Inducing Multiscale Clustering Using Multistage MAC Contention in CDMA Ad Hoc Networks

Xiangying Yang, *Member, IEEE*, and Gustavo de Veciana, *Senior Member, IEEE*

***Abstract*—This paper proposes a new principle for designing MAC protocols for CDMA-based ad hoc networks—inducing spatial clustering in contending transmitters/receivers. We first highlight the advantages of CDMA in handling quality of service (QoS) requirements, enhancing energy efficiency, and enabling spatial multiplexing of bursty traffic. Then, based on stochastic geometric models and simulation, we show how idealized contention resolution among randomly distributed nodes results in clustering of successful transmitters and receivers, in turn leading to efficient spatial reuse. This motivates the central idea of the paper which is to explicitly induce clustering among contending nodes to achieve even better spatial reuse. We propose two distributed mechanisms to realize such clustering and show substantial capacity gains over simple random access/ALOHA-like and even RTS/CTS-based protocols. We examine under what regimes such gains can be achieved, and how clustering and contention resolution mechanisms should be optimized to do so. We propose the design of ad hoc networks supporting hop-by-hop relaying on different spatial scales. By allowing nodes to relay beyond the set of nearest neighbors using varying transmission distances (scales), one can reduce the number of hops between a source and destination so as to meet end-to-end delay requirements. To that end we propose a multi-scale MAC clustering and power control mechanism to support transmissions with different ranges while achieving high spatial reuse. The considerations, analysis and simulations included in this paper suggest that the principle of inducing spatial clustering in contention has substantial promise towards achieving high spatial reuse, QoS, and energy efficiency in CDMA ad hoc networks.**

***Index Terms*—Ad hoc network, CDMA, clustering, contention, MAC.**

I. Introduction

RECENTLY, there has been extensive research towards understanding the asymptotic "capacity" scaling of fixed and mobile ad hoc networks, see e.g., [1], [2]. In practice, however, maximizing capacity is but one of many possible design goals. At least two other objectives are critical in some ad hoc network applications: energy efficiency and quality of service (QoS). In order to design such networks one must be able to appreciate tradeoffs among the various figures of merit. One should not, for example, consider system capacity without tieing it to energy efficiency, or, for some applications, consider "capacity" without an understanding of the QoS, e.g., delays, one will incur. In this paper we consider CDMA-based ad hoc networks. Our premise is that such networks are well suited to meet QoS and energy efficiency requirements while still providing good overall network capacity if suitable MAC and routing algorithms are devised. Our focus will be on ad hoc networks that are fairly *dense*, i.e., the number of nodes within a typical node's transmission range is high, and an individual traffic load generated by a node may be bursty but requires only a fraction of the total system "capacity". Below, we discuss our motivation for considering this scenario in more detail along with related work in this area.

Is Nearest-Neighbor Routing Optimal? One basic insight provided by recent work on the capacity scaling of ad hoc networks is that it is maximized by relaying traffic along nearest-neighbor paths to a destination. Indeed it turns out to be better to maximize the density of concurrent transmissions in order to achieve a maximum amount of forward progress, i.e., bit-meter/second. However, long-range relaying may still be appealing for a variety of reasons; see [3] for a thorough discussion of 18 points. In particular, consider the following aspects.

End-to-end delay: In practice, when nearest-neighbor routing is used, a packet may need to be relayed by a large number of nodes prior to reaching its destination. Each intermediate node would typically incur a delay, depending on the MAC protocol's contention overheads, making it difficult to meet end-to-end delay requirements.

Energy efficiency: Using nearest-neighbor routing, intermediate nodes would typically be switching among transmit, receive and idle modes, further decreasing the amount of time they can spend in the sleep mode. Depending on the actual energy characteristics of the nodes, the first three modes can be fairly energy-hungry [4]. An alternative would be to permit nodes to use longer transmission distances and relay through nodes that are in a larger "neighborhood". This would allow for paths with fewer hops between the source and destination, possibly enabling a larger number of nodes to dwell in the energy efficient sleep mode.

Route efficiency and reliability: More nodes are available within longer transmission range as candidates for relaying. Thus, routing protocols may exploit this to achieve better load balancing and route reliability, by choosing nodes with light traffic load, sufficient battery energy and slow/no mobility.

Why Use CDMA? Despite the above benefits, long-range relaying requires higher transmission power and increases interference. This could be very inefficient in a narrowband system

Manuscript received January 20, 2006; revised June 24, 2006, November 9, 2006 and March 25, 2007; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Borst. This work was supported in part by the National Science Foundation Award CNS-0435307 and CNS-0509355. Part of this work was presented at the Sixth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05), Urbana-Champaign, IL, 2005.

X. Yang is with the Mobility Group Wireless Standard and Technology Organization, Intel Corporation, Hillsboro, OR 97124 USA (e-mail: xiangying.yang@intel.com).

G. de Veciana is with the Wireless Networking and Communications Group (WNCG), Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: gustavo@ece.utexas.edu).

Digital Object Identifier 10.1109/TNET.2007.902690

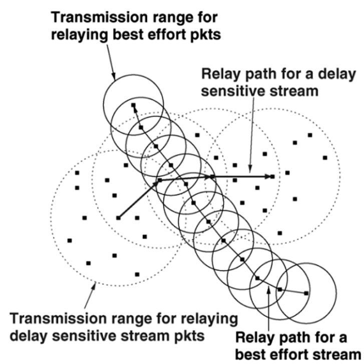


Fig. 1. An example of spatial multiplexing of heterogeneous traffic.

with bursty traffic, as additional contention is required to serialize transmissions, and might compromise network capacity and lead once again to poor delay performance. A physical layer that enables overlapping of concurrent transmissions and flexible resource allocation among traffic is thus desirable.

A CDMA-based physical layer has some key advantages in this regime—some of which are akin to those already exploited in cellular networks. Specifically consider a direct-sequence CDMA (DS-CDMA) system with spreading factor m. When a receiver de-spreads its received signal using the associated code, it roughly sees only $1/m$ of the interference before de-spreading.[1] As long as the received signal-to-interference-plus-noise ratio (SINR) after de-spreading exceeds a certain threshold, the transmission will be successful. This ability of a receiver to decode a signal in the presence of a substantial number of *concurrent* transmissions is referred as *interference averaging* and has the following advantages.

Spatial multiplexing: Because CDMA receivers can tolerate such fluctuations in interference, the network can statistically multiplex *concurrent* overlapping bursty traffic, which otherwise would block each other in a narrowband network. An example of such spatial multiplexing is shown in Fig. 1, where applications sharing the network may have different QoS requirements and possibly require different relay scales, e.g., delay-sensitive applications may prefer to use long relay distances and high transmission power to achieve low end-to-end delays while best effort traffic can use shorter relay distances and low transmission power enhancing the overall throughput. Such multiscale spatial multiplexing achieves efficient spatial/spectral reuse.

Power control: The ability to average interference can also help in managing spatial inhomogeneities in nodes' locations, and permit distributed power control with graceful degradation—e.g., when the power levels among various transmit nodes are not optimally selected due to the lack of central management/global information in an ad hoc network.

Robustness: Critical problems such as hidden or exposed terminals in narrowband systems are not as prominent in CDMA-based ad hoc networks. Indeed, their impact is mitigated by spreading gains and proper power control. These advantages may potentially simplify MAC design and operation.

[1]We consider codes generated by PN sequences which are quasi-orthogonal and conservatively assume that PN code cross-correlation is $1/m$ [5].

MAC Design Challenges and Related Work: There has been a resurgence of interest in CDMA-based ad hoc networks [6], [7] since last decade, perhaps driven by the above mentioned advantages. However, a CDMA physical layer using different relay distances requires a MAC design overcoming several challenges. First, the MAC must achieve efficient spatial reuse, yet "long" transmission distances compromise capacity by causing interference to a region with area that is quadratic in the distance [1], [8], [9]. Second, nodes should use power control in order to realize heterogeneous relay distances and energy savings. Yet heterogeneous power levels cause carrier sensing, commonly used in MAC for collision avoidance purposes, to be unreliable. We will discuss this more in Section VI.

In [7], an ALOHA-like random access protocol is proposed, which is simple but, as shown later in this paper, inefficient from a spatial reuse perspective. In [10], joint power control and scheduling is considered to achieve optimal spatial reuse but assuming a centralized scheduler. Various MAC designs using local signaling among nodes have been proposed [11]–[14]. They exchange request-to-send(RTS)/clear-to-send(CTS) messages or broadcast busy tone pulses over a common code/frequency control channel, realize transmitter–receiver hand-shaking, facilitate power control and result in improved spatial reuse. However, these approaches require extra hardware complexity and usually assume a contention free control channel.

"Clustering" in ad hoc networks has been extensively studied in the literature. In particular, joint network clustering and MAC design are studied in [15]–[17]. The scope of these papers is quite different from ours in that they deal with realizing a clustered network management structure *above MAC layer* to be used over relatively long time scales. One of the novelties of this work is *inducing* clustering in MAC contention processes so as to enhance performance for CDMA-based ad hoc networks. Our method dynamically structures clusters in the transmission patterns during each packet transmission slot, i.e., on much shorter time scales.

Main Contributions and Organization of the Paper: In this paper, we investigate fundamental strategies to devise efficient MAC protocols and exploit the tradeoffs among QoS, energy efficiency and spatial reuse in CDMA ad hoc networks. Detailed implementations are not in the scope of this paper. In Section II, we introduce an appropriate model and notation, which will serve to highlight the role interference plays in a CDMA system. In Section III, we review some basic MAC designs and discuss their ability to realize tradeoffs among QoS, energy and capacity. Then, by studying an idealized contention resolution mechanism, we identify a key feature of efficient MAC protocols for CDMA-based ad hoc networks—clustered packing of concurrent transmissions. We motivate the study of a novel MAC design approach based on inducing spatial clustering among contending nodes to further enhance spatial reuse. To this end, we propose and analyze two distributed algorithms that achieve such clustering in Sections IV and V. Through analysis and simulation we are able to show substantial capacity gains over simple random access/ALOHA-like and RTS/CTS-based protocols, when clustering mechanisms are properly optimized. In Section VI, we propose a multi-scale

MAC clustering mechanism to enable transmissions with different distances and support heterogeneous applications QoS requirements, while still achieving efficient spatial reuse. We present simulation results in Section VII and conclude in Section VIII.

