Coded Categorization in Massive Random Access

Ryan Song, Kareem M. Attiah, and Wei Yu
Electrical and Computer Engineering Department
University of Toronto, Toronto, Canada
E-mails: r.song@mail.utoronto.ca, kattiah@ece.utoronto.ca, weiyu@ece.utoronto.ca

***Abstract*—This paper considers a massive random access scenario in which a small set of k users out of a large number of n potential users are active at any given time, and a central base-station wishes to send a common message to the active users in order to label them into a finite number of categories. Specifically, given c possible categories, the base-station wishes to send label ℓ to a set of k_ℓ users, where $\ell \in \{1,\dots,c\}$ and $\sum_{\ell=1}^{c} k_\ell = k$. Assuming that $n, k_1, \dots, k_c$ are fixed, we ask: what is the minimum rate of the common message that the base-station needs to send so that the correct label is received at each of the k active users? This paper shows that instead of a conventional scheme of listing the indices of the users followed by their labels, which requires a common message rate of $k\left(\log(n) + H\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)\right)$ bits, it is possible to construct a fixed-length common message code with a rate of just $kH\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)$ bits plus a term that scales in n as $O(\log\log(n))$ for fixed $k_1,\dots,k_c$, where $H(\cdot)$ is the entropy of a probability distribution. If a variable-length code is permitted, the minimum common message rate is characterized as $kH\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)+O(1)$ bits, with no dependence on n. Finally, if $k_1,\dots,k_c$ deviate from the values for which the common message is designed, an additional cost per user equal to a Kullback-Leibler divergence term would be incurred.**

I. Introduction

This paper is motivated by the massive machine-type communications (mMTC) scenario in which a massive number of n devices are connected to a base-station (BS). Due to the sporadic nature of the traffic, only a small random subset of $k \ll n$ users are typically active at any given time [1]. Given a set of k active users, this paper considers a task in which the BS needs to send a common message to the k active users in order to categorize them into c categories. Specifically, the BS needs to send label ℓ to k_ℓ users where $\ell \in \{1,\dots,c\}$ and $\sum_{\ell=1}^{c} k_\ell = k$. We ask: what is the minimum rate of the common message that the BS needs to send so that the correct label is received at each of the k active users?

A naive scheme of constructing such a common message is to index each of the n users, then to transmit the indices of each of the k active users followed by their respective labels. Assuming that the sizes of the categories $k_1,\dots,k_c$ are fixed, since each label ℓ is transmitted with frequency $\frac{k_\ell}{k}$, the labels can be compressed using $H\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)$ bits per label, where $H(\cdot)$ is the entropy of a probability distribution. Further, indexing each one out of the n users costs $\log(n)$ bits. Thus, this naive scheme would require a common message rate of $k\left(\log(n) + H\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)\right)$ bits.

This paper shows that the above naive scheme, which uses $k\log(n)$ bits to explicitly identify each of the k active users, is far from optimal. The main result of this paper is that it is possible to construct a fixed-length common message code to communicate the categorization to the k active users at a rate of $kH\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)$ bits plus a term that scales in n as $O(\log\log(n))$ for fixed k. If a variable-length code is permitted, the minimum rate is characterized to be $kH\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)+O(1)$ bits. This implies that the labels can be communicated to the active users using a code with an average rate of essentially $H\left(\frac{k_1}{k},\dots,\frac{k_c}{k}\right)$ bits per active user, without having to broadcast the identities of the active users explicitly. Surprisingly, this is true regardless of n.

The main reason for the large saving in the optimal rate of coded categorization as compared to the rate of the naive scheme stems from the following two key observations. First, since only k out of n users are active, it is only necessary to design decoding functions for the k active users, and not for all n potential users. Second, among the active users, it is not necessary to reveal every user's label to everyone else, because each active user is only interested in learning its own label. In other words, the information about the other users' labels is superfluous. Removing this extra information results in a significant saving in the common message rate.

These ideas have been exploited in the earlier work [2], where it is shown that the minimum common message rate required to schedule k active users into k slots in a collision-free manner is approximately $\log(e)$ bits per active user, plus an $\Theta(\log\log(n))$ term for the fixed-length code case and an $O(1)$ term for the variable-length code case. While this paper shares similarities with [2] in terms of how the rate bounds are derived, the two problem settings are quite different. In [2], the common message ensures that each user is scheduled into a unique slot but does not control which user goes into which slot, whereas this paper considers a problem of communicating a specific label to each of the active users.

