SelectCast: Scalable Data Aggregation Scheme in Wireless Sensor Networks

Cheng Wang, Changjun Jiang, *Member*, *IEEE*, Shaojie Tang, and Xiang-Yang Li, *Senior Member*, *IEEE*

Abstract—In this work, for a wireless sensor network (WSN) of n randomly placed sensors with node density $\lambda \in [1, n]$, we study the tradeoffs between the *aggregation throughput* and *gathering efficiency*. The gathering efficiency refers to the ratio of the number of the sensors whose data have been gathered to the total number of sensors. Specifically, we design two efficient aggregation schemes, called *single-hop-length* (SHL) scheme and *multiple-hop-length* (MHL) scheme. By novelly integrating these two schemes, we theoretically prove that our protocol achieves the optimal tradeoffs, and derive the optimal aggregation throughput depending on a given threshold value (lower bound) on gathering efficiency. Particularly, we show that under the MHL scheme, for a practically important set of symmetric functions called *divisible perfectly compressible* (DPC) functions, including the mean, max, and various kinds of indicator functions, etc., the data from $\Theta(n)$ sensors can be aggregated to the sink at the throughput of a constant order $\Theta(1)$, implying that, our MHL scheme is indeed scalable.

Index Terms—Wireless sensor networks, data aggregation, percolation theory, aggregation capacity.

1 Introduction

Data aggregation is a key energy saving functionality in wireless sensor networks (WSNs) for both data gathering applications and event-based applications, since the communication cost is often the higher order of the computation cost [1], [2], [3]. Various data aggregation schemes have been proposed, e.g., cluster-based structures [4], [5], [6] and tree-based structures [7], [8], [9]. In data gathering applications, such as environment and habitat monitoring [10], [11], [12], sensors periodically send the sensed data to the sink. When the traffic pattern and network topology are assumed to be invariable, the *structure-based* [7], [8] methods need low-maintenance overhead and are thus applicable for such application scenarios.

It has been shown in the literature that the achievable minimum data rate among all sensor nodes is severely limited for random WSNs if we insist data from *all* sensors should be collected. In this paper, we design structure-based aggregation schemes for WSNs to achieve the optimal tradeoffs between the *aggregation throughput* and *gathering efficiency*. Here, the gathering efficiency refers to the ratio of the number of the sensor nodes whose data are gathered successfully to the total number of sensor nodes in the network. Collecting data from a subset of sensor nodes is reasonable because of the potential spatial correlations among sensed environment. In our protocol, for the neighborhood of every node, we will approximately select Ψ portion of nodes and aggregate their data to the sink. Such a sampling scheme will achieve high aggregation throughput while maintaining the spatial coverage by the sampled sensors.

For data gathering, we focus on an important set of symmetric functions called *divisible perfectly compressible* (DPC) functions, such as the mean, max, and kinds of indicator functions [13] that will be used to compute the data aggregation. Two characteristics of this work are extracted as following:

- To meet specific application requirement, e.g., full coverage, k-coverage, connectivity, etc., the node density (number of nodes per unit area) can be treated as a varible within a large range. Thus, we consider a random deployed WSN with a general density, where n sensors constitute a network with node density λ, $1 \leq \lambda \leq n$, rather than the special *random dense networks* [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] or *random extended networks* [20], [21], [23], [24], where $\lambda = n$ and $\lambda = 1$, respectively. Depending on the requirement of gathering efficiency, we determine the thresholds of the density λ by which the aggregation throughput and tradeoffs are divided into different regimes.
- As the node density is decreasing and the area of deployment region is thus increasing, we need to rely on some long links to ensure the network connectivity. For those long links, it is unrealistic to set the link rate to be a constant order as under the *protocol model* and *physical model* [14]. Hence, we design efficient protocols under a more realistic

- *C. Wang and C. Jiang are with the Department of Computer Science and Engineering, Tongji University, and the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Building of Electronics and Information Engineering, NO. 4800, Caoan Road, Shanghai 201804, China. E-mail: 3chengwang@gmail.com, cjjiang@tongji.edu.cn.*
- *S. Tang is with the Department of Computer Science, Illinois Institute of Technology, 10, West 31st Street, Chicago, IL 60616. E-mail: stang7@iit.edu.*
- *X.-Y. Li is with the Department of Computer Science and Engineering, Tongji University, and the Tsinghua National Laboratory for Information Science and Technology (TNLIST), Tsinghua University, and the Department of Computer Science, Illinois Institute of Technology, 10, West 31st Street, Chicago, IL 60616. E-mail: xli@cs.iit.edu.*

Manuscript received 3 Aug. 2011; revised 3 Dec. 2011; accepted 19 Dec. 2011; published online 30 Dec. 2011.
Recommended for acceptance by K. Li.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2011-08-0510.
Digital Object Identifier no. 10.1109/TPDS.2011.312.

communication model called *generalized physical model* [20], [22], [24], [25], rather than under the protocol model [15], [17] or physical model [13].

Under the structure-based aggregation schemes for a random WSN, the aggregation throughput for a specific type of function is mainly limited by the following two factors:

1 **Outliers**: In random networks, given a proper threshold (upper bound) on the length of links, there is a giant *connected component* in which any pair of nodes can be connected by the link of length below the threshold [26]. While, there might be some nodes, called *outliers*, [27], outside a specific connected component. To reach them, some links longer than the threshold are needed, which possibly leads to a lower link rate.
2 **Dense Components**: Given a deterministic routing, in the conflict graph modeling link interferences, there might be some cliques (complete subgraphs) of high-order size. Then, the scheduling of corresponding links might become a bottleneck.

To address these limitations and challenges, we design two efficient protocols to improve the tradeoffs between throughput and gathering efficiency.

- **Single-Hop-Length (SHL) Scheme**: The routing is nonhierarchical and consists of the links with similar lengths. By selecting a certain number of sensors in local regions depending on the given lower bound on gathering efficiency, we improve the aggregation throughput by deliberating the bottleneck produced by the second limitations, i.e., dense components.
- **Multiple-Hop-Length (MHL) Scheme**: The routing is hierarchical and consists of the links with various lengths. By selecting a fixed number of sensors from local regions and limiting the length of those long links, we improve the aggregation throughput by deliberating the bottleneck produced by both the outliers and dense components.

In summary, our main contributions are as follows:

1. Scalability is an important metric when designing the network protocol. We prove that under the MHL scheme, the measurements from $\Theta(n)$ sensors can be aggregated into the sink at the throughput of order $\Theta(1)$, which means that the MHL aggregation scheme is indeed scalable. To the best of our knowledge, our MHL scheme is the first scalable structure-based aggregation scheme.
2. Combining the schemes SHL and MHL, we derive the optimal tradeoffs between the aggregation throughput and gathering efficiency for DPC functions in the random WSN with general density λ, $1 \le \lambda \le n$. When we set the gathering efficiency to be 1 and the node density λ to be $\Theta(n)$, the resulted aggregation throughput is specified into the ordinary aggregation throughput for random dense WSNs [13], [15], [16], [17], [18], [19].
3. For some DPC functions, called *type-threshold DPC* functions, we introduce the block coding technique [17] into the local aggregation of SHL scheme, by which the throughput and tradeoffs are significantly improved. Then, combining with the MHL scheme, we derive the optimal throughput and tradeoffs as in Theorem 5.

The rest of the paper is organized as follows: In Section 2, we introduce the system model and formulate the problem. In Section 3, we propose two aggregation schemes for random WSNs with general density. We derive the achievable aggregation throughput and the tradeoffs between it and the gathering efficiency in Section 4. In Section 5, we derive the results by introducing block coding. In Section 6, we review the related work. In Section 7, we draw some conclusions and future perspective.

2 System Model

2.1 Network Model

We consider a random WSN, where sensors are deployed on the 2-dimension plane according to a Poisson point process of density λ with $\lambda \in [1, n]$. We focus on a square $\mathcal{A}(\lambda, n) = [0, \sqrt{n/\lambda}] \times [0, \sqrt{n/\lambda}]$. Then, according to Chebyshev's Inequality (Lemma A in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.312), the number of sensors in $\mathcal{A}(\lambda, n)$ is within $[(1-\varepsilon)\cdot n, (1+\varepsilon)\cdot n]$ with high probability. To simplify the description, we assume that the number of nodes is exactly n, without changing the final results in order sense. Denote $\mathcal{S}(n) = \{s_0\} \cup \{s_1, s_2, \ldots, s_{n-1}\}$, where s_0 is the sink node and $s_i, i \in [1, n-1]$ are the ordinary sensor nodes. We denote such a random WSN by $\mathcal{N}(\lambda, n)$.

Next, we examine the scaling characteristics of $\mathcal{N}(\lambda, n)$ as $n \to \infty$, according to the relation between λ and n. First, we recall an existing result about random euclidean minimal spanning tree (EMST) from [28]. Denote the length of the longest edge of EMST built on the set of n nodes in $\mathcal{N}(\lambda, n)$ by $\mathcal{L}(\lambda, n)$. According to the result in [28, (1)], by a simple scaling, we have

Lemma 1. *For the random variable $\mathcal{L}(\lambda, n)$, and for any real number $\nu(n)$, it holds that*

$$\lim_{n\to\infty} \Pr\Big(\pi \cdot \lambda \cdot (\mathcal{L}(\lambda, n))^2 - \log n \le \nu(n)\Big) = e^{-e^{-\nu(n)}}.$$

From Lemma 1, let $\nu(n) = -\ln\ln n$, we get that $\mathcal{L}(\lambda, n) = \Omega(\sqrt{\log n/\lambda})$ with high probability, e.g., $1 - \frac{1}{n}$; and let $\nu(n) = \ln n$, we obtain that $\mathcal{L}(\lambda, n) = O(\sqrt{\log n/\lambda})$ with high probability, e.g., at least $1 - \frac{1}{n}$. Hence, we have

$$\lim_{n\to\infty} \Pr\left(\mathcal{L}(\lambda, n) = \Theta\left(\sqrt{\log n/\lambda}\right)\right) \ge 1 - 1/n. \quad (1)$$

Now, we can define the criteria of the *extended scaling* versus *dense scaling* networks according to the order of the length of $\mathcal{L}(\lambda, n)$, i.e., $\Theta(\sqrt{\log n/\lambda})$.

Definition 1. *Given a random WSN $\mathcal{N}(\lambda, n)$, it is* dense scaling *if $\sqrt{\log n/\lambda} = O(1)$, i.e., $\lambda = \Omega(\log n)$; otherwise, it is* extended scaling.