II. Modeling Interference and Spatial Reuse in CDMA Systems

A. Network and Traffic Model

We assume each network node has a single transceiver which can either transmit or receive, but not both. Further we assume explicit signaling is only feasible between a source-destination pair, i.e., other nodes cannot decode signaling messages that are not intended for them, but they can measure interference. Explicit signaling among all neighboring nodes will be costly to implement or incur significant overheads in CDMA ad hoc networks. In Section III, we also consider MAC designs leveraging GPS capability.

We adopt a simple stochastic geometric model for transmitters and receivers in a multi-hop ad hoc network, in which we assume there is no mobility in the time scale of transmissions. We consider a homogenous offered load, captured by a homogenous distribution of transmitters including source and relay nodes. We assume that the set of transmitters are spatially distributed according to a homogenous Poisson point process [18], denoted by $\Pi = \{X_i, i \in \mathbb{N}\}$, with intensity λ. Nodes are interchangeably referred to/by their locations. Each transmitter is assumed to be sending to a receiver, that is assumed to be always available at distance d. This assumption on receiver availability is reasonable for the scenario where the transmit range exceeds the typical nearest-neighbor distance, i.e., a dense network. Later we consider heterogeneous network with different relay distances in Sections VI and VII. The choice of d captures the tradeoff between QoS and energy consumption. For delay sensitive traffic, d should be large enough to ensure delay due to multi-hop relaying sufficiently small. For the best-effort traffic, d should be the minimal distance that can maintain connectivity with neighbors.

We capture wireless channel using a basic path loss model where if a transmitter uses a power level ρ the receive power at distance d is given by $\rho_r(d) = \rho \times d^{-\alpha}$. The path loss exponent α is typically assumed to be between 3 and 5. We consider all transmissions using a fixed data rate with an associated target SINR $\zeta = \beta/m$, where β is some constant threshold associated with the data rate and m is the spreading factor. We model the power control such that the receive power is a constant ρ_0. Therefore, transmission power depends only on the relay distance. Initially we consider a homogeneous network with fixed relay distance d and transmission power $\rho = \rho_0 d^{\alpha}$.

Throughout this paper, we assume a slotted system with synchronous contention and transmission. We believe synchronization is desirable to support efficient MAC scheduling in ad hoc networks. First, data transmission and acknowledgements (ACKs) are well protected after handshaking, which eliminates the need for maintaining states, e.g., NAVs in 802.11 and [14]. Second, synchronous contention provides better priority access mechanism enabling better QoS support than asynchronous implementations, see e.g., [19]. Refs. [20] and [21] have shown that network-wide synchronization is indeed feasible for a resolution of a few microseconds, with or even without GPS enabled reference points, and can be maintained for hours without refresh. Although synchronization incurs extra overheads, such as inter-frame spacing and state transition guard time, these are typically measured in only a few microseconds versus milliseconds for frame/slot length. Moreover, similar MAC inefficiencies exist with asynchronous contention resolution in which the required carrier sensing usually causes conservative back-off both spatially and temporally [22]. Therefore, without significantly compromising capacity or energy consumption, synchronization offers many benefits that will warrant these overheads. See also [7], [19], [23] for representative protocols based on synchronized contention and the case for this not being excessively costly.

In terms of performance, in this paper we will focus on spatial reuse as measured by the transmission capacity [8]. It corresponds to the density of concurrent successful transmissions achieved after MAC contention resolution. Transport capacity, as defined by [1], can be obtained by transmission capacity by with the data rate and mean relay distance. Other performance metrics are not quantified in this paper. QoS support is addressed by classifying different flows into different MAC scheduling classes. Energy efficiency is addressed indirectly by reducing outage and protocol overheads.

B. Characterizing Interference and Outage

A receiver sees the interference powers from other concurrent transmitters. For conventional DS-CDMA, outage happens when the SINR at the receiver does not exceed a certain threshold ζ. Conditioning on a typical receiver at the origin O, the outage probability for a typical receiver $p_o(\lambda, d)$, which depends on the intensity of transmitters and transmission distance, is given by

$$p_o(\lambda, d) \approx \mathbb{P}\left(\frac{\rho_r(d)}{\sum_{X_i \in \Pi} \rho |X_i|^{-\alpha}} \leq \zeta\right) \tag{1}$$

where $|X_i|$ denotes the distance from interferer i to the receiver located at the origin. Note we have neglected the role of ambient noise since the capacity of a dense network is mostly interference constrained. The outage probability for a transmission is a critical to performance, because it not only represents a failed transmission, but also, wasted energy, possible violation in QoS requirements, and additional load on the network induced by retransmissions.

The outage probability is difficult to compute because the interference term defined in (1), i.e., $\sum_{X_i \in \Pi} \rho |X_i|^{-\alpha}$ does not generally have a closed-form distribution. There are, however, more intuitive ways to understand interference and outage in this model. In particular, let us consider the rough geometry of transmission and interference ranges.

Note that the aggregate interference is the sum of independent, but non-identically distributed, random variables. Neighboring interferers nearby a typical receiver contribute very strong interference. Remote interferers may still potentially contribute enough aggregate interference to cause

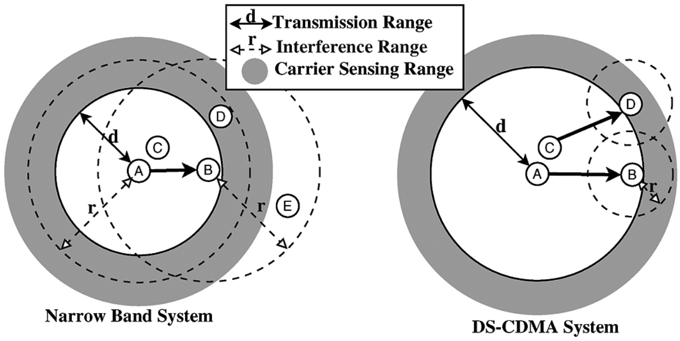


Fig. 2. The transmission range, interference range, and carrier sensing ranges for an idealized narrowband and CDMA system. Since $A \rightarrow B$ requires no concurrent transmission in the critical interference range r around $B, C \rightarrow D$ is not allowed in a narrowband system, but may be allowed in the CDMA system.

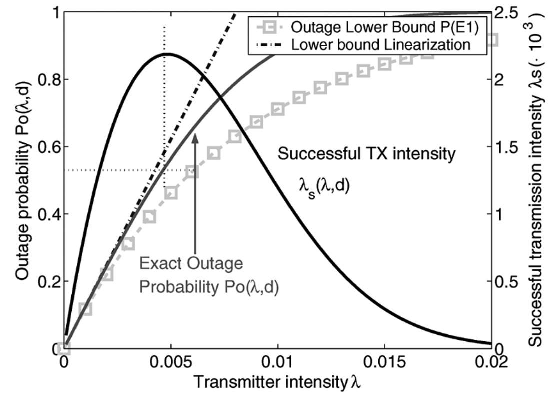


Fig. 3. Effective capacity λ_s is maximized when $p_o \approx 0.5$. The calculation is based on the exact analysis for $\alpha = 4$ case in [24]. Also shown are our outage lower bound and its linear approximation, which are both close to the exact analytical result in the low outage regime of $p_o < 0.5$.

an outage although each transmitter only contributes a small amount due to path loss. To quantify the neighborhood covering nearby strong interferers, we define the *critical interference range* r to be the largest distance from a receiver at which a *single* interferer could be located and still cause an outage, i.e., the largest r such that

$$\frac{\rho d^{-\alpha}}{\rho r^{-\alpha}} \leq \zeta \Longrightarrow r = (\zeta)^{\frac{1}{\alpha}} d. \tag{2}$$

In other words, a disc of radius r around each successful receiver should contain no transmitter. Note that the suppression range of radius r is only a *necessary* condition for successful reception because remote interferers are ignored. However, as discussed in the sequel, a suppression range of radius r is indeed close to a sufficient condition.

There is a fundamental difference in considering interference in a narrow band versus a CDMA system. As shown in Fig. 2, r is usually larger than the transmission distance d in a narrowband system. By contrast, in a CDMA system, r is smaller than d if the spreading factor m is large. This allows one to schedule concurrent overlapped transmissions, as long as there is no interfering transmitter within a range r of another receiver. Note that r depends on the transmission distance and associated power control strategies and the level of allowable overlapping depends on m.

Dumbbell Model—Modeling Concurrent Transmissions and Spatial Reuse in CDMA Ad Hoc Networks: Interference from nearby interferers within a distance r from a receiver, see (2), provides a reasonable approximate abstraction for the relevant source of interference, particularly when α is large. To see this, let $\mathbf{B}(O, r)$ denote a ball centered at the origin O with a radius r. We let the event $\mathbf{E_1}$ denote the occurrence that *at least* one interferer is within $\mathbf{B}(O, r)$ which in turn would necessarily cause an outage for a receiver at the origin. It follows that the outage probability for a typical receiver $p_o(\lambda, d)$ is such that

$$p_o(\lambda, d) \geq \mathbb{P}(\mathbf{E_1}). \tag{3}$$

For a Poisson point process with intensity λ, the probability of $\mathbf{E_1}$ is given by

$$p_o(\lambda, d) \geq \mathbb{P}(\mathbf{E_1}) = 1 - e^{-\lambda\pi r^2} = 1 - e^{-\lambda\pi\zeta^{\frac{2}{\alpha}} d^2}.$$

As shown in Fig. 3, in the low outage regime, the outage lower bound associated with $\mathbb{P}(\mathbf{E_1})$ is very accurate. One can show the lower bound becomes tighter as α increases. In the sequel we shall use a further approximate outage probability based on linearizing this lower bound, i.e., $p_o(\lambda, d) \approx \pi d^2 \lambda \zeta^{2/\alpha}$. The accuracy of the lower bound and its linearization are exhibited in Fig. 3. The key point here is to support the intuitive abstraction for interference and outage in a network where nodes are randomly distributed and transmit at the same power level: nearby interferers within an interference range r, given by (2), contribute most of the outage, and thus considering only "nearby" interferers is reasonably accurate. This allows to simplify visualizing how contention occurs among transmitters-receivers pairs. As shown on the left panel in Fig. 4, each transmission corresponds to a dumbbell with disks of radius $r/2$ at the transmitter and receiver, connected by a bar of length d. A successful transmission is modeled by a dumbbell without prohibited overlaps, i.e., no transmit disc overlaps with a receive disc on either end.[2] Thus, as shown in Fig. 4, among the three contending transmissions, only two transmissions can be successful. Spatial reuse corresponds to realizing a high spatial density of dumbbells subject to *at least* satisfying the rules on overlaps. We only use this dumbbell model to illustrate contention and later clustering phenomena among transmissions. Subsequent MAC designs will *not* be based on this dumbbell model and our simulations will factor the actual interference seen by a receiver.

