Coding against delayed adversaries

Bikash Kumar Dey
IIT Bombay
bikash@ee.iitb.ac.in

Sidharth Jaggi
CUHK
jaggi@ie.cuhk.edu.hk

Michael Langberg
Open University of Israel
mikel@openu.ac.il

Anand D. Sarwate
UCSD
asarwate@ucsd.edu

***Abstract*—In this work we consider the communication of information in the presence of a *delayed* adversarial jammer. In the setting under study, a sender wishes to communicate a message to a receiver by transmitting a codeword $\mathbf{x} = (x_1, \ldots, x_n)$ over a communication channel. The adversarial jammer can view the transmitted symbols x_i one at a time, but must base its action (when changing x_i) on x_j for $j \leq i - \Delta n$, where $\Delta \in [0, 1]$ is a *delay* parameter. In this work, we study codes for a class of delayed adversaries, and for any delay $\Delta > 0$ present a single letter characterization of the achievable communication rate in the presence of such adversaries.**

I. Introduction

In this paper we study a class of communication channels whose output $\mathbf{y}$ is the result of an adversary maliciously tampering with the channel input $\mathbf{x}$. The adversary is constrained in two ways: (a) it must satisfy certain *causality* or *delay* conditions and (b) it must satisfy certain *power* constraints. For (a), we restrict our adversary to only see the channel input after a delay equal to a fraction of the total transmission time. For (b), we allow the adversary to tamper with at most a predefined fraction of the input symbols transmitted over time. For example, for binary channels with modulo-2 additive adversarial interference, for a delay parameter $\Delta \in [0, 1]$, in a total block-length of n channel uses, the adversary must base its action at time i on the channel input up to time $i - \Delta n$. Moreover, for a parameter $p \in [0, 1]$ the total number of symbols that can be tampered with is assumed to be bounded from above by pn.

In this work for a class of communication problems (including, for instance the modulo-2 additive adversarial interference problem mentioned above) we present single letter characterizations of the communication capacity in the presence of such adversaries. Roughly speaking, if randomization at the encoder is allowed, we show that for the problems in our class, for any $\Delta > 0$ the capacity is equal to that of a certain discrete memoryless channel (DMC) induced by an adversary whose actions are i.i.d. and independent of the channel input. It is known [6] that for $\Delta = 0$ this is not true.

Communication schemes that protect data transmission against adversarial jammers have been studied in many different fields of electrical engineering and computer science. Much of classical coding theory is based on minimum distance considerations and provides guarantees against all errors of up to a given Hamming weight. Another way to see this is that the errors are generated adversarially (in a worst-case manner) with respect to both the code and the actual codeword transmitted. The information-theoretic framework for these questions is the theory of arbitrarily varying channels (AVCs) [1] with constraints [2]. In terms of delay, the AVC literature has considered the case where the adversary has no knowledge of the transmitted codeword, can see the codeword strictly causally, or has access to the full codeword prior to deciding its actions [3]. The works most closely related to this one are [4]–[8], which study causally delayed binary adversarial channels and channels with large alphabets.

There are many phenomena which can cause delay in the adversary's observations. There may be delay due to physical distance and the difference in signal propagation times. Computational overhead can also cause delay between when the signal is received and when the adversary can take action based on the information (see [9], [10] for a perspective on coding against computationally limited adversaries). If data transmission is packetized, then delay can be induced by the packet size itself.

In this paper we consider a class of channels in which the adversary may opt to do nothing, in which case $y_i = x_i$, or can impose a permutation π so that $y_i = \pi(x_i)$. We call such channels *permutation channels*, and for a subset of permutation channels we give a tight characterization of the rate region. The subset we study includes for example *additive channels* in which the alphabet is assumed to have an algebraic structure and the errors imposed by the channel are added onto the information transmitted. Our coding techniques are fairly general and we believe they can be extended to a much broader class of channels. The details are deferred to the full version of this paper.

II. Channel model

We generally model the channel as an arbitrarily varying channel (AVC) $\{W(y|x, z) : z \in \mathcal{Z}\}$ with finite input, output, and state alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ respectively. The state $z_i \in \mathcal{Z}$ at time i is chosen by an adversary who wishes to prevent reliable communication across the channel. Over a block of n symbols with input $\mathbf{x} \in \mathcal{X}^n$ and state $\mathbf{z} \in \mathcal{Z}^n$, the probability of an output sequence $\mathbf{y} \in \mathcal{Y}^n$ is given by $W^n(\mathbf{y}|\mathbf{x}, \mathbf{z}) = \prod_{i=1}^{n} W(y_i|x_i, z_i)$. Exponentials and logarithms are base 2 unless otherwise specified.

