Aalborg Universitet

Content-based Wake-up for Top-k Query in Wireless Sensor Networks

Shiraishi, Junya ; Yomo, Hiroyuki ; Huang, Kaibin; Stefanovic, Cedomir; Popovski, Petar

Published in:
IEEE Transactions on Green Communications and Networking

DOI (link to publication from Publisher):
10.1109/TGCN.2020.3033844

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

Content-based Wake-up for Top-k Query in Wireless Sensor Networks

Junya Shiraishi, Hiroyuki Yomo, *Member, IEEE*, Kaibin Huang, *Senior Member, IEEE*, Čedomir Stefanović, *Senior Member, IEEE*, and Petar Popovski, *Fellow, IEEE*

***Abstract*—This paper proposes content–based wake–up control for top–k query in wireless sensor networks, where the sink attempts to collect information on top–k nodes or top–k values from sensor nodes employing wake–up receivers. The wake–up procedure is designed with a goal of waking up only the subset of nodes which have the relevant data observations to construct the desired top–k set. This prevents the sensors with less relevant data from waking up and wasting energy. Assuming p–persistent carrier sense multiple access (CSMA) as the medium access protocol, the proposed scheme is analyzed theoretically in terms of data collection delay and total energy consumption. The performance of the proposed wake–up control is examined over a wide variety of parameters with practical considerations and compared to the conventional identity (ID)–based wake–up (IDWu) such as unicast wake–up (UCWu) and broadcast wake–up (BCWu). The obtained numerical results show that the proposed CDCoWu outperforms IDWu in terms of data collection delay and energy consumption for the maximum ratio of top–data to number of sensor nodes ranging from 0.1 to 0.5 when the number of sensor nodes is equal to or more than 20.**

***Index Terms*—Wireless Sensor Networks, Wake–up Receiver, Energy–Efficiency, Top–k query, Medium Access Control**

I. Introduction

Wireless sensor networks (WSNs) play a key role in supporting diverse Internet of Things (IoT) applications [2] [3]. A key requirement on WSNs is *energy efficiency* that directly affects the lifetime of networks consisting of battery-powered nodes. A technology enabler in that direction is the *wake–up radio* [4], which is a secondary, low–power radio that supplements the primary radio interface used for data transmission. The sensor node is in a dormant state when it does not require communication, where its wake–up receiver is active and primary radio is switched off. When a node needs to communicate to a dormant sensor node, it first transmits a wake–up signal, which is received by the wake–up receiver and initiates activation of the primary radio interface. This allows for an *on–demand operation*, in which the main radio interface spends energy only when actually communicating. The original concept of a wake–up radio is developed with *identity (ID)–based* wake–up [4], in which a target node ID is embedded into the wake–up signal and thus the addressed node activates its main radio [5][6].

The complexity of the sensor network design increases significantly when the wake–up procedure is not based on the ID of the target nodes, but on the query that the data sink performs over the data contained in the nodes. This is denoted as *content–based wake–up* and a typical example is top–k query, where a user is interested in the top–k observations, i.e., k nodes with the highest sensor readings in a sensing field [7][8]. A conventional approach would be to separate the query from the wake–up process, such that the data sink wakes up all sensors to collect the data and then identifies the top–k observations. This results in higher energy consumption. Our previous work proposed a content–based wake–up scheme (CoWu) to reduce the energy consumption in data collection [1]. In CoWu, only the sensor nodes storing sensing data larger than a specified threshold are woken up. Based on that, a scheme specialized for top–k query was proposed, termed CountDown CoWu (CDCoWu), which was shown to be superior compared to ID–based wake–up (IDWu) such as unicast wake–up (UCWu) and broadcast wake–up (BCWu) in terms of data collection delay and total energy consumption. UCWu individually activates all sensor nodes with unicast ID while BCWu wakes up all nodes simultaneously with broadcast ID for collecting their data. However, CDCoWu works under the assumption that the observed data follows the uniform distribution. Furthermore, the top–k query from [1] is *top–k node–set* query, where the information on *k nodes* with the highest readings is collected. A more exact approach is to carry out a *top–k value–set* query, where the sink collects *top–k values* in the sensing field [7]. To illustrate, assume that there are 5 nodes A, B, C, D, E, which respectively have the sensing data of 16, 25, 30, 32, 30. In this case, the information to be collected with the top–3 node–set query is C, D, E and their readings of 30, 32, 30 while that with the top–3 value–set query is 25, 30, 32, 30 and their corresponding ID of B, C, D, E.

This paper extends CDCoWu from the conference version in [1] to deal with both node– and value–set query. Specifically,

The work of Junya Shiraishi and Hiroyuki Yomo was supported by JSPS KAKENHI Grant Number 19K04401. The work of Kaibin Huang was supported by Hong Kong Research Grants under the Grants 17208319 and 17204917. The work of Čedomir Stefanović and Petar Popovski was supported by the European Research Council (ERC) under the Horizon 2020 research and innovation program (ERC Consolidator Grant Nr. 648382 WILLOW) and Danish Council for Independent Research (Grant Nr. 8022-00284B SEMIOTIC).

J. Shiraishi and H. Yomo are with Graduate School of Science and Engineering, Kansai University, Japan (e-mail: k980158@kansai-u.ac.jp, yomo@kansai-u.ac.jp).

K. Huang is with Dept. of EEE, The University of Hong Kong, Hong Kong (e-mail: huangkb@eee.hku.hk).

Č. Stefanović and P. Popovski are with Aalborg University, Denmark (e-mail: cs@es.aau.dk, petarp@es.aau.dk).

A part of this work was presented at the 17th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOPT 2019, June, 2019 [1].

the new contributions of this paper in comparison to [1] are as follows.

- While only top–k node–set query was considered in [1], this paper proposes an enhanced algorithm of CDCoWu, which is applicable to both node– and value– set query. Furthermore, this paper analyzes the data collection delay and energy consumption considering both node–set and value–set query by assuming that the activated sensors use medium access control (MAC) with p–persistent CSMA. The impact of packet loss caused by channel impairments is also investigated.
- In [1], it was assumed that the observed data follows uniform distribution. This paper also considers the case in which the observed data follows the exponential and normal distributions.
- In [1], the value observed by each sensor is continuous. This paper considers the practical case of digital sensors with quantized, discrete data, and investigates the relationship between the wake–up control and granularity of sensing.
- This paper compares the performance of the proposed scheme to UCWu for different ratio of k to the number of sensors, which were not considered in [1]. The possible extensions of IDWu to incorporate a scheduled data MAC (SDMAC) [9] are also discussed[1]. In addition, this paper considers how the step of countdown and practical settings of timeout values and distribution of sensing data affect the obtained performance.

