Multiple Round Random Ball Placement: Power of Second Chance

XiangYang Li[*] Yajun Wang[†] Wangsen Feng[‡]

ABSTRACT

In a pioneering work, Gupta and Kumar [8] studied the critical transmission range needed for the connectivity of random wireless networks. Their result implies that, given a square region of $\sqrt{n} \times \sqrt{n}$, the asymptotic number of random nodes (each with transmission range 1) needed to form a connected network is $\Theta(n \ln n)$ with high probability. This result has been used as cornerstones in deriving a number of asymptotic bounds for random multi-hop wireless networks, such as network capacity [7, 10, 11, 14]. In this paper we show that the asymptotic number of nodes needed for connectivity can be significantly reduced to $\Theta(n \ln \ln n)$ if we are given a "second chance" to deploy nodes. More generally, under some deployment assumption, if we can deploy nodes in k rounds (for a constant k) and the deployment of the ith round can utilize the information gathered from the previous $i-1$ rounds, we show that the number of nodes needed to provide a connected network with high probability is $\Theta(n \ln^{(k)} n)$. (See Eq (1) for the definition of $\ln^{(k)} n$.) Similar results hold when we need deploy sensors such that the sensing regions of all sensors cover the region of interest.

Keywords

Random Deployment, wireless ad hoc networks, second chance.

1. INTRODUCTION

In wireless ad hoc networks, autonomous nodes form self-organized networks without centralized control or infrastructure, i.e. there are no wired infrastructure or cellular networks. In the last few years, there has been a big interest in ad hoc wireless networks, especially, wireless sensor networks, as they have tremendous military and commercial applications. Ad hoc wireless networks can be deployed quickly anywhere and anytime as they eliminate the complexity of the infrastructure setup. In all such application scenarios, quick and easy deployment of a connected network is essential for wireless ad hoc networks, especially wireless sensor networks.

There are two ways to deploy a wireless sensor network: random deployment and precise deployment. Random deployment means setting positions of wireless sensor nodes randomly and independently in the target area. On the other hand, in precise deployment, nodes are set at exact positions one by one according to the communication range of the nodes. Usually, the positions are chosen to minimize the number of nodes required to achieve certain deployment goal. Clearly, precise deployment method is time consuming however costing the least number of nodes. Random deployment method is fast in practice though costs a relatively larger number of nodes to achieve the same deployment goal. Considering practical application scenarios, random deployment is a feasible and practical method, and sometimes it is the only feasible strategy (such as deploying sensors in a hostile environment).

Consider a square region of size $a \times a$. We assume all nodes have the same communication range r. Each wireless sensor node can communicate directly with all neighboring nodes within its communication range. One natural question is that how many nodes are needed to construct a connected network if all nodes are placed in the region randomly and uniformly. Gupta and Kumar [8] studied the critical transmission range (CTR) for connectivity: what is the CTR for connectivity if we randomly deploy n sensor nodes in a unit square. Their pioneering work showed that if n nodes V are uniformly randomly deployed in a unit square and the transmission range r satisfies that $n\pi r^2 \geq c_1 \ln n$ for some constant c_1, then the network $G(V, r)$ (two nodes of V are connected in G iff their Euclidean distance is at most r) will be connected with probability 1 when $n \to \infty$. This implies that, when all nodes have transmission range r, the asymptotic number m of nodes needed to deploy in a square of size $a \times a$ such that the resultant network is connected satisfies that $m\pi(\frac{r}{a})^2 = \Theta(\ln m)$. Let $n = (a/r)^2$. Therefore $m = \Theta(n \ln n)$. Another similar problem is to study the number of random nodes to deploy in the square region such that the entire region is covered by the disks centered at the sensors with radius r (called sensing disk for each node). Previous results [17, 18] proved that it costs $m = \Theta(n \ln n)$ nodes to cover the target square area with high probability when all sensor nodes are deployed randomly. Consequently, if we randomly deploy $\Theta(n \ln n)$ nodes in the square region, the resultant network will be connected and the square region will be covered by the sensors with high probability. On the other hand, if precise deployment is available, only $\Theta(n)$ sensors is needed to achieve both connectivity and coverage requirements. In other words, the cost overhead introduced by the

[*]Department of Computer Science, Illinois Institute of Technology, 10 W. 31st Street, Chicago, IL 60616, USA. Email: xli@cs.iit.edu

[†]Department of Computer Science, HongKong University of Science and Technology, HongKong, P.R.China. Email: yalding@cse.ust.hk

[‡]Department of Computer Science, Peking University, BeiJing, P.R.China. Email: fengws@pku.edu.cn

convenient random deployment is that we need an $\Omega(\ln n)$ factor of more sensors. Observe that, if we require the resultant network to be κ-connected or g-coverage for some constants $\kappa \geq 1$ and $g \geq 1$, the number of nodes needed by a random deployment is still in the order of $\Omega(n \ln n)$ [16, 17, 19, 20]. As we mentioned, convenient deployment of wireless sensor networks is critical for some applications, however, reducing the cost (i.e., the number of sensor nodes needed by the applications) is also important. For example, currently the price of a wireless sensor typically ranges from \$100 to \$500, e.g., ITS400 - IMOTE2 sensor board by crossbow costs about \$249 (as of Jan. 25th, 2008). Thus, it is important to reduce the number of sensors while do not sacrifice the convenience of random deployment.

In this paper, we will employ the concept of "second chance" to design a random deployment method. In our second chance deployment approach, we first *randomly* deploy a certain number of wireless sensor nodes $\mathcal{S}_1$ into the targeted deployment region. We then use the information collected from the nodes in $\mathcal{S}_1$ to decide where to deploy sensors $\mathcal{S}_2$ in the second round of deployment. The sensors $\mathcal{S}_1 \cup \mathcal{S}_2$ deployed in both rounds will form the final deployment.

We assume the target deployment region is a square region $\mathcal{R}$ of size $a \times a$. Intuitively, in the first round, each node is uniformly and randomly placed into the region. In following rounds, however, we assume the ability to reduce the region of placement for random nodes. Assume each node placed in the covers a disk with radius r and the union of all suck disks from the sensors in the first $i-1$ round is $\mathcal{R}_1$. In ith round, we assume the ability to randomly deploy nodes in region $\mathcal{R} \setminus \mathcal{R}_1$. In practice, though, it is hard to acquire $\mathcal{R}_1$. Instead, if we appropriately choose the number of nodes deployed in the first round, there exists a unique large connected component which touches the boundary. Therefore, it is easy to detect this large component simply by querying the sensors on the boundary of the deployment region. Localization algorithms [1, 2, 5, 6, 15] exist to find the location of the wireless nodes in this component, which gives us a subset $\mathcal{R}'$ of $\mathcal{R}_1$. We assume we are able to randomly deploy wireless sensor nodes in the region $\mathcal{R} \setminus \mathcal{R}'$. In this paper, we study both cases.

Formally, in a "second chance" deployment (or more generally a k-phrase deployment), we find a *giant* component formed by the sensors in $\mathcal{S}_1$ in first round (or the first $i-1$ rounds). Let $\mathcal{R}'$ be the region covered by this *giant* component. We then deploy $\mathcal{S}_2$ in the region excluding $\mathcal{R}'$.

Assume the communication range for the wireless nodes is r. Let $n = (a/r)^2$. We first prove that, by properly choosing the size of $\mathcal{S}_1$, there is a giant connected component formed by $\mathcal{S}_1$ which touches the boundary of the deployment region. We then theoretically prove that our second chance deployment strategy costs $\Theta(n \ln \ln n)$ nodes to form a connected network with high probability, and costs $\Theta(n \ln \ln n)$ nodes to cover the target area with high probability. This bound is tight for any "second chance" method, which means that one cannot achieve the same goal with asymptotically less than $\Theta(n \ln \ln n)$ nodes by *any "second chance" method*. Our results, together with the power of multiple choices (by Mitzenmacher [13]) show that the system performances can be greatly improved if the choices (temporal choices used in this paper and spacial choices used in [13]) are only slightly relaxed. Our second chance deployment will significantly reduce the number of sensors required to achieve connectivity and coverage. Assume that $n = 1000$, the traditional one-phase random deployment will need around $n \log n \simeq 10,000$ sensors, while our second chance deployment strategy only requires about $n \log \log n \simeq 3322$ sensors. Thus, the second-chance deployment strategy could save us about 66% number of sensors in this case. Our extensive simulations confirm our theoretical findings.

The results developed here could be used to enhance our study of the performances of (random) wireless networks in several aspects. Firstly, using a *second chance* deployment (or more generally, a k-phase deployment), we can significantly reduce the number of wireless devices required to accomplish certain missions, *e.g.*, forming a connected network, providing full monitoring coverage of a certain region. The saving on the number of devices is already in the order of $\Theta(\frac{\log n}{\log \log n})$ even if we randomly deploy the wireless sensors nodes in two rounds. Secondly, when a wireless network is deployed using a "second chance" deployment, we will inherently improve the per-node unicast (or multicast) capacity because the network nodes density is reduced compared with the traditional random deployment. The findings in this paper can be viewed as the first step towards the study of trade-offs between network performances and the complexity of deploying the network. It remains as a future work to study the asymptotic network capacity (and other properties such as the critical transmission range for certain localized routing methods) of such randomly deployed networks using "second chance" deployment.

