課程博士

2023 年 3 月

関西大学審査学位論文

Energy-efficient Content-based Wake-up Control for Wireless Sensor Networks

Junya Shiraishi

理工学研究科　総合理工学専攻

電気電子情報工学分野　情報通信工学領域

20D6004　白石 順哉

Abstract

Energy-efficiency of sensor nodes is one of the essential key requirements in Wireless Sensor Networks (WSNs) where the sensor nodes operate with batteries. In order to realize high energy-efficiency of sensor nodes, this thesis focuses on wake-up radio technology that remotely activates sensor nodes using radio waves only when needed, which is a technology for reducing the wasteful energy consumption during idle periods of sensor nodes. Wake-up radio technology can be realized by installing an ultra-low power wake-up receiver, which is dedicated to the operation waiting for communication requests, into each sensor node. Conventional Identity-based Wake-up (IDWu) specifies the node to be activated by IDentification (ID). However, this method cannot suppress wasteful wake-up of nodes storing undesired data for the sink because it cannot activate the sensor nodes based on their observations (sensing data). In order to solve this problem, this thesis proposes Content-based Wake-up (CoWu), which realizes wake-up control based on the observation of each node. In CoWu, the sink transmits a wake-up signal embedding the information on the target of content, i.e., the threshold. This signal is detected at wake-up receivers, and only nodes that observe desired data by the sink, i.e., nodes observing data more than a given threshold wake up and transmit data, while others suppress their wake-up, which saves wasteful energy consumption. We apply CoWu to four types of data collection scenarios: (A) data collection in the environment where the sink requests information on nodes that observe specific values, (B) data collection in the environment where the observed data have temporal correlation, (C) data collection in the environment where the observed data have spatial correlation, and (D) data collection in the environment where high freshness is required for the collected data. We investigate the interaction of CoWu and characteristics of content observed by sensors, such as the distribution of observed data, resolution of sensors, and spatio-temporal correlation of observed data. We evaluate the effectiveness of CoWu with a variety of setting, and analyze the gain of CoWu theoretically. With obtained numerical results, we reveal the energy-efficiency achieved by CoWu in WSNs.

The thesis is organized as follows. Chapter 1 gives introduction of this thesis, including the positioning of CoWu within the framework of WSNs. Then, we describe the related work and research objective. Finally, we briefly introduce the outline of the thesis. In chapter 2, we explain the background of this study, including WSNs, the necessity of data-oriented communication, and wake-up radio technology. In chapters 3 to 6, we investigate characteristics of CoWu by introducing it to the four different scenarios mentioned above. Chapter 3 applies CoWu to top-k query scenario and proposes Count-down CoWu (CDCoWu) designed

for top-k query. We investigate the characteristics of data collection employing CDCoWu and show its effectiveness by comparing with the conventional wake-up control. In Chapter 4, we extend the scenario in Chapter 3 to more general case, i.e., repeated top-k query. We propose a wake-up control exploiting temporal correlation of sensing data, which aims to realize both high energy efficiency and top-k ranking accuracy. Furthermore, we conduct experiments using a prototype of sensor node employing wake-up receiver to confirm the practicality of our proposed scheme. Chapter 5 focuses on the data collection environment where the observation of sensing data have spatial correlation, and the sink conducts the identification of multiple emission sources by collecting data from the sensor nodes. In chapter 6, we apply CoWu to a scenario where the freshness is required for the collected data. We investigate the effect of timing of CoWu signaling with regard to the deadline of data collection with theoretical analysis considering the evolution of physical process. We clarify the effectiveness of CoWu against the base line scheme in terms of the speed of physical process. In Chapter 7, we conclude this thesis regarding energy-efficient CoWu.

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Related Work . . . 4
  - 1.2.1 Power Reduction Technology during no Communication Period 4
  - 1.2.2 Energy-efficient Wake-up Control . . . 5
  - 1.2.3 Top-k Query in WSNs . . . 5
  - 1.2.4 Energy-efficient Protocol Designs Exploiting the Nature of Spatio-temproal Correlation of Observed Data . . . 6
  - 1.2.5 Information Freshness in the Context of IIoT . . . 7
- 1.3 Research Objective . . . 8
- 1.4 Thesis Outline . . . 8

2 Energy-efficient Data Collection in WSNs **11**
- 2.1 WSNs . . . 11
  - 2.1.1 Introduction to WSNs . . . 11
  - 2.1.2 Frequency Bands Used in WSNs . . . 12
  - 2.1.3 Communication Standards Used in WSNs . . . 13
- 2.2 Data Collection Scenario Considered in This Thesis . . . 16
  - 2.2.1 Data-oriented WSNs . . . 16
  - 2.2.2 Top-k Query in WSNs . . . 16
  - 2.2.3 Multiple Emission Sources Identification in WSNs . . . 17
  - 2.2.4 Data Collection Considering Timeliness in Wireless Sensor and Actuator Networks (WSAN) . . . 18
- 2.3 On-demand WSNs . . . 19
  - 2.3.1 Energy Saving with On-demand WSNs . . . 20
  - 2.3.2 Wake-up Receiver . . . 20
  - 2.3.3 Conventional Wake-up Control . . . 22
- 2.4 Problem Definition . . . 25
  - 2.4.1 The Problem Applying Conventional IDWu to a Single Round of Top-k Query . . . 25

2.4.2 The Problem Conducting Periodical Top-k Query in On-demand WSNs with Temporally Correlated Data 26
2.4.3 The Problem for the Multiple Emission Sources Identification in On-demand WSNs 27
2.4.4 The Problem of Data Collection Considering the Information Freshness and Energy Efficiency 27
2.5 Summary 28

3 Content-based Wake-up for Top-k Query in WSNs 29
3.1 System Model 29
3.2 Proposed Content-based Wake-up (CoWu) 31
3.2.1 Content-based Wake-up 31
3.2.2 Count-Down Content-based Wake-up (CDCoWu) 31
3.3 Theoretical Analysis of IDWu and CDCoWu 35
3.3.1 One-Shot Data Collection with p-persistent CSMA 35
3.3.2 Analysis of Conventional and Proposed Wake-up Control . . 35
3.4 Numerical Results 40
3.4.1 Approximate Analysis with MCMC 40
3.4.2 Impact of Different Parameters on the Performance of CDCoWu 42
3.4.3 Comparison of Different Wake-up Schemes 49
3.4.4 Impact of k and N 56
3.5 Summary 57

4 Wake-up Control for Repeated Top-k Query Exploiting Temporal Correlation of Sensing Data 59
4.1 Motivation to Consider Approximate Top-k Query 59
4.2 System Model 60
4.3 Definition of Accuracy of Top-k Set 61
4.3.1 Root Mean Square Error (RMSE) 61
4.3.2 Generalized Footrule Distance (GF-distance) 61
4.4 Proposed Scheme 62
4.4.1 Mode Control Signal 63
4.4.2 Positioning-based Filtering Phase (PF-phase) 64
4.4.3 Threshold-based Filtering Phase (TF-phase) 67
4.4.4 Validation Phase 68
4.5 Numerical Results 70
4.5.1 Trade-off Between Average Energy Consumption and GF-distance through β 71
4.5.2 Comparison of Different Wake-up Control 72
4.6 Feasibility Study of OTC-Wu with Experiments 74

4.6.1 Experimental Model 75
4.6.2 Experimental Results 77
4.7 Summary 79

5 Wake-up Control for Identifying Multiple Emission Sources of Spatially Correlated Observations 80
5.1 System Model 80
5.2 Identification Method for Multiple Emission Sources 81
5.2.1 LM-based Approach 82
5.2.2 CNN-based Approach 82
5.3 Selection of Nodes to Activate with Wake-up Control 84
5.3.1 UCWu based Random Sampling (UCWu-RS) 85
5.3.2 CountDown Content-based Wake-up 85
5.4 Numerical Evaluation 85
5.4.1 Simulation Model 85
5.4.2 Comparison of Different Identification Methods 86
5.4.3 Comparison of Different Wake-up Control 88
5.5 Summary 89

6 Wake-up Control for Data Collections by a Deadline 91
6.1 Introduction 91
6.2 System Model and Problem Definition 92
6.2.1 Scenario and Objective 92
6.2.2 CoWu Transmission Model 94
6.3 Analysis of CoWu 94
6.4 Numerical Results 97
6.4.1 Baseline Scheme for Evaluation: Round-robin Scheduling . . 97
6.4.2 Comparison Between CoWu and Round-robin Scheduling . . 98
6.5 Summary 100

7 Conclusions and Future Work 102
7.1 Conclusions 102
7.2 Future Work 105

Acknowledgments 107

List of Publications 109

List of Figures

1.1 Positioning of this research. 3
1.2 Outline of this thesis. 8

2.1 An example of operation of CSMA/CA. 14
2.2 A flowchart of IEEE 802.15.4 MAC operations. 15
2.3 An example of operation of top-k query. 17
2.4 An example of observations of sensors in a scenario where multiple emission sources exist. 18
2.5 The basic function of a WSAN. 19
2.6 An example of operations of on-demand WSNs. 21
2.7 An example of configuration of wake-up receiver to detect frame length. 21
2.8 An example of detection of frame length. 22
2.9 An example of data collection employing UCWu. 23
2.10 An example of communication in data collection employing UCWu. 23
2.11 An example of data collection employing BCWu. 24
2.12 An example of communication in data collection employing BCWu. 25
2.13 An example of the problem applying IDWu for top-k query. 26

3.1 Mechanism to decide wake-up frame length in CoWu. 32
3.2 An example of CoWu operations. 32
3.3 An example of CDCoWu operations. 33
3.4 Theoretical analysis model of BCWu. 36
3.5 Theoretical analysis model of UCWu. 37
3.6 Theoretical analysis model of CDCoWu. 37
3.7 Data collection delay against CD_{step} for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$). 43
3.8 Total energy consumption against CD_{step} for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$). 44
3.9 Data collection delay against b_q for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$). 47

3.10 Total energy consumption against b_q for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$). 47
3.11 Data Collection delay against α for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = 10V_{step}$, and $p = 0.0606$). 49
3.12 Data collection delay against k for N-CDCoWu and V-CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$). 50
3.13 Total energy consumption against k for N-CDCoWu and V-CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$). 50
3.14 Delay against the number of sensor nodes for BCWu and UCWu. . 52
3.15 Total energy consumption against the number of sensor nodes for BCWu and UCWu. 53
3.16 Total energy consumption against the number of sensor nodes for BCWu and CDCoWu. 54
3.17 A zoomed view of Fig. 3.16 in terms of CDCoWu. 54
3.18 Total energy consumption against the number of sensor nodes for UCWu and CDCoWu. 55
3.19 Maximum k-N ratio of CDCoWu to UCWu in terms of data collection delay and total energy consumption. 58

4.1 A block diagram of our proposed scheme. 62
4.2 An example to set a transmission suppression range based on the information on the ranking ($\beta = 0.5$). 65
4.3 An example of dynamic quantization and mapping in accordance with the distribution of top-k set. 66
4.4 An example of collected data at the beginning of the OV-phase. . . 69
4.5 An example of collected data at the beginning of the FV-phase. . . 69
4.6 Average energy consumption and GF-distance of OTC-Wu against β. 72
4.7 Average energy consumption of each wake-up control against the degree of temporal correlation. 74
4.8 Energy reduction ratio against σ_{DOV}. 75
4.9 Experimental setup with its network feature. 76
4.10 wake-up success ratio against different values of T_{step}. 78

5.1 An overview of model of observation values ($N = 10, N_p = 2$). . . . 81
5.2 CNN model adapted in our study. 83
5.3 Operations of each wake-up control ($N = 9$, $N_p = 2$). 84
5.4 The achieved set of energy consumption and accuracy for different types of wake-up control. 89

6.1 An example of data collection employing CoWu for data collections by a deadline. 93

6.2 Accuracy of CoWu against ζ_T. 99
6.3 Accuracy of CoWu and round-robin scheduling against q. 100

List of Tables

2.1 Frequency band usage conditions. 13
2.2 Modulation scheme of IEEE 802.15.4g. 13

3.1 Parameters employed for Numerical Evaluations. 43

4.1 Parameters employed for computer simulations. 71
4.2 Measured WRT for each wake-up control. 79

5.1 Parameters employed for computer simulations. 86
5.2 Estimation accuracy of multiple sources of sensing data using trained labels. 88

List of Acronyms

ACK ACKnowledgment

ADC Analog-to-Digital Converter

AI Artificial Intelligence

AoCI Age of Changed Information

AoI Age of Information

AoII Age of Incorrectness Information

AoS Age of Synchronization

AR Auto Regression

BCWu BroadCast Wake–up

BCWuID Broadcast Wake–up ID

BCWuSg BCWu signal

BFSK Binary Frequency Shift Keying

BPF Band Pass Filter

CCA Clear Channel Assessment

CoUD Cost of Update Delay

CDCoWu Count-Down CoWu

CoMoSg CoWu Mode Signal

CNN Convolutional Neural Network

CoWu Content-based Wake-up

CPU Central Processing Unit

CSMA Carrier Sense Multiple Access

CSMA/CA Carrier Sense Multiple Access/ Collision Avoidance

EXTOK EXact Top-k

F-distance Footrule distance

FILA FILtering Approach

FN False Negative

F-node Filtering node

FP False Positive

FV-phase Final Validation phase

GF-distance Generalized Footrule distance

ID IDentification

IDWu Identity-based Eake-up

I/F InterFace

IIoT Industrial Internet of Things

IoT Internet of Things

ISM Industrial Science Medical

ISO-OSI International Standardization Organization - Open Systems Interconnection

LM Local Maxima

LNA Low Noise Amplifier

LPF Low Pass Filter

MAC Media Access Control

MC Monte Carlo

MCMC Markov Chain Monte Carlo

MCU Micro Controller Unit

N-CDCoWu Node–set CDCoWu

OOK On-Off-Keying

OSD One-Shot Data

OTC-Wu Order and Temporal Correlation-based Wake-up

OV-phase Ordering Validation Phase

PAoI Peak AoI

PDF Probability Density Function

PF–phase Positioning–based Filtering Phase

PMF Probability Mass Function

PoMo Positioning Mode

PoMoSg PoMo Signal

PSM Power Saving Mode

PRIM PRIority–based top-k Monitoring

RI Receiver-Initiated

RMSE Root Mean Square Error

RSSI Received Signal Strength Indicator

TAG Tiny AGgregation

TF-phase Threshold–based Filtering Phase

TM-node Temporal Monitoring node

TI Transmitter-Initiated

UAV Unmanned Aerial Vehicles

UCWu UniCast Wake-up

UCWuID UniCast Wake–up ID

UCWu-RS UCWu based Random Sampling

UCMoSg UCWu Mode Signal

V-CDCoWu Value–set CDCoWu

VoI Value of Information

WRT Wake-up Response Time

WSAN Wireless Sensor and Actuator Networks

WSNs Wireless Sensor Networks

WuS Wake-up Signal

YOLO You Only Look Once

Chapter 1

Introduction

1.1 Background

As the concept of Internet of Things (IoT) has pervaded, the expectation for the upcoming IoT society is expanded, by which our quality of life is improved dramatically. In IoT society, the information on the physical world observed by sensors, such as temperature, humidity, and pressure, are collected in real-time and analyzed with Artificial Intelligence (AI) technology. These analyzed data are expected to bring a solution for a variety of social problems, including medical, agriculture, and logistics, and contribute to the efficiency of various businesses and the improvement of services. The IoT technology is also important in the context of Industrial Internet of Things (IIoT), where the closed loop operation among sensor nodes, sink node, environment, and actuator exist, and real-time operations are required. The fundamental technology for data collection to support these emerging applications is Wireless Sensor Networks (WSNs). Applications of WSNs are in a wide variety of fields, including medical, agricultural, military, and logistics, which will expand further in the future. Along with the increase of applications of WSNs and the diversification of types of collected data, it becomes challenging to satisfy the requirement from the user or network. For example, collecting data from all sensor nodes in every data collection instance is clearly undesirable because it increases amount of traffic. A more attractive strategy is the introduction of data (content)-oriented communication, in which the sink allocates limited wireless resources, such as energy and bandwidth, only to communication of data that has a value and meaning for the sink. In this aspect, data-oriented protocol designs considering the information on the application layer, including the distribution of observed data, temporal or spatial correlation of observed data [1], and freshness [2] become important.

The protocol design considering energy aspects is also crucial in WSNs, as

sensor nodes operates with batteries. To this end, this thesis aims to clarify the energy efficient wireless technology to be applied for data-oriented communication. Fig. 1.1 shows the positioning of this research. There are two types of directions to realize energy-efficient technology: (i) power reduction technology for communication and (ii) power reduction technology during idle period, i.e., no communication period. This thesis focuses on (ii), i.e., we aim to reduce wasteful power consumption in idle period. The basic approach of (ii) is duty-cycling in which main radio of each node is periodically turned on/off. This, however, is not able to realize both low-latency and high energy-efficiency of sensor nodes as there is an inherent trade-off between them: lower duty-cycle achieves smaller energy consumption but suffers from higher latency, while higher duty cycling achieves lower latency but suffers from increasing energy consumption. As a more attractive solution, this thesis introduces the *wake-up radio* technology into WSNs, which we call on-demand WSNs. In on-demand WSNs, an ultra-low power wake-up receiver is installed into each sensor node. When there is no communication request, each node turns off its main radio InterFace (I/F) and keeps only the wake-up receiver to be active. The wake-up receiver activates its main radio I/F via the host Central Processing Unit (CPU) only if it detects communication requests from the sink. By activating the main radio I/F only when communication is needed, we can reduce the wasteful energy consumption of idle period and realize energy-saving of the overall network. As a wake-up control in on-demand WSNs, we assume wake-up control exploiting frame length, i.e., the length of the energy burst, proposed in [3]. In the wake-up control exploiting the frame length, each node decides whether it should wake up or not by comparing its detected frame length and its wake-up frame length and activates its main radio I/F only if a specific wake-up condition is satisfied.

The conventional wake-up control investigated in literature is Range-based wake-up and Identity (ID)-based Wake-up (IDWu). In the Range-based wake-up, the wake-up receiver activates its main radio I/F when it detects a signal level (energy level) higher than a predetermined threshold. On the other hand, IDWu is a wake-up control that specifics node to wake up according to the IDentification (ID), e.g., Media Access Control (MAC) address, of each sensor node. In IDWu, the sink transmits wake-up signal embedding the information of target ID. The wake-up receiver detecting the wake-up signal addressing itself activates its main radio I/F and transmits data. In the example of the wake-up control exploiting the length of frame, to realize the operation of IDWu, we first need to prepare a mapping table between frame length and ID. Then, the sink generates and transmits a wake-up signal whose length corresponds to the target node ID. However, in IDWu, the sink cannot activate nodes in accordance with the observation values (sensing data), which causes wasteful energy consumption of nodes activated, but observing data without any value for the sink. To solve this problem, this thesis

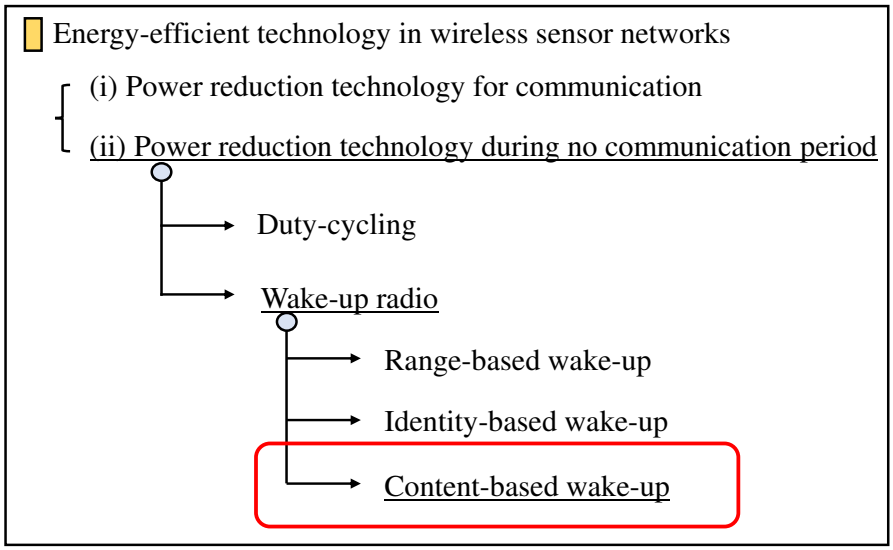


Figure 1.1: Positioning of this research.

proposes Content-based Wake-up control (CoWu) for on-demand WSNs, which enables the sink to wake up only nodes that have its desired data. In CoWu, the sink transmits a wake-up signal embedding the information on target content, i.e., the threshold of sensing data. The wake-up receiver detecting this wake-up signal compares the threshold with its sensing data, and only sensor node holding its desired data activates its main radio I/F and transmits data to the sink. This thesis applies CoWu to four types of data collection scenarios: (A) data collection in the environment where the sink requests information on nodes that observe specific values (c.f. Chapter 3), (B) data collection in the environment where the observed data have temporal correlation (c.f. Chapter 4), (C) data collection in the environment where the observed data have spatial correlation (c.f. Chapter 5), and (D) data collection in the environment where high freshness is required for the collected data (c.f. Chapter 6). These four types of data collection scenarios are important in terms of data-oriented communication, in which the importance of data to be collected depends on the user's requirements. Therefore, when collecting data in these scenarios, it is desirable for the sink to allocate limited radio resources in accordance with the observed data of each node. We believe data collection by employing CoWu plays an important role in these applications for reducing the wasteful energy consumption of nodes, and a thorough investigation of the performance of CoWu in these four types of scenarios is required. The brief outline for the study in the considered scenarios is as follows. First, in scenario (A), we propose a data collection algorithm, in which CoWu is applied for grasping the top-k node-set or value-set, i.e., the set of nodes and/or values observing top-k

highest readings, and characterize the performance with theoretical analysis and computer simulations. Next, in scenario (B), we propose a wake-up control that realizes both high top-k ranking accuracy and high energy efficiency by exploiting temporal correlation of observed values, and evaluate its effectiveness with computer simulations. Furthermore, we investigate the practicality of our proposed scheme by experiments with the prototype of sensor node including a wake–up receiver. Next, in scenario (C), we apply CoWu to data collection for the identification of multiple emission sources of sensing data, and investigate methods and wake-up control realizing high identification accuracy and high energy efficiency with computer simulations. Finally, in scenario (D), we apply CoWu to data collections for robots and actuators, in which the sink needs to grasp the specific information with high freshness at a specific instance. We optimize the timing of transmission of the wake-up signal through theoretical analysis, and clarify the gain of CoWu. Based on the above results, we confirm that CoWu improves the utilization efficiency of the limited radio resources (energy and frequency) in comparison to conventional control and algorithms.

1.2 Related Work

In this section, we describe related work to research problems tackled in this thesis [4][5][6][7].

1.2.1 Power Reduction Technology during no Communication Period

The basic approach for improving the energy-efficiency of wireless networks is duty-cycling [8] with a periodic on/off switching of the main radio. For instance, a Power Saving Mode (PSM) is introduced in cellular networks, where terminals in the idle (sleep) state periodically monitor the paging channel to check for incoming data [9]. In duty-cycling, idle listening leads to significant overall energy consumption. Furthermore, since no information can be exchanged until the nodes switch to active period, duty-cycling suffers from the increased latency. On the other hand, in the case of recent concept of wake-up radio [10][11], the power consumption of secondary radio module is order of magnitude lower than the main radio. Hence, the wake-up radio significantly reduces the energy for idle listening. Moreover, as the main radio is activated through wake-up signaling when it is needed, the latency is reduced. The existing studies on wake-up receivers [11] treat hardware design, routing and MAC protocols. This thesis focuses on the design of wake-up receiver considering the interaction with PHY/MAC protocols to be applied for WSNs.

1.2.2 Energy-efficient Wake-up Control

The MAC protocols with wake-up signaling can be Transmitter-Initiated (TI) or Receiver-Initiated (RI). In TI protocols [12], a transmitter sends wake-up signal to the receiver. In RI protocols, the data receiver sends wake-up signals to possible transmitters; this is suitable for the considered scenario of data collection from a field with dormant sensors. The RI protocol where each sensor node is woken up with an individual ID is called ID-based wake-up [10] and its performance is evaluated in [13]. As each request wakes up a single node, there is no contention among activated nodes. A similar type of ID-based RI protocol is employed in opportunistic wake-up MAC [14]. In [15], nodes are simultaneously woken up through broadcast ID and their transmissions are controlled to alleviate the contention. An RI protocol based on broadcast wake-up combined with contention resolution is used for neighbor discovery before unicast-based wake-up signaling in [16]. This thesis focuses on content-based RI protocol, i.e., CoWu. Note that a wake-up process activating nodes according to their sensed data, which can be considered as CoWu, was used for clustering in [17]. However, [17] is limited to clustering of nodes with similar readings. On the other hand, our work considers different scenarios and applies CoWu to the data collection in WSNs, such as top-k query as described in Sec. 1.1.

1.2.3 Top-k Query in WSNs

A top-k query aims to find the extreme values within a data set, and it has been widely studied in database processing and management for distributed systems [18][19]. For top-k query in WSNs [20][21][22], the main challenge is to reduce the communication overhead by avoiding transmissions that are unnecessary for constructing the top-k data set. [23] uses an approach termed tiny aggregation (TAG) to exploit data aggregation, by which each intermediate node along a constructed route compares its own data with the received data, forwarding only the data that are likely to be included into top-k data set. There are also solutions based on arithmetic filters, e.g., filtering approach (FILA) [24], exact top-k (EXTOK) [25], and their enhanced versions, (e.g., [26][27][28]), where each sensor node transmits its sensing data only if its value is within a filter. These schemes are useful for repeated top-k monitoring where sensed data are temporally correlated. EXTOK is a filtering-based data collection/aggregation to be applied to the upper layer of protocol stack, and no wake-up and medium access control is considered. Local interactions among the sensors and/or spatial correlation among the readings are utilized in [29][30]. Priority-based top-k monitoring (PRIM) [31] schedules the transmission of sensing data in the order of sensor readings and suspends transmissions after the data satisfies the top-k condition. Top-k monitoring

can be integrated with duty-cycling and other sleeping mechanisms, see [32] and the references therein. More recent studies on top-k query in WSNs consider the privacy and integrity issues [33][34], the problems related to node mobility [35][36], compression and correctness [37], and extensions to multiattribute query [38]. Although there are several studies considering energy reduction for top-k query in WSNs as described above, no work considers the wake-up control. In this thesis, we address the problem of top-k query with wake-up signaling.

1.2.4 Energy-efficient Protocol Designs Exploiting the Nature of Spatio-temproal Correlation of Observed Data

As the observed data of WSNs tend to be correlated for both time and space domains, efficient protocols can be designed by exploiting these characteristics [1]. For instance, EXTOK [25] and FILA [24] install arithmetic filter into sensor nodes, which takes advantage of the temporal correlation to reduce the number of transmissions, as described in Sec. 1.2.3. In these filtering approaches, a node transmits data only when its current observation values violate the filter set for each node. The spatial correlation of the sensor reading is used to reduce redundant transmissions in [30]. There have been many investigations for extracting spatial characteristics of sensing data in WSNs, such as (A) contour map data collection, (B) event detection, and (C) source detection. Contour map data collections aim to grasp the information on the map, where the same observation value is connected with lines [39][40]. Their examples include residual energy monitoring [41] and surveillance of coil miner [42]. Event detections aim to gather the information on nodes that detect a specific event or on the event locations, whose applications include intruder detection [43] and identification of acoustic sources exploiting the acoustic sensors [44]. Source detections attempt to identify the positions where the sensing observations are emitted by using the observation values collected by the sink. The applications of source detection include the detection of outflow of radioactive material [45] and chemical gas [46].

As mentioned above, the reduction of data traffic in WSNs by exploiting the nature of spatio-temporal correlation is considered in the literature. On the other hand, we aim to reduce the wasteful energy consumption of sensor nodes focusing on wake-up control exploiting the nature of temporal and spatial correlation, which has not been considered in the existing literature. As the former study, we introduce transmission suppression mechanism i.e., installing arithmetic filter at wake-up receivers aiming to suppress wasteful wake-up of nodes whose current observed value does not contribute to improving the accuracy of the required set. Then, as the latter study, we consider the (C) source detection applications, specifically identification of multiple emission sources, in which we investigate the

performance of wake-up control to activate only a subset of nodes to detect the locations of multiple emission sources efficiently in terms of energy.

1.2.5 Information Freshness in the Context of IIoT

Age of Information (AoI) [47] has attracted much attention among the research community as a new metric that quantitatively measures the freshness of information instead of the conventional throughput and latency since it was introduced in the vehicular network in [48].

Most of works related to AoI assume push-based communication, in which a node generating data autonomously transmits its data to a data collection node. There have also been studies combining AoI and random access networks, such as slotted Aloha-like protocol [49] and Carrier Sense Multiple Access (CSMA) [50]. There are variety of AoI related metrics, such as Peak AoI (PAoI) [51], Age of Incorrectness Information (AoII) [50], Value of Information (VoI) [52][53], Cost of Update Delay (CoUD) [54], Age of Synchronization (AoS) [55], Age of Changed Information (AoCI) [56], and so forth. PAoI tracks the peak point of AoI. AoII considers the penalty of the estimation error at the destination along with the time penalty in monitoring application. VoI is defined as a measure of uncertainty reduction from the information set of the receiver if the transmission is successful [53]. In [54], the authors consider non-linear age. AoS evaluates desynchronization of information between the receiver and source. AoCI considers not only the time lag of the update, but also changes in the content of these updates.

In the context of WSNs-related research, AoI has been introduced in the scenario of data collection using Unmanned Aerial Vehicles (UAV) [57], and sleep-wake sensor in IIoT scenario [58], in which the sleep time can be set based on the physical processes and the importance of sensors with age-penalty function introduced.

Contrast to the push-based AoI metrics, pull-based model, in which data collection node initiates data communication by sending queries to collect data from data generation nodes, considering AoI plays an important role and becomes attractive especially when the user's interest is in the freshness of information at the specific time instance [59] [2]. The AoI of pull-based transmission strategies has previously been studied in [2] and the related VoI metric was analyzed in [60]. Although pull-based communication is important with respect to information freshness, none of the work integrates AoI and wake-up radio. This work focuses on pull-based timely data collection exploiting wake-up radio technology and clarifies the importance of the timing of wake-up signaling with respect to the deadline for the data collection.

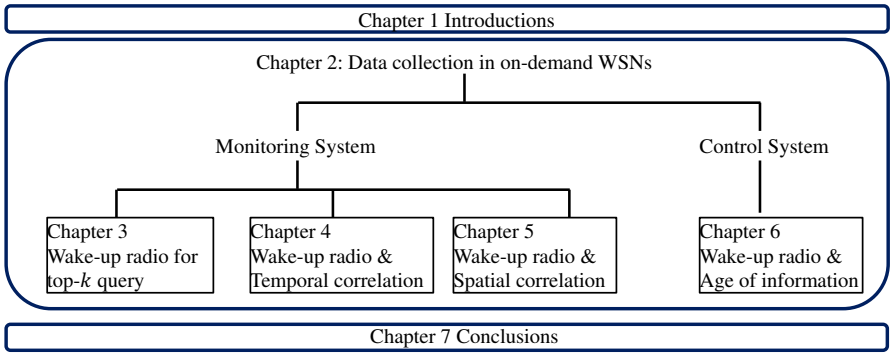


Figure 1.2: Outline of this thesis.

1.3 Research Objective

This thesis aims to investigate and characterize the performance of CoWu and clarify the efficiency of CoWu against the conventional wake-up control. This thesis will answer the following research questions in on-demand WSNs employing CoWu:

- Q1: How can the sink collect top-k data set while achieving both energy-efficiency and low latency?
- Q2: How can the sink exploit the temporal correlation of observed data for repeated top-k query to improve the energy efficiency of sensor nodes without sacrificing the accuracy of the collected set?
- Q3: How can the sink activate only the subset of nodes by exploiting spatial correlation of their data, whose observed values are important to improve the accuracy of identification of multiple emission sources?
- Q4: How can the sink collect data with high energy-efficiency and information freshness when the data are needed at a specific instance?

1.4 Thesis Outline

Fig. 1.2 shows the structure of this thesis. Chapter 2 outlines the background related to WSNs and discusses the necessity of data-oriented communication control. Then, we describe the technology of on-demand WSNs employing wake-up receivers with the operations of conventional wake-up control. Furthermore, we describe the problem definitions. In the following chapters, we apply CoWu to two types of systems: Monitoring system (Chapters 3-5) and Control system (Chapter 6).

Chapter 3 focuses on monitoring system in WSNs and proposes content-based wake-up and describes its detailed operations. Then, its enhanced version of CountDown CoWu (CDCoWu) designed for top-k query is proposed with a data collection algorithm, in which the threshold of CoWu is gradually reduced. Then, we derive the theoretical equations of CDCoWu and conventional IDWu in terms of data collection delay and total energy consumption. We apply the approximation methods called Markov Chain Monte Carlo (MCMC) for the analysis of CDCoWu to solve the complexity of calculations, and show validity of this approach by using commuter simulations. Furthermore, we investigate the interaction between CoWu and information of application layer, such as distribution of observed data, resolutions of sensor, and required size of k in top-k query, and evaluate the effectiveness of CoWu against conventional IDWu in terms of both data collection delay and total energy consumption with a variety of parameter setting.

In Chapter 4, we propose a wake-up control for repeated top-k query in monitoring system, where observed data have temporal correlation. We introduce a new wake-up control called Order and Temporal Correlation-based Wake-up (OTC-Wu), aiming to realize both high energy efficiency and high-ranking accuracy. OTC-Wu aims to only wake up nodes that cause the ranking variation of top-k set between different data collection rounds. To this end, we introduce the transmission suppression mechanism, in which each node suppresses its wake-up and transmission of data if the observation of node falls into a transmission suppression range. Furthermore, the metric called Generalized Footrule distance (GF-distance) is newly introduced into WSNs to evaluate the ranking and element accuracy between true top-k set and its estimated top-k set. We evaluate the trade-off between ranking accuracy and energy-efficiency through the range of transmission suppression and show its effectiveness against the conventional wake-up control, especially when the degree of temporal correlation is high.

Chapter 5 investigates a wake-up control suited for identifying the multiple emission sources of sensing data, e.g., toxic gas, which are spatially spread over a sensing field in terms of energy efficiency and identification accuracy. This scenario is classified as a monitoring system where a user is interested in grasping the existence of gas emissions. We investigate how to estimate the location of multiple emission sources of sensing data by using the collected subset of data and clarify that Convolutional Neural Network (CNN) provides good estimation accuracy. We evaluate the identification accuracy of different wake-up control, UCWu and CDCoWu, by computer simulations, and clarify wake–up control suited for energy–efficient identification of multiple emission sources.

In Chapter 6, we apply CoWu to a scenario where high freshness of collected data and high energy efficiency are both required. This scenario is classified into a control system, such as data collection for the actuator in the context of the

IIoT. With the clarification of our scenario and objective, we derive the theoretical equations of the accuracy at the specific time achieved by CoWu considering the evolution of the physical process. Then, with numerical evaluations, we clarify the trade-off between the reliability and freshness of collected data with respect to the timing of CoWu signaling when the data is required to collect by a deadline. Furthermore, we investigate the performance of CoWu against the speed of the physical process, characterize the gain of CoWu, and investigate the robustness against the inaccurate estimation of the physical process.

Finally, in Chapter 7, we conclude this thesis and describe the future work.

Chapter 2

Energy-efficient Data Collection in WSNs

In this chapter, we introduce data-oriented WSNs, which plays an important role in the context of IoT data collection, and wake-up radio technology realizing high energy efficiency of sensor nodes. In addition, problem definition considered in this thesis is clearly defined.

2.1 WSNs

2.1.1 Introduction to WSNs

WSNs are networks comprising of many sensors with a wireless communication function deployed over a sensing environment. The measurements observed by the sensor, such as temperature, humidity, illuminance, etc., are collected in one centralized server, e.g., a base station via wireless communication. By introducing these WSNs system, users such as companies and individuals can collect desired information (sensing data) without actually visiting the place of interests, reducing the human cost and quickly detecting problems that may occur in the sensing area. The applications of WSNs are pervasive in a wide variety of areas, including disasters, crime prevention, health management, logistics management in factories, and medicine.