The rest of the paper is organized as follows. Sec. II describes the related work. The system model is presented in Sec. III. In Sec. IV, the proposed content–based wake–up scheme is presented. In Sec. V, the performance of the conventional scheme and the proposed wake–up scheme is theoretically analyzed in terms of data collection delay and total energy consumption. Numerical results are provided in Sec. VI, followed by a conclusion in Sec. VII.

II. Related Work

The basic approach for improving the energy–efficiency of wireless networks is duty–cycling [10] 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 [11]. 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 [4][12], 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 [12] treat hardware design, routing and MAC protocols; the latter is the focus of this paper.

The MAC protocols with wake–up signaling can be transmitter–initiated (TI) or receiver–initiated (RI). In TI protocols [13], 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 [4] and its performance is evaluated in [5]. 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], the 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].

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 [17][18]. For top–k query in WSNs [19][20][21], the main challenge is to reduce the communication overhead by avoiding transmissions that are unnecessary for constructing the top–k data set. [22] 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) [23], exact top–k (EXTOK) [7], and their enhanced versions, (e.g., [24], [25], [26]), 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 is temporally correlated. Specifically, EXTOK [7] deals with the exact top–k query similar to the value–set query considered in this paper. However, 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 [27][28]. Priority–based top–k monitoring (PRIM) [29] 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 [8] and the references therein. More recent studies on top–k query in WSNs consider the privacy and integrity issues [30][31], the problems related to node mobility [32][33], compression and correctness [34], and extensions to multiattribute query [35].

To the best of our knowledge, our previous research work is the first to address the problem of top–k query with wake–up signalling [1]. Note that a wake–up process activating nodes according to their sensed data, which can be considered

[1]Originally, SDMAC was developed for cellular networks, for which licensed/dedicated frequency bands are commonly allocated. In contrast, this work focuses on sensor networks operating over unlicensed frequency bands, where the interference from different systems is inevitable. Since more elaborations and considerations are required for the integration of SDMAC with wake–up signaling over unlicensed bands as discussed in Sec. VI-E, we keep the fair comparison between CDCoWu and SDMAC outside of the scope of this paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

as content–based wake–up, was used for clustering in [36]. However, [36] is limited to clustering of nodes with similar readings and is thus not directly applicable to the considered scenario of top–k data collection.

III. System Model

This paper 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. 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 has both CSMA and CA mechanisms, which is suited for the approximation of the practical MAC protocol following IEEE 802.15.4 standard [37]. 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]. This is a way to extend the model in [1] and account for errors, while not explicitly including a propagation model for the attenuation of transmitted signals. When a packet is lost, each node detects the absence of acknowledgement (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 [7][8]. The top–k query has been considered to be employed for a large number of IoT applications, e.g., environment/infrastructure monitoring [7][30], smart-city [38], and network management [23]. 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 [7][39][40], this paper assumes 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.

This paper considers the wake–up control scheme, where the sink node does not require an additional hardware as it sends the wake–up sequence by reusing the main radio. The wake–up receiver uses a low power scheme based on non–coherent envelope detection and on–off–keying (OOK) demodulation, such that it can interpret the frame length, but not its header/payload. The configuration of this type of wake–up receiver is shown in Fig. 1 [41]. 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, micro controller unit (MCU) calculates the length of observed frame. 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 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 [12]. The conventional wake–up signaling is the ID–based wake–up (IDWu). For instance, a mapping can be made between different wake–up IDs and frame lengths, which is shared by all nodes [6]. For IDWu, this paper considers two types of IDs: Broadcast wake–up ID (BCWuID) and Unicast wake–up ID (UCWuID). BCWuID is a common ID to all sensor nodes, which implements BCWu and triggers all nodes to wake up. UCWuID is unique to each node, and the sink can specify a single target node with a specific ID for UCWu. As the ID does not contain information about the sensing data, it is

[2]Although this paper employs 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

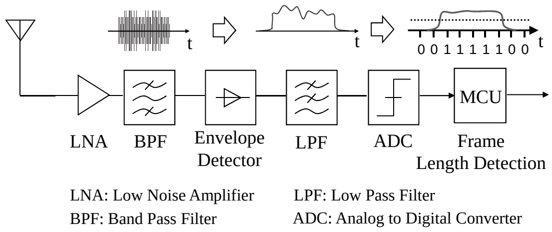

Fig. 1. An example of configuration of wake–up receiver to detect frame length.

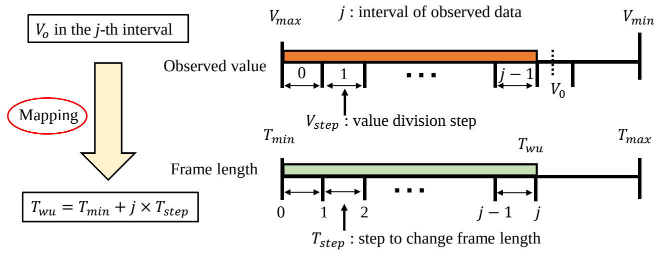

Fig. 2. Mechanism to decide wake–up frame length in CoWu.

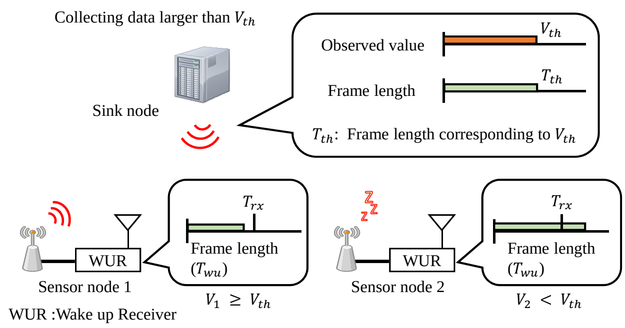

Fig. 3. An example of CoWu operations.

possible that IDWu scheme will wake up nodes that are not contributing to the top–k query, resulting in energy waste. In order to address this issue, CoWu was proposed in [1], which is discussed in Sec. IV.

