Analytic Bounds on Data Loss Rates in Mostly-Covered Mobile DTNs

Max Brugger, Kyle Bradford, Samina Ehsan, Bechir Hamdaoui, Yevgeniy Kovchegov
Oregon State University, Corvallis, OR 97331
bruggerm,bradfork,ehsans,hamdaoub,kovchegy@onid.orst.edu

***Abstract*— We derive theoretical performance limits of densely covered delay-tolerant networks (DTNs). In the DTN model we study, a number of fixed (data collector) nodes are deployed in the DTN region where mobile (data generator) nodes move freely in the region according to Brownian motion. As it moves, each mobile node is assumed to continuously generate and buffer data. When a mobile node comes within the communication coverage range of a data collector node, the mobile node immediately and completely uploads its buffered data to the data collector node, and then resumes generating and buffering its data. In this paper, we first derive analytic bounds on the amount of time a mobile node spends without communication coverage. Then, using these derived bounds, we derive sufficient conditions on node density that statistically guarantee that the expected amount of time spent in the uncovered region remains below a given threshold. Additionally, we derive sufficient conditions on node density to keep the probability of buffer overflow below a given tolerance.**

I. Introduction

Delay-tolerant networks (DTNs) are a class of networks that are, by nature, partially covered or intermittently connected. As a consequence, traditional end-to-end routing paradigms may not be the most effective in delivering data across nodes, due to the absence of multi-hop paths. In such sparse networks data delivery is only possible through the *store-carry-and-drop* routing approach, which relies on node mobility to carry data. Therefore, applications supported by these networks are typically delay insensitive/tolerant, as data packets are expected to experience some delay before reaching their destinations. DTNs have recently attracted significant interest in the context of mobile sensor networks (e.g., event/data collection [1–3], animal monitoring/tracking [4, 5], mobile ubiquitous LAN extensions [6, 7]), and continue to find new applications, for instance in vehicular networks (e.g., [8–10]).

Due to their importance and wide range of applications, there has been considerable research focus on DTNs, ranging from protocol design [11–14] to connectivity analysis [15–17] and delay modeling and characterization [16, 18–22]. The work in [16] uses continuum percolation theory [23] to show how delays in large wireless networks scale with the Euclidean distance between the sender and the receiver. Speed of information propagation has recently also been studied analytically for static [18, 19] as well as mobile [20–22] DTNs. The authors in [19] derived upper bounds on the maximum propagation speed in large-scale wireless networks, and those in [21] derived analytic upper bounds on information delay in large-scale DTNs with possible mobility and intermittent connectivity. Network connectivity has also been intensively studied, but mostly in the context of large-scale networks only. In [15], the authors derived an upper bound on the delay sufficient for disconnected networks to become connected through node mobility. The work in [16] derived the minimum node density required to ensure connectivity in large static networks.

This work was supported, in part, by the National Science Foundation (NSF).

In contrast, this work aims at deriving analytic upper bounds on the expected time a mobile node spends without communication coverage in mostly, but not fully, covered DTNs as a function of the communication coverage ratio; i.e, DTNs whose coverage ratio is close to one. Intermeeting times, defined as the time a mobile node spends before running into another node, have been derived in [24] for the generalized hybrid random walk mobility model. Additionally, La [24] shows that the distribution of intermeeting times can be approximated by an exponential distribution when mobile nodes move independently from one another and when the probability of establishing communication links among nodes is relatively low. This result provides support for our use of a 2-D Brownian Motion model of a mobile node, which we will show also has approximately exponential intermeeting times under the assumptions of the Poisson Clumping Heuristic [25] (and described in more detail in Subsection III-A).

Cai et al. [26] show that when removing the boundaries in a two-dimensional random walk model, intermeeting times follow power-law distributions, but not exponential ones. This does not conflict with our model. It is well-known that the 2-D random walk is only recurrent on bounded regions and transient otherwise, whereas 2-D Brownian Motion is always recurrent. Given a small region about the origin, this is the difference between the mobile node returning only finitely often and returning infinitely often to the region, respectively, and explains why the distribution of 2-D random walk intermeeting times has a shorter tail (lower probability of high values) than the distribution of 2-D Brownian Motion hitting times.

In the DTN model we study, a number of fixed nodes (also referred to as *access points*) are deployed in the DTN region, where mobile nodes (also referred to as *data generators*) move freely in the region by following a Brownian motion. As it moves, each mobile node is assumed to continuously generate and buffer data. When a mobile node comes within the communication coverage range of a data collector node, the mobile node immediately and completely uploads its buffered data to the data collector node, and then resumes generating and buffering its data. If still within the communication coverage range the mobile node generates data but uploads it virtually instantaneously.

In this work, we first use the Poisson Clumping Heuristic [25] to provide analytic bounds on the expected hitting time, the time a mobile node spends without communication coverage. Then, using these derived bounds, we derive sufficient conditions on node density that ensure that the expected hitting times are guaranteed to be below a given time threshold and that the probability of buffer overflow is below a given tolerance. Finally, using simulations, we validate the sufficiency of our conditions.

Our contributions in this paper are the following:

- Derive analytic bounds on the expected time mobile nodes spend without communication coverage.
- Provide sufficient node density conditions ensuring that the expected time mobile nodes spend without coverage remains below a fixed threshold.
- Derive analytic bounds on the rate at which mobile nodes drop data due to buffer overflow.
- Provide sufficient node density conditions ensuring the rate at which mobile nodes drop data, equivalently the probability of the hitting time exceeded that required for buffer overflow, stays below a given tolerance.
- Validate/verify the derived results via simulations.

The rest of the paper is organized as follows. In Section II, we state our network model. To introduce our methods, we first derive results for a one-dimensional model in Section III. We then derive and present our analytic results in Section IV. In Section V, we validate via simulations the derived models/bounds. Finally, we conclude the paper in Section VI.

II. DTN Model

In this paper, we analyze the performances of mostly covered DTNs for both one-dimensional (1-d) and two-dimensional (2-d) node deployment models. For each node deployment model, a number of fixed nodes (data collectors) are deployed in the DTN region, where mobile nodes (data generators) move freely in the region, following a Brownian motion. As they move, mobile nodes are assumed to continuously generate and buffer data independently from one another, at a rate c. When a mobile node comes within the communication coverage range of a data collector node, the mobile node immediately and completely uploads its buffered data to the data collector node, and then resumes generating and buffering its data. Each mobile node is assumed to have a buffer space of size B bits, and when the buffer is full, data is dropped.

