On the distance to the giant component along a straight line in a two-dimensional percolation model.

Olivier Dousse
School of Computer and
Communication Sciences
EPFL
CH-1015 Lausanne, Switzerland
olivier.dousse@epfl.ch

Christina Tavoularis
School of Electrical and
Computer Engineering
Cornell University
Ithaca, NY, USA
crt23@ece.cornell.edu

Patrick Thiran
School of Computer and
Communication Sciences
EPFL
CH-1015 Lausanne, Switzerland
patrick.thiran@epfl.ch

Abstract— **The supercritical regime of a percolation model refers to the range of probabilities (discrete) or densities (continuous) above a critical value for which there exists a unique unbounded cluster almost surely. In this paper, we provide an upper bound to the linear distance from the origin to this giant connected component for both the discrete and the continuous (Boolean) model in two-dimensions. By modeling a dense wireless sensor network with a supercritical Boolean model, our result bounds the distance traveled by a target moving in a straight line before it is detected by a node who can relay the alert through a multihop path to the sink. This result incorporates a solidified definition of detection requiring that the intrusion alert successfully reach the central authority.**

I. INTRODUCTION

Consider a wireless multihop sensor network whose task is to detect the intrusion of a moving object in the area in which the sensors are deployed. The first question that comes to mind is to know how far the object can move in the monitored area before being detected by a sensor. When the object moves along a straight line, stochastic geometry provides the answer under the name of *linear contact (or hit) distribution function* (see e.g. [1], p. 80). For example, when the sensor spatial distribution is Poisson, this function decreases exponentially with the distance traveled by the object.

The sensor that has detected the intruder needs however to convey this information to a monitoring station that collects the sensed data, the *sink*. Nodes between the detecting sensor and the sink act as relays for the message. However, various sources of noise, battery failures in the nodes or simply their random locations, will inevitably result in having some sensors disconnected from the network. If the sensor that detected the intruder is not connected to the sink, the intruder can continue to progress in the monitored area without actually being detected by the central sink. Consequently, a second, more difficult question is to know the distance that can be traveled by the intruder until the sink (and not only any sensor) can be notified. This paper addresses precisely this question.

We assume that the sensor network can be described by a 2-dimensional Poisson Boolean model: nodes (sensors) are distributed according to a 2-dimensional homogeneous Poisson process of intensity λ, their sensing range is identical to their transmission range, and is randomly distributed, independently from all other nodes. In such a model, it is well known (see. e.g. [2]) that given the distribution of the ranges, there is a critical intensity $\lambda_c > 0$ such that all clusters of connected nodes are almost surely finite if $\lambda < \lambda_c$ (subcritical phase) and a single giant, unbounded cluster of connected nodes appears almost surely if $\lambda > \lambda_c$ (supercritical phase). We assume in this paper that $\lambda > \lambda_c$, and that the sink belongs to the giant cluster.

Suppose that the intruder starts at the origin and moves along a straight line in any arbitrary direction. How far can it move until it hits the giant cluster? After having briefly reviewed the state of the art in Section II, we answer this question first in a discrete bond percolation setting in Section III. Using this construction, we then answer the question in the Boolean continuous model in Section IV. Section V concludes the paper.

II. RELATED WORK

Tracking and detecting a moving object is an important application of a sensor network, and has thus received some attention. Most of the work is so far devoted to the problem of computing the linear contact (or hit) distribution function, i.e. the distance traveled by the intruder until detection by a sensor without checking that the sensor that spawns the alarm is actually connected to the network. The simplest case is to compute this distance when the network is modeled by a Boolean model; it is known to be exponentially distributed (see e.g. [1], p.80). More results from this approach are derived in [3]. When the motion of the object does not follow a straight line, but a Brownian motion, explicit formulas and/or bounds are obtained in [4]. To save energy, nodes periodically switch

The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

off their batteries, which implies tradeoffs between the speed of detection of an intruder and the energy savings incurred by rendering many nodes into sleep mode. This tradeoff is examined in [5].

The same detection problem has been addressed in [6], but here the propagation speed of alarm messages is studied when nodes alternate between sleeping and active cycles. To the best of our knowledge, this paper is the first characterization of the distance traveled by an intruder before the alarm reaches the sink, and not only the first sensor in the network.