The results of this paper have potential applications to a wide class of problems in mMTC, such as that of sending positive and negative acknowledgements to the users following activity detection [3]–[6]. In [7], it is shown that by allowing a bounded probability of error, the rate of user acknowledgement message can be reduced. While the focus of [7] is on labeling all users yet allowing some finite false-positive errors, this paper focuses on labeling a subset of users with zero-error code. Our aim is to characterize the fundamental limit of the common message rate in this lossless setting.

The notations used in this paper are as follows. We use lowercase letters, e.g. ℓ, to denote scalars, and lowercase boldface letters, e.g., $\mathbf{k}$, to denote vectors. Further, we use

978-1-6654-2159-1/22/$31.00 ©2022 IEEE

uppercase boldface letters, e.g., $\mathbf{X}_\ell$, to denote sets of vectors, uppercase boldface letters with an overbar, e.g., $\bar{\mathbf{X}}$, to denote ordered tuples of sets, and finally calligraphic letters, e.g., $\mathcal{D}$, to denote sets of ordered tuples. We use $(\cdot)^\mathsf{T}$ to dentoe transpose of a vector. We use $[n]$ to denote $\{1, 2, \ldots, n\}$ and $\binom{[n]}{k}$ to denote the set of all k-element subsets of $[n]$. We use $\log(\cdot)$ for logarithm in base 2 and $\ln(\cdot)$ for natural logarithm. We use $H(\cdot)$ to denote the entropy of a discrete random variable or a probability distribution, and $D(\cdot||\cdot)$ to denote the Kullback-Leibler divergence. We also use the shorthand $a^{\underline{b}} = a(a-1)\cdots(a-b+1)$.

II. Problem Formulation

Consider a massive random access scenario in which a random subset of k users out of a massive number of n users become active, and the BS needs to send a common message to these active users in order to categorize them into c possible categories. Specifically, let $\mathbf{A}$ denote the random set of active users with $|\mathbf{A}| = k$. The BS knows the identities of active users in $\mathbf{A}$, but the active users do not know the identities of each other. A *categorization* of $\mathbf{A}$ is defined to be an ordered tuple of $\mathbf{X}_\ell$'s, denoted as $\bar{\mathbf{X}} \triangleq (\mathbf{X}_1, \ldots, \mathbf{X}_c)$, that partitions $\mathbf{A}$, i.e.,

$$\bigcup_{\ell=1}^{c} \mathbf{X}_\ell = \mathbf{A} \tag{1}$$

and for $\ell, \ell' \in [c]$ we have

$$\mathbf{X}_\ell \bigcap \mathbf{X}_{\ell'} = \emptyset, \qquad \forall \ell \neq \ell'. \tag{2}$$

Here, $\mathbf{X}_\ell$, $\ell \in [c]$, is the subset of users in $\mathbf{A}$ who should receive the category label ℓ. We use $k_\ell = |\mathbf{X}_\ell|$ to denote the size of the category ℓ, and $\mathbf{k} \triangleq [k_1 \ldots k_c]^\mathsf{T}$ to denote the vector of category sizes.

We assume that the values of $n, k_1, \ldots, k_c$ are fixed, but the set of active users and the categorization are random. We aim to design a code to communicate the categorization to the active users for arbitrary realization of $\mathbf{A}$ and $\bar{\mathbf{X}}$. The key observations here are that only the active users are categorized and further each user is only interested in learning its own category label and not necessarily the labels of other users.

To this end, the BS encodes the categorization into a common broadcast message, which is transmitted over a noiseless downlink channel and decoded by the active users. Mathematically, define $\mathcal{D}^{(n,\mathbf{k})}$ as the set of all possible categorizations of all possible active user sets $\mathbf{A}$ of size k, where the category sizes are given by $\mathbf{k} = [k_1 \ldots k_c]^\mathsf{T}$, with $\sum_{\ell=1}^{c} k_\ell = k$, i.e.,

$$\mathcal{D}^{(n,\mathbf{k})} = \left\{ \bar{\mathbf{X}} \;\middle|\; \mathbf{X}_\ell \cap \mathbf{X}_{\ell'} = \emptyset, \forall \ell \neq \ell' \in [c], |\mathbf{X}_\ell| = k_\ell \right\}. \tag{3}$$

