On the Design of Universal Schemes for Massive Uncoordinated Multiple Access

Austin Taghavi, Avinash Vem, Jean-Francois Chamberland, Krishna R. Narayanan
Department of Electrical and Computer Engineering
Texas A&M University, College Station, TX 77843-3128
{taghavi,vemavinash,chmbllnd,krn}@tamu.edu

Abstract—Future wireless access points may have to support sporadic transmissions from a massive number of unattended machines. Recently, there has been a lot of interest in the design of massive uncoordinated multiple access schemes for such systems based on clever enhancements to slotted ALOHA. A close connection has been established between the design of the multiple access scheme and the design of low density generator matrix codes. Based on this connection, optimal multiple access schemes have been designed based on slotted ALOHA and successive interference cancellation, assuming that the number of users in the network is known at the transmitters. In this paper, we extend this work and consider the design of universal uncoordinated multiple access schemes that are agnostic to the number of users in the network. We design Markov chain based transmission policies and numerical results show that substantial improvement to slotted ALOHA is possible.

I. INTRODUCTION

The emergence of machine-driven wireless communication is poised to alter the wireless landscape within the next few years. In addition to traditional traffic generated by individuals who interact with personal phones and mobile computers, a sizable portion of wireless resources will be consumed by legions of unattended devices that seek to disseminate information sporadically. This is an important paradigm shift, as it changes the profile of subscribers and may impact the performance of currently deployed protocols. It requires the design of multiple access schemes where a massive number of users need to act in an uncoordinated fashion.

To address this challenge, several researchers have shifted their interest from information-centric resource allocation algorithms and scheduling schemes to random access protocols. While the performance of traditional random access such as slotted ALOHA scales poorly with the number of users (the throughput is limited by $1/e \approx 0.37$ [1]), recent results have shown that by using clever retransmission strategies and successive interference cancellation, the performance of slotted ALOHA can be substantially improved [2], [3], [4]. By establishing a connection between successive interference cancellation and message-passing decoding on bipartite graphs, Liva et al. optimized the transmission strategy numerically and showed that substantial improvement over slotted ALOHA is possible [2]. In [3], the optimal transmission strategy is derived analytically and it is shown that uncoordinated large-scale slotted ALOHA can be as efficient as coordinated access.

Existing results on improving the performance of slotted ALOHA assume that the number of users is known at the transmitters in order to implement the random access strategy. We point out that, in [5], the authors consider the setting of unknown number of users and design a protocol which uses multiple rounds of transmission where estimation of the number of users and resolution of the transmitted packets are performed jointly. In this paper, we consider a single round of transmission and consider the important question of whether it is possible to design universal, adaptive massive uncoordinated random access schemes that are agnostic to the number of users in the network. Our contribution is threefold. First, we introduce a novel formulation of the slotted multiple access problem with successive interference cancellation. In our formulation, the access point does not need to estimate the number of active devices in the system prior to a communication round, and the number of time slots within a round is determined dynamically. Second, we propose a Markov framework for slot selection by devices. Within this framework, every device independently elects to send a copy of its own message during a slot based on two variables: the time elapsed since the beginning of the communication round, and the number of times the device has transmitted copies of its packet thus far. Finally, we derive necessary and sufficient conditions for a sequence of probability distributions on the number of packets transmitted within the first $t$ slots to be achievable, and we show that suitable families of distributions lie within the feasible set. Altogether, our framework yields a universal access scheme where devices gracefully adapt to current conditions without having to collect and broadcast information prior to every communication round.

II. SYSTEM MODEL

Figure 1 depicts the main components of random multiple access over wireless infrastructures, as discussed in this paper. The communication process takes place on the uplink, and transmission opportunities are divided into rounds. One communication round features a collection of synchronous time slots. In the envisioned framework, the access point broadcasts a beacon on the downlink that is used for coarse slot synchronization by active devices. At the onset of every round, the access point indicates the start of a communication cycle and invites devices to begin transmitting messages. Every
active device repeats its message over a random subset of available slots. The messages are encoded with a code at the physical layer such that when a packet is received at the access point without interference, it is assumed to be decodable. We also assume that the header of each packet contains the seed of a random number generator that is used by the user to determine which slots to transmit in. Once a packet is decoded, the access point can then determine all the time slots where the user will transmit. Successive interference cancellation is then applied to slots where collisions occurred, enabling the iterative decoding of additional codewords. This process continues until the access points signals the end of the round on the downlink, potentially also informing devices about successfully decoded packets. As the round evolves, the access point can obtain a good estimate of the number of active devices, but we do not discuss these specifics herein.

