1 Introduction

Recent years have seen the burgeon of wireless Ad hoc networks, devices with wireless interfaces can be connected to each other in pure ad hoc fashion or through infrastructures. As one of the most basic concerns, the capacity of such networks have received much attention from the researchers, especially after the pioneer works done by Gupta and Kumar [1].

Current researches generally follow either of the two tracks, one is the theoretical track, which meanly concern what is the theoretical bounds of the capacity, or in other words, how fast the data can be sent from source nodes to their destinations in such networks. They strive to find both upper and lower bounds and these results may serve as guidelines on evaluating how good a real network is.

The other track emphasizes more on the practical aspects, which concentrates on constructing optimal scheduling mechanisms to achieve the theoretical bound, although capacity is not necessarily the only goal of optimization. In fact, it has been found recently that capacity is in fact in contradict with other optimization goals such as connectivity [46], delay [43] and fairness [47].

In this survey, we will mainly follow the theoretical track and put great emphasise on the models used to facilitate analysis in state of art works.

1.1 Taxonomy

Due to the diversity of the underlying networks, the analytical models used to derive capacity results differ from each other. So we will study these models by classifying them according to the following criteria:

- Measurement of Capacity

*Throughput Capacity vs. Transport Capacity* When we try to make a clear definition of network capacity, what we will discuss throughout the paper, we found there is no single definition. In fact, during the studies trying to discover the maximum data transmission ability of the network, researchers have defined different measurements from different aspects. We will discuss these definitions in Section 2.1

- Interference Model

*Protocol Model vs. Physical Model* Many models of successful receiving conditions under interference have been introduced. Most of them follow the original models proposed by Gupta and Kumar [1], known as *Protocol Model and Physical Model*. We will discuss them and different variations in Section 2.2

- Network Structure
Flat Network vs. Hybrid Network We roughly classify the analytical models into flat networks and hybrid networks based on network structures. In Flat Networks, nodes connect with each other with wireless links, although each node may choose different transmission power levels or different sub-channels, they basically share a common bandwidth and interfere with nearby nodes. However, in Hybrid Networks, nodes can rely on some powerful nodes to facilitate communication, where the powerful nodes are connected with each other by high bandwidth communication links and do not interfere with the Ad hoc nodes. Data can be forwarded in hybrid fashion, through both infrastructure nodes and ad hoc nodes by multi-hopping. Network structures are discussed in Section 2.3.

- Traffic Pattern

When we take traffic patterns into account, we found that current works fall into two categories: networks with random traffic or with arbitrary traffic.

Random Traffic vs. Arbitrary Traffic For random traffic, the source nodes choose its destination randomly in the network, that will result an average distance of the order as the square root of network area, in two dimension case. For Arbitrary traffic, the destination node is chosen based on some specific criteria, for example, the distance to the destination may be power law distributed.

Symmetric vs. Asymmetric Traffic The number of sources and destinations also has impact on the capacity. When the of sources are equal or almost equal to the destinations, the traffic tend to distribute evenly in the network, given the source-destination pairs are randomly chosen. Most study fall into this category. However, if the sources are much more than the destinations, bottleneck may form around the destinations, making a limitation on the capacity. This is especially true in wireless sensor networks. On the other extreme, if the destinations are much more than the sources, this will result a multicast or broadcast traffic in the network. The first case is seen as the symmetric traffic and both the latter two cases are classified as asymmetric traffic in this study.

Different traffic patterns are discussed in Section 2.4.

- Mobility

Static Network vs. Mobile Network Recently, researchers found that mobile nodes may bring some advantages to the network, such as increasing capacity and eliminating bottlenecks.

Although most of the current works are based on Static Network, we can reasonably expect more works will be done in mobile ad hoc networks, considering capacity issues. The difficulties of analyzing the capacity of mobile ad hoc networks are multi-fold. First, we are not clear about the capacity of general wireless channels with interference, even when the nodes are static. Second, although we have identified many mobility model, few of them are analytically plausible, preventing researchers getting capacity result with them. Current works have shown that mobility models have substantial impact on the capacity of Ad hoc network, as will be discussed in Section 2.5.

- Density

Dense Network vs. Extended Network Current works assume either fixed area network or extensible area network. Fixed area network occupies a fixed length of distance for one dimension networks, a fixed area for two dimension networks or a fixed volume for three dimension networks, when the number of nodes in the network grows, and this setting will result a Dense Network whose density goes to infinity as node number increases. In contrary, an extensible area network usually assume minimum distance between nodes or constant density of nodes. Under either case, the network will expand its occupation as more nodes joining the network, resulting Extended Network.

We will discuss them in Section 2.6.

Although the two models are different, we found they have the same ability in capturing the characteristics of the network, moreover, the results based on two models are interchangeable.
1.2 Notations

Frequently used notations are listed below, others will be explained in the context.

\( n \): Number of ad hoc nodes.
\( m \): Number of infrastructure nodes.
\( T \): Set of all concurrent senders.
\( d_{ij} \): Euclidean distance between node \( i \) and \( j \).
\( A \): Area of the network.
\( W \): Wireless transceiver bandwidth.
\( P_i \): Transmission power of transmitter \( i \).
\( N \): Power of Gaussian white noise.
\( \alpha \): Path loss attenuation exponent.
\( C_T \): Transport capacity.
\( C_I \): Information capacity.

The asymptotic notations used in this paper are defined as follows:

\[ f(n) = O(g(n)) \]:\( \exists n_0, \exists M > 0, \) such that \( f(n) \leq M g(n) \) for all \( n > n_0 \).
We say \( f(n) \) is asymptotically upper bounded by \( g(n) \), or simply upper bounded by \( g(n) \).

\[ f(n) = \Omega(g(n)) \]:\( \exists n_0, \exists M > 0, \) such that \( f(n) \geq M g(n) \) for all \( n > n_0 \).
We say \( f(n) \) is asymptotically lower bounded by \( g(n) \), or simply lower bounded by \( g(n) \).

\[ f(n) = \Theta(g(n)) \]: \( f(n) = O(g(n)) \) and \( f(n) = \Omega(g(n)) \).
We say \( f(n) \) is asymptotically tight bounded by \( g(n) \), or simply tight bounded by \( g(n) \).

2 Analytical Models of Ad hoc Networks

In this section we will discuss the models developed so far to facilitate analysis of Ad hoc network capacity based on our taxonomy described in the previous section. The classification covers most of the important factors that need to be concerned when trying to solve the capacity problem of Ad hoc networks.

2.1 Measurement of Capacity

2.1.1 Throughput Capacity

- **Channel Capacity**

  *Channel Capacity* or *Information Capacity* was originally defined by Shannon as the maximum information that can be reliably transmitted from source to destination over a communication channel. One important point in the definition is that the maximization is with respect to the input distribution. Shannon found that the capacity so defined can be calculated as the mutual information between the input and output of the channel. Thus, this definition reflects the interaction of the source and the channel, and indicates what the source should do to get most out of the channel.

  The capacity of a discrete channel is defined as:

  \[
  C = \lim_{T \to \infty} \frac{N(T)}{T} \tag{1}
  \]

  where \( N(T) \) is the maximum allowed number of signals in duration \( T \).

  This definition concern only highest possible rate between the source and destination, neglecting the role of distance plays in the data transmission process.

