An Empirical Study on the Connectivity of Ad Hoc Networks

Ao Tang, Cédric Florens, Steven. H. Low
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125
626-395-8820
Email: aotang@caltech.edu

Abstract—This paper[1] discusses the probability of connectivity of ad hoc networks. An empirical formula is proposed to fit the simulation results. The parameters of the formula are determined for different cases and the asymptotic behavior is discussed. Finally, a new metric is proposed to quantify the connectivity of an ad hoc network.

TABLE OF CONTENTS

1. INTRODUCTION

A wireless ad hoc network consists of a collection of mobile nodes sharing a wireless channel without any centralized control or established communication backbone. Typical applications of ad hoc networks are scenarios where setting up a communication infrastructure is difficult (because of mobility) or very expensive (because of terrain).

An ad hoc network should be well connected in order to be useful. Due to limited transmission power, typically it is impossible for every two nodes to communicate with each other directly. So it may be necessary to relay a packet over multiple radio units to reach the destination. A prerequisite is that there is at least a path connecting every two nodes. Such a network is called fully connected.

One of the fundamental problems of ad hoc networks is connectivity. Part of the existing literature concentrates on the asymptotic behavior of connectivity when the number of nodes goes to infinity. In [1], it is proved that for the one-dimensional case, there exists a critical transmission range, which is the minimum transmission range of each node for the network to be connected with probability one, as the number of nodes in the network goes to infinity. The author also conjectured the critical transmission range for the two-dimensional case, which was studied by [2][3][4], and more recently by [5].

We are interested in the case where the number of nodes is finite, which is more valuable for engineering designing. To the best of our knowledge, there is still no exact solution to this problem. This has led to another approach to the problem where upper and lower bounds are derived for the probability. One can find quite a few results in [6]. Typically, these bounds are tight only when the probability that any two nodes can communicate directly is near 0 or 1, which is not the case in many applications.

In this paper, by simulation, we provide an empirical equation by fitting the data. The formula can be used to estimate the probability of connectivity for ad hoc networks, given node density, deployment surface area and intended transmission range. This is useful in designing or analyzing ad hoc networks.

2. BASIC MODEL AND SIMULATION METHOD

Consider an L-by-L square. We assume the transmitter density to be D. Therefore the number n of nodes is DL^2. Each node can communicate within a circle of radius R, which is called transmission range (Fig. 1). Furthermore we assume nodes are uniformly distributed over the area, which is a reasonable assumption in typical scenarios. In Fig. 1, a line connects two nodes that can communicate directly.

[2] IEEEAC paper #95

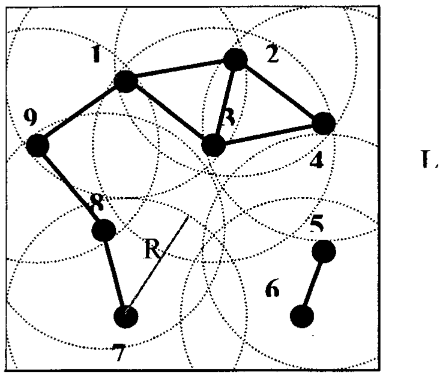


Fig.1. Instance of ad hoc network topology with n=9. It comprises two connected components.

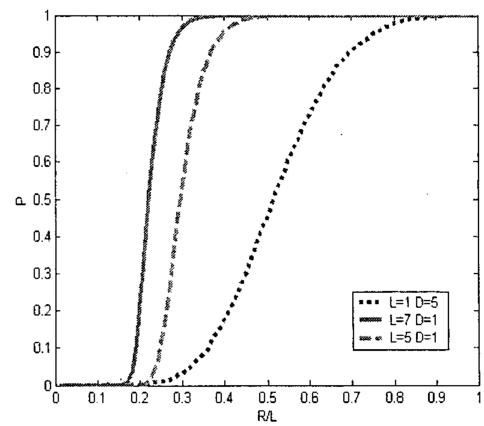


Fig. 2.Connectivity probability as a function of normalized transmission range for networks of size n=5, 25, 49.