III. Characteristics of MAC Contention

Let us reconsider the previous analysis results to understand basic MAC protocols based on random access and contention resolution schemes, from the perspective of the overall system capacity, outage probability, and associated overheads. We will show that they tend to be very inefficient. Further by considering an idealized contention resolution mechanism, we will motivate a novel approach based on inducing clustering among MAC layer contenders.

[2]If two such disks overlap, then the associated nodes are within r of each other.

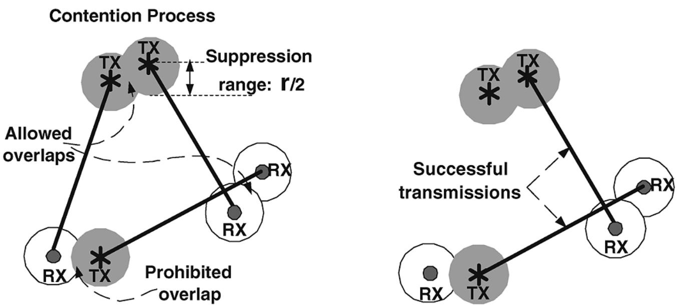


Fig. 4. On the left, contentions among three concurrent transmissions. On the right, after contentions, only those two transmissions whose receivers do not have prohibited overlap survive.

A. Random Access and Simple Contention Resolution MAC Protocols Achieve Poor Spatial Reuse

We consider a simple model for a random access protocol akin to those in [7]–[9], [24], with no carrier sensing or handshaking. The lack of a coordination phase among concurrent transmitter–receivers reduces overheads but increases the likelihood of outage. Fig. 3 exhibits the tradeoff between the outage probability $p_o(\lambda, d)$, the intensity of successful transmissions $\lambda_s(\lambda, d) = \lambda(1 - p_o(\lambda, d))$ and the intensity of contenders λ. By using the linear approximation discussed earlier we get the following simple expression:

$$\lambda_s(\lambda, d) \approx \lambda \left(1 - \pi d^2 \lambda \zeta^{\frac{2}{\alpha}}\right). \tag{4}$$

The contention intensity λ which maximizes $\lambda_s(\lambda, d)$, say λ^*, is roughly given by

$$\lambda^*(d) = \frac{1}{2\pi d^2}\left(\frac{1}{\zeta}\right)^{\frac{2}{\alpha}} \quad \text{and} \quad \lambda_s(\lambda^*, d) = \frac{1}{4\pi d^2}\left(\frac{1}{\zeta}\right)^{\frac{2}{\alpha}}. \tag{5}$$

Note from Fig. 3, to achieve a maximal capacity, one incurs a high outage probability, roughly 0.5. This observation also holds for the analysis in [9], wherein transmitters use a exponentially distributed transmission power and outage probability at the optimal contender intensity maximizing the capacity is roughly $1 - e^{-1} \approx 0.63$. The key observation here is that maximizing capacity using a random access MAC will require a high density of transmitters resulting in high spatial reuse but also a high likelihood of outage.

When energy efficiency or delay is a concern, it is reasonable to limit outage to a small likelihood. Define ε to be the maximally tolerable outage probability. This places a limit on the intensity of contenders λ and intensity of successful transmissions λ_s. Our previous work [8] showed that

$$\frac{\alpha - 1}{\alpha}\frac{\varepsilon}{\pi d^2}\zeta^{-\frac{2}{\alpha}} + O(\varepsilon^2) \leq \lambda_s \leq \frac{\varepsilon}{\pi d^2}\zeta^{-\frac{2}{\alpha}} + O(\varepsilon^2)$$

when all transmissions employ transmission power ρ for a fixed transmission distance d. For small ε, one obtains a fairly low capacity which is only reasonable for networks supporting low traffic loads. To summarize, we stress that for MAC protocols based purely on random access can only achieve a moderate spatial reuse but will incur a high outage probability and thus poor energy efficiency.

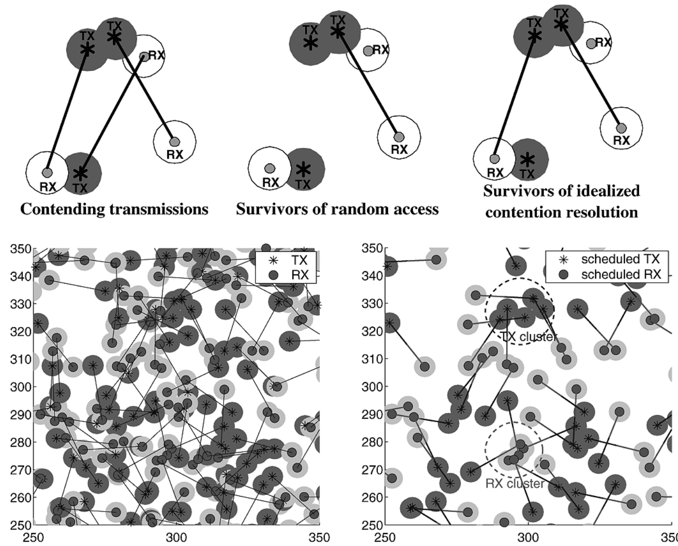


Fig. 5. On the top left panel, an initial contending pattern of three transmissions. On the top middle panel, only one successful transmissions under random access protocol. On the top right panel, two successful transmissions under idealized contention resolution scheme. On the bottom left panel, a realization of contending transmissions in our simulation, and on the bottom right, transmissions surviving an *idealized* contention resolution of prohibited overlaps.

To resolve the energy efficiency problem of random channel access, more sophisticated protocols introduce carrier sensing and contention resolution phases prior to data transmission. Contention resolution defers certain contending transmitters and ensures the receiver availability by exchanging signaling messages, e.g., RTS and CTS, to achieve hand-shaking between a transmitter and receiver and to suppress potential neighboring interfering transmissions. Signaling messages can be quite small relative to data packets and thus such mechanisms are worthwhile to reduce outages on data transmission and save on energy. However, a simple signaling scheme may not help improve spatial reuse in a dense network. More improvement on spatial reuse can be realized by proper back-off strategies which are effective under light loads. However, considering the much reduced interference range when using CDMA, see Fig. 2, simple variations of RTS/CTS or carrier sensing based narrowband MAC protocols are too conservative, leading to a poor spatial reuse.

B. Idealized Contention Resolution Achieves Efficient Spatial Reuse With a Clustering Pattern

Intuitively MAC contention resolution schemes "remove" transmissions with prohibited overlaps. For example, consider the realization of the contenders on the top left panel of Fig. 5. In this case two receivers have a prohibited overlap and only one transmission will be successfully scheduled by the RTS/CTS handshaking scheme discussed earlier—see the middle figure on the top of Fig. 5. Yet a sophisticated contention resolution mechanism could achieve a better spatial reuse. For example, an *idealized* contention resolution process might allow at least one of prohibited overlaps to survive—i.e., remove dumbbells with prohibited overlaps one at a time until no such overlaps are left. A possible result with two successful transmissions of such an idealized scheme is shown on the top right panel of

Fig. 5. A simulation of such an idealized scheme is shown at the bottom of Fig. 5: on the left is a realization of contenders while on the right is a subset of successful transmissions. Successful transmitters and receivers tend to be clustered. Note that this idealized scheme is much better in terms of spatial reuse than simply removing all transmitter–receiver pairs with prohibited overlaps, yet it would not be straightforward to implement in a distributed system.

Our simulations systematically exhibit two key aspects of contention-based MACs that have perhaps not been fully appreciated. First, as shown in our earlier analysis to achieve a high density of successful transmissions, i.e., a dense packing of dumbbells, one needs to have a high density of contenders. As a result, a significant number of transmitters will need to defer due to contention resolution or see a high outage probability under random access MAC protocol—roughly 50%. The second observation is that successful transmitters and receivers are *clustered*, in particular when efficient spatial reuse is achieved—see the right panel in Fig. 5. This is a unique property of CDMA-based ad hoc networks where receivers are capable of interference averaging and thus can tolerate certain level of neighboring interference. This suggests that by explicitly inducing spatial clustering in contention mechanisms, one might further improve spatial reuse. We consider this next.

IV. Inducing Clustering Based on a Virtual Grid

One can consider inducing spatial clustering of transmitters in many ways. To show the benefit of inducing clustering, let us first consider an idealized deterministic placement. Following this, we propose an approach assuming that nodes contend synchronously and are aware of their locations. These capabilities are used to directly generate a spatial clustering of transmitters and receivers. Finally, in the next section, we assume nodes are able to monitor interference levels to roughly infer relative locations of other contending nodes and use signaling stages to achieve clustering of transmissions. We use these two representative distributed mechanisms to exhibit the benefits of *inducing spatial clustering* in a practical system.

A. Idealized Deterministic Clustering

The pioneering work of [25] showed that optimal spatial reuse can be achieved by placing transmitters on a regular triangle grid. We shall extend their result to a CDMA ad hoc network where optimal spatial reuse is achieved by clustering. Although it is non-trivial to find the optimal clustered placement with interference from all sources being considered, the goal of this section is to show the deterministic clustered placement of transmitters/receivers on a regular grid is compact and thus leads to very efficient spatial reuse.