  In [4], the authors derived the information capacity for a particular network. Based on single source-destination pair assumption, the network can be modeled by a *Gaussian Relay Channel*, thus can be studied via classical information theoretical way.

  However in most of studies of Ad hoc network capacity, this definition is replaced by a generalized version, known as *Capacity Region*, which is able to describe the capacity of the network. We will discuss this definition later in this section.

- **Throughput Capacity**
Throughput Capacity was used by some authors as a measurement of the information transmission ability of the network. Note that the notion Throughput is the amount of data transmitted from source to destination during unit time, thus the maximum throughput is in fact the information capacity of the channel, where the maximum is taken for all source distributions.

The throughput capacity of a network is often defined as an average of the throughput capacity of all nodes in the network. For example, in [1], the authors proposed the following of definitions. Throughput is defined as the time average of bits that can be transmitted by each node to its destination.

\[ \lambda(n) := \lim_{T \to \infty} \frac{b_i(T)}{T} \text{ for every } 1 \leq i \leq n. \]

In this definition, the usage of probability is based on the random nature of the network. Although we cannot be sure that the capacity so defined is feasible to every realization of the network, we are confident that the exceptions are vanishingly small.

Similar definition was proposed in [30], [31] for Hybrid Network. For a hybrid network with \( n \) nodes and \( m \) base stations, an aggregate throughput of \( T(n, m) \) bits/sec is feasible if by transmitting data in the ad hoc or infrastructure mode, there is a spatial and temporal scheduling scheme that yields an aggregate network throughput of \( T(n, m) \) bits/sec on average. Here the aggregate throughput of the sum of the individual throughputs from each node to its chosen destination. The aggregate throughput capacity of a hybrid wireless network is of order \( \Theta(f(n, m)) \) bits/sec if there are deterministic constants \( c > 0, c' < +\infty \) such that

\[ \lim_{n \to \infty} \text{Prob}(\lambda(n) = cf(n) \text{ is feasible}) = 1 \quad (2) \]

\[ \lim_{n \to \infty} \inf \text{Prob}(\lambda(n) = c'f(n) \text{ is feasible}) < 1 \quad (3) \]

In [9], the authors gave the following set of definitions to describe the transmission ability of the network.

Let \( b_i(T) \) denotes the total amount of information (in bits) generated by node \( i \) and received by its destinations during interval \([0, T]\). The end-to-end throughput of node \( i \) is defined as follows:

\[ \lambda_i := \lim_{T \to \infty} \frac{b_i(T)}{T} \text{ for every } 1 \leq i \leq n. \]

Note there is a subtle difference in this definition between the definition of channel capacity, that is channel capacity is the maximum allowed signal number transmitted in \( T \) while throughput is signals actually transmitted, depending on how fast the source are sending them.

The per-node average end-to-end throughput, \( \lambda \), is defined as the arithmetic average of the end-to-end throughputs of all of the nodes; i.e.,

\[ \lambda := \frac{1}{n} \sum_{i=1}^{n} \lambda_i \]

Next, an end-to-end throughput \( \lambda_0 \) is said to be achievable by all nodes, if there exist (i) a mobility pattern of the nodes, (ii) a traffic pattern, (iii) a spatial-temporal transmission scheduling policy, and (iv) a temporal variation of transmission powers, such that \( \lambda_i \geq \lambda_0 \) for all \( 1 \leq i \leq n \). Similarly, an end-to-end throughput \( \lambda_0 \) is said to be achievable on average, if there exist (i), (ii), (iii) and (iv) from above such that \( \lambda \geq \lambda_0 \).

Next, the per-node end-to-end throughput capacity, \( \lambda_e \) defined as the supremum of all end-to-end throughputs that are achievable by all nodes. The per-node average end-to-end throughput capacity, \( \lambda_m \), is defined as the supremum of all end-to-end throughputs that are achievable on average.
Note that all of the definitions above are based on average value of the achievable per node throughput capacity, except $\lambda_e$, which is the minimum feasible throughput of all nodes. Two points in these definitions need to be clarified. Firstly, this per node throughput must be achievable, thus, it may not necessarily be a theoretical upper bound which is generally unknown till now. Secondly, the average of the throughput of all nodes only provides us limited information about the ability of the network.

_Say something about fairness here?_

- **Transmission Capacity**

  In [21], the authors defined a new measurement to the capacity, named as _Transmission Capacity_, which is defined as the maximum density of successful transmissions multiplied by their data rate, given an outage constraint.

  $$c^\varepsilon = \lambda^\varepsilon b(1 - \varepsilon)$$  \hspace{1cm} (8)

  Where $\lambda^\varepsilon$ denotes the maximum contention density such that at most a fraction of $\varepsilon$ of the attempted transmission are permitted to fail, and $b$ is the average rate that a typical successful user achieves in bit/sec/Hz.

  They found that the transmission capacity scales roughly linearly with the outage probability $\varepsilon$ for both frequency hopping and direct sequence CDMA. Note that their transmission capacity result is related to the _Transport Capacity_. As shown in the paper, the transmission capacity scales with transmission range $r$ as $\Theta(r^{-2})$. Furthermore, the transmission range scales with node number as $r^2 \propto \frac{1}{n}$. So the transport capacity can be got by $\lambda r$, which scales as $\Theta(\sqrt{n})$, the same as that reported by Gupta [1].

- **Capacity Region**

  When researchers began to study a network of channels, rather than a single channel, they found a measurement able to describe every channel was needed. Thus _Capacity Region_ was introduced by generalizing the _Channel Capacity_.

  The capacity region of a network is defined as the closure of all rate tuples that can be achieved simultaneously. Where a rate tuple specifies the rate for each of the desired source-destination pair. It is the network version of channel capacity that describes the data rate of every channel in the network and gives a joint achievable rate of these channels. If we find out the capacity region, we can understand the network better, since it describes all possible rate combinations of the source-destination pairs. Although researchers have found the capacity region for some special networks, the general capacity region for Ad hoc networks is still an open problem [6] [7] [26] [40].

2.1.2 Transport Capacity

All the definitions above considers only the data rate from the source to the destination, and they do not care how far they are from each other. Sometimes the distance between them is important especially when the channel is wireless where there are interferences between simultaneous transmissions.

The notion _Transport Capacity_ proposed by Gupta and Kumar has become popular since their paper [1] was published. It is defined based on the bit-meter notation, which means that one bit has been transported a distance of one meter toward its destination, and feasible throughput concept, as we have discussed before. Thus the transport capacity is actually the throughput capacity multiplying the average distance between the source and the destination.

A formal definition can be found in [10]. The network _Transport Capacity_ $C_T$ is:

$$C_T := \sup_{\{\lambda_i, R(i)\} \text{ feasible}, i \in T} \sum_{i \in T} \lambda_i R(i) d_i R(i)$$  \hspace{1cm} (9)

Similar definition can be found in [13] and [20].

The _Transport Capacity_ captures the physical difficulty to carry data to some distance away. However with such a definition, we know only an aggregated capacity of the network. From this point of view, this definition has the same ability as the average _Throughput Capacity_ described
above, neither measurement is able to reflect the transmission ability of each node, as *Capacity Region* is.

