Exploiting Data Fusion to Improve the Coverage of Wireless Sensor Networks

Rui Tan, *Member, IEEE*, Guoliang Xing, *Member, IEEE*, Benyuan Liu, *Member, IEEE*, Jianping Wang, *Member, IEEE*, and Xiaohua Jia, *Senior Member, IEEE*

***Abstract*—Wireless sensor networks (WSNs) have been increasingly available for critical applications such as security surveillance and environmental monitoring. An important performance measure of such applications is *sensing coverage* that characterizes how well a sensing field is monitored by a network. Although advanced *collaborative* signal processing algorithms have been adopted by many existing WSNs, most previous analytical studies on sensing coverage are conducted based on overly simplistic sensing models (e.g., the disc model) that do not capture the stochastic nature of sensing. In this paper, we attempt to bridge this gap by exploring the fundamental limits of coverage based on stochastic *data fusion* models that fuse *noisy* measurements of multiple sensors. We derive the scaling laws between coverage, network density, and signal-to-noise ratio (SNR). We show that data fusion can significantly improve sensing coverage by exploiting the collaboration among sensors when several physical properties of the target signal are known. In particular, for signal path loss exponent of k (typically between 2.0 and 5.0), $\rho_{\rm f} = \mathcal{O}(\rho_{\rm d}^{1-1/k})$, where $\rho_{\rm f}$ and $\rho_{\rm d}$ are the densities of uniformly deployed sensors that achieve full coverage under the fusion and disc models, respectively. Moreover, data fusion can also reduce network density for regularly deployed networks and mobile networks where mobile sensors can relocate to fill coverage holes. Our results help understand the limitations of the previous analytical results based on the disc model and provide key insights into the design of WSNs that adopt data fusion algorithms. Our analyses are verified through extensive simulations based on both synthetic data sets and data traces collected in a real deployment for vehicle detection.**

***Index Terms*—Coverage, data fusion, mobility, performance limits, target detection, wireless sensor network (WSN).**

I. Introduction

RECENT years have witnessed the deployments of wireless sensor networks (WSNs) for many critical applications such as security surveillance [1], environmental monitoring [2], and target detection/tracking [3]. Many of these applications involve a large number of sensors distributed in a vast geographical area. As a result, the cost of deploying these networks into the physical environment is high. A key challenge is thus to predict and understand the expected sensing performance of these WSNs. A fundamental performance measure of WSNs is *sensing coverage* that characterizes how well a sensing field is monitored by a network. Many recent studies are focused on analyzing the coverage performance of large-scale WSNs [4]–[10].

Despite the significant progress, a key challenge faced by the research on sensing coverage is the obvious discrepancy between the advanced information processing schemes adopted by existing sensor networks and the overly simplistic sensing models widely assumed in the previous analytical studies. On the one hand, many WSN applications are designed based on *collaborative* signal processing algorithms that improve the sensing performance of a network by jointly processing the noisy measurements of multiple sensors. In practice, various stochastic *data fusion* schemes have been employed by sensor network systems for event monitoring, detection, localization, and classification [1], [3], [11]–[16]. On the other hand, collaborative signal processing algorithms such as data fusion often have complex complications to the network-level sensing performance such as coverage. As a result, most analytical studies on sensing coverage are conducted based on *overly simplistic* sensing models [4]–[8], [10], [17]–[21]. In particular, the sensing region of a sensor is often modeled as a disc with radius r centered at the position of the sensor, where r is referred to as the *sensing range*. A sensor *deterministically* detects the targets (events) within its sensing range. Although such a model allows a geometric treatment to the coverage problem, it fails to capture the stochastic nature of sensing.

To illustrate the inaccuracy of the disc sensing model, we plot the sensing performance of an acoustic sensor in Fig. 1 using the data traces collected from a real vehicle detection experiment [11]. In the experiment, the sensor detects moving vehicles by comparing its signal energy measurement against a threshold (denoted by t). Fig. 1(a) plots the probability that the sensor detects a vehicle (denoted by $P_{\rm D}$) versus the distance from the vehicle. No clear cutoff boundary between successful and unsuccessful sensing of the target can be seen in Fig. 1(a). A similar result is observed for the relationship between the sensor's false alarm rate (denoted by $P_{\rm F}$) and the detection threshold shown in Fig. 1(b). Note that $P_{\rm F}$ is the probability of making a positive decision when *no* vehicle is present.

In this paper, we develop an analytical framework to explore the fundamental limits of coverage of large-scale WSNs based on stochastic data fusion models. To characterize the inherent

Manuscript received October 05, 2009; revised November 15, 2010; accepted June 26, 2011; approved by IEEE/ACM Transactions on Networking Editor K. Kar. Date of publication September 01, 2011; date of current version April 12, 2012. This work was supported in part by the National Science Foundation under Grants CNS-0954039 (CAREER) and ECCS-1029683.

R. Tan and G. Xing are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48823 USA (e-mail: tanrui@msu.edu; glxing@msu.edu).

B. Liu is with the Department of Computer Science, University of Massachusetts Lowell, Lowell, MA 01854 USA (e-mail: bliu@cs.uml.edu).

J. Wang and X. Jia are with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: jianwang@cityu.edu.hk; csjia@cityu.edu.hk).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2011.2164620

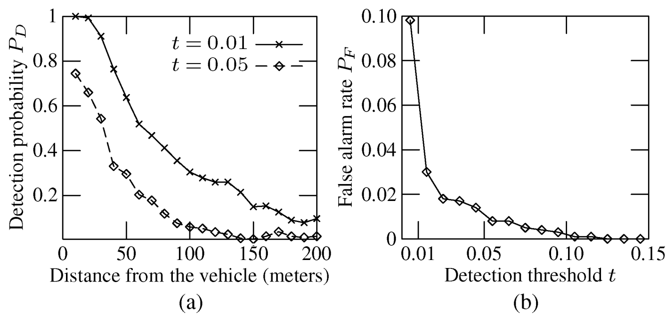


Fig. 1. Sensing performance of an acoustic sensor in detecting vehicle. (a) Detection probability versus the distance from the vehicle. (b) False alarm rate versus detection threshold.

stochastic nature of sensing, we propose a new coverage measure called (α,β)-coverage, where α and β are the upper and lower bounds on the system false alarm rate and detection probability, respectively. Compared to the classical definition of coverage, (α,β)-coverage explicitly captures the performance requirements imposed by sensing applications. For instance, the full (0.05, 0.9)-coverage of a region ensures that the probability of detecting any event occurring in the region is no lower than 90% and that no more than 5% of the network reports are false alarms.

The main focus of this paper is to investigate the fundamental scaling laws between coverage, network density, and signal-to-noise ratio (SNR). To the best of our knowledge, this work is the first to study the coverage performance of large-scale WSNs based on collaborative sensing models. Our results not only help understand the limitations of the existing analytical results based on the disc model, but also provide key insights into designing and analyzing the large-scale WSNs that adopt stochastic fusion algorithms. The main contributions of this paper are as follows.

- We derive the (α,β)-coverage of random networks under both data fusion and probabilistic disc models. The existing analytical results based on the classical disc model can be naturally extended to the context of stochastic event detection. With these results, we can compute the minimum network density before the deployment or turn on the fewest sensors of an existing network to achieve a desired level of sensing coverage.
- We study the fundamental scaling laws of (α,β)-coverage. Let ρ_d and ρ_f denote the minimum network densities for achieving full coverage under the disc and fusion models, respectively. For randomly deployed networks, we prove that $\rho_\mathrm{f} = \mathcal{O}((2r^2/R^2)\cdot\rho_\mathrm{d})$, where r is the radius of sensing disc and R is the fusion range within which the measurements of all sensors are fused.[1] As fusion range can be much greater than sensing range, ρ_f is much smaller than ρ_d. Furthermore, when the optimal fusion range is adopted, $\rho_\mathrm{f} = \mathcal{O}(\rho_\mathrm{d}^{1-1/k})$, where k is the signal's path loss exponent that typically ranges from 2.0 to 5.0. In particular, when $k = 2$ (which typically holds for acoustic signals), $\rho_\mathrm{f} = \mathcal{O}(\sqrt{\rho_\mathrm{d}})$. For regularly deployed networks (e.g., grid networks), we prove that $\rho_\mathrm{f} \le \rho_\mathrm{d}$. The above results show that data fusion can effectively reduce the network density compared to the disc model. Furthermore, the existing analytical results based on the disc model significantly overestimate the network density required for achieving coverage.
- We study the impact of SNR on the network density when full coverage is required. Both for randomly and regularly deployed networks, we prove that $\rho_\mathrm{f}/\rho_\mathrm{d} = \mathcal{O}(\mathrm{SNR}^{2/k})$. This result suggests that data fusion is more effective in reducing the density of low-SNR network deployments, while the disc model is suitable only when the SNR is sufficiently high.
- Besides static networks, we also study the coverage performance of mobile networks, in which mobile sensors relocate themselves to fill coverage holes after the initial deployment. We extend a relocation strategy that is based on the disc model [18] to the data fusion model. We show that data fusion results in lower network density without increasing the moving distance of mobile sensors.
- To verify our analyses, we conduct extensive simulations based on both synthetic data sets and real data traces collected from 20 sensors. The simulation results validate our analytical results under a variety of realistic settings.

Table I summarizes the main results of this paper.