We focus on uplink traffic, and we adopt a performance objective based on the number of successfully decoded packets as a function of time and device density. In spirit, this mechanism is a connectionless transmission protocol, with no guarantee of delivery. Data integrity is achieved through the judicious selection of error correcting codes at the network layer. Issues pertaining to sustained connections, packet ordering, and duplicate are relegated to higher layers. As mentioned in the introduction, this packet transmission mechanism is aimed at sporadic traffic generated by a massive amount of sensors and machines. It is not suitable for real-time traffic or streaming applications.

The setting considered earlier in [2], [3], [4] falls within the general framework considered here. An important difference is that the number of active devices is estimated prior to the beginning of every cycle and the access point broadcasts the slot count, $M$, for the upcoming communication round. For systems with a large number of active devices, near optimal performance can be achieved by having individual devices select the cardinality of their transmission subsets according to a soliton distribution, and then choose specific slots uniformly from the set $\{1, 2, \ldots, M\}$ [3]. In the absence of knowledge of $M$, such a policy cannot be implemented. To the best of our knowledge, good estimates on the achievable performance in the present case are not known.

### III. Abstract Framework

An uncoordinated transmission policy takes the elapsed time $t$ and the number of repeated transmissions of the packet since the beginning of the round as its arguments, and it outputs one or zero as an indication of whether to transmit or not during the upcoming slot. We are interested in scenarios where all active devices employ a same stochastic policy. We denote this policy by $\mu = (\mu_0(\cdot), \mu_1(\cdot), \ldots)$, where $\mu_t(m)$ represents the decision rule at time $t$ as a function of message count $m$. In general, the number of times a packet is transmitted by a specific device is a random process $X_t$ and it is governed by the recursive form

$$X_{t+1} = X_t + \mu_t(X_t) \quad t \geq 0. \quad (1)$$

Under a fixed policy $\mu$, we can use $p_t$ to denote the distribution of the ensuing process at time $t$. All devices share a same sequence of marginal distributions in the symmetric scenario, and their packet counts are independent from one another.

The recursive form of (1) naturally leads to a first-order Markov chain interpretation for $X_t$. The state of the chain corresponds to the number of packets transmitted so far and, as such, $X_0 = 0$. The chain can either stay in its current state or jump one level up because devices send at most one message per time slot. Transition probabilities are given by

$$\Pr(X_{t+1} = m|X_t = m) = \Pr(\mu_t(m) = 0)$$

$$\Pr(X_{t+1} = m+1|X_t = m) = \Pr(\mu_t(m) = 1).$$

A message is sent out during slot $t$ whenever $X_t - X_{t-1} = 1$. Using this interpretation, $p_t$ can be viewed as the probability distribution of Markov chain $X_t$ at time $t$. Under these two equivalent representations, it is possible to design uncoordinated transmission policies by working directly with $\mu$, or by constructing a Markov chain $X_t$ with suitable attributes and subsequently determining $\mu$ from the statistical properties of the chain. We center our attention on the latter approach.

To better understand how to select $X_t$ and appropriately shape the sequence of distributions $(p_0, p_1, \ldots)$, we briefly review the connection between successive interference cancellation and the iterative decoding of codewords on erasure channels [2], [4]. In this analogy, it is insightful to abstract the transmission and decoding processes as taking place on a bipartite graph with two sets of nodes $(V, C)$, where $V = \{v_1, \ldots, v_K\}$ denotes the collection of active devices and $C = \{c_1, \ldots, c_M\}$ represents the aggregate signals received during individual time slots. An edge appears between $v_k$ and $c_i$ whenever device $k$ transmits during slot $t$. Figure 1 also shows a sample communication process and its associated bipartite graph. The larger circle nodes portray variable nodes (messages) that need to be recovered. The smaller nodes denote the check nodes (or, generator nodes [6]), and they emphasize the fact that received signals $\{c_i\}$ are sums of codewords transmitted during individual time slots. A graph constructed in this manner is identical to a Tanner graph for
a low-density generator-matrix (LDGM) code. In the present case, the sum operation occurs in the space of signals, whereas in a typical LDGM code, the sum takes place over a finite field. Still, this discrepancy does not affect the decoding strategy nor its analysis.

