On the hop count statistics for randomly deployed wireless sensor networks

Stefan Dulman*

EWI Faculty,
University of Twente,
The Netherlands
E-mail: dulman@cs.utwente.nl
*Corresponding author

Michele Rossi

Department of Information Engineering,
University of Padova, Italy
E-mail: rossi@dei.unipd.it

Paul Havinga

EWI Faculty,
University of Twente,
The Netherlands
E-mail: havinga@cs.utwente.nl

Michele Zorzi

Department of Information Engineering,
University of Padova, Italy
E-mail: zorzi@dei.unipd.it

Abstract: In this paper we focus on exploiting the information provided by a generally accepted and largely ignored hypothesis (the random deployment of the nodes of an ad hoc or wireless sensor network) to design improved networking protocols. Specifically, we derive the relationship between the number of hops separating two nodes and the physical distance between them (one- and two- dimensional topologies). In this way, distance estimates between nodes are made available without the use of any distance measuring hardware. We conclude the paper showing how the obtained statistical results can be applied to improve the performance of distributed localisation protocols while simplifying their implementation.

Keywords: hop count statistics; multi-hop transmission techniques; randomly deployed sensor networks; statistically enhanced localisation algorithms; wireless sensor networks.

Reference to this paper should be made as follows: Dulman, S., Rossi, M., Havinga, P. and Zorzi, M. (2006) 'On the hop count statistics for randomly deployed wireless sensor networks', *Int. J. Sensor Networks*, Vol. 1, Nos. 1/2, pp.89–102.

Biographical notes: Stefan Dulman is a Postdoctoral Researcher at the University of Twente. At the same time, he acts as the chief software architect at Ambient Systems. His research interests include various aspects of wireless sensor networks as localisation algorithms and real-time operating systems for embedded systems. He received a PhD in Computer Science from the University of Twente.

Michele Rossi received a Laurea degree in Electrical Engineering (with honours) and a PhD in Information Engineering from the University of Ferrara in 2000 and 2004, respectively. Since 2000, he has been a Research Fellow at the Department of Engineering of the University of Ferrara. In 2003, he was on leave at the Center for Wireless Communications (CWC) at the University of California San Diego (UCSD), where he did research on wireless sensor networks. In November 2005, he joined the Department of Information Engineering of the University of Padova, Italy, where he is currently an Assistant Professor. Currently, he is a part of the EU funded Ambient Networks and eSENSE projects. His research interests include, TCP/IP protocols over wireless networks, performance analysis of link layer retransmission techniques, routing and access selection in heterogeneous wireless networks and MAC/routing algorithms for wireless sensor networks.

Paul Havinga is a Professor in the Embedded Systems Group of the EEMCS Faculty at the University of Twente. His research interests include energy-efficient architectures and protocols, wireless communication networks, personal communication systems, wireless sensor networks, multimedia communication and reconfigurable systems. He received his PhD from the University of Twente. He led several European and Dutch research projects (as Eyes, Smart Surroundings, etc.).

Michele Zorzi received a Laurea degree and a PhD in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. After spending three years at UCSD, in 1998 he joined the University of Ferrara, Italy, and in 2003 the University of Padova, Italy, where he is currently a Professor. His research interests include performance evaluation in mobile communications systems, ad hoc and sensor networks and energy constrained communications protocols. His professional service includes a three-year term as Editor-In-Chief of the IEEE Wireless Communications Magazine and many editorial board memberships.

1 Introduction and related work

Ad hoc and wireless sensor networks have gained a lot of interest lately. Owing to technological breakthroughs, it is now possible to build small sized and energy efficient devices that can organise themselves in ad hoc networks and help monitor various parameters of the surrounding world. This new technology brings also new challenges: the algorithms running in these networks have to be designed by taking into consideration the limited resources available, the large number of devices and the dynamics of the environment.

Owing to the lack of resources and computation power, each hypothesis on which the design of a sensor network is based needs to be taken into account. No matter how small is the amount of information brought, it needs to be exploited from the beginning, because rediscovering that information after the deployment of the network using distributed algorithms can simply be too expensive.

In this paper, we focus on exploiting the information provided by a generally accepted and largely ignored hypothesis: the random deployment of the nodes of a wireless sensor network. Although the vast majority of studies acknowledge this fact, only a few of them make use of the properties, in terms of connectivity statistics, of these networks. There is a close relationship between the density of the nodes in such a network (that is known at deployment time and can be easily checked and updated during the run time of these networks), the number of hops separating two nodes and the distance between these two nodes. Using statistical techniques we derive distance estimates between the nodes without using any sort of distance measuring hardware. This information can be further used to enhance and simplify various networking protocols (e.g. localisation protocols).

In the following we analyse the underlying statistics for uniformly random deployed ad-hoc networks in one- and two-dimensional cases and derive the relationships between the number of hops and the distance separating two nodes. For the one dimensional case, we give an exact recursive formula that can be used for the calculation of various parameters (e.g. the mean distance between two nodes separated by a known number of hops), while for the two dimensional case we propose several approximations. Finally, we show that applying these results in existing networking protocols (in particular in localisation algorithms), not only reduces their overhead but also leads to significantly improved results.

A substantial effort has already been put into characterising ad hoc networks from a statistical point of view. Kleinrock and Silvester (1978) derive an approximation for the expected progress in each hop for a two dimensional network deployed within a circular region. Moreover, they also obtain approximations for the average distance between any two nodes. Bettstetter and Eberspacher (2003) derive closed form formulas for the probability that two nodes can communicate within one or two hops. The nodes are deployed in a rectangular area and the results are based on the distribution of the distance between two random nodes in such a scenario, which was originally derived by Ghosh (1951). For hop counts larger than 2, simulation results are presented and analysed. Lower and upper bounds for the hop count distribution between two random nodes are derived based on the properties of the network as the node density grows to infinity. In Miller (2001), the author investigates the distance between two randomly positioned nodes, where the network scenario is represented by nodes distributed inside a rectangular region. Both the uniform and the Gaussian node placements are analysed. Also, it is shown that the shapes of the distance distributions for these scenarios are very similar when the width of the rectangular area in the first case is taken to be about three times the standard deviation of the distribution in the second scenario. Mullen (2003) continues the analysis, exploring the underlying spatial distributions in more depth. Mobility of the nodes is also taken into account, as an extension of the work done in Bettstetter and Wagner (2002). Average hop distance is investigated for other network topologies as well. For example, in Rose (1992) the average hop distance is investigated for ring and street network topologies. However, while most of the work done so far focuses on the calculation of first and second order moments, we instead investigate the complete statistics of the distance from a given sourcedestination pair once their hop count is known. In addition, in the last part of this paper, we advocate the exploitation of these statistics to improve the performance of existing network protocols.

The paper is organised as follows: Section 2 presents the hop count statistics for the One-dimensional (1D) scenario. The analysis continues with the Two-dimensional

(2D) scenario in Section 3. All the results we obtain are confirmed by simulations. In Section 4, we show how to exploit the obtained information to improve and simplify existing localisation protocols for randomly deployed ad hoc and sensor networks. Our conclusions are drawn in Section 5.

2 Hop count statistics in the 1D case

2.1 One-dimension network scenario

We model the one-dimensional network scenario through a set of nodes placed on the positive side of the real x axis, where positions are picked according to a one-dimensional Poisson process with known parameter λ (Stoyan et al., 1995) Among these nodes, we consider one device, denoted here as *source*, that is the originator of the hop count distribution procedure and that is placed at the origin of the real axis. The probability of having exactly k nodes within a given interval of length l is: $\text{Prob}(k) = e^{-\lambda l}(\lambda l)^k/k!$. Further, the distance between the *source* node and the first node to its right is a random variable (r.v.) X characterised by the cumulative distribution function (cdf) $F_X(x) = 1 - e^{-\lambda x}$ (Papoulis and Pillai, 2002). Finally, we assume that each node can communicate to all its neighbours whose distances are smaller than or equal to R, where R denotes the (constant) transmission range of the nodes.

