A Survey on the Characterization of the Capacity of Ad Hoc Wireless Networks

Paulo Cardieri\textsuperscript{1} and Pedro Henrique Juliano Nardelli\textsuperscript{2}  
\textsuperscript{1}University of Campinas  
\textsuperscript{2}University of Oulu  
\textsuperscript{1}Brazil  
\textsuperscript{2}Finland

1. Introduction

Even though the interest in ad hoc wireless networks has begun in the early 1970s, several technological difficulties, particularly those related to implementation, have postponed advances in this field until the 1990s, when important issues were investigated and solved, including medium access control, routing, energy consumption, among others. These advances have allowed for actual implementation and commercial deployment of wireless communication systems based on the ad hoc concept, including wireless sensor networks, Internet access in rural areas, etc. Despite the formidable advances in this field observed in the last two decades, one key problem remains open and is still subject to intense research effort: that of modeling and measuring the capacity of ad hoc networks (Andrews et al., 2008). The intrinsic characteristics of ad hoc networks, particularly the lack of a central coordination entity and its consequences, added to the peculiarities of the wireless communication channel, make the estimation of capacity of ad hoc networks a challenging task. Despite the mentioned difficulties, researchers have proposed a myriad of metrics for characterizing the capacity of ad hoc networks under different conditions and emphasizing different aspects of the network, as described throughout this chapter.

One of the first key results in this field was achieved by Kleinrock and Silvester (Kleinrock & Silvester, 1978) in late 1970’s, when they investigated the relationship between capacity and transmission radius in a network of packet radios operating under ALOHA protocol. Takagi and Kleinrock further investigated this relationship in (Takagi & Kleinrock, 1984). Both works were based on the metric so called expected forward progress, defined in such way to capture the tradeoff relating the one-hop throughput and the average one-hop length. In fact, decreasing the one-hop length has conflicting effects on throughput: it may increase throughput due to the resulting link quality improvement, but it may also decrease throughput, due to a larger traffic and a higher contention level caused by the consequent larger number of hops between source and destination. Subbarao and Hughes (Subbarao & Hughes, 2000) improved the model previously proposed, by including the effects of the transmission system, and introduced the concept of information efficiency, defined as the product of the expected forward progress and the spectral efficiency of the transmission system. Nardelli and Cardieri extended the concept of information efficiency by taking into account the effects of channel reuse and multi-hop transmissions, leading to a new metric, named aggregate multi-hop information efficiency (Nardelli & Cardieri, 2008a; Nardelli et al.,
Based on a similar concept as that of information efficiency, Weber et al. introduced the metric transmission capacity (Weber et al., 2005), which is related to the optimum density of concurrent transmissions that guarantees that outage constraints are met. Simply stated, transmission capacity is the area spectral efficiency of successful transmissions resulted from the optimal contention density. The capacity metrics cited above, to be described in Section 2, have in common their statistical basis, resulted from the statistical nature of several mechanisms related to wireless communications, such as the interaction among nodes sharing a given channel and the propagation effects.

Following a deterministic approach to characterizing capacity of ad hoc networks and focusing on the behavior of capacity scaling laws, Gupta and Kumar introduced the concept of transport capacity (Gupta & Kumar, 2000), which relates transmission rate and source-destination distance. Gupta and Kumar formulated the transport capacity from the perspective of the requirements for successful transmission, which were described according to two interference models: the Protocol Interference Model, which is geometric-based, and the Physical Interference Model, based on signal-to-interference ratio requirements. Gupta and Kumar investigated the behavior of the network capacity when the number of nodes grows (i.e., asymptotic capacity), to show that the per-node throughput decreases as $O(1/\sqrt{n})$, where $n$ is the number of nodes in the network. This approach was followed by several authors to investigate the asymptotic capacity of wireless ad hoc networks in a variety of scenarios, such as different transmission constraints (Xie & Kumar, 2004; 2006), and with directional antennas (Sagduyu & Ephremides, 2004). Grossglauser and Tse presented an important extension of the work of Gupta and Kumar by considering the effects of mobility on the capacity (Grossglauser & Tse, 2002). They showed that, in a network with mobile nodes operating under a 2-hop relaying transmission scheme, the per-node throughput capacity may remain constant as the number of nodes in the network increases, at the cost of unbounded packet transmission delay. This important result motivated other researchers to further investigate the tradeoff between capacity and delay in mobile wireless networks (El Gamal et al., 2006), (Herdtner & Chong, 2005), (Neely & Modiano, 2005). In Section 3 we will discuss the main results on network capacity evaluation from the perspective of scaling laws.

The brief review presented above is an evidence of the complexity of the problem of characterizing capacity of ad hoc networks, leading to a number of different metrics, with different focuses and perspectives. While this large number of metrics is also an evidence of the importance of this field, it may also mislead researchers looking for appropriate models and metrics for a particular application or scenario. This chapter therefore aims at providing readers with an overview of capacity metrics for wireless ad hoc networks, emphasizing the rationale behind the metrics.