III. THEOREM FORMULATION AND INSIGHT OF THE PROOF

We consider the set of vertices $\mathbb{Z}^2$, and denote by $\mathbb{L}^2$ the set of the edges joining all adjacent vertices (i.e. $\mathbb{L}^2 = \{\langle x, y\rangle \in \mathbb{Z}^4 : |x_1 - y_1| + |x_2 - y_2| = 1\}$). We construct the standard independent bond percolation model by declaring each edge *open* with probability p, independently of all the other edges, and *closed* with probability $1 - p$. We say that two vertices are *connected* if there exists a sequence of contiguous open edges that joins them. This definition yields the partitioning of $\mathbb{Z}^2$ into *connected components* (or *clusters*).

We know from percolation theory that if p is greater that the *percolation threshold* $p_c = 1/2$, then the probability $\theta(p)$ that the origin belongs to an infinite cluster is strictly positive. Furthermore, in this case, the infinite cluster (also called *giant cluster*) is unique a.s..

Our main result states that the distance between the origin and the infinite cluster is upper bounded by a geometric random variable. We denote by N the coordinate of the first vertex on the right part of the horizontal axis that belongs to the giant cluster. Thus, if the origin belongs to the giant cluster, we have $N = 0$.

Theorem 1: When $p > p_c$, there exist constants c_1 and $c_2 \in \mathbb{R}^+$ such that

$$P_p(N > n) \leq c_1 \exp(-c_2 n).$$

A. Construction and its independence properties

To prove Theorem 1, we will perform the following construction: we start at the origin, and look at the set of points of the half-plane $\mathbb{H} = \mathbb{Z}^+ \times \mathbb{Z} = \{x = (x_1, x_2) \in \mathbb{Z}^2 \mid x_1 \geq 0\}$ (all the points located on the right hand side of the vertical axis) that are connected to it. If this set is infinite, then $N = 0$. If not, we look at the size of this connected component. More precisely, if we denote by $\partial B^+(n)$ the set of vertices $x = (x_1, x_2)$ of $\mathbb{H}$ such that $x_1 = n$ or $|x_2| = n$, and by $\{0 \leftrightarrow S\}$ the event that the origin is connected to some element of the set S, we look at the events $\{0 \leftrightarrow \partial B^+(n)\}$, $n \in \mathbb{N}$, and find the first integer n such that this event is not true. Note that this can be done iteratively, starting with $n = 1$ and increasing its value one by one. Therefore, if we define the variable $M = \min\{n > 0 : 0 \not\leftrightarrow \partial B^+(n)\}$, the event $\{M = n\}$ only depends on the state of the edges inside the box $B^+(n) = [0, n] \times [-n, n]$.

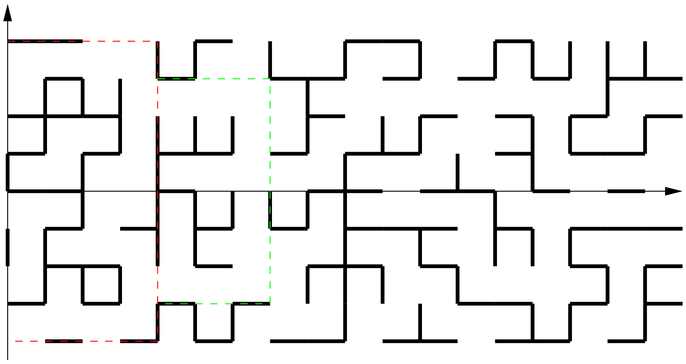

Fig. 1. The construction of the proof. Starting from the origin, the first cluster we meet is finite. We have that $0 \not\leftrightarrow \partial B^+(4)$. Thus, $M = 4$ in this example. Then we move to $(4, 0)$ and look only at the edges on the right, that are independent of M. The cluster at $(4, 0)$ is also finite, so we move to $(7, 0)$. The cluster at $(7, 0)$ is infinite, and therefore $(7, 0)$ belongs to the giant component. We conclude that $N' = 7$, and thus that $N \leq 7$.

Once we have found the value of M, we iterate the same procedure, starting at the point $(M, 0)$, and looking at the connected vertices in the half-plane $\mathbb{H} + (M, 0) = \{x = (x_1, x_2) \in \mathbb{Z}^2 \mid x_1 \geq M\}$. Clearly, the outcome of this second step is independent of what happened in the first step, because we only look at the half-plane located to the right of the box $B^+(n)$.

Then we iterate this algorithm until we hit an infinite connected component in the half-plane. The coordinate of the stopping point of the algorithm (denoted hereafter by N') gives an upper bound of N.