The rest of paper is organized as follows. In Section 2, we present new technical lemmas on the traditional bin and balls problem, which will be used to derive the results in this paper. These results could be of independent interest. In Section 3, we study the largest component of a random wireless network when m nodes are randomly deployed in a square region of n squarelets. We obtain tight bounds on the number of sensors nodes for the coverage problem in Section 4 and connectivity problem (and related critical range for connectivity) in Section 5 under the "second chance" deployment strategy. Extensive simulations are presented in Section 6. We conclude the paper in Section 7 with the discussion of some future problems.

2. TECHNICAL LEMMAS

Our analysis will be based on some key results from the balls and bins problem. In this section, we revisit the problem of randomly placing m *identical and indistinguishable* balls into n distinguishable (numbered) bins. Instead of placing balls in one round (*i.e.*, in one shot), we study the case that we are able to identify the empty bins after previous rounds of random placement and place balls into the remaining empty balls in the next round uniformly and randomly. We are interested in the asymptotic number of balls required to fill all the bins with high probability under our new model in k rounds. Here an event is said to happen *with high probability* (*w.h.p.*), if it happens with probability at least $1 - \frac{1}{n}$. Placing balls into bins can be used to study the random wireless sensor deployment by the following observation: Assuming that each sensor has communication range 1, we partition the deployment region into grids (with cell size $1/\sqrt{2}$); To cover the region using sensing disks defined by randomly placed sensors, it suffices that each bin is filled with at least one sensor. A similar analog can be made between the connectivity of the sensor networks and the number of *unit bins* (See Section 2.2).

2.1 Filling Empty Bins

Let n be the number of bins. We denote

$$\ln^{(k)} n = \underbrace{\ln \ln \ldots \ln}_{k\text{times}} n. \tag{1}$$

Notice that here $\ln^{(k)} n$ is only defined when $\ln^{(k-1)} n > 0$.

FACT 1. *For all values of n and t with $n \geq 1$ and $|t| \leq n$, it holds that*

$$e^t\left(1-\frac{t^2}{n}\right) \leq \left(1+\frac{t}{n}\right)^n \leq e^t.$$

DEFINITION 1 (k-ROUND BALL PLACEMENT (k-RBP)). *In a k round ball placement, balls are randomly placed into n bins in k rounds. Let $n_0 = n$ be the original number of empty bins and n_i be the number of empty bins after i rounds. In the ith round, m_i balls are randomly placed into n_{i-1} remaining empty balls. $m = \sum_{i=1}^k m_i$ is the total number of balls placed.*

Observe that a key requirement here for k-RBP is, after the first i rounds of placing balls randomly, we are able to determine the empty bins left and randomly place balls into the remaining empty balls. Let random variable Z be the number of empty bins when m balls are placed randomly into n bins in one round and μ be its expected value. Then

$$\mu = E[Z] = n\left(1-\frac{1}{n}\right)^m \quad \sim \quad ne^{-\frac{m}{n}}.$$

Define the function $H(m,n,z)$ as the probability that $Z = z$, *i.e.*, $H(m,n,z) = \Pr(Z = z)$. The following occupancy bounds were proved in [9].

LEMMA 1 (OCCUPANCY BOUND 1 [9]). *For any $\theta > 0$,*

$$\Pr[|Z-\mu| \geq \theta\mu] \leq 2\exp\left(-\frac{\theta^2\mu^2(n-1/2)}{n^2-\mu^2}\right) \tag{2}$$

LEMMA 2. (OCCUPANCY BOUND 2 [9]) *For $\theta > -1$,*

$$H(m,n,(1+\theta)\mu) \leq \exp(-((1+\theta)\ln[1+\theta]-\theta)\mu) \tag{3}$$

In particular, for $-1 \leq \theta < 0$,

$$H(m,n,(1+\theta)\mu) \leq \exp\left(-\frac{\theta^2\mu}{2}\right) \tag{4}$$

Here $\exp(x)$ denotes e^x for any x. The following lemma is directly implied by the *Occupancy Bound* 2 proved in [4,9].

LEMMA 3. *Assume that m balls are randomly placed into n bins. Let variable Z be the number of empty bins. We have:*

$$\mu = E[Z] = n(1-\frac{1}{n})^m \tag{5}$$

$$\Pr[Z=0] = \exp\left(-\frac{\mu}{2}\right) \tag{6}$$

2.1.1 Upper bound

We are interested in the number of balls required to fill n bins *w.h.p.* in any k *round ball placement*. In particular, we are interested in $m = \sum_i^k m_i$ so that with high probability there are no empty bins left.

We assume $k \geq 2$. In order to make our argument valid, we require that: $\ln^{(k)} n \geq 2$ which implies $n \geq e^{e^2} \geq 14$. We also require that

$$n(\ln^{(k)} n - 1) \geq 2(\ln n)^2. \tag{7}$$

This condition will hold when n is sufficiently large (depending on k). The following lemma discusses the case when we want to cover the bins by placing balls in one round.

LEMMA 4. *If we randomly place $(1+\epsilon)n\ln n$ balls into n bins, then all the n bins are occupied with probability at least $1-\frac{1}{n^\epsilon}$, where $0 < \epsilon \leq 1$ is a constant.*

PROOF. The probability p that there is at least one empty bin left is

$$p \leq \binom{n}{1}\left(1-\frac{1}{n}\right)^{(1+\epsilon)n\ln n} \leq n\left(\frac{1}{e}\right)^{(1+\epsilon)\ln n} = \frac{1}{n^\epsilon}$$

Thus, the probability that all the n bins are occupied is at least $1-\frac{1}{n^\epsilon}$. This implies that $2n\ln n$ balls (by setting $\epsilon = 1$) suffices to cover n bins with probability at least $1-\frac{1}{n}$. $\square$

For the simplicity of presentation later, we define function

$$f(n,k,\delta) = \frac{1}{\ln^{(k)} n}(1-\frac{\ln^{(k-1)} n \ln\delta}{n}).$$

LEMMA 5. *Let constant $\delta \in (0,1)$. If we randomly place $(1+f(n,k,\delta))n\ln^{(k)} n$ balls into n bins, then with probability at least $(1-\delta)$, the number of empty bins afterwards is at most $\frac{n}{\ln^{(k-1)} n}$.*

PROOF. To simply our notation, let $s = \ln^{(k-1)} n$ and then $\ln s = \ln^{(k)} n$. Let us consider the first $\frac{n}{s}$ bins. The probability that these bins are empty is:

$$(1-\frac{1}{s})^{(1+f(n,k,\delta))n\ln s} \leq e^{-\frac{(1+f(n,k,\delta))n\ln s}{s}}.$$

By union bound, the probability that there are more than $\frac{n}{s}$ empty bins is at most:

$$e^{-\frac{(1+f(n,k,\delta))n\ln s}{s}}\binom{n}{\frac{n}{s}} \leq e^{-\frac{(1+f(n,k,\delta))n\ln s}{s}}(e\cdot s)^{\frac{n}{s}}. \tag{8}$$

The inequality is based on the Sterling's approximation: $\forall 0 < m \leq n$,

$$\left(\frac{n}{m}\right)^m \leq \binom{n}{m} \leq \left(\frac{ne}{m}\right)^m.$$

Because $s^{1/s} = e^{\frac{\ln s}{s}}$, the above inequality (8) actually becomes $e^{-\frac{(1+f(n,k,\delta))n\ln s}{s}}(e\cdot s)^{\frac{n}{s}} \leq e^{\frac{n}{s}(1-f(n,k,\delta)\ln s)} = \delta$. This finishes the proof. $\square$

LEMMA 6. *Let $2 \leq l \leq k-1$. Randomly place $2n$ balls in $\frac{n}{\ln^{(l)} n}$ bins. With probability at least $(1-\frac{1}{kn})$, the number of empty bins afterwards is at most $\frac{n}{\ln^{(l-1)} n}$.*

PROOF. The probability p that the first $\frac{n}{\ln^{(l-1)} n}$ bins are empty is:

$$p = \left(1-\frac{\ln^{(l)} n}{\ln^{(l-1)} n}\right)^{2n} \leq e^{-\frac{2n\ln^{(l)} n}{\ln^{(l-1)} n}}.$$

By union bounds, the probability that there are more than $\frac{n}{\ln^{(l-1)}}$ empty bins is at most:

$$\begin{aligned}
p\cdot\binom{\frac{n}{\ln^{(l)} n}}{\frac{n}{\ln^{(l-1)} n}} &\leq p\left(\frac{e\ln^{(l-1)} n}{\ln^{(l)} n}\right)^{\frac{n}{\ln^{(l-1)} n}} \\
&\leq p\cdot e^{\frac{n}{\ln^{(l-1)} n}}(\ln^{(l-1)} n)^{\frac{n}{\ln^{(l-1)} n}} \\
&\leq e^{-\frac{2n\ln^{(l)} n}{\ln^{(l-1)} n}} e^{\frac{n}{\ln^{(l-1)} n}} e^{\frac{n\ln^{(l)} n}{\ln^{(l-1)} n}} \\
&\leq e^{\frac{n}{\ln^{(l-1)} n}(\ln^{(l)} n - 1)} \\
&\leq e^{-2\ln n} \leq \frac{1}{kn}.
\end{aligned}$$

The last inequality comes from eq. (7). This finishes the proof. $\square$

We are now ready to study the upper bound for the number of balls needed for covering n bins *w.h.p.* using k-RBP.