The encoder (Alice) wishes to transmit a message $m \in [2^{nR}]$ at rate R using a code of block-length n to a receiver (Bob) over the channel. To do this, Alice uses an *encoding map* $\Phi : [2^{nR}] \to \mathcal{X}^n$. In this paper we consider the case

Authors are in alphabetical order. Work supported in part by ISF grant 480/08, RGC GRF grants 412608 and 412809, RGC AoE grant on Institute of Network Coding, established under the University Grant Committee of Hong Kong, CUHK MoE-Microsoft Key Laboratory of Humancentric Computing and Interface Technologies, and the Bharti Centre for Communication in IIT Bombay, India.

where the encoding map Φ is *stochastic* and we will write it as $\Phi : [2^{nR}] \times [2^{nS}] \to \mathcal{X}^n$, where $[2^{nS}]$ represents nS bits corresponding to private random coins at the encoder. We stress that the encoder randomness is neither shared with the receiver Bob nor the adversary governing the channel states. In particular, Bob uses a decoding map $\Psi : \mathcal{Y}^n \to [2^{nR}]$ to decode the message sent by Alice.

We assume that the adversary knows the pair (Φ, Ψ) used by Alice and Bob, and may also eavesdrop on the channel with delay in order to choose its input $\mathbf{z}$. More formally, for a delay parameter $\Delta \in [0, 1]$, an *adversarial strategy* $\mathcal{A}$ *with delay* Δ is a sequence of maps $A_i : \mathcal{X}^{(i-\Delta n)^+} \to \mathcal{Z}$, where for $i \leq \Delta n$ there is no dependence on $\mathcal{X}$. In particular, the ith adversarial action A_i depends only on the channel inputs $x_1, \ldots, x_{i-\Delta n}$.[1] For instance, if $\Delta = 0$ then the adversary can base z_i on the current input symbol x_i, and if $\Delta = 1$ then the adversary must choose its action without knowledge of any input symbols.

Any distribution on $\mathcal{Z}$ is called an *action profile*. We write $T_{\mathbf{z}}$ for the type of a sequence $\mathbf{z}$. In this paper we constrain the actions of the adversary by insisting that its empirical action profile $T_{\mathbf{z}}$ in each length-n block be from an *admissible action set* $\mathcal{Q}$, that is, an open convex set of probability distributions. A special case of this is the cost-constrained AVC [2], [11]. Given a code (Φ, Ψ) we say an adversarial strategy $\mathcal{A}$ is *admissible* under $\mathcal{Q}$ if for all m and s it holds that $T(\mathcal{A}(\Phi(m, s))) \in \mathcal{Q}$. Let $\mathbf{A}(\mathcal{Q})$ denote the set of all admissible adversarial strategies. For example, for binary alphabets with $y_i = x_i \oplus z_i$ and an adversary constrained to flip no more than pn bits, $\mathcal{Q}$ is all distributions $(1-q, q)$ on $\mathcal{Z}$ with $q \leq p$.

The *maximal error* for a code (Φ, Ψ) and adversarial strategy $\mathcal{A}$ is given by $\varepsilon(\mathcal{A}) = \max_{m \in [2^{nR}]} \mathbb{P}\left(\Psi(\mathbf{Y}) \neq m \mid \Phi(m, \cdot), \mathcal{A}\right)$, where the probability is taken over the randomness in the encoder and channel. The maximal error for a code (Φ, Ψ) and the set of adversarial strategies is $\varepsilon(\mathcal{Q}) = \max_{\mathcal{A} \in \mathbf{A}(\mathcal{Q})} \varepsilon(\mathcal{A})$.

We say a rate R is *achievable* for maximal error under stochastic encoding against adversaries with delay Δ if for every $\delta > 0$ there exists a blocklength n and a code (Φ_n, Ψ_n) whose maximal error $\varepsilon(\mathcal{Q}) \leq \delta$. The supremum of achievable rates is the *capacity* $C_\Delta(\mathcal{Q})$.

A. Permutation channels

In this paper we consider what we call *permutation channels* in which $\mathcal{X} = \mathcal{Y}$ and $\mathcal{Z}$ is a subset of permutations on $\mathcal{X}$. Under an action/state vector $\mathbf{z} \in \mathcal{Z}^n$ we have $\forall i : y_i = z_i(x_i)$.

A special example of a permutation channel is a *modulo-additive channel* in which $\mathcal{Z} = \mathcal{X} = \mathcal{Y} = \{0, 1, \ldots, |\mathcal{X}| - 1\}$ and $y_i = x_i + z_i \mod |\mathcal{X}|$. In this case the permutations are all cyclic shifts over $\mathcal{X}$. For the binary channel $|\mathcal{X}| = 2$, there is only one shift in $\mathcal{Z}$ which is not the identity. A natural constraint is that the Hamming weight of the adversary's input $\mathbf{z}$ satisfy $w_H(\mathbf{z}) < pn$ for some fixed p.

[1] Here we assume that the adversarial strategy is deterministic. As our proofs take a *worst case* analysis over our adversarial model, our assumption is w.l.o.g.

