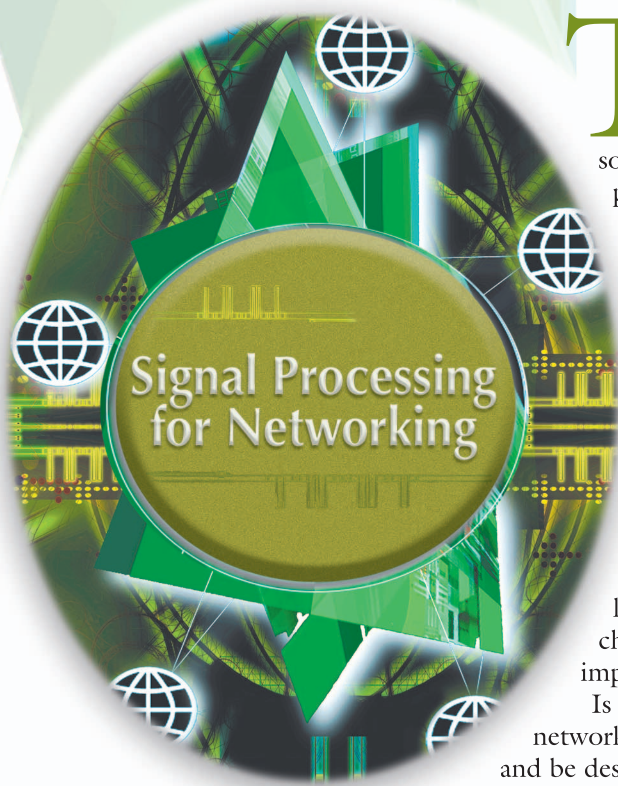
A cross-layer perspective in an uncharted path

Signal Processing in Random Access

Lang Tong, Vidyut Naware, and Parvathinathan Venkitasubramaniam



The design of medium access control (MAC) protocols has traditionally been separated from that of the physical (PHY) layer. To a MAC protocol designer, the PHY layer is a black box satisfying the so-called collision model: when only one user transmits, the packet arrives at the receiving node error free. But when transmissions are simultaneous, packets are lost due to collision. Until recently, the theory of random access was based on such an idealized model, and random access protocols were viewed as collision resolution or collision avoidance techniques.

In practice, the collision model is both optimistic and pessimistic: optimistic, for it ignores channel effects such as fading and noise on reception, and pessimistic, because it does not accommodate the possibility that packets may be successfully decoded in the presence of simultaneous transmissions. Given the advances in multiuser communications at the PHY layer, the collision model no longer represents all the characteristics of the PHY layer, missing some of its most important properties.

Is there a need to go beyond the collision model for wireless networks? Should the MAC layer assume a multiuser PHY layer and be designed with a cross-layer principle in mind? Is the gain of a cross-layer design significant enough to justify replacing a well-tested protocol with a more sophisticated one? Will the cross-layer design be too complicated to implement, and too sensitive to channel changes to be useful?

The idea of cross-layer design has been brought to the fore by the phenomenal growth in wireless applications and a continuing push for broadband access. The fundamental challenge, as noted by Gallager in 1985 [1] and more recently by Ephremides

1053-5888/04/$20.00©2004IEEE

and Hajek [2], lies in the choice of a proper model that interfaces the PHY layer and network layers. From an information theoretic viewpoint, multiaccess communications can be made error free via error control so long as the rates fall within the capacity region. The multiaccess channel is "smooth" in the sense that bits continuously flow from each user to the receiver; retransmissions are not necessary. From a network theoretic perspective, on the other hand, there is the basic notion of a packet. The multiaccess channel is "hard" in the sense that a packet either gets through or is lost in a collision; collisions need to be resolved by retransmissions. A major point of Gallager's paper [1] is that, in his words, "a better set of models and approaches are needed for multiaccess communication than collision resolution or information theory alone."

In this article, we consider the interactions between the PHY and MAC layers where there are reasonable models that interface the two. Specifically, we focus on roles of signal processing in random access networks: How does a multiuser PHY layer enabled by advanced signal processing affect the design of random access protocols? If a node is able to estimate the channel state via signal processing, how should that information be exploited at the protocol layer for higher efficiency?

We present two cases where cross-layer design has a quantifiable impact on system performance. The first is a small network (such as wireless LAN) where a few nodes with bursty arrivals communicate with an access point. The design objective is to achieve the highest throughput among users with variable rate and delay constraints. We examine the impact of PHY layer design on MAC protocols and illustrate a tradeoff between allocating resources to the PHY layer and to the MAC layer. The second case, in contrast, deals with large-scale sensor networks where each node carries little information but is severely constrained by its computation and communication complexity and, most importantly, battery power. The goal there is not necessarily achieving the highest throughput nor is it about assuring quality of service. The design must take into account the need of sensors transmitting with diminishing power. Here, the cross-layer design takes a different form where we illustrate the need to incorporate PHY layer parameters into MAC protocols.

A Brief Historical Perspective

The story of random access began with Abramson's landmark work [3]. In designing a multiaccess scheme that allows terminals in different islands access a central unit through a common wireless channel, Abramson devised a simple and elegant solution known as ALOHA: transmit when the terminal has a packet; back off randomly if the transmission is not successful. Transmissions fail because multiple users transmit simultaneously, and the randomized back-off resolves collisions probabilistically in a distributed fashion.

Throughput

The analysis of ALOHA by Abramson is equally remarkable. In its simplest form, it makes two fundamental assumptions that led to the magical throughput figure e^{-1}. The first is the collision model: any simultaneous transmissions lead to irrecoverable failure. In the second assumption, the aggregated traffic including both transmissions and retransmissions is Poisson. If the transmissions are slotted, then the average number of successfully received packets, the throughput λ, is given by

$$\lambda(G) = \Pr(\text{one and only one user transmits}) = Ge^{-G},$$

where G, the offered load, is the mean arrival of the aggregated traffic (packets/slot). (The slotted ALOHA is a variation by Roberts [4] of the original ALOHA by making all transmissions follow the same slot structure.) If each user chooses the retransmission probability wisely so that the offered load can be regulated to maximize $\lambda(G)$, we have

$$\lambda_{\text{ALOHA}} = \max_G \lambda(G) = e^{-1} \approx 0.36.$$

The Poisson assumption implicitly assumes the model in which infinitely many new arrivals and retransmissions, each with infinitesimally small rate, jointly form a traffic with mean G. In spite of the intuitive appeal of such an infinite population model, the offered traffic cannot be Poisson, and the model at best is a rough approximation under special circumstances. A variation of this assumption is the Bernoulli model with a finite number of users [3], where, in an M-user system, user i transmits with probability p_i. Again, under the collision model and the assumption that all users have packets waiting to be transmitted, the average number of successfully-received packets is given by

$$\begin{aligned}\lambda_M(\mathbf{p}) &= \Pr(\text{one and only one user transmits})\\ &= \sum_i p_i \prod_{j\neq i}(1-p_j).\end{aligned}$$

If all users are statistically the same and each user transmits with probability $p_i = G/M$ (so that the total traffic has mean G), then the throughput is given by

$$\lambda_M(G) = G\left(1-\frac{G}{M}\right)^{M-1} \to Ge^{-G} \text{ as } M\to\infty.$$

Again we arrive at the same expression and the same maximum throughput figure.

