Iterative Collision Resolution for Slotted ALOHA: An Optimal Uncoordinated Transmission Policy

Krishna R. Narayanan and Henry D. Pfister
Department of Electrical and Computer Engineering,
Texas A&M University
Email: {krn,hpfister}@tamu.edu

***Abstract*—We consider a multi-user wireless network in which each user has one packet of information to transmit to a central receiver. We study an uncoordinated paradigm where the users send their packet a random number of times according to a probability distribution. Instead of discarding the collided packets, the receiver performs iterative collision resolution. Recently, a few studies have shown that the iterative collision resolution process can be viewed as message-passing decoding on an appropriately defined Tanner graph. Using this equivalence, they used standard techniques to numerically optimize the probability distribution and demonstrated substantial throughput improvement over slotted ALOHA. In this paper, we show that the well-known soliton distribution is an optimal probability distribution and that the resulting throughput efficiency can be arbitrarily close to 1.**

***Index Terms*—Multiple-access, Collision resolution, Rateless Codes, Iterative Decoding**

I. Introduction

A. Problem Statement

We consider a multi-user system with K users where each user wishes to transmit one packet of information to a central receiver. The total time available for communication is split into M time slots and the duration of each time slot is assumed to equal the time required to transmit one packet. In the jth time slot, a subset of users transmit their packets. There is no coordination between the users and each user independently uses a policy that specifies whether or not they will transmit their packet in the jth time slot. We assume that the receiver knows the exact set of users who transmit during every time slot. There are many ways that this information can be shared with the receiver but this is not the focus of the paper and, hence, not discussed in detail. Specifically, we consider a paradigm which is similar to the one in [1], where the kth user generates a random variable $D_k \in \{1, \ldots, M\}$ according to a probability mass function f_D, i.e., $Pr(D_k = i) = f_D[i]$. Then, the user chooses D_k time slots uniformly at random without replacement from the set $\{1, \ldots, M\}$ and transmits during these slots.

In any given time slot, if exactly one user transmits a packet, then this packet is assumed to be decoded correctly. This is reasonable if good channel codes are used separately by each user. If more than one user transmits their packet in the same time slot, a collision results. In this scheme, the receiver subtracts all the previously decoded packets from the collided packets and if the receiver is able to subtract all but one of the users, then a single-user decoder can be used to recover the last user's packet. Otherwise, the received packet is saved in a buffer for future processing and every time a new packet is decoded, it is subtracted from all the packets in the buffer. We will refer to this as iterative interference cancellation since this is similar to interference cancellation in multi-user detection.

When the receiver has received all the K packets correctly, an acknowledgment is sent to terminate the transmission. Since a total of M time slots is required to successfully transmit all K packets of information, the throughput of the system is said to be $\eta = K/M$ packets/slot. Clearly, an upper bound on the throughput is $\eta = 1$ packet/slot. The main result in the paper is to derive the optimal probability mass function f_D for which this upper bound is achievable in the limit of $K \to \infty$, even when there is no coordination between the transmitters.

B. Background

It is well known that a standard slotted ALOHA scheme achieves a throughput efficiency of $1/e \approx 0.37$ and, hence, the proposed scheme is substantially better than standard slotted ALOHA. Recently, there have been a number of papers that consider collision resolution through iterative interference cancellation and thereby provide improved performance over slotted ALOHA. In [2] and [3], the time asymmetry between transmissions is exploited to bootstrap the iterative interference cancellation process. In the problem we consider, there is no time asymmetry and, hence, their results do not apply directly. The work that is most closely related to this paper is the work of Liva in [1] where a similar scheme as proposed here was considered and the author showed that an η of 0.965 can be achieved based on numerically optimizing the distribution f_D. In [4], the authors consider an extension of the work in [1], where the nodes encode their packet before transmission and, again, numerically optimize the distribution. However, the question of whether or not the upper bound on $\eta \leq 1$ is achievable is not addressed. The main observation of this paper is that the optimal distribution for this problem is the dual of the well-known soliton distribution [5] and that one can get arbitrarily close to the upper bound $\eta \leq 1$.

This material is based upon work supported, in part, by the National Science Foundation (NSF) under Grant No. 0830696. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

II. Graphical Representation and Density Evolution

It is convenient to think of the transmission and decoding process using a bipartite graph with two sets of nodes $(\mathcal{V}, \mathcal{C})$ where $\mathcal{V} = \{v_1, v_2, \ldots, v_K\}$ denotes the set of users and $\mathcal{C} = \{c_1, \ldots, c_M\}$ denotes the received signals during the time slots. An edge is placed between v_k and c_j if user k transmitted in the jth slot. Such a graph is shown in Fig. 1 for $K = 4$ and $M = 5$. In Fig. 1, we can also use the circle nodes to represent variable nodes which need to be recovered. The square nodes denote the check nodes (or, generator nodes [6]) which represent the fact that the received signal c_j during time slot j is the sum of the packets transmitted during time slot j. Such a graph can then be seen to be identical to a Tanner graph for a low-density generator-matrix (LDGM) code. Here, the sum operation at the check nodes is over the real field, whereas in a typical LDGM code, the sum is over a finite field. This difference does not affect the decoding algorithm and the analysis of the decoding algorithm.