Assume we are free to select *both* the locations and states, e.g., transmitter or receiver, of nodes. Specifically, as shown on the left in Fig. 6, we assume tight clusters of transmitters and receivers are placed at the centers of cells in a regular triangular grid of size $\frac{3\sqrt{3}}{4}d^2$ according to a checkerboard pattern. Each transmitter is assumed to transmit to a distinct receiver in one of the three neighboring cells at a fixed distance d. The number of nodes n within each cluster will be determined to ensure all transmissions are successful. Let π_n^d denote a set of locations in

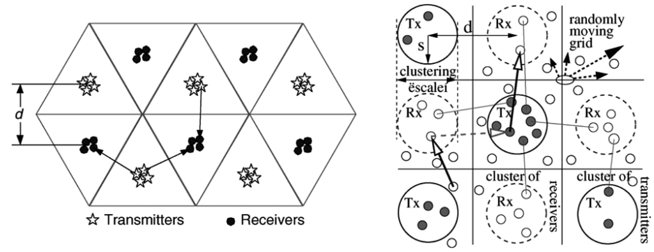


Fig. 6. On the left, an idealized deterministic placement on a triangular grid. On the right, a clustering of randomly located nodes in a random virtual square grid with an example of typical routing patterns marked with hollow-headed arrows.

$\mathbb{R}^2$, corresponding to n-clustered transmitters in this checkerboard configuration. In order for a receiver located at origin to successfully receive from a transmitter at X_0, $|X_0| = d$, one must have

$$\frac{\rho d^{-\alpha}}{\sum_{X_i \in \pi_n^d \setminus \{X_0\}} \rho |X_i|^{-\alpha}} \geq \zeta.$$

Fact 4.1, below, states this constraint in terms of a maximum allowable number of transmit nodes per cluster.

Fact 4.1: Under the clustered triangular grid placement of transmit nodes π_n^d, and $\alpha > 2$, a maximum number of nodes per cluster of $\left\lfloor \frac{1/\zeta + 1}{k(\alpha)} \right\rfloor$, where

$$k(\alpha) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left[\left(\left((\sqrt{3}i)^2 + (3j-1)^2\right)^{-\frac{\alpha}{2}}\right) + \left(\left(\sqrt{3}i + \frac{\sqrt{3}}{2}\right)^2 + \left(3j + \frac{1}{2}\right)^2\right)^{-\frac{\alpha}{2}}\right]$$

can be placed while ensuring no outage. This gives a density of successful transmissions of $\lambda_s = \frac{4}{3\sqrt{3}d^2}\left\lfloor \frac{1/\zeta + 1}{k(\alpha)} \right\rfloor$. □

This fact is shown through a brute force calculation of the aggregate interference offered by transmitter clusters in π_n^d at various distances from the origin, which equals $nk(\alpha)d^{-\alpha}\rho$. Since $k(\alpha) \approx 3$ this suggests we need only consider the interference due to the $3n$ nodes which are a distance d from the origin, in particular when path loss is severe. One can expect similar results for other regular grids, e.g., for a square grid $k(\alpha) \approx 4$ because each receiver is surrounded by four neighbor transmitter clusters. Thus, in the Section IV-B we will only focus on interference from the nearest clusters.

Comparing the best achievable spatial density of successful transmissions for randomly distributed MAC versus our ideally clustered grid, i.e., (5) to Fact 4.1, we have an approximate gain factor of $\frac{16\pi}{3\sqrt{3}k(\alpha)}\zeta^{\frac{2}{\alpha}-1}$, e.g., when $\alpha = 4$, $m = 128$ and $\beta = 10$ dB a 10-fold increase in capacity without outage!

B. Inducing Clustering Using GPS Capability

In practice, one cannot choose the placement of transmitters and receivers. Yet for a homogeneous distribution of transmitters wishing to send a distance of roughly d one can approximate this pattern. For example, suppose that nodes are location aware and can determine their location relative to a known *virtual grid* of span d whose location evolves in a "random" but known manner with time. This can be achieved with GPS-capable synchronized nodes, that share a randomization seed driving the evolution of the grid—[7], [23], [26] have used this shared seed idea to allow nodes to infer other nodes' states. Given this information and an *a priori* convention, a node can determine if it lies within a current transmitter/receiver cluster. For example, in a square grid, as shown on the right of Fig. 6, we assume for now that nodes within a transmitter cluster transmit/relay to receivers in a neighboring receiver cluster.

Furthermore, we let the parameter s determine the spatial scale of clustering and thus proximity of clustered nodes. Note that nodes that do not fall in either a transmitter or a receiver cluster region can defer, e.g., enter the sleep mode, unless they are sources or sinks.[3] "Random moving" of the grid happens in much slower time scale than transmission slots to balance loads and energy consumption. If s is too small, each cluster will contain but a few transmitters and we may under-utilize the available capacity. If it is too large, there may be too many transmitters and/or interference variability (due to increased proximity), resulting in outages at receivers. One may consider what is a good choice for s given the intensity λ of the Poisson point process Π of active nodes.

C. Performance of Virtual Grid Approach

Let us first evaluate the outage probability of a receiver at the center of a receiver cluster and use this as an approximate estimate for the outage probability of a typical receiver. As for the deterministic placement, we will focus on the nearest four transmit clusters as the source of interference. Using Campbell's theorem, see [18, ch. 4.1, 4.3], we can evaluate the mean and variance of the interference as follows.

Fact 4.2: Let Y denote the aggregate interference power level from the four transmit clusters closest to origin in a regular square grid, i.e.,

$$Y(\lambda, d, s) = \sum_{X_i \in \Pi} \mathbf{1}\left(X_i \cap A(s) \neq \emptyset\right) \times \rho |X_i|^{-\alpha}$$

where $A(s) = \bigcup_{i=-1,1} \bigcup_{j=-1,1} \mathbf{B}((i \times d, j \times d), s)$ is the union of these transmit cluster discs of radius s which are closest to the origin. Then

$$E[Y] = \lambda \int_{A(s)} \rho |x|^{-\alpha} dx, \quad \mathrm{Var}(Y) = \lambda \int_{A(s)} \rho^2 |x|^{-2\alpha} dx.$$

□

Assuming Y is approximately Gaussian,[4] the outage probability for a receiver located at the origin is given by

$$p_o(\lambda, d, s) = \mathbb{P}\left(\frac{\rho d^{-\alpha}}{Y(\lambda, d, s)} < \zeta\right) \approx \frac{1}{2} - \Phi\left(\frac{\frac{1}{\zeta} - E[Y]}{\sqrt{\mathrm{Var}(Y)}}\right). \tag{6}$$

[3]This requires a routing protocol to give special considerations to the first and last hops, e.g., the first hop of the typical route on the right of Fig. 6, which should also require a proper power control.

[4]Y is a linear sum of random variables, which represent the interference contributed by the strong interferers in neighbor cells. Because the expected number of such interferers is large in a dense CDMA network, we approximate Y's distribution to be Gaussian.

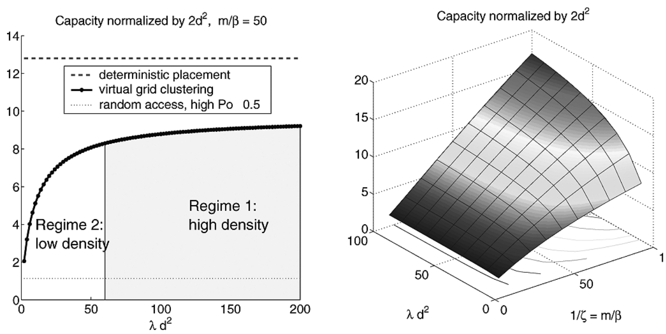


Fig. 7. On the left, a comparison of capacities of different schemes. On the right, the max capacity improves in both λd^2 and $1/\zeta$.

Note that $p_o(\lambda, d, s)$ increases in both λ and s, i.e., the number of interferers. Suppose each transmitter finds a distinct receiver and thus there are no collisions due to concurrent transmissions to a receiver. Then for a fixed λ and d one can consider optimizing the cluster scale s so as to maximize the mean number of successful transmitters per cluster. Let n^* denote the maximal mean number of successful transmissions per cluster, i.e., given by

$$n^*(\lambda, d) = \max_s \left\{ \lambda \pi s^2 \left(1 - p_o(\lambda, d, s)\right) | s > 0 \right\} \tag{7}$$

and let s^* denote the optimal s maximizing (7). The transmission capacity, see Section II, in the case of square grid is $\frac{n^*(\lambda, d)}{2d^2}$. We consider two regimes corresponding to different node density.

Regime 1—High Node Density: In this regime, $\lambda \pi d^2$ is larger than $1/\zeta$, resulting in an optimal clustering scale $s^* \ll d$, i.e., nodes are clustered closely around the center of each cell. This is akin to the deterministic placement considered earlier. However, each cluster will have a random, Poisson distributed, number of nodes with mean $\lambda \pi s^{*2}$. In this regime, Fact 4.2 gives

$$E[Y] \approx 4\lambda \pi s^{*2} \rho d^{-\alpha} \quad \text{and} \quad \mathrm{Var}(Y) \approx 4\lambda \pi s^{*2} \rho^2 d^{-2\alpha}.$$

We can in turn estimate the outage probability using (6). Note this optimization problem of (7) depends only on $\lambda \pi s^2$, the mean number of contending nodes per cluster.

As shown on the left panel of Fig. 7, in regime 1, the capacity achieved is close to the case of idealized deterministic placement. As shown on the right panel in Fig. 7, in this high density regime, we again obtain a transmission capacity that grows roughly linearly in $1/\zeta$ as in the case of idealized deterministic placement, leading to a significant improvement over random access/ALOHA protocols, on the order of $\zeta^{(2/\alpha)-1}$ or around 700% when $\zeta = 1/50$, $\alpha = 4$.

Regime 2—Low Node Density: In general if the intensity of active nodes is not high, the optimal choice of s^* will become comparable to d, i.e., one needs to increase the cluster scale so that a sufficient number of nodes can be scheduled. This case is difficult to study analytically because the distribution of the interference is affected by both the variability in the number of nodes per cluster, and the increased interference variability due to their larger set of possible locations. To better understand this regime, we numerically solved the optimization problem (7) when $\alpha = 4$. As shown on the left of Fig. 7 in the low density regime, if the spatial intensity λ is too small and there is not a sufficient number of nodes inside each cluster of size s, this negatively impacts the capacity. As λ grows larger, the capacity improves but the improvement eventually is marginal. Even in this regime, the achieved capacity is still significantly larger than random access/ALOHA. As shown on the right of Fig. 7, the capacity improves in $1/\zeta$ sub-linearly but closer to linearly when λd^2 is larger.

