On Message Authentication Channel Capacity Over a Wiretap Channel

Dajiang Chen, *Member, IEEE*, Shaoquan Jiang, Ning Zhang, *Senior Member, IEEE*, Lei Liu, *Member, IEEE*, and Kim-Kwang Raymond Choo, *Senior Member, IEEE*

***Abstract*— In this paper, a novel message authentication model using the same key over wiretap channel is proposed to achieve *information-theoretic security*. Specifically, in the proposed model, there is a discrete memoryless channel $W_1 : \mathcal{X} \to \mathcal{Y}$ between transmitter Alice and receiver Bob, while an attacker Oscar is connected with Alice via discrete memoryless channel $W_2 : \mathcal{X} \to \mathcal{Z}$. Alice encodes message M to codeword (S, X^n), using an encoding function with secret key K. Then, S is sent to Bob over a one-way noiseless channel (fully controlled by Oscar), and X^n is sent over the wiretap channel, say $X \to (Y, Z)$. Building on this model, a new message authentication scheme is proposed. The scheme incorporates a secure channel coding, which uses random coding techniques to detect man-in-the-middle (MITM) attacks. The authentication channel capacity is studied in a specific channel model when W_2 is not less noisy than W_1. We theoretically demonstrate that the authentication channel capacity is much larger than the secrecy capacity, since Bob does not need to recover information transmitted over the noisy channel.**

***Index Terms*— Physical layer security, multiple message authentication, wiretap channel.**

I. Introduction

MESSAGE authentication solutions approach users to validate that the message is truly from the claimed source (i.e., authenticity), and has not been altered during transmission (i.e., integrity). Generally, message authentication assumes that the channel between sender and receiver is noiseless, and this is also commonly referred to as the Simmons authentication model [1]. While message authentication has been widely studied, designing approaches that also guarantee information-theoretic security (ITS) remains challenging, particularly when we need to guarantee secrecy as attackers' computing capabilities scale up, for example due to technological advances [2], [3], [4], [5]. However, as noted in [6], Simmons model can cause an entropy loss of the authentication key when the same key is used to authenticate multiple messages. It has also been observed that after ℓ authentications in the noiseless channel model, an adversary can succeed with a probability higher than $2^{-H(K)/(\ell+1)}$ and this probability quickly approaches 1 as ℓ increases [6]. Therefore, to authenticate multiple messages with ITS, one needs to renew the authentication key frequently. However, because of the characteristics inherent of wireless networks (e.g., changing topology, limited battery supply and computation capability), it can be challenging to renew key frequently in wireless networks [7], [8], [9].

In this paper, we focus on solutions that can support multi-message authentication using the same authentication key to achieve ITS over other channel models. It is well known that the rate of leaked information over a wiretap channel asymptotically approaches zero as the length of the transmitted information increases, when the main channel is less noisy than the wiretapper's channel [12]. Later, Lai *et al.* [13] proposed an authentication scheme over channel $X \to (Y, Z)$ to realize multiple messages authentication with ITS when $I(X;Y) > I(X;Z)$.

Seeking to contribute to the literature, a novel message authentication model is presented over a wiretap channel. The model is as follows. A legitimate transmitter, say Alice, intends to send multiple messages $(M_1, M_2, \ldots, M_J)$ to a receiver Bob in the presence of an attacker Oscar. In addition, Alice will authenticate these messages with same authentication key K. As shown in Fig. 1(a), the channel model used in this paper is as follows. A discrete memoryless channel (DMC) $W_1 : \mathcal{X} \to \mathcal{Y}$ exists from Alice to Bob and *a one-way noiseless channel*[1] exists from Alice to Bob through Oscar. In addition, a DMC $W_2 : \mathcal{X} \to \mathcal{Z}$ is from Alice to Oscar, and a noiseless channel[2] is from Alice through Oscar to Bob. Here, (W_1, W_2) is a wiretap channel, and the outputs are determined

Manuscript received 13 May 2022; revised 11 July 2022; accepted 12 August 2022. Date of publication 24 August 2022; date of current version 7 September 2022. This work was supported in part by the NSFC under Grant 61872059 and Grant 61502085 and in part by the Project "The Verification Platform of Multi-Tier Coverage Communication Network for Oceans" under Grant LZC0020. The work of Kim-Kwang Raymond Choo was supported by the Cloud Technology Endowed Professorship. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dusit Niyato. *(Corresponding author: Ning Zhang.)*

Dajiang Chen is with the Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: djchen@uestc.edu.cn).

Shaoquan Jiang is with the Department of Computer Science, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: shaoquan.jiang@gmail.com).

Ning Zhang is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: ning.zhang@uwindsor.ca).

Lei Liu is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China, and also with the Xidian Guangzhou Institute of Technology, Guangzhou 510555, China (e-mail: tianjiaoliulei@163.com).

Kim-Kwang Raymond Choo is with the Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: raymond.choo@fulbrightmail.org).

Digital Object Identifier 10.1109/TIFS.2022.3201386

[1]In wireless communications, both noiseless and noisy channels between Alice and Bob are the same wireless medium, and the only difference is that the message transmitted over the former is with an error correcting code.

[2]Due to the fact that any noisy channel can be simulated with this noiseless channel by randomizing the transmitted signal, the assumption does not incur any loss of generality, and it even gives Oscar an advantage actually.

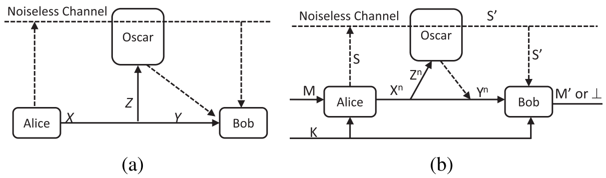

Fig. 1. (a) The communication model; (b) The proposed authentication model.

by the input X and the conditional probability $P_{YZ|X}$.[3] As shown in Fig. 1 (b), the proposed model includes an *encoder* $F: \mathcal{M}\times K \to \mathcal{S}\times X^n$ as well as a *decoder* $G: \mathcal{K}\times\mathcal{S}\times\mathcal{Y}^n\to\mathcal{M}\cup\{\perp\}$. If Alice plans to send a message M to Bob and authenticate it, she first calculates $(S, X^n) = F(M, K)$, and then transmits S over noiseless channel and X^n over (W_1, W_2). Bob can receive (S', Y^n), in which, n is the codeword length W_1 and S' might be different from S because of the potential attacks from Oscar. Based on (S', Y^n), Bob decides to accept(or reject) the authentication if $G(S', Y^n) \in \mathcal{M}$ (or $= \perp$).