At time $t$, the uncoordinated policy results in an ensemble of LDGM code equivalents. For this ensemble, we denote the variable-node degree distributions from the node and edge perspective by $L(x) = \sum_{i} L_{i} x^{i}$ and $\lambda(x) = \sum_{i} \lambda_{i} x^{i-1}$. By construction, $L_{i} = p_{t}(i)$ and, necessarily, $\lambda(x) = \frac{L(x)}{x^{1}}$. The check-node distributions are somewhat more involved to derive in the present scenario.

First, we note that the unconditioned probability that a device transmit during slot $t$ is governed by the Wasserstein (earth mover’s) distance between distributions $p_{t}$ and $p_{t+1}$,

$$\Pr(X_{t+1} = X_{t} + 1) = W_{1}(p_{t}, p_{t+1}) = \inf_{\gamma \in \Gamma(p_{t}, p_{t+1})} \sum_{(m, n) \in N_{0} \times N_{0}} d_{1}(m, n)\gamma(m, n)$$

where $\Gamma(p_{t}, p_{t+1})$ is the collection of joint probability measures on $N_{0} \times N_{0}$ with marginals $p_{t}$ and $p_{t+1}$ on their first and second factors, respectively. Thus, under this stochastic strategy, the unconditioned probability of $C_{t}$, the number of packets transmitted during slot $t$, is equal to

$$\Pr(C_{t} = k) = \binom{K}{k} (W_{1}(p_{t}, p_{t+1}))^{k} (1 - W_{1}(p_{t}, p_{t+1}))^{K-k}.$$ 

For large value of $K$ and small value of $W_{1}(p_{t}, p_{t+1})$, this expression is amenable to a Poisson approximation

$$\Pr(C_{t} = k) \approx \left(\frac{KW_{1}(p_{t}, p_{t+1})}{k!}\right)^{k} \exp(-KW_{1}(p_{t}, p_{t+1}))$$

IV. MARKOV CHAINS AND DISTRIBUTION SHAPING

The framework of Section III offers a valuable platform in the design of Markov processes and probability distributions for efficient decoding in massive uncoordinated multiple access. Still, the evolution of $X_{t}$ is governed by physical constraints. This Markov chain is monotone increasing and, at every point, it can either remain at its current level or jump up by a unit step. These outcomes correspond to the device electing to transmit its message during the next slot or choosing to remain silent. From a distribution viewpoint, the realm of possibilities is determined by stochastic dominance.

Let $X$ be a random variable with probability distribution $p_{X}$; and $Y$, a random variable with probability distribution $p_{Y}$. We use $X \preceq Y$ or, equivalently, $p_{X} \preceq p_{Y}$ to denote first-order stochastic dominance. By definition, we write $X \preceq Y$ whenever $\Pr(X > \ell) \leq \Pr(Y > \ell)$ for all $\ell \in \mathbb{R}$. As discussed below, this partial order underlies our ability to achieve sequences of distributions through monotone increasing Markov chains with at most unit increments [7], [8].

In the context of uncoordinated multiple access, we are especially interested in time-inhomogeneous, monotone increasing Markov chains over the non-negative integers, which we represent by $\mathbb{N}_{0}$. As mentioned above, prospective chains should only feature self-transitions and transitions to nearest neighbors on the right.

Proposition 1: Suppose that $(p_{0}, p_{1}, \ldots)$ is a sequence of discrete probability distributions over $\mathbb{N}_{0}$. This sequence of distributions can be achieved through a monotone increasing Markov chain containing solely self-transitions and transitions to nearest neighbors on the right if and only if

$$p_{t} \preceq p_{t+1} \quad \text{and} \quad p_{t+1} \preceq S p_{t} \quad \forall t \in \mathbb{N}_{0} \quad (2)$$

where $S$ denotes the standard right shift operator acting on one-sided infinite sequences of numbers.