The problem of communicating a categorization to the active users can be formulated as that of designing an encoding function that maps a categorization to a common message taking value in an index set $[T]$, i.e.,

$$f : \mathcal{D}^{(n,\mathbf{k})} \to [T], \tag{4}$$

and a set of decoding functions, one for each active user, that map the message into the appropriate category labels, i.e.,

$$g_u : [T] \to [c], \ \forall u \in \mathbf{A}, \tag{5}$$

so that the correct labels are recovered at their respective users, i.e., $\forall \bar{\mathbf{X}} \in \mathcal{D}^{(n,\mathbf{k})}$ and $\forall \ell \in [c]$, i.e.,

$$g_u(f(\bar{\mathbf{X}})) = \ell, \ \forall u \in \mathbf{X}_\ell. \tag{6}$$

The goal of this paper is to characterize the minimum rate of such a common message.

We consider both fixed-length and variable-length codes. For the fixed-length case, let $T^*(n, \mathbf{k})$ be the minimum T such that an encoding function (4) and a set of corresponding decoding functions (5) that satisfy the successful label recovery condition (6) can be found for all possible realizations of $\mathbf{A}$ and $\bar{\mathbf{X}}$, assuming fixed values of n and $\mathbf{k}$. The minimum rate of the fixed-length code is defined to be

$$R_{\mathrm{f}}^*(n, \mathbf{k}) \triangleq \log(T^*(n, \mathbf{k})). \tag{7}$$

For the variable-length case, we take $\mathbf{A}$ to be uniformly distributed over all possible sets of k active users out of n potential users, and further take $\bar{\mathbf{X}}$ to be uniformly distributed over all possible categorizations $\mathcal{D}^{(n,\mathbf{k})}$ for fixed $\mathbf{k}$. (These turn out to be the worst-case distributions.) We define f^* to be an encoder that minimizes $H(f(\bar{\mathbf{X}}))$ under the constraint that there exist decoders g_i^* that together with f^* satisfy the successful label recovery condition (6). The minimum rate of the variable-length code is given by

$$R_{\mathrm{v}}^*(n, \mathbf{k}) \triangleq H(f^*(\bar{\mathbf{X}})). \tag{8}$$

The rational here is that the encoder output $f^*(\bar{\mathbf{X}})$ is now a random variable over the set $[T]$. Hence, using entropy coding, one can achieve a rate of $H(f^*(\bar{\mathbf{X}}))$. Note that for the variable-rate code, it is advantageous to set T to be large. In this case, the choice of the encoder output to ensure successful label recovery is not necessarily unique, and the encoder may judiciously choose among these choices in order to minimize the output entropy.

III. Categorization Code

In this section, we introduce a codebook-based encoding and decoding scheme for communicating a categorization to a subset of k out of n users. The codebook is constructed a priori and known to the BS and all the users.

Definition 1: A categorization codebook $\bar{\mathbf{M}}$ is defined as an ordered tuple of vectors:

$$\bar{\mathbf{M}} = (\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots, \mathbf{m}^{(T)}), \tag{9}$$

where each codeword $\mathbf{m}^{(t)}$ is a length-n vector whose elements take on values in $[c]$, i.e., $\mathbf{m}^{(t)} \in [c]^n, \forall t \in [T]$.

The idea is that each vector $\mathbf{m}^{(t)} = [m_1^{(t)} \ldots m_n^{(t)}]^\mathsf{T}$ represents some fixed labeling of all n users, where the label of user $u \in [n]$ is given by $m_u^{(t)}$. Given a particular set of active users and the associated categorization that the BS wishes to communicate, the BS can simply look for t, such that the

categorization is described by $\mathbf{m}^{(t)}$, then send the index t. Each of the active users u can then recover its label by looking up the value of $m_u^{(t)}$.

Note that because in our particular problem setting only k out of n users are active, the same vector $\mathbf{m}^{(t)}$ can represent multiple categorizations, depending on the realizations of the set of active users. We capture this concept in the following.