Summary—Virtual Grid Mechanisms: Note that (not shown in Fig. 7) for both high density and low density regimes, the outage probability is significantly lower at the operating point achieving the highest transmission capacity, e.g., only about 5% compared with the outage probability of 50% in random access/ALOHA-like protocols, and this improves in m. When $\alpha \to 2$, the capacity gain becomes marginal and independent of m. However, the benefit of low outage remains significant.

An ad hoc network may have a non-homogenous spatial density of nodes or traffic. Thus, it would be desirable to let the clustering scale adapt to such inhomogeneities. A straightforward approach would be to modify our virtual grid mechanism such that each cell i has a different cluster scale s_i.

Thanks to GPS capability, the virtual grid approach achieves both good capacity and energy efficiency. For example, nodes that are not covered by the transmit or receive cluster areas, can put themselves to sleep, until the grid moves. Furthermore, the overheads are low because nodes can infer locality of traffic and thus contention or signaling are not required, except that collisions where two or more transmitters send to the same receiver must be avoided. Collisions should be unlikely however, since the corresponding routing protocol should take advantage of long relay distances to achieve load balancing. The achievable capacity is close to that achieved by an ideal deterministic placement. Note that relatively low spatial intensity compared with the spatial scale d may negatively impact the overall capacity because it prevents this mechanism from effectively inducing clustering.

V. Inducing Clustering via Multistage Contention

Contention resolution and handshaking-based protocols induce clustering among scheduled nodes; see Fig. 5. One might ask how practically these mechanisms should be designed and optimized for this purpose. We start with the case where nodes use a common transmission distance/power and then consider the case where they use heterogeneous transmission distance/power.

A. A Multistage Contention Protocol

Let us consider inducing clustering through a modified synchronous multi-stage RTS/CTS mechanism. Consider a two-stage example with the timing diagram shown in Fig. 8.

Stage 1 handshaking: In Stage 1 a subset of transmitters perform the three-way handshaking with their intended receivers, i.e., RTS, CTS, followed by an additional "confirmation" RTS message. Only transmitter–receiver pairs who successfully exchange the three messages survive the first stage. These survivor pairs serve as "seeds" for clusters in the subsequent handshaking stage(s).

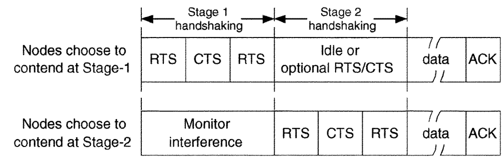


Fig. 8. Timing diagrams of a two-stage contention MAC with the top for Stage 1 transmitter/receiver and the bottom for Stage 2 transmitter/receiver.

Stage 1 monitoring: During Stage 1 contention, potential transmitters and receivers[5] not participating in the first stage handshaking process synchronously monitor interference levels, for which they can indeed distinguish RTS and CTS time slots. Doing so permits them to evaluate their proximity to surviving Stage 1 transmitters and receivers.

Stage 2 handshaking: In Stage 2, transmitters that sensed a "strong" (see below) CTS signal in Stage 1 do not participate in Stage 2, i.e., are suppressed since they would likely interfere with the a successful Stage 1 receiver. Similarly a Stage 2 receiver which successfully receives an RTS from a transmitter, will only send back a CTS, if during Stage 1 it did not sense a "strong" confirmation RTS signal. Thus, the role of Stage 1 "confirmation" RTS is to signal receivers in the Stage 2 that they will be interfered with and thus to suppress their CTS. This process can be carried out through multiple stages to achieve a higher level of spatial reuse. For example, survivors of Stage 1, might also concurrently participate in Stage 2 with RTS/CTS exchange, permitting actual Stage 2 contenders to estimate aggregate interference.

We need to formally define thresholds which are used to decide when signals should be deemed strong enough to result in suppression. The critical range analysis, see (2), suggests that a single interferer will cause outage for a transmitter–receiver pair using transmit power level ρ over a transmission distance d, if the interference as seen at the receiver exceeds $\frac{\rho d^{-\alpha}}{\zeta}$. To tolerate measurement uncertainty in the interference, we introduce a backoff factor c, where $0 < c \leq 1$ and thus a signal will be deemed strong if it exceeds $c \times \frac{\rho d^{-\alpha}}{\zeta}$. Note that c should be close to 1 otherwise we may be too conservative in utilizing available capacity. For purposes of visualizing this with 'dumbbells' and analytically studying clustering phenomenon later, we calculate the clearance range r_c around transmitters and receivers to be $r_c = c^{-\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha}} d$.

The assumption of equal transmission power is important for the protocol designs in Fig. 8. Our multistage contention protocol allows transmitters and receivers to contend in the same fashion in terms of transmission power and back-off threshold. Given a clustering pattern of successful transmissions, this symmetry in contention induce clustering among not only receivers but transmitters, which can be observed on the right panel of Fig. 9. Thus, transmissions in both directions, i.e., RTS/DATA and CTS/ACK, will succeed. Thus, we can allow Stage 1 contenders to participate in Stage 2 contention and we need only a single ACK slot for all transmissions contending at different stages. We will see in the sequel that when transmission power

[5]Those who will not be active at this cycle do not need to monitor, which is more efficient than [14] in which all nodes have to do consistent monitoring.

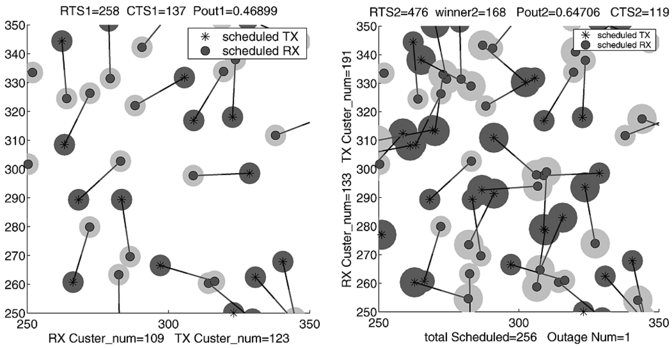


Fig. 9. On the left, contention result of successful transmitter–receiver pairs, which serve as cluster "seeds" for Stage 2. On the right the contention result after Stage 2, in which transmitters/receivers are indeed closely clustered with stage-1 transmitters/receivers and this significantly increases the overall clustering level.

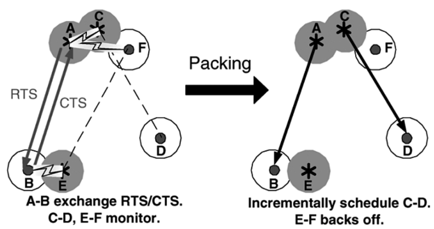


Fig. 10. Multistage contention achieves clustering by spatial packing.

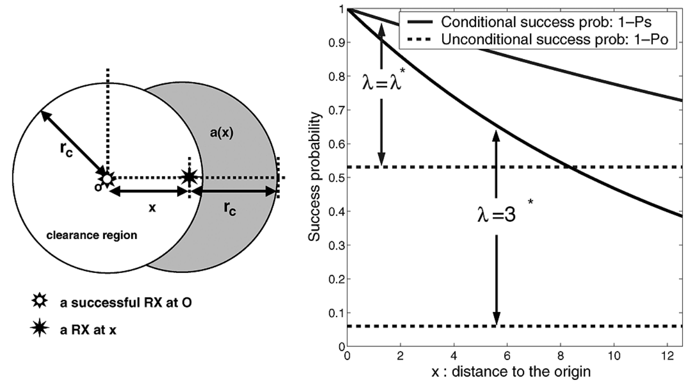


Fig. 11. On the left the area $a(x)$ for obtaining outage lower bound conditioning on a successful receiver at O. On the right given the intensity λ of contending transmitters, the upper bounds for $1 - p_s(x)$, the success probability of a receiver at distance x conditioning on a successful receiver at O, and the success probability $1 - p_o(\lambda, d)$ without conditioning. For $p_o(\lambda, d)$ and λ^* see (1), (5).

levels are heterogeneous, extra considerations are needed for the protocol design.

As exhibited by the simulation in Fig. 9, nodes that survive Stage 1 only achieve "weak" clustering while Stage 2 survivors are dense and clustered. In the sequel, we will analyze the multistage contention process and answer the following questions:

- How does Stage 1 simple contention realize initial weak clustering and interact with subsequent stages to realize these gains?
- What is the capacity gain provided by multistage contention process and how should it be optimized?

B. How Does Multistage Contention Lead to Clustering?

The basic intuition of inducing clustering with multistage contention is to let initial contenders in Stage 1 generate a spatial reuse pattern which serves as seeds for subsequent contenders to further enhance the clustering pattern. The back-off strategy based on RTS/CTS power level is visualized in Fig. 10 with our dumbbell model, i.e., subsequent transmitters or receivers will back-off if they have prohibited overlaps, e.g., E is too close to B and senses a strong CTS. On the other hand, subsequent transmitters/receivers clustering with existing transmitters/receivers, e.g., C is close to A, are likely to be able to contend and succeed. Thus, with multistage contention, we spatially "pack" subsequent transmissions by clustering them with existing transmissions scheduled in previous stages.

"Weak" Clustering From Stage 1: To quantify the clustering effect after Stage 1, we will consider, given a successful receiver at the origin, what is the intensity of other successful receivers around the origin after Stage 1. "Clustering" means the intensity of successful receivers should be higher close to the origin. Let $\lambda^{(1)}$ and $\lambda^{(2)}$ denote the intensity of contending transmitters in Stage 1 and 2 respectively. Consider a receiver that succeeds Stage 1 and suppose it is located at the origin O. Since it was successful during Stage 1 it must have cleared a disc of radius r_c of transmitters around it. Now conditioning on this receiver contending transmitters *outside* the disc are still homogenously distributed with intensity $\lambda^{(1)}$. Specifically let us evaluate the intensity of successful receivers within the ball $B(O, r_c)$ centered at the origin. Fact 5.1 summarizes the results in this regard.

