Keyless Authentication for AWGN Channels

Eric Graves, *Member, IEEE*, Allison Beemer, Jörg Kliewer, *Senior Member, IEEE*, Oliver Kosut, *Member, IEEE*, and Paul L. Yu, *Senior Member, IEEE*

***Abstract*— This work establishes that the physical layer can be used to perform information-theoretic authentication in *additive white Gaussian noise* (AWGN) channels, as long as the adversary is not omniscient. The model considered consists of an encoder, decoder, and adversary, where the adversary knows the message given to the encoder, has a non-causal noisy observation of the encoder's transmission and may use unlimited transmission power, while the decoder observes a noisy version of the sum of the encoder and adversary's outputs. A method to modify a generic existing channel code to enable authentication is presented. This method relies on injecting message-dependent noise into the transmission and accepting the transmission as authentic only if the correct noise levels for the decoded message are observed. One drawback to this method is that the encoder must still transmit a low-power signal in the case where there is no message to send. It is shown that this modification costs an asymptotically negligible amount of the coding rate, while still enabling authentication as long as the adversary's observation is not noiseless. Also notable is that this modification is not (asymptotically) a function of the statistical characterization of the adversary's channel and no secret key is required. We believe these features will pave the way for a robust practical implementation. Using these results, the channel-authenticated capacity is calculated and shown to be equal to the non-adversarial channel capacity. As our results will show, information-theoretic authentication in AWGN channels is possible without the need for the legitimate party to have a model-based advantage over the adversary. While this modular scheme is designed for use in the given channel model, it is applicable to a wide range of settings.**

***Index Terms*— Channel capacity, multiuser channels, authentication, information theory, security, wireless.**

I. Introduction

AUTHENTICATION, or the act of verifying the identity of an information source, is a crucial aspect of security; this is especially true in scenarios where the information leads to an observable action (e.g., calling in a missile strike or executing a stock-market trade). For a decoder to have information-theoretically secure authentication, it must be able to accept inputs deriving from the legitimate encoder while rejecting those that have been tampered with by an adversary, even when the adversary is computationally unbounded. Thus, the objectives of the decoder change depending on the presence of an adversary; if no adversary is present then observations should be accepted and decoded into the message given to the encoder, while if the adversary did tamper with the decoder's observation then the decoder only needs to identify this tampering.[1] We will consider information-theoretic authentication in a communication system where the encoder, decoder, and adversary are connected by noisy channels. In this context, information-theoretic authentication has traditionally been achieved by exploiting a feature of the communication model that is unique between the encoder and decoder that the adversary cannot imitate.

In the existing literature, two features are used: either exploiting the channel in such a way that the adversary cannot mimic a valid transmission, or by use of a secret key shared by encoder and decoder. This work is classified in the former category, as we will not allow the encoder and decoder to share a secret key. For readers interested in secret key-equipped information-theoretic authentication, see Perazzone *et al.* [1], [2] for an in-depth discussion on prior works and the best results to date.

In cases where no secret key is available, information-theoretic authentication can be obtained by exploiting (if possible) the uniqueness of the channel from the encoder to the decoder. This exploitation generally takes the form of choosing an encoder that produces channel outputs that cannot be reliably reproduced by the adversary. For this scenario, the information that the adversary can obtain, and how they may act given that information, is crucial to defining how capable the adversary will be at their task. Previous work [3], [4], [5], [6], [7], [8], [9], [10] on this topic is mainly differentiated by how the formulation defines the adversary's abilities. A few of these differentiating aspects are (without vs. with): the allowance of simultaneous transmission by adversary and encoder [3], [4], [6], [7] vs. [5], [8], [9], [10], side information about the encoder's message at the adversary [3], [4], [6], [7], [8] vs. [5], [9], [10], and a noisy copy of the encoder's output at the adversary [3], [4], [5], [7], [8], [10] vs. [6], [9]. It is not surprising then that most of this work is similar

Manuscript received 12 March 2021; revised 8 July 2022; accepted 26 August 2022. Date of publication 12 September 2022; date of current version 22 December 2022. This research was sponsored by the Combat Capabilities Development Command Army Research Laboratory under Cooperative Agreement Number W911NF-17-2-0183 and by the National Science Foundation under Grants No. 2107488, 2107370, and 2107526. *(Corresponding author: Eric Graves.)*

Eric Graves and Paul Yu are with the Computer and Information Sciences Division, U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: eric.s.graves9.civ@army.mil; paul.l.yu.civ@army.mil).

Allison Beemer is with the Department of Mathematics, University of Wisconsin–Eau Claire, Eau Claire, WI 54701 USA (e-mail: beemera@uwec.edu).

Jörg Kliewer is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07103 USA (e-mail: jkliewer@njit.edu).

Oliver Kosut is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: okosut@asu.edu).

Communicated by A. Beimel, Associate Editor for Sequences and Cryptography.

Digital Object Identifier 10.1109/TIT.2022.3205886

[1] Of course, in the case the adversary tampers with the decoder's observations and the decoder still produces a message estimate, this message estimate should be equal to the message given to the encoder.

in formulation, methodology, and results while still being diverse in terminology. Our work here makes all three "with" allowances. For simplicity, we will broadly characterize [3], [4], [5], [6], [7], [8], [9], [10].

To permit more formal discussion, let $p(y|x,v)$ be the conditional distribution of the decoder's channel observation (y) given the encoder's channel input (x) and the adversary's channel input (v), and let $\mathcal{Q}$ be the (model dependent) set of joint-probability distributions of encoder and adversary's channel inputs. Further, let $\emptyset$ be the symbol that represents the "not-transmitting" state; when the adversary chooses any state other than the "not-transmitting" state, it is considered tampering. As an example of how the adversary's definition affects $\mathcal{Q}$, consider the case where the encoder and adversary are not allowed simultaneous transmission. In this case, $\mathcal{Q}$ will contain only distributions such that $q(x,v) > 0$ only if $x = \emptyset$ or $v = \emptyset$. Continuing on, the previous literature divides the set of channels into sets based upon the property[2] that

$$\min_{q \in \mathcal{Q}} |p(y|x',\emptyset) - \sum_{x,v} p(y|x,v)q(v,x)|_1 > 0$$

for at least one encoder channel input x. To understand how this property equips channels for information-theoretic authentication, forget for the moment the need to transmit a message and focus solely on the detection of tampering. In particular, consider the case where the adversary wants the decoder to think that the encoder has input x' into the channel k times. Further, assume that the decoder initially believes that x' has been input into the channel k times, but wishes to verify this. If the adversary did not tamper, then the decoder's k channel observations should be distributed approximately according to $p(y|x',\emptyset)$. But, assuming that the k channel inputs were actually distributed according to $q(x)$, then the closest the adversary will be able to do in approximating this distribution is $w_{x'}(y) = \sum_{x,v} p(y|x,v)q(v,x)$. When $w_{x'}(y) \neq p(y|x',\emptyset)$, a binary hypothesis test can be constructed that will detect the adversary's tampering with a probability of $1 - e^{-O(k\sqrt{|w_{x'}(y) - p(y|x',\emptyset)|_1})}$.

As the previous example would suggest, some encoder-output distributions are better than others for authentication. That the output distribution would be good for both communicating information and authentication cannot be taken for granted. To circumvent this issue, much of the previous work, specifically [3], [4], [6], [7], [8], [10], [12], opts for a two-code concatenated approach. In this approach, one of the codes is a long (in terms of symbols) capacity-achieving code, while the other is a short low-rate code that provides information-theoretic authentication. The message is transmitted with the capacity-achieving code, while a randomly-generated number and a hash of that randomly-generated number and message are transmitted with the low-rate code. After decoding both codes, the hash (guaranteed to be authentic) can be used to authenticate the transmitted message. Of course, this two-stage approach requires that the adversary not be able to determine the hash prior to choosing their channel inputs. If the adversary was able to decode the information sent with the low-rate code before choosing their channel, they could alter the message sent with the high-rate code to one that is valid for the given hash.

Our work seeks to model communications in a wireless environment where the adversary is overwhelmingly powerful. There will exist an AWGN channel from the encoder to the adversary, and the encoder and adversary share an AWGN multiple access channel to the decoder. While both noise sources are assumed to be independent and have non-zero variance, no assumption is made about their relative values. More directly, the channel from the encoder to adversary can have a lower noise variance than the best-case channel from encoder to decoder. To ensure that the adversary is stronger than ever could be encountered in practice, we will allow the adversary to (i) non-causally view the encoder-to-adversary channel's output; (ii) know the message that is being transmitted by the encoder; (iii) use unlimited computational power; and (iv) use unlimited transmit power. Informally, the adversary can be thought of as sitting at unknown location with as much power as they need for computation and communication purposes, they know what the message the encoder wants to send, and they will be able to choose their outputs as a function of their noisy copy of the signal.

Our desire to model realistic channels under extremely adverse conditions costs us the traditional analysis. Indeed, recall the binary hypothesis testing analogy where the distribution of a specific set of symbols is tested. By considering channels with inputs taken from uncountable sets, as opposed to [3], [4], [5], [6], [7], [8], [9], [10], it becomes harder to guarantee that there does not exist a particular input from the adversary that allows them to mimic the correct distribution. Furthermore, the non-causal and less-noisy observations at the adversary prohibit the two-code approach. Even more than that, this prohibits the ability to transmit information that the adversary cannot decode.

Nevertheless, despite these adversarial advantages, our scheme will achieve information-theoretic authentication along with the following notable features:

- a construction that modifies a given deterministic channel code;
- does not require a shared secret key or common randomness;
- will detect an adversary's manipulation as long as the adversary's observations of the encoder outputs are not completely noiseless;
- achieves rates arbitrarily close to the non-adversarial channel capacity.

Informally then, our results prove that information-theoretic authentication is possible in AWGN channels without the legitimate party having an advantage over the adversary.[3] The main drawback to our scheme is that it requires the encoder output a low-power signal when not supplied with a message. This will be discussed further in Remarks 10 and 22. Still, despite the austere channel model, our scheme allows for a robust detection. Furthermore, by choosing to modify

[2]Typically this property is denoted by an "-able"-suffixed term, such as simulatable [3], [4], [6], [7] (pace Maurer [11]), or overwritable [8], [10]. U/I-overwritable [9]. The authors of [5] abstained from naming the channel condition.

[3]This statement will be slightly qualified in Section IV-A.

arbitrarily given channel codes we show that creating a code from scratch is not necessary. This should allow researchers to concentrate on modifying existing codes such as low-density parity-check codes, turbo codes, polar codes, or repetition codes.

Instead of relying on traditional information-theoretic security concepts, we build on the insights of Graves *et al.* [5] and Beemer *et al.* [9], who enabled authentication[4] by introducing artificial noise at the output of the encoder. For a simple example, consider a channel where the encoder can output 0 or 1, the adversary can output -1, 0, and 1, and where the decoder receives the sum of the two. Given any deterministic encoder $x : \mathcal{M} \to \{0,1\}^n$, if the adversary has knowledge of the transmitted message they may in turn choose their transmitted sequence as $z(M) = x(a) - x(M)$ so that the decoder receives

$$y(M) = x(M) + z(M) = x(a),$$

which is indistinguishable from the case where the encoder sends $a \in \mathcal{M}$ and the adversary does not interfere. Suppose, though, that the encoder passed their output through a binary symmetric channel, with positive crossover probability $p < 1/2$, before inputting it into the channel to the decoder. Now, the probability of false authentication decreases exponentially with the number of coordinates i that $z_i \neq 0$. Indeed, assume that $z_i = 1$, if the encoder inputs a 1 into the channel, which it does with probability at least p regardless of the message, then $y_i = 2$. This outcome is only possible if the adversary has tampered. Thus, it is easy to see that the probability of false authentication is at most $(1-p)^{|\{i|z_i \neq 0\}|}$. This probability can be made arbitrarily small by starting with a well-chosen channel code. The key is that adding the stochastic coding element produces events that the adversary cannot account for and hence leads to outputs that would not be expected were the adversary not trying to modify the transmission. Of course, our situation will be more complicated because the adversary will have their own observations, but the premise remains the same. Without complete knowledge of the encoder's output, the adversary's actions will result in the decoder observing unexpected channel outputs.

To take advantage of this insight, our code modification strategy consists of first adding carefully constructed message-dependent noise and then decimating the message set. The message-dependent noise is determined by a novel coding scheme that guarantees the adversary must always remove some of the noise added to the channel in order to forge a message. As long as the adversary's observations themselves are noisy, the adversary will not be able to completely eliminate the message-dependent noise the encoder added to the channel. In turn, whether or not there exists a remnant noise in the channel observations determines if the transmission has been tampered with. Specifically, this added noise will guarantee that the adversary cannot modify a message to a specific message of their own choosing. From there, decimating the message set (a concept borrowed from Ahlswede and Dueck's local strong converse [13]) extends this guarantee to ensure a small maximum probability of false authentication.

To begin the formal treatment of this problem, the notation, model, and operational measures will be presented in Section II. The results will be presented in Section III, with many of the proofs being removed to the appendices for readability. Section IV includes a discussion of topics for further investigation, as well as a comparison of our scheme to secret key-based authentication schemes. Conclusions are presented in the Section V.

II. Model and Notation

A. Notation

Uppercase letters will denote random variables, lowercase constants, and script sets. In particular $\mathcal{R}$ denotes the set of real numbers.

Bold font always denotes n-fold Cartesian products, with n to be later defined as the block length of the code. For a given $\boldsymbol{x}$, x_i is the i-th coordinate so that $\boldsymbol{x} = \times_{i=1}^n x_i$. While Cartesian products of random variables and constants may have unique coordinates, a set which is a Cartesian products of sets will not (i.e., $\boldsymbol{\mathcal{X}} = \times_{i=1}^n \mathcal{X}$).

Throughout the paper, $\boldsymbol{G_\rho} = \times_{i=1}^n G_{\rho,i}$ will be used to denote Cartesian product of n independent Gaussian random variables with mean 0 and where the i-th coordinate has variance ρ_i. When all the variances are equal, (i.e., $\boldsymbol{\rho} = \times_{i=1}^n \rho$) just the single variance will be listed (i.e., $\boldsymbol{G}_\rho$). Sometimes $\boldsymbol{G}_{\text{some qualitative value}}$ will be used in place of $\boldsymbol{G}_{\boldsymbol{\rho}_{\text{some qualitative value}}}$ so that it is easier to specify the source of this randomness in the math. Finally, all values of G, unless otherwise explicitly stated, should be assumed independent.

All logarithms are natural, and the following functions will be used:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\mathcal{R}} x f_X(x) \mathrm{d}x \\
\mathbb{D}(X||Y) &= \int_{\mathcal{R}} f_X(x) \log \frac{f_X(x)}{f_Y(x)} \mathrm{d}x \\
\mathbb{D}_2(a||b) &= a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b} \\
\mathbb{H}_2(a) &= -a \log a - (1-a) \log(1-a) \\
\mathbb{I}_2(a||b) &= b\mathbb{D}_2(a||b) + (1-b)\mathbb{D}_2\left(b\frac{1-a}{1-b}\middle\|b\right) \\
&= \mathbb{H}_2(b) - b\mathbb{H}_2(a) - (1-b)\mathbb{H}_2\left(b\frac{1-a}{1-b}\right) \\
\mathbb{1}_{\mathcal{A}}(b) &= \begin{cases} 1 & \text{if } b \in \mathcal{A} \\ 0 & \text{else} \end{cases} \\
\Phi(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t
\end{aligned}$$

where f_X is used to denote the probability density function of X. Furthermore, we will use $\binom{\mathcal{A}}{b}$ to denote the set of all b-element subsets of $\mathcal{A}$. For instance

$$\binom{\{1,2,3\}}{2} = \{\{1,2\},\{1,3\},\{2,3\}\}.$$

[4]In fact, both works show something even more surprising: there exist channels that *require* stochastic codes for information-theoretic authentication.

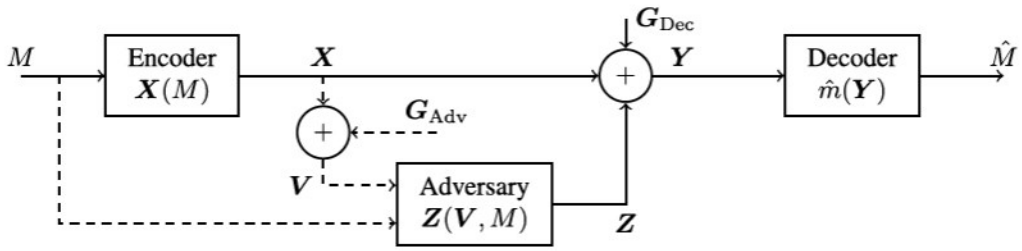

Fig. 1. Channel with encoder $\boldsymbol{X} : \mathcal{M} \to \boldsymbol{\mathcal{R}}$ and decoder $\hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\}$, where $\boldsymbol{G}_{\mathrm{Dec}} \sim \mathrm{Gaussian}(0, \rho_{\mathrm{Dec}})$ and $\boldsymbol{G}_{\mathrm{Adv}} \sim \mathrm{Gaussian}(0, \rho_{\mathrm{Adv}})$. The dashed lines represent non-causal links.

B. Model

The model for communications (pictured in Figure 1) studied here consists of three entities: an encoder, decoder, and an adversary. In this model, the encoder is tasked with sending a message M to the decoder, where the message is assumed to be uniform[5] over $\mathcal{M}$. To do this the encoder will map the message to an n-symbol sequence $\boldsymbol{X}(M)$ and send it across the communications channel. It is important to note that the code is allowed to be a random function of the message, and all parties will know the code.

When the encoder sends its codeword, the adversary will non-causally receive a noisy version of the encoder's output,

$$\boldsymbol{V} = \boldsymbol{X}(M) + \boldsymbol{G}_{\mathrm{Adv}},$$

where $\rho_{\mathrm{Adv}} \in (0, \infty)$ represents variance of the adversary's noise. Using this received information, the adversary will craft their own n-symbol sequence to add to the channel. This function will be modeled by $\boldsymbol{Z} : \boldsymbol{\mathcal{R}} \times \mathcal{M} \to \boldsymbol{\mathcal{R}}$. It is worth noting[6] that $\boldsymbol{X}(M)$ and $\boldsymbol{Z}(\boldsymbol{V}, M)$ are independent given $\boldsymbol{V}, M$. For discussion purposes, it can be generally assumed that the adversary chooses this function to maximize the probability of having the decoder produce a false message.

On the other hand, the decoder will receive a noisy copy of the combination of n-symbol sequences sent by encoder and adversary,

$$\boldsymbol{Y} = \boldsymbol{X}(M) + \boldsymbol{Z}(\boldsymbol{V}, M) + \boldsymbol{G}_{\mathrm{Dec}},$$

where $\rho_{\mathrm{Dec}} \in (0, \infty)$ represents the noise variance at the decoder. From there, the decoder will attempt to estimate the message the encoder sent as $\hat{m}(\boldsymbol{Y})$ or will output ! to indicate that the adversary has altered the transmission.

The encoder, adversary, and the decoder know the value of ρ_{Dec}, while only the adversary knows[7] the value of ρ_{Adv}.

C. Operational Parameters

The objective of this work is to construct a good code for authenticated communications.

*Definition 1 (**Code**):* A *code* is a set of paired functions $\boldsymbol{X} : \mathcal{M} \to \boldsymbol{\mathcal{R}}$, $\hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\}$ representing the *encoder* and *decoder* respectively. The symbol ! specifically represents the case that the decoder labels the observation as not authentic.

Remark 2: A code not designed for authenticated communications can be considered as a special case of the code defined here where $\hat{m}(\boldsymbol{y}) \neq !$ for all $\boldsymbol{y} \in \boldsymbol{\mathcal{R}}$.

Remark 3: Codes are assumed to have block length (number of channel uses) n, unless otherwise stated.

Codes will be measured by the rate at which they can send information, the power required to do so, the reliability that the information is decoded when there is no adversarial interference, and the likelihood the adversary can manipulate the decoder into accepting a false message. Formal definitions for the first three follow.

*Definition 4 (**Rate**):* The *rate* of a code $\mathcal{H} = (\boldsymbol{X}, \hat{m})$ is

$$r_{\mathcal{H}} = \frac{1}{n} \log |\mathcal{M}|.$$

*Definition 5 (**Power Constraint**):* The *power constraint* of a code $\mathcal{H} = (\boldsymbol{X}, \hat{m})$ is

$$\omega_{\mathcal{H}} = \max_{m \in \mathcal{M}} \sum_{i=1}^{n} \frac{1}{n} \mathbb{E}\left[X_i^2(m)\right].$$

*Definition 6 (**Error Probability**):* For code $\mathcal{H} = (\boldsymbol{X}, \hat{m})$ the *arithmetic-average error probability* at noise variance $\rho_{\mathrm{Dec}} \in (0, \infty)$ is

$$\varepsilon_{\mathcal{H}}(\rho_{\mathrm{Dec}}) = \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|} \Pr\left(\hat{m}(\boldsymbol{X}(m) + \boldsymbol{G}_{\mathrm{Dec}}) \neq m\right).$$

Note that the error probability is indeed a measure of reliability when not under adversarial influence, since if $\boldsymbol{Z}(\boldsymbol{V}, M) = \boldsymbol{0}$ then $\boldsymbol{Y} = \boldsymbol{X}(M) + \boldsymbol{G}_{\mathrm{Dec}}$.