Proof: We establish this result in a constructive manner, building a Markov chain $\{X_{t}\}$ with the desired properties. We emphasize that the resulting Markov chain can be and is often time-inhomogeneous. To begin, this Markov chain should have distribution $p_{0}$ at zero time. Also, $\{X_{t}\}$ should only feature two types of transitions: self-transitions and jumps to nearest neighbors on the right. Based on these requirements, we can analyze the progression of the chain over time. On the left boundary, at $m = 0$, we have

$$p_{t+1}(0) = \Pr(X_{t+1} = 0 | X_{t} = 0) p_{t}(0) \approx (1 - \gamma_{m}^{(t)}) p_{t}(0). \quad (3)$$

Furthermore, for any state $m > 0$, we can write

$$p_{t+1}(m) = \Pr(X_{t+1} = m | X_{t} = m - 1) p_{t}(m - 1) + \Pr(X_{t+1} = m | X_{t} = m) p_{t}(m) \approx \gamma_{m-1}^{(t)} p_{t}(m - 1) + (1 - \gamma_{m}^{(t)}) p_{t}(m). \quad (4)$$

Rearranging these expressions, we obtain generic conditions in the form of a system of linear equations,

$$\gamma_{m}^{(t)} p_{t}(m) = p_{t}(m) - p_{t+1}(m) + \gamma_{m-1}^{(t)} p_{t}(m - 1) = \sum_{\ell=0}^{m} p_{t}(\ell) - \sum_{\ell=0}^{m} p_{t+1}(\ell) \quad (5)$$

where $m$ takes value in $\mathbb{N}_{0}$. Since $\gamma_{m}^{(t)}$ can be interpreted as a conditional probability, it must lie within $[0, 1]$. This translates into two necessary and sufficient conditions. The first one corresponds to $\gamma_{m}^{(t)}$ being non-negative. This is achieved by having the difference between the two cumulative sums in (5) remain non-negative for all $m \in \mathbb{N}_{0}$ or, equivalently, $p_{t} \preceq p_{t+1}$. When this is satisfied, the second requirement ensures that $\gamma_{m}^{(t)}$ is no greater than one. We can express this requirement as

$$\sum_{\ell=0}^{m} p_{t}(\ell) - \sum_{\ell=0}^{m} p_{t+1}(\ell) \leq p_{t}(m) \quad (6)$$

which yields $\sum_{\ell=0}^{m-1} p_{t}(\ell) \leq \sum_{\ell=0}^{m} p_{t+1}(\ell)$ after reorganizing terms yields. Using the standard right shift operator acting on one-sided infinite sequences of numbers, we get the compact expression $p_{t+1} \preceq S p_{t}$. Thus, whenever the conditions in (2) are satisfied, it is possible to define (time-dependent) transition probabilities for $\{X_{t}\}$ such that the desired distributions are
achieved at every step. Conversely, given a Markov chain with admissible transition probabilities, i.e., self-transitions and transitions to nearest neighbors on the right, the corresponding sequence of discrete distributions \((p_0, p_1, \ldots)\) must satisfy the stochastic dominance conditions in (2).

Altogether, the partial order induced by stochastic dominance is key in being able to identify which sequences of distributions are realizable using first-order Markov chains. When the partial orders in (2) are fulfilled, explicit form for the transition probabilities can be inferred from Proposition 1.

**Definition 1 (Uncoordinated Markov Strategy):** Suppose that \((p_0, p_1, \ldots)\) is a sequence of discrete probability distributions over \(\mathbb{N}_0\) such that the conditions in (2) are satisfied. For integers \(m, t \in \mathbb{N}_0\), define constants

\[
\mu_t(m) = \gamma^*_m(t) = \begin{cases} \frac{\sum_{\ell=0}^{\ell=m} p_\ell(t) - \sum_{\ell=0}^{\ell=m+1} p_{\ell+1}(t)}{p_m(m)} & p_m(m) > 0 \\ 0 & p_m(m) = 0 \end{cases} (7)
\]

Let \(\{X_t\}\) be a first-order Markov chain with initial probability distribution \(p_0\), and transition probabilities

\[
\Pr(X_{t+1} = m | X_t = m) = 1 - \gamma^*_m(t)
\]

\[
\Pr(X_{t+1} = m+1 | X_t = m) = \gamma^*_m(t)
\]

where \(m, t \in \mathbb{N}_0\). Necessarily, \(\Pr(X_{t+1} = n | X_t = m) = 0\) whenever \(n \notin \{m, m+1\}\). Then \(\{X_t\}\), or equivalently \(\mu\), forms an admissible uncoordinated transmission policy.