We use Monte Carlo method to compute the connectivity probability P. For varying transmission range R ($0<R<L$), we randomly generate the locations of the n nodes according to a uniform distribution with density D, then determine whether they form a fully connected network or not. The standard shortest path algorithm is used to find the shortest path between every two nodes. If the length of the path is infinity, then we know the two nodes are not connected and the network is not fully connected. We repeat that process M times for each R. If the network is connected m times, then we say the connectivity probability for that R is m/M. We use M=20000 for our simulation, which can roughly guarantee accuracy to the order of 0.01.

In [7], the authors proved that the average throughput for each node decreases with the increasing number of users for a given transmission range. And they reached the conclusion that designers should target their efforts at networks for smaller number of users, rather than try to develop large wireless networks. For this reason, our simulation focuses on the case where n does not exceed 125.

3. SIMULATION RESULTS AND ANALYSIS

Typically, the probability curve looks as in Fig. 2. It is clear that the curves have generally the same shape. The more users, the sharper the transition from 0 to 1 looks. Through extensive simulations we found that, for $P \in [0.5,0.99]$

$$P \approx \frac{\exp(\frac{R-Rc}{E})}{1+\exp(\frac{R-Rc}{E})} \qquad (1)$$

where Rc and E are model parameters. [0.5, 0.99] is the range of most interest. Indeed when P is small, the network is not well connected, and when P is near 1, we can use the existing bounds to get a satisfactory approximation.

From (1), we get:

$$\frac{R-Rc}{E} \approx \ln(\frac{P}{1-P}) \qquad (2)$$

In order to determine Rc and E, we fit $\ln(P/(1-P))$ and R by a linear equation. The reciprocal of the slope gives E. And Rc is determined by the interception.

To get an idea on how Rc and E vary with D and L, we first fix D and vary L, and then vice versa. Table I and II are cases where D=1 and L=1, respectively. ρ is the linear correlation coefficient. We can see the linearity is very good.

Table I (D=1)

L	Rc	E	ρ
2	1.0964	0.1729	0.9965
3	1.3142	0.1735	0.9993
4	1.4096	0.1757	0.9997
5	1.4608	0.1783	0.9992
6	1.4912	0.1789	0.9990
7	1.5091	0.1800	0.9981
9	1.5459	0.1778	0.9987
11	1.5830	0.1686	0.9994

Table II (L=1)

D	Rc	E	ρ
5	0.5200	0.0788	0.9985
8	0.4549	0.0612	0.9991
10	0.4220	0.0549	0.9995
20	0.3233	0.0385	0.9983
30	0.2894	0.0332	0.9983
40	0.2383	0.0272	0.9992
60	0.1968	0.0231	0.9992
90	0.1642	0.0187	0.9981

To scale the problem, we set

$$Rc = aL \tag{3}$$

$$E = bL \tag{4}$$

We found a and b are only dependent on n. The curves are shown in Fig.3.

We claim that a is approximately a linear function of $\sqrt{\ln n / n}$ and $bn\ln^2 n$ is approximately a linear function of n, which is shown in Fig.4. By fitting the data, we get (5) and (6). The correlation coefficients are 0.9988 and 0.9977 for the two equations, respectively.

$$a \approx 1.0362\sqrt{\frac{\ln n}{n}} - 0.073 \tag{5}$$

$$b \approx \frac{0.3743n - 0.3331}{n\ln^2 n} \tag{6}$$

Equations (1) and (3)-(6) summarize our solution to the problem, which is based on the simulation with n between 3 and 125. If the number of nodes is not too large, then we can safely apply (3)-(6) to find E and R_c and finally use (1) to get a good estimate.

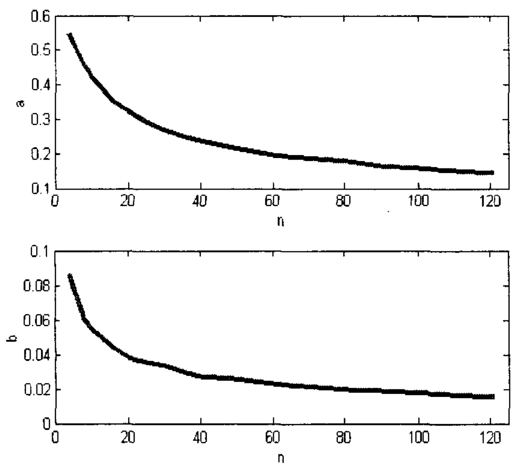


Fig.3. Plots of a and b versus n.