Definition 2: Given a fixed $\mathbf{m} \in [c]^n$, we define $\mathcal{C}(\mathbf{m})$ to be the set of all categorizations represented by $\mathbf{m}$, i.e.,

$$\mathcal{C}(\mathbf{m}) = \{\bar{\mathbf{X}} \mid m_u = \ell, \forall u \in \mathbf{X}_\ell, \forall \ell \in [c],\ \bar{\mathbf{X}} \in \mathcal{D}^{(n,\mathbf{k})}\}, \tag{10}$$

where $\mathbf{m} = [m_1 \dots m_n]^\mathsf{T}$ and $\bar{\mathbf{X}} = (\mathbf{X}_1, \dots, \mathbf{X}_c)$.

The encoding and decoding process based on the codebook $\bar{\mathbf{M}}$ can now be described mathematically as follows. To communicate a categorization $\bar{\mathbf{X}}$, the encoder finds a vector $\mathbf{m}^{(t)}$ in $\bar{\mathbf{M}}$ which is consistent with the categorization, i.e.,

$$f(\bar{\mathbf{X}}) = t \quad \text{s.t.} \quad \bar{\mathbf{X}} \in \mathcal{C}(\mathbf{m}^{(t)}). \tag{11}$$

The BS then transmits the index t as the common message. Upon receiving the common message, each of the active users u recovers its label ℓ as the u-th entry of the vector $\mathbf{m}^{(t)}$, i.e.,

$$g_u(t) = m_u^{(t)}. \tag{12}$$

For this encoding and decoding scheme to work with zero error, all possible categorizations in $\mathcal{D}^{(n,\mathbf{k})}$ must be covered by at least one $\mathbf{m}^{(t)}$ in $\bar{\mathbf{M}}$. Thus, a valid codebook $\bar{\mathbf{M}}$ must satisfy

$$\bigcup_{t=1}^{T} \mathcal{C}(\mathbf{m}^{(t)}) = \mathcal{D}^{(n,\mathbf{k})}. \tag{13}$$

Note that this codebook-based construction is completely general in the sense that any valid encoding and decoding scheme can be described this way. Thus, it is without loss of generality to restrict attention to this construction.

For fixed-length coding, the problem of finding the minimum rate common message now becomes that of finding the minimal T such that there exists an $\bar{\mathbf{M}}$ that satisfies (13).

For variable-length coding, we need to find an $\bar{\mathbf{M}}$ that satisfies (13) as well as to define an encoding function that minimizes $H(f(\bar{\mathbf{X}}))$. Note that the output $f(\bar{\mathbf{X}})$ is not necessarily unique, so a judicious choice of the output may reduce the output entropy.

IV. Achievable Rate

We now derive achievability bounds on the rate required to communicate a categorization to a subset of k out of n users using either fixed-length or variable-length codes.

A. Random Code Construction

The achievability results are derived based on a random codebook construction. In particular, we fix a distribution $\mathbf{q} = [q_1, \dots, q_c]^\mathsf{T}$ (i.e., $q_\ell \geq 0$ and $\sum q_\ell = 1$) over the category labels, then construct the T codewords in the codebook independently in the following fashion. For each codeword $\mathbf{m}^{(t)}$, the codeword elements $m_1^{(t)}, \dots, m_n^{(t)}$ are independently and identically generated according to $\mathbf{q}$, i.e., they take on values $\ell \in [c]$ with probability q_ℓ. We denote such a codebook as $\bar{\mathbf{M}}_\mathbf{q} = (\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(T)})$.

For a codeword $\mathbf{m}^{(t)}$ generated randomly according to $\mathbf{q}$, the probability that it is consistent with an arbitrary categorization $\bar{\mathbf{X}} = (\mathbf{X}_1, \dots, \mathbf{X}_c)$ of some arbitrary fixed set of active users $\mathbf{A} = \bigcup \mathbf{X}_\ell$ can be computed as follows. Since the probability that $\mathbf{m}^{(t)}$ assigns the correct labels to each active user is q_ℓ, we have

$$\Pr\left(\bar{\mathbf{X}} \in \mathcal{C}(\mathbf{m}^{(t)})\right) = \prod_{\ell=1}^{c} q_\ell^{k_\ell} \triangleq p. \tag{14}$$

This value p is a crucial quantity and is useful in the proofs of both the fixed-length and variable-length cases.

B. Fixed-Length Code

We now establish an achievable rate to communicate a categorization to a subset of k out of n users using a fixed-length code. The aim is to show that when T exceeds a threshold, there must exist at least one codebook that covers all categorizations in $\mathcal{D}^{(n,\mathbf{k})}$.

