The Cluster Density of a Distributed Clustering Algorithm in Ad Hoc Networks

Christian Bettstetter
DoCoMo Euro-Labs, Future Networking Lab, Munich, Germany
Technische Universität München, Institute of Communication Networks, Munich, Germany (previously)

***Abstract*— Given is a wireless multihop network whose nodes are randomly distributed according to a homogeneous Poisson point process of density ρ (in nodes per unit area). The network employs Basagni's *distributed mobility–adaptive clustering* (DMAC) algorithm to achieve a self–organizing network structure. We show that the cluster density, i.e., the expected number of clusterheads per unit area, is $\rho_c = \frac{\rho}{1+\mu/2}$, where μ denotes the expected number of neighbors of a node. Consequently, a clusterhead is expected to incorporate half of its neighboring nodes into its cluster. This result also holds in a scenario with mobile nodes and serves as a bound for inhomogeneous spatial node distributions.**

***Index Terms*— Ad hoc networking, sensor networks, self-organization, clustering, cluster density, leader election.**

I. Introduction

The aim of distributed clustering in ad hoc networks is to dynamically organize all mobile nodes into groups, called *clusters*. The obtained cluster structure provides the basis for a hierarchical network organization. Although many clustering algorithms have been proposed in the literature [1–15], only little analytical work has been done to obtain a comprehensive understanding of their behavior [14, 15]. This lack is our motivation to derive closed–form expressions for the *cluster density* and *cluster order* created by the well–known *distributed mobility–adaptive clustering* (DMAC) algorithm [8]. The cluster density is defined as the expected number of clusters per unit area, and the cluster order is the number of nodes belonging to a cluster. These two values are fundamental design and performance criteria of any clustering algorithm.

The content of this paper is organized as follows: Section II reviews the design and communication protocol of the DMAC algorithm, and shows the uniqueness of the obtained cluster structure. Section III takes a simulation–based approach to obtain the cluster density in a random ad hoc network. From the results of these simulations, Section IV deduces a closed–form expression of general validity. Section V addresses the cluster order. Finally, Section VI concludes.

II. The DMAC Algorithm

The DMAC algorithm is a promising and yet simple clustering algorithm suitable for mobile ad hoc networks. It employs the concept of electing *clusterheads* among the nodes. Typically, these clusterheads have special functions, such as maintaining routing information or creating a "virtual backbone." When a node appears in the network, it executes an initialization process to determine its role, i.e., whether it should create a new cluster (become a clusterhead) or affiliate with an existing cluster (become an *ordinary node*). Each node has a *weight* w that determines its chance to become a clusterhead; the larger the weight of a node, the better it is suited to be a clusterhead. These weights may be assigned randomly or according to certain characteristics of the node (e.g., its IP address, transmission power). We assume that each node has a unique weight, at least among all nodes within a distance of two hops.

A. Design Rules

Two nodes are denoted as *neighbors* if they have a direct wireless link between each other. The number of neighbors of a node is denoted as its *degree* d.

A valid cluster structure is defined by three rules [8]: (a) every ordinary node has at least one clusterhead as neighbor; (b) every ordinary node is affiliated with exactly one clusterhead, namely the neighboring clusterhead that has the largest weight; and (c) clusterheads must not be neighbors. The first rule ensure that each ordinary node has access to a clusterhead using a single–hop transmission. The second rule ensures that the node has access to the "best available" clusterhead. The third rule creates a "well scattered" set of clusterheads. Figure 1 gives an example of a valid DMAC structure. Figure 2 illustrates some further examples.