We shall note that, since most of the time, data is supposed to be generated by sources evenly distributed in network, both spatially and temporally, we can translate the aggregated capacity into per-node capacity by simply dividing it by the number of sources. Thus, the results presented below are in the aggregated version, exceptions will be pointed out explicitly. Furthermore, if the average distance between the sources and destinations is on the order of 1 m, the transport capacity takes the same value as the throughput capacity.

## 2.2 Interference Models

In this section, we will discuss the interference models used, based on two models proposed by [1], known as *Protocol Model* and *Physical Model*.

### 2.2.1 Protocol Model

For *Protocol Model*, the transmission rate is a binary function of the distance between a receiver and all the senders, including both the desired sender and other senders that are transmitting at the same time, to receivers other than currently considered one. If the condition is met, the transmit will succeed at a specific rate. Otherwise, the transmission will fail. This model is simple to analyze but unable to capture the fact that the receiver is able to receive under interference, only at a lower rate, offered by current technology.

To decide if the transmission will succeed, *Protocol Model* was proposed in [1] in the following form:

\[
d_{k,R(i)} \geq (1 + \Delta)d_{i,R(i)} \tag{10}
\]

for each simultaneously transmitting node \( k \), where \( \Delta \) is the width of the protecting range around the receiver, as illustrated in Fig. 1.

By setting \( d_{i,R(i)} \) to constant, it models the case when all transmitters take a common transmission range.

- **Transmitter Based Protocol Model**

In later paper [2], the authors proposed a variant of the *Protocol Model*. In this variant, the transmission from node \( i \) to its receiver \( R(i) \) is successful if:

\[
d_{k,R(i)} \geq (1 + \Delta)d_{k,R(k)} \tag{11}
\]

for every simultaneously transmitting node \( k \). While here \( \Delta \) models a guard zone around the transmitter rather than a guard zone around the receiver as in the original Protocol Model, as illustrated in Fig. 2.

There is another variant of the *Protocol Model* which is able to model the effect of directed antennas. In this model, successful receiving occurs when either the receiver \( R(i) \) lies outside the beam spread of other simultaneous transmitters \( k \), or the condition of the *transmitter based protocol model* holds.

- **Generalized Protocol Model**

In [8], the authors made further generalization on the *Protocol Model*, enabling it to model more kinds of interferences.

Let \( \{(k, R(k)) : k \in T\} \) denote the set of active transmitter-receiver pairs. Associate each pair an *interference region*, \( I_k \), and assume \( R(k) \in I_k \). Then the transmission from \( k \) to \( R(k) \) is successful if

\[
R(k) \notin I_j, \forall j \in T, j \neq k. \tag{12}
\]
By finding the interference region for specific cases of interests, we can decide the maximum number of successful simultaneous transmissions, thus derive capacity bounds for the network.

The Protocol Model is simple thus analytically attractive, and many works are based on this model, such as [5] [12] [29] [30] [31] [35] [41].

2.2.2 Physical Model

To model more realistic situations, researchers employ the Physical Model, with which we are able to incorporate different physical layer technologies such as CDMA, UWB and different modulation schemes into analysis. Physical Model usually includes a propagation model, an SINR model and a successful receiving model.

Propagation model reflects how much energy of the transmitted signal can be received by a receiver at certain distance. With this model, we can integrate many factors into the capacity analysis, such as absorption and different path loss exponents. SINR model reflects the signal to interference and noise ratio at the receiver, and is used by the successful receiving model to decide if the transmitter can transmit data successfully or at what rate the data can be transmitted. Generally, we have three kinds of receiving models: threshold based, rate based and probability based models. Under threshold based models, if the SINR is beyond the threshold, the transmitter can send successfully to the receiver at a specific data rate, otherwise, it can not send any. In contrast, the rate based models determine the transmission rate at which the transmitter is able to send its data to the receiver reliably, based on receiver SINR. For probability based receiving model, the receiver can receive successfully if the probability that SINR is below the threshold is less than a certain value.

Now we discuss the Physical Model based on its three parts.

**Propagation Model**  The propagation model defines the received signal power, with respect to the transmission power of the signal, the distance it traveled and different types of attenuation it is subjected to.
• **Power Law Propagation**

A simple propagation function may take the following form, as defined in [1]:

$$P_{R(i)} = P_i d_{i,R(i)}^{-\alpha}$$

(13)

where $2 < \alpha < 4$ is usually assumed.

This model was also used in [11] [21] [22] [33] [38] [37]. However, it will give unbounded receiving power when the receiver is infinitesimally near to the transmitter, somewhat unrealistic. As we will discuss later, whether the receiving power is bounded at origin does matter to both network connectivity and network capacity. So some modifications are made to this model.

• **Bounded Power Law Propagation**

If we want to bound the receiving power at origin, the option is the bounded power law attenuation function as proposed in [9]:

$$a(x) = (1 + x)^{-\alpha}, x \geq 0$$

(14)

Based on this attenuation function, the propagation model is defined by:

$$P_{R(i)}(t) = P_i(t) a(d_{i,R(i)}(t))$$

(15)

This model makes sense even when the distances between nodes are very small. Note the proposed model is a function of time, which reflects the time varying characteristics of the channel. In [13] and [26], the authors also made use of model.

• **Propagation with Absorption**

To incorporate the impact of other types of attenuation, the authors of [10] proposed a more complex propagation model which takes the following form:

$$P_{R(i)} = e^{-\gamma d_{i,R(i)}} P_i d_{i,R(i)}^\alpha$$

(16)

Where $\gamma$ is the absorption exponent. By setting different combinations to $\alpha$ and $\gamma$, we can model different attenuation situations. Following the same model, [20] and [24] studied the impact of different $\alpha$. The authors of [17] also made use of this model.

• **Ergodic Phase Fading**

In [34], the authors introduced the *Ergodic Phase Fading* in their propagation model, which takes the following form:

$$P_{R(i)}(t) = e^{j\theta_{i,R(i)}(t)} d_{i,R(i)}^\alpha P_i(t)$$

(17)

where $j^2 = -1$ and $\theta \sim U[-\pi, \pi]$.

• **Propagation with Shadowing**

In [6], a propagation model with shadowing effect was used. The model was defined as:

$$P_{R(i)} = KS_{i,R(i)} \left( \frac{d_0}{d_{i,R(i)}} \right)^\alpha P_i$$

(18)

Where $K$, $d_0$ are constants used to normalize the function and $S_{i,R(i)} = S_{R(i),i}$ is the shadowing factor, different $S$ are independently identically distributed random variables drawn from a lognormal distribution.

• **Propagation with General Fading**
The capacity under Rayleigh fading channel was studied in [7] with a propagation model in the following form:

\[ P_{R(i)} = \xi_{i,R(i)} d_{i,R(i)}^{-\alpha} P_i \]  \hspace{1cm} (19)

The difference is now \( \xi_{i,R(i)} \) are complex iid. Gaussian random variables following the standard distribution, i.e. \( \xi_{i,R(i)} \sim \mathcal{CN}(0,1) \).