THEOREM 7. *Let $k \geq 2$ be an integer constant. There exists a* k round ball placement *in n bins, such that with $(1 + f(n, k, \frac{1}{kn}) + \frac{2(k-1)}{\ln^{(k)} n}) n \ln^{(k)} n$ balls, it is sufficient to fill all the bins with probability at least $(1 - \frac{1}{n})$.*

PROOF. We first place $(1+f(n, k, \frac{1}{kn}))n \ln^{(k)} n$ balls randomly into n empty bins. From Lemma 5, with probability at least $(1 - \frac{1}{kn})$, the number of empty bins after first round is at most $\frac{n}{\ln^{(k-1)} n}$. In ith round, for $2 \leq i \leq k-1$, we place $2n$ balls randomly into the empty bins after $(i-1)$'th round. By Lemma 6, after $k-1$ round, with probability at least $(1 - \frac{1}{k \cdot n})^{k-1} \geq 1 - \frac{k-1}{kn}$, the number n_{k-1} of empty bins left is at most $\frac{n}{\ln n}$.

In the k-th round, we randomly place another $2n$ balls. We show that all the left n_{k-1} empty bins are filled with high probability. From Lemma 4, in order to achieve probability at least $(1 - \frac{1}{kn})$, it requires $(1 + \frac{\ln(kn)}{\ln n_{k-1}}) n_{k-1} \ln n_{k-1} \leq 2n$ balls (when $n_{k-1} \leq \frac{n}{\ln n}$ and $k \leq \ln n$). Thus if the number n_{k-1} of empty bins left before the last round k is at most $\frac{n}{\ln n}$, randomly placing $2n$ balls suffices to fill all the empty balls with probability at lest $(1 - \frac{1}{kn})$.

In total, we place $(1 + f(n, k, \frac{1}{kn}) + \frac{2(k-1)}{\ln^{(k)} n}) n \ln^{(k)} n$ balls. The probability of success is at least $(1 - \frac{1}{k \cdot n})^k \geq (1 - \frac{1}{n})$. This finishes the proof. $\square$

Remark: Note that $f(n, k, \frac{1}{kn} + \frac{2(k-1)}{\ln^k n}) = o(1)$ when $n \to \infty$ and k remains constant. Essentially, it is sufficient to use $n \ln^{(k)} n$ balls to fill n bins with high probability by a k *round ball placement*.

2.1.2 Lower bound

We now show that the bound of $n \ln^{(k)} n$ in Theorem 7 is tight. In particular, if the number of balls is $(1 - o(1)) n \ln^{(k)} n$, any k *round ball placement* will always have empty bins left with constant probability. We assume $\ln^{(k+1)} n \geq 1$. We first prove the case when $k = 1$.

LEMMA 8. *Randomly place $n \ln n$ balls into n bins. With constant probability, there is at least one empty bin.*

PROOF. It is easy to show that the expected number of empty bins is at least:

$$n(1 - \frac{1}{n})^{n \ln n} \geq 1 - \frac{(\ln n)^2}{n} \geq 1/2.$$

From Lemma 3, the probability that there is no empty bin left is at most $e^{-1/4}$. Thus, with constant probability (at least $1 - e^{-1/4}$), there is at least one empty bin. $\square$

We then prove the lower bound for any k-RBP. Compared with the upper bound on the balls required, the difficulty here is that we do not have any constraints on the distribution of m balls over k rounds of the deployment. We essentially have to show that, given a certain number of m balls, regardless of the choices of m_i $(1 \leq i \leq k)$, there are empty bins with constant probability when m is less than some number.

LEMMA 9. *Let $k > 1$ be an integer constant. Randomly place $(1 - \frac{4 \ln^{(k+1)} n}{\ln^{(k)} n}) n \ln^{(k)} n$ balls into n bins in one round. With constant probability, there are at least $\frac{n (\ln^{(k)} n)^2}{\ln^{(k-1)} n}$ empty bins afterwards.*

PROOF. Let $\epsilon = \frac{4 \ln^{(k+1)} n}{\ln^{(k)} n}$. Let Z be the random variable for the number of empty bins. The expected number of empty bins is:

$$\mu = E[Z] = n(1 - \frac{1}{n})^{(1-\epsilon) n \ln^{(k)} n}.$$

Because $1 - \frac{(1-\epsilon) \ln^{(k)} n}{n} \geq 1/2$, by Fact 1,

$$\mu \geq \frac{n}{2} e^{-(1-\epsilon) \ln^{(k)} n}.$$

Since $\ln^{(k+1)} n \geq 1$, we have $\ln^{(k)} n \geq e > 2$. Also $e^{-(1-\epsilon) \ln^{(k)}} = \frac{(\ln^{(k)} n)^4}{\ln^{(k-1)} n} \geq \frac{4 (\ln^{(k)} n)^2}{\ln^{(k-1)} n}$, we have

$$\mu \geq \frac{2n (\ln^{(k)} n)^2}{\ln^{(k-1)} n}.$$

By the occupancy bound, the probability that $Z \geq \mu/2$ is at least:

$$\Pr[Z \geq \frac{\mu}{2}] \geq 1 - \Pr[|Z - \mu| \geq \frac{\mu}{2}] \geq 1 - 2e^{-\frac{\mu^2 (n - 1/2)}{n^2 - \mu^2}}.$$

As $n \geq (\ln n)^4$, we have $\mu^2 \geq 2n$. Consequently,

$$\Pr[Z \geq \frac{\mu}{2}] \geq 1 - 2e^{-\frac{\mu^2}{2n}} \geq 1 - \frac{2}{e} > 0.$$

This finishes the proof. $\square$

LEMMA 10. *Let $2 \leq l \leq k$. Randomly place $n \ln^{(k)} n$ balls into $\frac{n (\ln^{(l)} n)^2}{\ln^{(l-1)} n}$ bins in one-round. With constant probability, the number of remaining empty bins afterwards is at lest $\frac{n (\ln^{(l-1)} n)^2}{\ln^{(l-2)} n}$.*

PROOF. Because $\ln^{(k)} n \leq \ln^{(l)} n$ when $l \leq k$, let us assume we place $n \ln^{(l)} n$ balls instead. Because we actually increase the number of balls. The argument on the lower bound still holds with fewer balls.

Let Z be the random variable for the number of empty bins. The expected number of empty bins is:

$$\mu = E[Z] = \frac{n (\ln^{(l)} n)^2}{\ln^{(l-1)} n} (1 - \frac{\ln^{(l-1)} n}{n (\ln^{(l)} n)^2})^{n \ln^{(l)} n}.$$

Let $t = \frac{\ln^{(l-1)} n}{\ln^{(l)} n}$. We assume $n \geq 2 (\ln^{(l-1)} n)^2$, which implies that $(1 - \frac{t^2}{n \ln^{(l)} n}) \geq 1/2$. By Fact 1, we have

$$\mu \geq \frac{n (\ln^{(l)} n)^2}{\ln^{(l-1)} n} (1 - \frac{t^2}{n \ln^{(l)} n}) e^t.$$

We assume n is sufficiently large such that $\ln^{(l-1)} n \geq \ln 4 + 3 \ln^{(l)} n$. Hence $\mu \geq \frac{2n (\ln^{(l-1)} n)^2}{\ln^{(l-2)} n}$. Let $m = \frac{n (\ln^{(l)} n)^2}{\ln^{(l-1)} n}$. From the occupancy bound,

$$\Pr[Z \geq \frac{\mu}{2}] \geq 1 - \Pr[|Z - \mu| \geq \frac{\mu}{2}] \geq 1 - 2e^{-\frac{\mu^2 (m - 1/2)}{m^2 - \mu^2}}.$$

We assume n is sufficiently large such that $n \geq (\ln n)^4$, $\mu^2 \geq 2m$.

$$\Pr[Z \geq \frac{\mu}{2}] \geq 1 - 2e^{-\frac{\mu^2}{2m}} \geq 1 - \frac{2}{e} > 0.$$

This finishes the proof. $\square$

THEOREM 11. *Let k be an integer constant. Randomly place $m = (1 - \frac{4 \ln^{(k+1)} n}{\ln^{(k)} n}) n \ln^{(k)} n$ balls into n bins by **any** k-RBP. With constant probability, there exists at least one empty bin.*

PROOF. Based on Lemma 9 and Lemma 10, before last round, the number of empty bins is at least $\frac{n(\ln\ln n)^2}{\ln n}$ with constant probability (at least $(1-\frac{2}{e})^k$ for a constant k), even we randomly place $m = (1 - \frac{4\ln^{(k+1)} n}{\ln^{(k)} n})n\ln^{(k)} n$ balls into n bins in round 1, and m balls (that is less than $n\ln^{(k)} n$) in rounds 2 to $k-1$.

