Connectivity of Wireless Sensor Networks with Constant Density

Sarah Carruthers and Valerie King

University of Victoria, Victoria, BC, Canada

Abstract. We consider a wireless sensor network in which each sensor is able to transmit within a disk of radius one. We show with elementary techniques that there exists a constant c such that if we throw cN sensors into an $n \times n$ square (of area N) independently at random, then with high probability there will be exactly one connected component which reaches all sides of the square. Its expected size is a constant fraction of the total number of sensors, and it covers a constant fraction of the area of the square. Furthermore, the other connected components of sensors and uncovered contiguous portions of the square are each very small ($O(\log^2 n)$ and $O(\log N)$ respectively), so that their relative size grows smaller as N increases. We also discuss some algorithmic implications.

1 Introduction

In wireless sensor networks, the connectivity of the network is established via radio transmission between sensors. For two sensors to be able to communicate, they must be within some critical range of each other, as transmission capability is finite. The connectivity of the entire network is composed of these sensor-to-sensor links. Depending on the disbursement of the sensors, it is possible for islands of sensors to be isolated from the rest of the network; there need only exist a gap that isolates them. We will model the wireless sensor network as a set of sensors placed independently at random in a $n \times n$ square.

We show that there exists a c such that if cN (hereafter $N = n^2$) sensors are placed independently at random in a $n \times n$ square, then with high probability there exists a large connected component of size $O(N)$ which covers at least one point on every side of the square. Furthermore, with high probability, we can bound the number of sensors of any other connected component by $O(\log^2 N)$ and any uncovered contiguous area by $O(\log N)$. Thus, if the density (number of sensors per area) is a constant, the size of each contiguous portion of the square which is uncovered or unconnected to the main connected component diminishes relative to the area of the square, as the square grows larger.

Our technique is to decompose the square into a grid of small squares (which we call *boxes*) and then use simple counting arguments to prove the necessary bounds. The boxes are small enough to ensure that any disconnected set of sensors must be surrounded by a path of empty boxes. We first prove a bound on the number of empty boxes. In Lemma **??**, we bound the probability that there exists a long path of empty boxes. This implies that exactly one component touches all four sides of the square.

We then reapply Lemma **??** to bound the number of sensors which are surrounded by sensors in the large component, but are themselves cut off from transmitting to these sensors by a large encircling gap. Imagine the large component as a giant piece of Swiss cheese or a donut. There are holes in its interior, and these holes may in fact contain connected components of sensors which can transmit to each other, but not to the large component. For the remainder of the paper, we will refer to such isolated connected components as *timbits*.

2 Related Work

Recently, there has been much work done in the area of connectivity of wireless networks. In 1998, Gupta and Kumar [**?**] showed that with limited transmission strength, a wireless network achieves asymptotic connectivity. Similarly, in 2003 Shakkottai, Srikant and Shroff [**?**] show this is also true for networks in which some sensors may fail. Xue and Kumar showed in 1998 that a sensor need be connected to $\Theta(\log N)$ nearest neighbours in order for the network to be asymptotically connected [**?**], i.e., $\Theta(N \log N)$ sensors are necessary for every sensor to be connected in a square of area N.

In 2003, Jennings and Okino [**?**] showed bounds on the probability that a network with n sensors in a unit square, each with a fixed radius of transmission, is connected. What they fail to prove, however, is the size of the largest connected component. They corroborate their results with simulations. In these simulations, up to 100 sensors were placed randomly in a square of area $A = 600^2$. The transmission radii of the sensors were chosen as a proportion of the length of the square, with a fixed radii picked for a particular graph. These simulations show that the size of the largest connected component is large ($.9A$), when the radius of transmission is $.3\sqrt{A}$. In contrast, our results show that in a square of area A, for any constant $d < 1$, a *constant* radius of transmission suffices for the largest connected component to cover an expected area of dA, in a network of constant density.

Very recently it has come to our attention that in 1995, Penrose considered a more general problem in a sophisticated analysis using percolation theory [**?**]. This work is cited in another paper written in 2000 by Diaz, Penrose, Petit and Seina [**?**]. See also [**?**].