B. The random adversary

A simple adversarial strategy is to generate $\mathbf{z}$ i.i.d. from some distribution $Q \in \mathcal{Q}$. Since Q is in the interior of open set $\mathcal{Q}$, with high probability, for sufficiently large n the realization of $\mathbf{z}$ satisfies the constraints on the adversarial empirical action profile. We call this a *random adversary*. A random adversary induces an *average channel*

$$W_Q(y|x) = \sum_{z \in \mathcal{Z}} W(y|x, z) Q(z). \tag{1}$$

For a given input distribution P, we can calculate the mutual information $I(P, W_Q)$ between X and Y with distribution $P(x) W_Q(y|x)$. It can be directly verified that the capacity of the channel in the presence of such random adversaries is

$$C(\mathcal{Q}) = \max_{P(x)} \min_{Q \in \mathcal{Q}} I(P, W_Q). \tag{2}$$

The minimization over all of $\mathcal{Q}$ is justified by the continuity of mutual information; we can approach any point on the boundary of $\mathcal{Q}$ by a suitable sequence.

C. Uniformizable admissible action set $\mathcal{Q}$

We denote the set of channels (as in (1)) the adversary can induce by using action profiles from $\mathcal{Q}$ by $\mathcal{W}_{\mathcal{Q}}$, i.e., $\mathcal{W}_{\mathcal{Q}} = \{W_Q | Q \in \mathcal{Q}\}$. If for a given set $\mathcal{Q}$, the uniform distribution P_u on $\mathcal{X}$ achieves the maximum in (2), we call $\mathcal{Q}$ (and the channel set corresponding to $\mathcal{Q}$) *uniformizable*. Our results in this paper will be presented for the class of channels for which $\mathcal{Q}$ is uniformizable. For example, the additive modulo-channel mentioned above has a corresponding set $\mathcal{Q}$ which is uniformizable. Additional examples are deferred to the full version of this paper.

The reader familiar with the literature on arbitrarily varying channels (AVCs) will note that the quantity in (2) is also the randomized coding capacity for the AVC [11]. For the class of permutation channels, (2) is also the same as the deterministic coding capacity under average error [2].

III. Main results

Our main result is a characterization of the capacity for uniformizable permutation channels whose adversaries have positive delay Δ. In particular, we show that the capacity is the same as under the worst-case random adversary satisfying the constraints.

Theorem 1 (Uniformizable permutation channels)**.** *Consider a uniformizable permutation channel with adversarial constraints $\mathcal{Q}$. Then for $\Delta > 0$ we have $C_\Delta(\mathcal{Q}) = C(\mathcal{Q})$, where $C(\mathcal{Q})$ is defined in (2).*

Corollary 1 (Binary Δ-delay)**.** *For $p \leq 1/2$, consider the binary modulo-additive channel where $\mathcal{Q} = \{(q, 1-q) : q < p\}$ imposes a constraint on the fraction of "additions of 1" in $\mathbf{z}$. Then for $\Delta > 0$ we have $C_\Delta(\mathcal{Q}) = 1 - H(p)$.*

IV. Analysis

In this section we prove Theorem 1. Let (P,Q) denote a saddle-point in the max-min expression (2), so that Q is a minimizing random adversarial strategy and P is a maximizing input distribution. (Recall that we assume throughout that $P = P_u$ is uniform[2].) Let P^n be the natural product probability distribution over $\mathcal{X}^n$. Without loss of generality we assume that the delay Δ is rational.

In our proof we use strong typicality and we need the following facts about strongly typical sequences. For a vector $\mathbf{x} \in \mathcal{X}^k$ and $x \in \mathcal{X}$, let $N_x(\mathbf{x})$ denote the number of times x appears in $\mathbf{x}$ and recall $T_{\mathbf{x}}$ is the type of $\mathbf{x}$. For a distribution P with $\min_x P(x) > 0$, the ϵ-typical set is $T^{(\epsilon,k)}(P) = \{\mathbf{x} \in \mathcal{X}^k : \|T_{\mathbf{x}} - P\|_\infty \leq \epsilon\}$. The size of the typical set $T^{(0,k)}(P)$ is [12, p. 39]: $|\{\mathbf{x} : T_{\mathbf{x}} = P\}| = \exp(k(H(P) + o(1)))$. From [13] we have that if $\|P - P'\|_\infty < \epsilon$ then $|H(P) - H(P')| = O(\epsilon \log \epsilon^{-1})$, so we have $|T^{(\epsilon,k)}(P)| \leq \exp\left(k(H(P) + O(\epsilon \log \epsilon^{-1}))\right)$. We denote the set of all types of sequences of length k by $\mathcal{T}_k(\mathcal{X})$.

We now show that $C_\Delta(\mathcal{Q}) \geq C(\mathcal{Q})$ (the achievability part of the proof). The converse proof $C_\Delta(\mathcal{Q}) \leq C(\mathcal{Q})$ is straightforward and deferred to the full version of this paper due to space limitations.