Stability

Hidden in the above throughput analysis is the notion of stability. For the finite user model, had each node stored infinitely many packets to start with, the throughput figure above would indeed have been the maximum number of packets that can be extracted by the central unit. But if each user's buffer size is finite, then we are faced with the constraint that the number of packets stored in each user's buffer must not go unbounded. To avoid buffer overflow, we need to reduce the arrival rate to some equilibrium point above which overflow is inevitable but below which the probability of overflow can be made arbitrarily small. Typically, the probability of a stable queue being empty is nonzero. (This is true for queues forming an irreducible and aperiodic Markov chain.) Thus, for a stable random access protocol, we have to deal with the case when there are no packets to transmit from some of the users. For the infinite population model, stability is closely connected with the number of users waiting to transmit their packets (backlogs). It turns out that, if each user retransmits with fixed probability in an infinite user system, a simple drift analysis [5] shows that the stable throughput is zero! While in practice, the network instability can be coped with by rebooting the network and flushing out all backlogged packets, this sharp discrepancy between 0 and $1/e$ underscores the need for a fundamental understanding of the queuing dynamics of random access protocols.

The failure to stabilize ALOHA in an infinite user system has a simple intuitive explanation: the greater the number of users having packets to transmit, the smaller their transmission probability must be to make the system stable. If each user transmits with a constant probability and by chance there are more users than the transmission probability can handle, more users will be backlogged and the system will become unstable. But what if we can estimate the number of users having packets to transmit by looking at how often collisions occur and adjust transmission probability accordingly? It was shown later that the maximum stabilized throughput is, almost miraculously, what was obtained in the first place: $1/e$ [6].

It was not until 1979 that Tsybakov and Mikhailov [7] tackled the stability issue for the finite user ALOHA rigorously. Under the collision model, they obtained, for the two-user case, the complete characterization of the stability region of slotted ALOHA. This region, shown in Figure 1, clearly demonstrates the suboptimality of ALOHA when compared with the deterministic scheduling protocol time division multiple access (TDMA). By time sharing the common channel, TDMA makes any rate pair on the line connecting the two single user rates (point A and B in Figure 1) stable. For ALOHA, on the other hand, the stability region is not convex, which means that if one user is to increase his rate, the other user may have to sacrifice disproportionally.

Unfortunately, the characterization of the stability region of ALOHA for the general M-user case remains an open problem with only inner and outer bounds available [8]–[10]. An exception is the symmetrical case when all users have the same rate. In that case, Tsybakov and Mikhailov [7] showed that an arrival rate $\lambda(M)$ is stable if

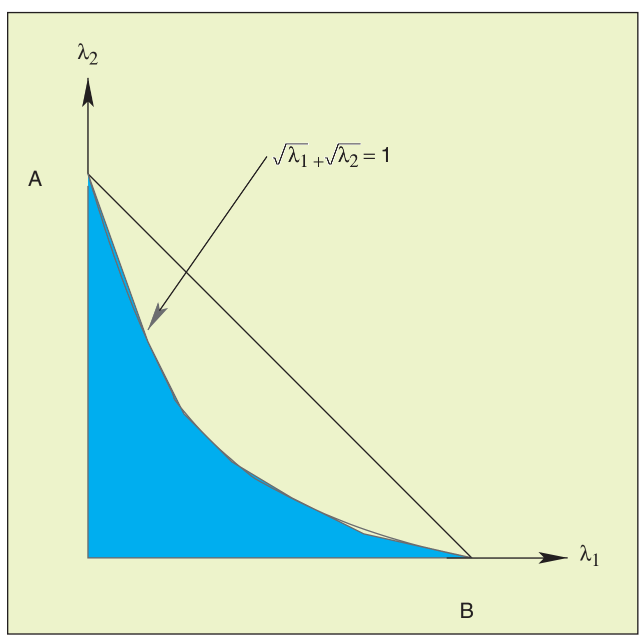
$$\lambda(M) < \left(1 - \frac{1}{M}\right)^{M-1}$$

and unstable if the inequality is reversed. Letting $M \to \infty$, we again obtain the throughput $1/e$. It should be pointed out that the limiting case of the finite user model does not represent the infinite population model even though the maximum stabilizable throughput is the same.

Collision Resolution

A breakthrough was made by Capetanakis [11] and independently by Tsybakov and Mikhailov by recognizing the fact that the sequence of collision events tell us something about the state of the network, and inferences can be made about who may or may not have packets. Using the idea of splitting the transmissions of collided users, Capetanakis developed a random access protocol that offers a stable throughput of 0.43. This discovery ignited an intensive search for stable random access protocols that achieve the maximum throughput. Despite having established upper and lower bounds on the achievable throughput, the maximum stable throughput remains elusive to this day.

The idea of making inferences about the state of the network, i.e., the users who have packets to transmit, was crystallized by Berger et al. [12] and Wolf [13] when they found a connection between collision resolution and classical group testing. During World War II, the U.S. military needed to administer syphilis tests for millions of inductees. Given the probability that a person having the disease is small, it would have been grossly inefficient had each person been tested individually (in a TDMA fashion). According to Wolf, Dorfman [14] suggested pooling blood samples from a group of persons and applying the



▲ *1. The stability region of a two-user ALOHA system: $A = B = 1$.*

so-called Wasserman test. The test was sensitive enough to give a positive reading if and only if one or more among the tested group had the disease. Thus, persons with the disease were identified by a sequence of tests referred to as the test plan. The connection between group testing and collision resolution is therefore natural: the person with a packet to transmit is a user with the disease (to talk, perhaps?); to enable a set of users to transmit is to perform a group testing; a collision in the transmission means a positive test, silence the negative. An efficientcollision resolution algorithm is an efficient group testing plan that, on the average, has the smallest number of tests. It turns out that many ideas developed for random access protocols were directly related with statistical methods of devising group testing plans [15].

The connection between group testing and collision resolution tells us more than the protocol side of the story; it also gives us insight into the difficulty of extending the collision model to more elaborate models discussed in following sections. For example, were the outcome of a (group) test random, repeatable only by the law of large numbers, the test plans in [15] would not have worked, nor would the collision resolution protocols under probabilistic reception models. This is one of the reasons that the extension to multipacket and probabilistic receptions is much more challenging, requiring perhaps an entirely new way of treating random access.

Multipacket Reception: A Step Toward Cross-Layer Design

The collision model represents a simplistic PHY layer that leaves the MAC layer to handle the difficult task of separating users via scheduling. But there is no fundamental reason that collided transmissions cannot be recovered by other means such as coding and signal processing. In the information theoretic setting for a multiaccess channel, the capacity region is achieved by making all users transmit at the same time, and it is the decoder that is responsible for untangling each user's transmission.