IV. Proposed Content–based Wake–up Control

In ID–based wake–up, 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.

A. 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. 2. 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 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 $T_{wu} < T_{rx}$, and thus wakes up, while sensor node 2 stays in a sleep mode.

B. Countdown Content–based Wake–up for Top–k Query

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. IV-A until the sink manages to collect the desired set of sensing data. The operation of the sink in the proposed CDCoWu depends on the considered query, which are called as node–set CDCoWu (N–CDCoWu) and value–set CDCoWu (V–CDCoWu).

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. III. 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 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 set is increased as follows. In N–CDCoWu, the sink increases the size of node–set by 1 whenever it

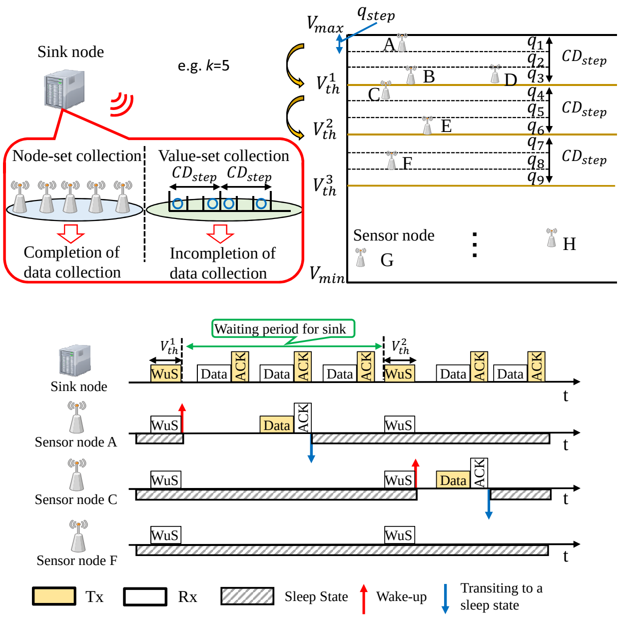

Fig. 4. An example of CDCoWu operations.

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 the collected set by 1 only if the received value is not already included in $\mathbb{V}$. 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. 4 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. 4, 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. VI.

V. Theoretical Analysis of IDWu and CDCoWu

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

A. 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 [37]. Under OSD model, each node enters a sleep state after the successful transmission of its packet and does not contend for the channel. In [37], 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

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

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

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

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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

B. 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 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 paper 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.

1) Analysis of BCWu scheme: 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

$$T_d^{BCWu}(N) = T_d(N) + T_{BCWu}, \tag{3}$$

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

$$E_{total}^{BCWu}(N) = E_{total}(N), \tag{4}$$

where $E_{total}(N)$ is given by eq. (2).

2) Analysis of UCWu scheme: 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

$$\begin{aligned} T_d^{UCWu}(N) &= N{\cdot}T_d(1) + \sum_{i=0}^{N-1} T_{UCWu}(i) \\ &= N{\cdot}T_d(1) + T_{min}{\cdot}N + \frac{T_{step}}{2}N(N-1), \end{aligned} \tag{5}$$

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

$$E_{total}^{UCWu}(N) = N{\cdot}E_{total}(1). \tag{6}$$

[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.

3) Analysis of CDCoWu scheme: As described in Sec. IV-B, 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 [42]

$$\begin{aligned} &\mathbb{P}(X_1 = q_1, \ldots, X_{I_{max}} = q_{I_{max}}) \\ &= \begin{cases} N! \prod_{n=1}^{I_{max}} \frac{P_{int(n)}^{q_n}}{q_n!} & (\sum_{n=1}^{I_{max}} q_n = N) \\ 0 & \text{(otherwise)}, \end{cases} \end{aligned} \tag{7}$$

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

$$P_{int(n)} = \begin{cases} \frac{q_{step}}{V_{max}-V_{min}} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{(V_{max}-V_{min})-(I_{max}-1)q_{step}}{V_{max}-V_{min}} & (n = I_{max}). \end{cases} \tag{8}$$

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

$$P_{int(n)} = \begin{cases} \frac{\int_{V_{max}-nq_{step}}^{V_{max}-(n-1)q_{step}} \mathrm{e}^{\alpha x} dx}{\int_{V_{min}}^{V_{max}} \mathrm{e}^{\alpha x} dx} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{\int_{V_{min}}^{V_{max}-(I_{max}-1)q_{step}} \mathrm{e}^{\alpha x} dx}{\int_{V_{min}}^{V_{max}} \mathrm{e}^{\alpha x} dx} & (n = I_{max}). \end{cases} \tag{9}$$

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

$$P_{int(n)} = \begin{cases} \frac{\int_{V_{max}-nq_{step}}^{V_{max}-(n-1)q_{step}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{\int_{V_{min}}^{V_{max}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx} & (n = 1, 2, \ldots, I_{max}-1) \\ \frac{\int_{V_{min}}^{V_{max}-(I_{max}-1)q_{step}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{\int_{V_{min}}^{V_{max}} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx} & (n = I_{max}). \end{cases} \tag{10}$$

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

$$x_\zeta^i = \sum_{l=m'(\zeta-1)+1}^{m'\zeta} q_l^i. \tag{11}$$

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. (1) and (7), mean total duration required for data transmissions with p–persistent CSMA (excluding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

periods of wake–up signal transmissions) can be calculated as

$$T_{CDCoWu}^{Data}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \sum_{\zeta=1}^{n_w^i} T_d(x_\zeta^i), \quad (12)$$

where M is the total number of realizations of multinomial distribution given in eq. (7). Note that n_w^i differs for N–CDCoWu or V–CDCoWu.