Two measures of the adversary's ability to interfere will be considered. The weaker of these two measures considers the adversary's ability to have the decoder accept a specific message.

*Definition 7 (**Maximum Probability of Targeted False Authentication**):* The *maximum probability of targeted false authentication* for code $\mathcal{H} = (\boldsymbol{X}, \hat{m})$ with decoder noise variance $\rho_{\mathrm{Dec}} \in (0, \infty)$ and adversary noise variance $\rho_{\mathrm{Adv}} \in (0, \infty)$ is

$$\alpha^*_{\mathcal{H}}(\rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) = \sup_{\substack{\boldsymbol{Z}: \boldsymbol{\mathcal{R}} \times \mathcal{M} \to \boldsymbol{\mathcal{R}} \\ a \in \mathcal{M} \\ b \in \mathcal{M} \setminus \{a\}}} \Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\mathrm{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) = b\right),$$

[5]The distribution of this message will not play a role in the results.

[6]This observation will be critical to our authentication scheme as it will preclude the adversary from canceling out stochastic $\boldsymbol{X}(M)$.

[7]Results can be improved in the case that the encoder and decoder are also aware of ρ_{Adv}, see Section IV-D for a brief discussion of this in the context of a comparison of secret key authentication and our results in this work.

where

$$\boldsymbol{V} = \boldsymbol{X}(a) + \boldsymbol{G}_{\text{Adv}}.$$

A small probability of targeted false authentication does not guarantee the decoder will not output a false message, instead it guarantees that the adversary cannot *choose* which message it is. This weaker metric will only play a brief role in this study, with the main goal being to obtain codes that perform well under the following, stronger metric.

*Definition 8 (**Maximum Probability of False Authentication**):* The *maximum probability of false authentication* for code $\mathcal{H} = (\boldsymbol{X}, \hat{m})$ with decoder noise variance ρ_{Dec} and adversary noise variance ρ_{Adv} is

$$\alpha_{\mathcal{H}}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) = \sup_{\substack{\boldsymbol{Z}:\mathcal{R}\times\mathcal{M}\to\mathcal{R} \\ a\in\mathcal{M}}} \Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\text{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) \notin \{a, !\}\right),$$

where

$$\boldsymbol{V} = \boldsymbol{X}(a) + \boldsymbol{G}_{\text{Adv}}.$$

Unlike the targeted false authentication probability, a vanishing probability of false authentication does asymptotically guarantee the decoder will not output a false message in the presence of an adversary.

Remark 9: To better understand the relationship between the two metrics observe that

$$\Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\text{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) \notin \{a, !\}\right) = \sum_{b\in\mathcal{M}\setminus\{a,!\}} \Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\text{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) = b\right),$$

from which it is clear that having a small maximum probability of targeted false authentication does not guarantee a small maximum probability of false authentication, but a small maximum probability of false authentication does guarantee a small maximum probability of targeted false authentication.

Remark 10:* Readers familiar with information-theoretic authentication literature[8] may be wondering why we have not defined the impersonation attack. For those unfamiliar, an impersonation attack is one where the adversary does not wait for the encoder to produce an output, but directly sends a value to the decoder.

We will not define this metric separately, since an appropriately defined code can already account for a "no-message-to-transmit state." That is, we may assume that $\mathcal{M}$ contains an element (call it $\emptyset$) that has the semantic meaning that a message has not been input to the encoder. With this "no-message" message built into the code, the maximum probability of false authentication upper bounds the probability of an impersonation attack since

$$\alpha_{\mathcal{H}}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) \geq \sup_{\boldsymbol{Z}:\mathcal{R}\times\mathcal{M}\to\mathcal{R}} \Pr\left(\hat{m}(\boldsymbol{X}(\emptyset) + \boldsymbol{G}_{\text{Dec}} + \boldsymbol{Z}(\boldsymbol{G}_{\text{Adv}}, \emptyset)) \notin \{\emptyset, !\}\right).$$

In most settings the encoder does not produce a signal (i.e., $\boldsymbol{X}(\emptyset) = 0$.) for the "no-message" message. Here though, this practice leads to non-operational codes. Indeed, if $\boldsymbol{X}(\emptyset) = \boldsymbol{0}$ then the adversary could fool the decoder into producing any given $m' \in \mathcal{M}$ by simply setting $\boldsymbol{Z}(\boldsymbol{G}_{\text{Adv}}, \emptyset) = \boldsymbol{X}(m')$.

Given this observation, it should be unsurprising that our coding scheme will require the encoder produce a non-zero output for the "no-message" message. As such[9] if the original code did not produce a signal for the "no message" message, then our modified code will produce low-power noise for the "no message" message.

One of the primary goals of our work will be to characterize the authenticated capacity. Intuitively, the authenticated capacity is the maximum rate possible under a given power constraint and the requirement that the probability of error and maximum probability of false authentication converge to zero. In order to present the exact definition, the notation $\boldsymbol{X}_{(n)} : \mathcal{M}_{(n)} \to \boldsymbol{\mathcal{R}}_{(n)}, \hat{m}_{(n)} : \boldsymbol{\mathcal{R}}_{(n)} \to \mathcal{M}_{(n)} \cup \{!\}$ will be used to denote codes with block length n.

*Definition 11 (**Authenticated Capacity**):* The *authenticated channel capacity* is

$$c(\rho, \rho_{\text{Dec}}, \rho_{\text{Adv}}) = \sup\left\{ r \in \mathcal{R} \;\middle|\; \begin{array}{l} \exists \mathcal{H}_{(n)} = \{\boldsymbol{X}_{(n)}, \hat{m}_{(n)}\}_{n=1}^{\infty} \\ \quad \text{such that} \\ \qquad \limsup_{n\to\infty} \omega_{\mathcal{H}_{(n)}} \leq \rho \\ \qquad \liminf_{n\to\infty} r_{\mathcal{H}_{(n)}} \geq r \\ \qquad \limsup_{n\to\infty} \varepsilon_{\mathcal{H}_{(n)}}(\rho_{\text{Dec}}) = 0 \\ \qquad \limsup_{n\to\infty} \alpha_{\mathcal{H}_{(n)}}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) = 0 \end{array} \right\}.$$

Remark 12: If the authentication requirement were removed, then only the reliability requirement would remain and hence the capacity would be

$$\frac{1}{2}\log\left(1 + \frac{\rho}{\rho_{\text{Dec}}}\right),$$

following from Shannon [15].

On a final note, as mentioned in the introduction, Section III's code construction results will be presented in terms of a given channel code. These results will, however, require the initial channel code be deterministic, i.e., the encoder output is not random given the message.

III. Results

In this section we will build a number of consecutive results which lead to the conclusion that information-theoretic authentication is possible in AWGN channels without the need for the encoder to have an advantage over the adversary. Not only this, but we will also show that arbitrary existing codes can be equipped with information-theoretic authentication at a small cost to the rate, power, and error probability of the code. This may be surprising to readers familiar with information-theoretic security where the legitimate party typically requires some exploitable advantage over the adversary. Usually this advantage takes the form of a secret key or a channel with certain desirable characteristics. These advantages will not be required here as the legitimate party already possesses a slight advantage over the adversary with regards

[8]Primarily, the information-theoretic authentication literature whose genesis is Simmons [14]. These works are primarily based on use of a secret key, and thus are markedly different than ours.

[9]Recalling that our methodology is to take a traditional channel code and modify it into one that allows for authentication.

to the operational objectives. The legitimate parties are successful if the correct message is decoded *or* the adversary is detected, while the adversary is *only* successful if a non-legitimate message is produced.

Instead we will generally enable information-theoretic authentication with two complementary code modifications. The first of these modifications will (literally) add to the encoder's output a type of code which enables detection of targeted authentication attacks. The second modification will eliminate messages in the process, converting codes that perform well under targeted false authentication into ones that perform well under general authentication. With these results in hand, we show that the costs asymptotically vanish while the ability to detect manipulation remains. Furthermore, we show that this is true regardless of the difference between the adversary and decoder's noise variance, instead only requiring that the noise variance at the adversary be non-zero.

In pursuit of the modular scheme, we begin by constructing a new type of code, termed an *overlay code*. Conceptually, these codes are used to control the amount of a persistent[10] resource at each channel use for a given message. The overlay code's purpose is to guarantee that a portion of this persistent resource must be removed by the adversary before they can falsify a message. If the adversary is unable to remove the persistent resource, then its presence can be used to detect the intrusion. For the given channel model, the persistent resource will take the form of Gaussian noise, and the adversary will have to attempt noise cancellation in order to remove the persistent resource's presence.

Before introducing overlay codes in Definition 13, it will be helpful to introduce the intuition behind their conception. These codes are structured to enable basic statistical testing practices to detect the overabundance of the persistent resource. This is done by first limiting to a discrete set the possible levels of persistent resource added per channel use. All channel uses that have a given amount of persistent resource (e.g., all channel uses which have had half of the maximum amount of resource added) can be thought of as the "test sets" since these sets will eventually form the sets over which we perform hypothesis testing in order to determine the presence of an adversary. The most important property of the overlay code is that for any given message and any alternative message, one of the test sets for the given message will correspond to channel uses whose persistent resource level is always less than or equal to (with a certain amount guaranteed to be strictly less than) the persistent resource level of the alternative message. Consider this set-up in the context of authentication, where the alternative message represents the actual transmitted message and the given message the one produced by the decoder. In this case, one of the test sets for the given (decoded) message will correspond to a set of channel uses where the encoder has added more of the persistent resource for the alternative (transmitted) message. If the adversary cannot remove this resource efficiently enough, then its presence will alert the decoder. We now define the overlay code.

[10]By "persistent" we mean that it is difficult to remove.

Definition 13: Given finite set $\mathcal{K} \subset [0,1)$, and $\tilde{\mathcal{K}} = \mathcal{K} \cup 1$, positive real number r, and $\gamma \in \left(\frac{1}{2}, 1\right)$, a function $\boldsymbol{f} : \mathcal{M} \to \tilde{\mathcal{K}}$ is a $(r, \mathcal{K}, \gamma)$-*overlay code* when

- $$\frac{1}{n} \log |\mathcal{M}| \geq r;$$
- $$\sum_{i=1}^{n} \mathbb{1}_{\{k\}} \left(f_i(m)\right) = \ell := \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|} \right\rfloor$$
  for all $m \in \mathcal{M}$ and $k \in \mathcal{K}$;
- and for each distinct $m, m' \in \mathcal{M}$ there exists a $k \in \mathcal{K}$ such that
  $$\sum_{i=1}^{n} \mathbb{1}_{\{k\}} \left(f_i(m)\right) \mathbb{1}_{\{k\}} \left(f_i(m')\right) \leq \gamma \ell$$
  and for all $j \in \mathcal{K}$ such that $j < k$
  $$\sum_{i=1}^{n} \mathbb{1}_{\{k\}} \left(f_i(m)\right) \mathbb{1}_{\{j\}} \left(f_i(m')\right) = 0.$$

Uniform overlay codes are overlay codes with $\mathcal{K} = \left\{0, |\tilde{\mathcal{K}}|^{-1}, \ldots, 1 - |\tilde{\mathcal{K}}|^{-1}\right\}$.

Remark 14: For the remainder of the paper, let $\tilde{\mathcal{K}} := \mathcal{K} \cup 1$ and $\ell := \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|} \right\rfloor$.

Remark 15: If $\boldsymbol{f}$ is an $(r, \mathcal{K}, \gamma)$-overlay code, then for each $\tilde{\mathcal{M}} \subset \mathcal{M}$ the function $\tilde{\boldsymbol{f}} : \tilde{\mathcal{M}} \to \boldsymbol{\mathcal{R}}$ defined by $\tilde{\boldsymbol{f}}(m) = \boldsymbol{f}(m)$ is a $(\frac{1}{n} \log |\tilde{\mathcal{M}}|, \mathcal{K}, \gamma)$-overlay code.

Remark 16: It would certainly be possible to define overlay codes to allow the persistent resource levels a non-uniform number of channel uses. That we did not do so is merely for the sake of simplicity.

Note, fewer resource levels $|\tilde{\mathcal{K}}|$ implies more symbols share each level, hence fewer levels implies that there are more channel uses for each particular resource level. Obviously though, fewer resource levels also means fewer unique output sequences for the overlay code, hence the overlay code will support fewer messages. To quickly see this, observe that if $\tilde{\mathcal{K}}$ consisted of two elements, then there would be at most 2^n different possible code combinations.

The a priori existence of overlay codes should not be taken for granted. For instance, consider a traditional random coding argument where for each message the encoder outputs are chosen at random from a predefined distribution. For any two messages a and b, let $F_i(a)$ and $F_i(b)$ denote the randomly chosen value of i-th coordinate resource level for messages a and b. Observe that $\Pr\left(F_i(a) < F_i(b)\right) = \frac{1 - \sum_{k \in \tilde{\mathcal{K}}} \Pr(F_i(a) = k)^2}{2}$. Thus for any choice of distribution other than a deterministic one, $\Pr\left(F_i(a) < F_i(b)\right) > 0$, and hence when rate $r > -n^{-1} \log \Pr\left(F_i(a) < F_i(b)\right)$ this construction will (with near certainty) produce a code such that for every message a, there exists a message b whose resource levels are always greater than or equal to a's. Increasing the size of $\mathcal{K}$ would exacerbate this problem. Nevertheless, overlay codes do exist given certain conditions outlined in Theorem 17 and Corollary 18.

Theorem 17: There exists $(r, \mathcal{K}, \gamma)$-overlay code for all positive real numbers r, finite $\mathcal{K} \subset [0,1)$, and $\gamma \in \left(\frac{1}{2}, 1\right)$ such that

$$r \leq \frac{1}{n} \sum_{k \in \mathcal{K}} n_k \left| \mathbb{I}_2 \left(\gamma \middle\| \frac{\ell}{n_k} \right) - \frac{4}{3n_k} - \frac{2}{n_k} \log n_k \sqrt{\ell} \right|^+,$$

where $n_k = n - \ell |\{j \in \mathcal{K} | j < k\}|$.

Corollary 18: For all $\gamma \in \left(\frac{1}{2}, 1\right)$ and finite $\mathcal{K} \subset [0,1)$, if positive number

$$r < \gamma \log(|\tilde{\mathcal{K}}|) - \gamma - \mathbb{H}_2(\gamma),$$

then for large enough n there exists a $(r, \mathcal{K}, \gamma)$-overlay code.

Proof Sketch: The full proofs of Theorem 17 and Corollary 18 can be found in Appendix B. Also to be found in Appendix B is a detailed example of the overlay code construction.

We prove the theorem using an iterated random coding procedure. First we represent $\mathcal{M}$ as in bijection with a product of smaller sets, that is $\mathcal{M} = \times_{i \in \{1,\ldots,|\mathcal{K}|\}} \mathcal{M}_i$. Next, independently for each $m_1 \in \mathcal{M}_1$ we randomly select an ℓ-coordinate subset out of the total n coordinates. These ℓ coordinates are those that the overlay code outputs the smallest resource concentration (i.e., the minimum value in $\mathcal{K}$). This process is repeated for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, with the difference being that ℓ coordinates are selected from the set of $n - \ell$ coordinates not selected for m_1 is used, and that the second-smallest resource concentration is output for these coordinates. This process of removing the selected coordinates and then randomly selecting a new set of coordinates is repeated until there are fewer than ℓ coordinates remaining; at this point the remaining coordinates are assigned an overlay output of 1.

From this process, the resulting form of Theorem 17 should be clear. The summand over each $k \in \mathcal{K}$ is simply the maximum rate that our analysis guarantees that two messages match on at most $\gamma\ell$-chosen coordinates.

To see why this method works, consider the following. For any two messages $m, m' \in \mathcal{M}$ there exist representations $(m_1, \ldots, m_{|\mathcal{K}|})$ and $(m'_1, \ldots, m'_{|\mathcal{K}|})$ respectively. Clearly, there exists a smallest value $j \in \{1, \ldots, |\mathcal{K}|\}$ such that $m_j \neq m'_j$. Now, for m and m' the overlay code coordinates corresponding to the 1st through $(j-1)$th resource levels will be equal since $(m_1, \ldots, m_{j-1}) = (m'_1, \ldots, m'_{j-1})$. For the jth level though, the two messages will have different coordinates. Furthermore, whenever the output overlay concentration for message m is equal to the jth level, the concentration for message m' must be greater than or equal to the jth level since all coordinates for resource levels less than that level are shared. Using the appropriate random coding techniques, we can then guarantee a certain percentage of coordinates that do not share a level for m_j and m'_j.

□

Remark 19: Of extreme importance here is that for a fixed rate r and fixed γ, there is a fixed $|\mathcal{K}|$ that guarantees the existence of a overlay code for large enough n. Thus, the value of $|\mathcal{K}|$ should be intuitively viewed as a constant when dealing with asymptotic results.

Remark 20: We will not be concerned with choosing the optimal values for inclusion in $\mathcal{K}$ in this paper. This is primarily because the optimal values will depend on the adversary's noise variance, and we wish to have our code construction be independent of this knowledge. We will return to this discussion in Section IV.

Given the existence of overlay codes, we now go about applying them to arbitrary codes to enable authentication. Importantly, a secret key is not necessary in this application, since authentication is enabled by the persistence of the resource added. For our communication model, the persistent resource is additive Gaussian noise. Our code modification will make use of the overlay code to determine the variance of the Gaussian noise added to the encoder's output. For primarily clerical reasons, another message-dependent signal, $\boldsymbol{t}(M)$, will also be added to the output of the encoder. We strongly suspect it is not necessary for most practical codes, although it is necessary for a result that is agnostic of the original code.

Code Modification 21:
Suppose

- a deterministic code $\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}}$, $\hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M}$,
- injection noise power $\rho_\Delta \in (0, \infty)$,
- tolerance $\delta \in (0,1)$, and
- an $\left(\frac{1}{n} \log |\mathcal{M}|, \mathcal{K}, \gamma\right)$-overlay code $\boldsymbol{f} : \mathcal{M} \to \tilde{\mathcal{K}}$, for some finite $\mathcal{K} \subset [0,1)$ and $\gamma \in \left(\frac{1}{2}, 1\right)$,

are given.

Independently for each $m \in \mathcal{M}$ and $i \in \{1, \ldots, n\}$ randomly choose $t_i(m) \in \mathcal{R}$ according to a Gaussian distribution with mean 0 and variance $(1 - f_i^2(m))\rho_\Delta$. Define the modified encoder $\boldsymbol{X}' : \mathcal{M} \to \boldsymbol{\mathcal{R}}$ by

$$\boldsymbol{X}'(M) = \boldsymbol{x}(M) + \boldsymbol{t}(M) + \boldsymbol{f}(M) \cdot \boldsymbol{G}_\Delta,$$

where $\cdot$ is the coordinate-wise product and $\boldsymbol{G}_\Delta = \boldsymbol{G}_{\rho_\Delta}$. Define the modified decoder $\hat{m}' : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\}$ by

$$\hat{m}'(\boldsymbol{y}) = \begin{cases} \hat{m}(\boldsymbol{y}) & \text{if } \forall k \in \mathcal{K} \\ & \displaystyle\sum_{i \in \mathcal{I}_k} \frac{[y_i - t_i(\hat{m}(\boldsymbol{y})) - x_i(\hat{m}(\boldsymbol{y}))]^2}{k^2 \rho_\Delta + \rho_{\text{Dec}}} \leq \ell(1+\delta), \\ ! & \text{else} \end{cases}$$

where $\mathcal{I}_k$ is the set of coordinates i such that $f_i(\hat{m}(\boldsymbol{y})) = k$.

The resulting modified code is defined by $\boldsymbol{X}', \hat{m}'$.

Remark 22:* Here we reiterate Remark 10 from Section II-B. When given a code $\boldsymbol{x}(\emptyset) = \boldsymbol{0}$, Code Modification 21 produces a code where $\boldsymbol{X}'(\emptyset) = \boldsymbol{t}(\emptyset) + \boldsymbol{f}(\emptyset) \cdot \boldsymbol{G}_\Delta$. Thus, in this situation, the code modification produces encoders that, in the absence of a message, send noise. Later it will be shown that this noise should be extremely low power. In practice, this means that the modified encoder would need to be continuously transmitting.