The advent of multiaccess techniques such as code division multiple access (CDMA) and multiuser detection led to a new examination of random access under a multiuser PHY layer. In 1988, Ghez et al. [16] made a fundamental change in the collision model that has been the foundation of virtually all PHY protocols. They offered the generalization that, when there are simultaneous transmissions, the reception can be described by conditional probabilities (instead of deterministic failure). They proposed the multipacket reception (MPR) model defined by the MPR matrix

$$\mathbf{C} = \begin{bmatrix} C_{10} & C_{11} & & \\ C_{20} & C_{21} & C_{22} & \\ \vdots & & & \ddots \end{bmatrix}, \qquad (1)$$

where C_{ij} is the conditional probability that, given i users transmit, j out of i transmissions are successful. Given k users transmit at the same time, the average number of successfully received packets is given by

$$C_k = \sum_j jC_{kj}.$$

Note that the model used by Ghez et al. captures only certain symmetrical channel models. Specifically, C_{kj} models only j success out of k transmissions but not which j transmissions are successful. Such a model is inadequate for some practical spatial diversity systems in which one set of j users may be quite different from the other. A more detailed model described later is needed.

The MPR model describes a multiuser PHY layer for random access. Focusing on ALOHA under MPR model, Ghez et al. illustrated that a major reason for the infinite user ALOHA to have zero stable throughput for any fixed retransmission probability comes from the classical collision model. In particular, they showed that ALOHA under MPR achieves stable throughput $\lim_{k\to\infty} C_k$ assuming the limit exists. The significance of that finding became even greater when Zorzi and Rao [17] and later Hajek [18] showed that under the capture model in fading, $\lim_{k\to\infty} C_k$ does exist and is positive. The capture model assumes that if a user's signal to interference and noise ratio (SINR) exceeds a threshold β, then his packet will be successfully received. For a very general class of fading models where the fading has distribution with rolloff factor δ, Hajek showed succinctly that

$$\lim_{k\to\infty} C_k = \begin{cases} \beta^{-\delta}\frac{\sin(\pi\delta)}{\pi\delta}, & 0 < \delta < 1, \\ 0, & \delta \geq 1. \end{cases}$$

The contribution of Ghez et al. is more than a technical one. By allowing a much broader class of PHY layers to be considered together with random access, they brought to light the interaction between PHY and MAC layers. Their work begged the question of how to take advantage of the enhanced PHY layer that can separate users and whether the notion of collision resolution is still a valid one.

Allowing MPR at the PHY layer changed the random access problem considerably. For example, the inference from the reception outcome of a slot is no longer simple. In the collision model, if a packet is successfully received, it implies that no one else transmitted in that slot. The probabilistic modeling of MPR means that the state of a user cannot be inferred with certainty until either an empty slot occurs when the user is scheduled to transmit or a packet from that user is received successfully; the successful reception of other users' packet does not imply that this user has not transmitted. It is such uncertainty that makes the splitting idea, so crucial in the classical collision resolution protocols, difficult to apply for MPR channels.

While splitting the transmissions of colliding users according to collision events is not directly applicable to the MPR model, the idea of inferring the state of the users and scheduling their transmissions is still valid. In [19], the so-called service room protocol is developed that performs a posteriori estimates of network state based on the entire reception history. The protocol then grants an optimal subset of users access to the MPR channel. The dynamic queue protocol [20] is a random access scheme designed also for the MPR channel, which offers a much simpler implementation and limited performance degradation. By and large, however, exploiting MPR for random access is a research area with many open theoretical and practical problems.

Signal Processing Versus Scheduling: A Cross-Layer View

In this section, we revisit the classical work of Tsybakov and Mikhailov [7] and Rao and Ephremides [8] on finite user ALOHA, this time under a general multipacket reception model. We know that, for the single user collision model, centralized scheduling is superior to a simple and distributed protocol such as ALOHA. At the other extreme, if the PHY layer completely decouples simultaneous transmissions, for example using orthogonal CDMA, then there is little advantage having any scheduling. It is not surprising that putting resources in the PHY layer will simplify the MAC design and vice versa. It is more instructive, however, to see where and how such transitions happen. Details of our discussion here can be found in [21].

A General MPR Model for Packet Switched Multiaccess

The MPR model of Ghez et al. has several limitations. It assumes a symmetrical model with indistinguishable users analogous to the the classical urn model with indistinguishable balls. Such a model does not allow the specification of a region where each user has his own rate. We now describe a more general MPR model that differentiates users. Such a generalization is necessary for systems with spatial diversity because users in different geographical locations interfere with each other differently.

We consider a wireless LAN of N users communicating with an access point as illustrated in Figure 2. The access point may have multiple antennas for beamforming or for other diversity techniques, thus allowing simultaneous transmissions to be received. The ith user generates packets at the rate of λ_i. Each user has a buffer for arriving and backlogged packets. The channel is slotted, and the slot duration equals the packet transmission time. The state of the queue in slot t is defined by the number of packets Q_i^t in the queue at the beginning of the slot.

For a system involving a set of N users $\mathcal{U} = \{1, 2, \ldots, N\}$, we consider a multiuser PHY layer defined by a set of conditional probabilities. For any subset of users $\mathcal{T} \subseteq \mathcal{U}$ transmitting in a slot, the probability of successfully receiving packets from $\mathcal{R} \subseteq \mathcal{T}$ is given by

$$P_{\mathcal{R},\mathcal{T}} = \Pr\{\text{only packets from users in } \mathcal{R} \text{ are successfully received}|\text{users in } \mathcal{T} \text{ transmit}\}. \tag{2}$$

$\mathcal{P} = \{P_{\mathcal{R},\mathcal{T}},\ \mathcal{T} \subseteq 2^{\mathcal{U}},\ \ \mathcal{R} \subseteq \mathcal{T}\}$, the set of conditional probabilities completely specifies the probability space for reception. Note also that all PHY layer characteristics are summarized by $\mathcal{P}$ in the sense that, for each PHY configuration, there is a corresponding probability space $\mathcal{P}$.

It turns out that [21] the achievable rate and the stability regions depend only on a set of marginal conditional probabilities

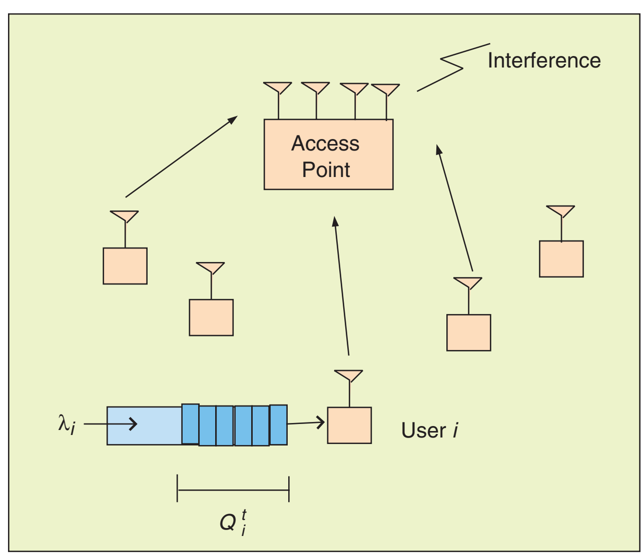
$$P_{i|\mathcal{T}} \triangleq \sum_{\mathcal{R}\subseteq\mathcal{T}:i\in\mathcal{R}} P_{\mathcal{R},\mathcal{T}},$$

where $P_{i|\mathcal{T}}$ is the probability that the ith user is successfully received given users in $\mathcal{T}$ transmit. The achievable rate for user i should depend on the probability $P_{i|\mathcal{T}}$ of its own successful transmission. However, since interference from simultaneous transmissions results in a system of interacting queues, it is less obvious that only the marginal probabilities are needed to specify the rate region.