As mentioned in Sec. IV-B, 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

$$T_{CDCoWu}^{WuS}(\zeta) = T_{min} + T_{step}(m\zeta - 1). \quad (13)$$

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

$$T_{CDCoWu}^{\Sigma WuS}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \sum_{\zeta=1}^{n_w^i} T_{CDCoWu}^{WuS}(\zeta). \quad (14)$$

Thus, mean total delay to complete the collection of top–k observations with CDCoWu is the sum of eqs. (12) and (14)

$$T_d^{CDCoWu}(N) = T_{CDCoWu}^{Data}(N) + T_{CDCoWu}^{\Sigma WuS}(N). \quad (15)$$

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

$$E_{total}^{CDCoWu}(N) = \sum_{i=1}^{M} \mathbb{P}(s_i) \cdot \sum_{\zeta=1}^{n_w^i} E_{total}(x_\zeta^i). \quad (16)$$

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

$$W_{max} = \lceil \frac{V_{max} - V_{min}}{CD_{step}} \rceil. \quad (17)$$

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

$$\sum_{\zeta=1}^{j} x_\zeta^i \geq k. \quad (18)$$

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

$$y_\zeta^i = \sum_{l=m'(\zeta-1)+1}^{m'\zeta} u(q_l^i - 1), \quad (19)$$

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

$$u(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (otherwise). \end{cases} \quad (20)$$

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

$$\begin{cases} \sum_{\zeta=1}^{j} y_\zeta^i \; \geq \; k & (if \; \exists \; s.t. \sum_{\zeta=1}^{W_{max}} y_\zeta^i \; \geq \; k) \\ \sum_{\zeta=1}^{j} x_\zeta^i \; = \; N & (otherwise). \end{cases} \quad (21)$$

VI. Numerical Results and Discussions

A. Approximate analysis with MCMC

The exact calculation of mean delay and energy consumption of CDCoWu, see eqs. (12), (14), and (16), requires consideration of all realizations of multinomial distribution given in eq. (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 paper resorts to Markov Chain Monte Carlo (MCMC) method [43] to obtain approximate results. Specifically, this work employs Metropolis algorithm [43] 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. (7).[4]
- STEP 3: [Update state] Generate a random number $r \in [0, 1]$, following uniform distribution, and decide the next state as follows:

$$X^{(t+1)} = \begin{cases} X' & (if \; r \leq \frac{\mathbb{P}(X')}{\mathbb{P}(X^{(t)})}) \\ X^{(t)} & (\text{otherwise}). \end{cases} \quad (22)$$

- 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. V, 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

$$\widehat{T}_{delay}^{CDCoWu} = \frac{1}{Z} \sum_{z=1}^{Z} t^{(z)} \;, \quad (23)$$

$$\widehat{E}_{total}^{CDCoWu} = \frac{1}{Z} \sum_{z=1}^{Z} e^{(z)} \;. \quad (24)$$

[4]More details on transition cost can be found in [43].

A unique feature of Metropolis method is that each sample to calculate the expectations of function in eqs. (12), (14), and (16) is selected from the 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.

B. 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. 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. III as well as IDWu and CDCoWu explained in Sec. V 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 I. Further, the values of l (see Sec. V-B3) 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 [41][44]. 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 [44]. 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.

1) Impact of CD_{step}: Figs. 5 and 6 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 [37]. Obviously, the figures show that the results obtained with the approximate analysis coincide with simulation results very well, validating the approach employed in this paper. Fig. 5 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. V. For $CD_{step} < CD_{step}^{opt}$, x_ζ^i in eq. (11) or y_ζ^i in eq. (19) becomes smaller, by which n_w^i satisfying eq. (18) or (21) becomes larger. This makes the duration required for

TABLE I
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 [45]
Power consumption in Receive state ξ_R	50 mW [45]
Distribution of observed value $[V_{min}, V_{max}]$	$[0, 50]$
Interval of wake–up frame length T_{step}	0.16 msec [44]
Minimum wake–up frame length T_{min}	10.8 msec [44]
Wake–up Frame length corresponding to BCWuID T_{BCWu}	10.8 msec

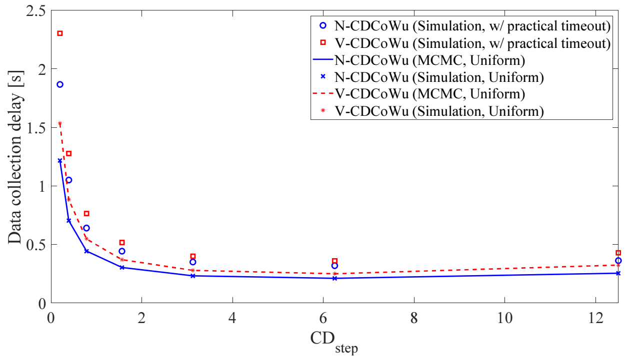

Fig. 5. Data collection delay against CD_{step} for N–CDCoWu and V–CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$).

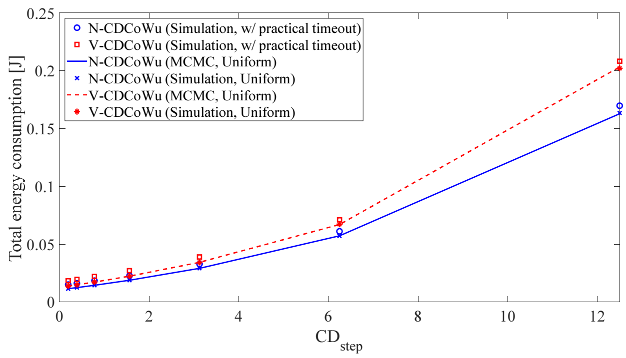

Fig. 6. Total energy consumption against CD_{step} for N–CDCoWu and V–CDCoWu ($N = 100$, $k = 25$, $b_q = 8$, and $p = 0.0606$).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

transmitting wake–up signals in eq. (14) larger. While this negative effect is alleviated as the value of CD_{step} increases, the delay for transmitting data expressed in eq. (12) becomes larger since x^i_ζ in $T_d(x^i_\zeta)$, which is a monotonically increasing function against x^i_ζ as given in eq. (1), increases. Due to this negative effect with increasing CD_{step}, data collection delay starts to increase again. On the other hand, Fig. 6 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. 5 and 6 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. 5 and 6, 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 time 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 [37]. 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 [37]. 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^{opt}_{step}, 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.

