On Determining Cluster Size of Randomly Deployed Heterogeneous WSNs

Cüneyt Sevgi and Altan Koçyiğit

***Abstract*— Clustering is an efficient method to solve scalability problems and energy consumption challenges. For this reason it is widely exploited in Wireless Sensor Network (WSN) applications. It is very critical to determine the number of required clusterheads and thus the overall cost of WSNs while satisfying the desired level of coverage. Our objective is to study cluster size, i.e., how much a clusterhead together with sensors can cover a region when all the devices in a WSN are deployed randomly. Therefore, it is possible to compute the required number of nodes of each type for given network parameters.**

***Index Terms*— Cluster size, random deployment, wireless sensor networks (WSNs).**

I. Introduction

SCALABILITY and energy consumption are among the most important challenges for WSN applications. Several hierarchical architecture and protocols are proposed to tackle these challenges. Devices in WSNs form clusters where the clusterheads aggregate and fuse data to conserve energy.

In this paper, we consider a randomly deployed heterogeneous WSN. A mixture of two different types of devices, clusterheads and ordinary sensors (or simply sensors), is assumed to be scattered over a region of interest. In this scenario, randomly deployed clusterheads form clusters with the nearby sensors. For such a network, we determine the cluster size, which is the area covered by a clusterhead together with the sensors connected to it.

The remainder of the paper is organized as follows. Section II presents the coverage and connectivity equations for randomly deployed sensors based on the Boolean coverage disk model. In Section III, we derive the expected value of the cluster size for randomly deployed heterogeneous WSNs. Section IV concludes the paper.

II. Coverage and Connectivity

Coverage is one of the fundamental issues in WSNs. A point is said to be covered if it is within the sensing range of at least one sensor. This coverage would be meaningful only when the sensor is able to transmit its data to the sinks. Therefore, coverage and connectivity should be analyzed jointly.

Manuscript received November 19, 2007. The associate editor coordinating the review of this letter and approving it for publication was S. Buzzi.

C. Sevgi is with the Department of Computer Technology and Information Systems, Bilkent University, Ankara, Turkey (e-mail: csevgi@bilkent.edu.tr).

A. Kocyigit is with the Department of Information Systems, Middle East Technical University, Ankara, Turkey (e-mail: kocyigit@metu.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2008.071942.

Recently, some research has been carried out to describe the relationship between coverage and connectivity. In [2] and [3], it is independently proven that "transmission range at least twice the sensing range" is the sufficient condition for connectivity as long as full coverage is guaranteed for a convex region. Both of the studies focus on analyzing the condition for a fully covered network to guarantee connectivity.

While in some WSN applications the goal is to gather data about an entire region of interest (full coverage), for many other, partial coverage is realistic and feasible since full coverage in randomly deployed WSNs reveals asymptotic behavior. This is because when the number of sensors scattered or the sensing range are increased beyond a threshold value, the coverage increases only marginally. Therefore, we primarily consider the partial coverage of a randomly deployed mixture of clusterheads and sensors.

In this letter, we adopt the following network model.

- The WSN consists of two different types of devices: sensors and clusterheads.
- We assume that N_H clusterheads and N_S sensors are deployed randomly over a planar region D.
- Both sensors and clusterheads have sensing capabilities and their sensing range is r_s.
- Sensors can only transmit their sensing data to a clusterhead and clusterheads transmit data to sinks. Communication among sensors is not allowed.
- A sensor can communicate with a clusterhead if it is within the communication range r_t of the clusterhead. The clusterheads are assumed to be connected to the sink.

In the following sections, coverage and connectivity equations are presented separately. In Section III, cluster size and connected coverage equations are derived by using these equations.

A. Coverage

Suppose a large planar area D is to be covered by N identical sensors which are scattered randomly over the area according to a Poisson point process. Suppose the area sensed by each sensor is a perfect disk with radius r_s and λ is the average number of sensors per unit area. The probability of a point in D being covered P_{cov} can be found as [1]:

$$P_{cov} = 1 - e^{-\lambda \pi r_s^2} \quad (1)$$

In this letter, all subsequent discussions are based on the similar approach used for the derivation of Eqn. 1 by Koskinen in [1].