However, we have to check that this algorithm stops after a finite number of steps a.s. whenever an infinite cluster exists in the whole plane. In fact, it is enough to show that the probability $\theta^+(p)$ that the origin belongs to an infinite cluster in $\mathbb{H}$ is strictly positive (in this case, the number of steps before stopping is a geometric random variable). This result follows directly from Theorem 7.2 in [7]. Let $B(k) = [-k, k]^d$ be the box with side length $2k$ centered at the origin.

Theorem 2 (7.2 in [7]): For $d \geq 2$, let F be an infinite connected subset of $\mathbb{Z}^d$ with percolation threshold $p_c(F) < 1$. For each $\eta > 0$ there exists an integer k such that $p_c(2kF + B(k)) \leq p_c + \eta$.

By choosing $F = \mathbb{H}$ we have $p_c(2kF + B(k)) \to p_c$ as $k \to \infty$ and the theorem implies that this is true when $p_c(F) = p_c$ (this result is also given in [7], page 162). In other words, the percolation threshold is the same in the right half plane $\mathbb{H}$ as in the entire plane $\mathbb{Z}^2$.

B. The size of finite clusters

According to our algorithm, the random variable N' is the sum of a random number (the number of steps) of i.i.d variables (identical to the variable M above). In this section, we derive a bound on the probability that M is larger than a given number n. Note that this amounts to bounding the radius of a finite cluster in the half-plane.

Let C^+ be the open cluster at the origin on the half-plane $\mathbb{H}$ and consider the event of an open path from the origin to the surface $\partial B^+(n)$ when C^+ is finite. We claim that there exist finite constants $k_1, k_2 > 0$ for $p_c < p < 1$ such that

$$P_p(0 \leftrightarrow \partial B^+(n), |C^+| < \infty) \leq k_1 n^2 e^{-k_2 n} \text{ for all } n.$$

This result is a fairly straightforward extension of Theorems 8.18 and 8.21 in [7]. For the open cluster C at the origin of $\mathbb{Z}^2$ and the box $B(n) = [-n, n]^d$ with surface $\partial B(n)$, the theorems state that there exist finite constants $A(p, d), \sigma(p) > 0$ for $d \geq 2$ and $p_c < p < 1$ such that

$$P_p(0 \leftrightarrow \partial B(n), |C| < \infty) \leq A(p, d) n^2 e^{-\sigma(p) n} \text{ for all } n.$$

In our case, the cluster C^+ in $\mathbb{H}$ results from the restriction of the cluster C in $\mathbb{Z}^2$ to its vertices and edges lying in $\mathbb{H}$.

We proceed by modifying the construction proofs for Theorem 8.18 and Lemma 8.27 of [7], so that the result applies to the the cluster C^+ when $|C^+| < \infty$. Rather than repeat the entire proof here, we refer the reader to [7] and highlight the necessary modifications. Firstly, we replace C in the proof with the cluster C^+ where $x_1 \geq 0$ for every $(x_1, x_2) \in C^+$. The cluster has minimum and maximum extremities in the horizontal coordinate direction $L_1 = \min\{x_1 : (x_1, x_2) \in C^+\}$ and $R_1 = \max\{x_1 : (x_1, x_2) \in C^+\}$ (similarly L_2 and R_2 for the vertical coordinate) and diameter $\text{diam}(C^+) = \max\{R_i - L_i : 1 \leq i \leq 2\}$. We follow the same construction of two finite clusters whose widths in the horizontal coordinate direction are m and n such that the clusters only lie in the half-plane, one to the right of the other. Following the proofs of Lemma 8.27 and Theorem 8.18, we obtain the same upper and lower bounds for $\text{diam}(C^+)$ as for $\text{diam}(C)$. In particular, we find the bound

$$P_p(\text{diam}(C^+) = k) \leq \frac{d^2}{p^2(1-p)^{2d-2}} (2k+1)^d e^{-(k+2)\sigma(p)}.$$

Since the diameter of C^+ is at least as large as its radius, for the half box $B^+(n) = [0, n] \times [-n, n]$ we have

$$\begin{aligned} & P_p(0 \leftrightarrow \partial B^+(n), |C^+| < \infty) \\ & \quad \leq \quad P_p(n \leq \text{diam}(C^+) < \infty) \\ & \quad \leq \quad \frac{d^2}{p^2(1-p)^{2d-2}} \sum_{m=n}^{\infty} (2m+1)^d e^{-(m+2)\sigma(p)}. \end{aligned}$$

This equation gives us the constants k_1 and $k_2 = \sigma(p)$.