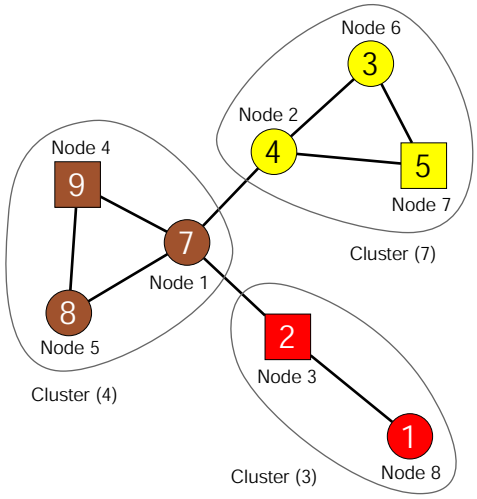
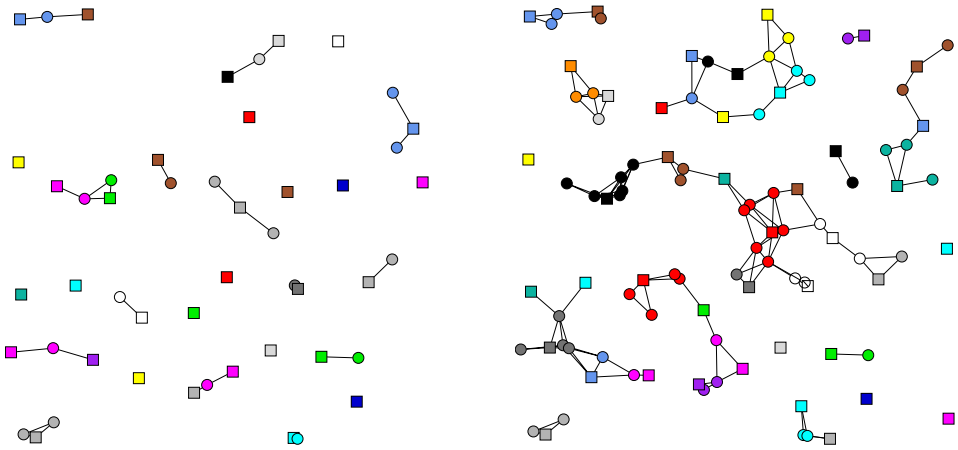
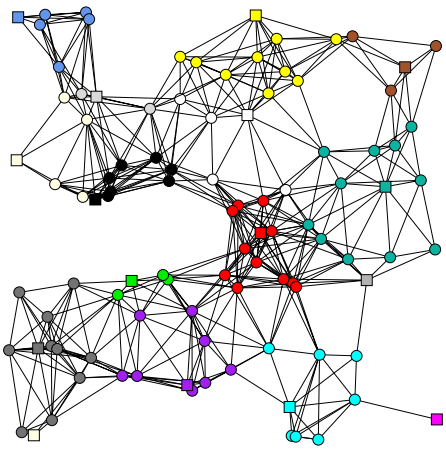


Fig. 1. Example network with DMAC clustering. The numbers within the nodes indicate their weights. Clusterheads are shown as squares and ordinary nodes as circles.

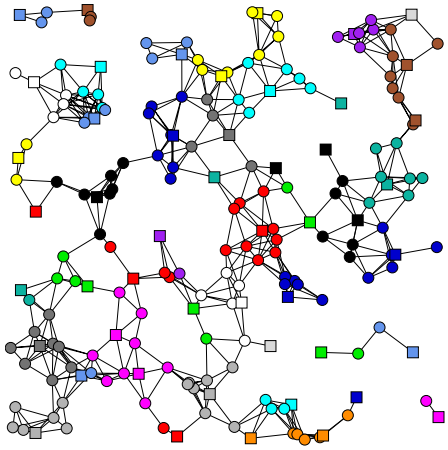


(a) $n = 50$ with $r_0 = 0.1a$ (32 clusters)

(b) $n = 100$ with $r_0 = 0.1a$ (42 clusters)



(c) $n = 100$ with $r_0 = 0.2a$ (16 clusters)



(d) $n = 200$ with $r_0 = 0.1a$ (49 clusters)

Fig. 2. Example topologies resulting from DMAC clustering on n uniformly distributed nodes on square area of length a.

B. Communication Protocol

Let us now explain how the nodes should act and communicate to achieve a structure that fulfills the DMAC rules [8]. The role decision of a node is only based on its knowledge about its neighbors. A new node decides to join a cluster if there is already a neighboring clusterhead with a weight that is higher than its own weight; otherwise it decides to become clusterhead itself. After making this decision, the node informs all its neighbors of its role. It sends out a `Join` message if it joins a cluster, or a `Clusterhead` message if it becomes a clusterhead. The algorithm is message driven and executed at each node. To react properly and consistently, each node has to know the weight and role of each of its current neighbors.