TABLE I
NETWORK DENSITIES FOR ACHIEVING FULL COVERAGE

Deployment	Random		Regular / Mobile
Density relationship	Fixed R	Optimal R	$\rho_f \le \rho_d$
	$\rho_f = \mathcal{O}\left(\frac{2r^2}{R^2}\cdot\rho_d\right)$	$\rho_f = \mathcal{O}\left(\rho_d^{1-1/k}\right)$	
Impact of SNR	$\frac{\rho_f}{\rho_d} = \mathcal{O}\left(\mathrm{SNR}^{2/k}\right)$		

II. RELATED WORK

Many sensor network systems have incorporated various data fusion schemes to improve the system performance. In the surveillance system based on MICA2 motes [1], the system false alarm rate is reduced by fusing the detection decisions made by multiple sensors. In the DARPA SensIT project [11], advanced data fusion techniques have been employed in a number of algorithms and protocols designed for target detection [3], [13], localization [14], [15], and classification [11], [12]. Despite the wide adoption of data fusion in practice, the performance analysis of large-scale fusion-based WSNs has received little attention.

There is vast literature on stochastic signal detection based on multisensor data fusion. Early works [22], [23] focus on small-scale powerful sensor networks (e.g., several radars). The theories on decentralized detection are surveyed in [24]. Recent studies on data fusion have considered the specific properties of WSNs such as sensors' spatial distribution [11], [12], [16], limited sensing/communication capability [13], and sensor failure [25]. However, these studies focus on analyzing the optimal fusion strategies that maximize the system performance

[1]We adopt the following asymptotic notation: 1) $f(x) = \mathcal{O}(g(x))$ means that $g(x)$ is the asymptotic upper bound of $f(x)$; 2) $f(x) = \Theta(g(x))$ means that $g(x)$ is the asymptotic tight bound of $f(x)$.

of a given network. In contrast, this paper explores the fundamental limits of sensing coverage of WSNs that are designed based on existing data fusion strategies.

As one of the most fundamental issues in WSNs, the coverage problem has attracted significant research attention. Previous works fall into two categories—namely, coverage maintenance algorithms/protocols and theoretical analysis of coverage performance. These two categories are reviewed briefly as follows, respectively.

Early work [26]–[28] quantifies sensing coverage by the length of target's path where the accumulative observations of sensors are maximum or minimum [26]–[28]. However, these works focus on devising algorithms for finding the target's paths with certain level of coverage. Several algorithms and protocols [9], [29] are designed to maintain sensing coverage using the minimum number of sensors. However, the effectiveness of these schemes largely relies on the assumption that sensors have circular sensing regions and deterministic sensing capability. Several recent studies [30]–[33] on the coverage problem have adopted probabilistic sensing models. The numerical results in [33] show that the coverage of a network can be expanded by the cooperation of sensors through data fusion. However, these studies do not quantify the improvement of coverage due to data fusion techniques. Different from our focus on analyzing the fundamental limits of coverage in WSNs, all of these studies aim to devise algorithms and protocols for coverage maintenance.

Theoretical studies of the coverage of large-scale WSNs have been conducted in [4]–[8], [10], and [19]–[21]. Most works [5]–[8], [20], [21] focus on deriving the asymptotic coverage of WSNs. The critical conditions for full k-coverage (i.e., any physical point is within the sensing range of at least k sensors) over a bounded square area [5]–[8] or barrier area [20], [21] are derived for various sensor deployment strategies. The coverage of randomly deployed networks is studied in [10]. The existing theoretical results on coverage for both static and mobile sensors/targets are surveyed in [4]. However, all the above theoretical studies are based on the deterministic disc model. In this paper, we compare our results obtained under a data fusion model against the results from [4] and [10].

Recent works [18], [34], [35] have exploited sensor mobility to reduce network density in achieving coverage. In such a scheme, randomly distributed mobile sensors can relocate themselves to fill coverage holes in the initial network deployment. A sensor relocation strategy is proposed in [18] to achieve full coverage with bounded moving distance of mobile sensors. In this paper, we extend the strategy to the data fusion model. The coverage of mobile WSNs with random sensor mobility has been studied based on the disc model in [36]. In this paper, we focus on quantifying the improvement of coverage in the mobile networks with limited sensor mobility due to data fusion.

III. Background and Problem Definition

In this section, we first describe the preliminaries of our work, which include sensor measurement, network, and data fusion models. We then introduce the problem definition.

A. Sensor Measurement and Network Models

We assume that sensors perform detection by measuring the energy of signals emitted by the target.[2] The energy of most physical signals (e.g., acoustic and electromagnetic signals) attenuates with the distance from the signal source. Suppose sensor i is d_i meters away from the target that emits a signal of energy S. The attenuated signal energy s_i at the position of sensor i is given by $s_i = S \cdot w(d_i)$, where $w(\cdot)$ is a decreasing function satisfying $w(0) = 1$, $w(\infty) = 0$, and $w(x) = \Theta(x^{-k})$. The $w(\cdot)$ is referred to as the *signal decay function*. Depending on the environment, e.g., atmosphere conditions, the signal's path loss exponent k typically ranges from 2.0 to 5.0 [15], [37]. We note that the theoretical results derived in this paper do not depend on the closed-form formula of $w(\cdot)$. We adopt the following signal decay function in the simulations conducted in this paper:

$$w(x) = \frac{1}{1+x^k}. \tag{1}$$

The sensor measurements are contaminated by additive random noises from sensor hardware or environment. Depending on the hypothesis that the target is absent (H_0) or present (H_1), the measurement of sensor i, denoted by y_i, is given by

$$H_0 : y_i = n_i \quad H_1 : y_i = s_i + n_i$$

where n_i is the energy of noise experienced by sensor i. We assume that the noise n_i at each sensor i follows the normal distribution, i.e., $n_i \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are the mean and variance of n_i, respectively. We assume that the noises, $\{n_i | \forall i\}$, are spatially independent across sensors. Therefore, the noises at sensors are independent and identically distributed (*i.i.d.*) Gaussian noises. In the presence of target, the measurement of sensor i follows the normal distribution, i.e., $y_i | H_1 \sim \mathcal{N}(s_i + \mu, \sigma^2)$. Due to the independence of noises, the sensors' measurements, $\{y_i | \forall i, H_1\}$, are spatially independent but *not* identically distributed as sensors receive different signal energies from the target. We define the peak signal-to-noise ratio (PSNR) as $\delta = S/\sigma$, which quantifies the noise level. The symbols used in this paper are summarized in the supplementary file containing appendixes of this paper.[3]

The above signal decay and additive *i.i.d.* Gaussian noise models have been widely adopted in the literature of multisensor signal detection [10], [13]–[16], [22], [23], [28], [32], [33] and also have been empirically verified [15], [37]. In practice, the parameters of these models (i.e., S, $w(\cdot)$, μ, and σ^2) can be estimated using the training data collected by the existing WSN or several *in situ* sensors before the large-scale deployment. The normal distribution might be an approximation to the real noise distribution in practice. As discussed in Appendix-E of the supplementary file, the assumption of *i.i.d.* Gaussian noises can be relaxed to any *i.i.d.* noises.

[2]Several types of sensors (e.g., acoustic sensor) only sample *signal intensity* at a given sampling rate. The *signal energy* can be obtained by preprocessing the time series of a given interval, which has been commonly adopted to avoid the transmission of raw data [11]–[15].

[3]Due to space limitations, all appendixes are omitted and can be found in the supplementary file of this paper.

Sensors are deployed in a vast two-dimensional geographical region. We consider two deployment schemes, i.e., *random* and *regular* networks. In the random networks, the positions of sensors are uniformly and independently distributed in the region. Such a deployment scenario can be modeled as a stationary two-dimensional Poisson point process. Let ρ denote the density of the underlying Poisson point process. The number of sensors located in a region A, $N(A)$, follows the Poisson distribution with mean of $\rho\|A\|$, i.e., $N(A) \sim \mathrm{Poi}(\rho\|A\|)$, where $\|A\|$ represents the area of the region A. We note that the uniform sensor distribution has been widely adopted in the performance analysis of large-scale WSNs [4]–[7], [10]. Therefore, this assumption allows us to compare our results to previous analytical results. In the regular networks, sensors are deployed at grid points in the vast region. Note that the grid deployment has also been widely adopted in theoretical studies [7], [18] and real systems [38].

B. Data Fusion Model

Data fusion can improve the performance of detection systems by jointly considering the noisy measurements of multiple sensors. There exist two basic data fusion schemes—namely, *decision fusion* and *value fusion*. In decision fusion, each sensor makes a *local* decision based on its measurements and sends its decision to the cluster head, which makes a *system* decision according to the local decisions. The optimal decision fusion rule has been obtained in [22]. In value fusion, each sensor sends its measurements to the cluster head, which makes the detection decision based on the received measurements. In this paper, we focus on value fusion, as it usually has better detection performance than decision fusion [23]. We will discuss how to extend the results of this paper to address a decision fusion model in Appendix-E of the supplementary file. Under the assumptions made in Section III-A, the optimal value fusion rule is to compare a weighted sum of sensors' measurements, i.e., $\sum_i (s_i/\sigma) \cdot y_i$, to a threshold. The derivation of this optimal value fusion rule is in Appendix-B of the supplementary file. However, as sensor measurements contain both noise and signal energy, the weight s_i/σ, i.e., the SNR received by sensor i, is unknown. A practical solution is to adopt equal constant weights for all sensors' measurements [13], [16], [33]. Since the measurements from different sensors are treated equally, the sensors far away from the target should be excluded from data fusion as their measurements suffer low SNRs. Therefore, we adopt a fusion scheme as follows.

For any physical point p, the sensors within a distance of R meters from p form a cluster and fuse their measurements to detect whether a target is present at p. R is referred to as the *fusion range*, and $\mathbf{F}(p)$ denotes the set of sensors within the fusion range of p. The number of sensors in $\mathbf{F}(p)$ is represented by $N(p)$. A cluster head is elected to make the detection decision by comparing the sum of measurements reported by member sensors in $\mathbf{F}(p)$ against a detection threshold T. Let Y denote the *fusion statistic*, i.e., $Y = \sum_{i \in \mathbf{F}(p)} y_i$. If $Y \geq T$, the cluster head decides H_1; otherwise, it decides H_0.