To show that $k_2 > 0$ when $p > p_c$, we must consider Theorem 8.21 in [7] (which, for the case $d = 2$, is implied by Theorem 11.24), and show that it is also true for the finite cluster on the half-plane. We refer the reader to the proof of Theorem 11.24 which shows that $0 < \xi(p) = \xi(1-p) < \infty$ when $p > p_c$, where $\xi(p) = \sigma(p)^{-1}$ is known as the correlation length. We adapt the proof by considering the whole plane $\mathbb{Z}^2$ but treating edges on the left half of the plane as closed with probability 1 to imply results on $\mathbb{H}$. Consider the vertex $e_n = (n, 0)$ and let $\tau_p^{f+}(0, e_n)$ denote the probability that the origin lies in a finite open cluster containing the vertex e_n in $\mathbb{H}$. We need to show an upper bound to this probability. We slightly modify the definition of the event A_n (page 298): denoting by $Y^+ = \{(-\frac{1}{2}, k + \frac{1}{2}) : k \geq 0\}$ and by $Y^- = \{(-\frac{1}{2}, -k - \frac{1}{2}) : k \geq 0\}$ the two vertical half-axes, we define the events A_n (respectively B_n) that some vertex in Y^+ (respectively Y^-) is joined by a closed path in the dual lattice to some vertex in X_d^+. Using the BK inequality, we have that

$$\tau_p^{f+}(0, e_n) \leq P_p(A_n) P_p(B_n) = P_p(A_n)^2,$$

because $P_p(A_n) = P_p(B_n)$. Now,

$$\begin{aligned} P_p(A_n) \quad & \leq \quad \sum_{k=0}^{\infty} \sum_{l=n}^{\infty} P_p((-\frac{1}{2}, k + \frac{1}{2}) \text{ joined to } (l + \frac{1}{2}, \frac{1}{2}) \\ & \qquad\qquad \text{by a closed dual path}) \\ & \leq \quad \sum_{k=0}^{\infty} \sum_{l=n}^{\infty} P_{1-p}\left((0, k) \leftrightarrow (l, 0)\right). \end{aligned}$$

However, $1 - p < p_c$ since $p > p_c$, and so

$$P_{1-p}\left((0, k) \leftrightarrow (l, 0)\right) \approx e^{-(l+k)/\xi(1-p)}$$

as $k + l \to \infty$, from which we deduce, as in [7],

$$\sigma(p) = \liminf_{n \to \infty} \left\{ -\frac{1}{n} \log \tau_p^{f+}(0, e_n) \right\} \geq \frac{2}{\xi(1-p)}.$$

C. Final computation

In the two previous subsections, we found that there exist $k_1, k_2 > 0$ such that

$$P_p(M > n) \leq k_1 n^2 e^{-k_2 n} \text{ for all } n. \tag{1}$$

As $k_2 > 0$, we can choose a new constant $e^{-k_2} < p_1 < 1$, and find a constant $k_3 \in \mathbb{N}$ such that

$$k_1 n^2 e^{-k_2 n} \leq (1 - p_1) p_1^{n - k_3}. \tag{2}$$

Let us define a sequence of i.i.d geometric random variables $\{X_i\}_{i \in \mathbb{N}}$ characterized by $P(X_i \geq n) = p_1^n$. Combining (1) and (2), we have that

$$P(M > n) \leq P(X_i + k_3 > n).$$

Moreover, the variable N' (defined in Section III-A) is the sum of K independent variables M_i, identically distributed as M (note that $M_1 = M$), where K is the number of steps in the algorithm:

$$N' = \sum_{i=1}^{K} M_i.$$

As the algorithm stops when we hit an infinite cluster in $\mathbb{H}$, and this happens with independent probability at each step (denoted by $\theta^+(p)$), we have $P(K = n) = \theta^+(p)(1 - \theta^+(p))^n$. Thus, K is also a geometric random variable with parameter

$p_2 := 1 - \theta^+(p)$. We can write

$$\begin{aligned} P(N' > n) &= P(\sum_{i=1}^{K} M_i > n) \\ &\le P(\sum_{i=1}^{K} (X_i + k_3) > n) \\ &= P(\sum_{i=1}^{K} X_i + k_3 K > n). \end{aligned} \tag{3}$$

The rest of this section is devoted to the computation of the above probability, and is purely technical.