2.2 Limits of the hop count intervals

We say that a node has a Hop Count (HC) equal to n, $n \in \mathbb{N}$, if and only if it is n hops away from the *source* (the shortest path between the two nodes has length n hops). Moreover, we say that a given node with HC $n \geq 0$ and in position $\overline{x} \geq 0$ is the *last node* in its HC if all nodes in $x > \overline{x}$ have a larger hop count value. Let $\Omega = (\tau_0, \tau_1, \tau_2, \ldots, \tau_n, \ldots)$ denote the ordered positions of the nodes, $\tau_i \in \mathbb{R}^+$. Moreover, let $\Omega' = (\xi_0, \xi_1, \xi_2, \ldots, \xi_m, \ldots)$ be the ordered set containing the positions of the last nodes in each HC, where ξ_i is the position of the last node in HC i. Observe that $\Omega' \subseteq \Omega$ and that the *source*, placed in $\tau_0 = \xi_0 = 0$, is considered to be the last node in HC 0. Moreover, it can be easily verified that a node belongs to HC n if and only if it is within the transmission range of the last node in HC $n-1$ and is not in the transmission range of the last node in HC $n-2$. Exceptions to this rule are HC 0 (that contains the source node only) and HC 1 (that contains all the nodes in the transmission range of the source). On the basis of these observations, we can write the following inequalities:

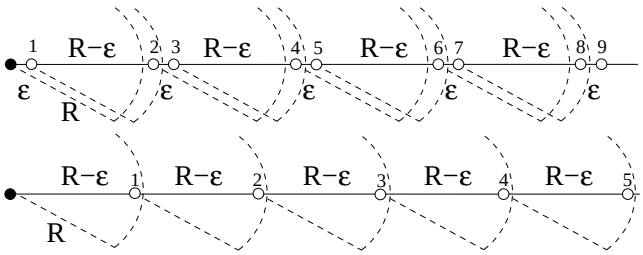
$$\begin{aligned} \xi_0 &= 0 \\ \xi_0 < \xi_1 &\leq \xi_0 + R \\ \xi_0 + R < \xi_2 &\leq \xi_1 + R \\ &\cdots \\ \xi_{n-2} + R < \xi_n &\leq \xi_{n-1} + R \\ &\cdots \end{aligned} \tag{1}$$

Now, we are interested in finding the boundaries of the interval containing all the feasible positions for those nodes in HC n. The inferior ($\mathcal{D}_{\min}$) and superior ($\mathcal{D}_{\max}$) limits of the interval containing all the nodes with HC n are found from (1):

$$\begin{aligned} \mathcal{D}_{\min}[n] &= \left\lfloor \frac{n}{2} \right\rfloor R \\ \mathcal{D}_{\max}[n] &= nR \end{aligned} \tag{2}$$

This result can be also justified by simple geometrical observations (see Figure 1). The fact that the distance between the two consecutive points ξ_i and ξ_{i+2} needs to be larger than or equal to R, leads to the inferior limit. The superior limit comes from the fact that ξ_i and ξ_{i+1}, in the limiting case, can be spaced by at most R.

Figure 1 Inferior (top) and superior (bottom) limits for the hop intervals



To proceed with our analysis, we introduce the following notation: let $\Gamma_{l_1}^{l_2}(x)$, $l_1, l_2 \in \mathbb{R}$, $l_1 < l_2$, be defined as:

$$\Gamma_{l_1}^{l_2}(x) = \begin{cases} 1 & \text{for } l_1 < x \leq l_2 \\ 0 & \text{otherwise} \end{cases}$$

2.3 Punctual distribution function of ξ_n

For each hop count $i \in \mathbb{N}$, we define a random variable E_i denoting the position of the last point in that hop. In the following, we compute the probability density functions (pdfs) for these random variables, $f_{E_i}(\xi_i) = f(\xi_i)$, where the joint densities of $E_1, E_2, \ldots, E_n$ is indicated as $f(\xi_1, \xi_2, \ldots, \xi_n)$. By the definition of conditional pdf we can write the following relations:

$$\begin{aligned} f(\xi_0) &= \delta(\xi_0) \\ f(\xi_1, \xi_0) &= f(\xi_1|\xi_0) f(\xi_0) \\ &\cdots \\ f(\xi_n, \xi_{n-1}, \ldots, \xi_0) &= f(\xi_n|\xi_{n-1}, \ldots, \xi_0) \cdots f(\xi_1|\xi_0) f(\xi_0) \end{aligned} \tag{3}$$

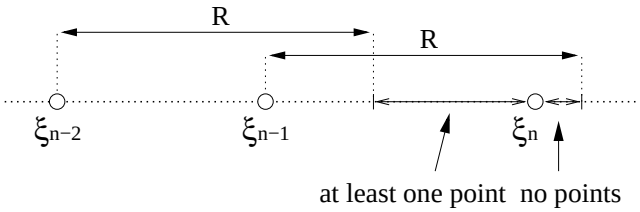
where $\delta(x)$ is the Dirac impulse and the above densities make sense in the intervals where the nodes can actually be placed, that is, according to the constraints in Equation (1) otherwise the functions equal 0. Moreover, due to the structure of the problem, $\forall\, n \in \mathbb{N}$, $n \geq 2$, when ξ_{n-1} and ξ_{n-2} are given ξ_n

does not depend on ξ_i with $i < n-2$. We can rewrite the set of Equation (3) as:

$$\begin{aligned} f(\xi_0) &= \delta(\xi_0) \\ f(\xi_1, \xi_0) &= f(\xi_1|\xi_0) f(\xi_0) \int_R^{\xi_0+R} \Gamma(\xi_1) \\ &\dots \\ f(\xi_n, \xi_{n-1}, \xi_{n-2}) &= f(\xi_n|\xi_{n-1}, \xi_{n-2}) f(\xi_{n-1}, \xi_{n-2}) \int_{\xi_{n-2}+R}^{\xi_{n-1}+R} \Gamma(\xi_n) \end{aligned} \quad (4)$$

The marginal density function $f(\xi_n)$ can therefore be computed by double integration of $f(\xi_n, \xi_{n-1}, \xi_{n-2})$ with respect to ξ_{n-2} and ξ_{n-1}. The conditional density $f(\xi_n|\xi_{n-1}, \xi_{n-2})$ can be computed based on the fact that, in order for ξ_n to be the position associated with the last node in hop count n, the following two events *have to be jointly verified* (see Figure 2).

Figure 2 Feasible region for ξ_n once ξ_{n-2}, and ξ_{n-2} are given



A) 'The interval $(\xi_{n-2} + R, \xi_n]$ is not empty' (from Equation (1) we have that $\xi_n > \xi_{n-2} + R$).

B) 'There are no points in the interval $(\xi_n,\ \xi_{n-1} + R]$'.

The events A and B are independent because the points (nodes) are placed according to a Poisson process (memoryless). We can therefore write that:

$$\begin{aligned} \text{Prob(A)} &= \text{Prob}\{k > 0 \text{ in } (\xi_{n-2} + R, \xi_n]\} \\ &= 1 - e^{-\lambda(\xi_n - \xi_{n-2} - R)} \\ \text{Prob(B)} &= \text{Prob}\{k = 0 \text{ in } (\xi_n, \xi_{n-1} + R]\} \\ &= e^{-\lambda(\xi_{n-1} + R - \xi_n)} \end{aligned} \quad (5)$$

where k is the number of nodes. We are now in the position of deriving the $F(\xi_n|\xi_{n-1}, \xi_{n-2})$ cdf and its derivative $f(\xi_n|\xi_{n-1}, \xi_{n-2})$:

$$\begin{aligned} F(\xi_n|\xi_{n-1}, \xi_{n-2}) &= \text{Prob(A)Prob(B)} \int_{\xi_{n-2}+R}^{\xi_{n-1}+R} \Gamma(\xi_n) \\ f(\xi_n|\xi_{n-1}, \xi_{n-2}) &= \lambda\, e^{-\lambda(\xi_{n-1} + R - \xi_n)} \int_{\xi_{n-2}+R}^{\xi_{n-1}+R} \Gamma(\xi_n) \end{aligned} \quad (6)$$

From Equations (4) and (6), the distribution function $f(\xi_n)$ can be obtained by induction as reported in Equation (7) at the top of the next page. In other words, we have decomposed $f(\xi_n)$ as the product of two terms, where the first term accounts for the network connectivity (λ), whereas the second term is independent of λ and for a given ξ_n, accounts for the feasible region of the remaining $n-1$ points $\xi_1, \dots, \xi_{n-1}$, according to the constraints in Equation (1). More compactly, $f(\xi_n)$ can be written as:

$$f(\xi_n) = \lambda^n e^{-\lambda(nR - \xi_n)} \times G_n(\xi_n) \quad (9)$$

where $G_n(x)$ replaces the succession of integrals in Equation (7). Moreover, $G_n(x)$ can be expressed as (see Section 2.4):

$$G_n(x) = \sum_{i=\lfloor \frac{n}{2} \rfloor}^{n-1} \left\{ g_n^i(x) \int_{iR}^{(i+1)R} \Gamma(x) \right\} \quad (10)$$

where the $g_n^i(x)$ are polynomials with the following properties (easily verified by inspecting the number of integrals, the terms that are being integrated and the continuity properties of the punctual distribution functions):

1 the coefficients of $g_n^i(x)$ do not depend on λ

2 $g_n^i(x)$ is a polynomial of order $n-1$

3 $g_n^i((i+1)R) = g_n^{i+1}((i+1)R)$.