Remark 23: Note the modified decoder is the original decoder with the extra requirement that

$$\sum_{i \in \mathcal{I}_k} \frac{[y_i - t_i(\hat{m}(\boldsymbol{y})) - x_i(\hat{m}(\boldsymbol{y}))]^2}{k^2 \rho_\Delta + \rho_{\text{Dec}}} \leq \ell(1+\delta)$$

for all $k \in \mathcal{K}$. In this sense, the modified decoder can be viewed as first using the original decoder to decode the message, and then checking for manipulation by ensuring that the extra requirement is met. For reference purposes, we shall

adopt this two-stage decoder view, and refer to the checking of the extra requirement as the *detector*.

To see why Code Modification 21 provides a small probability of *targeted* false authentication, consider the steps an adversary would have to perform in order to fool the decoder into authenticating a particular message. First, the adversary, given their a priori knowledge of the message and codebook, would subtract out the output of the unmodified encoder for the transmitted message as well as the t term. Next they would add in the unmodified encoder's output and the t term for the alternative message they wished the decoder to accept. Finally, the adversary would try to ensure that the correct amount of noise is applied to the correct symbols so as to avoid detection.

But, while the adversary knows the variance of the encoder-added noise per symbol, they will not know the exact value of this added noise since their measurement is itself noisy. As the injected noise power becomes smaller, the variance of the adversary's estimate will become increasingly large relative to the variance of the encoder-added noise. Eventually, the adversary's estimate will be so poor that if the adversary tries to cancel out the added noise, the resulting variance would not be significantly less than that of the encoder-added noise alone. Thus the scheme protects against any message being forged into a different particular message, since this different message will be guaranteed to have a set of coordinates that have less noise variance per symbol than the adversary can manage. Later we will extend this scheme using Code Modification 33/Code Modification Corollary 26 to protect against all types of attacks.

While this does provide a form of information-theoretic authentication, adding noise to the output of the encoder will degrade the signal-to-noise ratio. In turn, this decrease in the signal-to-noise ratio will reduce the maximum achievable rate or, alternatively, increase the probability of decoding error. Our analysis favors the increase in the probability of error. Additionally, the increase in noise will increase the power needed by the encoder. But, as Theorem 32/Corollary 24 formally shows, these costs can vanish while still allowing detection of *targeted* authentication attacks.

Corollary 24 (Theorem 32): Given injection noise power $\rho_\Delta = o(1)$ and tolerance $\delta = o(\rho_\Delta)$, then for all

- deterministic codes $\mathcal{H} = (\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M})$,
- $(r_\mathcal{H}, \mathcal{K}, \gamma)$-uniform overlay code $\boldsymbol{f} : \mathcal{M} \to \tilde{\boldsymbol{\mathcal{K}}}$ for any $\gamma \in (1/2, 1)$ and viable $|\mathcal{K}|$,
- and large enough n ,

Code Modification 21 yields with high probability a code $\mathcal{J} = (\boldsymbol{X}' : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m}' : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\})$ such that

$$\begin{aligned} r_\mathcal{J} &= r_\mathcal{H} \\ \omega_\mathcal{J} &\le \omega_\mathcal{H} + O(\sqrt{\rho_\Delta}) \\ \varepsilon_\mathcal{J}(\rho_{\mathrm{Dec}}) &\le \varepsilon_\mathcal{H}(\rho_{\mathrm{Dec}} + \rho_\Delta) + e^{-O(n\delta^2)} \\ \alpha^*_\mathcal{J}(\rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) &\le e^{-O(n\rho_\Delta^2)}. \end{aligned}$$

Remark 25: Corollary 24 is a corollary of Theorem 32 located in Appendix A. Theorem 32, unlike the above corollary, does not fix the injection noise power, tolerance, or the values in $\mathcal{K}$. Furthermore, Theorem 32 specifies the error terms instead of using order terms.

Proof Sketch: The proof of Theorem 32 is found in Appendix C, and Corollary 24 trivially follows.

Proving the rate is immediate, since it is unchanged from the original code.

For the average power, we have to deal with the deterministic value of $\boldsymbol{t}(M)$ added to the code, in particular analyzing the probability that a spurious value of $\boldsymbol{t}(m)$ is chosen with a large amount of correlation with the related $\boldsymbol{x}(m)$.

For the probability of error, we have to consider both the probability of error of the original decoder with the added noise and $\boldsymbol{t}$ as well as the probability of error introduced with the detector. To upper bound the probability of error of the original decoder, we use the fact that the randomly chosen value of $\boldsymbol{t}$ plus the message-dependent additive white Gaussian noise terms is effectively a message-independent additive white Gaussian noise term with variance ρ_Δ. Hence, the error averaged over all possible choices of $\boldsymbol{t}(M)$ is $\varepsilon_\mathcal{H}(\rho_{\mathrm{Dec}} + \rho_\Delta)$. Using Hoeffding's inequality, it follows that the random choice of $\boldsymbol{t}$ must yield a probability of error close to the average. On the other hand, the probability of error of the detector is straightforward to calculate since, under no manipulation, the detector is checking to see if a sum of independent random variables has the correct mean.

Finally for the probability of targeted false authentication, we note that if the adversary does try to attack, then the distribution of the received sequence at the decoder will consist of independent Gaussian random variables where the variance of the ith coordinate is

$$\tau_i(m) = \tau^\star(f_i(m)) := \frac{f_i^2(m)\rho_\Delta\rho_{\mathrm{Adv}}}{f_i^2(m)\rho_\Delta + \rho_{\mathrm{Adv}}} + \rho_{\mathrm{Dec}},$$

and the mean is of the adversary's choosing. For visualization purposes, note that when ρ_Δ becomes small this variance term converges to $f_i^2(m)\rho_\Delta + \rho_{\mathrm{Dec}}$. By properties of the overlay code though, for each message and alternative message, there exists one set of overlay output coordinates whose output for the decoded message is less than or equal to an alternative message. The probability of detecting this increase in noise variance (under the assumption that the decoded message is not the one transmitted by the encoder) is calculated and used to determine the probability of detecting the adversary's manipulation.

□

While Code Modification 21 does not allow the adversary to impersonate any specific message, it does not guarantee that the adversary cannot impersonate a message. This difference is made plain by referring to the operational definitions and observing again that

$$\begin{aligned} &\Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\mathrm{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) \notin \{a, !\}\right) \\ &\quad = \sum_{b \in \mathcal{M}\setminus\{a,!\}} \Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\mathrm{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) = b\right). \quad (1) \end{aligned}$$

While Code Modification 21 produces codes such that each summand $\Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\mathrm{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) = b\right)$ is small, it does not guarantee the production of a code for which the sum itself is small.

Some reflection, though, shows that the case where the summand is small but this sum is not can only occur if there is (in some sense) a densely packed set of decoding regions.

Under this notion, it makes sense to randomly decimate the message set, similar to how (and why) Ahlswede and Dueck [13] chose to demonstrate the local strong converse. While this does reduce the rate of the code, only a negligible reduction will be needed.

We will resume with a slightly more formal description of why this works after we introduce the coding modification. For now, we must mention that the amount of decimation the message set needs is dependent on operational measures of the underlying code. Therefore, to improve readability we have opted to produce a simplified version of the code modification here, and leave the more precise result for Appendix A.

Code Modification Corollary 26 (Code Modification 33): Suppose

- deterministic code $\mathcal{H} = (\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M})$,
- injection noise power $\rho_\Delta = o(1)$,
- tolerance $\delta = o(\rho_\Delta)$,
- $\left(\frac{1}{n}\log|\mathcal{M}|, \mathcal{K}, \gamma\right)$-overlay code $\boldsymbol{f} : \mathcal{M} \to \tilde{\boldsymbol{\mathcal{K}}}$, for finite $\mathcal{K} \subset [0,1)$ and $\gamma \in \left(\frac{1}{2}, 1\right)$,

are given.

First apply Code Modification 21 to code $\mathcal{H}$, to obtain code $\boldsymbol{X}' : \mathcal{M} \to \boldsymbol{\mathcal{R}}$, $\hat{m}' : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\}$. Next, select $\mathcal{M}^\ddagger$ uniformly at random from $\binom{\mathcal{M}}{\lfloor \exp(nr^\ddagger) \rfloor}$, where

$$r^\ddagger = r_{\mathcal{H}} - O\left(\rho_\Delta^2 + \frac{\log n}{n}\right).$$

Define the modified encoder $\boldsymbol{X}^\ddagger : \mathcal{M}^\ddagger \to \boldsymbol{\mathcal{R}}$ by

$$\boldsymbol{X}^\ddagger(M) \ = \boldsymbol{X}'(M).$$

Define the modified decoder $\hat{m}^\ddagger : \boldsymbol{\mathcal{R}} \to \mathcal{M}^\ddagger \cup \{!\}$ by

$$\hat{m}^\ddagger(\boldsymbol{Y}) = \begin{cases} \hat{m}'(\boldsymbol{Y}) & \text{if } \hat{m}'(\boldsymbol{Y}) \in \mathcal{M}^\ddagger \\ ! & \text{else} \end{cases}.$$

The resulting modified code is given by $\boldsymbol{X}^\ddagger, \hat{m}^\ddagger$.

Remark 27: Code Modification Corollary 26 is a corollary of Code Modification 33 located in Appendix A. There, the decimation terms are made explicit.

Remark 28:* Decimating the message set reduces the rate of the code.

We now return to a more formal description of why this works, which follows from two important facts. First, decimating the message set will not impact the maximum probability of targeted false authentication for any two non-decimated messages. Second, by decimating the message set to $\mathcal{M}^\ddagger$, the probability of false authentication for a given encoded message $a \in \mathcal{M}$ and fixed adversary function $\boldsymbol{Z}$ can be written as

$$\sum_{b \in \mathcal{M} \setminus \{a,!\}} \mathbb{1}_{\mathcal{M}^\ddagger}(b) \Pr\left(\hat{m}(\boldsymbol{X}(a) + \boldsymbol{G}_{\text{Dec}} + \boldsymbol{Z}(\boldsymbol{V}, a)) = b\right). \tag{2}$$

Equation (2), when considered jointly with the decimated message set $\mathcal{M}^\ddagger$ being randomly chosen, takes a form whose concentration is analytically tractable. More specifically Equation (2) should with high probability be close to the mean, which is at most $|\mathcal{M}^\ddagger|/|\mathcal{M}|$ since this is the probability a message is not decimated.

The above intuition is overly-simplistic because all possible attacks must be simultaneously considered. Nevertheless, the technique is sufficient to prove the next theorem/corollary.

Corollary 29 (Theorem 34): Setting injection noise power $\rho_\Delta = o(1)$ and tolerance $\delta = o(\rho_\Delta)$, then for all

- deterministic codes $\mathcal{H} = (\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M})$ with rate $r_{\mathcal{H}} = \Omega(n^{-1}\log n)$,
- $(r_{\mathcal{H}}, \mathcal{K}, \gamma)$-uniform overlay code $\boldsymbol{f} : \mathcal{M} \to \tilde{\boldsymbol{\mathcal{K}}}$ for any $\gamma \in (1/2, 1)$,
- and large enough n

Code Modification 33 with high probability yields a code $\mathcal{J} = (\boldsymbol{X}^\ddagger : \mathcal{M}^\ddagger \to \boldsymbol{\mathcal{R}}, \hat{m}^\ddagger : \boldsymbol{\mathcal{R}} \to \mathcal{M}^\ddagger \cup \{!\})$ such that

$$\begin{aligned} r_{\mathcal{J}} &\geq r_{\mathcal{H}} - O\left(\rho_\Delta^2 + \frac{\log n}{n}\right) \\ \omega_{\mathcal{J}} &\leq \omega_{\mathcal{H}} + O(\sqrt{\rho_\Delta}) \\ \varepsilon_{\mathcal{J}}(\rho_{\text{Dec}}) &\leq \varepsilon_{\mathcal{H}}(\rho_{\text{Dec}} + \rho_\Delta) + e^{-O(n\delta^2)} \\ \alpha_{\mathcal{J}}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) &\leq e^{-O(n\rho_\Delta^2)}. \end{aligned}$$

Remark 30: Corollary 29 is a corollary of Theorem 34 located in Appendix A. Theorem 34, unlike the above corollary, does not fix the injection noise power, tolerance, or the values in $\mathcal{K}$. Furthermore, Theorem 34 specifies the error terms instead of using order terms.

Proof Sketch: The proof of Theorem 34 is found in Appendix D; note that it relies on elements of the proof of Theorem 32 since Code Modification 33 relies on Code Modification 21.

The rate and power for the new code are straightforward, while the probability of error calculation essentially follows from Hoeffding's inequality.

The major difficulty in the proof is bounding the probability of false authentication. As the first step in proving this bound, we recall a result from the proof of Theorem 32; specifically, that the decoder's observation when conditioned on a particular message, adversary observation, and adversary attack is equal to a sequence of independent random variables with the mean of the adversary's choosing but the variance fixed, i.e.,

$$\boldsymbol{Y}|\{M, \boldsymbol{Y}, \boldsymbol{Z} = m, v, z\} = \boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{u}(m, v, z),$$

where $\boldsymbol{u}(m, v, z)$ is an arbitrary function (whose specification is unimportant for this proof) and

$$\tau_i(m) = \frac{f_i^2(m)\rho_\Delta\rho_{\text{Adv}}}{f_i^2(m)\rho_\Delta + \rho_{\text{Adv}}} + \rho_{\text{Dec}}$$

for each symbol $i \in \{1, \ldots, n\}$. Clearly, we can effectively ignore the values of $\boldsymbol{V}$ and $\boldsymbol{Z}$ by jointly considering $\boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}$ for all $m \in \mathcal{M}$ and $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$.

Now for any given $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$ and $m \in \mathcal{M}^\ddagger$, we start by noting the probability of false authentication can be written

$$\sum_{b \in \mathcal{M} \setminus \{m,!\}} \mathbb{1}_{\mathcal{M}^\ddagger}(b) \Pr\left(\hat{m}'(\boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}) = b\right), \tag{3}$$

where $\hat{m}'$ is the modified decoder resulting from the application of Code Modification 21 in Code Modification 33. Using a modified version of the Hoeffding lemma we then bound the concentration of equation (3). The problem that remains is to

extend the above concentration to simultaneously work for all $\mu \in \mathcal{R}$.

Here we take a divide-and-conquer approach by separately considering the sets of $\mu \in \mathcal{U}^\dagger$ and $\mu \notin \mathcal{U}^\dagger$, where $\mathcal{U}^\dagger$ is a bounded interval on the real number line. These bounds are set sufficiently large so that $\mu \notin \mathcal{U}^\dagger$ guarantees that for the coordinates where $\mu_i \notin \mathcal{U}^\dagger$, the probability of passing the detector for each message is less than $e^{-nr_{\mathcal{H}}-O(n\rho_\Delta^2)}$, and hence the probability of passing any message detector is less than $e^{-O(n\rho_\Delta^2)}$. For $\mu \in \mathcal{U}^\dagger$, we show that there exists a finite set $\mathcal{U}^\ddagger \subset \mathcal{R}$ such that bounding all $\mu \in \mathcal{U}^\ddagger$ will suffice to bound all $\mu \in \mathcal{U}^\dagger$. From there, we use the union bound to simultaneously guarantee the concentration of all $\mu \in \mathcal{U}^\ddagger$ (hence all $\mu \in \mathcal{U}^\dagger$) and all $m \in \mathcal{M}$.

□

At this point, it is important to reflect on the form of Theorem 34/Corollary 29. Specifically, consider Corollary 29 where δ is chosen such that $\lim n\delta^2 = \infty$. For example $\rho_\Delta = \sqrt[4]{n}\log n$ and $\delta = \sqrt[4]{n}$. In this case, the code modifications have necessitated a loss in rate, an increase in power, and require the code to be operational at a larger noise level than the original code. However, each of these changes disappear as n increases, and hence the rate converges back to the original rate, the new power converges to the original power, and the level of noise the code must be robust against converges to the original noise level. Suppose then we start with a capacity-achieving sequence of codes with average power $\omega - O(\sqrt{\rho_\Delta})$, and which are robust to a noise variance of $\rho_\Delta + \rho_{\mathrm{Dec}}$. Applying Theorem 34/Corollary 29 should give us a sequence of codes with rate

$$\lim_{n\to\infty} \frac{1}{2}\log\left(1 + \frac{\omega - O(\sqrt{\rho_\Delta})}{\rho_\Delta + \rho_{\mathrm{Dec}}}\right) - o(1) = \frac{1}{2}\log\left(1 + \frac{\omega}{\rho_{\mathrm{Dec}}}\right),$$

which is capacity. At the same time, plugging the values into the maximum probability of false authentication yields

$$\lim_{n\to\infty} \alpha_{\mathcal{J}}(\rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) \leq \lim_{n\to\infty} e^{-O(n\rho_\Delta^2)} = 0,$$

and thus we have the ability to authenticate. This essentially proves the following theorem.

Theorem 31:

$$c(\rho, \rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) = \begin{cases} \frac{1}{2}\log\left(1 + \frac{\rho}{\rho_{\mathrm{Dec}}}\right) & \text{if } \rho_{\mathrm{Adv}} > 0 \\ 0 & \text{else.} \end{cases}$$

Proof Sketch: The proof Theorem 31 is found in Appendix E and is essentially a more formal version of the discussion preceding the theorem. Additionally, we show that if $\rho_{\mathrm{Adv}} = 0$ then the capacity is zero. This is somewhat obvious since the adversary knows the encoder's output perfectly in this case. □

Notice that the capacity experiences a sharp jump at $\rho_{\mathrm{Adv}} = 0$, but is otherwise independent of the value. From a practical perspective, this is ideal. A perfect continuous channel is a physical impossibility. Our result therefore implies that in practical wireless scenarios, information-theoretic authentication is possible without use of a secret key. It is also important to observe that the code modifications themselves do not rely on knowledge of the adversary's channel.

Interestingly, our results indicate that obtaining information-theoretic authentication from a channel differs significantly from obtaining information-theoretic secrecy from a channel. Indeed, all practically relevant schemes for the wiretap channel, dating back to Wyner's seminal work [16], require both knowledge of the adversary's channel as well certain guarantees on this channel which make implementation a difficult proposition. In the relevant analog to our model,[11] information-theoretic secrecy cannot be guaranteed when the noise to the adversary is less than the noise to the decoder. This is not an impediment to information-theoretic authentication though, as our results demonstrate; importantly, knowledge of the message is distinct from knowledge of the transmitted sequence.

In the next section we will discuss the path forward in more detail. Among other things, we will discuss unexplored alternatives for implementation, barriers to practical implementation, difficulties in other channels, and different implementation scenarios.

IV. Discussion & Future Directions

While we derive a scheme that leads to information-theoretic authentication, there remains much to be done. It is worth discussing these remaining questions with some candor, so that those so motivated have a clear understanding of areas for improvement. We also provide here further discussion on the distinction (beyond the obvious) between secret key-based authentication and what we accomplish here.

A. Legitimate Party Advantage

At various points in the paper we have made note that the legitimate party does not need an advantage over the adversary for the scheme to work. While this statement is true for pedagogical purposes, it is also true that the method of authentication is similar in spirit to the works [3], [4], [5], [6], [7], [8], [9], [10] mentioned in the introduction. Also as mentioned in the introduction, these works present the legitimate party as being able to establish information-theoretic authentication whenever the channels fit a given statistical characterization. Belonging to the set of channels that allow for information-theoretic authentication can, and should, be viewed as the advantage of the legitimate party.

Extracting this advantage from our results then leads to the conclusion that, in our model, the legitimate party's advantage is that the channels have noise variance greater than zero. We hope you, dear reader, will agree that calling this an "advantage" would only make sense after our results were known. So while the statistical channel characterizations of [3], [4], [5], [6], [7], [8], [9], [10] can provide insight into how authentication is achieved, it would be inappropriate here.

B. Overlay Code Improvements

When first formulating overlay codes, the goal was to ensure the unique relationship of the output symbols for different messages. In the construction, there were a number of different

[11] Specifically, from Figure 1 remove the message side information given to the adversary and remove the adversary's output.

parameters that could have been varied. In particular: the number of coordinates for a given output concentration, the overlap amount per coordinate, and the output levels themselves (i.e., $\mathcal{K}$). To simplify our analysis, we chose to fix the first two considerations, while leaving $\mathcal{K}$ variable. Surprisingly, the actual values for $\mathcal{K}$, while they do impact the efficiency of the authentication scheme, are actually rather immaterial to achieving authentication.

