# CAPACITY AND SCALE-FREE DYNAMICS OF EVOLVING WIRELESS NETWORKS

A Thesis

by

BHARAT VISHWANATHAN IYER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2003

Major Subject: Electrical Engineering

CAPACITY AND SCALE-FREE DYNAMICS OF EVOLVING

WIRELESS NETWORKS

A Thesis

by

BHARAT VISHWANATHAN IYER

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

| | |
|---|---|
| Dmitri Loguinov<br>(Co-Chair of Committee) | A. L. Narasimha Reddy<br>(Co-Chair of Committee) |
| Shankar Bhattacharyya<br>(Member) | Deepa Kundur<br>(Member) |
| | Chanan Singh<br>(Head of Department) |

December 2003

Major Subject: Electrical Engineering

# ABSTRACT

Capacity and Scale-Free Dynamics of Evolving Wireless Networks.

(December 2003)

Bharat Vishwanathan Iyer, B.E., Maharaja Sayajirao University, Baroda, India

Co-Chairs of Advisory Committee: Dr. Dmitri Loguinov
Dr. A. L. Narasimha Reddy

Many large-scale random graphs (e.g., the Internet) exhibit complex topology, non-homogeneous spatial node distribution, and preferential attachment of new nodes. Current topology models for ad-hoc networks mostly consider a uniform spatial distribution of nodes and do not capture the dynamics of evolving, real-world graphs, in which nodes "gravitate" toward popular locations and self-organize into non-uniform clusters. In this thesis, we first investigate two constraints on scalability of ad-hoc networks – network reliability and node capacity. Unlike other studies, we analyze network resilience to node and link failure with an emphasis on the growth (i.e., evolution) dynamics of the entire system. Along the way, we also study important graph-theoretic properties of ad-hoc networks (including the clustering coefficient and the expected path length) and strengthen our generic understanding of these systems. Finally, recognizing that under existing uniform models future ad-hoc networks cannot scale beyond trivial sizes, we argue that ad-hoc networks should be modeled from an evolution standpoint, which takes into account the well-known "clustering" phenomena observed in all real-world graphs. This model is likely to describe how future ad-hoc networks will self-organize since it is well documented that information content distribution among end-users (as well as among spatial locations) is non-uniform (often heavy-tailed). Results show that node capacity in the proposed evolution model scales to larger network sizes than in traditional approaches, which suggest that non-uniformly clustered, self-organizing, very large-scale ad-hoc networks may become feasible in the future.

To my beloved family

# ACKNOWLEDGMENTS

I sincerely thank my research advisor, Dr. Dmitri Loguinov, for his valuable insights and regular discussions on various issues pertaining to this work. Further, I would also like to thank him for his reviews and suggestions in the writing of this thesis. I would also like to acknowledge Dr. Nitin Vaidya for introducing me to the research issues pertaining to wireless ad-hoc networks. I would like to thank Dr. Garth Hillman for setting an example of dedication and consistency towards one's work. Finally, I am grateful to my mother for supporting and encouraging me throughout my work.

# TABLE OF CONTENTS

# LIST OF FIGURES

# I. INTRODUCTION

Recent modeling efforts to understand the capacity of wireless networks have painted a rather bleak future for ad-hoc networks. As the size of these networks increases, the throughput available to each node asymptotically decays to zero under all practical conditions (several exceptions exist [16]; however, their mobility models and delay constraints are hardly practical). It appears that research in this area has reached a point of finally examining a fundamental question of *whether future ad-hoc networks will be deployed with millions of nodes in a single, very large-scale system, or will always be limited to trivial sizes (e.g., wireless LANs)?*

We find that the answer to this question heavily depends on the underlying modeling assumptions and characteristics of user traffic in these networks. It appears that if user traffic patterns are completely *random* and perfectly *uniform*, then large-scale ad-hoc networks are impractical; however, the good news is that very few real-world networks exhibit completely uniform all-to-all communication patterns. This observation leads us to developing a new paradigm for future ad-hoc networks in which wireless nodes self-organize around important sources of "information." The term "information" is rather abstract and can represent a variety of concepts including interesting speakers at a conference, large cities in a given country, several users in peer-to-peer networks that share the majority of files, popular websites such as CNN or Google, etc.

In the first half of the thesis, we examine the properties of existing uniform models

---

This thesis follows the style and format of *IEEE/ACM* Transactions on Networking.

of wireless ad-hoc networks. Besides the available node capacity often studied in this area, another important metric related to scalability of a wireless network is its *reliability* (i.e., resilience to node/edge failure). Wireless connectivity is often compromised due to varying channel conditions, finite node lifetime, and node mobility, all of which can result in partitioning of the network into small disjoint clusters. Unfortunately, network resilience often comes at the expense of throughput since an increased transmission range not only leads to more neighbors and higher resilience to node/edge failure, but also reduces spatial concurrency and available node throughput.

In the second half of the thesis, we model the growth process of ad-hoc networks in an attempt to understand topology characteristics of very large-scale self-organizing networks. We realize that real networks of humans [39] and smart devices generally exhibit a non-homogenous spatial distribution of nodes and non-uniform traffic patterns. Humans in social gatherings, public places (e.g. shopping malls, conference halls, entertainment centers) exhibit *preferential connectivity* and *clustering*. The nodes in such gatherings tend to cluster around "popular" nodes or important information sources. Furthermore, it is well documented [37] that shared (i.e., publicly-available) information is *not* uniformly distributed and is typically concentrated in a small number of users/locations. Another study [27] on mobile users suggested that there is spatial locality in information querying, i.e., users with similar interests are physically close to each other. These observations motivate us to study an *information-centric* model where ad-hoc nodes form clusters and exhibit preferential traffic exchange *only* with the nodes that belong to the same cluster. We show that this model scales to an arbitrarily large number of nodes and provides an accurate reflection of typical scenarios of human interaction, as well as smart-device deployment (e.g., in certain sensor networks). We study the formation of clusters by developing a scale-free network evolution model, analyze cluster size distribution, and understand the scalability issues of two types of node-attachment functions. In the end, we discuss a new paradigm of “bringing services to the user” that allows ad-hoc traffic to be localized within each cluster.

The rest of the thesis is organized as follows. In section II, we present some background and previous work on ad-hoc topology modeling and network scalability. In section III, we explain the topology models and assumptions considered in this thesis. Sections IV and V discuss important topology characteristics of traditional wireless ad-hoc networks. In section VI, we explain our information-centric paradigm for large-scale wireless ad-hoc networks. Finally, in section VII we conclude the thesis.

## II. BACKGROUND

Applications of ad-hoc networks are just beginning to be realized. Ad-hoc networks provide a valuable alternative in situations where cost and time constraints do not allow for deployment of fixed network infrastructure (e.g., battlefields and catastrophe-control situations). Another possible application of ad-hoc node formations is a collection of "smart homes" where computers, home appliances, door locks, water sprinklers, and other "smart devices" are interconnected by a wireless network.

Wireless ad-hoc networks as shown in Figure 1 consist of a number of nodes that communicate in a decentralized and self-organizing manner. Such networks do not rely on a fixed infrastructure and hence, can be deployed quickly and with minimal cost. In addition to low-overhead deployment, ad-hoc topologies exhibit increased spatial concurrency, frequency reuse, and total network capacity compared to regular cellular networks. These advantages make ad-hoc topologies a cost-effective option for large wireless networks.

Several other options also exist as topology models for large-scale wireless data networks. In a wireless *cellular* network or wireless *LAN*, nodes communicate with each other through base stations (often called *access points*). These base stations are assumed to be connected by a high-bandwidth wired network and act as relays for the wireless nodes. An access point is reachable in a *single* hop by any node in the cell.

In contrast with traditional cellular models, *hybrid* networks allow *multi-hop* cells, in which data is forwarded in a multi-hop fashion over wireless nodes in the same cell and through the base-station infrastructure to nodes belonging to different cells. A disadvantage of the infrastructure-based model is the cost and deployment time of base-stations. Additional problems with cellular wireless networks arise when mobile radio propagation is hampered by localized co-channel interference or signal shadowing from certain natural and/or man-made obstacles resulting in deterioration of network connectivity. As shown in Figure 2, node *x* does not have a direct link to the base-station due to an obstacle in the signal path. In such situations, spatial diversity allows the node to connect to the base-station through one of multiple nodes in the neighborhood.

