Node Clustering in Mobile Peer-to-Peer Multihop Networks

Chansu Yu†, Kang G. Shin‡, Ben Lee¶, Seung-Min Park* and Heung-Nam Kim*

†Department of Electrical and Computer Engineering, Cleveland State University,
Cleveland, OH 44115-2425, c.yu91@csuohio.edu
‡Department of Electrical Engineering and Computer Science, University of Michigan
Ann Arbor, MI 48109, kgshin@eecs.umich.edu
¶School of Electrical Engineering and Computer Science, Oregon State University
Corvallis, OR 97331, benl@eecs.orst.edu
*Embedded Software Division, Electronics and Telecommunications Research Institute
161 Gajeong, Yuseong, Daejeon, Korea, {minpark, hnkim}@etri.re.kr

Abstract

In mobile peer-to-peer (MP2P) networks, nodes tend to gather together rather than scattered uniformly across the network area. This paper considers the clustering of peer nodes and its performance impact in MP2P networks. The model for node clustering based on a heavy-tail distribution is first introduced and then a topology generation method that produces a clustered network is presented. Experiments based on ns-2 simulation with AODV routing protocol and IEEE 802.11 MAC reveal that the clustered layout significantly degrades the network performance and the main trouble comes from the MAC layer mechanisms. Node clustering results in as much as 77.6% lower packet delivery ratio compared to random node distribution. Moreover, it results in larger variation in packet delivery service, and thus has a serious impact on QoS, which is important in MP2P networks.

1. Introduction

Mobile peer-to-peer (MP2P) communication over *mobile ad hoc networks* (MANETs) will become an essential part of future computing environment with advances in wireless communication and portable system technology. Providing consistent performance to each participant is highly desirable in this type of networks. However, this may not be the case when nodes are cluttered rather than scattered uniformly across the geographic area. Node clustering occurs when nodes tend to move closer to some particular landmarks. In other word, some areas have a high concentration of nodes while other areas have only few nodes. We refer to this type of node placement as *clustered layout*. In contrast to the random distribution of nodes, the clustered layout can significantly affect network performance.

Understanding and modeling of clustered layout is the main theme of this paper. The profound impact of node clustering on network performance has not been addressed until recently [11, 19]. Kawadia and Kumar noted the performance degradation due to non-homogeneous distribution of nodes and proposed CLUSTERPOW and MINPOW algorithms to mitigate the problem [11]. Wang and Li pointed out the possibility of node clustering but focused on the corresponding network partition problem [19]. Lee and Campbell also observed that the performance degrades due to the existence of hub areas, which experience excessive contention, congestion, and resource depletion [4]. Thus, nodes in these areas become the bottleneck in terms of network performance. In contrast to these prior works, our effort focuses more on in-depth study of node clustering to better understand the problem and to offer a basis for improvements.

In this paper, we model the clustered layout based on a *heavy-tail distribution* and develop the *topology generation method* based on one used in modeling the Internet [14]. This synthetic network model is then used to investigate how node clustering degrades the performance of an MP2P network via *ns-2* simulation [15]. This paper assumes to use *Ad-hoc One-demand Distance Vector* (AODV) [16] and *IEEE 802.11* [9] as the network and MAC layer protocols, respectively. Our evaluation shows that the clustered layout results in as much as 77.6% lower packet delivery ratio than the random layout based on input parameters used in our simulation study. More importantly, the clustered layout exhibits a larger variation in packet delivery service, which is critically important for QoS provisioning in MP2P networks. It also exhibits non-negligible number of "blackout" source nodes that fail to deliver any data packets to the intended destination. In order to find out the cause of the trouble with the clustered layout, *MAC* (*medium access control*) layer parameters such as success ratio of *RTS-CTS* (*Request-to-send* and *Clear-to-send*) handshake and *contention window size* are monitored during the simulation.

The organization of the paper is as follows. Section 2 discusses previous mobility models and their presumed random layout of nodes. In addition, the characteristics of clustered layout of nodes as well as its generation method are introduced. Section 3 presents the simulation results on the performance impact of the clustered layout. Finally, Section 4 concludes the paper and discusses future work.

