Scaling Laws for Throughput Capacity and Delay in Wireless Networks – A Survey

Ning Lu, *Student Member, IEEE,* Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—The capacity scaling law of wireless networks has been considered as one of the most fundamental issues. In this survey, we aim at providing a comprehensive overview of the development in the area of scaling laws for throughput capacity and delay 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 and delay under a variety of mobility models. In addition, the capacity and delay 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. Finally, recent studies on scaling law for throughput capacity and delay in emerging vehicular networks are introduced.**

***Index Terms*—Fundamental limits, scaling laws, throughput capacity, delay, wireless networks.**

I. Introduction

WIRELESS networks have received a myriad of research attentions over the past decades, including medium access control, routing, security, cooperation, and energy-efficiency, among others. Despite significant advances in the field of wireless networking, a fundamental question remains unsolved: how much information can a wireless network transfer? To answer this question, we should resort to the study of network capacity which is a central concept in the field of network information theory [1]. Intuitively, if the capacity of a wireless network could be known, the network limit of information transfer would be obtained. Moreover, having such knowledge would shed light on what the appropriate architectures and protocols were for operating wireless networks. Although significant efforts have been put on the investigation of network capacity, developing a general theory of such a fundamental limit for wireless networks is a long standing open problem [2]. In [3], Claude Shannon successfully determined the maximum achievable rate, called the capacity, for a point-to-point communication channel, below which the reliable communication can be implemented while above which the reliable communication is impossible. However, general wireless networks with sources and destinations sharing channel resources are much more complex, making the quest for fundamental limits of wireless networks a formidable task. For example, even for a simple-looking three-node relay channel [4], the exact capacity still has yet to be determined.

As a retreat when exact fundamental limits are out of reach, capacity scaling laws, first investigated by Gupta and Kumar in [5], characterize the trend of node throughput behavior when the network size increases. The most salient feature of capacity scaling laws is to depict the capacity as a function of the number of nodes in the network, without distractions from minor details of network protocol. This approach is quite different from that of studying network information theory, which is to determine exact capacity region of wireless networks. The seminal work [5] not only provides an alternative and tractable way to study the network capacity, but also obtains insightful capacity results. Great efforts have been made thereafter to derive capacity scaling laws for different paradigms of wireless networks. Scaling laws for network delay and its tradeoff with the capacity have also been investigated.

The study of scaling laws can lead to a better understanding of intrinsic properties of wireless networks and theoretical guidance on network design and deployment [6]. Moreover, the results could also be applied to predict network performance, especially for the large-scale networks [7]. We provide the following illustration. We consider to deploy a large-scale sensor networks for a certain geographic area. Scaling laws show that the network scales poorly when the number of sensors grows, i.e., the throughput of each sensor would decrease. In order to enhance the throughput capacity, we may need to adopt some advanced technologies, such as directional antennas and network coding. However, scaling laws show that exploiting network coding cannot change the trend of throughput capacity; whereas exploiting directional antennas can introduce capacity gains (refer to Section III-A, Table I). Furthermore, suppose we have deployed a sensor network of 100 sensors with directional antennas. Typically we can obtain the throughput performance (denoted by λ_A) of the network through real measurement. If we need to extend the network to a larger one of 1000 sensors, with the same network settings, by capacity scaling results (denoted by $f(N)$), we are able to have a rough idea that how much throughput (denoted by λ_B) can be supported by the network that we will deploy, i.e., $\lambda_B = \lambda_A \cdot f(1000)/f(100)$.

This paper aims to provide a comprehensive survey of the state of the arts in the area of throughput capacity and delay scaling studies in wireless networks, which serves the following purposes.

- There has been a large body of research on capacity scaling laws. For new researchers in this area, confusion may rise since similar capacity bounds may be derived

Manuscript received February 7, 2012; revised June 14, 2013.
N. Lu and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada (e-mail: n7lu@uwaterloo.ca; sshen@uwaterloo.ca).
Digital Object Identifier 10.1109/SURV.2013.081313.00039

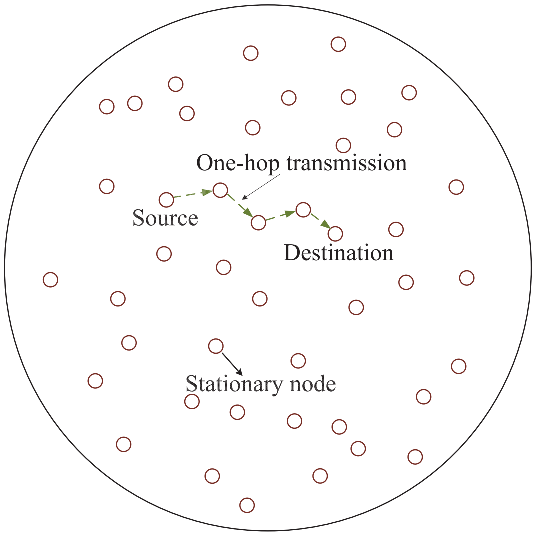


Fig. 1. A static ad hoc network in a unit disk.

for networks with different settings; while for the same network, different methodologies or techniques adopted in the study often yield different results. This paper is a modest attempt to summarize this field and provide rapid access to research results scattered over many papers.

- The research of scaling laws has undergone phenomenal growth in wireless communication and networking community. Since this research topic is also of practical significance, it should be accessible to general readers. We try to provide an overview of capacity and delay scaling laws in this regard. The premier is to show what the basic problem is and how different technologies and network settings affect scaling results, instead of demonstrating detailed theoretical derivations.

The remainder of this paper is organized as follows: Section II provides preliminaries of capacity scaling from Gupta and Kumar's ground-breaking work [5], including the notion of throughput capacity and random networks. Section III elaborates the advanced strategies to improve throughput capacity of ad hoc networks, and other factors that affect capacity scaling laws. Section IV presents the fundamental tradeoffs between throughput capacity and network delay for ad hoc networks under a variety of mobility models. Section V particularly surveys the capacity and delay for hybrid wireless networks. Section VI introduces the recent studies on capacity and delay scaling of emerging vehicular networks. Section VII discusses the future work and concludes the paper.

II. 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 [5], where Gupta and Kumar introduced two new notions of network capacity: *transport capacity* and *throughput capacity*. In this survey, 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.

A. 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 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 this 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 [8].

B. 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. 1 [5]. 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 a 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.

C. 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 [5]); while under Physical Model, the throughput capacity is given by $\Theta(\frac{W}{\sqrt{N\log N}}) \le \lambda(N) < \Theta(\frac{W}{\sqrt{N}})$ (see main result 4 in [5]). An explanation of the results is as follows. For Protocol Model, the lower bound and upper bound

[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 $f_2(n)$.

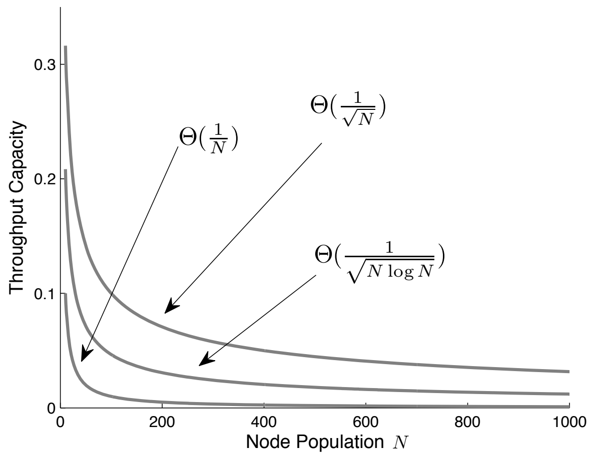


Fig. 2. Examples of showing throughput capacity trend in the order sense.