By relaxing the constrains on the fading coefficient \( \xi \), we can describe more types of channel fading. In [32][44], the authors assumed that the expectation \( E[\xi_{ij}] = 1 \), and that \( \xi_{ij} = \xi_{ji} \). Distinct fading coefficients are independent and identically distributed (iid). They also assume that their complementary cumulative distribution function \( F_c(x) \) has a thin, exponentially decaying tail, formally:

\[ F_c(x) = P[\xi_{ij} > x] \leq \exp[-qx], \forall x > x_1 \]  \hspace{1cm} (20)

for some real and positive parameters \( q, x_1 \). In addition, they assume that there is a median value \( \xi_m > 0 \) such that \( P[\xi_{ij} \geq \xi_m] \geq \frac{1}{2} \). These assumptions are satisfied by most distributions used to model fading, for example the Nakagami, Ricean, Rayleigh and the trivial distribution.

- **General Propagation Model**

However, we can also assume only a general propagation function as in [25]. This model only assumes that the receiving power of signal follows a general function:

\[ P_{R(i)}(d_{i,R(i)}) P_i \]  \hspace{1cm} (21)

where \( l(\cdot) \) is a general propagation function.

The propagation model is a basic part of the Physical Model and it provides us the knowledge of the power of the received signals, both desired and undesired ones. Thus we need to know how strong the desired signals compared with the undesired ones. SINR models are developed for this purpose.

**SINR Model**

Based on the propagation model, SINR can be defined in various ways.

- **SINR with Common Noise**

A very simple SINR model was used by Gupta and Kumar in their paper [1], with only AWGN and path loss attenuation considered. The model was defined as follows:

\[ SINR_j = \frac{G_{ij} P_i}{N + \sum_{k \in T \setminus i} G_{kj} P_k} \]  \hspace{1cm} (22)

with \( G_{ij} = d_{ij}^{-\alpha} \). Where \( P_i \) is the transmission power of node \( i \), \( N \) denotes the power of additive white Gaussian noise. [11][32][44][33][25] are also based on this model.

- **SINR with Different Noises**

Similarly in [22], the authors used a physical model in the following form:

\[ SINR_j = \frac{G_{ij} P_i}{N_j + \sum_{k \in T, k \neq i} G_{kj} P_k} \]  \hspace{1cm} (23)

With the relaxation that different nodes may suffer from AWGN with different power.

The SINR models used in [6][21] take a slightly different form, but basically the same as the one described above.

- **Time Variant SINR with Channel Gain**

In [9], the SINR is given by

\[ SINR_j(t) = \frac{P_{ij}(t)}{N_j(t) + \frac{G}{T} \sum_{k \in T, k \neq i} P_k(t)} \]  \hspace{1cm} (24)

where \( T \) denotes the set of simultaneous transmitters at time \( t \), and \( G \) is the processing gain which is able to model wideband systems as CDMA. Similar SINR models are used in [37][38].
Successful Receiving Condition  Generally we have three types of successful receiving condition models: threshold based, rate based and probability based.

- **Threshold Based Receiving Model**

For threshold based receiving models, we have a predefined threshold $\beta$ such that if the receiver SINR is beyond it, the transmission will succeed at a specific data rate, otherwise, the transmission will fail, i.e. the transmission from node $i$ is successfully received by $j$ if $SINR_j \geq \beta$. This model was used in [1] [2] [9] [11] [22] [37] [38].

- **Rate Based Receiving Model**

If the transmission rate is assumed to be a continuous function of the SINR, rather than a binary function, we can model the channel that can transmit at a certain rate depending on the channel condition.

In [6], the authors assumed that the transmission rate to be a general continuous function of SINR. While in [32] [44], they assumed a special function which takes the following form:

$$f(SINR_j) = W \log(1 + \frac{1}{\Gamma} SINR_j)$$  \hspace{1cm} (25)

When $\Gamma = 1$, the receiver achieves the Shannon’s capacity, when $\Gamma > 1$, the receiver can achieve the maximum rate that meets a given BER requirement under a specific modulation and coding scheme. However, if we take an information theoretic approach, we can simply model the received signal (possibly with noises and interferences), and the achievable rate becomes the consequence of the derivation. This reflects the fundamental position of information theory. A common practice is to assume that the received signal (usually called message in information theory) as follows: for a receiver $j$, the received message $Y_j$ is:

$$Y_j = \sum_{i \in T} G_{ij} X_i + Z_j$$  \hspace{1cm} (26)

where $G_{ij} = g(d_{ij})$ is the channel gain function, $X_i$ denotes the message sent by node $i$. $Z_j$ is a circular symmetric complex Gaussian random variable with unit variance, and is independent with different $j$. Then the total transmission power of all nodes is constrained as:

$$\sum_{i \in T} E(|X_i|^2) \leq nP$$  \hspace{1cm} (27)

These assumptions model a set of Gaussian channels between nodes, then known results about this channel can be applied to derive the capacity of the network. [4] [7] [10] [13] [17] [24] [26] [34] are works taking this approach.

- **Probability Based Receiving Model**

In [21], the authors proposed a probability based receiving model for both FH-CDMA and DS-CDMA. For FH-CDMA, they assumed the bandwidth $W$ is divided into $M$ sub-channels each with a bandwidth of $\frac{W}{M}$, thus, for each channel, the ambient noise power is $\eta = N \frac{W}{M}$. For DS-CDMA, the noise power is just $M\eta$. The receiving model for FH-CDMA is such that a receiver $j$ can receive successfully if:

$$P(SINR_j \leq \beta) \leq \varepsilon$$  \hspace{1cm} (28)

and for DS-CDMA,

$$P\left(SINR_j \leq \frac{\beta}{M}\right) \leq \varepsilon$$  \hspace{1cm} (29)

where the $\varepsilon$ is the outage constraint.
2.3 Network Structure

Network structure has fundamental impacts on capacity. Generally, Ad hoc networks may have two kinds of structures, pure Ad hoc network, which we name it Flat Network; and network with Ad hoc nodes and infrastructure nodes that aide the Ad hoc nodes, we name it Hybrid Network. We will discuss them in this section. Note that pure infrastructure based wireless networks such as cellular networks are out of the scope of this paper.

2.3.1 Flat Ad hoc Network

In [1], Gupta and Kumar studied the capacity of a typical flat Ad hoc network. They proposed two network models for this kind of network: Arbitrary Network and Random Network. In an arbitrary network, nodes are located arbitrarily and each of the senders can choose an arbitrary transmission power, for each transmission. For random networks, nodes are randomly distributed according to independent uniform distribution and nodes take a common transmission power. The difference between these two models is that in arbitrary network, we have higher degree of freedom to manipulate the network to maximize its capacity, as a result, in arbitrary network, the capacity can be higher than in a random network by a factor of $\sqrt{\log n}$ according to [1].

The flat Ad hoc networks are mainly studied under these two models and we will discuss some important results here.

In [1], the asymptotic scaling laws of aggregated throughput capacity when the number of nodes $n$ goes to infinity were reported, as shown in Table 1.

<table>
<thead>
<tr>
<th>Protocol Model</th>
<th>Arbitrary Network</th>
<th>Random Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Model</td>
<td>$\Theta (\sqrt{n})$</td>
<td>$\Theta (\sqrt{\frac{n}{\log n}})$</td>
</tr>
</tbody>
</table>

The lower bound for random network under physical model was later improved to meet the upper bound in [11], making the $\sqrt{n}$ bound tight.