In WSNs, the deployment of the sensors is generally dense so that a user can obtain spatially dense data. Therefore, observations of geographically-neighboring nodes have spatial correlation. Typical data of the sensors observed in applications, such as environmental monitoring, are related to physical phenomena such as temperature, humidity, and illuminance, therefore, there is correlation between adjacent data on the time series observed by the sensor. WSNs consist of a sink

node and sensor nodes. Each sensor node has a wireless communication function, data processing function, sensor, and battery, while the sink node has a communication function, data processing function, and power supply. The sensor nodes sense surrounding environments and transmit them to the sink, while the sink collects data from the sensor nodes and processes them for analysis. The sink, in general, is expected to supply power from a stable power source such as a power cable because it requires high computational power. On the other hand, the sensor node operates with batteries in order to reduce the complexity caused by wire connections and deployment cost, which enables flexible node installation.

One of the problems caused by battery-powered sensor nodes in WSNs is the energy efficiency of sensor nodes. The remaining battery capacity will continue to decrease unless it is recharged. When the battery runs out, the sensor node becomes inoperable and cannot fulfill its function. In this case, battery replacement is required, but the replacement cost would be high, especially when the number of nodes is large. Therefore, the sensor node should operate with as low power consumption as possible, and it is imperative to design the system from this point of view.

2.1.2 Frequency Bands Used in WSNs

Unlicensed frequency bands currently available for WSNs include 315 MHz band, 426 MHz band, 429 MHz band, 920 MHz band, 1200 MHz band, and 2.4 GHz band. Table 2.1 shows the usage conditions of these bands. The 315 MHz band is not suitable for data transmission as the communication power is 25mW, which results in a short communication range. The 426 MHz, 429 MHz, and 1200MHz bands have a narrow maximum occupied bandwidth. Thus, their power consumption increases due to longer communication time. The 2.4 GHz band is called the Industrial Science Medical (ISM) band, which is used for industry, science, and medical equipment. Currently, the 2.4GHz band is not only used for WSNs, but also in various devices, for instance, wireless LAN, Bluetooth, and microwave ovens. Therefore, radio interference caused by these devices has become a severe problem in the 2.4GHz band. In order to avoid this interference, it is desirable to consider the use of other frequency bands. The 920MHz band is an alternative and attractive frequency band to be applied for WSNs. Since the frequency of the 920 MHz band is lower than the 2.4GHz band, the communication distance is long with good diffraction characteristics. Currently, the 920 MHz band is used by many electronic tag systems. Although these devices emit strong radio waves, the environment in which they are actually used is limited to specific environments such as distribution warehouses. For this reason, compared to the 2.4 GHz, the number of interfering units and its influence is considered to be small. Thanks to these attractive characteristics, the 920MHz band is currently gaining attention

Table 2.1: Frequency band usage conditions.

Frequency band	315MHz	426MHz	429MHz	920MHz	1200MHz	2.4GHz
Bandwidth	3.25MHz	0.11MHz	0.11MHz	13.8MHz	1MHz	83.5MHz
Maximum occupied bandwidth	1000kHz	8.5kHz /16kHz	8.5kHz	1000kHz	16kHz /32kHz	26MHz
Maximum transmission power	25μW	1mW	10mW	20mW	10mW	10mW/MHz

Table 2.2: Modulation scheme of IEEE 802.15.4g.

	MR-FSK	MR-O-QPSK	MR-OFDM
modulation/ demodulation circuit	simple	complicated	Very complicated
Data transmission rate	Middle	Low	High
Occupied bandwidth	Middle	Middle	Wide

as a frequency band suited for WSNs more than the 2.4GHz band, and this thesis assumes the usage of 920MHz band for WSNs.

2.1.3 Communication Standards Used in WSNs

This thesis assumes that the IEEE 802.15.4 standard is used for the WSN communication. The IEEE 802.15.4 standard is a short-range wireless communication standard mainly aimed at realizing WSNs, which was standardized by the IEEE 802.15.4 task group [61]. This standard has three features: low communication speed, low power consumption, and low cost. The scope of standardization of the IEEE 802.15.4 is only the physical and MAC layers. This section describes the overview of the physical layer standard of IEEE 802.15.4g and the MAC layer standard of IEEE 802.15.4, which are assumed to be used in this thesis.

IEEE 802.15.4g

IEEE 802.15.4g is a physical layer standard for the 920MHz band. The utility network of a smart meter is a specific application of this standard, in which a meter of gas, electricity, and water is attached to a wireless device, and its reading is collected. In IEEE 802.15.4g, three types of modulation methods are currently available; Multi-Rate, Multi-Regional FSK (MR-FSK), Multi-Rate, Multi-Regional O-QPSK (MR-O-QPSK), and Multi-Rate, Multi-Regional OFDM (MR-OFDM). Table 2.2 shows the comparison of three schemes. Among the three modulation schemes, MR-OFDM can achieve the highest transmission rate. However, it requires a very wide occupied bandwidth and is difficult to realize under the Japanese radio regulation, which requires us to operate in a narrow bandwidth.

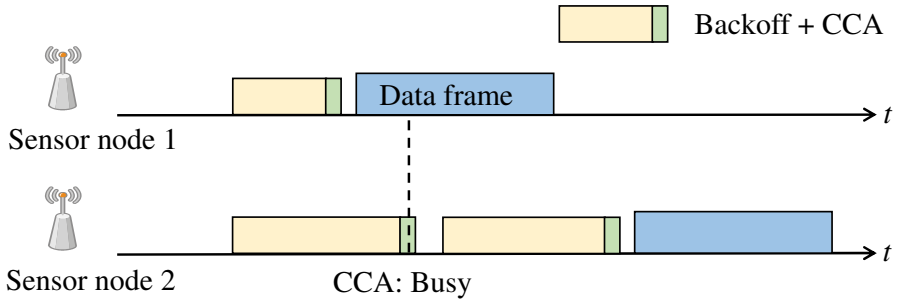


Figure 2.1: An example of operation of CSMA/CA.

On the other hand, both MR-FSK and MR-O-QPSK do not require wide-occupied bandwidth, and a high data transmission rate can be achieved. Comparing MR-FSK with MR-O-QPSK, we can see that MR-FSK can achieve higher transmission rate and simple circuit configuration. Considering this aspect, this thesis adopts MR-FSK as the modulation method.

IEEE 802.15.4

IEEE 802.15.4 is a MAC layer standard corresponding to the data link layer of the International Standardization Organization - Open Systems Interconnection (ISO-OSI) reference model. In IEEE 802.15.4, Carrier Sense Multiple Access / Collision Avoidance (CSMA/CA) is adopted as a medium access control. Fig. 2.1 shows an example of the operation of CSMA/CA. CSMA/CA is a collision avoidance protocol for multiple nodes to communicate data using the same channel. When each node attempts to transmit data, it first waits for randomly-chosen amount of time, called the backoff, followed by Clear Channel Assessment (CCA) to check the channel state. Fig. 2.1 shows an example where the sensor node 1 transmits data after CCA because the channel is idle, while sensor node 2 waits for second backoff and conducts CCA again since its first CCA result was busy. In this way, the collision is avoided by each node refraining from communicating while other nodes are communicating.

Fig. 2.2 shows a flow chart of the IEEE 802.15.4 MAC protocol, including the operation by the asynchronous Unslotted CSMA/CA, which is used in many WSNs [61]. In the figure, FR and NB are the number of retransmissions and re-backoffs, respectively, with an initial value of 0. Backoff Exponent (BE) indicates the backoff exponent whose initial value is set to the minimum value specified by macMinBE. Each sensor node first waits for backoff time, whose value can be calculated by multiplexing the unit time by the value chosen from 0 to 2^{BE}-1. After waiting for the backoff time, each node performs CCA and tries to send the data to the receiver if the channel is idle. If the transmitter receives an ACKnowledg-

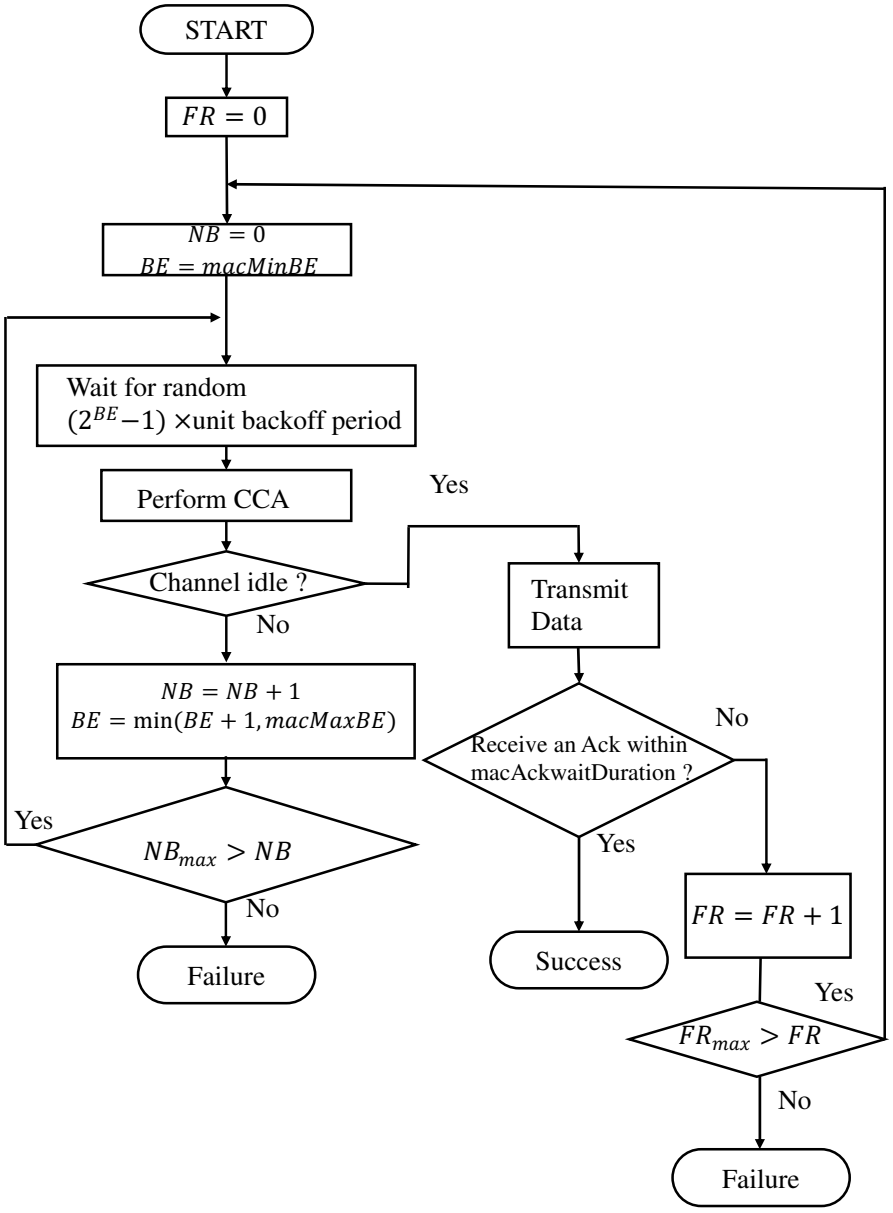


Figure 2.2: A flowchart of IEEE 802.15.4 MAC operations.

ment (ACK) response from the receiver during the maximum ACK response wait time (macAckwaitDuration) [61] after transmitting a packet, the transmission is considered to be successful, and the transmission processing is ended. If not, each node assumes that a collision occurred, FR is incremented by 1, and NB and BE are initialized under the condition that the value does not exceed the maximum number of possible retransmissions, and tries to retransmit the packet following

CSMA/CA operation. Then, if $FR > FR_{max}$ is satisfied, each node gives up its data transmission.

On the other hand, if the channel is busy when performing CCA, the value of NB is incremented by 1, and also the current BE value is incremented by 1 if the value of BE is less than the maximum number of BE specified as $macMaxBE$. Then, each node conducts back-off and CCA again until the number of back-offs reaches its upper limit NB_{max}. If NB is greater than NB_{max}, each node discards its packet, and regards it as a failure of transmission. As we mentioned, by randomly setting the waiting time before transmission, the collision caused by the simultaneous transmissions of frames from multiple sending nodes can be avoided.

2.2 Data Collection Scenario Considered in This Thesis

2.2.1 Data-oriented WSNs

With the upcoming IoT society, the amount of data exchanged via communication devices and the diversity of data itself are increasing, as represented by words like big data and data mining. In such an environment, collecting data from all sensor nodes by the sink node is not desirable from the viewpoint of energy and frequency utilization efficiency because this increases the amount of traffic. What is required for data collection in the future is to spend communication resources such as energy and frequency for collecting truly meaningful and valuable data. In other words, data-oriented communication that allocates communication resources considering the characteristics of the data observed by the sensor is attractive for future communication principles. As a data collection considering data content, we focus on top-k query, multiple emission sources detection, and timely collection of data within a specific range in the context of IIoT. Below, we explain the detail of these data collections.

2.2.2 Top-k Query in WSNs

Top-k data collection aims to grasp the information on the top-k set among the observations in a sensing field. Top-k query in WSNs has a variety of applications, including ecological observation of birds by ornithologists [21], forest fire monitoring [32], smart waste collection in smart city applications [22], and network management considering residual energy of sensor nodes [24]. There are two types of collection methods in top-k data collection in WSNs; node-set and value-set. Here, we explain the difference between the top-k node-set collection and top-k

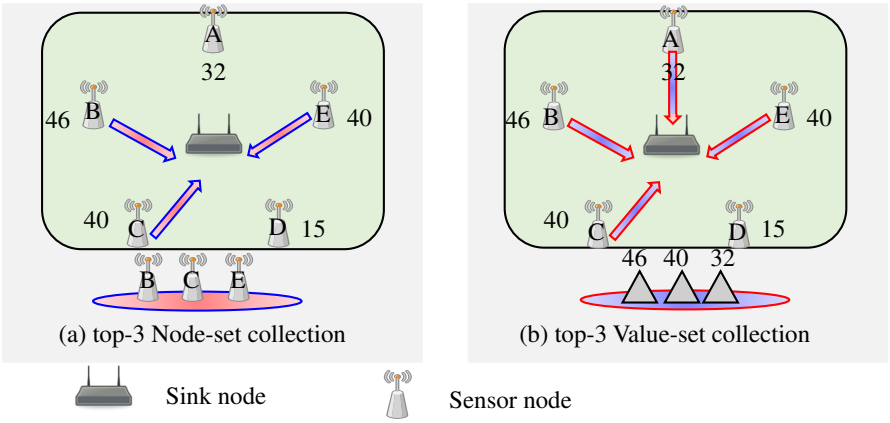


Figure 2.3: An example of operation of top-k query.

value-set collection using Fig. 2.3. Fig. 2.3 shows an example where a sink node conducts top-3 data collection from 5 sensor nodes $\{A, B, C, D, E\}$, each of which observes a value $\{32, 46, 40, 15, 40\}$. First, we explain the top-3 node-set query by using examples given in Fig. 2.3(a). Top-k node set collection is a method suitable when the sink node is interested in collecting the information on top-k node ID with higher reading. In an example of Fig. 2.3(a), as the top-3 node corresponds to sensor nodes B, C, and E, the sink can grasp top-k data if it can collect data from these three nodes. On the other hand, the top-k value-set collection is a suitable data collection method when the sink node is interested in grasping the top-k data among the data observed by the sensor node with its ID. Fig. 2.3(b) shows an example of top-3 value-set collections. In the example of Fig. 2.3(b), the data 40 observed by sensor nodes C and E is counted as one type of data because they observe the same value. Therefore, the top-3 value set consists of $\{46, 40, 32\}$ in descending order of observed values, and the corresponding nodes ID observing these values are $\{A, B, C, E\}$. The sink can grasp the top-3 value set if it collects data from these four nodes. Thus, there is a difference in the number of nodes required for grasping desired set of information between node-set collection and value-set collection.

2.2.3 Multiple Emission Sources Identification in WSNs

Identification of multiple emission sources aims to identify the positions of emission sources of sensing observations by using the data collected by the sink. In this scenario, the sensors located near the source positions observe higher values while the sensors located far from the sources observe smaller values. Fig. 2.4 shows an example of observations of sensors in the environment where multiple emission

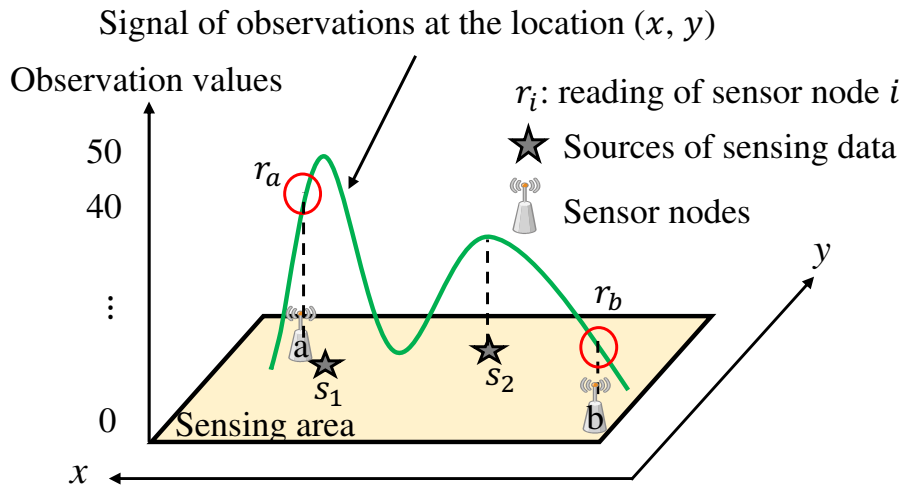


Figure 2.4: An example of observations of sensors in a scenario where multiple emission sources exist.

sources exist, in which sensor node $\{a\}$ observes a higher value as it is located near the emission source s_1, while sensor node $\{b\}$ observes smaller data as it is located far from both emission source s_1 and s_2. Thus, some nodes have valuable data for the identification of emission sources, for instance, the ones observing higher values (sensor node $\{a\}$ in Fig. 2.4), while others observe less relevant data (sensor node $\{b\}$ in Fig. 2.4). In terms of energy efficiency and data-oriented communication, the sink should collect data only from nodes with important data to improve accuracy. To this end, we need to clarify the important data for multiple emission sources identification and wake-up control suited for it in terms of identification accuracy and energy efficiency.

2.2.4 Data Collection Considering Timeliness in Wireless Sensor and Actuator Networks (WSAN)

Fig. 2.5 shows an example of the operation of WSAN, which uses a closed control loop. The actuator needs to accomplish some tasks, such as controlling the temperature of a room, fire extinguishing, and maintenance, and consequently modifies the environment whose features are measured by the sensor nodes. In order to take the appropriate action, the actuator needs to receive a command from a server at the network edge, also called sink node, at a specific time. In this scenario, the freshness of the information available to the sink node at the instant when it makes the decision is critical for the control task's performance, as a stale picture of the environment can result in suboptimal actions. Information freshness can be evaluated by using a metric called AoI [48][47], which measures the time

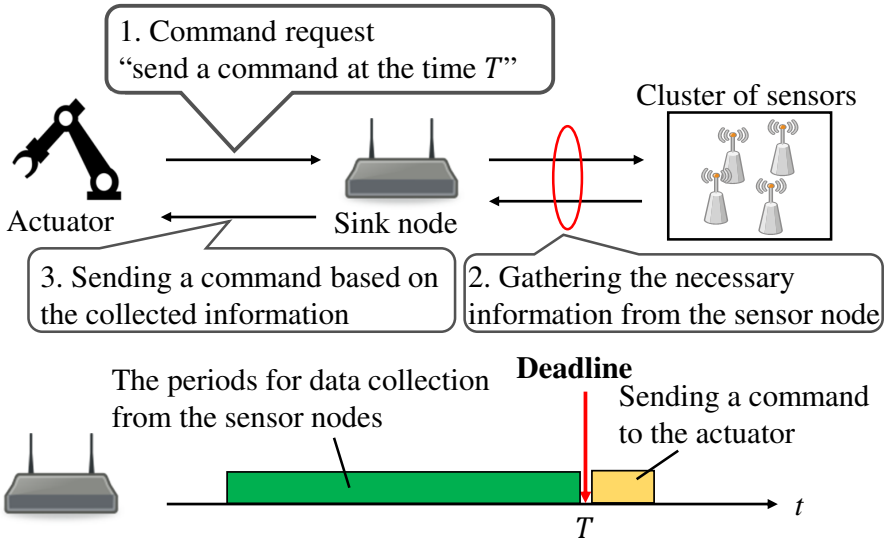


Figure 2.5: The basic function of a WSAN.

elapsed since the generation of the last measurement received by the sink node. Since the physical process that the WSAN measures and controls can change over time, both based on its own characteristics and due to the actions of the network's actuators, the uncertainty on its state increases with the AoI. In a WSAN, control performance can degrade significantly if the uncertainty on the state of the environment is high, as the controller can be forced to take conservative actions. In this scenario, the freshness of the information needs to be measured at the specific time when the decision is made, maximizing the accuracy at that precise moment [2]. On the other hand, a key requirement on both WSNs and WSANs is the energy efficiency of sensor nodes, which often have very limited batteries. Therefore, we need to design a wake-up control for WSAN so that both high information freshness at the query instance and high energy efficiency can be achieved.

2.3 On-demand WSNs

The main focus of this thesis is to develop wake-up radio technology for WSNs to save wasteful energy consumption of sensor nodes. This is realized by employing a wake-up receiver at each sensor node. Here, a wake-up receiver aims to reduce the wasteful energy consumption of the main radio interface of each node. This section describes a WSN system employing wake-up receivers, called on-demand WSNs.

2.3.1 Energy Saving with On-demand WSNs

There are various factors in the power consumption of sensor nodes, but among them, the consumption of the main radio I/F used for data communication is dominant. In general, the main radio I/F is operated not only during data communication but also during a non-communication period. The basic solution to this problem is introduction of duty-cycling [8], in which the main radio is periodically switched on/off. However, it is difficult for duty-cycling to avoid the increased latency and wasteful energy consumption during the idle listening. In order to solve this problem, we introduce the concept of *wake-up radio* into WSNs, which we call on-demand WSNs. In on-demand WSNs, each sensor node installs a wake-up receiver, which is operated with much smaller energy consumption than the main radio I/F. Fig. 2.6 shows an operation of on-demand WSNs. In on-demand WSNs, each sensor node consists of a wake-up receiver, Micro Controller Unit (MCU), and main radio I/F. When there is no need to communicate, each node waits for the transmission request with a sleep state, in which only the ultra low-power wake-up receiver stays active while its main radio I/F is switched off. As the power consumption of the wake-up receiver is much smaller than that of the main radio I/F, activating a wake-up receiver instead of the main radio I/F during the idle period can save the overall energy consumption of sensor nodes. In on-demand WSNs, the sink controls the node's activation so that each node transits to the communication-enabled mode only when the communicate is needed, which saves an overall power consumption of sensor nodes and realizes a long network lifetime. Here, in this thesis, we call the action that transits the nodes in a sleep state to the state where the communication is possible as *wake-up*.

When the sink wants to communicate to the target sensor node in a sleep state, it first transmits a wake-up signal toward it. Here, the wake-up signal aims to wake up nodes in a sleep state. The wake-up receiver detecting the wake-up signal to wake up itself outputs the wake-up command to the host CPU, which activates its main radio I/F. Then, the sensor node transmits data through the main radio I/F to the sink. If the sensor node receives the ACK from the sink after data transmission, it transits to the sleep state, as shown in Fig. 2.6.

2.3.2 Wake-up Receiver

There are two types of wake-up control proposed so far, which are classified as Range-based and ID-based [10]. In the former Range-based wake-up method, the wake-up receiver observes the energy level over the operating frequency and commands to wake up when it detects a signal level (energy level) higher than a predetermined threshold. In the latter ID-based wake-up, the information on the target node ID is embedded into the wake-up signal. In this thesis, we focus on

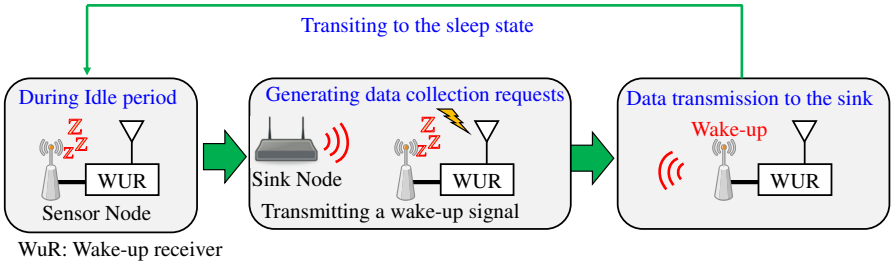


Figure 2.6: An example of operations of on-demand WSNs.

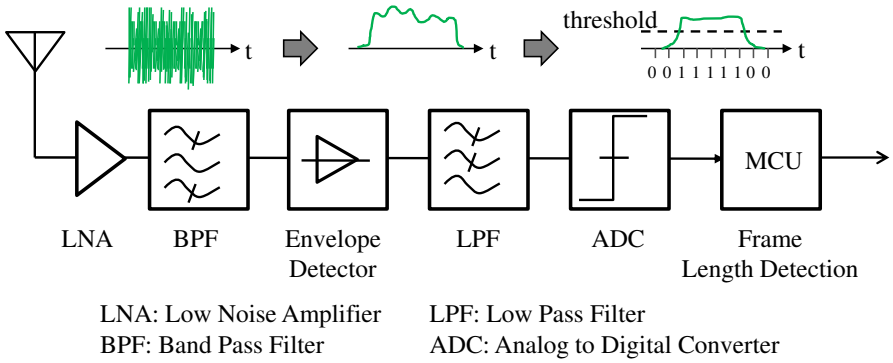


Figure 2.7: An example of configuration of wake-up receiver to detect frame length.

the method proposed in [3], in which the ID is mapped into the frame length of IEEE 802.15.4g standard. In this method, each wake-up receiver has its own ID, and the sink transmits a frame length corresponding to its ID as a wake-up signal. The wake-up receiver detecting this signal compares its received frame length to its wake-up frame length based on its allocated ID and decides whether it wakes up or not. The configuration of this type of wake-up receiver is shown in Fig. 2.7 [62]. As mentioned above, this wake-up receiver does not use a complicated decoding technique; rather, it only detects the frame length over the channel. As shown in the figure, the received signal is first amplified with low noise amplifier (LNA). Then, the signal over a target frequency band (e.g., 920MHz for IEEE 802.15.4g) is extracted with a Band Pass Filter (BPF). The output of envelope detector is smoothed with Low–Pass–Filter (LPF), and On–Off–Keying (OOK) detection is applied to detect the signal with its level larger than a threshold. By counting the continuous number of "1" s, MCU calculates the length of received frame. Fig. 2.8 shows an example of frame detection using our wake-up receiver. Let us denote the consecutive numbers of "1" as n and the bit detection interval as d as shown in Fig. 2.8. Then, the received frame length can be calculated as $(n-1) \times d$. Unlike the main radio following the existing standard such as IEEE 802.15.4, the wake–up receiver includes neither mixer/oscillator nor complicated signal processing. This

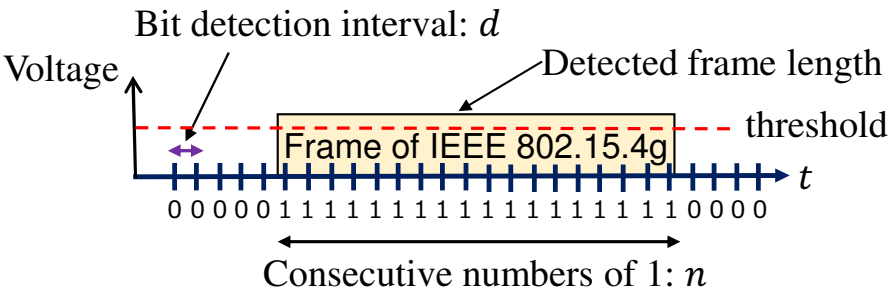


Figure 2.8: An example of detection of frame length.

simple configuration enables the wake–up receiver to achieve at most a few micro watts of power, which is several orders of magnitude lower than that of the main radio [11].

The advantage of the above-mentioned wake-up control exploiting the length of frame is as follows:

- The demodulation method of the wake-up receiver based on envelop detector and OOK demodulation can be realized with a small and simple circuit. Thus, we can introduce a wake-up receiver with low cost.
- The wake-up receiver can operate with ultra-low power consumption thanks to the simple circuit configuration based on envelop detector and OOK demodulation.
- Our wake-up receiver is designed so that it can decide whether it wakes up or not based on the detected frame length. This can be realized by the transmitter controlling the frame length via conventional wireless radio I/F. Therefore, the transmission of the wake-up signal is realized by only changing software without introducing a new hardware, such as a transmitter dedicated to the transmission of the wake-up signal.

Considering the advantages mentioned above, in this thesis, we apply the frame length method as a wake-up control in on-demand WSNs.

2.3.3 Conventional Wake-up Control

In order to realize the conventional IDWu with control employing frame length, we first need to prepare a mapping between the different IDs assigned to each node and the frame length. In this section, as IDWu, we describe a UniCast Wake-up (UCWu) and BroadCast Wake–up (BCWu).

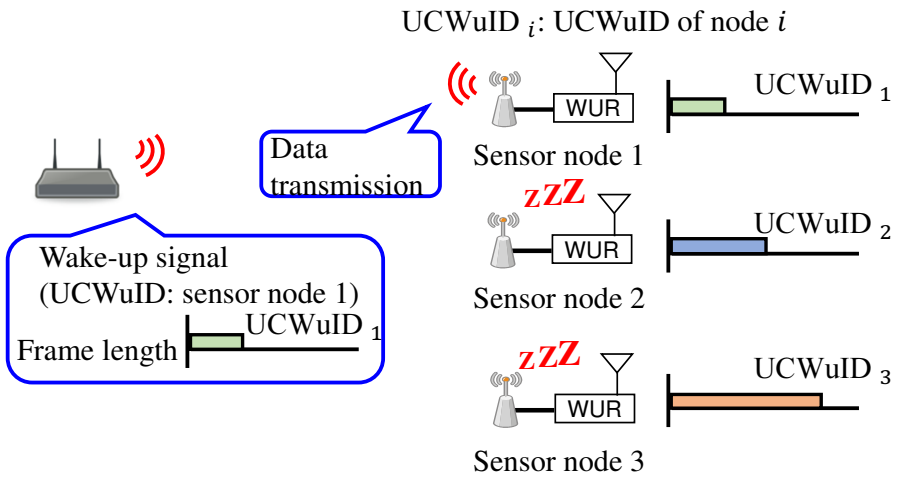


Figure 2.9: An example of data collection employing UCWu.

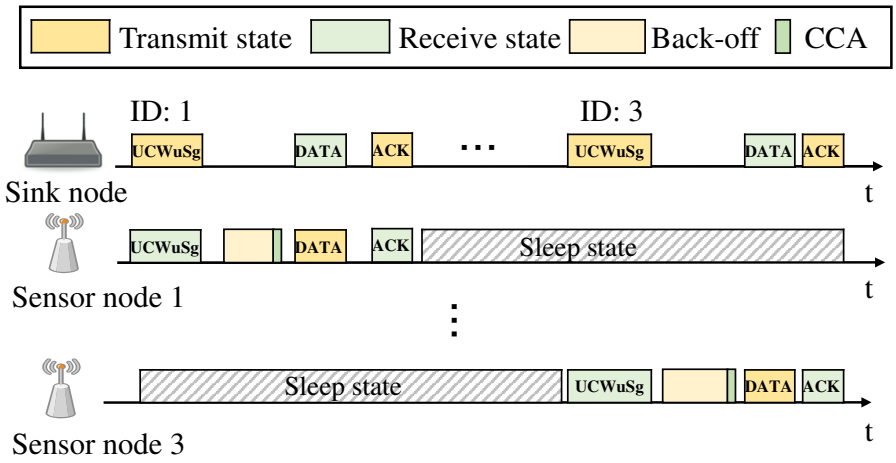


Figure 2.10: An example of communication in data collection employing UCWu.

Unicast wake-up (UCWu)

Figs. 2.9 and 2.10 respectively show an example of data collection and communication using UCWu. In UCWu, each node has a unique wake-up ID, called UniCast Wake–up ID (UCWuID), and the sink transmits a frame length corresponding to the target node ID as a wake-up signal. The wake-up receiver detecting the wake-up signal compares its received frame length and wake-up frame length corresponding to its UCWuID and activates its main radio I/F only if the length of these frames coincide. The wake-up node transmits data following the CSMA/CA operation. In UCWu, only the node whose UCWuID coincides with the ID embedded into the wake-up signal wakes up, while the nodes whose node IDs are not

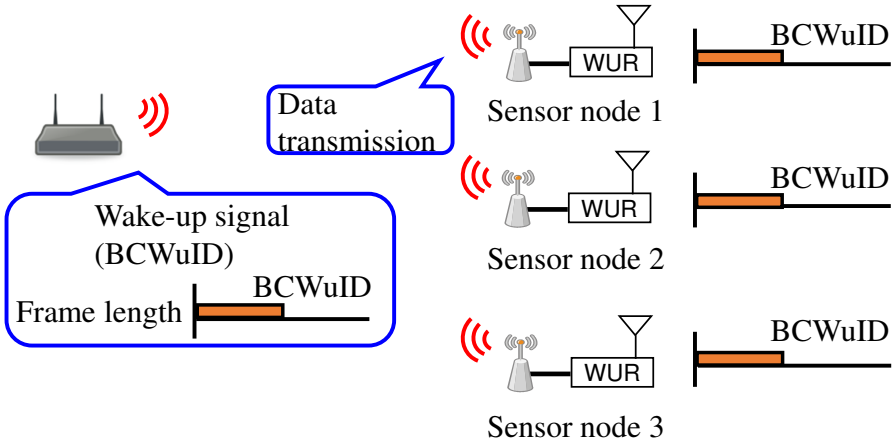


Figure 2.11: An example of data collection employing BCWu.

specified by the wake-up signal remain in a sleep state and do not transmit data to the sink. In the example of Fig. 2.9, as the sink transmits a wake-up signal whose length corresponds to the node ID 1, sensor nodes 2 and 3 stay in a sleep state. After receiving data from sensor node 1, the sink transmits ACK to sensor node 1, as shown in Fig. 2.10. After receiving ACK, sensor node 1 transits to the sleep state. The sink continues to collect data by transmitting a wake-up signal corresponding to node ID 2 and 3 subsequently. In UCWu, if the sink wants to collect data from N sensor nodes, it is necessary to specify a unique ID assigned to each node and wake up one by one by sending wake-up signal N times. At the cost of this wake-up overhead, UCWu can collect data from all nodes without congestion.

Broadcast Wake-up (BCWu)

Figs. 2.11 and 2.12 respectively show an example of data collection and communication using BCWu. In BCWu, all nodes have a common broadcast wake-up ID, called BroadCast Wake–up ID (BCWuID), and the sink transmits a frame length corresponding to BCWuID as a wake-up signal. The wake-up receivers detecting the frame length corresponding to the BCWuID activate their main radio I/F and transmit data following the CSMA/CA operation, as shown in Fig. 2.12. Unlike UCWu, BCWu uses a common BCWuID to specify all nodes within its communication range, so the sink does not need to know the unique ID information of each node within the communication range in advance. BCWu also has the advantage that wake-up overhead is small, as it can activate all nodes only by a single wake-up trial.

On the other hand, a drawback of the BCWu is the congestion caused by

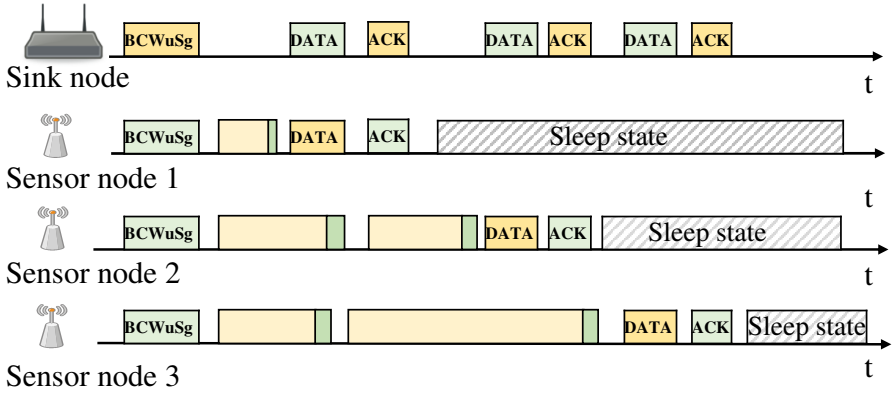


Figure 2.12: An example of communication in data collection employing BCWu.

simultaneous wake-up and data transmission from multiple nodes. Fig. 2.11 shows an example where three nodes (nodes 1, 2, and 3) are simultaneously woken up by BCWu Signal (BCWuSg). After backoff operations, each node attempts to send a packet to the sink, during which the channel needs to be contended and shared by multiple nodes. If multiple nodes use the channel simultaneously, packet collisions may occur, which increases delay, which is the time to complete data transmission. Each node first waits for the backoff time and then performs CCA following the CSMA/CA operations. If the channel is busy, it tries backoff again if the condition $NB < NB_{max}$ is satisfied, according to the flowchart shown in Fig. 2.2. In the example of Fig. 2.12, nodes 2 and 3 conduct CCA and transmit packet again after the end of the first backoff. In BCWu, congestion becomes a problem, which becomes more severe as the number of nodes increases. As a result, the number of nodes whose main radio I/F is active for a long period of time increases, by which the delay and wasteful power consumption of sensor nodes increases.