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. Fig. 2 gives three examples to show the trend of throughput capacity in the order sense.

The throughput capacity is studied asymptotically, i.e., capacity scaling law results hold with high probability when the population of nodes is larger than some threshold; on the other hand, results may not hold, or hold with small probability if the population of nodes is small. 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 such 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.

III. Throughput Capacity of Ad Hoc Networks

A. 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 $\Theta(\frac{1}{\sqrt{N}})$ [9]. 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 [10] 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. A capacity gain of $\Theta(\log N)$ is proved in [11]. Peraki *et al.* in [12] 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 [13] 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 [14]–[16].

By means of long-range multiple-input multiple-output (MIMO) communications with local cooperations as proposed in [17], significant improvement of throughput capacity scaling in random networks is attainable [18], 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 [19] and [20]. 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 [18] is inspiring but still controversial. Franceschetti *et al.* in [21] 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 [18]. While using Maxwell's equations without any artificial assumptions, Lee *et al.* in [22] 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 [18] and [21] is resolved. It is worth noting that even if such physical limits in [21] 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 [23], 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 [24]–[27].

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 [30], Negi and Rajeswaran showed that under the limiting case $W \to \infty$, the throughput capacity is lower bounded by $\Omega\big(P_0\sqrt{N^{\alpha-1}/(\log N)^{\alpha+1}}\big)$ and

TABLE I
SUMMARY: STRATEGIES TO IMPROVE THROUGHPUT CAPACITY

Strategies	Throughput Capacity Gain [Compared to $\Theta(\frac{1}{\sqrt{N\log N}})$]
Power Control	$\Theta((\log N)^{1/2})$ [9]
Directional Antenna	$\Theta((\log N)^{2})$ [12]
Multi-Packet Reception	$\Theta(\log N)$ for Protocal Model;
	$\Theta((\log N)^{\frac{\alpha-2}{2\alpha}})$ for Physical Model [13], $\alpha \geq 2$
Hierarchical Cooperation	Almost $\Theta((N\log N)^{1/2})$ [18]
Network Coding	Constant gain [23]–[27]
Ultra-wideband	$\Theta(N^{\alpha/2}(\log N)^{1/2})$ [28], $\alpha \geq 2$
Mobility	$\Theta((N\log N)^{1/2})$ [29]

* α is the path-loss component.

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 [31]. 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 [28] 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 node mobility on throughput capacity scaling was first investigated by Grossglauser and Tse in [29]. By applying an i.i.d mobility model (see Section IV) 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. 3(c)) 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. 3(a)) 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. 3(b)), 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 [29], extensive works have been done to investigate capacity scaling in mobile ad hoc networks. In [32], Diggavi *et al.* have shown that even one-dimensional mobility benefits capacity scaling. Restricted to move on a great circle, each node can attain a constant throughput. In [33], 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. A summary of capacity gains by using each strategy is given in Table I.

B. Other Factors Affecting Scaling Laws

The random network considered in [5] 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 Section III-A 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 [5], 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 [34] and [35], Kyasanur and Vaidya derived the capacity scaling laws for a general multi-channel networks with $l \leq c$, where c is the number of channels, and l is the number of interfaces per node. It was shown that different ratios between c and l 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 [36] 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 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 [37]. 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.* [38] for a single-hop scenario and the lower and upper bounds of throughput capacity were derived. The random connection model was considered in [39] and [40], i.e., the signal strengths of the connections between nodes are independent from each other and follow a common distribution. In [39], a throughput capacity of $\Omega((\log N)^{-d})$ is proved for some $d > 0$. In [40], Cui *et al.* showed that a constant throughput is achievable by relaxing some constraints of the connection model. In [41], 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.

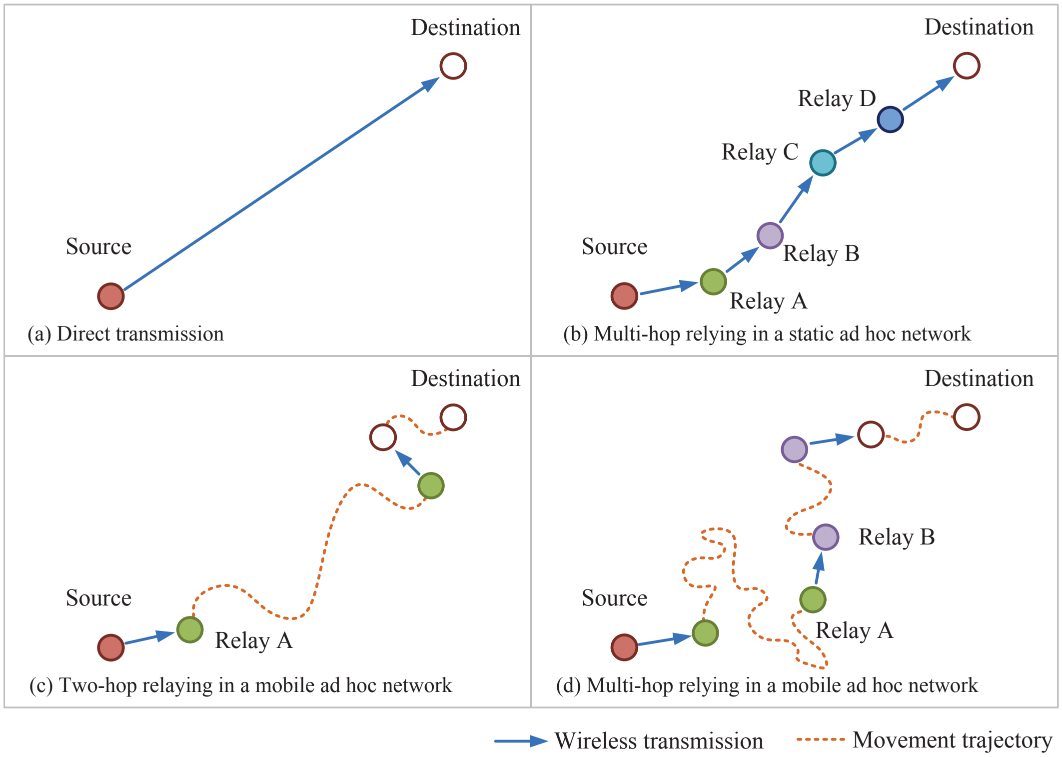


Fig. 3. An illustration of packet transmission strategies.

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 [42], 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. It is proved that the capacity lower bound scales as $\Theta(\sqrt{\log N/N})$ and $\Theta(\frac{1}{\sqrt{N \log N}})$ with independent and correlated erasure channels, respectively.