Our focus in this work is on the study of dense DTNs. That is, DTNs that are mostly covered, but not fully. Hence, the network formed by the data collector nodes is assumed to be unconnected, and the communication coverage ratio is assumed to be close to 1. In these dense DTNs, as mobile nodes move, they will eventually traverse a data collector's communication coverage area, and can then upload their buffered data. To this end, the *coverage ratio*[1] is assumed to be close to 1 throughout this paper, and all the mathematical analysis in this work depends heavily on this assumption.

We use node deployment models in the 1-d and 2-d models with regularly repeating patterns which we will describe fully in the forthcoming sections. Suffice it to say for now that in the 1-d model, each access node is exactly the same distance apart. The 2-d model is laid out on a square grid (with data collectors at each intersection) and extends in each direction, we assume, forever. The assumption that the grid continues forever is unnecessary, but considering the boundary behavior would add little to our analysis. The assumption of a square grid is also unnecessary. It is not hard to extend our analysis to other regularly-repeating patterns like a triangular or hexagonal grid. The hardest part is working out the trigonometry.

[1]The coverage ratio is defined as the fraction of the area covered by collector nodes' communication ranges to that of the total DTN area.

III. 1-d Example

This section demonstrates our analysis by a simple example of a DTN with a single user modeled by a one-dimensional diffusion on a line. This helps to clarify our methods of analysis, though the predictions of the model are significantly more accurate in two dimensions than in one. As we will discuss in Subsection III-A, this is partly due to the requirement of the Poisson Clumping Heuristic that the stochastic process have a nonzero drift away from the set we are interested in, in this paper this is the uncovered area. In one dimension, standard Brownian motion has zero drift ($\mu = 0$) and it seems unrealistic to impose a value. Therefore, the results derived in this section are only illustrative and thus there was no need to validate them in Section V.

We describe the movement of a mobile node by a Brownian motion, X_t, on a line of length n with the endpoints mapped to each other, so as to avoid issues with the boundary. Geometrically, this is equivalent to a Brownian motion on a circle of circumference n. Let w denote the number of access points, each having radius r. As shown in Fig. 1, let the positions of the access points be $\{\dots, -\frac{2n}{w}, -\frac{n}{w}, 0, \frac{n}{w}, \frac{2n}{w}, \dots\}$. The distance between two neighboring access points is then $\frac{n}{w}$. We assume that the DTN is mostly covered, meaning that the coverage ratio is close to 1 (i.e., $2rw/n \approx 1$). We also assume that X_t has drift $\mu = 0$ and variance $\sigma^2 = 1$.

Since the regions between any two access points are identical, it suffices to consider just one uncovered area between two neighboring access points. In particular, we consider the set $\mathcal{C}_0$ occurring while X_t is between the access points located at positions 0 and n/w.

Definition 3.1: We define $\mathcal{C}_j$ for some $j \in \mathbb{N}$ to be the set of times when the mobile node is within the uncovered area between the access points located at 0 and n/w. Formally,

$$C_j = \left\{ t : r < X_t < \frac{n}{w} - r \right\}.$$

If the mobile node exits the interval $[0, n/w]$ we record the times when the mobile node is in the uncovered area in a new set of times, C_{j+1}.

Consider two sets of times when the mobile node is in the uncovered area, C_j and C_k for some $j, k \in \mathbb{N}$. We observe that $C_j \neq C_k$ if and only if there exists times t^*, $t_j \in C_j$ (so $r < X_{t_j} < n/w - r$), and $t_k \in C_k$ (so $r < X_{t_k} < n/w - r$) such that $t_j < t^* < t_k$ and $X_{t^*} = 0$ or n/w.

We would like to point out that for times $t_0 < t_1 < t_2$ it is possible for the mobile node to move such that $X_{t_0} \in [r, n/w - r]$, $X_{t_1} \in (0, r)$, and $X_{t_2} \in [r, n/w - r]$ and $X_t \in (0, n/w)$ for all $t \in [t_0, t_2]$. This is the case that the mobile node is initially in the uncovered region, moves into either of the two neighboring covered regions and returns to the uncovered region without ever hitting 0 or n/w, and thus we would not start recording a new set

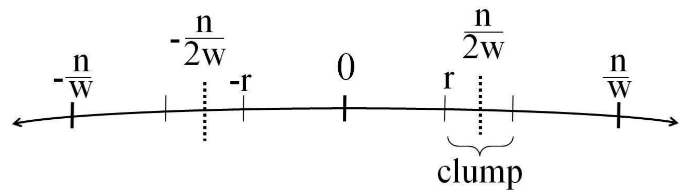

Fig. 1. Linear geometry: w access points, each with a communication coverage area of radius r, are located on a circle of length n at positions $\{\ldots, -\frac{2n}{w}, -\frac{n}{w}, 0, \frac{n}{w}, \frac{2n}{w}, \ldots\}$; $n/(2w)$ is the furthest distance of a mobile node from an access point; thus the clump is the set of times (*not* the set of points in space) when the mobile node is in the uncovered region, marked "clump." We suppose $n/(2w)$ is close to r.

of times even though the mobile node left the uncovered region. Therefore it is possible that C_j is a disconnected set.

For a Brownian motion, the expected hitting time to the endpoints of an interval $[a, b]$, starting at a point x such that $a < x < b$ is given by $\mathbb{E}T(a, x, b) = (b - x)(x - a)$ [25]. Let T_1 be the random amount of time it takes a mobile node to hit an access point from the time t immediately following $\mathcal{C}_0$ (so $X_t = r$ or $n/w - r$ and in either case the nearest access points are at 0 and n/w). Thus, $\mathbb{E}T_1 = (n/w - r)r$, and if we let T_2 be the random amount of time it takes a mobile node to return back to clump $\mathcal{C}_0$ immediately after leaving an access point, $\mathbb{E}T_2 = r^2$. Hence, the expected length of the total block $T_0 = T_1 + T_2$, or the expected amount of time required for the mobile node to walk from an access point to an uncovered area and back to an access point is

$$\mathbb{E}T_0 = rn/w. \tag{1}$$

Under specific assumptions, we can say that T_0 is approximately exponentially distributed with parameter $\frac{w}{rn}$ when $\frac{rw}{n} \leq \frac{1}{2}$ is sufficiently large [25]. These specific assumptions are explained in the following subsection, Subsection III-A.