MAC Capacity and Stability Characterizations

MAC Capacity

We begin by allowing centralized scheduling: the access point knows the complete network state including the queue state of each user. We use the notion of MAC capacity in the sense of Abramson [3] and Kleinrock [22], [23]. Assuming each node always has packets for



▲ 2. *A wireless LAN.*

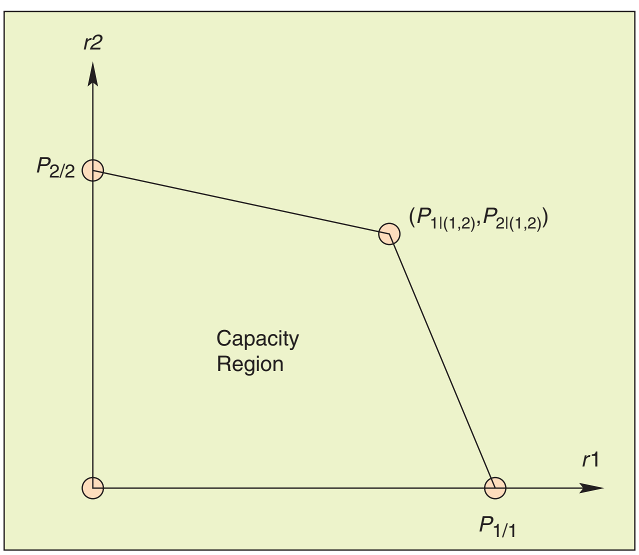
transmission, we define the MAC capacity region as the set $\mathcal{C}$ of (departure) rate vectors such that for each rate vector $\mathbf{r} = [r_1, \cdots, r_N] \in \mathcal{C}$, there exists a MAC protocol with which the rate of successfully received packets from user i is higher than r_i. For the N user wireless LAN system, under the reception model given in (2), the MAC capacity region can be obtained as a special case of [24]. In words, the capacity region is given by the convex hull of points $(\mathbf{s}(\mathcal{T})), \mathcal{T} \subseteq \{1, 2, \cdots N\}$ whose coordinates are given by

$$s_i(\mathcal{T}) = \begin{cases} P_{i|\mathcal{T}}, & \text{if } i \in \mathcal{T}, \\ 0, & \text{otherwise.} \end{cases}$$

Illustrated in Figure 3 is the two-user case. Notice that the capacity region is bounded by straight lines, a fact from taking the convex hull of vertices involving marginal reception probabilities. The linear boundaries have the implication of scheduled time-sharing among users. The MAC capacity region resembles the Shannon capacity region for multiaccess channels [25] except that the linear boundaries are slanted due to the lack of perfect packet separation.

MAC Stability

A primary concern in MAC design is stability. By stability we mean that the probability of buffer overflow can be made arbitrarily small by making the buffer size sufficiently large [26], [7]–[9]. (Let $\mathbf{Q}^t = (Q_1^t, \dots, Q_N^t)$ be the queue lengths of an N user system. The system is stable if for $\mathbf{x} \in \mathbb{N}^N$, $\lim_{t\to\infty} \Pr\{\mathbf{Q}^t < \mathbf{x}\} = F(\mathbf{x})$ and $\lim_{\mathbf{x}\to\infty} F(\mathbf{x}) = 1$.) The stability region is a set $\mathcal{S}$ of (arrival) rate vectors such that for each rate vector $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_N] \in \mathcal{S}$, there exists a MAC protocol that makes all queues stable. The stability region for a particular MAC protocol is the set of rates that are stable under that MAC. For example, the stability region of ALOHA, $\mathcal{S}_{\text{ALOHA}}$ is the set of all arrival rate vectors that are stable under ALOHA.



▲ 3. *Capacity region of a two-user system.*

The capacity region $\mathcal{C}$ and the stability region $\mathcal{S}$ are not always the same [24]. In general, $\mathcal{S} \subseteq \mathcal{C}$. Only in rare cases does the stability region of a specific MAC protocol coincide with the capacity region. For example, under the collision channel model, the MAC capacity $\mathcal{C}$ of a two-user system is the convex hull obtained from three rate vectors $(0, 0)$, $(0, P_{2|2})$, $(P_{1|1}, 0)$. The stability region of ALOHA [7] under the collision channel, in contrast, is $\mathcal{S}_{\text{ALOHA}} = \{(\lambda_1, \lambda_2), \lambda_i \geq 0, \sqrt{\lambda_1} + \sqrt{\lambda_2} < 1\}$, which is strictly inside $\mathcal{C}$.

ALOHA with Spatial Diversity May Achieve MAC Capacity

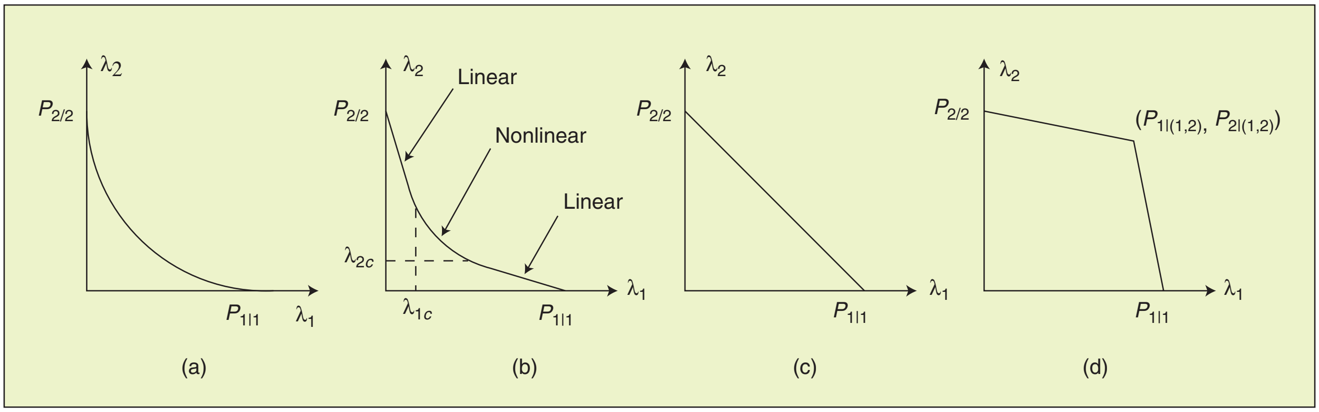
The ALOHA protocol can be parameterized by a vector of transmission probabilities $\mathbf{p} = [p_1, \cdots, p_N]$. Node i transmits a packet with probability p_i if its queue is not empty. If $\boldsymbol{\lambda} \in \mathcal{S}_{\text{ALOHA}}$, then there exists a $\mathbf{p}$ such that all queues are stable. The ALOHA stability region for an N user system is unknown in general. For the collision channel, characterizing stability region has been a long standing open problem. There is, however, a simple case when an antenna array is used. If a zero-forcing antenna array is used to eliminate all interfering nodes, we then have N independent MAC channels, albeit each may have a much smaller stability region due to noise enhancement of the zero-forcing operation.