Each node reacts to changes in the surrounding neighborhood and alters its role and cluster membership accordingly. Whenever a link breaks down between a clusterhead and one of its member nodes, the clusterhead removes the membership of the node from its cluster, and the member node must determine its new role as described above. A new link between two nodes is handled as follows: If a node notices the presence of a new neighbor and determines that this new neighbor is a clusterhead with a larger weight than its current clusterhead, it will join the new clusterhead. Similarly, if a clusterhead gets a new neighboring clusterhead with a higher weight, it will give up its clusterhead role and affiliate with the new neighbor. If a node receives a `Clusterhead` message from a neighbor, it has to check whether it has to affiliate with this neighboring clusterhead or not. If a clusterhead receives a `Join` message it has to check whether the sending node joins its cluster or a different cluster. If an ordinary node receives a `Join` message from its own clusterhead, it knows that this clusterhead gave up its role; hence, the ordinary node must re–decide its role and cluster membership.

C. Correctness and Uniqueness of the Steady–State Structure

Using this communication protocol, any multihop network converges within a finite time interval to a steady–state cluster structure that fulfills the above design rules [8].

Furthermore, for a given network and given weight assignments, the algorithm always produces the same cluster structure, independently of the chronology of the nodes' role decisions, re–decisions, and their messages. A proof for this

uniqueness property can be given by regarding the steady state of a connected component of the network: The node with the highest weight of all nodes in this component is always a clusterhead. All neighbors of this clusterhead are affiliated with it. From the remaining nodes, again the node with the highest weight is a clusterhead, all neighbors have joined it, and so on. This holds for all connected components.

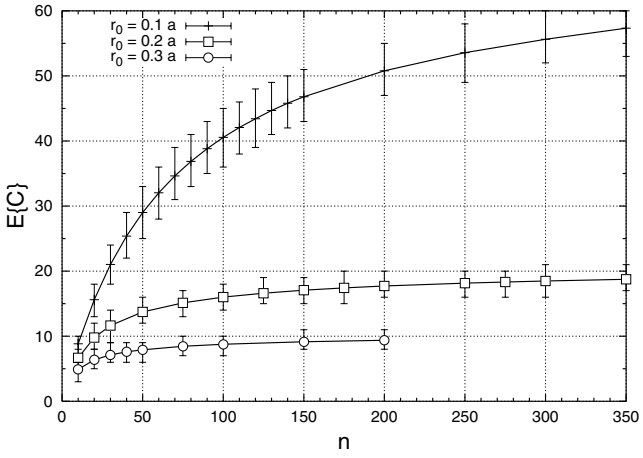
D. Why DMAC?

The reasons why we have chosen the DMAC algorithm for our investigations are as follows. First, most of the algorithms proposed in the early days of ad hoc networking do not allow the nodes to move during the initial clustering and are thus suited for rather static networks. The DMAC algorithm, however, is designed in a manner that nodes can always move. Second, the algorithms proposed more recently are much more complex than the DMAC algorithm. We believe that the mobile and distributed environment of ad hoc networks requires a rather simple clustering solution that can react to topology changes very quickly (only by knowing its one–hop neighbors). Other algorithms (e.g., [3]) require knowledge of all two–hop neighbors. Another argument for choosing DMAC is its detailed documentation in [8].

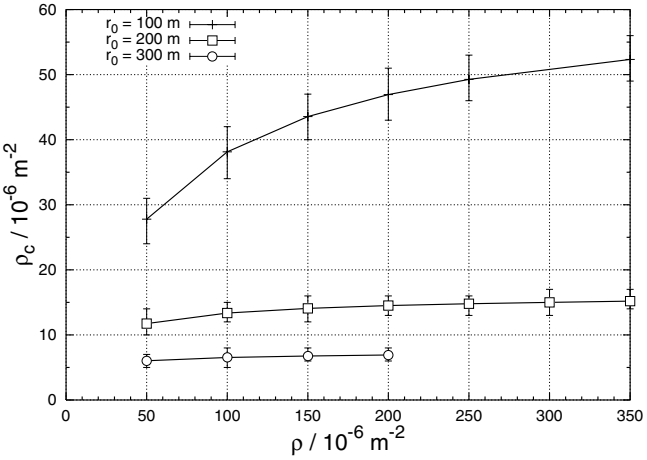
III. Cluster Density: Simulation–Based Approach