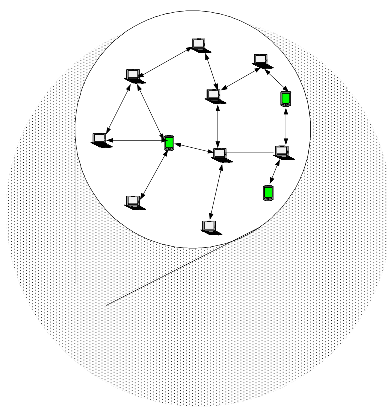

**Figure 1 An ad-hoc network with $N$ nodes in an area $A$.**

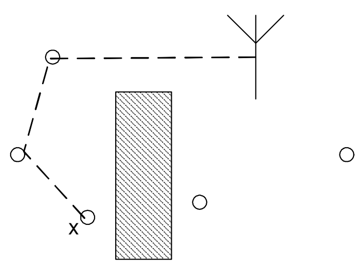


**Figure 2 Node $x$ does not have a direct connection with the base station.**

Several authors [36], [23], [23] have studied scalability and capacity constraints of ad-hoc networks. In [36], the authors showed that none of the current routing protocols scale well with node mobility and network size. In [23], Kawadia *et al.* consider the problem of power control and spatial reuse when nodes are non-homogeneously distributed in space. In [17], Gupta *et al.* study throughput capacity of ad-hoc nodes, uniformly distributed in a given area, under the assumption that each node continuously sends data to a randomly chosen destination. Given channel capacity $W$ and fixed transmission range, the throughput capacity per-node is shown to be $\Theta(W / \sqrt{N \log N})$, where $N$ is the number of nodes in the network. This result demonstrates that traffic rate per source-destination pair does not scale well in large ad-hoc networks and asymptotically decays to zero.

In contrast to a fixed ad-hoc network, Grossglauser *et al.* [16] show that by restricting the traffic path length, node capacity can be kept constant. In this model, traffic is spread to intermediate relay nodes to allow multiple 'two-hop routes' between each source and destination. The drawback of the approach is that it requires a complete mixing of trajectories of wireless nodes in the network and the delay experienced by the packets is large (the delay further grows with the size of the network).

1. Capacity and Scalability

Recently, as wireless technology has matured and wireless connectivity increased, scalability and capacity constraints of ad-hoc networks have come into focus. In [17], [20], [19], [23], the authors studied the throughput capacity of ad-hoc nodes uniformly distributed in a given area. Under the assumption that each node continuously sends data to a randomly chosen destination, Gupta *et al.* [17] concluded that the per-node throughput scales as $\Theta(1/\sqrt{N \log N})$, where $N$ is the number of nodes in the network. This result demonstrates that traffic rate per source-destination pair does not scale well in large ad-hoc networks and asymptotically decays to zero.

In contrast to fixed ad-hoc networks, the authors in [16], [33], [3] have shown that by restricting the traffic path length, node capacity could be kept constant. Grossglauser *et al*. [16] presented a model where traffic is spread to intermediate relay nodes to allow multiple "two-hop routes" between each source and destination. The drawback of this approach is that it requires a complete mixing of trajectories of wireless nodes in the network and the delay experienced by the packets is large (the delay further grows with the size of the network).

In [23], [21] the authors analyzed the hybrid model for wireless networks. Liu *et al*. [23] have shown a linear increase in total network capacity with the number of base-stations in the network. However, they do not focus on per-node capacity and consider aggregate network throughput while observing the effect of the number of base-stations on capacity. Several other authors [36], [43], [30], [22], [4] have examined physical layer and protocol enhancements for capacity optimizations. However, their work does not provide a precise model for large, scalable ad-hoc networks.

In [7], [8], [32] the authors adopt a graph-theoretic approach to understanding connectivity and other topology characteristics of wireless ad-hoc model. However, several interesting properties such as node/edge expansion and the clustering coefficient have not been studied. Another aspect that is missing from current literature on wireless ad-hoc networks is an evolving network model where the network grows continuously due to addition of new nodes. We provide such a model and show that under certain traffic assumptions, the node capacity in such networks scales with network size.

2. Large-Scale Networks and Models

Large-scale, complex systems such as social networks, scientific-collaboration networks and computer networks have been analyzed and modeled by a number of authors [6], [5], [11], [9]. Faloutsos [15] studied the degree distribution in the Internet and showed that its node degree follows a power law with an exponent close to 1.2. In addition, Simon [38], [39] examined the segregation of individuals/objects in groups and showed that the empirical distribution of city sizes by population approximates a power-law relationship with scaling exponent close to 2. In [5], Barabasi suggested that the heavy-tailed nature of node degree in networks is due to two generic mechanisms shared by many real networks: growth of the network due to addition of new nodes and preferential attachment[1] of new nodes to existing nodes. We examine the underlying common characteristics of such large-scale, self-organizing systems with respect to ad-hoc topology in V.

[1] This observation is a result of the “popularity is attractive” and “rich gets richer” phenomena observed in the real world [24].

## III. MODEL PRELIMINARIES

In mathematical terms, a network is represented by a graph, $G=\{V, E\}$, where $V$ is the set of nodes and $E$ is the set of edges.

*Ad-hoc Network Model*: The model considered in the thesis is based on the following assumptions.

Let $N = |V|$ be the number of nodes deployed in some area $A$. Without loss of generality, assume that $A$ is a unit disc of radius $1/\sqrt{\pi}$ and all $N$ nodes are independently and uniformly distributed in $A$. Let $\rho = N / A = N$ be the average network density. Then, the expected (average) number of nodes inside any area $A_1$ ($A_1 \subset A$) is $\rho A_1$.

Assume that all nodes in the network employ a common range $r$ for all their transmissions. As the network size increases, the average degree $k$ of the network is $\Theta(\log N)$ and the node transmission radius $r$ is $\Theta\left(\sqrt{\frac{\log N}{\pi N}}\right)$.

For the interference model, we adopt the protocol model[2] as developed in [17]. A transmission from node $X_i$ is successfully received by node $X_j$ if the following two conditions are satisfied:

Node $X_j$ is within the transmission range $r$ of node $X_i$:

$$\left|X_i - X_j\right| \le r. \tag{1}$$

where $\left|X_i - X_j\right|$ represents the distance between node $X_i$ and node $X_j$.

[2] We choose this interference model over the physical model to clearly realize the results without going into physical layer technicalities.

For every other node $X_k$ that is simultaneously transmitting over the same channel, we require

$$\left|X_k - X_j\right| \geq (1+\Delta)\left|X_i - X_j\right|. \tag{2}$$

This condition guarantees a guard zone around the receiving node and prevents a neighboring node from transmitting at the same time. The radius of the guard zone is $(1+\Delta)$ times the separation between the sender and receiver. The parameter $\Delta$ defines the size of the guard zone and governs the number of simultaneous transmissions in the network.

*Random Graph Model* ($G(N, p)$ *model*): This model as introduced by Erdős and Renyi [13] consists of a graph on $N$ vertices in which every pair of vertices is connected with a fixed probability $p$. The expected node degree in this model, if self-loops are not allowed, is $p(N-1)$ and the expected number of edges in the graph is $pN(N-1)/2$.

# IV. TOPOLOGY PROPERTIES

Wireless ad-hoc networks and other peer-to-peer architectures (e.g., Chord, CAN) are characterized by frequent changes in node connectivity. This observation is attributed to node mobility or node on/off in wireless networks and finite node lifetime in peer-to-peer networks. As the network grows in size, the number of link breakages increase quickly which proves to be a deterrent for network scalability. In this section, we analyze the resilience of ad-hoc networks and the available per-node capacity in different network topologies. The emphasis is on understanding the constraints on scalability of ad-hoc networks under current topology assumptions and explaining the need for a better large-scale network model.

## 1. Graph Expansion and Resilience

Graph expansion determines how fast a graph finds "unknown" nodes. Specifically, graph edge expansion gives us useful information about the graph resilience when edges are expected to fail (i.e., 4 edges are better than 2 as depicted in Figure 3). In wireless networks, path loss fluctuations due to mobility, large-scale and small-scale fading can result in frequent edge failure.

As noted in [27], edge expansion determines the number of edges that link a group of vertices to the rest of the graph (or similarly the number of edges inside the group of nodes); however it does not tell us the number of nodes on the other side of the cut between this group of vertices and the rest of the graph. In Figure 3, there are 4 edges leaving the neighborhood of node $v$, but they only link to 2 unique nodes, thus leading to a good amount of overlap. Thus, regardless of the number of edges crossing the cut, the number of

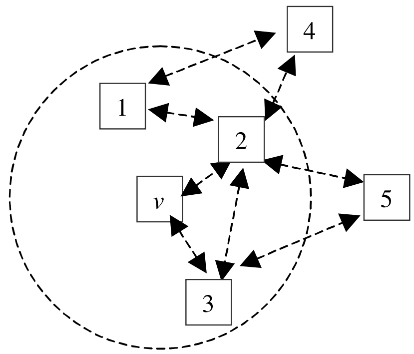


**Figure 3 Edge expansion and path overlap due to high clustering.**

nodes on the other side of the cut determines the strength of the neighborhood. "*Node failure*" due to node mobility or power on/off is also a well-observed fact in ad-hoc networks. Hence, we also examine *node* expansion of the network.