2) 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 5 to 10, and the performance of data collection delay and total energy consumption of N–CDCoWu and V–CDCoWu is evaluated. Figs. 7 and 8 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. 7 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. 8 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. V-B3, 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.VI-B. 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. 7 and total energy consumption in Fig. 8 decrease when the value of b_q changes from 9 to 10.

Finally, Figs. 7 and 8 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].

3) Impact of distribution of observed values: Here, the impact of distribution of observed values on delay and energy

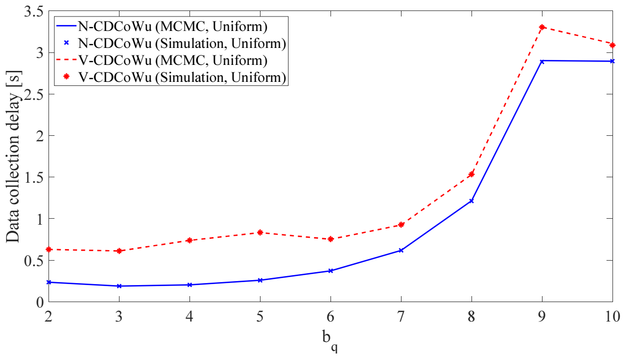

Fig. 7. Data collection delay against b_q for N–CDCoWu and V–CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$).

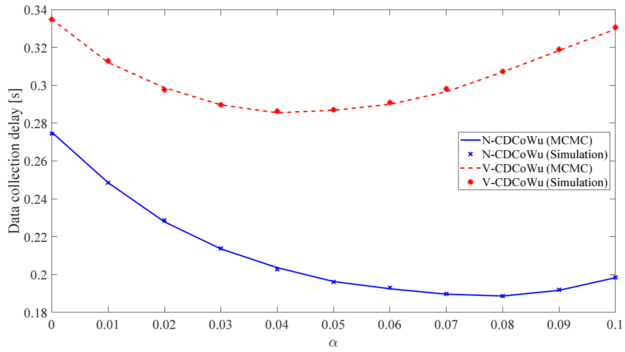

Fig. 9. Data Collection delay against α for N–CDCoWu and V–CDCoWu ($N = 100$, $k = 25$, $CD_{step} = 10V_{step}$, and $p = 0.0606$).

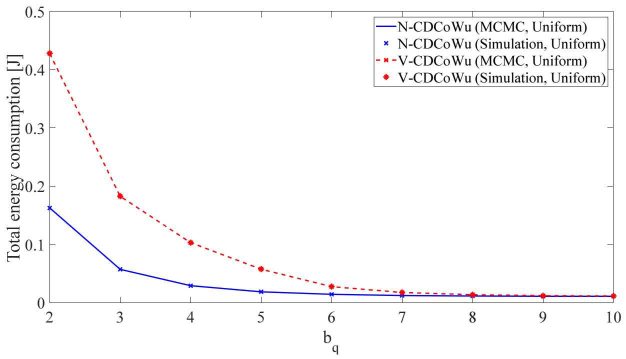

Fig. 8. Total energy consumption against b_q for N–CDCoWu and V–CDCoWu ($N = 100$, $k = 25$, $CD_{step} = V_{step}$, and $p = 0.0606$).

consumption of CDCoWu is evaluated by taking exponential distribution as an example, where the value of α in eq. (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. 9 shows data collection delay of N–CDCoWu and V–CDCoWu against α. From Fig. 9, 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. V, as similarly explained for the impact of CD_{step} in Fig. 5. As the value of α becomes larger, from eq. (9), it can be seen that the probability that each node observes higher values becomes higher, which increases x^i_ζ against the smaller ζ. This makes n^i_w satisfying eq. (11) or (19) smaller, which makes the duration required for transmitting wake–up signals in eq. (14) smaller. On the other hand, the delay for transmitting data given in eq. (12) becomes larger since x^i_ζ in eq. (12) for smaller ζ can become sufficiently large to cause severe congestion. Therefore, depending on whether delay for data transmissions in eq. (12) or that for transmitting wake–up signals in eq. (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.

4) Impact of k: Figs. 10 and 11 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 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 wireless sensor networks 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 [20] and forest fire monitoring [8] 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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

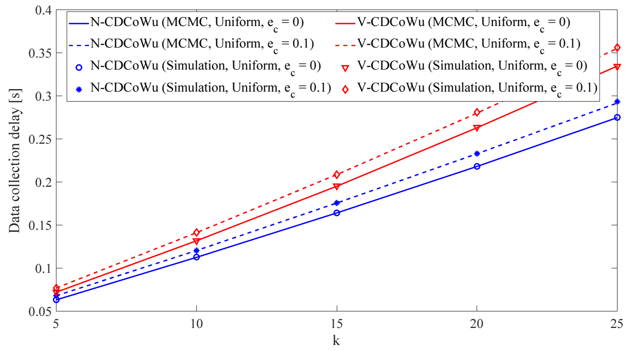

Fig. 10. Data collection delay against k for N–CDCoWu and V–CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$).

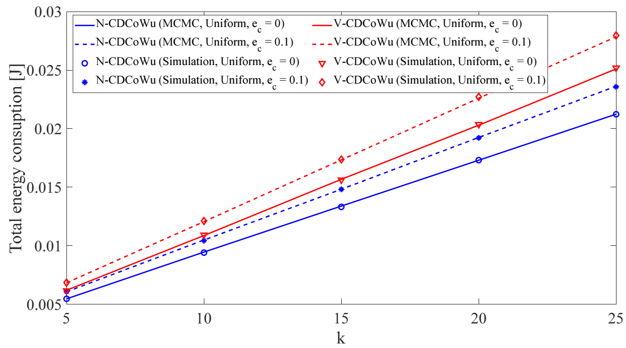

Fig. 11. Total energy consumption against k for N–CDCoWu and V–CDCoWu ($N = 100$, $CD_{step} = 10V_{step}$, $b_q = 8$, and $p = 0.0606$).

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).

C. 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 high energy efficiency at the cost of delay by setting CD_{step} to small value, as confirmed by the evaluations in Sec. VI-B. 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.[5] 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 $\tau^{opt}_{BCWu}(N)$.[5] 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

$$\mathbb{U}(N) = \min_{CD_{step},p} \{E^{CDCoWu}_{total}(N)\} \; s.t. \; \tau \leq \tau^{opt}_{UCWu}(N) \;, \quad (25)$$

$$\mathbb{B}(N) = \min_{CD_{step},p} \{E^{CDCoWu}_{total}(N)\} \; s.t. \; \tau \leq \tau^{opt}_{BCWu}(N) \;, \quad (26)$$

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

Note that this paper 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.