Network topology: The shape of geographic area where the network is deployed has a significant impact on capacity scaling laws. Hu *et al.* [43] investigated the effect of various geometries, including the strip, triangle, and three-dimensional cube. The main implication from [43] 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 addition to two-dimensional (2-D) networks, several efforts have been put on investigation of three-dimensional (3-D) networks. In [44], a throughput capacity of $\Theta\left(\frac{1}{N^{\frac{1}{3}}(\log N)^{\frac{2}{3}}}\right)$ and $\Theta(1/N^{\frac{1}{3}})$ is reported for 3-D random networks under Protocol Model and Physical Model, respectively. In [45], Li *et al.* respectively derived the capacity bound for the 3-D network with regularly and heterogeneously 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 [47]–[49], 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 [50]–[56] 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 [57] reported a short survey on multicast capacity scaling. A unifying study was provided by Wang *et al.* [46], in which how information is disseminated is generally modeled by the (N, m, k)-casting. In this particular context, 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 [46] for each type of traffic pattern. A summary for this subsection is given in Table II.

IV. 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 [6]. In [29], 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. 3(c). Basically, there are two ways to transfer an information packet from the source to the destination: wireless transmission and node movement.

TABLE II
SUMMARY: OTHER FACTORS AFFECTING SCALING LAWS

Factors	Main Results on Throughput Capacity (λ)
Multi-channel Multi-interface*	No capacity gain when $c/l = O(\log N)$;
	Capacity loss when $c/l = \Omega(\log N)$ [35]
Channel Model	$\lambda = \Theta(\frac{1}{\log N} \cdot \frac{1}{\sqrt{N \log N}})$ under general fading model [37]
	$\lambda = \Theta(\log N/N)$ under Rayleigh fading (single-hop scenario) [38]
	$\lambda = \Omega((\log N)^{-d})$ under random connection model for some $d > 0$ [39]
	$\lambda = \Theta(1)$ under random connection model for some cases [40]
	$\lambda = \Theta(1/(\log N)^4)$ under hybrid random connection model [41]
	$\lambda = \Omega(\sqrt{\log N/N})$ under independent erasure channel model [42]
	$\lambda = \Omega(\frac{1}{\sqrt{N \log N}})$ under correlated erasure channel model [42]
Network Topology	Symmetric topology yields a high throughput capacity [43]
	$\lambda = \Theta(\frac{1}{N^{\frac{1}{3}} (\log N)^{\frac{2}{3}}})$ under Protocol Model for 3-D networks [44]
	$\lambda = \Theta(1/N^{\frac{1}{3}})$ under Physical Model for 3-D networks [44]
Traffic Pattern**	A unified framework for (N, m, k)-casting [46]

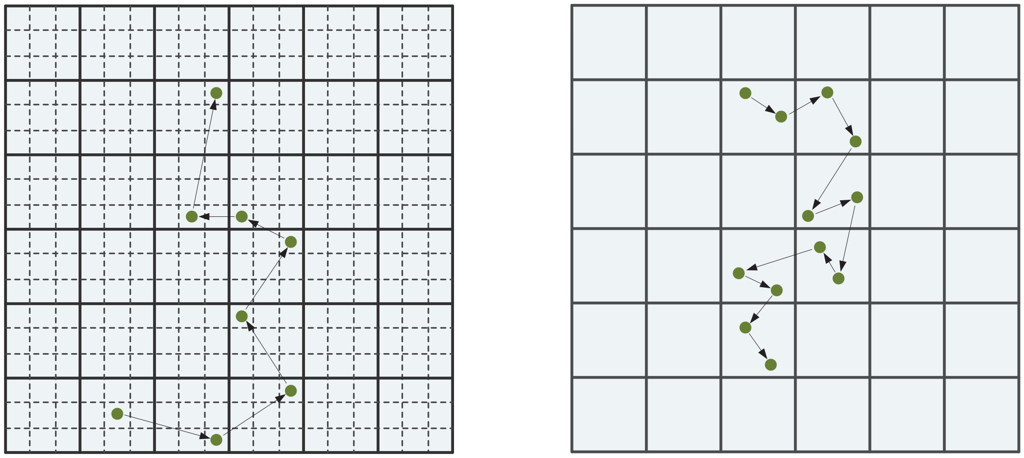
* c denotes the number of channels and l denotes the number of interfaces per node.
** m denotes the number of intended recipients of a source packet and k denotes the number of successful recipients.

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 component 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. 3(d)), it is very difficult for the node to move a long distance in one direction. To understand the tradeoffs between capacity and delay of wireless networks, a large body of research studies have been done under a variety of mobility models.

A. 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 [58] to study the relationship between delay and throughput capacity from a global perspective. We introduce these mobility models in the following and give a brief summary in Table III as well.

- 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 [59]. 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 [60].
- 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 [61]. It is reported that Lévy mobility has certain statistical similarity to human mobility and some animals' hunting patterns [62].
- Hybrid Random Walk Models: A family of hybrid random walk models is considered in [58] 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. 4(a). 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 [58]



(a) Hybrid random walk (b) Random direction model

Fig. 4. An example of trajectories of hybrid random walk (a) and random direction model (b), respectively.

TABLE III
SUMMARY OF MOBILITY MODELS

Mobility models	Key features
i.i.d Mobility	No motion constraints
Random walk	Next position is chosen from current neighboring positions
Randomly way-point	Randomly selected destination, speed, and pause duration
Brownian motion	The limit form of random walk
Lévy walk	Heavy-tailed distribution of flights; constant flight speed
Lévy flight	Heavy-tailed distribution of flights; constant flight time
Hybrid random walk	The hybrid of i.i.d mobility and random walk
Discrete random direction	The hybrid of randomly way-point and discrete Brownian motion

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. 4(b). 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.

B. Tradeoffs Between Throughput Capacity and Delay

The throughput capacity and delay under the i.i.d. mobility model were reported by Neely and Modiano [63] 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 [64]. Different time scales of node 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 $O(\sqrt{\mathcal{D}/N})$ given a delay constraint $\mathcal{D}$; while a per-node capacity is $O(\sqrt[3]{\mathcal{D}/N})$ under i.i.d. slow mobility, which is a tighter bound than $O(\sqrt[3]{\mathcal{D}/N}\log N)$ obtained in [65].

In [66], 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 [68], 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 [29], where σ^2 is related to the Brownian mobility model. More importantly, they demonstrated that it is impossible to reduce a large amount of delay without dropping the throughput to $O(\frac{1}{\sqrt{N}})$. From [66] and [68], 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 [67] analyzed the average delay of the two-hop relaying scheme in a network of N nodes following random way-point mobility.

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

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

TABLE IV
SUMMARY OF CAPACITY-DELAY TRADEOFFS FOR RANDOM AD HOC NETWORKS

	Two-hop delay	Critical delay	Any tradeoff?
i.i.d Mobility	$\Theta(N \log N)$ [66]	$\Theta(1)$ [65]	Yes
Random walk	$\Theta(N \log N)$ [66]	$\Theta(N \log N)$ [66]	No
Random way-point	$\Theta(N)$* [67]	$\approx \Theta(\sqrt{N})$ [58]	Yes
Brownian motion	$\approx \Theta(N)$ [58]	$\Theta(N)$ [61]	No
Discrete random direction (α)	$\Theta(N)$ [58]	$\approx \Theta(N^{\alpha+0.5})$ [58]	Yes
Hybrid random walk (β)	$\Theta(N)$ [58]	$\Theta(N^{2\beta} \log \log N)$ [58]	Yes
Lévy walk (γ)	$\approx \Theta(N)$ [61]	$\Theta(N^{\frac{1}{2}})$ for $0 < \gamma < 1$; $\Theta(N^{\frac{\gamma}{2}})$ for $1 \leq \gamma < 2$ [61]	Yes
Lévy flight (γ)	$\approx \Theta(N)$ [61]	$\Theta(N^{\frac{\gamma}{2}})$ [61]	Yes

* The result is for the case in which the velocity does not scale with the network size.
** $\alpha \in (0, 0.5)$, $\beta \in (0, 0.5)$ and $\gamma \in (0, 2]$.