The adversary model in this paper is shown as follows. Oscar can observe message authentications in polynomial number (in n), and launch a polynomial number (in n) of man-in-the-middle (MITM) attacks with infinite computing resources. We consider two types of adversary attacks i.e., Type-I attack: Oscar can modify S to arbitrary S' when Alice authenticates M to Bob with codeword (S, X^n); and Type-II attack: Oscar can adaptively transmit $\hat{S}$ and $\hat{Y}^n$ to Bob noiselessly, even if Alice does not authenticate a message.

The main goal of this paper is to ensure that the success probability of adversarial attacks is *negligible* [4] under the adversary model above. An authentication scheme is *secure* if the authentication properties *completeness*, *authentication*, and *key security* are satisfied (see Sec. IV). Completeness states that Bob would accept M with negligible probability of error when Oscar is not present. Authentication means that Oscar has a negligible probability of success in the above two types of attacks in polynomial times (in n). Key security states that the vanishing key leakage can be achieved after polynomially bounded (in n) attacks. A performance metric of a secure authentication scheme is defined as *authentication rate* $\rho_{\textbf{auth}} = \frac{1}{n} log|\mathcal{M}|$, in which, $|\cdot|$ denotes the cardinality of a set, and $log|\mathcal{M}|$ means the bit length of $M \in \mathcal{M}$.

We also focus on minimizing the wiretap channel usage while keeping Oscar's successful attack probability arbitrarily small under a polynomial (in n) number of attacks. In particular, a specific authentication model (F, G) with $F(M, K) = (M, Enc(Hash(M, K)))$ is studied, in which, $Hash$ is a hash function from $\mathcal{M} \times \mathcal{K}$ to $\mathcal{T}$, and Enc is an encoder from $\mathcal{T}$ to X^n. In this case, $\rho_{\textbf{auth}} = \rho_{\textbf{tag}} \cdot \rho_{\textbf{chan}}$, where $\rho_{\textbf{tag}} = \frac{log|\mathcal{M}|}{log|\mathcal{T}|}$ and $\rho_{\textbf{chan}} = \frac{log|\mathcal{T}|}{n}$. $\rho_{\textbf{chan}}$ is called authentication channel coding rate. To achieve a higher authentication rate, one need to improve both the tag rate and the authentication channel rate. As tag rate $\rho_{\textbf{tag}}$ is determined by traditional cryptographic techniques (which is out of the scope of this work), we mainly focus on authentication channel coding rate $\rho_{\textbf{chan}}$. The largest $\rho_{\textbf{chan}}$ is called authentication channel capacity.

The authentication channel capacity is obtained over the specific authentication model. Concretely, it is proved that if W_2 is *not less noisy than* [5] W_1, then the authentication channel capacity of wiretap channel (W_1, W_2) is $\max_{U\to X\to YZ} H(U|Z)$. Specifically,we present an authentication channel capacity achievable scheme as follows. Alice first computes a tag T ($\in \mathcal{T}$) from M and one part of authentication key K_0 by using an *ϵ-almost strongly-universal hash functions* (see Sec. III and Corollary 1). And then, she encodes T to X^n as follows: X^n is chosen form $\mathcal{C}_{K_1 T}$ uniformly at random where K_1 is the part of authentication key, and $\{\mathcal{C}_{ij}\}_{ij}$ are generated by leveraging a novel coding scheme (as shown in Th. 1). Finally, she transmits the codeword (M, X^n) to Bob, where M is for the noiseless channel, and X^n is for (W_1, W_2). After receiving (M', Y^n), Bob checks the consistency of M' and Y^n for message authentication. Compared with existing schemes, our scheme can achieve higher authentication rate. The reason is that the existing schemes require to recover T, while it is unnecessary in the proposed scheme. Actually, Bob can obtain tag T from $M' = M$ and K in the case of no attack; and when Oscar launches an attack, Bob can detect if M' and Y^n are inconsistent, and then rejects M' (if inconsistent). The main challenge is how to detect the inconsistent without fully recovering T. We address this issue by designing a novel channel coding based on random coding techniques (Sec. V).

II. Related Works

Wyner [12] first studies the secure channel coding over a wiretap channel, which is generalized by Csiszár and Körner [14]. After that, secure communication over a wiretap channel has been extensively studied ([9], [15], [16], [17]). In particular, secret key agreement over noisy channels was studied in [18], and secret transmission over fading channels was discussed in [20], [21], [22], and [19]. Wiretap codes achieving ITS were proposed in [23], [24], and [25]. Moreover, in [26], a Physical-Layer-based transmitter identification method is proposed by utilizing multiple channel-based features. However, not much attention was devoted to *message authentication*. In [10], the authentication problem was investigated in a noiseless channel with a noisy initialization (or simply in the source model [29]), in which, a trusted center is used to broadcast a random string to legitimate users over noisy channels such that the correlated data can be obtained by the transmitter, the receiver and the adversary. Later, keyless authentication problem over a MIMO fading wiretap channel was studied in [33] and [34], in which, an authenticated channel from Alice to Bob is assumed to exist for sending the preliminary data.

In [13], multiple messages authentication problem over a wiretap channel $X \to (Y, Z)$ is studied. It is assumed that Alice and Bob pre-share a secret key K and there is a noiseless channel between the adversary and Bob. In [30] and [31], the multiple message authentication protocols for special wiretap channel model were proposed by using a secure polar code and secure LDPC code, respectively. In addition, the authentication problem over noisy channel model without using a preshared

[3] In this paper, we do not specify the joint conditional distribution of YZ given X, but only the marginal distributions of $P_{YZ|X}$ ever enter our consideration.

[4] A function $negl(n) : \mathbb{N} \to \mathbb{R}$ is *negligible*, if for every positive integer c there exists an integer N_c such that for all $n > N_c$, $|negl(n)| < n^{-c}$.

[5] W_2 is not less noisy than W_1 if there exists a Markov chain $U \to X \to YZ$ such that $I(U; Y) > I(U; Z)$.

key was discussed in [37], [38], and [39]. The *keyless authentication capacity* of binary symmetric wiretap channel was discussed in [37], while the keyless authentication capacity of a non-interactive authentication over general wiretap channel was presented in [38]. The *keyless authentication channel capacity* was obtained in [39], which is equal to channel capacity when the simulatability condition is not satisfied; and will be zero, otherwise. In [40], a physical payer authentication method was proposed for non-coherent massive SIMO-enabled wireless communications in industrial IoT.

