Chapter 2
Capacity Scaling Laws of Wireless Networks

The capacity scaling law of wireless networks has been considered as one of the most fundamental issues. In this chapter, we aim at providing a comprehensive overview of the development in the area of scaling laws for throughput capacity in wireless networks. We begin with background information on the notion of throughput capacity of random networks. Based on the benchmark random network model, we then elaborate the advanced strategies adopted to improve the throughput capacity, and other factors that affect the scaling laws. We also present the fundamental tradeoffs between throughput capacity, delay, and mobility. Finally, the capacity for hybrid wireless networks are surveyed, in which there are at least two types of nodes functioning differently, e.g., normal nodes and infrastructure nodes.

2.1 Preliminaries: Milestone of Throughput Capacity Scaling

Capacity scaling laws offer fundamental understanding on how per-node capacity scales in an asymptotically large network. The line of investigation began with [22], where Gupta and Kumar introduced two new notions of network capacity: *transport capacity* and *throughput capacity*. In this monograph, we focus on the throughput capacity. We first introduce the notion of throughput capacity and the capacity result for random networks, as preliminaries for reading the remaining sections.

2.1.1 Notion of Throughput Capacity

Let N denote the number of nodes in a network. The per-node throughput of the network, denoted by $\lambda(N)$, is the average transmission rate, measured in bits or packets per unit time, that can be supported uniformly for each node to its destination in the network. A per-node throughput of $\lambda(N)$ bits per second

N. Lu and X. Shen, *Capacity Analysis of Vehicular Communication Networks*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-1-4614-8397-7_2, © The Author(s) 2014

is said to be ***feasible*** if there exists a spatial and temporal scheme for scheduling transmissions, such that each node can send $\lambda(N)$ bits per second on average to its destination node. The throughput capacity of the network is said of order $\Theta(f(N))$[1] bits per second if there are deterministic constants $c_1 > 0$ and $c_2 < \infty$ such that

$$\lim_{N\to\infty} \mathbf{Pr}\big(\lambda(N) = c_1 f(N) \text{ is feasible}\big) = 1$$

$$\liminf_{N\to\infty} \mathbf{Pr}\big(\lambda(N) = c_2 f(N) \text{ is feasible}\big) < 1.$$

Therefore, vanishingly small probabilities are allowed for in such definition of "throughput capacity" when considering the randomness involved in the network, such as the location and the destination of each node. Note that the notion of throughput capacity is different from the information-theoretic capacity notion that describes the exact region of simultaneous rates of communications from many senders to many receivers in the presence of interference and noise [31].

2.1.2 Random Networks

A wireless random network consisting of N identical immobile nodes randomly located in a disk of unit area in the plane and operating under a ***multi-hop*** fashion of information transfer, is shown in Fig. 2.1 [22]. Each node having a randomly chosen destination is capable of transmitting at W bits per second over a common wireless channel. The requirements for successful transmission are described as per two interference models: (i) the Protocol Model, which is a binary model, i.e., the transmission is successful if there is enough spatial separation from simultaneous transmissions of other nodes otherwise fails; and (ii) the Physical Model, based on signal-to-interference ratio requirements. In such static random ad hoc network, all the nodes are assumed to be homogeneous, i.e., all transmissions employ the same range or power, and wish to transmit at a common rate.

[1] Since studies of throughput capacity focus on the scaling behavior instead of a specific value, the order notation is involved to describe how capacity scales with the number of nodes N. Specifically, the following Knuth's notation is used throughout all the papers on scaling laws: given nonnegative functions $f_1(n)$ and $f_2(n)$, $f_1(n) = O(f_2(n))$ means $f_1(n)$ is asymptotically upper bounded by $f_2(n)$; $f_1(n) = \Omega(f_2(n))$ means $f_1(n)$ is asymptotically lower bounded by $f_2(n)$; and $f_1(n) = \Theta(f_2(n))$ means $f_1(n)$ is asymptotically tight bounded by $f_2(n)$; $f_1(n) = \omega(f_2(n))$ means $f_1(n)$ is asymptotically dominant with respect to $f_2(n)$; $f_1(n) = o(f_2(n))$ means $f_1(n)$ is asymptotically negligible with respect to $f_2(n)$.

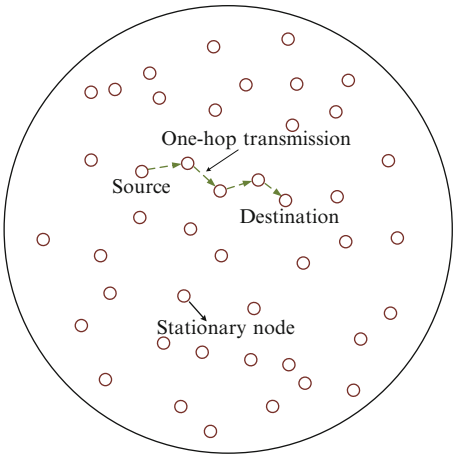


Fig. 2.1 A static ad hoc network in a unit disk

2.1.3 Throughput Capacity of Random Networks

For random networks, the order of the throughput capacity is $\lambda(N) = \Theta(\frac{W}{\sqrt{N \log N}})$ under Protocol Model (see main result 3 in [22]); while under Physical Model, the throughput capacity is given by $\Theta(\frac{W}{\sqrt{N \log N}}) \leq \lambda(N) < \Theta(\frac{W}{\sqrt{N}})$ (see main result 4 in [22]). An explanation of the results is as follows. For Protocol Model, the lower bound and upper bound are of the same order such that there exists a sharp order estimation of the throughput capacity; for Physical Model, a throughput of order $\Theta(\frac{W}{\sqrt{N \log N}})$ is feasible, while $\Theta(\frac{W}{\sqrt{N}})$ is not.