We define the random variable $N'' = \sum_{i=1}^{K} X_i$. Its generating function is $G_{N''}(z) = G_K(G_X(z))$, where G_K and G_X denote the generating function of K and X_i respectively. We have thus

$$\begin{aligned} G_{N''}(z) &= \frac{p_2}{1 - (1-p_2)\frac{p_1}{1-(1-p_1)z}} \\ &= p_2 + (1-p_2)\frac{p_3}{1-(1-p_3)z}, \end{aligned}$$

where $p_3 = p_1 p_2/(1 - p_1 + p_1 p_2)$. We see that N'' is a geometric variable of parameter p_3, with some extra weight on $\{N'' = 0\}$. To obtain the probability in (3), we must consider the sum of N'' and $k_3 K$. The tail of the distribution of this sum decreases exponentially, as fast as the slowest term. Here, depending on the value of p_1, p_2 and k_3, one can dominate the other. Anyway, there exists $k_4 \in \mathbb{N}$ such that

$$P(N'' + k_3 K > n) \le p_4^{n-k_4},$$

with $p_4 = \max\{p_3, p_2^{1/k_3}\} < 1$. Finally as $P(N > n) \le P(N' > n) \le P(N'' + k_3 K > n)$, we have

$$P(N > n) \le p_4^{n-k_4} = c_1 \exp(-c_2 n),$$

for $c_1 = p_4^{-k_4}$ and $c_2 = -\log(p_4)$.

IV. EXTENSION TO THE CONTINUOUS CASE

In this section, we address a different percolation model – the Poisson Boolean model– defined as follows: we consider a homogeneous Poisson point process of intensity λ over the plane $\mathbb{R}^2$. On each point of the process we center a ball, whose radius ρ is also random, independent of and identically distributed as the other balls' radii. This divides the plane into an *occupied region*, and a *vacant region*, which is not covered by any ball.

Given the distribution of ρ such that $E[\rho^2] < \infty$, there exists a critical density λ_c such that if the actual density λ of the point process is greater than λ_c, then the probability that the origin belongs to an unbounded occupied connected component is strictly positive. Furthermore, above the threshold, the unbounded occupied connected component is unique a.s..

In the following theorem, we show that the distance between the origin and the unbounded component along a given direction can be bounded from above by an exponential random variable.

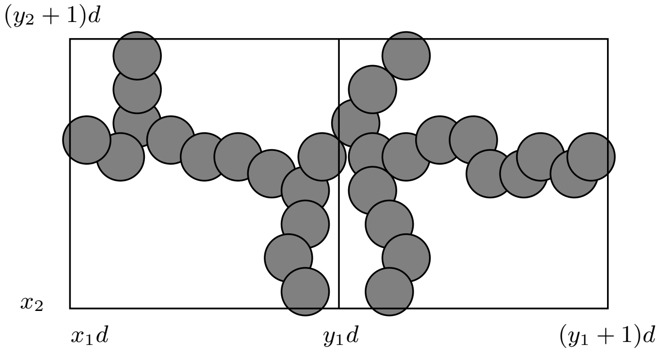

Fig. 2. A horizontal edge $\langle x, y\rangle$ that fulfills the three conditions.

Theorem 3: Let D denote the coordinate of the first point on the positive part of the horizontal axis that belongs to the unbounded component ($D = 0$ if the origin is already in the unbounded component). There exists c_3 and $c_4 \in \mathbb{R}^+$ such that

$$P(D > x) \le c_3 \exp(-c_4 x).$$

We prove this theorem by mapping the continuous model onto a (discrete) bond percolation model, and applying Theorem 1.

A. Mapping

We declare each edge $\langle x, y\rangle \in \mathbb{L}^2$ *open* if the conditions below are fulfilled in the Boolean model (we assume without loss of generality that $x_1 \le y_1$ and $x_2 \le y_2$, see Fig. 2):

- There exists an occupied crossing in the rectangle $[x_1 d, (y_1+1)d] \times [x_2 d, (y_2+1)d]$ in the direction of its longest edge.
- There exists a vertical and an horizontal crossing in the square $[x_1 d, (x_1+1)d] \times [x_2 d, (x_2+1)d]$.
- There exists a vertical and an horizontal crossing in the square $[y_1 d, (y_1+1)d] \times [y_2 d, (y_2+1)d]$.