We assume that the cluster head makes a detection based on snapshot measurements from member sensors without using temporal samples to refine the detection decision. Note that such a snapshot scheme is widely adopted in previous works on target surveillance [13]–[16], [33]. Fusion range R is an important design parameter of our data fusion model. As SNR received by sensor decays with distance from the target, fusion range lower-bounds the quality of information that is fused at the cluster head. In Section V-B, we will discuss how to choose the optimal fusion range. The above data fusion model is consistent with the fusion schemes adopted in [13], [16], and [33]. If more efficient fusion models are employed, the scaling laws proved in this paper still hold as discussed in Appendix-E of the supplementary file.

We assume that the target keeps stationary after appearance and the position of a possible target can be obtained through a localization algorithm. For instance, the target position can be estimated as the geometric center of a number of sensors with the largest measurements. Such a simple localization algorithm is employed in the simulations conducted in this paper. The localized position may not be the exact target position, and the distance between them is referred to as *localization error*. We assume that the localization error is upper-bounded by a constant ϵ. The localization error is accounted for in the following analyses. However, we show that it has no impact on the asymptotic results derived in this paper.

The above data fusion model can be used for target detection as follows. The detection can be executed periodically or triggered by user queries. In a detection process, each sensor makes a snapshot measurement, and a cluster is formed by the sensors within the fusion range from the possible target to make a detection decision. The cluster formation may be initiated by the sensor that has the largest measurement. Such a scheme can be implemented by several dynamic clustering algorithms [39]. The fusion range R can be used as an input parameter of the clustering algorithm. The communication topology of the cluster can be a multihop tree rooted at the cluster head. As the fusion statistic Y is an aggregation of sensors' measurements, it can be computed efficiently along the routing path to the cluster head. In Section VIII-B, we will discuss the delay of aggregating sensors' measurements.

C. Problem Definition

The detection of a target is inherently stochastic due to the noise in sensor measurements. The detection performance is usually characterized by two metrics—namely, the false alarm rate (denoted by P_{F}) and detection probability (denoted by P_{D}). P_{F} is the probability of making a positive decision when *no* target is present, and P_{D} is the probability that a present target is correctly detected. In stochastic detection, positive detection decisions may be false alarms caused by the noise in sensor measurements. In particular, although the detection probability can be improved by setting lower detection thresholds, the fidelity of detection results may be unacceptable because of high false alarm rates. Therefore, P_{F} together with P_{D} characterize the sensing quality provided by the network. For a physical point p, we denote the probability of successfully detecting a target located at p as $P_{\mathrm{D}}(p)$. Note that P_{F} is the probability of making positive decision when *no* target is present, and hence is location-independent.

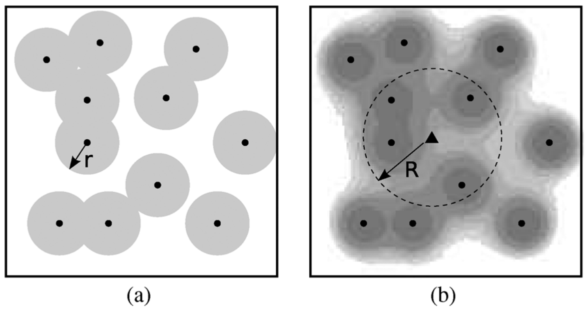


Fig. 2. (a) Coverage under the disc model, where $r = 17$ m that is computed by (2). (b) Coverage under the fusion model, where grayscale represents $P_{\rm D}$.

Our focus is to study the coverage of large-scale WSNs. We introduce a concept called (α, β)-*coverage* that quantifies the fraction of the surveillance region where $P_{\rm F}$ and $P_{\rm D}$ are bounded by α and β, respectively.

Definition 1 ((α, β)-Coverage): Given two constants $\alpha \in (0,1)$ and $\beta \in (0,1)$ where $\alpha < \beta$, a physical point p is (α, β)-*covered* if the false alarm rate $P_{\rm F}$ and detection probability $P_{\rm D}(p)$ satisfy

$$P_{\rm F} \leq \alpha \quad P_{\rm D}(p) \geq \beta.$$

The (α, β)-coverage of a region is defined as the fraction of points in the region that are (α, β)-covered.

The *full coverage* of a region refers to the case where the (α, β)-coverage of the region approaches one, i.e., the false alarm rate is below α and the probability of detecting a target present at *any* location is above β. Under the Gaussian noise assumption made in Section III-A, $P_{\rm D} \geq P_{\rm F}$ for any detector. Hence, the requirement of α and β where $\alpha \geq \beta$ can be always met, which results in the meaningless full coverage. The details of this issue will be discussed in Section IV. In practice, mission-critical surveillance applications [1], [11], [38], [40] require a low false alarm rate ($\alpha < 5\%$) and a high detection probability ($\beta \gg 50\%$).

We now illustrate the (α, β)-coverage by an example, where $\delta = 1000$ (i.e., 30 dB), $\alpha = 5\%$, $\beta = 95\%$, and $R = 50$ m. Fig. 2 illustrates the coverage under the disc and fusion models. In Fig. 2(b), when a target (represented by the triangle) is present, the sensors within the fusion range from it fuse their measurements to make a detection. The gray area is (α, β)-covered, where grayscale represents the value of $P_{\rm D}$ at each point. As shown in Fig. 2(a), the covered region under the disc model is simply the union of all sensing discs. As a result, when a high level of coverage is required, a large number of extra sensors must be deployed to eliminate small uncovered areas surrounded by sensing discs. In contrast, data fusion can effectively expand the covered region by exploiting the collaboration among neighboring sensors.

In the rest of this paper, we consider the following problems.

1) Although a number of analytical results on coverage [4]–[10], [29] have been obtained under the classical disc model, are they still applicable under the definition of (α, β)-coverage that explicitly captures the stochastic nature of sensing? To answer this question, we propose a probabilistic disc model such that the existing results can be naturally extended to the context of stochastic detection (Section IV).
2) How to quantify the (α, β)-coverage when sensors can collaborate through data fusion? Answering this question enables us to evaluate the coverage performance of a network. Moreover, it allows us to deploy or turn on fewest sensors for achieving a given level of coverage (Section V).
3) What are the scaling laws between coverage, network density, and SNR under both the disc and fusion models? The results will provide important insights into understanding the limitation of analytical results based on the disc model and the impact of data fusion on the coverage of large-scale random (Section VI) and regular/mobile (Section VII) WSNs.

IV. Coverage Under Probabilistic Disc Model

As the classical disc model deterministically treats the detection performance of sensors, existing results based on this model [4]–[10], [29], cannot be readily applied to analyze the performance or guide the design of real-world WSNs. In this section, we extend the classical disc model based on the stochastic detection theory [23] to capture several realistic sensing characteristics and study the (α, β)-coverage under the extended model.

In the *probabilistic disc model*, we choose the sensing range r such that: 1) the probability of detecting any target within the sensing range is no lower than β; and 2) the false alarm rate is no greater than α. As the probabilistic disc model ignores the detection probability outside the sensing range of a sensor, the detection capability of a sensor under this model is lower than in reality. However, this model preserves the *boundary* of sensing region defined in the classical disc model. Hence, the existing results based on the classical disc model [4]–[10], [29] can be naturally extended to the context of stochastic detection.

We now discuss how to choose the sensing range r under the probabilistic disc model. The optimal Bayesian detection rule for a single sensor i is to compare its measurement y_i to a detection threshold t [23]. If y_i exceeds t, sensor i decides H_1; otherwise, it decides H_0. Therefore, the $P_{\rm F}$ and $P_{\rm D}$ of sensor i are given by $P_{\rm F} = \mathbb{P}(y_i \geq t | H_0) = Q((t-\mu)/\sigma)$ and $P_{\rm D} = \mathbb{P}(y_i \geq t | H_1) = Q((t - \mu - s_i)/\sigma)$, where $\mathbb{P}(\cdot)$ is the probability notation and $Q(\cdot)$ is the complementary cumulative distribution function of the standard normal distribution, i.e., $Q(x) = (1/\sqrt{2\pi}) \int_x^{\infty} {\rm e}^{-t^2/2} {\rm d}t$. As $P_{\rm D}$ is nondecreasing function of $P_{\rm F}$ [23], it is maximized when $P_{\rm F}$ is set to be the upper bound α. Hence, the optimal detection threshold can be solved from $P_{\rm F} = \alpha$ as $t_{\rm opt} = \mu + \sigma Q^{-1}(\alpha)$, where $Q^{-1}(\cdot)$ is the inverse function of $Q(\cdot)$. By solving $P_{\rm D} = \beta$ where $t = t_{\rm opt}$ and $s_i = S \cdot w(r)$, we have

$$r = w^{-1}\left(\frac{Q^{-1}(\alpha) - Q^{-1}(\beta)}{\delta}\right) \quad (2)$$

where $w^{-1}(\cdot)$ is the inverse function of $w(\cdot)$. If the target is more than r meters from the sensor, the detection performance requirements, i.e., α and β, cannot be satisfied by setting any detection threshold. Note that a similar definition of sensing range is proposed in [33] for stochastic detection. From (2), the sensing range of a sensor varies with the user requirements (i.e.,

α and β) and PSNR δ. For instance, the sensing range r is 3.8 m if $\alpha = 5\%, \beta = 95\%, \delta = 50$ (i.e., 17 dB), and $w(\cdot)$ is given by (1) with $k = 2$. Note that the setting $\delta = 50$ is consistent with the measurements from the vehicle detection experiments based on MICA2 [40] and ExScal [38] motes. As $w(\cdot)$ is a decreasing function, $w^{-1}(\cdot)$ is also a decreasing function. Therefore, r increases with the PSNR δ according to (2). This conforms to the intuition that a sensor can detect a farther target if the noise level is lower (i.e., a greater δ). Note that if $\alpha \geq \beta$, by setting $t = t_{\rm opt}$, the requirement $P_{\rm F} \leq \alpha$ and $P_{\rm D} \geq \beta$ can be always satisfied where the target can appear at any position. Hence, the sensing range r is infinity, which results in the meaningless full coverage. Therefore, we have the constraint $\alpha < \beta$ in Definition 1.