First denote $a(x)$ to be the area of $B(x, r_c) \setminus B(o, r_c)$ as shown on the left of Fig. 11, which is the interference region of a receiver located at x, uncovered by the interference region of a given successful receiver at O. One can calculate

$$a(x) = x^2 \left[\omega \sin\theta + 2\omega^2(\pi - \theta)\right] - \pi r_c^2$$

with $\omega = r_c/x$ and $\cos\theta = x/(2r_c)$.

Fact 5.1: Consider Stage 1 contention of the multi-stage mechanism described above. Conditioning on a successful receiver at the origin O, the intensity of other successful receivers $\lambda_s^{(1)}(x)$ within the disc $B(O, r_c)$ at distance x from the origin is roughly given (upper-bounded) by $\lambda_s^{(1)}(x) = \lambda^{(1)}(1 - p_s(x))$, where $p_s(x) = 1 - e^{-\lambda^{(1)} a(x)}$ is the probability that a receiver a distance x, $0 < x < r_c$, from the origin, is suppressed by one or more Stage 1 transmitters within the area $a(x)$.

Proof: The above result is similar to that used to compute the outage lower bounds. Consider a receiver located a distance x from the origin. The receiver will survive Stage 1, if it has no transmitters within a ball of radius r_c of itself. As shown in Fig. 11, part of this ball has already been cleared of transmitters, since a successful Stage 1 receiver lies at the origin. Thus, our candidate receiver will be successful if there are no transmitters within the region $B((x, 0), r_c) \setminus B(O, r_c)$ whose area is given by $a(x)$, with the probability of this occurring given by 1 minus the probability that a homogenous Poisson point process places no points in a region of area $a(x)$. It then follows that the intensity of Stage 1 receivers within $B(O, r_c)$ which are sent RTS's by transmitters outside this disc is also homogenous and has intensity $\lambda^{(1)}$ for $r_c < d/2$. ■

Fact 5.1 shows that *given* a successful receiver at the origin, the probability that another Stage 1 receiver at distance x away from it, is successful decreases quickly with distance. Therefore, the *conditional* intensity for the point process of other *successful* receivers within $B(O, r_c)$ has a non-homogenous intensity $\lambda_s^{(1)}(x)$ decaying in x as shown on the right in Fig. 11, leading to clustering of successful receivers. This decay is more significant when the intensity is higher. Since Stage 1 is essentially a random access stage, which generates seeds for additional clustering in subsequent stages, one can try to optimize the overall clustering by first generating a maximal number seeds, e.g., using the optimal contention intensity in (5).

Optimizing the Additional Clustering Realized by Stage 2: Once Stage 1 transmitter–receiver pairs are scheduled, they suppress nearby Stage 2 receivers and transmitters respectively. Thus, the point process of Stage 2 transmitters that will contend depends on the process of successful Stage 1 receivers—it no longer corresponds to a homogeneous Poisson point process. Nevertheless, in the sequel, we assume a homogeneous Poisson point process for the purpose of computing the void probability, and show via simulation that this is a fairly good approximation. Consider Stage 1 successful transmitters with density $\lambda_s^{(1)}$, and a successful Stage 1 receiver located at the origin. *Assuming* that the Stage 2 transmitters which contend outside the ball $B(o, r_c)$, are reasonably well approximated by a homogenous Poisson point process with intensity $\lambda^{(2)}$, we can reuse the approach underlying Fact 5.1. Specifically the expected number of successful Stage 2 receivers clustered within the ball $B(O, r_c)$ of a Stage 1 receiver is approximately

$$\int_{o}^{r_c} \lambda^{(2)} e^{-\left(\lambda^{(2)}+\lambda_s^{(1)}\right)a(x)} 2\pi x dx. \tag{8}$$

To maximize (8), i.e., maximize spatial reuse, one can determine the optimal $\lambda^{(2)} \approx 5/8r_c^2$. Subsequently, the mean number of successful Stage 2 receivers around a successful Stage 1 receiver is roughly 0.9, i.e., we get roughly a 90% improvement on the capacity at the second stage and the expected number of successful receivers per cluster to be roughly 1.9. Our simulation results match this analysis quite well.

In summary, to maximize the capacity and achieve optimal clustering, the density of the contending transmitters should be

$$\lambda^{(1)} \approx \frac{1}{2\pi d^2}\left(\frac{1}{\zeta}\right)^{\frac{2}{\alpha}} \quad \text{and} \quad \lambda^{(2)} \approx \frac{5}{8r_c^2} \tag{9}$$

respectively. Depending on the choice of c, $\lambda^{(2)}$ is typically 2–3 times of $\lambda^{(1)}$. Because Stage 1 creates "seeds" for Stage 2 clustering, the second stage is able to handle a fairly high density of contenders. Overall, unlike Stage 1, the system capacity is not as sensitive to the intensity of Stage 2 contenders, because only transmitters who know they will not severely interfere a Stage 1 receivers will perform Stage 2 signaling. The above results exhibit the robustness of the proposed two stage signaling.

In the implementation of a multi-stage contention MAC protocol, to achieve an optimal contention density at each stage without centralized control, nodes participate in each stage based on certain probabilities as in a random access slotted ALOHA and update their contention probabilities according to interference levels or contention results. The rule of thumb should be to manage the contention probability of Stage 1 such that collision probability is below 0.5 and then choose Stage 2 contention probability 2–3 times of the Stage 1 contention probability. Thanks to the multi-stage design, performance is not very sensitive to suboptimal choices in contention probabilities as shown in our simulations. The overhead of multistage contention is comparable to simple RTS/CTS-based protocols, because most of the performance gain is achieved by including only two stages as shown in our simulations.

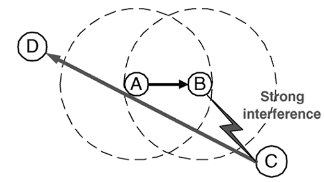


Fig. 12. A hidden terminal problem when aggressive power control exists.

VI. Handling Multi-Class or Non-Homogenous Traffic and Node Distributions With Multi-Scale Contention and Clustering

A. Why Contention With Heterogeneous Power is Challenging

A realistic network may support transmissions with different relay distances for the following reasons. First, spatial intensity of nodes may be heterogenous and nodes may need to use different distances to maintain connectivity. Second, applications sharing the network may have different QoS requirements and possibly require different relay strategies, e.g., relaying on different spatial scales, see the discussion of spatial multiplexing and Fig. 1 in Section I. Third, the network may consist of heterogeneous devices with different transceiver capabilities. In this case nodes should use power control to choose transmit power levels corresponding to the desired relay distances. Note that such open-loop power control geared at achieving performance or service differentiation is initiated by users/applications and it is different from the power control based on closed-loop feedback which is intended to ensure successful transmission by compensating for channel variations.

Inducing spatial clustering to achieve high spatial reuse faces additional difficulties in this context. First, monitoring nodes cannot correctly infer what the interference regions of contending nodes in the previous stage(s) are, given a mixture of heterogeneous transmissions with different power levels. This suggests transmissions with similar power levels should contend together. Second, contention at different power level does not fully solve the hidden node problem. As shown in Fig. 12, transmissions with low power and short range, e.g., $A \rightarrow B$, perform RTS/CTS handshaking before transmitting actual data. However, long-range transmissions with strong power like $C \rightarrow D$ may still severely interfere with low-power transmissions because C may not sense RTS/CTS from A or B. This suggests transmissions with higher power should contend before transmissions with lower power.

B. A Multi-Class Multi-Stage Contention Protocol

One approach to deal with this problem is via a multi-class and multi-stage contention process. The basic idea is to allow

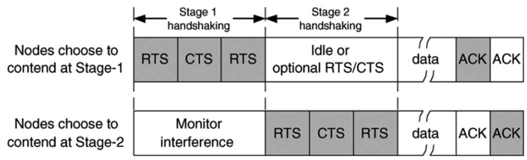


Fig. 13. Multiscale multistage contention protocol requires separate ACK slots for each contention stage.

transmissions with higher transmission power to perform handshaking first so as to enable transmitters and receivers in subsequent stages to detect their RTS/CTS and correctly estimate interference regions. Specifically, consider a network where nodes use one of k possible relay distances d_i, $i = 1, \ldots k$ satisfying $d_1 > d_2 \ldots > d_k$ each with an associated transmit power level $\rho_i^t, i = 1 \ldots k$. Suppose these power levels are pre-defined such that typically the receive power are the same, e.g., $\rho_i^t = \rho_0 d_i^\alpha$. Note that this is an idealized model and we will discuss imperfect power control versus relay distance later. In the sequel we refer to nodes which choose relay distance d_i to be of class i. Our new contention protocol is a variation on the multi-stage RTS/CTS/RTS process considered earlier. As shown in Fig. 13, we assume that only class i nodes perform handshaking at Stage i based on monitoring interference levels for stages $1, \ldots i - 1$ and thus inferring whether they will interfere with, or be interfered by, contenders in previous stages, by taking into account predefined power levels used at each stage.

The intuition for this choice is that by allowing only a given class to contend at a particular stage, nodes monitoring the process can obtain reasonable estimates of the proximity of contenders based on *a priori* knowledge of their transmitting power levels and the received interference levels. Furthermore the ordering in which classes contend (based on transmission distances) is enforced because if $i < j$ then the packing achieved by class i is likely to be less dense than that of class j. This ensures that long-range transmissions are effectively packed prior to committing to short range ones. This is akin to packing large "objects" first and then squeezing the smaller ones as appropriate within the gaps. This ordering also ensures a contender will hear the signaling of all relevant contenders with higher transmission power in previous stages, which solves the hidden terminal problem even under heterogeneous transmission power levels.

This approach achieves a multi-scale clustering and high spatial reuse of successful transmissions. Fig. 14 exhibits a realization of a two-class scenario. Transmissions with long relay distances and larger interference regions are scheduled in Stage 1, while transmissions with shorter relay distance are scheduled in Stage 2. As can be seen, in addition to clusters of receivers for both classes of transmitters, additional fine scale clustering of short range transmission fill the void area remaining after Stage 1.