2. Random and clustered layout of nodes

This section presents the random and the clustered layout of nodes in a MANET. Since node distribution is dictated by the underlying mobility pattern, we first review the existing mobility models developed for MANETs. We then discuss the clustered layout and its modeling and generation methodologies.

2.1. Random layout of nodes

Since node mobility significantly affects the performance of a MANET, there has been active research on characterizing the general motion behavior and developing mobility models [8,10] to be used in simulation or analysis of MANETs. One important observation in all the aforementioned mobility models is that the static property of node placement is almost identical even though they differ on how a node's or a group of nodes' dynamic movement behavior is determined. These models all produce *random layout* of nodes where nodes are well balanced and scattered across the entire MANET area. For example, in the Random Waypoint Model [10], initial positions of mobile nodes are randomly and independently selected. Even though they move, their locations in the MANET area are also quite random because the target waypoint is randomly and independently selected.

Now, consider the spatial distribution of nodes in a MANET based on the random layout. Assume that the entire area is divided into a number of equal-sized subareas. Each node is positioned in a particular subarea with independent probability p, which is the reciprocal of the number of subareas, s. The probability p_k that a subarea has exactly k nodes is given by the *binomial distribution*,

$$p_k = \binom{n}{k} p^k (1-p)^{n-k},$$

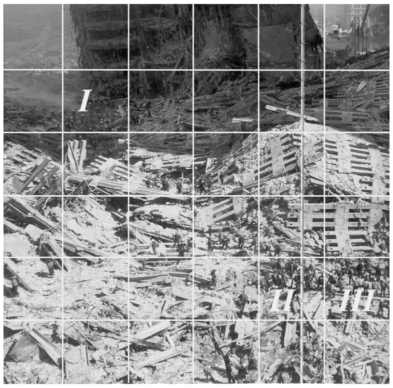
where n is the total number of nodes. As a limiting case, this becomes the well-known *Poisson distribution*

$$p_k = \frac{z^k e^{-z}}{k!},$$

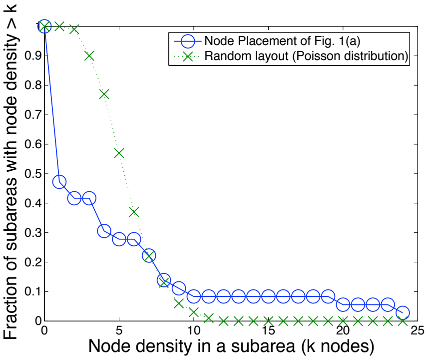
where z is the mean number of nodes in a subarea, or n/s. Both binomial and Poisson distributions are strongly peaked about the mean z, and have a large-k tail that decays rapidly as a function of $1/k!$ [7]. In other words, with the random layout of nodes, the majority of subareas have similar number of nodes and significant deviations from the average case, e.g., a subarea with a large fraction of nodes, is extremely rare.

2.2. Clustered layout of nodes

In a real network of mobile nodes, however, the node distribution can be very different from the Poisson distribution. For example, Fig. 1(a) shows an example of a disaster area where the infrastructure-less ad hoc network is well suited for supporting communication. Many rescue team members gather at three hot spot subareas, denoted as *I*, *II* and *III* in the figure, which may be a base camp or have many casualties. The three subareas out of 36 (s=*36*) include about the half of the total rescue team members (66 out of 137). Fig. 1(b) shows the node density distribution of the disaster area in Fig. 1(a) as well as that of the random layout that follows the Poisson distribution. It is clear from Fig. 1(b) that the random layout does not model the node distribution of a real ad hoc network situation. Even in the presence of node mobility, node clustering would persist because, for example in Fig. 1(a), a mobile node (i.e., a rescue team member) leaving a hot spot subarea is most likely to move to another hot spot subarea.



(a) Rescue team at Ground Zero [18]



(b) Node density distribution

Figure 1: Clustered layout in an example MP2P.