A. Poisson Clumping Heuristic

Given a time-dependent stochastic process, and a set A, if the arrival times to A are memoryless (specifically, the distribution of arrival times to A is exponential) and the process intersects the set A rarely (relative to the set of times the process could arrive at A which is the total time interval), then we can approximate the behavior of this process' arrivals to the set by the Poisson Process. In the Poisson Process, the inter-arrival times, denoted T, of the process to the set A are exponentially distributed, with parameter λ, and the random variable for the number of times the mobile node hits an access point up to time t is Poisson, with parameter λt. In the language of the heuristic, λ is called the clump rate, so named because the random sets of times, denoted C, that the process spends in the area A appear to "clump" together. The approximations given by the Poisson Clumping Heuristic improve if the process is unlikely to return to A immediately after leaving A; there should typically be some drift away from A. Let $\pi(A)$ be the probability (with respect to the stationary distribution) that the process is in A. The main result of the heuristic is:

$$\pi(A) = \lambda \mathbb{E}C. \tag{2}$$

The assumption that the interarrival times follow an exponential distribution additionally gives us that $\lambda = 1/T$, where T is the hitting time [25]. In our particular application to a DTN with a single user modeled by a one-dimensional diffusion on a line, this is analogous to the total block size, T_0, with expected value given by Equation 1.

The heuristic itself does not prescribe a universal definition of what does and does not constitute a clump, choosing instead to require "sufficiently long inter-arrival times" and arrivals to A being "sufficiently rare." The reason for this is that at any scale, whether we look at the time interval $[0, 100]$ or $[0, 10^9]$ the Poisson process yields the same clumping behavior: the process returns to the area A after a "long time," relative to the scale of the time interval, and the clump size $|C|$ is small (but nontrivial), relative to the scale of the time interval.

In practice, the process being modeled by the heuristic will not be approximated equally closely at all scales. So even though the heuristic does not call for it, we indirectly affect the length of the interarrival times by specifying a condition on the element of the model over which we have the most direct control: as in Definition 3.1 we define the exact point the mobile node has to reach before we say it has exited the clump for good. After the process hits this point, and subsequently returns to the set, we record the times when the mobile node is in the uncovered area in a new set of times.

B. Performance Analysis

Let $\nu = w/n$ and $\eta = 2rw/n$ denote respectively the access point density and the communication coverage ratio, where again r represents the radius/range of the communication coverage area of an access point. We begin by deriving some useful statistics (expected time without coverage and bit loss rate) for this toy example and use these to derive a function relating the density of access points to the expected amount of time spent without coverage and to the data loss rate.

Proposition 3.2: For sufficiently large η, the expected time a mobile node spends without communication coverage is $\frac{r}{\nu} - 2r^2$.

Proof: The average clump rate, λ_0, is the inverse of the expected time $\mathbb{E}T_0$ of the total block size, and thus it follows from Eq. (1) that $\lambda_0 = \frac{1}{\mathbb{E}T_0} = \frac{\nu}{r}$. The expected clump size, $\mathbb{E}C_0$, can then be expressed as

$$\mathbb{E}\mathcal{C}_0 = \frac{\pi(A_0)}{\lambda_0} = \frac{\pi(A_0)r}{\nu}$$

where $\pi(A_0)$ is the probability of being in the clump. Now, by noting that the probability of being in our clump, $\pi(A_0)$, can also be expressed as

$$\pi(A_0) = \frac{n - 2rw}{n} = 1 - 2r\nu$$

we can write $\mathbb{E}\mathcal{C}_0 = r/\nu - 2r^2$. ■

Let us now assume that each mobile node has a data buffer of capacity B bits, and that when the buffer becomes full, data is dropped (due to buffer overflow). Recall that data is generated at a constant rate c, and is fully uploaded when the mobile node reaches an access point's communication range. In what follows, let $\tau = B/c$, which represents the amount of time required to overflow the buffer of the mobile node.

Corollary 3.3: For a sufficiently small threshold τ, the expected time a mobile node spends without communication coverage is guaranteed to remain below τ if the density ν of access points is above $r/(\tau + 2r^2)$.

Proof: Proof follows from Proposition 3.2. ∎

Data loss occurs when the buffer of the mobile node overflows, and the buffer of the mobile node overflows when the time the mobile node spends without communication coverage in the interval $(r, n/w - r)$ exceeds τ. Denote the first hitting time of the process to r or $n/w - r$ by $T^* = T_r \wedge T_{n/w-r}$. Hence, the data loss rate equals $P(T^* > \tau)$.

Proposition 3.4: For sufficiently large η, the data loss rate of a mobile node is bounded above by $(r/\tau)(1/\nu - 2r)$.

Proof: Denote the first hitting time of the process to r or $n/w - r$ by $T^* = T_r \wedge T_{n/w-r}$. Since the interval $(r, n/w - r) \subset [0, n/w]$, we observe that $P(T^* > \tau) < P(|C_0| > \tau)$.

Now, using Markov's Inequality, we can write $P(|C_0| > \tau) < \mathbb{E}C_0/\tau$, yielding an upper bound of $(r/\tau)(1/\nu - 2r)$ on the achievable data loss rates. ∎

We provide the following sufficient condition on the density of access points to guarantee data loss rates due to buffer overflow remain below some condition on the rate ϵ.

Corollary 3.5: For sufficiently large η, data loss rates are guaranteed to remain below the rate condition ϵ if the density ν of access points satisfies $r/(\epsilon\tau + 2r^2)$.

Proof: Recall that the data loss rate is bounded above by $(r/\tau)(1/\nu - 2r)$. Hence, it suffices that $(r/\tau)(1/\nu - 2r) \le \epsilon$ or equivalently $\nu \ge r/(\epsilon\tau + 2r^2)$ to ensure that the data loss rate does not exceed the threshold ϵ. ∎

IV. 2-D DTN Model

We consider that mobile nodes follow a Brownian motion and move in a 2-dimensional plane. Collector nodes are placed in the plane to form a grid. We assume that each of the collector nodes has a circular coverage region with radius κ. The spacing distance D between two neighboring collector nodes is assumed to be larger than $\sqrt{2}\kappa$. This distance is also assumed to be smaller than 2κ so as to ensure that the DTN contains regions of no coverage, referred to as uncovered regions, that are disconnected. As shown in Fig. 2, we then draw a square of side length D around each uncovered region, with the center of each region placed in the middle of the square. Each corner of each square corresponds to one collector node.

We make two observations before proceeding with our derivation and analysis. First, the problem is symmetric, and hence studying one square suffices. Note that once a mobile node reaches the edge of a square coming from the edge of an uncovered region, returning back to the same region or another one makes no difference vis-a-vis of our clumping analysis. Second, because the uncovered region has an odd shape and our boundary region has a square shape, it is too difficult to derive the exact clump rate. Instead, we derive bounds on the clump rate.

We inscribe the largest possible circle in the uncovered region, centered at the center of the square, and denote the radius of the circle ρ_1. We also circumscribe the smallest possible circle around the uncovered region, centered at the center of each region, and denote the radius of this circle ρ_2. The geometry of this grid is shown in Fig. 2.