To further investigate the impact of node mobility on throughput capacity and delay, Sharma *et al.* [58] proposed two general classes of mobility models, i.e., hybrid random walk models and discrete random direction models, incorporating mobility models aforementioned in [63], [66], [68]. 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 [58], it was shown that tradeoffs are negligible under random walk model and Brownian motion, as also shown in [66] and [68], respectively; However, the tradeoff between delay and capacity is quite smooth under i.i.d. mobility and random way-point model. In [61], 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 IV. Fig. 5 also shows delay-capacity tradeoff regions under different mobility models.

C. 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.* [69] 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 [70] showing homogeneous node density, Garetto *et al.* have done a series of research [71]–[75] 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 [73] and [74]. In [75], 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.* [59] 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 a significant impact on asymptotic network performance and it is possible to achieve better delay and throughput performance than that shown in [64].

D. Delay and Capacity Scaling without Exploiting Mobility

In [66], 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 $\mathcal{D} = \Theta(N\lambda)$, where λ and $\mathcal{D}$ are respectively the throughput and delay. Following [18], throughput and delay tradeoff by means of hierarchical cooperation has been studied in [76], showing that $\mathcal{D} = N\log^2(N)\lambda$ for λ between $\Theta(\frac{1}{\sqrt{N}\log N})$ and $\Theta(\frac{1}{\log N})$.

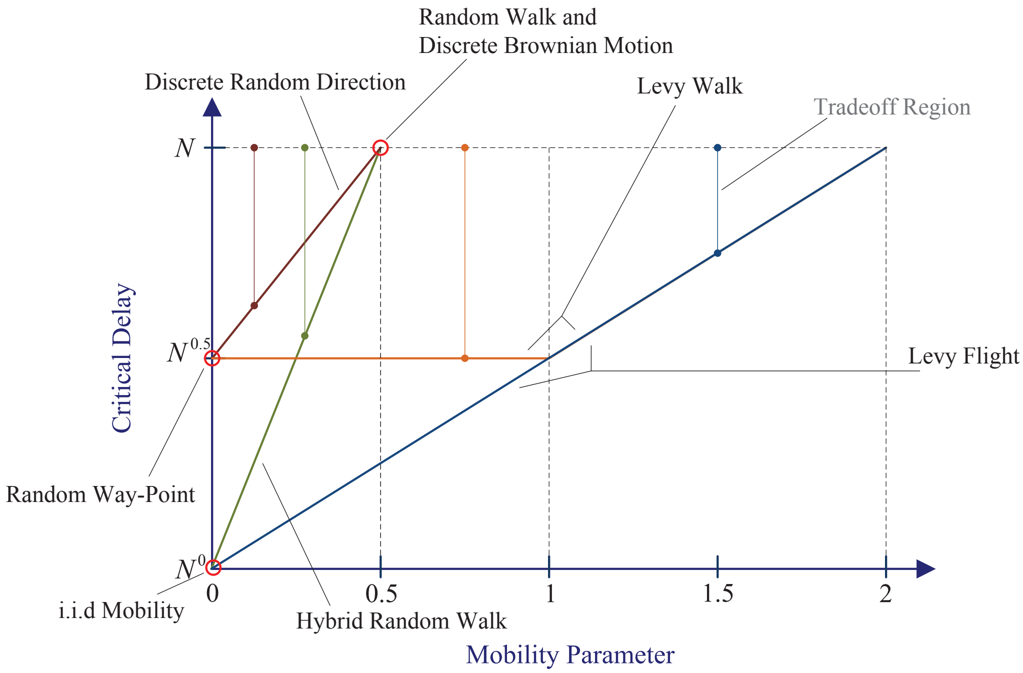


Fig. 5. Tradeoff regions for a particular mobility parameter under different mobility models.

To serve delay sensitive traffic, Comaniciu and Poor [77] 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.

V. Infrastructure Matters: Capacity and Delay 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 [5], 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.

A. 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.* [78] 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 [78], access points in [79] 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. Allowing power control, a constant throughput of $\Theta(1)$ is reported in [80]. In [81], 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 [78]. In [82], 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 V. 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.* [83] 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 VI) 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 VI. Impacts of

TABLE V
SCALING REGIMES SHOWN IN [82]

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)	$M \gtrsim N/\log N$	$\Theta(1/\log N)$

TABLE VI
IMPACT OF NETWORK GEOMETRY [83]

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

both network topology and traffic pattern were considered in [84]. 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 [84].

An important implication of results shown in [78], [79], [81]–[84] 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 a "threshold" comes from the underutilization of the capability of base stations, Shila *et al.* [85] provided a better capacity and delay scaling, as shown in Table VII. 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 yields a sublinear capacity scaling with the number of base stations.

Li *et al.* [86] 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 relying 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 this type of networks. By using the L-maximum-hop routing strategy as well, Zhang *et al.* [87] 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 [88], Shin *et al.* investigated the capacity scaling

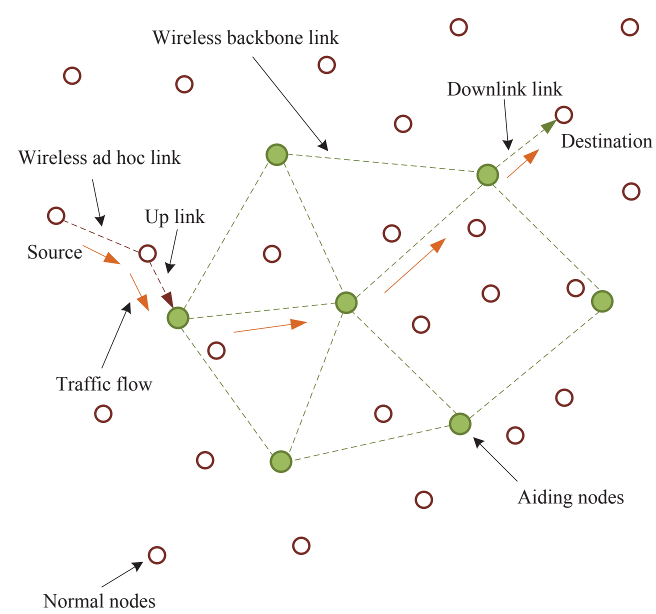


Fig. 6. Ad hoc network supported by wirelessly connected aiding nodes

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. Wang *et al.* in [89] considered the impact of fading impairments when operating hybrid wireless networks where base stations are deployed to support long-range communications between ad hoc nodes. The throughput capacity of mobile hybrid networks was reported in [90], in which the mobility model considered is similar to that in [75].

B. Ad Hoc Networks with Wireless Aiding Nodes

Deploying wired infrastructure to support ad hoc networks may incur a prohibitive cost which is always an important con-

TABLE VII
SCALING REGIMES SHOWN IN [85]

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)

cern 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. 6. A natural question arises in the context: how much capacity gain can be achieved? To answer this question, Li *et al.* [91] studied the throughput capacity of ad hoc networks with the deployment of wireless helping nodes. Other specific network features considered in [91] 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 [91] 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 [92], Zhou *et al.* provided another promising solution of wireless mesh structures. In such a 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 a mesh network can achieve the same throughput capacity as that of a hybrid infrastructure-based network, however, with a much lower cost. Literature [93] investigates a special scenario in which there exists only one active source-destination communication pair, and all remaining nodes act as aiding nodes. A constant capacity scaling is proved for that particular case.