### 2. Statistical-based capacity metrics

The inherent random nature of ad hoc networks suggests a statistical approach to quantify capacity of such networks. Specifically, a statistical approach is very useful for the design of practical communication systems, when a set of quality requirements is imposed by the user application in mind. In this section we will discuss some statistical-based capacity metrics found in the literature, namely expected forward progress, information efficiency, transmission capacity and aggregate multi-hop information efficiency metrics. The specificities of each metric will be discussed and their application scenario will be pointed out.
2.1 Expected forward progress
As already mentioned, the work done by Kleinrock and Silvester (Kleinrock & Silvester, 1978) in the late 1970’s was one of the first attempts to model capacity of ad hoc wireless networks (Kleinrock & Silvester, 1978). They proposed the metric expected forward progress (EFP), measured in meters and defined as the product of the distance traveled by a packet toward its destination and the probability that such packet is successfully received. Formally, 
\[ EFP = d \times (1 - P_{\text{out}}), \] 
where \( d \) is the transmitter-receiver separation distance and \( P_{\text{out}} \) is the outage probability, i.e., the probability that the bit error rate (or other related metric) is higher than a given threshold. In (Kleinrock & Silvester, 1978) the authors introduced the idea of modeling network as a collection of nodes following a spatial point process, allowing for the use of tools and properties of Stochastic Geometry (Baddeley, 2007), making possible to derive analytical formulation relating several network parameters, such node density, propagation channel parameters, number of hops, packet error probability, etc. In fact, a plethora of analysis was performed based on the metric EFP (e.g. (Sousa & Silvester, 1990), (Sousa, 1990), (Zorzi & Pupolin, 1995)).

2.2 Information efficiency
Subbarao and Hughes (Subbarao & Hughes, 2000) extended the work done by Silvester and Kleinrock by including in the model the spectral efficiency of the transmission system, resulting in a new metric, named information efficiency (IE), which is formally defined as the product of EFP and the spectral efficiency \( \eta \) of the link connecting transmitter and receiver nodes, or
\[ \text{IE} = \eta \times d \times (1 - P_{\text{out}}). \]
Roughly speaking, IE quantifies how efficiently the information bits can travel towards its destination.

In order to understand the tradeoff captured by the information efficiency, let us consider a transmission system in which modulation and error-correcting coding techniques should be selected to optimize the IE of the network. If a modulation technique with large cardinality is used, then the spectral efficiency of the system increases, at expenses of a higher minimum required signal-to-interference plus noise ratio (SINR) to achieve a given packet error probability. This higher required SINR clearly increases the outage probability \( P_{\text{out}} \). Error correcting coding also plays an important role in this tradeoff, as it can reduce the minimum required SINR, at the expenses of a higher bandwidth, reducing therefore the spectral efficiency of the transmissions. These tradeoffs are captured by the information efficiency metric, allowing for a joint system design involving modulation, coding, transmission range, among other parameters. Following this approach, the performance of different transmission schemes was investigated, such as, discrete sequence spread spectrum (Subbarao & Hughes, 2000), frequency hopping (Liang & Stark, 2000), direct sequence mobile networks (Chandra & Hughes, 2003), direct sequence code-division multiple access with channel-adaptive routing (Souryal et al., 2005) and coded MIMO frequency hopping CDMA (Sui & Zeidler, 2009).

It should be noted that, from the perspective of the whole network, the information efficiency of a link does not tell us much about how efficiently the channel is being reused throughout the network area. We will return to this point when discussing the next two metrics.
2.3 Transmission capacity
Weber et al. proposed in (Weber et al., 2005) the transmission capacity (TmC) metric of single-hop ad hoc networks. TmC is defined as the product of the density of successful links and their communication rates, subject to a constraint on the outage probability. Formally,

$$\text{TmC} = \eta \times \lambda \times (1 - P_{\text{out}}),$$  \hspace{1cm} (3)

where $\lambda$ is the density of active links in the network. Therefore, TmC quantifies the spatial spectral efficiency of the network, capturing in its formulation the effects of active links density on the outage probability. In fact, with a high density of concurrent transmissions, information flow in the network is also higher, which is indicated by a high TmC. However, the downside of a high density of active links is an increase in the interference level, leading to a higher outage probability and, consequently, a lower transmission capacity. This tradeoff, together with the ones previously presented, are the basis of the TmC framework, which can be used to evaluate several transmission strategies with different focuses. For instance, TmC was used to study frequency hopping spread spectrum (Weber et al., 2005), interference cancelation (Weber, Andrews, Yang & de Veciana, 2007), threshold transmissions and channel inversion (Weber, Andrews & Jindal, 2007), power control (Jindal et al., 2008), among many others. In fact, TmC is one of the most flexible metrics to study single-hop ad hoc networks. However, in multi-hop links scenarios, TmC is not an appropriate metric, as it does not take into account the expected forward progress of packets, making this metric unsuitable to study, for instance, the effects of different routing strategies.

2.4 Aggregate multi-hop information efficiency
In (Mignaco & Cardieri, 2006), Mignaco and Cardieri extended the work done by Subbarao and Hughes by including the effects of spatial reuse in the definition of the IE, leading to a new metric named aggregate information efficiency (AIE). This new metric is defined as the sum of the IE of active links in the network per unit area. Nardelli and Cardieri further improved the network model used to define AIE, by including the effects of retransmissions (Nardelli & Cardieri, 2008a) and outage constraints (Nardelli & Cardieri, 2008b). Particularly, in (Nardelli & Cardieri, 2008b) the authors make the AIE an extension of the metric TmC, where the distance traveled by a packet is explicitly considered. Nonetheless, the metric AIE does not yet take into account the effects of multi-hop communication links. In (Nardelli et al., 2009), Nardelli et al. addressed such limitation and proposed the metric aggregate multi-hop information efficiency (AMIE). The idea behind the evolution from AIE to AMIE is to abstract multi-hop links and evaluate the AMIE based on the end-to-end performance of multi-hop links. Formally, the aggregate multi-hop information efficiency is defined as

$$\text{AMIE} = d \times \eta \times \lambda \times (1 - P_{\text{out}})^h, \hspace{1cm} (4)$$

where $h$ is the average number of hops between source and destination, and $d$, $\eta$, $\lambda$, and $P_{\text{out}}$ were already defined. The main advantage of the AMIE is to be more flexible and general than other similar metrics. Based on this metric, several transmission schemes and network scenarios have been investigated, such as M-QAM modulation with Reed-Solomon coding scheme and ARQ retransmissions (Nardelli et al., 2009), different access protocols with limited number of retransmissions and back-offs (Nardelli et al., 2010; Kaynia et al., 2010) and different hopping strategies (Nardelli & Cardieri, 2010).
3. Capacity scaling laws