A. The code

For any given Δ we choose an integer N such that $\Delta > 1/N$ and prove the achievability for delay $1/N$. So, without loss of generality, we consider a normalised delay of the form $\Delta = 1/N$ and we then consider code-lengths n which are multiples of N. Our code divides the total block length of n into chunks of size $k = \Delta n$ and let $\ell = 1/\Delta$ be the number of chunks. By assumption, both ℓ and k are integral. Let $\mathbf{x}_j$, $\mathbf{z}_j$, and $\mathbf{y}_j$ denote the j-th chunks of the codeword $\mathbf{x}$, adversary's sequence $\mathbf{z}$, and the output $\mathbf{y}$ respectively. We assume the secret $s \in [2^{nS}]$ to consist of ℓ parts $s_1, s_2, \ldots, s_\ell \in [2^{kS}]$. For an input distribution P, rate R, and secret size S, we define the following random variable $\mathbf{\Phi}$ taking values in the set of (stochastic) codes.

1) Generate the codeword chunks $\{\mathbf{X}_j(m, s_j) \in \mathcal{X}^k : j \in [\ell],\ m \in [2^{nR}]\},\ s_j \in [2^{kS}]\}$ i.i.d. according to P^k.
2) To encode the message-secret pair $(m, s) \in [2^{nR}] \times [2^{nS}]$, we set $\mathbf{X}(m, s) = (\mathbf{X}_1(m, s_1), \ldots, \mathbf{X}_\ell(m, s_\ell))$.

For a fixed realization Φ of $\mathbf{\Phi}$ with codewords $\{\mathbf{x}(m, s)\}$, the decoder Bob decodes using the following procedure.

1) Let $\epsilon > 0$. For each feasible type $Q \in \mathcal{Q} \cap \mathcal{T}_n(\mathcal{Z})$ and for each decomposition of Q into $\vec{Q} = Q_1, Q_2, \ldots, Q_\ell \in \mathcal{T}_k(\mathcal{Z})$ such that $Q = \frac{1}{\ell}\sum_{j=1}^{\ell} Q_j$, Bob defines a set of *lists* $L_j(\vec{Q})$. The initial list is defined to be $L_0(\vec{Q}) = [2^{nR}]$. For each $j = 1, 2, \ldots, \ell$, the list $L_j(\vec{Q})$ is the set of messages $m \in L_{j-1}(\vec{Q})$ such that there exists an $s_j \in [2^{kS}]$ such that
   a) The chunk $\mathbf{x}_j(m, s_j) \in T^{(\epsilon,k)}(P)$.
   b) $(\mathbf{x}_j(m, s_j), \mathbf{y}_j) \in T^{(|\mathcal{Z}|\epsilon,k)}(P \times W_{Q_j})$ where W_{Q_j} is given by (1).
2) If $\bigcup_{\vec{Q} \in (\mathcal{T}_k(\mathcal{Z}))^l, Q \in \mathcal{Q}} L_l(\vec{Q}) = \{\hat{m}\}$ then Bob declares $\hat{m}$ as the transmitted message. If the set is empty or contains more than one messages then he declares an error.

We now prove that setting R to be $C(\mathcal{Q}) - \delta$, S to be $\delta/2$, and ϵ such that $\epsilon = O(\delta^2)$ our encoding/decoding scheme succeeds with high probability. More precisely, we show that with high probability over our encoder construction, the resulting code (Φ, Ψ) allows communication with high probability. In Lemma 1 we show that the transmitted codeword, adversarial input, and channel output satisfy some joint typicality conditions with high probability. In Lemma 2 we show that the size of the lists excluding the correct message go to 0 with high probability. In Lemma 3 we show that the true message is in some list with high probability.

B. Technical lemmas

We now need a few technical lemmas to show that the random code $\mathbf{\Phi}$ satisfies certain structural properties which prove useful in the proof of the main theorems. Define $\mathbf{\Phi}_{n,k}$ as a random variable with the distribution of the code $\mathbf{\Phi}$ in a single chunk, with $k = \Delta n$. That is, $\mathbf{\Phi}_{n,k}$ assigns to a pair $(m, s_j) \in [2^{nR}] \times [2^{kS}]$ an i.i.d. codeword $\mathbf{X}(m, s_j) \in \mathcal{X}^k$ according to P^k. The following lemma shows that w.h.p. over the code $\mathbf{\Phi}_{n,k}$ in a particular chunk (say j), the transmitted message m survives in the output list L_j if it is in the input list L_{j-1} and if the decoder's assumed action profile and the true action profile of the adversary are the same in that chunk. The proof is a direct application of concentration inequalities and is omitted.

Lemma 1. *Let $0 < \epsilon < 1$. Let the block-length n be sufficiently large, and $k = \Theta(n)$. Then with probability greater than $1 - \exp(-\exp(k(S - 2\epsilon^2)))$, the realization of the code $\mathbf{\Phi}_{n,k}$ satisfies, for all $Q \in \mathcal{T}_k(\mathcal{Z})$, all adversarial actions $\mathbf{z} \in \mathcal{Z}^k$ of type Q and all messages m,*

$$\mathbb{P}\left((\mathbf{x}(m, s), \mathbf{y}) \notin T^{(|\mathcal{Z}|\epsilon,k)}(P \times W_Q)\right) \leq \exp(-k\epsilon^2/2), \tag{3}$$

where the probability is taken over the secret randomness $s \in [2^{kS}]$ of the encoder. Here $\mathbf{y} = \mathbf{z}(\mathbf{x}(m, s))$ denotes the received vector when $\mathbf{x}(m, s)$ is transmitted and the adversary acts by $\mathbf{z}$.