In our model, the distribution of hitting times satisfies the assumptions in Subsection III-A regarding the rarity with which

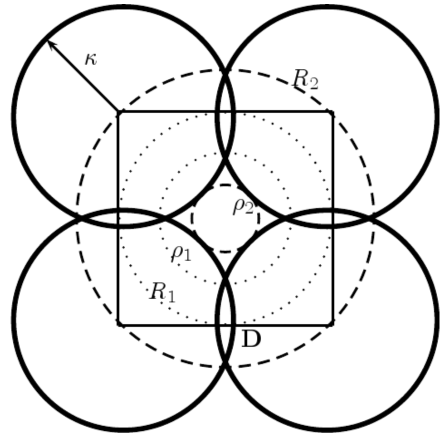

Fig. 2. Grid geometry: the uncovered area (the star shape) is bounded by two circles, one of radius ρ_2 from within and one of radius ρ_1 from without, which will be used to calculate bounds on the expected time a mobile node spends outside the communication coverage area. D is the distance between two neighboring collector nodes; κ is the radius of the communication coverage area of a collector node.

the mobile node hits the uncovered area because of the assumption of a high coverage ratio, and the fact that the drift for the radial part of Brownian Motion, given by the Bessel Process, has drift $\mu(r) = 1/(2r)$, where r is the Euclidean distance of the Brownian Motion from the origin.

Let $\pi(C)$ be the probability (given by the stationary distribution) that the process is in the uncovered region and $\mathbb{E}C$ the expected amount of time spent in the uncovered region. We recall the main result of the heuristic, $\pi(C) = \lambda\mathbb{E}C$, described in Subsection III-A.

The time a mobile node spends with communication coverage corresponds to the time it takes a mobile node to reach the edge of the square from the edge of an uncovered region, and then to return to one of the uncovered regions again.

To derive an upper bound, we first investigate a radial diffusive process on an inner disk with radius $\rho_1 = \frac{1}{2}D - \sqrt{\kappa^2 - \frac{1}{4}D^2}$ and an outer disk with radius $R_1 = \frac{1}{2}D$ centered at the same point (these boundaries are shown in Fig. 2). We then investigate a radial diffusive process on an inner disk with radius $\rho_2 = \frac{\sqrt{2}}{2}D - \kappa$ and an outer disk with radius $R_2 = \frac{\sqrt{2}}{2}D$ to find a lower bound on the clump rate. We can see that with these two disks, we are inscribing the square within a disk and another within the square, as is shown in Fig. 2. The geometric properties of the square allow for a probabilistic coupling construction, where the radial Brownian motion on a square region is stochastically dominated from above and below by the radial Brownian motions (Bessel processes) whose boundary conditions are respectively the inner and outer circle. A similar geometric approach to provide upper and lower bounds on the hitting time, using the same function derived in Proposition 4.1, is used in [27].

Since we are investigating the radial diffusive process on a disk with radius R_i centered at the center of a circle of radius ρ_i where $i \in \{1, 2\}$, the Brownian motion in either case can be modeled as a Bessel process with parameter 2, i.e., with drift $\mu(r) = \frac{1}{2r}$ and variance $\sigma^2 = 1$ [28].

A. Inner and Outer Disks of Radii ρ_i and $R_i > \rho_i$

Let us now consider a disk of radius ρ_i centered in a disk of radius $R_i > \rho_i$ as shown in Fig. 2, where $i \in \{1, 2\}$. Let us assume that a mobile node moves inside the disk of radius R_i, and it bounces back when it hits the boundary of the disk. We define then the smaller disk of radius ρ_i to be our clump, and the hitting time to be the time between two consecutive clump visits. The hitting time is then the time it takes a mobile node to reach the boundary of the outer disk of radius R_i given it just left the inner disk of radius ρ_i plus the time it takes a mobile node to hit back the inner disk given that it just bounced back from hitting the boundary. The expected value of this hitting time is derived from Lemmas 4.2 and 4.3 and stated in the following proposition.

Proposition 4.1: For a smaller disk of radius ρ_i centered in the larger disk of radius R_i, the expected hitting time is $h(\rho_i, R_i)$, where $h(\rho, R) = R^2 \ln \left|\frac{R}{\rho}\right|$.

For one pair of inner and outer circles, of radii ρ and R, the hitting time is defined as the sum of the amount of time the mobile node spends traveling from the edge of the uncovered region to the boundary of the larger disk (with expected value denoted $\mathbb{E}_\rho[T_R]$), and the time spent returning from the edge of the larger disk to the boundary of the communication disk (with expected value denoted $\mathbb{E}_R[T_\rho]$). We calculate these two means in the following lemmas.

Lemma 4.2: $\mathbb{E}_\rho[T_R] = \frac{1}{2}\left(R^2 - \rho^2\right)$.

Proof: Consider the radial interval $[a, R]$, where ρ is in this interval. Let $\mathbb{E}_\rho[\min\{T_a, T_R\}]$ be the expected time that it takes to hit either a or R given that we start at ρ. Because the Bessel process has a drift away from the origin that becomes large as we approach the origin, then the expected time $\mathbb{E}_\rho[T_R]$ it takes to hit the boundary given we are at ρ is equal to $\lim_{a\to 0} \mathbb{E}_\rho[\min\{T_a, T_R\}]$.

From Eq. (15.3.12) in [28], it follows that the expected hitting time $\mathbb{E}_\rho[\min\{T_a, T_R\}]$ equals

$$\int_a^\rho 2m(y)\frac{(S(R)-S(\rho))(S(y)-S(a))}{S(R)-S(a)}\,dy + \int_\rho^R 2m(y)\frac{(S(\rho)-S(a))(S(R)-S(y))}{S(R)-S(a)}dy$$

where $S(x) = \ln|x|$ and $m(x) = x$ are the speed function and speed density of the Bessel process. Integrating the above expression yields $\mathbb{E}_\rho[\min\{T_a, T_R\}]$ to be equal to

$$\frac{2}{\ln(R/a)}\left[\ln(R/\rho)a^2\left(\frac{1}{2}\left(\frac{\rho}{a}\right)^2\left(\ln(\rho/a)-\frac{1}{2}\right)+\frac{1}{4}\right) - \ln(\rho/a)R^2\left(-\frac{1}{4}-\frac{1}{2}\left(\frac{\rho}{R}\right)^2\left(\ln(\rho/R)-\frac{1}{2}\right)\right)\right].$$

When a goes to 0, the first term goes to $\rho^2\ln\left(\frac{R}{\rho}\right)$ and the second term goes to $R^2\left(\frac{1}{2}+\left(\frac{\rho}{R}\right)^2\left[\ln\left(\frac{\rho}{R}\right)-\frac{1}{2}\right]\right)$. Since $\mathbb{E}_\rho[T_R] = \lim_{a\to 0}\mathbb{E}_\rho[\min\{T_a, T_R\}]$, it then follows that $\mathbb{E}_\rho[T_R] = \frac{1}{2}(R^2 - \rho^2)$. ■

Lemma 4.3: $\mathbb{E}_R(T_\rho) = R^2\ln|\frac{R}{\rho}| - \frac{1}{2}\left(R^2-\rho^2\right)$.