This section takes a simulation–based approach to study the cluster density in a random scenario. We generate a random network $G_i(r_0, n)$ by placing n nodes uniformly on a square area of size $A = a \times a$ and adding links between any two nodes located within the *radio transmission range* r_0 from each other. Each node gets a weight sampled from a uniform random distribution. After initial node placement, all nodes remain static and run the usual DMAC algorithm to determine their roles. Within a finite time interval, the network converges to a steady–state and valid cluster structure. Four example topologies with different parameters are given in Figure 2. Once this structure is achieved, we determine the number c_i of clusterheads. Repeating this experiment over many random topologies ($i = 1 \ldots \Omega$, with large Ω), the mean of the samples $\{c_i\}_{i=1\ldots\Omega}$ is an estimate for the *expected number of clusters* $E\{C\}$ on the given area. We also compute the $5\,\%$ and $95\,\%$ quantiles of the samples to obtain a measure for the variance of C.

Figure 3 a shows the result for $E\{C\}$ over n for three different values of r_0. We first interpret the impact of the number of nodes n for fixed r_0. For low n, the number of clusterheads increases with increasing n. In this region of the curve, the nodes are sparsely distributed over the area such that the level of connectivity is low. Several nodes are isolated from other nodes, thus being clusterheads without any member nodes. Consequently, additional nodes result in more clusterheads. This behavior changes, however, for large n. Now, the curve levels off because additional nodes typically join existing clusters or force a resignation of existing clusterheads. Finally, the number of clusters converges toward an asymptotic value $\lim_{n\to\infty} E\{C\}(n, r_0/a) = E\{C\}^{\infty}(r_0/a)$. In this asymptotic case, we obtain a connected continuum of nodes, where



(a) Expected number of clusterheads $E\{C\}$ on $a \times a$ area with border effects



(b) Density of clusterheads ρ_c without border effects

Fig. 3. Density of DMAC clusters with uniformly distributed nodes. Each simulation result (each dot) is based on $\Omega = 30000$ random topologies and was simulated with $a = 1000\,\mathrm{m}$. An error bar shows the interval in which $90\,\%$ of the measured samples are located ($5\,\%$ and $95\,\%$–quantiles).

the circular transmission areas of the clusterheads cover the entire system area.[1] The qualitative impact of r_0 on the number of clusters is also obvious. The higher r_0, the fewer clusters are produced, and the earlier the curve levels off. This is because clusterheads with a higher r_0 cover a larger area.

Notice that we must be aware of border effects in our results. Obviously, nodes at the border have a higher isolation probability. Hence, although the nodes are uniformly distributed within the system area, clusterheads are more likely to be found at the border in a sparsely connected network. Thus, we perform a second simulation series in which these effects are avoided by using a cyclic (toroidal) distance model for link formation [18].

[1] A loose bound for $E\{C\}^{\infty}$ is given as follows. The number of clusterheads in an area is larger than or equal to the minimum number of overlapping circles of radius r_0 that cover the entire area. A lower bound for the latter value is given in [16, 17]. This yields $E\{C\}^{\infty} \geq$ # covering circles $> \left\lceil \frac{1}{3^{3/2}} \left(\frac{2a^2}{r_0^2} + \frac{a}{r_0} \right) \right\rceil$.

With this model, nodes close to the border of the simulation area can establish links across the borderline to nodes located at the opposite side of the area. We also ensure that $a \gg r_0$. This setup approximates a subarea of a network between nodes that are distributed according to a two–dimensional Poisson point process with a constant node density $\rho = n/A$ on an infinite area.

Given this spatial node distribution and the uniform distribution of the weights among the nodes, it follows that the clusterheads are now homogeneously distributed with a constant *cluster density* ρ_c given by

$$\rho_c = \rho_c(\rho, r_0) = \frac{E\{C\}}{A} . \tag{1}$$

Figure 3 b shows the corresponding simulation result. In this case, less clusters per unit area are created, since nodes can now join clusterheads via the borderline of the area. The curves again converge toward a value $\lim_{\rho\to\infty} \rho_c(\rho, r_0) = \rho_c^{\infty}(r_0)$.