III. Notations and Preliminaries

For any positive integer s, $[s] = \{1, 2, \ldots, s\}$. *Distance* between random variables (RVs) U and U' is $\mathsf{D}(U; U') = \sum_u |P_U(u) - P_{U'}(u)|$. *Conditional distance between* U *given* V and U' is

$$\mathsf{D}(U|V; U') = \sum_{v,u} P_V(v)|P_{U|V}(u|V) - P_{U'}(u)|. \quad (1)$$

Let $\mathcal{H}$ be a finite hashing family from $\mathcal{M}$ to $\mathcal{T}$. $\mathcal{H}$ is ϵ*-almost strongly-universal* (ϵ-ASU) if (1) $|\{h \in \mathcal{H} : h(x) = t\}| = \frac{|\mathcal{H}|}{|\mathcal{T}|}$, $\forall x \in \mathcal{M}, \forall t \in \mathcal{T}$; and (2) $|\{h \in \mathcal{H} : h(x_1) = t_1, h(x_2) = t_2\}| \leq \frac{\epsilon|\mathcal{H}|}{|\mathcal{T}|}, \forall x_1, x_2 \in \mathcal{M}\ (x_1 \neq x_2), \forall t_1, t_2 \in \mathcal{T}$. $\mathcal{H}$ is called 1-*universal* if only property 1 is satisfied.

A. Discrete Memoryless Channel

A *discrete memoryless channel* (DMC) $W : \mathcal{X} \to \mathcal{Y}$ is defined by a stochastic matrix $W = \{W(y|x)\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$, in which, $W(\cdot|x)$ is the distribution of channel output Y by inputting $X = x$. If the input is x^n and the output is y^n, then $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i) = \prod_{i=1}^n W(y_i|x_i)$. For simplicity, $W(y^n|x^n)$ is used to denote $\prod_{i=1}^n W(y_i|x_i)$.

A n-length code with codebook $\mathcal{C}$ for a DMC $W : \mathcal{X} \to \mathcal{Y}$ with message space $\mathcal{M}$ is an encoder $f : \mathcal{M} \to \mathcal{X}^n$ and a decoder $g : \mathcal{Y}^n \to \mathcal{M} \cup \{\perp\}$, where $\mathcal{C} = f(\mathcal{M})$ and $\perp$ denotes the decoding error. For $m \in \mathcal{M}$, $f(m) \in \mathcal{X}^n$ is a *codeword*. If a sender plans to send a message m, he/she transmits codeword $f(m)$. After receiving vector $y^n \in \mathcal{Y}^n$, the receiver decodes it to $g(y^n)$. If $g(y^n) \neq m$, an error occurs. The *probability of error* of a code (f, g) is defined $e(\mathcal{C}) = P(g(Y^n) \neq M)$, where Y^n is the channel output with message M that is uniformly distributed over $\mathcal{M}$.

B. Typical Sequences

The distribution $P_{x^n}(\cdot)$ is called the *type* of the sequence x^n over $\mathcal{X}$, where $P_{x^n}(a)$ is the fraction of occurrences of a in x^n. *Type Set* T^n_P of type P over $\mathcal{X}$ is the set of sequences with type P and length n.

Definition 1: Let X be an RV over $\mathcal{X}$. $x^n \in \mathcal{X}^n$ is ϵ*-typical* if $P_{x^n}(a) = 0$ for any a with $P_X(a) = 0$; and $|P_{x^n}(a) - P_X(a)| \leq \frac{\epsilon}{|\mathcal{X}|}$, otherwise. $\mathsf{T}^n_{[X]_\epsilon}$ is used to denote the set of ϵ-typical sequences. If $x^n = (y^n, z^n)$ is ϵ-typical for $X = (Y, Z)$, then (y^n, z^n) is *jointly* ϵ*-typical*. $\mathsf{T}^n_{[YZ]_\epsilon}$ is used to denote the set of jointly ϵ-typical sequences for Y and Z.

Definition 2: Let X (rep. Y) be RVs over $\mathcal{X}$ (rep. $\mathcal{Y}$). $y^n \in \mathcal{Y}^n$ is *conditionally* ϵ*-typical* given $x^n \in \mathcal{X}^n$, if $P_{x^n y^n}(a, b) = 0$ for any a, b with $P_{XY}(a, b) = 0$; and $|P_{x^n y^n}(a, b) - P_{x^n}(a) P_{Y|X}(b|a)| \leq \frac{\epsilon}{|\mathcal{X}| \cdot |\mathcal{Y}|}$, otherwise. $\mathsf{T}^n_{[Y|X]_\epsilon}(x^n)$ is used to denote the set of conditionally ϵ-typical sequences for Y, given x^n. $\mathsf{T}^n_{[W]_\epsilon}(x^n)$ is used to denote $\mathsf{T}^n_{[Y|X]_\epsilon}(x^n)$ for DMC $W : \mathcal{X} \to \mathcal{Y}$.

IV. The Authentication Framework

A. Authentication Syntax Model

The channel model considered here including a noiseless channel from Alice to Bob which is fully controlled by Oscar, a noiseless channel from Oscar to Bob, and a wiretap channel ($W_1 : \mathcal{X} \to \mathcal{Y}, W_2 : \mathcal{X} \to \mathcal{Z}$) from Alice to Bob and Oscar, respectively. Based on this channel model, we propose a message authentication model, which includes an *encoder* $F : \mathcal{M} \times \mathcal{K} \to \mathcal{S} \times \mathcal{X}^n$ and a *decoder* $G : \mathcal{K} \times \mathcal{S} \times \mathcal{Y}^n \to \mathcal{M} \cup \{\perp\}$, where $\mathcal{M}$ is the message space. The authentication syntax is described as follows.

(1) If Alice plans to send $M \in \mathcal{M}$ to Bob and authenticate it, she first calculates $(S, X^n) = F(M, K)$. And then, S and X^n are sent over a noiseless channel to Bob and over a wiretap channel (W_1, W_2), respectively. Bob receives S' from the noiseless channel through Oscar. Bob and Oscar receive Y^n and Z^n through W_1 and W_2, respectively.

(2) Upon (S', Y^n), Bob calculates $M' = G(K, S', Y^n)$. If $M' \neq \perp$, he accepts M'; otherwise, he rejects it.