2.4 Problem Definition

2.4.1 The Problem Applying Conventional IDWu to a Single Round of Top-k Query

We consider the case where the sink conducts a single cycle of top-k query in on-demand WSNs, employing conventional IDWu. In IDWu, as shown in Fig. 2.13, first, the sink needs to collect data from all sensor nodes by wake-up signaling since it is not aware of observations of all nodes beforehand (step 1 in Fig. 2.13). Then, the sink grasps the desired set (top-k node set or top-k value set) based on collected data (step 2 in Fig. 2.13). In IDWu, as shown in step 2 of Fig. 2.13, the nodes whose observed values do not belong to the current top-k set need to wake

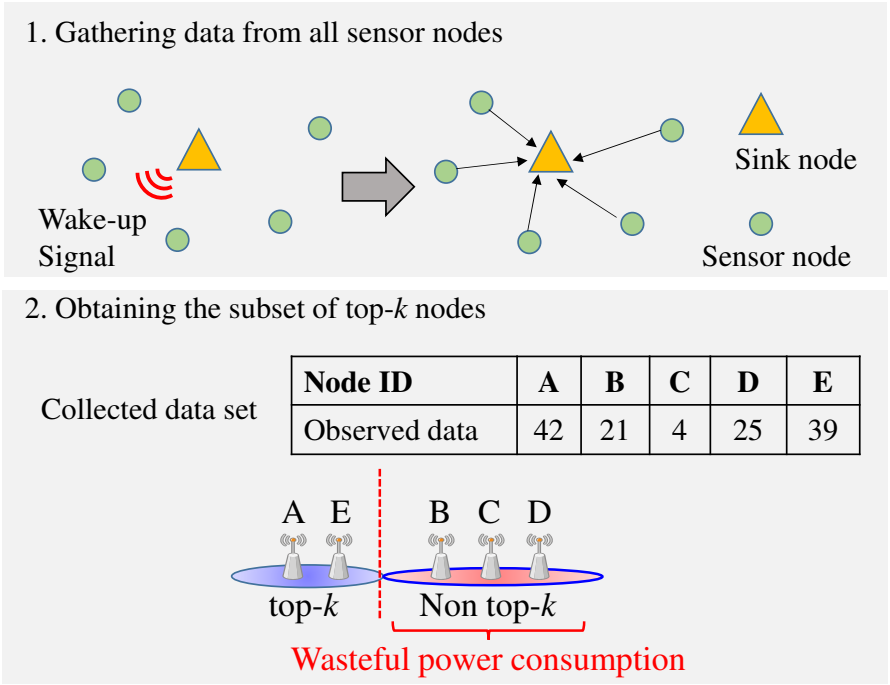


Figure 2.13: An example of the problem applying IDWu for top-k query.

up and transmit data to the sink at least once, which is clearly wasteful energy consumption. Furthermore, this is undesirable from the viewpoint of spectrum efficiency since the limited frequency band is wasted on the transmission of data that has no value. Therefore, it is desired to design a protocol that suppresses the wake-up of nodes that do not belong to top-k set and transmits only the relevant data to the sink.

2.4.2 The Problem Conducting Periodical Top-k Query in On-demand WSNs with Temporally Correlated Data

When the observed data have temporal correlation, the variation of observed data between consecutive rounds of query is relatively low, i.e., the nodes that belong to the previous top-k set are highly likely to belong to the top-k set at the current round. Since the sink only aims to grasp the top-k set, it is sufficient for the sink to grasp the information on nodes whose observed data contribute to the changes of top-k set from the previous round. In such cases, many sensor nodes do not need to consume much energy to transmit their observations to the sink. The sink should perform wake-up control by utilizing the knowledge of the past data to satisfy the desired accuracy of the top-k set and maximize the energy utilization

efficiency of the sensor node.

2.4.3 The Problem for the Multiple Emission Sources Identification in On-demand WSNs

In order to realize the identification of multiple emission sources of sensing data based on the data collected by the sink with high energy-efficiency while realizing high identification accuracy, we need to consider two problems. First, we need to clarify the methods to identify multiple sources based on the data collected by the sink. If the sink grasps the number of emission sources in a sensing area beforehand, it can infer the unknown parameters, such as the locations of sources, data values of their locations, and variance, e.g., by using maximum likelihood. However, if the information on the number of emission sources is unavailable for the sink, the approach like a maximum likelihood is not applicable. This motivates us to develop the identification methods of multiple sources. Second, we need to investigate the wake-up control realizing the selection of sensor nodes to be activated by the sink. In order to realize an energy-efficient identification of sources, it is desirable for the sink to collect data from the subset of nodes whose data contribute to the increased accuracy of identifications. However, it has not been clarified which characteristics of sensing data are important to increase the accuracy and which conventional wake-up control is suitable for the identification of multiple emission sources with respect to the identification accuracy and energy efficiency. To this end, this thesis investigates the performance of different approaches of identification methods with several conventional wake-up control with respect to energy consumption and the identification accuracy by numerical evaluations.

2.4.4 The Problem of Data Collection Considering the Information Freshness and Energy Efficiency

As described in Sec. 2.2.4, in WSAN scenarios, the sink needs to collect informative data by a deadline. When there is a deadline for data collection, the timing to transmit a wake-up signal plays an important role: early transmission leads to high reliability in data collection, but the received data may become obsolete by the deadline, while later transmission ensures a higher timeliness of the sensed data, but some nodes might not succeed in data transmission by the deadline. Therefore, we need to carefully design a timing to transmit a wake-up signal considering the deadline.

2.5 Summary

In this chapter, we have described the characteristics of WSNs with their MAC protocol and introduced data-oriented communication and its practical applications. Then, we have outlined the energy-saving technology called on-demand WSNs, in which a wake-up receiver is installed into each sensor node. We have also explained the operations of conventional UCWu and BCWu. Furthermore, we have clarified the problems of four types of data collection scenarios. In the next chapter, we will introduce a new wake-up control for a single cycle of top-k query and clarify its effectiveness and characteristics with theoretical analysis.

Chapter 3

Content-based Wake-up for Top-k Query in WSNs

This chapter introduces Content-based wake-up and its enhanced version of Count-Down Content-based wake-up (CDCoWu) for top-k query and investigates their performance with theoretical analysis and computer simulations[4][63][1].

3.1 System Model

This chapter considers a scenario where a number of sensor nodes are deployed over a sensing field, and a sink attempts to collect their observations. A star network topology is assumed, in which each sensor node directly communicates with the sink. Each sensor stores the last periodic measurement for potential reporting. A sensor has a main radio interface used to transmit the observation to the sink, and a wake-up receiver. In the absence of communication requests from the sink, main radio of each sensor is switched off, keeping only the wake-up receiver active. The sink collects information by first sending a request through a dedicated wake-up signaling, upon which the target sensor activates its main radio interface and sends a packet with its observation. The main radio is assumed to operate over unlicensed frequency bands, e.g., following IEEE 802.15.4 standard, and to use a p-persistent CSMA protocol for transmitting each packet. Here, we assume p-persistent CSMA as a MAC protocol for simplified analysis. Note that p-persistent CSMA protocol can prevent interference with the other signals (e.g., interference with the other systems sharing the same unlicensed frequency band) by conducting carrier sense and avoid collisions by transmitting data with probability p when the channel is sensed to be free, which can imitate the back-off (collision avoidance: CA) mechanism in IEEE 802.15.4. Thus, p-persistent CSMA

[1]The content of this chapter was published in [4][63].

has both CSMA and CA mechanisms, which is suited for the approximation of the practical MAC protocol following IEEE 802.15.4 standard [64]. The MAC channel is slotted, with slot length δ $[s]$, and each node with a packet to transmit conducts carrier sensing at the beginning of a slot. If the channel is sensed to be free, the node transmits the packet with probability p, otherwise it attempts again in the next slot. For simplicity, it is assumed that packets can be lost due to collisions or due to channel impairments with independent and identical probability e_c for each packet[2]. When a packet is lost, each node detects the absence of ACK and retransmits the packet also with p-persistent CSMA. If ACK is received, a sensor switches its main radio off and enters a sleep state. The ACK transmission from the sink is assumed to be collision-free and error-free. All nodes, including the sink, are assumed to be located within communication/wake-up/carrier-sensing range of each other and there are no hidden terminals.

The sink collects data through a top-k query [25][32]. The top-k query has been considered to be employed for a large number of IoT applications, e.g., environment/infrastructure monitoring [25][33], smart-city [22], and network management [24]. The fundamental and common role of the top-k query in these applications is to identify the nodes with the highest readings of interest, e.g., various types of sensing data or energy level of each sensor node. Since the synthetic data following the ideal probability distributions, such as uniform, exponential, and normal distributions, are commonly employed for the analysis of query processing [25][65][66], we assume that the observed values of sensor nodes follow uniform, exponential, or normal distribution between the minimum and maximum values of V_{min} and V_{max}, respectively, and considers both top-k node-set and value-set query. The exponential distribution follows Probability Density Function (PDF) $p(x) = \frac{e^{\alpha x}}{\int_{V_{min}}^{V_{max}} e^{\alpha x} dx}$. With $\alpha = 0$, $p(x)$ is reduced to the uniform distribution while the observed values tend to have higher values with larger α. On the other hand, PDF of a truncated Gaussian distribution with its mean of μ and variance of σ^2 is employed for normal distribution. The observed data at each sensor node is quantized with the uniform interval q_{step}, determined by the resolution of Analog-to-Digital Converter (ADC) of the sensor. Assuming that the quantization bit rate of each sensor is b_q [bits], q_{step} is calculated as $q_{step} = \frac{V_{max} - V_{min}}{2^{b_q}}$. Note that, with larger (or smaller) q_{step}, multiple sensor nodes report the same quantized value to the sink with high (or low) probability.

[2]Although we employ a fixed e_c for all nodes in the analysis and evaluations for simplicity, it has been confirmed by simulations that the results with a fixed e_c coincide with those obtained when the error probability for each node is randomly varied with its mean set to be e_c following, e.g., uniform or truncated Gaussian distribution.

3.2 Proposed Content-based Wake-up (CoWu)

In IDWu, the sink first needs to wake up all nodes by employing either BCWu or UCWu, and run data collection process through their main radio interfaces. This inevitably leads to energy waste, as the nodes storing data that are out of top-k range also need to wake up at least once. This problem is addressed by running a top-k query with CoWu, described next.

3.2.1 Content-based Wake-up

In CoWu, each sensor node determines its wake-up frame length based on its sensing data. Specifically, the range of observable value [V_{min},V_{max}] is mapped to the frame length employed for wake-up control [T_{min},T_{max}], see Fig. 3.1. Note that T_{min} is selected such that it is longer than the frame length commonly employed for data transmissions. This is to prevent false wake-up caused by background data traffic, as the wake-up receiver regards only the frames larger than T_{min} as wake-up frames. The value division step is defined as V_{step}, which is a unit-size to divide the range of observable value. Furthermore, the step to change the frame length is defined as T_{step}. Note that the minimum value of T_{step} is given by the resolution to change the duration of frame transmitted by the main radio interface. Then, when the observed value V_o belongs to j-th interval, i.e., $V_{max} - (j+1)\times V_{step} < V_o \leq V_{max} - j\times V_{step}$, its wake-up frame length T_{wu} is set to $T_{wu} = T_{min} + j\times T_{step}$. Furthermore, the wake-up receiver of CoWu is designed such that only nodes satisfying $T_{wu} \leq T_{rx}$ wake up, where T_{rx} is the frame length to be detected at the wake-up receiver. The sink transmits a wake-up signal with the frame length of T_{th}, which corresponds to the value threshold of V_{th}. Thus, CoWu activates only nodes with data satisfying the condition specified by the sink.

Fig. 3.2 shows an example of data collection with CoWu where the sink collects data from nodes storing the observed data larger than V_{th}. Here, only sensor 1 with its observed value $V_1 \geq V_{th}$ has $T_{wu} \leq T_{rx}$, and thus wakes up, while sensor node 2 stays in a sleep mode.

3.2.2 Count-Down Content-based Wake-up (CDCoWu)

The scheme CDCoWu is designed such that CoWu can be applied to both top-k node-set and value-set query. The basic idea of CDCoWu is to gradually decrease the threshold of CoWu described in Sec. 3.2.1 until the sink manages to collect the desired set of sensing data. The operation of the sink in the proposed CD-CoWu depends on the considered query, which are called as node-set CDCoWu (N-CDCoWu) and value-set CDCoWu (V-CDCoWu).

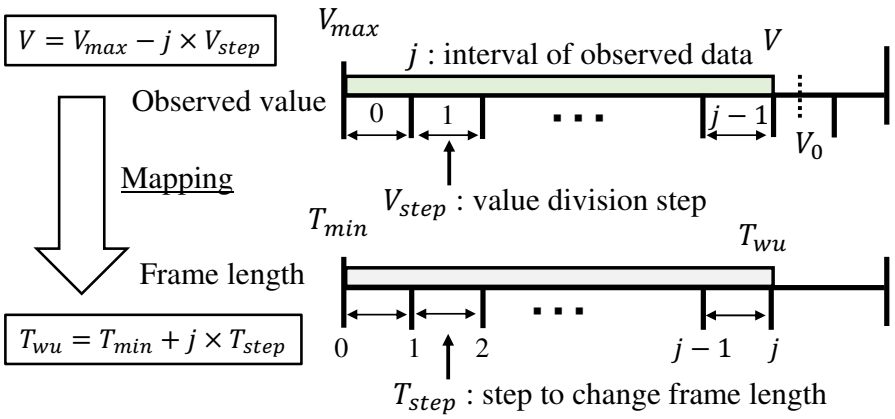


Figure 3.1: Mechanism to decide wake-up frame length in CoWu.

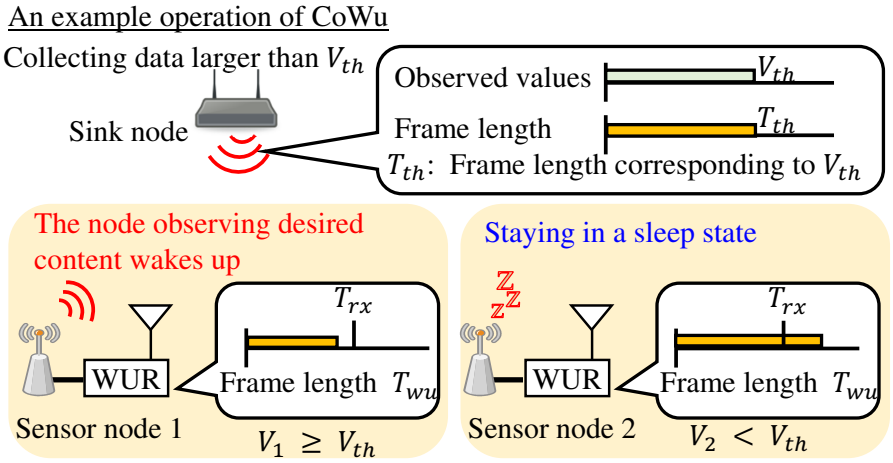


Figure 3.2: An example of CoWu operations.

In the main flow of CDCoWu, the sink reduces the threshold of CoWu from an initial value by a parameter called countdown step (CD_{step}), by which it enlarges the range of value for data collections step-by-step. Here, the initial threshold of V_{th}^1 is set to $V_{max} - CD_{step}$. The sink first generates a Wake-up Signal (WuS) whose length corresponds to V_{th}^1 and broadcasts it to sensor nodes. The nodes which have T_{wu} shorter than the length corresponding to V_{th}^1 (i.e., nodes storing value higher than V_{th}^1) wake up after detecting the wake-up signal, and attempt to send a data packet to the sink with p-persistent CSMA described in Sec. 3.1. As the sink does not know how many nodes will wake up with each wake-up trial, it needs to set a time-out and wait until there are no more responses. Following a successful data transmission, a node transits to a sleep state and is configured not to wake up for a certain period of time, preventing it from being woken up again

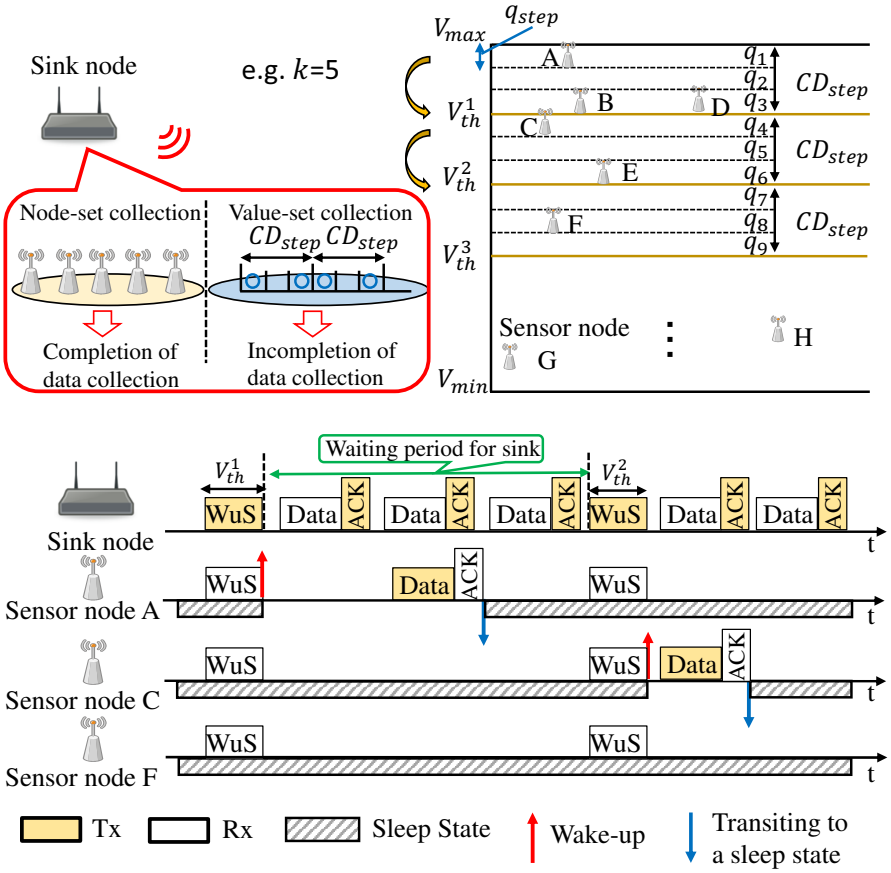


Figure 3.3: An example of CDCoWu operations.

by the immediate wake-up signals from the same round. After the time-out, the sink calculates the size n_k of the collected set and, if $n_k < k$, the sink lowers the threshold to V_{th}^2 ($V_{th}^2 = V_{th}^1 - CD_{step}$) and transmits a new wake-up signal whose length corresponds to V_{th}^2. This operation continues until $n_k \geq k$.

The size of collected set is increased as follows. In N-CDCoWu, the sink increases the size of node-set by 1 whenever it successfully receives the data from a node. In V-CDCoWu, the sink stores the received values in the set of $\mathbb{V}$ and increases the size of collected set by 1 only if the received value is not already included in $\mathbb{V}$. Here, $\mathbb{V}$ represents the collected set in which the sink stores information on the collected node IDs and their observations, and at the beginning of top-k data collection, $\mathbb{V} = \emptyset$. Through the above operations, the sink collects the information on top-k nodes/values without waking up nodes that do not belong to the range of values specified by the sink.

Fig. 3.3 shows an example operation of CDCoWu with $k = 5$, where the sink completes the first and second wake-up trials. At the bottom of Fig. 3.3, the

temporal behavior of sensor nodes together with that of the sink is also shown, which clarifies the timing of wake-up and transitions between active and sleep states. In the first wake-up trial with V_{th}^1, sensor nodes A, B, and D wake up and transmit data to the sink while the other nodes, such as nodes C and F, keep their sleep states. Each activated node transits to the sleep state after receiving ACK. After the second wake-up trial with V_{th}^2, the sink collects 5 data readings from nodes A to E, completing top-5 node-set query. However, top-5 value-set query is not completed, as only 4 different values (q_1, q_3, q_4, q_6) are gathered. Hence, the sink needs to further decrease the threshold by CD_{step} and continue the wake-up trials. Thus, it is likely that more nodes need to be woken up in V-CDCoWu than in N-CDCoWu, especially when many nodes observe the same values. The probability of this event depends on the distribution of observed data and quantization step q_{step}.

The complexity of V-CDCoWu is higher than N-CDCoWu due to a larger number of nodes to be woken up. In V-CDCoWu, the sink executes the input of data into the set of $\mathbb{V}$ after comparing the collected data with those already included into $\mathbb{V}$. The number of comparisons to be made for each collected data depends on the size of $\mathbb{V}$. Since the maximum size of $\mathbb{V}$ is k, let us assume a worst case in terms of complexity where the sink needs to always make comparisons with k values for each collected data. Assuming that top-k query is completed after collecting data from n nodes, the sink needs to execute nk operations of comparison and k operations of inputs. However, the time complexity of these operations is considered to be low considering that the sink is in general assumed to have high computational capability with a stable power-source. On the other hand, the communication delay, i.e., the time required for the sink to collect data from n nodes, is dominant from a viewpoint of latency to complete top-k query with n nodes. Therefore, this work only considers communication delay to evaluate the latency to complete top-k query.

The parameter CD_{step} affects the energy efficiency and data collection delay of CDCoWu. With larger CD_{step}, more nodes are simultaneously woken up with a single wake-up signal, which increases the congestion level and the number of unnecessarily activated nodes, thereby increasing the total energy consumption. This can be avoided by employing smaller CD_{step}, which, however, may increase data collection delay due to more wake-up trials with no replying node. Therefore, CD_{step} should be optimized based on the target delay or energy efficiency, which will be further discussed in Sec. 3.4.

3.3 Theoretical Analysis of IDWu and CDCoWu

This section analyzes the data collection delay and energy efficiency performance of IDWu and CDCoWu.

3.3.1 One-Shot Data Collection with p-persistent CSMA

In BCWu and the proposed CDCoWu, it can happen that multiple nodes attempt to transmit their packets to acknowledge the wake-up request from the sink, where each contending node holds only a single packet to transmit. This traffic model is called One-Shot Data (OSD) model [64]. Under OSD model, each node enters a sleep state after the successful transmission of its packet and does not contend for the channel. In [64], data collection delay and energy consumption of nodes operating with p-persistent CSMA are analyzed when OSD model is employed for the collision channel. These equations are extended to take account of the packet errors due to channel impairments with independent and identical probability of e_c for each packet. The data collection delay $T_d(n_o)$ [s], defined as duration for n_o nodes ($n_o \geq 1$) with OSD model to complete their transmissions, is

$$T_d(n_o) = \sum_{n=1}^{n_o} \frac{L - (L-1)(1-p)^n}{(1-e_c)np(1-p)^{n-1}}\delta, \tag{3.1}$$

where L [slot] is the packet length in slots. As a special case, this work assumes $T_d(0) = 0$. Further, the total energy consumption of the sensor nodes, $E_{total}(n_o)$ [J], can be computed as

$$\begin{aligned} E_{total}(n_o) = &\sum_{n=1}^{n_o} \xi_R \delta \frac{L - (L-1)(1-p)^{n-1}}{(1-e_c)p(1-p)^{n-2}} \\ &+ \xi_T \delta \frac{L}{(1-e_c)(1-p)^{n-1}}, \end{aligned} \tag{3.2}$$

where ξ_R [W]/ξ_T [W] is the power consumption of a node in the receive/transmit state. This work also assumes $E_{total}(0) = 0$. By using the above results, equations expressing data collection delay and energy consumption are derived for different wake-up control schemes applied to top-k query.

3.3.2 Analysis of Conventional and Proposed Wake-up Control

This subsection first analyzes delay and total energy consumption of conventional BCWu and UCWu, followed by the analysis of the proposed CDCoWu. As already

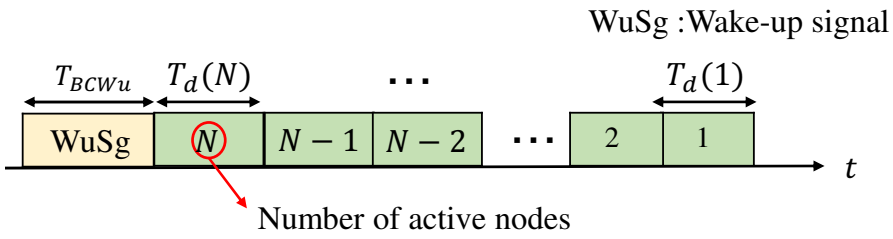


Figure 3.4: Theoretical analysis model of BCWu.

noted, the BCWu and UCWu are required to activate all sensor nodes to collect top-k observations (i.e., top-k node-set or top-k value-set), as they do not conduct wake-up signaling according to requested/sensed data. Here, delay is defined as time required to collect top-k observations from sensor nodes. Regarding energy consumption, this chapter focuses on the total energy spent by all sensor nodes during a single cycle of top-k query. Considering that the wake-up receiver is always active during a cycle[3], the energy consumed by the wake-up receiver is assumed to be same for all wake-up schemes. Therefore, for simplicity, the energy consumed by the wake-up receiver is neglected, and only energy consumed by the main radio is calculated. Furthermore, in the following analysis, all wake-up signals are assumed to be transmitted by the sink with $p = 1$ in the operation of p-persistent CSMA, since no node is supposed to contend with the sink. That is, the downlink transmissions from the sink to the sensor nodes are assumed to be error-free.

Analysis of BCWu scheme

Fig. 3.4 shows the model of theoretical analysis for BCWu. In BCWu, the sink first sends a wake-up signal whose length corresponds to BCWuID in order to wake up all sensor nodes. Assuming that the total number of sensor nodes is N and that all nodes detect the wake-up signal correctly, N nodes wake up and attempt to transmit their packets with p-persistent CSMA and OSD model. Thus, delay of BCWu, $T_d^{BCWu}(N)$, is expressed as

$$T_d^{BCWu}(N) = T_d(N) + T_{BCWu}, \tag{3.3}$$

where T_{BCWu} [s] is the frame length corresponding to BCWuID. The energy consumed by the main radios of sensor nodes in BCWu is

$$E_{total}^{BCWu}(N) = E_{total}(N), \tag{3.4}$$

[3]The energy management, e.g., adaptive switch on/off of the wake-up receiver, can be also applied. However, the energy consumption of the wake-up receiver is so small that the impact of such an energy management on total energy consumption would be negligible.

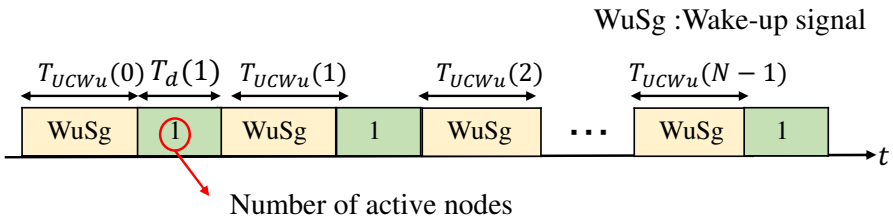


Figure 3.5: Theoretical analysis model of UCWu.

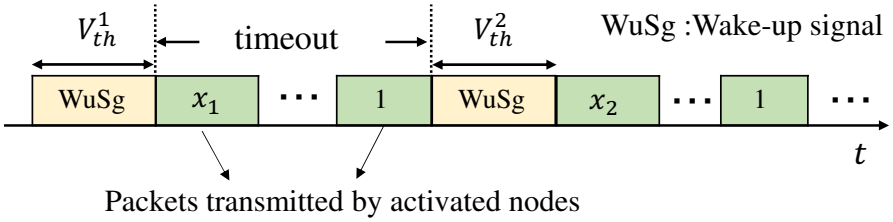


Figure 3.6: Theoretical analysis model of CDCoWu.

where $E_{total}(N)$ is given by eq. (3.2).

Analysis of UCWu scheme

Fig. 3.5 shows the model of theoretical analysis for UCWu. In the case of top-k query employing UCWu, the sink individually activates each node by sending the wake-up signal whose length corresponds to UCWuID. Hence, data collection delay of UCWu is given by

$$\begin{aligned} T_d^{UCWu}(N) &= N{\cdot}T_d(1) + \sum_{i=0}^{N-1} T_{UCWu}(i) \\ &= N{\cdot}T_d(1) + T_{min}{\cdot}N + \frac{T_{step}}{2}N(N-1), \end{aligned} \tag{3.5}$$

where $T_{UCWu}(i)$ [s] is the frame length corresponding to the UCWuID assigned to sensor node i. Here, frame length of UCWu is designed as $T_{UCWu}(i) = T_{min} + T_{step}{\times}i$ $(i = 0, 1, 2, \ldots)$. As for total energy consumption of UCWu, considering that N nodes separately transmit their data in reply to each wake-up request, it can be expressed as

$$E_{total}^{UCWu}(N) = N{\cdot}E_{total}(1). \tag{3.6}$$

Analysis of CDCoWu scheme

Fig. 3.6 shows the model of theoretical analysis for CDCoWu. As described in Sec. 5.3.2, both N-CDCoWu and V-CDCoWu transmit (multiple) wake-up signals

until the data collection of top-k observations is completed. For each wake-up signal, the number of activated nodes randomly varies depending on the distribution of observed value and CD_{step}. Here, CD_{step} and V_{step} are respectively set as $CD_{step} = mV_{step}$ and $V_{step} = lq_{step}$, where m and l are positive integers. Hereafter, lm is denoted as m'. On the other hand, q_{step} is determined as $q_{step} = \frac{V_{max}-V_{min}}{I_{max}}$, where I_{max} is the total number of quantization intervals calculated as $I_{max} = 2^{b_q}$. For each wake-up signal specifying a value threshold, different number of activated nodes try to send their packets with p-persistent CSMA. For simplicity of analysis, it is assumed that timeout for each wake-up, during which the sink waits for replies from the activated nodes, can be ideally set to a value that is required for all activated nodes to complete their transmissions. The sink continues to send wake-up requests until the size of collected set, n_k, reaches k. Denote the probability that a sensor node observes a value in the n-th quantization interval by $P_{int(n)}$, and by random variables X_n $(n = 1, ..., I_{max})$ the number of nodes included in the n-th quantization interval. X_n follow multinomial distribution, whose Probability Mass Function (PMF) is expressed as [67]

$$\begin{aligned}&\mathbb{P}(X_1 = q_1, \ldots, X_{I_{max}} = q_{I_{max}}) \\ &= \begin{cases} N! \prod_{n=1}^{I_{max}} \frac{P_{int(n)}^{q_n}}{q_n!} & (\sum_{n=1}^{I_{max}} q_n = N) \\ 0 & (\text{otherwise}), \end{cases}\end{aligned} \tag{3.7}$$

where N is the number of sensor nodes. Here, $P_{int(n)}$ depends on the assumed distribution of observed value. In the case of uniform distribution, $P_{int(n)}$ is given by

$$P_{int(n)} = \begin{cases} \frac{q_{step}}{V_{max}-V_{min}} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{(V_{max}-V_{min})-(I_{max}-1)q_{step}}{V_{max}-V_{min}} & (n = I_{max}). \end{cases} \tag{3.8}$$

In the case of exponential distribution, whose PDF is defined in Sec. 3.1, $P_{int(n)}$ is expressed as

$$P_{int(n)} = \begin{cases} \frac{\int_{V_{max}-nq_{step}}^{V_{max}-(n-1)q_{step}} \mathrm{e}^{\alpha x} dx}{\int_{V_{min}}^{V_{max}} \mathrm{e}^{\alpha x} dx} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{\int_{V_{min}}^{V_{max}-(I_{max}-1)q_{step}} \mathrm{e}^{\alpha x} dx}{\int_{V_{min}}^{V_{max}} \mathrm{e}^{\alpha x} dx} & (n = I_{max}). \end{cases} \tag{3.9}$$

Furthermore, for normal distribution, $P_{int(n)}$ is expressed as

$$P_{int(n)} = \begin{cases} \frac{\int_{V_{max}-nq_{step}}^{V_{max}-(n-1)q_{step}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{\int_{V_{min}}^{V_{max}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{\int_{V_{min}}^{V_{max}-(I_{max}-1)q_{step}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{\int_{V_{min}}^{V_{max}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx} & (n = I_{max}). \end{cases} \tag{3.10}$$

Given the i-th realization (sample) of $\{X_n\}$ as $s_i = \{q_1^i, q_2^i, \ldots, q_{I_{max}}^i\}$, the number of wake-up nodes at ζ-th wake-up trial with $CD_{step} = m'q_{step}$ is given by

$$x_\zeta^i = \sum_{l=m'(\zeta-1)+1}^{m'\zeta} q_l^i. \tag{3.11}$$

The number of wake-up signals required to complete the collection of top-k observations in i-th realization is denoted by n_w^i. By using eqs. (3.1) and (3.7), mean total duration required for data transmissions with p-persistent CSMA (excluding periods of wake-up signal transmissions) can be calculated as

$$T_{CDCoWu}^{Data}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \sum_{\zeta=1}^{n_w^i} T_d(x_\zeta^i), \tag{3.12}$$

where M is the total number of realizations of multinomial distribution given in eq. (3.7). Note that n_w^i differs for N-CDCoWu or V-CDCoWu.

As mentioned in Sec. 5.3.2, the sink reduces the threshold of CoWu by CD_{step} for each wake-up trial. Then, the frame length of the ζ-th wake-up signal is set as

$$T_{CDCoWu}^{WuS}(\zeta) = T_{min} + T_{step}(m\zeta - 1). \tag{3.13}$$

Therefore, the mean total duration required for transmitting wake-up signals is expressed as

$$T_{CDCoWu}^{\Sigma WuS}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \sum_{\zeta=1}^{n_w^i} T_{CDCoWu}^{WuS}(\zeta). \tag{3.14}$$

Thus, mean total delay to complete the collection of top-k observations with CDCoWu is the sum of eqs. (3.12) and (3.14)

$$T_d^{CDCoWu}(N) = T_{CDCoWu}^{Data}(N) + T_{CDCoWu}^{\Sigma WuS}(N). \tag{3.15}$$

Analogously, the mean of total energy consumption of CDCoWu, $E_{total}^{CDCoWu}(N)$ can be calculated as

$$E_{total}^{CDCoWu}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \cdot \sum_{\zeta=1}^{n_w^i} E_{total}(x_\zeta^i). \tag{3.16}$$

As already mentioned, the required number of wake-up trials, n_w^i, is different for N-CDCoWu and V-CDCoWu. Its maximum value denoted as W_{max}, is expressed as

$$W_{max} = \lceil \frac{V_{max} - V_{min}}{CD_{step}} \rceil. \tag{3.17}$$

In the case of N-CDCoWu, the total number of wake-up trials to complete top-k node-set collection is minimum j ($1 \leq j \leq W_{max}$) satisfying

$$\sum_{\zeta=1}^{j} x_{\zeta}^{i} \geq k. \tag{3.18}$$

On the other hand, with V-CDCoWu, the sink increases the size of collected set by 1 when there exists at least one node observing the data within a quantization interval. Thus, for a sample s_i, the size of collected set at ζ-th wake-up trials is

$$y_{\zeta}^{i} = \sum_{l=m'(\zeta-1)+1}^{m'\zeta} u(q_l^i - 1), \tag{3.19}$$

where $u(x)$ is a step function, defined as

$$u(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (otherwise). \end{cases} \tag{3.20}$$

With V-CDCoWu, it can happen that the size of collected set does not exceed k even if the data collections from all sensor nodes deployed over the sensing field are completed, especially when the observed value is concentrated into a certain range. Therefore, two conditions for finishing top-k value-set query at j-th wake-up trial are defined: one is when the size of collected set reaches k and the other is when completing data collections from all sensor nodes, which are expressed as

$$\begin{cases} \sum_{\zeta=1}^{j} y_{\zeta}^{i} \ \geq \ k & (if\ \exists\ s.t. \sum_{\zeta=1}^{W_{max}} y_{\zeta}^{i} \ \geq \ k) \\ \sum_{\zeta=1}^{j} x_{\zeta}^{i} \ = \ N & (otherwise). \end{cases} \tag{3.21}$$

3.4 Numerical Results

3.4.1 Approximate Analysis with MCMC

The exact calculation of mean delay and energy consumption of CDCoWu, see eqs. (3.12), (3.14), and (3.16), requires consideration of all realizations of multinomial distribution given in eq. (3.7), whose number is $\binom{I_{max}+N-1}{N}$. This calculation becomes intractable when N and/or I_{max} become large. For this reason, this thesis resorts to MCMC method [68] to obtain approximate results. Specifically, this work employs Metropolis algorithm [68] that generates sequence of samples (realizations), as described below:

- STEP 0: [Set initial state] Set an initial state $X^{(0)} = \bigcup_{n=1}^{I_{max}} q_n^{(0)}$, where $q_n^{(0)}$ is the number of nodes belonging to the n-th quantization interval under the initial state. The initial state is set so that the number of nodes in each interval is distributed as fairly as possible, with a constraint that the total number of nodes is N, i.e., $\sum_{n=1}^{I_{max}} q_n^{(0)} = N$. This constraint is the same in any arbitrary state (i.e., state $X^{(t)}$ is set s.t. $\sum_{n=1}^{I_{max}} q_n^{(t)} = N$).