Proof: Let Δ be small and assume that whenever our process hits R that it jumps instantaneously to $R-\Delta$. Taking the limit as Δ approaches zero will make R a reflecting boundary. By letting

$$\begin{aligned} p_\Delta &= P_{R-\Delta}(T_\rho < T_R) \\ a_\Delta &= \mathbb{E}_{R-\Delta}(T_\rho | T_\rho < T_R) \\ b_\Delta &= \mathbb{E}_{R-\Delta}(T_R | T_R < T_\rho) \end{aligned}$$

one can write $\mathbb{E}_{R-\Delta}[\min\{T_\rho, T_R\}] = p_\Delta a_\Delta + (1-p_\Delta)b_\Delta$.

By Wald's Theorem, we find that $\mathbb{E}_{R-\Delta}(T_\rho)$

$$\begin{aligned} &= p_\Delta a_\Delta + (1-p_\Delta)p_\Delta(a_\Delta + b_\Delta) \\ &\quad + (1-p_\Delta)^2 p_\Delta(a_\Delta + 2b_\Delta) + \cdots \\ &= a_\Delta + \frac{b_\Delta(1-p_\Delta)}{p_\Delta} \\ &= \frac{1}{p_\Delta}\mathbb{E}_{R-\Delta}(T_{\rho\wedge T_R}) \end{aligned}$$

From Eq. (15.3.10) [28], it follows

$$P_{R-\Delta}(T_\rho < T_R) = \frac{S(R)-S(R-\Delta)}{S(R)-S(\rho)}.$$

and from Eq. (15.3.12) [28], it follows $\mathbb{E}_{R-\Delta}[\min\{T_\rho, T_R\}]$ equals

$$\int_\rho^{R-\Delta}\frac{2(S(R)-S(R-\Delta))(S(y)-S(\rho))}{S(R)-S(\rho)}m(y)dy + \int_{R-\Delta}^R\frac{2(S(R-\Delta)-S(\rho))(S(R)-S(y))}{S(R)-S(\rho)}m(y)dy.$$

From this, we can write $\mathbb{E}_{R-\Delta}(T_\rho)$ as

$$2\int_\rho^{R-\Delta}(S(y)-S(\rho))m(y)dy + \frac{2(S(R-\Delta)-S(\rho))}{S(R)-S(R-\Delta)}\int_{R-\Delta}^R(S(R)-S(y))m(y)dy.$$

Substituting in our functions and solving, we can write $\mathbb{E}_{R-\Delta}(T_\rho)$ as

$$\frac{\rho^2\ln|\frac{R-\Delta}{R}|\left[(\frac{R-\Delta}{\rho})^2(\ln|\frac{R-\Delta}{\rho}|-\frac{1}{2})+\frac{1}{2}\right]}{\ln|\frac{R-\Delta}{R}|} - \frac{\ln|\frac{R-\Delta}{\rho}|\left[(\frac{R-\Delta}{R})^2(\ln|\frac{R-\Delta}{R}|-\frac{1}{2})+\frac{1}{2}\right]}{\ln|\frac{R-\Delta}{R}|}.$$

Now, by taking the limit as Δ goes to zero, we find $\mathbb{E}_R(T_\rho) = R^2\ln|\frac{R}{\rho}| - \frac{1}{2}\left(R^2-\rho^2\right)$. ■

We can now combine our two lemmas to prove Proposition 4.1.

Proof of Prop. 4.1: For one pair of inner and outer circles, of radii ρ and R, the hitting time is the sum of $\mathbb{E}_\rho[T_R]$ and $\mathbb{E}_R[T_\rho]$, derived respectively in Lemmas 4.2 and 4.3; therefore $h(\rho, R) = R^2\ln|\frac{R}{\rho}|$. ■

The following corollary, which is an immediate application of the hitting time function, h, is used throughout to derive the main results of this paper.

Corollary 4.4: For a star-shaped inner region and a square boundary of side length D, the expected hitting time is lower bounded by $h(\rho_1, R_1)$ and upper bounded by $h(\rho_2, R_2)$.

Looking back over the results we have derived, in order to perform our analysis on a different node deployment structure of regularly-repeating shapes (e.g. triangles, squares, hexagons), we just need to calculate the radii of the four circles we used: the radius ρ_2 of the largest circle contained in the uncovered region, the radius ρ_1 of the smallest circle containing the uncovered region, the radius R_1 of the largest circle that can be contained in one of the shapes and the radius R_2 of the smallest circle containing one of the shapes, each in terms of κ and D. Then, computing the upper bounds and sufficient conditions can be done by applying $h(\rho_i, R_i)$ for each i and proceeding through the calculations in the proofs.

B. Upper Bounds and Sufficient Conditions

We now derive upper bounds on the expected hitting times, and provide sufficient conditions on node density that guarantee that the expected time the mobile node spends without coverage does not exceed a given threshold. We define the node density ν to be equal to $1/D^2$.

Given that each mobile node has a data buffer with a limited capacity of B bits, it may happen that the buffer overflows, causing data to be dropped. Again, let $\tau = B/c$ represent the time required to overflow the buffer of the mobile node.

Proposition 4.5: The expected amount of time a mobile node spends without communication coverage, $\mathbb{E}C$, is bounded above by

$$\frac{\kappa}{\sqrt{2\nu}}\left(1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}}\right).$$

Proof: The probability of being in the uncovered region, $\pi(C)$, is the ratio of the expected amount of time the mobile node spends in the clump, $\mathbb{E}C$, to the expected amount of time the mobile node spends between clumps, $\mathbb{E}T$. More formally, $\pi(C) \approx \frac{\mathbb{E}C}{\mathbb{E}T}$. From Corollary 4.4, it follows then that

$$\pi(C) \cdot h(\rho_1, R_1) \leq \mathbb{E}(C) \leq \pi(C) \cdot h(\rho_2, R_2).$$

From a geometric argument, we can find the area of the uncovered region and divide it by the area of the square surrounding it to find $\pi(C)$,

$$\pi(C) = 1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}} - \pi\kappa^2\nu + 4\cos^{-1}\left(\frac{1}{2\sqrt{\kappa^2\nu}}\right) \cdot \kappa^2\nu,$$

where the π in the right hand side of the equation is the constant and the π on the left hand side of the equation is the terminology for the stationary distribution.