C. 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, and connectivity. To justify the benefit of augmenting access networks with multihop wireless links, Law *et al.* [94] 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 this 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 [95] 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 [95], 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.

D. 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 [96]. 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, capacity and delay scaling laws of cognitive radio networks need to be investigated carefully.

By only allowing single-hop communication between a pair of SUs, Vu and Tarok [97] 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 [98], 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 [98] is that SUs know the locations of primary RXs. However, such prior knowledge is typically unavailable in practical scenarios. Instead, Yin *et al.* [99] 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 [98]. Huang and Wang [100] 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 [101] 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 a kind of cognitive radio networks.

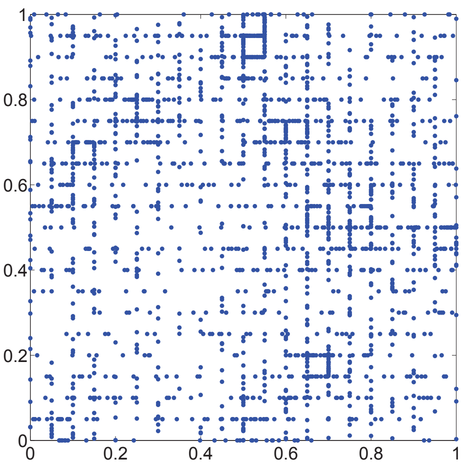


Fig. 7. A distribution of vehicles in a city grid of unit area following socialized mobility model.

VI. Capacity and Delay Scaling for Vehicular Networks

Due to growing urbanization and environmental pressures, improving efficiency and safety of road transportation has been increasingly pressing to alleviate transportation problems, including traffic accidents, congestions, and air pollution, among others, especially in the developing world. There has been increasing interest and significant progress in the domain of emerging VehiculAr NETworks (VANETs)[4], which target to incorporate wireless communications and informatics technologies into the road transportation system, enabling the evolution to next generation Intelligent Transportation Systems (ITS).

The capacity scaling laws of VANETs are desirable since unlike generic mobile ad hoc networks, VANETs present unique characteristics, which impose distinguished challenges on networking. i) *Large scale*: the VANET is a large-scale mobile network, which is deployed in a wide geographic area with a vast amount of vehicles and roadside infrastructures; ii) *Cars on the road:* the movement of vehicles should follow certain street pattern, different from generic mobile ad hoc networks in which nodes typically move in a free space; iii) *Cars on wheels:* the vehicle mobility is related to the social life of the driver; iv) *Spatio-temporal dynamics:* there are spatio-temporal variations of vehicle density and link quality due to vehicle mobility and unstable wireless channels, respectively; and v) *Diversified applications:* VANET applications are of a large variety and with different quality of service (QoS) requirements. All these features dramatically complicate scaling laws studies.

There have been a few efforts to investigate the capacity of VANETs. Pishro-Ni *et al.* [102] initiated the study of capacity scaling for vehicular networks with an emphasis on the impact of road geometry on the network capacity. Nekoui *et al.* [103] specially developed a novel notion of capacity for safety applications, which is called *Distance-Limited Capacity*. That is the capacity of VANETs when a pair of vehicles can only communicate if the two vehicles reside in a certain distance of each other. Both [102] and [103] showed that the road geometry has an important role in determining the capacity of vehicular networks. As the demand of public information dissemination is high in vehicular networks, multicast flows, in which one source is associated with a set of destinations, may be viable to be deployed for practical applications. In [104], Zhang *et al.* analyzed multicast capacity of hybrid VANETs, in which base stations are deployed to support communications between vehicles. It was assumed that each vehicle is equipped with a directional antenna. By respectively applying one-dimensional and two-dimensional i.i.d. mobility model to vehicles, they derived bounds of the multicast throughput capacity under certain end-to-end delay constraint. In [105], Wang *et al.* studied the uplink capacity of hybrid VANETs, where each vehicle, following random way-point mobility, is required to send packets to regularly placed sink roadside units (RSUs). The basic routing strategy adopted in [105] is to distribute source packets to as many RSUs as possible to increase concurrent uploading opportunities.

One of the limitations of [104] and [105] is that the specific mobility features of vehicles are not fully considered. The i.i.d mobility is not practical for vehicular scenarios. Moreover, the assumption that vehicles are uniformly distributed in the network is also unrealistic. For urban areas, vehicle densities in different regions may be highly diverse. Therefore, we investigated the throughput capacity and average packet delay of social-proximity vehicular networks in [106], [107], considering inhomogeneous vehicle densities. Specifically, we modeled urban area as a scalable grid with equal length road segments and a set of social spots. Each vehicle has a restricted mobility region around a specific social spot, and transmits via a unicast flow to a destination vehicle which is associated with the same social spot. Moreover, the spatial distribution of the vehicle decays following a power-law distribution from the central social spot towards the border of the mobility region. Fig. 7 shows an example of vehicle densities in the network following the socialized mobility model aforementioned. In [107], it was shown that although the throughput and delay may degrade in high density areas, it is still possible to achieve almost constant per-vehicle throughput and constant delay.

[4]To deemphasize the ad hoc nature of vehicular networks, we redefine the term VANETs, which is traditionally the acronym of vehicular ad hoc networks.

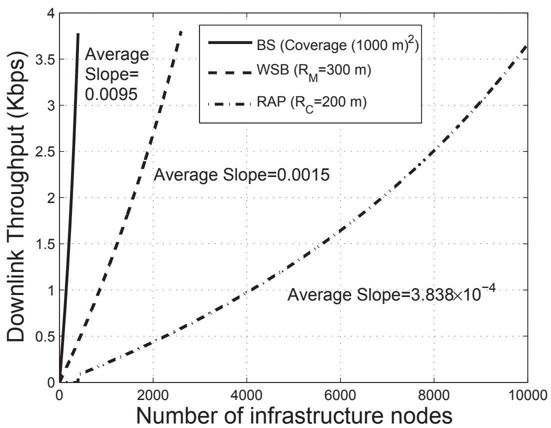


Fig. 8. Comparison of number of deployed infrastructure nodes in the hybrid mode (R_M and R_C are the coverage radii of MN and RAP, respectively).

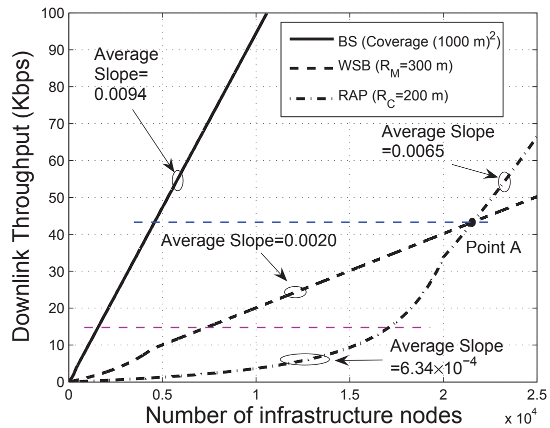


Fig. 9. Comparison of number of deployed infrastructure nodes in the infrastructure mode.