First, we formally define graph node/edge expansion and then derive expressions for the same in ad-hoc networks. Consider a graph $G = (V, E)$ and a set of vertices $S \subset V$. Define the set of all edges between $S$ and $V \backslash S$ to be $\partial_e S = \{(u, v): u \in S, v \in V \backslash S\}$. The set $\partial_e S$ is called the *edge boundary* of $S$ and further define edge expansion $i(S)$ as the ratio of the size of $\partial_e S$ to the size of $S$:

$$i(S) = \frac{|\partial_e S|}{|S|}. \qquad (3)$$

*Definition 1*: Graph edge expansion $i(G)$ (sometimes called the isoperimetric number of the graph) is the minimum of $i(S)$ for all sets $S \subset V, |S| \leq |V|/2$.

*Definition 2*: Consider graph $G = (V, E)$ and set $S \subset V$. Define the node boundary of S to be $\partial_n S = \{v : (u, v) \in E, u \in V, v \in V \setminus S\}$. Node expansion $h(G)$ of the graph is defined as:

$$h(G) = \min_{\{S:|S| \le |V|/2\}} \frac{|\partial_n S|}{|S|}. \tag{4}$$

*Lemma 1*: Node expansion $h(G)$ of wireless ad-hoc networks is upper limited by the following:

$$h(G) \le 2\sqrt{\pi}(\sqrt{2}r + \sqrt{\pi}r^2). \tag{5}$$

*Proof*: Consider $S$ to be the set of nodes that lie within distance $x$ units from a node $v$ (i.e., a ball of radius $x$). Then, the size of $S$ is:

$$|S| = N\pi x^2. \tag{6}$$

Hence, the set $\partial_n S$ consist of nodes that lie between $x$ and $(x+r)$ units from node $v$. Thus,

$$\begin{aligned} |\partial_n S| &\le N\pi(x+r)^2 - N\pi x^2 \\ &= N\pi r(2x+r). \end{aligned}$$

From (4), we obtain that node expansion for set $S$ is no more than:

$$h(S) \le \frac{r(2x+r)}{x^2}. \tag{7}$$

Noting that the nodes lie in a unit area, the minimum of (5) over all sets $S$ $(|S| \le |V|/2 = N/2)$ occurs for $x = \sqrt{1/2\pi}$, which covers half the area and contains $N/2$ nodes.

$$h(G) \le 2\sqrt{\pi}(\sqrt{2}r + \sqrt{\pi}r^2).$$

As the number of nodes $N$ increases in a given area, radius $r$ decreases as a function of $\sqrt{\log N/\pi N}$ resulting in a decrease in $h(G)$ for the entire graph as shown in Figure 4.

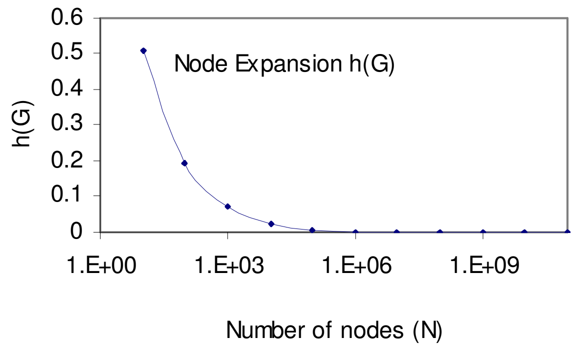


**Figure 4 Node expansion for ad-hoc networks with increasing number of nodes $N$ . $x$ -axis is shown in log scale.**

*Lemma 2*: Edge expansion of wireless ad-hoc networks is bounded by:

$$i(G) \le N\pi^{3/2}\left(\sqrt{2}r^3 + \sqrt{\pi}r^4\right).$$

*Proof*: The edge boundary for the set $S$ as defined in (6) is given by:

$$\begin{aligned} \left|\partial_e S\right| &\le N\pi r(2x+r) \times N(\pi r^2/2) \\ &= \frac{1}{2}N^2 r^3 \pi^2 (2x+r). \end{aligned}$$

The term $N(\pi r^2/2)$ gives the expected edges for a single node across the node boundary.

Hence, edge expansion for set $S$ is given by:

$$i(S) \le \frac{Nr^3\pi(2x+r)}{2x^2}. \tag{8}$$

Again taking the minimum over all possible set $S$ $(|S| \le |V|/2)$, we get

$$i(G) \le N\pi^{3/2}\left(\sqrt{2}r^3 + \sqrt{\pi}r^4\right). \tag{9}$$

Graph edge expansion is shown in Figure 5 as a function of the number of nodes. As observed, edge expansion drops fast with increasing network size. This is attributed to the fact that the radius $r$ decreases with increase in the number of nodes in the network as $\sqrt{\log N / \pi N}$. Graph bisection width $bw(G)$ is a well known graph resilience property and can be related to edge expansion as:

$$bw(G) = i(G)\frac{N}{2} \qquad (10)$$

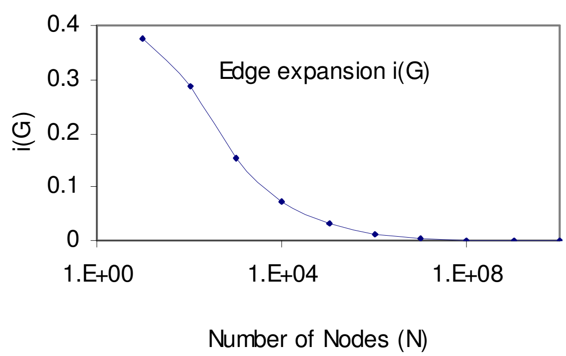


**Figure 5 Graph edge expansion with increasing number of nodes $N$. $x$-axis is shown in log scale.**

2. Node Capacity

The capacity of the network depends on many aspects of the network: network architecture, power and bandwidth constraints, routing strategy, radio channel characteristics, etc. We examine the available node capacity in the uniform ad-hoc model and hybrid network model in order to understand the capacity limitations of these topologies. We derive a precise expression of *per-node* capacity in hybrid networks, which is missing from [23].

Recall that for $N$ nodes in the network with random traffic patterns, the achieved per-node capacity is necessarily smaller than that for the optimal network. An optimal network is defined as a network with optimally placed nodes, optimally assigned node range and traffic pattern. From [17], we obtain an upper bound on per-node throughput capacity $\lambda$ as:

$$\lambda = O\left(\frac{W}{\sqrt{N}}\right). \tag{11}$$

We know that as the number of nodes in the network increases, the radius decreases as $\Theta\left(\sqrt{\log N / \pi N}\right)$ resulting in the average path length in hops to increase as $\Theta(\sqrt{N})$. This translates into an increase in the relay load on the nodes of the network and a decrease in the useful node capacity given by (11). Thus, one option to improve the node capacity in ad-hoc networks involves reducing the average path length of the network as achieved by hybrid networks.

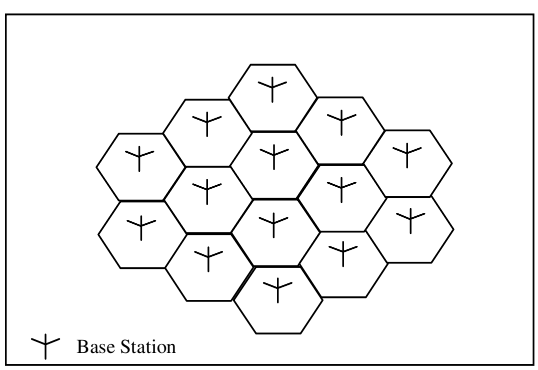


**Figure 6 Hybrid Wireless Network Model.**

Suppose that the base-stations (as shown in Figure 6) have the same transmission range $r$ as other nodes in the network and that the channel capacity $W$ is allocated to three non-interfering sub-channels of capacity $W_1$ – for ad-hoc mode transmissions within a cell and $W_2$ (=$W_3$) – for uplink (downlink) transmissions through the base station infrastructure[3].

$$W = W_1 + W_2 + W_3, \quad W_2 = W_3. \tag{12}$$

For a hybrid network of $N$ nodes uniformly and independently distributed in $m$ cells, the probability that the source-destination pair lie in the same cell is $1/m^2$. Hence, the nodes employ ad-hoc mode transmission with probability $1/m^2$ and infrastructure mode transmission with probability $1-1/m^2$.