In the case where nodes use heterogeneous transmission power levels, we can no longer allow all nodes to contend or ACK together as we did in a homogeneous network as shown in Fig. 8. This is because the desired multiscale clustering pattern for nodes with heterogeneous power levels is biased toward inducing receiver clusters. By clustering receivers along with proper power control corresponding to different transmission

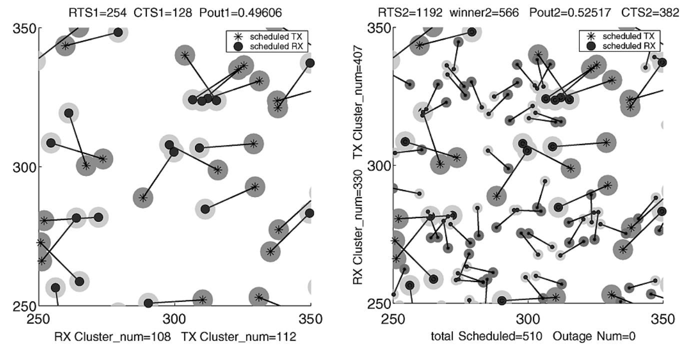


Fig. 14. On the left, the resulting transmitter–receiver pairs of a multi-stage multi-class contention protocol's Stage 1 contention among nodes relaying delay sensitive traffic, i.e., using longer relay distances and thus having larger interference ranges. On the right the resulting transmitter–receiver pairs for Stage 1 and 2 for a multi-class multi-stage contention protocol. Note how the short-range transmissions cluster around Stage 1 receivers as well as independently in voids left during Stage 1.

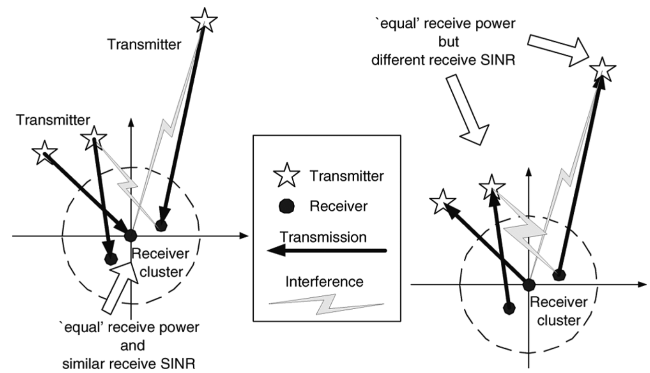


Fig. 15. A typical multi-scale clustering pattern, in which receivers are likely clustered but transmitters may not. On the left, when receiving RTS/DATA, all receivers achieve similar receive signal power and SINR with power control. On the right, when receiving CTS/ACK, transmitters achieve uneven SINR if power control is in place.

distances, i.e., $\rho^t = \rho_0 \times d^\alpha$, receive signal power and interference power across nodes will be similar. Thus, receivers will achieve similar SINR, as shown on the left of Fig. 15, and all successfully receive, assuming a fixed data rate for all nodes. However, for CTS/ACK signaling messages in the reverse direction, concurrent transmissions with power control lead to uneven SINR at the destinations, with nodes close by the receiver cluster overwhelmed by high power interference and suffering a very poor SINR, as shown on the right of Fig. 15. For this reason, we need separate contention and ACK slots for each class.

In summary, the multi-class multi-stage contention protocol can also handle the general case where transmission distances are randomly distributed, e.g., nodes in class i may have any relay distances in $[d_{i-1}, d_i]$. The variation of transmission power/distance even within the same class can negatively impact performance. We mitigate this problem by sending CTS/ACK at some maximal power level, e.g., $\rho_0 d_i^\alpha$ for class i. Thus, CTS/ACK will be received by transmitters with roughly equal SINR and succeed with high probability. Indeed, here we mimic the power control strategy used by cellular networks with receiver clusters corresponding to base stations, transmitters corresponding to mobile terminals, RTS/DATA corresponding to reverse link using variable power and CTS/ACK corresponding to forward link using equal power. Simulations show that this typically leads to more than 10% performance gain.

C. Overhead Analysis

In a multi-class scenario, since each contention stage leads to a significant improvement of spatial reuse, the overhead from multi-stage contention has negligible impact on system capacity. However, the energy consumption associated with contention overhead should be considered carefully. Here we present a brief discussion. Define $t_{\rm cont}$ to be the duration of each contention stage, t_{pl} to be the duration of payload transmission time and $t_{\rm ack}$ to be the duration of ACK slot. Assume the energy consumptions for receiving/monitoring and transmission are all 1 unit, which represents the worst case since monitoring typically consumes less energy than transmission. For class j, nodes spend a fraction of $\frac{j\cdot t_{\rm cont}}{j\cdot t_{\rm cont}+t_{pl}+t_{\rm ack}}$ in total energy consumption to perform multi-stage contention. Overall, the fraction of energy consumption associated with contention overhead in the proposed protocol is sub-linear in the number of QoS classes supported in the network. The interesting observations here are: 1) increasing payload duration t_{pl}, i.e., frame length, reduces overheads but incurs longer delay per hop, and 2) reducing the number of stages/classes also reduces overheads but leads to heterogeneity in each class and thus hurts capacity. Once again, a network design has to achieve a good tradeoff among capacity, QoS and energy consumption.

VII. Performance Evaluation

We shall compare the performance of multistage contention with existing MAC protocols.

Our multistage contention scheme in a network with homogeneous traffic has three stages, with the first two stages being identical to the two-stage version discussed in Section V-A and the last stage consists of retries by those who fail in the previous two stages.

Our multi-class multi-stage contention scheme in a network with heterogeneous traffic of two classes, has the same implementation discussed in Section VI-B.

Simple random channel access is the same as discussed in Section III, where all intended transmissions transmit together without signaling or contention resolution.

Serialized contention resolution is an idealized scheme derived from [14]. We assume transmissions contend one at a time. Each transmission in contention checks the SINR margin of existing transmissions and the contention is successful if this transmission can succeed without causing outage to existing transmissions. We further assume transmissions never stop. This can be visualized as packing transmissions one by one into the system by checking their feasibility. This scheme can be shown to be Pareto optimality and it offers an optimistic upper bound on the performance of [14], because it assumes signaling messages are always successfully received, the network is saturated with ongoing transmissions, and thus ACK protection can be ignored.

We fix the path loss exponent $\alpha = 4$ and the SINR threshold $\zeta = \beta/m = 10/512$. We generate transmitters according to a

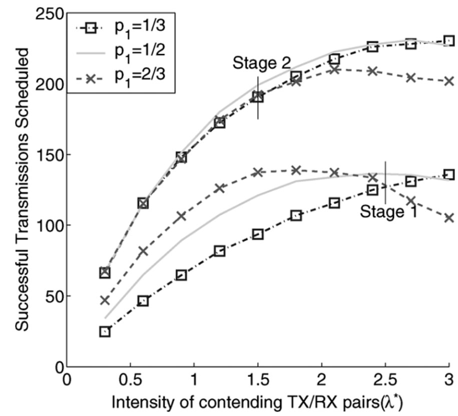


Fig. 16. Performance comparison of different configurations for contention probability in a homogeneous network.

Poisson point process in a 600×600 rectangle area, and associate corresponding receivers at certain distance with random angle. At each round, different MAC schemes are applied to the same realization of nodes. Each performance point is an average of 50 rounds. We only consider nodes falling inside a margin of 150 to eliminate edge effects.

We shall first use simulation to show the robustness of our multi-stage protocol when varying contention probabilities at each stage. We consider a simple two-stage contention protocol in a homogeneous network with the same transmission distance $d = 20$. We increase the density of contending transmissions from $0.25\lambda^*$ to $3.5\lambda^*$, see (5). We assume when a node chooses to contend, with probability p_1 it will contend in Stage 1 and otherwise it will contend in Stage 2. We consider the following cases $p_1 = 1/3$, $p_1 = 1/2$, and $p_1 = 2/3$. We plot the number of successful transmissions after both Stage 1 and 2. As shown in Fig. 16, the performance is indeed insensitive to suboptimal contention probabilities particularly when the network is lightly loaded. Overall, it conforms to the analysis that Stage 2 should allow more contenders than Stage 1 for the best performance. It appears that $p_1 = 1/3$ is a good choice and we will use this in the sequel.

In the sequel, we examine the spatial reuse achieved by different schemes in the following three scenarios.

Scenario 1. A homogeneous network where all transmissions have the same relay distance $d = 20$. We increase the density of contending transmissions from $0.25\lambda^*$ to $3\lambda^*$, see (5). Nodes randomly choose Stage 1 with probability 1/3 and Stage 2 with probability 2/3. Those who fail both Stage 1 and 2 retry in Stage 3.

Scenario 2. A heterogeneous network with two classes of traffic. Class 1 has relay distance $d_1 = 20$ and class 2 has relay distance $d_2 = 10$. We fix the density of class 1 contending transmitters to be $\lambda^*(d_1)$. We vary the density of class 2 contending transmitters from 0 to $4\lambda^*(d_2)$, i.e., increasing the heterogeneity in the network.

Scenario 3. A heterogeneous network with transmission distances uniformly distributed between [10,30]. We assign transmissions with distance between [10,20] to be class 1 and the rest to be class 2. All other settings are the same as in Scenario 2. Scenario 3 captures the behavior of a realistic network where receiver availability causes variation among relay distance.