- STEP 1: [Generate new sample] Create a new sample $X^{'}$ from the current state $X^{(t)}$. In each state, there are I_{max} intervals, from which one interval i with number of nodes n_i satisfying $n_i \geq 1$ is randomly chosen and n_i is decreased by 1. Then, another interval j with number of nodes n_j satisfying $n_j < N$ is randomly selected and n_j is increased by 1. Through these operations, new sample $X^{'}$ is generated.

- STEP 2: [Calculate transition cost] Calculate transition cost as $\frac{\mathbb{P}(X^{'})}{\mathbb{P}(X^{(t)})}$, by using PMF given by eq. (3.7).[4]

- STEP 3: [Update state] Generate a random number $r \in [0, 1]$, following uniform distribution, and decide the next state as follows:

$$X^{(t+1)} = \begin{cases} X^{'} & (if\ r \leq \frac{\mathbb{P}(X^{'})}{\mathbb{P}(X^{(t)})}) \\ X^{(t)} & (\text{otherwise}). \end{cases} \tag{3.22}$$

- STEP 4: [Go back to STEP 1].

One round of STEP 1 to STEP 4 defines a Monte Carlo (MC) step, which is repeated Z times. For each MC step, by using the distribution of nodes represented by the state of $X^{(t)}$, delay and energy consumption are calculated through the equations derived in Sec. 3.3, and stored as $t^{(z)}$ and $e^{(z)}$, respectively. After repeating MC step Z times, the approximate values of mean delay and energy consumption are respectively calculated as

$$\widehat{T}_{delay}^{CDCoWu} = \frac{1}{Z}\sum_{z=1}^{Z} t^{(z)} \ , \tag{3.23}$$

$$\widehat{E}_{total}^{CDCoWu} = \frac{1}{Z}\sum_{z=1}^{Z} e^{(z)} \ . \tag{3.24}$$

A unique feature of Metropolis method is that each sample to calculate the expectations of function in eqs. (3.12), (3.14), and (3.16) is selected from the

[4]More details on transition cost can be found in [68].

region where its probability is relatively high, while searching the other regions with probability r. After repeating a sufficient number of MC steps, it can be expected to obtain approximate results that are close to exact solutions.

3.4.2 Impact of Different Parameters on the Performance of CDCoWu

This subsection investigates the performance of N-CDCoWu and V-CDCoWu for different parameters such as CD_{step}, b_q, α, and k via numerical results obtained by the theoretical analysis and computer simulations. We compare the results of theoretical analysis with those of computer simulations in order to assess the validity of the derived equation and approximate analysis. The simulations are conducted by a custom-made simulator created with Matlab. First, a set of observed data of sensor nodes is randomly generated following the uniform, exponential, or normal distribution. Then, the procedures of wake-up/data collections following the system model given in Sec. 3.1 as well as IDWu and CDCoWu explained in Sec. 3.3 are simulated, including the actual operations of p-persistent CSMA protocol. The data collection delay and energy consumption of each node are recorded when the sink completes top-k query. These processes are repeated for different samples of observed data of sensor nodes, whose averaged results are plotted in the following figures. The values of the other parameters used for numerical evaluations are shown in Table 3.1. Further, the values of l (see Sec. 3.3.2) are set to 2^{b_q-9} if b_q is larger than 9, and otherwise set to 1. This restriction comes from constraints on range and resolution of wake-up frame length, specified by considering the standard IEEE 802.15.4g [62][3]. The maximum size of IEEE 802.15.4g frame is 2055 bytes while the sizes from 28 to 135 bytes are commonly employed for their data transmissions. Therefore, the length corresponding to the range between 135 bytes and 2055 bytes is used for the wake-up frame length. With the assumption that the transmission rate is 100 kbps and the step to vary frame length is 1 byte, the temporal resolution to change the wake-up frame length is 0.08 ms. However, considering the detection accuracy, the step of 2 bytes is employed as the interval to change the wake-up frame length [3]. This allows us to apply 960 different frame length for mapping with the observed value. If the number of quantization bits is equal to or less than 9, i.e., if the number of quantization intervals is less than or equal to 960, the appropriate mapping can be achieved by setting $l = 1$. Otherwise, the mapping cannot be properly made unless the value of l is appropriately adjusted.

Table 3.1: Parameters employed for Numerical Evaluations.

Parameters	Values
Data transmission rate	100 kbps
Length of packet in time slots L	10
Time slot length δ	320 μsec
Power consumption in Transmit state ξ_T	55 mW [69]
Power consumption in Receive state ξ_R	50 mW [69]
Distribution of observed value $[V_{min}, V_{max}]$	$[0, 50]$
Interval of wake-up frame length T_{step}	0.16 msec [3]
Minimum wake-up frame length T_{min}	10.8 msec [3]
wake-up Frame length corresponding to BCWuID T_{BCWu}	10.8 msec

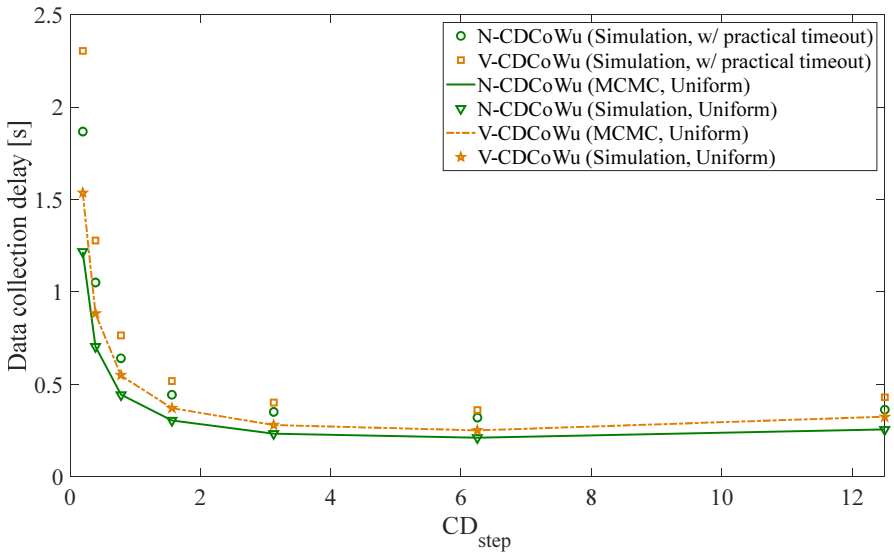


Figure 3.7: Data collection delay against CD_{step} for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$).

Impact of CD_{step}

Figs. 3.7 and 3.8 show data collection delay and total energy consumption of N-CDCoWu and V-CDCoWu against CD_{step}, respectively, where the number of MC steps, Z, is set to 1,000,000 (this assumption will hold hereafter unless otherwise stated). This evaluation sets $N = 100$, $k = 25$, $b_q = 8$, and $e_c = 0$, assuming uniform distribution of observed values. Here, $p = 0.0606$ is employed, which corresponds to the practical size of back-off window of 32 [64]. Obviously, the figures show that the results obtained with the approximate analysis coincide

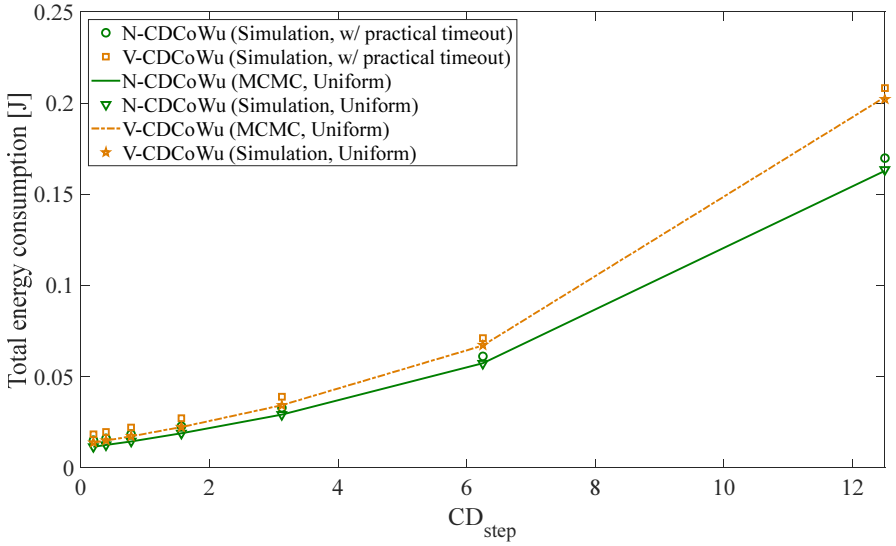


Figure 3.8: Total energy consumption against CD_{step} for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$).

with simulation results very well, validating the approach employed in this thesis. Fig. 3.7 shows that the data collection delay is a convex function of CD_{step}, such that there is a minimal value with CD_{step}^{opt}. When $CD_{step} < CD_{step}^{opt}$, the delay increases because a smaller number of nodes observe the value within each CD_{step}, which requires increased number of transmissions of wake-up signals. For $CD_{step} > CD_{step}^{opt}$, the delay increases due to the congestion caused by the increased number of nodes that observe values within each CD_{step}, simultaneously wake up and contend for the shared channel. Note that the convexity of data collection delay against CD_{step} can be also observed in theoretical equations derived in Sec. 3.3. For $CD_{step} < CD_{step}^{opt}$, x_ζ^i in eq. (3.11) or y_ζ^i in eq. (3.19) becomes smaller, by which n_w^i satisfying eq. (3.18) or (3.21) becomes larger. This makes the duration required for transmitting wake-up signals in eq. (3.14) larger. While this negative effect is alleviated as the value of CD_{step} increases, the delay for transmitting data expressed in eq. (3.12) becomes larger since x_ζ^i in $T_d(x_\zeta^i)$, which is a monotonically increasing function against x_ζ^i as given in eq. (3.1), increases. Due to this negative effect with increasing CD_{step}, data collection delay starts to increase again. On the other hand, Fig. 3.8 shows that total energy consumption increases as CD_{step} becomes larger. This is because the probability that a node wakes up in a trial increases as CD_{step} becomes larger, which increases the amount of energy consumed for each node to resolve contention.

Figs. 3.7 and 3.8 also show that V-CDCoWu has larger delay and energy consumption than N-CDCoWu. This is because, even if the number of wake-up

nodes exceeds k in V-CDCoWu, the size of collected set does not necessarily reach k. In this case, the sink lowers the value threshold in order to collect new data, which increases the number of wake-up trials and activated nodes, resulting in larger delay and energy consumption.

Note that, in Figs. 3.7 and 3.8, simulation results of CDCoWu are also plotted when we employ a practical timer for the sink to wait for the replies from activated nodes instead of ideally assuming that the sink can set the timer to the value required for all activated nodes to complete their transmissions. The timer is set every time the sink completes the reception of a packet from a sensor node. Once the timer expires before detecting a packet of the other nodes, the sink decides that there is no longer any node to respond to the wake-up request. Here, the timer should be sufficiently longer than the period to accommodate a packet transmission of a single node, including back-off period. Two different timers are prepared: one is the value used for wake-up trials before the sink grasps top-k set, and the other is the value employed for the last wake-up trial in which the size of top-k set reaches k. The former value is set to be 32 slots, as $p = 0.0606$ corresponds to the back-off window size of 32 [64]. The latter value is set to be sufficiently larger than the former one, i.e., 320 slots, in order for the sink to make sure that there is no more node belonging to the exact top-k set. Note that, practically, this timer can be much shorter since back-off period of each node is upper-bounded, which is controlled by maximum size of back-off element [64]. For instance, in IEEE 802.15.4 MAC protocol, maximum size of back-off element is commonly set to be 5, which makes each node transmit a packet at least within 32 slots after the last busy period of channel. In this work, 320 slots are needed because the back-off period of p-persistent CSMA, which is employed just for simplifying the analysis, is not upper-bounded. The results with a practical timer are marked with "Simulation w/ practical timeout" in the figures. From these results, it can be seen that data collection delay and total energy consumption are increased with the practical timeout. The longer timer than the ideal case results in the increase of data collection delay. On the other hand, the shorter timer results in the increase of total energy consumption since not all activated nodes complete data transmissions within the corresponding wake-up trial, and they remain to be active until the next wake-up trials. However, focusing on the results for the range close to CD_{step}^{opt}, it can be seen that the increase of data collection delay and total energy consumption is not significant, which indicates that the assumption on the ideal timeout is reasonable.

Impact of quantization step

The number of discrete values that each sensor node can observe and report changes in accordance with quantization bit rate b_q of each sensor. In order to

investigate the impact of the value of b_q, b_q is varied from 2 to 10, and the performance of data collection delay and total energy consumption of N-CDCoWu and V-CDCoWu is evaluated. Figs. 3.9 and 3.10 show results with uniform distribution of observed values, for $N = 100$, $k = 25$, $CD_{step} = V_{step}$, $p = 0.0606$, and $e_c = 0$. Note that, this evaluation employed the smallest possible CD_{step} in order to clarify the difference of impact of b_q on N-CDCoWu and V-CDCoWu.

Fig. 3.9 shows that V-CDCoWu requires larger delay than N-CDCoWu. With the smaller value of b_q (i.e., with the larger quantization step), the number of nodes waking up against a wake-up signal becomes larger. Therefore, as b_q becomes smaller, the sink can complete top-k data collection with relatively smaller number of wake-up signals, resulting in smaller data collection delay. On the other hand, as the value of b_q increases, the required number of wake-up trials increases. That is why data collection delay of CDCoWu monotonically increases until b_q reaches 9. Fig. 3.10 shows that total energy consumption of V-CDCoWu and N-CDCoWu decreases as the value of b_q increases. Note that, as mentioned in Sec. 3.3.2, the value of V_{step} is changed in accordance with b_q. Under the setting of $CD_{step} = V_{step}$, smaller values of b_q lead to worse resolution in terms of quantization, and many nodes tend to be included into the same V_{step} that becomes larger with the decrease of b_q. This increases the probability that many nodes wake up in each wake-up trial, resulting in severe congestion and a larger energy consumption. Note that, if b_q is larger than 9, V_{step} takes the same value due to the upper bound resulting from mapping design of the frame length as mentioned in Sec. 3.4.2. However, total number of observable data increases, therefore, the larger b_q becomes, the more values the sink can collect for the same size of CD_{step}. Therefore, for V-CDCoWu, data collection delay in Fig. 3.9 and total energy consumption in Fig. 3.10 decrease when the value of b_q changes from 9 to 10.

Finally, Figs. 3.9 and 3.10 show that data collection delay and total energy consumption of V-CDCoWu tend to converge into those of N-CDCoWu. As b_q becomes larger, the resolution of quantization becomes better, and the probability for different sensors to observe the same value becomes smaller. With the extreme condition of $b_q \to \infty$, this probability becomes negligibly small. In this case, the set to be collected by N-CDCoWu and V-CDCoWu are almost the same, and the data collection delay and total energy consumption of V-CDCoWu and N-CDCoWu tend to become equal. For instance, for b_q of 20, it has been confirmed with simulations that data collection delay of V-CDCoWu and N-CDCoWu are 2.8974 [s] and 2.8939 [s], respectively, and total energy consumption of V-CDCoWu and N-CDCoWu are both 0.0111 [J].

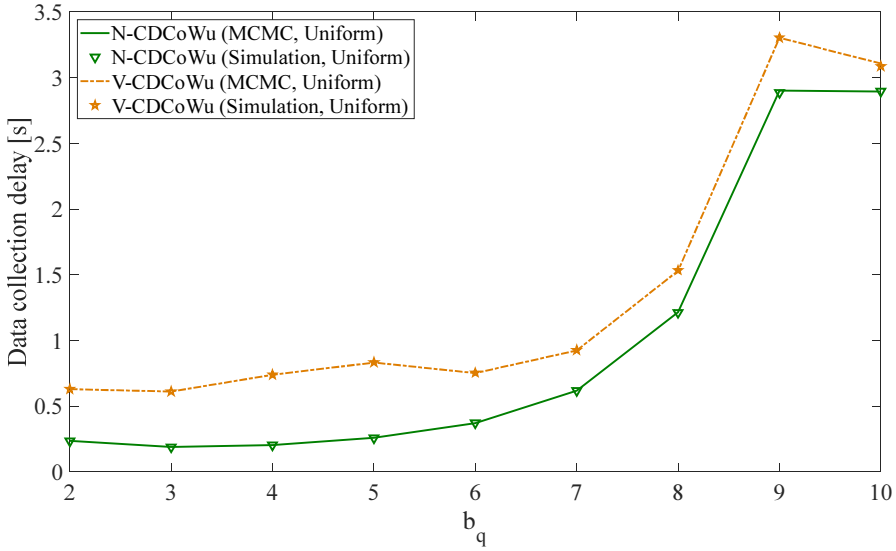


Figure 3.9: Data collection delay against b_q for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$).

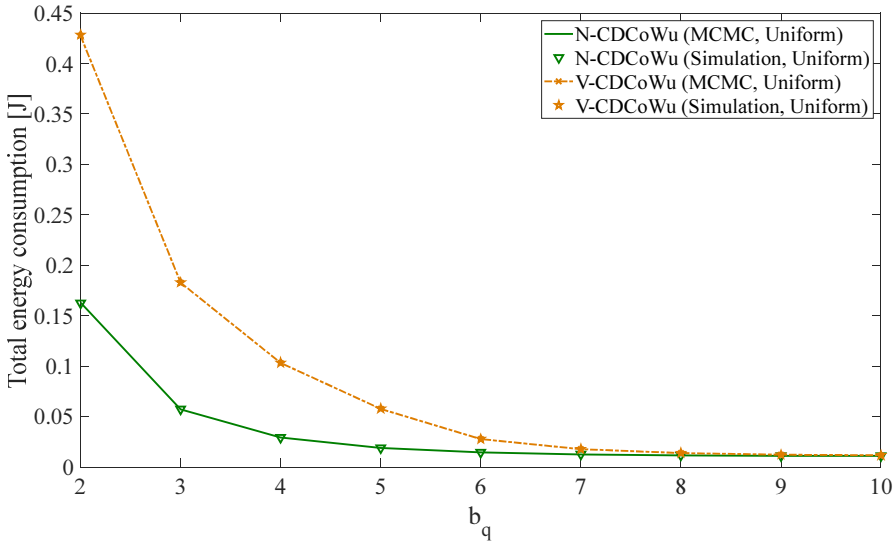


Figure 3.10: Total energy consumption against b_q for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$).

Impact of distribution of observed values

Here, the impact of distribution of observed values on delay and energy consumption of CDCoWu is evaluated by taking exponential distribution as an example, where the value of α in eq. (3.9) is changed from 0 (uniform distribution) to 0.1

with the step of 0.01, for $N = 100$, $k = 25$, $CD_{step} = 10V_{step}$, $p = 0.0606$, $e_c = 0$, and $b_q = 8$. Fig. 3.11 shows data collection delay of N-CDCoWu and V-CDCoWu against α. From Fig. 3.11, one can see that data collection delay of N-CDCoWu and V-CDCoWu have the minimum value for $\alpha = 0.04$ and $\alpha = 0.08$, respectively. As the value of α is increased from 0, the distribution of observed values becomes non-uniform, and more biased toward higher values. Therefore, the size of collected set reaches k with smaller number of wake-up signals, which decreases data collection delay of both N-CDCoWu and V-CDCoWu. However, the data collection delay is increased after α increases over the optimal value, as the negative impact of congestion caused by simultaneous wake-up of nodes observing higher values becomes more dominant than the reduction in the number of wake-up trials. For high values of α, many nodes tend to hold the same value and the size of collected set of V-CDCoWu becomes smaller against the number of nodes with their data collected, which then requires more number of wake-up trials to collect new value. Therefore, the positive effect to reduce the number of wake-up trials with increased α is smaller for V-CDCoWu than that for N-CDCoWu; this also explains why the value of α at which data collection delay takes the minimum value for V-CDCoWu happens for smaller value of α.

The convexity described above can also be observed in the analysis and equations derived in Sec. 3.3, as similarly explained for the impact of CD_{step} in Fig. 3.7. As the value of α becomes larger, from eq. (3.9), it can be seen that the probability that each node observes higher values becomes higher, which increases x_ζ^i against the smaller ζ. This makes n_w^i satisfying eq. (3.11) or (3.19) smaller, which makes the duration required for transmitting wake-up signals in eq. (3.14) smaller. On the other hand, the delay for transmitting data given in eq. (3.12) becomes larger since x_ζ^i in eq. (3.12) for smaller ζ can become sufficiently large to cause severe congestion. Therefore, depending on whether delay for data transmissions in eq. (3.12) or that for transmitting wake-up signals in eq. (3.14) is dominant, data collection delay increases or decreases, which gives its convexity against the value of α.

It has been also confirmed that total energy consumption of both N-CDCoWu and V-CDCoWu monotonically increase as α becomes larger due to an increased level of congestion and that N-CDCoWu consistently requires smaller energy consumption than V-CDCoWu as the number of nodes to collect the data from are smaller.

Impact of k

Figs. 3.12 and 3.13 show data collection delay and total energy consumption of N-CDCoWu and V-CDCoWu against k, respectively. This evaluation sets $N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, $p = 0.0606$, assuming uniform distribution of observed

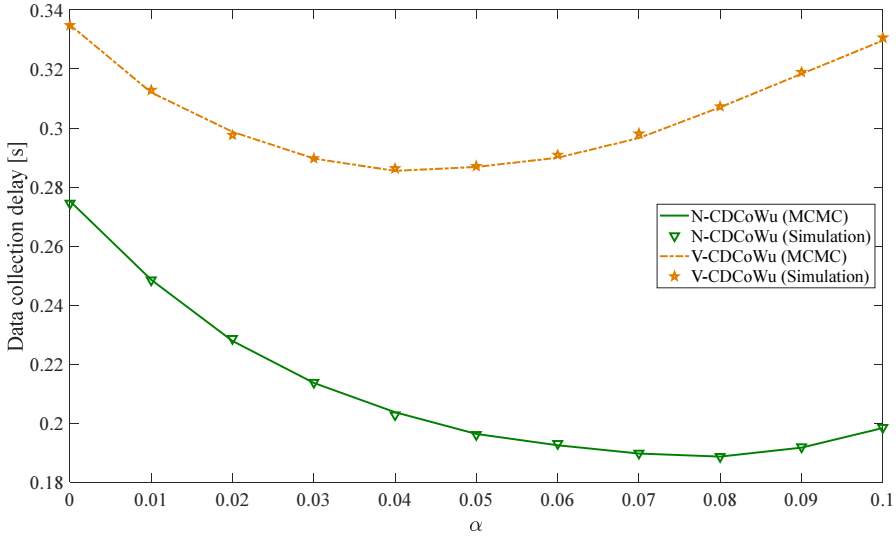


Figure 3.11: Data Collection delay against α for N-CDCoWu and V-CDCoWu ($N = 100$, $k = 25$, $CD_{step} = 10V_{step}$, and $p = 0.0606$).

values. The results with different e_c are plotted, which shows the agreement between approximate and simulation results. It can also be seen that the performance is degraded as the probability of packet losses due to channel impairments becomes higher. Focusing on the impact of k, it can be noticed that the selection of the smallest value of k is desirable in terms of both data collection delay and total energy consumption. However, practically, the application scenario should be also considered to decide an appropriate value of k. For instance, top-k query in WSNs is useful when the resources required to execute some actions against the sensing targets are limited. In this case, the prioritized usage of the resources is desired. The examples include ecology observation of the bird species in a forest by ornithologists [21] and forest fire monitoring [32] where a user dispatches drones to the fire environment to put out the fire. In these scenarios, the optimal value of k depends on the type and number of resources available for users. On the other hand, if top-k query is simply applied for monitoring purpose, the optimal value of k is the maximum value satisfying the constraints of communication characteristics such as data collection delay and total energy consumption (i.e., battery lifetime).

3.4.3 Comparison of Different Wake-up Schemes

Now data collection delay and total energy consumption achieved by different wake-up schemes are compared. The performance of each wake-up scheme depends on different parameters such as p and CD_{step}. For instance, CDCoWu can achieve

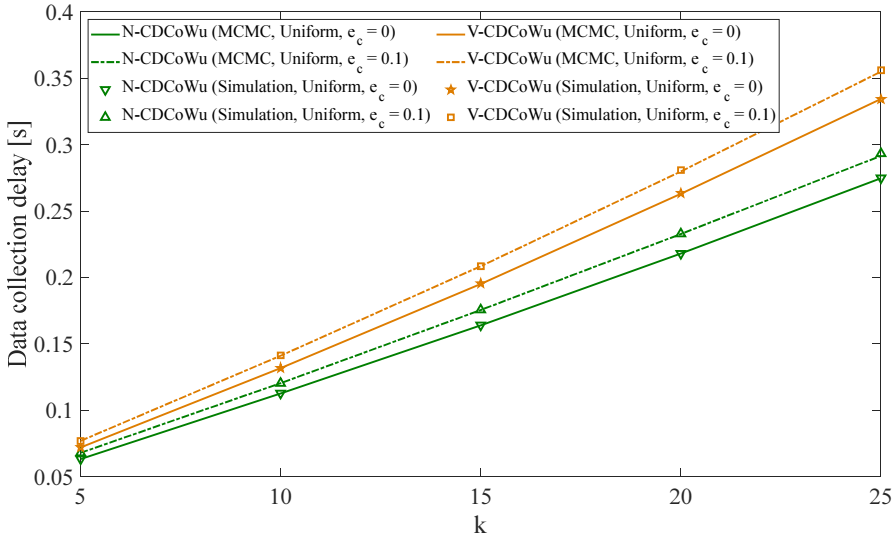


Figure 3.12: Data collection delay against k for N-CDCoWu and V-CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$).

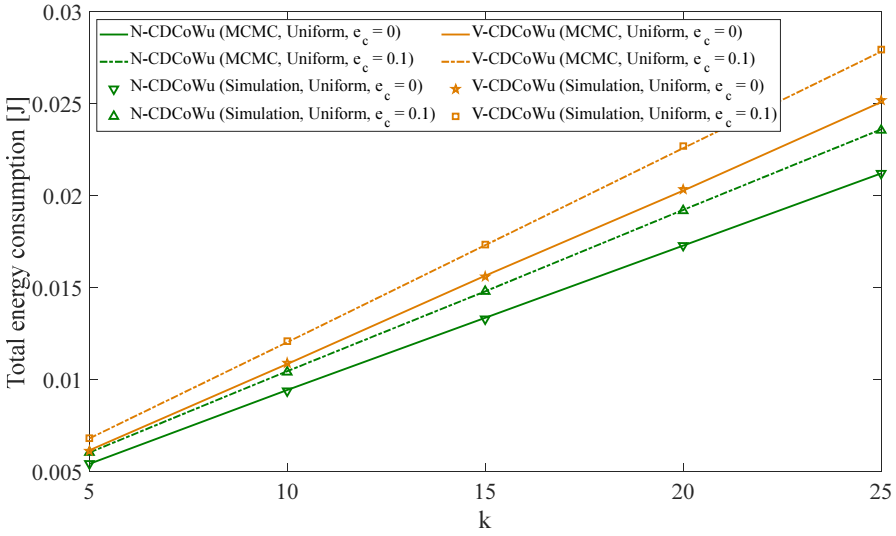


Figure 3.13: Total energy consumption against k for N-CDCoWu and V-CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$).

high energy efficiency at the cost of delay by setting CD_{step} to small value, as confirmed by the evaluations in Sec. 3.4.2. Because of such trade-offs, it is difficult to compare different wake-up schemes unless either target delay or energy is fixed. Thus, the effectiveness of CDCoWu is investigated by comparing its total energy consumption with IDWu on condition that delay of CDCoWu does not exceed that

of IDWu. This upper bound on delay is obtained with the optimized transmission probability p of each IDWu (BCWu, UCWu), i.e., minimum data collection delay for different number of nodes. The optimal parameters are chosen as follows:

- UCWu: In UCWu, only a single node wakes up for each wake-up trial conducted by the sink, which means that there are no collisions. Therefore, the transmission probability of UCWu is set as $p = 1$. The optimized data collection delay of UCWu for different number of sensor nodes N is defined as $\tau^{opt}_{UCWu}(N)$.

- BCWu: In BCWu, the optimal transmission probability depends on the number of sensor nodes. The optimal probability p_{opt} is obtained for different number of sensor nodes N. Here, the optimized data collection delay of BCWu for different number of sensor nodes is defined as $\tau^{opt}_{BCWu}(N)$.

- CDCoWu: In CDCoWu, the achievable performance depends not only on the transmission probability p but also on CD_{step}. The set of p and CD_{step} values is obtained, which minimizes the total energy consumption for different number of sensor nodes on condition that delay does not exceed $\tau^{opt}_{UCWu}(N)$ or $\tau^{opt}_{BCWu}(N)$. Two sets of parameters are defined for $\{CD_{step}, p\}$: $\mathbb{U}(N)$, the optimized set for comparison with UCWu, and $\mathbb{B}(N)$, for comparison with BCWu, which are required to satisfy

$$\mathbb{U}(N) = \min_{CD_{step},p} \{E^{CDCoWu}_{total}(N)\} \; s.t. \; \tau \; \leq \; \tau^{opt}_{UCWu}(N) \; , \tag{3.25}$$

$$\mathbb{B}(N) = \min_{CD_{step},p} \{E^{CDCoWu}_{total}(N)\} \; s.t. \; \tau \; \leq \; \tau^{opt}_{BCWu}(N) \; , \tag{3.26}$$

  where τ is data collection delay of CDCoWu for the optimized set of parameters, employed in the evaluations.

Note that this thesis resorts to a heuristic approach to obtain the optimal values of the above-mentioned parameters. That is, approximate results of mean delay and total energy consumption are evaluated for a wide range of parameters, and the appropriate values of CD_{step} and p are selected. This is because of the difficulty to derive the expressions of mean delay and total energy consumption of the proposed scheme in closed form. This means that the employed parameters may not be strictly optimal.

Comparison between BCWu and UCWu

Before investigating the efficiency of CDCoWu, we evaluate and compare delay and total energy consumption achieved by BCWu and UCWu. Note that, with

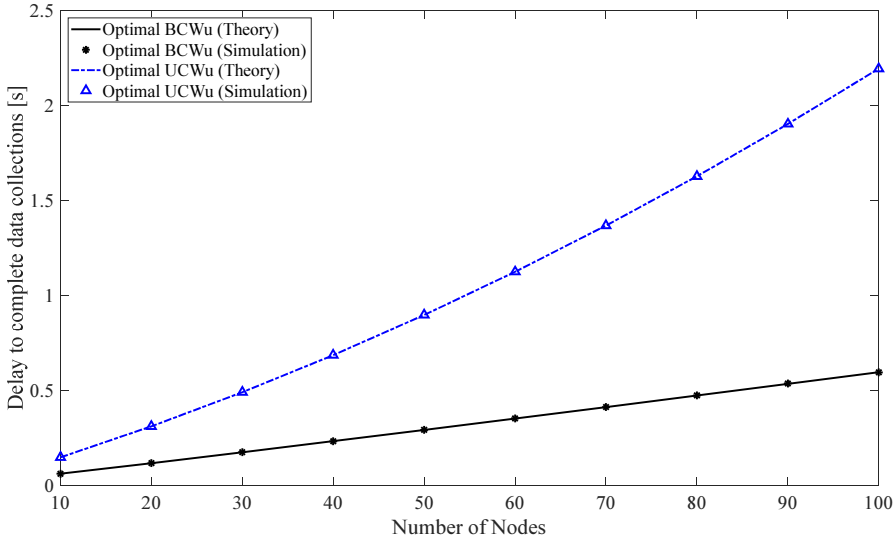


Figure 3.14: Delay against the number of sensor nodes for BCWu and UCWu.

BCWu and UCWu, all nodes are activated, which means that its performance does not depend on k of top-k data collection. Figs. 3.14 and 3.15 respectively show delay and total energy consumption of BCWu and UCWu. First, from these figures, we can clearly see the agreement between theoretical results obtained with eqs. (3.3) - (3.6) and simulation results. Thus, we can confirm the validity of the derived equations. From Fig. 3.14, under the condition of optimal p, it can be seen that UCWu has worse performance than BCWu in terms of data collection delay. This is due to large amount of wake-up overhead, i.e., the transmission of wake-up signal is required before every data transmission in the case of UCWu. On the other hand, BCWu realizes better performance than UCWu because all sensor nodes can be woken up with only a single transmission of wake-up signal from the sink, and thanks to the optimization of p, nodes hardly experience collisions (note that collision is the major cause to deteriorate data collection delay with BCWu). Next, looking at Fig. 3.15, we can see that the total energy consumption of BCWu becomes significantly larger than that of UCWu as the number of nodes increases. This is because many nodes need to stay awake during contention process of p-persistent CSMA with BCWu while, with UCWu, only a single node wakes up and can transmit data and quickly go back to sleep without spending wasteful energy for contention process. From these results, we can confirm that BCWu is superior to UCWu in terms of delay while UCWu outperforms BCWu in terms of total energy consumption.

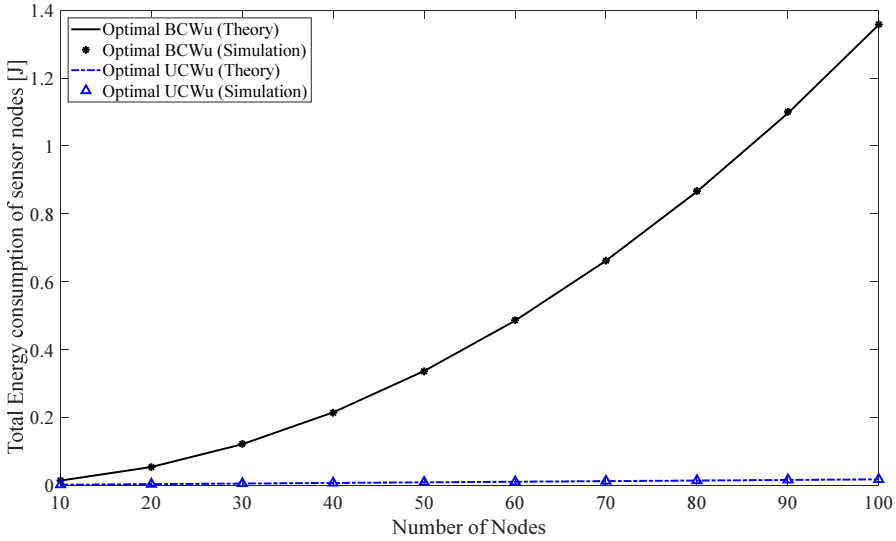


Figure 3.15: Total energy consumption against the number of sensor nodes for BCWu and UCWu.