Wireless infrastructure, such as Wi-Fi access points and cellular base stations, plays a vital role in providing pervasive Internet access to vehicles. However, the deployment costs of different access infrastructure are highly variable. In [108], we made an effort to investigate the capacity-cost tradeoffs for vehicular access networks to better understand this issue. We first analyzed the downlink capacity of vehicles, i.e., the maximum average downlink rate achieved *uniformly* by vehicles from the access infrastructure. To provide pervasive Internet access, two operation modes of the network were considered: *infrastructure mode*, in which the model city is fully covered by infrastructure nodes, i.e., all the vehicles are within the coverage of the infrastructure, and hence only the infrastructure-to-vehicle (I2V) communication is used to deliver the downlink traffic; and *hybrid mode*, in which the model city is not fully covered and the downlink flow is relayed to the vehicles outside the coverage of infrastructure nodes by means of *multi-hop* vehicle-to-vehicle (V2V) communications. A lower bound of the downlink capacity was derived for the network with deployment of cellular base stations (BSs), wireless mesh backbones (WMBs) (a network of mesh nodes (MNs), including one mesh gateway), and roadside access points (RAPs), respectively. To examine the capacity-cost tradeoffs of different deployments, we presented a case study based on a perfect city grid of 400 km^2 with 0.4 million vehicles. It was shown that in the hybrid mode, to achieve the same downlink throughput, the network roughly needs X BSs, or $6X$ MNs, or $25X$ RAPs[5], as shown in Fig. 8; while in the infrastructure mode, if it is expected to improve the downlink throughput by the same amount for each deployment, we roughly need to additionally deploy X BSs, or $5X$ MNs, or $1.5X$ RAPs, as shown in Fig. 9. By explicitly taking capital expenditures and operational expenditures of access infrastructures into consideration, the deployment of BSs or WMBs is cost-effective to offer a low-speed downlink rate; nonetheless, when providing a high-speed Internet access, the deployment of RAPs outperforms the other two in terms of deployment costs. Such implications could provide a basis for the choice of access infrastructures for the automobile and telecommunication industry. Particularly, as automotive industry gears for supporting high-bandwidth applications, non-cellular access infrastructure will play an increasingly important role in offering a cost-effective data pipe for vehicles.

Despite recent studies on scaling laws of VANETs, there are still many open issues. For example, when jointly considering more complex street patten and inhomogeneous vehicle densities, it might be difficult to determine the throughput capacity and network delay. Moreover, due to the emergent and public nature of safety applications, broadcasting plays an important role in disseminating safety messages to vehicles in proximity. The study of broadcast capacity should be another interesting topic.

[5] X is used to represent a ratio relationship rather that a specific value.

VII. Concluding Remarks

We have surveyed the existing literature for scaling laws of throughput capacity in wireless networks. A comprehensive overview of capacity-delay tradeoffs under a variety of mobility models and scaling laws for hybrid wireless networks have also been presented. In addition, recent progress in throughput capacity of emerging vehicular networks has been introduced.

We close this survey with our thoughts on future research directions in this field.

- The design, analysis and deployment of wireless networks necessitate a general understanding of capacity scaling laws. Existing works often adopt different methodologies and sets of assumptions and models in developing capacity scaling laws, which may yield custom-designed solutions without universal properties that can be applied to other types of wireless networks. To better understand the impact of various settings and techniques on capacity scaling laws, it would be useful to provide a unified framework. Two research works have been performed toward this end: the study of capacity scaling laws under a generalized physical model [109] and the establishment of a simple set of criteria that can be used to determine the capacity for various physical layer technologies under the protocol model [110].

- The Shannon capacity was achieved by considering arbitrarily delay and vanishingly small error probability. In [2], Andrews *et al.* referred to a throughput-delay-reliability (TDR) triplet, since these quantities are interrelated. Thus, the throughput capacity of wireless networks would likely be constrained by these two fundamental quantities—delay and reliability jointly. Actually, the link reliability has been considered in studies of transmission capacity [111]–[113] which is the spatial intensity of attempted transmissions under a target outage of wireless links. The tradeoff between throughput capacity, delay, and reliability should be investigated, however this is much more challenging.
- Investigations on throughput capacity and network delay of emerging wireless networks are also promising. Particular characteristics of networks being studied often make the problem very challenging, such as road geometry and vehicle density in vehicular networks. In addition to the aforementioned cognitive radio networks and vehicular networks, femtocell networks [114] and smart grid have also gained much interest recently, both of which have complex network architecture and heterogenous communication devices, making the study of scaling laws a demanding task.

Acknowledgement

The authors wish to thank Xiugang Wu from University of Waterloo for helpful discussions. The authors are also grateful to the anonymous reviewers for their valuable comments.

References

[1] A. El Gamal and Y. Kim, *Network information theory*. Cambridge University Press, 2011.

[2] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar *et al.*, "Rethinking information theory for mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 12, pp. 94–101, 2008.

[3] C. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[4] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[5] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[6] A. Goldsmith, M. Effros, R. Koetter, M. Médard, A. Ozdaglar, and L. Zheng, "Beyond shannon: the quest for fundamental performance limits of wireless ad hoc networks," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 195–205, 2011.

[7] P. Li, M. Pan, and Y. Fang, "The capacity of three-dimensional wireless ad hoc networks," in *Proc. IEEE INFOCOM*, Shanghai, China, April 2011.

[8] T. Cover, J. Thomas, J. Wiley *et al.*, *Elements of information theory*. Wiley Online Library, 1991, vol. 306.

[9] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, 2007.

[10] S. Yi, Y. Pei, and S. Kalyanaraman, "On the capacity improvement of ad hoc wireless networks using directional antennas," in *Proc. ACM MobiHoc*, 2003.

[11] P. Li, C. Zhang, and Y. Fang, "The capacity of wireless ad hoc networks using directional antennas," *IEEE Trans. Mobile Computing*, vol. 10, no. 10, pp. 1374–1387, 2011.

[12] C. Peraki and S. Servetto, "On the maximum stable throughput problem in random networks with directional antennas," in *Proc. ACM MobiHoc*, 2003.

[13] H. Sadjadpour, Z. Wang, and J. Garcia-Luna-Aceves, "The capacity of wireless ad hoc networks with multi-packet reception," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 600–610, 2010.

[14] J. Garcia-Luna-Aceves, H. Sadjadpour, and Z. Wang, "Challenges: towards truly scalable ad hoc networks," *Proc. ACM MobiCom*, 2007.

[15] Z. Wang, H. Sadjadpour, and J. Garcia-Luna-Aceves, "The capacity and energy efficiency of wireless ad hoc networks with multi-packet reception," in *Proc. MobiHoc*. ACM, 2008, pp. 179–188.

[16] ——, "Fundamental limits of information dissemination in wireless ad hoc networks–part II: Multi-packet reception," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 803–813, 2011.

[17] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2044–2059, 2007.

[18] A. Ozgur, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, 2007.

[19] J. Ghaderi, L. Xie, and X. Shen, "Hierarchical cooperation in ad hoc networks: Optimal clustering and achievable throughput," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3425–3436, 2009.

[20] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3959–3982, 2009.

[21] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: information-theoretic and physical limits," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3413–3424, 2009.

[22] S. Lee and S. Chung, "Capacity scaling of wireless ad hoc networks: Shannon meets maxwell," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1702–1715, 2012.