Note that the cases of $\lambda = n$ and $\lambda = 1$ correspond to the well-known *random dense networks* [14], [20] and *random extended networks* [20], [24], respectively.

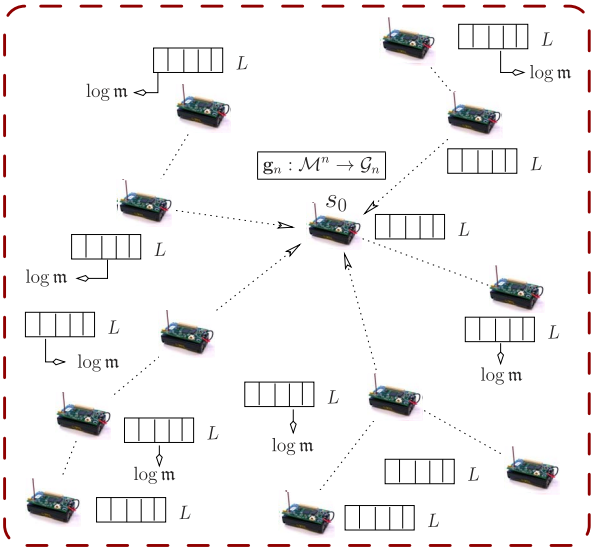


Fig. 1. System model. An *aggregation unit* consists of L measurements. Each measurement can be indicated by log m bits.

2.2 Communication/Interference Model

Generally, there are three widely used types of communication/interference models: the *protocol model* [14], *physical model* [14], and *generalized physical model* [25]. We adopt the generalized physical model as it is more realistic than the other two, [20], [22], [24], [25].

Let $\mathcal{K}_t$ denote a *scheduling set* in which all links are scheduled simultaneously in time slot t. Specifically,

Definition 2. *Under generalized physical model, when a scheduling set $\mathcal{K}_t$ is scheduled, the rate of link $i \in \mathcal{K}_t$ is of*

$$R_{i,t} = B \times \mathbf{1} \cdot \{i \in \mathcal{K}_t\} \times \log(1 + \mathrm{SINR}(i,t)), \qquad (2)$$

where

$$\mathrm{SINR}(i,t) = \frac{P \cdot \ell(|\mathbf{t}_i - \mathbf{r}_i|)}{N_0 + \sum_{j \in \mathcal{K}_t / i} P \cdot \ell(|\mathbf{t}_j - \mathbf{r}_i|)};$$

$\mathbf{t}_i$ and $\mathbf{r}_i$ denote the transmitter and receiver of link i, respectively; $|\cdot|$ represents the euclidean distance between two points; $\ell(\cdot)$ denotes the power attenuation function that is assumed to depend only on the distance between the transmitter and receiver [14], [20], [24], [29]; $\ell(|\cdot|) := |\cdot|^{-\alpha}$ for dense scaling networks, and $\ell(|\cdot|) := \min\{1, |\cdot|^{-\alpha}\}$ for extending scaling networks [20].

2.3 Aggregation Throughput for Wireless Sensor Networks

As in the models of most related works [13], [17], [18], [19], every sensor node $s_i, i \in [0, n-1]$, periodically generates measurements of the environment, which belong to a fixed finite set $\mathcal{M}$ with $|\mathcal{M}| = \mathrm{m}$, and the function of interest is then required to be computed periodically for the measured data. Then, each measurement can be indicated by $\log \mathrm{m}$ bits. Note that all the logs in this paper are to the base 2. Please see the illustration in Fig. 1. Take the temperature monitoring, for example, $\mathcal{M}$ can be defined as $\{-50°\mathrm{C}, -49.9°\mathrm{C}, \ldots, +49.9°\mathrm{C}, +50°\mathrm{C}\}$, and $\mathrm{m} = 1{,}000$. Define the function of interest to sink node as $\mathbf{g}_n{:}\mathcal{M}^n \to \mathcal{G}_n$; furthermore, for any node $k \in [1, n]$, define the function of the sensor measurements as $\mathbf{g}_k{:}\mathcal{M}^k \to \mathcal{G}_k$, where $\mathcal{G}_k$ is the range of $\mathbf{g}_k$. Take function max for instance, $\mathbf{g}_k$ can be defined as the maximum value among all the data from k's descendants (w.r.t. an aggregation tree, [30]). Suppose that each sensor has an associated block of L readings, known a priori [17]. We call L rounds of measurements an *aggregation unit*. We again take temperature monitoring as an example, assume that the sensor samples the temperature every 5 mins, we thus can treat six rounds of readings (e.g., 30 mins) as one aggregation unit by setting $L = 6$. Notice that the aggregation operation can only be applied to the data from the same round. Before formulating the definition of aggregation throughput, we introduce some notations in Table 1.

TABLE 1
Some Notations

Notations	Meaning
$\phi(n) \sim [\phi_0(n), \phi_1(n)]$	$\phi(n) = \Omega(\phi_0(n))$ and $\phi(n) = O(\phi_1(n))$.
$\phi(n) \sim (\phi_0(n), \phi_1(n))$	$\phi(n) = \omega(\phi_0(n))$ and $\phi(n) = o(\phi_1(n))$.
$\mathcal{A}(\lambda, n)$	the square region $[0, \sqrt{n/\lambda}]^2$.
$\mathcal{N}(\lambda, n)$	a random network composed of n sensors with density λ.
$\mathrm{m} := \lvert\mathcal{M}\rvert$	the size of a fixed finite set $\mathcal{M}$ containing all measurements.
L	block-length, i.e., the size of aggregation units.
$\mathrm{M}^{n\times L} \in \mathcal{M}^{n\times L}$	a $n \times L$ matrix of measurements.
$\mathrm{M}^{n\times L}(i,j)$	the j-th measurement of sensor node s_i.
$\mathrm{M}^{n\times L}(i,\cdot)$	a block of L consecutive measurements of s_i.
$\mathrm{M}^{n\times L}(\cdot,j)$	a set of the j-th measurements of n sensors.
$\mathbf{g}_k(\mathrm{M}^k)$	$:= \mathbf{g}_k(\mathrm{M}_1, \mathrm{M}_2, ..., \mathrm{M}_k)$, for any k-vector $\mathrm{M}^k = [\mathrm{M}_1, \mathrm{M}_2, ...\mathrm{M}_k]^T \in \mathcal{M}^k$.
$\mathbf{g}_k^{\mathrm{L}}(\mathrm{M}^{k\times L})$	$:= \left(\mathbf{g}_k(\mathrm{M}^{k\times L}(\cdot,1)), \cdots, \mathbf{g}_k(\mathrm{M}^{k\times L}(\cdot,L))\right)$, for a given matrix $\mathrm{M}^{k\times L}$.
$\Psi := \Psi(n)$	gathering efficiency.
$\mathcal{S}(\Psi \cdot n)$	a subset consisting of $\Psi \cdot n$ sensors.
$\mathscr{A}(n, L, \mathcal{S}(\Psi \cdot n))$	a scheme that can aggregate the measurements from sensors in $\mathcal{S}(\Psi \cdot n)$.
$\Lambda := \Lambda(\lambda, n)$	the achievable aggregation throughput.
$\Phi := \Phi(\lambda, n)$	tradeoff between throughput and gathering efficiency.

2.3.1 Aggregation Functions of Interest

We focus on an important class of symmetric functions called DPC functions [13]. Functions such as the mean, max (or min), and various kinds of indicator functions all belong to this category. A necessary condition of the DPC functions is provided in Lemma 2.

Lemma 2. *If an aggregation function $\mathbf{g}_k$, $1 \le k \le n$, is* perfectly compressible, *then $|\mathcal{G}_k| = \Theta(\mathrm{m})$, where $\mathcal{G}_k$ is the range of function $\mathbf{g}_k$.*

Note that $|\mathcal{G}_k| = \Theta(\mathrm{m})$ is not a sufficient condition ensuring the aggregation function $\mathbf{g}_k$ to be perfectly compressible [13], [17]. For simplicity, we assume that $|\mathcal{G}_k| = \mathrm{m}$ for a perfectly

compressible aggregation function, without changing the order of the derived throughput. Functions like max, min, etc., indeed belong to this category.

2.3.2 Achievable Aggregation Throughput

Denote an aggregation scheme by $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$, where

- L denotes the block-length, which determines a sequence of message passings between sensors and computations at sensors;
- $\mathcal{S}(\Psi \cdot n) \subseteq \mathcal{S}(n)$, $\Psi \in (0, 1]$, is a subset of sensors which will be used to measure the gathering efficiency;
- input any $\mathrm{M}^{(\Psi \cdot n) \times L} \in \mathcal{M}^{(\Psi \cdot n) \times L}$ from all sensors in $\mathcal{S}(\Psi \cdot n)$, output $\mathbf{g}^{\mathrm{L}}_{(\Psi \cdot n)}(\mathrm{M}^{(\Psi \cdot n) \times L})$ at the sink node.

Next, we define the *achievable aggregation throughput*.

Definition 3. *For a given aggregation function:* $\mathbf{g}_n$*:*$\mathcal{M}^n \to \mathcal{G}_n$*, we say a throughput of* $\Lambda(n) = \frac{L \cdot \log \mathrm{m}}{T}$ *bps* Ψ-achievable*, if there exists an aggregation scheme* $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$ *under which there is a subset* $\mathcal{S}(\Psi \cdot n)$ *such that the corresponding* $\mathrm{M}^{(\Psi \cdot n) \times L} \in \mathcal{M}^{(\Psi \cdot n) \times L}$ *can be aggregated into* $\mathbf{g}^{\mathrm{L}}_{(\Psi \cdot n)}(\mathrm{M}^{(\Psi \cdot n) \times L})$ *at the sink node within* T *seconds.*

Based on Definition 3, we have

- The ordinary achievable throughput [13], [17], [18], [19] is indeed 1-achievable.
- We say a Ψ-achievable throughput *asymptotically 1-achievable* if $\liminf_{n \to \infty} \Psi = 1$. We also directly call it *asymptotically achievable*.
- We call the ratio Ψ the *gathering efficiency* of a specific aggregation scheme $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$.