Computing the distribution of the position of the last node in hop n makes sense only in the cases where the structure is connected up to that hop (i.e. there are no *connectivity holes* larger than R up to the end of hop n). Observe that, for a given hop number n, the probability of having a connected structure is a constant quantity that is used to scale the joint punctual distribution function $f(\xi_n)$ to obtain the conditional distribution $f_c(\xi_n) = f(\xi_n|\,\text{Structure connected up to hop } n)$. Formally, $f_c(\xi_n)$ is written as

$$f_c(\xi_n) = \frac{\lambda^n e^{-\lambda(nR - \xi_n)} \times G_n(\xi_n)}{\int_{\lfloor n/2 \rfloor R}^{nR} \lambda^n e^{-\lambda(nR - \xi_n)} \times G_n(\xi_n)\, \mathrm{d}\xi_n} \quad (11)$$

As the calculation of the integral in Equation (11) is rather tedious, we approximate it as follows. The problem of having a connected structure in the one dimensional case has been studied in Dousse et al. (2002), where the authors derived a close formula (Equation (8)) to obtain the probability $P_c(x)$ of having a connected structure up to any position $x \in \mathbb{R}^+$. The probability of having a connected structure up to the end of hop n can be approximated as $P_n = P_c(nR)$. $f_c(\xi_n)$ can therefore be approximated as

$$f_c(\xi_n) \approx \frac{\lambda^n e^{-\lambda(nR - \xi_n)} \times G_n(\xi_n)}{P_n} \quad (12)$$

Equation (12) gives the distribution of the last node in hop n. It is composed by two parts: an exponential function depending on the parameter λ and a set of polynomials whose coefficients are *independent* of λ. This formula is the central point for the analytical characterisation of the hop count statistics for the one dimensional case. In Figure 4 we plot the distribution calculated by means of Equation (12) against simulation points. The obtained results confirm the accuracy of the approximation.

$$f(\xi_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi_n, \xi_{n-1}, \xi_{n-2})\, \mathrm{d}\xi_{n-2}\, \mathrm{d}\xi_{n-1}$$
$$= \lambda^n e^{-\lambda(nR-\xi_n)} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\xi_{n-1}+R}_{\xi_{n-2}+R} \Gamma(\xi_n)\cdots \int_{-\infty}^{\xi_3+R}_{\xi_2+R} \Gamma(\xi_4) \int_{-\infty}^{\xi_2+R}_{\xi_1+R} \Gamma(\xi_3)\, \Gamma^{\xi_1+R}_{R}(\xi_2)\Gamma^{R}_{0}(\xi_1)\, \mathrm{d}\xi_1\, \mathrm{d}\xi_2 \cdots \mathrm{d}\xi_{n-2}\, \mathrm{d}\xi_{n-1} \quad (7)$$

$$P_c(x) = \begin{cases} 1 & x < R \\ \sum_{i=0}^{\lfloor \frac{x}{R} \rfloor} \left\{ \frac{[-\lambda e^{-\lambda R}(x-iR)]^i}{i!} \right\} - e^{-\lambda R} \sum_{i=0}^{\lfloor \frac{x}{R} \rfloor - 1} \left\{ \frac{[-\lambda e^{-\lambda R}(x-(i+1)R)]^i}{i!} \right\} & x \geq R \end{cases} \quad (8)$$

2.4 *Determination of the $G_n(x)$ coefficients*

In this section, we address the computation of the coefficients of the $g_n^i(x)$ polynomials that compose $G_n(x)$. As one can notice by inspecting Equation (7), the coefficients of $g_n^i(x)$ are functions of the transmission range R. This means that, for every particular scenario, they should be computed separately, as the transmission range R is unlikely to be the same in all cases. Fortunately, there is a way to avoid this. In fact, in Meester and Roy (1996), it has been proved that for a network whose nodes are distributed according to a Poisson process with parameter λ and R is the value of the transmission range, one can vary R and modify λ, such that the properties of the underlying graph remain unchanged (the reverse situation where λ is varied while adapting R is also true). In more detail, if the distances are scaled by a ratio r/r' in the d-dimensional space, the λ parameter must be modified according to $\lambda' = (r/r')^d\lambda$.

In what follows, we consider the transmission range to be $R = 1$. Any particular application can then scale its distances such that the transmission range becomes 1 and vary λ accordingly. This means that any application could then use the coefficients of the polynomials as computed below at no additional expense. In other words, this makes the coefficients of $g_n^i(x)$ *constants independent of the particular deployment scenario* (R and λ). To keep the notation simple, in the following we consider that we already modified R to 1 such that λ is the new value obtained after the normalisation step. This will not have any influence on the final results.

Moreover, we observe that the Γ functions only take values in $\{0, 1\}$ and this means that their only influence is on the intervals over which the integration in Equation (7) takes place. In the form in which it is presented, Equation (7) cannot be integrated in close form as the integration variables are all linked together. To make this integration possible in numerical form, we adopt the following three–step procedure:

Step 1: From the definition of the Γ functions and keeping in mind that the variables we use conform to the inequalities described in Equation (1) we can write the following identity:

$$\Gamma^{\xi_{i-1}+1}_{\xi_{i-2}+1}(\xi_i) = \left[\Gamma^{\xi_{i+2}-1}_{\xi_{i+1}-1}(\xi_i)\Gamma^{i}_{\lfloor \frac{i}{2} \rfloor}(\xi_i)\right] \cdot \left[\Gamma^{\xi_{i+1}-1}_{\xi_i - 1}(\xi_{i-1})\Gamma^{i-1}_{\lfloor \frac{i-1}{2} \rfloor}(\xi_{i-1})\right] \cdot \left[\Gamma^{\xi_i - 1}_{\xi_{i-1}-1}(\xi_{i-2})\Gamma^{i-2}_{\lfloor \frac{i-2}{2} \rfloor}(\xi_{i-2})\right]$$

where $i \in \mathbb{N},\ i = 1, 2, \ldots, n$. Therefore, we can apply the previous transformation to each Γ function in $G_n(x)$. Observe that, in Equation (7), the terms ξ_{n+2} and ξ_{n+1} do not exist, their place being taken by their extreme values as defined in Equation (2).

Step 2: One can observe that each integration has as limits integer numbers. More than this, the integration can be performed by splitting the domain of the variable ξ_i into $(i - \lfloor i/2 \rfloor)$ disjoint integrals where the Γ functions to be integrated are evaluated; each integral corresponds to one interval of size 1 and the union of these disjoint intervals covers the valid region of existence $(\lfloor i/2 \rfloor, i)$ for ξ_i.

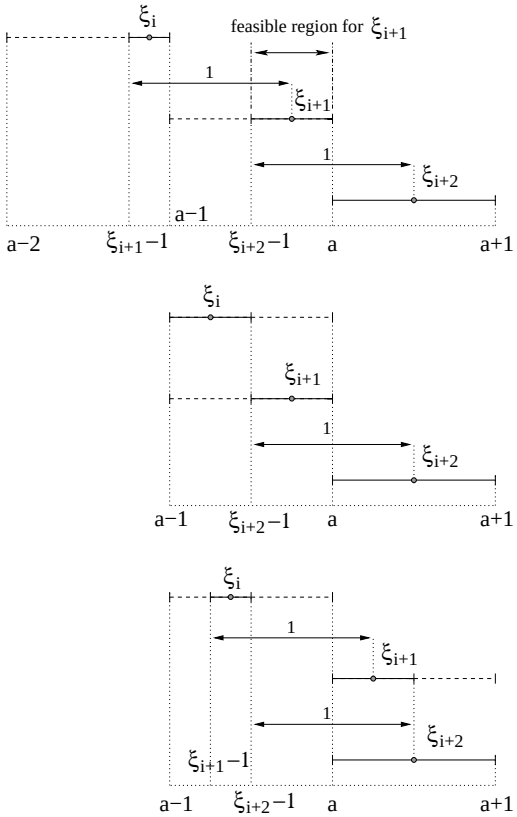
Step 3: For what concerns the combination of any three subsequent variables of the kind $(\xi_{i+2}, \xi_{i+1}, \xi_i)$, only one out of the three situations presented in Figure 3 is possible. That is, if a is an integer such that: $\lfloor (n+2)/2 \rfloor \leq a < n+2$, then:

1 $\xi_{i+2} \in (a, a+1) \quad \xi_{i+1} \in (a-1, a) \quad \xi_i \in (a-2, a-1)$

2 $\xi_{i+2} \in (a, a+1) \quad \xi_{i+1} \in (a-1, a) \quad \xi_i \in (a-1, a)$

3 $\xi_{i+2} \in (a, a+1) \quad \xi_{i+1} \in (a, a+1) \quad \xi_i \in (a-1, a)$

Figure 3 Possible configurations for the last nodes in each hop. $a \in \mathbb{N}$ is such that $\lfloor (n+2)/2 \rfloor \leq a < n+2$



The integration limits (feasible regions) in the previous three cases are accounted for by the following expressions (see Figure 3):

$$\begin{array}{ll} 1 & \Gamma\limits_{a}^{a+1}(\xi_{i+2})\Gamma\limits_{\xi_{i+2}-1}^{a}(\xi_{i+1})\Gamma\limits_{\xi_{i+1}-1}^{\xi_{i+2}-1}(\xi_i) \\ 2 & \Gamma\limits_{a}^{a+1}(\xi_{i+2})\Gamma\limits_{\xi_{i+2}-1}^{a}(\xi_{i+1})\Gamma\limits_{a-1}^{\xi_{i+2}-1}(\xi_i) \\ 3 & \Gamma\limits_{a}^{a+1}(\xi_{i+2})\Gamma\limits_{a}^{\xi_{i+2}}(\xi_{i+1})\Gamma\limits_{\xi_{i+1}-1}^{\xi_{i+2}-1}(\xi_i) \end{array} \tag{13}$$

By applying the three observations described above, one can perform the integration in a recursive manner. As an example, in Table 1 we report the $g_n^i(x)$ parameters for $n = 1, 2, \ldots, 6$. Figure 4 shows the punctual distribution function for the last node in each hop, for the first ten hops and for $\lambda = 6.7$. The theoretical distributions are plotted against simulation points.