Let $m_k = \frac{n(\ln\ln n)^2}{\ln n}$. Assume $\ln\ln n > 2$ and $\ln\ln n \leq \frac{\ln n}{2}$. Obviously, we have $m_k \ln m_k \geq n\ln\ln n \geq n\ln^{(k)} n$. By Lemma 8, $n\ln^{(k)} n$ balls will leave at least one empty bin with constant probability after the last round k. Thus, with probability at least $(1-\frac{2}{e})^k$, randomly placing m balls by any k-RBP strategy will leave at least one empty bin. $\square$

2.2 Number of Unit Bins in One Round

We denote *unit bins* as the bins that have exactly one ball inside. As shown later, the existence of unit bins is closely related to connectivity of the network after the random deployment.

In this section, we develop the bound on the number of unit bins after place m balls into n bins randomly, which is similar to the occupancy bound 1 [4,9]. In particular, we follow the scheme of proof as in [4], which apply the "bounded difference" method introduced in [4].

Let $Z_i (1 \leq i \leq n)$ be the indicating random variable which is 1 if the ith bin has exactly one ball (bin i is a *unit bin*) and 0 otherwise. Let $Z = \sum_{i=1}^n Z_i$ be the number of unit bins. We are interested in tail bounds on the distribution of Z. Clearly,

$$E[Z_i] = \binom{m}{1}\frac{1}{n}(1-\frac{1}{n})^{m-1} = \frac{m}{n}(1-\frac{1}{n})^{m-1}$$

and

$$E[Z] = \sum_{i=1}^n E[Z_i] = m(1-\frac{1}{n})^{m-1}.$$

We denote $[n] = \{1, 2, ..., n\}$. We first recall the following lemma by McDiarmid [12].

LEMMA 12 (MCDIARMID [12]). *Let $X_1, ..., X_n$ be independent random variables, variable X_i taking values in a finite set A_i for each $i \in [n]$, and suppose the function f satisfies the following "bounded difference" conditions: for each $i \in [n]$, there is a constant c_i such that for any $x_k \in A_k, k \in [i-1]$ and for $x_i, x_i' \in A_i$*

$$|E[f(\boldsymbol{X})|X_1 = x_1, ..., X_{i-1} = x_{i-1}, X_i = x_i] -$$

$$E[f(\boldsymbol{X})|X_1 = x_1, ..., X_{i-1} = x_{i-1}, X_i = x_i']| \leq c_i.$$

then

$$Pr[|f(\boldsymbol{X}) - E[f(\boldsymbol{X})]| > t] < 2\exp\left(-\frac{t^2}{2\sum_i c_i^2}\right).$$

THEOREM 13 (OCCUPANCY BOUND FOR UNIT BINS). *Randomly place m balls into n bins. Let Z be the number of* unit bins *afterwards. For any $\theta > 0$,*

$$\Pr[|Z-\mu| \geq \theta\mu] \leq 2\exp\left(-\frac{\theta^2\mu^2(2n-1)}{8(m+n)^2(1-(1-\frac{1}{n})^{2m})}\right)$$

where, $\mu = E[Z] = m(1-\frac{1}{n})^{m-1}$ is the expected number of unit bins.

PROOF. To get a tail probability estimate, we view the variable Z as $Z = Z(B_1, \ldots, B_m)$ where the random variable B_k takes values in the set $[n]$ indicating which bin the ball k occupies, for each $k \in [m]$. In order to use Lemma 12, we need to compute the difference

$$D = |E[Z|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i]$$

$$-E[Z|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i']|$$

for fixed $i \in [m]$, fixed $b_1, \ldots, b_{i-1}, b_i, b_i'$.

Note that for any $j \in [n]$ and $j \neq b_i, b_i'$,

$$E[Z_j|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i] =$$

$$E[Z_j|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i'].$$

Thus,

$$D = |E[Z_{b_i} + Z_{b_i'}|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i]$$

$$-E[Z_{b_i} + Z_{b_i'}|B_1 = b_1, \ldots, B_{i-1} = b_{i-1}, B_i = b_i']|.$$

Clearly, we are interested in the case $b = b_i \neq b_i' = b'$. Let $I = \{b_1, \ldots, b_{i-1}\}$, $B = \{B_1 = b_1\} \wedge \ldots \wedge \{B_{i-1} = b_{i-1}\} \wedge \{B_i = b_i\}$ be a multiset, $B' = \{B_1 = b_1\} \wedge \ldots \wedge \{B_{i-1} = b_{i-1}\} \wedge \{B_i = b_i'\}$. Because all expectations are non-negative, $D \leq \max\{E[Z_b + Z_{b'}|B], E[Z_b + Z_{b'}|B]\}$. For symmetry, we only study $E[Z_b + Z_{b'}|B]$. Clearly, we have

$$E[Z_b|B] = \begin{cases} 0 & b \in I \\ (1-\frac{1}{n})^{m-i} & \text{otherwise} \end{cases}$$

Let $n(b')$ be the number of times b' appears in I, *e.g.*, the number of balls in b'th bin after the first i balls are randomly placed.

$$E[Z_{b'}|B] = \begin{cases} 0 & n(b') \geq 2 \\ (1-\frac{1}{n})^{m-i} & n(b') = 1 \\ \frac{m-i}{n}(1-\frac{1}{n})^{m-i-1} & \text{otherwise} \end{cases}$$

Hence $E[Z_b + Z_{b'}|B] \leq 2(1-\frac{1}{n})^{m-i} + \frac{m-i}{n}(1-\frac{1}{n})^{m-i-1}$. Because $E[Z_b + Z_{b'}|B']$ has the same bound, we have

$$\begin{aligned} D &\leq 2(1-\frac{1}{n})^{m-i} + \frac{m-i}{n}(1-\frac{1}{n})^{m-i-1} \\ &= (2 + \frac{m-i}{n-1})(1-\frac{1}{n})^{m-i} \\ &\leq \frac{2(m+n)}{n}(1-\frac{1}{n})^{m-i} \end{aligned} \tag{9}$$

As a result, for $i \in [m]$, we have

$$D = |E[Z|B] - E[Z|B']| \leq c_i,$$

where

$$c_i = \frac{2(m+n)}{n}(1-\frac{1}{n})^{m-i}.$$

$$\begin{aligned} \sum_{i=1}^m c_i^2 &= \frac{4(m+n)^2}{n^2}\sum_{i=1}^m[(1-\frac{1}{n})^{m-i}]^2 \\ &= \frac{4(m+n)^2}{n^2} \times \frac{n^2[1-(1-\frac{1}{n})^{2m}]}{2n-1} \\ &= \frac{4(m+n)^2(1-(1-\frac{1}{n})^{2m})}{2n-1} \end{aligned} \tag{10}$$

The theorem follows directly by applying Lemma 12. $\square$

THEOREM 14. *Let $n \geq (12)^4$. Randomly place m balls into n bins, where $1 < m < n\ln n/4$. With constant probability, there exists one* unit bin *afterwards.*

PROOF. Let X be random variable of the number of *unit bins*. Then

$$E[X] = \mu = m(1 - \frac{1}{n})^{m-1}, \text{ and } E[X^2] = m^2(1 - \frac{1}{n})^{m-1}.$$

First we consider the case that $m \leq n/2$. By the *second moment method*: $\Pr[X = 0] \leq \frac{E[X^2]}{(E[X])^2} - 1 = 1/(1 - \frac{1}{n})^{m-1} - 1$. Note that by Fact 1, $(1 - \frac{1}{n})^{m-1} \geq (1 - \frac{1}{n})^m \geq (1 - \frac{m}{n^2})e^{-\frac{m}{n}}$. Because $m \leq n^2/10$, $(1 - \frac{1}{n})^m \geq \frac{9}{10}e^{-1/2}$,

$$\Pr[X = 0] \leq \frac{10e^{1/2}}{9} - 1 < 1.$$

Second, we consider the case that $m = cn$ where $c \in (1/2, \ln n/4)$. From Theorem 13 and $n \geq 324(\ln n)^4$, we have

$$\begin{aligned}
\Pr[X = 0] &\leq 2\exp(-\frac{\mu^2(2n-1)}{8(n+m)^2(1-(1-\frac{1}{n})^{2m})}) \\
&\leq 2\exp(-\frac{\mu^2 n}{8(n+m)^2}) \\
&\leq 2\exp(-\frac{m^2(1-\frac{1}{n})^{2m-2}n}{8(n+m)^2}) \\
&\leq 2\exp(-\frac{(1-\frac{1}{n})^{2m}n}{8(1+\frac{n}{m})^2}).
\end{aligned}$$

Because $(1 - \frac{1}{n})^{2m} \geq (1 - \frac{2m}{n^2})e^{-2m/n} \geq \frac{1}{2\sqrt{n}}$ and $(1 + \frac{n}{m})^2 \leq 9$. Hence

$$\Pr[X = 0] \leq 2\exp(-\frac{\sqrt{n}}{144}) \leq 2/e.$$

This finishes the proof. □

3. LARGEST CONNECTED COMPONENT

In order to deploy network in multiple rounds, we have to assume the ability to detect "empty areas"[1] after one round. This assumption is sometimes too strong to be true in practice. As we discussed in the introduction, there exits a largest connected component if we place an appropriate number of wireless sensor nodes in the first round. We will use the complement of the area covered by this largest connected component, which is a super-set of the empty area, as an estimation of the empty area. In the second round of the deployment of wireless sensors, we will *not* place sensors into the region covered by the largest component formed by sensors in the first round. This approach clearly will "waste" sensors that are not in the largest component from the first round deployment. Fortunately, we can later show that such "waste" is negligible: the asymptotic number of sensors required for a "second chance" deployment in two rounds will not increase.