Note that, if the noisy channel is not in use by Alice, then the proposed authentication model degenerates to a traditional one. So naturally, (F,G) is called *a message authentication code (MAC) scheme over channel* (W_1,W_2), and (S, X^n) is called the *codeword* of M. For the sake of convenience, a *decision bit* $D : \mathcal{K} \times \mathcal{S} \times \mathcal{Y}^n \to \{0,1\}$ is defined as $D(K,S',Y^n) = 0$ if $G(K,S',Y^n) = \perp$; and $D(K,S',Y^n) = 1$, otherwise.

B. Adversary Model

We focus on the adversary model as follows. (1) The adversary Oscar has infinite computing resources, and he can know the parameters configuration in the authentication scheme, except the pre-shared authentication key by Alice and Bob. (2) Oscar can receive the information transmitted over the noiseless channel from Alice to Bob, and the outputs of $W_2 : \mathcal{X} \to \mathcal{Z}$. (3) Oscar can arbitrarily modify information S transmitted over the noiseless channel from Alice to Bob for launching an attack. (4) Oscar can send any information to Bob over a noiseless channel to launch an attack. (5) Oscar is allowed to learn each decision bit of authentication from Alice to Bob.

Based on the adversary model, Oscar can eavesdrop the communication and obtain Z^n, S and the decision bit b, when Alice authenticates M to Bob with codeword (S, X^n). He can also launch an active attack in two different ways as follows. (1) Type-I attack: Oscar can modify S to arbitrary $S' \in \mathcal{S}$ when Alice authenticates M to Bob with codeword (S, X^n); (2) Type-II attack: Oscar can adaptively transmit $\hat{S} \in \mathcal{S}$ and $\hat{Y}^n \in \mathcal{Y}^n$ to Bob over a noiseless channel, even if Alice does not authenticate a message to Bob.

The main goal of this work is to ensure the success probability of adversarial attacks is *negligible*, even though the attacks number is polynomially bounded (in n). Throughout this work, a function $negl(n) : \mathbb{N} \to \mathbb{R}$ is *negligible*, if, for any positive integer c, there exists an integer N_c so that for all $n > N_c$, $|negl(n)| < n^{-c}$. he formal model of multiple attacks is as follows. Assume that Oscar has launched $t-1$ attacks, and Oscar's currently view is denoted by $View_{t-1}$.

Consider that the t-th attack is Type-I attack. Let M_t be the message authenticated by Alice, and (S_t, X_t^n) be codeword of M_t. When Oscar receives S_t, he can modify S_t to arbitrary $S_t' \in \mathcal{S}$, and send S_t' to Bob noiselessly, where S_t' is computed by Oscar based on $View_{t-1}$ and S_t. Oscar will obtain Z_t^n and learn the decision bit $b_t' = D(K, S_t', Y_t^n)$. This attack is successful if $b_t' = 1$. After this attack, Oscar's view $View_t = View_{t-1} \cup \{S_t, Z_t^n, b_t'\}$.

Consider that the t-th attack is Type-II attack. Oscar adaptively sends $\hat{S}_t \in \mathcal{S}$ and $\hat{Y}_t^n \in \mathcal{Y}^n$ to Bob over a noiseless channel, where $(\hat{S}_t, \hat{Y}_t^n)$ is computed by Oscar based on $View_{t-1}$. He will learn the decision bit $\hat{b}_t = D(K, \hat{S}_t, \hat{Y}_t^n)$. Oscar succeeds if $\hat{b}_t = 1$. After this attack, Oscar's view $View_t = View_{t-1} \cup \{\hat{b}_t\}$.

The model that allows Oscar to learn the verification result has been considered by Rei *et al.* in [36]. It is practical as Oscar can learn the decision bit by observing the receiver's action after rejecting or accepting a message. Actually, if Bob rejects a message, he could ask for a re-authentication on this message. In this case, Oscar leans that the decision bit is 0.

C. Authentication Property

Definition 3: Let n be the number of channel W_1 uses. A MAC scheme (F, G) over channel $W_1 : \mathcal{X} \to \mathcal{Y}, W_2 : \mathcal{X} \to \mathcal{Z}$ is *secure* if the following conditions holds:

1. **Completeness.** When Oscar is not present, Bob rejects the message with an exponentially (in n) small probability.
2. **Authentication.** Let **Succ** be the event that Oscar is successful by an attack. Pr(**Succ**) is negligible, when the number of Oscar's attacks is bounded polynomially in n.
3. **Key Security.** If the number of Oscar's attacks is bounded polynomially in n, the information leakage $I(K; View(Oscar))$ is arbitrarily small, where $Veiw(Oscar)$ is Oscar's view.

Remark 1: The number of Type-I attacks is polynomially bounded because each attack involves Bob (as a verifier), and it is impractical to require Bob to perform with complexity class beyond a polynomial. Restriction on the number of Type-II attacks is also inevitable, as Oscar can always choose a message M and impersonate with every possible $(s, y^n) \in \mathcal{S} \times \mathcal{Y}^n$ to Bob (since $\mathcal{S}$ and $\mathcal{Y}^n$ are finite sets, he can always succeed for some pair (s, y^n)). In a real secure communication scenario, the secret key is used not only for message authentication, but also for secure information transmission, digital signature, etc. Accordingly, it is necessary to ensure the security of the key during the message authentication process.

It is worth pointing out that multiple-time attack is much more powerful than a one-time attack, even for a traditional authentication system. For instance, let $F_{a,b}(M) = a + bM$ be an authentication system with secret key $(a, b) \in \mathbb{F}_p$ and source message $M \in \mathbb{F}_p$, in which, p is a prime number, and $\mathbb{F}_p$ is a finite field constructed from the integers modulo p. This system can be broken by eavesdropping two authentications (M_1, T_1) and (M_2, T_2), where $T_i = a + bM_i$ for $i = 1, 2$. However, it is impossible to forge an authentication tag of a new message with probability greater than $1/p$ after eavesdropping (M_1, T_1).

C_{11}	$\cdots$	C_{1j}	$\cdots$	$C_{1\mathbb{J}}$
$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
C_{i1}	$\cdots$	C_{ij}	$\cdots$	$C_{i\mathbb{J}}$
$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$C_{\mathbb{I}1}$	$\cdots$	$C_{\mathbb{I}j}$	$\cdots$	$C_{\mathbb{I}\mathbb{J}}$

Fig. 2. The codebook used in the authentication scheme.

Definition 4: For a secure MAC scheme (F, G) over (W_1, W_2), the ratio of the message length to the codeword length is called *authentication rate* of the MAC scheme, i.e, $\rho_{\mathbf{auth}} = \frac{1}{n} \log |\mathcal{M}|$, in which, $|\mathcal{M}|$ is the cardinality of the message space.