Comparison between IDWu and CDCoWu

Figs. 3.16 and 3.18 show the comparison of total energy consumption of CDCoWu with BCWu and with UCWu, respectively, when they are applied for top-10 query. Furthermore, Fig. 3.17 shows a zoomed view of Fig. 3.16 in terms of CDCoWu. The results are plotted for distributions of observed values of uniform and exponential distribution with $\alpha = 0.1$ and $e_c = 0$. Hereafter, the number of MC steps, Z, is set to 100,000. As already noted, delay of CDCoWu is upper-bounded by that of BCWu and UCWu in the figures. The plots show that performance of IDWu obtained with the analysis coincide with simulation results very well. In Fig. 3.16, results for energy consumption of CDCoWu for the number of nodes of 10 and 20 are not plotted. This is because top-k node/value set collections are not completed under the delay constraint of $\tau^{opt}_{BCWu}(N)$ for these sets of parameters. Specifically, in order to satisfy the delay constraint of optimized BCWu, CDCoWu needs to reduce the number of transmissions of wake-up signals for data collections. In CDCoWu, the length of wake-up frame determined by mapping rule described in Sec. 3.2.1 is designed to alleviate the negative impact of wake-up overhead for a scenario of top-k query, i.e., higher (lower) value is mapped to shorter (longer) frame length. However, each length of wake-up frame of CDCoWu is equal to or longer than that of BCWu, T_{BCWu}, which is set to T_{min}. When the number of nodes is 10 or 20, (i.e., when N is small), in order to realize smaller delay than BCWu, the sink needs to wake up all nodes with smaller number of wake-up trials. However, as mentioned above, the length of wake-up frame of CDCoWu are longer

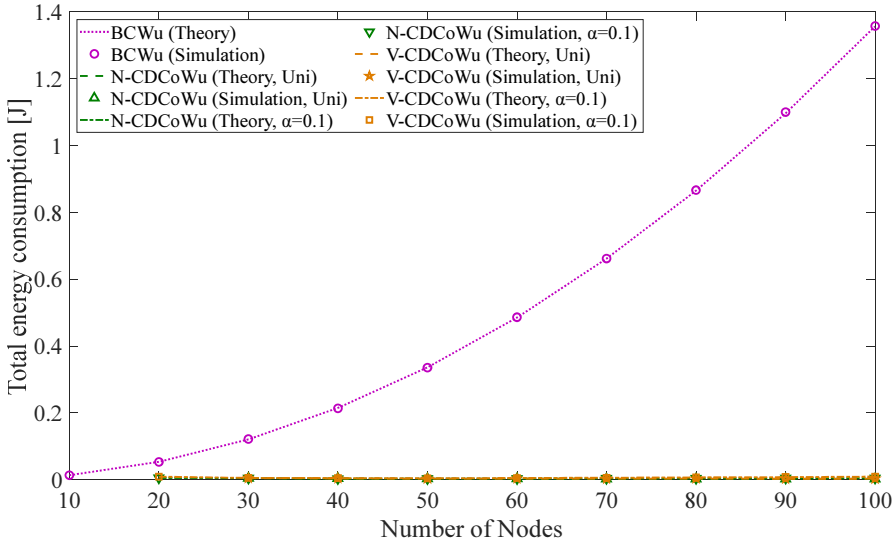


Figure 3.16: Total energy consumption against the number of sensor nodes for BCWu and CDCoWu.

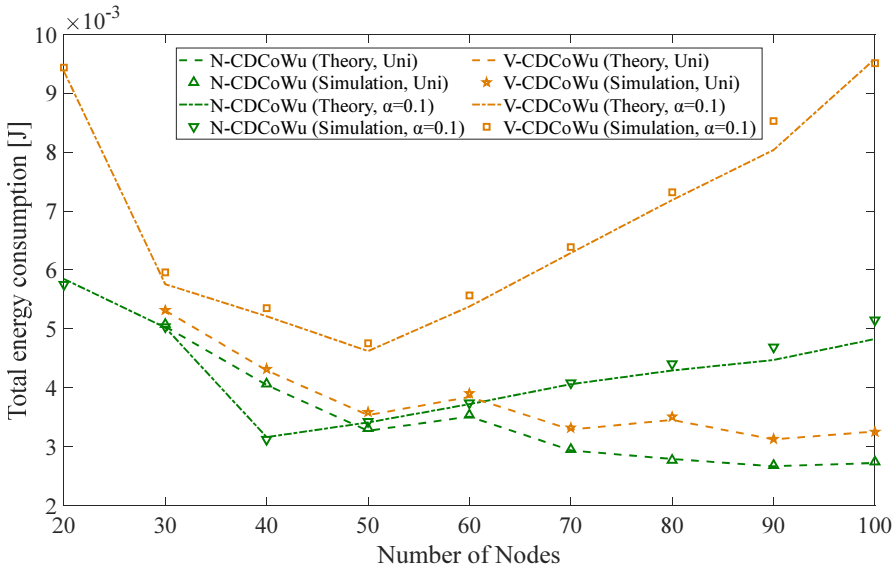


Figure 3.17: A zoomed view of Fig. 3.16 in terms of CDCoWu.

than that of BCWu. That is why CDCoWu can not satisfy the delay constraint of eq. (3.26), when the number of nodes is small.

Fig. 3.16 shows that total energy consumption of BCWu is much larger than that of CDCoWu since, with BCWu, the sink needs to aggregate data from all sensor nodes with relatively small transmission probability (e.g., when the num-

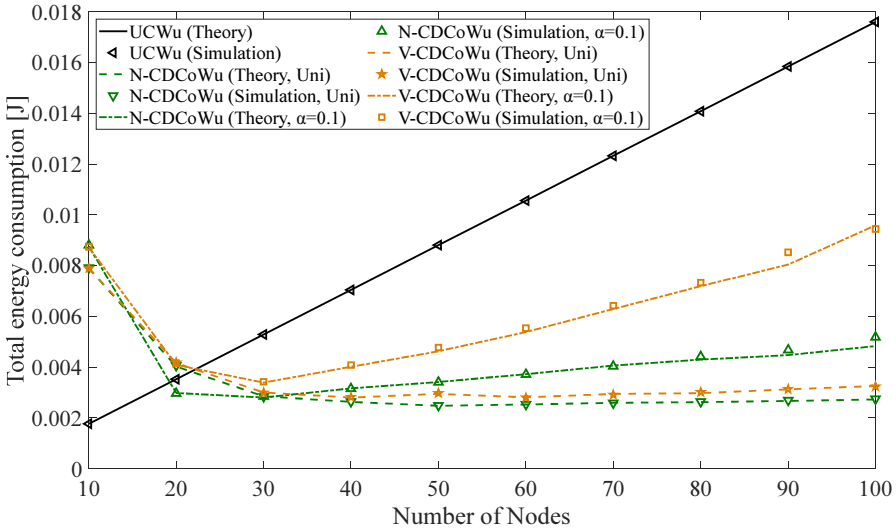


Figure 3.18: Total energy consumption against the number of sensor nodes for UCWu and CDCoWu.

ber of nodes $N = 100$, the transmission probability p needs to be set to 0.0111). In this case, many nodes need to stay awake for a long period of time to resolve contentions. On the other hand, CDCoWu can achieve small energy consumption, especially when the number of nodes is large. This is because, in CDCoWu, the sink can collect top-k observations just by sending wake-up signals a couple of times without waking up all sensor nodes. Under the constraint of data collection delay, total energy consumption of CDCoWu is minimized by employing relatively small CD_{step} with high transmission probability p, which makes small number of activated nodes succeed in data transmissions quickly. Furthermore, Fig. 3.17 shows a remarkable difference on total energy consumption of CDCoWu with the increasing number of nodes for uniform distribution and exponential distribution. Specifically, total energy consumption for uniform distribution becomes smaller as the number of nodes increases. On the other hand, total energy consumption for exponential distribution first decreases, and then increases after it reaches a certain point, against the number of nodes. The declining tendency for uniform distribution in Fig. 3.17 is mainly due to the adopted value of CD_{step}. As already discussed, when the number of nodes becomes larger, the sink can adopt relatively small CD_{step} values under the delay constraint, which reduces the number of wake-up nodes for each wake-up trial. Thus, total energy consumption for uniform distribution becomes smaller as the number of nodes increases thanks to the reduced level of congestion. The same effect can be observed for exponential distribution in Fig. 3.17, for a smaller number of nodes. However, for a larger number of nodes, more nodes make observations in the higher range of quantization

intervals for exponential distribution. The negative effect caused by simultaneous wake-up of nodes within the same interval becomes dominant, which is why total energy consumption for exponential distribution becomes larger over the range of $N \geq 40$ in N-CDCoWu and $N \geq 50$ in V-CDCoWu.

Fig. 3.18 shows that CDCoWu achieves smaller energy consumption than UCWu when the number of sensor nodes is equal to or more than 30. As mentioned in Sec. 3.4.3, UCWu realizes higher energy efficiency compared with BCWu, while requiring larger delay than BCWu to complete data collection. Thus, small values of CD_{step} are applied to CDCoWu under the constraint of delay for UCWu expressed in eq. (3.25), to make the impact of post wake-up congestion less significant. However, as shown in Fig. 3.18, total energy consumption for $N = 10$ is larger than that for $N = 20$. This is because the constraint on delay described in eq. (3.25) for $N = 10$ is so strict that CD_{step} should be set to relatively large value, which causes unacceptable level of congestion. On the other hand, for larger N, unnecessary wake-up of nodes is suppressed in CDCoWu while all sensor nodes need to be woken up at least once in UCWu. This makes it possible for CDCoWu to achieve much smaller energy consumption than UCWu, while satisfying the delay constraint.

In summary, it can be stated that CDCoWu is superior to IDWu, especially when the number of sensor nodes is large. The data collection delay and total energy consumption of N-CDCoWu and V-CDCoWu have also been compared with those of UCWu by simulations when we employ a practical dataset of temperature sensors, which is Intel Lab Data [70]. With this evaluation, it has been confirmed that N-CDCoWu and V-CDCoWu with $N = 50$ and $k = 5$ achieve smaller total energy consumption than UCWu by about 80% and 50%, respectively, while keeping smaller data collection delay than UCWu.

3.4.4 Impact of k and N

Figs. 3.16 and 3.18 show that there are some cases where CDCoWu does not outperform IDWu in terms of both data collection delay and total energy consumption. Specifically, whether CDCoWu will outperform IDWu depends on the ratio of k to N. This subsection investigates the phenomenon when k and N are varied, focusing on the superiority of CDCoWu to UCWu which has more comparable performance than BCWu.

This thesis provides insights on the percentage of top-data for a given N that motivates use of CDCoWu instead of IDWu. To this end, this work investigates k-N ratio ($\frac{k}{N} \in \mathbb{KNR} = \{0.1, 0.2, \ldots, 1\}$), where CDCoWu outperforms UCWu, i.e.,

$$\mathbb{U}_k(N) = \{\frac{k}{N} \mid \epsilon \leq \epsilon^{opt}_{UCWu}(N)\} , \tag{3.27}$$

where ϵ is minimum total energy consumption of CDCoWu under the optimized set of parameters satisfying the delay constraints of UCWu, and $\epsilon^{opt}_{UCWu}(N)$ is total energy consumption against the number of nodes under the optimized parameter p for UCWu. Below, maximum k-N ratio, i.e., max $\mathbb{U}_k(N)$ for different number of nodes, is investigated.

Fig. 3.19 shows maximum k-N ratio for UCWu against the number of nodes, where $e_c = 0$ and uniform distribution, exponential distribution with $\alpha = 0.1$, and normal distribution are employed for observed values. Here, the mean of normal distribution is set as $\mu = \frac{V_{max}+V_{min}}{2} = 25$. On the other hand, the standard deviation of normal distribution is set as $\sigma = 2.85$, which is the value observed in the Intel Lab Data set of temperature sensors [71]. The declining tendency of maximum k-N ratio for V-CDCoWu with exponential distribution in Fig. 3.19 is mainly due to the constraint on total energy consumption. As discussed in Sec. 3.4.2, total energy consumption of CDCoWu increases with α due to congestion. In the case of V-CDCoWu, the sink needs to collect data from more nodes in order to complete top-k value-set query. This decreases the maximum k-N ratio of V-CDCoWu for larger number of nodes, as shown in Fig. 3.19. Furthermore, from Fig. 3.19, one can see that maximum k-N ratio for normal distribution is the smallest among different distributions. The normal distribution has the observed values concentrated on smaller range than the other distributions. This makes it difficult for CDCoWu to control a trade-off between delay and energy consumption. That is, larger CD_{step} is required to quickly find the range with more data, but it increases the number of activated nodes for each wake-up trial, resulting in higher level of congestion. This becomes apparent especially for smaller number of nodes, where the delay constraint of UCWu is more severe, and CDCoWu cannot outperform UCWu for the number of nodes of 10. This problem can be addressed by adapting the size of CD_{step} in accordance with the distribution of observed data, which has been studied in [72]. However, from Fig. 3.19, it is evident that CDCoWu outperforms UCWu in terms of both data collection delay and total energy consumption for any distribution when the number of nodes is equal to or more than 20, and its maximum k-N ratio ranges from 0.1 to 0.5.

3.5 Summary

In this chapter, we introduced CoWu to solve the problem of IDWu, and proposed CDCoWu for both top-k node-set and value-set query, which realizes both lower data collection delay and high energy efficiency of nodes. Then, to evaluate these proposed schemes, we have conducted a theoretical analysis adopting p-persistent CSMA, which is a well-known MAC protocol that can approximate the operation of practical MAC protocol. With numerical evaluations, we have validated the

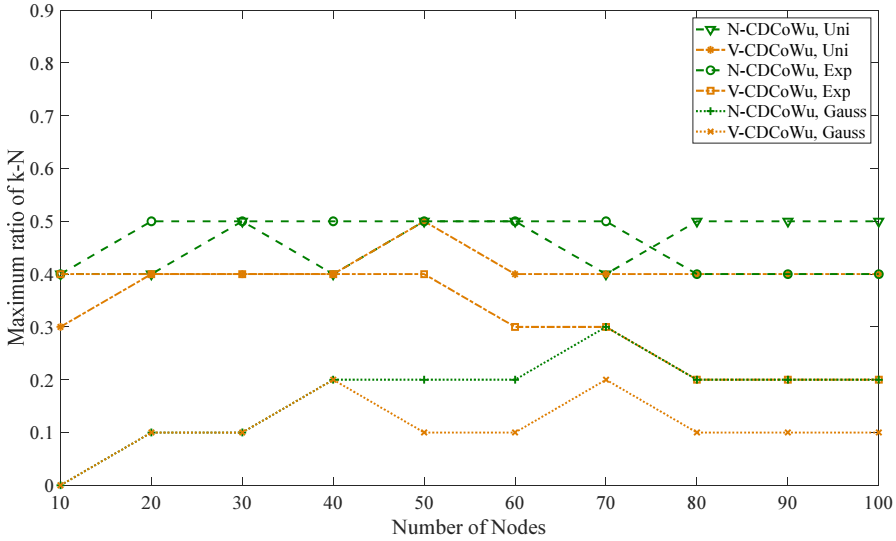


Figure 3.19: Maximum k-N ratio of CDCoWu to UCWu in terms of data collection delay and total energy consumption.

derived equations and the approach of approximate analysis applied in CDCoWu. We have also confirmed the efficiency of CDCoWu against UCWu and BCWu in terms of data collection delay and total energy consumption. In the next chapter, we will focus on a repeated top-k query and the characteristics of temporal correlation of observed data with the introduction of a wake-up control for approximate top-k ranking query.

Chapter 4

Wake-up Control for Repeated Top-k Query Exploiting Temporal Correlation of Sensing Data

This chapter focuses on repeated top-k query, in which the observed values of sensor have temporal correlation. We introduce a new wake-up control realizing both high ranking accuracy and high energy-efficiency [5][5].

4.1 Motivation to Consider Approximate Top-k Query

In the previous chapter, we proposed a wake-up control focusing on *exact top-k query*, in which the sink is interested in grasping the exact top-k set at the query instance. On the other hand, this chapter focuses on an *approximate top-k query*, in which a user aims to grasp the approximate results of the top-k set, e.g., grasping 80% of the information on nodes including the true top-k set. In many practical applications of top-k query in WSNs, 100 % accuracy, i.e., to collect exact top-k set is not necessary [21], i.e., it is sufficient to collect the approximate top-k set. The approximate top-k query is well studied as a query processing for distributed processing [73].

In WSNs, the approximate top-k query is studied in a variety of aspects, which can be categorized into three aspects, namely, (1) observation value [24], [74], (2) ranking [24], and (3) elements of the top-k set [24], [21], [75].

For the approximate top-k query, its accuracy can be sacrificed to improve the energy efficiency. For instance, a sensor node can suppress its data transmission

[5]The content of this chapter was published in [5].

when its observation value does not vary from the previously reported data. While this can reduce the energy consumption of sensor nodes, the accuracy of the top-k set is deteriorated. Thus, there is a trade-off between the energy consumption of sensor nodes and the accuracy of the top-k set. Therefore, for the approximate top-k query, it is important to introduce a wake-up control, which can control this trade-off appropriately. However, no existing work considers wake-up control for the approximate top-k query, which is the main scope of this chapter.

4.2 System Model

We consider a scenario, where N sensor nodes with wake-up receivers are deployed over a sensing field, and a sink conducts a periodical top-k query with wake-up control. Here, we define a single cycle of top-k query as *round*. Each sensor node performs periodical measurements, storing the last observation for reporting to the sink. We assume that each sensor observes one attribute, such as temperature, humidity and so on, having temporal correlations [25], [70] with a range of $[V_{min}, V_{max}]$. The temporal correlation of observed data is modeled by adopting the first order Auto Regression (AR) model [76]. Specifically, we generate an observation value of each node for a round based on Gaussian distribution whose mean is set to its observation in the previous round with variance of σ^2_{AR}. Here, the initial observation of sensor nodes is assumed to follow Gaussian distribution whose mean is set to $\frac{V_{max}+V_{min}}{2}$ with variance of σ^2_{DOV}[6]. Note that, for simplicity, we generate sensing data of each sensor independently without considering their spatial relationship and sensing ranges. In each round, the sink transmits wake-up signals aiming for waking up sensor nodes in a sleep state. After detecting the wake-up request transmitted by the sink, each sensor node activates its main radio I/F and transmits data to the sink following CSMA/CA protocol defined in IEEE 802.15.4 [61]. A sensor node that succeeds in data transmission, i.e., if it receives ACK from the sink, transits to a sleep state, and it is controlled not to wake up until the next round. Furthermore, all nodes including the sink are assumed to be located within communication/wake-up/carrier-sensing range of each other. For simplicity, we assume a collision channel, in which packets can be lost only due to collisions. We also assume that each sensor node is within the communication range of the sink and they can successfully transmit data to the sink as long as there is no collision. Thus, in this work, we do not assume any distribution of sensor deployment.

[6]This assumption is reasonable since sensing data observed in some practical examples of WSNs have been reported to follow Gaussian distribution, for example, a histogram of temperature reading of sensor nodes in the Intel Lab data set follows Gaussian distribution [77].

4.3 Definition of Accuracy of Top-k Set

As the accuracy of top-k set, we consider three elements, namely, (1) observation value, (2) ranking, and (3) elements of the top-k set. These are crucial for different applications, e.g., (1) for anomaly detection [25], (2) for network management considering residual energy of sensor nodes [24], and (3) for dispatching drones toward a wildfire area [32].

4.3.1 Root Mean Square Error (RMSE)

The accuracy of the observation value is evaluated by using RMSE [78]. Here, RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{|S|}\sum_{i=1}^{|S|}(y_i - \hat{y}_i)^2}, \tag{4.1}$$

where $\hat{y}_i (i = 1, 2, \ldots, |S|)$ are the observation values in the top-k set collected for each round while their true values are expressed as y_i, and $|S|$ is the number of nodes estimated to be in the top-k set.

4.3.2 Generalized Footrule Distance (GF-distance)

In general, the distance between the set considering the order (i.e., ranking) can be calculated as the Footrule distance (F-distance) [79]. F-distance is defined by the sum of the absolute values of the differences between the ranking vectors of different sets. Here, the ranking vector is defined as a vector that shows the ranking of each element within a set. For example, if the order of a set {A, B, C} is [B C A], then its ranking vector becomes [3 1 2]. Formally, F-distance $F(\mathbb{A}, \mathbb{B})$ between sets $\mathbb{A}$ and $\mathbb{B}$ is described as

$$F(\mathbb{A}, \mathbb{B}) = \sum_{i=1}^{Z} |\pi_A(i) - \pi_B(i)|, \tag{4.2}$$

where $\pi_A(i)$ and $\pi_B(i)$ are elements of the ranking vector of the set $\mathbb{A}$ and $\mathbb{B}$, respectively and Z is the size of the set.

When estimating a top-k set, False Positive (FP) and False Negative (FN) can occur in some cases. Here, FP refers to an event that erroneously estimates a node as the top-k set, which is not actually an element in the top-k set. On the other hand, FN refers to an event that misses a node in the estimated top-k set, which is actually an element in the top-k set. In the literature [80], the authors generalize the ranking accuracy for simultaneously considering the accuracy of

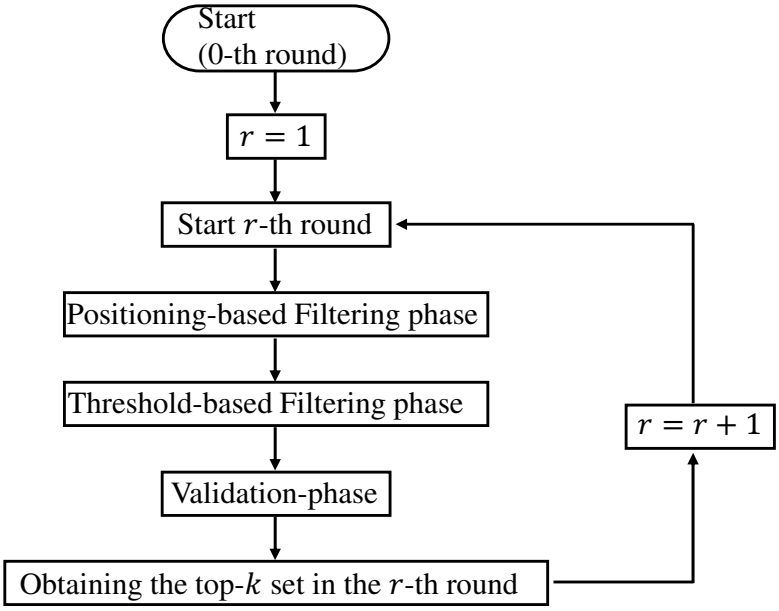


Figure 4.1: A block diagram of our proposed scheme.

elements in the top-k set, i.e., FP and FN, which we call Generalized Spearman's Footrule distance (GF-distance). Here, let $\mathbb{T}$ and $\mathbb{S}$ denote a true top-k set and its estimated top-k set, respectively. Then, according to [80], GF-distance of $\mathbb{T}$ and $\mathbb{S}$ can be calculated by applying the same processes as eq. (4.2) against the ranking vector of union elements $\mathbb{U} = \mathbb{T} \cup \mathbb{S}$. As for the elements of FP and FN for $\mathbb{T}$ and $\mathbb{S}$, the value of $k + 1$ is given as a ranking penalty [80]. From this setting, GF-distance between $\mathbb{T}$ and $\mathbb{S}$ are expressed as

$$\gamma_D = \frac{1}{|C|} \sum_{i \in \mathbb{U}} |\pi_T(i) - \pi_S(i)|. \tag{4.3}$$

Here, $|C|$ is a constant to normalize γ_D, whose value is $k(k + 1)$. By applying eq. (4.3), the ranking and element accuracy between the top-k set estimated by the sink and true top-k set can be calculated. The unique feature of the above metric is that the value γ_D increases not only when the ranking error between the set is large, but also when the FP and/or FN occur, by adding the penalty of $k + 1$.

4.4 Proposed Scheme

In this section, focusing on the approximate top-k query, where high accuracy for both ranking and elements of top-k set are required, we propose a wake-up control, named Order and Temporal Correlation-based wake-up (OTC-Wu). OTC-Wu

aims to grasp the information on the variations of top-k ranking from the previous round of query while only activating a small subset of sensor nodes. Here, we define nodes included in the top-k set in the previous round as Temporal Monitoring node (TM-node), and others as Filtering node (F-node), following the terminology in [25], [81].

Fig. 4.1 shows a block diagram of the operation of our proposed scheme. There are three phases, repeated for each round. Here, in an initial round (i.e, 0-th round) of OTC-Wu, the sink is assumed to employ CDCoWu [4] in order to collect an exact top-k set. For the following rounds, each sensor node is controlled to wake up when its observation value is likely to cause the ranking validation of the top-k set. To this end, each sensor node needs to know the ranking for itself and others in the previous round. Thus, in OTC-Wu, the sink first transmits the information of observation values of top-k set in the previous round to each wake-up receiver of TM-node. The wake-up receiver of each TM-node extracts the ranking relationship among nodes and sets an appropriate transmission strategy. Then, the wake-up receiver activates the main radio I/F and transmits data to the sink only when it decides that the current observation value can cause the ranking variation of the top-k set (Positioning-based Filtering phase described in Sec. 4.4.2). Next, in order to reduce the element errors of the top-k set, the sink checks the existence of F-nodes that are likely to newly join the top-k set by using CoWu (Threshold-based Filtering phase described in Sec. 4.4.3). Finally, in order to identify the current top-k set with high ranking accuracy, the sink transmits validation queries to confirm its ranking and elements (Validation phase described in Sec. 4.4.4).

4.4.1 Mode Control Signal

In order to realize the above-mentioned flow, we prepare three different modes of operations for each wake-up receiver: UCWu, CoWu, and Positioning Mode (PoMo). The operating modes are switched by three types of mode control signals, namely, UCWu, CoWu, and PoMo Signal (termed as UCMoSg, CoMoSg, and PoMoSg, respectively). Different length of the frames are assigned to these signals which are denoted as T_{UCMoSg}, T_{CoMoSg}, and T_{PoMoSg}, respectively. The wake-up receiver that detects UCMoSg (CoMoSg) regards the following received signals as UCWu (CoWu) signal. When the sink attempts to collect data employing UCWu (CoWu), it first transmits UCMoSg (CoMoSg), and then sends the UCWu (CoWu) signal corresponding to the target node-ID (target content). After detecting UCMoSg (CoMoSg), the wake-up receiver sets its own wake-up frame length T_{wu}, and is controlled to wake up only when it satisfies the condition on $T_{rx} = T_{wu}$ ($T_{wu} \leq T_{rx}$). On the other hand, as we will describe in Sec. 4.4.2, PoMoSg is used for the sink to share the information on the top-k set in the previous round with

wake-up receivers of TM-nodes.

4.4.2 Positioning-based Filtering Phase (PF-phase)

In PF-phase, the sink attempts to grasp the information on the current readings of TM-nodes by only activating nodes whose current readings are likely to cause the ranking variation from the previous top-k set.

Transmission suppression mechanism for TM-nodes

In this phase, the sink first notifies the wake-up receiver of each TM-node of the information on the previous top-k set. Since the wake-up receiver only has the capability to detect the frame length, this notification is conducted by exploiting the frame length corresponding to the observation values of TM-nodes. Based on the detected frame length, the wake-up receiver of TM-node obtains the ranking of its latest value $\hat{v}_i$ within the top-k set and the information on observation values of the other ranked nodes. The specific mechanism for obtaining the information on ranking at the wake-up receiver will be discussed later in this section. Each TM-node exploits this obtained information and sets a transmission suppression range, which works as a kind of arithmetic filter, i.e., if the current reading of TM-node falls within the transmission suppression range, it suppresses the wake-up and data transmission. Here, the transmission suppression range is designed so that each node transmits data only when its current observation value is likely to cause the variation of the ranking from the previous top-k set. This can be calculated by the relationship with the adjacent ranking nodes, i.e., $(i-1)$-th and $(i+1)$-th values $\hat{v}_{i-1}^{Notify}$ and $\hat{v}_{i+1}^{Notify}$. Fig. 4.2 shows a specific example of the collected top-k set in the previous round and an applied transmission suppression range for the TM-node i. Here, in Fig. 4.2, the latest reported value of TM-node i is $\hat{v}_i = 30$, and its adjacent ranking values are $\hat{v}_{i-1}^{Notify} = 38$ and $\hat{v}_{i+1}^{Notify} = 28$. The difference between the i-th and $(i-1)$-th values is $\Delta_{i \to i-1} = 8$ and that of the i-th and $(i+1)$-th value is $\Delta_{i \to i+1} = 2$. From this observation, we can say that, relatively, the change of observation value of TM-node i toward the direction for the value of $(i+1)$-th ($(i-1)$-th) value is more (less) likely to cause the ranking variation between them. Thus, it is desirable for TM-node i to set its transmission suppression range, say [29, 34] according to the relationship with the adjacent node as shown in Fig. 4.2, which is an average value between itself and adjacent node.

Through these considerations, we formulate the transmission suppression range of TM-nodes by using the $(i-1)$-th and $(i+1)$-th observation values. Here, the i-th transmission suppression range $[l_i, u_i]$ can be expressed as follows:

$$u_i = \hat{v}_i + \beta(\hat{v}_{i-1}^{Notify} - \hat{v}_i). \tag{4.4}$$

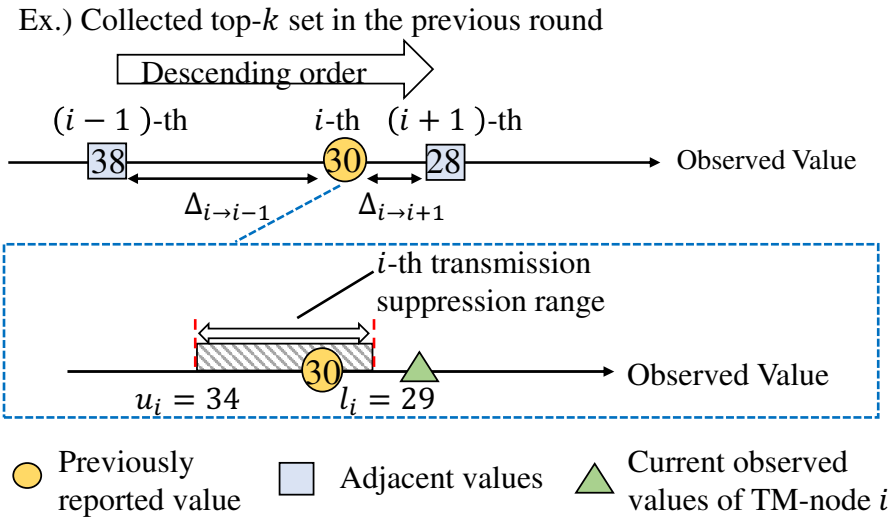


Figure 4.2: An example to set a transmission suppression range based on the information on the ranking ($\beta = 0.5$).

$$l_i = \hat{v}_i - \beta(\hat{v}_i - \hat{v}_{i+1}^{Notify}). \tag{4.5}$$

For special cases, we assume $u_1 = \infty$ and $l_k = V'_{min}$, where V'_{min} corresponds to the k-th observation value of the previous top-k set. Here, β is a parameter that controls the accuracy of the collected top-k set. Note that, when $\beta = 0$, OTC-Wu is reduced to the scheme proposed in [81], in which the authors attempt to grasp the exact top-k set exploiting temporal correlation. Hereafter, this thesis calls the scheme in [81] as EXACT–Wu. If l_i becomes lower than V'_{min}, l_i is rounded to V'_{min}. If v_i is located within the transmission suppression range of a TM-node i, i.e., if it satisfies the condition on $(l_i \leq v_i \leq u_i)$, it suppresses to wake up. On the other hand, the TM-node that does not satisfy the above condition wakes up and transmits its current observation value v_i to the sink. Fig. 4.2 is an example of $\beta = 0.5$ in eqs. (4.4) and (4.5), where the current observed value of TM-node i, depicted by the triangular mark, violates its transmission suppression range. Therefore, TM-node i is required to wake up and transmit data to the sink in an adequate timing. The timing for data transmission is determined based on the received frame length, which will be described below.

Sharing the previous top-k ranking information with the wake-up receiver

As described in Sec. 4.4.2, the sink needs to convey the information on previous top-k set by using the length of frame toward the wake-up receiver. To this end, the mapping between the length of frame and observed values is necessary, which

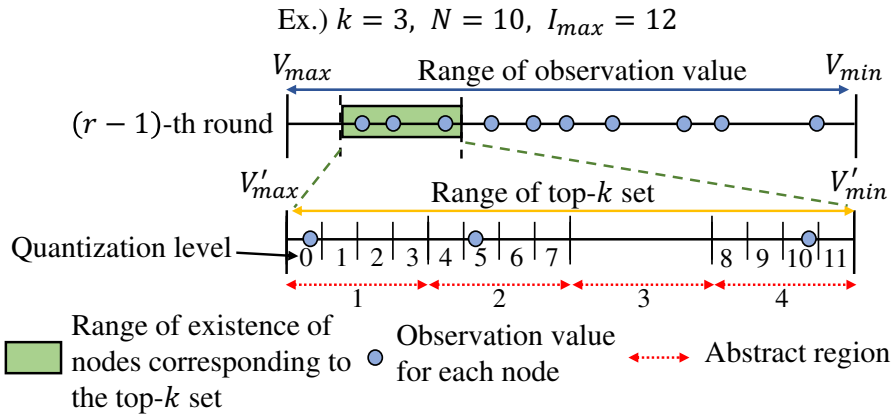


Figure 4.3: An example of dynamic quantization and mapping in accordance with the distribution of top-k set.

is a similar operation to CoWu. Here, we need to newly design a mapping rule between the frame length and observed value for ranking notifications, which aims at alleviating quantization error.

In order to utilize the limited number of quantization levels I_{max} for the mapping between the observations and the frame length effectively, the quantization is conducted only in the range for the ranking notification (i.e., where the observation values of TM-nodes exist). Here, the restriction for the value of I_{max} comes from constraints on the range and resolution of wake-up frame length, specified by considering the standard IEEE 802.15.4g [4]. Fig. 4.3 shows a specific example of the r-th round, where $N = 10$ and $k = 3$. The range of $[V_{min}, V_{max}]$ depicted in the blue arrow shows the range of value that each sensor can observe and the window of green square shows the range corresponding to the top-k set in the $(r-1)$-th round. The specific process is shown below.

- STEP 1: The sink derives the range of $[V'_{min}, V'_{max}]$, to which the first and k-th ranked data belong based on the results of the previous round of query. Next, based on the mapping rule explained in Sec. 3.2.1, the interval of V'_{min} and V'_{max}, denoted q_{v_k} and q_{v_1}, respectively, are derived. Then, the sink first transmits PoMoSg, followed by the transmissions of frame with length corresponding to q_{v_1} and q_{v_k}.

- STEP 2: The sink divides the range of $[V'_{min}, V'_{max}]$ by w, whose divided range is defined as abstract region in this thesis, and allocates unique number j from V'_{max} ($j \in 1, 2, \ldots, w$), which corresponds to the area marked by the dotted arrow in Fig. 4.3. Here, the event that some nodes exist in an abstract region j is expressed as $b_j = 1$, and otherwise $b_j = 0$, which respectively

corresponds to the abstract region $\{1, 2, 4\}$, and $\{3\}$ in Fig. 4.3. Then, the sink generates a binary number $s_2 = b_1 b_2 \ldots b_w$, which is converted into the decimal number s_{10}. Finally, the sink transmits a frame whose length corresponds to $T_{min} + T_{step} * s_{10}$ as a rule for quantization. For a special case, when the sink does not apply this quantization, s_{10} is set to be 0. Note that, due to the constraint of applicable quantization level, w is required to be chosen carefully. According to [4], we can apply about 960 different frame length, which means that the maximum value of s_{10} $(= 2^w)$ should be below 960. Thus, here, we apply $w = 9$.

- STEP 3: The sink allocates quantization levels to the only abstract regions where nodes exist, i.e., to the region j satisfying $b_j = 1$. In the example of Fig. 4.3, the sink allocates total quantization levels $I_{max} = 12$ only in the abstract regions 1, 2, and 4.

Based on the mapping set in STEP 3, the sink transmits a series of frames with their length corresponding to observation values in the previous top-k set in decreasing order toward wake-up receivers.

Each wake-up receiver, more precisely, the MCU attached to the wake-up receiver extracts the information of V'_{min}, V'_{max} and s_2 from the received frame length, and applies the notified mapping rule. Based on this rule and detected length of frame, each wake-up receiver calculates the information on the ranking in the previous top-k set, and records it for future utilization. By detecting the length of frame corresponding to the last reported value $\hat{v}_i$, the wake-up receiver grasps its ranking in the previous round of query as i. We set the wake-up timing for TM-node i, whose current reading violates its transmission suppression range, to the time when it receives the frame length corresponding to the next ranking value, i.e., $(i + 1)$-th value. Note that there can be a case where a frame length, say L_x, is transmitted more than once when multiple observations belong to the same quantization level. In this case, after receiving the frame L_{x+1}, which is longer than L_x, all nodes whose last reported values correspond to L_x are controlled to wake up and transmit data to the sink. If the sink fails to receive data from this subset of nodes with the same quantization level, it transmits a wake-up request again by using the frame length T_{min}. Here, TM-nodes are controlled to wake up and transmit data if they detect a frame with length T_{min} during the ranking notification.