For $\mathbf{x} \in \mathcal{X}^k$ and $\epsilon > 0$ define the set

$$\mathcal{D}(\mathbf{x}, Q, \epsilon) = \left\{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in T^{(|\mathcal{Z}|\epsilon,k)}(P \times W_Q)\right\}. \tag{4}$$

Let $\mathbf{z}(\mathbf{x})$ denote the output $\mathbf{y}$ formed by applying permutation maps $\mathbf{z}$ to an input $\mathbf{x}$.

The next lemma shows that with overwhelming probability over the code $\mathbf{\Phi}_{n,k}$ in a single chunk, the realization of the code is such that with high probability over the secret, the decoder's list size (excluding the transmitted message) will decrease by a certain amount. The decoder for chunk j assumes a particular action profile Q_j that is not necessarily the empirical type of the adversary action $\mathbf{z}_j$. We do a single

[2] We note that a significant portion of our analysis holds for general P, and thus we leave P as a parameter throughout our proofs. Our need for a uniform P appears solely in Lemma 2 in Section IV-B.

chunk analysis and omit the chunk index in the subscript in the lemma. Let $L \subseteq [2^{nR}] \setminus \{m\}$ denote the initial list of the decoder for the chunk excluding the transmitted message m. Since the output of a permutation channel is a deterministic function of the input and adversary's action, for each secret s the output $\mathbf{Y}$ only depends on $\mathbf{X}(m,s)$ and $\mathbf{z}$. The decoder, assuming an action profile Q, sets the new list as

$$\begin{aligned} L' &= L'(m, s, \mathbf{z}, Q) \\ &= \{m' \in L : \exists s' \text{ s.t. } \mathbf{z}(\mathbf{X}(m,s)) \in \mathcal{D}(\mathbf{X}(m', s'), Q, \epsilon)\} \end{aligned}$$

Lemma 2. *Let $L \subsetneq [2^{nR}]$ be a set of messages, and $\epsilon > 0$ sufficiently small. Then for sufficiently large n, and $k = \theta(n)$, with probability greater than $1 - \exp(-\exp(kS/2))$, the realization of the code $\mathbf{\Phi}_{n,k}$ satisfies, for every $m \in [2^{nR}] \setminus L$, $Q \in \mathcal{T}_k(\mathcal{Z}), \mathbf{z} \in \mathcal{Z}^k$,*

$$\begin{aligned} &\mathbb{P}(|L'| \leq |L| \exp(k(3S/2 - I(X;Y)))) \\ &\quad \geq 1 - 2^{-k(S/2 - \sqrt{\epsilon})} \end{aligned}$$

where the probability is taken over the secret randomness $s \in [2^{kS}]$ of the encoder. Here $I(X;Y)$ is computed with respect to the distribution $P \times W_Q$.

Proof: In what follows we use the notation defined above. Note that Q represents the action profile assumed by the decoder and $T_{\mathbf{z}}$ need not equal Q. Let $H(X)$, $H(X|Y)$, and $I(X;Y)$ be computed with respect to $P \times W_Q$ unless otherwise noted. From [12], [13] we see that

$$|\mathcal{D}(\mathbf{x}, Q, \epsilon)| \leq \exp\left(k(H(Y|X) + O(|\mathcal{Z}|\epsilon \log(|\mathcal{Z}|\epsilon)^{-1}))\right).$$

Consider generating all codewords $\mathbf{X}(m', s')$ for $m' \in L$ and all secrets s' via $\mathbf{\Phi}_{n,k}$ using the distribution $P^n = P_u^n$. Let us consider a fixed encoded message m. The code can be viewed as the union of the two sub-codes $\mathbf{F} = \{\mathbf{X}(m', s) : m' \neq m,\ m' \in L,\ s \in [2^{kS}]\}$ and $\mathbf{G} = \{\mathbf{X}(m,s) : s \in [2^{kS}]\}$.

Let us also fix an action vector $\mathbf{z}$ and a secret s. Since any component of $\mathbf{z}$ is a permutation and the components of $X(m,s)$ are uniformly i.i.d., the components of $\mathbf{z}(X(m,s))$ are uniformly iid over the output distribution induced by $W_{T_{\mathbf{z}}}$. So for a given realization of $\mathbf{F}$, and for any $m' \neq m,\ s'$, we have

$$\begin{aligned} &\mathbb{P}_{\mathbf{G}}(\mathbf{z}(X(m,s)) \in \mathcal{D}(\mathbf{x}(m', s'), Q, \epsilon)) \\ &\quad \leq \exp\left(k(-H(Y) + H(Y|X) + O(|\mathcal{Z}|\epsilon \log(|\mathcal{Z}|\epsilon)^{-1}))\right) \\ &\quad = \exp\left(-k(I(X;Y) - O(|\mathcal{Z}|\epsilon \log(|\mathcal{Z}|\epsilon)^{-1}))\right). \end{aligned} \tag{5}$$

Here the probability is taken over $\mathbf{G}$, in particular over $X(m,s)$.