V. The Proposed MAC Scheme

A secure MAC scheme with high efficiency is presented in this section. The main idea of our MAC scheme is as follows. Alice first computes a tag T by using message M and a part of secret key K_0. Then, she encodes T to a codeword $\hat{X}^n$ with another part of key K_1 and some randomness. Finally, Alice sends M across a noiseless channel and transmits $\hat{X}^n$ over the wiretap channel (W_1, W_2). The main challenge is to encode T to $\hat{X}^n$ with a strong authentication property. Specifically, Oscar could eavesdrop the communication between Alice and Bob and launch Type-I or Type-II attacks in polynomial times.

To encode M into $\hat{X}^n$, a simple idea is to generate a tag T and send it to Bob using a wiretap code. However, this method has two issues: (1) For Bob to recover T, the information rate for T is bounded by secrecy capacity. However, we only need to authenticate M and It is not necessary for Bob to recover T. If this issue is addressed, we could improve the channel efficiency (for T). However, if Bob cannot recover T, he might accept the W_1 channel output $\hat{Y}^n$ that corresponds to an invalid $\hat{X}^n$. This in turn increases the success probability of Oscar. (2) Even if Alice uses a secure wiretap code to encode M, Oscar can still modify M over the noiseless channel to M' and learn whether Bob accepts $(\hat{Y}^n, M')$. This means that Oscar can still learn information for each authentication. It is important to ensure that after multiple authentications, Oscar cannot gradually learn the secret key completely. In the following, we design a channel code with a rate larger than the secrecy capacity for (W_1, W_2) to address the above issues. The channel output of W_2 gives no information about the input. The channel output of W_1 allows to learn the partial information about input rather than the complete information. It also has one cryptographic property that allows us to ensure the authentication property of the proposed scheme.

A. The Random Coding Lemma

To design an secure MAC scheme with high efficiency, the existence of a random channel coding satisfying the following properties will be proved. Let T_P^n be a typical set with type P over $\mathcal{X}$. There exists a codebook $\mathcal{C} \subseteq \mathsf{T}_P^n$ so that: (1) $\mathcal{C}$ is divided into subsets $\{C_{ij}\}_{i \in [\mathbb{I}], j \in [\mathbb{J}]}$ (as shown in Fig. 2) such that each column $C_{1j} \cup \cdots \cup C_{\mathbb{I}j}$ is the codebook of a good channel code (f_j, g_j) for channel W_1; (2) If RV I is over $[\mathbb{I}]$ uniformly at random and an arbitrary RV J over $[\mathbb{J}]$ is independent of I, then for $\hat{X}^n$ chosen from C_{IJ} uniformly at random which is transmitted over (W_1, W_2), the output $\hat{Z}^n$

of wiretapper's channel W_2 is nearly independent of (I, J); (3) If RV J' depends on J (under certain constraints) and the output of main channel W_1 is $\hat{Y}^n$, it is negligible that $\hat{Y}^n$ can be decoded into a codeword in $\mathcal{C}_{IJ'}$ by using $g_{J'}$.

Note that the dependency between J and J' could be high. For example, in our authentication scheme, $J = F_{K_0}(M)$ and $J' = F_{K_0}(M')$ for known M and M'. Therefore, J and J' depend on an unknown secret K_0.

Lemma 1: Let (W_1, W_2) be a wiretap channel with $P_{Y|X} = W_1$ and $P_{Z|X} = W_2$ such that $P_X = P$ for a type P over $\mathcal{X}$ satisfying $P(x) > 0$ for all$(x \in \mathcal{X})$. If $I(X;Y) > I(X;Z) + \tau$ for a constant $\tau > 0$. Then, for any integer pairs $\mathbb{I}$ and $\mathbb{J}$ with

$$0 \le \frac{1}{n} \log \mathbb{I} < I(X;Y) - I(X;Z) - \tau, \tag{2}$$

$$0 \le \frac{1}{n} \log \mathbb{J} < H(X|Y) + \tau, \tag{3}$$

there exists disjoint subsets $\mathcal{C}_{ij} \subset \mathsf{T}^n_P$ (where $i \in [\mathbb{I}]$ and $j \in [\mathbb{J}]$) s.t. for a sufficiently large n, the properties hold as follows:

1. $\forall j$, $\mathcal{C}_{\cdot j} \stackrel{def}{=} \bigcup_i \mathcal{C}_{ij}$ is the codebook of a channel coding (f_j, g_j) for channel W_1 with an exponentially small average probability of error, in which f_j encodes message m to the m-th codeword in codebook $\mathcal{C}_{\cdot j}$;
2. for any RVs I over $[\mathbb{I}]$ and J over $[\mathbb{J}]$ by $P_{IJ} = \frac{P_J}{\mathbb{I}}$, if $\hat{Z}^n$ is the output of channel W_2 by inputting $\hat{X}^n \leftarrow_U \mathcal{C}_{IJ}$, then $I(I, J; \hat{Z}^n) \le 2^{-n\beta_2}$, for some $\beta_2 > 0$ (not depending on P_J);
3. for any J over $[\mathbb{J}]$ and I over $[\mathbb{I}]$ with $P_{IJ} = \frac{P_J}{\mathbb{I}}$, let $\hat{Y}^n$ be the output of channel W_1 with input $\hat{X}^n \leftarrow_U \mathcal{C}_{IJ}$. Assume RV J' over $[\mathbb{J}]$ satisfying (a) $J' \neq J$; (b) $\mathsf{D}(P_{J'J}; P_{J'J|I}) \le \delta_1$; (c) $J' \to IJ \to \hat{X}^n \to \hat{Y}^n$ is a Markov chain; (d) $\forall j, j'$, there exists a function $d(\cdot,\cdot)$ with $\sum_{j',j} d(j', j) < \delta_2$ and a constant $\omega \in (0, 1)$ s.t. $P_{J'J}(j', j) \le \frac{2^{n^\omega}}{\mathbb{J}(\mathbb{J}-1)} + d(j', j)$. Then,

$$P\Big(g_{J'}(\hat{Y}^n) \in \mathcal{C}_{IJ'}\Big) \le 2^{-n^\omega} + \delta_1 + \delta_2. \tag{4}$$

In what follows, the idea of the proof will be present. For the details of the proof, please refer to the Appendix.