Because $\frac{1}{4} \leq \kappa^2\nu \leq \frac{1}{2}$, we can then write

$$\pi(C) \leq 1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}}.$$

Now since

$$-\frac{1}{2\nu}\ln\left(1 - \sqrt{2\kappa^2\nu}\right) \leq \frac{\kappa}{\sqrt{2\nu}},$$

then,

$$\mathbb{E}(C) \leq \frac{\kappa}{\sqrt{2\nu}}\left(1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}}\right).$$

■

Corollary 4.6: For a sufficiently small threshold, τ, (where we require $\tau \leq 2\kappa^2$ for the square root to remain real), the expected time a mobile node spends without communication coverage is guaranteed to remain below the threshhold (i.e., $\mathbb{E}C \leq \tau$) if the density ν of collector nodes satisfies

$$\nu \geq 2 \cdot \left(\frac{\kappa}{\tau + \sqrt{4\kappa^4 - \tau^2}}\right)^2.$$

Proof: Proposition 4.5 provides an upper bound on $\mathbb{E}C$, so it suffices that

$$\frac{\kappa}{\sqrt{2\nu}}\left(1 - \sqrt{4\kappa^2\nu - 1}\right) \leq \tau$$

in order for the expected time to remain below τ.

By letting $\hat{\nu} = \kappa^2\nu$, the above inequality becomes

$$\frac{\kappa^2}{\sqrt{2\hat{\nu}}}\left(1 - 2\sqrt{\hat{\nu} - \frac{1}{4}}\right) \leq \tau.$$

Now, solving for $\hat{\nu}$, we find

$$\sqrt{\hat{\nu} - \frac{1}{4}} \geq -\frac{1}{2}\left(\frac{\sqrt{2\hat{\nu}}\tau}{\kappa^2} - 1\right).$$

Because $\frac{1}{4} \leq \hat{\nu} \leq \frac{1}{2}$ and $\tau \leq \kappa^2$, we have that

$$-\frac{1}{2}\left(\frac{\sqrt{2\hat{\nu}}\tau}{\kappa^2} - 1\right) \geq 0.$$

This implies that

$$\left(\frac{\tau^2}{2\kappa^4} - 1\right)\left(\sqrt{\hat{\nu}}\right)^2 - \frac{\sqrt{2}\tau}{2\kappa^2}\sqrt{\hat{\nu}} + \frac{1}{2} \leq 0.$$

The leading coefficient is negative, so this is a parabola in $\sqrt{\hat{\nu}}$ that opens downwards. Using the quadratic formula to find the roots of this polynomial in $\sqrt{\hat{\nu}}$, (note one root will be negative, which is impossible for a square root, so we only need to concern ourselves with the positive root),

$$\sqrt{\hat{\nu}} \geq \frac{\frac{\sqrt{2}\tau}{2\kappa^2} - \sqrt{\frac{\tau^2}{2\kappa^4} + 2\left(1 - \frac{\tau^2}{2\kappa^4}\right)}}{2\left(\frac{\tau^2}{2\kappa^4} - 1\right)}$$

which implies that

$$\hat{\nu} \geq 2 \cdot \left(\frac{1}{\frac{\tau}{\kappa^2} + 2\sqrt{1 - \frac{1}{4}\left(\frac{\tau}{\kappa^2}\right)^2}}\right)^2.$$

Replacing $\hat{\nu}$ by $\nu\kappa^2$ results in the sufficient node density stated in the corollary, which guarantees that the expected time a mobile node spends without communication coverage will be less than the threshold, τ. ■

We now provide a condition on the node density that is sufficient to keep the data loss rate below a threshhold, ϵ. Note that ϵ is a condition on the rate at which buffer overflow can occur, and τ is a condition on the amount of time the mobile node can spend in the uncovered region.

The mobile node overflows its buffer exactly when the mobile node spends an amount of time outside the communication range that exceeds τ. Denote by T^* the first time the mobile node hits the communication range. The data loss rate is then $P(T^* > \tau)$. We do not deal directly with the case that the mobile node hits the covered region while in C but note that our sufficient condition still holds, because $P(T^* > \tau) < P(|C| > \tau)$.

Proposition 4.7: For sufficiently large communication coverage ratio (and requiring $\frac{1}{4} \leq \kappa^2\nu \leq \frac{1}{2}$), the data loss rate of a mobile node is bounded above according to

$$P(|C| > \tau) < \frac{\kappa}{\tau\sqrt{2\nu}} \left(1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}}\right).$$

Proof: Since the uncovered region is contained in Markov's Inequality yields

$$P(|C| > \tau) < \mathbb{E}(C)/\tau.$$

And, from Proposition 4.5, we have

$$\mathbb{E}(C) \leq \frac{\kappa}{\sqrt{2\nu}} \left(1 - 2\sqrt{\kappa^2\nu - \frac{1}{4}}\right).$$

Combining these two inequalities yields the stated upper bound on the data loss rate. ∎

The following sufficient condition guarantees that data loss rates remain below the threshold rate ϵ.

Corollary 4.8: The data loss rates are guaranteed to remain below the threshold rate, $P(|C| > \tau) < \epsilon$ if the density ν of collector nodes satisfies

$$\nu \geq 2 \cdot \left(\frac{\kappa}{\tau\epsilon + \sqrt{4\kappa^4 - (\tau\epsilon)^2}}\right)^2.$$

Proof: If we require $P(|C| > \tau) < \epsilon$, then it is sufficient to require (by Markov's Inequality)

$$P(|C| > \tau) < \frac{\mathbb{E}C}{\tau} < \epsilon$$

and we showed in Corollary 4.6 and Proposition 4.7 that $\mathbb{E}C < \epsilon\tau$ if the node density ν satisfies the resulting inequality. ∎

V. Verification and Validation

Through simulations, we first validate the use of the Poisson Clumping Heuristic by measuring the expected hitting times and comparing them against the derived bounds, and then verify the derived sufficient conditions on node density by mimicking and simulating a Brownian motion.