Network coding was used by [4] to achieve higher transport capacity of a flat network. The authors showed that in a network with single source-destination pair, the capacity of the source scales as $\Theta(\log n)$ if arbitrarily complex network coding allowed. As compared with simple point-to-point coding scheme, for example, $\Theta(\sqrt{n})$ scaling law discussed above, which results $\Theta(1)$ scaling when applied to the same scenario, network coding changed the scaling character fundamentally. However their results also suggests that pure network coding is unable to achieve maximum throughput for the whole network, since only one source-destination in the network achieves $\Theta(\log n)$ scaling.

Upper bound of transport capacity was got in [13] via random cut-set approach. The authors show the upper bound should be $O(n \log n)$ if channel state information(CSI) is not available, and can be as high as $O(n \log^2 n)$ if CSI is available to facilitate sophisticated receiving scheme.

Note that all the results are presented as throughput capacity in bits/sec, and can be translated into transport capacity with average source-destination distance of 1m.

Later in [5], the authors further improved the bounds by exploiting directional antennas. Their results showed that for arbitrary network, by using a directional transmitter but an omnidirectional receiver, the transport capacity scales as $\frac{1}{\sqrt{\beta_T}} \sqrt{n}$ bit-meters/sec. Compared with Gupta’s result, the capacity is scaled by a factor of $\sqrt{\frac{2\pi}{\beta_T}}$, where $\beta_T$ is the circular sector of the transmitter. If omnidirectional transmitters and directional receivers are to be used, the capacity will be scaled by a factor of $\sqrt{\frac{2\pi}{\beta_R}}$, where $\beta_R$ is the sector angle of the receiver antenna. If both transmitters and receivers are directional, the capacity will be scaled by $\frac{2\pi}{\sqrt{\beta_T \beta_R}}$, compared with omnidirectional case. For random networks, compared with all antennas are omnidirectional, the throughput gains are $\frac{2\pi}{\beta_T}$, $\frac{2\pi}{\beta_R}$ and $\frac{4\pi^2}{\beta_T \beta_R}$, under conditions that the transmitter antennas are directional, the receiver antennas are directional and both antennas are directional, respectively. In [8], general
interference models are proposed which is able to model different types interferences including directional antennas.

The throughput capacity of wireless networks with ultra wide bandwidth was studied in [14]. For such a network, each transmitter subjects to a power constraint $P_0$ and the bandwidth of each communication link is very large, i.e. $W \to \infty$, resulting a very low spectral efficiency, $\frac{P_0}{W} << 1$. The per node throughput was shown to be upper bounded to $O((n \log n)^{\frac{1}{2}})$ and lower bounded to $\Omega((n \log n)^{-\frac{1}{2}})\log n^{-\frac{1}{2}}$.

Realistic link layer was considered in [18]. Since in real network, there exists non-zero probability of packet loss at each link during the packet forwarding process, the cumulative effects will not be negligible. This study showed that this will actually force the per node throughput scales as only $O\left(\frac{n}{n}\right)$. However, a proper scheduling scheme was developed as a countermeasure to this effect. The proposed scheduling scheme is able to achieve a scaling on the same order of the theoretical upper bound $(n \log n)^{-\frac{1}{2}}$, with only a constant fraction of capacity loss.

It has been proved that dividing the bandwidth to multiple sub-bands will not improve the capacity in [1]. Furthermore, [16] studied the impacts of the number of channels and interfaces to capacity. They found that for Arbitrary Network, when the number of interfaces is less than channel number, there will be capacity loss compared with when each channel has a dedicated interface. However, for Random Network, the capacity loss happens only when the ratio of channel number to interface number is over $\log n$. Thus it is feasible to build a multi-channel Ad hoc network with single interface that achieves the $\sqrt{n \log n}$ upper bound.

Under a General Fading Model (will be discussed later), the throughput scales as $\Theta\left(\sqrt{\frac{n}{\log n}}\right)$, shown in [44].

The scaling laws were also extended to three dimension space in [2]. The aggregated throughput capacity is shown in Table 2.

Table 2: Throughput Capacity Bounds for Three Dimensional Network.

<table>
<thead>
<tr>
<th>Protocol Model</th>
<th>Arbitrary Network</th>
<th>Random Network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Theta(n^{\frac{1}{2}})$</td>
<td>$\Theta\left(\frac{n}{n \log n}\right)^{\frac{1}{2}}$</td>
</tr>
<tr>
<td>Physical Model</td>
<td>$\Omega(n^{\frac{1}{2}})$ and $O(n^{\frac{1}{n}\frac{1}{2}})$ with $\alpha &gt; 3$.</td>
<td>$\Omega\left(\frac{n}{n \log n}\right)^{\frac{1}{2}}$ and $O(n^{\frac{1}{2}})$</td>
</tr>
</tbody>
</table>

Besides the works mentioned above, many others are based on the flat network model, such as [3], [6], [7], [9], [10], [12], [17], [20]-[29]. These works made further assumptions about the networks from different aspects and got different results. For example, in [3], the authors studied the transport capacity with different types of average source-destination distance distributions, in flat networks, under four type of network layouts: single sender-receiver pair, one-dimension network (chain of nodes), regular lattice network and random network. We will discuss their models and results in more detail in Section 2.4.1. In [3], the authors reported the transport capacity scaling of $\Theta(1)$ based on a more realistic attenuation function which is reasonable even when the nodes are arbitrarily close to each other, as we have discussed in Section 2.2.2.

2.3.2 Hybrid Networks

On the other hand, Ad hoc networks with infrastructure support are also studied and representative works are [30], [31], [32], [33], [35], [34]. When both infrastructure nodes and Ad hoc nodes present in the network, their relationship becomes important. Almost all of the current works concentrated on this issue.

The capacity problem under hybrid ad hoc network was studied in [30]. They assumed that in a hybrid Ad hoc network, the Ad hoc nodes are independently and uniformly distributed and all have the same transmission power. The base stations are regularly placed, such that they divide the area into hexagon tessellations, as shown in Fig. 3. The bandwidth between nodes and between
nodes and base stations are both \( W \). Two routing strategies were proposed, namely, k-nearest-cell routing and probabilistic routing. We briefly describe them here.

**k-nearest-cell routing**: If the destination is within the \( k \) nearest neighboring cells as the source, the data will be forwarded in ad hoc mode (do not rely on base stations), otherwise, the infrastructure nodes will be used to forward the data.

**probabilistic routing**: Transmission mode is chosen by each source independently such that ad hoc mode is chosen with probability \( p \), and infrastructure is used with probability \( 1 - p \).

The authors found that for k-nearest-cell routing strategy, if the number of base stations \( m \) grows slower than \( \sqrt{n} \), the throughput capacity scales with \( n \) and \( m \) as \( \Theta \left( \left( n \log \frac{n}{m^2} \right)^{-\frac{1}{2}} \right) \), the benefit of adding \( m \) is insignificant. If \( m \) grows faster than \( \sqrt{n} \), the throughput capacity scales as \( \Theta \left( n \right) \), as Fig. 4 shows. For probabilistic routing strategy, if \( m \) grows slower than \( \sqrt{\frac{n}{\log n}} \), the throughput capacity scales as \( \Theta \left( \sqrt{n \log n} \right) \), the same as pure ad hoc network. If \( m \) grows faster than \( \sqrt{\frac{n}{\log n}} \), the throughput capacity scales as \( \Theta \left( m \right) \), as shown in Fig. 5.