4.4.3 Threshold-based Filtering Phase (TF-phase)

In this phase, leveraging CoWu, the sink attempts to grasp the information on F-nodes that are likely to newly move into the current top-k set. The lower bound

of the current top-k set, τ', is estimated by the sink from the results of the previous top-k set. Since F-nodes, which observe values within the transmission suppression range from the first to k-th ranked values, are likely to become the elements of the current top-k set, it is desirable for the sink to set the lower bound $\tau' = l_k$. The sink first transmits CoMoSg in order to notify the employment of CoWu to the wake-up receiver, followed by the transmission of CoWu signal with the threshold τ'. By this wake-up trial, only F-nodes, which observe values equal to or more than τ', wake up and transmit data to the sink.

4.4.4 Validation Phase

Here, we define the F-nodes that observe values equal to or more than τ' and transmit data in TF-phase as Join nodes and the TM-nodes that observe values less than τ' as Leave nodes. The number of Join (Leave) nodes is denoted as $|J|$ ($|L|$).

Ordering Validation Phase (OV-phase)

In PF-phase and TF-phase, each sensor node reports its current observation value to the sink only when it violates the transmission suppression range or exceeds a threshold τ'. If this reported value is within the transmission suppression range of the other nodes, the sink needs to check the current observed value of these nodes to clarify the ranking among all reported values. Fig. 4.4 shows an example of collected data at the beginning of the OV-phase, where $k = 3$ and $\beta = 0.5$. This example shows that sensor nodes A, B, and C, are included in the top-k set in the $(r-1)$-th round, and in the r-th round, the observation value of sensor node B moves into the range of $[l_1, u_1]$. In this case, while the sink exactly knows the information on sensor node B in the range of $[l_1, u_1]$, it does not know the current value of sensor node A. Therefore, the sink tries to clarify the ranking relationship between sensor node A and B by collecting data from sensor node A with UCWu. Here, let us denote the number of collected data whose values are more than l_i as c_i. Then, for $\beta \leq 0.5$, in order to suppress wasteful wake-up, the sink only transmits validation query for the nodes whose last reported value is larger than l_m. Here, the value of m is the minimum value satisfying $c_m \geq k$, and if such value does not exist, m is set to be k. On the other hand, for $\beta > 0.5$, m is always set to be k to prevent the increase of GF-distance.

Final Validation phase (FV-phase)

This phase is conducted only when $|J| < |L|$. Through the above phases, the sink has grasped the ranking information of nodes that observe values equal to or

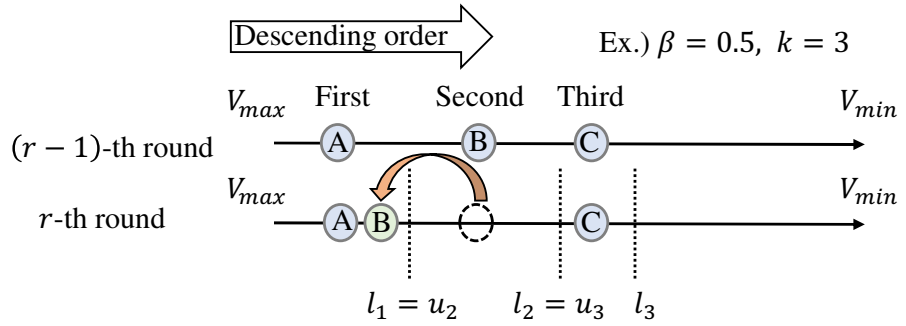


Figure 4.4: An example of collected data at the beginning of the OV-phase.

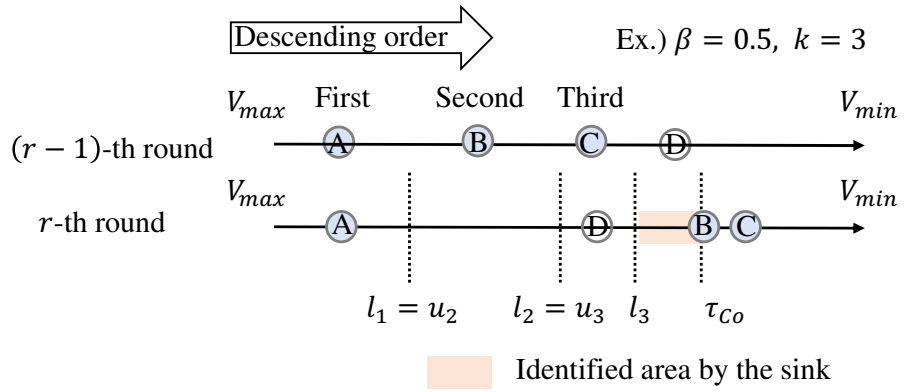


Figure 4.5: An example of collected data at the beginning of the FV-phase.

larger than l_k. On the other hand, the ranking of nodes that observe values less than l_k have not been determined. Therefore, the sink needs to transmit queries for identifying the current top-k set. A specific example with $k = 3$ is shown in Fig. 4.5. In Fig. 4.5, the Join node corresponds to sensor node D ($|J| = 1$), and the Leave nodes correspond to the sensor nodes B and C ($|L| = 2$). In this setting, the sensor node A and D are identified as the top-k set in the r-th round, while the sensor node B, which owns the $(|L| - |J|)$-th value among the nodes, has not been determined to be top-k set. This is because the sink does not have any knowledge on F-nodes that observe values less than τ' (l_3 in the example of Fig. 4.5) as this range has not been explored with CoWu. In this case, the sink transmits a CoWu signal whose threshold τ_{Co} corresponds to the value of sensor node B. By this wake-up trial, the sensor nodes that observe the values in the range of $[\tau_{Co}, l_k]$ wake up and transmit data to the sink. In the example of Fig. 4.5, since none of the F-nodes observe the values in the range of $[\tau_{Co}, l_3]$, the sensor node B is identified as an element of the current top-k set.

4.5 Numerical Results

In this section, we show numerical results of average energy consumption and accuracy of top-k set for different wake-up control. The parameters employed in this evaluation are shown in Table 4.1. Here, average energy consumption [J/round] and average GF-distance are respectively defined as the averaged amount of total energy consumed by all nodes[7] and averaged γ_D expressed in eq. (4.3) in each round. In this evaluation, we neglect the power consumption of the wake-up receiver since it is always switched on and its consumed energy is the same for all wake-up control. The simulation model follows the system model described in Sec. 5.1. The simulations are conducted by a custom-made simulator created with Matlab. Furthermore, as shown in Table 4.1, we set the minimum length employed for wake-up signaling to 10.80 ms ($T_{PoMoSg} = 10.80$ ms), and prepare the interval $T_{step} = 0.16$ ms for different signals i.e., different mode control signal and/or wake-up signal. These values are chosen according to the standard of IEEE 802.15.4g operating over 920 MHz frequency band [3]. The data size from 28 to 135 bytes are commonly employed for data transmissions over the considered standard of IEEE 802.15.4g. In order to avoid interference between wake-up signals and data transmissions, we use the length equal to or more than 135 bytes (which corresponds to 10.80 ms in time) for wake-up signal and mode control signal[8]. In addition, in order to avoid the errors in frame length detection, a sufficiently large interval between different wake-up signals are required, which motivates us to employ large T_{step}. Here, T_{step} of 2 bytes is employed to offer sufficient reliability of frame length detection [4].

As reference schemes for comparison, we employ CDCoWu [4] and EXACT-Wu [81]. In CDCoWu, the sink gradually reduces the threshold of CoWu by CD_{step} from $V_{max} - CD_{step}$ until the number of collected values reaches k. This process of wake-up and data collections is conducted for every round of top-k query, which means that CDCoWu does not exploit the temporal correlation of observation values. Here, CD_{step} is set to be a minimum possible value in order to suppress the effect of congestion [4]. On the other hand, in EXACT-Wu, the sink exploits the temporal correlation of observation values and attempts to collect data from the nodes whose observation values are varied from the previous round [81].

[7]For calculating average energy consumption, only overall energy consumption of each node, which includes energy consumed at circuit electronics and power amplification, is required. Therefore, we do not separately specify power consumption for circuit electronics and power amplification, but only specify the overall power consumption of each node in Table 4.1.

[8]We have confirmed that the variation of values of mode control signals does not give significant impact on the obtained simulation results.

Table 4.1: Parameters employed for computer simulations.

Parameters	Values
Data transmission rate	100 kbps [3]
Access Control	CSMA/CA [3]
Maximum number of back-offs	4 [61]
macMinBE	3 [61]
macMaxBE	5 [61]
Maximum number of retransmission	3 [61]
Size of transmitted packet	10 bytes
T_{PoMoSg}	10.80 ms
T_{UCMoSg}	10.96 ms
T_{CoMoSg}	11.12 ms
Minimum frame length for wake-up signals T_{min}	11.28 ms
Step to change wake-up frame length T_{step}	0.16 msec [3]
wake-up frame length for UCWu	$T_{min} + j \times T_{step}$ (j: ID number)
$[V_{min}, V_{max}]$	$[0, 50]$
CD_{step}	0.1953
k	10
Number of Nodes	100
Power consumption in Transmit state	55 mW [4]
Power consumption in Receive state	50 mW [4]
Number of simulations	1000
Number of rounds	100

4.5.1 Trade-off Between Average Energy Consumption and GF-distance through β

Fig. 4.6 shows average energy consumption and GF-distance of OTC-Wu against β in eqs. (4.4) and (4.5), where $\sigma_{AR} = 0.1$ and $\sigma_{DOV} = 2.85$ [71]. From this figure, we can first see that there is a trade-off between these two metrics through β. The reduction of average energy consumption with the increase of β is mainly due to the decreasing energy consumption at PF-phase, which is caused by more suppressions of wake-up and transmissions of TM-nodes. On the other hand, average GF-distance monotonically increases after β exceeds 0.5. This is because the transmissions of TM-nodes are suppressed even with the large variation of observation values, which causes the ranking variation undetected by the sink.

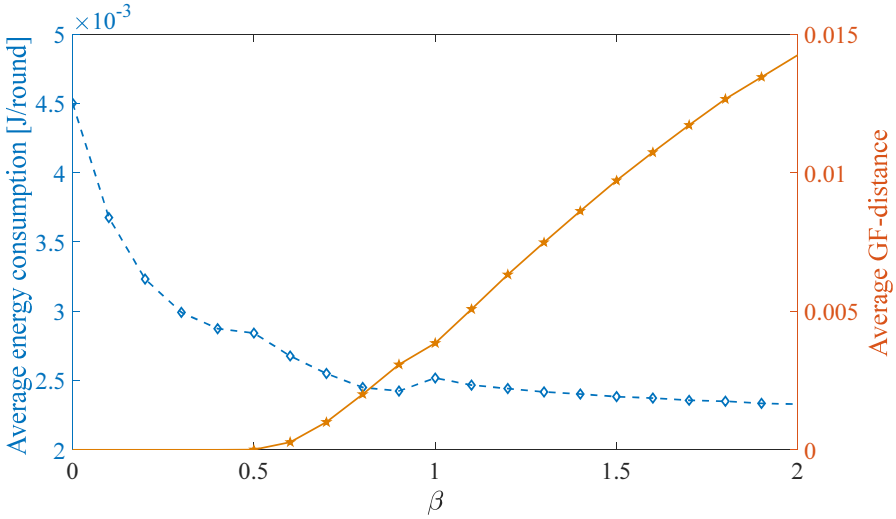


Figure 4.6: Average energy consumption and GF-distance of OTC-Wu against β.

From these results, we can say that OTC-Wu can control the gain of energy saving and the level of accuracy by the parameter β, which is a desired feature of the approximate top-k query. Finally, from Fig. 4.6, we can emphasize that, for $\beta < 0.5$, OTC-Wu achieves large energy saving while keeping the high accuracy of ranking and elements of top-k set.

4.5.2 Comparison of Different Wake-up Control

Fig. 4.7 shows average energy consumption of different wake-up control against the degree of temporal correlation σ_{AR}, where $\beta = 0.5$ and $\sigma_{DOV} = 2.85$. From Fig. 4.7, we can first see that average energy consumption of CDCoWu decreases as σ_{AR} increases. This is because, as a result of variations of observation values over time, the deviation of observation values becomes smaller as round passes, which reduces the number of waking up nodes for each wake-up trial and causes the reduction of congestion level. On the other hand, EXACT-Wu, which exploits the correlation of top-k results between consecutive rounds, can identify the current top-k set by waking up a smaller number of nodes than CDCoWu, leading to smaller energy consumption for higher degree of correlation. However, when σ_{AR} exceeds a certain value, the performance of EXACT-Wu becomes worse than that of CDCoWu. As the value of σ_{AR} increases, the variation of top-k set from the previous round becomes larger, which means many TM-nodes (F-nodes) are likely to leave (join) the top-k set. Thus, EXACT-Wu collects data from the nodes that do not belong to the current top-k set, which leads to higher energy consumption of sensor nodes. Finally, we can see that OTC-Wu achieves smaller energy consumption than any

other scheme when the degree of correlation is high. This is because OTC-Wu only activates a subset of nodes contributing to the variation of the ranking, for which each wake-up receiver sets an appropriate transmission suppression range based on the top-k set in the previous round. When σ_{AR} exceeds 0.45, the average energy consumption of OTC-Wu becomes larger than that of CDCoWu. As the value of σ_{AR} increases, the correlation of observation values between rounds becomes small. In this case, the variation of observation values between the consecutive rounds tends to increase, by which the difference in terms of ranking and element of top-k set between consecutive rounds becomes larger. Our OTC-Wu is designed so that the sink and sensor nodes spend their resource to grasp these difference. When σ_{AR} is large, the sink needs to activate many nodes by wake-up signaling in order to collect the information on difference between consecutive rounds, i.e., the difference of top-k set. Specifically, for larger σ_{AR}, the current readings of TM-nodes are more likely to violate their transmission suppression range due to the variation of observation value, which requires them to wake up and transmit data to the sink in PF-phase, thereby consuming more energy. Likewise, the number of activated F-nodes, whose current readings exceed the threshold τ' set in TF-phase, increases. These F-nodes need to wake up and transmit data in TF-phase, whose congestion level is assumed to be relatively high. Thus, with larger σ_{AR}, more resource of energy need to be spent for OTC-Wu. On the other hand, in CDCoWu, as mentioned above, the sink can collect the current top-k set only by activating nodes around the top-k set with smaller congestion level even when σ_{AR} increases. That is why the average energy consumption of OTC-Wu becomes larger than that of CDCoWu with a larger σ_{AR}. Note that, as a special case we have investigated the performance of CDCoWu, EXACT-Wu, and OTC-Wu, when we generate the observation value of each node randomly for each round according to the Gaussian distribution whose mean is set to be $\frac{V_{min}+V_{max}}{2}$ with variance σ^2_{DOV}, which corresponds to the case of $\sigma_{AR} = \infty$. With this evaluation, we have obtained the average energy consumption of CDCoWu, ECACT-Wu, and OTC-Wu as 0.00547, 0.875, and 0.869 [J/round], respectively. From this result, we can conclude that our proposed OTC-Wu operates effectively in terms of average energy consumption when σ_{AR} is relatively small, i.e., when there is sufficiently large correlation for the sensing data between the consecutive rounds. The average GF-distance and RMSE of OTC-Wu for the range of σ_{AR} in Fig. 4.7 are 4.35×10^{-6} and 0.22, respectively, while those of CDCoWu and EXACT-Wu are both 0. This means that OTC-Wu realizes energy efficient top-k data collection while realizing high accuracy of approximate top-k set.

Finally, Fig. 4.8 shows energy reduction ratio $(1-\frac{E_{OTC}}{E_{EXACT}})$ against the deviation of initial observation value, σ_{DOV}, where E_{OTC} and E_{EXACT} denote average energy consumption of OTC-Wu and EXACT-Wu, respectively. Here, we set $\sigma_{AR} = 0.1$

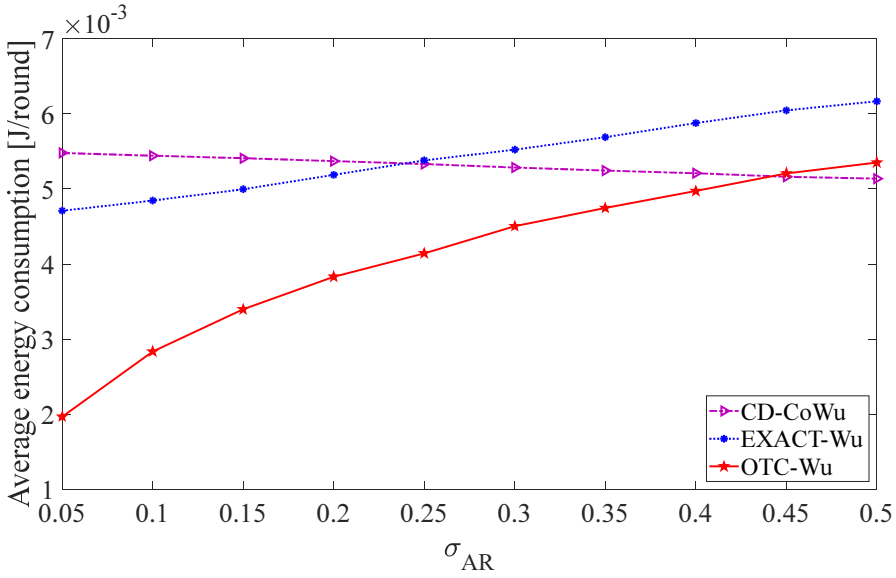


Figure 4.7: Average energy consumption of each wake-up control against the degree of temporal correlation.

and $\beta = 0.5$. Fig. 4.8 shows that the energy reduction ratio becomes larger as σ_{DOV} increases. In OTC-Wu, the transmission suppression range installed to each node becomes larger when the difference of observation values in the top-k set enlarges for larger σ_{DOV}. In this case, OTC-Wu can suppress many redundant wake-ups of nodes that do not contribute to the ranking variation of the top-k set, which achieves more energy-efficient operations than EXACT-Wu.

4.6 Feasibility Study of OTC-Wu with Experiments

In this section, we show the results of the experiment using our prototype of sensor node including wake-up receiver and discuss the practicality of OTC-Wu. The experiments using the prototypes of sensor node including wake-up receiver are only conducted in this chapter. As we will discuss below, the evaluation of the practicality of OTC-Wu is important because it requires much more complex operations at low-power wake-up receivers than the other schemes proposed in this thesis. Although we have confirmed the effectiveness of OTC-Wu with computer simulations, the practical complexity of applying OTC-Wu is unknown.

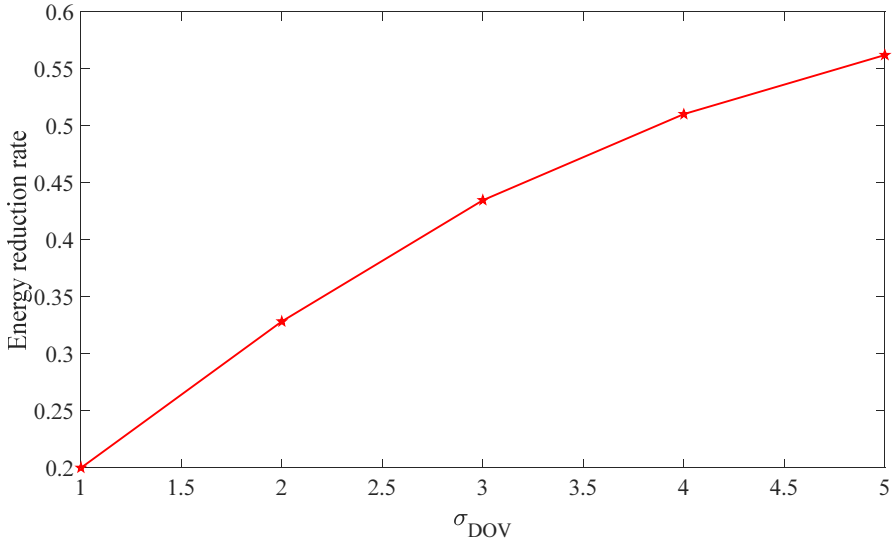


Figure 4.8: Energy reduction ratio against σ_{DOV}.

4.6.1 Experimental Model

In order to investigate the practicality of our proposed scheme, we conduct an experiment using our prototype of sensor node including a wake-up receiver. Here, MCU used for our wake-up receiver is an off-the-shelf component of Renesas RL78/G12 [82]. In [3], the implementation of IDWu has been conducted for our prototype of sensor node with wake-up receiver, and its practicality has been confirmed. The wake-up receiver employing IDWu only has a single operation, in which MCU checks whether the received frame length matches with the predetermined length corresponding to its ID. On the other hand, our proposed OTC-Wu requires the wake-up receiver to perform additional operations, such as the decision on mapping rule, dynamic allocation of a quantization level, extraction of previous top-k set, setting of a transmission suppression range, and so on, besides the simple operation of wake-up process. To the best of our knowledge, there is no work that reports the feasibility of the above operations with wake-up receiver.

The experiment is conducted in a shield room to avoid the effect of interference from other radio equipment. Fig. 4.9 shows the detailed experimental setup. As shown in Fig. 4.9, we prepare a pair of sink and sensor node equipped with a wake-up receiver. The sensor node is assumed to be TM-node whose previous ranking is i, and its current observation value is manually assigned. Here, we assign the previous values of $(i-1)$-th, i-th, and $(i+1)$-th nodes to 23.5, 21.45, and 19.95 respectively, and the current observation value of TM-node is set to be 23.7. In this experiment, the sink transmits six frames, each of which plays different role, over a channel in 920 MHz frequency band following the standard of IEEE 802.15.4g. The

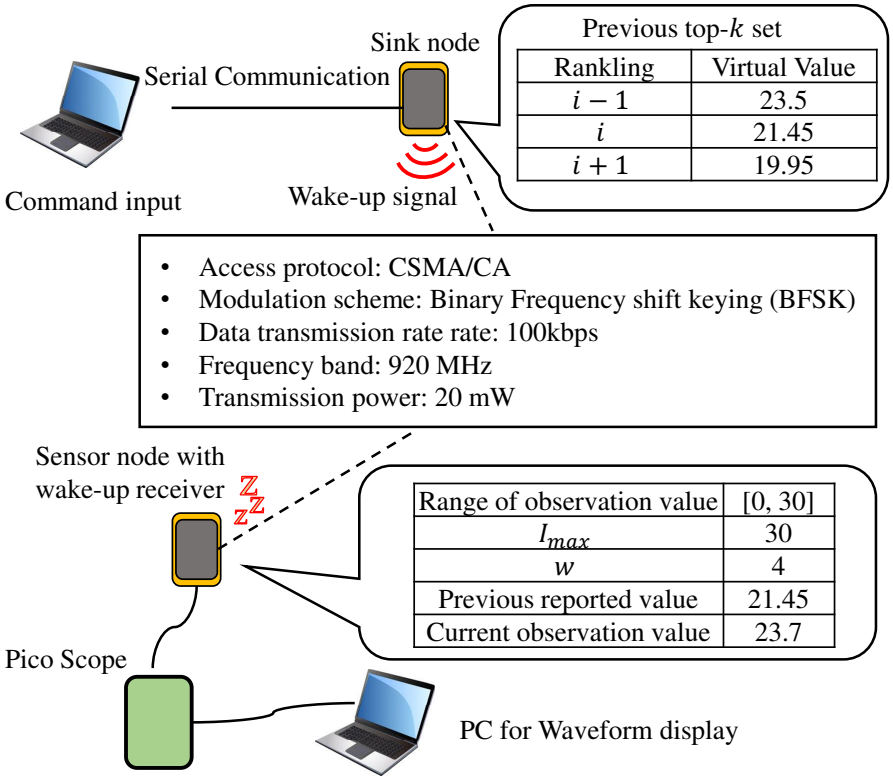


Figure 4.9: Experimental setup with its network feature.

sensor node transmits data following CSMA/CA specified by IEEE 802.15.4. The transmitted data are modulated by Binary Frequency Shift Keying (BFSK) over a channel in 920 MHz frequency band, whose transmission power is 20 mW, with its data transmission rate set to 100 kbps. The sink notifies the range of the top-k set in the previous round with the first two frames. Based on these frames, the wake-up receiver sets its new range for the mapping. Following this, the mapping rule regarding dynamic quantization level (i.e., s_{10}) is transmitted with a single frame. Lastly, three frames for notifying the ranking, each of which corresponds to the observation value of $(i-1)$-th, i-th, and $(i+1)$-th ranks, respectively, are transmitted. The wake-up receiver extracts its ranking and the information on the last reported values of the adjacent ranks from the received frame length. Then, it sets the transmission suppression range, compares it with the current reading, and activates the main radio of sensor node if the current observation value violates its current suppression range. The frame length transmitted by the sink are 5.84, 6.32, 6.64, 5.68, 7.28 and 8.08 ms. Here, T_{min} and T_{step} are set to be 4.88 ms and 0.16 ms, respectively, and w is assumed to be 4. We also assume $[V_{min}, V_{max}] = [0, 30]$ and $I_{max} = 30$.

4.6.2 Experimental Results

Evaluation of practicality of our proposed scheme

We checked the snapshot of output of an oscilloscope observing the envelope of radio signals over the operating channel in 920 MHz band [5]. In this evaluation, we observed six frames transmitted by the sink, followed by the data transmitted by the activated TM-node, whose current reading is set to the outside of transmission suppression range according to the operation of the proposed OTC-Wu. This data transmission of TM-node (whose ranking is i) indicates that the wake-up receiver can successfully extract information on the mapping rule and set a transmission suppression range based on the detected frame length. Therefore, we can conclude that OTC-Wu is practical to be implemented for WSNs employing wake-up receivers.

Impact of T_{step}

Under the practical environments involving unreliable wireless channel, the received frame length at the wake-up receiver can differ from the one transmitted by the sink, which causes the wake-up receiver to falsely wake up its main radio (false positive) or ignore it when it is actually needed (false negative). These failures can be avoided if we set a large interval between different frame length used for the mapping, i.e., by employing large T_{step}. This motivates us to evaluate the impact of T_{step} on the wake-up success ratio, which is defined as the number of successful wake-ups to that of wake-up trials. Here, we employ UCWu and vary the value of T_{step} from 1 byte to 4 bytes. The basic experimental setting is the same as Sec. 4.6.1. The experiment is conducted 10 times for each T_{step}. Fig. 4.10 shows wake-up success ratio against different value of T_{step}. From this figure, we can see that the wake-up success ratio improves as T_{step} becomes larger, which reaches 100% of wake-up success ratio at the step of 2 bytes. This indicates that, if we employ T_{step} equal to or larger than 2 bytes, the sink can reliably convey the wake-up signal toward the wake-up receiver. Note that, though larger T_{step} increases the wake-up success ratio, it sacrifices the granularity of mapping (i.e., it reduces the maximum number of quantization intervals), since there is a restriction for available range of frame length, e.g., specified by the standard IEEE 802.15.4g. Therefore, it is reasonable to set T_{step} to 2 bytes, which validates the parameter used in our simulation.

Comparison in terms of wake-up response time

As mentioned above, in general, the computational power of MCU installed into the wake-up receiver is limited to realize its low power operation. Therefore,

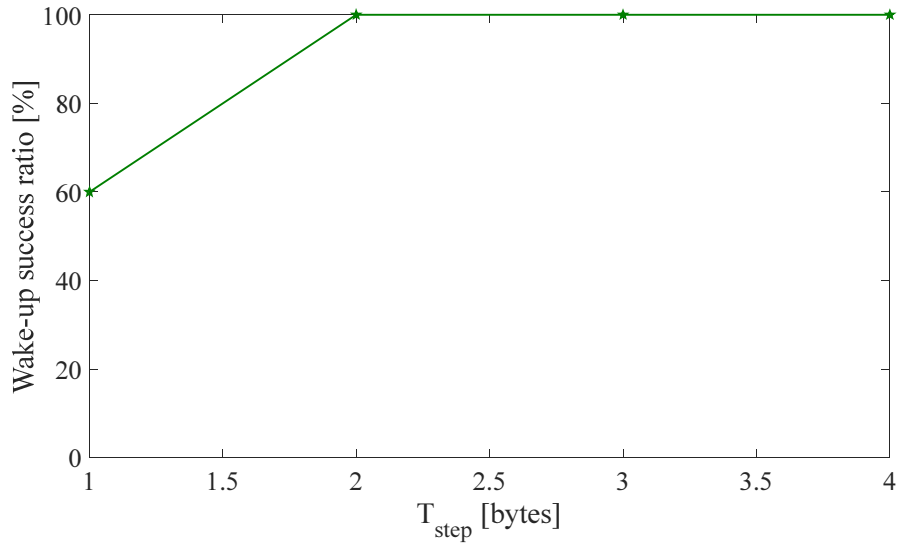


Figure 4.10: wake-up success ratio against different values of T_{step}.

OTC-Wu can cause large wake-up delay due to high computational complexity, which gives us a motivation to compare Wake-up Response Time (WRT) among different wake-up control. Here, WRT is defined as the time from completing the reception of the last frame length at the wake-up receiver to the time when data transmission is started by a main radio after its activation. We only compare the WRT of a TM-node of OTC-Wu with that of EXACT-Wu and the WRT of a node that is supposed to be included in the current top-k set of CDCoWu. This is because the computational complexity of the wake-up receiver required by the operation of the PF-phase is considered to be high. The basic experimental setting for OTC-Wu is the same as Sec. 4.6.1. In EXACT-Wu [81], the sink checks the variation of observation value of TM-node by transmitting UCWu. Therefore, we evaluate WRT when we employ UCWu at the wake-up receiver. On the other hand, CDCoWu [4], as described in Sec. 4.5, does not exploit the temporal correlation of observation values. Thus, we focus on a node, whose observation value is included in the current top-k set, and we evaluate the WRT when the sink employs CoWu. Table 4.2 shows the results on WRT for these wake-up control. As expected, WRT of OTC-Wu is slightly larger than those of EXACT-Wu and CDCoWu. However, the maximum difference is 10 ms, which corresponds to the increase in terms of WRT by about 3%, which is marginal. From these results, we can conclude that our proposed OTC-Wu is practical enough to be implemented in low-complexity receivers.

Table 4.2: Measured WRT for each wake-up control.

wake-up control	WRT ms
CDCoWu	340
EXACT-Wu	340
OTC-Wu	350

4.7 Summary

This chapter proposed OTC-Wu for periodical top-k ranking query, exploiting temporal correlation. With numerical evaluations, we revealed that OTC-Wu could control the trade-off between top-k ranking accuracy and energy efficiency by appropriately selecting the transmission suppression range β. In addition, we confirmed that it could grasp the approximate top-k set high energy efficiency while achieving higher ranking accuracy than CDCoWu and EXACT-Wu, especially when the degree of temporal correlation is high. Furthermore, we confirmed the practicality of our proposed scheme with experiments. In the next chapter, we will apply CoWu in the environment where observed data have spatial correlation and will clarify the wake-up control suited for the identification of multiple emission sources in terms of energy-efficiency and identification accuracy of emission sources.

Chapter 5

Wake-up Control for Identifying Multiple Emission Sources of Spatially Correlated Observations

This chapter focuses on the nature of spatial correlation of observed data in WSNs. Specifically, we focus on the identification of multiple emission sources and clarify the wake-up control that realizes high identification accuracy and high energy efficiency[9].

5.1 System Model

In this study, we consider a scenario where the sink conducts identification of multiple emission sources of sensing observations through wake-up control and collections of sensing data. We denote the number of sensor nodes as N and the size of sensing area as l_{SA} [m] $\times$ l_{SA} [m]. We assume that the number of emission sources, denoted as N_p, is more than 1, which is unknown for the sink. We model the observation value of each position (x, y) as a function $f(x, y)$ by adopting the Gaussian mixture [45][83] as shown in Fig. 5.1. Specifically, the observation at a location (x, y) is expressed as follows:

$$f(x,y) = \sum_{i=1}^{N_p} \exp\{\frac{(x-\mu_x^i)^2 + (y-\mu_y^i)^2}{\sigma_i^2}\} r_i, \tag{5.1}$$

where $(\mu_x^i,\ \mu_y^i)$, σ_i^2 and r_i denote the coordinates of positions, covariance and observation value at the position of the i-th source of sensing data, respectively.

[9]The content of this chapter will appear in [6].

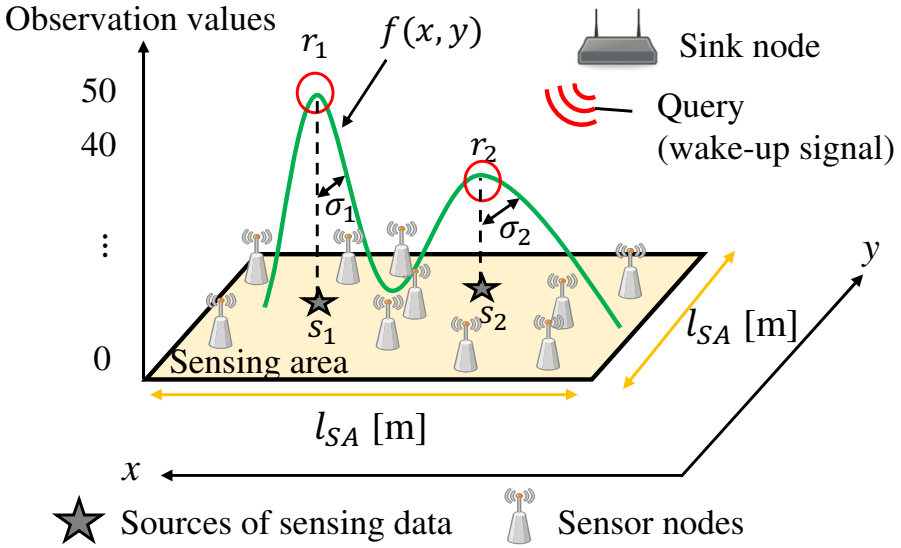


Figure 5.1: An overview of model of observation values ($N = 10, N_p = 2$).

We assume that each sensor node installs a wake-up receiver. The sink transmits a wake-up signal to collect data from the nodes in a sleep state, in which a primary radio is turned off. After detecting the wake-up signal at the wake-up receiver, each node wakes up (i.e., turns on its primary radio) and transmits data to the sink following CSMA/CA protocol defined in IEEE 802.15.4 [61]. A sensor node that succeeds in data transmission, i.e., if it receives an ACK from the sink, transits back to the sleep state, and it is controlled not to wake up for a certain period of time. Furthermore, all nodes including the sink are assumed to be located within communication/wake-up/carrier-sensing range of each other. For simplicity, we assume a collision channel, in which packets are assumed to be lost only due to collisions.

5.2 Identification Method for Multiple Emission Sources

In this work, we take an approach to regard the observed value at each position as a pixel value of the gray-scale image, which is used to identify the information of multiple sources. To this end, we consider two types of identification methods: Local Maxima (LM) based approach (c.f. Sec. 5.2.1) and CNN based approach (c.f. Sec. 5.2.2). Note that the gray-scale image constructed by data collected from sensor nodes is "sporadic" in a sense that only locations with sensors deployed have their pixel values.

5.2.1 LM-based Approach

When the distance between emission sources is large, the position of source is likely to approach to the peak point (local maxima) over the 2-dimensional field shown in Fig. 5.1. Therefore, it is reasonable to use LM-based approach as an identification method, in which the sink regards the local maxima point over the 2-dimensional space as the position of sources. This approach consists of two steps. First, as step 1, the sink increases the resolution of sporadic gray-scale image generated by the data collected from sensor nodes by using an interpolation method. Then, in step 2, the sink outputs the local maxima of two dimensional interpolated data. As an interpolation method, we resort to the linear or cubic interpolation, whose performance will be evaluated in Sec. 5.4.2. The condition for a pixel to be a local maxima is to own the value of pixel, which is equal to or more than its neighboring pixels. In this thesis, we define the connectivity of pixels as an 8-connected neighbor, i.e., the 8 pixels surrounding a target pixel are regarded as a neighbor.

5.2.2 CNN-based Approach

In order to identify multiple emission sources, it is sufficient for the sink to identify whether each source exists in a certain point of an image (a sensing area) given the set of observations of sensor nodes. In other words, the objective is for the sink to choose a plausible combination of sources that gives the input data of observations. This is a classification problem that is widely considered in the field of machine learning. In this study, we consider an approach using CNN in particular, where the sink inputs the collected observations as a gray-scale image into the trained classifier. The classifier estimates and outputs the multiple sources of sensing observations as a label. Note that CNN has been applied for the method of source detection, such as estimation of a position of transmitting node based on Received Signal Strength Indicator (RSSI) value observed by sensor nodes [84], point source detection in astronomy [85], and so on.