We assume the communication range for all wireless sensors is r. The target deployment region is a square of size $a \times a$. Let $n = (a/r)^2$. Carruthers and King [3] proved that there exists an unique large connected component which covers at least a constant portion of the square in expectation, if the number of random nodes deployed is $\Theta(n)$. Based on this result, we show that, by increasing the number of nodes slightly, the portion of the area covered by the largest connected component will be $1 - o(1)$. We first present some technical lemmas.

[1]Here a point x is called "uncovered" if it is not inside the transmission range of any deployed wireless nodes. The empty area after a random deployment of some wireless nodes is the collection of all uncovered points in the region.

LEMMA 15. *Let $n \geq 4(\ln n)^5$. Randomly place $2n \ln\ln n - n\ln 2$ balls into n bins. With probability at least $1 - \frac{1}{2n}$, the number of empty bins is at most $\frac{4n}{(\ln n)^2}$.*

PROOF. Let $m = 2n\ln\ln n - n\ln 2$ and Z be the number of empty bins. By Fact 1, $\mu = E[Z] = n(1 - \frac{1}{n})^m \in (\frac{n}{(\ln n)^2}, \frac{2n}{(\ln n)^2})$. From the Lemma 1, we have

$$\Pr[Z \geq 2\mu] \leq \Pr[|Z - \mu| \geq \mu] \leq 2\exp(-\frac{\mu^2(n-1/2)}{n^2 - \mu^2}).$$

Because $n \geq 4(\ln n)^5$ and $\mu \geq \frac{n}{(\ln n)^2}$, we have

$$\begin{aligned}
\Pr[Z \geq 2\mu] &\leq 2\exp(-\frac{n/2}{(\ln n)^4}) \\
&\leq 2\exp(-2\ln n) = \frac{2}{n^2} \leq \frac{1}{2n}.
\end{aligned}$$

□

In our "second chance" deployment, we first place $O(n \ln\ln n)$ wireless nodes randomly. Based on Carruthers and King's result, we have an unique large connected component. Localization algorithms [1,2,5,6,15] exist to find the location of wireless nodes. This location information can then be used to calculate the area covered by this connected component. Because this component is large, by a few probes, we can contact this component and retrieve its area. Recall that it was already proved in [3] that, which high probability, the largest connected component will touch the boundary of the deployment region. Hence this component can also be contacted by simply querying the sensor nodes along the boundary. After detecting the area covered by the largest connected component, we then deploy wireless sensor nodes in the area not covered by the largest component.

Let $r_0 = \frac{a}{\lfloor a/r \rfloor}$. Note that $r/2 \leq r_0 \leq r$. We divide the $a \times a$ square region into cells with size $r_0/3 \times r_0/3$. Let $n_0 = 9a^2/r_0^2$ be the number of cells. A cell is called *empty* if it does not contain any node. Two empty cells are connected by an edge if they share a side-segment. We denote a *path of empty cells* a "Manhattan" path in the sense that it consists of up, down, left and right turns. As observed in [3], if two wireless nodes are disconnected (there is no path connecting them in the wireless communication graph), there exists a path of empty cells that separate one from the other. This is a direct consequence of our setting of $r_0/3$ as the cell size[2]. The following lemma characterizes the length of paths of empty cells, which is directly implied by the Lemma 3 from [3]. The bounds could be improved in our setting, though this improvement will not give us better results.

LEMMA 16 (EMPTY PATH LENGTH LEMMA [3]). *Let G be a grid of size n, with $\lfloor en \rfloor$ random empty boxes, then the probability that there is a path of empty cells of length $l \geq \ln n$ is less than $1/n$.*

THEOREM 17 (LARGEST CONNECTED COMPONENT). *Assume the communication range for the wireless nodes is r. The target deployment region is a square of size $= a \times a$. If $n = a^2/r^2 \geq 36$, by randomly placing $36n\ln\ln(36n)$ wireless nodes in the square, the largest connected component covers area with size more than $(1 - \frac{2}{\ln(36n)})a^2$ with probability at least $1 - \frac{1}{n}$.*

[2]For any node u from one connected component and its closest v from another connected component, clearly we have $\|u - v\| > r$. Then we need 3 adjacent cells to reach them since $\frac{r}{\sqrt{2}r_0/3} > 2$. The middle cell must be empty one.

PROOF. Let $r' = \frac{a}{\lceil a/r \rceil} \in (1/2, 1)$. We divide the $a \times a$ square into cells with size $r'/3 \times r'/3$. We denote $n' = 9a^2/r'^2 \in (9n, 36n)$.

Because $36n \ln\ln(36n) \geq n' \ln\ln n'$. From Lemma 15 and Lemma 16, with probability at least $1 - \frac{2}{n'}$, the number of empty cells after the deployment is at most $\frac{n'}{(\ln n')^2}$ and there is no path of empty cells with length more than $\ln n'$.

Let us consider the connected components C_i formed by the wireless nodes, where $0 \leq i \leq k$ for some k. The non-existence of long path of empty cells implies the existence of an unique largest component, which we denote as C_0.

Because C_i for $1 \leq i \leq k$ is small, it is separated by a path of empty cells with length l_i at most $\ln n'$ (possibly together with the boundary). C_i spans a square with side length at most l_i. Clearly, $\sum_i l_i \leq \frac{n'}{(\ln n')^2}$. Let A_i be the number of cells that contain C_i. We have $A_i \leq l_i^2$. Then

$$\sum_i A_i \leq \sum_i l_i^2 \leq \ln n' \sum_i l_i \leq \frac{n'}{\ln n'}.$$

On the other hand, the number of empty cells is at most $\frac{n'}{(\ln n')^2} \leq \frac{n'}{\ln n'}$. Then the number of cells that are filled by the largest connected component is at least $(1 - \frac{2}{\ln n'})a^2 \geq (1 - \frac{2}{\ln 36n})a^2 \geq (1 - \frac{1}{\ln n})a^2$, with probability at least $1 - \frac{2}{n'} \geq 1 - \frac{1}{n}$. $\square$

By deploying $\Theta(n \ln\ln n)$ number nodes into the square region $a \times a$, Theorem 17 states that almost surely the region (except an infinite small portion) is already covered. Thus, the second round of deployment only has to spend a small number of wireless nodes to cover the remaining empty portion.

4. COVERAGE

In this section, we discuss the number of nodes required to cover a $a \times a$ square region in our "second chance" deployment model. We assume the area the wireless network is deployed to be a square of size $a \times a$. As before, we assume each wireless node has communication range r and $n = (a/r)^2$.

In traditional one round random deployment, wireless nodes are randomly placed in the square. It is shown that we need $\Theta(n \ln n)$ nodes [17,18] to provide a coverage with high probability. We first assume we can deploy nodes outside the region covered by wireless nodes deployed in previous rounds. Based on the technical lemma derived in Section 2, we obtain tight bounds for any "second chance" deployment strategy in k rounds. Then, we assume we only able to deploy nodes outside the largest connected component from the first round. We obtain tight bounds for any "second chance" deployment strategy for two rounds.

We assume that we can detect empty areas after previous round deployment and selectively place wireless nodes only in the empty area randomly in the next round. By applying the result in Section 2, we show that in a k round random deployment, $\Theta(n \ln^{(k)} n)$ nodes sufficiently cover the region *w.h.p.*.

THEOREM 18. *Assume the communication range for wireless nodes is r. The target deployment region is a square of size $a \times a$. Let $n = (a/r)^2$. Assume we can randomly deploy wireless nodes outside the covered region in previous rounds. With $\Theta(n \ln^{(k)} n)$ wireless nodes, we can cover the square in k rounds with probability at least $1 - \frac{1}{n}$.*

PROOF. We first show $O(n \ln^{(k)} n)$ is sufficient. Let $r' = \frac{a}{\lceil \sqrt{2}a/r \rceil} \in [\sqrt{2}r/4, \sqrt{2}r/2]$. We divide the square into cells with side length r' (see Figure 5(b) for partitioning). The number of cells is $n' = (a/r')^2 = O(n)$. By Theorem 7, $O(n' \ln^{(k)} n') = O(n \ln^{(k)} n)$ wireless nodes sufficiently fill all the cells with probability at least $1 - \frac{1}{n}$ in k rounds, which assures the coverage.