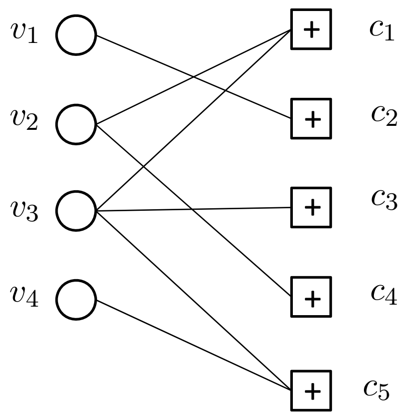

Fig. 1. Tanner graph representation of the multiple-access scheme

A. Degree Distributions

The encoding procedure described in the previous section with a chosen PMF f_D, results in an ensemble of LDGM codes (or, equivalently, Tanner graphs) whose degree distributions can be determined as follows [6]. Let $L(x) = \sum_i L_i x^i$ and $\lambda(x) = \sum_i \lambda_i x^{i-1}$ denote the variable-node degree-distributions from the node and edge perspectives, respectively. Then, $L_i = f_D[i]$ and $\lambda(x) = \frac{L'(x)}{L'(1)}$. Similarly, let $R(x) = \sum_i R_i x^i$ and $\rho(x) = \sum_i \rho_i x^{i-1}$ denote the check-node degree-distributions from the node and edge perspectives, respectively, and $\rho(x) = \frac{R'(x)}{R'(1)}$. The average bit-node degree $\mathtt{l}_{avg} = L'(1)$ and, hence, the average check-node degree is $\mathtt{r}_{avg} = \frac{K}{M}\mathtt{l}_{avg}$. If the decoding is successful, the efficiency of this scheme is simply

$$\eta = \frac{\mathtt{r}_{avg}}{\mathtt{l}_{avg}} = \frac{R'(1)}{L'(1)}. \tag{1}$$

In any given time slot j, since each of the K users transmits independently with probability $\mathtt{r}_{avg}/K$, the check-node degree-distribution from the node perspective is given by the binomial distribution $\text{Bin}\left(K, \frac{\mathtt{r}_{avg}}{K}\right)$. In the limit as $K \to \infty$, this binomial distribution becomes a Poisson distribution and, hence, $R(x) = e^{-\mathtt{r}_{avg}(1-x)}$. It is straightforward to verify that $\rho(x) = e^{-\mathtt{r}_{avg}(1-x)}$ as well.

B. Decoding and Density Evolution

The iterative interference cancellation sequence used to recover the information packets is identical to the peeling decoder schedule for the binary erasure channel (BEC) using the Tanner graph shown in Fig. 1. Specifically, each check node sends an outgoing message along each edge connected to the check node whose value is an erasure or the value of the information packet. At a check node of degree 1, the outgoing message along the edge is not an erasure as the value of that information packet can be determined from the received signal. At a check node of larger degree, the outgoing message is not an erasure if each of the incoming messages along the other edges connected to that check node is not an erasure. Otherwise, the outgoing message is an erasure. At a variable node, the outgoing message is an erasure if the incoming messages along all the other edges connected to that variable node is an erasure.

Let $\mathcal{G}(K, M, \lambda, \rho)$ denote the ensemble of Tanner graphs corresponding to the multiple-access scheme with K variable nodes, M check nodes, and the degree distribution pair $(\lambda(x), \rho(x))$. We will consider the decoding performance averaged over the ensemble of graphs $\mathcal{G}(K, M, \lambda, \rho)$ in the limit as $K, M \to \infty$. Let x_l and y_l denote the probability that an outgoing message from the variable node and check node, respectively, are erased during the lth iteration. Since the depth-$2l$ neighborhood of a randomly chosen edge in $\mathcal{G}(K, M, \lambda, \rho)$ is tree-like with probability one as $M, K \to \infty$, it follows that

$$\begin{aligned} \mathsf{y}_1 &= 1 - \rho(0) \\ \mathsf{x}_l &= \lambda(\mathsf{y}_l), \quad l \geq 1 \\ \mathsf{y}_{l+1} &= 1 - \rho(1 - \mathsf{x}_l), \quad l \geq 1 \end{aligned} \tag{2}$$