Effective node capacity $\lambda$ can be expressed as a sum of two components: *ad-hoc* mode capacity $\lambda_a$ and *infrastructure* mode capacity $\lambda_i$.

$$\lambda = (1/m^2)\lambda_a + (1-1/m^2)\lambda_i. \tag{13}$$

Now, the expected number of ad-hoc mode transmissions in single cell is $N/m^2$. Hence, from (11), we obtain the ad-hoc mode capacity as:

$$\lambda_a = O\left(\frac{W_1}{\sqrt{N/m^2}}\right). \tag{14}$$

Since all infrastructure mode traffic is directed to the base-station, allowing multi-hop communication with the base-station does not alleviate the bottleneck. Thus, we assume that the nodes within a cell can reach the base station in a single hop and that a

[3] As noted in [23], every data packet that uses the infrastructure mode transmission goes through an uplink and downlink transmission. Hence, we choose $W_2 = W_3$. For the same reason, capacity estimation includes either uplink or downlink transfer.

round-robin node schedule gives a total per-cell throughput of $W_2$. With the expected number of infrastructure mode transmissions in a single cell given by $N(1-1/m^2)/m$, the per-node infrastructure mode capacity is bounded by:

$$\lambda_i = \Theta\left(\frac{mW_2}{N(1-1/m^2)}\right). \tag{15}$$

Hence, from (13), the effective per-node capacity is:

$$\lambda = O\left(\frac{W_1}{m\sqrt{N}}\right) + \Theta\left(\frac{W_2}{N/m}\right). \tag{16}$$

The denominators for the two terms in (16) suggest that with increasing values of *m*, the first term goes to 0. Thus, for optimal node capacity, using (12), we require channel assignments such that $W_1/W \to 0$. Note that *m*=1 transforms the network to an ad-hoc topology requiring $W_2/W \to 0$. From (12), for large *m* we obtain a per-node capacity of:

$$\lambda = \Theta(\frac{mW}{N}). \tag{17}$$

Thus, (17) suggests that to maintain a constant per-node capacity with increasing network size, we require $N/m \sim \text{constant}$ i.e., the number of nodes in a cell remains fixed. This entails a linear increase in *m* with the number of nodes in the network. The node capacity improvement comes at an increased cost of network infrastructure and transforms the ad-hoc topology to a cellular structure. In the next section, we examine wireless ad-hoc topology properties in context with other large-scale networks.

# V. WIRELESS AD HOC NETWORKS AND SMALL WORLDS

We examine the underlying traits of large-scale, self-organizing systems with respect to ad-hoc topology in an attempt to identify a topology model for large wireless data networks. Many real-world complex systems (such as the chemical-reaction networks [1], social networks [2], scientific-collaboration networks [40], and the Internet [10]) build topologies that exhibit the "small world" phenomena. Recall that a graph is a *small-world* graph if, for a given number of nodes *N*, it has a much higher clustering coefficient and similar average distance compared to a *random* graph with the same number of nodes.

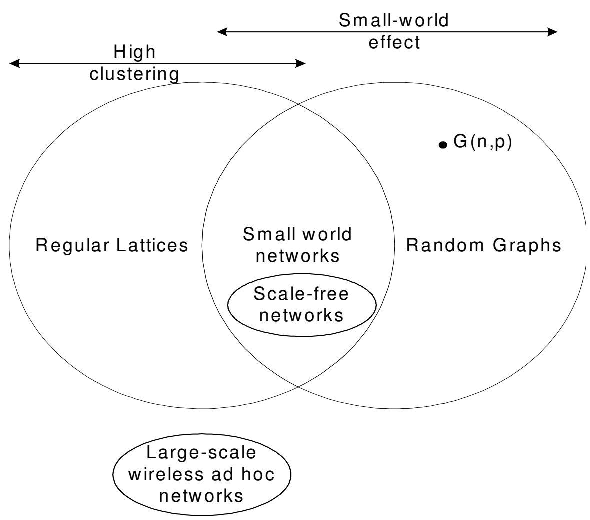


**Figure 7 Ad-hoc networks exhibit high clustering but large diameter and large average path length.**

As shown in Figure 7, small-world networks can be viewed as a superposition of regular lattices and purely random graphs [42]. These networks combine high clustering of regular lattices and the small-world effect (i.e., low diameter) typical to many random graphs. The main question studied in this thesis is *where do ad-hoc networks belong in* Figure 7 *(random, structured, or both)?* In fact, current uniform node-distribution models belong to the class of *random geometric* graphs $G(N,r)$ and have many properties very similar to those of the ER model.

Next recall that *scale-free* networks [6], [1], [5] emerge as the primary model of evolving graphs, in which new nodes connect preferentially to the popular (or highly connected) nodes in the network. Scale-free networks are also exhibit short node-separation distances [5], [10]. In this section, we explore clustering and path length characteristics of uniform ad-hoc networks to better understand the typical ad-hoc topology as compared to that of other existing complex networks.

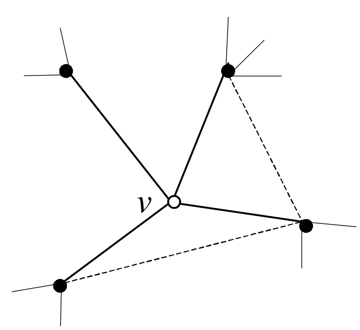


**Figure 8 Illustration for clustering coefficient. For the vertex $v$ shown, clustering coefficient is 2/6=1/3.**

1. Clustering Coefficient

Recall that the clustering coefficient $\gamma$ characterizes the density of connections in the local neighborhood of a vertex. Suppose that the network is undirected and one of its vertices $v$ has $z$ neighbors (this is shown in Figure 8 for $z = 4$). The maximal clustering is achieved if all $z(z-1)/2$ possible edges between the neighbors of $v$ exist. The clustering coefficient captures the probability that there exists an edge between two neighbors of a randomly chosen vertex and relates to the density of 3-cycles in the network.

*Definition 3*: Given a graph $G = (V, E)$, node $v \in V$, and its neighborhood $\Gamma(v) = \{u : (u, v) \in E\}$, the clustering coefficient of node $v$ is defined as the ratio of the number of links $L(\Gamma)$ that are entirely contained in $\Gamma(v)$ to the maximum possible number of such links.

$$\gamma(v) = \frac{L(\Gamma(v))}{|\Gamma(v)|(|\Gamma(v)|-1)/2}. \tag{18}$$

Graph clustering $\gamma(G)$ is the average of $\gamma(v)$ for all vertices with degree at least 2. We determine the clustering coefficient for wireless ad-hoc networks as follows.

*Lemma 3*: The expected clustering coefficient of wireless ad-hoc network is given by:

$$\gamma(G) = \frac{0.69(\bar{k}+1)-2}{\bar{k}-\pi r^2}. \tag{19}$$

where $\bar{k}$ is the expected node degree.

*Proof*: For any pair of neighbors $v \in V$ and $u \in \Gamma(v)$, let $A_u$ be the expected overlap area of circles centered at $v$ and $u$ as shown in Figure 9. Thus, $\rho A_u - 2$ gives the expected number of nodes in the set $C = \{w: w \in V, w \in \Gamma(v), w \in \Gamma(u)\}$, which contains neighbors common to both $u$ and $v$. Therefore, the expected number of undirected edges between the neighbors of node $v$ is given by:

$$C_v = \frac{1}{2}\sum_{u=1}^{\bar{k}}(\rho A_u - 2) = \frac{1}{2}\rho\sum_{u=1}^{\bar{k}} A_u - \bar{k}, \tag{20}$$

where the expected node degree is given by:

$$\bar{k} = \rho\pi r^2 - 1, \tag{21}$$

(the term $\rho\pi r^2$ in (21) represents the expected number of nodes in a circle of radius $r$). From Figure 9, the shaded overlap region between the two circles is given by $2(A - B)$, where $A = \widehat{acb}$ is the sector of the circle swept by angle $\theta = 2\cos^{-1}(x/2r)$, and $B$ is the area of the triangle $\Delta abc$. Now, notice that $A = 1/2\theta r^2$ and $B = r^2\cos(\theta/2)\sin(\theta/2) = 1/2r^2\sin\theta$, which gives the area of the overlap as $r^2(\theta - \sin\theta)$. The assumptions of a uniform spatial distribution of nodes suggests that the neighboring nodes are equally likely to lie at distance uniformly distributed between 0 and $r$ from node $v$. Hence, the expected area of the overlap, $A_u$, is obtained as:

$$A_u = \frac{1}{r}\int_0^r r^2(\theta - \sin\theta)dx = 2.16r^2. \tag{22}$$

Using (20), we get $C_v = \frac{1}{2}[\rho(2.16r^2) - 2\bar{k}]$ and the expected clustering coefficient of node $v$ is given by:

$$
\begin{aligned}
E[\gamma(v)] &= \frac{C_v}{E\left[|\Gamma(v)|(|\Gamma(v)|-1)/2)\right]} \\
&= \frac{C_v}{\frac{1}{2}E\left[|\Gamma(v)|^2-|\Gamma(v)|\right]} \\
&= \frac{C_v}{\frac{1}{2}E\left[|\Gamma(v)|^2\right]+Var\left(|\Gamma(v)|\right)-E\left[|\Gamma(v)|\right]}.
\end{aligned}
\tag{23}
$$

Recalling that node degree $|\Gamma(v)|$ of wireless ad-hoc networks is a binomial random variable $B(N,\ \pi r^2)$, we get from (23):

$$
E[\gamma(v)] = \frac{C_v}{\frac{1}{2}\left(\bar{k}^2-\bar{k}(\pi r^2)\right)}.
$$

Thus, the expected clustering coefficient is given as:

$$
\gamma(G) = E[\gamma(v)] = \frac{0.69(\bar{k}+1)-2}{\bar{k}-\pi r^2}. \tag{24}
$$

This completes the proof and leads to (19).