In this section, we study the capacity of wireless networks from the perspective of scaling laws, that is, we are now interested in understanding how capacity scales as the number of nodes in the network grows. This is an important subject to be investigated, as it exposes how several intrinsic aspects of wireless communication, such as interference, channel reuse and resource limitation, affect the performance of a network. Throughput, measured in bit per second, is a typical metric of capacity of communication networks and, as such, is one of the quantities considered in this section. However, in ad hoc wireless networks, in their most general configuration, source and destination nodes may be far apart, such that direct communication (single hop) is not possible, requiring a multi hop connection, with neighboring nodes acting as relays. Clearly, multi hop connections leads to a traffic increase, as a given packet is transmitted several times before reaching its final destination. Therefore, source-destination separation distance must be taken into account when characterizing capacity in wireless ad hoc networks. In this sense, a very popular capacity metric for ad hoc networks is the transport capacity, measured in bit·meter per second. Consider a network with transport capacity of $T$ bit·meter per second. This means that the rate between two nodes spaced one meter away from each other is $T$ b/s. If the distance between the nodes is doubled, then the rate decreases to $T/2$ b/s.

Gupta and Kumar (Gupta & Kumar, 2000) investigated the transport capacity and the throughput capacity of wireless networks, and derived bounds that describe the behavior of the network capacity when the number of the nodes in the network increases. Several other authors extended the work done by Gupta and Kumar, by including other aspects in the models or improving the formulation. In this section we will review the main results from the work of Gupta and Kumar and some of the extensions, particularly those presented in (Xue & Kumar, 2006).

Before discussing the models and the results of capacity scaling law, we will review some auxiliary concepts and models. We will begin with a review of asymptotic notation, commonly used to describe the asymptotic behavior of capacity as the number of nodes in the network increases.

3.1 Some auxiliary definitions

3.1.1 Asymptotic notation

In the asymptotic analysis of capacity of wireless network, the results are often presented using the asymptotic notation (or big O-notation) (Bruijn, 2010). In this section we briefly review the definition of some of the notation commonly used. In the following, we will assume that $f(n)$ and $g(n)$ are functions that map positive integers to positive real numbers.

Definition 1 We say that $f(n) = O(g(n))$ (or, more precisely, $f(n) \in O(g(n))$), or even $f(n)$ is $O(g(n))^{1}$, if there exists a constant $c$ and there exists an integer $n_0 \geq 1$ such that $f(n) \leq c \cdot g(n)$ for $n \geq n_0$ (see Figure 1(a)).

In other words, $f(n) = O(g(n))$ means that $g(n)$ grows at least as fast as $g(n)$.

---

$^{1}$Formally, we should write $f(n) \in O(g(n))$, and the form $f(n) = O(g(n))$ is considered an abuse of notation. In fact, the symmetry that the equals sign implicitly suggests does not exist in the statements involving asymptotic notation.
Definition 2. We say that \( f(n) = o(g(n)) \) if for any positive constant \( c \), there exists an integer \( n_0 \geq 1 \) such that \( f(n) \leq c g(n) \) for \( n \geq n_0 \) (see Figure 1(a)).

The difference between the definitions of \( O() \) and \( o() \) is that in the former there must exist at least one constant \( c \) such that \( f(n) \leq c g(n) \), while in the latter the relation \( f(n) \leq c g(n) \) must be true for any constant \( c \). Therefore, \( O() \) and \( o() \) provide tight and loose upper bounds, respectively.

Definition 3. We say that \( f(n) = \Omega(g(n)) \) if there exists a constant \( c \) and there exists an integer \( n_0 \geq 1 \) such that \( f(n) \geq c g(n) \) for \( n \geq n_0 \) (see Figure 1(a)).

Definition 4. We say that \( f(n) = \Theta(g(n)) \) if there exist positive constants \( c_1 \) and \( c_2 \), and there exists \( n_0 \geq 1 \) such that \( c_1 g(n) \leq f(n) \leq c_2 g(n) \), for \( n \geq n_0 \). Equivalently, \( f(n) = \Theta(g(n)) \) if \( f(n) = O(g(n)) \) and \( f(n) = \Omega(g(n)) \) (see Figure 1(b)).

Note that \( f(n) = \Theta(g(n)) \) means that \( g(n) \) is both a tight upper bound and a tight lower bound on \( f(n) \).

3.1.2 Capacity metrics

Definition 5 (Transport capacity). Let us suppose that node \( i \) successfully transmits to node \( j \) at rate \( \lambda_{ij} \) bits per second, and that the distance between \( i \) and \( j \) is \( d_{ij} \) meters. Therefore, we can say that the network transports \( \lambda_{ij} \times d_{ij} \) bit-meter per second. Note that this metric expresses the difficulty of transmitting to a longer distances. Transport Capacity \( T \) of a network is evaluated as \( \sum_{i \neq j} \lambda_{ij}d_{ij} \), where \( \lambda_{ij} \) is the feasible rate between nodes \( i \) and \( j \).