In order to check the accuracy of our empirical solution, we randomly choose some cases to compare the results from simulation and from the empirical equations. They are shown in Table III and show a good match.

Table III
Comparison of the simulation and empirical results

L, D	R	Simulation	Formula	Error
2, 10	0.6	0.909	0.8974	-1.47%
20, 0.05	8	0.8852	0.8589	-2.97%
6, 1	1.56	0.5697	0.5532	-2.9%
9, 1	1.8	0.8115	0.8390	3.39%
1, 60	0.21	0.6195	0.6364	2.73%

Finally we give a theoretical upper bound on the probability that any two nodes within a network of N nodes are connected. This bound, although loose for small values of the ratio R/L supports the results obtained through simulations. It may be derived in a straightforward way from [8] and [9]. Indeed it was proven in [9] that when the node locations have a two-dimensional gaussian distribution, the probability of an m-hop connection P_m satisfies:

$$P_m < e^{-(m-1)^2R^2/4\sigma^2} - e^{-m^2R^2/4\sigma^2} = P_{m\infty} \tag{7}$$

It was further proven in [8] that the distance distributions in scenarios where nodes are uniformly distributed in a rectangular area and where they are distributed over an unbounded area according to a 2-D gaussian distribution are very similar when the width of the rectangular area is taken to be about three times the standard deviation of the location distribution in the gaussian distribution. We make use of this fact to conclude that (7) holds for the scenario of interest here (that is uniform distribution, square deployment area L^2) if $\sigma = L/3$. Thus two nodes are connected with probability:

$$P \le \sum_{i\ge1} P_i = 1 - e^{-N^2R^2/4\sigma^2} \tag{8}$$

Here we would also like to point out a very recent work done by Desai and Manjunath [10]. They obtained the exact formula for the probability that the network is connected for a one-dimensional finite ad hoc network. They also extended the result to find a simple upper bound for the connectivity in a two-dimensional finite ad hoc network. The bound is not asymptotically tight.

4. DISCUSSION

We can now have a look at the asymptotic behavior of connectivity probability. Given L, when D increases, i.e. n increases, b and E will go to zero in (4) and (6). This explains the asymptotical behavior of the probability curve. When E is near zero, from (1), we can see that P approaches 1 for R greater than R_c and approaches zero for R less than R_c.

Clearly, R_c is the critical transmission range. The exact value of critical transmission range is still an open problem since Gilbert's work [4] published in 1961. It is conjectured to be $2\sqrt{\ln(DL^2)/\pi D}$ in [1]. But a reviewer of that paper conjectured it to be $\sqrt{2\ln(DL^2)/\pi D}$ and the author of [2] conjectured it to be $\sqrt{\ln(DL^2)/\pi D}$. If we notice that $\sqrt{\ln(DL^2)/D} = \sqrt{\ln n/n}L$, $2/\sqrt{\pi} \approx 1.128$ and our coefficient in (5) is 1.0362, we would lean to support the author of [1]'s conjecture.

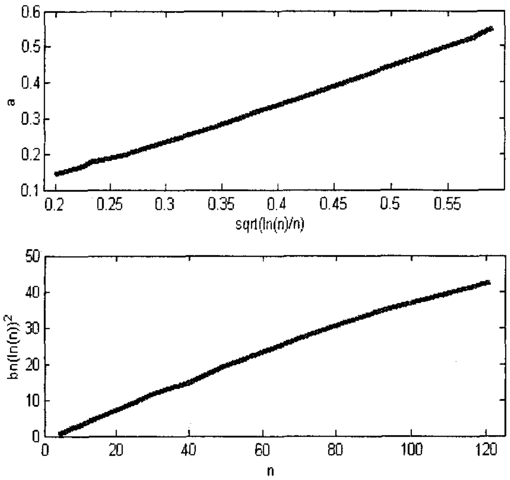


Fig.4. Linear fitting for *a* and *b*.

From the simulation results (Fig. 2), one can see that in order to achieve a fully connected network with high probability, a large transmission range is required, which is not always feasible or desirable. However when the network is mobile, due to changing topology, the distribution of the node locations, after some time, can be seen to be independent of the previous one. By taking advantage of the statistical independence of the network topologies at different points in time we need not always require full connectivity, as we shall explain later.

