets are transmitted through the AWGN channel. A practical system is also affected by fading and interference. Therefore, we could take large-scale and small-scale fading into consideration. We could also consider the impact of interference in multiple-user networks. In this case, a more general framework can be formalised to characterise the value of information in com-

plex networks.

The VoI in large-scale networks with multiple sensing devices may also receive much interest. In this thesis, the VoI concept is explored in the single-user network and is able to reflect the timeliness of data and the correlation properties of the underlying data source. When it turns to large-scale networks, the data packets sampled by two devices which are close to each other are spatially correlated. Thus, the usefulness of data is not only reflected in the time domain but also in the space domain. In this case, the value of information could be a more systematic metric which is suitable for spatial-temporal networks.

In addition, the VoI could be jointly optimised with other features. In Chapter 6, we maximise the VoI by optimising the mobility and communication strategy in IoT networks. However, in IoT-assisted data collection networks, energy is also a key point to support the mobility of the central entity. There should be a trade-off between the value of information at the destination and the energy consumption of the central entity. How to find VoI-optimal and energy-efficient path planning is a direction worth further study. This problem can also be formulated as an MDP, but the Q-learning method may not be suitable due to the large state and action space. Therefore, the deep Q-learning method may be needed.

7.3 Conclusions

This thesis proposed a value of information framework based on information theory, and analysed the VoI performance for various hidden Markov models. We showed that the VoI is able to capture the timeliness of data, the correlation properties of the underlying data source and the noise in the transmission environment. Therefore, compared with the traditional AoI metric, we believe that the proposed VoI metric is more appropriate to measure how valuable the data is in communication systems.

However, it is worth noting that VoI still faces challenges with the aim of implementation and commercialisation in the practical world. First, the practical physical process may not be perfectly fitted by Gauss-Markov random process model, thus the closed-form VoI expression may not hold for other process models. Next, the proposed VoI framework is proposed based on mutual information which is not conditioned on the realisation of the random variables but is an average definition. Then, the practical transmission channel may be more complex than the AWGN channel model due to fading and interference. The simplified modelling assumption can be a limitation of the VoI framework. Moreover, the calculation complexity of the VoI may be extremely high in large-scale networks. Thus, the VoI still needs to be improved in order to meet the demand for practical networks. This would require analysis of VoI for other mathematical models with latent variables to reflect real systems, and also require more efficient resource allocation algorithms to optimise the VoI in practical networks.

The usefulness of data is significant for emerging real-time communication systems. This thesis provides a theoretical study of the information value. Although there are still challenges, we believe that the results presented in this thesis could be an important starting point to facilitate the design of future timely control applications.

Bibliography

[1] K. Ashton, "That 'Internet of Things' thing," *RFID journal*, Jun, 2009.

[2] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2020.

[3] 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 and Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[4] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with rf energy harvesting: A contemporary survey," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 757–789, 2015.

[5] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless charging technologies: Fundamentals, standards, and network applications," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 2, pp. 1413–1452, 2016.

[6] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[7] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2021.

[8] O. Novo, "Blockchain meets IoT: An architecture for scalable access management in IoT," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1184–1195, 2018.

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

[10] A. Kosta, N. Pappas, and V. Angelakis, *Age of Information: A New Concept, Metric, and Tool*. Now Foundations and Trends, 2017.

[11] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the Internet of Things," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 72–77, 2019.

[12] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

[13] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8305–8324, 2019.

[14] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. 46th Conf. Inf. Sci. Syst. (CISS)*, pp. 1–6, 2012.

[15] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5215–5232, 2019.

[16] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 576–580, 2017.

[17] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling for a class of continuous Markov processes," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2456–2461, 2020.

[18] E. Gindullina, L. Badia, and D. Gündüz, "Age-of-information with information source diversity in an energy harvesting system," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 3, pp. 1529–1540, 2021.

[19] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, 2018.

[20] Z. Wang, X. Qin, B. Liu, and P. Zhang, "Joint data sampling and link scheduling for age minimization in multihop cyber-physical systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 765–768, 2019.

[21] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 193–204, 2018.

[22] M. A. Abd-Elmagid and H. S. Dhillon, "Average peak age-of-information minimization in UAV-assisted IoT networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 2003–2008, 2019.

[23] Z. Jia, X. Qin, Z. Wang, and B. Liu, "Age-based path planning and data acquisition in UAV-assisted IoT networks," in *Proc. IEEE ICC Workshops*, pp. 1–6, 2019.

[24] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[25] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[26] C. Kam, S. Kompella, and A. Ephremides, "Age of incorrect information for remote estimation of a binary Markov source," in *Proc. IEEE INFOCOM Workshops*, pp. 1–6, 2020.