2.5 Relation hop count number – distance statistics

In the following, we derive the relationships between the HC number and the actual position of any given node. The basic assumption is that we are dealing with a connected topology up to and including the node in question. Assume that a node finds out the number of hops it is separated from the source node. This information is enough for it to get an estimate of the distance between the two points. From the distribution of the distance, the node can compute the average distance and its associated standard deviation and use it, for example, in localisation algorithms, routing protocols, etc.

The distribution function of the distance conditioned on the number of hops has particular expressions for the case when the node lies in the first and second hop. For the third hop on, we provide a recursive formula for the computation of the distribution. In the following, by ξ_i we understand the position of the last node in hop i. By x_i we understand the position of any node in hop i. The punctual distribution function for the case where the node lies in the first hop (f_{d1}) is given by

$$f_{d1}(x_1) = \frac{1}{R}\Gamma\limits_{0}^{R}(x_1). \tag{14}$$

This result is quite obvious: a node is in the first hop if it is in the transmission range of the *source*. Within this interval, due to the fact that we are dealing with a Poisson process, the node has an equal probability of being at any given position, thus a uniform distribution. Assume now that a node belongs to the second hop. We can write the following expressions:

$$\begin{aligned} f(x_2|\xi_1) &= \frac{1}{\xi_1}\Gamma\limits_{R}^{\xi_1+R}(x_2) \\ f(\xi_1, x_2) &= f(x_2|\xi_1)f(\xi_1) \\ f_{d2}(x_2) &= \int_{-\infty}^{\infty} f(\xi_1, x_2)\, \mathrm{d}\xi_1 \end{aligned} \tag{15}$$

Node x_2 belongs to the second hop and this means that it is in the transmission range of the last node in the first hop and out of the transmission range of the *source*. Moreover, given that the position of the last node in the first hop is ξ_1, the node in second hop can belong only to the interval $(R, \xi_1 + R]$. Its distribution is again a uniform distribution due to the uniform deployment of the nodes. We can now compute $f(\xi_1, x_2)$ and integrate it over ξ_1 to get the marginal distribution $f_{d2}(x_2)$ (the expression for $f(\xi_1)$ has already been determined in Section 2.3). Performing these computations leads us to the following result:

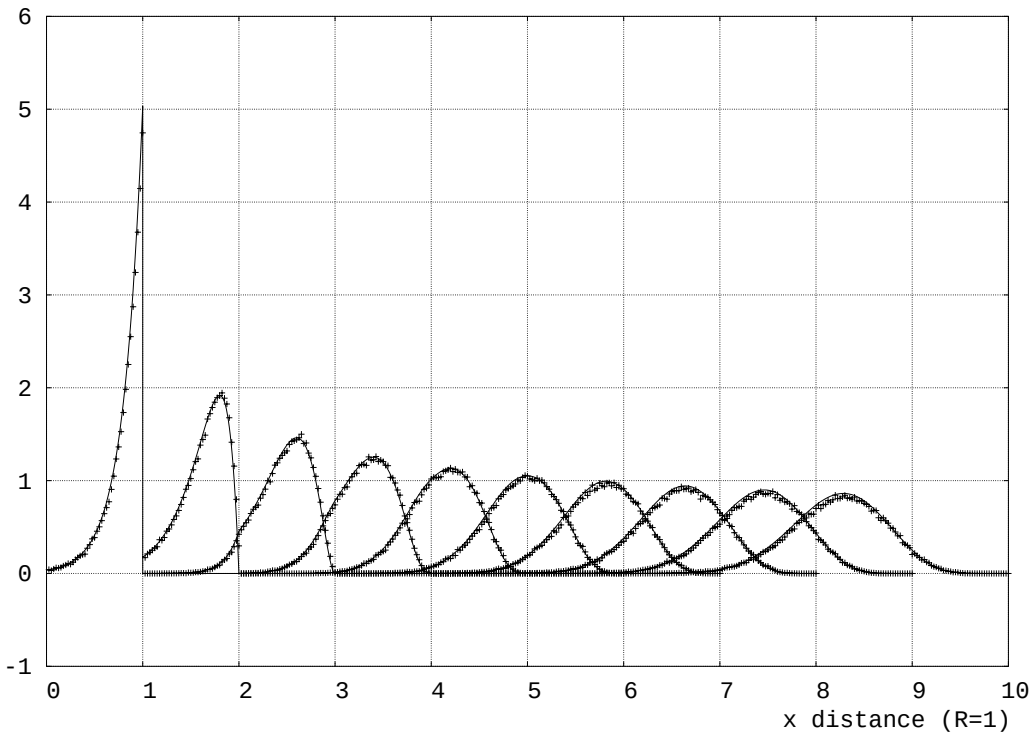
$$f_{d2}(x_2) = \lambda e^{-\lambda R}\int_{-\infty}^{+\infty}\Gamma\limits_{R}^{\xi_1+R}(x_2)\Gamma\limits_{0}^{R}(\xi_1)\frac{e^{\lambda\xi_1}}{\xi_1}\,\mathrm{d}\xi_1 \tag{16}$$

For the distributions of distances in hops further away than the second hop, we use the following reasoning (see Figure (2): a node belongs to hop n if it is in the transmission range of the last node in hop $n-1$ and is out of the transmission range of the last node in the $n-2$ hop. Therefore, within the interval $(\xi_{n-2} + R, \xi_{n-1} + R]$ the position of

Table 1 Coefficients of the $g_n^i(x)$ polynomials ($n = 1, 2, \ldots, 6$ and $i = \lfloor n/2 \rfloor, \ldots, n-1$)

$g_n^i(x)$	x^5	x^4	x^3	x^2	x^1	x^0
g_1^0	–	–	–	–	–	1.0000
g_2^1	–	–	–	–	−1.0000	2.0000
g_3^1	–	–	–	0.5000	−1.0000	0.5000
g_3^2	–	–	–	0.5000	−3.0000	4.5000
g_4^2	–	–	−0.3333	2.0000	−3.5000	1.6667
g_4^3	–	–	−0.1667	2.0000	−8.0000	10.6667
g_5^2	–	0.0833	−0.6667	2.0000	−2.6667	1.3333
g_5^3	–	0.1250	−1.5000	6.2500	10.1667	4.7083
g_5^4	–	0.0417	−0.8333	6.2500	−20.8333	26.0417
g_6^3	−0.0417	0.5833	−3.1667	8.3333	−10.6250	5.2500
g_6^4	−0.0333	0.6667	−5.0833	18.0000	−27.9583	12.7167
g_6^5	−0.0083	0.2500	−3.0000	18.0000	−54.0000	64.8000

Figure 4 Probability distribution functions of the last node in each hop ($n = 1, 2, \ldots, 10$, $\lambda = 6.7$)



the node is uniformly distributed. The marginal distribution $f_{dn}(x_n)$ is obtained as

$$f(x_n|\xi_{n-1}, \xi_{n-2}) = \frac{1}{\xi_{n-1} - \xi_{n-2}} \int_{\xi_{n-2}+R}^{\xi_{n-1}+R} \Gamma(x_n) \tag{17}$$

$$f(x_n, \xi_{n-1}, \xi_{n-2}) = f(x_n|\xi_{n-1}, \xi_{n-2}) f(\xi_{n-1}, \xi_{n-2})$$

$$f_{dn}(x_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_n, \xi_{n-1}, \xi_{n-2}) \, \mathrm{d}\xi_{n-2} \, \mathrm{d}\xi_{n-1}$$

The expression of $f(\xi_{n-1}, \xi_{n-2})$ has already been determined in Section 2.3. Numerical integration of these expression should follow the method described in Section 2.4 of rearranging the integration limits.

3 Hop count statistics in the 2D case

3.1 Network model

In the 2D case, we model the network as a graph $\mathcal{G} = (\mathcal{M}, \mathcal{L})$, where $\mathcal{M}$ is the set of nodes and $\mathcal{L}$ is the set of links between nodes. As in the one dimensional case, among the $|\mathcal{M}|$ nodes in $\mathcal{M}$, we consider a special device called *source*, which is the originator of the hop count distribution procedure. For what concerns connectivity, we adopt the Unit Disk model (Clark et al., 1991), where any two nodes can communicate if their distance is smaller than or equal to R, where R is the transmission range. Moreover, we assume that nodes are placed according to a planar Poisson process with density λ. Similarly to the one dimensional case, the probability to have n nodes within an area A is given by $\phi(n, A) = e^{-\lambda A}(\lambda A)^n/n!$ (Stoyan et al.,1995). The hop count value of a generic node $j \in \mathcal{M}$ is defined as the length, in terms of number of links, of the shortest path connecting node j to the source.