By setting the cells' side length to $r'' = \frac{2a}{\lfloor a/r \rfloor} \in [2r, 4r]$ and $n'' = (a/r'')^2 = \Theta(n)$, we need at least one node for each such cell to provide full coverage of the region. From Theorem 11, we need at least $\Omega(n'' \ln^{(k)} n'') = \Omega(n \ln^{(k)} n)$ wireless nodes so that all cells are filled with a positive constant probability. This finishes the proof. $\square$

The preceding theorem assumes that we can precisely determine which cells are not covered after previous rounds. However, as we mentioned, it could be hard to detect the empty area after one deployment. On the other hand, by Theorem 17, after one deployment with $O(n \ln\ln n)$ nodes, there is a unique large connected component. As mentioned in [3], the largest connected component will *w.h.p.* cover at least one point on the boundary of the deployment square region. This property ensures that we can easily probe this largest connected component by simply querying the sensors along the region boundary. If we assume we can randomly place additional wireless nodes outside this largest connected component, we then show that $\Theta(n \ln\ln n)$ suffices to cover the area in two rounds.

THEOREM 19. *Assume the communication range for wireless nodes is r. The target deployment region is a square of size $a \times a$. Let $n = (a/r)^2$. Assume we can randomly deploy wireless nodes outside the area covered by a large connected component in the first round. With $\Theta(n \ln\ln n)$ wireless nodes, we can cover the square region by a "second chance" random deployment with probability at least $1 - \frac{1}{n}$.*

PROOF. The lower bound of $\Omega(n \ln\ln n)$ comes from the last theorem. From Theorem 17, after deploy $36n \ln\ln(36n)$ wireless nodes, the largest connected component covers region with area at least $(1 - \frac{2}{\ln n})a^2$ with high probability. In other words, there are at most $\frac{2}{\ln n} n$ cells with side length $\frac{a}{\lceil a/r \rceil}$ are empty.

Applying localization technique, the area (hence the cells) covered by the largest connected component can be computed. After we compute the $n' = \frac{2(36n)}{\ln(36n)})$ cells not covered by the largest component, by Lemma 4, randomly placing another $2n' \ln n' \leq 148n$ wireless nodes in the area sufficiently covers the entire region with high probability. $\square$

We summarize the discussion in this section by following algorithm states our "second chance" deployment strategy. We show in next section that this algorithm actually achieves full connectivity as well.

Algorithm 1 Second Chance Deployment for Coverage and Connectivity

1: We first place $36n \ln\ln(36n)$ nodes randomly into the region.
2: The deployed nodes then apply some location techniques to find their locations.
3: We query nodes along the boundary of the deployment region and find the largest connected component. Based on this, we find the "empty" region that is not covered by nodes for this largest connected component.
4: We randomly deploy $148n$ nodes into the "empty" region.

The constants in the algorithm is quite large based on our theoretical result. However, in our simulation, we find small constants are sufficient to achieve the deployment goals with high probability.

5. CONNECTIVITY

In the previous section, we studied the asymptotic number of nodes needed to cover a region with high probability using a "second chance" random deployment (or generally k-RBP). It is easy to show that asymptotic upper bound on the number of nodes needed for achieving a connected network is the same with the coverage. In particular, if a set of nodes achieves coverage for a square region with communication range $r/2$, they form a connected network under communication range r. Hence $O(n \ln^{(k)} n)$ nodes is sufficient to assure connectivity by a "second chance" random deployment in k rounds and $O(n \ln\ln n)$ in two rounds. In this section, we study the asymptotic lower bound on the number of nodes needed for achieving a connected network with a "second chance" deployment strategy. As usual, we assume that the region is an $a \times a$ size square and the communication range is r for every wireless nodes. Let $n = (a/r)^2$.

Let $l = \frac{a}{\lceil \sqrt{5}a/r \rceil} \in (\sqrt{5}r/10, \sqrt{5}r/5]$. We divide the $a \times a$ square into grids with side length l. The number of grid cells is at most $10n$. For any two nodes from two adjacent cells, their Euclidean distance is at most r, hence they are connected in the communication graph. If we randomly deploy wireless nodes using a k-RBP strategy, such that each cell contains at least one node. Because nodes in adjacent cells are connected, the entire network $G = (V, E)$ (where two nodes are connected by an edge in E iff their Euclidean distance is at most r) is connected. Compared with the coverage case, the grid size is slight smaller, though the number of cells are both $\Theta(n)$. An upper bound on the number of nodes needed for producing a connected network when we randomly deploy wireless nodes using k-RBP strategy asymptotically is thus $O(n \ln^{(k)} n)$, which matches result for the coverage case. Thus, we have the following theorem.

THEOREM 20. *Assume the communication range for wireless nodes is 1. The target deployment region is a square of size $a \times a$. Let $n = a^2/r^2$. Assume we can randomly deploy wireless nodes outside the covered region of nodes in previous rounds. With $O(n \ln^{(k)} n)$ wireless nodes, we can produce a connected network in k rounds with probability at least $1 - \frac{1}{n}$.*

Again, if we are only able to deploy in region outside the largest connected component from the first round. We first deploy $36n \ln\ln n$ random wireless nodes in the square region, which will leave at most $72n/\ln(36n)$ empty grid cells with side length at most $r/3$. Since $r/3 \leq r/sqrt5$, if each such empty cell has at least one wireless node in the second round, the entire network is connected. Hence, another $148n$ wireless nodes in the second round is sufficient to achieve full connectivity. Hence the "second chance" deployment described in Algorithm 1 achieves a connected network with high probability.

Similar to the coverage case, our simulations suggest better performance for our "second chance" deployment than the theoretical upper bound.

The lower bound, however, requires an argument that is different with the coverage case. To prove the lower bound, we first exclude the case when we only deploy one node, which is a always connected network by itself.

THEOREM 21. *Assume the communication range for wireless nodes is r. The target deployment region is a square of size $a \times a$. Let $n = (a/r)^2$. Assume we can randomly deploy wireless nodes outside the region covered by the nodes in the first round. There exists a constant $c > 0$ such that with at most $cn \ln\ln n$ wireless nodes, the network is disconnected with a positive constant probability using any "second chance" deployment strategy in two rounds.*

PROOF. Let $r' = \frac{a}{\lfloor a/5r \rfloor}$. We first divide the square $a \times a$ into grid with cell side-length r. See Figure 5(c) for illustration. Note that $r' \geq 54$ and the number of cells is $n' = (a/r')^2 = \Theta(n)$. We deploy at most $(1 - \frac{4 \ln\ln n'}{\ln n})n' \ln\ln n' = \Theta(n \ln\ln n)$ nodes in the first round of deployment.

From Lemma 9 the number of empty cells after this round is at least $\frac{n'(\ln\ln n')^2}{\ln n'}$ with a probability at least $1 - \frac{2}{e}$.

Now we consider the second round of the deployment. Let $r'' = r'/5 \geq r$. We divide each cell further into smaller cells with side length r'', i.e., each original cell is further partitioned into 5 rows and 5 columns. See Figure 5(c) for an illustration. For each large empty cell, our second random deployment region will at least cover the centering $3r'' \times 3r''$ grid cells. We define each 9 empty centering cells as a "middle" cell, which is empty of nodes.

Define $n'' = \frac{n'(\ln\ln n')^2}{\ln n'}$ and $m = \frac{1}{4}n'' \ln n'' = \Theta(n(\ln\ln n)^2)$. Assume we deploy at most m nodes in the second round, it is clear that we deploy at most m nodes in the union of the "middle" cells. From Theorem 14, with a constant probability, there exists an *unit "middle" cell* after the second deployment, i.e., there is only one node placed into this "middle" cell. (See Figure 5(a) for illustration of unit bins.) On the other hand, for an *unit "middle" cell*, with probability $1/9$, the node will lie in the centering small-cell of the 9 small-cells. Consequently, this node is disconnected from the rest of the nodes, because the side length of the small-cells is at least r and all its adjacent 8 small-cells are empty. Consequently, the network is disconnected with a constant probability after a two-round random deployment if we only have $(1 - \frac{4 \ln\ln n'}{\ln n})n' \ln\ln n' = \Theta(n \ln\ln n)$ nodes overall. This finishes the proof. □

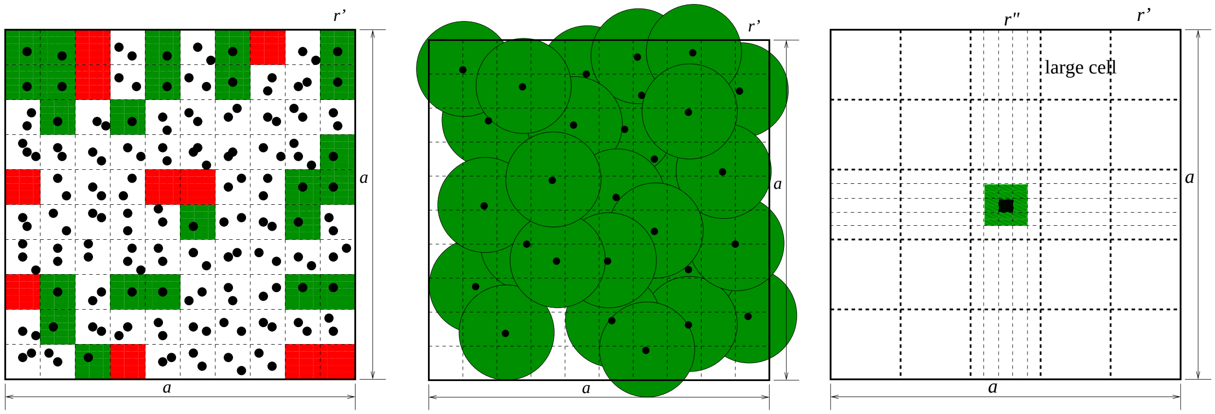
Notice that in the previous section and this section, we essentially show that, when the sensing range (or the transmission range respectively) of all nodes is r and the target deployment region is a square of size $a \times a$, having $\Theta(n \ln\ln n)$ nodes is sufficient and necessary condition to providing full coverage of the region (or a connected network respectively) with high probability using a second chance random deployment, where $n = (a/r)^2$. These results also can be used to derive the critical coverage range (or the transmission range respectively) when we have m wireless nodes ready for deploying using second chance deployment. Essentially, assuming $a = 1$, the critical range r is a solution of the following equation

$$m = \frac{1}{r^2} \ln\ln \frac{1}{r^2}.$$

Thus, $\ln\ln m = \ln\ln \frac{1}{r^2} + \ln\ln\ln\ln \frac{1}{r^2}$. This implies that $r \simeq \sqrt{\frac{\ln\ln m}{m}}$.