IV. Cluster Density: Analytical Approach

From the literature on stochastic point processes, we find out that the spatial distribution of the DMAC clusterheads belongs to the family of *hard–core point processes* [19]. In such a process, points are forbidden to lie closer together than a certain minimum distance. The Matérn hard–core process, for example, applies a thinning to a homogeneous Poisson point process in which a point is only retained if it has the highest weight w of all Poisson points located within a circle of radius r_0 around its position. The thinned out points are removed after all points have been classified as either retained or thinned out. There is a small but important difference to the DMAC clustering: A clusterhead may indeed have a neighboring (ordinary) node with a higher weight (see Fig. 1, Nodes 3 and 1). Hence, the Matérn process gives us only a subset of the set of clusterheads and yields a lower bound for the clusterhead density. We have $\rho_c = \rho_{\text{Matérn}} + \rho_c'$, with $\rho_{\text{Matérn}} = \frac{1}{r_0^2\pi}\left(1 - e^{-\rho\, r_0^2 \pi}\right)$ and ρ_c' being the density of clusterheads that have at least one higher–weighted neighbor. Although the qualitative behavior of $\rho_{\text{Matérn}}$ is the same as the one of ρ_c, it yields in general only a loose bound for ρ_c. Let us therefore take our own approach to give an equation for ρ_c.

To do so, we regard the problem from slightly different viewpoint. The expected *percentage of clusterheads* among all nodes in a random topology, i.e., the probability that a randomly chosen node is a clusterhead, is given by

$$P(\text{CH}) \;=\; \frac{E\{C\}}{n} \;=\; \frac{\rho_c}{\rho} . \tag{2}$$

The key to our solution is to realize that this probability is only a function of the expected value of a node's number of neighbors — i.e., the *expected node degree* $E\{D\}$ — in the given scenario. The values for $E\{D\}$ are (see Appendix A)

$$E\{D\} = n\,\frac{r_0^2\pi}{a^2}\left(1 - \frac{8}{3\pi}\,\frac{r_0}{a} + \frac{1}{2\pi}\,\frac{r_0^2}{a^2}\right) \tag{3}$$

for a square area, and $E\{D\} = \rho\, r_0^2\pi$ for a Poisson point process. Using the above simulation results, we can plot $P(\text{CH})$ over $E\{D\}$ as shown in Figure 4. The given curve holds with and without border effects. Two extreme cases can be easily understood:

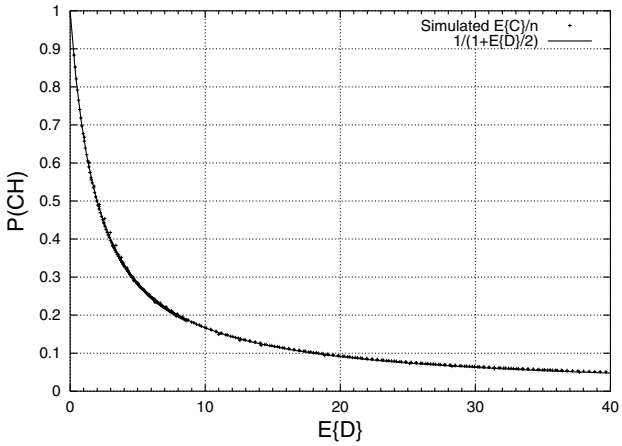


Fig. 4. Probability that a randomly chosen node is a clusterhead

- If $r_0 = 0$, the degree of each node is zero; each node becomes a clusterhead.
- If $n/A \to \infty$, the expected degree tends to infinity; hence, the probability of a node to be clusterhead tends to zero.

Furthermore, one out of two nodes becomes a clusterhead if $E\{D\} = 2$. With these observations, a regression function can be derived that matches exactly the simulation results; we have

$$P(\text{CH}) \;=\; \frac{1}{1 + \frac{E\{D\}}{2}} . \tag{4}$$

This general result enables us to compute the expected number of clusters $E\{C\} = n\,P(\text{CH})$ for different area shapes as long as we know the expected degree. The cluster density of Poisson distributed nodes is

$$\rho_c \;=\; \frac{1}{\frac{1}{2}\,r_0^2\pi + \frac{1}{\rho}} ; \quad \text{with } \rho_c^{\infty} = \frac{2}{r_0^2\pi} . \tag{5}$$

V. Cluster Order

Given this insight on the cluster density, we can deduce the expected number of ordinary nodes $E\{M\}$ in a randomly chosen cluster ($0 \le M \le n-1$). Assuming a homogeneous node distribution without border effects, each clusterhead has on average the same number of member nodes. This yields

$$E\{M\} = \frac{\rho - \rho_c}{\rho_c} \overset{(5)}{=} \frac{\rho\, r_0^2\pi}{2} = \frac{E\{D\}}{2} . \tag{6}$$

Interestingly, the number of nodes that a "typical" clusterhead incorporates into its cluster is half of its expected total number of neighbors. The expected cluster order is then $E\{M\} + 1$.