Higher capacity results were reported in [31]. In this paper, access points and Ad hoc nodes are
both uniformly distributed. The authors found that under such a network, the throughput capacity scales as $\Theta\left(\frac{n}{\log n}\right)$. They also showed that even the network is allowed to be not fully connected, the throughput capacity is still not able to scale as $\Theta(n)$. However, in a following paper [33], the authors found that it is possible to provide a $\Theta(1)$ scaling of throughput capacity to any fraction $f, (0 < f < 1)$ of nodes.

(We need to think what is the reason that these two papers gave out different results.)

Based on the Physical Model, the authors of [32] found that when $m$ scales as $n^c$, throughput capacity scales as $n^c$ if $\frac{1}{2} < c < 1$, and as $n^\frac{1}{2}$ if $0 < c < \frac{1}{2}$.

In a recent paper [35], the authors supposed that ad hoc nodes are randomly distributed and infrastructure nodes are arbitrarily deployed in the network. The scaling laws of throughput capacity are given as follows.

- When $m = O\left(\sqrt{\frac{n}{\log n}}\right)$, throughput scales as $\sqrt{\frac{n}{\log n}}$, the same as pure Ad hoc network.
- When $m = \Omega\left(\sqrt{\frac{n}{\log n}}\right)$ and $m = O\left(\frac{n}{\log n}\right)$, the throughput scales as $\Theta(m)$.
- When $m = \Omega\left(\frac{n}{\log n}\right)$, the throughput scales as $\frac{n}{\log n}$.

Fig. 6 shows that only when $m = \Omega\left(\sqrt{\frac{n}{\log n}}\right)$ and $m = O\left(\frac{n}{\log n}\right)$, Ad hoc nodes will benefit most from the infrastructure since the capacity scales linearly with the number of infrastructure nodes. In contrast, under the case either $m = O\left(\sqrt{\frac{n}{\log n}}\right)$ or $m = \Omega\left(\frac{n}{\log n}\right)$, increasing the number of infrastructure nodes will not increase the transport capacity, since the capacity is shown to depend solely on the number of Ad hoc nodes.

The effect of the infrastructure on the throughput is it keeps a constant number of hops from sources to destinations, in contrast to pure ad hoc networks where hop number between sources and destinations grows with $\sqrt{n}$. Apparently, hybrid networks are superior to flat network with respect to capacity scaling. However it is not always true according to [30] [35], as we have already discussed.

In [34], the authors studied a three-tire hierarchical network with an additional tire formed by relaying nodes whose only duty is to facilitate the data transmission by cooperating with other nodes. Capacity region was got for this network based on the multiple-access relay channel model. This paper did not study the capacity scaling laws with either Ad hoc nodes number or access points number.

2.4 Traffic Pattern

In this section, we consider two different but related aspects of traffic pattern in an Ad hoc network. They are average source-destination distance and ratio of source nodes to destination nodes. Based on the first consideration, we classify the models into random traffic model and arbitrary traffic model, then based on the second consideration, we classify the models into symmetric traffic model and asymmetric traffic model.
2.4.1 Random Traffic vs. Arbitrary Traffic

For random traffic, each source node chooses a random node as the destination of its data. This procedure will result in average distance between source-destination pairs on the same order as the square root of the network area, for networks with static nodes. By assuming a random traffic, we can expect that the traffics are distributed evenly in the network, implying that there exists a scheduling algorithm that enables \( \Omega(n) \) nodes to send simultaneously without interfering with each other. This effect is helpful in analyzing the network capacity when the underlying interference model assumes a simple receiving scheme that takes interference as noise. Studies with this assumption include [11] [30] [31] [32] [44] [33] [35].

However, when more sophisticated receiver operations are allowed, the even distribution of the traffic becomes less important since interference can be utilized to achieve cooperative transmission of information, thus, physically interfered nodes can be scheduled to transmit at the same time. On the other hand, if mobile Ad hoc networks are considered, it is not important how node choose its destination. Only assuming the source-destination pairs are bound is sufficient to get a random traffic in the network as long as nodes move independently and randomly. Since the data are sent to their destinations through at most two hops (by a relay node or directly transmitted from source to destination), the traffic in this kind of mobile Ad hoc networks is generally respected as local traffic. [36] [37] [38] [39] [40] are studies with this kind traffic model.

Different from random traffic assumption, some researchers assumed an arbitrary traffic dominates the network. [1] and [2] proposed to use arbitrary traffic to achieve an ideal and optimal case that gives the highest possible transport capacity. This method was used by [5] [8] and [9] to study the capacity under optimal cases. In contrast, random traffic assumption reflects a more realistic case.

The average distance between the source-destination pairs have impact on \( \text{throughput capacity} \). With this understanding, some researchers made assumptions directly on this value. In [3], the authors studied two distributions of source-destination distance \( L \) and their impacts on the scaling laws of throughput capacity. They found that the per node throughput capacity scales inversely with the average source-destination distance, i.e.

\[
\lambda < k L^{-1}
\]

for some constant \( k \).

When traffic is random and the network occupies a square of with area \( A \), the distance between source and destination was shown to be a random variable with the following probability density function (pdf):

\[
p(L) = \frac{L}{\int_0^{\sqrt{A}} t \, dt} \tag{31}
\]

Thus, the average distance can be retrieved by taking its expectation:

\[
\bar{L} = \int_0^{\sqrt{A}} L p(L) dL = \frac{2\sqrt{A}}{3} \tag{32}
\]

If the area of network \( A \) scales with the number of nodes \( n \), per node throughput capacity will scale as \( O\left(\frac{1}{\sqrt{n}}\right) \). This result shows that the random traffic pattern is not scalable since per node capacity goes to zero as the network increases.

However the authors showed that the capacity scales differently when \( L \) is power law distributed with the following pdf:

\[
p(L) = \frac{L^\delta}{\int_{\epsilon}^{\sqrt{A}} t^\delta \, dt} \tag{33}
\]

With the same approach as the random traffic case, we have Table [3].

Thus for each region of \( \delta \), we have a corresponding capacity result. It is interesting to point out that when \( \delta < -2 \), average distance between source-destination is constant, resulting constant throughput capacity scaling with number of nodes in extensible network. Note that the Transport Capacity is not affected by the average source-destination distance.
Table 3: Throughput Capacity with Different Average Distance between S-D Pairs.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$\lambda$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\delta+1}{\delta+2}$</td>
<td>$O(1)$</td>
<td>When $\delta &lt; -2$ and $A$ is large</td>
</tr>
<tr>
<td>$O\left(\frac{\log A}{\lambda} \right)$</td>
<td>$O\left(\frac{1}{\log A} \right)$</td>
<td>When $\delta = -2$</td>
</tr>
<tr>
<td>$\frac{2\sqrt{A}}{\log A}$</td>
<td>$O\left(\frac{\log A}{\sqrt{n}}\right)$</td>
<td>When $-2 &lt; \delta &lt; -1$ and $A$ is large</td>
</tr>
<tr>
<td>$\frac{2\sqrt{A}}{\sqrt{n}} \sqrt{A}$</td>
<td>$O\left(\frac{1}{\sqrt{n}} \right)$</td>
<td>When $\delta &gt; -1$</td>
</tr>
</tbody>
</table>

In [12], the authors also made assumptions directly on the distance between source-destination pairs such that the sum of all source-destination distances scales as $\Omega(n\sqrt{A})$, this assumption is basically the same as the random traffic assumption. The papers [1] [2] [5] [8] studied both random traffic and arbitrary traffic, providing us with good comparisons of the impacts that different traffic patterns.