A. Poisson Clumping Heuristic Validation

Recall that, as illustrated in Section III-A, the derived theoretical results are based on the assumption that the 2-D Brownian motion in dense networks yields approximately exponentially distributed intermeeting times (i.e., times without communication coverage are exponentially distributed), thus allowing us to use the Poisson Clumping Heuristic approach. In this section, we focus on validating the Poisson Clumping Heuristic approach by simulating and measuring the hitting times of a Brownian motion in the 2-D model, and comparing them with the theoretical upper bound (in Corollary 4.4) for a fixed value of κ. Recall that D can range, for fixed κ, from 2κ, where the circles of the covered regions just barely touch, to $\sqrt{2}\kappa$, where the circles overlap so that the uncovered region is at its smallest; the coverage ratio η varies respectively from about 0.7854 to 1. In our simulation, we set $\kappa = 5$.

We use Matlab to simulate 2-D Brownian motion in a square by generating a displacement and angle of the displacement (r_t, θ_t) for $t \in [0, \dots, \infty)$. We generate a normal random variable with

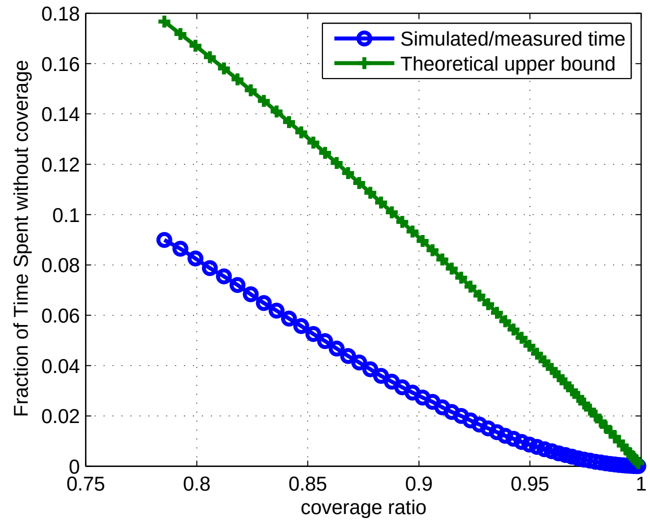

Fig. 3. Measured and theoretical hitting times for 2-d Brownian motion on a square with $\kappa = 5$ and D ranging from $\sqrt{2}\kappa$ to 2κ.

distribution ($\mu = 0, \sigma^2 = 1$) for the displacement of the mobile node, and a uniform random number selected from $[0, \pi]$ for the angle that the mobile node's path makes with the x-axis. Therefore the position of the mobile node at any time T, described in x- and y-coordinates is $(\sum_{i=0}^{T} r_i \cos(\theta_i), \sum_{i=0}^{T} r_i \sin(\theta_i))$.

Shrinking the displacement to zero increases the accuracy of the simulation in modeling simple Brownian motion, but it increases the computation time. We find $s = 0.2$ to be a good choice for the step-size.

Because of symmetry, simulating 2-D Brownian motion on a plane is equivalent to simulating it on a square of side length D with 4 collector nodes each located at one corner and an uncovered area located at its center. The covered region then is the area within κ of any corner. If the simulated Brownian motion exits the square, it is equivalent to continue the simulation with the mobile node placed back inside the box at the opposite position. Fig. 3 shows that the simulated times are well bounded by the derived upper bounds for a range of values for η, and as expected, the higher the coverage, the tighter the bound.

B. Sufficient Node Density Conditions Verification

In this section we test and verify the sufficient conditions on node density stated in Propositions 4.5 and 4.7 to ensure that the time required to overflow the buffer of the mobile node and data loss rate remain below a threshold.

We verify the sufficient condition for the expected time a mobile node spends without communication coverage to be guaranteed to remain below a threshold, τ, for a range of values of κ: 5, 20, 35 and 50, and for D near to $\sqrt{2}\kappa$ for each (1.45κ to 1.78κ). We calculate the sufficient density from Proposition 4.5, which is a function of τ and κ.

The assumption of the heuristic that the region is mostly covered imposes another condition on the scale of the times we can choose for our threshold τ. Our simulation will yield random times spent in the covered region between arrivals to the uncovered region, of which $\mathbb{E}C$ represents an average, and

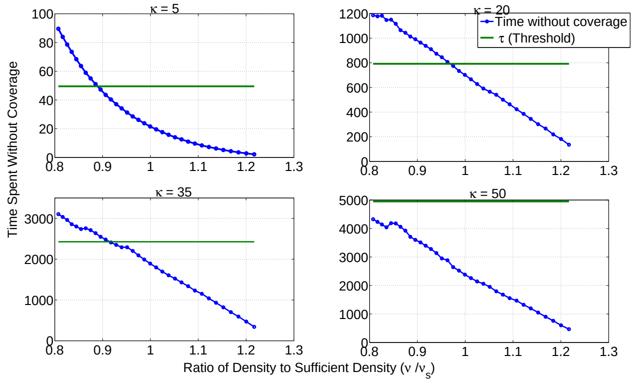

Fig. 4. The measured and the derived upper bound (Proposition 4.5) on the time a mobile node spends without communication coverage when varying D at values near $\sqrt{2}\kappa$, for $\kappa = (5, 20, 35, 50)$. The negative exponential relationship is most clear when $\kappa = 5$ and looks linear in the other graphs. For some values of κ, namely $\kappa = 5$ and 50, the time spent without coverage is below the threshold for values of D relatively small when compared to the sufficient condition. The sufficient condition is much tighter for $\kappa = 20$ and 35. The relationship between κ, D, τ, and the tightness of the sufficient condition is not well understood but seems to depend on the star shape of the uncovered region.

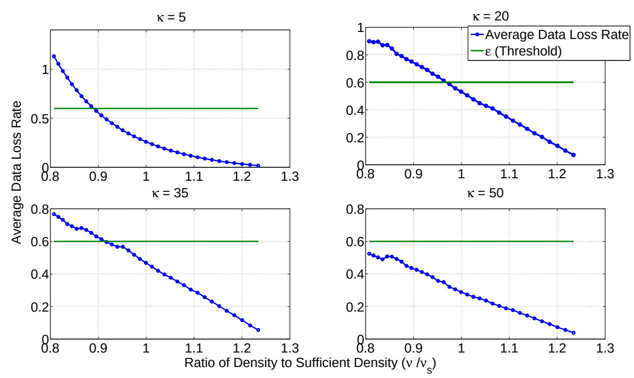

Fig. 5. The measured and the derived upper bound (Proposition 4.5) on the rate at which a mobile node drops data due to buffer overflow. D varies over values near $\sqrt{2}\kappa$, for $\kappa = (5, 20, 35, 50)$. Since the mobile node drops data exactly when the time spent without coverage exceeds the threshold, we see the same relationship with the ratio of the density to the sufficient density as in Figure 4.

we recall from the heuristic, $\mathbb{E}C \leq \pi(C)h(\rho_2, R_2)$, so we want to pick τ close to $\pi(C)h(\rho_2, R_2)$. The value $h(\rho_2, R_2)$ is the expected value of the (exponential) distribution of the time spent between arrivals to the uncovered region, but the heuristic is most accurate in the tail of the distribution, so we can improve our accuracy by picking a time T equal to the 99th percentile of the exponential distribution. Therefore we use T such that

$$0.99 = \int_{-\infty}^{T} \frac{1}{h(\rho_2, R_2)} e^{-\frac{1}{h(\rho_2, R_2)}x}\ dx.$$

Since we scale the radial displacement of the Brownian motion by $s = 0.2$, we want τ such that $2s^2\tau = \pi(C)T$. Therefore $\tau = (\pi(C)T)/0.08$.