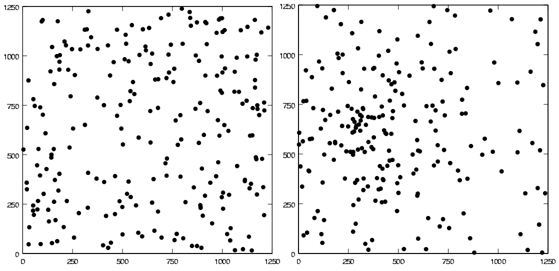
As evident in Fig. 1(b), the corresponding node distribution contains a heavy tail unlike the Poisson distribution and can be modeled by a *power-law distribution*. In general, a power-law distribution is one for which $Pr\{K>k\} \sim k^{-\alpha}$, where $0<\alpha<2$. A smaller value of α forms more concentrated clusters. If $\alpha<2$, the distribution has an infinite variance, and if $\alpha<1$, it has an infinite mean. This paper uses the simplest power-law distribution, called the *Pareto distribution*, to model the clustered layout in a MANET. In particular, we use the *Bounded Pareto distribution* in order to bound the minimum and maximum number of nodes in each subarea. This is to produce a connected network. If the upper and lower bounds are denoted as a and b, respectively, the Bounded Pareto distribution can be represented with the cumulative density function of

$$F(k) = \frac{1-(a/k)^{\alpha}}{1-(a/b)^{\alpha}},$$

where $a<k<b$, $0<\alpha<2$ [2].

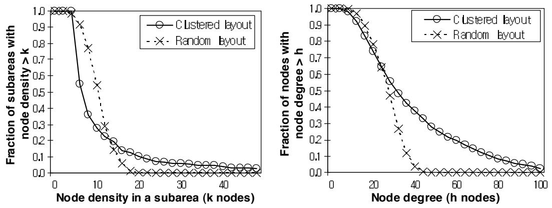
2.3. Topology generation of clustered layout

This subsection presents the procedure for generating the clustered layout in a MANET. First, the network area is divided into a number of square subareas. Then, the Bounded Pareto distribution is used to determine the number of nodes in each subarea. A subarea that happens to have a large number of nodes (heavy tail) can be considered a hot spot. Once the number of nodes in a particular subarea is determined, they are randomly positioned within that subarea. Fig. 2 shows examples of node distributions with the random and the clustered layout. The parameters used are n=250, s=25, α=1.1, a=3, and b=100, which are carefully chosen to exhibit reasonable degree of clustering (with α=1.1) and to have the average number of nodes in a subarea of 10 (250/25 with a=3 and b=100). Since the probability function $F(\cdot)$ is continuous but k must be an integer, the value p_k can be obtained by integrating the corresponding area under the probability density function derived from $F(\cdot)$.



(a) Random layout (b) Clustered layout

Figure 2: Examples node distributions (1250×1250m^2 area).



(a) Node density distribution (b) Node degree distribution

Figure 3: Comparison of topological properties.

As can be seen in Fig. 2(a), the random layout distributes the nodes to 25 subareas quite uniformly but the clustered layout produces a high concentration of nodes in a few subareas as in Fig. 2(b). To see the distribution more clearly, the average statistics is shown in Fig. 3. In Fig. 3(a), the node density distribution of the random layout decays rapidly as k increases while that of the clustered layout decays slowly with a non-negligible tail, which is also the case in Fig. 1(b). Distribution of *node degree*, defined as the number of nodes within direct communication range, is shown in Fig. 3(b). For this case, the radio transmission range of 250m and a network area of 1250×1250m^2 are assumed. With the random layout, most of the nodes have less than 40 neighboring nodes and around half of them have less than 20 neighboring nodes. With the clustered layout, however, about half of the nodes have more than 40 neighboring nodes and about 20% of the nodes have more than 80 neighboring nodes, which means these nodes are highly likely to interfere with their neighbors' communication.

While the aforementioned modeling technique is new in MANET research, a similar method has been studied to generate the Internet topology in [14]. It is also noted that the heavy-tail distributions have recently been observed from many measurement studies of computing and communication systems, where the exponential distribution has been traditionally assumed, e.g., network traffic [12], I/O traffic, Unix process lifetime [6], and file sizes in the Web [1], as well as from sociology of friendship connections [7] and Web page connectivity [3].

3. Performance evaluation

In this section, the performance of a MANET with the random and the clustered layout of nodes is evaluated using *ns-2* [15], which simulates node mobility, physical layer, radio network interfaces, the IEEE 802.11 MAC and AODV routing protocols.