3 Preliminaries

In our model, we say two sensors are *neighbours*, i.e., there is a transmission link between them, if they are within distance one of each other. Two sensors x and y are *connected* if there is a path of sensors $z_1, ..., z_k$ such that x is a neighbour of z_1, y is a neighbour of z_k, and for each $i < k$, z_i is a neighbour of z_{i+1}. A *connected component* is a maximal set of sensors which are connected. A point is *uncovered* if it is of distance greater than one from any sensor; otherwise the point is covered. A set of points is *uncovered* (*covered*) if every element of the set is uncovered (covered).

We toss cN sensors independently at random into an $n \times n$ square of area N. We partition the $n \times n$ square into $2\sqrt{2}n$ equally spaced rows and columns so that each small square (hereafter *box*) has size $1/(2\sqrt{2}) \times 1/(2\sqrt{2})$, for a total of $8N$ such boxes. This size of box is chosen so that if a box A contains a sensor, then this sensor has a link to any sensor in any box which is adjacent or diagonal to A. A box is empty if it contains no sensors; otherwise it is nonempty.

The first question we ask is: if we throw cN sensors into the $n \times n$ square, how many boxes will remain empty?

Lemma 1. *For any constant c_3, there exists a constant c such that if we throw cN sensors into an $n \times n$ square, the expected number of empty boxes is $N/5c_3$.*

Proof: We observe that the problem of determining the number of empty boxes is the same as the well-known balls in bins occupancy problem: Given r bins and s balls thrown in at random, how many bins are empty? Here, $r = 8N$ and $s = cN$. The expected number of empty bins is $r(1-1/r)^s \leq 8Ne^{-c/8}$ which is less than $N/5c_3$, for some $c \geq 24 + \ln c_3/8$.

Lemma 2. *Let c_1, d be any constants, where $c_1, d > 1$. There exists a constant c such that if cN sensors are thrown independently at random into the square, then the number of empty boxes is less than N/c_1 with probability less than $1/e^{dN}$.*

Proof: Fix $\lfloor N/c_1 \rfloor$ empty boxes. The probability that a sensor doesn't fall into one of these empty boxes is less than $(1 - 1/c_1)$. The probability that no sensors fall into these boxes is $(1 - 1/c_1)^{cN}$. Since there are $\binom{N}{\lfloor N/c_1 \rfloor}$ ways to choose the set of empty boxes, the probability of this occurring for any such set of boxes is bounded by:

$$\binom{N}{\lfloor N/c_1 \rfloor}(1-1/c_1)^{cN} \leq (c_1 e)^{N/c_1}(1/e)^{(c/c_1)N} \leq (1/e)^{dN} \quad , \tag{1}$$

when $c > c_1 d + \ln c_1 + 1$. We derive these inequalities from the following two facts: (1) $\binom{s}{t} \leq (se/t)^t$ and (2) $(1 - 1/t) \leq (1/e)^{1/t}$, where $1/t < 1$, which we will use repeatedly in this paper.

4 A Unique Largest Component

By a *path of empty boxes* we mean a "manhattan" path in the sense that it consists of up, down, left and right turns. We define a square of size r to be a square composed of r boxes, i.e. smaller squares of size $1/(2\sqrt{2}) \times 1/(2\sqrt{2})$, such that $\sqrt{r}$ is an integer value. We define a timbit as a connected component which covers portions of no more than two sides of the $n \times n$ square.

Observation 3. *If a set of sensors is unconnected to another set, there must be a path of empty boxes separating that set from the other. In particular, every timbit must be surrounded by a path composed of empty boxes, and possibly the boundary of the square.*

We now characterize the paths that a random set of less than $\lfloor r/c_1 \rfloor$ empty boxes can form in a square of size r. We show:

Lemma 4. *Given a square of size r, with $\lfloor r/c_1 \rfloor$ empty boxes occurring in it randomly, then the probability that there is a path empty boxes of length $l \geq 3\log_2 r$ is less than $(r/3)(3/(c_1-1))^l$.*

Proof: We give an upper bound on the probability by (over) counting the number of possible ways to place the empty boxes so that a path of length l is formed and then divide by the possible number of ways of placing the $\lfloor r/c_1 \rfloor$ empty boxes in the square. There are r ways to choose the start of the path and 3^{l-1} ways to continue it (we don't go back the way we came). The number of ways to pick the remaining sensors is $\binom{r-l}{\lfloor r/c_1 \rfloor - l}$. The total number of ways of placing r/c_1 empty boxes in the square is $\binom{r}{\lfloor r/c_1 \rfloor}$. Thus we have:

$$r3^{l-1}\frac{\binom{r-l}{\lfloor r/c_1 \rfloor - l}}{\binom{r}{\lfloor r/c_1 \rfloor}} \ , \tag{2}$$

which is

$$\leq (r/3)(3/(c_1-1))^l \ . \tag{3}$$

This expression is less than $1/r$ when $c_1 > 7$ and $l > 2\log_2(r/3)$).