and, hence, we obtain a recursion for y_{l+1} given by

$$\mathsf{y}_{l+1} = 1 - \rho\left((1 - \lambda(\mathsf{y}_l)\right). \tag{3}$$

We say that the decoding is successful if $\mathsf{y}_l \overset{l\to\infty}{\to} 0$. An equivalent condition for successful decoding is

$$\rho(1 - \lambda(y)) > 1 - y, \quad y \in (0, 1 - \rho(0)]. \tag{4}$$

Checking the derivative of this condition at $y = 0$ gives the *stability condition*, $\lambda_2 \rho'(1) \leq 1$, which is also required for convergence to 0. If the inequality is strict, then y_l converges to 0 exponentially with iteration for sufficiently large l.

III. Optimal Degree Distribution

The main result in the paper is the observation that the sequence of ensembles obtained by truncating the ideal soliton distribution to N terms has an efficiency $\eta \overset{N\to\infty}{\to} 1$. For example, it achieves successful decoding with high probability, as $M \to \infty$, for a randomly chosen code from the sequence and the efficiency of this sequence of ensembles is $\eta \overset{N\to\infty}{\to} 1$. Therefore, one can obtain an efficiency arbitrarily close to 1. The following lemma makes this precise.

Lemma 1: For any $a \in [0,1]$, consider the sequence (indexed by $N \in \mathbb{N}$) of node-perspective degree-distributions, $(L^N(x), R^N(x))$, where

$$L^N(x) = \frac{\sum_{i=2}^{N+1} \frac{x^i}{i(i-1)} - \frac{ax^2}{2}}{\sum_{i=2}^{N+1} \frac{1}{i(i-1)} - \frac{a}{2}}$$
$$R^N(x) = e^{-(H(N)-a)(1-x)}.$$

Here $H(N) = \sum_{i=1}^{N} \frac{1}{i}$ is the Nth Harmonic number. For every ensemble in this sequence, $\mathsf{y}_l \overset{l\to\infty}{\to} 0$ while $\eta^N = \frac{N}{N+1} - a \overset{N\to\infty}{\to} 1-a$.

Proof: Let $\lambda^N(x), \rho^N(x)$ denote the corresponding degree distributions from the edge perspective. Then, we have $\rho(x) = e^{-(H(N)-a)(1-x)}$ and

$$\lambda^N(x) = \frac{L'^N(x)}{L'^N(1)} = \frac{\sum_{i=1}^{N} \frac{x^i}{i} - ax}{H(N)-a}.$$

To show that the convergence condition in (4) holds, notice that the numerator of $\lambda^N(x)$ is closely related to a truncated version of the power series expansion of $-\ln(1-x) = \sum_{i=1}^{\infty} \frac{x^i}{i}$. Therefore, for $y \in (0,1)$, we have

$$-ay + \sum_{i=1}^{N} \frac{y^i}{i} < -ay - \ln(1-y).$$

Substituting $(\lambda^N(x), \rho^N(x))$ into (4) gives

$$\begin{aligned}\rho^N(1-\lambda^N(y)) &= e^{ay - \sum_{i=1}^{N} \frac{y^i}{i}} \\ &> e^{\ln(1-y)} e^{ay} \\ &> (1-y), \end{aligned} \quad (5)$$

for $y \in (0,1)$ and shows that (4) holds. This implies that $\mathsf{y}_l \overset{l\to\infty}{\to} 0$ and, hence, that successful decoding can be achieved with high probability. The efficiency of the resulting system is given by

$$\begin{aligned}\eta^N &= \frac{R'^N(1)}{L'^N(1)} = \frac{H(N)-a}{\frac{H(N)-a}{\sum_{i=2}^{N+1} \frac{1}{i(i-1)} - a}} \\ &= \sum_{i=1}^{N} \frac{1}{i(i+1)} - a = \frac{N}{N+1} - a. \end{aligned} \quad (6)$$

The last step can be readily verified by induction and we observe that $\eta^N \to 1-a$ as $N \to \infty$. ■

Remark 1: Setting $a = 0$ in the previous lemma shows that one can achieve an efficiency $\eta^N \to 1$ by using the truncated ideal soliton distribution. One problem with this choice is that, for $a = 0$, the ensemble satisfies the stability condition with equality. While this does not prevent the iteration from converging to 0, it does imply that convergence will be very slow. To rectify this, the parameter a is introduced. For a small penalty in efficiency, this causes y_l to converge to 0 exponentially fast. In particular, the iteration of the proposed ensemble can be expanded about $y = 0$ to get

$$\mathsf{y}_{l+1} = \lambda_2^N \rho'^N(1)\mathsf{y}_l + O(\mathsf{y}_l^2) = (1-a)\mathsf{y}_l + O(\mathsf{y}_l^2).$$

IV. Connection to Rateless Codes

It can be seen that the coding paradigm considered is similar to the fountain-coding paradigm of LT codes [5]. However, the important difference is that in the problem considered here, the data is not centrally available to be encoded using a rateless fountain code. From mathematical viewpoint, the main difference is that, for our problem, the check-node degree-distribution is Poisson and the bit-node degree-distribution is Poisson for LT codes. Fortunately, the structure of LDGM codes implies that the degree distribution remains optimal if the bit and check degree distributions are switched [6]. Therefore, the optimal degree distribution for $L(x)$ is the same soliton distribution which is used in the case of fountain codes.