We now extend the coverage of random networks [4], [10] derived under the classical disc model to (α, β)-coverage. Under both the classical and probabilistic disc models, a location is regarded as being covered if it is within at least one sensor's sensing range. Accordingly, the area of the union of all sensors' sensing ranges is regarded as being covered by the network. The coverage of random networks under the classical disc model has been extensively studied based on the stochastic geometry theory [4], [10]. Specifically, the coverage of a network deployed according to a Poisson point process of density ρ is given by

$$c = 1 - \mathrm{e}^{-\rho\pi r^2}. \tag{3}$$

If the sensing range r is chosen by (2), (3) computes the (α, β)-coverage of a random network under the probabilistic disc model. This result will be used as the basis for studying the impact of data fusion on network coverage in Section VI.

For regular networks, it has been shown in [18] that the necessary and sufficient condition for achieving full coverage under the classical disc model is $l = \sqrt{2}r$, where l is the grid side length. Hence, under the probabilistic disc model, the minimum network density for achieving full coverage of a regular network is

$$\rho = \frac{1}{l^2} = \frac{1}{2r^2} \tag{4}$$

where r is given by (2).

Although the probabilistic disc model captures the stochastic nature of sensing, it has two similar major limitations as the classical disc model. First, as the disc model ignores the sensing capability outside the sensing range of a sensor, it cannot accurately quantify the real sensing performance of a sensor. Second, as the disc model does not exploit the collaboration among sensors, the existing analytical results based on the disc model may significantly underestimate the system sensing performance that a WSN can achieve. The above two results [i.e., (3) and (4)] will be used as the baselines to study the impact of data fusion on coverage of random and regular networks in Sections VI and VII, respectively.

V. Coverage of Random Networks Under Data Fusion Model

In this section, we first derive the (α, β)-coverage of random networks under the fusion model, then illustrate the analytical results using numerical examples.

A. Deriving Coverage of Random Networks Under Data Fusion Model

We have the following lemma regarding the (α, β)-coverage of random networks. Due to space limitations, all proofs are omitted and can be found in Appendix-C from the supplementary file of this paper.

Lemma 1: The (α, β)-coverage of a uniformly deployed network under the data fusion model, denoted by c, is

$$c = \mathbb{P}\left(\frac{\sum_{i\in\mathbf{F}(p)} s_i}{\sqrt{N(p)}} \geq \sigma\left(Q^{-1}(\alpha) - Q^{-1}(\beta)\right)\right) \tag{5}$$

where p is an arbitrary physical point in the network.

As p is an arbitrary point in the network, $N(p)$ is a Poisson random variable, i.e., $N(p) \sim \mathrm{Poi}(\rho\pi R^2)$. Moreover, $\{s_i | i \in \mathbf{F}(p)\}$ are also random variables. However, we have no closed-form formula for computing (5) due to the difficulty of deriving the cumulative distribution function of $\sum_{i\in\mathbf{F}(p)} s_i/\sqrt{N(p)}$. We now give an approximation to (5) in the following lemma.

Lemma 2: Let μ_s and σ_s^2 denote the mean and variance of $s_i | i \in \mathbf{F}(p)$ for arbitrary point p, respectively. The (α, β)-coverage of a uniformly deployed network under the data fusion model can be approximated by

$$c \simeq Q\left(\frac{\gamma(R) - \rho\pi R^2}{\sqrt{\rho\pi R^2}}\right) \tag{6}$$

where

$$\gamma(R) = \left(\frac{Q^{-1}(\alpha)\sigma - Q^{-1}(\beta)\sqrt{\sigma_s^2 + \sigma^2}}{\mu_s}\right)^2.$$

The derivation of μ_s and σ_s^2 as well as the proof of Lemma 2 are in Appendix-C.2 of the supplementary file. The μ_s and σ_s^2 are given by $\mu_s = (2S/R^2) \cdot \int_0^R x w(x) \mathrm{d}x$ and $\sigma_s^2 = (2S^2/R^2) \int_0^R x w^2(x) \mathrm{d}x - \mu_s^2$. As Central Limit Theorem (CLT) is applied in the derivation of (6), this approximation is accurate when $N(p) \geq 20$ [41]. This condition can be easily met in many applications. For example, it is shown in [40] that the detection probability is only about 40% when four MICA2 motes are deployed in a 10×10 m^2 region. Suppose $R = 20$ m and the network density is the same as in [40], $N(p)$ will be about 50. With the approximate formula, we can evaluate the coverage performance of an existing network or compute the minimum network density to achieve the desired level of coverage under the fusion model. Our simulation results in Section IX show that (6) can provide accurate prediction of coverage under the fusion model. We note that the localization error has little impact on the accuracy of the approximate formula when $R \gg \epsilon$. Recent sensor network localization protocols can achieve a precision within 0.5 m in large-scale outdoor deployments [42].

We now derive the lower bound of (α, β)-coverage under the fusion model, which will be used in the derivations of scaling laws in Section VI. We denote $F_{\rm Poi}(\cdot|\lambda)$ as the cumulative distribution function of the Poisson distribution $\mathrm{Poi}(\lambda)$, which is formally given by $F_{\rm Poi}(x|\lambda) = \sum_{k=0}^{\lfloor x \rfloor} (\mathrm{e}^{-\lambda}\lambda^k/k!)$.

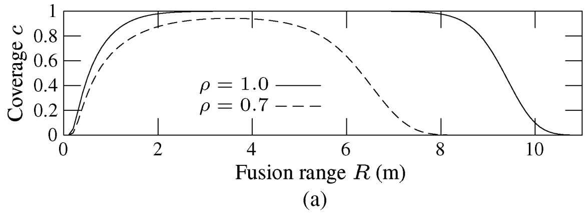


Fig. 3. Coverage versus fusion range ($\alpha = 5\%, \beta = 95\%$). (a) $\delta = 4$. (b) $\delta = 100$.

Lemma 3: The lower bound of (α, β)-coverage of a uniformly deployed network under the data fusion model, denoted by c_{L}, is given by

$$c_{\mathrm{L}} = 1 - F_{\mathrm{Poi}}\left(\Gamma(R) | \rho\pi R^2\right) \tag{7}$$

where

$$\Gamma(R) = \left(\frac{Q^{-1}(\alpha) - Q^{-1}(\beta)}{\delta}\right)^2 \cdot \frac{1}{w^2(R+\epsilon)}. \tag{8}$$

When $\rho\pi R^2$ is large enough

$$c_{\mathrm{L}} = Q\left(\frac{\Gamma(R) - \rho\pi R^2}{\sqrt{\rho\pi R^2}}\right). \tag{9}$$

It is important to note that $N(p) \geq \Gamma(R)$ is a sufficient condition that point p is (α, β)-covered.

B. Numerical Examples

In this section, we provide several numerical results to help understand the coverage performance of random networks under the data fusion model. We adopt the signal decay function given by (1) with $k = 2$. Fig. 3 plots the approximate coverage computed by (6) for various PSNRs. We can see from Fig. 3 that the coverage initially increases with fusion range R, but decreases to zero eventually. Intuitively, as the fusion range increases, more sensors contribute to the data fusion resulting in better sensing quality. However, as R becomes very large, the aggregate noise starts to cancel out the benefit because the target signal decreases quickly with the distance from the target. In other words, the measurements of sensors far away from the target contain low-quality information, and hence fusing them leads to lower detection performance. Moreover, we can see the same behavior for different settings of PSNR. An important question is thus how to choose the optimal fusion range (denoted by R_{opt}) that maximizes the coverage. First, the R_{opt} can be obtained through numerical experiments. Fig. 4 plots the optimal fusion ranges under different network densities, which are obtained by numerically maximizing the coverage. Second, it is possible to obtain the analytical R_{opt} by solving $\mathrm{d}c/\mathrm{d}R = 0$. For instance, when the signal decay function $w(\cdot)$ is given by (1) with $k = 2$, R_{opt} satisfies $R_{\mathrm{opt}} / \ln R_{\mathrm{opt}} = \Theta(\sqrt{\rho})$, and hence R_{opt} increases with network density ρ. The intuition behind this increasing relationship is as follows. Due to the increased network density, the effect to system sensing performance brought by the increased sensors in the ring area between the original fusion region (i.e., πR^2) and an enlarged fusion region (i.e., $\pi(R + \Delta R)^2$) is positive. Hence, increasing fusion range is beneficial when the network density is increased.

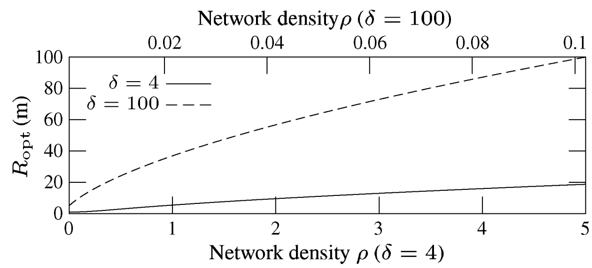


Fig. 4. Optimal fusion range versus density ($\alpha = 5\%, \beta = 95\%$).

VI. Impact of Data Fusion on Coverage of Random Networks

Many mission-critical applications require a high level of coverage over the surveillance region. As an asymptotic case, full coverage is required, i.e., *any* target/event present in the region can be detected with a probability of at least β while the false alarm rate is below α. For random networks, a higher level of coverage always requires more sensors. Therefore, the network density for achieving full coverage is an important cost metric for mission-critical applications.