From the lemma above, we see that with high probability there is no path of empty boxes in the square which goes from one side of the square to its opposite. All other connected components in the square either share a border on the outer boundary of the square, or are *timbits*.

Theorem 1 *With high probability, there is a single connected component which touches all four sides of the $n \times n$ square, (2) all other contiguous connected components have area $O(\log^2 N)$ and (3) all contiguous uncovered portions have area at most $O(\log N)$.*

Proof: Proof of (1) is immediate from Lemma **??**.

Proof of (2) follows from Lemma **??** as well, however, if the perimeter of all such other contiguous connected components is $O(\log N)$, then the area of these components is $O(\log^2 N)$.

Proof of (3) follows from a slight modification to Lemma **??**. In order to bound the area of the uncovered portion, we will treat it as a tree of length l embedded in the grid. This tree can be described by walk of length $2l$. We modify Lemma **??** as follows. Rather than having 3^{l-1} ways to continue that path, we have 4^{2l} ways, i.e., we may have to back-track. The result is that the probability that there is an uncovered portion of the $n \times n$ square is:

$$r4^{2l}\frac{\binom{r-l}{\lfloor r/c_1 \rfloor - l}}{\binom{r}{\lfloor r/c_1 \rfloor}} \ , \tag{4}$$

which is

$$\leq (r4^l)(4/(c_1-1))^l \ . \tag{5}$$

This expression is less than $1/r$ when $c_1 \geq 33$, and $l > 2 \log r$.

It remains to show that this single connected component contains $\Omega(N)$ sensors. The problem is that it may contain holes. We next bound the expected area which is not covered by the largest component.

5 The Largest Connected Component

We now show the expected size of the largest connected component. We prove an upper bound on the expected portion of all areas of the $n \times n$ square which may not be part of the largest component, i.e. all boxes containing elements of timbits and all empty boxes. These parts are bounded as follows:

- Lemma **??** bounds the expected area of all boxes which haven't too many empty boxes, and which contain timbits.
- Lemma **??** bounds the expected area of squares which have too many empty boxes, even though they have a sufficient number of sensors.
- Lemma **??** bounds the expected area of squares which have too many empty empty boxes because they receive too few sensors.
- Lemma **??** bounds the expected area of empty boxes.
- Observation 2 bounds the area which contains rectangles of abnormal size, when the $n \times n$ square is partitioned into r-squares.

We add this all up, and subtract this amount from N to get a lower bound on the expected area covered by the largest component (and the number of sensors in the largest component).

We do this by partitioning the $n \times n$ square into smaller squares. We first bound the expected area of squares which have long paths of empty boxes, then bound the expected area of squares with too many empty boxes.

5.1 Timbits

In order to bound the area of the $n \times n$ square which is occupied by timbits, we will decompose the $n \times n$ square into grids of squares of size $3^2, ..., \log^2 N$. The size of one of these squares is equal to the number of boxes it contains. An *r-square* is a square of size r. Note that we need only consider squares up to size $\log^2 N$ as we have already shown that with high probability, there are no paths of empty boxes of length $O(\log N)$. We then use these grids of squares to bound the number of timbits that can be contained (along with the path that encircles them) in a square of a particular size. Counting over all possible sizes, we can bound the total area occupied by timbits. In Lemma **??**, we end up over counting this area, because we can't say how a timbit may fall within a fixed $\sqrt{r} \times \sqrt{r}$ grid. What we can say however, is that any timbit will reside in at most four r-squares, for some r.

Note however, that when partitioning the $n \times n$ square into these grids, the last row or column of this grid may contain rectangles of area less than r. To account for this, we bound the size of the $n \times n$ square which can contain these *deviant* rectangles.

Observation 5. *The total area of these deviant rectangles is $O(r\log(n))$, which is less than $N/5c_3$ for sufficiently large N.*

Lemma 6. *For every timbit contained in the $n \times n$ square, there exists an s such that the timbit is contained in up to four grid squares of size s, one of which contains a path of length at least $(2\sqrt{s}+2)/4$.*

Proof: We define the length of a timbit as follows. Let x_1 (x_m) be the column number of the left-most (right-most) box in the timbit which contains a sensor. Similarly, let y_1 (y_m) be the row number of the top-most (bottom-most) box of the timbit which contains a sensor. Let the length of the timbit be: $max\{x_m - x_1, y_n - y_1\}$.