2.3.3 Tradeoffs between Aggregation Throughput and Gathering Efficiency

It is intuitive that there exists a tradeoff between the aggregation throughput $\Lambda(\lambda, n)$ and gathering efficiency $\Psi(n)$. Define such a tradeoff as

$$\Phi(\lambda, n) = \Lambda(\lambda, n) \cdot \Psi(n).$$

Obviously, $\Phi(\lambda, n) = O(1)$. Particularly,

Definition 4. *We say that an aggregation scheme* $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$ *is* scalable *if*

$$\Phi(\lambda, n) = \Lambda(\lambda, n) \cdot \Psi(n) = \Theta(1), \tag{3}$$

i.e., $\Lambda(\lambda, n) = \Theta(1)$ *and* $\Psi(n) = \Theta(1)$*, where* $\Lambda(\lambda, n)$ *is the throughput derived by* $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$.

3 Aggregation Schemes for Random Wireless Sensor Networks

Both proposed aggregation schemes are based on lattices. To simplify the description, we recall a notion called *scheme lattice* from [31].

Definition 5 (Scheme Lattice). *Partition a square region* $\mathcal{A} = [0, \mathrm{a}]^2$ *into a lattice consisting of square cells of side length* l*, we call the produced lattice* scheme lattice*, and denote it by* $\mathbb{L}(\mathrm{a}, \mathrm{l}, \theta, \lambda)$*, where* $\theta \in [0, \frac{\pi}{4}]$ *is the minimum angle between the boundaries of* $\mathcal{A}$ *and the sides of cells.*

TABLE 2
Some Predefined Constant Parameters

Range	Conditions
$c \in (0, \infty)$	$c^2 > \ln 6$
$\kappa \in (0, \infty)$	$\kappa > 2/(c^2 - \ln 6)$
$\varpi \in (0, \infty)$	$\varpi < (\kappa \cdot (c^2 - \ln 6) - 2)/(\ln(1 - e^{-c^2}) + c^2)$
$\varepsilon_1 \in (0, \infty)$	arbitrary
$\varepsilon_2 \in (0, 1)$	$\varepsilon_2 + (1 - \varepsilon_2)\ln(1 - \varepsilon_2) > 0$
$\varepsilon_3 \in (0, \infty)$	$(1 + \varepsilon_3)\ln(1 + \varepsilon_3) - \varepsilon_3 > 0$
$\varepsilon_4 \in (0, 1)$	$\varepsilon_4 + (1 - \varepsilon_4)\ln(1 - \varepsilon_4) > 1/z$, for $z \in (0, \infty)$
$\varepsilon_5 \in (0, \infty)$	$(1 + \varepsilon_5)\ln(1 + \varepsilon_5) - \varepsilon_5 > 1/z$, for $z \in (0, \infty)$
$\varepsilon_6 \in (0, \infty)$	$\sigma\lambda \cdot ((1 + \varepsilon_6) \cdot \ln(1 + \varepsilon_6) - \varepsilon_6) + \ln(\sigma\lambda) = o(\ln n)$
$\varepsilon_7 \in (0, \infty)$	arbitrary
$\varepsilon_8 \in (0, 1)$	$\varepsilon_8 = (1 + \varepsilon_3) \cdot (1 + \varepsilon_7) \cdot \left(e^{\varepsilon_6}/(1 + \varepsilon_6)^{(1 + \varepsilon_6)}\right)^{c^2}$
$\varepsilon_9 \in (0, 1)$	$\varepsilon_9 \geq \varepsilon_2 + \varepsilon_8$

3.1 Single-Hop-Length Aggregation Scheme

We design the scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ based on the scheme lattice $\mathbb{L}_1 = \mathbb{L}(\sqrt{n/\lambda}, \sqrt{z \cdot \ln n/\lambda}, 0, \lambda)$. According to Lemma D (Please refer to Appendix B, available in the online supplemental material), for all cells in $\mathbb{L}_1$, the number of sensors inside each cell is w.h.p. within $[(1 - \varepsilon_4) \cdot \ln n, (1 + \varepsilon_5) \cdot \ln n]$, where ε_4 and ε_5 are some constants depending on z and defined in Table 2. For simplicity, we ignore the details about the integer, and assume that the number of rows (or columns) $\sqrt{\frac{n}{z \cdot \ln n}}$ is always an integer, without changing the results in order sense. Taking the cell in bottom left corner as the origin with a 2-dimensional index $(0, 0)$, we give each cell in $\mathbb{L}_1$ an index in the order from left to right and bottom to top, i.e., the index of the cell in top right corner is (δ, δ), where $\delta = \delta(n) = \frac{\sqrt{n}}{\sqrt{z \cdot \ln n}} - 1$. Without loss of generality, we assume that the sink node s_0 is located in the cell (δ, δ). Choose randomly just one sensor from each cell as the *aggregation station*, we obtain a set, denoted by $\mathcal{B}$, consisting of $\frac{n}{z \cdot \ln n}$ sensors. Let $b_{i,j} \in \mathcal{B}$ denote the aggregation station in cell (i, j). Note that the sink node s_0 can be selected as the aggregation station of cell (δ, δ). Define a sequence of sets

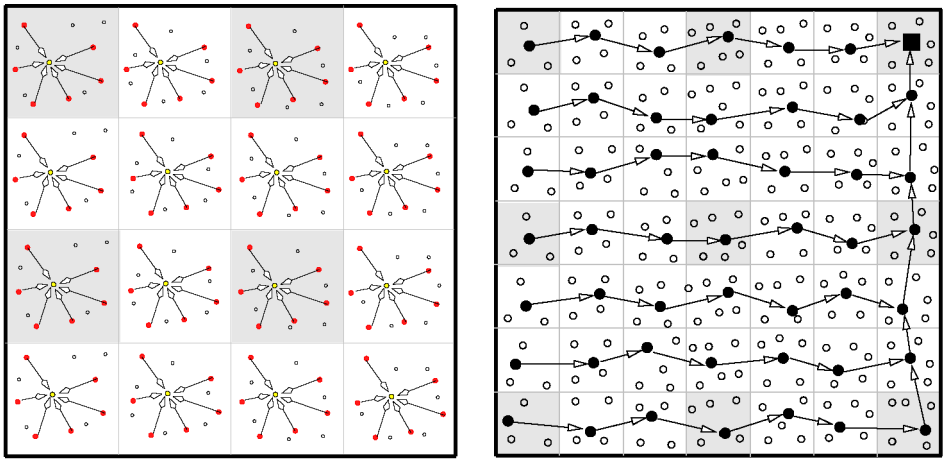
$$\mathcal{H}_{h,v} := \{b_{i,j} | (i \bmod 3 = h) \wedge (j \bmod 3 = v)\},$$

where $h \in \{0, 1, 2\}$ and $v \in \{0, 1, 2\}$. Then, the aggregation scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ is described as follows:

- **Local Aggregation**: In each cell of $\mathbb{L}_1$, $\beta \cdot \ln n$ sensors are selected, if applicable, where

$$\beta = \max\left\{\Psi \cdot (1 + \varepsilon_5), \frac{1}{\ln n}\right\}. \tag{4}$$

  L rounds of measurements from those sensors are aggregated to the aggregation stations by a single hop; all transmissions are scheduled by a 4-TDMA scheme, as illustrated in Fig. 2a.
- **Horizontal Backbone Aggregation**: L rounds of data held by each aggregation station are aggregated



(a) Local Selection and Aggregation (b) Aggregation Backbones

Fig. 2. Aggregation scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$. The shaded cells are simultaneously scheduled. (a) In each cell, $\beta \cdot \ln n$ sensors are selected, if applicable. (b) The black square is the sink node.

to the adjacent aggregation stations in the order from left to right in a *pipelined fashion* (Algorithm 1); all transmissions are scheduled by a 9-TDMA scheme, as illustrated in Fig. 2b.

- **Vertical Backbone Aggregation**: L rounds of data held by each aggregation station in the δth-column are aggregated to the adjacent aggregation stations in the order from bottom to top in a similar pipelined fashion to Algorithm 1; all transmissions are scheduled by a 3-TDMA scheme, as illustrated in Fig. 2b.

Algorithm 1. Horizontal Backbone Pipelined Aggregation

Input: L rounds of aggregated measurements at all aggregation stations.
Output: L rounds of aggregated data at all station $b_{i,\delta}$.
for $k = 1, 2, \cdots, L, L+1, \cdots, L+\delta-3$ **do**
  $k \to k'$;
  if $k > L$ **then** $L \to k$;
  else for $h = 0, 1, 2$ **do**
    for $v = 0, 1, 2$ **do**
      for $r = 1, \cdots, k$ **do**
        All $b_{i,j} \in \mathcal{H}_{h,v}$ are permitted to transmit;
        if *it holds that* $1 \le j \le \delta - 1$, *and (1)* $b_{i,j}$, $j \ge 1$, *has received the* r*-th round aggregated data from* $b_{i,j-1}$, *and (2)* $b_{i,j+1}$ *has not received the* r*-th round aggregated data from* $b_{i,j}$, **then** $b_{i,j}$ sends them to $b_{i,j+1}$;
        else if $j = 0$, *and* $b_{i,1}$ *has not received the* r*-th round aggregated measurements from* $b_{i,0}$, **then** $b_{i,0}$ sends them to $b_{i,1}$.
  $k' \to k$

3.2 Multiple-Hop-Length Aggregation Scheme

We design another aggregation scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$ based on the scheme lattice $\mathbb{L}_2 = \mathbb{L}(\sqrt{\frac{n}{\lambda}}, \frac{c}{\sqrt{\lambda}}, \frac{\pi}{4}, \lambda)$, where $c > 0$ is a constant and the specific value is determined in Table 2. Choose randomly a sensor from each nonempty cell, called *aggregation station*, then, we can build the *aggregation backbones* using the method in [20], [32]. Please see the illustration in Fig. 3. The *backbone stations*, i.e., the stations on the aggregation backbones, are connected by only *short* links, whereas every *peripheral station*, i.e., the stations other than backbone stations, can access a specific backbone station node in one-hop transmission.