The throughput capacity is studied asymptotically, i.e., the population of nodes continuously grows. The scaling result for random networks is pessimistic because the per-node throughput tends to zero similar to $\frac{1}{\sqrt{N \log N}}$ as the population of nodes goes to infinity, which indicates that static ad hoc networks are not feasible to scale to a large size. What causes those discouraging results? The fundamental reason is that every node in the network needs to share the channel resources or certain geographic area with other nodes in proximity, which constricts the capacity. Specifically, concurrent wireless transmissions in a wireless network limit its throughput capacity, because they create mutual interference so that nodes cannot communicate as that in the wireline network where much less mutual interference exists. This interpretation also demonstrates how desirable it is to mitigate the mutual interference in wireless communications, although it is very challenging.

2.2 Throughput Capacity of Ad Hoc Networks

2.2.1 Strategies to Improve Throughput Capacity

One natural question is if it is possible to improve throughput capacity of random networks by employing any advanced techniques or sophisticated strategies. After significant progress that has been made to further the investigation on throughput capacity scaling, the answer is positive.

First of all, by allowing both long-distance and short-distance transmissions, the throughput capacity can be improved slightly to $\Theta(\frac{1}{\sqrt{N}})$ [32]. The scheme constructed to achieve this throughput relies on multi-hop transmission, pairwise coding and decoding at each hop, and a time-division multiple access. The gain of throughput capacity can also be achieved by employing directional antennas. Yi et al. in [33] considered different beamform patterns, and showed that the throughput capacity can be achieved with a gain of $\frac{4\pi^2}{\alpha\beta}$ using directional transmission and reception, where α and β are antenna parameters. Peraki et al. in [34] further revealed that the maximum capacity gain is $\Theta((\log N)^2)$ by using directional antennas at the transmitters and receivers, corresponding to a throughput of $\Theta((\log N)^{3/2}/\sqrt{N})$. If nodes have multi-packet reception (MPR) capabilities, i.e., a receiver is capable of correctly decoding multiple packets transmitted concurrently from different transmitters, the capacity gain can also be achieved. Sadjadpour et al. in [35] showed that with MPR, the throughput capacity of random ad hoc networks can be improved at least by an order of $\Theta(\log N)$ and $\Theta((\log N)^{\frac{\alpha-2}{2\alpha}})$ under Protocol Model and Physical Model, respectively, where α is the path loss exponent in the Physical Model. Similar research efforts applying MPR can be found in [36–38].

By means of long-range multiple-input multiple-output (MIMO) communications with local cooperations as proposed in [39], significant improvement of throughput capacity scaling in random networks is attainable [40], i.e., almost constant per-node throughput of $\Theta(n^{-\epsilon})$ on average is achievable, where $\epsilon > 0$ can be arbitrarily small. This yields an aggregate throughput $(N\lambda(N))$ of $\Theta(N^{1-\epsilon})$ for the whole network, indicating almost linear capacity scaling in N. $\epsilon = \Theta(\frac{1}{\sqrt{\log N}})$ was explicitly obtained later in [41] and [42]. However, the capacity gain is at the cost of increased system complexity due to the intelligent hierarchical cooperation among nodes. Regardless the complexity of the constructed strategy, the result in [40] is inspiring but still controversial. Franceschetti et al. in [43] claimed that a throughput higher than $O((\log N)^2/\sqrt{N})$ cannot be achieved because of degrees of freedom limitation which is a result of laws of physics. Artificial assumptions and models lead to the impossible linear capacity scaling in [40]. While using Maxwell's equations without any artificial assumptions, Lee et al. in [44] established the capacity scaling laws for the line-of-sight (LOS) environments, which show that a linear scaling of the aggregate throughput is indeed possible for static random networks. Thus, the conflict between [40] and [43] was resolved. It is worth noting that even if such physical limits in [43] do exist and sophisticated

strategies like the hierarchical cooperation cannot further improve the per-node throughput ($\Theta(1/\sqrt{N})$) in the scaling limit sense, these strategies generally could be considerably beneficial in networks of any finite size. An example is the physical-layer network coding. In [45], it was shown that although the physical-layer network coding scheme does not change the scaling law, it improves throughput performance of the network in the sense by enlarging the constant component of the scaling result. The similar studies applying the network coding schemes can be found in [46–49].

Since the above research works are based on the assumption that the network is bandwidth-constrained, i.e., each node is only capable of transmitting at W bits per second, it is interesting to consider a scenario where each node has power constraint but can utilize unlimited bandwidth. Hence, there have been a few research efforts which focus on the ultra-wideband (UWB) techniques. In [50], Negi and Rajeswaran showed that under the limiting case $W \to \infty$, the throughput capacity is lower bounded by $\Omega\left(P_0\sqrt{N^{\alpha-1}/(\log N)^{\alpha+1}}\right)$ and upper bounded by $O(P_0(\sqrt{N\log N})^{\alpha-1})$, where α is the path loss exponent and P_0 is the maximum transmission power. The gap between the upper bound and the lower bound was closed by Tang and Hua in [51]. They showed that the throughput capacity of a UWB power-constrained ad hoc network is given by $\Theta(P_0(\sqrt{N/\log N})^{\alpha-1})$. A better result was obtained in [52] that the throughput capacity scales as $\Theta(P_0 N^{(\alpha-1)/2})$.

Without leveraging aforementioned advanced techniques in the static random network, what if nodes move? The effect of nodal mobility on throughput capacity scaling was first investigated by Grossglauser and Tse in [53]. By applying an i.i.d mobility model (see Sect. 2.3) to each node, they have shown that the per-node throughput of the mobile ad hoc network could remain constant, i.e., $\Theta(1)$, by using a ***two-hop relaying*** scheme (see Fig. 2.2c) and allowing finite but arbitrary delay. This result provides an interesting implication that dramatic gains in network capacity are possible when mobility is considered so that the nodes can exploit mobile relays to carry packets to distant nodes. Compared with such *store-carry-and-forward* communication paradigm, in the absence of mobility, direct transmission (see Fig. 2.2a) between distant nodes causes too much interference, or equivalently requires a large spatial area, so that the number of concurrent transmissions are reduced; on the other hand, if the network only allows the communication between nearest neighbors (see Fig. 2.2b), most of the packets will be delivered through multiple hops, resulting in the decrease in throughput capacity as well. Inspired by the promising result in [53], extensive works have been done to investigate capacity scaling in mobile ad hoc networks. In [54], Diggavi et al. have shown that even one-dimensional mobility benefits capacity scaling. In [55], Syed Ali Jafa explored the capacity of high mobile ad hoc networks in the presence of channel uncertainty, and has shown that high mobility introduces rapid channel fluctuations and hence limits the capacity of wireless networks.