The idea is to fix the generating distribution $\mathbf{q}$ and to construct a bound on the probability that the covering condition (13) holds for a random codebook $\bar{\mathbf{M}}_\mathbf{q}$. As long as the above probability is nonzero for a given value of T, it follows that there must exist one codebook with rate $R_\mathrm{f} = \log T$ that covers all categorizations. This rate R_f can be further minimized by optimizing over the generating distribution. The main result of this section is formally stated as follows.

Theorem 1: Fix the total number of users n, the number of active users k, and the category sizes $\mathbf{k} = [k_1 \dots k_c]^\mathsf{T}$, with $k = \sum_{\ell=1}^{c} k_\ell$. The minimum rate R_f^* of a fixed-length code for communicating a categorization of k out of n users is bounded above as

$$\begin{aligned} R_\mathrm{f}^*(n,\mathbf{k}) \leq\ & kH\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right) + \log(k) \\ & + \log\left(\ln\left(\frac{n}{k}\right) + \ln(2) H\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right) + 1\right). \end{aligned} \tag{15}$$

For fixed $\mathbf{k}$, the above expression scales as $O(\log\log(n))$.

Proof: Consider a random codebook $\bar{\mathbf{M}}_\mathbf{q}$ consisting of T codewords generated independently according to $\mathbf{q}$. Let $\bar{\mathbf{X}}$ be an arbitrary categorization. The probability that $\bar{\mathbf{X}}$ is not covered by any codeword in the $\bar{\mathbf{M}}_\mathbf{q}$ can be seen as

$$\Pr\left(\bar{\mathbf{X}} \notin \bigcup_{t=1}^{T} \mathbf{C}(\mathbf{m}^{(t)})\right) = (1-p)^T, \tag{16}$$

where p is defined as in (14).

Now, consider all possible categorizations in $\mathcal{D}^{(n,\mathbf{k})}$. The probability that the codebook fails to cover every categoriza-

tion in $\mathcal{D}^{(n,\mathbf{k})}$, i.e., it fails to satisfy the condition (13), can be upper bounded using the union bound:

$$\begin{aligned}\Pr\left(\mathcal{D}^{(n,\mathbf{k})} \neq \bigcup_{t=1}^{T} \mathbf{C}(\mathbf{m}^{(t)})\right) &\leq \sum_{\bar{\mathbf{X}}\in\mathcal{D}^{(n,\mathbf{k})}} \Pr\left(\bar{\mathbf{X}} \notin \bigcup_{t=1}^{T} \mathbf{C}(\mathbf{m}^{(t)})\right), \\ &= |\mathcal{D}^{(n,\mathbf{k})}|(1-p)^T, \\ &< |\mathcal{D}^{(n,\mathbf{k})}|e^{-pT}, \end{aligned} \tag{17}$$

where the inequality $(1-x) < e^{-x}, \forall x > 0$, is used in the last step. Now, the cardinality of $\mathcal{D}^{(n,\mathbf{k})}$ can be computed by first choosing k out of n active users then subsequently partitioning the active set into subsets of sizes $k_1, \ldots, k_c$, i.e.,

$$|\mathcal{D}^{(n,\mathbf{k})}| = \binom{n}{k}\binom{k}{k_1 \cdots k_c}. \tag{18}$$

Based on (17), if we let

$$\binom{n}{k}\binom{k}{k_1 \cdots k_c} e^{-pT} \leq \epsilon, \tag{19}$$

for $0 \leq \epsilon < 1$, then $\Pr\left(\mathcal{D}^{(n,\mathbf{k})} \neq \bigcup_{t=1}^{T} \mathbf{C}(\mathbf{m}^{(t)})\right) < \epsilon$, and the covering condition (13) holds for $\bar{\mathbf{M}}_{\mathbf{q}}$ with a non-vanishing probability of at least $1-\epsilon$. Accordingly, this implies the existence of a codebook of size T that covers all possible categorizations in $\mathcal{D}^{(n,\mathbf{k})}$, as long as

$$T \geq \frac{1}{p} \ln\left(\frac{1}{\epsilon}\binom{n}{k}\binom{k}{k_1 \cdots k_c}\right). \tag{20}$$

By taking logarithm on both sides, it follows that the rate

$$R \geq -\log(p) + \log\left(\ln\left(\frac{1}{\epsilon}\binom{n}{k}\binom{k}{k_1 \cdots k_c}\right)\right) \tag{21}$$

is achievable for any values of $\epsilon < 1$ and $\mathbf{q}$.