For each ratio of the node density to the sufficient density (for each value of D) we simulate the process and measure the time spent outside the coverage for each value of κ. The results are presented in Fig. 4

We do the same for the sufficient condition on the averate data loss rate, setting the acceptable probability of the mobile node overflowing its buffer at 60% (so $\epsilon = 0.6$) and τ chosen according to the process described above. The results are presented in Fig. 5.

To summarize, through simulations, we are able to support the use of the Poisson Clumping Heuristic techniques and illustrate the applicability of the sufficient conditions on the node density.

VI. Conclusion

In this paper, we derived theoretical bounds on data loss rates and packet delays in mostly covered DTNs for both the 1-d and the 2-d node deployment models. For each model, we first provided analytic bounds/approximations on the expected time mobile nodes spend without communication coverage, and derived sufficient conditions ensuring that these times are guaranteed to remain below a given threshold. We then analyzed the data loss rates that mobile nodes experience by providing upper bounds on the achievable data loss rates, and by deriving sufficient conditions that (statistically) guarantee that these rates remain below a given threshold. We verified our obtained models via simulations.

References

[1] Y. Wang and H. Wu, "DFT:MSN: the delay/fault-tolerant mobile sensor networks for pervasive information gathering," in *Proc. of INFOCOM*, 2006.

[2] Y. Wang, H. Wu, F. Lin, and N.-F. Tzeng, "Cross-layer protocol design and optimization for delay/fault-tolerant mobile sensor networks (dft-msn's)," *IEEE Journal of Selected Areas in Comm.*, June 2008.

[3] X. Xu, J. Luo, and Q. Zhang, "Delay tolerant event collection in sensor networks with mobile sink," in *Proc. of INFOCOM*, 2010.

[4] P. E. Clark, D. E. Johnson, M. A. Kniep, P. Jermann, B. Huttash, A. Wood, M. Johnson, C. McGillivan, and K. Titus, "An advanced low-cost GPS-based animal tracking system," *Rangeland Ecol. Manage*, May 2006.

[5] Y. Guo, P. Corke, G. Poulton, T. Wark, G. Bishop-Hurley, and D. Swain, "Animal behaviour understanding using wireless sensor networks," in *Proc. of IEEE Conf. on Local Computer Networks*, 2006.

[6] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULES: Modeling a three-tier architecture for sparse sensor networks," in *Proc. of IEEE SNPA Workshop*, 2003.

[7] D. Jea, A. A. Somasundara, and M. B. Srivastava, "Multiple controlled mobile elements (data mules) for data collection in sensor networks," in *Proc. of IEEE/ACM Int'l Conf. on Distributed Comp. in Sensor Systems*, 2005.

[8] D. Niyato, P. Wang, J. Chee, and M. Teo, "Performance analysis of the vehicular delay tolerant network," in *Proc. of WCNC*, 2009.

[9] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "Improving vehicular delay-tolerant network performance with relay nodes," in *Proc. of 5th Euro-NGI Conf. on Next-Generation Internet Networks*, 2009.

[10] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "Evaluating the impact of storage capacity constraints on vehicular delay-tolerant networks," in *Proc. of 2nd Int'l Conf. on Comm. Theory, Reliablity, and Quality of Service*, 2009.

[11] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *Proc. of INFOCOM*, 2010.

[12] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, 2007.

[13] P. Jacquet, B. Mans, P. Muhlethaler, and G. Rodolakis, "Opportunistic routing in wireless ad hoc networks: upper bounds on the packet propagation speed," *IEEE MASS*, October 2008.

[14] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Trans. on Mobile Computing*, August 2009.

[15] F. De Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac, "A graph-based model for disconnected ad hoc networks," in *Proc. of INFOCOM*, 2007.

[16] Z. Kong and E. M. Yeh, "Connectivity and latency in large-scale wireless networks with unreliable links," in *Proc. of INFOCOM*, 2008.

[17] Z. Kong and E. M. Yeh, "Characterizaton of the critical density for percolation in random geometric graphs," in *Proc. of IEEE ISIT*, 2007.

[18] R. Zheng, "Information dissemination in power-constrained wireless networks," in *Proc. of INFOCOM*, 2006.

[19] Y. Xu and W. Wang, "The speed of information propagation in large wireless networks," in *Proc. of INFOCOM*, 2008.
[20] Z. Kong and E. Yeh, "On the latency for information dissemination in mobile wireless networks," in *Proc. of ACM MOBIHOC*, 2008.
[21] P. Jacquet, B. Mans, and G. Rodolakis, "Information propagation speed in mobile and delay tolerant networks," in *Proc. of INFOCOM*, 2009.
[22] P. Jacquet, B. Mans, and G. Rodolakis, "On space-time capacity limits in mobile and delay tolerant networks," in *Proc. of INFOCOM*, 2010.
[23] R. Meester and R. Roy, "Continuum percolation," *New York: Cambridge University Press*, 1996.
[24] R. J. La, "Distributed convergence of intermeeting times under the generalized hybrid random walk mobility model," *IEEE Transactions on Mobile Computing*, September 2010.
[25] D.J. Aldous, *Probability approximations via the Poisson clumping heuristic*, Springer-Verlag New York, 1988.
[26] H. Cai and D. Y. Eun, "Crossing over the bounded domain: from exponential to power-law intermeeting time in mobile ad hoc networks," *IEEE/ACM Transactions on Networking*, October 2009.
[27] K. Bradford, M. Brugger, S. Ehsan, B. Hamdaoui, and Y. Kovchegov, "Data loss modelling and analysis in partially-covered delay-tolerant networks," in *Proc. of ICCCN*, 2011.
[28] S. Karlin and H.M. Taylor, *A second course in stochastic processes*, Academic press, 1999.