Recalling that $\bar{k} = \log N$ and $r = \sqrt{\log N / \pi N}$ , we obtain from (24),

$$
\gamma(G) = \frac{0.69(\log N+1)-2}{\log N - \frac{\log N}{N}}.
$$

For large values of $N$, the clustering coefficient of wireless ad-hoc networks asymptotically tends to:

$$
\gamma(G) \rightarrow 0.69. \tag{25}
$$

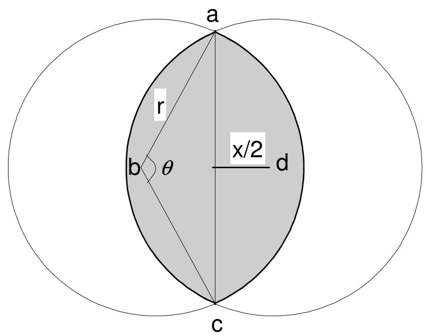


**Figure 9 Determination of clustering coefficient by estimation of overlap area of two circles.**

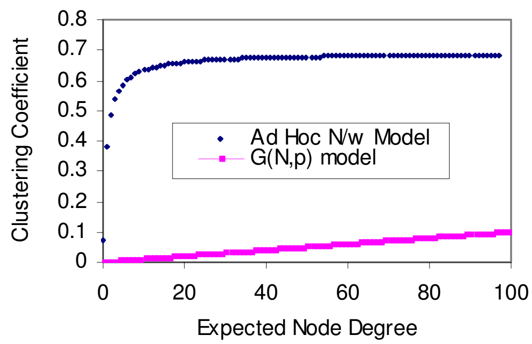


**Figure 10 Clustering coefficient for wireless ad-hoc networks and random graph model $G(N,p)$ with expected node degree $k$ for a network with $N$ =1000 nodes.**

In comparison, if we consider a node in a random graph $G(N,p)$ with expected node degree $\bar{k}$, the probability that two of its neighbors are connected is equal to the probability that two randomly selected nodes are connected. Consequently, the clustering coefficient of a random graph is given by:

$$\gamma_{rand} = p = \frac{\bar{k}}{N} \to 0. \quad (26)$$

We observe from (26) that $\gamma_{rand}$ tends to 0 for large values of *N*. Thus, the clustering coefficient of wireless ad-hoc networks is significantly higher than that of random graphs based on the *G*(*N*,*p*) model. The variation of the clustering coefficient in both models with increasing network size is shown in Figure 10.

2. Diameter and Average Path Length

The diameter of a graph is the maximal distance between any pair of its nodes. Strictly speaking, the diameter of a disconnected graph (i.e., one made up of several isolated clusters) is infinite, but it can be defined as the maximum diameter of its clusters. The average path length provides another way to characterize the spread of a network. For a uniform spatial distribution of nodes, the maximum and average paths (in hops) among nodes in a connected set of *N* nodes both increase as $\Theta(\sqrt{N})$ [17]. In particular, it is well known that the maximum path length across the entire network and the average path length *L* increase as $N^{1/2}$:

$$L = \Theta(\sqrt{A}) = \Theta(\sqrt{N}). \quad (27)$$

Unlike ad-hoc networks, random graphs based on the ER model tend to have *logarithmic* ($\log_k N$) diameters [1], which are substantially smaller than the bound in (27). The plot of the expected path length as a function of the number of nodes in the network is shown in Figure 11.

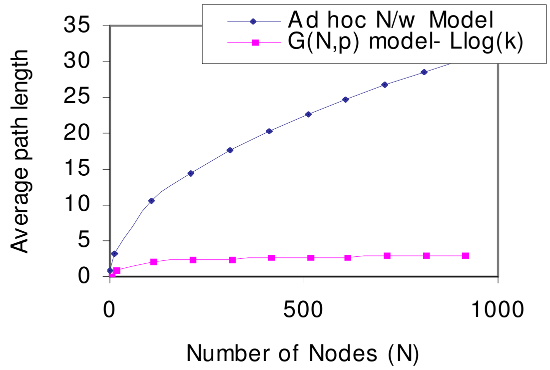


**Figure 11 Average path lengths for wireless ad-hoc networks and random graph model** $G(N, p)$**.**

Hence, unlike real, large-scale networks, the ad-hoc model exhibits high clustering and large average distance. The high average path length has implications on the route reliability and node capacity. The probability of route breakage increases with the size of the route. In addition, more hops on the data path result in increased relay load on the intermediate nodes of the route. This leads to a reduction of the useful per-node capacity as we have shown in section IV.2. We now develop a large-scale evolving model for ad-hoc networks, which gives a constrained average path length over all source-destination pairs.

# VI. LARGE-SCALE WIRELESS NETWORK TOPOLOGY

The above discussion on graph resilience and node capacity motivates us to develop an alternate model for wireless ad hoc networks. Our model attempts to capture the growth dynamics of a network consisting of mobile users. We consider two types of attachment functions for newly introduced nodes. In the first model, as new nodes join the network, they attempt to associate with large clusters - based on the reasoning that large clusters would most likely contain more information. A natural truncation or aging effect may be seen due to reduction in per-node capacity for significantly large sized clusters. In the second model, we consider a uniform node-attachment function wherein the new node is equally likely to attach to any of the existing clusters.

## 1. Preferential Attachment Evolving Model

Linear preferential linking has been suggested as a means to model many real networks [6], [10]. Any self-organizing group of nodes (humans with laptops, smart devices) exhibits non-uniform spatial distribution and traffic patterns. This is based on the observed phenomena that *popularity is attractive* and *information is non-uniformly distributed*. We base our model on this linear preferential attachment function. This model describes three events occurrences pertaining to the growth of the network: a new node joining an existing cluster, a new node forming a cluster, and transition of an existing node from one cluster to another. We start with $c_0$ initial clusters (of size 1) in a given region of the unit area. At each time step, we perform one of the following two operations.

With probability $p$ ($0 \le p < 1$), we add a new node to the system. With probability $q$, the new node forms a new cluster at a random location. With probability $1-q$, the new node joins one of the existing clusters. The new node selects its (*x,y*) coordinates within the transmission range of a random node in the chosen cluster. Let $\prod(c_i(t))$ denote the probability for choosing cluster *i,* where $c_i(t)$ is the size of cluster *i* at time *t*. In linear preferential attachment, $\prod(c_i(t))$ is defined as:

$$\prod(c_i(t)) = \frac{c_i(t)}{\sum_j c_j(t)}. \tag{28}$$

With probability $1-p$, a node moves from one cluster to another, which models mobility of wireless nodes. A randomly selected node from a cluster moves from its existing cluster to a new cluster based on the preferential attachment model. The node selects its (*x,y*) coordinates within the chosen cluster in a manner identical to that in step 1.