3.2 Introduction to the computation of hop count statistics for 2D random networks

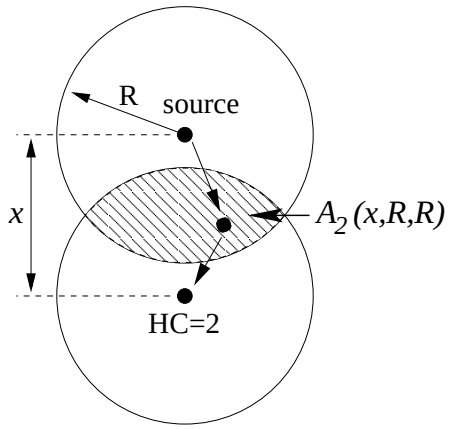
The aim of the analysis presented in the following is to find the HC distribution $f(n|x)$, defined as the probability that the hop count of a generic node $j \in \mathcal{M}$ is n given that its distance with respect to the source is x. Owing to the adopted connectivity model, $f(1|x)$ is one if $x \leq R$ and zero otherwise. The computation of the second hop distribution is slightly more complicated, but can be readily calculated considering the diagram in Figure 5. In fact, a target node placed at a distance x from the source has hop count equal to 2 if the common coverage area between the source and the target node contains at least one node. We refer to the common area between two circles at distance x as $A_2(x, R, R)$ (see Equation (19) below). Hence, the distribution $f(2|x)$ equals the probability that the area $A_2(x, R, R)$ (shaded area in Figure is not empty. Formally:

$$f(2|x) = \begin{cases} 0 & x > 2R \\ 1 - \phi(0, A_2(x, R, R)) & R < x \leq 2R \end{cases} \tag{18}$$

$A_2(x, R_1, R_2)$ is the area of the intersection of two circles of radius R_1 and R_2 whose centres are separated by x and is derived as:

$$A_2(x, R_1, R_2) = \begin{cases} 0 & R_2 \leq x - R_1 \\ \pi R_1^2 & R_2 \geq x + R_1 \\ \mathcal{A}_2(x, R_1, R_2) & \text{elsewhere} \end{cases} \tag{19}$$

Figure 5 Diagram for the computation of the second hop distribution



where

$$\mathcal{A}_2(x, R_1, R_2) = R_1^2 \cos^{-1}\left(\frac{x^2 + R_1^2 - R_2^2}{2xR_1}\right) + R_2^2 \cos^{-1}\left(\frac{x^2 + R_2^2 - R_1^2}{2xR_2}\right) - \frac{1}{2}\sqrt{[(R_1 + R_2)^2 - x^2][x^2 - (R_1 - R_2)^2]} \quad (20)$$

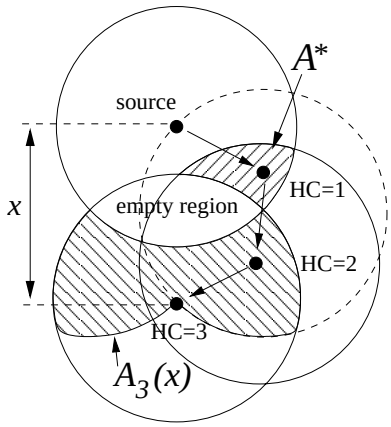
and assumed $R_1 \leq R_2$ (the other case can be obtained by simmetry). R_1 and R_2 are the transmission ranges of the source and the target node, respectively. In our case $R_1 = R_2 = R$. The diagram related to the third hop distribution $f(3|x)$ is shown in Figure 6. In this case, a node at distance x from the source has HC equal to 3 if the following conditions are jointly verified:

1 The intersecting region between the source and the target node coverage areas (empty region in Figure 6) must be empty in order to exclude all the cases where the target node has HC equal to 2.

2 Two further nodes must exist with the following properties. The first node (node with HC equal to 1 in Figure 6) must be in the transmission range of the source and must not be within the intersecting region between the source and the target node (empty region in Figure 6). Moreover, a further node (with HC equal to 2 in Figure 6) must exist in the transmission range of the target node and its position must be such that this node is also in the transmission range of the first node (but not of the source). In Figure 6 we report a possible scenario for this case. $A_3(x)$ is the feasible region for the second point. That is, the transmission area of a node placed within $A_3(x)$ has a non void intersection with the portion of the source transmission area where the first point can be placed (all coverage area minus the *empty region*, as specified above). Once the position for the second point has been fixed, the first point can be placed within an area A^* which is given by the intersection between the transmission area of the second point and the transmission area of the source minus the empty region.

The target node has hop count 3 if the two conditions above are verified, that is, if the minimum number of transmissions for a packet sent by the target node to get to the source is 3. As can be easily understood by the geometry of the problem, the resolution of this statistics is rather tedious. Moreover, due to the form of the involved integral it is also impossible to obtain it in exact close form (Niculescu and Nath, 2004a,b).

Figure 6 Diagram for the computation of the third hop distribution



In the next section, we present an approximate approach based on a Markov Chain representation of the underlying hop count assignment process, which is able to provide accurate approximations of the connectivity statistics.

3.3 A recursive algorithmic approach to the computation of hop count statistics

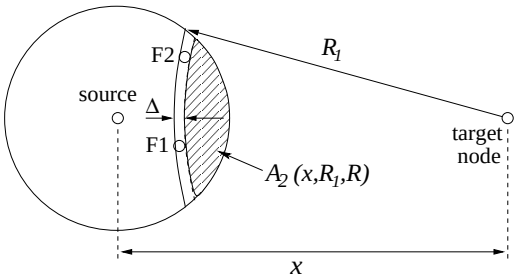
The hop count statistics is strictly related to the shortest hop path concept. The analysis that follows is mainly based upon this observation. In deeper detail, if a target node has hop count equal to n this means that the minimum number of intermediate nodes that connect it with the destination is $n - 1$. The key idea of our analysis is to build, starting at the source, a process that, at every step of the forwarding process, selects the set of nodes leading to the maximum advancement towards the target device. This concept can be better understood by referring to Figure 7, where we depict the first hop selection in our hop count assignment process. During this first step, we need to pick the nodes in the coverage area of the source that are the closest to the destination (target node at distance x). It shall be observed that, by geometric considerations, some nodes are equivalent for the purpose of the advancement. In particular, in the figure nodes $F1$ and $F2$ lead to the same advancement as they are both placed at the same distance R_1 from the target node. This equivalence is very important as it allows to model the hop count assignment process by only tracking the remaining distance towards the destination. The probability that the maximum advancement in the first step is, say a_1 ($R_1 = x - a_1$), given that the initial distance between the source and the target node is x, can be easily evaluated by multiplying the probability that no nodes are present within the intersecting area $A_2(x, R_1, R)$ (see Figure 7) by the probability that there is at least one node in the coverage

region of the source whose distance to the target node is R_1. For analytical tractability, we introduce the following two assumptions:

1 We subdivide the radio range R into a finite number of intervals N_Δ, of length $\Delta = R/N_\Delta$. This quantisation is introduced to enable a convenient and recursive calculation.

2 Only positive or zero advancements are possible. This assumption is equivalent to assuming that a node at a given distance x_1 can inherit its HC value only from nodes placed at a distance $x_2 \leq x_1$. This assumption is well verified in dense enough networks and therefore in most of the cases of practical interest. If this assumption is not verified this analysis can be seen as an approximation.

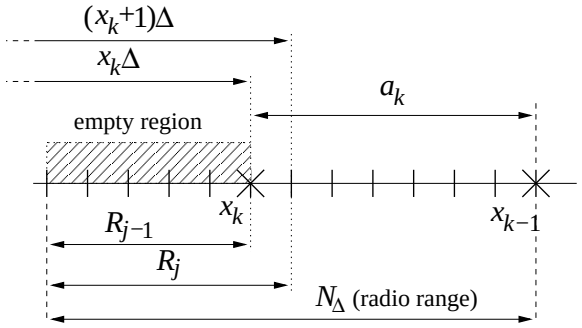
Figure 7 Circle intersection for the analysis depicting the first node selection stage of the hop count assignment process



In the following, we present in detail the recursive approach for the computation of the HC statistics. Due to the above assumptions, the HC statistics computed here is an approximation of the actual one. However, as will be shown in the following, the approach gives very accurate results. The approximate statistics is referred to as $f[n|x]$, $n \geq 1$, $n \in \mathbb{N}$ and x is the initial distance separating the source and the target node. The advancement process is tracked by means of a Markov Chain with a finite number of states. The generic state at the generic step k consists of the triple $\mathcal{X}(k) = (k, x_k, a_k)$, where k is the stage of the advancement process, that also corresponds to the HC assigned to the currently occupied node, x_k is the residual distance between the currently occupied node and the target node and a_k corresponds to the current advancement that is defined as $a_k = x_{k-1} - x_k$. As in Zorzi and Rao (2003) both a_k and x_k are represented in the quantised space, that is, their absolute values are given by $x'_k = x_k\Delta$ and $a'_k = a_k\Delta$, respectively. A graphical representation of the advancement process is shown in Figure 8.