Figs. 12 and 13 show the comparison of total energy consumption of CDCoWu with BCWu and with UCWu, respectively, when they are applied for top–10 query. 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. 12, 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. IV-A 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

[5]Due to lack of space, this paper does not show results corresponding to this optimization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

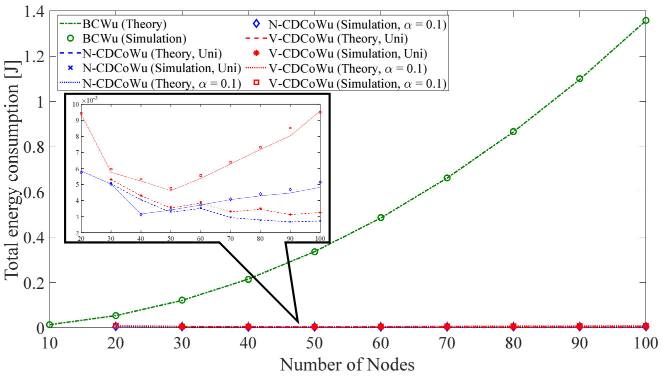

Fig. 12. Total energy consumption against the number of sensor nodes for BCWu and CDCoWu.

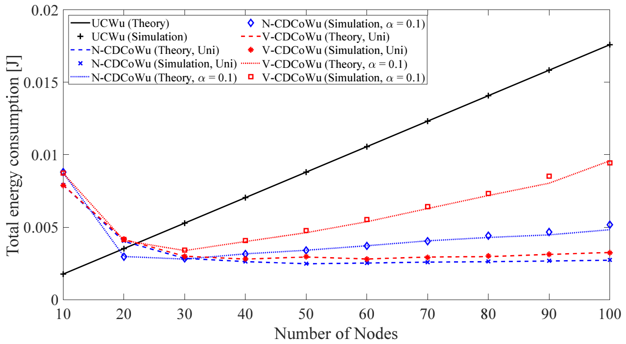

Fig. 13. Total energy consumption against the number of sensor nodes for UCWu and CDCoWu.

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 than that of BCWu. That is why CDCoWu can not satisfy the delay constraint of eq. (26), when the number of nodes is small.

Fig. 12 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 number 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. 12 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. 12 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. 12, 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. 13 shows that CDCoWu achieves smaller energy consumption than UCWu when the number of sensor nodes is equal to or more than 30. In [1], it was shown that 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. (25), to make the impact of post wake–up congestion less significant. However, as shown in Fig. 13, total energy consumption for $N = 10$ is larger than that for $N = 20$. This is because the constraint on delay described in eq. (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 [46]. Due to lack of space, this paper does not show results, but 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.

D. Impact of k and N

Figs. 12 and 13 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 paper provides insights on the percentage of top–data for a given N that motivates use of CDCoWu instead

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

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.,

$$\mathbb{U}_k(N) = \{\frac{k}{N} \mid \epsilon \leq \epsilon^{opt}_{UCWu}(N)\} \ , \qquad (27)$$

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. 14 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 [47]. The declining tendency of maximum k–N ratio for V–CDCoWu with exponential distribution in Fig. 14 is mainly due to the constraint on total energy consumption. As discussed in Sec. VI-B3, 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. 14. Furthermore, from Fig. 14, 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 is kept for future work. However, from Fig. 14, 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.

Fig. 15 plots maximum k–N ratio for UCWu against the number of nodes when the probability of packet loss due to channel impairments, e_c, is set to 0.1. The employed parameters are the same as Fig. 14. By comparing Figs. 15 and 14, it can be seen that the introduction of packet loss due to channel impairments does not cause significant change for achievable maximum k–N ratio. For CDCoWu, each node needs to spend time and energy for contention resolution to retransmit a lost packet. However, the increased delay of UCWu caused by packet losses allows CDCoWu to employ small CD_{step}, which reduces the negative impact of congestion. Overall, the proposed scheme achieves the same range of maximum k–N ratio as Fig. 14.

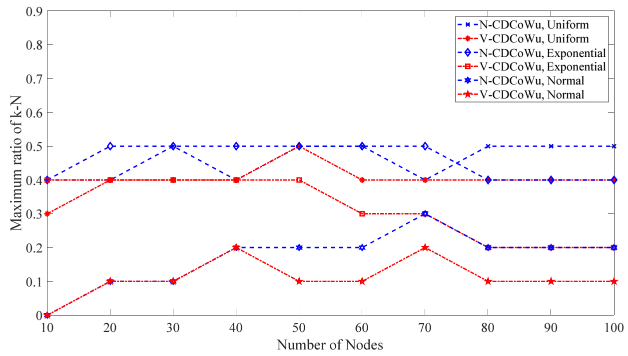

Fig. 14. Maximum k–N ratio of CDCoWu to UCWu in terms of data collection delay and total energy consumption.

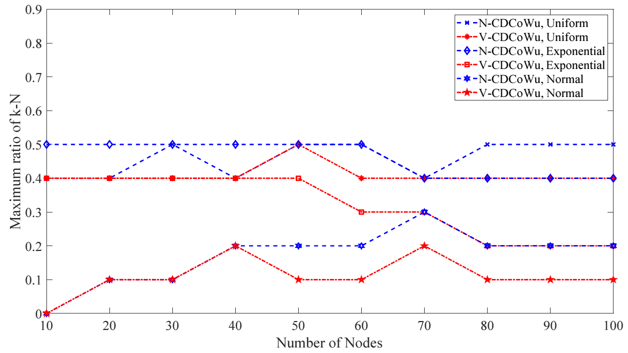

Fig. 15. Maximum k–N ratio of CDCoWu to UCWu in terms of data collection delay and total energy consumption with $e_c = 0.1$.