Design of CNN

CNN is a type of supervised learning, which mainly has strength in the feature extraction within an image, and is widely used in object classification such as You Only Look Once (YOLO) [86]. In this study, we design the simplest CNN with three layers, which is shown in Fig. 5.2. In this work, we assume 10×10 matrix as input data. A pixel value of each image takes the integer with the range of [0, 255] corresponding to the observed data of each location. The features of this input data are extracted in the convolution layer. In this study, we set the

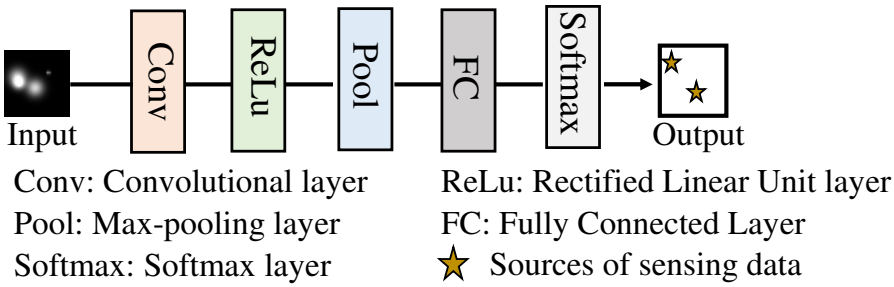
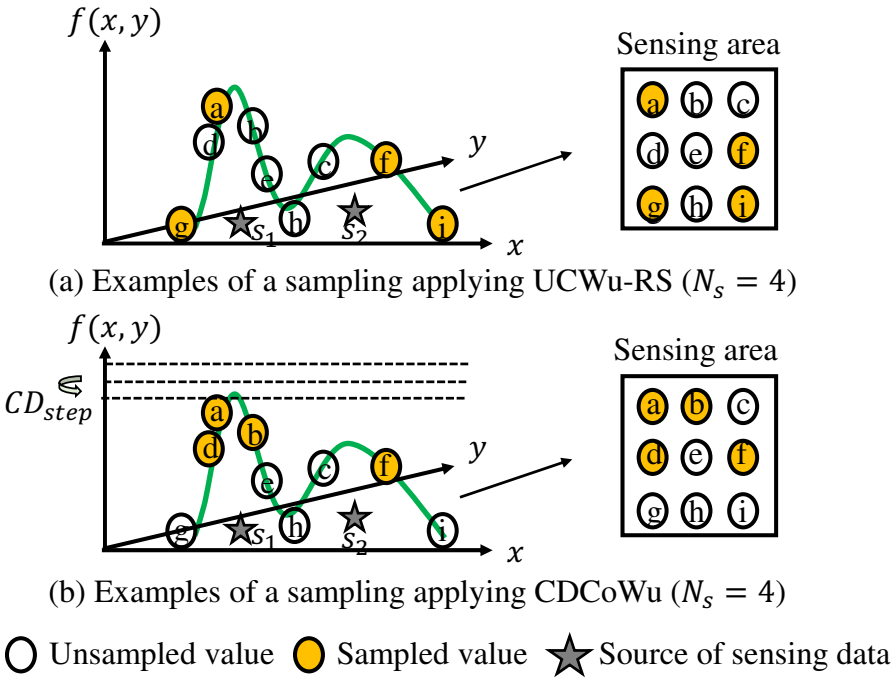


Figure 5.2: CNN model adapted in our study.

filter size used in the convolution layer to 5×5, and the number of filters to 20. The activation function is applied to the value after convolution in the Rectified Linear Unit Layer, which outputs a linear value when the input value exceeds 0 and otherwise returns 0. Then, unbiased feature extraction and downsampling are performed in the pooling layer. Here, max-pooling, which outputs the maximum value in each area, is used, and the size of the stride is set to be 2. Then, it is connected to the Fully Connected Layer, where the output, whose size corresponds to the total number of labels, is obtained. Finally, the value normalized by the softmax layer is output. Then, as a result of the classification, the positions of sources of sensing observations are estimated as a label.

Preparing a dataset for the learning

In order to accurately estimate the positions of observation sources, it is necessary to prepare all possible answers (i.e., all patterns of sources) as the input data. However, learning with all data prepared are complex because it requires a massive amount of calculation time. Therefore, for simplicity, we only consider 100 types of possible labels in this study. In each label, we choose the number of sources N_p from the range of [2, 5] and the position coordinates (x_i, y_i) on the 2-dimensional plane of the source i ($i \leq N_p$) from the range of [0, 100] based on uniform distribution, respectively. Here, we synthetically generate 100 types of gray-scale images for each label. Specifically, the observation values of position (x, y) for each label are generated based on eq. (5.1), where the covariance σ_i^2 and the data value r_i of the position (x_i, y_i) are respectively chosen from the range of [25, 100] and [25, 40] based on uniform distribution. Then, 70% of the created data set is allocated to training data, and the rest of 30% is set to be the test data. In addition, the maximum number of epochs and the initial learning rate are set to be 20 and 10^{-4}, respectively, and the stochastic gradient descent with the moment is used for updating the network weight.



(a) Examples of a sampling applying UCWu-RS ($N_s = 4$)

(b) Examples of a sampling applying CDCoWu ($N_s = 4$)

Figure 5.3: Operations of each wake-up control ($N = 9$, $N_p = 2$).

5.3 Selection of Nodes to Activate with Wake-up Control

This chapter aims to clarify the energy-efficient wake-up control to selectively activate sensor nodes suited for the identification of emission sources of sensing observations. To this end, we introduce two types of selection methods (c.f. Sec. 5.3.1 and Sec. 5.3.2), in which we apply two types of conventional wake-up control: UCWu and CoWu. The sink collects a subset of data within a network by applying these wake-up control, which we call "sampling" in this thesis. Hereafter, the number of collected data by sampling is denoted as N_s.

Below, we will explain each sampling method by using examples given in Fig. 5.3. In Fig. 5.3, there are 9 sensor nodes $\{a, b, c, d, e, f, g, h, i\}$ ($N = 9$), two emission sources $\{s_1, s_2\}$ ($N_p = 2$) with $N_s = 4$. In Fig. 5.3 –(a) and (b), the left figure shows the observation value corresponding to each position, while the right figure shows the node arrangement in the 2-dimensional sensing area and sampling examples of each method.

5.3.1 UCWu based Random Sampling (UCWu-RS)

With UCWu-Random Sampling (RS), the sink applies UCWu to collect data from nodes with their IDs randomly selected, in which a single sensor node is activated in turn without any congestion to be resolved. Fig. 5.3-(a) shows an example where the sink randomly samples four different data $\{a, f, g, i\}$ in the sensing area. In UCWu-RS, as shown on the left side of Fig. 5.3-(a), the sink randomly collects data regardless of the distribution of observed values. Therefore, while UCWu-RS is able to collect data uniformly both over the observation value and over the sensing area, it can miss some important data to identify the emission sources, e.g., large values close to each emission source. In UCWu-RS, the sink uses 2D linear interpolation to reconstruct the sensing field, which is reasonable because the sink has data uniformly spread in a sensing field.

5.3.2 CountDown Content-based Wake-up

As described in Sec. 5.3.2, in CDCoWu, the sink gradually reduces the threshold of CoWu by CD_{step}. The sink finishes to collect data if the size of collected set reaches n_s, i.e., we set $k = n_s$. As we have seen through Chap. 3, in CDCoWu, the nodes that wake up simultaneously need to contend the channel, which causes congestions to be resolved by activated nodes, by which energy consumption of nodes becomes higher. However, by applying CDCoWu, the sink can grasp the information close to the local maxima (peak) of observations, which are likely to be located around the emission sources. Fig. 5.3-(b) shows an example, where the sink collects data from 4 nodes $\{a, b, d, f\}$ that observe higher values. In this example, we can see that the sink can grasp the information on local maxima around s_1 by collecting data from nodes $\{a, b, d\}$.

5.4 Numerical Evaluation

5.4.1 Simulation Model

In this evaluation, we divide a sensing area of 100 [m] $\times$ 100 [m] into grids with their intervals of 10 [m] $\times$ 10 [m], where each sensor node is deployed at the center of each grid, which results in $N = 100$. The positions of emission sources are generated as explained in Sec. 5.2.2, where r_i and σ_i^2, are randomly generated from the range of [25, 40] and [25, 100] respectively. Note that, these generated data are not used for training of the classifier described in Sec. 5.2.2. We have prepared 100 types of labels and conducted 100 simulations for each, i.e., we conducted 10,000 simulations. The values of the other parameters for numerical evaluations are shown in Table 5.1. Here, we assume that the sink employs signaling, exploiting

Table 5.1: Parameters employed for computer simulations.

Parameters	Values
Data transmission rate	100 kbps
Access Control	CSMA/CA
Maximum number of back-offs	4
macMinBE	3
macMaxBE	5
Maximum number of retransmission	3
Range of observed values of a sensor node	$[0, 50]$
CD_{step}	0.1953
Size of transmitted packet	10 bytes
Power consumption in Transmit state	55 mW
Power consumption in Receive state	50 mW
Minimum wake-up frame length [4]	10.80 ms
Interval of wake-up frame length [4]	0.16 ms
UCWu signal [4]	$10.80 + i \times 0.16$ ms (i: node ID)
CoWu signal [4]	$10.80 + j \times 0.16$ ms (j: quantization interval)

the length of the frame [4], in which we prepare different frame lengths for mapping the different IDs in UCWu and observations in CoWu as shown in Table 5.1.

5.4.2 Comparison of Different Identification Methods

In this thesis, we introduce two metrics to evaluate the identification accuracy of each method described in Sec. 5.2: capture rate ω and travel cost τ_c. The capture rate is defined as a ratio of the number of trials where all emission sources are correctly detected (captured) against the number of total trials. The condition for capturing the emission source is that at least one estimation point exists within a cycle of radius of τ [m] with its center located at each emission source. If all emission sources are not captured, its trial is considered as failed. On the other hand, travel cost is defined as an averaged length to move from a base point to the estimated positions of different sources under the condition that all emission sources are correctly captured. In this study, we assume that a user moves from a specific base point (x_b, y_b) to the estimated position of each source to confirm the presence or absence of emission gas outflow, and conduct some actions against

the source if necessary. Here, finding the optimal travel route, i.e., considering the cost of travel between different emission sources is out of the scope of this thesis. Therefore, this thesis assumes that a user returns to the base point after each travel to the estimated position and evaluates the total travel distance as travel cost without considering the cost required for each action. The smaller the travel cost is, the better the estimation accuracy is. If the number of sources estimated by the sink node is simply increased in comparison to the actual number of sources, the capture rate ω increases. On the other hand, the larger number of estimated sources leads to the increase of the travel cost τ_c, which is not desirable considering the practical applications. Therefore, in addition to the capture rate of emission sources, travel cost is introduced to evaluate the accuracy in terms of the number and location of estimated positions. We first evaluate two types of identification methods described in Sec. 5.2 when the sink collects data from all sensor nodes. The base position (x_b, y_b) for calculating travel cost is set to [0, 0] and $\tau = 10\sqrt{2}$. We prepare two types of label sets considering different situations of source density as described below:

- Label set 1: N_p is selected from [2, 5], in which the distances among different sources are relatively large.
- Label set 2: N_p is selected from [5, 10], which corresponds to the case where the sources are densely located.

Table 5.2 shows the simulation results. From Table 5.2, first, we can see that, the capture rates of LM methods are decreased as N_p increases. As N_p becomes large, the probability that the emission sources are located within proximity increases, by which the local maxima in 2D field deviates from the position of emission source. In this case, the LM based approach cannot give the correct answer. By comparing the results of different interpolation methods in the LM based approach, we can see that the cubic interpolation achieves higher capture rate than the linear interpolation. However, we can clearly see that this requires us to accept much higher travel cost caused by many false positive. On the other hand, we can see that the travel cost of LM method with linear interpolation is much smaller than the other schemes. This is simply because, it cannot distinguish the different sources located nearby, i.e., there are many false negative, thereby causing lower capture rate.

Finally, we can see that CNN realizes both higher capture rate and smaller travel cost than those of LM methods. With prior learning, CNN can select the correct label with high probability, which enables the sink to grasp all the information on emission sources correctly. From these results, we can confirm the effectiveness of CNN in terms of the identification accuracy of multiple emission sources of sensing data.

Table 5.2: Estimation accuracy of multiple sources of sensing data using trained labels.

Identification method	ω ($N_p \in [2,5]$)	τ_c ($N_p \in [2,5]$)	ω ($N_p \in [5,10]$)	τ_c ($N_p \in [5,10]$)
LM-based (Linear interpolation)	0.830	2.18×10^2 [m]	0.470	4.15×10^2 [m]
LM-based (Cubic interpolation)	0.857	2.34×10^3 [m]	0.488	1.61×10^3 [m]
CNN	0.993	2.67×10^2 [m]	0.995	5.71×10^2 [m]

5.4.3 Comparison of Different Wake-up Control

In this section, we evaluate the identification accuracy of different sampling approaches employing wake-up control described in Sec. 5.3. Here, we adopt CNN as an identification method based on the results shown in Sec. 5.4.2, and will clarify the wake-up control suited for the identification of multiple emission sources, in terms of the estimation accuracy and the total energy consumption of sensor nodes. We are interested in how much the accuracy would be improved by an additional number of N_s with different node selection scheme. Therefore, hereafter, we define accuracy as the ratio of the number of trials where the estimated label by CNN is correct against the number of total trials. On the other hand, total energy consumption is defined as the total amount of energy consumed during the data collection exploiting each wake-up control described in Sec. 5.3. Note that the energy consumption of sensor nodes becomes larger as N_s increases, as more nodes wake up and transmit data by consuming the energy.

Fig. 5.4 shows the simulation results of the achieved set of total energy consumption and the estimated accuracy for UCWu-RS and CDCoWu. First, from this figure, we can see that, as an overall tendency, the estimation accuracy is increased with larger energy consumption caused by more number of sampling. Thus, we can say that increasing the number of collected data leads to the improved accuracy. In UCWu-RS, the sink collects data uniformly over space, not considering the importance of data. Thus, its accuracy monotonically increases with more number of sampling nodes at the cost of consumed energy.

Next, we can see that CDCoWu achieves higher accuracy with smaller energy consumption, i.e., with a fewer N_s. For example, with an energy of 0.005 [J], the accuracy reaches 0.65 using CDCoWu, at least six times higher than UCWu-RS. This is because CDCoWu enables the sink to grasp the rough information on the positions of emission sources, by which the classifier of CNN can choose the correct label with a higher probability. We can also see the difference in characteristics between CDCoWu and UCWu-RS. The accuracy of CDCoWu largely increases

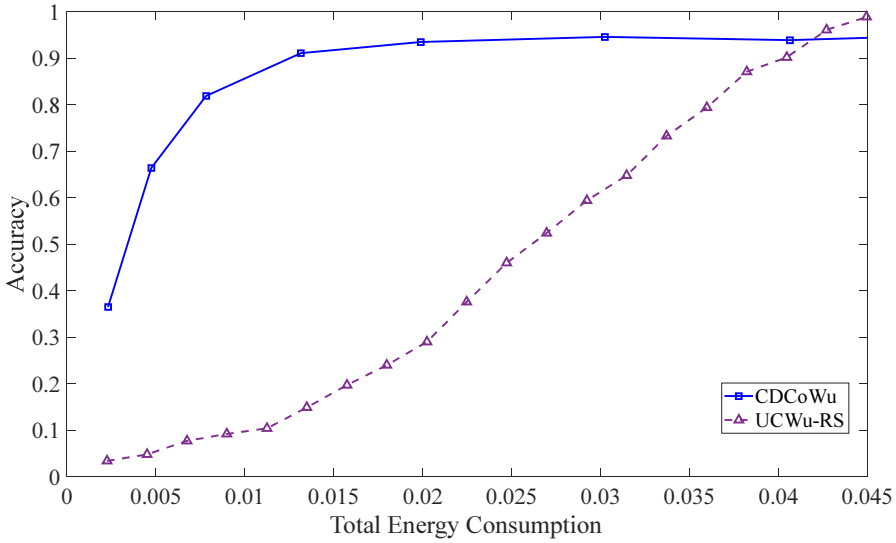


Figure 5.4: The achieved set of energy consumption and accuracy for different types of wake-up control.

at first with the additional samples of data. However, the increasing rate gradually becomes smaller against the increase of energy, which means that additional data samples obtained by CDCoWu are not necessarily beneficial to improve the accuracy after a certain number of data are collected. In order to realize larger accuracy with CDCoWu, the sink needs to collect a larger sample N_s by reducing the threshold of CoWu to small values. In this case, the effect of the congestion becomes severe as many nodes observing lower values located far from the position of sources wake up simultaneously, which increases total energy consumption. However, we can clearly see that CDCoWu achieves high energy efficiency and high estimation accuracy for most of the region compared with the results of UCWu-RS. From these observations, we can conclude that, in order to realize higher energy-efficiency and accuracy of identification of multiple emission sources of sensing data, it is desirable for the sink to collect data by using CDCoWu.

5.5 Summary

This chapter focused on the identification of multiple emission sources in WSNs. We investigated identification methods of multiple emission sources and clarified that the approach using CNN achieves high estimation accuracy. In order to realize energy-efficient data collection and high estimation accuracy, we investigated the characteristics of different wake-up control with respect to the number of sampling and clarified that CDCoWu achieves high accuracy with a smaller number

of sampling because it can grasp the characteristics of emission sources. In the next chapter, we will introduce CoWu to an IIoT scenario and investigate the importance of timing of CoWu signal transmission with theoretical analysis.

Chapter 6

Wake-up Control for Data Collections by a Deadline

This chapter focuses on an IIoT scenario, in which the sink must collect informative data by a deadline to send a command to the actuator. To this end, we apply CoWu and clarify the importance of query transmission timing with theoretical analysis[10].

6.1 Introduction

Low-power WSNs are expected to play a key role in supporting novel IIoT [87] applications, where the collected data may be used for real-time monitoring and diagnostics, and to control actuators. To support these applications, it is crucial for the sensor data to not only be available for the monitoring/control process when needed, but also be *timely*, i.e., represent the current state of the system. However, ensuring the timeliness of the data requires frequent transmissions, which in turn increases the power consumption of the nodes. An attractive strategy to ensure both timeliness and low power consumption is to operate in the *pull-based* communication regime, in which the desired data is directly requested by a sink node prior to a deadline, as opposed to being transmitted regularly by the sensors [2]. In addition to ensuring that the data is transmitted only when it is needed, this also allows the sink to request specific types of data, such as values within a certain range, which further reduces the total number of sensor transmissions and thus the power consumption of the WSN.

In this work, we propose a pull-based, timely communication scheme using wake-up radio [3][11]. We focus on the case where the sink aims to collect sensor

[10]The content of this chapter will appear in [7].

values within a certain range. To this end, we apply a CoWu strategy and characterize the timing of the wake-up signal: sending it too early increases the risk that the values will be outdated, i.e., no longer within the requested range, at the deadline, while sending it too late may prevent some of the sensors from successfully transmitting their values before the deadline. This trade-off is illustrated in Fig. 6.1(a) and 6.1(b), where the range $[V_L, V_U]$ is requested from the sensors ζ_T time slots prior to the deadline, and the sink transmits the wake-up signal in an early and late timing, respectively. Similarly, Fig. 6.1(c), illustrates an example of how ζ_T influences the likelihood that an observed process taking values in the interval $[1, 7]$, remains in the requested range $[4, 6]$ from the request time, t_s, until the deadline, T. We present a theoretical analysis considering a realistic MAC protocol, evaluate the performance of CoWu with the baseline scheme, and clarify the importance of the timing of the CoWu signaling and the performance of CoWu as a function of the speed of the physical process.

6.2 System Model and Problem Definition

6.2.1 Scenario and Objective

We study a time-slotted scenario comprising a sink node and N sensor nodes equipped with wake-up receivers. Each sensor observes an independent but identical, integer-valued physical process that takes values $1, 2, \ldots, M$, and evolves according to an M-state discrete-time, irreducible Markov process with transition matrix $\mathbf{Z}$. The time is indexed by the time slots, and we will denote the steady-state distribution of the Markov process by $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_M\}$. The process dynamics, but not the instantaneous values, are assumed to be known by the sink node.

The sink node receives sporadic requests, i.e., at unpredictable time instants, to collect data from the sensor nodes, which must be gathered before a given deadline T. The requests originate from an external entity, such as an actuator or monitoring software. To facilitate our analysis, in this work, we will ignore the time dependency between requests, i.e., we will assume that the time between consecutive requests is sufficiently long to allow the physical processes to reach steady-state conditions. We limit our focus to *range queries*, i.e., where the actuator requests sensor readings whose values are in the interval $[V_L, V_U]$ where $1 \leq V_L \leq V_U \leq M$.

The data is collected by the sink node by forwarding the range query to the sensors using CoWu at time t_s. As further elaborated in Sec. 6.2.2, the sensors measuring values within the requested range sample and transmit their observations to the sink through a contention-based access protocol. Since the physical

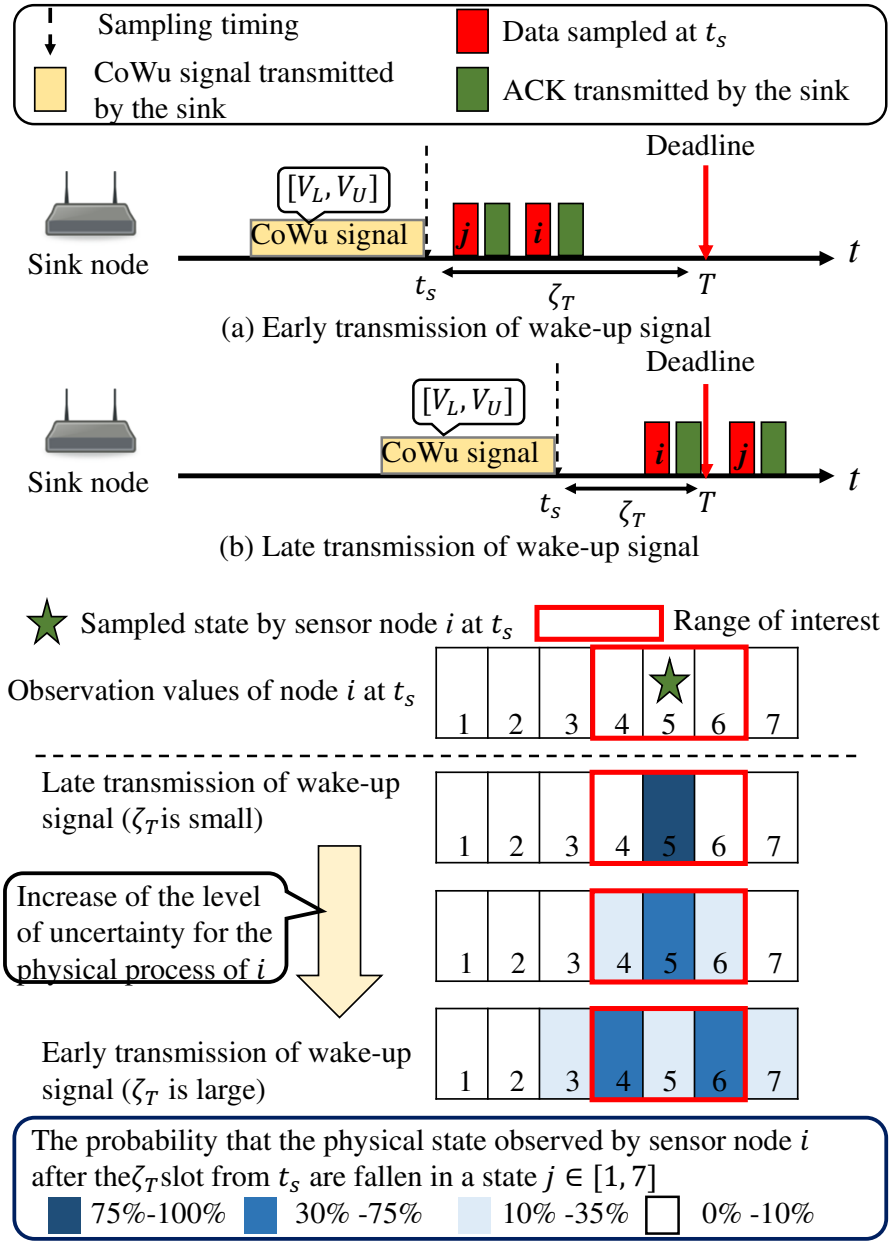


Figure 6.1: An example of data collection employing CoWu for data collections by a deadline.

processes generally evolve between the sampling time and the deadline, we define $\zeta_T = T - t_s$ as the time interval between the sampling time and the deadline, measured in time slots. The goal of the sink is to collect data that *remains timely*

at the deadline. To this end, we define the accuracy, γ, as the probability that the received sensor measurements are exactly the ones that are within the queried range at the time of the deadline. Formally, denoting by $\mathcal{T}$ the subset of nodes belonging to the interval $[V_L, V_U]$ at the deadline time T, and by $\mathcal{S}$ the subset of nodes whose data have been successfully transmitted to the sink by time T, we define the accuracy as

$$\gamma = \Pr(\mathcal{T} = \mathcal{S}). \tag{6.1}$$

Thus, the accuracy is one only if the received range set always coincides with the true set at the deadline. Our aim is to characterize the trade-off between ζ_T and γ.

6.2.2 CoWu Transmission Model

CoWu for the range query can be implemented by encoding the lower and upper interval limits, V_L and V_U, into the duration of the wake-up signal [4]. Specifically, the sink transmits first a wake-up signal of length proportional to the lower interval limit, V_L, and then transmits one proportional to the upper limit, V_U. Each wake-up receiver then checks whether its sensed value is within the range based on the signal length extracted by non-coherent OOK detection, and, if so, activates its main radio interface and transmits its observation; otherwise, it remains in a sleep state, keeping the main radio off.

The primary radio, used by the nodes that observe data within the specified range to transmit their observations to the sink, follows a p-persistent CSMA [4] protocol. We assume that each transmission occupies L slots, and that the sink transmits an error-free acknowledgment after each successful sensor transmission, so that nodes return to sleep after a successful transmission. The p-persistent model, in which each active node transmits in an idle slot with probability p and stays silent with probability $1 - p$, is an analytically tractable, yet good approximation, of many practical CSMA protocols, such as the one used in the IEEE 802.15.4 standard [4][61][64]. For simplicity, we assume that all nodes, including the sink, are located within each other's communication/wake-up/carrier-sensing range, i.e., there are no hidden terminals.

6.3 Analysis of CoWu

In this section, we derive the accuracy, γ, of the range query defined in eq. (6.1) for a given value of ζ_T. In CoWu, the probability of a node waking up depends on N, the distribution of physical process, and $[V_L, V_U]$. Let $P_w(V_{th})$ denote the probability of nodes waking up given the stationary distribution, $\boldsymbol{\pi}$, and the CoWu

threshold, $V_{th} = [V_L, V_U]$. We then have

$$P_w(V_{th}) = \sum_{i=V_L}^{V_U} \pi_i, \tag{6.2}$$

and the probability that w nodes wake up follows a binomial distribution

$$P_d(w) = \binom{N}{w} P_w(V_{th})^w (1 - P_w(V_{th}))^{N-w}. \tag{6.3}$$

To compute the distribution of the number of successful transmissions under p-persistent CSMA, we construct a two-dimensional Markov chain indexed by the current slot, in which the state represents the number of nodes that still need to transmit and the number of slots elapsed since the start of the ongoing transmission. Specifically, conditioned on the number of active nodes w and knowing that each transmission requires L slots, the state space is $\{(w,0),(w,1),\ldots,(w,L-1),(w-1,0),\ldots,(1,L-1),(0,0)\}$, where state (n,l) represents the case where n nodes have not completed their transmission, and the transmitting node(s) has been transmitting for l slots.

The transition probabilities are defined as follows. When the state is $(n,0)$, $n = 1,2,\ldots,w$, the channel is idle and one or more of the n remaining nodes can initiate a transmission, causing a transition to state $(n,1)$. This happens with probability $1-(1-p)^n$. On the other hand, if none of the nodes transmit, the Markov chain remains in state $(n,0)$, which happens with probability $(1-p)^n$. In states (n,l), $l = 1,2,\ldots,L-2$, the channel is busy and $L-l$ slots remain of the current transmission, so the Markov chain transitions to state $(n,l+1)$ with a probability 1. In state $(n,L-1)$, there are two cases depending on whether one or multiple users were transmitting in the previous slots. If only one node transmitted, the transmission is successful and the Markov chain transitions into state $(n-1,0)$. This happens with probability

$$S_n = \frac{np(1-p)^{n-1}}{1-(1-p)^n}, \tag{6.4}$$

which is conditioned on the event that at least one node is transmitting. If more than one user transmitted, which happens with probability $1-S_n$, all transmissions fail and the Markov chain returns to state $(n,0)$. Finally, state $(0,0)$ is an absorbing state representing the event that all w active users have successfully transmitted their measurement.

Using the Markov chain, we can obtain the distribution of the number of successful transmissions by state evolution from the initial state distribution $\Phi(0) = (1,0,0,\ldots,0)$ as

$$\Phi(t+1) = \Phi(t)\mathbf{R}, \tag{6.5}$$

where $\mathbf{R}$ is $(wL+1)\times(wL+1)$ transition matrix containing the transition probabilities defined above. The probability that w_s out of the w active nodes succeed by the deadline for a given ζ_T, denoted as $P_s(w_s,\zeta_T)$, is then

$$P_s(w_s,\zeta_T)=\begin{cases}\phi_{(0,0)}(\zeta_T) & \text{if } w_s=w\\ \sum_{l=0}^{L-1}\phi_{(w-w_s,l)}(\zeta_T) & \text{otherwise,}\end{cases} \tag{6.6}$$

where $\phi_{(i,j)}(\zeta_T)$ is the entry of $\Phi(\zeta_T)$ corresponding to state (i,j).

We can now derive the accuracy, defined as the probability that the set of nodes from which the sink has received values coincides with the set of nodes whose reading at the deadline, T, is in the requested interval $[V_L,V_U]$. Let us denote the state of the physical process at node i at time t_s as v_s^i and at the deadline T as v_T^i. In order for the sink to estimate the range set correctly at the deadline T, all of the following three conditions must be satisfied:

- Cond. A: For the nodes that succeed in data transmission, $(v_s^i\in[V_L,V_U])\wedge(v_T^i\in[V_L,V_U])$.
- Cond. B: For the nodes that wake up but fail their data transmission by the deadline, $(v_s^i\in[V_L,V_U])\wedge(v_T^i\notin[V_L,V_U])$.
- Cond. C: For the nodes that do not wake up and transmit data, $(v_s^i\notin[V_L,V_U])\wedge(v_T^i\notin[V_L,V_U])$.

Due to the symmetry of the physical processes, the probabilities of conditions A, B, and C are the same for all users, and denoted as $P_A(\zeta_T)$, $P_B(\zeta_T)$ and $P_C(\zeta_T)$, respectively. The probabilities are given as

$$P_A(\zeta_T)=\frac{\sum_{i\in[V_L,V_U]}\left[\pi_i\sum_{j\in[V_L,V_U]}[\mathbf{Z}^{\zeta_T}]_{i,j}\right]}{\sum_{s\in[V_L,V_U]}\pi_s}, \tag{6.7}$$

$$P_B(\zeta_T)=\frac{\sum_{i\in[V_L,V_U]}\left[\pi_i\sum_{j\notin[V_L,V_U]}[\mathbf{Z}^{\zeta_T}]_{i,j}\right]}{\sum_{s\in[V_L,V_U]}\pi_s}, \tag{6.8}$$

$$P_C(\zeta_T)=\frac{\sum_{i\notin[V_L,V_U]}\left[\pi_i\sum_{j\notin[V_L,V_U]}[\mathbf{Z}^{\zeta_T}]_{i,j}\right]}{\sum_{s\notin[V_L,V_U]}\pi_s}, \tag{6.9}$$

where $[\mathbf{Z}]_{i,j}$ is the (i,j)-th entry of $\mathbf{Z}$. Combining eqs. (6.3), (6.6), (6.7), (6.8), and (6.9), the accuracy of CoWu for a given ζ_T and $[V_L,V_U]$ can be computed as

$$\begin{aligned}\gamma_{CoWu}(\zeta_T)=\sum_{w=0}^{N}\sum_{w_s=0}^{w}&P_A(\zeta_T)^{w_s}P_B(\zeta_T)^{w-w_s}\\ &\times P_C(\zeta_T)^{N-w}P_s(w_s,\zeta_T)P_d(w).\end{aligned} \tag{6.10}$$

6.4 Numerical Results

In this section, we evaluate the performance of CoWu and compare it to a round-robin scheduling method. Throughout the evaluation, we assume that the physical process follows a truncated birth-death process where the probability that the value is incremented or decremented is q.

6.4.1 Baseline Scheme for Evaluation: Round-robin Scheduling

In round-robin scheduling, the nodes transmit their measurements according to a Time Division Multiple Access (TDMA)-like policy, whose transmission starts exactly NL slots prior to the deadline, denoted as t_{sch}, after detecting a wake-up signal triggering all nodes at each wake-up receiver. The nodes sample and transmit their measurements in order, i.e., node $j = 0, 1, \ldots, N-1$, samples and transmits its measurement at $t_{sch} + jL$.

Here, we derive the accuracy γ of round-robin scheduling for the range query. The distribution of physical process during the sampling period is assumed to be the stationary distribution, and sampled value evolves by the discrete Markov chain model until the deadline. The value to be taken at the deadline T to estimate the correct range-set at the sink depends on the value observed by each node in each slot. In round-robin scheduling, accuracy is perfect only if all nodes satisfy the following requirements:

- Cond. D: If $v_s^i \in [V_L, V_U]$, then $v_T^i \in [V_L, V_U]$
- Cond. E: If $v_s^i \notin [V_L, V_U]$, then $v_T^i \notin [V_L, V_U]$.

Here, the probabilities of conditions D and E are denoted as $P_D(\zeta_T)$ and $P_E(\zeta_T)$, respectively, and can be expressed as

$$P_D(\zeta_T) = \sum_{i \in [V_L, V_U]} \left[\pi_i \sum_{j \in [V_L, V_U]} [\mathbf{Z}^{\zeta_T}]_{i,j} \right], \tag{6.11}$$

$$P_E(\zeta_T) = \sum_{i \notin [V_L, V_U]} \left[\pi_i \sum_{j \notin [V_L, V_U]} [\mathbf{Z}^{\zeta_T}]_{i,j} \right]. \tag{6.12}$$

The accuracy of round-robin scheduling is then computed as

$$\gamma_{Sch} = \prod_{i=1}^{N} P_D\{(N-i+1)L\} + P_E\{(N-i+1)L\}. \tag{6.13}$$

6.4.2 Comparison Between CoWu and Round-robin Scheduling

Accuracy against ζ_T

Fig. 6.2 shows the accuracy of CoWu against ζ_T, where we set $N = 100$, $M = 100$, $[V_L, V_U] = [94, 98]$, $L = 10$ and $q = 0.0002$. We also plot the upper bound of CoWu, where we set $P_s(w_s, \zeta_T) = 1$ for all ζ_T if $w_s = w$, otherwise $P_s(w_s, \zeta_T) = 0$, which corresponds to the case where all active users succeed in the communication, and the results of round-robin scheduling, whose accuracy does not depend on ζ_T. We obtained the numerical results from the theoretical analysis presented in Sec. 6.3 and a Monte Carlo simulation over 10^4 transmission rounds. From this figure, we can see that the results obtained with theoretical analysis coincide with simulation results, which validates our analysis.

From this figure, we also see an optimal value in terms of accuracy. Let us denote the optimal value of ζ_T as ζ_T^{opt}. For $\zeta_T < \zeta_T^{opt}$, we can see that the accuracy becomes smaller as ζ_T decreases, because the number of nodes that fail their data transmission by the deadline T increases. For $\zeta_T > \zeta_T^{opt}$, most users complete their transmission, but the sensed values become obsolete at the deadline, also leading to a decreased accuracy. We also see that CoWu achieves a higher accuracy than round-robin scheduling, provided that ζ_T is selected appropriately. This result illustrates the importance of the timing of the wake-up signaling so as to maximize the accuracy at the deadline. Finally, we can see that CoWu approaches the upper bound as ζ_T becomes larger, which is because the error is dominated by obsolete values at the deadline as opposed to collisions.