Note that in the above equation, coverage probability is independent of the geometry of the (convex) region sensed by a sensor and if the sensing region covered by any sensor was A_S, the coverage probability would be:

$$P_{cov} = 1 - e^{-\lambda\ A_S} \tag{2}$$

Both sensors and clusterheads have sensing capability and their sensing region is a perfect disk. We have $N = N_H + N_S$ sensing devices scattered randomly over the area. Therefore, without considering connectivity, the coverage probability could be found as:

$$P_{cov} = 1 - e^{-\frac{(N_H+N_S)\pi r_s^2}{D}} \tag{3}$$

B. Connectivity

Like coverage, connectivity is another requirement to be satisfied in WSNs. If a sensor can reach the clusterhead directly, then this sensor is said to be connected. However, by using an approach similar to the approach employed in deriving coverage probability, the probability that a sensor is within the communication range of a clusterhead P_{con} could be derived as:

$$P_{con} = 1 - e^{-\frac{N_H \pi r_t^2}{D}} \tag{4}$$

III. Cluster Size in a Heterogeneous WSN

Up to this point, the concepts of connectivity and coverage have been described separately. However, we should only take into account the part of the sensing area covered by connected sensors. In order to find the actual coverage probability we first determine $S_{cluster}$, the expected value of area covered by each clusterhead together with the sensors connected to it. Then it is possible to find the actual coverage (i.e., connected coverage probability) by using Eqn. 2:

$$P_{cov} = 1 - e^{-\frac{N_H S_{cluster}}{D}} \tag{5}$$

Let the average number of sensors connected to a single clusterhead be n_s. Since there are N_S sensors and N_H clusterheads scattered over region D, n_s can be found using Eqn. 4 as follows:

$$n_s = \frac{N_S}{N_H}\left(1 - e^{-\frac{N_H \pi r_t^2}{D}}\right) \tag{6}$$

Therefore, in order to find $S_{cluster}$, we need to find the area covered by the clusterhead and n_s sensors connected to it. For the sake of simplicity, consider that a single clusterhead and a set of sensors are scattered over a region D (Fig. 1).

Since there should be n_s sensors in the communication range of the clusterhead, the number of sensors in the square, C_s, should be:

$$\frac{n_s}{C_s} = \frac{\pi r_t^2}{D} \Longleftrightarrow D = \frac{C_s}{n_s/\pi r_t^2} \Longleftrightarrow C_s = \frac{D n_s}{\pi r_t^2} \tag{7}$$

Since we already know that any point within r_s from the center is covered by the clusterhead at the center, any point outside the inner circle can only be covered by the sensors connected to the clusterhead. To find the probability of a point p outside the inner disc to be covered, there should be one or more "connected sensors" sensing the point p. That is, there should be one or more sensors in region $I(x)$, which is the shaded region in Fig. 1.

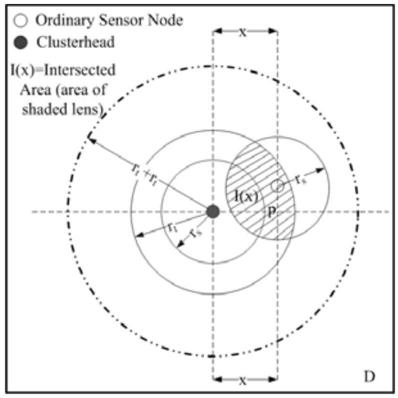

Fig. 1. Area covered by a clusterhead and a single sensor.

$I(x)$ is the region formed by the intersection of two discs, and its area is a function of the distance between the centers of these two discs (denoted by x) and their radii r_t and r_s. We can examine two different cases to find the area of the shaded region.