Under the disc model, the sensing regions of randomly deployed sensors inevitably overlap with each other when a high-level coverage is required. According to (3), we have $\mathrm{d}\rho = (1/\pi r^2) \cdot (1/(1-c)) \cdot \mathrm{d}c$. If c is close to 1, a large number of extra sensors (i.e., $\mathrm{d}\rho$) are required to eliminate a small uncovered area (i.e., $\mathrm{d}c$). Moreover, the situation gets worse when c increases. In this section, we are interested in how much network density can be reduced by adopting data fusion. Specifically, we study the asymptotic relationships between the network densities for achieving full coverage under the probabilistic disc and data fusion models. The results provide important insights into understanding the limitation of the disc model and the impact of data fusion on coverage of random networks.

A. Full Coverage of Random Networks Using Fixed Fusion Range

We first study the relationship between the network densities for achieving full coverage under the disc and fusion models when fusion range R is a constant. We have the following theorem.

Theorem 1: For uniformly deployed networks, let ρ_{d} and ρ_{f} denote the minimum network densities required to achieve the

(α, β)-coverage of c under the disc and fusion models, respectively. If the fusion range R is fixed, we have

$$\rho_{\mathrm{f}} = \mathcal{O}\left(\frac{2r^2}{R^2} \cdot \rho_{\mathrm{d}}\right), \qquad c \to 1. \tag{10}$$

Theorem 1 shows that in order to achieve full coverage, ρ_{f} is smaller than ρ_{d} if $R > \sqrt{2}r$. According to (2), sensing range r is a constant independent of network density. On the one hand, fusion range R is a design parameter of the fusion model, which is mainly constrained by the communication overhead. In practice, the condition $R > \sqrt{2}r$ can be easily satisfied. For instance, the acoustic sensor on MICA2 motes has a sensing range of 3–5 m if a high performance (e.g., $\alpha = 5\%$ and $\beta = 95\%$) is required [40]. On the other hand, the fusion range can be set to be much larger. For example, Fig. 4 shows that R_{opt} ranges from 5 to 100 m when network density increases from 1.5×10^{-3} to 0.1. Therefore, according to Theorem 1, the fusion model with the optimal fusion range can significantly reduce network density for achieving a high level of coverage.

B. Full Coverage of Random Networks Using Optimal Fusion Range

As discussed in Section V-B, we can obtain the optimal fusion range via numerical experiment or analysis. Data fusion with the optimal fusion range allows the maximum number of informative sensors to contribute to the detection. The scaling law obtained with optimal fusion range will help us understand the maximum performance gain by adopting the data fusion model. The following theorem shows that ρ_{f} further reduces to $\mathcal{O}(\rho_{\mathrm{d}}^{1-1/k})$ as long as the fusion range is optimal.

Theorem 2: For uniformly deployed networks, let ρ_{d} and ρ_{f} denote the minimum network densities required to achieve the (α, β)-coverage of c under the disc and fusion models, respectively. If the optimal fusion range R_{opt} is adopted, we have

$$\rho_{\mathrm{f}} = \mathcal{O}\left(\rho_{\mathrm{d}}^{1-1/k}\right), \qquad c \to 1. \tag{11}$$

Theorem 2 shows that if the optimal fusion range is adopted, the fusion model can significantly reduce the network density for achieving high coverage. In particular, from Theorem 2, the density ratio $\rho_{\mathrm{f}}/\rho_{\mathrm{d}} = \mathcal{O}(\rho_{\mathrm{d}}^{-1/k}) = 0$ when $c \to 1$, which means ρ_{f} is insignificant compared to ρ_{d} for achieving high coverage. Theorem 2 is applicable to the scenarios where the physical signal follows the power-law decay with path loss exponent k, which are widely assumed and verified in practice. We note that the path loss exponent k typically ranges from 2.0 to 5.0 [15], [37]. In particular, the propagation of acoustic signals in free space follows the inverse-square law, i.e., $k = 2$, and therefore $\rho_{\mathrm{f}} = \mathcal{O}(\sqrt{\rho_{\mathrm{d}}})$.

C. Impact of Signal-to-Noise Ratio on Random Networks

In this section, we study the impact of PSNR on the results derived in the previous sections. PSNR is an important system parameter that is determined by the property of target, noise level, and sensitivity of sensors. We have the following corollary.

Corollary 1: For uniformly deployed networks, if the fusion range R is fixed, we have $\rho_{\mathrm{f}}/\rho_{\mathrm{d}} = \mathcal{O}(\delta^{2/k})$ when $c \to 1$.

Corollary 1 suggests that for a fixed R, the relative cost between the fusion and disc models is affected by the PSNR δ. Specifically, the fusion model requires fewer sensors to achieve full coverage than the disc model if the PSNR is low. On the other hand, the disc model suffices only if the PSNR is *sufficiently* high. Intuitively, sensor collaboration is more advantageous when the PSNR is low to moderate. However, when the PSNR is *sufficiently* high, the detection performance of a single sensor is satisfactory, and the collaboration among multiple sensors may be unnecessary.

VII. IMPACT OF DATA FUSION ON COVERAGE OF REGULAR AND MOBILE NETWORKS

It has been shown that random network deployments can lead to undesirable overprovision of sensing coverage [18], i.e., many fully covered areas have redundant sensors. In Section VII-A, we will study the coverage of regular networks, in which sensors are deployed at grid points. Our analysis shows that the data fusion can still reduce the network density for achieving full coverage of regular networks. Recent works [18], [34], [35] show that mobility can be introduced to trade with network density in achieving coverage. In such a scheme, randomly distributed mobile sensors relocate themselves to fill coverage holes in the initial network deployment. In Section VII-B, we will extend a relocation strategy proposed in [18] to the data fusion model. Our analysis shows that data fusion results in lower network density without increasing the moving distance of mobile sensors.

A. Full Coverage of Regular Networks

We have obtained the minimum network density for achieving full coverage of regular networks under the probabilistic disc model [given by (4)]. In this section, we first derive the minimum network density required by the data fusion model, and then study the ratio of network densities under the disc and fusion models, respectively. The following lemma gives the upper bound of network density for achieving full coverage of regular networks under the fusion model.

Lemma 4: Let ρ_{f} denote the minimum network density for achieving full coverage of regular networks under the fusion model. The upper bound of ρ_{f} is given by

$$\rho_{\mathrm{f}} \le \frac{\left(Q^{-1}(\alpha) - Q^{-1}(\beta)\right)^2}{2 \cdot \delta^2 \cdot R^2 \cdot w^2(R+\epsilon)}. \tag{12}$$

With the result in Lemma 4, we can study the ratio of network densities under the disc and fusion models, respectively, which is given by the following theorem.

Theorem 3: Let ρ_{d} and ρ_{f} denote the minimum network densities for achieving full coverage of regular networks under the disc and fusion models, respectively. If the localization error ϵ is insignificant, there exists a fusion range R such that $\rho_{\mathrm{f}} \le \rho_{\mathrm{d}}$.

Theorem 3 shows that by choosing a proper fusion range R, the data fusion model requires fewer sensors to fully cover the surveillance region. Moreover, we can draw the following observations from the proofs of Lemma 4 and Theorem 3. First, Lemma 4 is based on the sufficient condition that point p is (α, β)-covered, i.e., $N(p) \ge \Gamma(R)$, where $\Gamma(R)$ is given by

(8). However, from the proof of Lemma 3, this sufficient condition is conservative, as it assumes that all sensors are R meters from the point p. Therefore, the bounds in both Lemma 4 and Theorem 3 are conservative, and the fusion model can save more sensors in practice than the prediction of above analytical results. In Section IX, simulations based on both synthetic data set and real data traces show that the fusion model only requires half of the sensors required by the disc model. Second, Theorem 3 is proved using a proper setting of R that increases with the SNR. However, as discussed in Section VI-A, the fusion range is constrained by the communication overhead. For higher SNR, fewer sensors are required to cover the region. However, at least one sensor is required within the fusion range. Therefore, the fusion model will saturate due to bounded fusion range when the SNR becomes sufficiently high. The following corollary gives the density ratio when the fusion model saturates in case of sufficiently high SNRs.

Corollary 2: For regular networks, if the fusion range R is fixed, the ratio of densities for achieving full coverage satisfies $\rho_{\rm f}/\rho_{\rm d} = \mathcal{O}(\delta^{2/k})$.

From Corollary 1 and 2, the network density ratio shows the same trend in the case of sufficiently high SNRs. Hence, both for the random and regular networks, the disc model suffices only if the SNR is sufficiently high.

B. Full Coverage of Mobile Networks With Limited Mobility

Recent works [18], [34], [35] have exploited limited sensor mobility to reduce the network density for achieving full coverage under the disc model. In such a scheme, randomly distributed mobile sensors relocate themselves to fill coverage holes in the initial network deployment. In this section, we first extend an existing mobile relocation strategy [18] to the probabilistic disc and data fusion models, respectively, such that the relocated networks provide full (α, β)-coverage. Moreover, we show that data fusion can reduce the network density without increasing the moving distance of sensors.

We first briefly describe the mobile relocation strategy for achieving k-coverage proposed in [18]. Under the classical disc model, a region is k-covered if every point in the region is within the sensing range of at least k sensors. Suppose a mobile sensor network is randomly deployed in a vast region of area L. After the initial deployment, the region is divided by grids of side length $\sqrt{2}r/\sqrt{k}$, and the mobile sensors relocate themselves such that each grid point has exactly one sensor in it. After the mobile relocation, the region is fully k-covered. Moreover, the maximum distance that any mobile sensor has to move is $\mathcal{O}((1/\sqrt{k})\log^{3/4}(kL))$.