We can now minimize the right-hand side of (21) by letting $\epsilon \to 1$ and setting $\mathbf{q} = \mathbf{q}^*$ that minimizes $-\log p$. Such a $\mathbf{q}^*$ can be found by solving the optimization problem

$$\underset{\mathbf{q}}{\text{minimize}} \quad -\log \prod_{\ell=1}^{c} q_\ell^{k_\ell} \tag{22}$$

$$\text{subject to} \quad \sum_{\ell=1}^{c} q_\ell = 1, \tag{23}$$
$$q_\ell \geq 0, \quad \forall \ell \in [c].$$

To solve (22), we express the objective as

$$-\log \prod_{\ell=1}^{c} q_\ell^{k_\ell} = \sum_{\ell=1}^{c} -k_\ell \log q_\ell = kH(\mathbf{p}) + kD(\mathbf{p}||\mathbf{q}), \tag{24}$$

where $\mathbf{p} = \left[\frac{k_1}{k} \ldots \frac{k_c}{k}\right]^{\mathsf{T}}$. Since $D(\mathbf{p}||\mathbf{q}) \geq 0$, with equality if and only if $\mathbf{q} = \mathbf{p}$, the optimum $\mathbf{q}^* = \left[\frac{k_1}{k} \ldots \frac{k_c}{k}\right]^{\mathsf{T}}$, and the minimum value of the objective is $kH\left(\frac{k_1}{k}, \ldots, \frac{k_c}{k}\right)$. Substituting this into (21), we conclude that any rate

$$R > kH\left(\frac{k_1}{k}, \ldots, \frac{k_c}{k}\right) + \log\left(\ln\left(\binom{n}{k}\binom{k}{k_1 \cdots k_c}\right)\right)$$

is achievable. This implies that the minimum rate R_{f}^* must be less than the right-hand side of the above. Finally, using the inequalities $\binom{n}{k} < \left(\frac{en}{k}\right)^k$ and $\binom{k}{k_1 \cdots k_c} \leq 2^{kH\left(\frac{k_1}{k}, \ldots, \frac{k_c}{k}\right)}$, we arrive at the desired result. ∎

C. Variable-Length Code

We derive an achievable rate for the variable-length case using a different strategy as compared to the fixed-length case. Instead of finding the smallest T such that the codebook covers every categorization, we let the codebook size be arbitrarily large, i.e., $T \to \infty$. Denote such a codebook as $\bar{\mathbf{M}}_{\mathbf{q}}^{\infty} = \left(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots\right)$. In this case, a given realization of the set of active users and a given categorization would be covered by infinitely many codewords in the codebook. To minimize the entropy of the encoder output, we design the encoding function to be one that searches through the codewords in the codebook sequentially and outputs the index of the first codeword that covers the given categorization, i.e.,

$$f_{\mathrm{g}}(\bar{\mathbf{X}}) = \min \quad t \quad \text{s.t.} \quad \bar{\mathbf{X}} \in \mathcal{C}(\mathbf{m}^{(t)}). \tag{25}$$

This greedy encoder is used in [2] to derive the achievability bounds on the variable-length code rate of a common feedback message for user scheduling. We use the same encoder herein due to its ability to produce highly skewed distributions (i.e., toward the first few codewords), which reduces the output entropy. Our main result for the variable-length case is:

Theorem 2: Fix the total number of users n, the number of active users k, and the category sizes $\mathbf{k} = [k_1 \ldots k_c]^{\mathsf{T}}$, with $k = \sum_{\ell=1}^{c} k_\ell$. The minimum rate R_{v}^* of a variable-length code for communicating a categorization of k out of n users is bounded above as

$$R_{\mathrm{v}}^*(n,\mathbf{k}) < \; kH\left(\frac{k_1}{k}, \ldots, \frac{k_c}{k}\right) + \log(e). \tag{26}$$

Proof: Let the set of active users and their categorization be uniform under fixed $n, k_1, \ldots, k_c$, (which is the worst case distribution). Fix a distribution $\mathbf{q}$ and construct an infinite-size codebook $\bar{\mathbf{M}}_{\mathbf{q}}^{\infty} = \left(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots\right)$. We use the greedy encoder in (25) and let $f_{\mathrm{g}}(\bar{\mathbf{X}})$ be a random variable denoting the output of such encoder. It can be seen that the encoder operation is equivalent to performing successive independent trials, with $f_{\mathrm{g}}(\bar{\mathbf{X}}) = t$ if and only if the first $t-1$ trials are unsuccessful and the t-th trial is successful, where the success probability is p as given in (14).