For properties 1 and 2, we consider independent and uniformly random partitions $\sigma_1 : \mathsf{T}^n_P \to \{1, \ldots, s_1\}$ and $\sigma_2 : \mathsf{T}^n_P \to \{1, \ldots, s_2\}$ for T^n_P. Then, $\sigma = (\sigma_1, \sigma_2)$ is a random partition of size $s_1 s_2$ for T^n_P, through $\mathcal{A}_{ij} = \sigma^{-1}(i, j)$. Then, by Lemma 8 (with s_1, s_2 defined properly), we have

$$|\mathcal{A}_{ij}| = \frac{|\mathsf{T}^n_P|}{s_1 s_2}(1 + \epsilon_{ij}) \tag{5}$$

$$\mathsf{D}(\hat{Z}^n | \sigma(\hat{X}^n); \hat{Z}^n) < 2^{-n\beta_1} \tag{6}$$

for $\beta_1 > 0$ and small $\epsilon_{ij} \ge 0$. As $\mathcal{A}_{\cdot j} := \mathcal{A}_{1j} \cup \cdots \cup \mathcal{A}_{s_1 j} = \sigma_2^{-1}(j)$ is a random subset of T^n_P (of size no larger than the constraint on s_1, s_2), by Lemma 13, most of $\mathcal{A}_{\cdot 1}, \ldots, \mathcal{A}_{\cdot s_2}$ are codebooks with small decoding errors. If all of $\mathcal{A}_{\cdot 1}, \ldots, \mathcal{A}_{\cdot s_2}$ are codebooks with small decoding errors and $\epsilon_{ij} = 0$, then properties 1-2 hold by defining $\mathcal{C}_{ij} = \mathcal{A}_{ij}$, as in this case, property 2 is just Eq. (6). For the general case, since ϵ_{ij} is small and most of $\mathcal{A}_{\cdot j}$'s are good codes, we can discard $\mathcal{A}_{\cdot j}$ (that is not a good code) and define $\mathcal{C}_{ij}$ to be $\mathcal{A}_{ij}$ (where $\mathcal{A}_{\cdot j}$ is a good code) except cutting off a small subset of $\mathcal{A}_{ij}$ (to make $\mathcal{C}_{ij}$ having an equal size). As the changes are minor, the resulting $\mathcal{C}_{\cdot j}$ will remain a codebook of a good code and satisfy property 2.

For property 3, we need to give an upper bound for $P(g_{J'}(\hat{Y}^n) \in \mathcal{C}_{IJ'})$. Since $g_{J'}(\cdot)$ is a typical decoding function, it follows that it is upper bounded by

$$P(\hat{Y}^n \in \mathsf{T}^n_{[W_1]_\epsilon}(\mathcal{C}_{IJ'})). \tag{7}$$

Based on the conditions in property 3, we can reduce Eq. (7) to $\delta_1 + \delta_2$ plus the same probability

$$P(\hat{Y}^n \in \mathsf{T}^n_{[W_1]_\epsilon}(\mathcal{C}_{IJ'})) \tag{8}$$

except that $I, (J, J'), \hat{X}^n$ are independent and uniform RVs in their respective domains (here the uniform randomness of (J, J') means a random pair in $\{1, \ldots, \mathbb{J}\}$). That is, $P_{IJJ'\hat{X}^n} = \frac{1}{r\mathbb{J}(\mathbb{J}-1)\mathbb{I}}$, where $r = |\mathcal{C}_{IJ'}|$. Since $\hat{Y}^n$ is typical with $\hat{X}^n$, the following approximation holds with high probability,

$$\begin{aligned} Eq.(8) &\approx P(\hat{Y}^n \in \mathsf{T}^n_{[W_1]_\epsilon}(\mathcal{C}_{IJ'}) \cap \mathsf{T}^n_{[W_1]_\epsilon}(\hat{X}^n)) \\ &\le \sum_{t=1}^{r} P(\hat{Y}^n \mathsf{T}^n_{[W_1]_\epsilon}(u_t) \cap \mathsf{T}^n_{[W_1]_\epsilon}(\hat{X}^n)) \\ &\approx \sum_{t=1}^{r} \frac{|\mathsf{T}^n_{[W_1]_\epsilon}(u_t) \cap \mathsf{T}^n_{[W_1]_\epsilon}(\hat{X}^n)|}{2^{nH(Y|X)}} \end{aligned} \tag{9}$$

where $\{u_1, \ldots, u_r\} = \mathcal{C}_{IJ'}$ with a random ordering. Note that when $(I, J, J') = (i, j, j')$, u_t is over $\mathcal{C}_{ij'}$ uniformly at random as u_t is indexed uniformly at random. Further because $\mathcal{C}_{ij'}$ and $\mathcal{A}_{ij'}$ are chosen from $\mathcal{A}_{ij'}$ and T^n_P uniformly at random, respectively, it follows that u_t is uniformly distributed in T^n_P. Similarly, $\hat{X}^n$ is uniformly random in T^n_P. Thus, from Lemma 36, the bound for Eq. (9) can be derived.

Discussion: The main purpose of the proposed code in Lemma 1 is to design a secure authentication scheme in Section IV.B. In the proposed authentication scheme, J is the authentication tag, and J' is the authentication tag corresponding to the fake source message generated by Oscar. If Oscar generates a fake source message M' by modifying M on the noiseless channel, then the tag will correspondingly change from J to J'. J' is not necessarily known to Oscar. However, for a bad tag function, Oscar might be able to create M' so that J' and J are related (although neither of them is known to Oscar). It becomes even more interesting when Oscar can learn the information on the secret key of the tag function through active attacks (introduced in our model).

B. The Proposed MAC Scheme

Let $W_1 : \mathcal{X} \to \mathcal{Y}$, $W_2 : \mathcal{X} \to \mathcal{Z}$ be a wiretap channel such that $I(X;Y) > I(X;Z) + \tau$ for some $\tau > 0$ and P_X be a type P over $\mathcal{X}$ satisfying $P(x) > 0$ $(\forall x \in \mathcal{X})$. Let $\mathcal{C}_{ij}$ $(i = \in [\mathbb{I}]$ and $j \in [\mathbb{J}])$ be the subsets of T^n_P obtained in Lemma 1. The details of the proposed scheme are shown in Scheme 1.

In the channel coding (f_j, g_j) on $\mathcal{C}_{\cdot j}$ in Lemma 1, the encoder f_j encodes message ℓ to the ℓ-th codeword of $\mathcal{C}_{\cdot j}$, and g_j decodes Y^n to the index of a codeword in $\mathcal{C}_{\cdot T'}$ or $\perp$. Since a codeword's index and its codeword is one-to-one mapping, it is assumed that Y^n is decoded to the codeword itself or $\perp$ with decoder g_j.