The generalization of the work of Tsybakov and Mikhailov [7] for MPR channels reveals important features masked previously by the collision model. It is shown in [21] and depicted in Figure 4 that there are only four possible stability regions, each corresponding to a different level of MPR capability: Figure 4(a) is the collision channel case with no MPR. The instability region is convex. Figure 4(b) corresponds to weak MPR where the point $(P_{1|(1,2)}, P_{2|(1,2)})$ is below the line connecting $(0, P_{2|2})$ and $(P_{1|1}, 0)$. The stability region in this case is bounded by linear boundaries in the low interference regions below certain critical rates λ_{ic} and by a quadratic curve (in $\sqrt{\lambda}$) in the strong interference region. Figure 4(c) is the critical MPR case when the stability region, for the first time, becomes convex and is bounded by straight lines. This case happens when $(P_{1|(1,2)}, P_{2|(1,2)})$ is on the line connecting $(0, P_{2|2})$ and $(P_{1|1}, 0)$. Figure 4(d) shows the strong MPR case when the capacity region is convex and bounded by straight lines. This happens when $(P_{1|(1,2)}, P_{2|(1,2)})$ is above the line connecting $(0, P_{2|2})$ and $(P_{1|1}, 0)$. To summarize, we have the following theorem.

Theorem 1 [21]

Assume that $P_{1|1} \geq P_{1|(1,2)}$, $P_{2|2} \geq P_{2|(1,2)}$. Then, the stability region of ALOHA coincides with the MAC capacity region, i.e., $\mathcal{S}_{\text{ALOHA}} = \mathcal{C}$ if and only if the following critical MPR condition is satisfied:

$$P_{2|(1,2)} \geq \left(P_{1|1} - P_{1|(1,2)}\right) \frac{P_{2|2}}{P_{1|1}}. \tag{3}$$



▲ 4. *The stability regions for different multiuser PHY layers. (a) The collision model. (b) The weak MPR model. (c) The critical MPR model. (d) The strong MPR model.*

The message to system designers from Theorem 1 is unmistakable: if one can provide MPR capability at the PHY layer beyond the critical level [Figure 4(c) and (d)], the optimal MAC layer is none other than the simplest—the ALOHA. There are several ramifications. First, it shows that no centralized scheduling is needed, and the best MAC is ALOHA. This appears to be the second instance in which the ALOHA stability region coincides with a certain capacity region; the first case was discovered by Massey and Mathys [27] who showed that, for the collision channel, the ALOHA stability region coincides with the (zero-error) capacity region of the collision channel without feedback. See also the work of Anantharam [28]. Second, the convexity of the ALOHA stability region has special theoretical significance. Such a property assures that any rate vector between two stable rate vectors is stable; the rate increases of some users can be accommodated by proportional rate decreases of others, which is useful for rate allocations in a dynamic environment.

But can the result in Theorem 1 be extended to the more general N user case? We do not yet know. We suspect that the complete characterization will be difficult for general MPR models. Still, there are reasons for optimism. For the general N user ALOHA when all users generate packets at the same rate, i.e., the symmetrical case, ALOHA is again optimal under strong MPR conditions. Next, if it is possible to decouple all users by zero forcing, ALOHA will again achieve capacity. One suspects that ALOHA remains optimal under strong MPR conditions. Nonetheless, inner bounds for the stability region can be developed [21] following the techniques by Rao and Ephremides [8], Lou and Ephremides [10], and Szpankowski [9].

Exploiting Queue Statistics in Signal Processing

We have seen that signal processing techniques that offer a multiuser PHY layer can have a significant impact on queuing behavior of random access protocols. We now examine this interaction in the reverse direction and explore the possibility of using queue statistics for enhanced signal processing.

As an example, we consider the problem of designing a beamformer that maximizes the stability region of ALOHA, an unconventional metric that involves cross-layer issues. As illustrated in Figure 5, the wireless channel is modeled as

$$\mathbf{y}[t] = \mathbf{H}[t]\mathbf{s}[t] + \mathbf{n}[t],$$

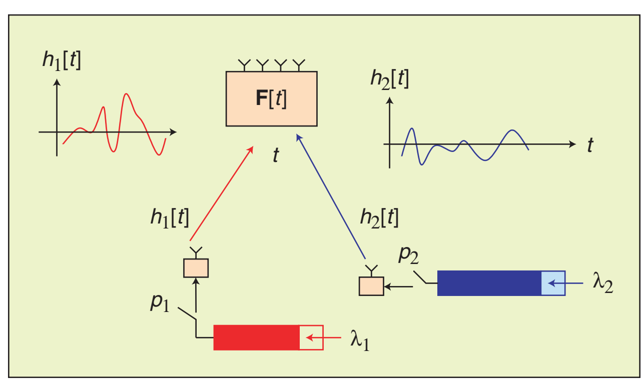
where t is the slot index, $\mathbf{y}[t]$ the received vector at the antenna elements, $\mathbf{H}[t]$ the channel matrix that models a MIMO random fading channel, $\mathbf{s}[t]$ the vector of user's transmitted signal, and $\mathbf{n}[t]$ the additive noise. We consider the class of adaptive linear beamformers $\mathbf{F}[t]$

$$\hat{\mathbf{s}}[t] = \mathbf{F}^H[t]\mathbf{y}[t],$$

where $\hat{\mathbf{s}}[t]$, assumed to be a stationary ergodic process, is the signal estimate. Under the SINR threshold model, conditional on $\mathbf{H}[t] = \mathbf{H}$, a packet of user i is successfully received if its SINR $\gamma_i(\mathbf{H})$ exceeds a threshold β_i. The choice of β_i depends, of course, on the coding and decoding strategy used for user i.

Given a set of transmitting users, and a beamformer $\mathbf{F}(t)$, the probability space for the reception is given by $\mathcal{P}^{\mathbf{F}} = \{P^{\mathbf{F}}_{i|\mathcal{T}}\}$ as defined in the previous section by

$$P^{\mathbf{F}}_{i|\mathcal{T}} = \Pr(\gamma_i(\mathbf{H}) \geq \beta_i).$$



▲ 5. *Beamforming in MAC.*

Thus, for every front end, there corresponds an ALOHA stability region ($\mathcal{S}_\mathbf{F}$). For a class $\mathcal{F}$ of beamformers, we define the stability region as

$$\mathcal{S} = \bigcup_{\mathbf{F}\in\mathcal{F}} \mathcal{S}_\mathbf{F}.$$

We can then evaluate a number of design strategies by examining their stability region. For this we assume a special model where the MIMO channel is given by $\mathbf{H}[t] = \mathbf{V}\mathbf{G}[t]$ where $\mathbf{V}$ is the deterministic and time invariant array response, and $\mathbf{G}[t] = \mathrm{diag}(G_1[t], \cdots, G_N[t])$ is a diagonal matrix with the ith entry modeling the fading gain of user i. We assume that $\mathbf{V}$ is known at the receiver but the fading $G_i[t]$ varies from slot to slot.