In order to describe the connectivity when the network is not fully connected. We define a connectivity index η as following:

$$\eta = \frac{\sum_i n_i(n_i - 1)}{\sum_i n_i(\sum_i n_i - 1)} \tag{9}$$

where n_i is the number of nodes in the *ith*-connected component. η ranges from 0 to 1. When n_i is 1 for all i, i.e. no node can communicate with other nodes, η is zero. When the whole network is fully connected, i.e. there is only one component with *n* nodes, η is one. For example, in Fig.1, there are two connected components with $n_1 = 7$ and $n_2 = 2$. So, η is 0.6111. When the number of nodes is large, equation (9) can be approximated by equation (10):

$$\eta \approx \frac{\sum_i n_i^2}{(\sum_i n_i)^2} \tag{10}$$

The typical shape of an η curve is shown in Fig. 5, where *L*=1 and *D*=5. We also show the standard deviation of η for that case in Fig. 6.

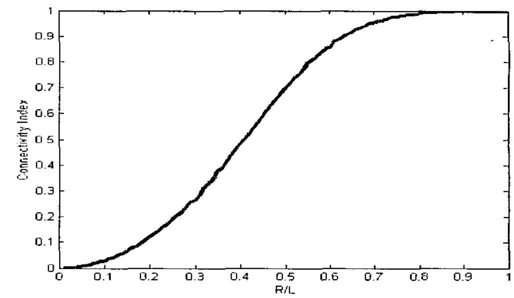


Fig.5. Connectivity index as a function of the normalized transmission range. (*L*=1, *D*=5)

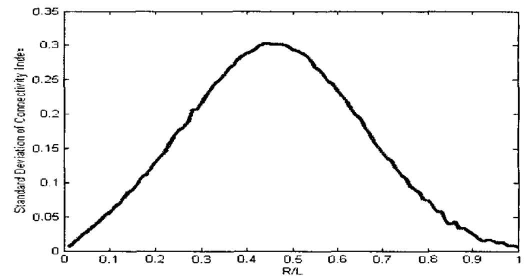


Fig.6. Standard deviation of connectivity index as a function of the normalized transmission range. (*L*=1, *D*=5)

Since $n_i - 1$ is the number of nodes with which one node in the *ith*-connected component can communicate, $(n_i - 1)/(\sum n_i - 1)$ is the probability that this node can communicate with a randomly chosen destination. Sum this probability with weight n_i and normalize it by the total number of nodes n, we get η. η can be interpreted as the average probability of successful communication between two nodes. Clearly, η is larger than *P* for the same *D*, *L* and *R* since *P* is the probability that all the nodes are connected which is a more stringent requirement. This can also be seen by comparing fig. 2 and fig. 5 for the case *L*=1, *D*=5.

Next we investigate how mobility impacts connectivity. In order to simplify our analysis we assume time is slotted. If after a certain number of time units, say *T*, the topology can be regarded as being independent, then after *kT* time units the probability P_s that any two nodes can communicate is:

$$P_s = 1 - \prod_{i=1}^{k}(1 - \eta_i) \tag{11}$$

where η_i is the connectivity index for time unit *i*. For simplicity of analysis, assume η_i is a constant η then (9) reduces to:

$$P_s = 1 - (1 - \eta)^k \tag{12}$$

Now we illustrate how we can take advantage of statistical independence of the network topologies at different times on a simple example where *k*=5. Even though the connectivity index is 0.7, any two nodes can still successfully communicate with a probability as high as 0.998 by (12). That means it is still possible to communicate 'reliably' with a proper scheme even if the network is not quite well connected. This can decrease the required transmission range dramatically. For example, for the case L=1 and D=5, we can achieve η = 0.7 by setting R=0.5 according to Fig. 5 while we need R=0.92 to guarantee P to be 0.998 according to Fig. 2. Of course, the scheme will lead to larger delays. How much the delay will increase depends on the changing rate of the topology, i.e. the mobility of the nodes.

5. CONCLUSIONS AND FUTURE WORK

In this article we provided a simple empirical formula to calculate the probability that an ad hoc network is fully connected. By simulation, we find that this formula is quite accurate at least for the case that the number of nodes does not exceed 125, which is enough in typical applications (like the ad hoc network formed by laptops in a conference hall). We have also examined the asymptotic behavior of the connectivity formula and gotten results that agree with the existing literature. Finally, we introduced the connectivity index metric to quantify the connectivity of a network.