2.4.2 Symmetric Traffic vs. Asymmetric Traffic

An other important factor that impacts the network capacity is the ratio of source nodes to destination nodes. When they are of the same number, the traffic will be evenly distributed across the network as long as the destination nodes are selected randomly. Most of the works we have discussed is under this assumption. However, when many source nodes send data to the same destination, the throughput of each node will be further limited by the aggregation of traffic, and bottlenecks may be formed near the destination. Wireless sensor networks are very good examples for this kind of network. On the other hand, if there are only a few source nodes but a lot of destination nodes, a broadcast traffic will be formed. In this section, we will discuss these two types of asymmetric traffic and their impacts on the network capacity. We will take a network point of view and concentrate on the capacity of multihop networks and will not discuss the capacity of classical multiple access and broadcast channels.

- Convergent Traffic

In [41], the authors studied the transport capacity of dense wireless sensor networks (WSNs), where the data always converge to a single sink. They found that the aggregated throughput capacity scales as $\Theta(1)$ when the number of nodes in a fixed area goes to infinity. Furthermore, the authors also showed that for WSNs, even the amount of data each node has to generate approaches to zero as the density of nodes increases, the total amount of data the network has to send back to the sink always goes to infinity at the same time, in order that the gathered data shall keep a predefined mean square error (MSE). This implies that no quantization-encoding-transport scheme is sufficient to achieve the desired MSE when $n$ becomes large.

Recently, researchers found that in typical WSN applications, the data originated from nearby sensors are correlated especially when the sensors are densely deployed. This is quite different from traditional Ad hoc networks, where the traffic are assumed to be uncorrelated. If the correlation can be exploited, it is possible to achieve higher capacity. In [42], the authors pointed out that the $\Theta(1)$ result is somewhat pessimistic. By exploiting data correlation between nearby sensors, they showed that $\Theta(\log n)$ scaling law of throughput capacity is achievable in dense WSNs.

Besides, in [32], the authors studied the general cases of asymmetric networks by assuming $n$ source nodes and $m$ destination nodes in the network. Each source node creates data that is supposed to be delivered to a randomly chosen destination node. The capacity depends on $n$ and $m$ under this case. They studied the capacity scaling law by varying the ratio between the two kinds of nodes and found that when $m = n^d$,

- with $0 < d < \frac{1}{2}$, bottlenecks form around the destinations, and aggregated throughput is only $n^d$,
- with $\frac{1}{2} < d < 1$, the aggregated throughput scales as $n^{\frac{d}{2}}$, and no bottlenecks formed.
This is a quite interesting result, since it includes both symmetric traffic and the single destination cases. When \( d \to 0 \), we have the result for a network with traffic converge to a constant number of destinations. When \( d \to 1 \), we have the result for a network with symmetric traffic.

- **Broadcast Traffic**

When traffic is to be broadcast in the network, the packets originated form one source should be received by all other nodes, thus we need to extend the capacity definition to fit the broadcast scenario.

*Broadcast Capacity* is defined as the maximum rate at which broadcast packets can be generated by the source nodes such that all other nodes receive the packets successfully in finite time \([27][28]\). Note that this definition defines the *Broadcast Capacity* as a kind of *Throughput Capacity*.

In \([23]\), the authors showed that the upper bound of broadcast capacity is \( O(1) \) under a constant density network with \( n \) nodes (We will discuss constant density network in Section ?? as a kind of Extended Network). Since all nodes except the source should receive the packet, the total path length of each packet is on the order of the network area \( A \), which scales as \( \Theta(n) \) in constant density networks. Together with the fact that at most an order of \( n \) transmissions can take place at the same time in the network, the \( O(1) \) upper bound could be derived. A later work \([24]\) studied the broadcast capacity in both Dense Networks and Extended Networks. The author showed that the broadcast capacity scales as \( \Theta((\log^{-\alpha} n)) \) when \( \alpha > 2 \), for extended network; while for dense network, the upper bound and the lower bound were shown to be \( O((\log(n))) \) and \( \Omega(1) \), respectively. However in \([25]\), the authors showed a different scaling law of broadcast capacity. They adopted a protocol model and assumed an homogeneous node distribution in \( d \) dimensional space. As a result, the scaling law is with respect to the dimension of space and the protecting region defined by the Protocol Model. The broadcast capacity was shown to scale as \( \Theta((\log^{-\alpha/2} n)) \) for dense networks, where \( \Delta \) denotes the protect region defined in protocol model. Furthermore, they found that mobility cannot significantly increase the broadcast capacity. With the help of mobility, the broadcast capacity changes at most by a factor of \( O((\log(n))) \). Note that the result is \( \Theta(1) \) when translated to scaling law with respect to \( n \).

### 2.5 Mobility

The studies on the capacity of mobile Ad hoc network began from the works of Grossglauser and Tse. They found that the mobility of nodes can be exploited to increase the network capacity \([30]\). We will discuss their works as well as others based on the mobility models they used.

#### 2.5.1 Random Uniform Mobility Model

The basic assumptions of *Random Uniform Mobility* model are that when the nodes are moving,

- the process of their positions is stationary and ergodic with uniform distribution and
- the mobility pattern is completely random, i.e. the trajectories of the nodes are completely mixed, independent and identically distributed.

In \([30][38]\), the authors proposed a two-phase scheduling policy in order to utilize the mobility. This mechanism enables each source node to split its packet stream to intermediate relay nodes. When the relay nodes are near enough to the destination, they forward the packets to it. By dispatching packets to several relaying nodes, the network is able to efficiently exploit the multi-user diversity provided by mobility. The two phase scheduling policy achieves a per source-destination throughput of \( \Theta(1) \). In contrast, if we forbid mobile relaying, the achievable per source-destination throughput goes to zero at least as fast as \( n^{-1/(1+\alpha/2)} \), for sufficiently large \( n \). Furthermore, in order to achieve a per source-destination throughput of \( \Theta(1) \), \( \Theta((\sqrt{n})) \) relaying nodes are necessary. One major constrain of the proposed mechanism is it incurs unbounded packet delay, which limits its application to only delay tolerable ones.

In \([44]\), the throughput capacity bounds are derived under certain delay constraint. Their results show that the throughput capacity of \( \sqrt{\frac{n}{\log n}} \) is achievable with constant delay and \( n^{\frac{d}{2} + \frac{1}{2}} (\log n)^{\frac{d}{2}} \) is achievable with an \( O(n^d) \) delay, where \( 0 < d < 1 \) is a trade-off coefficient.
2.5.2 Markovian Random Walk

Later results in [40] collaborated the $\Theta(1)$ scaling law. The authors assumed that $n$ users are roaming independently in the network, which is partitioned into $C$ cells. The nodes have the same steady state location distribution $\pi_c$ over cells, $c \in \{1, 2, \ldots, C\}$ (Note that the distribution is not necessarily uniform). They further assumed that the nodes can communicate only when they are in the same cell or in adjacent cells.