Definition 6 (Throughput capacity). It is the guaranteed rate, measured in bits per second, that can be supported uniformly for all source-destination pairs.

3.1.3 Interference models

Definition 7 (The protocol interference model). Let \( \{ (X_i, X_{R(i)}) : k \in T \} \) be the set of active transmitter-receiver pairs in the network. According to the protocol interference model, this transmission is successfully received if the distance between nodes \( X_{R(i)} \) (the intended receiver of node
Fig. 2. The protocol model: (a) Disk around receiver $X_{R(i)}$ must be free of interfering nodes for correct reception at node $X_{R(i)}$; (b) Two links are successful if the corresponding exclusion regions are disjoint.

$X_i$ transmission) and any other node $X_k$ transmitting on the same channel is larger than the distance between $X_i$ and $X_{R(i)}$, that is

$$|X_k - X_{R(i)}| \leq (1 + \Delta)|X_i - X_{R(i)}|,$$

where $|X_k - X_{R(i)}|$ indicates the distance between nodes $X_i$ and $X_{R(i)}$, and $\Delta > 0$ is the spatial protection margin. Figure 2(a) shows a geometric interpretation of this model. Now, let us consider two pairs of active nodes $X_i$ and $X_k$, with $X_i$ transmitting to $X_{R(i)}$ and $X_k$ transmitting to $X_{R(k)}$, and with both pairs operating under the protocol model, represented by expression (5). We can show that, in order to have both transmissions successfully received, we must have

$$|X_{R(k)} - X_{R(i)}| \geq \Delta \left( |X_k - X_{R(k)}| + |X_i - X_{R(i)}| \right).$$

This result indicates that circular exclusion regions around the receivers $X_{R(i)}$ and $X_{R(k)}$, of radius $\Delta(|X_i - X_{R(i)}|/2$ and $\Delta(|X_k - X_{R(k)}|/2$, respectively, are disjoint, as shown Figure 2(b). Therefore, exclusion regions around receivers of each successful transmission are mutually disjoint, and consume a portion of the network area.

**Definition 8 (The physical interference model)** Consider, as before, a set of active transmitter-receiver pairs $\{(X_i, X_{R(i)}) : i \in \mathcal{N}\}$, transmitting over the same channel, with a transmit power assignment $\{P_i\}$. According to the physical interference model, the transmission from node $X_i$ is successfully received by node $X_{R(i)}$ if the signal-to-interference plus noise ratio (SINR) at $X_{R(i)}$ is equal to or larger than a given threshold $\beta$, that is

$$\frac{P_i}{|X_i - X_{R(i)}|^\gamma} \geq \frac{P_k}{|X_k - X_{R(i)}|^\gamma},$$

where

$$\frac{1}{\sigma^2 + \sum_{k \in \mathcal{N}, k \neq i} \frac{P_k}{|X_k - X_{R(i)}|^\gamma}} \geq \beta,$$
where $\sigma^2$ is the additive noise power. The threshold $\beta$ depends on transmission parameters, such as modulation technique, error correcting coding and the minimum acceptable bit error rate.

### 3.2 Transport capacity in arbitrary networks with immobile nodes

We consider in this section a network of $n$ immobile nodes, which can act simultaneously as source, relay or destination. These $n$ nodes are arbitrarily located in a planar disk of unity area. This means that the positions of the nodes can be adjusted in order to satisfy the conditions for successful transmissions imposed by the interference model considered in the analysis. Every node selects randomly another node as the destination of its bits. The results of this analysis are presented in the sequel, for both the Protocol Interference model and the Physical Interference model.

#### 3.2.1 Capacity under the protocol interference model

The authors of (Gupta & Kumar, 2000) showed that the transport capacity $T_A$ of an arbitrary network with $n$ nodes under the Protocol Model is

$$T_A = \Theta(W\sqrt{n}) \text{ bit \cdot meter/s,}$$

This means that the transport capacity per node is $\Theta(W\sqrt{1/n}) \text{ bit\cdot meter/s}$, and goes to zero as the number of nodes increases. Following (Xue & Kumar, 2006), this result can be proved using the fact that, under the Protocol Interference model, disks of radius equals to $\Delta |X_i - X_{R(i)}|/2$ centered at receiver nodes of successful links are disjoint (see Definition 7). Therefore, each successful link consumes a fraction of the network area and the sum of the area of disks of all successful links is upper limited by the network area (see Figure 3). Neglecting the border effects (i.e., when nodes are close to the boundary of the network area), we can write

$$\sum_{i \in \mathcal{T}(t)} \pi \left(\frac{\Delta}{2} d_i\right)^2 \leq 1 \rightarrow \sum_{i \in \mathcal{T}(t)} d_i^2 \leq \frac{4}{\pi \Delta^2},$$

where $d_i$ is the T-R separation distance $|X_i - X_{R(i)}|$ of the $i$-th T-R pair, and $\mathcal{T}(t)$ is the set of successful links at time $t$. This expression can be interpreted as follows: a set of $n$ nodes is accommodated in such way\(^2\) that condition (9) is satisfied. It should be noted that, at any

\(^2\)Recall that we are dealing with the arbitrary network case.
given time $t$, at most $n/2$ nodes will be transmitting (the other $n/2$ nodes will be receiving).