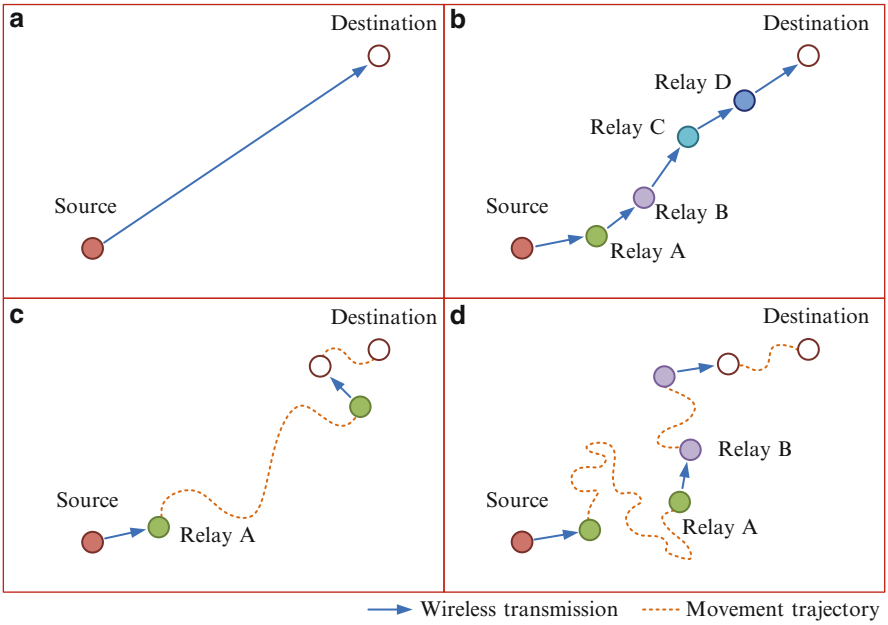


Fig. 2.2 An illustration of packet transmission strategies. (**a**) Direct transmission. (**b**) Multi-hop relying in a static ad hoc network. (**c**) Two-hop relaying in a mobile ad hoc network. (**d**) Multi-hop relying in a mobile ad hoc network

2.2.2 Other Factors Affecting Scaling Laws

The random network considered in [22] is a benchmark network model, in which nodes have basic communication capabilities (i.e., simple coding and decoding strategies implemented on the single radio), and the traffic model (symmetric unicast) and interference model (Protocol Model or Physical Model) are simplified. Besides the strategies mentioned in Sect. 2.2.1 to improve throughput capacity, significant research efforts have been made to study the impact of different modeling factors on capacity scaling laws.

Multi-channel multi-interface: In [22], it has been shown that with a single radio mounted on each vehicle, splitting the total bandwidth W into multiple sub-channels does not change the order of throughput capacity of random networks. However, in practice, a communication device may have multiple radio interface operating on one or different channels. What if each node is equipped with multiple radio interfaces? In [56] and [57], Kyasanur and Vaidya derived the capacity scaling laws for a general multi-channel networks with $m \leq c$, where c is the number of channels, and m is the number of interface per node. It was shown that different ratios between c and m yield different capacity bounds, and in general, the network

capacity is reduced except when c is upper bounded by $O(\log N)$. Kodialam and Nandagopal [58] also provided capacity trends for multi-channel multi-interface wireless mesh networks by considering channel assignment and scheduling.

Channel model: Most research efforts follow either Protocol Model (governed by geometry) or Physical Model (governed by path loss), which only characterize the deterministic behavior of wireless channel connection. To consider the channel randomness which is more realistic, several works have been done assuming different channel models. The impact of channel fading on capacity scaling was studied by Toumpis and Goldsmith [59]. They showed that a throughput of $\Theta(\frac{1}{\sqrt{N(\log N)^3}})$ is feasible under a general model of fading for static random networks. The Rayleigh fading was considered by Ebrahimi et al. [60] and the lower and upper bounds of throughput capacity were derived. The random connection model was considered in [61] and [62], i.e., the signal strengths of the connections between nodes are independent from each other and follow a common distribution. In [63], Gowaikar and Hassibi considered a *hybrid* channel connection model: for a short distance between transmitter and receiver, the channel strengths are governed by the random connection model; while for a long range, the channel strengths are governed by a Rayleigh distribution. They showed that a throughput of $\Theta\left(\frac{1}{(\log N)^4}\right)$ is achievable. The lower bound on the capacity of wireless erasure networks was reported by Jaber and Andrews [64], in which an erasure channel model is considered, i.e., each channel is associated with an erasure probability. Such a channel model incorporates erasure events which may correspond to packet drops or temporary outages when transmission is undergoing.

Network topology: The shape of geographic area where the network is deployed has a significant impact on capacity scaling laws. Hu et al. [65] investigated the effect of various geometries, including the strip, triangle, and three-dimensional cube. The main implication from [65] is that the symmetry of the network shape plays an important role. In other words, a high throughput capacity can be achieved if the network is symmetric. In [66], Li et al. respectively derived the capacity bound for the three-dimensional network with regularly and randomly deployed nodes.