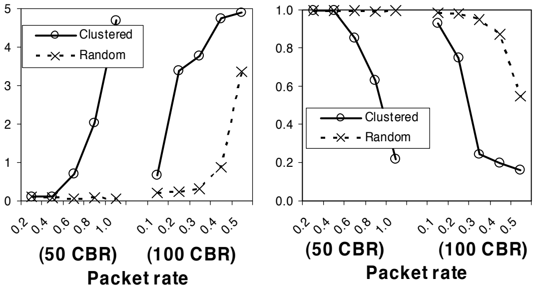
3.1. Simulation environment

Our evaluation is based on the simulation of 250 nodes located over an area of 1250×1250m^2. Since we are interested in network capacity, this paper assumes that these nodes do not move, as similarly assumed in earlier works [5, 13, 17]. The radio transmission range is assumed to be 250m and a *two-ray ground propagation channel* is assumed with a data rate of 1 Mbps. Node positions in the network area are randomly selected for the random layout of nodes as in Fig. 2(a). For the clustered layout, a Bounded Pareto distribution with parameters α=1.1, a=3, and b=100 is used to determine the number of nodes in each of 25 subareas (250×250m^2 each) as in Section 2.

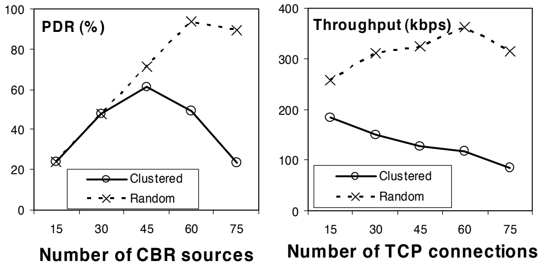
IEEE 802.11MAC protocol [9] is used with the conventional backoff scheme and RTS-CTS (Request-to-send and clear-to-send) exchange. AODV routing algorithm is used to find and maintain the routes between two end nodes. Data traffic simulated is *constant bit rate* (*CBR*) traffic: 25 to 125 CBR sources generate 256-byte data packet every 0.1~1 second. Source and destination nodes for the CBR traffic are randomly selected among the 250 mobile nodes. It is noted that simulation parameters are chosen to simulate a large-scale peer-to-peer network.

3.2. Simulation results and discussion

Network performance in terms of packet delay and *packet delivery ratio* (*PDR*) is measured during the simulation. Fig. 4(a) and 4(b) compare the average delay and PDR with 50 and 100 CBR sources. For the case of 100 CBR sources, each source transmits 0.1~0.5 packets per second. As shown in the figure, the network performance degrades faster with the clustered layout than the random layout of nodes: As much as 71.2% reduction in PDR is observed when the number of CBR sources is 100. If PDR of more than 85% is desired, the operation range is limited to 0.1 packets/s with the clustered layout while it can be increased to 0.4 packets/s with the random layout.



(a) Packet delay (sec.) (b) Packet delivery ratio (%)
Figure 4: Performance comparison (256-byte packets).

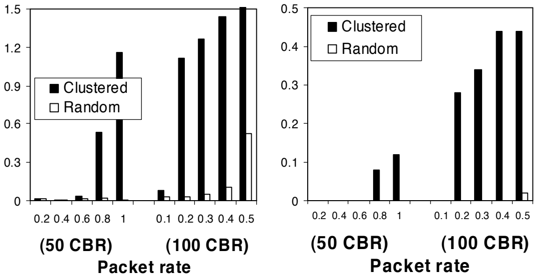


(a) With CBR traffic (b) With TCP traffic
Figure 5: Total end-to-end throughput
(0.8 packets/sec for CBR traffic, 256-byte packets).

The difference is more significant with less number (50) of CBR sources as shown also in Fig. 4. The random layout exhibits negligible degradation with the packet rate up to 1.0, while the cluster layout suffers significantly. As much as 77.6% reduction in PDR is observed. Here, higher packet rates (0.2~1.0 packets/s) are applied in order to provide the same traffic intensity as with the 100 CBR-source case. The comparison of the two cases shows that there is a noticeable performance difference between CBR sources of 50 and 100, in spite of having the same traffic intensity. This is mainly because data transmissions are more "controlled" in the 50 CBR-source case. In other words, two subsequent packets from the same source do not collide or compete with each other. The performance degrades as the number of data streams increases, which suggests that interference among the streams is a critical limiting factor in determining network capacity.