E. Discussions

This paper has considered IDWu based on UCWu/BCWu as reference wake–up schemes for comparison with CDCoWu. All of these schemes are designed such that the activated nodes can attempt to transmit packets at arbitrary timing based on CSMA protocol. On the other hand, wake–up control with scheduled transmission of packets can be also designed, e.g., by integrating wake–up control with scheduled data MAC (SDMAC) [9], which is called as Wu–SDMAC here, as follows:

- A wake–up signal to wake up all nodes is transmitted by the sink, whose length is set to be T_{min}.
- The timing for each node to wake up after detecting the wake–up signal is scheduled based on time division multiple access (TDMA)–like policy, i.e., the node ID j ($j = 0, 1, \cdots$) wakes up in jT_{sch} [s] after receiving the wake–up signal and transmits data.
- After completing data transmissions, each node transits to a sleep state.

The Wu–SDMAC involves only a single transmission of wake–up signal while avoiding contentions by nodes, which includes the advantages of BCWu and UCWu. Furthermore, it requires neither variable frame length for wake–up signaling nor optimizations of parameters such as CD_{step}, V_{step} and practical timer to detect the absence of replies from activated

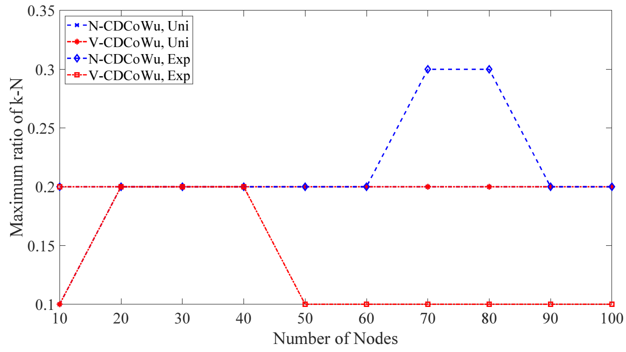

Fig. 16. Maximum k–N ratio of CDCoWu to Wu–SDMAC in terms of data collection delay and total energy consumption.

nodes, resulting in lower complexity than BCWu/UCWu. However, the period for each node to attempt packet transmission, i.e., T_{sch}, should be carefully designed such that it can accommodate the delay caused not only by the contention with possible interferers (e.g., devices sharing the same unlicensed frequency band), but also by retransmissions required when packets are lost due to channel impairments. The fair comparison between CDCoWu and Wu–SDMAC requires us to optimize T_{sch} considering possible congestion level over the shared channel and packet losses due to channel impairments, which is out of scope of this paper. However, when there is no possible interference (e.g., the operation over the dedicated, licensed frequency band) and no packet loss, T_{sch} can be set to $L\delta$, which is the period to accommodate a single packet. As an evaluation in this particular setting, the maximum k–N ratio for Wu–SDMAC is evaluated here. The equations to give data collection delay and total energy consumptions can be easily obtained by extending those derived for BCWu and UCWu in Sec. V.

Fig. 16 shows maximum k–N ratio for Wu–SDMAC against the number of nodes, where uniform and exponential distribution with $\alpha = 0.1$ are employed for observed values. Due to the difficulty to control trade–off between delay and energy consumption described in Sec. VI-D, CDCoWu does not outperform Wu–SDMAC for normal distribution, therefore, only results for uniform and exponential distributions are plotted. The main parameters employed for evaluations are the same as Fig. 14. From this figure, it can be seen that maximum k–N ratio realizing superiority of CDCoWu to Wu–SDMAC is decreased in comparison to that to UCWu. As the wake–up overhead of Wu–SDMAC is negligible, Wu–SDMAC achieves smaller data collection delay than UCWu. From Fig. 16, it can be seen that the proposed CDCoWu shows the superiority to Wu–SDMAC for k–N ratio ≤ 0.1, reaching 0.3 at maximum for these two types of distributions. The Wu–SDMAC improves performance after wake–up process while the proposed CDCoWu solves unnecessary wake–up of nodes irrelevant to top–k data set at the cost of contentions after the wake–up process. The thorough and fair comparison between CDCoWu and Wu–SDMAC in more general settings is kept for our future work.

VII. Conclusions

This paper investigated how wasteful wake–ups in WSNs exploiting wake–up receivers can be reduced when top–k node/value set collections are used. The enhanced algorithms for top–k node–set/value–set query employing CoWu were proposed, called N–CDCoWu and V–CDCoWu. The equations expressing data collection delay and total energy consumption of different wake–up schemes were derived in order to analyze the effectiveness of the proposed schemes, assuming the employment of p–persistent CSMA as MAC protocol. In order to obtain numerical results on the proposed CDCoWu, an MCMC method was used. Through comparison between theoretical and simulation results, the validity was confirmed for the derived equations and analysis based on the MCMC method. The performance of CDCoWu was thoroughly investigated over a wide variety and range of parameters, and valid insights related to its comparison to the performance of IDWu schemes were obtained. The obtained numerical results showed that the proposed CDCoWu outperforms IDWu in terms of data collection delay and energy consumption for the maximum ratio of top–data to number of sensor nodes ranging from 0.1 to 0.5 when the number of sensor nodes is equal to or more than 20.

The future work includes the investigation of the effectiveness of adapting of CD_{step} in accordance with the distribution of observed data of sensor nodes, as well as considerations on more realistic network and channel models with location–dependent channel impairments. The thorough and fair comparison between CDCoWu and Wu–SDMAC, which was outside of the scope of this paper, is also a topic for future work.

References

[1] J. Shiraishi, H. Yomo, K. Huang, C. Stefanovic, and P. Popovski, "Content-based Wake-up Control for Wireless Sensor Networks Exploiting Wake-up Receivers," *Proc. of The 17th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt 2019)*, pp. 1–8, June 2019.

[2] J. A. Stankovic, "Research Directions for the Internet of Things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, March 2014.

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

[4] 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.

[5] 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.

[6] H. Yomo, K. Abe, Y. Ezure, T. Ito, A. Hasegawa, and T. Ikenaga, "Radio-On-Demand Sensor and Actuator Networks (ROD-SAN): System Design and Field Trial," *Proc. of Globecom 2015*, Dec. 2015.

[7] 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.

[8] 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.