Now, we can use the Cauchy-Schwarz inequality to write

$$\sum_{i=1}^{n/2} d_i^2 \sum_{j=1}^{n/2} 1^2 \geq \left( \sum_{i=1}^{n/2} d_i \times 1 \right)^2,$$

or

$$\sum_{i=1}^{n/2} d_i \leq \sqrt{\frac{n}{2} \sigma^2} \sqrt{\frac{1}{\pi \Delta^2}}.$$

Therefore, we have found an upper bound on the sum of the T-R separation distances of successful links. Now, if we assume that all sources transmit at rate $W$, then the transport capacity $T_A$ of the network at a given time $t$ is upper bounded as

$$T_A = W \sum_{i \in T(t)} d_i \leq \sqrt{\frac{2W}{\pi \Delta \sqrt{n}}},$$

or, $T_A = O(W \sqrt{n})$ bit-meter/s. Now, we can also show that a transport capacity of $\frac{W}{1 + 2\Delta \sqrt{\frac{n-1}{2n}}}$ bit-meter/s is achievable under the Protocol Interference Model (see (Xue & Kumar, 2006) for details), completing the proof of (8).

Recalling that the network has $n$ nodes, we can conclude that the transport capacity per node is $\Theta\left(\frac{W}{\sqrt{n}}\right)$. This means that the transport capacity diminishes to zero as the number of users in the network increases. Note that we are assuming here that sources randomly select other nodes as their destinations and, therefore, the average source-destination separation distance does not depend on the number of nodes $n$. So, as $n$ increases, we have more and more nodes willing to send their bits over paths with the same average length, but sharing the same available bandwidth.

### 3.2.2 Capacity under the physical interference model

Now, if the Physical Interference model is adopted, Kumar and Gupta (Gupta & Kumar, 2000) showed that the transport capacity is

$$T_A = O(W n^{\frac{\alpha - 1}{\pi}}) \text{ bit \cdot meter/s.}$$

This upper bound can be proved recalling that, according to the Physical Interference model, a successful transmission requires that

$$\frac{P_id_i^{-\alpha}}{N + \sum_{j \in T, j \neq i} P_jd_j^{-\alpha}} \geq \beta.$$  \hspace{1cm} (11)

If we include the desired signal power in the summation in denominator, and isolate the term $d_i^{-\alpha}$, we get

$$d_i^{-\alpha} \leq \frac{(\beta + 1) P_i}{\beta \left( N + \sum_{j \in T} P_jd_j^{-\alpha} \right)}.$$  \hspace{1cm} (12)
Noting that the T-R separation distance $d_i$ is smaller than the diameter of the network area, i.e., $d_i \leq \frac{2}{\sqrt{\pi}}$, then

$$d_i^a \leq \frac{(\beta + 1) P_i}{\beta \left[ N + \left( \frac{\pi}{4} \right)^{a/2} \sum_{j \in T} P_j \right]} \leq \frac{(\beta + 1) P_i}{\beta \left( \frac{\pi}{4} \right)^{a/2} \sum_{j \in T} P_j}.$$  \hfill (13)

Now, summing the quantities $d_i^a$ of all active links, we get

$$\sum_{i \in T} d_i^a \leq \frac{(\beta + 1)}{\beta} \left( \frac{4}{\pi} \right)^{a/2}. \hfill (14)$$

Next, we use the Holder’s inequality, according to which, for $a, b > 0, p, q \geq 1$ and $1/p + 1/q = 1$,

$$\sum ab \leq \left( \sum a^p \right)^{1/p} \left( \sum b^q \right)^{1/q}. \hfill (15)$$

Therefore, recalling that there are at most $n/2$ links, then

$$\sum_{i \in T} d_i \leq \left( \sum_{i \in T} d_i^a \right)^{1/a} \left( \sum_{i \in T} 1^{\frac{a-1}{a}} \right)^{\frac{a-1}{a}} \leq \left( \sum_{i \in T} d_i^a \right)^{1/a} \left( \frac{n}{2} \right)^{\frac{a-1}{a}} \leq \left[ \frac{(\beta + 1)}{\beta} \left( \frac{4}{\pi} \right)^{a/2} \right]^{1/a} \left( \frac{n}{2} \right)^{\frac{a-1}{a}} \leq \frac{1}{\sqrt{\pi}} \left( \frac{2\beta + 2}{\beta} \right)^{1/a} n^{\frac{a-1}{a}}. \hfill (16)$$

Finally, if all sources transmit at rate $W$, the transport capacity is upper bounded as

$$T_A = W \sum_{i \in T} d_i \leq \frac{W}{\sqrt{\pi}} \left( \frac{2\beta + 2}{\beta} \right)^{1/a} n^{\frac{a-1}{a}}. \hfill (17)$$

Note that if capacity is equitably shared among all sources, the transport capacity per node is $T_A = O(W/n^{1/a})$, and goes to zero as $n$ increases. Note also that this bound indicates that a larger path loss exponent $\alpha$ leads to a higher capacity. This can be explained by noting that larger $\alpha$ means stronger signal attenuation and, therefore, reduced interference. Consequently, concurrent links can be packed together, increasing capacity.

### 3.3 Throughput capacity in random networks with immobile nodes

#### 3.3.1 Capacity under the protocol interference model

Gupta and Kumar also showed that the throughput capacity in bits per second of a random network under the Protocol Model is upper bounded by

$$\lambda(n) \leq \frac{c W}{\sqrt{n \log n}}. \hfill (18)$$

This result can be proved using again the argument that successful transmissions consume portions of the network area. Let us consider a network with $n$ nodes randomly placed on a
Network of unity area with n nodes

Fig. 4. The protocol model: (a) Disks around active receivers must be disjoint; (b) Average number of hops between source and destination.

disk of unity area. Let us also assume that all nodes transmit with a common transmission range \( r_n \). In order to guarantee that no node is isolated in the network, it can be shown that \( r_n \) must be asymptotically larger than \( \sqrt{\log n / \pi n} \) (Gupta & Kumar, 1998) (Penrose, 1997).