Traffic pattern: Besides symmetric unicast, i.e., each node is only the source of one unicast flow and the destination of another, dissemination of information in other fashions has been extensively studied in the literature. The broadcast capacity is reported in [67–69], which is the maximum per-node throughput of successfully delivered broadcast packets. For each broadcast packet, it is successfully delivered if all nodes in the network other than the source receive the packet correctly in a finite time. The multicast capacity has been widely investigated [70–76] considering different network settings. By employing multicast, each packet is disseminated to a subset of $N-1$ nodes which are interested in the common information from the source. Nie [77] reported a short survey on multicast capacity scaling. A unifying study was provided by Wang et al. [78], in which how information is disseminated is generally modeled by the (n, m, k)-casting. In this particular context, n denotes

the population of nodes, and m and k denote the number of intended recipients of a source packet and the number of successful recipients, respectively. For unicast, $m = k = 1$; for multicast, $k \leq m < n$; and for broadcast, $k \leq m = n - 1$. The capacity bounds were established in [78] for each type of traffic pattern.

2.3 Fundamental Tradeoffs: Capacity, Delay, and Mobility

For network performance, capacity is not the only metric. From applications point of view, network delay (its average, maximum, or distribution) is also an important design aspect [23]. In [53], it has been shown that striking performance gains in throughput capacity are achievable in mobile ad hoc networks, however at the expense of enlarged delay. With the same time scale of node mobility, the delay is incurred by the ***movements*** of the relay (transmitter) and the destination (receiver) since they have to be geographically close enough for transmission, as shown in Fig. 2.2c. Basically, there are two ways to transfer a piece of information from the source to the destination: wireless transmission and node movement. Since wireless transmission is typically at a much smaller time scale, the time spent on the relay movements towards the destination contributes to the major part of the delay. There is a tradeoff between capacity and delay: if an increase in throughput is desired, we should reduce the distance of wireless transmission to allow more concurrent transmissions in the network; while if a decrease in delay is desired, we should reduce the distance of relay movement towards the destination. However, it is impossible to reduce both distances simultaneously given a fixed distance between the source and the destination. Furthermore, intuitively, different mobility models may incur different delays, because the node movement pattern determines the time spent on the relay movements. For example, if a node always wanders around (see Relay A in Fig. 2.2d), it is very difficult for the node to move a long distance in one direction. Therefore, quite straightforward research studies have been done to understand the tradeoffs between capacity and delay of wireless networks under a variety of mobility models.

2.3.1 Mobility Models

The type of node mobility studied includes the i.i.d mobility, random walk model, random way-point model, Brownian motion, and Lévy mobility. Besides, there are two more general mobility models defined in [79] to study the relationship between delay and throughput capacity from a global perspective.

- i.i.d Mobility Model: In time-slotted system, at each time slot, each node selects a new position independently and identically distributed over all positions in the network. The position distributions of the nodes are independent between time

slots. The i.i.d mobility is also referred to as the reshuffling model [80]. Depicting an extreme mobility, the i.i.d mobility model is unrealistic but analytically tractable.

- Random Walk Model: Random walk can be described by Markovian dynamics from i.i.d mobility and is often considered symmetric, i.e., nodes select new positions for next time slot equally likely from the set of current neighboring positions.
- Random Way-Point Model: In random way-point model, at each time slot, the mobile node chooses a random destination in the network and moves toward it at a random speed. The node pauses for some random time after reaching the destination, and then repeats this process.
- Brownian Motion: Brownian motion is like the motion conducted by a small particle totally immersed in a liquid or gas. Brownian mobility has a strong connection with random walk model and is a limiting case when taking smaller and smaller steps in smaller and smaller time intervals in symmetric random walk [81].
- Lévy Walk and Lévy Flight: Lévy mobility is a special type of random walk in which the distribution of flights, i.e., step-lengths, is heavy tailed. In other words, the trajectory of Lévy mobility contains many short flights and an exponentially small number of long flights. The difference between Lévy walk and Lévy flight is that the former has constant flight speed and the latter has constant flight time [82]. It is reported that Lévy mobility has certain statistical similarity to human mobility and some animals' hunting patterns [83].
- Hybrid Random Walk Models: A family of hybrid random walk models is considered in [79] and characterized by a single parameter $\beta \in [0, \frac{1}{2}]$. The unit area of the network is divided into $N^{2\beta}$ equal-sized squares, each of which is further divided into $N^{1-2\beta}$ equal-sized sub-squares. At the beginning of each time slot, each node jumps from its current sub-square to a random sub-square of one uniformly selected neighboring square, as shown in Fig. 2.3a. It can be seen that the model turns to the i.i.d mobility model and the random walk model when $\beta = 0$ and $\beta = 1/2$, respectively.
- Discrete Random Direction Models: A family of discrete random direction models is also considered in [79] and characterized by a single parameter $\alpha \in [0, \frac{1}{2}]$. The unit area of the network is divided into $N^{2\alpha}$ squares with equal area. The movement of each node is of the following pattern: at the beginning of each time slot, the node jumps from its current square to a uniformly selected neighboring square; and during the time slot, the node moves from a start position to an end position at a certain velocity, as shown in Fig. 2.3b. The two positions are uniformly selected from all the positions in the square. It can be seen that the above mobility model turns to the random way-point model and the discrete Brownian motion when $\alpha = 0$ and $\alpha = 1/2$, respectively.

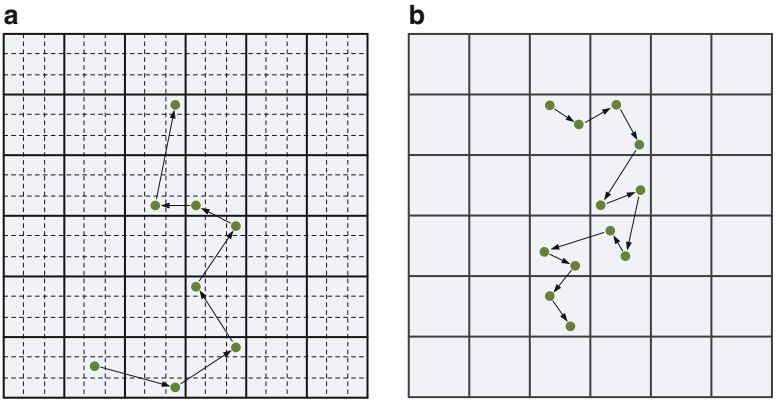


Fig. 2.3 An example of trajectories of hybrid random walk (**a**) and random direction model (**b**), respectively