▲ *Beamforming via matched filtering*. This is perhaps the simplest form of beamforming. Without access to queue and the fading states, the receiver beamformer is the simple nonadaptive matched filter $\mathbf{F} = \mathbf{V}^H$. The matched filter approach is, of course, not optimal unless there is only one user transmitting.

▲ *Pseudo-MMSE beamformer*. If the receiver assumes that all users have packets in their queues, a minimum mean-squared error receiver beamformer can be implemented either coherently (when fading states are known) or noncoherently (when the fading states are unknown). For the noncoherent beamformer, the receiver matrix $\mathbf{F}$ does not vary from slot to slot. Because the queues of the users are occasionally empty, such a beamformer is not the true MMSE beamformer.

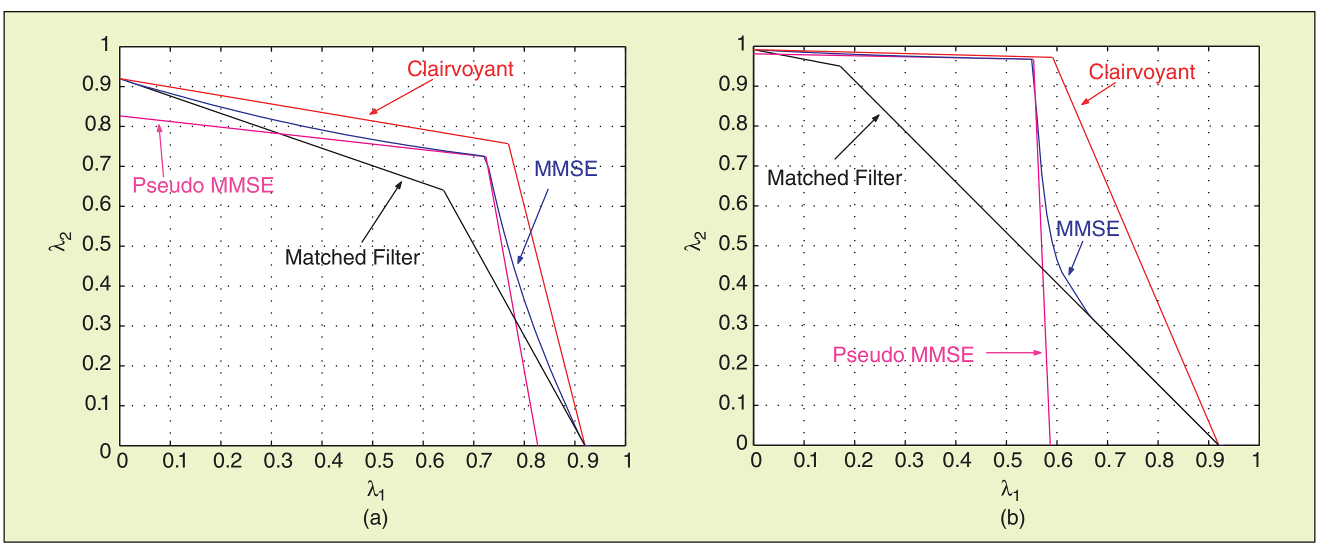
▲ *MMSE beamformer*. If we do not have access to the queue state but know instead the probability that a queue is empty, then the true MMSE beamformer can be implemented [29].

▲ *The clairvoyant MMSE beamformer*. We begin by making the assumption that the access point knows everything about the system from channel fading to queue state of each user. Unrealistic as it is, the "clairvoyant" receiver maximizes the SINR for each user and gives an outer bound on the stability region. Since fading and queue state both vary from slot to slot, the clairvoyant MMSE beamformer adaptively changes its weight in each slot.

Figure 6 illustrates two cases in which these beamformers are used. In (a) is the case in which the two users have equal power with well-conditioned channels. In this case, the simple nonadaptive matched filter receiver offers nearly as large a stability region as the clairvoyant beamformer. This indicates that the gain of using sophisticated signal processing may not justify the complexity of implementations. In (b) is the case when one user has significantly stronger power then the other. The difference among these receivers becomes evident. The simple matched filter offers a substantially smaller stability region than that of the clairvoyant beamformer. However, if both users operate in the lower-left corner of the rate region, the matched filter beamformer will perform reasonably well because it is often the case that only one user has packets to transmit. The pseudo-MMSE that assumes both users always have packets to transmit can perform poorly in the high rate region of the weaker user. In fact, despite its more complex implementation, the pseudo-MMSE becomes unstable when user 1 exceeds the rate of 0.6 packets/slot. In this region, the use of queue statistics becomes crucial. Note also that when queue statistics are used, the stability region of the true MMSE receiver is not much smaller than the clairvoyant receiver.

Incorporating Channel State in Random Access

A prime example of cross-layer design is the use of channel-state information in random access. Classical random-access protocols assume a static PHY layer. Even the MPR channel model that reveals more PHY layer characteristics is still a black box to the MAC layer. For wireless networks, channel fluctuations at the PHY layer provide



▲ *6. Stability regions of several beamformers. (a) Equal power users. (b) Unequal power users (near-far effects). User 2 is the stronger one.*

valuable information for random access. For example, if a user is in a deep fade and its transmission has little chance of being decoded successfully, then it is better that the user not transmit and wait for a better channel state.

We discuss in this section the use of channel state in random access, specifically in ALOHA. The idea of using channel state information was sparked by the work of Knopp and Humblet [30] where they showed that, under the information theoretic setting, maximizing the sum-rate under the average power constraint leads to scheduling only the best user to transmit. Exploiting channel state information induces multiuser diversity, and the performance improves with the increase of the number of users [31], [32]. Decentralized use of channel state was investigated by Telatar and Shamai under the metric of sum-rate [33]. Qin and Berry [34] proposed the use of channel state information to vary the transmission probability in ALOHA. They analyzed the system under the collision model, and demonstrated the effect of multiuser diversity on the throughput.

Opportunistic ALOHA

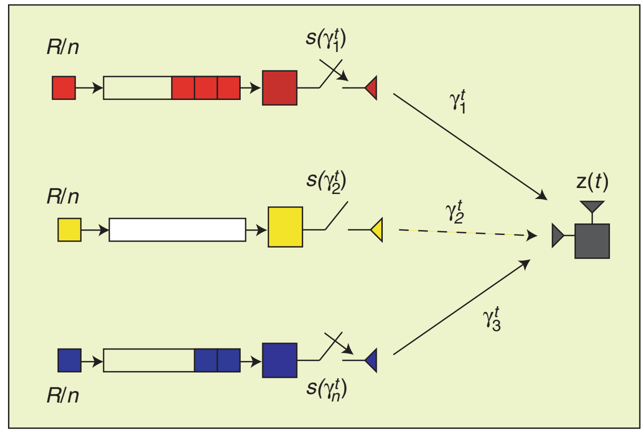
Consider again the ALOHA random access shown in Figure 7 where the channel between the node and the access point has random fading gain γ with cumulative distribution $F(\gamma)$. Suppose that the access point broadcasts a beacon, and each user measures the signal strength γ. Assuming reciprocity, the node transmits with probability $s(\gamma)$, which is a function of its own channel state. What then should be the best transmission probability $s(\cdot)$ that maximizes the stable throughput?