[23] K. Lu, S. Fu, Y. Qian, and H. Chen, "On capacity of random wireless networks with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 763–772, 2009.

[24] J. Liu, D. Goeckel, and D. Towsley, "The throughput order of ad hoc networks employing network coding and broadcasting," in *IEEE Proc. MILCOM*, Washington, USA, October 2006.

[25] ——, "Bounds on the gain of network coding and broadcasting in wireless networks," in *IEEE Proc. INFOCOM*, Anchorage, USA, May 2007.

[26] A. Keshavarz-Haddadt and R. Riedi, "Bounds on the benefit of network coding: Throughput and energy saving in wireless networks," in *IEEE Proc. INFOCOM*, Phoenix, USA, March 2008.

[27] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the throughput gain of network coding in unicast and multicast wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 582–592, 2009.

[28] H. Zhang and J. Hou, "Capacity of wireless ad-hoc networks under ultra wide band with power constraint," in *IEEE Proc. INFOCOM*, Miami, USA, March 2005.

[29] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, 2002.

[30] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *IEEE Proc. INFOCOM*, Hong Kong, China, March 2004.

[31] X. Tang and Y. Hua, "Capacity of ultra-wideband power-constrained ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 916–920, 2008.

[32] S. Diggavi, M. Grossglauser, and D. Tse, "Even one-dimensional mobility increases the capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3947–3954, 2005.

[33] S. Jafar, "Too much mobility limits the capacity of wireless ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3954–3965, 2005.

[34] P. Kyasanur and N. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proc. ACM MobiCom*, 2005.

[35] ——, "Capacity of multichannel wireless networks under the protocol model," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 515–527, 2009.

[36] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom*, 2005.

[37] S. Toumpis and A. Goldsmith, "Large wireless networks under fading, mobility, and delay constraints," in *Proc. IEEE INFOCOM*, Hongkong, China, March 2004.

[38] M. Ebrahimi, M. Maddah-Ali, and A. Khandani, "Throughput scaling laws for wireless networks with fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4250–4254, 2007.

[39] R. Gowaikar, B. Hochwald, and B. Hassibi, "Communication over a wireless network with random connections," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2857–2871, 2006.

[40] S. Cui, A. Haimovich, O. Somekh, H. Poor, and S. Shamai, "Throughput scaling of wireless networks with random connections," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3793–3806, 2010.

[41] R. Gowaikar and B. Hassibi, "Achievable throughput in two-scale wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1169–1179, 2009.

[42] R. Jaber and J. Andrews, "A lower bound on the capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6502–6513, 2011.

[43] C. Hu, X. Wang, Z. Yang, J. Zhang, Y. Xu, and X. Gao, "A geometry study on the capacity of wireless networks via percolation," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2916–2925, 2010.

[44] P. Gupta, P. R. Kumar *et al.*, "Internets in the sky: The capacity of three dimensional wireless networks," *Communications in Information and Systems*, vol. 1, no. 1, pp. 33–49, 2001.

[45] P. Li, M. Pan, and Y. Fang, "Capacity bounds of three-dimensional wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 4, pp. 1304–1315, 2012.

[46] Z. Wang, H. Sadjadpour, and J. Garcia-Luna-Aceves, "A unifying perspective on the capacity of wireless ad hoc networks," in *Proc. IEEE INFOCOM*, Phoenix, USA, April 2008.

[47] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proc. ACM MobiCom*, 2006.

[48] R. Zheng, "Asymptotic bounds of information dissemination in power-constrained wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 251–259, 2008.

[49] X. Li, J. Zhao, Y. Wu, S. Tang, X. Xu, and X. Mao, "Broadcast capacity for wireless ad hoc networks," in *Proc. IEEE MASS*, Atlanta, USA, September 2008.

[50] S. Li, Y. Liu, and X. Li, "Capacity of large scale wireless networks under gaussian channel model," in *Proc. ACM MobiCom*, 2008.

[51] X. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 950–961, 2009.

[52] C. Wang, X. Li, C. Jiang, S. Tang, Y. Liu, and J. Zhao, "Scaling laws on multicast capacity of large scale wireless networks," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

[53] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, pp. 1691–1700, 2010.

[54] U. Niesen, P. Gupta, and D. Shah, "The balanced unicast and multicast capacity regions of large wireless networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2249–2271, 2010.

[55] X. Wang, W. Huang, S. Wang, J. Zhang, and C. Hu, "Delay and capacity tradeoff analysis for motioncast," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1354–1367, 2011.

[56] X. Wang, L. Fu, and C. Hu, "Multicast performance with hierarchical cooperation," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 917–930, 2012.

[57] D. Nie, "A survey on multicast capacity of wireless ad hoc networks," 2009. [Online]. Available: http://iwct.sjtu.edu.cn/personal/xwang8/research/nieding/survey.pdf

[58] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: A global perspective," *IEEE/ACM Trans. Netw.*, vol. 15, no. 5, pp. 981–992, 2007.

[59] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad-hoc networks," in *Proc. IEEE INFOCOM*, San Diego, USA, March 2010.

[60] S. Ross, *Introduction to probability models*. Academic press, 2009.

[61] K. Lee, Y. Kim, S. Chong, I. Rhee, and Y. Yi, "Delay-capacity tradeoffs for mobile networks with Lévy walks and Lévy flights," in *Proc. IEEE INFOCOM*, San Diego, USA, March 2010.

[62] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, 2011.

[63] M. Neely and E. Modiano, "Capacity and delay tradeoffs for ad hoc mobile networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1917–1937, 2005.

[64] L. Ying, S. Yang, and R. Srikant, "Optimal delay–throughput tradeoffs in mobile ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4119–4143, 2008.

[65] X. Lin and N. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," in *Proc. 3rd Annual Mediterranean Ad Hoc Networking Workshop*, Bodrum, Turkey, June 2004.

[66] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks-part I: The fluid model," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2568–2592, 2006.

[67] G. Sharma and R. Mazumdar, "Scaling laws for capacity and delay in wireless ad hoc networks with random mobility," in *Proc. IEEE ICC*, Paris, France, June 2004.

[68] X. Lin, G. Sharma, R. Mazumdar, and N. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with brownian mobility," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2777–2784, 2006.

[69] P. Li, Y. Fang, J. Li, and X. Huang, "Smooth trade-offs between throughput and delay in mobile ad hoc networks," *IEEE Trans. Mobile Computing*, vol. 11, no. 3, pp. 427–438, 2012.

[70] P. Li, Y. Fang, and J. Li, "Throughput, delay, and mobility in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, San Diego, USA, March 2010.

[71] M. Garetto, P. Giaccone, and E. Leonardi, "On the capacity of ad hoc wireless networks under general node mobility," in *Proc. IEEE INFOCOM*, Anchorage, USA, May 2007.

[72] ——, "Capacity scaling of sparse mobile ad hoc networks," in *Proc. IEEE INFOCOM*, Phoenix, USA, April 2008.

[73] ——, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the super-critical regime," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1522–1535, 2009.

[74] ——, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: the subcritical regime," *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1888–1901, 2009.

[75] M. Garetto and E. Leonardi, "Restricted mobility improves delay-throughput tradeoffs in mobile ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5016–5029, 2010.

[76] A. Ozgur and O. Lévêque, "Throughput-delay tradeoff for hierarchical cooperation in ad hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1369–1377, 2010.