Next, we recall that, under the Protocol Interference model, successful transmissions require that disks of radius \( \Delta r_n / 2 \), centered at receivers, must be disjoint, as shown in Figure 4(a).

Therefore, the number of successful transmissions \( N_S \) within a disk of unity area is upper bounded as

\[
N_S < \frac{4}{\pi \Delta^2 r_n^2}.
\]  

(19)

Therefore, the aggregate number of bits transmitted per second in the network cannot be larger than

\[
\frac{4W}{\pi \Delta^2 r_n^2},
\]

where \( W \) is the common transmission rate of the individual transmissions.

Now, as before, let us consider that source nodes choose at random their destination nodes, and denote \( \bar{L} \) the average source-destination separation distance. Note that \( \bar{L} \) does not depend on the number of nodes in the network. Therefore, the average number of hops between source and destination is lower bounded by \( \bar{L} / r_n \) (see Figure 4(b)). If each source generates bits at rate \( \lambda(n) \), then the average number of bits transmitted by the whole network is given by \( n \lambda(n) \bar{L} / r_n \) and must satisfy

\[
\frac{n \lambda(n) \bar{L}}{r_n} \leq \frac{4W}{\pi \Delta^2 r_n^2}.
\]  

(20)

Finally, using \( r_n > \sqrt{\log n / \pi n} \), we complete the proof of (18).

In this same context, i.e., random networks under the Protocol Interference model, Xue and Gupta presented in (Xue & Kumar, 2006) a transmission scheme that achieves a throughput

\[
\lambda(n) \leq \frac{c W}{(1 + \Delta)^2 \sqrt{n \log n}}.
\]  

(21)

To demonstrate that (21) is valid, \( n \) nodes are randomly placed in a square of unity area. This area is tessellated by cells of side \( s_n = \sqrt{K \log n / n} \), as shown in Figure 5(a). We can
show that, with probability approaching one, each cell has at least one but no more than \( K \log n \) nodes (see (Xue & Kumar, 2006) for details). We suppose that nodes transmit with a common transmission range such that every node can transmit to any node located in its neighboring cells. In order to guarantee successful transmissions, by controlling interference, the following transmission scheme is used. We divide the cells into groups of \( M^2 \) adjacent cells (see Figure 5(a)). At each time-slot, one node from one cell of each group is allowed to transmit. Therefore, at each time-slot, there will be \( n/M^2 \) concurrent transmissions (or concurrent cells), as exemplified in Figure 5(b). Clearly, time is split into \( M^2 \) time-slots. Successful transmissions are guaranteed if concurrent cells are enough far apart, being the distance between concurrent cell controlled by the number \( M \). Note that the required value of \( M \) for successful transmission does not depend on \( n \), as only one node from each cell transmits at each time-slot. Therefore, under the Protocol Interference model, we can simply set \( M = c(1 + \Delta) \) (Xue & Kumar, 2006). Since, as before, each source node chooses at random its destination node, bits reach their destination by means of multi-hop routes. Therefore, every node transmits not only its own bits, but also bits from other nodes. Therefore, the number of bits each node pumps to the network (its own bits and those from other nodes) is related to the number \( N_R \) of multi hop routes crossing the cell to which the node belongs (see Figure 5(b)). This number \( N_R \), in turn, is related to the number of lines connecting a source and a destination that intersect a given cell. Xue and Gupta (Xue & Kumar, 2006) showed that, with probability approaching one, \( N_R \leq c \sqrt{n \log n} \). Therefore, the number of bits transmitted per second from a given cell is \( \lambda(n)c' \sqrt{n \log n} \), where \( \lambda(n) \) is the throughput per node. If \( W \) is the transmission rate in each time-slot, and recalling that there are \([c(1 + \Delta)]^2\) time-slots, then each cell transmits at rate \( W/[c(1 + \Delta)]^2 \). Therefore, the throughput per cell \( \lambda(n)c' \sqrt{n \log n} \) is feasible if

\[
\lambda(n)c' \sqrt{n \log n} \leq \frac{W}{[c(1 + \Delta)]^2},
\]
Fig. 6. Evaluation of the interference in a tesselated network under the Physical Interference model.

Concluding the proof of (21). It should be noted that one node in each cell can be designated to handle all relay traffic, while all other nodes act as sources or destinations.

Note that while (18) gives an upper bound on the throughput per node, (21) gives a feasible throughput, and we say that the order of the throughput of random networks under the Protocol Interference model is

$$\lambda(n) = \Theta\left(\frac{W}{\sqrt{n \log n}}\right).$$

(22)

As noted in (Xue & Kumar, 2006), the result in (22) suggests that the throughput of random networks is almost that achieved in the best case scenario (arbitrary networks), in which throughput is $O(1/\sqrt{n})$, despite the fact that nodes are optimally located.