Since the codewords in $\bar{\mathbf{M}}_{\mathbf{q}}^{\infty}$ are constructed independently, it follows that the encoder output is a geometrically distributed random variable with a success probability p, whose entropy is given by

$$\begin{aligned} H\left(f_{\mathrm{g}}(\mathbf{X})\right) &= -\log(p) - \frac{1-p}{p}\log(1-p), \\ &< -\log(p) + \log(e), \end{aligned} \tag{27}$$

where the inequality $-\frac{1-p}{p}\log(1-p) < \log(e)$, for $0 \leq p \leq 1$ is used in the last step. Note that (27) holds for any distribution $\mathbf{q}$, including $\mathbf{q}^* = \left[\frac{k_1}{k}, \ldots \frac{k_c}{k}\right]^{\mathsf{T}}$ that minimizes $-\log(p)$ as in (22). Substituting $\mathbf{q} = \mathbf{q}^*$ in $-\log(p)$ yields the desired result. ∎

V. Converse

We now present a converse which shows that the minimum rate required to communicate a categorization to a subset of k out of n users must have at least a linear scaling in k with a scaling factor equal to the entropy of the categorization for both the fixed-length and variable-length cases. This result is based on the volume bound, which is used in [2] to prove a converse for the scheduling problem. The main idea is to characterize the maximum number of categorizations that can be covered by a codeword and to use this to lower bound the minimum number of codewords needed to satisfy the covering condition (13).

Theorem 3: Fix the total number of users n, the number of active users k, and the category sizes $\mathbf{k} = [k_1 \dots k_c]^\mathsf{T}$, with $k = \sum_{\ell=1}^{c} k_\ell$. The minimum rate R_{f}^* of a fixed-length code for communicating a categorization of k out of n users is bounded below as

$$R_{\mathrm{f}}^*(n, \mathbf{k}) \geq kH\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right) - \log\left(\frac{n^k}{n^{\underline{k}}}\right). \tag{28}$$

The same bound also applies to the variable-length code, i.e., $R_{\mathrm{v}}^*(n, \mathbf{k})$.

Proof: To construct a volume bound, we need to characterize the maximum number of categorizations that can be covered by a single codeword $\mathbf{m} \in [c]^n$. We let $d_{\max}$ denote this maximum number. We also let $\mathbf{a} = [a_1, \dots, a_c]^\mathsf{T}$, where a_ℓ denotes the number of entries in $\mathbf{m}$ that are equal to ℓ and $\sum_{\ell=1}^{c} a_\ell = n$. We can explicitly compute $d_{\max}$ as follows

$$d_{\max} = \max_{\mathbf{a}} \prod_{\ell=1}^{c} \binom{a_\ell}{k_\ell} \leq \max_{\mathbf{a}} \prod_{\ell=1}^{c} \frac{a_\ell^{k_\ell}}{k_\ell!}. \tag{29}$$

Next, we find the optimal $\mathbf{a}^*$ that maximizes the upper bound on $d_{\max}$. Note that the optimization of the above upper bound has the same form as (22), so the optimal $a_\ell^* = \frac{nk_\ell}{k}$. Thus, $d_{\max}$ can be upper bounded as follows

$$d_{\max} \leq \prod_{\ell=1}^{c} \frac{\left(\frac{nk_\ell}{k}\right)^{k_\ell}}{k_\ell!}. \tag{30}$$

To complete the volume bound, we divide $|\mathcal{D}^{(n,\mathbf{k})}|$ by $d_{\max}$ to obtain a lower bound on the minimum number of codewords needed to satisfy the condition that the codebook covers $|\mathcal{D}^{(n,\mathbf{k})}|$, i.e., (13),

$$T^*(n, \mathbf{k}) \geq \binom{n}{k} \binom{k}{k_1 \cdots k_c} \frac{1}{d_{\max}}. \tag{31}$$