As reported in Figure 8, we make a further assumption to track the position of a node. That is, we approximate the position of each node within a given interval $[x_k\Delta, (x_k+1)\Delta)$ with that of the leftmost edge of the interval itself, $x_k\Delta$. Another possibility would be to assign the rightmost edge. We prefer to adopt the first solution due to its superior ability to fit the actual statistics.

Figure 8 Model for the single hop advancement process



The probability mass distribution function of the advancement a_k achieved at step k, given the position x_{k-1} at step $k-1$, and the advancement a_{k-1} at the previous step $k-1$, is referred to as $f[a_k|x_{k-1}, a_{k-1}]$. Further details on the computation of this function are reported in Sections 3.4 and 3.5. For now, we take this function as given and use it into the computation of the statistics of interest. We refer to the joint probability of position and advancement at step k as $f[x_k, a_k]$. This distribution can be recursively evaluated as:

$$f[x_k, a_k] = \begin{cases} f[a_1|x_0 = 0, a_0 = 0] & k = 1 \\ \sum_{a_{k-1}=0}^{N_\Delta} \left\{ f[x_{k-1}, a_{k-1}] \right. & \\ \qquad \left. \times f[a_k|x_{k-1}, a_{k-1}] \right\} & k > 1 \end{cases} \tag{21}$$

where $x_k \in \{0, 1, \ldots, \lceil x/\Delta \rceil\}$, $a_k \in \mathcal{S}_a = \{0, 1, \ldots, N_\Delta\}$ and $x_k = x_{k-1} - a_k$.[1] $f[n|x]$ is found by summing all the contributions connecting the source to the target node in exactly n hops. Observe that each of these paths is a shortest hop path. This derives directly from the way in which the path is built, that is, by selecting, at every stage, the hop leading to the maximum advancement toward the target node.[2] $f[n|x]$ can be formally written as follows:

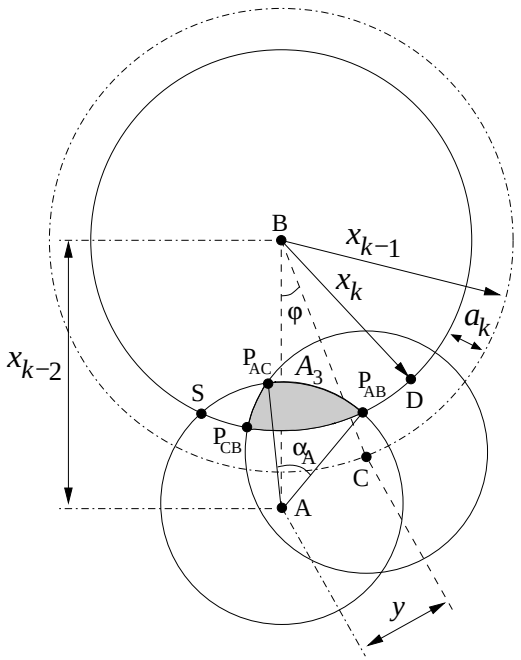
$$f[n|x] = \sum_{x_{n-1}=0}^{N_\Delta} \sum_{a_{n-1}=0}^{N_\Delta} f[x_{n-1}, a_{n-1}] \tag{22}$$

In the equation above, we sum all the terms that lead, in $n-1$ hops, to a residual distance between the $(n-1)$-st hop and the target node (nth hop) that is shorter than or equal to N_Δ, that is, within transmission range of the target node that is at a quantised distance of $\lceil x/\Delta \rceil$. Once the residual distance x_{n-1} between the current and target nodes is smaller than or equal to N_Δ, the current node is connected with the target with probability one and $f[n|x]$ is just obtained by summing the probabilities of these events to occur.

For the computation of the $f[a_k|x_{k-1}, a_{k-1}]$ function, we need to look at the forwarding process in more depth. In Figure 9, we report the related diagram. Here, we consider a forwarding sequence composed by three hops A, C and D. As clearly highlighted in the figure, the advancement at hop D (a_k) depends on the past forwarding history. In this example, for instance, A chooses C as its next hop. Of course,

this implies that the common region between the circle centered in A with radius R and the circle centered in B with radius x'_{k-1} is empty, that is, no relay nodes can be found in the area $A_2(x'_{k-2}, x'_{k-1}, R)$, where x' means $x\Delta$. Given that, at node D (step k), the area that was considered empty in the previous step $k-1$ (shaded area A_3 in the figure) *can not* be reconsidered for forwarding purposes. To sum up, the advancement at step k depends on the previous advancement at step $k-1$ and, in general, on all the previous advancements as well (it can be shown that for the two dimensional case, only the previous thirteen neighbors are of influence). Here, for tractability purposes, at the generic step k, we choose to consider the previous advancement at step $k-1$ only.

Figure 9 Detailed diagram for the advancement process



Note that area A_3 also depends on the angle φ reported in the figure. In fact, to fully specify the geometry of the three points constituting the advancement sequence we need to know x_{k-1}, a_{k-1}, a_k and the angle φ. In our Markov Chain $\mathcal{X}$, we precisely track the first three quantities, whereas the dependence on φ is accounted for in a statistical sense. The next two subsections are devoted to the presentation of two different approaches to obtain the $f[a_k|x_{k-1}, a_{k-1}]$ function and clarify the role of φ.

3.4 Computation of $f[a_k|x_{k-1}, a_{k-1}]$: approach I

In this first approach to the computation of $f[a_k|x_{k-1}, a_{k-1}]$ we neglect the dependence on the forwarding history two or more hops in the past. In this case, the forwarding process statistics is renewed at every forwarding step so that the forwarding decision is made independently of previously occupied locations. This is, of course, a rough approximation of the actual forwarding process, where indeed there may be a strong correlation between subsequent node selections. Nevertheless, as will be shown next, such an approximation becomes accurate as the node density increases.

Now, refer to the kth step of our forwarding process and consider that the position of the previous hop is x_{k-1}. Owning to assumption 2 of Section 3.3, the advancement at step k is constrained to be in the set $a_k \in \mathcal{S}_a = \{0, 1, \ldots, N_\Delta\}$. Considering the diagram in Figure 8, having a maximum advancement equal to a_k corresponds to having no nodes in the area delimited by R_{j-1} and at least one node in the interval $[x_k\Delta, (x_k+1)\Delta)$. This probability is found as a function of R_j and R_{j-1}, which can be written as $x_{k-1} - a_k + 1$ and $x_{k-1} - a_k$, respectively (see Figure 8). The probability that the advancement is a_k given x_{k-1} and a_{k-1} is derived as

$$f[a_k|x_{k-1}, a_{k-1}] = \begin{cases} \zeta[a_k|x_{k-1}, a_{k-1}] & a_k \in \mathcal{S}_a \\ 0 & \text{elsewhere} \end{cases} \tag{23}$$

where

$$\zeta[a_k|x_{k-1}, a_{k-1}] = e^{-\lambda\mathcal{A}_{j-1}} - e^{-\lambda\mathcal{A}_j} \tag{24}$$

and

$$\begin{aligned} \mathcal{A}_{j-1} &= A_2(x'_{k-1}, R, R'_{j-1}) \\ \mathcal{A}_j &= A_2(x'_{k-1}, R, R'_j) \\ R_j &= x_{k-1} - a_k + 1 \\ R_{j-1} &= x_{k-1} - a_k \\ x' &= x\Delta \end{aligned} \tag{25}$$

3.5 Computation of $f[a_k|x_{k-1}, a_{k-1}]$: approach II

In the approach presented in this subsection, we consider the dependence on the previous stage $(k-1)$ only. Our target is to compute the function $\zeta[a_k|x_{k-1}, a_{k-1}]$, which can be subsequently plugged into Equation (23) to obtain a better approximation for $f[a_k|x_{k-1}, a_{k-1}]$. As discussed above, to this end we need to account for the four quantities x_{k-1}, a_{k-1}, a_k as well as the angle φ. Note that φ is the only quantity that is not explicitly tracked by our Markov Chain. The knowledge of this angle is needed to compute the shaded area A_3 in Figure 9, that is, the common area arising from the intersection of the three circles centered in points A, B and C. Furthermore, the only quantities that we need to fully specify this area are x_{k-2}, x_{k-1}, φ, R and x_k. In the following, we refer to this area as $A_3(x_{k-2}, x_{k-1}, \varphi, R, x_k)$. As expected, the angle φ is the only unknown quantity in our analysis. In the next, we therefore proceed with its characterisation. From Figure 9 it is clear that, once x_{k-1} and x_{k-2} have been fixed, the feasible region for φ consists of the set $\varphi \in [-\varphi_{\max}, \varphi_{\max}]$, where:

$$\varphi_{\max} = \cos^{-1}\left[\frac{(x'_{k-2})^2 + (x'_{k-1})^2 - R^2}{2x'_{k-2}x'_{k-1}}\right] \tag{26}$$

where $x_{k-2} = x_{k-1} + a_{k-1}$ and $x' = x\Delta$.