This is clearer in Fig. 5, where the number of CBR and TCP sources varies from 15 to 75. More number of data streams shows worse performance after reaching a certain threshold, which suggests again the interference among streams is a critical limiting factor in determining network capacity. With 75 CBR sources, the MANET is still operational with the random layout because PDR is 87.8%. However, with the clustered layout, more than 70% of packets are lost. Similar observation can be made with TCP traffic as shown in Fig. 5(b).

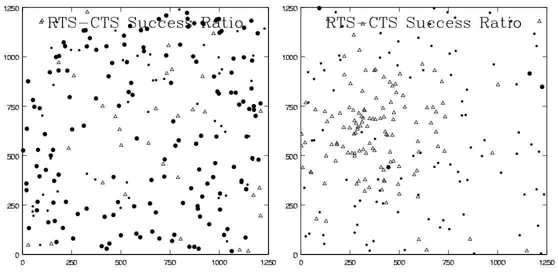
A more serious problem related to the clustered layout is QoS. It can be measured in many different ways but this paper focuses on variation in packet delivery service. Low PDR may not be a problem in certain applications but large variation in PDR limits the usability of the network especially in applications that require periodic services. Fig. 6(a) shows standard deviation of PDR for 50 and 100 CBR sources. As shown in the figure, the clustered layout results in significant variations in PDR compared to the random layout of nodes. This is an expected result because packets traversing across a hot spot area would experience severe interference, while those traversing in sparse areas would be routed with minimal contention. More importantly, we observed "blackout" CBR sources that could not deliver any packets during the simulation. Fig. 6(b) shows the percentage of these blackout sources among 50 and 100 designated sources. As many as 44% of the CBR sources are shut down with the clustered layout, while this effect is almost negligible with the random layout.



(a) Deviation of PDR (b) Ratio of "blackout" nodes
Figure 6: QoS performance (256-byte packets).

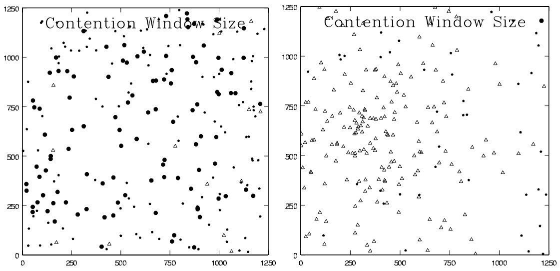
In order to investigate the cause of blackouts with the clustered layout, MAC layer parameters were monitored during the simulation study. Fig. 7 shows the success ratio of RTS-CTS handshake. When an RTS or CTS packet collides with other interfering signals, the actual data communication cannot happen. The percentage of the CTS receptions relative to the RTS transmissions is illustrated in Fig. 7(a) and 7(b) for the random and clustered layout, respectively. (Nodes that transmit less than 10 RTS packets are not included in this graph.) 100 CBR sources and 0.2 packet rate was used for this experiment. For the random layout, more than half of the nodes are successful in RTS-CTS handshaking more than 60% of the time (marked as large dots in Fig.7(a)). In comparison, for the clustered layout, most of the nodes receive a CTS packet less than 30% of the time in response to a RTS packet (marked as triangles in Fig.7(b)).

Another MAC layer parameter, the contention window size, was also monitored. When a packet collides, each node adjusts its contention window size to reduce the chance of further collisions. In our simulation study, the minimum window size is 32 and is doubled whenever a collision occurs until the maximum window size (1024) is reached. Fig. 8 shows the average contention window size of each node. This average is obtained by sampling the window size when each node decides to transmit a packet. As shown in Fig. 8(a) and 8(b), the contention window size is smaller than 64 for most of the nodes with the random layout (marked as large dots in Fig.8(a)), while it is mostly larger than 160 with the clustered layout (marked as triangles in Fig.8(b)). With Fig. 7 and 8, it can be concluded that the MAC layer protocol suffers when nodes are clustered rather than scattered in the network.



(a) Random layout (b) Clustered layout

Figure 7: Success ratio of RTS-CTS handshake (Triangle: <30%, small dot: 30~60%, large dot: >60%).