Scheme 1 The Proposed MAC Scheme

Setup:

- Set $\mathcal{K}_1 = \{1, \ldots, \mathbb{I}\}$, and $\mathcal{T} = \{1, \ldots, \mathbb{J}\}$.
- Let $K{=}(K_0, K_1) \in \mathcal{K}_0 \times \mathcal{K}_1$ be the authentication key shared by Alice and Bob.
- Let $\{h_{k_0} : \mathcal{M} \to \mathcal{T}\}_{k_0 \in \mathcal{K}_0}$ be a ϵ-ASU hashing family with the key space $\mathcal{K}_0$.

Scheme: If the sender Alice plans to authenticate message $M \leftarrow_{P_M} \mathcal{M}$ to Bob, then they perform:

- *Encoding.* Alice first calculates $T = h_{K_0}(M)$. Then Alice takes X^n from $\mathcal{C}_{K_1 T}$ uniformly at random. Finally, Alice obtains M's codeword (M, X^n), where M and X^n are sent over the noiseless channel and wiretap channel (W_1, W_2), respectively.
- *Decoding.* After receiving (M', Y^n), Bob calculates $T'{=}h_{K_0}(M')$. If $g_{T'}(Y^n) \in \mathcal{C}_{K_1 T'}$, he accepts M'; and he rejects it, otherwise, in which, g_j is the decoder of W_1 with codebook $\mathcal{C}_{\cdot j}$.

Remark 2: The main idea of our MAC scheme is as follows. If M is not modified (i.e., $M' = M$), then $T' = T$. Hence, Bob accepts the message. In contrast, if M' is modified, then $g_{T'}(Y^n) \notin \mathcal{C}_{K_1 T'}$, which is guaranteed by property 3 in Lemma 1. This is equivalent to $Y^n \notin \mathsf{T}^n_{[W_1]_\epsilon}(\mathcal{C}_{K_1 T'})$, as $g_{T'}$ is a typical decoding function. Hence, although there are several j's so that $\mathcal{C}_{K_1 j}$ contains some x^n that is typical with Y^n, Oscar cannot find M' so that T' $(= h_{K_0}(M'))$ happens to be one of such j. Note that Oscar does not know K_0 and thus cannot compute T'. But he could learn some information about K_0 through active attacks and learning the verification results from Bob.

Remark 3: With the proposed scheme, the larger the tag size $|\mathcal{T}|$ (i.e., $\mathbb{J}$) is, the higher the authentication rate is. From Lemma 1, for the proposed scheme, $\mathbb{J}$ can be large as $2^{n(H(X|Z)-\delta)}$ by taking τ close to $I(X;Y) - I(X,Z)$, where δ is nearly zero. The main reason for having such a large $\mathbb{J}$ is that, it is not necessary to decode Y^n to X^n according to Lemma 1. To demonstrate this advantage, we can obtain a naive MAC scheme by modifying the verification of the proposed MAC scheme as follows. Bob tries to decode X^n from Y^n. Since $\mathcal{C}_{ij}$'s are disjoint, he can find a unique T such that $X^n \in \mathcal{C}_{K_1 T}$. However, it is well-known that the tag size $\mathbb{J}$ of the naive scheme can be upper bounded by $2^{nI(X;Y)}$ (without considering the security of the naive scheme), which is not necessarily larger than $2^{n(H(X|Z)-\delta)}$.

Remark 4: In Lemma 1, property 2 guarantees that Oscar learns nothing about K_1 and T from his observation Z^n. Intuitively, after observing many authentication instances, this should hold essentially. However, this only states that Z^n is almost independent of K_1, T. In the security model, Oscar can obtain information much more than just Z^n. He can impersonate Alice to authenticate some $\tilde{M}$ to Bob, or modify M to M'. In any case, he can also learn the verification result by Bob. Given such active attacks, K_1, T are certainly not independent of the view by Oscar.

To understand the last remark, we give an example of $\{\mathcal{C}_{ij}\}_{ij}$ that satisfies properties 1 and 2 (but not property 3), where the scheme is not secure.

Let $\mathbb{I} = 2^{nc}$ and $\mathbb{J} = 2^{nc}$ for $0 < c \ll \frac{1}{2} I(X;Y)$ (also satisfying the constraint in Lemma 1). Let $\{\hat{\mathcal{C}}_{ij}\}_{(i,j)\in[\mathbb{I}_0]\times[\mathbb{J}_0]}$ be the subsets obtained by using Lemma 1. Notice that $|\mathsf{T}^n_{[Y]_\epsilon}| \approx 2^{nH(Y)}$. Since c is small, $\mathsf{T}^n_{[Y]_\epsilon} - \mathsf{T}_{[W_1]_\epsilon}(\cup_{ij}\mathcal{C}_{ij})$ is not empty and assume that $\hat{y}^n$ is such a sequence. Let $\{x^n_{ij} \mid i \in [\mathbb{I}], j \in [\mathbb{J}]\}$ be a set of sequences such that each x^n_{ij} is typical with $\hat{y}^n$. It is true as $\hat{y}^n$ has approximately $2^{nH(Y)}$ possible x^n that is typical with it while c is small. Let $\mathcal{C}_{ij} = \hat{\mathcal{C}}_{ij} \cup \{x^n_{ij}\}$. Then, $\{\mathcal{C}_{ij}\}_{ij}$ will satisfy properties 1 and 2 with slightly changed average error probability in property 1 and β_2 in property 2.

However, if this collection of $\mathcal{C}_{ij}$ is used in the proposed authentication scheme, Oscar can launch an active attack by revising Alice's noiseless channel message M to M'. Obviously, Bob will accept M' as $\hat{y}^n \in \mathsf{T}_{[W_1]_\epsilon}(x^n_{K_1 T'}) \subseteq \mathsf{T}_{[W_1]_\epsilon}(\mathcal{C}_{K_1 T'})$, where $T' = h_{K_0}(M')$.