[9] S. Garlapati, T. Kuruganti, M. R. Buehrer, and J. H. Reed, "Smac: A soft mac to reduce control overhead and latency in cdma-based ami networks," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2648–2662, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGCN.2020.3033844, IEEE Transactions on Green Communications and Networking

[10] R. C. Carrano, D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, "Survey and Taxonomy of Duty Cycling Mechanisms in Wireless Sensor Networks," *IEEE Communicaitons Surveys & Tutorials*, vol. 16, no. 1, pp. 181–194, Feb. 2014.

[11] 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.

[12] 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, Fourthquarter 2017.

[13] 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.

[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] 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.

[18] 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.

[19] 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.

[20] 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.

[21] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov, "Top–k Query Based Dynamic Scheduling for IoT-enabled Smart City Waste Collection," *Proc. of 2015 16th IEEE International Conference on Mobile Data Management*, vol. 2, pp. 50–55, June 2015.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] J. Blanckenstein, J. Klaue, and H. Karl, "Energy efficient clustering using a wake-up receiver," *Proc of 18th European Wireless Conference 2012*, pp. 1–8, April 2012.

[37] 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 (WiOpt08)*, vol. 4, pp. 152–161, April 2008.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

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

[43] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[44] H. Yomo *et al.*, "ROD-SAN: Energy-efficient and High-response Wireless Sensor and Actuator Networks employing Wake-up Receiver," *IEICE Transactions on Communications*, vol. E99-B, no. 9, pp. 1998–2008, Sept. 2016.

[45] N. Tamura and H. Yomo, "Low-overhead Wake-up Control for Wireless Sensor Networks employing Wake-up Receivers," *IEICE Transactions on Communications*, vol. E102-B, no. 4, pp. 732–740, April 2019.

[46] 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.

[47] 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.

PLACE PHOTO HERE

Junya Shiraishi received B.E. and M.E. degrees in Electrical, Electronic and Information engineering from Kansai University, Japan, in 2018 and 2020, respectively. He is now pursuing a Ph.D. degree in the Graduate School of Science and Engineering, Kansai University. His current research interests include energy-efficient PHY/MAC protocol designs for wireless sensor networks, data collection with content-based wake-up exploiting the nature of the spatio-temporal correlation of physical phenomenon, and theoretical analysis of the characterization of wake-up controls.

PLACE PHOTO HERE

Hiroyuki Yomo received B.S., M.S., and Ph.D. degrees in communication engineering from Osaka University, Japan, in 1997, 1999 and 2002, respectively. From April 2002 to March 2004, he was a Post-doctoral Fellow at Aalborg University, Denmark. From April 2004 to September 2004, he was at NEC Corporation, Japan. In October 2004, he joined Aalborg University, Denmark, as an Assistant Research Professor, and worked as an Associate Professor from February 2006 to March 2008. From April 2008 to March 2010, he was a senior researcher at ATR, Japan. In April 2010, he joined Kansai University, Japan, as an Associate Professor, and has been a Professor since April 2015. He is also affiliated with ATR, Japan, as a guest researcher. He received the 2010 Funai Academic Award from Funai Foundation for Information Technology. He has received several best paper awards, including the IEICE 2016 best paper award. His main research interests are access technologies, radio resource management, and link-layer techniques in the broad area of wireless communications. He is an editor of IEEE Wireless Communications Letters.

PLACE PHOTO HERE

Kaibin Huang (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from the National University of Singapore, and the Ph.D. degree from The University of Texas at Austin, all in electrical engineering. Presently, he is an associate professor in the Dept. of Electrical and Electronic Engineering at The University of Hong Kong. He received the IEEE Communication Society's 2019 Best Tutorial Paper Award, 2015 Asia Pacific Best Paper Award, and 2019 Asia Pacific Outstanding Paper Award as well as Best Paper Awards from IEEE GLOBECOM 2006 and IEEE/CIC ICCC 2018. Moreover, he received an Outstanding Teaching Award from Yonsei University in S. Korea in 2011. He has served as the lead chairs for the Wireless Comm. Symp. of IEEE Globecom 2017 and the Comm. Theory Symp. of IEEE GLOBECOM 2014 and the TPC Co-chairs for IEEE PIMRC 2017 and IEEE CTW 2013. He has edited special issues for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING, and IEEE Communications Magazine. He is/was an Associate Editor for several major journals in wireless including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is a Distinguished Lecturer of IEEE Vehicular Technology Society and an ISI Highly Cited Researcher.

PLACE PHOTO HERE

Čedomir Stefanović (S'04–M'11–SM'17) obtained his Dipl.-Ing., Mr.-Ing., and PhD degrees from the University of Novi Sad, Serbia. Since 2012, he has been affiliated with the Department of Electronic Systems, Aalborg University, where he currently holds the position of associate professor. He is a principal researcher on a number of European projects related to IoT, 5G and mission-critical communications. He co-authored more than 100 peer-reviewed publications. His research interests include communication theory and wireless communications. He serves as an editor for the IEEE IoT Journal.

PLACE PHOTO HERE

Petar Popovski (S'97–A'98–M'04–SM'10–F'16) is a Professor at Aalborg University, where he is heading the section on Connectivity. He received his Dipl. Ing and M. Sc. degrees in communication engineering from the University of Sts. Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005. He is a Fellow of IEEE, has over 400 publications in journals, conference proceedings, and edited books and he was featured in the list of Highly Cited Researchers 2018, compiled by Web of Science. He holds over 30 patents and patent applications. He received an ERC Consolidator Grant (2015), the Danish Elite Researcher award (2016), IEEE Fred W. Ellersick prize (2016), IEEE Stephen O. Rice prize (2018) and the Technical Achievement Award from the IEEE Technical Committee on Smart Grid Communications. He is currently a Steering Committee Member of IEEE SmartGridComm and IEEE Transactions on Green Communications and Networking. He previously served as a Steering Committee Member of the IEEE INTERNET OF THINGS JOURNAL. He is currently an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Prof. Popovski was the General Chair for IEEE SmartGridComm 2018 and IEEE Communication Theory Workshop 2019. From 2019, he is also a Member-at-Large of the Board of Governors of the IEEE Communications Society. His research interests are in the area of wireless communication, communication theory and Internet of Things. In 2020 he published the book "Wireless Connectivity: An Intuitive and Fundamental Guide".