Example. Suppose a sensor network is distributed for environmental monitoring with a density of $n/A = 0.001\,\text{m}^{-2}$ on an area of dimensions $1000 \times 2000\,\text{m}^2$. Each sensor is equipped with a transceiver capable of transmitting $r_0 = 90\,\text{m}$. The DMAC protocol is used to elect clusterheads that have the task of transmitting the sensed data to a central unit. Each sensor

collects about 8 kbit of data per second. What is the average transmission load of the clusterheads? With (8) the expected degree of a sensor is 24. Ignoring border effects each clusterhead has on average 12 affiliated nodes, which gives a total data traffic of 96 kbit/s to be transmitted.

The fact that $E\{M\}$ increases linearly with ρ for fixed r_0 is a drawback of the DMAC algorithm with respect to its scalability. As the node density increases, clusterheads might be overloaded with too many members and become a bottleneck. An extension that limits the cluster order and allows a splitting of an existing cluster into several smaller clusters would be useful. Such an extension requires us to waive the third DMAC rule "clusterheads must not be neighbors." Another possibility to solve the problem would be that the overloaded clusterhead automatically reduces its transmission power.

VI. Concluding Remarks

Using a heuristic approach, this paper gave equations for the cluster density and cluster order of homogeneously distributed nodes running the DMAC algorithm. Since the DMAC structure is unique, the equations also hold in a mobile scenario if the used mobility model retains the homogeneous distribution of the nodes (see [20] for a discussion on this topic). If the nodes are inhomogeneously distributed, the cluster density will decrease. In fact, the validity of our result is not restricted to the DMAC algorithm. It also holds for other algorithms that limit the cluster size to two hops. For example, our simulation study in [21] shows that the Lin/Gerla clustering algorithm [3] creates the same cluster density, although it does not employ the concept of clusterheads.

A topic for future research is the analytical study of algorithms that allow the clusterheads to be more sparsely distributed [11].

References

[1] D. J. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *IEEE Trans. Commun.*, vol. 29, pp. 1694–1701, Nov. 1981.

[2] M. Gerla and J. T.-C. Tsai, "Multicluster, mobile, multimedia radio network," *ACM/Kluwer Wireless Networks*, vol. 1, pp. 255–265, Mar. 1995.

[3] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1265–1275, Sept. 1997.

[4] P. Krishna, N. H. Vaidya, M. Chatterjee, and D. K. Pradhan, "A cluster–based approach for routing in dynamic networks," *ACM Computer Communications Review*, vol. 27, Apr. 1997.

[5] R. Ramanathan and M. Steenstrup, "Hierarchically–organized, multihop mobile wireless networks for quality-of-service support," *ACM/Kluwer Mobile Networks and Applications*, vol. 3, pp. 101–119, June 1998.

[6] B. Das and V. Bharghavan, "Routing in ad-hoc networks using minimum connected dominating sets," in *Proc. IEEE Intern. Conf. on Communications (ICC)*, (Monréal, Canada), June 1997.

[7] R. Sivakumar, B. Das, and V. Bharghavan, "Spine–based routing in ad hoc networks," *ACM/Baltzer Cluster Computing*, vol. 1, Nov. 1998.

[8] S. Basagni, "Distributed clustering for ad hoc networks," in *Proc. Intern. Symp. Parallel Architectures, Algorithms, and Networks (ISPAN)*, (Perth/Fremantle, Australia), June 1999.

[9] J. Wu and H. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," in *Proc. Intern. Workshop Discrete Algorithms and Methods for Mobile Computing and Communications (DialM)*, (Seattle, USA), Aug. 1999.

[10] A. B. McDonald and T. F. Znati, "A mobility–based framework for adaptive clustering in wireless ad hoc networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1466–1487, Aug. 1999.

[11] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max–min d–cluster formation in wireless ad hoc networks," in *Proc. IEEE Infocom*, (Tel Aviv, Israel), 2000.