(a) Random layout (b) Clustered layout

Figure 8: Average contention windows size (Triangle: >160 slots, small dot: 64~160 slots, large dot: <64 slots).

4. Conclusions and future work

This paper studied capacity scalability of a multihop ad hoc network when node distribution is not random. We characterized and modeled the clustered layout of nodes based on topology generation method with a heavy-tail distribution. Based on extensive simulation using ns-2 network simulator, it has been shown that the clustered layout resulted in a serious degradation not only in average performance, such as delay and packet delivery ratio, but also with QoS metrics such as variation in packet delivery service and the number of blackout nodes. It can be concluded that the clustered layout easily saturates a MANET with less traffic intensity and less number of data streams than the random layout. In-depth monitoring of MAC layer parameters revealed that implementation of adaptive capability according to the traffic intensity at the MAC layer is desirable in order to provide consistent performance irrespective of node distribution. We are currently investigating the effective measures to improve the network performance in the presence of node clustering.

References

[1] Arlitt, M. F. and Williamson, C. L., “Web server workload characterization: The search for invariants,” *ACM SIGMETRICS*, pp. 126-137, 1996.

[2] Bansal, N. and Harchol-Balter, M., “Analysis of SRPT Scheduling: Investigating Unfairness,” *ACM SIGMETRICS/Performance*, pp. 279-290, 2001.

[3] Barabasi, A.-L., *Linked: The New Science of Networks*, Perseus Publishing, Cambridge, MA, 2002.

[4] Lee, S. and Campbell, A, “HMP: Hotspot Mitigation Protocol for Mobile Ad Hoc Networks,” *11th IEEE/IFIP International Workshop on Quality of Service, Monterey, Canada*, June 2003.

[5] Gupta, P. and Kumar, P. R., “The Capacity of Wireless Networks,” *IEEE Transactions on Information Theory*, Vol. 46, No. 2, pp. 388-404, March 2000.

[6] Harchol-Balter, M. and Downey, A. B., “Exploiting Process Lifetime Distributions for Dynamic Load Balancing,” *ACM SIGMETRICS*, pp. 13-24, 1996.

[7] Watts, D. J. and Strogatz, S. H., “Collective Dynamics of Small-World Networks,” *Nature*, Vol. 393, pp. 440-442, 1998.

[8] Hong, X. *et al.*, “A Mobility Framework for Ad Hoc Wireless Networks,” *Mobility Data Management*, 2001.

[9] IEEE Std 802.11-1999, Local and Metropolitan Area Network, Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications.

[10] Johnson, D. and Maltz, D., “Dynamic Source Routing in Ad Hoc Wireless Networks,” *Mobile Computing*, edited by T. Imielinski and H. Korth, pp. 153-181, Kluwer Academic Pub., 1996.

[11] Kawadia, V. and Kumar, P. R., “Power Control and Clustering in Ad Hoc Networks,” *IEEE Infocom*, 2003.

[12] Leland, W. E. *et al.*, “On the Self-Similar Nature of Ethernet Traffic,” *IEEE/ACM Transactions on Networking*, No. 2, pp. 1-15, 1994.

[13] Li, J. *et al.*, “Capacity of Ad Hoc Wireless Networks,” *ACM/IEEE MobiCom* 2001, pp. 61-69, 2001.

[14] Medina, A. *et al.*, “On the Origin of Power Laws in Internet Topologies,” *ACM SIGCOMM Computer Communication Review*, Vol. 30, No. 2, Apr. 2000.

[15] ns-2 Network Simulator, http://www.isi.edu/nsnam/ns/.

[16] Perkins, C and Royer, E., “Ad-hoc On-Demand Distance Vector Routing,” *IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90-100, 1999.

[17] Shepard, T. J., “A Channel Access Scheme for Large Dense Packet Radio Networks,” *ACM SIGCOMM’96*, 1996.

[18] Sullivan, R. (Editor), *One Nation: America Remembers September 11, 2001*, Time Warner Trade Publishing, 2001.

[19] Wang, K. H., and Li, B., “Group Mobility and Partition Prediction in Wireless Ad-Hoc Networks,” *IEEE ICC 2002*, Vol. 2, pp. 1017-1021, 2002.