2.3.2 *Tradeoffs Between Throughput Capacity and Delay*

The throughput capacity and delay under the i.i.d. mobility model were reported by Neely and Modiano [84] for a cell-partitioned ad hoc network. They found that a general delay-throughput tradeoff can be established: the ratio of delay and throughput is at least $O(N)$ under different scheduling policies (i.e., two-hop or multi-hop relaying) with or without packet redundancy.[2] The optimality of delay-capacity tradeoffs under i.i.d. mobility model was studied in [85]. Different time scales of nodal mobility are taken into consideration: fast mobility, only allowing one-hop transmissions during a time slot after which the topology changes; and slow mobility, allowing packets to be delivered through multiple hops during a time slot since the mobility of nodes is much slower than packet delivery time. It was shown that under i.i.d. fast mobility, a per-node capacity is $O(\sqrt{\mathscr{D}/N})$ given a delay constraint $\mathscr{D}$; while a per-node capacity is $O(\sqrt[3]{\mathscr{D}/N})$ under i.i.d. slow mobility, which is a tighter bound than $O(\sqrt[3]{\mathscr{D}/N}\log N)$ obtained in [86].

In [87], El Gamal et al. studied the throughput and delay under random walk model. It was shown that the ratio of delay and throughput is $\Theta(N)$ for throughput of $O(\frac{1}{\sqrt{N\log N}})$, while the delay remains $\Theta(N\log N)$ for almost any throughput of a higher order, indicating an unsmooth tradeoff under random walk model. Similar insights can be obtained for Brownian motion. In [88], Lin et al. first derived a lower bound of $\Omega(\log N/\sigma^2)$ for average delay associated with capacity of $\Theta(1)$ by using the two-hop relaying scheme proposed in [53], where σ^2 is related to the Brownian mobility model. More importantly, they demonstrated that it is impossible to reduce

[2]Redundancy of the packet means extra copies of the original packet, which are issued by the source node.

Table 2.1 Summary of capacity-delay tradeoffs for random ad hoc networks[a]

	Two-hop delay	Critical delay	Any tradeoff?
i.i.d Mobility	$\Theta(N \log N)$ [87]	$\Theta(1)$ [86]	Yes
Random walk	$\Theta(N \log N)$ [87]	$\Theta(N \log N)$ [87]	No
Random way-point	$\Theta(N)$[b][89]	$\approx \Theta(\sqrt{N})$ [79]	Yes
Brownian motion	$\approx \Theta(N)$ [79]	$\Theta(N)$ [82]	No
Discrete random direction (α)	$\Theta(N)$ [79]	$\approx \Theta(N^{\alpha+0.5})$ [79]	Yes
Hybrid random walk (β)	$\Theta(N)$ [79]	$\Theta(N^{2\beta} \log\log N)$ [79]	Yes
Lévy walk (γ)	$\approx \Theta(N)$ [82]	$\Theta(N^{\frac{1}{2}})$ for $0 < \gamma < 1$;	
		$\Theta(N^{\frac{\gamma}{2}})$ for $1 \leq \gamma < 2$ [82]	Yes
Lévy flight (γ)	$\approx \Theta(N)$ [82]	$\Theta(N^{\frac{\gamma}{2}})$ [82]	Yes

[a] The result is for the case in which the velocity does not scale with the network size

[b] $\alpha \in (0, 0.5)$, $\beta \in (0, 0.5)$ and $\gamma \in (0, 2]$

a large amount of delay without dropping the throughput to $O(\frac{1}{\sqrt{N}})$. From [87] and [88], significant increase in delay cannot be circumvented if a larger throughput than $\Theta(\frac{1}{\sqrt{N}})$[3] is desired by using random walk mobility or Brownian motion. Without showing any tradeoff, Sharma and Mazumda [89] analyzed the average delay of the two-hop relaying scheme in a network of N nodes following random way-point mobility.

To further investigate the impact of nodal mobility on throughput capacity and delay, Sharma et al. [79] proposed two general classes of mobility models, i.e., hybrid random walk models and discrete random direction models, incorporating mobility models aforementioned in [84, 87, 88]. The objective of this systematical study is to inquiry how much delay the network has to bear to achieve a per-node capacity better than $\Theta(\frac{1}{\sqrt{N}})$ under different mobility models, resulting in the notion of *critical delay*. Considering that the worst performance in network delay is incurred by the two-hop relaying scheme (*two-hop delay*), however, with an optimal throughput, the room left for tradeoff is actually determined by these two important delays. In [79], it was shown that tradeoffs are negligible under random walk model and Brownian motion, as also shown in [87] and [88], respectively; However, the tradeoff between delay and capacity is quite smooth under i.i.d. mobility and random way-point model. In [82], Lee et al. studied the delay-capacity tradeoffs under Lévy mobility. By using the limiting features of the joint spatio-temporal probability density functions of Lévy models, they derived the critical delay under Lévy walk and Lévy fight, respectively. It was shown that smooth tradeoffs can be obtained and are determined by the distribution parameter related to Lévy mobility. A summary of delay-capacity tradeoffs for random ad hoc networks is given in Table 2.1. Figure 2.4 also shows delay-capacity tradeoff regions under different mobility models.

[3] The throughput is achievable in static random ad hoc networks.