These conditions depend on a parameter d, that determines the size of the rectangle. We know that if the Boolean model is supercritical, the probability that the three conditions are fulfilled tends to one when d tends to infinity (see e.g. Corollary 4.1 in [2]). However, in this mapping, the status of two contiguous edges are *not* independent. We constructed thus a dependent bond percolation model that has the desired property that any of its connected component correspond to a connected component in the Boolean model. Therefore, if the first point on the horizontal axis that belongs to the infinite cluster has the horizontal coordinate N in the discrete model, then there exists a point of the Boolean model in the interval $[Nd, (N+1)d]$ that belongs to the infinite cluster. It follows from this observation that if Theorem 1 is valid in this dependent bond percolation model, then Theorem 3 is also valid.

B. Domination of the dependent model

To validate Theorem 1 in the above dependent bond percolation model, we use the results in [8]. Let $(N, 0)$ be the first point of $\mathbb{Z}^+ \times \{0\}$ (i.e. the one with the smallest coordinate) that belongs to the infinite connected cluster. Consider the 1-dependent bond percolation model defined above, where edge e is open with some probability $0 \leq p' \leq 1$ independently of edges separated by distance greater than 1. The claim is that $P(N \leq n) \geq \alpha(n)$ for the 1-dependent model is implied when it is true for the independent model. That is, the probability to hit the infinite open cluster within n steps is at least as great on the dependent graph as on the independent graph.

We show this by employing Theorem 0.0 in [8] which states that the k-dependent random field of $\{0, 1\}$ random variables $(X_s)_{s\in\mathbb{Z}^2}$ stochastically dominates a translation invariant product random field $(Y_s)_{s\in\mathbb{Z}^2}$ of density $p'' > 0$, whenever p' is sufficiently large. Furthermore, $p'' \to 1$ when $p' \to 1$. If we consider the indicator function $1_{\{N\leq n\}}$ for the event $\{N \leq n\}$, where $1_{\{N\leq n\}}$ is an increasing function on our state space Ω, the theorem tells us that by stochastic domination

$$\int 1_{\{N\leq n\}} dP_{p'} \geq \int 1_{\{N\leq n\}} dP_{p''}. \tag{4}$$

The result of integration gives $P_{p'}(N \leq n) \geq P_{p''}(N \leq n)$. Finally, by increasing the mapping parameter d, one can make p' large enough, so that $p'' > 1/2$ and Theorem 1 becomes valid for the field $(Y_s)_{s\in\mathbb{Z}^2}$.

V. CONCLUSION

In this paper, we showed that the distance traveled by an object moving along a straight line in a supercritical percolation model until it hits the giant component is bounded from above by a (shifted) exponential variable. It was already known that the distance before hitting *any* connected component is exponentially distributed. With the results of this paper, we now have a lower and an upper bound for the distribution of the distance traveled before hitting the giant component.

This distribution is very relevant in the context of detection of an intruder in wireless sensor networks, as outlined in the introduction, as well as that of connectivity in mobile ad hoc networks. Indeed, finite connected components are essentially disconnected parts of the network, as they cannot exchange data with the vast majority of the other nodes (those who belong to the giant component). Therefore, from the point of view of a mobile user, the value of interest is the time before connecting to the giant component. The bound computed in this work provides some guarantee about connectivity in this very simple mobile scenario. Future work will address the case where the other nodes also move.

REFERENCES

[1] D. Stoyan, W. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Willey & Sons, Chichester, 1995.

[2] R. Meester and R. Roy. *Continuum percolation*. Cambridge University Press, 1996.

[3] B. Liu, P. Brass, O. Dousse, P. Nain, and D. Towsley. Mobility improves coverage of sensor networks. 2005. preprint.

[4] G. Kesidis, T. Konstantopoulos, and S. Phoha. Surveillance coverage of sensor networks under a random mobility strategy. In *Proc. IEEE Sensors*, October 2003.

[5] Chao Gui and Prasant Mohapatra. Power conservation and quality of surveillance in target tracking sensor networks. In *Proc. Mobicom*, Philadelphia, September 2004.

[6] O. Dousse, P. Mannersalo, and P. Thiran. Latency of wireless sensor networks with uncoordinated power saving mechanisms. In *Proc. Mobihoc*, Tokyo, May 2004.

[7] G. R. Grimmett. *Percolation*. Springer, 1999.

[8] T. M. Liggett, R. H. Schonmann, and A. M. Stacey. Domination by product measures. *Annals of Probability*, 25(1):71–95, 1997.