[27] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an effective age of information: Remote estimation of a Markov source," in *Proc. IEEE INFOCOM Workshops*, pp. 367–372, 2018.

[28] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, 2020.

[29] T. Z. Ornee and Y. Sun, "Sampling and remote estimation for the Ornstein-Uhlenbeck process through queues: Age of information and beyond," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 1962–1975, 2021.

[30] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal self-driven sampling for estimation based on value of information," in *International Workshop on Discrete Event Systems (WODES)*, pp. 183–188, 2016.

[31] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *Proc. IEEE SPAWC*, pp. 1–5, 2018.

[32] Z. Wang, M. A. Badiu, and J. P. Coon, "A value of information framework for latent variable models," in *Proc. IEEE GLOBECOM*, pp. 1–6, 2020.

[33] Z. Wang, M.-A. Badiu, and J. P. Coon, "On the value of information and mean squared error for noisy Gaussian models," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2023–2026, 2022.

[34] Z. Wang, M.-A. Badiu, and J. P. Coon, "A framework for characterizing the value of information in hidden Markov models," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, pp. 5203–5216, 2022.

[35] Z. Wang, M.-A. Badiu, and J. P. Coon, "Trading off SNR and the number of observations to improve the value of information in IoT networks," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, pp. 1–6, 2022.

[36] Z. Wang, M.-A. Badiu, and J. P. Coon, "Relationship between age and value of information for a noisy Ornstein–Uhlenbeck process," *Entropy*, vol. 23, no. 8, 2021.

[37] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, 2014.

[38] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surv. Tut.*, vol. 17, no. 4, pp. 2347–2376, 2015.

[39] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2569–2573, 2016.

[40] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2666–2670, 2012.

[41] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 1583–1587, 2014.

[42] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, 2016.

[43] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

[44] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 326–330, 2017.

[45] X. Zheng, S. Zhou, Z. Jiang, and Z. Niu, "Closed-form analysis of non-linear age of information in status updates with an energy harvesting transmitter," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4129–4142, 2019.

[46] J. Cho and H. Garcia-Molina, "Effective page refresh policies for web crawlers," *ACM Trans. Database Syst.*, vol. 28, p. 390–426, Dec. 2003.

[47] Y. Azar, E. Horvitz, E. Lubetzky, Y. Peres, and D. Shahaf, "Tractable near-optimal policies for crawling," *Proc. National Academy of Sciences*, vol. 115, no. 32, pp. 8099–8103, 2018.

[48] M. Bastopcu and S. Ulukus, "Information freshness in cache updating systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1861–1874, 2021.

[49] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.

[50] J. Hribar, M. Costa, N. Kaminski, and L. A. DaSilva, "Updating strategies in the Internet of Things by taking advantage of correlated sources," in *Proc. IEEE GLOBECOM*, pp. 1–6, 2017.

[51] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 321–325, 2017.

[52] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–8, 2019.

[53] T. Z. Ornee and Y. Sun, "Performance bounds for sampling and remote estimation of Gauss-Markov processes over a noisy channel with random delay," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2021.

[54] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6485–6499, 2021.

[55] A. Arafa, K. Banawan, K. G. Seddik, and H. Vincent Poor, "Timely estimation using coded quantized samples," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1812–1817, 2020.

[56] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6485–6499, 2021.

[57] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, 2020.

[58] A. Maatouk, M. Assaad, and A. Ephremides, "Semantics-empowered communications through the age of incorrect information," in *ICC 2022 - IEEE International Conference on Communications*, pp. 3995–4000, 2022.

[59] X. Zheng, S. Zhou, and Z. Niu, "Urgency of information for context-aware timely status updates in remote control systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7237–7250, 2020.

[60] X. Zheng, S. Zhou, and Z. Niu, "Beyond age: Urgency of information for timeliness guarantee in status update systems," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, pp. 1–5, 2020.

[61] X. Zheng, S. Zhou, and Z. Niu, "Context-aware information lapse for timely status updates in remote control systems," in *2019 IEEE GLOBECOM*, pp. 1–6, 2019.

[62] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[63] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[64] T. Soleymani, S. Hirche, and J. S. Baras, "Maximization of information in energy-limited directed communication," in *2016 European Control Conference (ECC)*, pp. 1001–1006, 2016.

[65] Y. Li, Y. Xu, Q. Zhang, and Z. Yang, "Age-driven spatially temporally correlative updating in the satellite-integrated internet of things via Markov decision

process," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13612–13625, 2022.

[66] Y. Liu, C. Dong, X. Qin, and X. Xu, "Efficient sensor scheduling strategy based on spatio-temporal scope information model," 2022.