[77] C. Comaniciu and H. Poor, "On the capacity of mobile ad hoc networks with delay constraints," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2061–2071, 2006.

[78] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE INFOCOM*, San Francisco, USA, March 2003.

[79] U. Kozat and L. Tassiulas, "Throughput capacity of random ad hoc networks with infrastructure support," in *Proc. ACM MobiCom*, 2003, pp. 55–65.

[80] A. Agarwal and P. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 3, pp. 71–81, 2004.

[81] S. Toumpis, "Capacity bounds for three classes of wireless networks: asymmetric, cluster, and hybrid," in *Proc. ACM MobiHoc*, 2004, pp. 133–144.

[82] A. Zemlianov and G. De Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 657–667, 2005.

[83] B. Liu, P. Thiran, and D. Towsley, "Capacity of a wireless ad hoc network with infrastructure," in *Proc. the 8th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2007, pp. 239–246.

[84] P. Li and Y. Fang, "Impacts of topology and traffic pattern on capacity of hybrid wireless networks," *IEEE Trans. Mobile Computing*, vol. 8, no. 12, pp. 1585–1595, 2009.

[85] D. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," in *Proc. IEEE INFOCOM*, Shanghai, China, April 2011.

[86] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 117–125, 2009.

[87] G. Zhang, Y. Xu, X. Wang, and M. Guizani, "Capacity of hybrid wireless networks with directional antenna and delay constraint," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2097–2106, 2010.

[88] W. Shin, S. Jeon, N. Devroye, M. Vu, S. Chung, Y. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5088–5102, 2011.

[89] X. Wang and Q. Liang, "On the throughput capacity and performance analysis of hybrid wireless networks over fading channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2930–2940, June 2013.

[90] W. Huang, X. Wang, and Q. Zhang, "Capacity scaling in mobile wireless ad hoc network with infrastructure support," in *Proc. IEEE ICDCS*, Genoa, Italy, June 2010.

[91] P. Li and Y. Fang, "The capacity of heterogeneous wireless networks," in *Proc. IEEE INFOCOM*, San Diego, USA, March 2010.

[92] P. Zhou, X. Wang, and R. Rao, "Asymptotic capacity of infrastructure wireless mesh networks," *IEEE Trans. Mobile Computing*, vol. 7, no. 8, pp. 1011–1024, 2008.

[93] O. Dousse, M. Franceschetti, and P. Thiran, "On the throughput scaling of wireless relay networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2756–2761, 2006.

[94] L. Law, K. Pelechrinis, S. Krishnamurthy, and M. Faloutsos, "Downlink capacity of hybrid cellular ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 243–256, 2010.
[95] P. Li, X. Huang, and Y. Fang, "Capacity scaling of multihop cellular networks," in *Proc. IEEE INFOCOM*, Shanghai, China, April 2011.
[96] N. Zhang, N. Lu, N. Cheng, J. W. Mark, and X. Shen, "Cooperative Spectrum Access Towards Secure Information Transfer for CRNs," *IEEE J. Sel. Areas Commun.*, to appear.
[97] M. Vu and V. Tarokh, "Scaling laws of single-hop cognitive networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4089–4097, 2009.
[98] S. Jeon, N. Devroye, M. Vu, S. Chung, and V. Tarokh, "Cognitive networks achieve throughput scaling of a homogeneous network," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5103–5115, 2011.
[99] C. Yin, L. Gao, and S. Cui, "Scaling laws for overlaid wireless networks: a cognitive radio network versus a primary network," *IEEE/ACM Trans. Netw.*, vol. 18, no. 4, pp. 1317–1329, 2010.
[100] W. Huang and X. Wang, "Capacity scaling of general cognitive networks," *to appear in IEEE/ACM Trans. Netw.*.
[101] Y. Li, X. Wang, X. Tian, and X. Liu, "Scaling laws for cognitive radio network with heterogeneous mobile secondary users," in *Proc. IEEE INFOCOM*, Orlando, USA, March 2012.
[102] H. Pishro-Nik, A. Ganz, and D. Ni, "The capacity of vehicular ad hoc networks," in *Proc. Allerton Conference*, 2007.
[103] M. Nekoui and H. Pishro-Nik, "Throughput scaling laws for vehicular ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2895–2905, 2012.
[104] G. Zhang, Y. Xu, X. Wang, X. Tian, J. Liu, X. Gan, H. Yu, and L. Qian, "Multicast capacity for hybrid VANETs with directional antenna and delay constraint," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 818–833, 2012.
[105] M. Wang, H. Shan, L. Cai, N. Lu, X. Shen, and F. Bai, "Throughput capacity of VANETs by exploiting mobility diversity," in *Proc. IEEE ICC*, Ottawa, Canada, June 2012.
[106] N. Lu, T. Luan, M. Wang, X. Shen, and F. Bai, "Capacity and delay analysis for social-proximity urban vehicular networks," in *Proc. IEEE INFOCOM*, Orlando, USA, March 2012.
[107] ——, "Bounds of asymptotic performance limits of social-proximity vehicular networks," *IEEE/ACM Trans. Netw.*, to appear.
[108] N. Lu, N. Zhang, N. Cheng, X. Shen, J. W. Mark, and F. Bai, "Vehicles meet infrastructure: Towards capacity-cost tradeoffs for vehicular access networks," *IEEE Trans. Intelligent Transportation Systems*, to appear.
[109] C. Wang, C. Jiang, X. Li, S. Tang, and P. Yang, "General capacity scaling of wireless networks," in *Proc. IEEE INFOCOM*, Shanghai, China, April 2011.
[110] C. Jiang, Y. Shi, Y. Hou, W. Lou, S. Kompella, and S. Midkiff, "Toward simple criteria to establish capacity scaling laws for wireless networks," in *Proc. IEEE INFOCOM*, Orlando, USA, April 2012.
[111] S. Weber, X. Yang, J. Andrews, and G. De Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4091–4102, 2005.
[112] A. Hunter, J. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5058–5071, 2008.
[113] S. Weber, J. Andrews, and N. Jindal, "An overview of the transmission capacity of wireless networks," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3593–3604, 2010.
[114] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, 2008.

Ning Lu (S'12) received the B.Sc. and M.Sc. degrees from Tongji University, Shanghai, China, in 2007 and 2010, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His current research interests include capacity and delay analysis, real-time scheduling, and cross-layer design for vehicular networks. Mr. Lu served as a Technical Program Committee Member for IEEE 2012 International Symposium on Personal, Indoor, and Mobile Radio Communications.

Xuemin (Sherman) Shen (IEEE M'97-SM'02-F'09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering.

He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. Dr. Shen's research focuses on resource management in interconnected wireless/wired networks, wireless network security, wireless body area networks, vehicular ad hoc and sensor networks. He is a co-author/editor of six books, and has published many papers and book chapters in wireless communications and networks, control and filtering. Dr. Shen served as the Technical Program Committee Chair for IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Co-Chair for Chinacom'07 and QShine'06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Trans. Wireless Commun.; an Associate Editor for IEEE Trans. Vehicular Technology, Computer Networks, and ACM/Wireless Networks; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, IEEE Commun. Mag., and ACM Mobile Networks and Applications.

Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, a Fellow of the Canadian Academy of Engineering, a Fellow of Engineering Institute of Canada, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.