Now, we refer to $p(\varphi)$ as the pdf of φ. The form of this function will be discussed later. For a given φ, the area $A_3(x_{k-2}, x_{k-1}, \varphi, R, x_k)$ is readily computed by geometric

arguments, whereas we express the probability mass function of the advancement, conditioned on φ, as $\eta[a_k|x_{k-1}, a_{k-1}, \varphi]$. $\zeta[a_k|x_{k-1}, a_{k-1}]$ can be evaluated as follows:

$$\zeta[a_k|x_{k-1}, a_{k-1}] = 2\int_0^{\varphi_{\max}} \eta[a_k|x_{k-1}, a_{k-1}, \varphi]\ p(\varphi)\ \mathrm{d}\varphi \tag{27}$$

Observe that the equation above holds as long as $\varphi_{\max} > 0$, for $\varphi_{\max} = 0$ it reduces to Equation (23) above. Hence, for a given (x_{k-1}, a_{k-1}) pair, the advancement mass function is found by statistically averaging over the admitted values of φ. $\eta[a_k|x_{k-1}, a_{k-1}, \varphi]$ is evaluated as

$$\eta[a_k|x_{k-1}, a_{k-1}, \varphi] = e^{-\lambda\mathcal{A}_{j-1}} - e^{-\lambda\mathcal{A}_j} \tag{28}$$

where:

$$\begin{aligned}
\mathcal{A}_{j-1} &= A_2(x'_{k-1}, R, R'_{j-1}) \\
&\quad - A_3(x'_{k-2}, x'_{k-1}, \varphi, R, R'_{j-1}) \\
\mathcal{A}_j &= A_2(x'_{k-1}, R, R'_j) - A_3(x'_{k-2}, x'_{k-1}, \varphi, R, R'_j) \\
x_{k-2} &= x_{k-1} + a_{k-1} \\
R_j &= x_{k-1} - a_k + 1 \\
R_{j-1} &= x_{k-1} - a_k \\
x' &= x\Delta
\end{aligned} \tag{29}$$

For what concerns the probability distribution function $p(\varphi)$, we consider here the simple case where φ is uniformly distributed in the interval $[-\varphi_{\max}, \varphi_{\max}]$. In this case $p(\varphi) = 1/(2\varphi_{\max})$. As will be shown in the next section, the results obtained with such an approach are in good agreement with simulation points, even for quite small values of λ.

3.6 Verification of the accuracy of the considered approach

In this section, we report some results to test the goodness of the methods described above. In Figures 10 and 11, we plot the hop count statistics $f[n|x]$ for $\lambda_n = \lambda\pi R^2 \in \{8, 15\}$ and $n \geq 2$. Simulation points are also reported for comparison. As can be shown in Figure 10 the second approach for the computation of the function $f[a_k|x_{k-1}, a_{k-1}]$ leads to a higher accuracy, whereas the first approach leads to overestimating the actual statistics. This is due to the fact that at every stage the full coverage area is made available independently on the past forwarding history and the complete statistics arises from the convolution of single stage advancement statistics.

However, as can be observed from Figure 11, as the node density increases, the two approaches become almost equivalent. Moreover, it is worth noting how the behaviour of the statistics suddenly changes for small probabilities. These regions are, in fact, the ones where the past forwarding history most affects the results. It can also be observed that the refined analysis is closer to the real behaviour of the curves thereby providing a better approximation. We note that an even better approximation can be achieved by further refining the $p(\varphi)$ function considered in the analysis in Section 3.5. Finally, once $f[n|x]$ has been computed, we observe that $f[x|n]$ can be easily found via Bayes' formula. Note that $f[n|x]$ can be used to estimate the geographical distance x between two nodes through maximum likelihood criteria when the hop count distance n is known. A first example in this sense will be given in section 4.

Figure 10 $f[n|x]$ for $\lambda_n = 8$

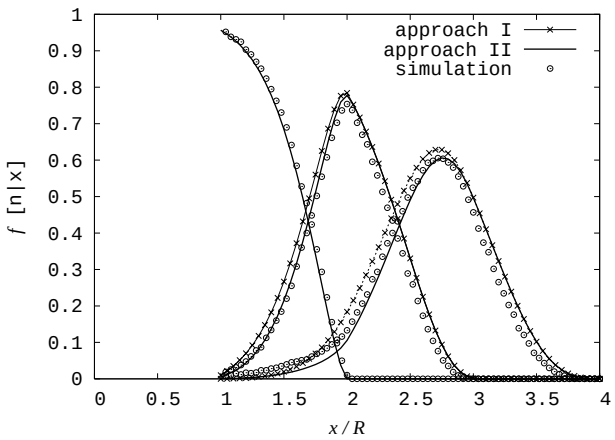
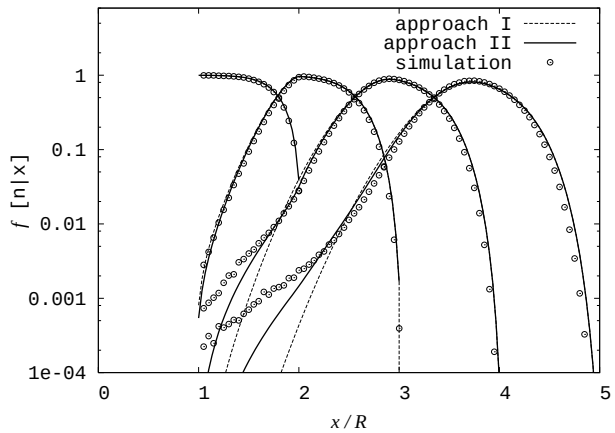


Figure 11 $f[n|x]$ for $\lambda_n = 15$



3.7 Approximation by means of maximum forwarding within radius routing

In this section, we report an alternative method to derive approximations for hop count related statistical values. In Bose et al. (2001) and Zorzi and Rao (2003) it has been shown that geographical routing provides good shortest path approximations. Refer now to Figure 12, where P is the originator of the hop count distribution procedure, whereas node Q is a target node with hop count n. Moreover, let us refer to the single hop progress as ξ and to the associated r.v. as E. If the distance between P and Q is large enough, we may neglect A_2 thereby having $A_1(\xi) + A_2(\xi) \approx A_1(\xi)$, where $A_1(\xi) = \arccos(\xi) - \xi\sqrt{1-\xi^2}$. In this case, the cdf of the maximum advancement $F_E(\xi)$ is achieved as Kleinrock

and Silvester (1978) $F_E(\xi) = \text{Prob}[E \leq \xi] = e^{-\lambda A_1(\xi)}$ and its pdf is found as:

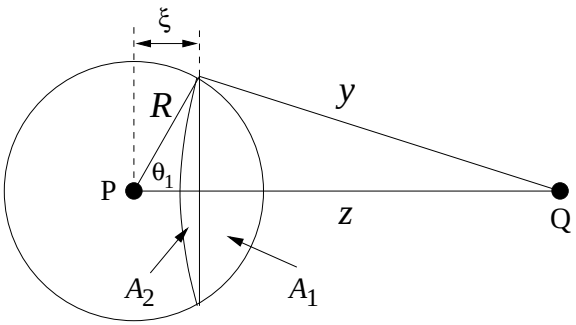
$$f_E(\xi) = 2\lambda\sqrt{1-\xi^2}e^{-\lambda A_1(\xi)} \tag{30}$$

Unfortunately, close form expressions for the mean $\overline{\xi}$ and the variance σ_ξ^2 of the advancement do not exist (Niculescu and Nath, 2004a,b). As in Niculescu and Nath (2004a,b), the mean and the variance of the distance between P and Q can be approximated by $\overline{\xi}(n) = n\overline{\xi}$ and $\sigma_\xi^2(n) = n\sigma_\xi^2$, respectively. The above method gives very accurate results for both mean and variance for high values of λ. However, the approximation fails for lower densities as:

1 the independence between subsequent advancements is no longer verified and

2 the means of Maximum forwarding within Radius (MFR) forwarding strategy is no longer a good approximation of the actual paths.

Figure 12 Single hop geographical advancement



4 Example application

In this section, as an example application of the results presented in this paper, we describe how they can be used to enhance existing localisation protocols. Firstly, we will use the results derived in Section 3 to compute a series of coefficients in the particular case of a random topology network. Secondly, we will modify the DVHop protocol (Niculescu and Nath, 2004a,b) to include these results and compare the performance of the new algorithm with the original one. For a more detailed description and further results on the applicability of these methods to other localisation protocols the reader is referred to Dulman and Havinga (2004).

4.1 Preliminaries

We note that a large number of publications (especially the ones related to localisation in wireless ad hoc or sensor networks) assume random topologies. Most of them, however, develop their solutions without taking advantage of the statistics governing the node positioning. We however think that such statistics, if available, can be successfully exploited to improve performance. Next, we give an example of this fact.