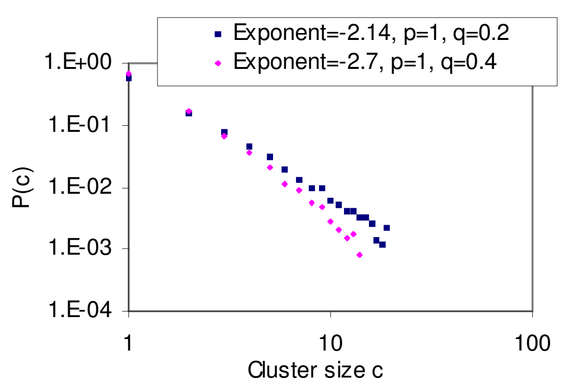


**Figure 12 Cluster size distribution for preferential attachment of nodes to existing clusters.**

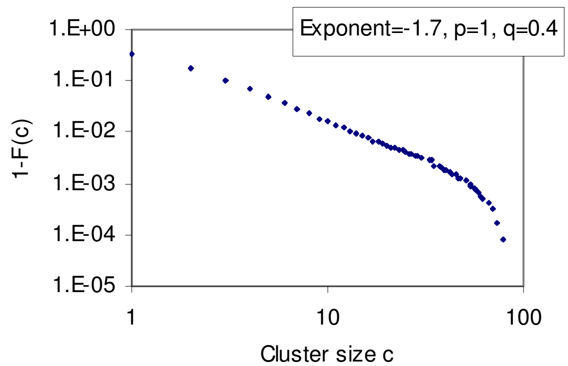


**Figure 13 Complementary cumulative distribution for cluster size - Truncation of power-law cluster size due to constraints on acceptable capacity and delay.**

2. Analysis

Using the continuum theory approach [1], we show that our model results in a graph with a power law cluster size distribution as shown in Figure 12 and Figure 13. We extend the scale-free model as introduced in [6] to model node clustering and mobility in addition to network growth. We also derive a precise relationship between the model parameters and the power law exponent.

*Theorem 1*: The cluster size distribution for the preferential attachment model follows a power law distribution with the exponent given by:

$$\alpha(p,q) = \frac{p}{1-pq}. \tag{29}$$

*Proof*: Assuming $c_i$ changes continuously with time, the probability $\prod(c_i(t))$ can be interpreted as the rate at which $c_i$ changes. The growth of the network from the addition of new nodes to existing clusters and node movement between clusters contribute to changes in $c_i$, while formation of new clusters does not affect the existing values of $c_i$. Hence, the contribution from formation of new clusters to rate of change of $c_i$ is 0.

- Addition of a new node to an existing cluster with probability $p(1-q)$ can be modeled as:

$$\frac{\partial c_i(t)}{\partial t} = \frac{c_i(t)}{\sum_j c_j(t)}. \tag{30}$$

- Addition of a new node to form a new cluster with probability $pq$ has no effect on $c_i$:

$$\frac{\partial c_i(t)}{\partial t} = 0. \tag{31}$$

- Transition of a node from an existing cluster to another one with probability $1-p$ results in:

$$\frac{\partial c_i(t)}{\partial t} = -\frac{1}{C(t)} + \frac{c_i(t)}{\sum_j c_j(t)}. \tag{32}$$

  where $C(t)$ is the total number of existing clusters.

The first term models the rate of decrease in the size of the cluster due to a "node leave", while the second term shows the rate of increase in size of the cluster due to a "node join".

Since these three processes take place simultaneously, we add their contributions:

$$\begin{aligned}\frac{\partial c_i(t)}{\partial t} &= p(1-q)\frac{c_i(t)}{\sum_j c_j(t)} + (1-p)\frac{c_i(t)}{\sum_j c_j(t)} - (1-p)\frac{1}{C(t)} \\ &= (1-pq)\frac{c_i(t)}{\sum_j c_j(t)} - (1-p)\frac{1}{C(t)}. \end{aligned} \tag{33}$$

The number of clusters $C(t)$ in the graph is simply $(pq)t+c_0$, where $c_0$ represents the initial number of clusters in the network. For large $t$, the constant term $c_0$ can be safely omitted from $C(t)$ as compared to $(pq)t$ - $C(t) = (pq)t$. The total number of nodes in all the clusters of the network varies with time and is $\sum_j c_j(t) = pt + c_0 \approx pt$. Thus, (33) transforms to:

$$\frac{\partial c_i(t)}{\partial t} = \frac{(1-pq)}{p}\frac{c_i(t)}{t} - \frac{(1-p)}{pq}\frac{1}{t} \tag{34}$$

Using the initial condition that the size of a cluster formed at time $t_i$ is $c_i(t_i)=1$, the solution of (34) for $c_i(t_i)$ is:

$$c_i(t) = \left(1-\frac{B(p,q)}{A(p,q)}\right)\left(\frac{t}{t_i}\right)^{A(p,q)} + \frac{B(p,q)}{A(p,q)}. \tag{35}$$

where

$$\begin{aligned} A(p,q) &= \frac{1-pq}{p}, \\ B(p,q) &= \frac{1-p}{pq}. \end{aligned} \tag{36}$$

Using (35), the probability that a cluster has size $c_i(t)$ smaller than $c$ can be written as:

$$P[c_i(t) < c] = P[t_i > C(p,q)t]. \tag{37}$$

where

$$C(p,q) = \left(\frac{1-\frac{B(p,q)}{A(p,q)}}{c-\frac{B(p,q)}{A(p,q)}}\right)^{\frac{1}{A(p,q)}}. \tag{38}$$

Since $t_i$ must satisfy the condition $0 \le t_i \le t$, we have three cases for $C(p,q)$:

a) If $C(p,q) > 1$, then $P[c_i(t) < c] = 0$. Thus, $c > 1$ is the condition for $P(c)$ to be non-zero. $P(c)$ is the probability that a cluster has size $c_i(t)$ equal to c.

b) If $C(p,q)$ is not real, then $P[c_i(t) < c]$ is not well defined. Thus, to be able to calculate $P$(c), we need $C(p,q) > 0$, for all $c > 1$, $1 - B(p,q)/A(p,q) > 0$. This condition translates into $p > 1/(1+q)$.

c) For $0 < C(p,q) \le 1$, we determine the cluster size distribution $P(c)$ below.

Defining the unit of time in the system as one cluster form/cluster join/node transition attempt, the probability density of $t_i$ is $P_i(t_i) = 1/(pqt + c_0)$. Thus,

$$P[\ c_i(t) < c\ ] = 1 - \ C(p,q)\frac{t}{pqt + c_0}. \tag{39}$$

As $P(c) = \dfrac{\partial P[\ c_i(t) < c]}{\partial c}$, we obtain

$$\begin{aligned} P(c) = {} & \frac{t}{pqt + c_0}\frac{1}{A(p,q)}\ \left(1 - \frac{B(p,q)}{A(p,q)}\right)^{\frac{1}{A(p,q)}} \\ & \times \left(c - \frac{B(p,q)}{A(p,q)}\right)^{-1-\frac{1}{A(p,q)}}. \end{aligned} \tag{40}$$

Thus, the cluster size distribution has a generalized power-law form.

$$P(c) \propto [\ c - D(p,q)\ ]^{-(1+\alpha(p,q))}. \tag{41}$$

where

$$D(p,q) = \frac{B(p,q)}{A(p,q)}. \tag{42}$$

and scaling exponent is given as $1+\alpha(p,q)$, where

$$\alpha(p,q)=\frac{1}{A(p,q)}.$$

This leads to (29) and completes the proof.

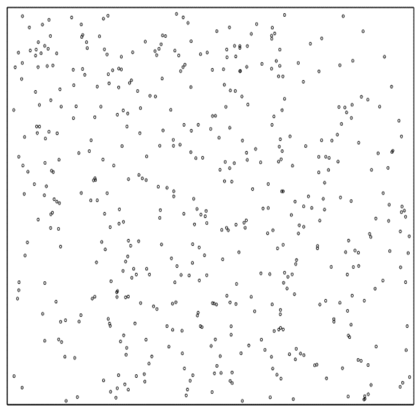

**Figure 14 Homogeneous spatial distribution of nodes.**

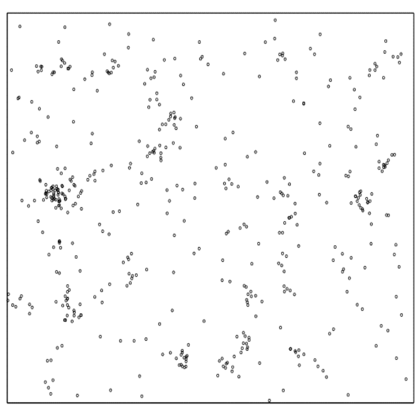

**Figure 15 Preferential attachment model: non-homogeneous spatial distribution of nodes.**

For values of $p < 1/(1+q)$ as $p \to 0$, growth of the network is suppressed and the model deviates from the power law distribution to approximate an exponential tail for the cluster size distribution. On the other end, $p = 1$ gives a power-law distribution $P(c) = \alpha(q)c^{-(1+\alpha(q))}$ with exponent $\alpha(q) = 1/(1-q)$, where $q$ is the probability of forming a new cluster. This is equivalent to the static network case where nodes do not transition between clusters. A snapshot of the network is shown Figure 14 for uniform distribution of nodes while Figure 15 shows the clusters in the preferential attachment model.