The empty path which surrounds the timbit must be at least $2t+6$ in length to completely enclose the timbit. Place around the timbit the smallest square, s, so that it completely contains both the timbit and the path surrounding it. Note that size of $s = (t+2)^2$. Regardless of how the timbit sits in this any grid (as described earlier), it will only reside in at most four of these squares. This path must be at least $2t+6$ in length, which in terms of s is: $2\sqrt{s}+2$.

Note that for any timbit, this implies that there are at most four squares of size s that contain an empty path of length greater than or equal to $(1/4)(2\sqrt{s}+2) = (\sqrt{s}+1)/2$. We must also consider the case that the timbit covers one or more of the borders of the $n \times n$ square. It is easy to see, however, that the above lemma still holds, as any timbit covering a border, must still be cut off from the rest of the components by a path, this path must be at least half the length of the timbit, and this timbit must lie in at least one square of size s which contains at least half of this path of empty boxes. This the path must be at least $(\sqrt{s}+1)/2$, in at least one square of size s, as required.

Lemma 7. *For any constant c_3, there exists a constant c_1, such that the expected sum over r of all areas of r-squares, which have at most r/c_1 empty boxes and contain timbits is less than $N/5c_3$.*

Proof: This follows from Lemma **??**. To account for all possible timbits, for a particular r, we multiply the probability of the path occurring $((r/3)(3/(c_1-1))^l)$, by the number of boxes in the square $(4r)$, times the area that squares of this size take up in the $n \times n$ square $(8N/r)$. Summed over all possible r-values, this gives us the expected size of the portion of the $n \times n$ square which is occupied by these timbits. So for a fixed r, the expected area of these timbits is:

$$\leq 4r(8N/r)(r/3)(e/c_1-1)^{(\sqrt{r}+1)/2} \quad . \qquad (6)$$

And over all possible r-values, the expected area of the $n \times n$ square occupied by these timbits is:

$$\leq \sum_{i=3}^{(\log N)} 4i^2(8N/i^2)(i^2/3)(3/(c_1-1))^{(\sqrt{i^2}+1)/2}$$

$$= 4N/3 \sum_{i=3}^{(\log N)} (i^2)(3/(c_1-1))^{(i+1)/2} \leq N/5c_3 \ , \tag{7}$$

for some c_1.

Lemma 8. *For any constants $c_1, d > 1$, there exists a constant c, such that the probability that an r-square receives cr sensors and ends up with greater than r/c_1 empty boxes is $1/e^{dr}$.*

Proof: This follows from Lemma **??**.

Lemma 9. *For any constants c_1, c_3, there exists a constant c such that for all r, the expected sum of area of all boxes in the $n \times n$ square which are contained in an r-square which (1) receives cr sensors and (2) ends up with more than r/c_1 empty boxes, is less than $N/5c_3$.*

Proof: For a fixed r, the area of r-squares which have too many empty boxes is just equal to the area of all r-squares in the $n \times n$ square, times the probability that, an r-square receives cr sensors and ends up with more than r/c_1 empty boxes. This follows from Lemma **??**, and is: $(8N/r)(r)(1/e^{dr}) = 8N/e^{dr}$. Summed over all possible r values, the total area of the $n \times n$ square which receives cr sensors, but ends up with more than r/c_1 empty squares is:

$$\sum_{i=3}^{\log N} 8N(1/e^{di^2})$$

$$= 8N \sum_{i=3}^{\log N} (1/e^{di^2}) \leq 8N/e^{d+2} \leq N/5c_3 \ . \tag{8}$$

Lemma 10. *For any constant c_3, there exists a constant c such that if cN sensors are thrown independently at random into a square of area N, then the expected sum of area of all boxes contained in any r-square which receives fewer than $c_2 r$ sensors, where $c_2 \geq (1 + \ln(c_3)/2)$, is less than $N/5c_3$.*