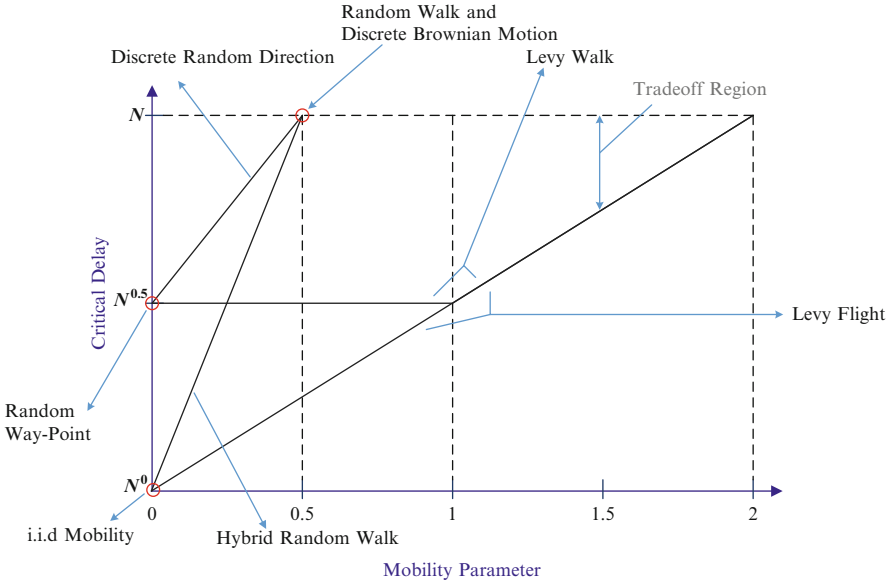


Fig. 2.4 Tradeoff regions for a particular mobility parameter under different mobility models

2.3.3 Impact of Restricted and Correlated Mobility

The mobility models considered in aforementioned delay-capacity studies rely on the following assumptions: (i) the mobility pattern of each node is identical; (ii) following certain ergodic mobility process, each node can visit the entire network area equally likely; and (iii) the movements of different nodes are independent. There have been several efforts made by follow-up investigations to relax these assumptions and then find the impact of restricted and correlated mobility on delay and throughput performance in ad hoc networks.

Restricted Mobility: By noticing that nodes often spend most of the time in proximity of a few preferred places within a localized area, some researchers have studied the throughput and delay under the restricted node mobility, which is more realistic to characterize mobility traces of humans, animals, and vehicles. Li et al. [90] investigated the impact of a restricted mobility model on throughput and delay of a cell-partitioned network. They found that smooth throughput-delay tradeoffs in mobile ad hoc networks can be obtained by controlling the mobility pattern of nodes. Unlike the network in [90] showing homogeneous node density, Garetto et al. have done a series of research [91–95] on the network with heterogeneous node density under restricted mobility model. The capacity scaling of a class of mobile ad hoc networks which show spatial inhomogeneities by considering a cluster mobility model was analyzed in [93] and [94]. In [95], Garetto and Leonardi demonstrated that the delay-throughput tradeoffs can be improved by restricting the

node mobility. They considered a restricted mobility that the node moves around a fixed home-point according to a Markov process, and the stationary distribution of the node location decays as a power law of exponent δ with the distance from the home-point. They showed that it is possible to exploit node heterogeneity under a restricted mobility model to achieve $\Theta(1/\log^2(N))$ throughput capacity and $O(\log^4(N))$ delay by using a sophisticated bisection routing scheme.

Correlated Mobility: Instead of exploring the full range of possible capacity-delay trade-offs, Ciullo et al. [80] studied the impact of correlated mobility on performance of delay and throughput capacity. They considered a mobility model in which nodes in the network are grouped and each group, occupying a disc area, moves following i.i.d mobility. Although each node visits uniformly the entire network, movements of different nodes belonging to the same group are not independent. It was shown that the correlated mobility pattern has significant impact on asymptotic network performance and it is possible to achieve better delay and throughput performance than that shown in [85].

2.3.4 Capacity and Delay Scaling without Exploiting Mobility

In [87], El Gamal et al. established delay-capacity tradeoffs for static ad hoc networks. It was shown that the tradeoff when applying multi-hop schemes is given by $\mathscr{D} = \Theta(N\lambda)$, where λ and $\mathscr{D}$ are respectively the throughput and delay. Following [40], throughput and delay tradeoff by means of hierarchical cooperation has been studied in [96], showing that $\mathscr{D} = N\log^2(N)\lambda$ for λ between $\Theta(\frac{1}{\sqrt{N}\log N})$ and $\Theta(\frac{1}{\log N})$. To serve delay sensitive traffic, Comaniciu and Poor [97] reported the delay-constrained capacity scaling of mobile ad hoc networks. Without taking advantage of mobility, they exploited multiuser detection among other signal processing techniques to enhance user capacity.

2.4 Infrastructure Matters: Capacity of Hybrid Wireless Networks

Unlike pure ad hoc networks of homogeneous nodes operating in the same manner, hybrid wireless networks consist of at least two types of nodes functioning differently. After [22], significant efforts have been made to investigate capacity and delay scaling considering node heterogeneity, i.e., for hybrid networks, including wireless ad hoc networks with infrastructure aiding nodes, ad hoc networks with wireless helping nodes, multihop acess networks, and cognitive radio networks, among others.

Table 2.2 Scaling regimes shown in [101]

Regime	Number of infrastructure nodes	Per-node throughput capacity
(i)	$M \lesssim \sqrt{N/\log N}$	$\Theta(1/\sqrt{N \log N})$
(ii)	$\sqrt{N/\log N} \lesssim M \lesssim N/\log N$	$\Theta(M/N)$
(iii)	$M \gtrsim N/\log N$	$\Theta(1/\log N)$

2.4.1 Ad Hoc Networks with Supportive Infrastructure

It has been shown that adding wired infrastructure nodes, such as base stations, to ad hoc networks can render significant benefits in terms of both throughput capacity and delay. In the context of related investigations, the fixed infrastructure supports the underlying ad hoc networks by relaying their packets, rather than access points to the Internet. The advantage of infrastructure nodes is to overcome geographic limitations since the packet can be relayed over a long distance through high-bandwidth wired links, as a complement of local ad hoc delivery.