3. Information Storage and Retrieval

In the model described in the previous subsection, connectivity is guaranteed within clusters. In situations where nodes require data objects from nodes in other clusters (or services from the Internet such as websites or email access), complete network connectivity (or connectivity to wired infrastructure) is needed. In such scenarios, we require the support of wireless relay nodes or wired base-stations between clusters, as shown in Figure 16. Initial access to data objects stored in nodes belonging to other clusters (or services from the Internet) would be achieved through these relay nodes between clusters (or base-stations). Further access to these objects can be optimized through collaborative caching [13] as observed in information sharing networks (peer-to-peer networks). Such requests would then be met from nodes within the requesting node's cluster.

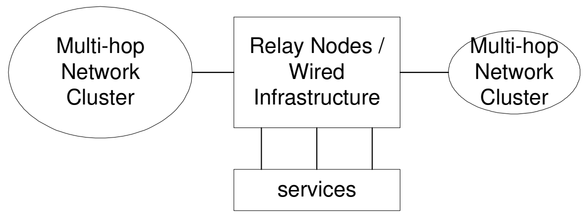


**Figure 16 Communication between clusters through relay nodes (e.g. wired infrastructure).**

To avoid creating hotspots in communication within the cluster and to distribute the load evenly amongst the nodes, we propose use of GHT (Geographic Hash Table) [34] for data storage and retrieval. Each data object in the cluster is associated with a unique key and GHT hashes the keys into geographic coordinates within the cluster. The nodes that are geographically nearest to those coordinates store the key-value pair. This system achieves high data availability by replicating stored data locally, thus providing resilience against node failure.

Next, we obtain an estimate for node capacity for the proposed network model.

4. Node Capacity

Assume, as before that each node is capable of transmitting at $W$ bits per second. Since the information distribution governs the formation and size of clusters, the node that preferentially attaches itself to one the existing cluster constricts its traffic to the nodes belonging to the same cluster. Hence, each node sends data to a chosen destination *within its cluster*. Under the above assumptions, we obtain an expression for per-node capacity $\lambda$ in the above model.

*Lemma 4*: The per-node capacity in the preferential attachment model is constant with increasing network size and is given by:

$$\lambda \sim \frac{W}{\sqrt{\beta}}\frac{2\alpha}{2\alpha+1}.$$

*Proof*: Considering the power law cluster size distribution given by

$$F(c) = 1-\beta^{\alpha}c^{-\alpha}. \qquad (43)$$

where $\beta$ is the minimum cluster size (generally 2), $c \geq \beta$, and $\alpha = \dfrac{p}{1-pq} > 1$.

Recall that the expected cluster size for a pareto distribution is given by

$$\bar{c} \sim \beta\frac{\alpha}{\alpha-1}. \qquad (44)$$

From (11), we obtain that per-node capacity is bounded by $\lambda = O(W/\sqrt{N})$. Hence, the optimal per-node capacity, $\lambda_i$, within a cluster $i$ is a function of the size of the cluster $c_i$ and can be estimated as:

$$\lambda_i \sim \frac{W}{\sqrt{c_i}}.$$

Hence, the per-node capacity for any node in the network is given as:

$$\lambda = E\left(\lambda_i\right) = \frac{W}{E\left(\sqrt{c_i}\right)}.$$

We obtain the cumulative distribution function of $Y = 1/\sqrt{c_i}$ as:

$$F(y) = \beta^{\alpha}y^{2\alpha}. \qquad (45)$$

From (45), we obtain the expectation of $Y$ as:

$$E[Y] = E[1/\sqrt{c_i}\,] = \frac{1}{\sqrt{\beta}} \frac{2\alpha}{2\alpha+1}. \tag{46}$$

Hence, the optimal per-node throughput capacity is:

$$\lambda \sim \frac{W}{\sqrt{\beta}} \frac{2\alpha}{2\alpha+1}. \tag{47}$$

*Hence, the node capacity in the system remains constant as the number of nodes in the system increases.* The estimate of node capacity is a function of system characteristics, which include the probability of cluster formation and node transition frequency between clusters.

5. Uniform Attachment Evolving Model

Natural development of power-law degree distribution in large-scale networks indicates that growth and preferential attachment are important features of network evolution. We briefly consider a network model that does not incorporate preferential attachment. We aim to show that such a model also scales to large number of nodes. Hence, it can provide for systems that are not self-organizing and where nodes exhibit controlled behavior. We show that if the nodes entering the network attach uniformly to existing clusters, the system consists of exponentially distributed cluster sizes. The model is defined as follows:

*Growth:* Starting with a small number of clusters ($c_0$), at every time step a new node enters the system and forms a new cluster with probability $p$ and joins an existing cluster with probability $1-p$.

*Uniform Attachment:* The new node joins existing clusters with equal probability, independent of the cluster size:

$$\Pi(c_i) = \frac{1}{c_0 + pt}. \tag{48}$$

Figure 17 shows that unlike the power-law form for preferential attachment model, probability $P(c)$ has an exponential form. Using the continuum theory arguments mentioned in 2, we analytically obtain the expression for $P(c)$. The rate of change of the size of cluster $i$ in this case is given by:

$$\frac{\partial c_i(t)}{\partial t} = (1-p)\Pi(c_i(t)) = (1-p)\frac{1}{pt + c_0}. \tag{49}$$

Solving the above equation for $c_i(t)$, and taking into account that $c_i(t_i) = 1$, we obtain:

$$c_i(t) = 1 + a(\log t - \log t_i). \tag{50}$$

where

$$a = \frac{1-p}{p}. \tag{51}$$

indicating that $c_i$ has a logarithmic increase with time.

The probability that cluster $i$ has size $c_i(t)$ smaller than $c$ is:

$$P(c_i(t) < c) = P\left(t_i > t\exp\left(-\frac{c-1}{a}\right)\right). \tag{52}$$

Assuming that we add the nodes uniformly to the system, we get

$$P\left(t_i > t\exp\left(-\frac{c-1}{a}\right)\right)$$

$$
\begin{aligned}
&= 1 - P\left( t_i \le t \exp\left( -\frac{c-1}{a} \right) \right) \\
&= 1 - \frac{t \exp\left( -\frac{c-1}{a} \right)}{pt + c_0}.
\end{aligned}
\tag{53}
$$

As $P(c) = \dfrac{\partial P[\ c_i(t) < c]}{\partial c}$, we obtain for large values of *t:*

$$
P(c) = \frac{e^{\frac{1}{a}}}{a} \exp\left( -\frac{c}{a} \right). \tag{54}
$$

This indicates that the cluster size has an exponential distribution with expected cluster size $ae^{1/a}$. The expected node capacity is constant and lower bounded by $W / \sqrt{ae^{1/a}}$. This suggests that with uniform attachment of nodes to clusters, the maximum cluster size is bounded and cannot be infinitely large. A snapshot of the cluster formation process for uniform attachment model is shown in Figure 18.

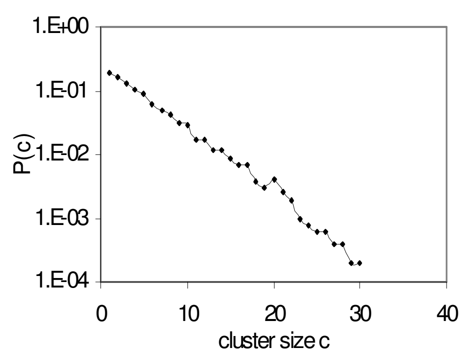


**Figure 17 Cluster size distribution for uniform attachment model for probability of new cluster formation $p = 0.2$.**

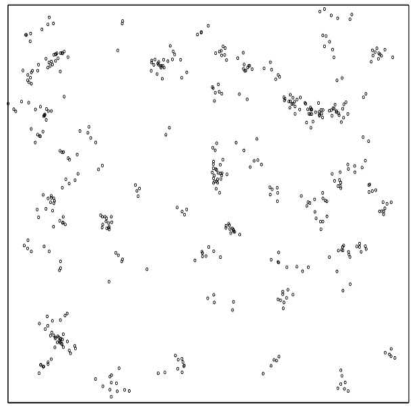

**Figure 18 Uniform attachment model: non-homogeneous distribution of nodes.**

# VII. CONCLUSION

The applications for ad-hoc networks are only beginning to be recognized. As applications for large-scale information/file sharing in the wireless domain develop, ad-hoc network are expected to provide a cost-effective solution for increasing and complementing wired connectivity.

We have seen that large networks with a uniform spatial distribution of nodes do not scale well in terms of reliability of the network and node capacity. One approach to improve the scalability of an ad-hoc network involves creating a "small world" effect by setting up base stations and dividing the nodes into single-hop cells. This results in a node capacity $\lambda=\Theta(1/\text{number of nodes per cell})$. In this case, a cellular topology consisting of a collection of small LANs connected together by base stations would be required to seamlessly cover a large region.