For a given constant $\kappa > 0$, partition the scheme lattice $\mathbb{L}_2$ into horizontal (vertical) rectangle slabs with the horizontal (vertical) width of $\sqrt{\frac{n}{\lambda}}$ and the vertical (horizontal) width of

$$w_{\mathrm{R}} = (\kappa \ln m) \cdot c\sqrt{2/\lambda}, \tag{5}$$

where $m = \frac{\sqrt{n}}{\sqrt{2}c}$. We assume that $\frac{m}{\kappa \ln m}$, denoting the number of rectangle slabs, is an integer. Then, according to [20, Theorem 5], we have

Lemma 3. *For any constants* c, κ *satisfying*

$$0 < \frac{2}{c^2 - \ln 6} < \kappa < \infty, \tag{6}$$

there exists a constant ϖ *depending on* κ *and* c *such that for all horizontal (or vertical) slabs, there are w.h.p. at least* $\varpi \cdot \ln m$ *horizontal (or vertical) aggregation backbones, where*

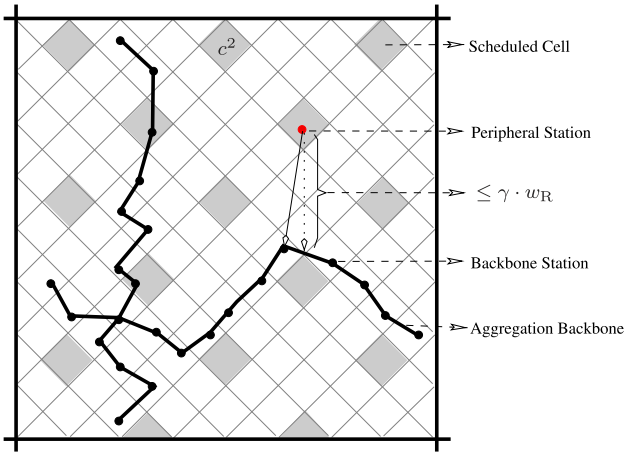


Fig. 3. Aggregation backbones under the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$.

$$0 < \varpi < \frac{\kappa \cdot (c^2 - \ln 6) - 2}{\ln(1 - e^{-c^2}) + c^2}. \quad (7)$$

When the aggregation backbones are built, the cells in each slab can be assigned averagely to $\varpi \cdot \ln n$ aggregation backbones. For instance, each slab is further divided into $\varpi \cdot \ln n$ slices, and each slice is mapped to a specific backbone. Anyway, the distance between a peripheral station and the corresponding backbone station is within $(0, w_{\mathrm{R}}]$. Now, we give the aggregation scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$. The involved constants are all defined in Table 2.

- **Selection**: Choose a subset of cells that contain aggregation stations, denoted by $\mathbb{C}(\Psi)$, in which the aggregation stations are at distance of at most $\gamma \cdot w_{\mathrm{R}}$ to the corresponding aggregation backbones, where

$$\gamma = \max\left\{\frac{\Psi}{1 - \varepsilon_9} - \frac{\varpi}{\kappa}, 0\right\}. \quad (8)$$

- **Local Aggregation**: In each cell in $\mathbb{C}(\Psi)$, choose randomly at most $\mathbf{c} = \lceil (1 + \varepsilon_6) \cdot c^2 \rceil$ sensors, if applicable, where $\varepsilon_6 > 0$ is defined in Table 2 by letting $\sigma\lambda = c^2$; L rounds of measurements from those chosen sensors are aggregated to the aggregation station by a single hop; all transmissions are scheduled by a 4-TDMA scheme based on the scheme lattice $\mathbb{L}_2$.
- **Draining Aggregation**: All peripheral stations in $\mathbb{C}(\Psi)$ drain the L rounds of data into the corresponding backbone stations by a single hop of distance at most $\gamma \cdot w_{\mathrm{R}}$; all transmissions are scheduled by a K^2-TDMA scheme based on $\mathbb{L}_2$, where

$$K = 2 \cdot \left(\left\lceil \frac{\gamma \cdot w_{\mathrm{R}}}{c/\sqrt{\lambda}} \right\rceil + 1\right). \quad (9)$$

- **Horizontal Backbone Aggregation**: L rounds of data held by each backbone station are horizontally aggregated to the adjacent backbone stations in the order from left to right in a similar pipelined fashion to Algorithm 1, until the data are aggregated into the backbone stations on the backbones passed through by the sink node s_0, denoted by b_{s_0}; all transmissions are scheduled by a 9-TDMA scheme, as illustrated in Fig. 2b.
- **Vertical Backbone Aggregation**: L rounds of data held by each backbone station in the backbone b_{s_0} are aggregated to the adjacent aggregation stations in the order from bottom to top in a similar pipelined fashion to Algorithm 1; all transmissions are scheduled by a 3-TDMA scheme.

4 Ψ-Achievable Aggregation Throughput

4.1 Ψ-Achievable Throughput under SHL Scheme

First, according to Step 1 of $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$, the number of sensors is at least $\Psi \cdot n$, then, it is easy to get that

Lemma 4. *Under the scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$, the derived throughput is Ψ-achievable.*

Theorem 1. *Under the scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ with $L = \Omega(\frac{\sqrt{n}}{\sqrt{\log n}})$, the achievable throughput for DPC functions is of order*

$$\Lambda_1(\lambda, n) = \begin{cases} \Omega\left(\frac{1}{\beta \cdot \log n}\right) & \text{when} \quad \lambda \sim [\log n, n] \\ \Omega\left(\frac{\lambda^{\frac{\alpha}{2}}}{\beta \cdot (\log n)^{1+\frac{\alpha}{2}}}\right) & \text{when} \quad \lambda \sim [1, \log n], \end{cases}$$

where β is defined in (4).

As mentioned in Section 2.3.2, the ordinary achievable throughput [13], [17], is indeed the 1-achievable throughput under Definition 3. Giridhar and Kumar [17] designed an aggregation scheme under the *fixed-range protocol model* [14] by which the achievable throughput of a special case of random dense scaling WSNs (RDS-WSN), i.e., $\mathcal{N}(n, n)$, is of $\Omega(\frac{1}{\log n})$.

Next, we prove Theorem 1 in two cases, i.e., random dense scaling WSN and random extended scaling WSN (RES-WSN).

4.1.1 Proof for RDS-WSN

By using a 4-TDMA, the total rate of each cell is sustained of $R^{\mathrm{d}}_{\mathrm{LA},1} = \Omega(1)$. Then, it takes

$$T^{\mathrm{d}}_{\mathrm{LA},1} \le \beta \cdot \ln n \cdot L \cdot \log \mathrm{m} / R^{\mathrm{d}}_{\mathrm{LA},1} = O(\beta \cdot L \cdot \log n),$$

to finish the local aggregation of L rounds of measurements.

By using the 9-TDMA scheme, each link can sustain a rate of $R^{\mathrm{d}}_{\mathrm{HBA},1} = \Omega(1)$. Then, by Algorithm 1, it takes

$$T^{\mathrm{d}}_{\mathrm{HBA},1} \le 9 \cdot (L + \delta(n) - 3) \cdot \log \mathrm{m} / R^{\mathrm{d}}_{\mathrm{HBA},1},$$

to finish the horizontal backbone aggregation of L rounds of data. Since $L = \Omega(\delta(n))$, $T^{\mathrm{d}}_{\mathrm{HBA},1} = O(L)$. Similarly, we get that it takes $T^{\mathrm{d}}_{\mathrm{VBA},1} = O(L)$ to finish the vertical backbone aggregation of L rounds of data.

Hence, under the scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$, the Ψ-achievable throughput is of order

$$\Lambda_1(\lambda, n) = \frac{L \cdot \log \mathrm{m}}{T^{\mathrm{d}}_{\mathrm{LA},1} + T^{\mathrm{d}}_{\mathrm{HBA},1} + T^{\mathrm{d}}_{\mathrm{VBA},1}} = \Omega\left(\frac{1}{\beta \cdot \log n}\right),$$

which proves the lemma.

4.1.2 Proof for RES-WSN

By using a 4-TDMA, the total rate of each cell is sustained of $R^{\mathrm{e}}_{\mathrm{LA},1} = \Omega((\frac{\log n}{\lambda})^{-\frac{\alpha}{2}})$. It takes

$$T^{\mathrm{e}}_{\mathrm{LA},1} \le \frac{\beta \cdot \ln n \cdot L \cdot \log \mathrm{m}}{R^{\mathrm{e}}_{\mathrm{LA},1}} = O\left(\frac{\beta \cdot L \cdot (\log n)^{1+\frac{\alpha}{2}}}{\lambda^{\frac{\alpha}{2}}}\right),$$

to finish the local aggregation of L rounds of measurements.

By using the 9-TDMA scheme, each link can also sustain a rate of $R^{\mathrm{e}}_{\mathrm{HBA},1} = \Omega((\frac{\log n}{\lambda})^{-\frac{\alpha}{2}})$. Then, it takes

$$\begin{aligned} T^{\mathrm{e}}_{\mathrm{HBA},1} &\le 9 \cdot (L + \delta(n) - 3) \cdot \log \mathrm{m} / R^{\mathrm{e}}_{\mathrm{HBA},1} \\ &= O\left(L \cdot \left(\frac{\log n}{\lambda}\right)^{\frac{\alpha}{2}}\right), \end{aligned}$$

to finish the horizontal backbone aggregation of L rounds of data. Similarly, we get that it takes $T^{\mathrm{e}}_{\mathrm{VBA},1} = O(L \cdot (\frac{\log n}{\lambda})^{\frac{\alpha}{2}})$ to finish the vertical backbone aggregation of L rounds of data.

Hence, under the scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$, the Ψ-achievable throughput is of order

$$\Lambda_1(\lambda, n) = \frac{L \cdot \log \mathrm{m}}{T^{\mathrm{e}}_{\mathrm{LA},1} + T^{\mathrm{e}}_{\mathrm{HBA},1} + T^{\mathrm{e}}_{\mathrm{VBA},1}} = \Omega\left(\frac{\lambda^{\frac{\alpha}{2}}}{\beta \cdot (\log n)^{1+\frac{\alpha}{2}}}\right),$$

which proves the lemma.

4.2 Ψ-Achievable Throughput under MHL Scheme

First, we have the following result.

Lemma 5. *Under the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the derived throughput is Ψ-achievable.*

Proof. It is clear that the total area of the cells passed through by the horizontal backbones is at least of $\frac{\varpi}{\kappa} \cdot \frac{n}{\lambda}$. According to Step 1 of $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$ (Selection), there is a region of area $\sqrt{n/\lambda} \times \gamma \cdot \sqrt{n/\lambda}$ in which all sensors are considered to be gathered.