THEOREM 22. *Assume that we randomly deploy n nodes using a second chance deployment strategy in a unit square. The critical sensing range for providing full coverage and the critical transmission range for achieving a connected network is $\Theta(\sqrt{\frac{\ln\ln m}{m}})$.*

Recall that, Gupta and Kumar [8] proved that the critical range for achieving connectivity using one-round random deployment of m nodes is $\Theta(\sqrt{\frac{\ln m}{m}})$. Our bound is asymptotically smaller than this bound. Observe that, the optimal deterministic deployment of m nodes that provides the full coverage requires the sensing range to be of order $\Theta(\sqrt{\frac{1}{m}})$. Our second-chance deployment strategy



(a) empty and unit cells (b) coverage (c) lower bound for connectivity

Figure 1: (a) empty cells (colored red) and unit cells (colored green) after random deployment. (b) Randomly placing sensors for achieving full coverage. Here a sensor with transmission range r in a $\frac{\sqrt{2}r}{2} \times \frac{\sqrt{2}r}{2}$ cell can cover the entire cell. (c) Lower bound on the number of randomly placed sensors needed for achieving a connected network with at least a constant probability. Here the large cells have side-length $r' \geq 5r$ and the smaller cells have side-length $r" \geq r$. The shaded region is a "middle" cell formed by 9 smaller cells.

moves the critical range needed for coverage and connectivity significantly closer to the deterministic optimal solution using only a small cost.

6. SIMULATIONS

In this section, we conduct extensive simulations to illustrate the effectiveness of our method. In our simulations, the target region is a square of size 1×1. We varies the communication range for the wireless nodes deployed to study the asymptotic behavior when the relative field size $N = (1/r)^2$ increases. Each sensor is connected to the neighbors within its communication range.

6.1 Largest Connected Component

Our result assumes the existence of a large connected component after the first round deployment. Let the communication range for the wireless nodes is r. In our simulation, we set r to be in $\{0.1, 0.05, 0.0333, 0.025, 0.02\}$.

In Theorem 17, we prove that by placing $36N \ln \ln(36N)$ random nodes in one round the largest connected component covers most area of the square region. Therefore, in the simulation, we randomly place $cN \ln \ln N$ nodes, where c is a constant in $\{0.5, 0.8, 1.0, 1.2\}$ in our setting. We vary the constant c to examine the impact of the node density on the largest component.

For each pair of values r and c, we run the simulation for 50 rounds. When $c = 0.5$, the largest connected component accounts 20% of the network for $r = 0.1$ and 6% for $r = 0.02$. Notice that in this case, although the size of the largest connected component increases when N increase, the larger component takes a smaller portion of the network.

On the other hand, for all other larger values of c we tested, the largest connected component takes a larger portion of the network when N increases. In particular, when $c = 1$, the largest connected component accounts 85% of the network for $r = 0.1$ and 99% for $r = 0.025$. See Figure 3 for a largest connected component instance when $c = 1$ and $r = 0.025$. The growth of the portion taken by the largest connected component matches the prediction of Theorem 17. However, the case of $c = 0.5$ indicates that this trend does not hold for small c values.

6.2 Coverage

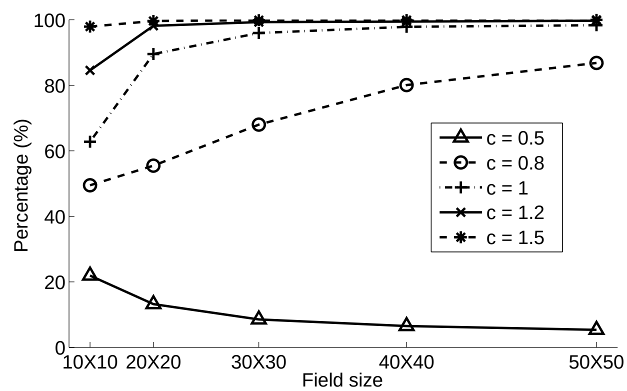


Figure 2: The portion of the largest connected component in the entire networks.

We test the number of wireless nodes needed to achieve 99%+ coverage over a square region. Again, we set the square region to be of size 1×1 and vary the communication range of the wireless nodes. Instead of directly computing the area of the union of the sensing disks for the wireless nodes, we simply place a dense grid in the square region and count the portion of the grid vertices covered by the deployed wireless nodes. This treatment saves us from boundary issues in computing the size of covered area.

In simulating the traditional one time deployment, we randomly place node one by one until 99% of the region is covered. The number of nodes needed to achieve coverage increases nearly linearly with the field size $N = (1/r)^2$ in Figure 4. The growth approximately matches the theoretical bound $\Theta(N \ln N)$.

In our "second chance" deployment, we deploy $cN \ln \ln N$ nodes in the first round where c is a chosen constant. We then randomly deploy wireless nodes one by one outside the region covered by the largest connected component in the first round. In our simulation, we simply reject each new random node if its sensing disk is inside

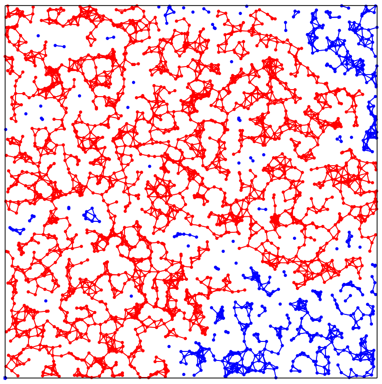


Figure 3: **The largest connected component for case $c = 1$ and $r = 0.025$. The blue nodes are the nodes outside the largest component.**

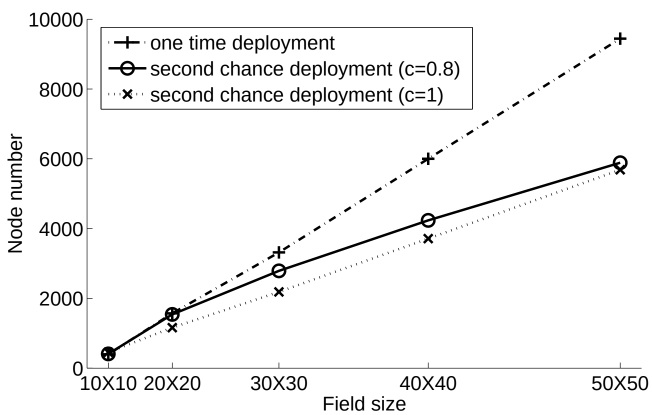


Figure 4: The number of nodes to achieve 99% coverage of the region.

the sensing regions formed by the nodes in the largest connected component from the first round.

In Algorithm 1, we deploy $36N \ln\ln(36N)$ wireless nodes in the first round and place another $148N$ wireless nodes in the second round. In our simulation, we find that this theoretical bound is too pessimistic in practice. In Figure 4, we show the number of nodes required to achieve 99% coverage in one time random deployment and our "second chance" deployment with $c = 0.8$ and 1. In particular, we find that placing $N \ln\ln N$ nodes in the first round actually costs the smallest number nodes to achieve the coverage of the region, i.e., c should be set to 1.

On average, our "second chance" deployment strategy saves about 40% on the number of nodes. The gap between the two method is expected to widen due to their different theoretical growth speeds. Figure 5 gives an instance of the node distribution in our "second chance" deployment. Red nodes denote the largest connected component in the first round. Blue nodes are the nodes in the first round that are not in the largest component, and black nodes are deployed in the second round.

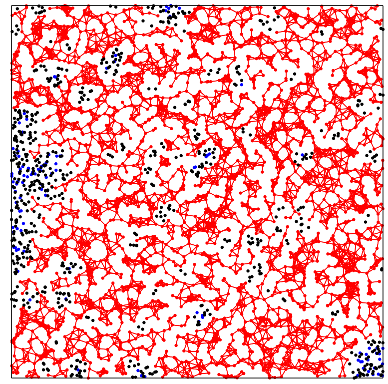


Figure 5: An example for the "second chance" deployment for coverage with $c = 1$ and field size $N = 40 \times 40$. Red nodes are the largest connected component in the first round. Blue nodes are the nodes outside the largest connected component in the first round, and black nodes are deployed in the second round.