Since ad-hoc networks must exist with no or little infrastructure support, a realistic topology of large ad-hoc networks is a collection of connected components or clusters. We contend that since information is clustered or concentrated in a small number of users, cluster formation is a realistic situation.

In this thesis, we have studied cluster formation as a result of an evolving model for ad-hoc networks. The cluster size distribution has been shown to depend on how the nodes attach to existing clusters and the formation of new clusters. Preferential attachment of nodes to existing clusters results in a truncated power-law cluster size distribution. The truncated distribution ensures that an infinitely large cluster does not exist. The per-node capacity for the power-law distribution of cluster sizes is shown to scale with network size though ad-hoc connectivity is only maintained within clusters. On the other hand, uniform attachment of nodes gives an exponential distribution of cluster sizes. Since per-node capacity in both the models scales with network size, we contend that any networks with intermediate node attachment behavior would also scale well for large network sizes. Communication between clusters still requires support from a collection of relay nodes (base-stations) distributed between clusters.

Thus, this thesis developed a model for growing ad-hoc networks that demonstrates that wireless ad-hoc networks are likely to be scalable in practical settings where users' communication patterns are highly clustered.

# REFERENCES


[1] R. Albert and A. Barabasi, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics,* vol. 74, 2002.

[2] U. Alon, M. Surette, N. Barkai and S. Leibler, "Robustness in Bacterial Chemotaxis," *Nature* vol. 397, pp. 168-171, 1999.

[3] N. Bansal and Z. Liu, "Capacity, Delay and Mobility in Wireless Ad-Hoc Networks," *INFOCOM*, 2003.

[4] S. Bansal, R. Gupta, R. Shorey and A. Misra, "Energy Efficiency and Throughput for TCP Traffic in Multi-Hop Wireless Networks," *INFOCOM*, 2002.

[5] A. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science,* vol. 286, pp. 509-512, 1999.

[6] A. Barabasi, R. Albert and H. Jeong, "Mean-Field Theory for Scale-Free Random Networks," *Physica A.* vol. 272, 173-187, 1999.

[7] C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multi-hop Network," *MobiHoc*, 2002.

[8] C. Bettstetter, "On the Connectivity of Wireless Multihop Networks with Homogeneous and Inhomogeneous Range Assignment," *VTC*, 2002.

[9] L. Breslau, P. Cao, L. Fan, G. Phillips and S. Shenker, "Web Caching and Zipf-like Distributions: Evidence and Implications," *INFOCOM*, 1999.

[10] T. Bu and D. Towsley, "On Distinguishing between Internet Power Law Topology Generators," *INFOCOM*, 2002.

[11] R. Cohen and S. Havlin, "Scale Free Networks Are Ultra Small," *Phys. Rev. Letters* vol. 90, 2003.

[12] O. Dousse, P. Thiran and M. Hasler, “Connectivity in Ad-Hoc and Hybrid Networks,” *INFOCOM*, 2002.

[13] M. Duska, D. Marwood and J. Freeley, “The Measured Access Characteristics of World-Wide-Web Client Proxy Caches,” *USITS*, 1997.

[14] P. Erdős and A. Renyi, “On Random Graphs,” *Publicationes Mathematicae (Debrecen)* vol. 6, pp. 290-297, 1959.

[15] M. Faloutsos, P. Faloutsos and C. Faloutsos, “On Power-law Relationships of the Internet Topology,” *SIGCOMM*, 1999.

[16] M. Grossglauser and D. Tse, “Mobility Increases the Capacity of Ad-hoc Wireless Networks,” *INFOCOM*, 2001.

[17] P. Gupta and P. Kumar, “The Capacity of Wireless Networks,” *IEEE Transactions on Information Theory*, March 2000.

[18] P. Gupta and P. Kumar, “Towards an Information Theory of Large Networks: An Achievable Rate Region,” *IEEE International Symposium on Information Theory ISIT*, 2001.

[19] P. Gupta and P. Kumar, “Internets in the Sky: The Capacity of Three Dimensional Wireless Networks,” *Communications in Information and Systems*, Jan. 2001.

[20] P. Gupta, R. Gray and P. Kumar, “An Experimental Scaling Law for Ad Hoc Networks,” http://black.csl.uiuc.edu/~prkumar/ps_files/exp.ps, May 2001.

[21] H. Hsieh and R. Sivakumar, “Towards a Hybrid Network Model for Wireless Packet Data Networks,” *ISCC*, 2002.

[22] K. Kar, M. Kodialam, T. Lakshman and L. Tassiulas, “Routing for Network Capacity Maximization in Energy-constrained Ad-hoc Networks,” *INFOCOM*, 2003.

[23] V. Kawadia and P. Kumar, “Power Control and Clustering in Ad Hoc Networks,” *INFOCOM*, 2003.

[24] M. Levy, “Market Efficiency, the Pareto Wealth Distribution, and the Levy Distribution of Stock Returns,” The Hebrew University of Jerusalem Technical Report, 2001.

[25] H. Li and D. Yu, “Comparison of Ad Hoc and Centralized Multihop Routing,” Siemens Technical Report, May 2001.

[26] J. Li, C. Blake, D. S. J. De Couto, H. Lee and R. Morris, "Capacity of Ad Hoc Wireless Networks," *MOBICOM*, 2001.

[27] M. Lindsey, M. Papadopouli, F. Chinchilla, A. Singh, “Measurement and Analysis of the Spatial Locality of Wireless Information and Mobility Patterns in a Campus,” UNC Technical Report, March 2003.

[28] B. Liu, Z. Liu and D. Towsley, “On the Capacity of Hybrid Wireless Networks,” *INFOCOM*, 2003.

[29] D. Loguinov, A. Kumar, V. Rai and S. Ganesh, “Graph-Theoretic Analysis of Structured Peer to Peer Systems: Routing Distances and Fault Resilience,” *SIGCOMM*, 2003.

[30] S. Mengesha and H. Karl, "Relay Routing and Scheduling for Capacity Improvement in Cellular WLANs", *WiOpt,* 2003.

[31] M. Papadopouli and H. Schulzrinne, "Effects of Power Conservation, Wireless Coverage and Cooperation on Data Dissemination among Mobile Devices," *MobiHoc*, 2001.

[32] M. Penrose, "On k-connectivity for a Geometric Random Graph," *Random Structures and Algorithms*, vol. 15, pp. 145-164, 1999.

[33] E. Perevalov and R. Blum, "Delay Limited Capacity of Ad Hoc Networks: Asymptotically Optimal Transmission and Relaying Strategy," *INFOCOM*, 2003.

[34] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, S. Shenker, "GHT: A Geographic Hash Table for Data Centric Storage," *WSNA*, 2002.

[35] S. Ratnasamy, P. Francis, M. Handley and R. Karp, "A Scalable Content-Addressable Network," *SIGCOMM*, 2001.

[36] C. Santivanez, B. McDonald, I. Stavrakakis and R. Ramanathan, "On the Scalability of Ad Hoc Routing Protocols," *INFOCOM*, 2002.

[37] S. Saroiu, P. Gummadi and S. Gribble, "A Measurement Study of Peer-to-Peer File Sharing Systems," MMCN, 2002.

[38] H. A. Simon, "On a Class of Skew Distribution Functions," *Biometrika,* vol. 42, pp. 425-440, 1955.

[39] H. A. Simon, *Models of Man*, John Wiley, New York, 1957.

[40] A. Van Raan, "Fractal Dimension of Co-citations, " *Nature* vol. 347, pp. 626, 1990.

[41] S. Wasserman and K. Faust, "*Social Network Analysis*," Cambridge University Press, Cambridge, U.K, 1994.

[42] D. Watts and S. Strogatz, "Collective Dynamics of "Small-World" Networks." *Nature* vol. 393, pp. 440-442, 1998.

[43] S. Yi, Y. Pei and S. Kalyanaraman, "On the Capacity Improvement of Ad Hoc Wireless Networks Using Directional Antennas," *MobiHoc*, 2003.

# VITA

Bharat Vishwanathan Iyer was born in Indore, India, on March 13, 1978. In May 2000, he received a Bachelor of Engineering degree in electrical engineering from the Maharaja Sayajirao University, Baroda, India. In August 2000, he began his graduate studies at Texas A&M University in pursuit of a Master of Science in electrical engineering. While pursuing his Master's degree, he worked as a research assistant under Dr. Nitin Vaidya and Dr. Dmitri Loguinov in the area of mobile computing. From January 2002, he also worked for a year at Advanced Micro Devices as a co-op engineer in the Personal Connectivity Solutions Division. His permanent address is: C/209 Sitaram Complex, Subhanpura, Baroda, Gujarat, India-390023.