From Lemma D in Appendix B, available in the online supplemental material, there exist w.h.p. some cells containing $\omega(1)$ sensors, while we only choose at most $\mathbf{c}$ (a constant) sensors in Step 2 of $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$ (Local aggregation). That means that we have to discard some sensors. Combining Lemma D and Lemma E in Appendix B, available in the online supplemental material, we get that the number of omitted sensors inside the region of area $\sqrt{n/\lambda} \times \gamma \cdot \sqrt{n/\lambda} + \frac{\varpi}{\kappa} \cdot \frac{n}{\lambda}$, i.e., $(\gamma + \frac{\varpi}{\kappa}) \cdot \frac{n}{\lambda}$, is at most of $\varepsilon_8 \cdot (\gamma + \frac{\varpi}{\kappa}) \cdot n$. From Lemma D in Appendix B, available in the online supplemental material, the number of sensors inside this region is at least of $(1 - \varepsilon_2) \cdot (\gamma + \frac{\varpi}{\kappa}) \cdot n$. Then, $(1 - \varepsilon_9) \cdot (\gamma + \frac{\varpi}{\kappa}) \cdot n \geq \Psi \cdot n$, which completes the proof. □

Theorem 2. *Under the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$ with $L = \Omega(\sqrt{n})$, the Ψ-achievable throughput for DPC functions is of order*

$$\Lambda_2(\lambda, n) = \begin{cases} \Omega\left(\frac{1}{(\gamma \cdot \log n)^2 + 1}\right) & \text{when} \quad \lambda \sim [(\gamma \log n)^2, n] \\ \Omega\left(\frac{\lambda^{\frac{\alpha}{2}}}{(\gamma \cdot \log n)^{\alpha+2} + \lambda^{\frac{\alpha}{2}}}\right) & \text{when} \quad \lambda \sim [1, (\gamma \log n)^2], \end{cases}$$

where γ is defined in (8).

Before proving Theorem 2, we propose the following lemma,

Lemma 6. *Based on the scheme lattice $\mathbb{L}_2$ (Fig. 3), for any integer $D > 0$, there exists an $R(D)$, such that in each cell there is a sensor that can transmit w.h.p. at a rate $R(D)$ to any destination located within a distance $D \cdot \sqrt{2} \cdot \frac{c}{\sqrt{\lambda}}$, where*

- $R(D) = \Omega(D^{-2})$, *when* $D = O(\sqrt{\lambda})$,
- $R(D) = \Omega(\frac{\lambda^{\frac{\alpha}{2}}}{D^{\alpha+2}})$, *when* $D = \Omega(\sqrt{\lambda})$.

Due to the similarity to the proofs of [20, Theorem 3 and Theorem 4], we omit the detailed proof. The following corollary clearly holds.

Corollary 1. *During the draining aggregation phase of the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, it holds that $D = \Theta(\gamma \cdot \log n)$.*

Now, we begin to prove Theorem 2 in two cases when $\gamma \cdot \log n = O(\sqrt{\lambda})$ and when $\gamma \cdot \log n = \Omega(\sqrt{\lambda})$.

4.2.1 When $\gamma \cdot \log n = O(\sqrt{\lambda})$

According to Lemma 6 during the local aggregation phase, the total rate of each cell can be sustained of order $R_{\mathrm{LA},2} = \Omega(1)$. Then, it takes

$$T_{\mathrm{LA},2} \leq \lceil (1 + \varepsilon_6) \cdot c^2 \rceil \cdot L \cdot \log \mathrm{m} / R_{\mathrm{LA},2} = O(L),$$

to finish the local aggregation of L rounds of measurements.

Similarly, during the horizontal backbone aggregation phase, each link can sustain a rate of $R_{\mathrm{HBA},2} = \Omega(1)$. Then, by Algorithm 1, it takes

$$T_{\mathrm{HBA},2} \leq 9 \cdot \left(L + \sqrt{2n}/c\right) \cdot \log \mathrm{m} / R_{\mathrm{HBA},2} = O(L),$$

to finish the horizontal backbone aggregation of L rounds of data; and we get that it takes $T_{\mathrm{VBA},2} = O(L)$ to finish the vertical backbone aggregation of L rounds of data.

For the draining aggregation phase, by Corollary 1 and Lemma 6, the link in this phase can sustain a rate of $R_{\mathrm{DA},2} = \Omega(\frac{1}{(\gamma \cdot \log n)^2})$. Thus, it takes

$$T_{\mathrm{DA},2} \leq L \cdot \log \mathrm{m} / R_{\mathrm{DA},2} = O((\gamma \cdot \log n)^2 \cdot L),$$

to finish the draining aggregation of L rounds measurements.

Hence, under the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the Ψ-achievable throughput is of order $\Lambda_2(\lambda, n) = \Omega(\frac{1}{(\gamma \cdot \log n)^2 + 1})$, where $\gamma = \max\{\frac{\Psi}{1 - \varepsilon_9} - \frac{\varpi}{\kappa}, 0\}$.

4.2.2 When $\gamma \cdot \log n = \Omega(\sqrt{\lambda})$

There is no change on the link rate except for the draining aggregation phase. From Lemma 6, the link rate is achieved of order $R_{\mathrm{DA},2} = \Omega(\frac{\lambda^{\frac{\alpha}{2}}}{(\gamma \cdot \log n)^{\alpha+2}})$. By a similar procedure to the previous case, we can obtain that under the scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the Ψ-achievable throughput is of order $\Lambda_2(\lambda, n) = \Omega(\frac{\lambda^{\frac{\alpha}{2}}}{(\gamma \cdot \log n)^{\alpha+2} + \lambda^{\frac{\alpha}{2}}})$.

4.3 Tradeoffs between Throughput and Gathering Efficiency

Based on Theorem 1 and Theorem 2. We get that

Theorem 3. *Under the schemes $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ and $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the optimal Ψ-achievable throughput for DPC functions, denoted by $\Lambda := \Lambda(\lambda, n)$, and the tradeoff between the aggregation throughput and gathering efficiency, denoted by $\Phi := \Phi(\lambda, n)$, are summarized as follows:*

- *When* $\Psi(n) = (1 - \varepsilon_9) \cdot \frac{\varpi}{\kappa} + O(\frac{1}{\log n})$ *or* $\Psi(n) \in (0, (1 - \varepsilon_9) \cdot \frac{\varpi}{\kappa}]$,

$$\Phi = \Lambda = \Omega(1) \text{ for all } \lambda \sim [1, n].$$

- *When* $\Psi(n) - (1 - \varepsilon_9) \cdot \frac{\varpi}{\kappa} \sim (\frac{1}{\log n}, \frac{1}{\sqrt{\log n}}]$,

$$\Phi = \Lambda = \begin{cases} \Omega\left(\frac{\lambda^{\alpha/2}}{(\gamma \cdot \log n)^{\alpha+2}}\right) & \text{when} \quad \lambda \sim [1, (\gamma \log n)^2] \\ \Omega\left(\frac{1}{(\gamma \cdot \log n)^2}\right) & \text{when} \quad \lambda \sim [(\gamma \log n)^2, n], \end{cases}$$

where $\gamma = \frac{\Psi(n)}{1 - \varepsilon_9} - \frac{\varpi}{\kappa}$.

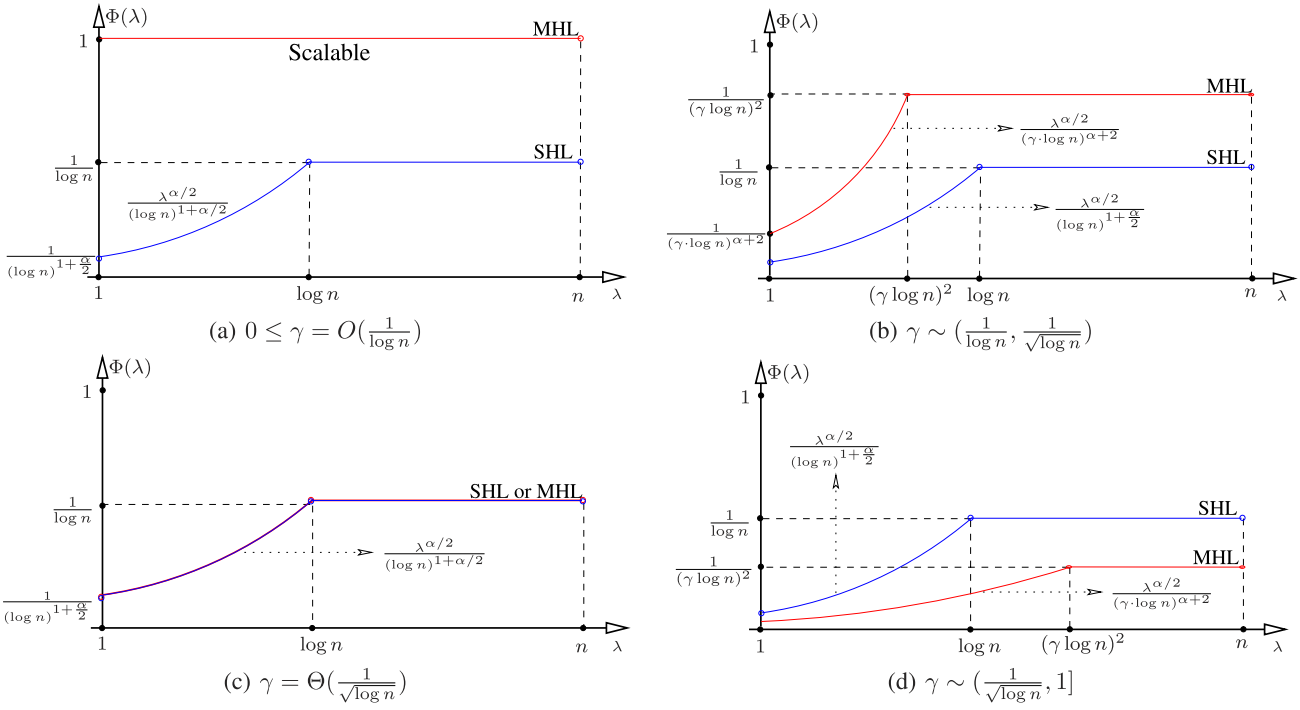


Fig. 4. Tradeoffs $\Phi(\lambda, n)$ between aggregation throughput $\Lambda(\lambda, n)$ and gathering efficiency $\Psi(n)$ for ordinary DPC functions under the schemes $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ and $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, denoted by **SHL** and **MHL**, respectively. Here, $\gamma = \max\{\frac{\Psi(n)}{1-\varepsilon_9} - \frac{\varpi}{\kappa}, 0\}$, and the constant parameters ε_9, ϖ, and κ are defined in Table 2.