[67] R. Singh, G. K. Kamath, and P. R. Kumar, "Optimal information updating based on value of information," in *Proc. Allerton Conf. Commun. Control and Comput.*, pp. 847–854, 2019.

[68] M. Giordani, T. Higuchi, A. Zanella, O. Altintas, and M. Zorzi, "A framework to assess value of information in future vehicular networks," in *Proc. 1st ACM MobiHoc Workshop on Technologies, MOdels, and Protocols for Cooperative Connected Cars*, TOP-Cars '19, (New York, NY, USA), p. 31–36, Association for Computing Machinery, 2019.

[69] B. Abolhassani, J. Tadrous, A. Eryilmaz, and E. Yeh, "Fresh caching for dynamic content," in *Proc. IEEE INFOCOM*, pp. 1–10, 2021.

[70] R. D. Yates, "The age of gossip in networks," in *Proc. IEEE Int. Symp. Inf. Theory,*, pp. 2984–2989, 2021.

[71] B. Buyukates, M. Bastopcu, and S. Ulukus, "Age of gossip in networks with community structure," in *Proc. IEEE SPAWC*, pp. 326–330, 2021.

[72] J. Zhong, R. D. Yates, and E. Soljanin, "Two freshness metrics for local cache refresh," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1924–1928, 2018.

[73] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an "effective age" concept," in *2018 IEEE 19th International Work-*

shop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 1–5, 2018.

[74] M. Bastopcu and S. Ulukus, "Minimizing age of information with soft updates," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 233–243, 2019.

[75] M. Bastopcu and S. Ulukus, "Age of information with soft updates," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 378–385, 2018.

[76] H. Zhang, Z. Jiang, S. Xu, and S. Zhou, "Error analysis for status update from sensors with temporally and spatially correlated observations," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 2136–2149, 2021.

[77] P. D. Mankar, M. A. Abd-Elmagid, and H. S. Dhillon, "Spatial distribution of the mean peak age of information in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4465–4479, 2021.

[78] H. H. Yang, C. Xu, X. Wang, D. Feng, and T. Q. S. Quek, "Understanding age of information in large-scale wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 3196–3210, 2021.

[79] W. J. Stewart, *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling.* Princeton, NJ, USA: Princeton University Press, 2009.

[80] M. Moltafet, M. Leinonen, and M. Codreanu, "Average age of information for a multi-source M/M/1 queueing model with packet management," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1765–1769, 2020.

[81] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 131–135, 2017.

[82] C. Kam, S. Kompella, G. D. Nguyen, and A. Ephremides, "Effect of message transmission path diversity on status age," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1360–1374, 2016.

[83] E. Najm and R. Nasser, "Age of information: The Gamma awakening," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2574–2578, 2016.

[84] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6419–6428, 2018.

[85] B. Wang, S. Feng, and J. Yang, "To skip or to switch? Minimizing age of information under link capacity constraint," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2018.

[86] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Controlling the age of information: Buffer size, deadline, and packet replacement," in *Proc. IEEE MILCOM*, pp. 301–306, 2016.

[87] S. K. Kaul and R. D. Yates, "Age of information: Updates with priority," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2644–2648, 2018.

[88] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 1681–1685, 2015.

[89] K. Chen and L. Huang, "Age-of-information in the presence of error," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2579–2583, 2016.

[90] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 571–575, 2017.

[91] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for age of information for the D/G/1 queue," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 130–135, 2018.

[92] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization," *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7477–7489, 2021.

[93] J. P. Champati, H. Al-Zubaidy, and J. Gross, "On the distribution of AoI for the GI/GI/1/1 and GI/GI/1/2* systems: Exact expressions and bounds," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 37–45, 2019.

[94] M. A. Abd-Elmagid and H. S. Dhillon, "Closed-form characterization of the MGF of AoI in energy harvesting status update systems," *IEEE Transactions on Information Theory*, vol. 68, no. 6, pp. 3896–3919, 2022.

[95] M. A. Abd-Elmagid and H. S. Dhillon, "Distributional properties of age of information in energy harvesting status update systems," in *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, pp. 1–8, 2021.

[96] Q. He, D. Yuan, and A. Ephremides, "On optimal link scheduling with min-max peak age of information in wireless systems," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–7, 2016.

[97] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.

[98] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. Allerton Conf. Commun. Control and Comput.*, pp. 486–493, 2017.

[99] J. Lou, X. Yuan, S. Kompella, and N. Tzeng, "AoI and throughput tradeoffs in routing-aware multi-hop wireless networks," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pp. 476–485, 2020.