Proof: Note that this is the same as the ball and bins problem. Dividing the $n \times n$ square into r-squares, we have $8N/r$ bins, into which we throw, independently at random, cN balls (sensors). We bound the probability (for a fixed r) that an r-square receives fewer than $c_2 r$ sensors. The expected number of sensors received by an r-square is $\mu = \frac{cN}{8N/r} = cr/8$. Setting $c = 16c_2$. We use the following Chernoff bound [**?**]:

$$Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2} \ . \tag{9}$$

Setting $\delta = 1/2$, and $(1-\delta)\mu = c_2 r$. This gives:

$$Pr(X \leq c_2 r) \leq e^{-c_2 r/4} \ . \tag{10}$$

Let $\delta = 1/2$. The expected number of boxes in the $n \times n$ square, for fixed r, with too few sensors is $(r)(8N/r)(e^{-c_2 r/4}) = 8Ne^{-c_2 r/4}$. Summed over all feasible r, this area is then:

$$\sum_{i=3}^{\log N} \frac{8N}{e^{i^2 c_2/4}}$$

$$= 8N \sum_{i=3}^{\log N} \frac{1}{e^{i^2 c_2/4}} \leq 8N/e^{c_2+2} \leq N/5c_3 \ , \quad (11)$$

where $c_2 \geq (1 + \ln(c_3)/2)$.

Lemma 11. *For any constants c_1, c_3, there exists a constant c such that if cN sensors are thrown independently at random into a square of area N, then the expected area of all boxes contained in any r-square with more than r/c_1 empty boxes, for any r is less than $2N/5c_3$.*

Proof: Let a be the constant c from Lemma **??**. Let $b = max\{a, c_2\}$, where c_2 is the constant from Lemma **??**. Now, from Lemma **??**, there exists a constant c such that if cN sensors are thrown into a square of area N, the expected area of all boxes contained in any r-square which receives fewer than br sensors is less than $N/5c_3$.

From Lemma **??**, the expected area of all boxes contained in any r-square receiving br sensors with more than r/c_1 empty boxes is less than $N/5c_3$. Since a box in an r-square is either in an r-square with fewer than br sensors, or in a r-square with more than br sensors, the lemma follows.

Theorem 12. *For any constant c_3, there exists a constant c such that if cN sensors are thrown into a square of area N independently at random, then the area covered by the largest component is $N - N/c_3$.*

Proof: Every point which is not covered by the largest component is contained in either a box containing an element of a timbit, or an empty box. We choose c to be the maximum of the c's in Lemmas **??**, **??** and **??**. Along with the amount from Observation 2, the area covered by the largest component is $N - N/c_3$, for any constant c_3.

6 Discussion

This result has algorithmic implications. For example, since the small components contain less than $O(\log^2 N)$ sensors, a sensor can detect if it is contained in the large connected component by a flooding transmission which takes $c \log^2 N$ steps. If after $c \log^2 N$ steps there continue to be new sensors discovered, then the sensor "knows" that it is in the largest component. Another implication is that the shortest path between any two connected sensors is at most $O(n \log N)$, with high probability.

Further analytical and/or empirical work is needed to determine the bounds on the values of the constants.

7 Acknowledgements

We would like to thank Kui Wu for suggesting this problem.

References

1. E.H. Jennings, C.M. Okino. On the Diameter of Sensor Networks. *Proceedings of the 2002 IEEE Aerospace Conference*, Big Sky, Montana, Mar. 2002.
2. P. Gupta, P. R. Kumar. Critical Power for Asymptotic Connectivity in Wireless Networks. in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming* (W. M. McEneany, G. Yin, and Q. Zhang, eds), pp. 547-566, Boston, MA: Birkhauser, March 1998.
3. F. Xue and P. R. Kumar. The number of neighbours needed for connectivity of wireless networks. Manuscript, 2002.
4. S. Shakkottai, R. Srikant, N. Shroff. Unreliable Sensor Grids: Coverage, Connectivity and Diameter. In *IEEE INFOCOM 2003*.
5. M. D. Penrose. Single Linkage Clustering and Continuum Percolation. *Journal of Multivariate Analysis 53*, pp. 94-109, 1995.
6. J. Diaz, M. D. Penrose, J. Petit, M. Seina. Convergence Theorems for Some Layout Measures on Random Lattice and Random Geometric Graphs. *Combinatorics, Probability and Computing 9*, pp. 489-511, 2000.
7. M. D. Penrose. Random Geometric Graphs. Oxford University Press, 2003.
8. R. Motwani, P. Raghavan. *Randomized Algorithms*, p 70, Cambridge Press, 1995.