- *When* $\Psi(n) - (1-\varepsilon_9)\cdot\frac{\varpi}{\kappa} \sim [\frac{1}{\sqrt{\log n}}, 1]$,

$$\Phi = \Lambda = \begin{cases} \Omega\left(\frac{\lambda^{\alpha/2}}{(\log n)^{1+\alpha/2}}\right) & when \quad \lambda \sim [1, \log n] \\ \Omega\left(\frac{1}{\log n}\right) & when \quad \lambda \sim [\log n, n]. \end{cases}$$

Detailed tradeoffs under two schemes are presented in Fig. 4.

From Theorem 3.we have the following observations:

1. The SHL scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ is not scalable.
2. Under the MHL scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, when

$$\Psi(n) = (1-\varepsilon_9)\cdot\left(\frac{\varpi}{\kappa} + O\left(\frac{1}{\log n}\right)\right),$$

the Ψ-achievable throughput is of order $\Theta(1)$, which means that the MHL scheme is indeed *scalable*.

We give an example of feasible solutions for the constants in Table 2:

$$c = 5, \quad \kappa = 10, \quad \varpi = 9.2032, \quad \varepsilon_2 = 10^{-5},$$
$$\varepsilon_3 = 10^{-6}, \quad \varepsilon_6 = 3, \quad \varepsilon_7 = 10^{-6}, \quad \varepsilon_8 = 10^{-4}.$$

Then, we can set $\varepsilon_9 = 2 \times 10^{-4}$, which means that the 0.92-achievable aggregation throughput is of order $\Theta(1)$.

3. When

$$\frac{\Psi(n)}{1-\varepsilon_9} - \frac{\varpi}{\kappa} = \omega\left(\frac{1}{\sqrt{\log n}}\right),$$

the tradeoff under the $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$ is better than that under the $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$ (Fig. 4d); otherwise, it is not better, as illustrated in Figs. 4a, 4b, and 4c.

5 Block Coding for Type-Threshold Functions

Now, we consider a special case of DPC functions, so-called *type-threshold DPC* functions. Intuitively, type-threshold functions are ones whose outcome can be computed even if knowing only a fixed number of known arguments (see [17] for a formal definition). The max function, min function, and range function belong to this types of functions. The technique of block coding can significantly improve the throughput for type-threshold functions in the collocated network whose conflict graph is a complete graph [17].

During the local aggregation under the SHL scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$, the communication graph in each cell is indeed a collocated network consisting of $\Theta(\beta \cdot \log n)$ vertexes, because any two links in a cell cannot be scheduled simultaneously. Then, we introduce the block coding technique into the SHL scheme $\mathcal{A}_1(n, L, \mathcal{S}(\Psi \cdot n))$; denote the modified scheme by $\mathcal{A}_1^{\mathrm{B}}(n, L, \mathcal{S}(\Psi \cdot n))$. Based on Theorem 1, by using the similar procedure to the proofs of [17, Theorem 4 and Theorem 5], we obtain that

Theorem 4. *By using the scheme* $\mathcal{A}_1^{\mathrm{B}}(n, L, \mathcal{S}(\Psi \cdot n))$ *with*

$$L = \Omega\left(\frac{\sqrt{n}}{\sqrt{\log n}}\right),$$

the achievable aggregation throughput for the type-threshold aggregation functions is of order

$$\Lambda_1^{\mathrm{B}}(\lambda, n) = \begin{cases} \Omega\left(\frac{1}{\log(\beta \cdot \log n)}\right) & when \quad \lambda \sim [\log n, n] \\ \Omega\left(\frac{\lambda^{\alpha/2}}{(\log n)^{\alpha/2} \cdot \log(\beta \cdot \log n)}\right) & when \quad \lambda \sim [1, \log n], \end{cases}$$

where β *is defined in (4).*

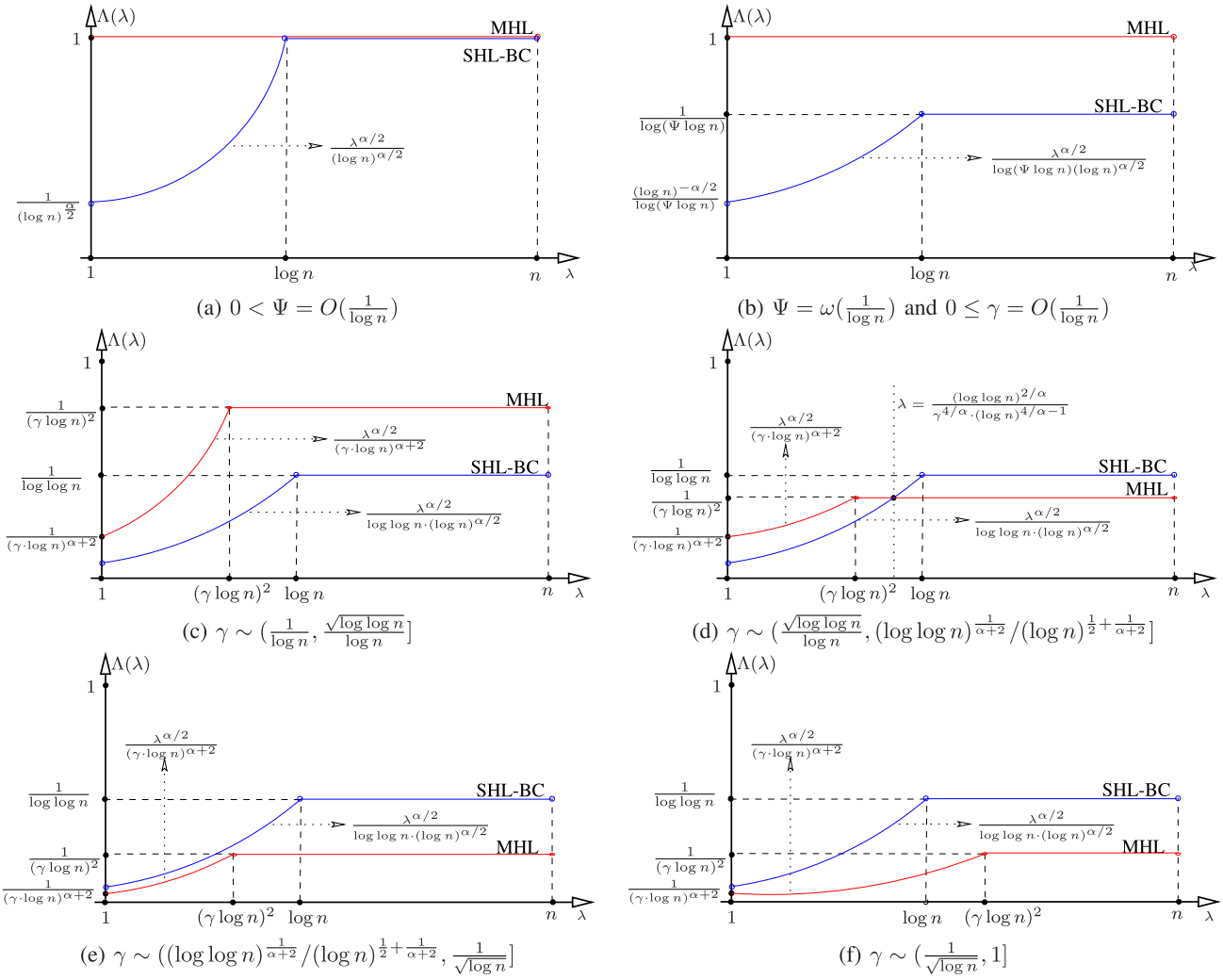


Fig. 5. Ψ-achievable aggregation throughput for type-threshold functions under the schemes $\mathcal{A}_1^{\mathrm{B}}(n, L, \mathcal{S}(\Psi \cdot n))$ and $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, denoted by **SHL-BC** and **MHL**, respectively. Here, $\gamma = \max\{\frac{\Psi(n)}{1-\varepsilon_9} - \frac{\varpi}{\kappa}, 0\}$, and the constant parameters ε_9, ϖ and κ are defined in Table 2.

Note that although during the local aggregation under the MHL scheme $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the communication graph in each cell is also a collocated network, the block coding cannot improve the throughput due to the finite number of vertexes in each collocated network.

Combining Theorem 2 and Theorem 4, we get that

Theorem 5. *Under the schemes $\mathcal{A}_1^{\mathrm{B}}(n, L, \mathcal{S}(\Psi \cdot n))$ and $\mathcal{A}_2(n, L, \mathcal{S}(\Psi \cdot n))$, the Ψ-achievable aggregation throughput for type-threshold DPC functions is described in Fig. 5; and the optimal Ψ-achievable throughput $\Lambda := \Lambda(\lambda, n)$ and trade-off between the aggregation throughput and gathering efficiency, $\Phi := \Phi(\lambda, n)$, are then summarized as following:*

- *When $\gamma = O(\frac{1}{\log n})$: Figs. 5a and 5b,*

$$\Phi = \Lambda = \Omega(1) \text{ for all } \lambda \sim [1, n].$$

- *When $\gamma \sim (\frac{1}{\log n}, \frac{\sqrt{\log\log n}}{\log n}]$: Fig. 5c*

$$\Phi = \Lambda = \begin{cases} \Omega\left(\frac{\lambda^{\alpha/2}}{(\gamma\cdot\log n)^{\alpha+2}}\right) & when \quad \lambda \sim [1, (\gamma\log n)^2] \\ \Omega\left(\frac{1}{(\gamma\cdot\log n)^2}\right) & when \quad \lambda \sim [(\gamma\log n)^2, n]. \end{cases}$$

- *When $\gamma \sim (\frac{\sqrt{\log\log n}}{\log n}, \frac{(\log\log n)^{\frac{1}{\alpha+2}}}{(\log n)^{\frac{1}{2}+\frac{1}{\alpha+2}}}]$: Fig. 5d,*

$$\Phi = \Lambda = \begin{cases} \Omega\left(\frac{\lambda^{\alpha/2}}{(\gamma\cdot\log n)^{\alpha+2}}\right) & when \quad \lambda \sim [1, (\gamma\log n)^2] \\ \Omega\left(\frac{1}{(\gamma\cdot\log n)^2}\right) & when \quad \lambda \sim [(\gamma\log n)^2, \lambda_0] \\ \Omega\left(\frac{(\frac{\lambda}{\log n})^{\alpha/2}}{\log\log n}\right) & when \quad \lambda \sim [\lambda_0, \log n] \\ \Omega\left(\frac{1}{\log\log n}\right) & when \quad \lambda \sim [\log n, n], \end{cases}$$

where

$$\lambda_0 = \frac{(\log\log n)^{\frac{2}{\alpha}}}{\gamma^{\frac{4}{\alpha}}(\log n)^{\frac{4}{\alpha}-1}}.$$