Further analysis showed the optimal values of $\mathcal{K}$ depend on the value of ρ_{Adv}. As a result, we chose not to optimize over $\mathcal{K}$ since important to our claims is that the value of ρ_{Adv} need not be known when constructing the code. During the review process, the question of the optimal value of $\mathcal{K}$ was raised. To that end, when $\rho_\Delta = o(1)$ the optimal choice of $\mathcal{K} = \{0, k_{(1)}, \ldots, k_{(|\mathcal{K}|-1)}\}$ converges to

$$k_{(a)} = \left[\frac{(1-c\gamma)}{c(1-\gamma)}\right]^{a-1} \frac{\rho_{\text{Dec}}}{\rho_\Delta},$$

where

$$c = \frac{\sqrt[|\mathcal{K}|]{\rho_{\text{Dec}}}}{\gamma \sqrt[|\mathcal{K}|]{\rho_{\text{Dec}}} + (1-\gamma)\sqrt[|\mathcal{K}|]{\rho_\Delta + \rho_{\text{Dec}}}},$$

and yields a maximum probability of false authentication (subject to our analysis) of essentially

$$\exp\left(-\frac{1}{8}(1-\gamma)[1-(1+\delta)c]^2\right).$$

Still, use of this asymptotically optimal value did not simplify our analysis and hence was not instituted.

This does, however, raise the question of what is being lost (in terms of authentication ability) by choosing sub-optimal values for $\mathcal{K}$. Additionally, it remains an open question whether allowing variable γ and variable coordinates per output symbol could further improve the final results.

C. Practical Implementation of Code Modifications

To enable authentication, message-dependent noise must be added and then certain distance properties between the codewords must be ensured. These two tasks appear here as Code Modifications 21 and 33. Our original intent was practicality in these code modifications; we were moderately successful with regards to Code Modification 21, but not so with 33. That Code Modification 21 could be reasonably implemented guided our decision to include here the non-asymptotic versions of Theorem 32 and 34. Still, it is worthwhile to discuss alternatives to our code modifications that could allow for analytical bounds on the operational parameters, as well as a practical implementation.

For Code Modification 21, the only real concern in terms of practicality is the construction of the t function. Indeed, since the initial decoder is used in the first stage of the updated decoder, the output of the decoder can be used to determine what the appropriate value of t should be for the estimated message. It is worth mentioning that we suspect that setting t equal to zero will suffice in most cases. Our suspicion derives from the fact that t is only needed to ensure that the code appears to have uniform noise across all coordinates. In practical decoders though, less noise per symbol is usually to the decoder's benefit. Setting t to zero would yield $\omega_{\mathcal{J}} \leq \omega_{\mathcal{H}} + \rho_\Delta$, with the rate and probability of targeted false authentication remaining as in Theorem 32. On the other hand, the average arithmetic error could be estimated empirically. Hence, this should result in a practical implementation of Code Modification 21 for which Theorem 32 is relevant.

Code Modification 33, on the other hand, cannot be directly implemented as currently stated. Choosing such a large subset uniformly at random from the set of all such subsets is clearly impossible in practice. There may of course be feasible alternatives. For instance, the subset selection could be accomplished using a universal hash function, and the Hoeffding concentration analysis replaced with one deriving from the leftover hash lemma. Alternatively, it may be possible to show that some codes do not actually require a rate reduction. Indeed, our analysis for Code Modification 33 relies heavily on the maximum probability of targeted false authentication established by Code Modification 21. But the adversary can only obtain this maximum by choosing a very specific output, and cannot obtain it for multiple alternative messages at one time. As a result, it seems likely that a more sophisticated analysis, using the amount of perturbation from the optimal output, could yield a maximum distance between codewords required for there to be a successful attack. Ensuring that the code's minimum distance was greater than this maximum would be sufficient to skip Code Modification 33 entirely.

D. Comparison With Secret Key-Based Authentication

The major advantage our authentication scheme has over one that is secret key-dependent is that the secret key becomes a finite resource when the channel to the adversary is better than the channel to the decoder. Hence, in some channel models, our scheme could operate in perpetuity while one which is key-based would have a finite life span. That this is particularly true in any case where the adversary has a better channel is shown in Graves *et al.* [2] whose converse proves the key has a finite duration of use.

On the other hand, secret key-based authentication still allows for two advantages over the non-secret key-based authentication of this paper. First, it is still operational when there is no noise over the channel to the adversary[12] and when the adversary knows, and can therefore cancel, the decoder's noise.[13] Second, and more important, secret key-based authentication experiences a better trade-off between rate loss and how quickly the probability of false authentication converges to zero.

For secret key-based authentication, we know that there must exist a trade-off between the channel capacity and the exponent for the probability of false authentication due to the converse results from Graves and Wong [17] and Graves *et al.* [2]. For some measures of false authentication, this trade-off is linear, and in that sense the message rate and probability of false authentication must share the channel capacity.

Our results do not allow for this type of trade off. That is, while our results require a reduction in rate from the channel capacity in order to achieve authentication, the exponent for

[12] A physical impossibility.

[13] Also a physical impossibility.

the probability of false authentication is at most $\exp(-o(n))$ whereas secret key-based authentication allows $\exp(-O(n))$. If we assume that the encoder knows the channel to the adversary[14] then it is possible to also achieve $\exp(-O(n))$ with our results. Indeed, this is because in this case we do not need ρ_Δ to vanish, but instead just be sufficiently small. Regardless, even under this unfair comparison, and further assuming the more generous result on the power constraint raised in Section IV-C and that the second code modification was unnecessary, to obtain a maximum probability of false authentication of $\exp(-O(n\rho_\Delta^2))$ requires that the difference between the maximum rate and capacity be at least

$$\frac{1}{2}\log\left(1+\frac{\rho}{\rho_{\text{Dec}}}\right)-\frac{1}{2}\log\left(1+\frac{\rho-\rho_\Delta}{\rho_{\text{Dec}}+\rho_\Delta}\right)=\sum_{i=1}^{\infty}\frac{1}{i}(c\rho_\Delta)^i$$

where

$$c=\frac{\rho+\rho_{\text{Dec}}}{\rho+\rho_{\text{Dec}}+\rho_\Delta\left(1+\frac{\rho}{\rho_{\text{Dec}}}\right)}.$$

Hence, a loss of rate does not lead to a linear increase in the exponent of maximum probability of false authentication using our scheme.

E. Higher Order Wireless Channel Approximations

While Gaussian channels are great approximations for free-space fixed point single antenna communications,[15] there exist other scenarios of wireless communications with their own corresponding best channel approximations. Some of these alternative channels consider *multi-input multi-output* (MIMO) antenna arrays, fading channels, and multi-path channels.

Outright, we do not see any reason that the overlay code concept cannot be modified and applied to these channels to create codes that provide information-theoretic authentication. However, any such modification will be highly dependent on the assumptions placed on the encoder and decoder with regards to knowledge of their own channels.

V. Conclusion

In this work, we have shown that physical layer authentication is possible for a channel that models wireless communication. Not only is physical layer authentication possible, but our scheme can be used to detect any adversary as long as the block length is sufficiently large and the adversary does not have access to a completely noiseless copy of the encoder's output. Our scheme achieves this by adding artificial noise into the system using the novel concept of overlay codes. This approach allows for authentication by forcing the adversary to remove the added noise when they attempt to insert a fake message of their own.

[14]This comparison to secret key-based authentication is not entirely fair, since knowledge of the channel to the adversary is not needed in that case.

[15]This point is discussed by Massey [18] regarding deep-space communications.

Although random coding elements were used in the proofs, many of the usual difficulties in practical implementation do not exist in our modular scheme. That is, only the encoder needs to be constructed, since part of the concept of the modular scheme is that the message can still be decoded using the original decoder (see Section IV-C). Furthermore, our modular scheme establishes that every deterministic channel code can be modified to provide physical layer authentication. We expect this to lower the implementation barrier since we therefore do not require a completely new channel code be added to the system design.

Open problems include those outlined in Section IV, as well as investigating further scenarios where adding artificial noise can provide authentication.

Appendix A
Non-Asymptotic Versions of Corollaries 24 and 29 and Code Modification Corollary 26

Theorem 32: For all

- deterministic codes $\mathcal{H}=(\boldsymbol{x}:\mathcal{M}\to\boldsymbol{\mathcal{R}},\hat{m}:\boldsymbol{\mathcal{R}}\to\mathcal{M})$,
- injection noise power $\rho_\Delta\in(0,\infty)$,
- tolerance $\delta\in(0,1)$,
- and $\left(\frac{1}{n}\log|\mathcal{M}|,\mathcal{K},\gamma\right)$-overlay codes $\boldsymbol{f}:\mathcal{M}\to\tilde{\boldsymbol{\mathcal{K}}}$, for finite $\mathcal{K}\subset[0,1)$ and $\gamma\in\left(\frac{1}{2},1\right)$,

Code Modification 21 with high probability yields a code $\mathcal{J}=(\boldsymbol{X}':\mathcal{M}\to\boldsymbol{\mathcal{R}},\hat{m}':\boldsymbol{\mathcal{R}}\to\mathcal{M}\cup\{!\})$ such that

$$\begin{aligned}r_{\mathcal{J}}&=r_{\mathcal{H}}\\ \omega_{\mathcal{J}}&\le\omega_{\mathcal{H}}+2\sqrt{2\omega_{\mathcal{H}}\rho_\Delta(r_{\mathcal{H}}+1)}\\&\quad+\rho_\Delta\left(1+8|\tilde{\mathcal{K}}|\left[r_{\mathcal{H}}+1+\frac{\log|\mathcal{K}|}{n}\right]\right)\\ \varepsilon_{\mathcal{J}}(\rho_{\text{Dec}})&\le\varepsilon_{\mathcal{H}}(\rho_{\text{Dec}}+\rho_\Delta)+\sqrt{\frac{n}{2}e^{-nr_{\mathcal{H}}}}+|\mathcal{K}|e^{-\frac{1}{8}\ell\delta^2}\\ \alpha^*_{\mathcal{J}}(\rho_{\text{Dec}},\rho_{\text{Adv}})&\le e^{-\frac{1}{8}\ell(1-\gamma)\lambda^2}+e^{-\frac{1}{8}\ell\gamma\lambda^2}\end{aligned}$$

where

$$\begin{aligned}\lambda&=\max\left(0,\min_{k\in\mathcal{K}}1-\frac{(1+\delta)(k^2\rho_\Delta+\rho_{\text{Dec}})}{\gamma\tau^\star(k)+(1-\gamma)\tau^\star(d_k)}\right)\\ \tau^\star(a)&=\frac{a^2\rho_\Delta\rho_{\text{Adv}}}{a^2\rho_\Delta+\rho_{\text{Adv}}}+\rho_{\text{Dec}}\\ d_k&=\min\{d\in\tilde{\mathcal{K}}|d>k\}.\end{aligned}$$

Code Modification 33: Suppose

- deterministic code $\mathcal{H}=(\boldsymbol{x}:\mathcal{M}\to\boldsymbol{\mathcal{R}},\hat{m}:\boldsymbol{\mathcal{R}}\to\mathcal{M})$,
- injection noise power $\rho_\Delta\in(0,\infty)$,
- tolerance $\delta\in(0,1)$,
- $\left(\frac{1}{n}\log|\mathcal{M}|,\mathcal{K},\gamma\right)$-overlay code $\boldsymbol{f}:\mathcal{M}\to\tilde{\boldsymbol{\mathcal{K}}}$, for finite $\mathcal{K}\subset[0,1)$ and $\gamma\in\left(\frac{1}{2},1\right)$,

are given.

First, apply Code Modification 21 to code $\mathcal{H}$ and let $\boldsymbol{X}':\mathcal{M}\to\boldsymbol{\mathcal{R}}$, $\hat{m}':\boldsymbol{\mathcal{R}}\to\mathcal{M}\cup\{!\}$ be the result. Next, select $\mathcal{M}^\ddagger$ uniformly at random from $\binom{\mathcal{M}}{\lfloor\exp(nr^\ddagger)\rfloor}$, where $r^\ddagger$

$$r^\ddagger=(1-n^{-1})r_{\mathcal{H}}-\frac{(1-\gamma)\ell}{4n}\lambda^2-\frac{2+\log 2\theta}{n}$$

and λ, $\tau^\star$, and d_k are as defined in Theorem 32, while

$$\theta = \max\left(1, \sqrt{3n\left[\omega_{\mathcal{K}} + (\rho_\Delta + \rho_{\text{Dec}})\left(1 + \delta + 2\lambda^2 + 2\, r_{\mathcal{J}}\right)\right]}\right).$$

Define the modified encoder $\boldsymbol{X}^\ddagger : \mathcal{M}^\ddagger \to \boldsymbol{\mathcal{R}}$ by

$$\boldsymbol{X}^\ddagger(M) \;= \boldsymbol{X}'(M).$$

Define the modified decoder $\hat{m}^\ddagger : \boldsymbol{\mathcal{R}} \to \mathcal{M}^\ddagger \cup \{!\}$ by

$$\hat{m}^\ddagger(\boldsymbol{Y}) = \begin{cases} \hat{m}'(\boldsymbol{Y}) & \text{if } \hat{m}'(\boldsymbol{Y}) \in \mathcal{M}^\ddagger \\ ! & \text{else} \end{cases}.$$

The new modified code is $\boldsymbol{X}^\ddagger, \hat{m}^\ddagger$.

Theorem 34: For all

- deterministic codes $\mathcal{H} = (\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}},\ \hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M})$, with rate
$$r_{\mathcal{H}} \geq \frac{(1-\gamma)\ell}{4(n-1)}\lambda^2 + \frac{2 + \log 4n\theta}{n-1}$$
- injection noise power $\rho_\Delta \in (0, \infty)$,
- tolerance $\delta \in (0, 1)$,
- and $\left(\frac{1}{n}\log|\mathcal{M}|, \mathcal{K}, \gamma\right)$-overlay codes $\boldsymbol{f} : \mathcal{M} \to \tilde{\boldsymbol{\mathcal{K}}}$, for finite $\mathcal{K} \subset [0, 1)$ and $\gamma \in \left(\frac{1}{2}, 1\right)$,

with high probability Code Modification 33 yields a code $\mathcal{J} = (\boldsymbol{X}^\ddagger : \mathcal{M}^\ddagger \to \boldsymbol{\mathcal{R}},\ \hat{m}^\ddagger : \boldsymbol{\mathcal{R}} \to \mathcal{M}^\ddagger \cup !)$ such that

$$\begin{aligned} r_{\mathcal{J}} &\geq r_{\mathcal{H}} - \frac{(1-\gamma)\ell}{4n}\lambda^2 - \frac{r_{\mathcal{H}} + 2 + \log 4n\theta}{n} \\ \omega_{\mathcal{J}} &\leq \omega_{\mathcal{H}} + 2\sqrt{2\omega_{\mathcal{H}}\rho_\Delta(r_{\mathcal{H}} + 1)} \\ &\quad + \rho_\Delta\left(1 + 8|\tilde{\mathcal{K}}|\left[r_{\mathcal{H}} + 1 + \frac{\log|\mathcal{K}|}{n}\right]\right) \\ \varepsilon_{\mathcal{J}}(\rho_{\text{Dec}}) &\leq \varepsilon_{\mathcal{H}}(\rho_{\text{Dec}} + \rho_\Delta) + \sqrt{2ne^{-nr_{\mathcal{H}}}} + |\mathcal{K}|e^{-\frac{1}{8}\ell\delta^2} \\ \alpha_{\mathcal{J}}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) &\leq \left(2n + \frac{1}{2\sqrt{n\rho_{\text{Dec}}}}\right)e^{-\frac{1-\gamma}{8}\ell\lambda^2}, \end{aligned}$$

where θ is defined in Code Modification 33 and λ is defined in Theorem 32.

Appendix B
Theorem 17 and Corollary 18

The proof of theorem and corollary rely on the following random code construction.

Code Construction 35: Suppose finite set $\mathcal{K} \subset [0, 1)$ and positive number $\gamma \in (1/2, 1)$ are given.

For convenience, for each $k \in \mathcal{K}$ set

$$\begin{aligned} n_k &= n - \ell|\{j \in \mathcal{K} | j < k\}|, \\ \mathcal{N}_k &= \{1, \ldots, n_k\}, \\ \mathcal{S}_k &= \binom{\mathcal{N}_k}{\ell} \\ r_k &= \frac{\log\left\lfloor \exp\left(n_k\left|\mathbb{I}_2\left(\gamma\middle|\middle|\frac{\ell}{n_k}\right) - \frac{1}{3n_k} - \frac{2}{n_k}\log n_k\sqrt{\ell}\right|^+\right)\right\rfloor}{n_k}. \end{aligned}$$

For $k = 1$, let n_k and $\mathcal{N}_k$ be defined as above, but let $\mathcal{S}_k = \binom{\mathcal{N}_k}{n_k}$ and $r_k = 0$.

Next for each $k \in \mathcal{K}$ let $\mathcal{M}_k = \{1, \ldots, e^{n_k r_k}\}$, allowing that $\mathcal{M} = \times_{k\in\mathcal{K}}\mathcal{M}_k$.

Independently for each $k \in \mathcal{K}$ and each $m_k \in \mathcal{M}_k$, choose a set $\mathcal{S}_k(m_k)$ uniformly at random from $\mathcal{S}_k$. Then, for each $m = \times_{k\in\mathcal{K}} m_k \in \mathcal{M}$ and $j \in \{1, \ldots, n\}$ set

$$f_j(m) = k \quad \Leftrightarrow \quad j \in g_{\mathcal{N}_k \to \mathcal{I}\left(\times_{j\in\mathcal{K}|j<k} m_j\right)}(\mathcal{S}_k(m_k)),$$

where $\mathcal{I}\left(\times_{j\in\mathcal{K}|j<k} m_j\right)$ is defined recursively by

$$\begin{aligned} &\mathcal{I}\left(\times_{j\in\mathcal{K}|j\leq k} m_j\right) \\ &\quad = \mathcal{I}\left(\times_{j\in\mathcal{K}|j<k} m_j\right) - g_{\mathcal{N}_k \to \mathcal{I}\left(\times_{j\in\mathcal{K}|j<k} m_j\right)}(\mathcal{S}_k(m_k)) \end{aligned}$$

with $\mathcal{I}(\emptyset) = \{1, \ldots, n\}$, and where $g_{\mathcal{A}\to\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ is the lexicographical order-preserving mapping between two equal size sets of natural numbers.

Prior to using this code construction to prove the theorem and corollary, we will present an example to make the construction more clear, as well as present a technical lemma in order to streamline the proof.

A. Example Overlay Code Construction

Suppose $n = 9$ and $\mathcal{K} = \{0, 1/2\}$, (hence $\ell = \left\lfloor \frac{9}{3} \right\rfloor = 3$) are given. For simplicity, let rates r_0, $r_{1/2}$ and r_1 be such that $e^{9r_0} = 4$, $e^{6r_{1/2}} = 3$, and $e^{3r_1} = 1$, yielding a total of $4 \cdot 3 \cdot 1 = 12$ different messages, or a rate of $\frac{1}{9}\log 12$. Note, that we do not need to specify γ in this case since its only involvement in the code construction is choosing values for the rates.