Case 1: When r_t-r_s<x≤r_t+r_s the area $I(x)$ can be found as in [4]:

$$I(x) = r_s^2 \cos^{-1}\left(\frac{x^2 + r_s^2 - r_t^2}{2xr_s}\right) + r_t^2 \cos^{-1}\left(\frac{x^2 + r_t^2 - r_s^2}{2xr_t}\right) - \frac{1}{2} \sqrt{(r_s + r_t - x)(x + r_s - r_t)(x - r_s + r_t)(x + r_s + r_t)} \tag{8}$$

Case 2: When $x \le r_t$-r_s,

$$I(x) = \pi r_s^2 \tag{9}$$

Let $P_{p1}(x)$ be the probability that point p is not sensed by a single sensor connected to the clusterhead at the center. In order for a point p not to be sensed, the sensor node should not be in the intersected area. Therefore,

$$P_{p1}(x) = \left(1 - \frac{I(x)}{D}\right) \tag{10}$$

We have C_s sensors in region D. Therefore, the probability of having no sensor in the shaded area is:

$$P_{p-nc}(x) = \left(1 - \frac{I(x)}{D}\right)^{C_s} \tag{11}$$

As C_s goes to infinity we have:

$$P_{p-nc}(x) = \lim_{C_s \to \infty}\left(1 - \frac{\frac{I(x) n_s}{\pi r_t^2}}{C_s}\right)^{C_s} = e^{-\frac{I(x) n_s}{\pi r_t^2}} \tag{12}$$

Therefore, the probability that a point p which is x units away from the clusterhead at the center is sensed by "at least one sensor" connected to the clusterhead at the center can be found as:

$$P_{pc}(x) = 1 - e^{-\frac{I(x) n_s}{\pi r_t^2}} \tag{13}$$

By using the individual point connectivity probabilities derived above, we can find "the expected value of the area

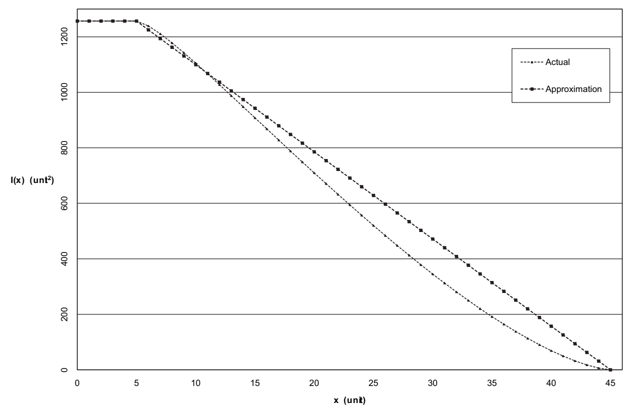

Fig. 2. $I(x)\ vs.\ x$ (x is the distance between the centers of the two discs) where $r_t = 25\ unit$ and $r_s = 20\ unit$.

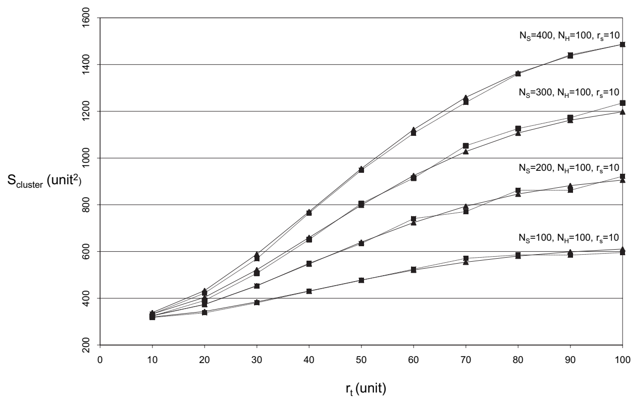

Fig. 3. $S_{cluster}\ vs.\ r_t$ where $D = 1000 \times 1000\ unit^2$ and $r_s = 10\ unit$.

covered" by the clusterhead and the sensors connected to it. For this purpose, we can integrate in cylindrical coordinates as:

$$S_{cluster} = \int_{x=r_s}^{r_s+r_t} \int_{\phi=0}^{2\pi} x(1-e^{-\frac{I(x)n_s}{\pi r_t^2}}) d\phi dx$$
$$= 2\pi \int_{x=r_s}^{r_s+r_t} x(1-e^{-\frac{I(x)n_s}{\pi r_t^2}}) dx \qquad (14)$$