We now extend the above strategy to the probabilistic disc and data fusion models, respectively. First, under the probabilistic disc model, by replacing k with 1 and r with (2), the relocated mobile network provides full (α, β)-coverage, and the maximum moving distance is $\mathcal{O}(\log^{3/4} L)$. Note that the grid side length is $l = \sqrt{2}r$. Second, under the data fusion model, by replacing k with $\Gamma(R)$ and r with the fusion range R, at least $\Gamma(R)$ sensors are within the fusion range of any point in the region after relocation. As a result, the region is fully (α, β)-covered. Note that the grid side length is $l = \sqrt{2}R/\sqrt{\Gamma(R)}$. The illustration of the two extended relocation strategies is in Appendix-D of the supplementary file. As the relocated network is a grid network, the results in Section VII-A also apply. Moreover, the maximum moving distance under the data fusion model is $\mathcal{O}((1/\sqrt{\Gamma(R)})\log^{3/4}(\Gamma(R) \cdot L))$, which has the same order with respect to the network size L as under the probabilistic disc model, i.e., $\mathcal{O}(\log^{3/4} L)$. In Section IX, we will conduct simulations to evaluate the maximum moving distance that any sensor has to move as well as the total moving distance of all sensors.

VIII. Implications of Results and Discussions

In this section, we first summarize the implications of the theoretical results in this paper, and then discuss several issues that have not been addressed.

A. Implications of Results

1) Limitations of Disc Model: According to Theorem 2, when the coverage of random networks approaches one, $\rho_{\rm d}$ increases significantly faster than $\rho_{\rm f}$, especially for a small decay exponent. For instance, when $k = 2$ (which typically holds for acoustic signals), $\rho_{\rm f} = \mathcal{O}(\sqrt{\rho_{\rm d}})$. This result implies that the existing analytical results based on the disc model (e.g., [4]–[6], [8], [10]) significantly overestimate the network density required for achieving full coverage of random networks. Moreover, Theorem 3 shows that the disc model also leads to redundant deployment for regular networks. On the other hand, Corollaries 1 and 2 show that the disc model may lead to similar or even lower network density than the fusion model if PSNR is sufficiently high. The noise experienced by a sensor in real systems comes from various sources, e.g., the random disturbances in the environment and the electronic noise in sensor circuit. In practice, the PSNR in the applications based on low-cost sensors is usually low. For instance, the PSNRs in the vehicle detection experiments based on MICA2 [40] and ExScal [38] motes are about 50 (i.e., 17 dB). In such a case, in order to achieve a high level of coverage with random networks, $\rho_{\rm d} \geq 2\rho_{\rm f}$ if R is set to be greater than 8 m.

2) Design of Data Fusion Algorithms: Our results provide several important guidelines on the design of data fusion algorithms for large-scale WSNs. First, data fusion is very effective in improving sensing coverage and reducing network density. In particular, Corollaries 1 and 2 suggest that the performance gain of data fusion increases when the PSNR is lower. Therefore, data fusion should be employed for low-SNR deployments when a high level of coverage is required. Second, Theorems 1 and 2 suggest that fusion range plays an important role in the achievable performance of data fusion. As discussed in Section V-B, the optimal fusion range that maximizes the coverage of random networks increases with network density and can be numerically computed. Although we have not derived the optimal fusion range for regular networks, as discussed in Section VII-A, the fusion range should also increase with SNR. However, a larger fusion range may lead to longer transmission distances and more sensors that take part in data fusion. Investigating the optimal fusion range under both coverage and communication constraints is left for future work.

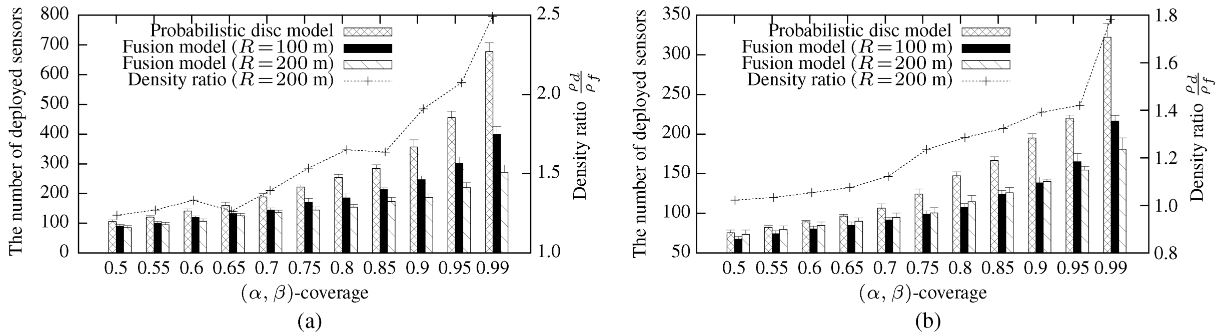


Fig. 5. Number of deployed sensors versus achieved (α, β)-coverage. (a) Random networks. (b) Regular networks.

B. Discussions

Under the fusion model, the fusion statistic Y is the aggregate function of the readings of the sensors taking part in fusion. Aggregating readings of multiple sensors in a wireless network may cause communication contention and packet delivery delay. In particular, we are interested in the detection latency caused by the aggregation process, which determines the minimum period with which the system can perform detection repeatedly. Our analysis shows that the average detection latency is $\Theta(\log(\rho R^2))$, and a higher level of coverage is always achieved at the price of longer detection latency. Besides the issue of detection latency, we have extended several results in this paper to other noise models, signal decay laws, and fusion models. However, due to space limitations, the details of the detection latency analysis and the extensions are omitted here and can be found in Appendix-E of the supplementary file.

IX. Simulations

In this section, we conduct extensive simulations based on real data traces as well as synthetic data to evaluate the coverage performance in nonasymptotic and asymptotic cases, respectively.

A. Trace-Driven Simulations

We first conduct simulations using the data traces collected in a real vehicle detection experiment [11]. In the experiments, 75 WINS NG 2.0 nodes are deployed to detect military vehicles driving through the surveillance region. We refer to [11] for detailed setup of the experiments. The data set used in our simulations includes the ground truth data and the acoustic time series recorded by 20 nodes when a vehicle drives through. The ground truth data include the positions of sensors and the trajectory of the vehicle.

As this paper is focused on the coverage performance of large-scale random and regular networks, the data traces collected by a limited number of sensors with a particular placement cannot be directly used to evaluate our results. In order to extensively evaluate our results, we reuse the sensors and rearrange their locations in our experiments to simulate various network densities and sensor distributions. The details of our simulation methodology are as follows. Sensors' sensing ranges under the probabilistic disc model are determined individually to meet the detection performance requirements ($\alpha = 5\%, \beta = 95\%$). Specifically, according to (2), $s_i \geq \sigma(Q^{-1}(\alpha) - Q^{-1}(\beta))$ within the sensing range of sensor i. In the simulation, the sensing range is estimated as the maximum distance from the target satisfying $y_i \leq \sigma(Q^{-1}(\alpha) - Q^{-1}(\beta))$. The resulted sensing ranges are from 22.5 to 59.2 m with the average of 43.2 m. Such a significant variation is due to several issues including poor calibration and complex terrain. In our simulation, we deploy random or regular networks with size of 1000×1000 m^2. Each sensor in the simulation is associated with a real sensor chosen at random. For each deployment, we evaluate the (α, β)-coverage under both the disc and fusion models. We divide the region into 1000×1000 grids. Under the disc model, the coverage is estimated as the ratio of grid points that are covered by discs. Under the fusion model, the coverage is estimated as the ratio of (α, β)-covered grid points. Specifically, for a target that appears at a grid point, each sensor makes a measurement that is set to be the sum of a random noise and the energy gathered by the associated real sensor at a similar distance to vehicle in the data trace. From Lemma 1, in the simulation, a grid point p is regarded to be (α, β)-covered if $(\sum_{i \in \mathbf{F}(p)} y_i / \sqrt{N(p)}) \geq \sigma(Q^{-1}(\alpha) - Q^{-1}(\beta))$. Therefore, our simulation methodology only needs the noise variance σ^2. The noise variance is estimated as 4×10^{-6} using the sensor measurements when no vehicle is present.

Fig. 5(a) plots the the numbers of uniformly deployed sensors under the disc and fusion models as well as the corresponding density ratio versus the achieved (α, β)-coverage. We can see that the disc model suffices if a moderate level of coverage is required. However, the fusion model is more effective for achieving high coverage. In particular, the fusion model with a fusion range of 200 m saves more than 50% sensors when the coverage is greater than 0.75. We note that the average number of sensors taking part in data fusion is within 30 and hence will not introduce high communication overhead. According to Theorem 1, the limit of $\rho_{\mathrm{d}}/\rho_{\mathrm{f}}$ is $R^2/2r^2$ when the coverage approaches one. We will evaluate the coverage performance in asymptotic case through simulations based on synthetic data in Section IX-B. Fig. 5(b) plots the results of regular networks. We can see that Fig. 5(b) shows a similar trend as Fig. 5(a). Moreover, the density ratio for achieving a high level of coverage (i.e., 0.99) is consistent with Theorem 3. The network density under the fusion model is 1×10^{-4} to 4×10^{-4} in Fig. 5(a), and 0.7×10^{-4} to 2.5×10^{-4} in Fig. 5(b). These densities are in the same order with that of the real deployment in [11],

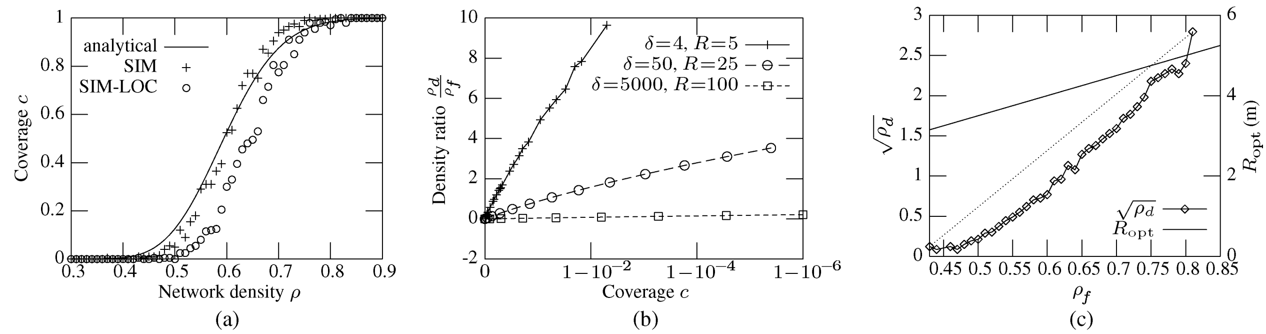


Fig. 6. Simulation results of random networks. (a) Coverage versus network density ($\delta = 4, R = 5$ m). (b) Density ratio $\rho_{\rm d}/\rho_{\rm f}$ versus coverage in $\log_{10}$ scale with various PSNRs. (c) $\sqrt{\rho_{\rm d}}$ versus $\rho_{\rm f}$ with optimal fusion range $R_{\rm opt}$ ($\delta = 4$).

i.e., 2.8×10^{-4}. Note that the network density can be up to 1×10^{-2} [38] in practice.