Suppose the randomly selected subsets, $\mathcal{S}_0(i) \subset \{1, \ldots, 9\}$ for $i \in \mathcal{M}_0$ and $\mathcal{S}_{1/2}(j) \subset \{1, \ldots, 6\}$ for $j \in \mathcal{M}_{1/2}$, are

$$\begin{array}{ll} \mathcal{S}_0(1) = \{2,7,8\} & \\ \mathcal{S}_0(2) = \{1,2,6\} & \text{and} \\ \mathcal{S}_0(3) = \{2,6,9\} & \\ \mathcal{S}_0(4) = \{1,5,9\} & \end{array} \qquad \begin{array}{ll} \mathcal{S}_{1/2}(1) = \{2,4,5\} & \\ \mathcal{S}_{1/2}(2) = \{3,4,6\} & \\ \mathcal{S}_{1/2}(3) = \{1,3,5\} & \end{array}$$

then the resulting code constructed is

$$\begin{array}{lccccccccc} \boldsymbol{f}(11) = & 1 & 0 & 1/2 & 1 & 1/2 & 1/2 & 0 & 0 & 1 \\ \boldsymbol{f}(12) = & 1 & 0 & 1 & 1/2 & 1/2 & 1 & 0 & 0 & 1/2 \\ \boldsymbol{f}(13) = & 1/2 & 0 & 1 & 1/2 & 1 & 1/2 & 0 & 0 & 1 \\ \boldsymbol{f}(21) = & 0 & 0 & 1 & 1/2 & 1 & 0 & 1/2 & 1/2 & 1 \\ \boldsymbol{f}(22) = & 0 & 0 & 1 & 1 & 1/2 & 0 & 1/2 & 1 & 1/2 \\ \boldsymbol{f}(23) = & 0 & 0 & 1/2 & 1 & 1/2 & 0 & 1 & 1/2 & 1 \\ \boldsymbol{f}(31) = & 1 & 0 & 1/2 & 1 & 1/2 & 0 & 1/2 & 1 & 0 \\ \boldsymbol{f}(32) = & 1 & 0 & 1 & 1/2 & 1/2 & 0 & 1 & 1/2 & 0 \\ \boldsymbol{f}(33) = & 1/2 & 0 & 1 & 1/2 & 1 & 0 & 1/2 & 1 & 0 \\ \boldsymbol{f}(41) = & 0 & 1 & 1/2 & 1 & 0 & 1/2 & 1/2 & 1 & 0 \\ \boldsymbol{f}(42) = & 0 & 1 & 1 & 1/2 & 0 & 1/2 & 1 & 1/2 & 0 \\ \boldsymbol{f}(43) = & 0 & 1/2 & 1 & 1/2 & 0 & 1 & 1/2 & 1 & 0 \end{array}.$$

In more detail, take for example $\boldsymbol{f}(32)$ which corresponds to $\mathcal{S}_0(3) = \{2, 6, 9\}$ and $\mathcal{S}_{1/2}(2) = \{3, 4, 6\}$. Here $\boldsymbol{f}(32)$ is constructed by first assigning a 0 to all indices in $\mathcal{S}_0(3)$, after which these indices are removed from the pool of possible indices $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ leaving indices $\{1, 3, 4, 5, 7, 8\}$. Now considering the remaining indices ($\{1, 3, 4, 5, 7, 8\}$) as an ordered set, of these the $\mathcal{S}_{1/2}(2)$ indices (the $\{3, 4, 6\}$-th smallest, i.e., $\{4, 5, 8\}$) are assigned a value of $\frac{1}{2}$, and all remaining unassigned indices ($\{1, 3, 7\}$) are given 1.

Also from this example, the important aspect of the overlay code can be observed. Namely, for any fixed message one

of the sets of coordinates for the message which produce the same output (e.g., for 33, $\{2,6,9\}$ produce 0, $\{1,4,7\}$ produce $1/2$, and $\{3,5,8\}$ produce 1) is strictly not greater than the corresponding outputs produced for any alternative message.

B. Technical Lemma

Lemma 36: For integers a, b, c, such that $a > b > c \geq \max(b-(a-b), 1)$,

$$-\log \frac{\binom{b}{c}\binom{a-b}{b-c}}{\binom{a}{b}} \geq a\mathbb{I}_2\left(\frac{c}{b}\middle\|\frac{b}{a}\right) - \frac{1}{3} - 2\log a.$$

Proof: This lemma follows nearly directly from Robbins' remark[16] on Stirling's formula [19], along with some basic algebra. More specifically

$$\begin{aligned}
&-\log \frac{\binom{b}{c}\binom{a-b}{b-c}}{\binom{a}{b}} \\
&\quad = c\log\frac{c}{b} + (b-c)\log\frac{b-c}{b} \\
&\qquad + (a-2b+c)\log\frac{a-2b+c}{a-b} + (b-c)\log\frac{b-c}{a-b} \\
&\qquad - b\log\frac{b}{a} - (a-b)\log\frac{a-b}{a} \\
&\qquad - \zeta' - \log\frac{b(a-b)}{(b-c)\sqrt{2\pi ac(a-2b+c)}}, \qquad (4)
\end{aligned}$$

for some ζ' such that

$$\zeta' \leq \frac{1}{12b} + \frac{1}{12(a-b)} + \frac{1}{12(a-b)+1} + \frac{1}{12b+1},$$

by Robbins' remark. Clearly though $\zeta' \leq \frac{1}{3}$ since $a > b \geq 1$, while

$$\log\frac{b(a-b)}{(b-c)\sqrt{2\pi ac(a-2b+c)}} \leq 2\log a$$

due to the constraints placed on a, b, c in the lemma statement. To simplify the remainder of the statement recognize that

$$b\log\frac{b}{a} = (b-c)\log\frac{b}{a} + c\log\frac{b}{a} \qquad (5)$$

$$(a-b)\log\frac{a-b}{a} = (a-2b+c)\log\frac{a-b}{a} + (b-c)\log\frac{a-b}{a}, \qquad (6)$$

hence

$$\begin{aligned}
&c\log\frac{c}{b} + (b-c)\log\frac{b-c}{b} \\
&+ (a-2b+c)\log\frac{a-2b+c}{a-b} + (b-c)\log\frac{b-c}{a-b} \\
&- b\log\frac{b}{a} - (a-b)\log\frac{a-b}{a} \\
&\quad = b\left[\frac{c}{b}\log\frac{\frac{c}{b}}{\frac{b}{a}} + \left(1-\frac{c}{b}\right)\log\frac{1-\frac{c}{b}}{1-\frac{b}{a}}\right] \\
&\quad + (a-b)\left[\left(1-\frac{b-c}{a-b}\right)\log\frac{1-\frac{b-c}{a-b}}{1-\frac{b}{a}} + \frac{b-c}{a-b}\log\frac{\frac{b-c}{a-b}}{\frac{b}{a}}\right] \qquad (7) \\
&\quad = a\left[\frac{b}{a}\mathbb{D}_2\left(\frac{c}{b}\middle\|\frac{b}{a}\right) + \left(1-\frac{b}{a}\right)\mathbb{D}_2\left(\frac{b-c}{a-b}\middle\|\frac{b}{a}\right)\right] \qquad (8)
\end{aligned}$$

and thus proving the lemma. ■

C. Proof of Theorem 17

Proof:

Once again $m = \times_{k\in\mathcal{K}} m_k$ and $\mathcal{M} = \times_{k\in\mathcal{K}}\mathcal{M}_k$.

The theorem will be proven by showing that Code Construction 35 can produce $(r, \mathcal{K}, \gamma)$-overlay codes with non-zero probability. Note, the fact that it can produce a code with non-zero probability directly implies the existence of such a code. Also note that the code construction near directly provides two of the $(r, \mathcal{K}, \gamma)$-code requirements. Indeed,

$$\begin{aligned}
\log|\mathcal{M}| &= \sum_{k\in\mathcal{K}} \log|\mathcal{M}_k| \qquad (9) \\
&\geq \sum_{k\in\mathcal{K}} n_k \left|\mathbb{I}_2\left(\gamma\middle\|\frac{\ell}{n_k}\right) - \frac{4}{3n_k} - \frac{2}{n_k}\log n_k\sqrt{\ell}\right|^+ \qquad (10) \\
&\geq r \qquad (11)
\end{aligned}$$

since $\left\lfloor e^{|a|^+}\right\rfloor \geq e^{|a-1|^+}$. While

$$\sum_{j=1}^{n} \mathbb{1}_{\{k\}}(f_j(m)) = \ell \qquad (12)$$

for each $k \in \mathcal{K}$ and $m \in \mathcal{M}$ is already directly implied by the code construction. What therefore remains to prove is that for each distinct pair of messages m, m' there exists a $j \in \mathcal{K}$ such that

$$\sum_{i=1}^{n} \mathbb{1}_{\{j\}}(f_i(m))\,\mathbb{1}_{\{j\}}(f_i(m')) \leq \gamma\ell \qquad (13)$$

and for all $t < j$

$$\sum_{i=1}^{n} \mathbb{1}_{\{t\}}(f_i(m))\,\mathbb{1}_{\{t\}}(f_i(m')) = \ell. \qquad (14)$$

It should be noted that Equation (14) implies the third overlay-code requirement of

$$\sum_{i=1}^{n} \mathbb{1}_{\{j\}}(f_i(m))\,\mathbb{1}_{\{t\}}(f_i(m')) = 0 \qquad (15)$$

for all $t < j$.

To this end consider any $m = \times_{k\in\mathcal{K}} m_k \in \mathcal{M}$ and $m' = \times_{k\in\mathcal{K}} m'_k \in \mathcal{M}$ such that $m \neq m'$. Specifically, let[17] $j \in \mathcal{K}$ be the minimum value such that $m_k \neq m'_j$, and note that for all $t < j$ if $f_i(m) = t$ then $f_i(m') = t$ since

$$\begin{aligned}
&g_{\mathcal{N}_t \to \mathcal{I}\left(\times_{i\in\mathcal{K}|i<t} m_i\right)}(\mathcal{S}_t(m_t)) \\
&\quad = g_{\mathcal{N}_t \to \mathcal{I}\left(\times_{i\in\mathcal{K}|i<t} m'_i\right)}(\mathcal{S}_k(m'_t)).
\end{aligned}$$

[16]For all positive integers n, $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\zeta}$ for some ζ such that $\frac{1}{12n+1} \leq \zeta \leq \frac{1}{12n}$.

[17]The value j is used here since it will become the value of j that satisfies Equations (13) and (14).

Clearly then for all $t < j$

$$\sum_{i=1}^{n} \mathbb{1}_{\{t\}}\left(f_i(m)\right)\mathbb{1}_{\{t\}}\left(f_i(m')\right) = \ell. \tag{16}$$

What remains is to show that

$$\sum_{i=1}^{n} \mathbb{1}_{\{j\}}\left(f_i(m)\right)\mathbb{1}_{\{j\}}\left(f_i(m')\right) \leq \gamma\ell, \tag{17}$$

which will be done via random coding arguments. In particular, for each $k \in \mathcal{K}$ we will show that with probability greater than zero the random choice of subsets in Code Construction 35 yields a code such that

$$|\mathcal{S}_k(a) \cap \mathcal{S}_k(b)| \leq \gamma\ell \tag{18}$$

for all $k \in \mathcal{K}$, $a \in \mathcal{M}_k$, and $b \in \mathcal{M}_k \setminus \{a\}$. If a code with property (18) is produced, then

$$\sum_{i=1}^{n} \mathbb{1}_{\{j\}}\left(f_i(m)\right)\mathbb{1}_{\{j\}}\left(f_i(m')\right) \leq |\mathcal{S}_j(m_j) \cap \mathcal{S}_j(m'_j)| \leq \gamma\ell, \tag{19}$$

since the combination of $g_{\mathcal{N}_j \to \mathcal{I}\left(\times_{i\in\mathcal{K}|i<j} m_i\right)}$ being an invertible mapping and

$$g_{\mathcal{N}_j \to \mathcal{I}\left(\times_{i\in\mathcal{K}|i<j} m_i\right)} = g_{\mathcal{N}_j \to \mathcal{I}\left(\times_{i\in\mathcal{K}|i<j} m'_i\right)}.$$

imply that

$$\begin{aligned} &f_i(m) = f_i(m') = j \\ &\quad \Leftrightarrow g^{-1}_{\mathcal{N}_j \to \mathcal{I}\left(\times_{a\in\mathcal{K}|a<j} m_i\right)}(\{i\}) \in \mathcal{S}_j(m_j) \cap \mathcal{S}_j(m'_j). \end{aligned}$$

To prove a code with Property (18) can be produced from the code construction, consider any $k \in \mathcal{K}$, and without loss of generality assume $\mathcal{M}_k = \{1, 2, \ldots, |\mathcal{M}_k|\}$. Further let $S_k(a)$ be the random variable representing the randomly chosen subset of $\mathcal{N}_k$ particular to each $a \in \mathcal{M}_k$. Observe that the probability Code Construction 35 generates a code satisfying (18) is

$$\Pr\left(\cap_{a=2}^{|\mathcal{M}_k|} Q_k(a)\right) = \prod_{a=2}^{|\mathcal{M}_k|} \Pr\left(Q_k(a) \middle| \cap_{b=1}^{a-1} Q_k(b)\right) \tag{20}$$

where for each $a \in \mathcal{M}_k$

$$Q_k(a) = \bigcap_{b=1}^{a-1} \{|S_k(a) \cap S_k(b)| \leq \gamma\ell\}.$$

But,

$$\begin{aligned} &\Pr\left(Q_k(a) \middle| \cap_{b=1}^{a-1} Q_k(b)\right) \\ &\quad = 1 - \Pr\left(\cup_{c=1}^{a-1} |S_k(a) \cap S_k(c)| > \gamma\ell \middle| \cap_{b=1}^{a-1} Q_k(b)\right), && (21) \\ &\quad \geq 1 - \sum_{c=1}^{a-1} \Pr\left(|S_k(a) \cap S_k(c)| > \gamma\ell \middle| \cap_{b=1}^{a-1} Q_k(b)\right) && (22) \\ &\quad \geq 1 - (a-1)\Pr\left(|S_k(a) \cap S_k(1)| > \gamma\ell\right); && (23) \end{aligned}$$

where (21) follows by De Morgan's Law; (22) is the union bound; and (23) is because $\{S_k(a)\}_{a\in\mathcal{M}_k}$ are independent and identically distributed. Therefore, from combining equations (20), (23), and the independence of the layer construction, it follows that if

$$\log|\mathcal{M}_k| + \log\Pr\left(|S_k(1) \cap \mathcal{S}_k(2)| > \gamma\ell\right) < 0 \tag{24}$$

for all $k \in \mathcal{K}$, then the probability Code Construction 35 produces a code with property (18) for all values of $k \in \mathcal{K}$ is greater than 0. To prove Equation (24) is indeed true, observe that

$$\begin{aligned} &\log\Pr\left(|S_k(1) \cap \mathcal{S}_k(2)| > \gamma\ell\right) \\ &\quad = \log \sum_{i=\lceil\gamma\ell\rceil}^{\ell} \Pr\left(|S_k(1) \cap \mathcal{S}_k(2)| = i\right) && (25) \\ &\quad = \log \sum_{i=\lceil\gamma\ell\rceil}^{\ell} \frac{\left|\binom{\mathcal{S}_k(2)}{i}\right|\left|\binom{\mathcal{N}_k - \mathcal{S}_k(2)}{\ell - i}\right|}{\left|\binom{\mathcal{N}_k}{\ell}\right|} && (26) \\ &\quad = \log \sum_{i=\lceil\gamma\ell\rceil}^{\ell} \frac{\binom{\ell}{i}\binom{n_k - \ell}{\ell - i}}{\binom{n_k}{\ell}} && (27) \\ &\quad \leq \log\max\left(\sum_{i=\lceil\gamma\ell\rceil}^{\ell} e^{-n_k\left(\mathbb{I}_2\left(\frac{i}{\ell}\middle\|\frac{\ell}{n_k}\right) - \frac{1}{3n_k} - \frac{2}{n_k}\log n_k\right)}, 1\right) && (28) \\ &\quad \leq -n_k\left|\mathbb{I}_2\left(\gamma\middle\|\frac{\ell}{n_k}\right) - \frac{1}{3n_k} - \frac{2}{n_k}\log n_k\sqrt{\ell}\right|^{+} && (29) \\ &\quad < -\log|\mathcal{M}_k|; && (30) \end{aligned}$$

where (28) is from Lemma 36 and because the probability of an event is at most 1; (29) is because $\gamma \geq \frac{1}{2} \geq \frac{\ell}{n_k}$ for all k, in turn implying

$$\mathbb{I}_2\left(\frac{i}{\ell}\middle\|\frac{\ell}{n_k}\right) \geq \mathbb{I}_2\left(\frac{\lceil\gamma\ell\rceil}{\ell}\middle\|\frac{\ell}{n_k}\right) \geq \mathbb{I}_2\left(\gamma\middle\|\frac{\ell}{n_k}\right)$$

and because a summation is always less than the maximum summand multiplied by total number of summands; finally (30) is by code construction. ■

D. Proof of Corollary 18

Proof:

Given $\gamma \in (\frac{1}{2}, 1)$ and finite $\mathcal{K} \subset [0, 1)$, recall that for all $k \in \mathcal{K}$

$$n_k = n - j_k\ell$$

where

$$\ell = \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|} \right\rfloor \quad \text{and} \quad j_k = |\{a \in \mathcal{K} | a < k\}|.$$

As a first step, observe that $\ell n_k \leq n^2$, hence

$$\begin{aligned} &\sum_{k\in\mathcal{K}} n_k\left|\mathbb{I}_2\left(\gamma\middle\|\frac{\ell}{n_k}\right) - \frac{4}{3n_k} - \frac{2}{n_k}\log\ell n_k\right|^{+} \\ &\quad \geq -|\mathcal{K}|\left(\frac{4}{3} + 4\log n\right) + \sum_{k\in\mathcal{K}} n_k\mathbb{I}_2\left(\gamma\middle\|\frac{\ell}{n_k}\right) && (31) \end{aligned}$$

and what remains is to lower bound $n_k \mathbb{I}_2\left(\gamma \middle\| \frac{\ell}{n_k}\right)$. To this end observe that

$$\begin{aligned} & n_k \mathbb{I}_2\left(\gamma \middle\| \frac{\ell}{n_k}\right) \\ & \geq \ell \mathbb{D}_2\left(\gamma \middle\| \frac{\ell}{n_k}\right) && (32) \\ & \geq \ell\left[\gamma \log\left(|\tilde{\mathcal{K}}| - j_k\right) + (1-\gamma)\log\left(\frac{n - j_k\ell}{n-(j_k+1)\ell}\right)\right] \\ & \quad - \ell \mathbb{H}_2(\gamma) && (33) \\ & \geq \ell\gamma \log\left(|\tilde{\mathcal{K}}| - j_k\right) - \ell \mathbb{H}_2(\gamma); && (34) \end{aligned}$$

where (32) is by the fact that the KL divergence is always greater than zero; (33) is because

$$\log\left(\frac{n}{\left\lceil \frac{n}{|\tilde{\mathcal{K}}|}\right\rceil} - j_k\right) \geq \log(|\tilde{\mathcal{K}}| - j_k);$$

and (34) is because

$$n - j_k\ell \geq n - (j_k+1)\ell \geq 0.$$

Using this bound,

$$\begin{aligned} & \sum_{k\in\mathcal{K}} n_k \mathbb{I}_2\left(\gamma \middle\| \frac{\ell}{n_k}\right) && (35) \\ & \geq \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|}\right\rfloor \left(-|\mathcal{K}|\mathbb{H}_2(\gamma) + \gamma \sum_{j=0}^{|\mathcal{K}|-1} \log(|\tilde{\mathcal{K}}| - j)\right) && (36) \\ & = \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|}\right\rfloor \left(-|\mathcal{K}|\mathbb{H}_2(\gamma) + \gamma \log(|\tilde{\mathcal{K}}|)!\right) && (37) \\ & \geq \left\lfloor \frac{n}{|\tilde{\mathcal{K}}|}\right\rfloor \left(-|\mathcal{K}|\mathbb{H}_2(\gamma) + \gamma|\tilde{\mathcal{K}}| \log\frac{|\tilde{\mathcal{K}}|}{e}\right) && (38) \\ & \geq (n - |\mathcal{K}| - 1)\left[\gamma \log|\tilde{\mathcal{K}}| - \gamma - \mathbb{H}_2(\gamma)\right], && (39) \end{aligned}$$

where (38) is a consequence of Stirling's Approximation of the factorial. Combining Equations (31), (38) with some further basic algebra and Theorem 17 yields the corollary statement. ■

Appendix C
Theorem 32

A technical lemma (which is essentially the Hoeffding lemma [20]), a basic calculation, an intuitively obvious lemma, and a bookkeeping lemma will be useful in the proof of Theorem 32. These are presented first, as to help streamline the proofs of the theorem.

Lemma 37:

$$\Pr\left(\sum_{i=1}^n G_{\rho,i}^2 \gtrless n(1\pm c)\rho\right) \leq \begin{cases} e^{-\frac{1}{8}c^2 n} & \text{if } c \leq 1 \\ e^{-\frac{1}{8}cn} & \text{else} \end{cases}$$

for all $c \geq 0$.