[100] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *2015 Information Theory and Applications Workshop (ITA)*, pp. 25–31, 2015.

[101] W. He, C. Guo, and X. Wang, "Age of information aware resource allocation and packet sampling control in vehicular networks," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2245–2249, 2022.

[102] X. Qin, Y. Xia, H. Li, Z. Feng, and P. Zhang, "Distributed data collection in age-aware vehicular participatory sensing networks," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14501–14513, 2021.

[103] M. Liu, D. Li, H. Wu, F. Lyu, and X. Shen, "Real-time search-driven caching for sensing data in vehicular networks," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12219–12230, 2022.

[104] Y. Huang, C. Xu, C. Zhang, M. Hua, and Z. Zhang, "An overview of intelligent wireless communications using deep reinforcement learning," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 15–29, 2019.

[105] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[106] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[107] J. Liu, X. Wang, B. Bai, and H. Dai, "Age-optimal trajectory planning for UAV-assisted data collection," in *Proc. IEEE INFOCOM Workshops*, pp. 553–558, 2018.

[108] P. Tong, J. Liu, X. Wang, B. Bai, and H. Dai, "UAV-enabled age-optimal data collection in wireless sensor networks," in *Proc. IEEE ICC Workshops*, pp. 1–6, 2019.

[109] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep reinforcement learning for minimizing age-of-information in UAV-assisted networks," in *Proc. IEEE GLOBECOM*, pp. 1–6, 2019.

[110] M. Sun, X. Xu, X. Qin, and P. Zhang, "AoI-energy-aware UAV-assisted data collection for IoT networks: A deep reinforcement learning method," *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17275–17289, 2021.

[111] J. Liu, P. Tong, X. Wang, B. Bai, and H. Dai, "UAV-aided data collection for information freshness in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2368–2382, 2021.

[112] C. Zhou, H. He, P. Yang, F. Lyu, W. Wu, N. Cheng, and X. Shen, "Deep RL-based trajectory planning for AoI minimization in UAV-assisted IoT," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, 2019.

[113] W. Li, L. Wang, and A. Fei, "Minimizing packet expiration loss with path planning in UAV-assisted data sensing," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1520–1523, 2019.

[114] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep reinforcement learning for minimizing age-of-information in UAV-assisted networks," in *Proc. IEEE GLOBECOM*, pp. 1–6, 2019.

[115] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: John Wiley & Sons, 2006.

[116] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. New York, NY, USA: John Wiley & Sons, 2015.

[117] J. L. Doob, "The Brownian movement and stochastic equations," *Annals of Mathematics*, vol. 43, no. 2, pp. 351–369, 1942.

[118] K. I. Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Cham, Switzerland: Springer, 2018.

[119] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[120] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Information acquisition with sensing robots: Algorithms and error bounds," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6447–6454, 2014.

[121] B. Allévius, "On the precision matrix of an irregularly sampled AR(1) process," 2018.

[122] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*. Boca Raton, FL, USA: CRC press, 2005.

[123] J. Yun, C. Joo, and A. Eryilmaz, "Optimal real-time monitoring of an information source under communication costs," in *Proc. IEEE Conf. Decision and Control*, pp. 4767–4772, 2018.

[124] H. Tang, J. Wang, L. Song, and J. Song, "Scheduling to minimize age of information in multi-state time-varying networks with power constraints," in *Proc. Allerton Conf. Commun. Control and Comput.*, pp. 1198–1205, 2019.

[125] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of Information: A New Metric for Information Freshness*. San Rafael, CA, USA: Morgan & Claypool, 2019.

[126] Y. Wei, X. Jiang, Z. Jiang, and S. Shon, "Determinants and inverses of perturbed periodic tridiagonal Toeplitz matrices," *Advances in Difference Equations*, vol. 2019, no. 1, p. 410, 2019.

[127] B. Zuo, Z. Jiang, and D. Fu, "Determinants and inverses of Ppoeplitz and Ppankel matrices," *Special Matrices*, vol. 6, no. 1, pp. 201–215, 2018.

[128] A. Papoulis, *Probability, random variables, and stochastic processes*. McGraw-Hill, 1991.

[129] J. Conway and R. Guy, *The Book of Numbers*. Springer New York, 1998.

[130] L. Andrews, *Special Functions of Mathematics for Engineers, Second Edition*. Oxford University Press and SPIE Optical Engineering Press, 1998.

[131] Z. Cui, K. Guan, C. Briso-Rodríguez, B. Ai, and Z. Zhong, "Frequency-dependent line-of-sight probability modeling in built-up environments," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 699–709, 2020.

[132] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. New York, NY, USA: John Wiley & Sons, 2014.