By rearranging the terms, we have:

$$S_{cluster} = \pi r_s^2 + 2\pi \int_{x=r_s}^{r_s+r_t} x(1-e^{-\frac{I(x)n_s}{\pi r_t^2}}) dx \qquad (15)$$

A. Linear Approximation for Cluster Size

To find $S_{cluster}$, the complex integral in Eqn. 15 should be performed. So as to simplify the integration, $I(x)$ can be approximated by a line segment whose equation is (see Fig. 2):

$$I(x) = -(\pi r_s/2)x + (\pi r_s/2)(r_s + r_t) \qquad (16)$$

Case 1: If $r_t \leq 2r_s$ then area covered by a clusterhead and sensors connected to it can be found as:

$$S_{cluster} = \pi r_s^2 + 2\pi \int_{x=r_s}^{r_s+r_t} x(1-e^{-\frac{\left[-\left(\frac{\pi r_s}{2}\right)x+\left(\frac{\pi r_s}{2}\right)(r_s+r_t)\right]n_s}{\pi r_t^2}}) dx \qquad (17)$$

Then $S_{cluster}$ can be found as:

$$S_{cluster} = \pi(r_s+r_t)^2 + \frac{2\pi}{\alpha}\left[(\frac{1}{\alpha} - r_s)(1-e^{-\alpha r_t}) - r_t\right] \qquad (18)$$

where

$$\alpha = \frac{n_s r_s}{2r_t^2} \qquad (19)$$

Case 2: If $r_t > 2r_s$ then area covered by a clusterhead and sensors connected to it can be found as:

$$S_{cluster} = \pi r_s^2 + 2\pi \int_{r_s}^{r_t-r_s} x(1-e^{-\frac{\pi r_s^2 n_s}{\pi r_t^2}}) dx + 2\pi \int_{r_t-r_s}^{r_s+r_t} x(1-e^{-\frac{\left[-\left(\frac{\pi r_s}{2}\right)x+\left(\frac{\pi r_s}{2}\right)(r_s+r_t)\right]n_s}{\pi r_t^2}})) dx \qquad (20)$$

Then $S_{cluster}$ can be found as:

$$S_{cluster} = \pi(r_s+r_t)^2 - \pi r_t(r_t - 2r_s)e^{-2\alpha r_s} - \frac{2\pi}{\alpha^2}\left[(\alpha(r_t-r_s)-1)(1-e^{-2\alpha r_s}) + 2\alpha r_s\right] \qquad (21)$$

We performed extensive simulations to validate derived cluster size equations by using a simulator we developed in Java. The number of experiments for each cluster size value is determined according to a confidence interval of $\pm 5\%$ with the probability of 0.95. Fig. 3 shows that there is at most 2% discrepancy between simulation results and the analytical findings. This variation is due to the edge effect because Fig. 3 demonstrates that for smaller values of r_t, analytical and simulation values do not deviate significantly. However, as the r_t values get larger, the variation increases.

IV. Conclusion

In this letter, we derived general equations (Eqn. 18 and Eqn. 21) for cluster size in randomly deployed heterogeneous WSNs and we validated the equations through simulations. The result obtained in this letter enables one to compute the number of required devices and therefore the optimum proportion of the different types of devices. Issues related to finding the optimum number of different types of devices are currently under study.

V. Acknowledgment

The authors would like to thank to Syed Amjad Ali and Nil Korkut for valuable comments and proofreading.

References

[1] H. Koskinen, "On the coverage of a random sensor network in a bounded domain," in *Proc. ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, 2004.

[2] H. Zhang and J. C. Hou, "Maintaining sensing coverage and connectivity in large sensor nodes," *Wireless Ad Hoc and Sensor Networks: An International Journal*, vol. 1, pp. 89-124, Jan. 2005.

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

[4] E. W. Weisstein, "Circle-circle intersection" *MathWorld–A Wolfram Web Resource* http://mathworld.wolfram.com/Circle-CircleIntersection.html.