Fountain codes are also LDGM codes and when interpreted this way, the optimal degree distributions for the fountain code corresponds to $\rho(x) = -\ln(1-x)$, whereas in our case, the optimal $\lambda(x) = -\ln(1-x)$. This duality is the result of the fact that in the fountain code case, the information bits are uniformly chosen at random and, hence, the variable node degree distribution is a Poisson distribution. In our case, since the data is not centrally available we cannot choose the information bits at random to form linear combinations.

Our result shows that one can circumvent this problem in an optimal way by choosing the time slots for transmission uniformly at random, thereby inducing a Poisson distribution on the check nodes. It is interesting that the degree distribution $-\ln(1-x)$ is perfectly matched to the Poisson distribution regardless of which is used as $\lambda(x)$. In some ways, this degree distribution is actually better suited to this problem than to LT codes. For example, unlike in the LT codes case, here an outer code is not required to recover the bits that are not covered by the encoding process.

V. Future Work - Extension to Noisy Channels, Finite Length Effects, etc.

In this paper, we have abstracted the physical layer into error free bit pipes and hence, we have not considered the effect of noise, energy consumption and coding at the physical layer. The design of an uncoordinated transmission scheme in the case of a noisy channel should carefully consider the tradeoff between the total energy that is used for the transmissions, latency and the complexity of processing at the receiver.

Suppose each user is allowed an average transmit energy of E_s Joules per channel use. If the objective is to minimize E_s for a given (bandwidth) efficiency without consideration of computational complexity, we can consider the channel as a K user multiple-access channel (MAC) with ML channel uses, where L is the length of the packet in symbols and design codes that will achieve capacity of the MAC. In this case, the uncoordinated transmission scheme can be trivial - namely all the users transmit all the time. That is, energy efficiency can be traded for bandwidth efficiency optimally entirely through the choice of coding at the physical layer. However, in this case, the physical layer coding is spread across all the time slots that are available for transmission and hence, the latency and decoding complexity may be large.

At the other extreme of decoding complexity, we can use codes that achieve capacity during each single user transmission and decode each slot individually and use the retransmission scheme and the interference cancellation scheme considered here in Section II-B. The decoding complexity is only that of decoding each user individually; however, this is not energy efficient since the energy used in the retransmissions may not be optimally used. The design of efficient coding and retransmissions schemes that cleverly tradeoff complexity for energy efficiency is an interesting topic of research. The design of codes that permit decoding, using a joint Tanner graph whose complexity scales linearly with the block length, is one viable option. Exploiting the distribution of the number of times a packet is received in order to improve the energy efficiency is another option. Indeed, the first approach may be a particularly good option in a fixed wireless network where each transmitter has only a regulatory power constraint and no battery considerations. But, a detailed study of this is left for future work.

Other important directions for future work include a detailed study of the performance of the soliton distribution for finite lengths, a comparison with the performance of numerically optimized distributions in [1], and the design of protocols for implementing the iterative interference cancellation algorithm within the framework of other MAC protocols such as the IEEE 802.11.

VI. Acknowledgement

The authors would like to thank the reviewers for pointing some interesting directions for future work.

References

[1] G. Liva, "Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA," *IEEE Trans. on Commun.*, vol. 59, no. 2, pp. 477–487, 2011.

[2] S. Gollakota and D. Katabi, "ZigZag decoding: Combating Hidden Terminals in Wireless Networks," in *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, 2008, pp. 159–170.

[3] A. Saber Tehrani, A. Dimakis, and M. Neely, "SigSag: Iterative Detection through Soft Message-Passing," in *Proc. IEEE Conf. on Computer Commun. (INFOCOM)*, 2011, pp. 1017–1025.

[4] E. Paolini, G. Liva, and M. Chiani, "Graph-Based Random Access for the Collision Channel without Feedback: Capacity Bound," *Arxiv preprint arXiv:1201.1835*, 2012.

[5] M. Luby, "LT Codes," in *Proc. IEEE Symp. on the Foundations of Computer Science (FOCS)*, 2002, pp. 271–280.

[6] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge Univ Press., 2008.