6.3 Connectivity

We also test the number of nodes needed to achieve full connectivity of the entire network while preserving more than 90% coverage of the field. The simulation setup is the same with previous sections.

In one time deployment, we randomly place node one by one until 90% of the region is covered and the network is connected. From Figure 6, the number of nodes required to achieve connectivity is roughly the same with the coverage case in one time deployment, which grows nearly linearly.

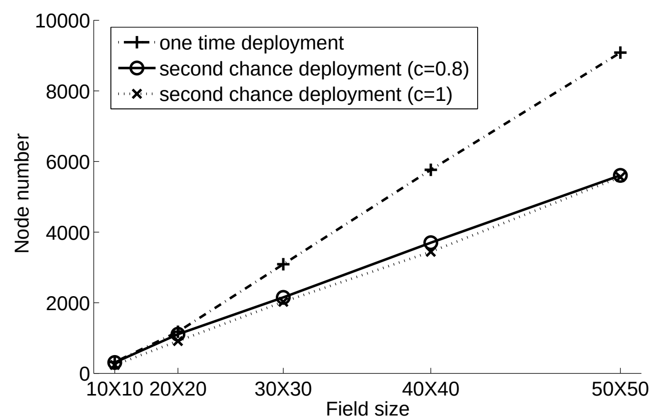


Figure 6: The number of nodes to achieve full connectivity and 90% coverage of the region.

In our "second chance" deployment, again, we deploy $cN \ln\ln N$ nodes in the first round. We test the cases that $c = 0.8$ and 1 as in the coverage case. Our choice of the second round deployment is the same with the coverage case. Though the condition for termination is to achieve full connectivity and 90% coverage instead. The number of nodes to achieve the deployment goal of connectivity and the savings on the number of nodes over the one time deployment is roughly the same with the coverage case, as indicated in Figure 6.

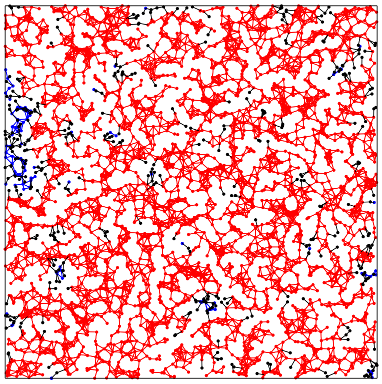


Figure 7: An example for the "second chance" deployment for connectivity with $c = 1$ and field size $N = 40 \times 40$. Red nodes are the largest connected component in the first round. Blue nodes are the nodes outside the largest connected component in the first round, and black nodes are deployed in the second round.

Figure 7 gives an instance of the node distribution in our "second chance" deployment to achieve connectivity. Red nodes denote the largest connected component in the first round. Blue nodes are the nodes in the first round that are not in the largest component, and black nodes are deployed in the second round to achieve full connectivity.

7. CONCLUSIONS

In this paper, we studied the number of nodes needed to provide full coverage and the number of nodes needed to provide connectivity, when we have a second chance (or multiple chances) in random deployment. Under some deployment assumption, we showed that the number of nodes needed is $\Theta(n \ln \ln n)$ for both cases in 2 rounds where $n = (a/r)^2$, which is much smaller than the $\Theta(n \ln n)$ bound needed when we are only able to deploy nodes in one-shot. Our results also imply the critical sensing range for providing coverage and the critical transmission range for providing connectivity when we randomly deploy m nodes using our second-chance deployment is asymptotically $\Theta(\sqrt{\frac{\ln \ln m}{m}})$, which is also much smaller than the critical range using one-round deployment [8]. We further showed that our second-chance deployment is practical by showing that there is a giant largest connected component, and that it touches one of the deployment boundary with high probability. Therefore, it is possible to acquire deployment information in the first round by querying this large component.

This paper is the first step towards a more complete study of performance limits of multi-hop wireless networks produced by random deployment. In particular, this result can be viewed as a trade-off between the deployment complexity and the resulting network quality. There are a number of interesting and challenging questions left for future research. First, we would like to close the constant gaps between the lower bounds and upper bounds. We conjecture that the actual coefficients of all formulas are either 1 or sufficiently close to 1. Second, we would like to study the exact critical range for coverage and the critical range for connectivity when m nodes are randomly placed using a second-chance deployment strategy. We conjecture that the critical range to be exactly $\sqrt{\frac{\ln \ln m}{m}}$ for connectivity. Third, we would like to study the network capacity of random networks produced a "second chance" deployment. Recall that, for random networks of m nodes produced by one-round random deployment, the pioneering work by Gupta and Kumar [7] proved that the per unicast flow capacity (when there are m random flows) is $\Theta(\sqrt{\frac{1}{m \ln m}})$. We conjecture that, the per-unicast flow capacity for random networks produced by a "second chance" deployment will be $\Theta(\sqrt{\frac{1}{m \ln \ln m}})$.

8. REFERENCES

[1] BASU, A., GAO, J., MITCHELL, J. S. B., AND SABHNANI, G. Distributed localization using noisy distance and angle information. In *MobiHoc '06: Proceedings of the seventh ACM international symposium on Mobile ad hoc networking and computing* (New York, NY, USA, 2006), ACM Press, pp. 262–273.

[2] BISWAS, P., AND YE, Y. Semidefinite programming for ad hoc wireless sensor network localization. In *In Feng Zhao and Leonidas Guibas, editors, Proceedings of Third International Workshop on Information Processing in Sensor Networks* (Berkeley, CA, April 2004).

[3] CARRUTHERS, S., AND KING, V. Connectivity of wireless sensor networks with constant density. In *ADHOC-NOW, LNCS 3158* (2004), pp. 149–157.

[4] DUBHASHI, D. Simple proofs of occupancy tail bounds. *Random Structure Algorithms 11*, 2 (1997), 119–123.

[5] EREN, T., GOLDENBERG, D., WHITLEY, W., YANG, Y., MORSE, S., ANDERSON, B., AND BELHUMEUR, P. Rigidity, computation, and randomization of network localization. In *Proceedings of IEEE INFOCOM '04* (Hong Kong, China, Apr. 2004).

[6] GOLDENBERG, D. K., BIHLER, P., YANG, Y. R., CAO, M., FANG, J., MORSE, A. S., AND ANDERSON, B. D. O. Localization in sparse networks using sweeps. In *MobiCom '06: Proceedings of the 12th annual international conference on Mobile computing and networking* (New York, NY, USA, 2006), ACM Press, pp. 110–121.

[7] GUPTA, P., AND KUMAR, P. Capacity of wireless networks. *IEEE Transactions on Information Theory IT-46* (1999), 388–404.

[8] GUPTA, P., AND KUMAR, P. R. Critical power for asymptotic connectivity in wireless networks. *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming, W. M. McEneaney, G. Yin, and Q. Zhang (Eds.)* (1998).

[9] KAMATH, A., MOTWANI, R., PALEM, K., AND SPIRAKIS, P. Tail bounds for occupancy and the satisfiability threshold conjecture. In *IEEE Foundation of Computer Science (FOCS)* (1994).

[10] KYASANUR, P., AND VAIDYA, N. H. Capacity of multi-channel wireless networks: impact of number of channels and interfaces. In *MobiCom '05: Proceedings of*

the 11th annual international conference on Mobile computing and networking (New York, NY, USA, 2005), ACM Press, pp. 43–57.

[11] Li, X.-Y., Tang, S.-J., and Ophir, F. Multicast capacity for large scale wireless ad hoc networks. In *ACM Mobicom* (2007).

[12] McDiarmid, C. On the method of bounded differences. *Surveys in Combinatorics 141* (1989), 148–188.

[13] Mitzenmacher, M. The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distrib. Syst. 12*, 10 (2001), 1094–1104.

[14] Shakkottai, S., Liu, X., and Srikant, R. The multicast capacity of ad hoc networks. In *Proc. ACM Mobihoc* (2007).

[15] So, A. M.-C., and Ye, Y. Theory of semidefinite programming for sensor network localization. In *In Proc. 16th ACM-SIAM Symposium on Discrete Algorithms* (2005), pp. 405–414.

[16] Wan, P.-J., and Yi, C.-W. Asymptotic critical transmission radius and critical neighbor number for k-connectivity in wireless ad hoc networks. In *MobiHoc '04: Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing* (New York, NY, USA, 2004), ACM Press, pp. 1–8.

[17] Wan, P.-J., and Yi, C.-W. Coverage by randomly deployed wireless sensor networks. *IEEE Transactions on Information Theory*, 52 (2006).

[18] Wan, P.-J., and Yi, C.-W. Coverage by randomly deployed wireless sensor networks. *IEEE/ACM Transactions on Networking 14* (2006), 2658–2669.

[19] Wan, P.-J., and Yi, C.-W. Coverage by randomly deployed wireless sensor networks. *IEEE/ACM Trans. Netw. 14*, SI (2006), 2658–2669.

[20] Xue, F., and Kumar, P. R. On the θ-coverage and connectivity of large random networks. *IEEE/ACM Trans. Netw. 14*, SI (2006), 2289–2299.