The use of channel state at the transmitter creates an "illusion" at the receiver. For example, if all the users transmit only if their channel gain is greater than a threshold γ_0, then the receiver will never receive signals from those users with poor channel states, and the reception performance is only determined by the group of strong users (users with channel state greater than γ_0). In other words, by allowing the transmission probability to depend on the channel state, the prior channel distribution $F(\gamma)$ given by nature is shaped into a different a posteriori channel state distribution $G(\gamma)$. Conditioned on a user transmitting [with probability $s(\gamma)$], the a posteriori distribution (probability density function as viewed by the receiver) of the channel is

$$dG(\gamma) = \frac{dF(\gamma)s(\gamma)}{\int s(\gamma)dF(\gamma)}.$$

The key point is that the reception performance is determined not by the prior channel distribution F but by the a posteriori distribution G, which can be designed with a judicious choice of $s(\gamma)$.

For example, for the MPR reception model given in (1), the average number of successfully received packets given that k users transmit is then $C_k(G)$, which is a function of the a posteriori distribution G that can be manipulated by the transmission probability $s_n(\gamma)$,



▲ 7. Opportunistic ALOHA.

where n denotes the size of the network. It can then be shown [35] that an n user ALOHA system is stable if the total arrival rate R satisfies

$$R < \sum_{k=1}^{n} \binom{n}{k} \left(1 - p_{s_n}\right)^{n-k} p_{s_n}^{k} C_k(G) \triangleq \lambda_n(s_n), \qquad (4)$$

where $p_{s_n} = \int_0^\infty s_n(\gamma)dF(\gamma)$ is the unconditional probability that a node transmits. If the inequality is reversed, the system becomes unstable. The design of transmission control $s_n(\gamma)$ affects both p_{s_n} and $C_k(G)$ and therefore the maximum throughput.

Optimizing the maximum throughput (λ_n) for n users over all possible transmission probability functions $[s_n(\cdot)]$, is in general, difficult. For large networks, however, it is more appropriate to consider the asymptotic throughput $\lambda_\infty(s_n) = \lim\inf_{n\to\infty} \lambda_n(s_n)$. For a specific class of transmission controls [35], it can be shown that the maximum asymptotic stable throughput has the form

$$\lambda_\infty = \sup_{x,G} e^{-x} \sum_{k=1}^{\infty} \frac{x^k}{k!} C_k(G), \qquad (5)$$

where x is a design parameter that regulates the average number of transmissions per slot and G the a posteriori channel distribution. Thus, the problem of choosing the optimal transmission probability becomes selecting the best a posteriori channel distribution G. We next apply this strategy to large-scale sensor networks, where the benefit of exploiting channel state for random access becomes evident.

Opportunistic ALOHA in Large-Scale Sensor Networks

Consider the problem of information retrieval in a large-scale sensor network. Suppose that the PHY layer uses orthogonal CDMA with K codes, and each sensor chooses randomly one of the orthogonal codes for transmission. We assume that there is a mobile access point which travels over the sensor network sending beacons to sensors, to collect the observed data as shown in Figure 8. Each sensor synchronizes with the beacon, measures the strength of the beacon γ, and transmits opportunistically with probability $s_n(\gamma)$, where n is the size of the network.

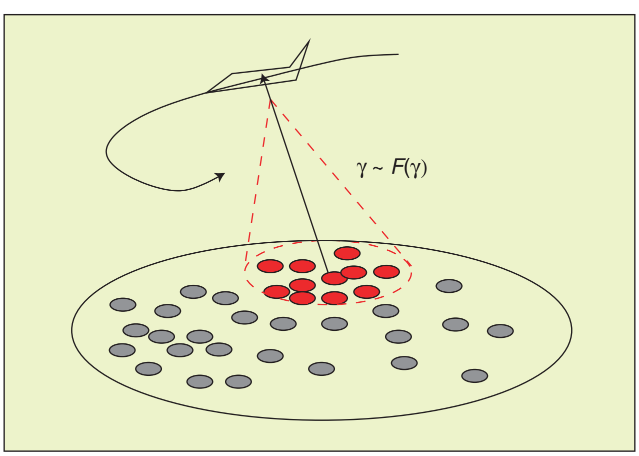
At the mobile access point, a bank of matched filters is used; the transmission of a sensor is successful if its channel gain γ is such that the SINR at the receiver exceeds a certain threshold β. If $\beta > 1$, then at most one packet can be received successfully per orthogonal code [18]. It is therefore clear that the maximum throughput achievable from such a model is at most the spreading gain K. A centralized scheduler could easily achieve such a throughput by choosing the top K users for transmission.

Interestingly, the throughput of K packets per slot can in fact be closely approximated by opportunistic ALOHA. Suppose that the prior channel state has the Rayleigh distribution. By choosing the transmission control

$$s_n(\gamma) = \begin{cases} \min\{\frac{\delta P_T e^{\gamma/P_T}}{\gamma^{\delta+1}} \frac{x}{n}, 1\}, & \gamma > \gamma_0 \\ 0, & \text{otherwise} \end{cases}$$

with sufficiently small γ_0, one can show [35], [36] that the a posteriori channel distribution has a heavy tail distribution with rolloff δ. Again x is the average number of transmissions per slot. By setting δ small and letting the average number of transmitting users x be sufficiently large, there will be a user with channel gain significantly larger than the other transmitting users for each orthogonal code. Thus, one packet will be successfully received for each code, and the overall throughput will be equal to that of the centralized scheduler.

A less obvious result (and one more relevant for low-power sensor networks) is that the throughput K can be achieved with arbitrarily small transmission power P_T. The price paid for this is that there must be enough sensors in the network and the access point must be mobile to provide a rich fading environment. Apart from channel state, the protocol can be used to incorporate other parameters like location into the transmission probability. It is possible to incorporate such information either using the actual realization of the parameter or using its distribution function [36].



▲ *8. Opportunistic ALOHA in sensor networks with mobile access.*

An interesting feature of sensor networks is the possibility of collaborative transmission, where all sensors decide upon the transmission of a particular message. In such a situation, it is possible to design cross-layer coding schemes for reliable transmission by embedding coding in the random access. Examples of such schemes are discussed in [36].

Conclusion

We presented a cross-layer view for roles of signal processing in random access and vice versa. Our discussion is by no means comprehensive and is meant to stimulate interest in what we believe to be a promising research area. (See also [37].) For example, the idea of cross-layer design applies also to the PHY layer described in an information theoretic setting, taking the view that the PHY layer provides an information flow determined by power allocations and scheduling. A particular relevant paper is by Telatar and Gallager [38] where the stability of the multiaccess systems is analyzed using the processor share model. See also the paper by Berry and Yeh [39].

The proliferation of various communication devices makes the sharing of wireless medium and coping with interference from coexisting communication systems inevitable. In a wireless world where connections are ad hoc and users asynchronous, signal processing will be crucial for efficiency and reliability. A key point we hoped to make in this article is that the design of signal processing algorithms must take into account the role of MAC and the nature of random arrivals and bursty transmissions. It is our hope that cross-layer design will provide a few useful ideas as we tread this uncharted path.