Taking the logarithm and applying (30), after some algebra, we derive a lower bound on $R_{\mathrm{f}}^*(n, \mathbf{k})$ as

$$R_{\mathrm{f}}^*(n, \mathbf{k}) \geq kH\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right) - \log\left(\frac{n^k}{n^{\underline{k}}}\right). \tag{32}$$

For the variable-length case, consider the optimal encoder f^*. Let $\mathbf{r}^* = [r_1^*, \dots, r_\infty^*]^\mathsf{T}$ denote the distribution of the output of encoder f^* given that the input categorizations are distributed uniformly over $\mathcal{D}^{(n,\mathbf{k})}$ (which can be shown to maximize the lower bound of the encoder output entropy), i.e., r_t^* denotes the probability that the encoder f^* outputs t. Since each codeword can only cover up to $d_{\max}$ different categorizations, we have

$$r_t^* \leq \frac{d_{\max}}{\binom{n}{k}\binom{k}{k_1 \cdots k_c}}. \tag{33}$$

This gives a lower bound on the entropy of the output of encoder f^* as follows:

$$\begin{aligned} H(\mathbf{r}^*) &= -\sum_t r_t^* \log(r_t^*) \geq -\sum_t r_t^* \log\left(\frac{d_{\max}}{\binom{n}{k}\binom{k}{k_1 \cdots k_c}}\right) \\ &= \log\left(\binom{n}{k}\binom{k}{k_1 \cdots k_c}\frac{1}{d_{\max}}\right). \end{aligned} \tag{34}$$

The term inside the logarithm on the right-hand side of (34) is identical to the right-hand side of (31). Therefore the converse for $R_{\mathrm{f}}^*(n, \mathbf{k})$ also holds for $R_{\mathrm{v}}^*(n, \mathbf{k})$. ∎

Comparing the upper bounds (15) and (26) with the lower bound (28), we see that the minimum rate for communicating a categorization is essentially $H\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right)$ bits per active user. The scaling coefficients in the leading terms in the upper and lower bounds match exactly. Note that for the lower bound, in the regime of $n \to \infty$ with fixed k, $\log\left(\frac{n^k}{n^{\underline{k}}}\right) \to 0$.

VI. Concluding Remarks

In a conventional multiuser system, categorizing k users into c categories with $k_1, \dots, k_c$ users in each category requires the transmission of k labels. We can compress the labels according to their frequencies of occurrences, resulting in an overall rate of $kH\left(\frac{k_1}{k}, \dots, \frac{k_c}{k}\right)$ bits. We refer to this quantity as the *categorization entropy*.

The main result of this paper is that in a massive random access scenario, where a *random* subset of k active users among a *massive* number of n potential users are to be categorized, a BS can communicate the categorization to the active users using a code of essentially the same rate, plus an overhead of $O(\log\log(n))$ bits for the fixed-length code case and $\log(e)$ bits for the variable-length code case. In other words, an arbitrary categorization of the active users can be communicated without having to explicitly broadcast the identities of the active users, which would have incurred a cost of $k\log(n)$ bits. This result can be thought of as a counterpart to *coded scheduling* result in [2], which shows that scheduling k out of n users into k distinct slots requires essentially only $k\log(e)$ bits, also without the $k\log(n)$ cost.

As a final remark, the development of this paper assumes that the sizes of the categorization $k_1, \dots, k_c$ are known and the codebook is generated to match this distribution. What if the codebook is designed according to $\mathbf{q}$, while the true category sizes are $\mathbf{p}$? An examination of the proofs reveals that it would incur an extra $kD(\mathbf{p}||\mathbf{q})$ cost, as evident in (24). Interestingly, this is exactly the same extra cost as in the classical data compression theory when the true source distribution is $\mathbf{p}$, while the compression codebook is designed according to a different distribution $\mathbf{q}$.

REFERENCES

[1] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[2] J. Kang and W. Yu, "Minimum feedback for collision-free scheduling in massive random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, pp. 8094–8108, Dec. 2021.

[3] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance based massive random access with massive MIMO," *IEEE Trans. Inf. Theory*, vol. 6, no. 3, pp. 1696–1715, Mar. 2022.

[4] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.

[5] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[6] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, 2018.

[7] A. E. Kalør, R. Kotaba, and P. Popovski, "Common message acknowledgements: Massive ARQ protocols for wireless access," [Online]. Available: https://arxiv.org/abs/2201.03907, 2022.