At present, we are extending this work by exploring the effect of different connectivity indices on the performance of the network, such as throughput and delay. It would be interesting to design some schemes to combat the possibility of disconnection. Work is also needed to calculate the connectivity index under different areas and user densities.

ACKNOWLEDGMENT

The authors would like to thank Jiantao Wang for his suggestions.

REFERENCES

[1] Philippe Piret, "On the connectivity of Radio Networks," *IEEE Transactions on Information Theory*, vol.37, pp.1490-1492, September 1991.

[2] Thomas K Philips, Shivendra S. Pandwar and Asser N. Tantawi, "Connectivity Properties of a Packet Radio Network Model," *IEEE Transactions on Information Theory*, vol.35, pp.1044-1047, September 1989.

[3] Yuan-Chieh Cheng and Thomas G. Robertazzi, "Critical Connectivity Phenomena in Multihop Radio Models," *IEEE Transactions on Communications*, vol.37, pp.770-777, July 1990.

[4] E. N. Gilbert, "Random plane networks", *J.Soc.Indust. Appl.Math*, vol.9, pp.533-543, December 1961.

[5] Piyush Gupta, P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," pp. 547-566, in Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming. Edited by W. M. McEneany, G.Yin, and Q.Zhang, Birkhauser, Boston, 1998.

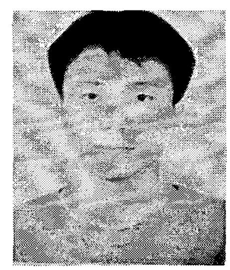
[6] Charles J. Colbourn, The Combinatorics of Network Reliability, Oxford University Press, 1987.

[7] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol.46, pp.388-404, March 2000

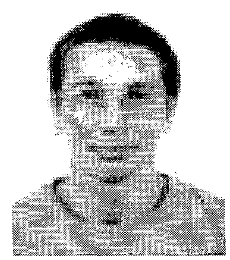
[8] L. E. Miller, "Distribution of link Distances in a wireless network," Journal of Research of the National Institute of Standards and Technology, 2001.

[9] L.E. Miller, "Probability of a two-hop connection in a random mobile network," Conference on information Sciences and Systems, the John Hopkins University, March 2001.

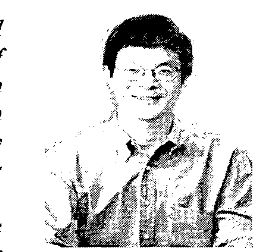
[10]M. Desai and D. Manjunath, " On the connectivity in Finite Ad Hoc Networks", IEEE Communication Letters, vol.6, no.10, pp.437-439, Oct.2002



Ao Tang *is a second year graduate student in Dept. of EE, Caltech. He is working in networking lab and his advisor is Prof. Steven. H. Low. He got his BE and ME in Tsinghua University, Beijing in 1999 and 2001, respectively. He got his MSEE in Caltech in 2002. Currently he is working toward PhD and his main research interest lies in wireless network and Internet congestion control.*



Cedric Florens *is a PhD student at the California Institute of Technology. His interests are in discrete mathematics and communications. He is working with Prof. Robert J. McEliece. His current research focuses on ad-hoc networks and sensory networks. He has a MSEE from Caltech.*



Steven. H .Low *received his B.S. degree from Cornell University and PhD from the University of California -- Berkeley. both in electrical engineering. He was with AT&T Bell Laboratories, Murray Hill, from 1992 to 1996, and was with the University of Melbourne, Australia, from 1996 to 2000, and is now an Associate Professor at the California Institute of Technology, Pasadena. He has held visiting academic positions in the US and Hong Kong,and has consulted with companies and government in the US*

and Australia. He was a co-recipient of the IEEE William R. Bennett Prize Paper Award in 1997 and the 1996 R&D 100 Award. He is on the editorial board of IEEE/ACM Transactions on Networking. He has been a guest editor of the IEEE Journal on Selected Areas in Communications, on the program committee of several conferences. His research interests are in the control and optimization of communications networks and protocols. His home is netlab.caltech.edu and email is slow@caltech.edu.