Consider a localisation example where a set of nodes is randomly deployed within a given geographical area. The position of a node can be determined based on distance estimates from other nodes that have already acquired their position (let us call these nodes *anchors*). As we have previously discussed, with a simple hop count assignment, all nodes can be informed about the distance, in number of hops, between them and any other node in the network. In particular, we might think of calculating the number of hops separating each node from a set of *anchors*, that are assumed to know their position with a sufficient precision. Note that the hop count information can be easily obtained, as both MAC and routing protocols either natively exploit it or can disseminate it together with data traffic at a low additional cost. Focusing on a given node, it is reasonable to assume that anchors further away should have a diminished influence than closer anchors when we have to derive position estimates through, for example, triangulation techniques. This is due to the fact that precision is affected by distance. The central point of our enhancement to DVHop is to reflect this intuition into the localisation protocol. In practice, we would like closer anchors to play a more determinant role in the position estimation procedure. As said above, any given node can obtain a list of hop counts separating it from each anchor in the network. Hence, exploiting the statistics for the 2D case and as a function of the hop counts only, we can further obtain the statistics of the distance separating a given node from each anchor. In particular, we can calculate the variance of the distance related to each hop count value (anchor) and use it to calculate a weight for that specific hop count (anchor). A possible way to calculate the weights is as follows: $w_i = 1/\sigma_i^2$, where i is the hop count value. In the next section, we show how this intuition can be used to enhance the localisation protocol.

4.2 Statistically enhanced localisation protocol

The DVHop protocol (Niculescu and Nath, 2004a,b) consists of three phases:

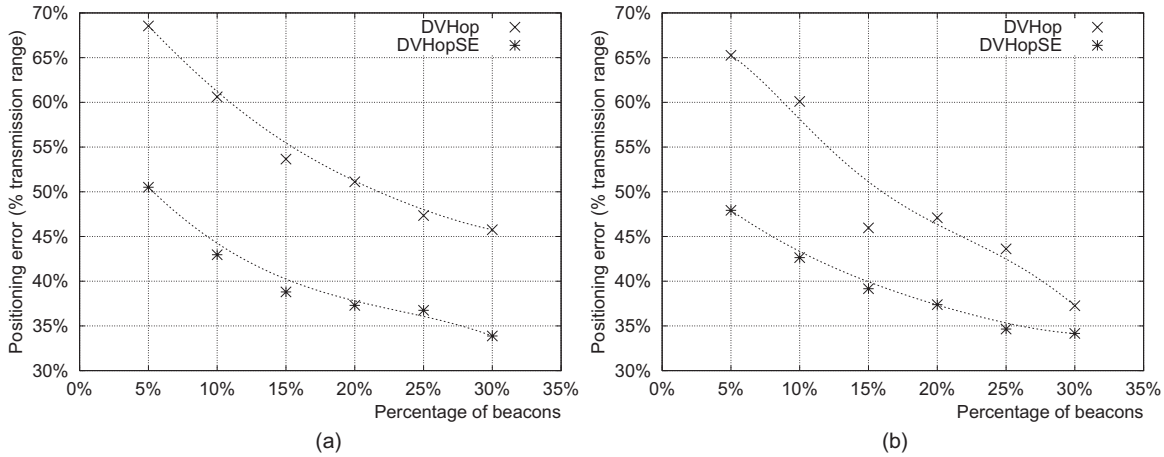
1 nodes get distances in number of hops to the anchors

2 each anchor computes an average hop length based on the number of hops towards the other anchors and

3 anchors distribute the average hop length and nodes are able to compute their position through triangulation.

By exploiting the results in the previous section, we can distribute the value of the density as part of the first phase of the algorithm. The number of hops towards the anchors are now enough for the nodes to compute their position. Instead of the Least Squares Method used when computing the position via triangulation, nodes can now make use of Weighted Least Squares Method with the weights as specified above. In other words, the new protocol (DVHopSE) will make use of only phases 1 and 3 from the previous description (leading to smaller traffic in the network) and use a set of precomputed weights.

Figure 13 Comparison DVHop - DVHopSE: (a) mean value of the positioning error and (b) standard deviation of the positioning error



Example results are presented in Figure 13. The setup consisted in 200 nodes randomly deployed in a square area (1x1 unit^2), with the nodes having a transmission range of 0.115 units leading to situations where each node had on average approximately seven neighbours (all simulations have been performed in Matlab).

The number of beacons was varied from 5% to 30% of the total number of nodes and it can be seen that an improvement between 22% and 29% is obtained in the mean value of the positioning error. The results given by DVHopSE are more precise, and also the associated standard deviation is improved (improvement spans from 8% to 22%).

The reason for this improvement lies in the fact that the closer anchors had a bigger influence in the calculation of position estimates than anchors being further away. DVHop computes hop estimates that best fit the local topology: in practice if nodes are denser in a certain part of the network, they will get a smaller value for the average hop distance than nodes in a less dense part of the network. DVHopSE averages these values to only one. Simulation results show that the weights related to how far the anchors are located play a bigger role than local densities. Similar results were obtained for distance-based localisation algorithms (called DVDistance, see Niculescu and Nath, 2004a,b), see, for example, Dulman and Havinga (2004). We note that the *random deployment* assumption, which is present in the vast majority of the scenarios considered so far in the literature, is mostly unexploited. However, the underlying random topology has certain properties that can actually be used within localisation protocols, instead of being 'rediscovered' at the cost of additional resources and time (as for phase 2 of the DVHop protocol).

5 Conclusions

Although the large majority of applications consider randomly deployed networks, only a few of them make use of the properties resulting from the underlying statistics, as most protocols try to 'rediscover' things that should be taken as granted. In this paper we focused on these underlying statistics in the case of uniform random deployment of networks in the one dimensional and two dimensional cases and derived the relationship between the number of hops and the distance separating two nodes of a uniformly distributed network. For the one dimensional case, we were able to give an exact recursive formula for the computation of various parameters, while for the two dimensional case we proposed several approximations. We have shown that applying these results in already existing algorithms, not only reduces their overhead but also leads to significantly better performance.

References

Bettstetter, C. and Wagner, C. (2002) 'The spatial node distribution of the random waypoint mobility model', *German Workshop on Mobile Ad Hoc Networks (WMAN02)*, Ulm, Germany.

Bettstetter, C. and Eberspacher, J. (2003) 'Hop distances in homogenous ad hoc networks', *Proceedings of IEEE Vehicular technologies conference(VTC)*, Korea.

Bose, P., Morin, P., Stojmenović, I. and Urrutia, J. (2001) 'Routing with guaranteed delivery in ad hoc wireless networks', *ACM Wireless Networks*, Vol. 7, pp.609–616.

Clark, B.N., Colbourn, C.J. and Johnson, D.S. (1991) 'Unit Disk Graphs', *Discrete Mathematics*, Vol. 86, Nos. 1–3, pp.165–177.

Dousse, O., Thiran, P. and Hasler, M. (2002) 'Connectivity in ad hoc and hybrid networks', in *IEEE Infocom*, Vol. 2, pp.1079–1088.

Dulman, S. and Havinga, P. (2004) 'Statistically enhanced localization schemes for randomly deployed wireless sensor networks', *Proceedings of DEST International Workshop on Signal Processing for Sensor Networks*, Melbourne, Australia.

Ghosh, B. (1951) 'Random distances within a rectangle and between two rectangles', *Bulletin Calcutta Mathematical Society*, Vol. 43, pp.17–24.

Kleinrock, L. and Silvester, J. (1978) 'Optimum transmission radii for packet radio newtorks or why six is a magic number', *Proceedings of IEEE National Telecommunications Conference*, Birminghan, Alabama, pp.4.3.1–4.3.5.

Meester, R. and Roy, R. (1996) *Continuum Percolation*, Cambridge University Press.

Miller, L. (2001) 'Distribution of link distances in a mobile network', *Journal of Research of the National Institute of Standards and Technology*.

Mullen, J. (2003) 'Robust approximations of, the distribution of link distances in a mobile network', *Mobile Computing and Communications Review*, Vol. 72, No. 2.

Niculescu, D. and Nath, B. (2004a) 'Error characteristics of ad hoc positioning systems (APS)', *Proceedings of MobiHoc '04*, Roppongi, Japan, May.

Niculescu, D.B. Nath, P. (2004b) 'Position and orientation in ad hoc networks', *Elsevier Journal of Ad Hoc Networks*, Vol. 2, No. 2, pp.133–151.

Papoulis, A. and Pillai, S. (2002) *Probability, Random Variables and Stochastic Processes*, 4th Edition, McGraw–Hill Higher Education.

Rose, C. (1992) 'Mean internodal distance in regular and random multihop networks', *IEEE Transactions on Communications*, Vol. 40, August.

Stoyan, D. Kendall, W. S. and Mecke, J. (1995) *Stochastic Geometry and its Applications*, 2nd Edition, John Wiley and Sons.

Zorzi, M. and Rao, R.R. (2003), 'Geographic random forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance', *IEEE Transactions on Mobile Computing*, Vol. 2, October–December, pp.337–348.

Notes

[1]See Figure 8. Note that the inequality $x_k \leq x_{k-1}$ holds as x_k is by definition the (quantised) distance that remains to be covered at stage k to reach the target node.

[2]This fact is taken into account by the distribution $f[a_k|x_{k-1}, a_{k-1}]$.