### 3.3.2 Capacity under the physical interference model

When the Physical Interference model is used, it can be shown that throughput per bits per second

$$\lambda(n) = \Theta\left(\frac{W}{\sqrt{n \log n}}\right)$$

(23)

is feasible. This result can be derived using the same transmission scheme used in Section 3.3.1. We just need to show that $M$ can be selected such that transmissions can achieve $SINR \geq \beta$, as required by the Physical Interference model for successful transmission (Xue & Kumar, 2006). In order to show that, let us consider the transmission from node $X_i$ to receiver $X_{R(i)}$ in a network tesselated as before, as shown in Figure 6. This transmission is disturbed by transmissions from nodes located in the concurrent cells, which are arranged according to tiers of $8k$ cells, with $k = 1, 2, \ldots$. Using simple geometric arguments, we see that in the worst-case scenario, the distance between $X_i$ and $X_{R(i)}$ is $2\sqrt{2}sn$, and the distances between receiver $X_{R(i)}$ and interferers of the $k$-th tier are larger than $kMs - 2sn$. The aggregate interference power
can therefore be upper bounded as
\[
\sum_{k \in \mathcal{N}, k \neq i} P_k |X_k - X_{R(i)}|^\alpha \leq \sum_{k=1}^{\infty} 8k \frac{P}{(kMsn - 2sn)^\alpha} \leq \frac{8P}{(Msn)^\alpha} \sum_{k=1}^{\infty} k (k - 2/M)^\alpha.
\]

It can be shown that \(\sum_{k=1}^{\infty} \frac{k}{(k - 2/M)^\alpha}\) converges when \(\alpha > 2\) (Xue & Kumar, 2006), and therefore there is a value of \(M\) sufficiently large that guarantees \(SINR \geq \beta\) at the receiver. Therefore, the throughput
\[
\lambda(n) = \frac{cW}{\sqrt{n \log n}}
\]
is feasible in a random network under the Physical Interference model as well.

An upper bound on the throughput for random network under the Physical Interference model can be derived using the upper bound on the throughput for the case under the Protocol Interference model. In fact, successful links \((X_i, X_{R(i)})\) in a random network under the Physical Interference mode are also successful under the Protocol Model, for appropriate values of \(\Delta\) and \(\beta\). Therefore, an upper bound on the throughput for the Protocol Model also holds for the Physical Interference model. Therefore, for a random network under the Physical Interference model the throughput is upper bounded as
\[
\lambda(n) < \frac{c W}{\sqrt{n}}.
\] (24)

### 3.4 Capacity with directional antennas

In the previous sections we assumed that transmitters and receivers are equipped with omnidirectional antennas. However, it is well known that directional antennas can reduce interference and, consequently, increase capacity. Yi et al. (Yi et al., 2007) extended the work done by Gupta and Kumar by including directional antennas in the model, and investigated the effects of directional antennas on the capacity scaling laws. The radiation pattern adopted by Yi et al. is modeled as a sector with beamwidth \(\alpha\), for the transmit antenna, and \(\beta\), for the receive antenna. This is a rather optimistic model as it assumes that the energy irradiated outside the main beam is zero (i.e., sidelobes have zero gain). Following the same reasoning as in (Gupta & Kumar, 2000), the authors in (Yi et al., 2007) show that the throughput capacity per node for an arbitrary network under the Protocol Interference model scales as
\[
\lambda(n) = O\left(\frac{1}{\sqrt{n\alpha\beta}}\right).
\] (25)

Therefore, capacity increases as beamwidth decreases, what can be explaining by the fact that directional antennas reduces the overall interference, and more concurrent transmissions can be accommodated at a given time. However, even though the use of directional antennas may increase capacity, it does not change the form of the scaling law of capacity. That would be possible if \(\alpha\) and \(\beta\) decreased as fast as \(1/\sqrt{n}\), leading to a constant throughput per node as the size \(n\) of the network increases.

Spyropoulos and Raghavendra (Spyropoulos & Raghavendra, 2003) also investigated the effects of directional antennas on the capacity scaling laws of ad hoc networks, but using more
general antenna models. First, they considered an idealized radiation pattern with beamwidth \( \theta \) with unity gain, and constant sidelobe with gain \( G_{\text{side}} \), as shown in Figure 7(a). When this radiation pattern is assumed at both transmitters and receivers, the use of the Protocol Interference model results in an exclusion region as shown in Figure 7(b), in which \( R_1 \) and \( R_2 \) are given by

\[
R_1 = \left( \frac{P}{P_{\text{th}}} G_{\text{side}} \right)^{1/\alpha} \quad \text{and} \quad R_2 = \left( \frac{P}{P_{\text{th}}} G_{\text{side}}^2 \right)^{1/\alpha}.
\]

(26)

Therefore, small gain \( G_{\text{side}} \) leads to small exclusion area, which, in turn, leads to a large number of concurrent transmissions. In fact, Spyropoulos and Raghavendra showed that the throughput capacity per node is upper bounded as

\[
\lambda(n) \leq \frac{cW}{\sqrt{n \log n}} \frac{1}{\theta G_{\text{side}} + (2\pi - \theta) G_{\text{side}}^2}.
\]

(27)

In the directional antenna model adopted by Spyropoulos and Raghavendra, a narrow beam is steered towards the intended node, and out of the main beam, the antenna gain is constant. This, however, is not an appropriate model for the so called smart antenna, which are capable of not only steering a narrow beam towards a given direction, by also steering strong attenuation (nulls) towards some directions, in order to mitigate the signal from known interfering transmitters. In order to evaluate the effects of a smart antenna on the network capacity, Spyropoulos and Raghavendra considered that a smart antenna with \( N \) elements can steer a beam of gain \( G_{\text{max}} = 1 \) towards the desired direction, and gains \( G_{\text{null}} \ll 1 \) towards at most \( N - 2 \) different directions. Now, the use of this antenna model together with the Protocol Interference model allows for the accommodation of at most \( N - 2 \) receiving nodes within a circle of radius \( R = (P/P_{\text{th}})^{1/\alpha} \), and the throughput capacity per bits/sec per node is upper bounded as

\[
\lambda(n) \leq \frac{cW(N - 2)}{\sqrt{n \log n}}.
\]