**Corollary 1:** The discrete random process \(\{X_t\}\) specified in Definition 1 is a valid Markov chain and it possesses discrete distribution \(p_t\) at time \(t\).

**Proof:** This result follows from the proof of Proposition 1. First, \(p_t \preceq p_{t+1}\) guarantees that the constants defined in (7) are non-negative. Second, \(p_{t+1} \preceq S_{p_t}\) ensures that the constants remain bounded by one. Random process \(\{X_t\}\) is then characterized as a first-order Markov chain through its initial distributions and transition probabilities. Furthermore, by construction, it only features self-transitions and transitions to nearest neighbors on the right. Finally, through the recursive equations (3) and (4), we deduce that the probability distribution of the Markov process at time \(t\) is \(p_t\), as desired.

As a first example for the proposed methodology, we consider the soliton sequence of distributions. Soliton distributions have played an instrumental role in the development of rateless codes. Through duality, they arise naturally in the traditional setting for uncoordinated multiple access where \(K\) is known beforehand. As seen below, it is possible to create a Markov chain whose distribution at every time instant falls within the soliton family.

**Example 1 (Soliton Distributions):** The soliton distributions fulfill the requirements of Proposition 1 and, as such, the corresponding sequence can serve as a basis for an uncoordinated Markov strategy. Recall that, for \(t \in \mathbb{N}\), the soliton distribution is given by

\[
\Pr(X_t = 1) = \frac{1}{t+1}
\]

Checking the first condition in Proposition 1, we get

\[
\sum_{\ell=0}^{m} p_\ell(t) - \sum_{\ell=0}^{m-1} p_{\ell+1}(t) = \frac{1}{t} - \frac{1}{t+1} = 1 - \frac{1}{t+1}
\]

for \(m = 0, \ldots, t\); the difference vanishes for \(m \geq t + 1\). Thus, \(p_{\text{sol}(t)} \preceq p_{\text{sol}(t+1)}\). Likewise, verifying the second condition in Proposition 1, we start with

\[
\sum_{\ell=0}^{m} p_{\ell+1}(t) - \sum_{\ell=0}^{m-1} p_{\ell}(t) = \frac{1}{t+1} \geq 0
\]

for \(m = 1\); moving forward, we have

\[
\sum_{\ell=0}^{m} p_{\ell+1}(t) - \sum_{\ell=0}^{m-1} p_{\ell}(t) = \frac{1}{(m-1)m} - \frac{1}{t(t+1)} \geq 0
\]

for \(m = 2, \ldots, t\). Again, the difference vanishes for \(m \geq t + 1\) and hence \(p_{\text{sol}(t+1)} \preceq S_{p_{\text{sol}(t)}}\). Jointly, these results imply that the soliton probability distributions ordered according to parameter \(t\) form an admissible sequence for an uncoordinated Markov strategy. Specifically, consider the sequence

\[
p_0 = e_0, p_1 = p_{\text{sol}(1)}, \ldots, p_t = p_{\text{sol}(t)}, \ldots
\]

where \(e_0\) is the canonical unit vector with unit mass at zero. We note briefly that \(e_0 \preceq p_1\) and \(p_1 = S_{e_0}\), which preserve the admissibility of the sequence. By applying Corollary 1 to these distributions, we gather that the transition probabilities for the Markov strategy become

\[
\gamma^*_m(t) = \begin{cases} \frac{1}{t+1} & m = 1 \\ \frac{(m-1)m}{t(t+1)} & m = 2, \ldots, t \\ 0 & \text{otherwise} \end{cases}
\]

for \(t \in \mathbb{N}\). Explicitly, the transition probabilities associated with time-inhomogeneous Markov chain \(\{X_t\}\) for \(t \in \mathbb{N}\) are given by

\[
\Pr(X_{t+1} = m | X_t = m) = 1 - \frac{(m-1)m}{t(t+1)}
\]

\[
\Pr(X_{t+1} = m+1 | X_t = m) = \frac{(m-1)m}{t(t+1)}
\]

for \(m = 2, \ldots, t\); and

\[
\Pr(X_{t+1} = 2 | X_t = 1) = \frac{1}{t+1}
\]

\[
\Pr(X_{t+1} = 1 | X_t = 1) = \frac{t}{t+1}.
\]