Proof:

What follows is essentially the derivation of Hoeffding [20, Equation (2.1)] followed by a loosening of the bound. Proving

$$\Pr\left(\sum_{i=1}^n G_{\rho,i}^2 > n(1+c)\rho\right) \leq e^{-\frac{1}{8}nc^2}$$

for all $c \geq 0$ since a bound for $\Pr\left(-\sum_{i=1}^n G_{\rho,i}^2 > -n(1-c)\rho\right)$ follows with the same steps. Thus, the lemma can be derived as follows

$$\begin{aligned} & \Pr\left(\sum_{i=1}^n G_{\rho,i}^2 > n(1+c)\rho\right) \\ & = \min_{t>0} \Pr\left(e^{t\sum_{i=1}^n G_{\rho,i}^2} > e^{tn(1+c)\rho}\right) && (40) \\ & \leq \min_{t>0} e^{-tn(1+c)\rho} \mathbb{E} e^{t\sum_{i=1}^n G_{\rho,i}^2} && (41) \\ & = \min_{t>0} e^{-tn(1+c)\rho} \prod_{i=1}^n \frac{1}{\sqrt{1-2t\rho}} && (42) \\ & = \min_{t>0} e^{-tn(1+c)\rho} (1-2t\rho)^{\frac{-n}{2}} && (43) \\ & = e^{-\frac{nc}{2}} (1+c)^{\frac{n}{2}} && (44) \\ & \leq \begin{cases} e^{-\frac{1}{8}c^2 n} & \text{if } c \leq 1 \\ e^{-\frac{1}{8}cn} & \text{else} \end{cases} && (45) \end{aligned}$$

where (40) is Bernstein's trick; (41) is Markov's inequality; (42) is because $G_{\rho,i}$ is independent for each i and hence

$$\mathbb{E} e^{t\sum_{i=1}^n G_{\rho,i}^2} = \prod_{i=1}^n \mathbb{E} e^{tG_{\rho,i}^2}$$

while

$$\begin{aligned} \mathbb{E}\left[e^{tG_{\rho,i}^2}\right] &= \int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\rho}} e^{-\frac{x^2}{2\rho} + tx^2} \mathrm{d}x \\ &= \int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\rho}} e^{-(1-t2\rho)\frac{x^2}{2\rho}} \mathrm{d}x \\ &= \frac{1}{\sqrt{1-2t\rho}} \int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\tilde{\rho}}} e^{-\frac{x^2}{2\tilde{\rho}}} \mathrm{d}x \\ &= \frac{1}{\sqrt{1-2t\rho}} \end{aligned}$$

where $\tilde{\rho} = \frac{\rho}{1-2t\rho}$; (44) is the result of solving the minimization problem, and then substituting the minimum $t = \frac{c}{2(1+c)\rho}$ back in; finally (45) is because

$$c - \log(1+c) \geq \min_{c_0\in[0,c]} \frac{c^2}{2}\frac{1}{1+c_0} \geq \begin{cases} \frac{1}{4}c^2 & \text{if } c \leq 1 \\ \frac{1}{4}c & \text{else} \end{cases}$$

for all $c \geq 0$ by Taylor's theorem. ■

Calculation 38: If $X = G_\rho$ then

$$X|\{X + G_a = z\} = G_{\frac{\rho a}{\rho+a}} + \frac{\rho}{\rho+a} z.$$

Proof: Letting $Y = G_a$ and $Z = X + Y$, and $f_{X,Y,Z}$ denote the probability density functions of the various random variables, the calculation follows

$$\begin{aligned} & f_{X|Z}(x|z) \\ & = \frac{f_{Z|X}(z|x) f_X(x)}{f_Z(z)} \\ & = \sqrt{\frac{\rho+a}{2\pi\rho a}} \exp\left(-\frac{(z-x)^2}{2a} - \frac{x^2}{2\rho} + \frac{z^2}{2(\rho+a)}\right) \\ & = \sqrt{\frac{\rho+a}{2\pi\rho a}} \exp\left(-\frac{x^2 - 2\frac{\rho}{a+\rho}xz + \frac{\rho^2}{(a+\rho)^2}z^2}{2\frac{\rho a}{a+\rho}}\right) \end{aligned}$$

$$= \sqrt{\frac{1}{2\pi\frac{\rho a}{\rho+a}}} \exp\left(-\frac{\left(x-\frac{\rho}{\rho+a}z\right)^2}{2\frac{\rho a}{a+\rho}}\right). \tag{46}$$

■

Lemma 39: Let $\boldsymbol{G}$ be independent (but not identical) Gaussian RVs with mean 0 and finite (but otherwise arbitrary) variance, and let $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$ be fixed. For all fixed $a > 0$

$$\Pr\left(\sum_{i=1}^{n}(G_i+\mu_i)^2 \le a\right) \le \Pr\left(\sum_{i=1}^{n} G_i^2 \le a\right).$$

Proof: To prove the lemma, we need to show

$$\Pr\left((G_i+\mu_i)^2 \le a\right) \le \Pr\left(G_i^2 \le a\right) \tag{47}$$

since the more general lemma will then follow from repeated use of the following observation that uses Equation (47)

$$\begin{aligned}
&\Pr\left(\sum_{i=1}^{n}(G_i+\mu_i)^2 \le a\right)\\
&= \int \Pr\left((G_1+\mu_1)^2 \le a-b\right) \mathrm{d}\Pr\left(\sum_{i=2}^{n}(G_i+\mu_i)^2 \le b\right)\\
&\le \int \Pr\left(G_1^2 \le a-b\right) \mathrm{d}\Pr\left(\sum_{i=2}^{n}(G_i+\mu_i)^2 \le b\right)\\
&= \Pr\left(G_1^2 + \sum_{i=2}^{n}(G_i+\mu_i)^2 \le a\right).
\end{aligned}$$

To prove Equation (47), it is helpful to simplify it to

$$\begin{aligned}
&\Pr\left(-\mu_i-\sqrt{a} \le G_i \le -\mu_i+\sqrt{a}\right)\\
&\quad\le \Pr\left(-\sqrt{a} \le G_i \le \sqrt{a}\right),
\end{aligned} \tag{48}$$

or even more directly

$$\begin{aligned}
&\Phi\left(\frac{-\mu_i+\sqrt{a}}{\sqrt{\rho_i}}\right) - \Phi\left(\frac{-\mu_i-\sqrt{a}}{\sqrt{\rho_i}}\right)\\
&\quad\le \Phi\left(\sqrt{\frac{a}{\rho_i}}\right) - \Phi\left(-\sqrt{\frac{a}{\rho_i}}\right),
\end{aligned} \tag{49}$$

by taking square roots and then using basic algebraic manipulation. Equation (49) can be validated by showing that $\mu_i = 0$ maximizes

$$\Phi\left(\frac{-\mu_i+\sqrt{a}}{\sqrt{\rho_i}}\right) - \Phi\left(\frac{-\mu_i-\sqrt{a}}{\sqrt{\rho_i}}\right). \tag{50}$$

Using the basic calculus approach, the derivative of Equation (50) is

$$\frac{\partial(50)}{\partial\mu_i} = \frac{-1}{\sqrt{2\pi\rho_i}}\left[e^{-\frac{(-\mu_i+\sqrt{a})^2}{2\rho_i}} - e^{-\frac{(-\mu_i-\sqrt{a})^2}{2\rho_i}}\right]. \tag{51}$$

Setting the derivative equal to zero and solving gives $|\mu_i + \sqrt{a}| = |-\mu_i+\sqrt{a}|$, which can be further simplified to $2\mu_i = 0$ since $a > 0$. Furthermore, the second derivative at $\mu_i = 0$ is

$$-\sqrt{\frac{2a}{\pi\rho_i^3}}e^{-\frac{a}{2\rho_i}} < 0,$$

thus guaranteeing that $\mu_i = 0$ is the global maximum in turn proving Equation (47) and the lemma. ■

Lemma 40: Suppose that $\tau_i \ge \alpha > 0$ for $i \in \{1,\dots,n\}$, and that $\beta \ge \alpha$. Then for all positive real numbers b, c, and γ, where $\gamma \le \frac{1}{n}|\{i \in \{1,\dots,n\} | \tau_i < \beta\}|$,

$$\Pr\left(\sum_{i=1}^{n} G_{\boldsymbol{\tau},i}^2 \le n(1+c)b\right) \le e^{-\frac{1}{8}n\gamma\lambda^2} + e^{-\frac{1}{8}n(1-\gamma)\lambda^2},$$

where

$$\lambda = \max\left(0, 1 - \frac{(1+c)b}{\gamma\alpha + (1-\gamma)\beta}\right).$$

Proof: Choose any $\mathcal{B} \subset \{1,\dots,n\}$ such that $|\mathcal{B}| = n\gamma$ and all coordinates in $\mathcal{B}$ correspond to $\tau_i < \beta$, i.e.,

$$\mathcal{B} \subseteq \{i \in \{1,\dots,n\} | \tau_i < \beta\}.$$

Let $\bar{\mathcal{B}} = \{1,\dots,n\} \setminus \mathcal{B}$.

Now the proof is trivial for $\lambda = 0$, otherwise when $\lambda > 0$ the results follows as so.

$$\begin{aligned}
&\Pr\left(\sum_{i=1}^{n} G_{\boldsymbol{\tau},i}^2 \le n(1+c)b\right)\\
&\quad= \Pr\left(\sum_{i\in\mathcal{B}} G_{\boldsymbol{\tau},i}^2 + \sum_{i\in\bar{\mathcal{B}}} G_{\boldsymbol{\tau},i}^2 \le n(1+c)b\right) &(52)\\
&\quad\le \Pr\left(\sum_{i\in\mathcal{B}} \frac{\alpha}{\tau_i} G_{\boldsymbol{\tau},i}^2 + \sum_{i\in\bar{\mathcal{B}}} \frac{\beta}{\tau_i} G_{\boldsymbol{\tau},i}^2 \le n(1+c)b\right) &(53)\\
&\quad\le \Pr\left(\sum_{i\in\mathcal{B}} \frac{\alpha}{\tau_i} G_{\boldsymbol{\tau},i}^2 \le n\gamma(1-\lambda)\alpha\right)\\
&\qquad+ \Pr\left(\sum_{i\in\bar{\mathcal{B}}} \frac{\beta}{\tau_i} G_{\boldsymbol{\tau},i}^2 \le n(1-\gamma)(1-\lambda)\beta\right) &(54)\\
&\quad\le e^{-\frac{1}{8}n\gamma\lambda^2} + e^{-\frac{1}{8}n(1-\gamma)\lambda^2} &(55)
\end{aligned}$$

where (53) is because $\alpha/\tau_i \le 1$ for all $i \in \mathcal{B}$ and $\beta/\tau_i \le 1$ for all $i \in \bar{\mathcal{B}}$; (54) is by using the inequality

$$\begin{aligned}
\Pr(A+B \le a+b) &\le \Pr(\{A \le a\} \text{ or } \{B \le b\})\\
&\le \Pr(A \le a) + \Pr(B \le b)
\end{aligned}$$

in conjunction with

$$n(1+c)b \le n\gamma(1-\lambda)\alpha + n(1-\gamma)(1-\lambda)\beta;$$

and (54) is by Lemma 37 and because $\frac{\alpha}{\tau_i}G_{\boldsymbol{\tau},i}^2 = G_\alpha^2$ and $\frac{\beta}{\tau_i}G_{\boldsymbol{\tau},i}^2 = G_\beta^2$. ■

A. Proof of Theorem 32

Proof Sketch: Let $\mathcal{H} = (\boldsymbol{x} : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m}' : \boldsymbol{\mathcal{R}} \to \mathcal{M})$ be the original code, and let $\mathcal{J} = (\boldsymbol{X}' : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m}' : \boldsymbol{\mathcal{R}} \to \mathcal{M} \cup \{!\})$ be the modified code obtained from Code Modification 21. By $\mathcal{I}_k(m)$, for each $k \in \tilde{\mathcal{K}}$ and $m \in \mathcal{M}$, denote all coordinates $i \in \{1,\dots,n\}$ such that $f_i(m) = k$.

Both the power constraint and the average arithmetic probability of error arguments will rely on random coding (due to the random choice of $\boldsymbol{t} : \mathcal{M} \to \boldsymbol{\mathcal{R}}$). Because of this, let $\boldsymbol{T} : \mathcal{M} \to \boldsymbol{\mathcal{R}}$ be the random variable representing the randomly chosen value of $\boldsymbol{t}$ in the code construction.

The random coding construction will proceed by showing that the random choice of $\boldsymbol{T}$ with probability greater than

$$1-\left(1+\sqrt{\frac{2}{\pi}}\right)e^{-n}$$

yields a code with the stated power constraint, and with probability greater than

$$1-e^{-n}$$

yields a code with the stated average arithmetic probability of error. Clearly, this also implies that the random choice of $\boldsymbol{T}$ yields a code which satisfies both the power constraint and the average arithmetic probability of error bound with probability greater than

$$1-\left(2+\sqrt{\frac{2}{\pi}}\right)e^{-n}.$$

For readability, we have separated the derivation of each bound by a dividing line.

(Rate)
Encoders for $\mathcal{J}$ and $\mathcal{H}$ have the same domain hence

$$r_{\mathcal{J}}=r_{\mathcal{H}}.$$

(Power) Towards the power constraint observe that for each message $m\in\mathcal{M}$

$$\sum_{i=1}^{n}\frac{1}{n}\mathbb{E}\left[(X_i'(m))^2\right]$$

$$=\sum_{i=1}^{n}\frac{1}{n}\mathbb{E}\left[(x_i(m)+t_i(m)+f_i(m)G_{\Delta,i})^2\right] \tag{56}$$

$$=\sum_{i=1}^{n}\frac{1}{n}\left(x_i^2(m)+t_i^2(m)+2t_i(m)x_i(m)+f_i^2(m)\rho_\Delta\right); \tag{57}$$

$$\leq\omega_{\mathcal{H}}+\frac{1}{n}\sum_{i=1}^{n}2\ t_ix_i+\frac{1}{n}\sum_{k\in\tilde{\mathcal{K}}}\sum_{i\in\mathcal{I}_k(m)}\left(t_i^2(m)+k^2\rho_\Delta\right); \tag{58}$$

where (57) is because $\mathbb{E}\left[G_{\Delta,i}^2\right]=\rho_\Delta$ and $\mathbb{E}\left[G_{\Delta,i}\right]=0$; and (58) is by definition of the power constraint. Thus, we need to bound the tail probability for choosing large values of $t_i^2(m)$ and $2t_i(m)x_i(m)$.

To this end, for each $m\in\mathcal{M}$ and $k\in\mathcal{K}$

$$\Pr\left(\sum_{i\in\mathcal{I}_k(m)}T_i^2(m)\geq\ell\,(1+c)\,(1-k^2)\rho_\Delta\right)\leq e^{-n[r_{\mathcal{H}}+1]-\log|\mathcal{K}|}, \tag{59}$$

where

$$c=8\frac{n}{\ell}\left[r_{\mathcal{H}}+1+\frac{\log|\mathcal{K}|}{n}\right]>1,$$

by Lemma 37. That with probability $1-e^{-n}$ a $\boldsymbol{T}=\boldsymbol{t}$ is chosen such that

$$\sum_{i=1}^{n}(t_i^2(m)\ +k^2\rho_\Delta)\leq n\left(1+8\frac{n}{\ell}\left[r_{\mathcal{H}}+1+\frac{\log|\mathcal{K}|}{n}\right]\right)\rho_\Delta \tag{60}$$

for all m follows by applying the union bound to extend (59) to simultaneously consider all $m\in\mathcal{M}$ and $k\in\mathcal{K}$ (while observing that $t_i(m)=0$ for all $i\in\mathcal{I}_1(m)$).

The other term in the summation, $\sum_{i=1}^{n}2\ x_i(m)T_i(m)$, follows directly from basic laws of probability. Specifically, for each $m\in\mathcal{M}$ we have

$$\Pr\left(\sum_{i=1}^{n}2\ x_i(m)T_i(m)\geq n\ 2\sqrt{2\omega_{\mathcal{H}}(r_{\mathcal{H}}+1)\rho_\Delta}\right)$$

$$=\Phi\left(\frac{-n2\sqrt{2\omega_{\mathcal{H}}(r_{\mathcal{H}}+1)\rho_\Delta}}{\sqrt{\sum_{i=1}^{n}4\ x_i^2(m)(1-f_i^2(m))\rho_\Delta}}\right) \tag{61}$$

$$\leq\sqrt{\frac{2}{\pi}}\exp\left(-\frac{n^2\omega_{\mathcal{H}}(r_{\mathcal{H}}+1)}{\sum_{i=1}^{n}x_i^2(m)(1-f_i^2(m))}\right) \tag{62}$$

$$\leq\sqrt{\frac{2}{\pi}}e^{-n(r_{\mathcal{H}}+1)}; \tag{63}$$

where (61) is because $2x_i(m)T_i(m)$ is a sum of independent Gaussian random variables by the code construction; (62) is because $\Phi(t)\leq\sqrt{\frac{2}{\pi}}e^{-\frac{t^2}{2}}$ for $t\leq 0$; and (63) is because $0\leq f_i(m)\leq 1$ for all coordinates $i\in\{1,\ldots,n\}$ and $m\in\mathcal{M}$. Once again, that with probability greater than $1-\sqrt{\frac{2}{\pi}}e^{-n}$ a $\boldsymbol{T}=\boldsymbol{t}$ is chosen such that

$$\sum_{i=1}^{n}2\ x_i(m)t_i(m)\leq n\ 2\sqrt{2\omega_{\mathcal{H}}(r_{\mathcal{H}}+1)\rho_\Delta} \tag{64}$$

follows by applying the union bound to Equation (63) as to consider all m jointly.

Combining Equations (58), (60), (64), and that $\frac{n}{\ell}\leq|\tilde{\mathcal{K}}|$ shows that with probability

$$1-\left(1+\sqrt{\frac{2}{\pi}}\right)e^{-n}$$

a $\boldsymbol{T}=\boldsymbol{t}$ is chosen such that

$$\omega_{\mathcal{J}}\leq\omega_{\mathcal{H}}+2\sqrt{2\omega_{\mathcal{H}}(r_{\mathcal{H}}+1)\rho_\Delta}+\left(1+8|\tilde{\mathcal{K}}|\left[r_{\mathcal{H}}+1+\frac{\log|\mathcal{K}|}{n}\right]\right)\rho_\Delta. \tag{65}$$

(Average arithmetic probability of error)
To prove the bound on the average arithmetic error probability, observe that the condition for error given $M=m$,

$$\hat{m}'(\boldsymbol{x}(m)+\boldsymbol{t}(m)+\boldsymbol{f}(m)\cdot\boldsymbol{G}_\Delta+\boldsymbol{G}_{\mathrm{Dec}})\neq m, \tag{66}$$

occurs if and only

$$\hat{m}(\boldsymbol{x}(m)+\boldsymbol{t}(m)+\boldsymbol{f}(m)\cdot\boldsymbol{G}_\Delta+\boldsymbol{G}_{\mathrm{Dec}})\neq m, \tag{67}$$

or if there exists a $k\in\mathcal{K}$ such that

$$\sum_{i\in\mathcal{I}_k(m)}\frac{(kG_{\Delta,i}+G_{\mathrm{Dec},i})^2}{k^2\rho_\Delta+\rho_{\mathrm{Dec}}}\geq\ell(1+\delta). \tag{68}$$

Hence

$$\varepsilon_{\mathcal{J}}(\rho_{\text{Dec}}) \leq \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} a_m(\boldsymbol{t}) + \frac{1}{|\mathcal{M}|}\sum_{\substack{k\in\mathcal{K}\\ m\in\mathcal{M}}} b_m(k), \quad (69)$$

where

$$a_m(\boldsymbol{t}) = \Pr\left(\hat{m}(\boldsymbol{x}(m)+\boldsymbol{t}(m)+\boldsymbol{f}(m)\cdot\boldsymbol{G}_\Delta + \boldsymbol{G}_{\text{Dec}}) \neq m\right)$$

$$b_m(k) = \Pr\left(\sum_{i\in\mathcal{I}_k(m)} \frac{(kG_{\Delta,i}+G_{\text{Dec},i})^2}{k^2\rho_\Delta+\rho_{\text{Dec}}} \geq \ell(1+\delta)\right),$$

by the union bound. Note from the code construction that random variables $a_m(\boldsymbol{T})$ and $a_{m'}(\boldsymbol{T})$ are independent when $m \neq m'$, and that $0 \leq a_m(\boldsymbol{T}) \leq 1$ for all $m \in \mathcal{M}$, and that

$$\mathbb{E}[a_m(\boldsymbol{T})] = \Pr\left(\hat{m}(\boldsymbol{x}(m)+\boldsymbol{G}_{\rho_\Delta+\rho_{\text{Dec}}}) \neq m\right) \quad (70)$$

because $T_i(m)+f_i(m)G_{\Delta,i}+G_{\text{Dec},i}$ has a Gaussian distribution with mean 0 and variance $\rho_\Delta+\rho_{\text{Dec}}$ for each coordinate i. Therefore

$$\Pr\left(\sum_{m\in\mathcal{M}} \frac{a_m(\boldsymbol{T})}{|\mathcal{M}|} \geq \varepsilon_{\mathcal{H}}(\rho_\Delta+\rho_{\text{Dec}}) + \sqrt{\frac{n}{2|\mathcal{M}|}}\right) \leq e^{-n} \quad (71)$$

follows from Hoeffding's inequality because

$$\sum_{m\in\mathcal{M}} \frac{\Pr\left(\hat{m}(\boldsymbol{x}(m)+\boldsymbol{G}_{\rho_\Delta+\rho_{\text{Dec}}}) \neq m\right)}{|\mathcal{M}|} = \varepsilon_{\mathcal{H}}(\rho_\Delta+\rho_{\text{Dec}}).$$

On the other hand

$$b_m(k) \leq e^{-\frac{1}{8}\ell\delta^2} \quad (72)$$

comes directly from Lemma 37 since $\frac{kG_{\Delta,i}+G_{\text{Dec},i}}{\sqrt{k^2\rho_\Delta+\rho_{\text{Dec}}}}$ are independent Gaussian random variables with mean 0 and variance 1 for each coordinate i. Therefore, with probability greater than $1-e^{-n}$, a function $\boldsymbol{T}=\boldsymbol{t}$ will be chosen such that

$$\varepsilon_{\mathcal{J}}(\rho_{\text{Dec}}) \leq \varepsilon_{\mathcal{H}}(\rho_{\text{Dec}}+\rho_\Delta) + \sqrt{\frac{n}{2|\mathcal{M}|}} + |\mathcal{K}|e^{-\frac{1}{8}\ell\delta^2}. \quad (73)$$

(Maximum probability of targeted false authentication)
To prove the bound on the maximum probability of targeted false authentication, fix messages $M=m$ and $m' \neq m$ with the intention that m' is the targeted message, let $k \in \mathcal{K}$ be the index such that

$$\sum_{i=1}^{n} \mathbb{1}_{\{k\}}(f_i(m))\,\mathbb{1}_{\{k\}}(f_i(m')) \leq \gamma\ell,$$

and

$$\sum_{i=1}^{n} \mathbb{1}_{\{k\}}(f_i(m))\,\mathbb{1}_{\{j\}}(f_i(m')) = 0$$

for all $j<k$. The probability that the decoder will produce m' is always less than

$$\Pr\left(\sum_{i\in\mathcal{I}_k(m')} \frac{(Y_i-t_i(m')-x_i(m'))^2}{k^2\rho_\Delta+\rho_{\text{Dec}}} \leq \ell(1+\delta)\right), \quad (74)$$

where as a reminder

$$\boldsymbol{Y} = \boldsymbol{x}(m)+\boldsymbol{t}(m)+\boldsymbol{f}(m)\cdot\boldsymbol{G}_\Delta + \boldsymbol{Z}(\boldsymbol{V},m),$$

due to the code modification to the decoder.