Liu et al. [98] initiated the study on capacity scaling of hybrid wireless networks. By placing N stationary nodes and M base stations in the network, they found that the throughput capacity increases linearly with M if $M = \omega(\sqrt{N})$, otherwise the improvement is negligible. Different from the hexagonal cell structure of base station in [98], access points in [99] are randomly distributed in the network and the results show that it is possible to achieve a throughput of $\Theta(\frac{1}{\log N})$ under the condition that the number of ad hoc nodes associated with each access point is upper bounded. In [100], Toumpis derived capacity bounds of hybrid wireless networks assuming randomly located access points and a general fading channel model and reported very similar results to those in [98]. In [101], Zemlianov et al. provided upper bounds of per-node throughput capacity for the network of randomly distributed ad hoc nodes and base stations placed in any deterministic fashion. By allowing power control of base stations, they determined three scaling regimes as shown in Table 2.2. It can be seen that there is no need to deploy any infrastructure for regime (i), since the throughput is achievable by only leveraging ad hoc communications; and for regime (iii), adding more infrastructure nodes does not make any improvement in throughput, at least in the order sense.

By noting that previous studies usually consider a two-dimensional square or disk network area, Liu et al. [102] investigated the impact of network geometry on capacity scaling by exploring one-dimensional networks and two-dimensional strip networks with regularly placed base stations. The main implications of theirs results (shown in Table 2.3) are: (i) for the one-dimensional network, even a small number of supportive base stations can significantly increase the per-node throughput capacity; and (ii) for a two-dimensional strip network, depending on the width of the strip, the behavior of capacity scaling is the same as that of either the one-dimensional network or the two-dimensional square network. The upper bound of average packet delay for each type of network was also derived, as shown in Table 2.3. Impacts of both network topology and traffic pattern were considered

Table 2.3 Impact of network geometry [102]

Network geometry	Number of base stations	Throughput capacity	Average delay
1-D network & 2-D strip with strip width of $o(\log N)$	$M \log M = O(N)$	$\Omega(M/N)$	$O(N/M \log N)$
	$M \log M = \omega(N)$	$\Omega(1/\log M)$	$O(N/M \log N)$
2-D square & 2-D strip with strip width of $\Omega(\log N)$	$M = O(\sqrt{N})$	$\Omega(1/\sqrt{N})$	$O(\sqrt{N})$
	$M = \omega(\sqrt{N})$	$\Omega(\min\{M/N, 1/\log M\})$	$O(\sqrt{N/M \log N})$

Table 2.4 Scaling regimes shown in [104]

Regime	Number of base stations	Throughput capacity	Average delay
(i)	$M = O(N/\log N)$	$\Omega(\sqrt{\frac{M}{N \log N}})$	$\Omega(\sqrt{\frac{N}{M \log N}})$
(ii)	$M = \Omega(N/\log N)$	$\Omega(M/N)$	O(1)

in [103]. Traffic patterns differ from each other in number of destination nodes in the network. The capacity scaling is determined by the number of base stations, the shape of network area, and the traffic pattern. Moreover, the impact of base station placement, i.e., regular or random placement, was also considered in [103].

An important implication of results shown in [98–103] is that capacity gain will be insignificant if the number of infrastructure nodes placed in a square or disk network area grows asymptotically slower than certain threshold. By pointing out that such "threshold" comes from the underutilization of the capability of base stations, Shila et al. [104] provided a better capacity and delay scaling, as shown in Table 2.4. The basic strategy they adopted is to deliver a packet to the nearest base station through multiple hops, in contrast to the one-hop transmission from the node to the associated base station assumed in previous studies, which results in a sublinear capacity scaling with the number of base stations.

Li et al. [105] revisited capacity and delay scaling in hybrid wireless networks by exploiting an *L-maximum-hop routing strategy*. Specifically, if the destination can be reached within L hops, packets from the source are delivered without reliance on any infrastructure node. More importantly, it was shown that without degrading throughput, network delay can be improved substantially, however, at the expense of built infrastructure. It is possible to achieve both constant throughput and delay in such type of networks. By using the L-maximum-hop routing strategy as well,

Zhang et al. [106] studied the throughput capacity for a network of N randomly distributed nodes, each of which is equipped with a directional antenna, and M regularly placed base stations. By analyzing the relationship between L, M and directional antenna beamwidth θ, they showed a "threshold" result on impacts of directional antenna, i.e., throughput gain can be achieved by implementing directional antenna only when the number of base stations grows slower than certain threshold. Multiantenna systems were also considered. In [107], Shin et al. investigated the capacity scaling in the network with supportive base stations, at each of which the number of antennas scales at arbitrary rates relative to N. It is beneficial to exploit the spatial dimension of infrastructure by deploying multiple antennas, which enable simultaneous uplinks, at each base station. The throughput capacity of mobile hybrid networks was reported in [108], in which the mobility model considered is similar to that in [95].

2.4.2 Ad Hoc Networks with Wireless Aiding Nodes

Deploying wired infrastructure to support ad hoc networks may incur a large amount of investment. Cost is always an important concern of building real-world communication networks. Moreover, under some emergency (e.g., earthquake) or extreme (e.g., underwater) circumstances, infrastructure is typically unavailable. Therefore, a potential substitute is to deploy a set of aiding nodes which are wirelessly connected and more powerful than normal nodes, as shown in Fig. 2.5. A natural question arises in the context: how much capacity gain can be achieved? To answer this question, Li et al. [109] studied the throughput capacity of ad hoc networks with the deployment of wireless helping nodes. Other specific network features considered in [109] are rectangular network area, both regular and random placement of helping nodes, and asymmetric traffic in which the number of destination nodes can scale at a lower rate than $\Theta(N)$, all of which have large impacts on throughput capacity. The main result of [109] illustrates that it is possible to achieve higher per-node throughput than that of pure ad hoc networks when the allocated bandwidth of helping nodes scales at a much higher rate than $\Theta(1)$. In [110], Zhou et al. provided another promising solution of wireless mesh structures. In such hierarchical wireless mesh network, mesh clients (normal nodes) are uniformly distributed, and mesh routers (aiding nodes) constitute a wireless mesh backbone, some of which can function as infrastructure gateways. Asymptotic throughput was derived and represented by the number of mesh clients, the number of mesh routers and the number of mesh gateways. Relying on only a small number of mesh gateways, it was shown that such mesh network can achieve the same throughput capacity as that of a hybrid infrastructure-based network, however, with a much lower cost.