B. Simulations Based on Synthetic Data

1) Numerical Settings: In addition to trace-driven simulations, we also conduct extensive simulations based on synthetic data. These simulations allow us to evaluate the theoretical results in a wide range of settings. We adopt the signal decay function in (1) with $k = 2$. Both the mean and variance of the Gaussian noise generator, μ and σ^2, are set to be 1. We set the orginal energy of target, S, to be 4, 50, and 5000 so that the SNRs in the simulations are consistent with several real experiments, i.e., 0–10 dB in [43], around 17 dB in [40] and [38], and over 30 dB in [11].

As proved in Lemma 1, it suffices to measure the probability that a point is covered for evaluating the coverage of a random network. Hence, we let the target appear at a fixed point p and deploy random networks with size of $4R \times 4R$ centered at p. For each deployment, $P_{\rm D}(p)$ is estimated as the fraction of succesful detections. The (α, β)-coverage is estimated as the fraction of deployments whose $P_{\rm D}(p)$ is greater than β. We also evaluate the impact of localization error by integrating a simple localization algorithm. Specifically, for each detection, if a sensor's reading exceeds $S \cdot w(R) + \mu$, it will take part in the target localization. The target is localized as the geometric center of the sensors participating in the localization.

For a regular network, it suffices to measure the fraction of covered area in a grid for evaluating the coverage of the whole network. In our simulations, we find the minimum network density with which 10×10 points in the grid are covered.

2) Simulation Results of Random Networks: We first present the simulation results if sensors are randomly deployed. The first set of simulations evaluate the accuracy of the approximate formula given in Lemma 2. Fig. 6(a) plots the analytical and measured coverage versus network density. The curves labeled with SIM-LOC and SIM represent the measured results with and without accounting for localization error, respectively. We can see that the simulation result matches well with the analytical result given by (6). A network density of 0.8 is enough to provide high coverage under the fusion model, where the SNR is very low ($\delta = 4$). When there is localization error, a maximum deviation of about 0.2 from the analytical result can be seen from Fig. 6(a). The coverage decreases in the presence of localization error as sensors received weaker signals when the target cannot be accurately localized. However, the impact of localization error diminishes when $c \rightarrow 1$.

The second set of simulations evaluates the impact of SNR on the asymptotic network densities. Fig. 6(b) plots the network density ratio $\rho_{\rm d}/\rho_{\rm f}$ versus the achieved coverage under various PSNRs, where $\rho_{\rm d}$ is computed by (3) and $\rho_{\rm f}$ is obtained in simulations, respectively. The x-axis is plotted in $\log_{10}$ scale. We can see that the density ratio increases with the coverage, i.e., the fusion model becomes more effective for achieving higher coverage. Moreover, the density ratio decreases with the PSNR, which conforms to the result of Corollary 1. For instance, to achieve a coverage of 0.99, the density ratio $\rho_{\rm d}/\rho_{\rm f}$ is about 8 when $\delta = 4$. The density ratio decreases to about 2 when $\delta = 50$. When $\delta = 5000$, the disc model suffices. These results are consistent with the analysis in Section VI-C.

The third set of simulations evaluates the asymptotic relationship between $\rho_{\rm d}$ and $\rho_{\rm f}$ when the fusion range is optimized. In Fig. 6(c), the x- and y-axis of each data point represent the required network densities for achieving the same coverage that approaches to one under the disc and fusion models, respectively. Note that the y-axis is plotted in square-root scale. The optimal fusion range $R_{\rm opt}$ plotted in Fig. 6(c) is computed for each given $\rho_{\rm f}$ by numerically maximizing (6). We can see from Fig. 6(c) that the relationship between $\sqrt{\rho_{\rm d}}$ and $\rho_{\rm f}$ is convex and therefore conforms to the theoretical result $\rho_{\rm f} = \mathcal{O}(\sqrt{\rho_{\rm d}})$ according to Theorem 2. Moreover, $R_{\rm opt}$ increases with $\rho_{\rm f}$, which is also consistent with the analysis in Section VI-B.

3) Simulation Results of Regular and Mobile Networks: We now present the simulation results when sensors are deployed at grid points. We first measure the minimum network densities for achieving full coverage under disc and fusion models, respectively. For the fusion model, we numerically find the optimal fusion range $R_{\rm opt}$ that minimizes the network density for achieving full coverage. Fig. 7(a) plots the network density ratio and the corresponding $R_{\rm opt}$ versus the PSNR. We can see that the fusion model can reduce 100% of sensors compared to the disc model in a wide range of PSNRs, i.e., from 15 to 37 dB. The fusion model is more effective in the case of low PSNRs, i.e., from 4 to 10 dB. Note that the $R_{\rm opt}$ increases with the PSNR. The above results are consistent with the analysis in Section VII-A.

We then evaluate the moving distances of mobile sensors in relocation with various settings of SNR. We employ a greedy relocation algorithm [18], [35]. Specifically, in each round, the

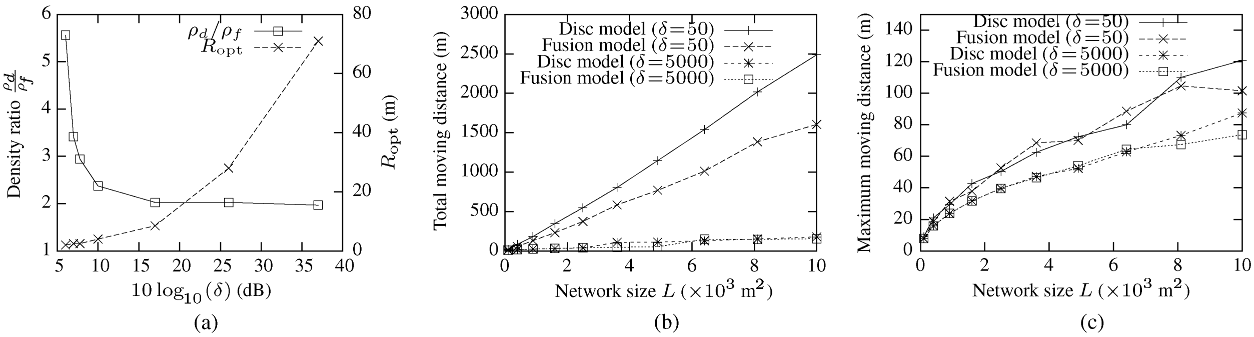


Fig. 7. Simulation results of regular and mobile networks. (a) Density ratio of regular networks versus PSNR with optimal fusion range. (b) Total moving distance of mobile sensors versus network size. (c) Maximum moving distance of mobile sensors versus network size.

sensor that is closest to any grid point is relocated to the grid point. Fig. 7(b) plots the total moving distance of all sensors versus the network area. We can see that the fusion model outperforms the disc model when SNR is low, and the two models are comparable when SNR is high. Fig. 7(c) plots the maximum moving distance of sensors. We can see that the maximum moving distances under the two models are in the same order with respect to network size. This result is consistent with the analysis in Section VII-B.

X. Conclusion

Sensing coverage is an important performance requirement of many critical sensor network applications. In this paper, we explore the fundamental limits of coverage based on stochastic data fusion models that jointly process noisy measurements of sensors. The scaling laws between coverage, network density, and SNR are derived. Data fusion is shown to significantly improve sensing coverage by exploiting the collaboration among sensors. Our results help understand the limitations of the existing analytical results based on the disc model and provide key insights into the design and analysis of WSNs that adopt data fusion algorithms. Our analyses are verified through simulations based on both synthetic data sets and data traces collected in a real deployment for vehicle detection.

References

[1] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-efficient surveillance system using wireless sensor networks," in *Proc. MobiSys*, 2004, pp. 270–283.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. ACM WSNA*, 2002, pp. 88–97.

[3] D. Li, K. Wong, Y. H. Hu, and A. Sayeed, "Detection, classification and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 17–29, Mar. 2002.

[4] P. Brass, "Bounds on coverage and target detection capabilities for models of networks of mobile sensors," *Trans. Sensor Netw.*, vol. 3, no. 2, 2007, Article no. 9.

[5] S. Kumar, T. H. Lai, and J. Balogh, "On k-coverage in a mostly sleeping sensor network," in *Proc. ACM MobiCom*, 2004, pp. 144–158.

[6] P.-J. Wan and C.-W. Yi, "Coverage by randomly deployed wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2658–2669, Jun. 2006.