At time zero, the Markov chain starts in state \(X_0 = 0\), and deterministically transitions to \(X_1 = 1\). The emission probability for a particular device is governed by

\[
\sum_{i=1}^{t} \gamma^*_i(t) p_{\text{sol}(t)}(i) = \frac{1}{t+1} \quad \text{for } t \geq 2.
\]

While the soliton distribution is optimal for the case when \(K\) is known at the transmitters, the sequence of distributions does not yield an efficient scheme for universal uncoordinated multiple access. One issue that degrades performance with the
A soliton Markov strategy is the fact that devices are collectively more likely to send packets at the onset of the communication round, as their behavior is correlated; this fails to induce a suitable right degree distribution. To mitigate this issue, we introduce two new classes of distributions, and we adopt a mixture of these distributions as a basis for universal access.

Example 2 (Stateless Distributions): The stateless family of distributions is characterized by a device using emission probabilities determined solely by the time elapsed since the onset of the communication round. Initially, the stateless distribution is simply $p_0 = e_0$. Given $p_t$, the distribution at the next time step is governed by

$$p_{t+1} = (1 - \gamma^{(t)})p_t + \gamma^{(t)}Sp_t,$$

where $S$ is the standard right shift operator. For the problem at hand, candidates for the emission probabilities include

$$\gamma^{(t)} = \frac{c}{t} \quad \text{and} \quad \gamma^{(t)} = 1 - \exp\left(\frac{c\log(t)}{t}\right)$$

where $c$ is a tuning parameter. The first function maximizes the probability of getting a check node of degree one towards the end of the round, whereas the second function sets the probability of an empty slot at the end of a round to $e$. It is too possible to define an alternate version of this distribution where the number of transmitted packets at time $t$ is upper bounded by a prescribed maximum $\tau$.

Example 3 (Skewed Distributions): In contrast to the previous example, the skewed family of distributions favors nodes that have transmitted several packets in the past. Again, at the onset of the communication process $p_0 = e_0$. Given an average emission target $\bar{\tau}^{(t)}$, the state transition probabilities at time $t$ become

$$\gamma^{(t)}_{m} = \begin{cases} 
0, & \sum_{i=0}^{m} p_t(i) < 1 - \bar{\tau}^{(t)} \\
1 - \frac{\sum_{i=m+1}^{m} p_t(i)}{p_t(m)}, & \sum_{i=m+1}^{m} p_t(i) \leq \bar{\tau}^{(t)} \\
\text{otherwise} & \end{cases}$$

A slightly different version of the skewed distributions limits the number of packets a device can send by time $t$.

In both cases, these distributions seek to tradeoff the probability of getting a single packet per time slot towards the end of a round with the likelihood of empty slots. The former class favors transmission behavior where the number of packets sent by a device rapidly concentrates around the mean. The latter classes features a heavier tail with a select few devices being responsible for a large portion of the traffic. Not too surprisingly, traffic shaping can affect the decoding process significantly.

V. NUMERICAL RESULTS AND DISCUSSION

The results in Fig. 2 showcase mean throughput, measured as the average number of decoded packets per time slot for the Markov access schemes based on a mixture of the stateless and skewed distributions. The mean emission probability is employed to control the degree of check nodes as a function of time. The weighting factor between these two distributions acts as a means to control the variance in the degree distribution of the variable nodes. Graphs are shown for $K = 1000$ devices with a weighting factor of 0.85 in favor of the stateless distributions; the various curves correspond to different tuning parameters. The maximum efficiency is approximately equal to 69 percent, and it is achieved with $N = 1250$ and $c = 1.2$. This level of performance is encouraging, as it substantially exceeds the performance of traditional ALOHA. Similar performance is observed under the same Markov access schemes, but with different device count $K$.

With perfect knowledge of the number of users, close to 80 percent efficiency can be obtained and, hence, there appears to be a penalty for not having precise knowledge of the number of users at the transmitters. Yet, tight upper bounds on the average number of messages per slot are not known for universal uncoordinated multiple access. It is an interesting open question to understand the behavior of universal multiple access in the limit of large $K$.

REFERENCES