2.5.3 One-Dimensional Mobility Model

In [37], the authors developed a mobility model which is believed more realistic. In this model,
- each node moves along a great circle on a unit sphere and
- each node chooses its great circle randomly and independently and
- the location of node in its great circle is a stationary and ergodic random process with uniform distribution.

They showed that the throughput scales as constant with respect to the number of nodes $n$, for almost all configurations of the great circles.

2.5.4 High Mobility

On the other hand, the results reported in [39] showed that in Ad hoc networks with high mobility, the nodes may not be able to get capacity gain by exploiting the mobility.

The authors assume the nodes move very fast in the network, and under such harsh conditions, communications between nodes can take place only if they are equipped with multiple antennas which enables the nodes to send and receive simultaneously, giving the nodes potential benefits from channel feedbacks. But due to the effect of high mobility, the channel may fluctuate too fast to be tracked by the transmitter or receiver. If the channel coherent time can not be kept beyond a certain level, feedback can not be retrieved. Furthermore, under such case, the transmitters will not be able to decide which node is its nearest neighbour, thus unable to adjust transmission power to limit interference as one does in static Ad hoc networks. The mobility was modeled based on the resulting CSI states, such that if CSI is available, the mobility is regarded as low mobility, otherwise if CSI is unavailable, the mobility is regarded as high mobility.

Exact capacity region was determined under this case for any given partition of source, destination and relaying nodes. The throughput capacity is shown to be upper bounded by $\log \log(n)$. Even if the channel status can be captured perfectly, due to the inability of senders to track the network topology, the throughput capacity still can not scale faster than $\log(n)$.

The reason behind the capacity loss in case of high mobility is believed to be two fold. Firstly, due to high mobility, the channels between nodes vary rapidly and channel estimation is not possible, this is known as channel uncertainty. Secondly, high mobility makes it impossible for the transmitter to distinguish between the nearest nodes and distant nodes, network seems more homogeneous compared with the static case. The above two negative effects of high mobility lead to the loss of degree of freedom in a multiple antenna network, making CSI unavailable and decreasing the network capacity dramatically.

The results presented above are very different from those of static networks, which implies that the mobility changes the network characteristics fundamentally. Although only certain types of mobility model are analyzed, they show a great diversity that the capacity maybe higher or lower than that of a static network. Besides, one caveat of utilizing mobility to achieve higher capacity is the longer delay thus incurred. This problem has triggered many research works recently trying to trade-off between capacity and packet delay in mobile Ad hoc networks.

2.6 Density

The network models used in the study of capacity can be classified into Dense Network or Extended Network, based on node density.
If a network occupies a fixed area, when the number of nodes increases, the distance between nodes decreases and the density of nodes increases. One consequence of this procedure is more nodes interfere with each other, as long as the transmission range of each node is fixed. This kind of network is usually referred to as Dense Network. In order to decrease interference between nodes in a Dense Network, we have to decrease the transmission range. However, this will shorten the distance each hop travels, making packets have to take more hops to arrive their destinations.

On the contrary, if the area of network scales with the number of nodes sufficiently fast, the nodes that interfere with each other will not increase when more nodes join the network, even they have fixed transmission range. However, the occupation of the network becomes larger. This kind of network is usually referred to as Extended Network, it includes Minimum Distance Network and Constant Density Network. For the former, the distances between nodes are lower bounded by a constant and for the latter, the density of nodes keeps constant. The extension of the network area will result more hops from sources to destinations. The countermeasure is to keep the traffic locally, i.e. prevent the average distance between the source and destination to scale with the network area. This is the basic idea why we exploit infrastructure or mobility to increase transport capacity.

From the discussions above, we find there is no fundamental difference between the two models. We can translate from the result derived under Dense Network to that under Extended Network, by identifying how the area scales with the number of nodes. For example, in [10], the aggregated transport capacity of all nodes under high attenuation case is shown to scale as $\Theta(n)$, under minimum inter-node distance assumption. While the capacity was also reported to scale as $\Theta(\sqrt{An})$, which is the optimal case can be achieved. Thus, we see the two results are the same when $A$ scales as $\Theta(n)$.

Furthermore, both assumptions have their roots in real world applications. For example, we may want to expand the coverage of an existing network by deploying new nodes in the uncovered area with node density same as the original network, or we may also encounter the situation that people gathering to a conference hall with their laptops, forming an Ad hoc network with increasing density.

Among previous works, [1], [2], [3], [4], [5], [6], [7], [8], [9], [11], [12], [22], [23], and [30]-[35] assumed a Dense Network, and [10], [13], [21], [26] assumed a Minimum Distance Network. While the works [3], [17], [25] assumed a Constant Density Network. However, there are also works with no assumption about the network density such as [6], [7], [21], and [44], they only derived the information capacity of the network without consideration on the distance the information can be transmitted in unit time, this kind of information capacity can be translated into transport capacity by multiplying the source-destination distance.

3 Conclusion

We identify some facts about the state of art research of Ad hoc network capacity based on the analytical models we have discussed in this paper. These facts provide us important information about future research directions and methodologies.

Firstly, we find the analytical models are of great diversity. Each reflects a certain set of aspects of real network. However, one interesting question is “Are there some commonly used models”? Or “Is it better to provide a ‘Standard’ model”? We realize that if there are standard models, researchers will be able to work on the same basis and compare their results conveniently.

Secondly, we find that the models are becoming more and more complex, and researchers are faced with great challenge in analyzing them. Furthermore, the results are becoming less general and less meaningful. It maybe helpful to think about what level of details should the models say about reality?

Thirdly, great efforts have been taken to reveal capacity of Ad hoc network, some important results are available, however, much haven’t been discovered. A brief review gives us the following facts about the progress ever made and what remains unknown.

Several fundamental problems need to study:

- Why capacity problems are different for wired and wireless networks? Basically, it is the result of shared medium, but wired LANs are also working in a shared medium, are they
have something in common?

In wired shared medium, signals are always strong enough to destroy each other when they are sent simultaneously, the nodes sharing the medium basically form an one hop network, i.e. every node can send to any other node in the shared medium. However in wireless shared medium, thanks to rather strong path loss of the magnetic weave, communication can take place under interference. The consequence of this characteristic is we can hope to allow more transmissions at the same time, achieving some form of spacial multiplexing. While on the other hand, this will force the network to be a multihop network, and some of the ability to detect collisions are lost. In such a network, only receivers can detect the collision reliably, senders may not detect all the collisions taken place.

The analogy in wired LAN is using switches to separate several LAN in order to limit the contention within each LAN.

Unfortunately, we do not have similar devices as a switch that is able to separate wireless signals in shared space from interfering with each other, till now. However, we have methods that makes the wireless signals themselves not interfere at all, such as TDMA, FDMA or CDMA, at the cost of throughput, hardware complexity, etc.

In fact, neither wired nor wireless networks in shared medium scales to large number of nodes.

- Traditional source-channel coding is not enough? Network coding increases capacity?

To this end, we first figure out most important factors of the currently used models, then we propose a standard model.
References