- *When $\gamma \sim (\frac{(\log\log n)^{\frac{1}{\alpha+2}}}{(\log n)^{\frac{1}{2}+\frac{1}{\alpha+2}}}, 1]$: Figs. 5e and 5f*

$$\Phi = \Lambda = \begin{cases} \Omega\left(\frac{1}{\log\log n}\right) & when \quad \lambda \sim [\log n, n] \\ \Omega\left(\frac{\lambda^{\alpha/2}}{(\log n)^{\frac{\alpha}{2}}\cdot\log\log n}\right) & when \quad \lambda \sim [1, \log n]. \end{cases}$$

6 Literature Review

Earlier, the aggregation capacity for large-scale WSNs was done by Marco et al. [15]. They considered random dense WSNs ($\lambda = n$) under the protocol model [14]. In [17], Giridhar and Kumar also focused on dense WSNs, and investigated the more general problem of computing and communicating symmetric functions of the sensor measurements. They showed that for type-sensitive and type-threshold functions, the aggregation capacities for random dense WSNs under the protocol model are of order $\Theta(\frac{1}{\log n})$ and $\Theta(\frac{1}{\log\log n})$, respectively. Moscibroda [13] derived the aggregation capacity scaling laws of DPC functions for *worst case* networks. They showed that under the protocol model and physical model [14], the aggregation capacity for worst case dense networks can be achieved of order $\Omega(\frac{1}{n})$ and $\Omega\Big(\frac{1}{(\log n)^2}\Big)$, respectively. For the extended network, Wang et al. [33] showed that under the generalized physical model, the aggregation capacities for type-sensitive and type-threshold functions are of order $\Theta((\log n)^{-\frac{\alpha}{2}-1})$ and $\Theta(\frac{(\log n)^{-\alpha/2}}{\log\log n})$, respectively, where $\alpha > 2$ denotes the power attenuation exponent in the generalized physical model. Gao et al. [34] studied the problem of aggregating data from a sparse set of nodes in a wireless sensor network where the sensors are deployed to detect relatively rare events. Ying et al. [19] studied the optimization problem of the total transmission energy for computing a symmetric function, subject to the constraint that the computation is w.h.p. correct. Li et al. [35] proposed the efficient algorithms for data collection and data aggregation for a multihop wireless sensor network of n nodes, and studied the complexities of time, message, and energy cost in various processing operations. Chen et al. [36] studied theoretical limitations of data collection and data aggregation in terms of delay and capacity for a random wireless sensor network of n sensors. They investigated some representative communication scenarios, such as those with single sink or multiple sinks, with regularly-deployed or randomly-deployed sinks, with or without aggregation. Under the cooperative paradigm, Zheng and Barton [23] demonstrated that the upper bound of the capacity of data collecting for extended WSNs ($\lambda = 1$) is of order $\Theta(\frac{\log n}{n})$ when operating in fading environments with power path-loss exponents that satisfy $2 < \alpha < 4$ or $\alpha > 4$, respectively. The work considered the aggregation functions without no *in-network aggregation* [37], e.g., *data downloading problem* [17]. In [38], Wang et al. studied the capacity, delay and their tradeoffs in k-source converge-cast networks. They considered both stationary and mobile networks, and showed that the essential key controlling the capacity is the redundancy, with inverse relationship. In [39], [40], by introducing the visual MIMO technique, Fu and Wang et al. designed two different cooperative schemes for static and mobile ad hoc wireless networks, respectively.

7 Discussion and Conclusion

We study the data aggregation of DPC functions for random WSNs with general density. We design two protocols, called SHL and MHL schemes, to derive the optimal aggregation throughput depending on a given gathering efficiency, and provide the optimal tradeoffs. Particularly, we show that the MHL scheme is scalable. The following are some important and interesting issues to be studied.

7.1 Gathering Efficiency and Coverage Rate

One of main applications of WSNs is to provide proper coverage of their deployment regions [41], [42], [43], [44]. Recall that for a random WSN of area n/λ, the sufficient condition ensuring k-coverage of a square by a Poisson point process with density λ and unit-area coverage range [41], [45], [46], [47] is that

$$\lambda + \ln\lambda + (k+2)\ln\ln\lambda \geq \ln n + (k+2)\ln\ln n + \xi(n/\lambda),$$

where $\xi(n/\lambda) = \omega(1)$ if $\lambda = o(n)$. Then, based on this result, how to build the relation between gathering efficiency and k-coverage rate, i.e., the ratio of the area of k-covered region to the total area, should be an important and interesting issue. A straightforward result is that: under our schemes, we can select uniformly at random a Ψ portion of sensors, then the density degenerates to $\Psi \cdot \lambda$. That means that the network remains k-coverage, if the following condition holds:

$$\begin{aligned}&\Psi\lambda + \ln(\Psi\lambda) + (k+2)\ln\ln(\Psi\lambda)\\ &\quad\geq \ln n + (k+2)\ln\ln n + \xi\Big(\frac{n}{\Psi\lambda}\Big).\end{aligned}$$

In addition, for the selection, the unique constraint is essentially the given number of selected sensors; then, there should be an optimal selection scheme in terms of coverage rate.

7.2 Gathering Efficiency and Monitoring Accuracy

Given a DPC function $\mathbf{g}$, the *monitoring accuracy* Υ under a scheme $\mathcal{A}(n, L, \mathcal{S}(\Psi \cdot n))$ can be defined as: for a random variable $\mathrm{M}^n \in \mathcal{M}^n$, let

$$\Upsilon(\mathrm{M}^n) := 1 - \left|\frac{\mathbf{g}_n(\mathrm{M}^n) - \mathbf{g}_{\Psi n}(\mathrm{M}^{\Psi n})}{\mathbf{g}_n(\mathrm{M}^n)}\right|,$$

we say the monitoring accuracy is achieved of Υ if

$$\liminf_{n\to\infty} \Pr(\Upsilon(\mathrm{M}^n) \geq \Upsilon) = 1.$$

As a special scenario, when M^n follows a uniform distribution in the region $\mathcal{M}^n$ and $\Psi \cdot n$ sensors are uniformly selected, we conjecture that $\lim_{n\to\infty} \Upsilon = 1$ for some functions, such as *average* function, if $\Psi = \Theta(1)$.

The rigorous analysis can be implemented based on *statistical theory*, since the $\Psi \cdot n$ sensors are selected as a *sample* of all n sensors.

7.3 Improving Practicality of Schemes

7.3.1 Challenges in Practical Implementation

In this work, we mainly focus on the issue of scaling laws, i.e., scaling of the network performance in the limit when the network gets large, [17], [21]. We aim to investigate the qualitative and fundamental properties of WSNs in terms of

aggregation capacity, [17], [33], [40]. We devise two types of architectural schemes that provide some architectural guidelines on how to design the practical aggregation schemes for WSNs that scale well. It should be emphasized that when implementing the specific schemes under the proposed architecture, we need to take into account some technique details, e.g., the tuning of many parameters and optimization of some preconstants (Table 2) in the system throughput.

7.3.2 Centralized Schemes versus Decentralized Schemes

Since sensor nodes are prone to failure in WSNs [48], a distributed architecture is usually more suitable than a centralized one for the implementation of WSNs due to the greater reliability. While, the schemes proposed in this work are centralized. Then, how to modify these schemes into the ones in the decentralized pattern, if possible, will be an important step toward making the proposed schemes more practical.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments. The research of authors is partially supported by the National Science Foundation for Postdoctoral Scientists of China under grant No. 2012M510118, the National Basic Research Program of China (973 Program) under grants No. 2010CB328101 and No. 2010CB334707, the Program for Changjiang Scholars and Innovative Research Team in University, the Shanghai Key Basic Research Project under grant No. 10DJ1400300, the Expo Science and Technology Specific Projects of China under grant No. 2009BAK43B37, the US National Science Foundation (NSF) CNS-1035894, the Program for Zhejiang Provincial Key Innovative Research Team, and the Program for Zhejiang Provincial Overseas High-Level Talents (One-hundred Talents Program).

REFERENCES

[1] K. Fan, S. Liu, and P. Sinha, "Scalable Data Aggregation for Dynamic Events in Sensor Networks," *Proc. ACM Fourth Int'l Conf. Embedded Networked Sensor Systems (SenSys '06),* 2006.

[2] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks," *ACM Trans. Information and System Security,* vol. 11, no. 4, pp. 1-43, 2008.

[3] T. Gao, R. Jin, T. Xu, and L. Wang, "Energy-Efficient Hierarchical Routing for Wireless Sensor Networks," *Adhoc & Sensor Wireless Networks,* vol. 11, nos. 1/2, pp. 35-72, 2011.

[4] A. Chandrakasan, W. Heinzelman, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *Proc. 33rd Hawaii Int'l Conf. System Sciences (HICSS),* 2000.

[5] A. Chandrakasan, A. Smith, W. Heinzelman, and W. Heinzelman, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Trans. Wireless Comm.,* vol. 1, no. 4, pp. 660-670, Oct. 2002.

[6] Y. Zhuang, J. Pan, and G. Wu, "Energy-Optimal Grid-Based Clustering in Wireless Microsensor Networks with Data Aggregation," *Int'l J. Parallel, Emergent and Distributed Systems,* vol. 25, no. 6, pp. 531-550, 2010.

[7] W. Zhang and G. Cao, "Optimizing Tree Reconfiguration for Mobile Target Tracking in Sensor Networks," *Proc. IEEE INFOCOM '04,* 2004.

[8] W. Zhang and G. Cao, "DCTC: Dynamic Convoy Tree-Based Collaboration for Target Tracking in Sensor Networks," *IEEE Trans. Wireless Comm.,* vol. 3, no. 5, pp. 1689-1701, Sept. 2004.

[9] J. Zhang, X. Shen, G. Dai, Y. Feng, S. Tang, and C. Lv, "Energy-Efficient Lossy Data Aggregation in Wireless Sensor Networks," *Adhoc & Sensor Wireless Networks,* vol. 11, nos. 1/2, pp. 111-135, 2011.

[10] C. Liu and G. Cao, "Minimizing the Cost of Mine Selection via Sensor Networks," *Proc. IEEE INFOCOM '09,* 2009.

[11] L. Mo, Y. He, Y. Liu, J. Zhao, S. Tang, X. Li, and G. Dai, "Canopy Closure Estimates with GreenOrbs: Sustainable Sensing in the forest," *Proc. Seventh ACM Conf. Embedded Networked Sensor Systems (SenSys '09),* 2009.