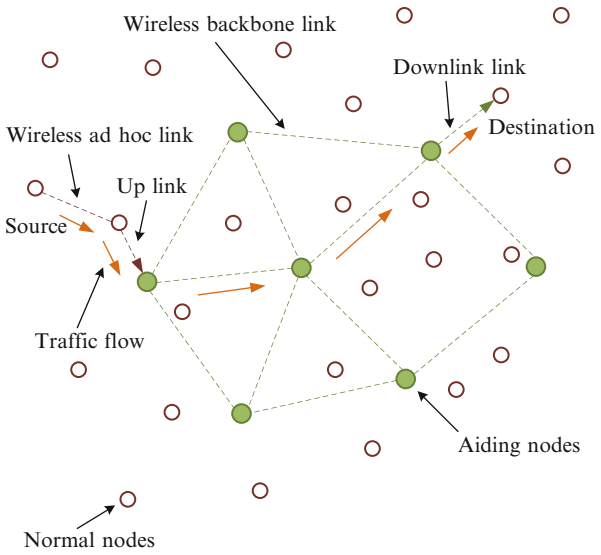


Fig. 2.5 Ad hoc network supported by wirelessly connected aiding nodes

2.4.3 Multihop Access Networks

Unlike ad hoc networks with supportive infrastructure nodes which do not generate or consume any data traffic, multihop access networks consist of infrastructure gateways bring/routing data traffic from/to the outside, such as Internet. Moreover, ad hoc transmissions between normal nodes are enabled and expected to enhance performance of such access networks, including capacity, coverage, connectivity, among others. To justify the benefit of augmenting access networks with multihop wireless links, Law et al. [111] investigated the downlink capacity of multihop cellular networks with regular placement of normal nodes and base stations. Due to poor spatial reuse, it was shown that one-dimensional multihop cellular networks yield almost no capacity gain compared to pure cellular networks. However, it is possible to significantly improve capacity of hexagonal hybrid network by exploiting multihop wireless links. By analyzing mathematically, they also found that capacity scaling in such type of networks mainly depends on the coverage of the base station, the transmission range of ad hoc links, and bandwidth allocation between different types of links. As a follow-up effort, Li et al. in [112] investigated capacity scaling for multihop cellular networks of randomly placed base stations and normal nodes distributed following a general inhomogeneous poisson process (IPP). In addition, throughput capacity was analyzed under different fairness constraints: (i) throughput-fairness, making throughput equal over all the nodes; and (ii) bandwidth-fairness, which guarantees that each node has equal allocated bandwidth. A "$\log_2 N$" result was shown in [112], i.e., multihop cellular networks with

regular placement of nodes and base stations achieve higher per-node throughput than pure cellular networks by a scaling factor of $\log_2 N$, regardless the underlying fairness constraint. For the network with heterogeneous node distribution, it is possible to obtain the "$\log_2 N$" result under certain conditions.

2.4.4 Cognitive Radio Networks

Nowadays, the demand on the frequency spectrum is increasingly difficult to meet due to scarce and underutilized spectrum resources. Cognitive radio is a paradigm created in an attempt to enhance spectrum utilization, by enabling unlicensed users to opportunistically utilize the spectrum bands owned by licensed users [113]. In cognitive radio networks, licensed users and unlicensed users are referred to as primary users (PUs) and secondary users (SUs), respectively. With overlapping primary and secondary networks operating simultaneously, behaviors of capacity and delay scaling of cognitive radio networks need to be investigated carefully.

By only allowing single-hop communication between a pair of SUs, Vu and Tarok [114] showed that the aggregate throughput of SUs can scale linearly with the number of SUs in the presence of a single or multiple pairs of primary transmitter (TX) and receiver (RX). In [115], Jeon et al. considered an ad hoc primary network of N randomly distributed PUs overlapped with an ad hoc secondary network of M randomly distributed SUs. Assuming M is much larger than N, they showed that an aggregate throughput of $\Theta(\sqrt{N})$ is achievable for the primary network, and in the meantime, the aggregate throughput of the secondary network is $\Theta(M^{\frac{1}{2}-\delta})$, for any arbitrarily small fraction of outage δ. The main implication of their result is that both two networks have almost the same capacity scaling as if each were a single network, given that one is much denser than the other. Another assumption made in [115] is that SUs know the locations of primary RXs. However, they are typically unavailable in practical scenarios. Instead, Yin et al. [116] studied capacity scaling of cognitive radio networks on the assumption that the locations of primary TXs are available to SUs and obtained very similar results to those in [115]. Huang and Wang [117] considered a more general model of cognitive radio networks, where the primary network can be different types, including classic static network, network with random walk mobility, and hybrid network, among others. Within this scope, they showed that the secondary network can attain the same asymptotic capacity and delay as standalone networks. The literature [118] is different from previous works in twofold. First, SUs are mobile and follow a specific heterogeneous speed-restricted mobility model. Second, cooperative communications are enabled so that SUs are allowed to relay packets for PUs. By exploiting the mobility heterogeneity of SUs, it was shown that almost constant capacity and delay scalings (except for poly-logarithmic factors) are possible in such kind of cognitive radio networks.

2.5 Summary

This chapter has surveyed the existing literature for scaling laws of throughput capacity for both ad hoc wireless networks and wireless networks with communication infrastructure. It has also presented a comprehensive overview of capacity-delay tradeoffs under a variety of mobility models. Extensive comparisons of existing results have been done to reach a better understanding.



http://www.springer.com/978-1-4614-8396-0

Capacity Analysis of Vehicular Communication Networks

Lu, N.; Shen, X.S.

2014, IX, 82 p. 26 illus., Softcover
ISBN: 978-1-4614-8396-0