(28)
3.5 Networks with mobile nodes

Grossglauser and Tse (Grossglauser & Tse, 2002) extended in another direction the work done by Gupta and Kumar, by introducing mobility in the model. As discussed in previous sections, throughput in a network with immobile nodes decays as $1/\sqrt{n}$ due to the traffic increase caused by multi hop connections between sources and destinations. Alternatively, one could use large transmission ranges in order to reduce the number of hops between source and destination. However, this strategy limits the number of concurrent transmissions, limiting the capacity of the network. Other alternative would be to restrict transmissions to neighbors. However, only a small fraction of sources are close enough to their destination nodes, limiting capacity as well. In the light of this observation, and considering a network of mobile nodes, Grossglauser and Tse (Grossglauser & Tse, 2002) developed a 2-hop relaying transmission scheme with two phases, described in the following as exemplified in Figure 8:

- Phase 1: A packet generated by a node is either directly transmitted to the corresponding destination node, or relayed to a intermediate (relay) node. In the former case, the transmission session is concluded.

- Phase 2: If the packet is sent to a relay node, the packet is buffered until the relay node is close enough to the destination node, when the packet is eventually sent to its final destination.

Note that an essential aspect of this scheme is that, due to mobility, the relay node and the destination nodes will eventually be close enough to each other to allow communication between them. Based on this model, Grossglauser an Tse showed that the average long-term throughput per S-D pair remains constant as $n$ increases, that is, throughput scales as $\Theta(1)$. An important aspect of this analysis is that the mobility model adopted assumes that, at a given time, a node is equally likely to be in any part of the network, meaning that the network topology completely changes over time. Clearly, this mobility model is an oversimplification of a real scenario, but the results obtained under this model can be viewed as upper bound on the performance.
Grossglaucer and Tse pointed out that throughput remains constant as $n$ increases at the expenses of an increasing delay. This has motivated several studies of the tradeoff between delay and throughput in ad hoc networks (El Gamal et al., 2006), (Herdtner & Chong, 2005), (Lin et al., 2006), (Neely & Modiano, 2005), (Sharma et al., 2007). For instance, El Gamal et al. (El Gamal et al., 2006) investigated this tradeoff not only for mobile networks, but also for static networks. For mobile networks, they considered a network operating under the same 2-hop relaying transmission scheme adopted by Grossglauser and Tse, and assumed a mobility model named random-walk model, according to which nodes move a distance $1/\sqrt{n}$ per unit time. They then showed that the throughput scales as $\Theta(1)$, as in (Grossglauser & Tse, 2002), but the delay scales as $\Theta(n \log n)$. For static network, El Gamal et. al showed that, at throughput $\Theta(1/\sqrt{n \log n})$ (as in the work done by Gupta and Kumar), the average delay is $\Theta(\sqrt{n}/\log n)$.

Another important extension of the work done by Grossglaucer and Tse is the one carried out by Herdtner and Chong (Herdtner & Chong, 2005) in which the authors showed that mobility alone does not increase capacity of ad hoc networks. Specifically, they showed that if the buffer size of nodes is finite and limited to $\Theta(1)$, i.e., it remains constant as $n$ increase, then the throughput capacity is only $O(1/\sqrt{n})$, instead of $\Theta(1)$. Therefore, a scaling law for throughput in a mobile network in the form $\Theta(1)$ is only possible if the buffer size increases as $n$ increases.

Lin et al. (Lin et al., 2006) investigated the tradeoff between capacity and delay in a mobile wireless network, assuming a Brownian motion model. A key parameter in this mobility model is the variance $\sigma^2$, which is related to the time required by a node to move to different parts of the network. Large $\sigma^2$ means that the node will take a short amount of time to move. The authors of (Lin et al., 2006) showed that, under the 2-hop relaying transmission scheme proposed by Grossgluaser and Tse, throughput of $\Theta(1)$ is achieved at the expenses of an average delay of $\Omega(\log n / \sigma^2)$, showing how the node speed affects the delay.

4. Summary

This chapter provided an overview of metrics for capacity evaluation of ad hoc wireless networks. The peculiarities of wireless ad hoc networks make the estimation of capacity of this kind of networks a complex task, which is evidenced by the variety of capacity metrics found in the literature.

The capacity metrics discussed in this chapter can be classified into two groups: metrics based on a statistical approach, and metrics focused on the network scalability. In the first group, discussed in Section 2, capacity metrics incorporate aspect from the physical layer (e.g. modulation parameters, spectral efficiency, etc.) and from the network layer (e.g. spatial reuse, number of hops, etc.). Therefore, these metrics are suitable for network design and parameter optimization.

The metrics in the second group, discussed in Section 3, essentially describe how network capacity behaves when the number of nodes in the network grows. As can be noted from the discussion presented in Section 3, the scaling laws derived are closely related to the particular network model and transmission scheme assumed. Therefore, even though the resulting scaling laws are rather pessimistic (per-node capacity vanishes as the size of the network increases), the results can be used as guideline for the design of more appropriate transmission schemes, that would hopefully result in non-vanishing capacity.
5. References


URL: 10.1109/TCOMM.2002.807607


URL: 10.1109/TNET.2002.801403


URL: 10.1109/MILCOM.2000.904053


A Survey on The Characterization of the Capacity of Ad Hoc Wireless Networks