Taking the union bound over s', we have, for each message $m' \neq m$ the probability that there is a codeword $\mathbf{x}(m', s') \in \mathbf{F}$ (for some s') such that $\mathbf{z}(X(m,s)) \in \mathcal{D}(\mathbf{x}(m', s'), Q, \epsilon)$ is bounded from above by $\exp(-k(I(X;Y) - S - \gamma))$, where $\gamma = O(|\mathcal{Z}|\epsilon \log(|\mathcal{Z}|\epsilon)^{-1})$. This immediately implies that the expected list size for a given $\mathbf{F}$ satisfies

$$\mathbb{E}_{\mathbf{\Phi}_{n,k}}[|L'|\ |\mathbf{F}] \leq |L| 2^{kS} 2^{-kI(X;Y)} 2^{k\gamma}. \tag{6}$$

Denote the right side of (6) by $\bar{L}$.

Note that L' depends on the transmitted message m, a fixed secret s in that chunk, the codeword $\mathbf{X}(m,s)$, the sequence $\mathbf{z}$, the assumed type Q, and the remainder of the codebook $\mathbf{F} = \{\mathbf{X}(m', s') : m' \neq m,\ s' \in [2^{kS}]\}$, but not on the codewords $\mathbf{G} = \{\mathbf{X}(m, s') : s' \neq s\}$. For every $s \in [2^{kS}]$ define a variable

$$B_{m,z,Q}(s) = \mathbf{1}\left(|L'(m, s, \mathbf{z}, Q)| \geq \bar{L} 2^{k(S/2 - \gamma)}\right). \tag{7}$$

Now, for fixed m, z, Q, and conditioned on a fixed $\mathbf{F}$, the value of $B_{m,z,Q}(s)$ depends only on $X(m,s)$. In particular, there is a subset $E \subseteq \mathcal{X}^k$, s.t. $B_{m,z,Q}(s) = 1$ if and only if $X(m,s) \in E$. Since $\{X(m,s) : s \in [2^{kS}]\}$ are i.i.d., $\{B_{m,z,Q}(s) : s \in [2^{kS}]\}$ are also i.i.d. (here we stress that we are conditioning on $\mathbf{F}$).

By Markov's inequality,

$$\mathbb{P}_{\mathbf{\Phi}_{n,k}}(B_{m,z,Q}(s) = 1 | \mathbf{F}) \leq 2^{-k(S/2 - \gamma)}.$$

From the above, the conditional mean $\mathbb{E}_{\mathbf{\Phi}_{n,k}}(\sum_s B_{m,z,Q}(s)|\mathbf{F})$ is at most $2^{kS/2 + k\gamma}$. So by the Chernoff bound [14], $\mathbb{P}_{\mathbf{\Phi}_{n,k}}(\sum_s B_{m,z,Q}(s) \geq 2 \cdot 2^{kS/2 + k\gamma}|\mathbf{F})$ is at most $\exp\left(-\frac{1}{3}\exp(kS/2 + k\gamma)\right)$. Since the bound holds for all $\mathbf{F}$, we get $\mathbb{P}_{\mathbf{\Phi}_{n,k}}\left(\sum_s B_{m,z,Q}(s) \geq 2 \cdot 2^{kS/2 + k\gamma}\right)$ is at most $\exp\left(-\frac{1}{3}\exp(k(S/2 + \gamma))\right)$.

That is, with doubly-exponential probability over $\mathbf{\Phi}_{n,k}$, for at most $2 \cdot 2^{kS/2 + k\gamma}$ secrets s the list size corresponding to m will be larger than $\bar{L} 2^{k(S/2 - \gamma)} = |L| 2^{k(3S/2 - I(X;Y))}$. Using a union bound over all messages m, state sequences $\mathbf{z}$, and types Q, plus the facts that $\gamma \to 0$ as $\epsilon \to 0$ and $k = \Theta(n)$ for n sufficiently large we conclude that

$$\begin{aligned} \mathbb{P}_{\mathbf{\Phi}_{n,k}}\Big(\forall m, \mathbf{z}, Q : \quad &\mathbb{P}_s(B_{m,z,Q}(s) = 1) \leq 2 \cdot 2^{-kS/2 + k\gamma}\Big) \\ &\geq 1 - \exp(-\exp(kS/2)), \end{aligned}$$

which suffices to prove our assertion for sufficiently small $\epsilon > 0$ (that satisfies $\sqrt{\epsilon} > \gamma$). ■

Let m be the transmitted message, and assume $m \in L$. Our final technical lemma shows that with high probability over the code $\mathbf{\Phi}_{n,k}$ in a single chunk, the realization of the code is such that with high probability over the secret, if the decoder guesses a type Q_j which happens to be equal to the the empirical type $T_{\mathbf{z}}$ of the adversarial action $\mathbf{z}$, then the list L' includes m. The lemma below is a direct consequence of Lemma 1.