[7] S. Shakkottai, R. Srikant, and N. Shroff, "Unreliable sensor grids: Coverage, connectivity and diameter," in *Proc. IEEE INFOCOM*, 2003, pp. 1073–1083.

[8] H. Zhang and J. Hou, "On deriving the upper bound of α-lifetime for large sensor networks," in *Proc. ACM MobiHoc*, 2004, pp. 121–132.

[9] G. Xing, X. Wang, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration for energy conservation in sensor networks," *Trans. Sensor Netw.*, vol. 1, no. 1, pp. 36–72, 2005.

[10] B. Liu and D. Towsley, "A study on the coverage of large-scale sensor networks," in *Proc. IEEE MASS*, 2004, pp. 475–483.

[11] M. Duarte and Y. H. Hu, "Vehicle classification in distributed sensor networks," *J. Parallel Distrib. Comput.*, vol. 64, no. 7, pp. 826–838, 2004.

[12] M. Duarte and Y. H. Hu, "Distance based decision fusion in a distributed wireless sensor network," in *Proc. IPSN*, 2003, pp. 556–556.

[13] T. Clouqueur, K. K. Saluja, and P. Ramanathan, "Fault tolerance in collaborative sensor networks for target detection," *IEEE Trans. Comput.*, vol. 53, no. 3, pp. 320–333, Mar. 2004.

[14] X. Sheng and Y. H. Huc, "Energy based acoustic source localization," in *Proc. IPSN*, 2003, pp. 551–551.

[15] D. Li and Y. H. Hu, "Energy based collaborative source localization using acoustic micro-sensor array," *EUROSIP J. Appl. Signal Process.*, vol. 2003, no. 4, pp. 321–337, 2003.

[16] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 4, pp. 462–472, 2005.

[17] N. Bisnik, A. Abouzeid, and V. Isler, "Stochastic event capture using mobile sensors subject to a quality metric," in *Proc. ACM MobiCom*, 2006, pp. 676–692.

[18] W. Wang, V. Srinivasan, and K. C. Chua, "Trade-offs between mobility and density for coverage in wireless sensor networks," in *Proc. ACM MobiCom*, 2007, pp. 39–50.

[19] C. Gui and P. Mohapatrac, "Power conservation and quality of surveillance in target tracking sensor networks," in *Proc. ACM MobiCom*, 2004, pp. 129–143.

[20] S. Kumar, T. H. Lai, and A. Arora, "Barrier coverage with wireless sensors," in *Proc. ACM MobiCom*, 2005, pp. 284–298.

[21] B. Liu, O. Dousse, J. Wang, and A. Saipulla, "Strong barrier coverage of wireless sensor networks," in *Proc. ACM MobiHoc*, 2008, pp. 411–420.

[22] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.

[23] P. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1996.

[24] J. N. Tsitsiklis, "Decentralized detection," *Adv. Stat. Signal Process.*, vol. 2, pp. 297–344, 1993.

[25] W. Tay, J. Tsitsiklis, and M. Win, "On the impact of node failures and unreliable communications in dense sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2535–2546, Jun. 2008.

[26] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Coverage problems in wireless ad-hoc sensor networks," in *Proc. IEEE INFOCOM*, 2001, pp. 1380–1387.

[27] X. Li, P. Wan, and O. Frieder, "Coverage in wireless ad hoc sensor networks," *IEEE Trans. Comput.*, vol. 52, no. 6, pp. 753–763, Jun. 2003.

[28] S. Meguerdichian, F. Koushanfar, G. Qu, and M. Potkonjak, "Exposure in wireless ad-hoc sensor networks," in *Proc. ACM MobiCom*, 2001, pp. 139–150.

[29] T. Yan, T. He, and J. A. Stankovic, "Differentiated surveillance for sensor networks," in *Proc. ACM SenSys*, 2003, pp. 51–62.

[30] M. Hefeeda and H. Ahmadi, "A probabilistic coverage protocol for wireless sensor networks," in *Proc. IEEE ICNP*, 2007, pp. 41–50.

[31] S. Ren, Q. Li, H. Wang, X. Chen, and X. Zhang, "Design and analysis of sensing scheduling algorithms under partial coverage for object detection in sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 3, pp. 334–350, Mar. 2007.

[32] N. Ahmed, S. S. Kanhere, and S. Jha, "Probabilistic coverage in wireless sensor networks," in *Proc. IEEE LCN*, 2005, pp. 672–681.

[33] W. Wang, V. Srinivasan, K.-C. Chua, and B. Wang, "Energy-efficient coverage for target detection in wireless sensor networks," in *Proc. IPSN*, 2007, pp. 313–322.

[34] G. Wang, G. Cao, and T. L. Porta, "Movement-assisted sensor deployment," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 640–652, Jun. 2006.

[35] S. Chellappan, W. Gu, X. Bai, D. Xuan, B. Ma, and K. Zhang, "Deploying wireless sensor networks under limited mobility constraints," *IEEE Trans. Mobile Comput.*, vol. 6, no. 10, pp. 1142–1157, Oct. 2007.

[36] B. Liu, P. Brass, O. Dousse, P. Nain, and D. Towsley, "Mobility improves coverage of sensor networks," in *Proc. ACM MobiHoc*, 2005, pp. 300–308.

[37] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. VT-29, no. 3, pp. 317–325, Aug. 1980.

[38] L. Gu, D. Jia, P. Vicaire, T. Yan, L. Luo, A. Tirumala, Q. Cao, T. He, J. Stankovic, T. Abdelzaher, and H. Bruce, "Lightweight detection and classification for wireless sensor networks in realistic environments," in *Proc. ACM SenSys*, 2005, pp. 205–217.

[39] W.-P. Chen, J. C. Hou, and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 3, pp. 258–271, Mar. 2004.

[40] B. P. Flanagan and K. W. Parker, "Robust distributed detection using low power acoustic sensors," MITRE Corporation, McLean, VA, 2005, Tech. Rep..

[41] "NIST/Sematech e-Handbook of statistical methods," NIST, Gaithersburg, MD, 2011 [Online]. Available: http://itl.nist.gov/div898/handbook/

[42] C. Taylor, A. Rahimi, J. Bachrach, H. Shrobe, and A. Grue, "Simultaneous localization, calibration, and tracking in an ad hoc sensor network," in *Proc. IPSN*, 2006, pp. 27–33.

[43] S. Coleri, S. Cheung, and P. Varaiya, "Sensor networks for monitoring traffic," in *Proc. Allerton Conf. Commun., Control, Comput.*, 2004, pp. 32–40.

Rui Tan (M'10) received the B.S. and M.S. degrees in automation from Shanghai Jiao Tong University, Shanghai, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from City University of Hong Kong, Hong Kong, in 2010.

He is currently a Visiting Professor with Michigan State University, East Lansing. His research interests include collaborative sensing in wireless sensor networks and cyber-physical systems.

Guoliang Xing (M'06) received the B.S. degree in electrical engineering and M.S. degree in computer science from Xi'an Jiao Tong University, Xi'an, China, in 1998 and 2001, respectively, and the M.S. and D.Sc. degrees in computer science and engineering from Washington University in St. Louis, St. Louis, MO, in 2003 and 2006, respectively.

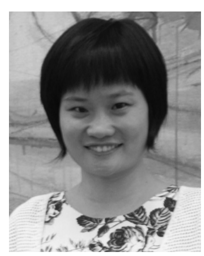
He is an Assistant Professor with the Department of Computer Science and Engineering, Michigan State University, East Lansing. From 2006 to 2008, he was an Assistant Professor of computer science with City University of Hong Kong, Hong Kong. His research interests include wireless sensor networks, mobile systems, and cyber-physical systems.

Dr. Xing was an NSF CAREER Award recipient in 2010. He received the Best Paper Award at the 18th IEEE International Conference on Network Protocols (ICNP) in 2010.

Benyuan Liu (M'05) received the Ph.D. degree in computer science from the University of Massachusetts, Amherst, in 2003.

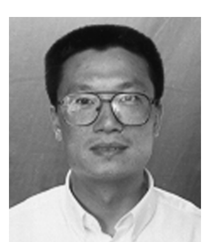
He is currently an Associate Professor of computer science with the University of Massachusetts Lowell, Lowell. His research interests are in the area of protocol design and performance modeling of wireless ad hoc and sensor networks.

Dr. Liu served as Technical Program Co-Chair of WASA 2009 and Cross-layer Design and Optimization Track of IEEE ICCCN 2008 and served on the technical program committees for numerous computer networking conferences.



Jianping Wang (M'03) received the B.S. and M.S. degrees in computer science from Nankai University, Tianjin, China, in 1996 and 1999, respectively, and the Ph.D. degree from the University of Texas at Dallas in 2003.

She is currently an Assistant Professor of computer science with City University of Hong Kong, Hong Kong. Prior to joining City University of Hong Kong, she was with the University of Mississippi, Oxford, and Georgia Southern University, Statesboro. Her research interests are optical networks, multicast, dependable networking, wireless networks, and the integration of optical networks and wireless networks.



Xiaohua Jia (M'00–A'01–SM'01) received the B.Sc. and M.Eng. degrees in computer science from the University of Science and Technology of China, Hefei, China, in 1984 and 1987, respectively, and the D.Sc. degree in information science from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently the Chair Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong. His research interests include distributed systems, computer networks, wireless sensor networks, and mobile wireless networks.

Prof. Jia is an Editor of the IEEE TRANSACTION ON PARALLEL AND DISTRIBUTED SYSTEMS, *Wireless Networks*, *World Wide Web*, *Journal of Combinatorial Optimization*, etc. He was General Chair of ACM MobiHoc 2008, Technical Program Committee Co-Chair of IEEE MASS 2009, and the Area-Chair of IEEE INFOCOM 2010.