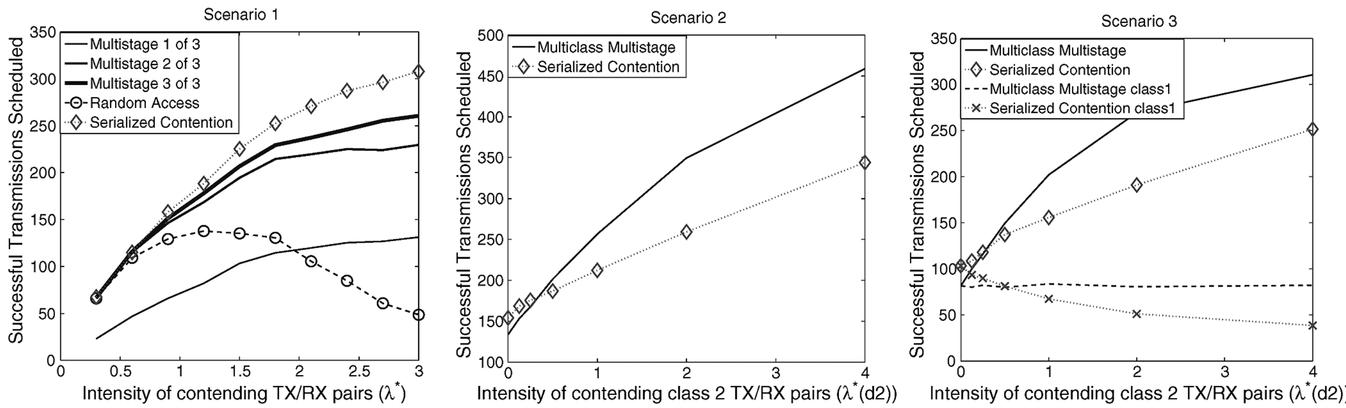


Fig. 17. Performance comparison of different MAC schemes in a homogeneous network.

In Scenario 1, as shown on the left panel in Fig. 17, we plot the number of successful transmissions achieved by different schemes. For multistage contention scheme, we plot the *overall* successful transmissions achieved at the end of each stage. The serialized contention resolution scheme has the best performance and random channel access has the worst performance. As discussed before, Random channel access is sensitive to the contention intensity and there is some optimal contention intensity to achieve the best performance, e.g., in Fig. 17 this happens when the contention intensity is roughly $1.2\lambda^*$, see (5). Beyond this point, throughput collapse happens for random channel access. The performance of multistage contention scheme is within 85% of the serialized contention resolution but remains much better than random channel access, which is very impressive because it does not rely on the assumptions made for the serialized contention scheme. In this simulation, multistage contention is properly configured by choosing the stage 2 contention intensity to be approximately twice of that in stage 1. Stage 2 achieves most of the performance gain, which indicates that our multistage protocol can be implemented with only two or three stages, i.e., a relatively low overhead, without compromising potential performance.

In Scenario 2, we only compare the serialized contention resolution and our multi-class multi-stage contention protocol discussed in Section VI-B. We measure the total number of successful transmissions achieved by both protocols. As shown in the middle panel of Fig. 17, when there is more heterogeneity in the network with more class 2 transmissions, our multi-class multi-stage contention protocol outperforms the serialized contention resolution significantly, which is in contrast with the first simulation in which all transmissions have the same relay distance. The reason is that when transmissions with heterogeneous relay distances contend in a random order, even the serialized contention resolution no longer leads to a desired spatial reuse pattern, e.g., a multi-scale clustering shown in Fig. 14, as transmissions with different relay distances tend to block each other. Thus, in a heterogeneous network, the multi-class multi-stage contention protocol is particularly attractive to achieve efficient spatial reuse by inducing multiscale spatial clustering.

In Scenario 3, as shown in the right panel of Fig. 17, we again observe our multi-class multi-stage contention protocol significantly outperforms serialized contention resolution. Note that due to variations in the transmission distances across and within classes, the number of successful transmissions is lower than Scenario 2. Also observe that multi-class multi-stage contention provides better fairness and QoS guarantees. When we increase class 2 traffic, the performance of class 1 is compromised in serialized contention, while class 1 is not affected in our multi-class multi-stage contention thanks to synchronous contention.

VIII. Conclusion

In this paper, we have considered MAC design for ad hoc networks based on a CDMA physical layer. Our interests in considering such networks are multiple, and involve effective ways of addressing QoS, energy efficiency, and efficient spatial multiplexing of bursty traffic, etc. We are specifically interested in the regime where the transmission distance of a node exceeds nearest-neighbor scales and thus is likely to include a number of other potential transmitters. In other words, the transmission distance may in fact exceed several nearest-neighbor hops. In this case, a properly designed MAC and CDMA physical layer will permit overlaying of concurrent transmissions without resulting in outage. The key observation in this paper is that for such a physical layer technology an efficient packing of transmitter–receiver pairs will be one exhibiting clustering of receive and transmit nodes. As such, we propose a novel approach to MAC design whereby clustering is induced in the contention process with a view on enhancing capacity, and sketch two distributed approaches that achieve this end.

References

[1] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

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

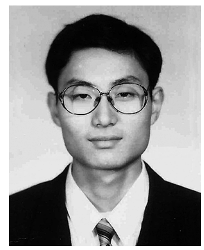
[3] M. Haenggi and D. Puccinelli, “Routing in ad hoc networks: A case for long hops,” *IEEE Commun. Mag.*, vol. 43, no. 10, pp. 93–101, Oct. 2005.

[4] M. Stemm and R. H. Katz, “Measuring and reducing energy consumption of network interfaces in hand-held devices,” *IEICE Trans. Commun.—Special Issue Mobile Comput.*, vol. E80-B, no. 8, pp. 1290–1302, 1997.

[5] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, “On the capacity of a cellular CDMA system,” *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, May 1991.

[6] M. Pursley, “The role of spread spectrum in packet radio networks,” *Proc. IEEE*, vol. 75, no. 1, pp. 116–134, Jan. 1987.

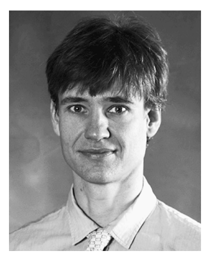
[7] T. J. Shepard, "A channel access scheme for large dense packet radio networks," in *Proc. ACM SIGCOMM*, Aug. 1996, pp. 219–230.
[8] S. Weber, X. Yang, G. de Veciana, and J. G. Andrews, "Transmission capacity of ad hoc networks with outage constraint," *IEEE Trans. Inf. Theory*, vol. 13, no. 6, pp. 1275–1288, Dec. 2005.
[9] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "An ALOHA protocol for multihop mobile wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 421–436, Feb. 2006.
[10] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," in *Proc. IEEE INFOCOM*, 2002, pp. 976–984.
[11] J. J. Garcia-Luna-Aceves and J. Raju, "Distributed assignment of codes for multihop packet-radio networks," in *Proc. IEEE MILCOM*, 1997, pp. 450–454.
[12] M. Joa-Ng and I. Lu, "Spread spectrum medium access protocol with collision avoidance in mobile ad hoc wireless network," in *Proc. IEEE INFOCOM*, 1999, pp. 776–783.
[13] J. P. Monks, V. Bharghavan, and W. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. IEEE INFOCOM*, 2001, pp. 219–228.
[14] A. Muqattash and M. Krunz, "CDMA-based MAC protocol for wireless ad hoc networks," in *Proc. ACM MOBIHOC*, 2003, pp. 153–164.
[15] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 7, pp. 1265–1275, Sep. 1997.
[16] V. Kawadia and P. Kumar, "Power control and clustering in ad hoc networks," in *Proc. IEEE INFOCOM*, 2003, pp. 459–469.
[17] B. Tavli and W. B. Heinzelman, "MH-TRACE: Multihop time reservation using adaptive control for energy efficiency," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 942–953, Jun. 2004.
[18] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*. Chichester, U.K.: Wiley, 1995.
[19] J. Stine and G. de Veciana, "A paradigm for quality of service in wireless ad hoc networks using synchronous signaling and node states," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 7, pp. 1301–1321, Sep. 2004.
[20] J. Elson, L. Girod, and D. Estrin, "Fine-grained network time synchronization using reference broadcasts," in *Proc. ACM OSDI*, 2002, pp. 147–163.
[21] J. Stine, G. de Veciana, K. H. Grace, and R. C. Durst, "Orchestrating spatial reuse in wireless ad hoc networks using synchronous collision resolution (SCR)," *J. Interconnection Networks*, vol. 3, no. 3–4, pp. 167–198, Sep.–Dec. 2002.
[22] A. Muqattash and M. Krunz, "A single-channel solution for transmission power control in wireless ad hoc networks," in *Proc. ACM MOBIHOC*, 2004, pp. 210–221.
[23] R. Rozovsky and P. R. Kumar, "SEEDEX: A MAC protocol for ad hoc networks," in *Proc. ACM MOBIHOC*, 2001, pp. 67–75.
[24] E. S. Sousa and J. A. Silvester, "Optimum transmission ranges in a direct-sequence spread-spectrum multihop packet radio network," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 762–771, Jun. 1990.
[25] L. Kleinrock and J. Silvester, "Spatial reuse in multihop packet radio networks," *Proc. IEEE*, vol. 751, no. 1, pp. 156–167, Jan. 1987.
[26] L. Bao and J. J. Garcia-Luna-Aceves, "Transmission scheduling in ad hoc networks with directional antennas," in *Proc. ACM MOBICOM*, 2002, pp. 48–58.



Xiangying Yang (S'01–M'05) received the B.S. degree in electrical engineering from Tsinghua University, China, in 1998 and the M.S. and Ph.D. degrees, both in electrical and computer engineering, from the University of Texas at Austin in 2000 and 2005, respectively.

He is now a Research Scientist/System Engineer with Intel Corporation, Mobility Group Wireless Standard and Technology Organization in Hillsboro, OR. His research interests include wireless broadband access networks, end-to-end performance and mobility management, wireless mesh networks, web caching and peer-to-peer applications.

Dr. Yang was a recipient of the Texas Telecommunications Engineering Consortium (TxTEC) Fellowship in 2000 and 2002.



Gustavo de Veciana (S'88–M'94–SM'01) received the B.S., M.S, and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1987, 1990, and 1993, respectively.

He is currently a Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin. His research focuses on the design, analysis, and control of telecommunication networks. His current interests include: measurement, modeling and performance evaluation; wireless and sensor networks; architectures and algorithms to design reliable computing and network systems.

Dr. de Veciana has been an editor for the IEEE/ACM TRANSACTIONS ON NETWORKING. He was the recipient of General Motors Foundation Centennial Fellowship in Electrical Engineering and a 1996 National Science Foundation CAREER Award, co-recipient of the IEEE William McCalla Best ICCAD Paper Award for 2000, and co-recipient of the Best Paper in *ACM Transactions on Design Automation of Electronic Systems*, January 2002–2004.