Acknowledgments

This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, Army Research Laboratory CTA on Communication and Networks Grant DAAD19-01-2-0011, and National Science Foundation Contract CCR-0311055.

Lang Tong is a professor in the School of Electrical and Computer Engineering, Cornell University. He received the B.E. degree from Tsinghua University, Beijing, in 1985 and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame. He was a postdoctoral research affiliate at the Information Systems Laboratory, Stanford University, in 1991. He was also the 2001 Cor Wit Visiting Professor at the Delft University of Technology. He received Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society. His areas of interest include statistical signal processing, wireless communications, communication networks and sensor networks, and information theory. He is a Senior Member of the IEEE.

Vidyut Naware received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay in 2000. Since then, he has been in the M.S/Ph.D program in the School of Electrical and Computer Engineering at Cornell University. His research interests broadly include wireless communications, information theory and networking.

Parvathinathan Venkitasubramaniam received his B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2002. He joined the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, in 2002, and he is working toward his Ph.D. degree. His research interests include random-access protocols, sensor networks, and information theory.

References

[1] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.

[2] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2416–2434, Oct. 1998.

[3] N. Abramson, "The aloha system–another alternative for computer communications," in *Proc. Fall Joint Comput. Conf., AFIPS Conf.*, Montvale, NJ, 1970, vol. 44, pp. 281–285.

[4] L. Roberts, "Aloha packet system with and without slots and capture," *Comput. Commun. Rev.*, vol. 5, pp. 28–42, Apr. 1975.

[5] R.G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[6] B.E. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Automat. Contr.*, vol. 27, pp. 559–569, June 1982.

[7] B. Tsybakov and W. Mikhailov, "Ergodicity of slotted ALOHA systems," *Probl. Inform. Transmission*, vol. 15, no. 4, pp. 73–87, 1979.

[8] R. Rao and A. Ephremides, "On the stability of interacting queues in a multi-access system," *IEEE Trans. Inform. Theory*, vol. 34, pp. 918–930, Sept. 1988.

[9] W. Szpankowski, "Stability conditions for some multiqueue distributed systems: Buffered random access systems," *Adv. Appl. Probab.*, vol. 26, pp. 498–515, 1994.

[10] W. Luo and A. Ephremides, "Stability of *N* interacting queues in random-access Systems," *IEEE Tran. Inform. Theory*, vol. 45, no. 5, pp. 1579–1587, 1999.

[11] J. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, pp. 505–515, Sept. 1979.

[12] T. Berger, N. Mehravari, D. Towsley, and J. Wolf, "Random Multiple access communications and Group Testing," *IEEE Trans. Commun.*, vol. COM-32, pp. 769–779, July 1984.

[13] J. Wolf, "Born again group testing: Multiaccess communications," *IEEE Trans. Inform. Theory*, vol. 31, pp. 185–191, Mar. 1985.

[14] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Statist.*, vol. 14, pp. 436–440, Dec. 1943.

[15] M. Sobel and P.A. Groll, "Group testing to eliminate efficiently all defectives in a binomial sample," *Bell Syst. Tech. J.*, vol. 38, pp. 1178–1252, Sept. 1959.

[16] S. Ghez, S. Verdú, and S. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 640–649, July 1988.

[17] M. Zorzi and R. Rao, "Capture and retransmission control in mobile radio," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1289–1298, Oct. 1994.

[18] B. Hajek, A. Krishna, and R.O. LaMaire, "On the capture probability for a large number of stations," *IEEE Trans. Commun.*, vol. 45, pp. 254–260, Feb. 1997.

[19] Q. Zhao and L. Tong, "A multiqueue service room MAC protocol for wireless networks with multipacket reception," *IEEE/ACM Trans. Networking*, vol. 11, pp. 125–137, Feb. 2003.

[20] Q. Zhao and L. Tong, "A dynamic queue protocol for multiaccess wireless networks with multipacket reception," *IEEE Transactions on Wireless Communications*, to be published, http://acsp.ece.cornell.edu/pubJ.html. See also *Proc. 2000 Asilomar Conf. Sig., Sys., and Computers.*

[21] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite user slotted ALOHA with multipacket reception," *IEEE Trans. Inform. Theory*, to be published, http://acsp.ece.cornell.edu/pubJ.html. See also *Proc. 30th Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2002.

[22] L. Kleinrock, *Queueing Systems Volume II: Computer Applications*. New York: Wiley, 1976.

[23] J. Silvester and L. Kleinrock, "On the capacity of multihop slotted ALOHA networks with regular structure," *IEEE Trans. Commun.*, vol. 31, pp. 974–982, Aug. 1983.

[24] G. Mergen and L. Tong, "Stability and capacity of wireless networks with probabilistic receptions," *IEEE Trans. Inf. Theory*, to be published, http://acsp.ece.cornell.edu/pubJ.html. See also *Proc. 2002 Conf. Information Sciences and Systems*, Princeton, NJ, March 2002.

[25] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[26] R.M. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Proc. Cambridge Philos. Soc.*, vol. 58, pp. 497–520, 1962.

[27] J. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. Inform. Theory*, vol. 31, pp. 192–204, Mar. 1985.

[28] V. Anantharam, "The stability region of the finite-user slotted ALOHA protocol," *IEEE Trans. Inform. Theory*, vol. 37, pp. 535–540, May 1991.

[29] V. Naware and L. Tong, "Using queue statistics for beamforming in ALOHA," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, Nov. 2003, pp. 212–215.

[30] R. Knopp and P. Humblet, "Information capacity and power control in single cell multi-user communications," in *Proc. Intl Conf. Comm.*, Seattle, WA, June 1995, pp. 331–335.

[31] D.N.C. Tse and S.V. Hanly, "Multiaccess fading channels : I polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.

[32] P. Viswanath, D.N.C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[33] I.E. Telatar and S. Shamai, "Some information theoretic aspects of decentralized power control in multiple access fading channels," in *Proc. Info. Theory and Networking Workshop*, Piscataway, NJ, 1999, pp. 23–23.

[34] X. Qin and R. Berry, "Exploiting multiuser diversity in wireless ALOHA networks," in *Proc. Allerton Conf. Communication, Control and Computing*, Allerton, IL, Oct. 2001, pp. 793–794.

[35] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. Information Theory*, to be published. See also Tech. Rep. ACSP-TR-11-02-01, Cornell University, Nov. 2002. http://acsp.ece.cornell.edu/pubR.html.

[36] P. Venkitasubramaniam, S. Adireddy, and L. Tong, "Sensor networks with mobile agents: Optimal random access and coding," *IEEE J. Sel. Areas in Comm.: Special Issue on Sensor Networks,* to be published. http://acsp.ece.cornell.edu/pubJ.html.

[37] G. Dimić, R. Zhang, and N.D. Sidiropoulos, "Medium access control–physical cross-layer design," *IEEE Signal Processing Mag.*, vol. 21, pp. 40–58, Sept. 2004.

[38] I.E. Telatar and R.G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 963–969, Aug. 1995.

[39] R. Berry and E. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Processing Mag*, vol. 21, pp. 59–68, Sept. 2004.