[12] N. Labraoui, M. Guerroui, M. Aliouat, and T. Zia, "Data Aggregation Security Challenge in Wireless Sensor Networks: A Survey," *Adhoc & Sensor Wireless Networks,* vol. 12, nos. 3/4, pp. 295-324, 2011.

[13] T. Moscibroda, "The Worst-Case Capacity of Wireless Sensor Networks," *Proc. ACM/IEEE Sixth Int'l Conf. Information Processing in Sensor Networks (IPSN '07),* 2007.

[14] P. Gupta and P.R. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. Information Theory,* vol. 46, no. 2, pp. 388-404, Mar. 2000.

[15] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the Many-to-One Transport Capacity of a Dense Wireless Sensor Network and the Compressibility of Its Data," *Proc. ACM Second Int'l Conf. Information Processing In Sensor Networks (IPSN '03),* 2003.

[16] A. Boulis, S. Ganeriwal, and M. Srivastava, "Aggregation in Sensor Networks: An Energy-Accuracy Trade-Off," *Ad Hoc Networks,* vol. 1, nos. 2/3, pp. 317-331, 2003.

[17] A. Giridhar and P. Kumar, "Computing and Communicating Functions over Sensor Networks," *IEEE J. Selected Areas in Comm.,* vol. 23, no. 4, pp. 755-764, Apr. 2005.

[18] A. Giridhar and P. Kumar, "Toward a Theory of In-Network Computation in Wireless Sensor Networks," *IEEE Comm. Magazine,* vol. 44, no. 4, pp. 98-107, Apr. 2006.

[19] L. Ying, R. Srikant, and G. Dullerud, "Distributed Symmetric Function Computation in Noisy Wireless Sensor Networks," *IEEE Trans. Information Theory,* vol. 53, no. 12, pp. 4826-4833, Dec. 2007.

[20] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the Gap in the Capacity of Wireless Networks via Percolation Theory," *IEEE Trans. Information Theory,* vol. 53, no. 3, pp. 1009-1018, Mar. 2007.

[21] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks," *IEEE Trans. Information Theory,* vol. 53, no. 10, pp. 3549-3572, Oct. 2007.

[22] C. Wang, X.-Y. Li, and C. Jiang, "Scaling Laws on Multicast Capacity of Large Scale Wireless Networks," *Proc. IEEE INFOCOM '09,* 2009.

[23] R. Zheng and R. Barton, "Toward Optimal Data Aggregation in Random Wireless Sensor Networks," *Proc. IEEE INFOCOM '07,* 2007.

[24] S. Li, Y. Liu, and X.-Y. Li, "Capacity of Large Scale Wireless Networks under Gaussian Channel Model," *Proc. ACM MobiCom '08,* 2008.

[25] A. Agarwal and P. Kumar, "Capacity Bounds for Ad Hoc and Hybrid Wireless Networks," *ACM SIGCOMM Computer Comm. Rev.,* vol. 34, pp. 71-81, 2004.

[26] G. Grimmett, *Percolation.* Springer Verlag, 1999.

[27] M. Ding, Q. Liang, X. Cheng, and M. Al-Rodhaan, "A Model-Based Approach for Outlier Detection in Sensor Networks," *Adhoc & Sensor Wireless Networks,* vol. 12, nos. 3/4, pp. 275-293, 2011.

[28] M. Penrose, "The Longest Edge of the Random Minimal Spanning Tree," *Annals of Applied Probability,* vol. 7, pp. 340-361, 1997.

[29] C. Chau, M. Chen, and S. Liew, "Capacity of Large-Scale Csma Wireless Networks," *Proc. ACM MobiCom '09,* 2009.

[30] J.-C. Bermond and M.-L. Yu, "Optimal Gathering Algorithms in Multi-Hop Radio Tree-Networks with Interferences," *Adhoc & Sensor Wireless Networks,* vol. 9, nos. 1/2, pp. 109-127, 2010.

[31] C. Wang, X.-Y. Li, C. Jiang, S. Tang, and Y. Liu, "Multicast Throughput for Hybrid Wireless Networks under Gaussian Channel Model," *IEEE Trans. Mobile Computing,* vol. 10, no. 6, pp. 839-852, June 2011.

[32] V. Raman and I. Gupta, "Performance Tradeoffs among Percolation-Based Broadcast Protocols in Wireless Sensor Networks," *Int'l J. Parallel, Emergent and Distributed Systems,* vol. 25, no. 6, pp. 509-530, 2010.

[33] C. Wang, C. Jiang, Y. Liu, X.-Y. Li, S. Tang, and H. Ma, "Aggregation Capacity of Wireless Sensor Networks: Extended Network Case," *Proc. IEEE INFOCOM '11,* 2011.

[34] J. Gao, L. Guibas, N. Milosavljevic, and J. Hershberger, "Sparse Data Aggregation in Sensor Networks," *Proc. ACM Sixth Int'l Conf. Information Processing in Sensor Networks (IPSN '07),* 2007.
[35] X. Li, W. Ya-Jun, and Y. Wang, "Complexity of Data Collection, Aggregation, and Selection for Wireless Sensor Networks," *IEEE Trans. Computers,* vol. 60, no. 3, pp. 386-399, Mar. 2011.
[36] S. Chen, Y. Wang, X. Li, and X. Shi, "Capacity of Data Collection in Randomly-Deployed Wireless Sensor Networks," *Wireless Networks,* vol. 17, no. 2, pp. 305-318, 2011.
[37] P. Wan, C. Scott, L. Wang, Z. Wan, and X. Jia, "Minimum-Latency Aggregation Scheduling in Multihop Wireless Networks," *Proc. ACM MobiHoc '09,* 2009.
[38] X. Wang, L. Fu, X. Tian, Y. Bei, Q. Peng, X. Gan, H. Yu, and J. Liu, "Converge-Cast: On the Capacity and Delay Tradeoffs," *to appear in IEEE Trans. Mobile Computing,* 2011.
[39] L. Fu, Y. Qin, X. Wang, and X. Liu, "Converge-Cast with MIMO," *Proc. IEEE INFOCOM '11,* 2011.
[40] L. Fu, Y. Qin, X. Wang, and X. Liu, "Throughput and Delay Analysis for Convergecast with MIMO in Wireless Networks," *IEEE Trans. Parallel and Distributed Systems,* vol. 23, no. 4, pp. 768-775, Apr. 2012.
[41] P. Wan and C. Yi, "Coverage by Randomly Deployed Wireless Sensor Networks," *IEEE/ACM Trans. Networking,* vol. 14, no. SI, pp. 2658-2669, June 2006.
[42] C. Xu, X. Bai, L. Ding, J. Teng, S. Chellappan, and D. Xuan, "Directed Coverage in Wireless Sensor Networks: Concept and Quality," *Adhoc & Sensor Wireless Networks,* vol. 11, nos. 3/4, pp. 173-197, 2011.
[43] X. Li, A. Nayak, D. Simplot-Ryl, and I. Stojmenovic, "Sensor Placement in Sensor and Actuator Networks," *Wireless Sensor and Actuator Networks,* pp. 263-294, John Wiley & Sons, 2010.
[44] K. Islam and S.G. Akl, "A Local Algorithm to Compute Multiple Connected Dominating Sets in Wireless Sensor Networks," *Int'l J. Parallel, Emergent and Distributed Systems,* vol. 26, no. 5, pp. 369-380, 2011.
[45] H. Zhang and J. Hou, "On the Upper Bound of α-Lifetime for Large Sensor Networks," *ACM Trans. Sensor Networks,* vol. 1, no. 2, pp. 272-300, 2005.
[46] S. Kumar, T. Lai, and J. Balogh, "On k-Coverage in a Mostly Sleeping Sensor Network," *Wireless Networks,* vol. 14, no. 3, pp. 277-294, 2008.
[47] W. Wang, V. Srinivasan, and K.-C. Chua, "Trade-offs Between Mobility and Density for Coverage in Wireless Sensor Networks," *Proc. ACM MobiCom '07,* 2007.
[48] A. Nayak and I. Stojmenović, *Wireless Sensor and Actuator Networks: Algorithms and Protocols for Scalable Coordination and Data Communication.* Wiley-Interscience, 2009.

Cheng Wang received MS degree at Department of Applied Mathematics from Tongji University in 2006 and the PhD degree in Department of Computer Science at Tongji University in 2011. He is currently a postdoc fellow in the Department of Computer Science and Engineering at University of Ottawa, Canada. His research interests include wireless communications and networking, social network analysis, and distributed computing.

Changjun Jiang received the PhD degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995 and conducted postdoctoral research at the Institute of Computing Technology, Chinese Academy of Sciences, in 1997. Currently he is a professor with the Department of Computer Science and Engineering, Tongji University, Shanghai. He is also a council member of China Automation Federation and Artificial Intelligence Federation, the vice director of Professional Committee of Petri Net of China Computer Federation, the vice director of Professional Committee of Management Systems of China Automation Federation, and an Information Area Specialist of Shanghai Municipal Government. His current areas of research are concurrent theory, Petri net, and formal verification of software, concurrency processing, and intelligent transportation systems. He is a member of the IEEE.

Shaojie Tang received the BS degree in radio engineering from Southeast University, China, in 2006 and is working toward the PhD degree in Computer Science Department at the Illinois Institute of Technology since 2006. His current research interests include algorithm design and analysis for wireless ad hoc networks, wireless sensor networks, and online social networks.

Xiang-Yang Li (M'99-SM'08) received the bachelor's degree at the Department of Computer Science and bachelor's degree at the Department of Business Management from Tsinghua University, China, both in 1995, the MS (2000) and PhD (2001) degrees at the Department of Computer Science from the University of Illinois at Urbana-Champaign. He is an associate professor of Computer Science at the Illinois Institute of Technology. He serves as an editor of several journals, including "*IEEE Transactions on Parallel and Distributed Systems (TPDS)*," and "*Networks: An International Journal*." He also serves in Advisory Board of "*Ad Hoc & Sensor Wireless Networks: An International Journal*" from 2005, and "*IEEE CN*" from 2011. He was a guest editor of several special issues, e.g., "*ACM Mobile Networks and Applications*," "*IEEE JSAC*." He published a monograph "*Wireless Ad Hoc and Sensor Networks: Theory and Applications*," in June 2008 by Cambridge University Press. He also coedited the book "*Encyclopedia of Algorithms*," by Springer publisher, as the area editor for mobile computing. He is a senior member of the IEEE.