[12] K. M. Alzoubi, P.-J. Wan, and O. Frieder, "Message–optimal connected dominating sets in mobile ad hoc networks," in *Proc. ACM Intern. Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc)*, (Lausanne, Switzerland), June 2002.

[13] Y. P. Chen and A. L. Liestman, "Approximating minimum size weakly-connected dominating sets for clustering mobile ad hoc networks," in *Proc. ACM Intern. Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc)*, (Lausanne, Switzerland), June 2002.

[14] P.-J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed construction of connected dominating set in wireless ad hoc networks," in *Proc. IEEE Infocom*, (New York, USA), June 2002.

[15] J. Sucec and I. Marsic, "Clustering overhead for hierarchical routing in mobile ad hoc networks," in *Proc. IEEE Infocom*, (New York, USA), June 2002.

[16] S. Verblunsky, "On the least number of unit circles which can cover a square," *J. London Math. Soc.*, vol. 24, no. 164-170, 1949.

[17] L. F. Tóth, "Review of [16]." http://www.ams.org/mathscinet-getitem?mr=11,455g, cited August 2003.

[18] C. Bettstetter and O. Krause, "On border effects in modeling and simulation of wireless ad hoc networks," in *Proc. IEEE Intern. Conf. on Mobile and Wireless Commun. Networks (MWCN)*, (Recife, Brazil), Aug. 2001.

[19] S. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. Wiley, 2nd ed., 1995.

[20] C. Bettstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and border effects," *ACM Mobile Computing and Communications Review*, vol. 5, no. 3, 2001.

[21] C. Bettstetter and J. Xi, "Mobility modeling and analysis of adaptive clustering algorithms in ad hoc networks," in *Proc. European Personal Mobile Commun. Conf. (EPMCC)*, (Vienna, Austria), Feb. 2001.

[22] B. Ghosh, "Random distances within a rectangle and between two rectangles," *Bull. Calcutta Math. Soc.*, vol. 43, pp. 17–24, 1951.

Appendix

The probability density function of the distance S between two nodes that are independently uniformly distributed (at random) in a rectangular area of size $A = a \times b$, $a \geq b$, is $f_S(s) = \frac{4s}{a^2 b^2}\, f_0(s)$ with [22]

$$f_0(s) = \begin{cases} \frac{\pi}{2}\, ab - as - bs + \frac{1}{2} s^2 \quad \text{for } 0 \leq s \leq b \\ ab\, \arcsin \frac{b}{s} + a\sqrt{s^2 - b^2} - \frac{1}{2}\, b^2 - as \\ \quad \text{for } b < s < a \\ ab\, \arcsin \frac{b}{s} + a\sqrt{s^2 - b^2} - \frac{1}{2}\, b^2 - \\ \quad ab\, \arccos \frac{a}{s} + b\sqrt{s^2 - a^2} - \frac{1}{2}\, a^2 - \frac{1}{2}\, s^2 \\ \quad \text{for } a \leq s \leq \sqrt{a^2 + b^2} \\ 0 \quad \text{otherwise} \end{cases} . \quad (7)$$

The probability that two uniformly distributed nodes, each having a transmission range r_0, are neighbors is therefore given by the integral $P_0 = \int_0^{r_0} f_S(s)\, ds$. Thus, the expected degree of a node in a network with $n \gg 1$ uniformly distributed nodes, each node with transmission range r_0, is $E\{D\} = nP_0$. This yields for $r_0 < b$ the expression

$$E\{D\} = \frac{n\, r_0^2 \pi}{ab} \left(1 - \frac{4}{3\pi} \left(\frac{r_0}{a} + \frac{r_0}{b} \right) + \frac{1}{2\pi} \frac{r_0^2}{ab} \right) . \quad (8)$$

Using a square area of size $a \times a$ we obtain

$$E\{D\} = n\, \hat{r}_0^2 \pi \left(1 - \frac{8}{3\pi}\, \hat{r}_0 + \frac{1}{2\pi}\, \hat{r}_0^2 \right) . \quad (9)$$

with the normalized range $\hat{r}_0 = r_0/a$.

Acknowledgment

The author would like to thank the German research foundation (DFG) for supporting his research. He would also like to thank Prof. Jörg Eberspächer (TUM) and the five reviewers.