Lemma 3. *Let $L \subset [2^{nR}]$ be a set of messages, and $\epsilon > 0$. Then for sufficiently large n, and $k = \theta(n)$, with probability greater than $1 - \exp(-\exp(k(S - 2\epsilon^2)))$, the realization of the code $\mathbf{\Phi}_{n,k}$ satisfies, for every $m \in L, \mathbf{z} \in \mathcal{Z}^k$, and $Q = T_{\mathbf{z}}$, $\mathbb{P}(m \in L') \geq 1 - 2\exp(-k\epsilon^2/2)$. where the probability is taken over the secret randomness $s \in [2^{kS}]$ of the encoder.*

C. Achievability

In order to prove the correctness of our code, we must characterize the capabilities of the adversary. First, we assume that the adversarial strategy can be based on the transmitted message m. This is justified by the maximal error criterion, which requires vanishing error probability for every message. Second, since the adversary is delayed by $k = \Delta n$ time steps, its action in the j-th chunk can only depend on the codewords $\{\mathbf{x}_i(m, s_i) : i \leq j-1\}$. We use the facts that the sub-codebook $\mathbf{\Phi}_{n,k}$ used in chunk j is drawn independently of the preceding sub-codebooks and that the secret s_j is independent of s_i for $i < j$. This implies that $\mathbf{z}_j$ cannot depend on s_j. Furthermore, we note that the overall type of $\mathbf{z} = \mathbf{z}_1, \ldots, \mathbf{z}_\ell$ chosen by the adversary must lie in $\mathcal{Q}$.

Fix a sufficiently small $\delta > 0$ and set $R = C(\mathcal{Q}) - \delta$. Set $\epsilon > 0$ so that $\delta > 4\ell\sqrt{\epsilon}$ and $S = \delta/2$. Fix an m and let $s_1, s_2, \ldots, s_\ell$ be i.i.d. and uniformly distributed on $[2^{\Delta n S}]$. Let $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_\ell$ be the (random) sub-codes for chunks $1, \ldots, \ell$. Each of these codes is independent and distributed identically to $\mathbf{\Phi}_{n,k}$. Let $\mathbf{y} = \mathbf{y}_1, \ldots, \mathbf{y}_\ell$ be the received word. Consider the decoding process defined in Section IV-A. Namely, fix an overall profile $Q \in \mathcal{Q}$ and decomposition $\vec{Q} = (Q_1, Q_2, \ldots, Q_\ell) \in (\mathcal{T}_k(\mathcal{Z}))^l$ of Q (that satisfies $Q = \frac{1}{\ell}\sum_{j=1}^{\ell} Q_j$). There are fewer than $(k+1)^{|\mathcal{Z}|\ell}$ possible values for $\vec{Q}$, so the number of decompositions is at most polynomial in n.

We set $k = \Delta n$ and consider a given small $\epsilon > 0$. Throughout the proof we say that a sub-code $\mathbf{\Phi}_i$ is *good* with respect to $L_i \subseteq [2^{nR}]$ if for all $m \in L_i$, $\mathbf{z} \in \mathcal{Z}^k$, (i) Lemma 2 is satisfied with $L = L_i \setminus \{m\}$ for all $Q \in \mathcal{T}_k(\mathcal{Z})$ and (ii) Lemma 3 is satisfied with $L = L_i$.

Consider the decoder's action after the first chunk. The decoder starts with a list $L_0 = L_0(\vec{Q}) = L = [2^{nR}]$, and obtains a new list $L_1(\vec{Q})$ (denoted by L' in Lemma 2). The code $\mathbf{\Phi}_1$ is good with probability $1 - \exp(-\exp(\Delta n S/2))$. From this point on we assume the realization Φ_1 is indeed good and treat Φ_1 as fixed. For Φ_1, with probability at least $1 - 2^{-\Delta n(S/2 - \sqrt{\epsilon})}$ over the value of s_1, it holds that $|L_1 \setminus \{m\}| = |L_1(\vec{Q}) \setminus \{m\}| \leq |L_1(\vec{Q}) \setminus \{m\}| \leq 2^{n(R + 3\Delta S/2 - \Delta I(P, W_{Q_1}))}$.