Assume for now (we will come back to prove this after finishing the proof, see after break) that

$$(74) \leq \Pr\left(\sum_{i\in\mathcal{I}_k(m')} G^2_{\boldsymbol{\tau}(m),i} \leq \ell(1+\delta)(k^2\rho_\Delta+\rho_{\text{Dec}})\right), \quad (75)$$

where

$$\tau_i(m) = \tau^\star(f_i(m)) := \frac{f_i^2(m)\rho_\Delta\rho_{\text{Adv}}}{f_i^2(m)\rho_\Delta+\rho_{\text{Adv}}} + \rho_{\text{Dec}}.$$

With Equation (74) assumed, the following properties of the overlay-code allow for application of Lemma 40:

- $$|\mathcal{I}_k(m')| = \ell,$$
- $$|\mathcal{I}_k(m') \cap \mathcal{I}_k(m)| \leq \gamma\ell,$$
- $$\tau_i(m) = \tau^\star(k) = \frac{k^2\rho_\Delta\rho_{\text{Adv}}}{k^2\rho_\Delta+\rho_{\text{Adv}}} + \rho_{\text{Dec}}$$
  for all $i \in \mathcal{I}_k(m') \cap \mathcal{I}_k(m)$,
- and
  $$\tau_i(m) \geq \tau^\star(d_k) = \frac{d_k^2\rho_\Delta\rho_{\text{Adv}}}{d_k^2\rho_\Delta+\rho_{\text{Adv}}} + \rho_{\text{Dec}},$$
  where $d_k = \min\{a \in \tilde{\mathcal{K}} | a > k\}$, for all $i \in \mathcal{I}_k(m') \setminus \mathcal{I}_k(m)$.

With these overlay-code properties we can directly apply Lemma 40 to upper-bound the right-hand side of Equation (75) in turn yielding

$$(74) \leq e^{-\frac{1}{8}\ell\gamma\lambda_k^2} + e^{-\frac{1}{8}\ell(1-\gamma)\lambda_k^2}, \quad (76)$$

where

$$\lambda_k = \max\left(0, 1-\frac{(1+\delta)(k^2\rho_\Delta+\rho_{\text{Dec}})}{\gamma\tau^\star(k)+(1-\gamma)\tau^\star(d_k)}\right).$$

Recall now that Equation (74) assumed a fixed m and $m' \neq m$ and that the maximum probability of targeted false authentication is a maximum over all pairs of $m, m' \neq m$. Thus the maximum, over all m and $m' \neq m$, of the right-hand side of Equation (76) is also an upper bound on $\alpha^\star_{\mathcal{H}}(\rho_{\text{Adv}}, \rho_{\text{Dec}})$. Clearly though, the maximum of the right-hand side of Equation (76) corresponds to the minimum value of λ_k. Hence the final result

$$\alpha^\star_{\mathcal{H}}(\rho_{\text{Adv}}, \rho_{\text{Dec}}) \leq e^{-\frac{1}{8}\ell\gamma\lambda^2} + e^{-\frac{1}{8}\ell(1-\gamma)\lambda^2}, \quad (77)$$

where

$$\lambda = \max\left(0, \min_{k\in\mathcal{K}} 1-\frac{(1+\delta)(k^2\rho_\Delta+\rho_{\text{Dec}})}{\gamma\tau^\star(k)+(1-\gamma)\tau^\star(d_k)}\right).$$

We now return to prove Equation (75). Here we will primarily use the inequality $\Pr(\cdot) \leq \sup_a \Pr(\cdot|A=a)$ along with calculation 38. To that end note

$$\boldsymbol{Y} - \boldsymbol{x}(m') - \boldsymbol{t}(m') | \{\boldsymbol{V}, \boldsymbol{Z} = \boldsymbol{v}, \boldsymbol{z}\} = \boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}(\boldsymbol{v}, \boldsymbol{z}) \tag{78}$$

where

$$\begin{aligned}\mu_i(\boldsymbol{v}, \boldsymbol{z}) &= z_i - x_i(m') - t_i(m') \\ &\quad + \frac{f_i^2(m)\rho_\Delta(v_i - x_i(m) - t_i(m))}{f_i^2(m)\rho_\Delta + \rho_{\text{Adv}}} \\ \tau_i(m) &= \frac{f_i^2(m)\rho_\Delta\rho_{\text{Adv}}}{f_i^2(m)\rho_\Delta + \rho_{\text{Adv}}} + \rho_{\text{Dec}},\end{aligned}$$

as a consequence of calculation 38. Hence (74) must itself be less than

$$\sup_{\boldsymbol{v},\boldsymbol{z}} \Pr\left(\sum_{i \in \mathcal{I}_k(m')} \left(G_{\boldsymbol{\tau}(m),i} + \mu_i(\boldsymbol{v}, \boldsymbol{z})\right)^2 \leq c \middle| \boldsymbol{V}, \boldsymbol{Z} = \boldsymbol{v}, \boldsymbol{z} \right) \tag{79}$$

where $c = \ell(1+\delta)(k^2\rho_\Delta + \rho_{\text{Dec}})$. Applying Lemma 39 to (79), and recognizing that the resulting probability is independent of $\boldsymbol{V}, \boldsymbol{Z}$ proves

$$(74) \leq \Pr\left(\sum_{i \in \mathcal{I}_k(m')} \left(G_{\boldsymbol{\tau}(m),i}\right)^2 \leq c \right), \tag{80}$$

which is exactly Equation (75).

□

APPENDIX D
THEOREM 34

The proof of the Theorem 34 will rely on the following technical lemmas.

The first of these technical lemmas will be used to create a finite subset of points that when bounded also bound the original set.

Lemma 41: Let $\mathcal{X}^* \subseteq [a, b]$, for real numbers a and $b > a$. For a given positive real number c, there exists an $\boldsymbol{\mathcal{X}}^\dagger \subset \boldsymbol{\mathcal{R}}$ such that $|\boldsymbol{\mathcal{X}}^\dagger| = \lfloor (b-a)c \rfloor^n$ and

$$\begin{aligned}&\sup_{\boldsymbol{x}^* \in \boldsymbol{\mathcal{X}}^*} \Pr\left(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x}^* \in \mathcal{D}\right) \\ &\quad \leq \max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}^\dagger} \Pr\left(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x}^\dagger \in \mathcal{D}\right) + c^{-1}\sqrt{\sum_{i=1}^n \frac{1}{2\rho_i}}\end{aligned}$$

simultaneously for all $\mathcal{D} \subseteq \boldsymbol{\mathcal{R}}$.

Proof;

First we identify $\boldsymbol{\mathcal{X}}^\dagger$, where

$$\mathcal{X}^\dagger = \left\{a + c^{-1}, a + 2c^{-1}, \ldots, a + \lfloor (b-a)c \rfloor c^{-1}\right\},$$

as the set guaranteed in the lemma. It is immediate that $|\boldsymbol{\mathcal{X}}^\dagger| = \lfloor (b-a)c \rfloor^n$.

Now, for each $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}^*$ consider the corresponding $\boldsymbol{x}^\dagger = \arg\min_{\boldsymbol{x}^\dagger \in \boldsymbol{\mathcal{X}}^\dagger} |\boldsymbol{x} - \boldsymbol{x}^\dagger|$, and note that $\boldsymbol{x}^\dagger \in \boldsymbol{\mathcal{X}}^\dagger$ by definition. Here, $|x_i - x_i^\dagger| \leq c^{-1}$ for all coordinates $i \in \{1, \ldots, n\}$. Hence, for $\boldsymbol{x}$ and corresponding $\boldsymbol{x}^\dagger$ it follows that

$$\begin{aligned}&\left|\Pr\left(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x} \in \mathcal{D}\right) - \Pr\left(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x}^\dagger \in \mathcal{D}\right)\right| \\ &\quad \leq \sqrt{\frac{1}{2}\mathbb{D}(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x} || \boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x}^\dagger)} \qquad (81) \\ &\quad \leq c^{-1}\sqrt{\sum_{i=1}^n \frac{1}{2\rho_i}}; \qquad (82)\end{aligned}$$

where (81) is by Pinsker's inequality and the convexity of the KL divergence; while (82) is because

$$\mathbb{D}\left(\boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x} \middle|\middle| \boldsymbol{G}_{\boldsymbol{\rho}} + \boldsymbol{x}^\dagger\right) = \sum_{i=1}^n \frac{(x_i - x_i^\dagger)^2}{2\rho_i} \leq c^{-2}\sum_{i=1}^n \frac{1}{2\rho_i}.$$

This proves the lemma since for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}^*$ there is a corresponding $\boldsymbol{x}^\dagger \in \boldsymbol{\mathcal{X}}^\dagger$ such that Equation (82) holds independent of $\mathcal{D}$.

■

Next we provide a corollary of the well known Hoeffding Lemma. While we will prove the corollary, we point readers to [20, Section 5] for proof of the lemma, and note that uniformly selecting m values without replacement is equivalent to uniformly selecting a subset of size m.

Lemma 42 ([20, Section 5]): Let $\mathcal{A}$ be a finite set, β an integer less than $|\mathcal{A}|$, and let $p : \mathcal{A} \to [0, 1]$. If B is uniform over $\binom{\mathcal{A}}{\beta}$ then

$$\begin{aligned}\Pr\left(\sum_{a \in \mathcal{A}} \mathbb{1}_{\{B\}}(a)\, p(a) \geq c\beta\mu\right) &\leq \exp(-\beta\mathbb{D}_2(c\mu||\mu)), \\ &\leq \exp(-2\beta[(c-1)\mu]^2),\end{aligned}$$

where $\mu = |\mathcal{A}|^{-1}\sum_{a \in \mathcal{A}} p(a)$, for all real numbers $c \in (1, \mu^{-1})$.

Corollary 43: Additionally if $\max_{a \in \mathcal{A}} p(a) = \eta < 1$ then

$$\begin{aligned}&\Pr\left(\sum_{a \in \mathcal{A}} \mathbb{1}_{\{B\}}(a)\, p(a) \geq c\beta\mu\right) \\ &\quad \leq \exp\left(-\beta\mathbb{D}_2\left(c\frac{\mu}{\eta}\middle|\middle|\frac{\mu}{\eta}\right)\right) \\ &\quad \leq \exp\left(-\frac{c\beta\mu}{\eta}\left(\log(c) - \frac{1}{2} - \frac{1}{2\left(1 - c\frac{\mu}{\eta}\right)}\right)\right)\end{aligned}$$

for all real numbers $c \in (1, \eta\mu^{-1})$.

Proof: The first inequality comes from substituting $\frac{p(a)}{\eta}$ for $p(a)$ (and subsequently $\frac{\mu}{\eta}$ for μ) in Lemma 42.

The second inequality comes from recognizing that

$$\left(1 - \frac{\mu}{\eta}c\right)\log\frac{1 - \frac{\mu}{\eta}c}{1 - \frac{\mu}{\eta}} \geq \left(1 - \frac{\mu}{\eta}c\right)\log\left(1 - \frac{\mu}{\eta}c\right)$$

and that if $a \in [0, b]$, where $0 \leq b \leq 1$, then

$$(1-a)\log(1-a) \geq -a - \frac{a^2}{2(1-b)} \geq -a - \frac{a^2}{2(1-a)}$$

by Taylor's theorem.

■

A. Proof of Theorem 34

Proof:

Since Code Modification 33 builds on Code Modification 21, let $\mathcal{H} = (x : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m} : \boldsymbol{\mathcal{R}} \to \mathcal{M})$ be the original encoder and decoder, and let $\mathcal{L} = (\boldsymbol{X}' : \mathcal{M} \to \boldsymbol{\mathcal{R}}, \hat{m}' : \boldsymbol{R} \to \mathcal{M} \cup \{!\})$ be the code after Code Modification 21. We will assume that the operational measures of $\mathcal{L}$ are bounded as in Theorem 32.

From the Proof of Theorem 32 it is important to recall that for each m, v, z there exists some $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$ such that

$$\boldsymbol{Y}|\{M, \boldsymbol{V}, \boldsymbol{Z} = m, v, z\} = \boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu} \tag{83}$$

where

$$\tau_i(m) = \tau^\star(f_i(m)) = \frac{f_i^2(m)\rho_\Delta \rho_{\text{Adv}}}{f_i^2(m)\rho_\Delta + \rho_{\text{Adv}}} + \rho_{\text{Dec}}.$$

Key to the proof of the upper-bound on the maximum probability of false authentication is that the decimation of the message set will not change the above.

The proof will rely on the random selection of the new (decimated) message set, $\mathcal{M}^\ddagger$, for the final code $\mathcal{J} = (\boldsymbol{X}^\ddagger : \mathcal{M}^\ddagger \to \boldsymbol{\mathcal{R}}, \hat{m}^\ddagger : \boldsymbol{\mathcal{R}} \to \mathcal{M}^\ddagger \cup \{!\})$. To represent this random selection, let $M^\ddagger$ be the random variable representing the chosen value of $\mathcal{M}^\ddagger$ in the code construction of 33. The random code construction will be useful in calculating bounds for both the average arithmetic error and the maximum probability of false authentication. In particular, we will show that with probability $1-e^{-n}$ the randomly chosen value of $\mathcal{M}^\ddagger$ yields a code with the stated average arithmetic error bound, and with probability $1 - e^{-n/2}$ yields a code with the stated maximum probability of false authentication bound. Note then the probability of selecting a code which satisfies both bound simultaneously must be at least $1 - e^{-n} - e^{-n/2}$ due to the union bound.

Once again for readability, we have separated by a dividing line the bound for each for the operational measures.

(Rate)

For the rate, first assume that $r_{\mathcal{J}} \geq \frac{\log 2\ n}{n}$. In this case

$$\begin{aligned} r_{\mathcal{J}} &= n^{-1} \log \left\lfloor \exp\left(nr^\ddagger\right) \right\rfloor \\ &\geq n^{-1} \log \left(\exp\left(nr^\ddagger\right) - 1\right) && (84) \\ &= r^\ddagger + n^{-1} \log\left(1 - \exp(-nr^\ddagger)\right) && (85) \\ &\geq r^\ddagger + n^{-1} \log\left(1 - \exp(-nr_{\mathcal{J}})\right) && (86) \\ &\geq r^\ddagger - n^{-1} \log(2n) && (87) \end{aligned}$$

where the last line is from the assumption. Plugging in the definition of $r^\ddagger$ yields

$$r_{\mathcal{J}} \geq r_{\mathcal{H}} - \frac{(1-\gamma)\ell}{4n}\lambda^2 - \frac{r_{\mathcal{H}} + 2 + \log 4n\theta}{n} \tag{88}$$

where

$$\lambda = \max\left(0, \min_{k \in \mathcal{K}} 1 - \frac{(1+\delta)(k^2\rho_\Delta + \rho_{\text{Dec}})}{\gamma\tau^\star(k) + (1-\gamma)\tau^\star(d_k)}\right)$$

$$\theta = \max\left(1, \sqrt{3n\left[\omega_{\mathcal{K}} + (\rho_\Delta + \rho_{\text{Dec}})\left(1 + \delta + 2\lambda^2 + 2\ r_{\mathcal{H}}\right)\right]}\right).$$

What remains is to prove the assumption, to that end observe

$$\begin{aligned} &\lfloor \exp(nr^\ddagger) \rfloor \\ &= \left\lfloor \exp\left((n-1)r_{\mathcal{H}} - \frac{(1-\gamma)\ell}{4}\lambda^2 - 2 - \log 2\theta\right)\right\rfloor && (89) \\ &\geq \lfloor \exp(\log 2n) \rfloor = 2n && (90) \end{aligned}$$

since

$$(n-1)r_{\mathcal{H}} \geq \frac{(1-\gamma)\ell}{4}\lambda^2 + 2 + \log 4n\theta.$$

(Power)

For the power constraint,

$$\begin{aligned} \omega_{\mathcal{J}} &\leq \omega_{\mathcal{L}} && (91) \\ &\leq \omega_{\mathcal{H}} + 2\sqrt{2\omega_{\mathcal{H}}\rho_\Delta(r_{\mathcal{H}}+1)} \\ &\quad + \rho_\Delta\left(1 + (8|\mathcal{K}|+1)\left[r_{\mathcal{H}} + 1 + \frac{\log|\mathcal{K}|}{n}\right]\right) && (92) \end{aligned}$$

since $\boldsymbol{X}^\ddagger(m) = \boldsymbol{X}'(m)$ whenever $m \in \{M^\ddagger\}$.