Accuracy against q

Next, we compare the accuracy of CoWu and round-robin scheduling against the speed of the physical process, characterized by q. Furthermore, we study the impact of imperfect knowledge about q. To this end, we denote the assumed value of q by $\hat{q}$, and use $\hat{q}$ to optimize ζ_T in the CoWu scheme. Fig. 6.3 shows the accuracy of CoWu and round-robin scheduling against the true value of q for the same parameters as in Sec. 6.4.2. From the figure, we see that the accuracy of both schemes decreases as q increases, as the data becomes obsolete at the deadline. Next, we see that CoWu can realize higher accuracy than round-robin scheduling across the entire considered range of q. In round-robin scheduling, to realize reliable data transmission from all nodes by the deadline, some nodes need to transmit data very early, and their measurements become obsolete at the deadline. This penalty increases as q gets larger. On the other hand, in CoWu only the subset of the nodes that observe values in the requested range at t_s wake up and transmit data. With the optimized transmission timing for the wake-

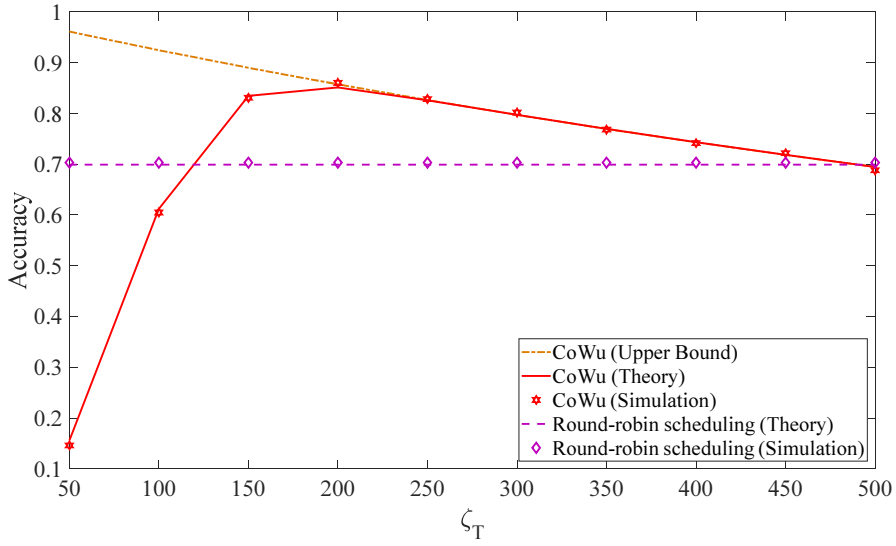


Figure 6.2: Accuracy of CoWu against ζ_T.

up signal, each node can complete data collection by the deadline and convey the timely data toward the sink, thereby obtaining a higher accuracy than with round-robin scheduling.

Finally, we focus on the results of imperfect knowledge. When $\hat{q} = 0.2 \times 10^{-3}$, i.e., the assumed value is lower than the true q, we see that the accuracy deteriorates compared to the one with perfect knowledge as q increases, and thus the difference between the assumed and true q becomes larger. The primary reason for this is that for small $\hat{q}$, the sink chooses a relatively large ζ_T, because the measurements are unlikely to change before the deadline, and it is more important to ensure that more sensor nodes successfully transmit their measurements. However, because the true q is larger than $\hat{q}$, the data collected at the deadline is likely to be out of date, leading to poor accuracy. When $\hat{q} = 4.2 \times 10^{-3}$, i.e., larger than the true q, we see that the accuracy is deteriorated when q is small. In that case, the sink sets a small ζ_T to ensure that the received data is timely at the deadline at the cost of a relatively high probability of transmission failures. This leads to an accuracy that is smaller than the case with perfect knowledge of q. Note that, despite this deterioration for the cases with imperfect knowledge, the accuracy of CoWu is still higher than that of round-robin scheduling for most cases.

Energy consumption of nodes

Here, we compare the total energy consumption of CoWu and round-robin scheduling by Monte Carlo simulation employing the same parameters as in Sec. 6.4.2, running for 10^4 rounds. The power consumption of the transmitting and receiving

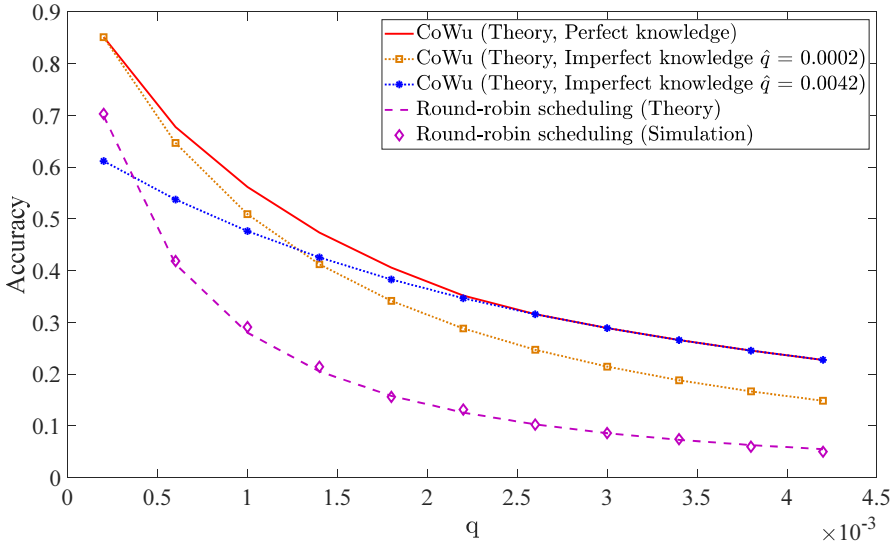


Figure 6.3: Accuracy of CoWu and round-robin scheduling against q.

state is set to 55 mW and 50 mW, respectively [69], and the duration of each time slot is defined as 320 μs [4]. We only consider the energy consumed by the main radio and ignore the energy consumed by the wake-up receiver, as its power consumption is normally much smaller than that of the main radio. Here, the total energy consumption is defined as the total amount of energy consumed by sensor nodes during the data collection period. We assume that the node activated by a wake-up signal continues the attempt of data transmission until it succeeds in data transmission. Thus, the value of total energy consumption only depends on the $[V_L, V_U]$, not on the value of ζ_T. The numerical results show that the average total energy consumption of round-robin scheduling is 1.76×10^{-2} J, while that of CoWu is 4.50×10^{-3} J. CoWu can then reduce energy consumption by about 75% against round-robin scheduling. This is achieved by the selective activation of the specific nodes that have informative data, i.e., those in the interval $[V_L, V_U]$, while all nodes are activated in round-robin scheduling regardless of the importance of their data.

6.5 Summary

This chapter focused on a scenario, where the freshness for the collected data and high energy efficiency are required for WSNs. To this end, we applied CoWu and investigated the query timing considering the deadline for data collections. By exploiting the statistical knowledge about the observed physical process, we confirmed that the sink can maximize the accuracy of the query by transmitting

a wake-up signal at an appropriate timing, and also confirmed the superiority of CoWu to round-robin scheduling. In the next chapter, we will conclude this thesis.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis focused on wake-up radio technologies to improve the overall network life time of WSNs. In order to solve the fundamental problem of conventional IDWu, i.e., inefficiency regarding energy and frequency resource for data-oriented communications, we proposed CoWu, in which the sink only activates nodes that stores the desired data by wake-up signaling.

First, we focused on a single cycle of top-k query in on-demand WSNs and proposed CDCoWu, in which the sink collects data only from the nodes belonging to the top-k set. Specifically, we proposed N-CDCoWu for top-k node-set query and V-CDCoWu for top-k value set query. N–CDCoWu (V–CDCoWu) realizes top-k data collection from the view point of nodes (observed values) with low delay and low power consumption by collecting higher values of data with higher priority. In order to confirm the efficiency of our proposed scheme, we derived theoretical equations of CDCoWu in terms of data collection delay and total energy consumption, assuming p-persistent CSMA as a MAC protocol and we revealed the following points:

- Data collection delay and total energy consumption of BCWu/UCWu calculated based on theoretical equations assuming p-persistent CSMA coincided with those of computer simulations. Furthermore, approximate analysis of data collection delay and total energy consumption of CDCoWu using MCMC, which is applied to solve the difficulty of exact calculation, agreed with the computer simulation results. From these results, we confirmed the validity of the derived equations and the approach of approximate analysis for CDCoWu.

- Regarding IDWu (BCWu and UCWu) for a single cycle of top-k query, BCWu is superior to UCWu in terms of data collection delay because BCWu can

collect data with small wake-up overhead, while UCWu is superior to BCWu in terms of total energy consumption because UCWu can collect data without congestions.

- In CDCoWu, applying small CD_{step} realizes smaller energy consumption while it increases data collection delay due to the wake-up overhead. On the other hand, large CD_{step} in CDCoWu realizes lower data collection delay by reducing the number of transmissions of wake-up signal, which, however, increases total energy consumption due to the severe congestion. Therefore, in order to reduce both data collection delay and total energy consumption, we need to optimize CD_{step}.

- In data collection employing CDCoWu, the number of activated devices changes according to the observed distribution of sensor nodes. Therefore, it is necessary to optimize CD_{step} considering the distribution of observed values in the sensing.

- Comparing N-CDCoWu and V-CDCoWu, V-CDCoWu requires more data collection delay and total energy consumption than N-CDCoWu due to the difference in the size of collected set. However, as the number of quantization bits increases, the difference in the characteristics of both methods becomes smaller.

- Obtained numerical results showed that the proposed CDCoWu outperforms IDWu for a single cycle of top-k query in terms of data collection delay and energy consumption for the maximum k-N ratio ranging from 0.1 to 0.5 when the number of sensor nodes is equal to or more than 20.

Then, we focused on the periodical top-k query in WSNs employing wake-up receivers, in which users attempt to collect a top-k set with high ranking accuracy. We proposed a wake-up control that exploits the temporal correlation of observed data, in which we introduced a transmission suppression mechanism to suppress wake-up and data transmission of nodes whose observed values do not contribute to the change of the top-k set from the previous round. We investigated the effectiveness and practicality of OTC-Wu with computer simulations and experiments and confirmed the following points:

- By choosing the adequate transmission suppression range characterized by the parameter of β, OTC-Wu can finely control trade-off between the accuracy and energy saving.

- Obtained numerical results showed that the proposed OTC–Wu can grasp the current top-k set with high energy efficiency and ranking accuracy, com-

pared with other wake-up control, especially when the degree of temporal correlation is high.

- With experiments, we confirmed that the operations of OTC-Wu at the wake-up receiver do not cause large wake-up delay compared with other schemes, which validates the practicality of OTC-Wu.

Next, we focused on energy-efficient wake-up control realizing the identification of multiple emission sources of spatially-correlated sensing data. We investigated identification methods of multiple emission sources based on the collected data and the wake-up control that achieves both high identification accuracy and energy efficiency. With computer simulations, we confirmed the following points:

- An approach employing CNN achieves high estimation accuracy for identifying multiple emission sources with respect to capture rate and travel cost.

- The simulation results showed that CDCoWu can achieve high identification accuracy with a smaller energy consumption than UCWu-RS, as CDCoWu can grasp the characteristics of multiple emission sources by collecting data from the subset of nodes observing higher value with higher priority. Thus, the energy-efficiency identification of multiple emission sources is achieved by employing CDCoWu.

Finally, we applied CoWu to a scenario in which the timeliness of data at the query timing is crucial. In order to investigate the performance of CoWu in this scenario, we analytically derived the accuracy for both CoWu and round-robin scheduling, by considering the evolution of physical process. With theoretical analysis, we clarified the following points:

- With theoretical analysis and computer simulation, we revealed that there is an optimal point that maximizes the accuracy of range-set at the deadline timing with respect to the timing of CoWu signaling. Earlier transmission of CoWu signaling results in lower accuracy at the deadline timing because the received data becomes obsolete. On the other hand, later transmission results in lower accuracy because some nodes may fail to transmit data by the deadline due to congestion.

- By setting the appropriate transmission timing of CoWu signal in accordance with the speed of physical process, CoWu can achieve high estimation accuracy at the deadline than round robin-scheduling while saving energy consumption by about 75 %, which clarified the superiority of CoWu to round-robin scheduling in terms of accuracy and total energy consumption.

- We investigated the case where the statistical prior on the physical process of the sink is imperfect, and confirmed the robustness of the CoWu to the estimation error by comparing the case where the sink has a perfect knowledge on the physical process.

Through these observations, we can conclude that CoWu is an energy-efficient technology, by which we only can allocate limited and precious wireless resource to important data for users of WSNs.

The proposed CoWu has integrated two important concepts: wake-up radio and data-oriented communications. As we have seen so far, we designed CoWu to reduce nodes' wasteful energy consumption, specifically, energy consumption during the standby period and energy consumption of nodes observing less important data. As mentioned above, we confirmed CoWu could reduce the large amount of energy consumed by sensor nodes. This indicates the important principle for protocol design for WSNs to allocate limited wireless resources based on the data content, the importance of data-oriented communication. This principle would play a vital role in WSNs systems, such as wireless power transfer systems, and other types of wireless communication systems, in which the concept of data-oriented communication has not been introduced yet. Furthermore, the developed principle is in line with the emerging communication paradigm of semantic communication, in which the data should be transmitted considering the meaning, timing, value, and relevance. The data-oriented communication protocol designs as embodied by CoWu in this thesis would play a crucial role in future wireless systems, not only for IoT data collection. This thesis brings about a new wave for future protocol design principles for wireless communication systems with respect to data-oriented communications.

7.2 Future Work

Throughout this thesis, we focused on the collision model, in which the packet is lost only by the collision, and channel impairments with independent and identical probability for simplified analysis. As a next step, it is desirable to consider more realistic network and channel models, such as distant–dependent channel impairments including multi-hop scenario and interference caused by the other systems that share the same frequency band. For example, the design of wake-up control for top-k query in multi-hop scenario is an interesting direction, in which we need to consider routing mechanism, including data aggregation at the intermediate nodes and the selection of relay nodes, and how to alleviate the effect of the hidden terminal on the wake-up signaling. Theoretical analysis considering realistic network and channel model, and the operation of practical MAC protocol is also a future work.

In Chap. 4, we investigate the energy efficiency and practicality of OTC-Wu with computer simulation and experiments. In the future work, theoretical analysis for the gain obtained by exploiting temporal correlation with a simplified scenario is needed.

In Chap. 5, we investigated the wake-up control suited for identification of multiple emission sources, which realizes high identification accuracy and high energy efficiency. The possible future work includes the introduction of a transmission suppression mechanism exploiting the spatial correlation of sensing data, in which the node suppresses its wake-up if the neighbouring nodes transmitted data representing its observed value. Furthermore, the design of wake-up control applying both UCWu and CoWu, in which the sink estimates the important data or nodes whose data significantly improves the estimation accuracy and only collects important data with an appropriate wake-up control is also an interesting direction for future work.

Possible avenues for future work related to the work described in Chap. 6 include (a) analysis of top-k query considering both information freshness of top-k set at deadline timing and energy-efficiency of sensor nodes, (b) the investigation of a case where the sink has a imperfect knowledge on the physical process, in which a new sampling policy based on the observed physical process is introduced, and (c) designing a wake-up control to collect data from the subset of nodes that improves the accuracy of the model for the sink to estimate the current value of the physical process.

Acknowledgments

The doctoral research was far more challenging and profound than I had initially imagined. I often felt frustrated and got stuck. However, thanks to many people's support, I am about to complete my Ph.D. journey. Looking back, I am amazed to find myself far away from where I was before.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Hiroyuki Yomo, for his continuous support and for giving me many opportunities for a variety of experiences through the research activity. The discussion with Prof. Yomo, his critical comments, and his valuable advice is definitely an asset to me. He always made me aware of the shallowness of my thinking, the immaturity of my understanding, and the things I should improve in various aspects. I will never forget what I learned from him during six years of my bachelor's, master's, and doctoral studies. I appreciate him for everything with my heart.

I would like to express my sincere appreciation to Prof. Petar Popovski for hosting my research stay at Aalborg University and for giving me a lot of insightful comments. It was a precious period for me, and I learned a lot during my research stay. I would also like to thank his group member for kindly welcoming me. I especially appreciate Dr. Anders E. Kalør, Dr. Federico Chiariotti, and Dr. Israel Leyva-Mayorga for taking a lot of time for the discussion and for providing a lot of valuable comments.

I also appreciate Prof. Miki Yamamoto and Prof. Kouji Hirata for their review and comments to my thesis. Their comments help me to refine my thesis.

I would like to extend my sincere thanks to the laboratory members I have spent time with, especially those who joined the laboratory at the same time and worked together, encouraging and improving each other.

This work is supported financially from JST SPRING, Grant Number JPMJSP2150, and JSPS Overseas Challenge Program for Young Researchers. I also appreciate everyone involved in "A Doctoral Development Project for Pioneering Research Students" at Kansai University, Japan.

Last, but not least, I would like to thank my parents for always taking care of me warmly and supporting me. I dedicate this thesis to my parents.

Junya Shiraishi
Kansai University, November 30, 2022

List of Publications

Academic Article with Peer Reviewed

1. Junya Shiraishi, Hiroyuki Yomo, Kaibin Huang, Čedomir Stefanović and Petar Popovski, "Content-Based Wake-Up for Top-k Query in Wireless Sensor Networks," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 362-377, March 2021.

2. Junya Shiraishi and Hiroyuki Yomo, "Periodical Top-k Ranking Query in Wireless Sensor Networks Exploiting Wake-Up Receivers," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23698-23710, Oct. 2021.

3. Junya Shiraishi, Anders E. Kalør, Federico Chiariotti, Israel Leyva-Mayorga, Petar Popovski, and Hiroyuki Yomo,"Query Timing Analysis for Content-based Wake-up Realizing Informative IoT Data Collection," *IEEE Wireless Communications Letters* (Accepted).

International Conference (with Peer Reviewed)

1. Junya Shiraishi, Hiroyuki Yomo, Kaibin Huang, Čedomir Stefanović and Petar Popovski,"Content-based Wake-up Control for Wireless Sensor Networks Exploiting Wake-up Receivers," *Proc. of 17th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt2019)*, June 2019.

2. Junya Shiraishi and Hiroyuki Yomo,"Wake-up Control for Wireless Sensor Networks Collecting top-k Data with Temporal Correlation," *Proc. of IEEE Vehicular Technology Communication (VTC2020-Fall)*, October 2020.

3. Hiroyuki Yomo, Junya Shiraishi, and Hitoshi Kawakita, "Content-based Wake-up for Aggregate Query in Wireless Sensor Networks," *Proc. of International Conference on Emerging Technologies for Communications 2020*, Dec. 2020.

4. Hiroyuki Yomo, Yuta Minami, Hitoshi Kawakita, and Junya Shiraishi, "Wake-up Control with Kernel Density Estimation for Top-k Query in Wireless Sensor Networks," *Proc. of 2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021.

5. Junya Shiraishi and Hiroyuki Yomo,"Wake-up Control for Energy-efficient Identifications of Multiple Emission Sources in Wireless Sensor Networks," *International Conference on Computing, Networking and Communications (ICNC)*, 2023 (Accepted).

Non-Refereed Technical Paper

1. Junya Shiraishi and Hiroyuki Yomo, "Theoretical Analysis of Data Collection Performance of p-persistent CSMA-based Wireless Sensor Networks employing Wake-up Receiver," IEICE Technical report, vol. 118, no. 127, ASN2018-48, pp. 147-152, July 2018.

2. Junya Shiraishi and Hiroyuki Yomo, "Theoretical Analysis of Wake-up Control for Wireless Sensor Networks," The 5th International Symposium on Electrical Engineering and Computer Science, Feb. 2019.

3. Junya Shiraishi, Hiroyuki Yomo, Kaibin Huang, Čedomir Stefanović, and Petar Popovski, "Top-k Data Collection in Wireless Sensor Networks Using Wake-up Receivers," IEICE Technical report, vol. 119, no. 53, SeMI2019-16, pp. 219-224, May 2019.

4. Junya Shiraishi and Hiroyuki Yomo, "Top-k Data Collection Exploiting Temporal Correlation in Wireless Sensor Networks with wake-up Receivers," The 6th International Symposium on Electrical Engineering and Computer Science, Feb. 2020.

5. Junya Shiraishi, Shun Hirano, and Hiroyuki Yomo, "Top-k Data Collection for Wireless Sensor Networks Employing Wake-up Receiver Having Temporal Correlation of Observed Data," The 2020 IEICE General Conference, B-15-41, March 2020.

6. Junya Shiraishi and Hiroyuki Yomo, "An Energy-Efficient Top-k Data Collection Considering Ranking Accuracy for Wireless Sensor Networks Using Wake-up Receivers," IEICE Technical report, vol. 120, no. 297, NS2020-97, pp. 54-59, December 2020.

7. Junya Shiraishi and Hiroyuki Yomo, "A Study on Energy-Efficient Identification of Multiple Sources of Sensing Data for Wireless Sensor Networks,"

IEICE Technical report, vol. 121, no. 393, pp.40-45, SRW2021-77, March 2022.

8. Takuya Murakami, Junya Shiraishi, and Hiroyuki Yomo, "Multi-hop Wake-up Control for Top-k Query in Wireless Sensor Networks," IEICE Technical report, vol. 122, no. 164, RCS2022-113, pp. 90-95, August 2022.

9. Junya Shiraishi, Anders E. Kalør, Federico Chiariotti, Israel Leyva-Mayorga, Petar Popovski, and Hiroyuki Yomo, "Content-based Wake-up Considering Deadline for Data Collection," Joint IEEE SPS and EURASIP Summer School on Defining 6G, August 2022.

10. Junya Shiraishi, Anders E. Kalør, Federico Chiariotti, Israel Leyva-Mayorga, Petar Popovski, and Hiroyuki Yomo, "Content-based Wake-up Control for Wireless Sensor Networks Considering Information Freshness at Query Instance," IEICE Technical report, vol. 122, no. 278, SeMI2022-62, pp. 55-56, November 2022.

Bibliography

[1] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.

[2] F. Chiariotti et al., "Query age of information: Freshness in pull-based communication," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1606–1622, 2022.

[3] H. Yomo et al., "ROD-SAN: Energy-efficient and high-response wireless sensor and actuator networks employing wake-up receiver," *IEICE Trans. on Comms.*, vol. E99-B, no. 9, pp. 1998–2008, Sept. 2016.

[4] J. Shiraishi, H. Yomo, K. Huang, Č. Stefanović, and P. Popovski, "Content-based wake-up for top-k query in wireless sensor networks," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 362–377, 2021.

[5] J. Shiraishi and H. Yomo, "Periodical top-k ranking query in wireless sensor networks exploiting wake-up receivers," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23 698–23 710, 2021.

[6] ——, "Wake-up control for energy-efficient identifications of multiple emission sources in wireless sensor networks," *Accepted for International Conference on Computing, Networking and Communications (ICNC).*

[7] J. Shiraishi, A. E. Kalør, F. Chiariotti, I. Leyva-Mayorga, P. Popovski, and H. Yomo, "Query timing analysis for content-based wake-up realizing informative iot data collection," *Accepted for IEEE Wireless Communications Letters.*

[8] R. C. Carrano, D. Passos, L. C. Magalhaes, and C. V. Albuquerque, "Survey and taxonomy of duty cycling mechanisms in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 181–194, 2013.

[9] L. Feltrin, G. Tsoukaneri, M. Condoluci, C. Buratti, T. Mahmoodi, M. Dohler, and R. Verdone, "Narrowband iot: A survey on downlink and uplink perspectives," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 78–86, 2019.

[10] I. Demirkol, C. Ersoy, and E. Onur, "Wake-up receivers for wireless sensor networks: Benefits and challenges," *IEEE Wireless Communications*, vol. 16, no. 4, pp. 88–96, August 2009.

[11] R. Piyare, A. L. Murphy, C. Kiraly, P. Tosato, and D. Brunelli, "Ultra low power wake-up radios: A hardware and networking survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2117–2157, 2017.

[12] D. Ghose, F. Y. Li, and V. Pla, "MAC protocols for wake-Up radio: Principles, modeling and performance analysis," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 2294–2306, May 2018.

[13] J. Oller, I. Demirkol, J. Casademont, J. Paradells, G. U. Gamm, and L. Reindl, "Has time come to switch from duty-cycled MAC protocols to wake-up radio for wireless sensor networks?" *IEEE Transactions on Networking*, vol. 24, no. 2, pp. 674–687, April 2016.

[14] F. A. Aoudia, M. Gautier, and O. Berder, "OPWUM: Opportunistic MAC protocol leveraging wake-up receivers in WSNs," *Journal of Sensors*, vol. 2016, p. 9 pages, 2016, article ID 6263719. [Online]. Available: https://doi.org/10.1155/2016/6263719

[15] L. Guntupalli, D. Ghose, F. Y. Li, and M. Gidlund, "Energy efficient consecutive packet transmissions in receiver-initiated wake-up radio enabled WSNs," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4733–4745, June 2018.

[16] T. H. Le, A. Pegatoquet, and M. Magno, "Asynchronous on demand MAC protocol using wake-up radio in wireless body area network," *Proc. of 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pp. 228–233, June 2015.

[17] J. Blanckenstein, J. Klaue, and H. Karl, "Energy efficient clustering using a wake-up receiver," in *European Wireless 2012; 18th European Wireless Conference 2012*. VDE, 2012, pp. 1–8.

[18] B. Babcock and C. Olston, "Distributed top-k monitoring," *Proc. of the 2003 ACM SIGMOD International Conference on Management of Data*, pp. 28–39, June 2003.

[19] R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," *Proc. of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 102–113, 2001.

[20] D. Zeinalipour-Yazti *et al.*, "The threshold join algorithm for top-k queries in distributed sensor networks," *Proc. of the 2nd International Workshop on Data Management for Sensor Networks*, pp. 61–66, August 2005.

[21] A. S. Silberstein, R. Braynard, C. Ellis, K. Munagala, and Jun Yang, "A sampling-based approach to optimizing top-k queries in sensor networks," *Proc. of 22nd International Conference on Data Engineering (ICDE'06)*, pp. 102–113, April 2006.

[22] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov, "Top – k query based dynamic scheduling for iot-enabled smart city waste collection," in *2015 16th IEEE International Conference on Mobile Data Management*, vol. 2, 2015, pp. 50–55.

[23] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: A tiny aggregation service for ad-hoc sensor networks," *SIGOPS Oper. Syst. Rev.*, vol. 36, no. SI, pp. 131–146, Dec. 2002.

[24] M. Wu, J. Xu, X. Tang, and W. Lee, "Top-k monitoring in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 7, pp. 962–976, July 2007.

[25] B. Malhotra, M. A. Nascimento, and I. Nikolaidis, "Exact top-k queries in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1513–1525, Oct. 2011.

[26] Q. Pan, M. Li, M. Wu, and W. Shu, "Optimization of accurate top-k query in sensor networks with cached data," *Proc. of 2007 IEEE Wireless Communications and Networking Conference*, pp. 4233–4238, March 2007.

[27] B. Chen, W. Liang, and J. X. Yu, "Online time interval top-k queries in wireless sensor networks," *Proc. of 2010 Eleventh International Conference on Mobile Data Management*, pp. 177–182, May 2010.

[28] B. Chen, W. Liang, R. Zhou, and J. X. Yu, "Energy-efficient top-k query processing in wireless sensor networks," *Proc. of the 19th ACM International Conference on Information and Knowledge Management*, pp. 329–338, Oct. 2010.

[29] B. Chen and G. Min, "Robust top-k query evaluation in wireless sensor networks," *Proc. of 2010 10th IEEE International Conference on Computer and Information Technology*, pp. 660–667, Oct. 2010.

[30] Y. Cho, J. Son, and Y. D. Chung, "POT: An efficient top-k monitoring method for spatially correlated sensor readings," *Proc. of the 5th Workshop on Data Management for Sensor Networks*, pp. 8–13, August 2008.

[31] M. H. Yeo, D. O. Seong, and J. S. Yoo, "PRIM: Priority-based top-k monitoring in wireless sensor networks," *Proc. of International Symposium on Computer Science and its Applications*, pp. 326–331, Oct. 2008.

[32] C. Zhu, L. T. Yang, L. Shu, V. C. M. Leung, T. Hara, and S. Nishio, "Insights of top- k query in duty-cycled wireless sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 2, pp. 1317–1328, Feb. 2015.

[33] X. Ma, J. Liang, S. Yang, Y. Li, Y. Li, W. Ma, and T. Wang, "Sls-stq: A novel scheme for securing spatial-temporal top-k queries in twsns-based edge computing systems," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 093–10 104, 2019.

[34] X. Kui, J. Feng, X. Zhou, H. Du, X. Deng, P. Zhong, and X. Ma, "Securing top-k query processing in two-tiered sensor networks," *Connection Science*, vol. 0, no. 0, pp. 1–19, 2020.

[35] H. Yang, C. Chung, and M. H. Kim, "An efficient top-k query processing framework in mobile sensor networks," *Elsevier Data & Knowledge Engineering*, vol. 102, pp. 78–95, March 2016.

[36] Y. Sasaki, T. Hara, and Y. Ishikawa, "Efficient framework for processing top-k queries with replication in mobile ad hoc networks," *Geoinformatica*, vol. 23, pp. 591–620, 2019.

[37] Y. Tsou and Y. Chen, "Ccpt: Compression and correctness-preserving top-k query for wireless sensor networks," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7749–7758, 2018.

[38] J. Tang, Z. Zhou, and L. Wang, "Answering multiattribute top-k queries in fog-supported wireless sensor networks leveraging priority assignment technology," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4507–4518, 2018.

[39] X. Meng, T. Nandagopal, L. Li, and S. Lu, "Contour maps: Monitoring and diagnosis in sensor networks," *Computer Networks*, vol. 50, no. 15, pp. 2820–2838, 2006.

[40] M. Li and Y. Liu, "Iso-map: Energy-efficient contour mapping in wireless sensor networks," *IEEE transactions on knowledge and data engineering*, vol. 22, no. 5, pp. 699–710, 2009.

[41] Y. J. Zhao, R. Govindan, and D. Estrin, "Residual energy scan for monitoring sensor networks," in *2002 IEEE Wireless Communications and Networking Conference Record. WCNC 2002 (Cat. No. 02TH8609)*, vol. 1. IEEE, 2002, pp. 356–362.

[42] W. Xue, Q. Luo, L. Chen, and Y. Liu, "Contour map matching for event detection in sensor networks," in *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, 2006, pp. 145–156.

[43] A. Hussain and Y. Luo, "Decentralized source localization without sensor parameters in wireless sensor networks," *arXiv preprint arXiv:2009.01062*, 2020.

[44] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE transactions on Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[45] M. R. Morelande and A. Skvortsov, "Radiation field estimation using a gaussian mixture," in *2009 12th International Conference on Information Fusion*. IEEE, 2009, pp. 2247–2254.

[46] L. Shu, M. Mukherjee, X. Xu, K. Wang, and X. Wu, "A survey on gas leakage source detection and boundary tracking with wireless sensor networks," *IEEE Access*, vol. 4, pp. 1700–1715, 2016.

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

[48] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. IEEE, 2011, pp. 350–358.

[49] H. Chen, Y. Gu, and S.-C. Liew, "Age-of-information dependent random access for massive iot networks," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2020, pp. 930–935.

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

[51] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1583–1587.

[52] F. Alawad and F. A. Kraemer, "Value of information in wireless sensor network applications and the iot: A review," *IEEE Sensors Journal*, 2022.

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

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

[55] H. Tang, J. Wang, Z. Tang, and J. Song, "Scheduling to minimize age of synchronization in wireless broadcast networks with random updates," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4023–4037, 2020.

[56] X. Wang, W. Lin, C. Xu, X. Sun, and X. Chen, "Age of changed information: Content-aware status updating in the internet of things," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 578–591, 2021.

[57] J. Liu, P. Tong, X. Wang, B. Bai, and H. Dai, "Uav-aided data collection for information freshness in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2368–2382, 2020.

[58] J. Wang, X. Cao, B. Yin, and Y. Cheng, "Sleep–wake sensor scheduling for minimizing aoi-penalty in industrial internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 9, pp. 6404–6417, 2021.

[59] F. Li, Y. Sang, Z. Liu, B. Li, H. Wu, and B. Ji, "Waiting but not aging: Optimizing information freshness under the pull model," *IEEE/ACM Transactions on Networking*, vol. 29, no. 1, pp. 465–478, 2020.

[60] F. Chiariotti, A. E. Kalør, J. Holm, B. Soret, and P. Popovski, "Scheduling of sensor transmissions based on value of information for summary statistics," *IEEE Networking Letters*, vol. 4, no. 2, pp. 92–96, 2022.

[61] "Ieee standard for local and metropolitan area networks–part 15.4: Low-rate wireless personal area networks (lr-wpans)," *IEEE Std 802.15.4-2011 (Revision of IEEE Std 802.15.4-2006)*, pp. 1–314, 2011.

[62] H. Yomo, Y. Kondo, N. Miyamoto, S. Tang, M. Iwai, and T. Ito, "Receiver design for realizing on-demand WiFi wake-up using WLAN signals," *Proc. of IEEE GLOBECOM 2012*, Dec. 2012.

[63] J. Shiraishi, H. Yomo, K. Huang, Č. Stefanović, and P. Popovski, "Content-based wake-up control for wireless sensor networks exploiting wake-up receivers," in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2019, pp. 1–8.

[64] K. Yedavalli and B. Krishnamachari, "Enhancement of the 802.15.4 MAC protocol for scalable data collection in dense sensor networks," *Proc. of 6th Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt' 08)*, vol. 4, pp. 152–161, April 2008.

[65] M. Ye, K. C. K. Lee, W. Lee, X. Liu, and M. Chen, "Querying uncertain minimum in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 12, pp. 2274–2287, 2012.

[66] D. Amagata, Y. Sasaki, T. Hara, and S. Nishio, "Efficient multidimensional top-k query processing in wireless multihop networks," *Mobile Information Systems*, vol. 2015, no. 657431, 2015.

[67] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*. Prentice Hall, 2007.

[68] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[69] N. Tamura and H. Yomo, "Low-overhead wake-up control for wireless sensor networks employing wake-up receivers," *IEICE Transactions on Communications*, vol. 102, no. 4, pp. 732–740, 2019.

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

[71] A. Abid, A. Kachouri, and A. Mahfoudhi, "Outlier detection for wireless sensor networks using density-based clustering approach," *IET Wireless Sensor Systems*, vol. 7, no. 4, pp. 83–90, 2017.

[72] H. Yomo, Y. Minami, H. Kawakita, and J. Shiraishi, "Wake-up control with kernel density estimation for top-k query in wireless sensor networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 01–06.

[73] S. Michel, P. Triantafillou, and G. Weikum, "Klee: A framework for distributed top-k query algorithms," in *Proceedings of the 31st international conference on Very large data bases*, 2005, pp. 637–648.

[74] M. Wu, J. Xu, and X. Tang, "Processing precision-constrained approximate queries in wireless sensor networks," in *7th International Conference on Mobile Data Management (MDM'06)*. IEEE, 2006, pp. 31–31.

[75] H. Jiang, J. Cheng, D. Wang, C. Wang, and G. Tan, "A general framework for efficient continuous multidimensional top-k query processing in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1668–1680, 2012.

[76] D. Tulone and S. Madden, "Paq: Time series forecasting for approximate query answering in sensor networks," in *European Workshop on Wireless Sensor Networks*. Springer, 2006, pp. 21–37.

[77] J. Wang, Y. Liu, and S. K. Das, "Improving information quality of sensory data through asynchronous sampling," in *2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE, 2009, pp. 1–6.

[78] X. Tang and J. Xu, "Adaptive data collection strategies for lifetime-constrained wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 6, pp. 721–734, 2008.

[79] P. Diaconis and R. L. Graham, "Spearman's footrule as a measure of disarray," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 2, pp. 262–268, 1977.

[80] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," *SIAM Journal on discrete mathematics*, vol. 17, no. 1, pp. 134–160, 2003.

[81] J. Shiraishi and H. Yomo, "Wake–up control for wireless sensor networks collecting top–k data with temporal correlation," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*. IEEE, 2020, pp. 1–5.

[82] "Compact, low power microcontrollers for general purpose applications ideal for sub-mcus," accessed: 2022-11-16. [Online]. Available: https://www.renesas.com/jp/en/

products/microcontrollers-microprocessors/rl78-low-power-8-16-bit-mcus/rl78g12-compact-low-power-microcontrollers-general-purpose-applications-ideal-sub-mcus

[83] J. Pohjankukka, S. Tuominen, and J. Heikkonen, "Utilizing remote sensing data in forest inventory sampling via bayesian optimization," *arXiv preprint arXiv:2009.08420*, 2020.

[84] C. Zhan, M. Ghaderibaneh, P. Sahu, and H. Gupta, "Deepmtl: Deep learning based multiple transmitter localization," in *2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2021, pp. 41–50.

[85] D. Mariia, B. Anton, U. Andrey, and M. Denys, "A study of neural networks point source extraction on simulated fermi/lat telescope images," *Astronomische Nachrichten*, vol. 341, no. 8, pp. 819–826, 2020.

[86] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[87] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial internet of things: Challenges, opportunities, and directions," *IEEE Transactions on industrial informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.