Now consider a random code $\mathbf{\Phi}_2$. The number of different values for $L_1(\vec{Q})$ can be bounded from above by $2^{nR}|\mathcal{Y}|^k(k+1)^{|\mathcal{Z}|}2^{kS}$, which is less than doubly-exponentially large in k. Therefore a union bound shows that $\mathbf{\Phi}_2$ is good (for every such $L_1(\vec{Q})$ and sufficiently large n) with probability at least $1 - \exp(-\exp((\Delta n S/2) - 1))$. Thus a good realization Φ_2 exists (here we use the fact that the code $\mathbf{\Phi}_2$ is independent of the fixed code Φ_1). As before, fix the code Φ_2. Thus, with probability at least $1 - 2^{-\Delta n(S/2 - \sqrt{\epsilon})}$ over the value s_2, $L_2 \setminus \{m\} = L_2(\vec{Q}) \setminus \{m\}$ satisfies

$$|L_2(\vec{Q}) \setminus \{m\}| \leq 2^{n(R + 6\Delta S/2 - \Delta I(P, W_{Q_1}) - \Delta I(P, W_{Q_2}))}.$$

We continue in a similar manner for all the chunks. Thus with probability at least $1 - \ell \cdot 2^{-\Delta n(S/2 - \sqrt{\epsilon})}$ over the secret s, $|L_\ell \setminus \{m\}| = |L_\ell(\vec{Q}) \setminus \{m\}|$ is bounded from above by $2^{n(R + \frac{3\ell\Delta S}{2} - \Delta \sum_{j=1}^{k} I(P, W_{Q_j}))}$.

By our choice of parameters, this probability is at least $1 - \ell 2^{-k\epsilon^2/2}$ (for sufficiently large n). Since mutual information is convex in the channel and since $\Delta = 1/\ell$, $\Delta \sum_{j=1}^{k} I(X, W_{Q_j}) \geq I\left(P, \Delta \sum_{j=1}^{l} W_{Q_j}\right) \geq \min_{Q \in \mathcal{Q}} I(P, W_Q)$. Hence $|L_\ell(\vec{Q}) \setminus \{m\}|$ is at most

$$\exp\left(-n\left(C(\mathcal{Q}) - R - \frac{3S}{2}\right)\right) \leq \exp\left(-n(\delta - 3S/2)\right).$$

Therefore by our setting of $S = \delta/2$ we obtain that for sufficiently large n, $|L_\ell(\vec{Q}) \setminus \{m\}| < 1$, so every list either contains m or is empty. By Lemma 3, the true message is in a list for some $\vec{Q}$ for a $1 - 2\exp(-k\epsilon^2/2)$ fraction of secrets, so the probability (over the secret) that the secret sequence guarantees m is in $L_\ell(\vec{Q})$ for one $\vec{Q}$ is at least $1 - 2\ell\exp(-k\epsilon^2/2)$.

All in all, with probability at least $1 - 2\ell \cdot 2^{-k\epsilon^2/2}$ over the secret s, the resulting list is either empty or contains m. Union bounding over $\vec{Q}$ of the decoder we obtain our assertion.

V. Conclusion

Our results in this paper are limited to delays which grow linearly with the blocklength, which contrasts with previous results such as [7], in which results are shown for adversaries with delay $\Delta = \log(n)/n$. This is in part due to some looseness in our analysis; in principle we believe it should be possible to show the same capacity results for limited delay. For permutation channels, our capacity region corresponds to both the AVC capacity under stochastic encoding and maximal error and the AVC capacity under deterministic coding and average error. In general, these capacities are not the same; extending our results to those cases will shed some insight into how the adversary is weakened by the delay.

References

[1] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *Annals of Mathematical Statistics*, vol. 31, pp. 558–567, 1960.

[2] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited : Positivity, constraints," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[3] A. Sarwate, "Robust and adaptive communication under uncertain interference," Ph.D. dissertation, University of California, Berkeley, July 2008.

[4] A. Smith, "Scrambling adversarial errors using few random bits, optimal information reconciliation, and better private codes," in *Proceedings of the 2007 ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*, 2007.

[5] M. Langberg, "Oblivious channels and their capacity," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 424–429, 1988.

[6] M. Langberg, S. Jaggi, and B. Dey, "Binary causal-adversary channels," in *Proceedings of the 2009 International Symposium on Information Theory (ISIT)*, 2009.

[7] B. Dey, S. Jaggi, and M. Langberg, "Codes against online adversaries," in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, 2009.

[8] V. Guruswami and A. Smith, "Explicit capacity-achieving codes for worst-case additive errors," arXiv:0912.0965v1 [cs.IT].

[9] S. Micali, C. Peikert, M. Sudan, and D. Wilson, "Optimal error correction against computationally bounded noise," in *Proceedings of the Second Theory of Cryptography Conference (TCC)*, 2005.

[10] K. Jung and D. Shah, "On computationally bounded adversarial capacity," in *Proceedings of the 2006 Information Theory and Applications Workshop*, 2006.

[11] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 27–34, 1988.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémi Kiadó, 1982.

[13] K. Eswaran, A. Sarwate, A. Sahai, and M. Gastpar, "Zero-rate feedback can achieve the empirical capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 1, January 2009.

[14] D. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge, UK: Cambridge University Press, 2009.