(Average arithmetic probability of error)

Next, for the average arithmetic error, let

$$a(m) = \Pr\left(\hat{m}'(\boldsymbol{X}'(m) + \boldsymbol{G}_{\text{Dec}}) \neq m\right)$$

so that the average arithmetic probability of error for $M^\ddagger = \mathcal{M}^\ddagger$ can be written

$$e^{-nr^\ddagger} \sum_{m \in \mathcal{M}} \mathbb{1}_{\mathcal{M}^\ddagger}(m)\, a(m).$$

From Lemma 42 though

$$\begin{aligned} &\Pr\left(e^{-nr^\ddagger}\sum_{m\in\mathcal{M}} \mathbb{1}_{\{M^\ddagger\}}(m) a(m) \geq \varepsilon_{\mathcal{K}}(\rho_{\text{Dec}}) + \sqrt{\frac{n}{2}e^{-nr^\ddagger}}\right) \\ &\leq e^{-n} && (93) \end{aligned}$$

since $e^{-nr}\sum_{m\in\mathcal{M}} a(m) = \varepsilon_{\mathcal{K}}(\rho_{\text{Dec}})$. Thus with probability greater than $1 - e^{-n}$ the chosen $\mathcal{M}^\ddagger$ will yield

$$\begin{aligned} \varepsilon_{\mathcal{J}}(\rho_{\text{Dec}}) &\leq \varepsilon_{\mathcal{L}}(\rho_{\text{Dec}}) + \sqrt{\frac{n}{2}e^{-nr^\ddagger}} && (94) \\ &\leq \varepsilon_{\mathcal{H}}(\rho_{\text{Dec}} + \rho_\Delta) + \sqrt{2ne^{-nr^\ddagger}} + |\mathcal{K}|e^{-\frac{1}{8}\ell\delta^2}. && (95) \end{aligned}$$

(Maximum probability of false authentication)

Finally for the probability of false authentication recall from the proof of Theorem 32 that for each $M^\ddagger = m$, $\boldsymbol{V} = v$, and attack $\boldsymbol{Z}(\boldsymbol{V}, m) = z$ there exists a $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$ such that

$$\boldsymbol{Y}|\{M^\ddagger, \boldsymbol{V}, \boldsymbol{Z} = m, v, z\} = \boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}.$$

Therefore, by letting

$$b_{m,\boldsymbol{\mu}}(c) = \Pr\left(\hat{m}'(\boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}) = c\right)$$

the maximum probability of false authentication of code $\mathcal{J}$ can be expressed as

$$\max_m \sup_{\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}} \sum_{c \in \mathcal{M}\setminus\{m\}} \mathbb{1}_{\mathcal{M}^\ddagger}(c)\, b_{m,\boldsymbol{\mu}}(c) \tag{96}$$

since $\hat{m}^\ddagger(y) = \hat{m}'(y)$ when $\hat{m}'(y) \in \mathcal{M}^\ddagger$. An upper bound on $\alpha_{\mathcal{J}}$ can therefore be obtained by computing an upper bound on $\sum_{c\in\mathcal{M}-\{m\}} \mathbb{1}_{\mathcal{M}^\ddagger}(c)\, b_{m,\boldsymbol{\mu}}(c)$ that holds simultaneously for

all $m \in \mathcal{M}$ and $\boldsymbol{\mu} \in \boldsymbol{\mathcal{R}}$. To that end we will employ a divide and conquer approach based on if

$$\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\dagger} \quad \text{where} \quad \mathcal{U}^{\dagger} = (-\theta, \theta).$$

To begin, for $\boldsymbol{\mu} \notin \mathcal{U}^{\dagger}$ there must be a coordinate of $i \in \{1, \ldots, n\}$ such that $|\mu_i| \geq \theta$. But

$$G_{\boldsymbol{\tau}(m),i} \lessgtr x_i(c) + t_i(c) - \mu_i \pm \sqrt{\ell(1+\delta)(f_i^2(c)\rho_{\Delta} + \rho_{\text{Dec}})} \tag{97}$$

is required for $\hat{m}'(\boldsymbol{G}_{\boldsymbol{\tau}(m)} + \boldsymbol{\mu}) = c$ thanks to Code Modification 21. Clearly the event in (97) is the probability that a Gaussian random variable lies in a particular interval. If $\mu_i \geq \theta$, then the key inequality is

$$G_{\boldsymbol{\tau}(m),i} \leq x_i(c) + t_i(c) + \sqrt{\ell(1+\delta)(f_i^2(c)\rho_{\Delta} + \rho_{\text{Dec}})} - \mu_i,$$

while for $\mu_i < -\theta$ it is

$$G_{\boldsymbol{\tau}(m),i} \geq x_i(c) + t_i(c) - \sqrt{\ell(1+\delta)(f_i^2(c)\rho_{\Delta} + \rho_{\text{Dec}})} - \mu_i.$$

Indeed, if $\mu_i \geq \theta$ then it follows that

$$\begin{aligned}
&x_i(c) + t_i(c) + \sqrt{\ell(1+\delta)(f_i^2(c)\rho_{\Delta} + \rho_{\text{Dec}})} - \mu_i \\
&\quad \leq \sqrt{n\omega_{\mathcal{K}}} + \sqrt{\ell(1+\delta)(\rho_{\Delta} + \rho_{\text{Dec}})} \\
&\qquad - \sqrt{3n\left[\omega_{\mathcal{K}} + (\rho_{\Delta} + \rho_{\text{Dec}})\left(1 + \delta + 2\lambda^2 + 2\, r_{\mathcal{H}}\right)\right]} && (98) \\
&\quad \leq \sqrt{n\omega_{\mathcal{K}}} + \sqrt{\ell(1+\delta)(\rho_{\Delta} + \rho_{\text{Dec}})} \\
&\qquad - \sqrt{n\omega_{\mathcal{K}}} - \sqrt{n(1+\delta)(\rho_{\Delta} + \rho_{\text{Dec}})} \\
&\qquad - \sqrt{2n(\rho_{\Delta} + \rho_{\text{Dec}})\left[\lambda^2 + r_{\mathcal{H}}\right]} && (99) \\
&\quad \leq -\sqrt{2n(\rho_{\Delta} + \rho_{\text{Dec}})\left[\lambda^2 + r_{\mathcal{H}}\right]} && (100)
\end{aligned}$$

where (98) is plugging in the maximum values of $x_i(c) + t_i(c)$ (itself due to the power constraint), μ_i, and $f_i(c)$; (99) is because the concavity of the square root implies $\sqrt{3(a+b+c)} \geq \sqrt{a} + \sqrt{b} + \sqrt{c}$; (100) is because $n \geq \ell$. Hence when $\mu \geq \theta$ we can use the key inequality to bound the probability of (97) as follows

$$\begin{aligned}
&\Phi\left(\frac{x_i(c) + t_i(c) + \sqrt{\ell(1+\delta)(f_i^2(c)\rho_{\Delta} + \rho_{\text{Dec}})} - \mu_i}{\sqrt{\tau_i(m)}}\right) \\
&\quad \leq \Phi\left(-\sqrt{\frac{(2nr_{\mathcal{H}} + 2n\lambda^2)(\rho_{\Delta} + \rho_{\text{Dec}})}{\tau_i(m)}}\right) && (101) \\
&\quad \leq \Phi\left(-\sqrt{2\, nr_{\mathcal{H}} + 2n\lambda^2}\right) && (102) \\
&\quad \leq \sqrt{\frac{2}{\pi}} e^{-nr_{\mathcal{H}} - n\lambda^2} && (103)
\end{aligned}$$

where (101) is because of (100) and $\Phi(a) > \Phi(a')$ if and only if $a > a'$; (102) is because $0 \leq f_i^2(m) \leq 1$ for all i and m implies

$$\frac{\rho_{\Delta} + \rho_{\text{Dec}}}{\tau_i(m)} = \frac{\rho_{\Delta} + \rho_{\text{Dec}}}{\frac{f_i^2(m)\rho_{\Delta}\rho_{\text{Adv}}}{f_i^2(m)\rho_{\Delta} + \rho_{\text{Adv}}} + \rho_{\text{Dec}}} \geq 1;$$

finally (103) follows from $\Phi(a) \leq \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}a^2}$ for $a \leq 0$. A similar derivation follows for the $\mu_i \leq -\theta$ case. Thus Equation (103) proves

$$b_{m,\boldsymbol{\mu}} \leq \sqrt{\frac{2}{\pi}} e^{-nr_{\mathcal{H}} - n\lambda^2} \tag{104}$$

for all $m \in \mathcal{M}^{\dagger}$ and $\boldsymbol{\mu} \notin \boldsymbol{\mathcal{U}}^{\dagger}$, and hence

$$\sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\mathcal{M}^{\ddagger}}(c)\, b_{m,\boldsymbol{\mu}}(c) \leq \sqrt{\frac{2}{\pi}} e^{-n\lambda^2} \tag{105}$$

for all $\boldsymbol{\mu} \notin \boldsymbol{\mathcal{U}}^{\dagger}$ regardless of the choice of $\mathcal{M}^{\ddagger}$.

We now move on to the case that $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\dagger}$. Let $\boldsymbol{\mathcal{U}}^{\ddagger}$ be the set guaranteed by Lemma 41 with respect to $\boldsymbol{\mathcal{U}}^{\dagger}$ and positive constant $2e^{-\frac{1-\gamma}{8}\ell\lambda^2}$. It will be important for later to note that

$$|\boldsymbol{\mathcal{U}}^{\ddagger}| = e^{n\log(2\theta) + n\frac{1-\gamma}{8}\ell\lambda^2}. \tag{106}$$

From here, our strategy is to show that with high probability

$$\sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\{M^{\ddagger}\}}(c)\, b_{m,\boldsymbol{\mu}}(c) \leq 2\, ne^{-\frac{1}{8}\ell(1-\gamma)\lambda^2} \tag{107}$$

for all $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\ddagger}$ and $m \in \mathcal{M}$. With this result in hand

$$\sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\{M^{\ddagger}\}}(c)\, b_{m,\boldsymbol{\mu}}(c) \leq \left(2\, n + \frac{1}{2\sqrt{n\rho_{\text{Dec}}}}\right) e^{-\frac{1}{8}\ell(1-\gamma)\lambda^2} \tag{108}$$

for all $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\dagger}$ and $m \in \mathcal{M}$ follows by Lemma 41. To prove (107) first note that for any given $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\ddagger}$ and $m \in \mathcal{M}$ we have $e^{nr_{\mathcal{J}}} \geq 2\, n$ by Equation (90),

$$b_{m,\boldsymbol{\mu}}(c) \leq \alpha_{\mathcal{L}}^{*}(\rho_{\text{Dec}}, \rho_{\text{Adv}}) \leq 2e^{-\frac{1}{8}\ell(1-\gamma)\lambda^2}$$

by the assumptions that code $\mathcal{L}$ satisfies the operational bounds set forth in Theorem 32, and

$$\sum_{c \in \mathcal{M}} \frac{1}{|\mathcal{M}|} e^{-nr_{\mathcal{J}}} b_{m,\boldsymbol{\mu}}(c) \leq e^{-nr_{\mathcal{J}}}$$

because $\sum_{m \in \mathcal{M}} b_{m,\boldsymbol{\mu}} \leq 1$. Hence we also have

$$\begin{aligned}
&\Pr\left(\sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\{M^{\ddagger}\}}(c)\, b_{m,\boldsymbol{\mu}}(c) \geq 2\, ne^{-\frac{1}{8}\ell(1-\gamma)\lambda^2}\right) \\
&\quad \leq \exp\left(-n^2\left[r_{\mathcal{H}} - r_{\mathcal{J}} - \frac{(1-\gamma)\ell}{8n}\lambda^2 - \frac{3}{2n}\right]\right) && (109)
\end{aligned}$$

by Corollary 43 and simple algebra. Using the union bound (recalling (106)) and that

$$r_{\mathcal{J}} \leq (1 - n^{-1})r_{\mathcal{H}} - \frac{(1-\gamma)\ell}{4n}\lambda^2 - \frac{2 + \log 2\theta}{n}$$

yields

$$\Pr\left(\max_{\substack{\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\ddagger} \\ m \in \mathcal{M}}} \sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\{M^{\ddagger}\}}(c)\, b_{m,\boldsymbol{\mu}}(c) \geq 2\, ne^{-\frac{1-\gamma}{8}\ell\lambda^2}\right) \leq e^{-\frac{n}{2}}. \tag{110}$$

So, as discussed prior, Equation (110) shows that Equation (107) is true for all $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\ddagger}$ and $m \in \mathcal{M}$ with exponentially high probability, hence (108) is true for all $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\dagger}$ and $m \in \mathcal{M}$ with exponentially high probability as a consequence of Lemma 41.

As a final step in the proof, we note that the upper bound on the probability of false authentication for $\boldsymbol{\mu} \in \boldsymbol{\mathcal{U}}^{\dagger}$, Equation (108), is greater than the upper bound on the probability

of false authentication given $\mu \notin \mathcal{U}^{\ddagger}$, Equation (105). Hence combining the two, we have

$$\alpha_{\mathcal{J}}(\rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) = \sup_{\substack{\mu \in \mathcal{R} \\ m \in \mathcal{M}}} \sum_{c \in \mathcal{M} \setminus \{m\}} \mathbb{1}_{\mathcal{M}^{\ddagger}}(c)\, b_{m,\mu}(c) \quad (111)$$

$$\leq \left(2n + \frac{1}{2\sqrt{n\rho_{\mathrm{Dec}}}}\right) e^{-\frac{1-\gamma}{8}\ell\lambda^2}, \quad (112)$$

finishing the proof. ∎

Appendix E
Theorem 31

Proof: First let

$$\rho_{\Delta} = o(1) \quad \text{and} \quad \delta = o(\rho_{\Delta}) \quad (113)$$

be sufficiently large and note by the channel capacity theorem of Shannon [15], for all positive real finite numbers a and b there exists a sequence of codes $\mathcal{H}_{(n)} = (\boldsymbol{x}_{(n)} : \mathcal{M}_{(n)} \to \mathcal{R}_{(n)}, \hat{m}_{(n)} : \mathcal{R}_{(n)} \to \mathcal{M}_{(n)})$ such that

$$\lim_{n\to\infty} r_{\mathcal{H}_{(n)}} = \frac{1}{2}\log\left(1 + \frac{a}{b}\right)$$
$$\lim_{n\to\infty} \omega_{\mathcal{H}_{(n)}} = a$$
$$\varepsilon_{\mathcal{H}_{(n)}}(b) = 0.$$

Letting $a = \rho - O(\sqrt{\rho_{\Delta}})$ and $b = \rho_{\Delta} + \rho_{\mathrm{Dec}}$ and applying Theorem 34 yields a sequence of codes $\mathcal{J}_n$ such that

$$\lim_{n\to\infty} \omega_{\mathcal{J}_{(n)}} \leq \lim_{n\to\infty} \omega_{\mathcal{H}_{(n)}} + O(\sqrt{\rho_{\Delta}}) = \rho \quad (114)$$

$$\lim_{n\to\infty} \varepsilon_{\mathcal{J}_{(m)}}(\rho_{\mathrm{Dec}}) \leq \lim_{n\to\infty} \varepsilon_{\mathcal{H}}(\rho_{\mathrm{Dec}} + \rho_{\Delta}) + e^{-O(n\delta^2)} \quad (115)$$

$$= 0 \quad (116)$$

$$\lim_{n\to\infty} \alpha_{\mathcal{J}_{(n)}}(\rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) \leq e^{-O(n\rho_{\Delta}^2)}. = 0, \quad (117)$$

while

$$\lim_{n\to\infty} r_{\mathcal{J}_{(n)}}$$
$$\geq \lim_{n\to\infty} r_{\mathcal{H}_{(n)}} - O\left(\rho_{\Delta}^2 + \frac{\log n}{n}\right) \quad (118)$$
$$= \lim_{n\to\infty} \frac{1}{2}\log\left(1 + \frac{\rho - O(\sqrt{\rho_{\Delta}})}{\rho_{\Delta} + \rho_{\mathrm{Dec}}}\right) - O\left(\rho_{\Delta}^2 + \frac{\log n}{n}\right) \quad (119)$$
$$= \frac{1}{2}\log\left(1 + \frac{\rho}{\rho_{\mathrm{Dec}}}\right) \quad (120)$$

This proves $\frac{1}{2}\log\left(1 + \frac{\rho}{\rho_{\mathrm{Dec}}}\right)$ is achievable, and by [15] it is also an upper bound, hence

$$c(\rho, \rho_{\mathrm{Dec}}, \rho_{\mathrm{Adv}}) = \frac{1}{2}\log\left(1 + \frac{\rho}{\rho_{\mathrm{Dec}}}\right) \quad (121)$$

for $\rho_{\mathrm{Adv}} > 0$.

On the other hand if $\rho_{\mathrm{Adv}} = 0$, then $\boldsymbol{V} = \boldsymbol{X}(M)$ for any code $\boldsymbol{X}$, $\hat{m}$. Hence the adversary may choose $\boldsymbol{Z}(\boldsymbol{V}, M) = \boldsymbol{X}(m') - \boldsymbol{V}$, to produce

$$\boldsymbol{Y} = \boldsymbol{X}(m') + \boldsymbol{G}_{\mathrm{Dec}}.$$

Thus

$$\alpha_{\boldsymbol{X},\hat{m}}(\rho_{\mathrm{Dec}}, 0) \geq 1 - \varepsilon_{\boldsymbol{X},\hat{m}}(\rho_{\mathrm{Dec}}, 0),$$

and consequently

$$c(\rho, \rho_{\mathrm{Dec}}, 0) = 0. \quad (122)$$

∎

Acknowledgment

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Combat Capabilities Development Command Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes not withstanding any copyright notation here on.

References

[1] J. B. Perazzone, E. Graves, P. Yu, and R. Blum, "Secret key-enabled authenticated-capacity region, Part I: Average authentication," *IEEE Trans. Inf. Theory*, early access, Jun. 23, 2022, doi: 10.1109/TIT.2022.3185933.
[2] E. Graves, J. B. Perazzone, P. Yu, and R. Blum, "Secret key-enabled authenticated-capacity region, Part II: Typical-authentication," *IEEE Trans. Inf. Theory*, early access, Jun. 27, 2022, doi: 10.1109/TIT.2022.3186443.
[3] S. Jiang, "Keyless authentication in a noisy model," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 1024–1033, Jun. 2014.
[4] S. Jiang, "On the optimality of keyless authentication in a noisy model," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1250–1261, Jun. 2015.
[5] E. Graves, P. Yu, and P. Spasojevic, "Keyless authentication in the presence of a simultaneously transmitting adversary," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2016, pp. 201–205.
[6] O. Gungor and C. E. Koksal, "On the basic limits of RF-fingerprint-based authentication," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4523–4543, Aug. 2016.
[7] W. Tu and L. Lai, "Keyless authentication and authenticated capacity," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3696–3714, May 2018.
[8] O. Kosut and J. Kliewer, "Authentication capacity of adversarial channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2018, pp. 1–5.
[9] A. Beemer, E. Graves, J. Kliewer, O. Kosut, and P. Yu, "Authentication with mildly myopic adversaries," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 984–989.
[10] N. Sangwan, M. Bakshi, B. Kumar Dey, and V. M. Prabhakaran, "Multiple access channels with Byzantine users," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2019, pp. 1–5.
[11] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography*. Boston, MA, USA: Springer, 1994, pp. 271–285.
[12] A. Beemer, O. Kosut, J. Kliewer, E. Graves, and P. Yu, "Structured coding for authentication in the presence of a malicious adversary," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 617–621.
[13] R. Ahlswede and G. Dueck, "Every bad code has a good subcode: A local converse to the coding theorem," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 34, no. 2, pp. 179–182, 1976.
[14] G. J. Simmons, "Authentication theory/coding theory," in *Advances in Cryptology–(CRYPTO)*, Santa Barbara, CA, USA: Springer, Aug. 1984, pp. 411–431.
[15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[16] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.
[17] E. Graves and T. F. Wong, "Inducing information stability to obtain information theoretic necessary requirements," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 835–864, Feb. 2020.
[18] J. L. Massey, "Deep-space communications and coding: A marriage made in heaven," in *Advanced Methods for Satellite and Deep Space Communications*. Berlin, Germany: Springer, 1992, pp. 1–17.
[19] H. Robbins, "A remark on Stirling's formula," *Amer. Math. Monthly*, vol. 62, no. 1, pp. 26–29, 1955.
[20] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

Eric Graves (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Florida in 2008, 2011, and 2013, respectively. He is currently a Civilian Research Scientist at the U.S. Army Research Laboratory, Adelphi, MD, USA. His research interests include information theory and security.

Allison Beemer received the B.A. degree in mathematics from the Whitman College in 2012, and the M.S. and Ph.D. degrees in mathematics from the University of Nebraska–Lincoln in 2015 and 2018, respectively. From 2018 to 2020, she held a Post-Doctoral Research Associate position jointly at Arizona State University and the New Jersey Institute of Technology, funded by the U.S. Army Research Laboratory. In 2020, she became an Assistant Professor with the Department of Mathematics, University of Wisconsin–Eau Claire. Her research interests include secure and authentic communication, information privacy, graph-based codes and decoding algorithms, and applied discrete mathematics.

Jörg Kliewer (Senior Member, IEEE) received the Dr.Ing. (Ph.D.) degree in electrical engineering from the University of Kiel, Germany, in 1999. From 1993 to 1998, he was a Research Assistant at the University of Kiel, where he was a Senior Researcher and a Lecturer from 1999 to 2004. In 2004, he visited the University of Southampton, U.K., for one year. From 2005 to 2007, he was with the University of Notre Dame, Notre Dame, IN, USA, as a Visiting Assistant Professor. From 2007 to 2013, he was with New Mexico State University, Las Cruces, NM, USA, most recently as an Associate Professor. He is currently a Professor with the New Jersey Institute of Technology, Newark, NJ, USA. His research interests include information theory, machine learning, graphical models, and statistical algorithms with applications to secure and private communication and data storage. He has been a member of the Editorial Board of the IEEE Information Theory Society Newsletter since 2012. He was a recipient of the Leverhulme Trust Award, the German Research Foundation Fellowship Award, the IEEE GLOBECOM Best Paper Award, and the Fulbright Scholarship. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2014 and an Area Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2015 to 2021. He has also been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2017 to 2020. Since 2021 he has been serving as an Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

Oliver Kosut (Member, IEEE) received the B.S. degree in electrical engineering and mathematics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2004, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2010. Since 2012, he has been a Faculty Member with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA, where he is currently an Associate Professor. Previously, he was a Post-Doctoral Research Associate with the Laboratory for Information and Decision Systems, MIT, from 2010 to 2012. His research interests include information theory, particularly with applications to security and machine learning and power systems. He received the NSF CAREER Award in 2015.

Paul L. Yu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at College Park, College Park, MD, USA. He is currently an Electronics Engineer with the U.S. Army Research Laboratory (ARL). His research interests include signal processing and security for wireless tactical networking. He has several patents in these areas.