Example 1: Let W_1 be a noiseless channel and W_2 be a binary erasure channel (BEC) with erasure probability $e_1 > 0$. Then, the subsets $\{\mathcal{C}_{ij}\}_{i,j}$ can be trivially designed as follows. Let $\mathcal{C}_{11}$ be the codebook of a channel code of wiretapper's channel with code length n and exponential decoding error, and $\{\mathcal{C}_{ij}\}_{i,j}$ $(i \in [\mathbb{I}], j \in [\mathbb{J}])$ be the set of coset of $\mathcal{C}_{11}$. Note that, for any sufficiently small positive number δ, the rate of channel code on $\mathcal{C}_{11}$ can be $1 - e_1 - \delta$ when n is large enough (e.g., the large-girth LDPC codes deigned in [23]). In this case, the number of coset is larger than $2^{n(e_1-\delta/2)}$. Let $|\mathbb{I}| = 2^{n\delta/2}$, and $|\mathbb{J}| = 2^{n(e_1-\delta)}$. It is easy to show that $\{\mathcal{C}_{ij}\}_{i,j}$ satisfy the Properties (1)-(3) in Lemma 1. Thus, $\{\mathcal{C}_{ij}\}_{i,j}$ can be used in Scheme 1 to authenticate the message.

VI. Security Analysis

A. Technical Lemmas

We begin with two lemmas (Lemma 2 and 3). The first lemma states that Oscar obtains no significant amount of information about secret key (K_0, K_1), after *eavesdropping* J times of authentications which give Oscar information $M_1 Z^n_1, \ldots, M_J Z^n_J$. The main idea is as follows. Let $T_j = h_{K_0}(M_J)$, and it holds that $I(K_1 T_j; Z^n_j | M_j = m_j) \approx 0$, according to Lemma 1 (property 2). Note that given $M_j = m_j$, $K_0 K_1 \to K_1 T_j \to Z^n_j$ forms a Markov chain as Z^n_j depends on $K_1 T_j$ and some randomness independent of $K_0 K_1$. Therefore, by data processing inequality, $I(K_0 K_1; Z^n_j | M_j = m_j) \le I(K_1 T_j; Z^n_j | M_j = m_j) \approx 0$. As $K_0 K_1$ is independent of M^J,

$$I(K_0 K_1; M^J Z^n_1 \cdots Z^n_J) = I(K_0 K_1; Z^n_1 \cdots Z^n_J | M^J). \quad (10)$$

Finally, by standard information theory techniques, we can show that Eq. (10) is bounded by

$$\sum_{j=1}^{J} I(K_0 K_1; Z^n_j | M^J = m^J),$$

which is now known small. The lemma follows by averaging on M^J.

Lemma 2: Let (K_0, K_1) be the RVs uniformly at random from key space $\mathcal{K}_0 \times \mathcal{K}_1$ and $M_1, \ldots, M_J$ be arbitrary J messages in $\mathcal{M}$. For $j = 1, \ldots, J$, let Z^n_j be the output of

W_2 when Alice sends X_j^n (w.r.t. M_j). Then, there exists a constant $\beta_2 > 0$ s.t. when n is large enough,

$$I(K_0K_1; M_1Z_1^n \cdots M_J Z_J^n) \le J \cdot 2^{-n\beta_2}. \quad (11)$$

Proof: For $j = 1, \ldots, J$, let $T_j = h_{K_0}(M_j)$. Define $M^J = M_1 \cdots M_J$ and $m^J = m_1 \cdots m_J$ for $m_j \in \mathcal{M}$. Then, $P_{K_1T_j|M^J}(k_1, t_j|m^J) = P_{K_1}(k_1)P_{T_j|M^J}(t_j|m^J)$ as K_1 is independent of M^J and K_0. Hence, we have $P_{K_1T_j|M^J=m^J}(k_1, t_j) = P_{K_1}(k_1)P_{T_j|M^J=m^J}(t_j) = P_{T_j|M^J=m^J}(t_j)/|\mathcal{K}_1|$. If we rewrite the joint distribution of K_1, T_j when $M^J = m^J$, as $P_{K_1T_j}^{m^J}$, then $P_{K_1T_j}^{m^J} = P_{T_j}^{m^J}/|\mathcal{K}_1|$. Hence, by Lemma 1 (property 2), when n is sufficiently large, we have $I(K_1, T_j; Z^n|M^J = m^J) \le 2^{-n\beta_2}$ for some $\beta_2 > 0$, where β_2 does not depend on m^J.

Given $M^J = m^J$, we have that $(K_0, K_1) \to (T_j, K_1) \to Z_j^n$ forms a Markov chain. Hence, it holds that $I(K_0K_1; Z_j^n|M^J = m^J) \le I(K_1T_j; Z_j^n|M^J = m^J) \le 2^{-n\beta_2}$ by data processing inequality. Thus, $I(K_0K_1; Z_j^n|M^J) \le I(K_1T_j; Z_j^n|M^J) \le 2^{-n\beta_2}$ by averaging over m^J.

For any $j \in [J]$, let $M^J = m^J$. Since X_j^n is fully determined by (K_0K_1, m^J) and X_j^n is selected from $\mathcal{C}_{K_1T_j}$ uniformly at random, and Z_j^n is determined by X_j^n and the noise in channel W_2, we have that $Z_1^n \cdots Z_{j-1}^n \to K_0K_1 \to Z_j^n$ forms a Markov chain. Accordingly,

$$\begin{aligned}I(K_0K_1; Z_j^n|Z_1^n \cdots Z_{j-1}^n, M^J = m^J)& \\ \le I(K_0K_1; Z_j^n|M^J = m^J).&\end{aligned}$$

Averaging over m^J, we have

$$I(K_0K_1; Z_j^n|Z_1^n \cdots Z_{j-1}^n M^J) \le I(K_0K_1; Z_j^n|M^J). \quad (12)$$

Hence, from chain rule of mutual information, we have

$$\begin{aligned}&I(K_0K_1; Z_1^n \cdots Z_J^n M^J) \\ &= I(K_0K_1; M^J) + I(K_0K_1; Z_1^n \cdots Z_J^n|M^J) \\ &= I(K_0K_1; Z_1^n \cdots Z_J^n|M^J), \quad (K_0K_1 \text{ is independent of } M^J) \\ &\le \sum_j I(K_0K_1; Z_j^n|M^J) \le J2^{-n\beta_2}.\end{aligned}$$

This completes the proof. ■

The following result will be utilized to show that the conditional distribution of key on the decision bit is almost uniform.

Lemma 3: Let V and K be RVs over $\mathcal{V}$ and $\mathcal{K}$, respectively. Then, for any $v \in \mathcal{V}$ and any $\mathcal{K}_v \subseteq \mathcal{K}$,

$$|P_{K|V=v}(\mathcal{K}_v) - P_K(\mathcal{K}_v)| \le \frac{1}{2}\mathsf{D}(P_{K|V=v}; P_K). \quad (13)$$

Proof: As $\mathsf{D}(P_{X_1}; P_{X_2}) = 2\max_{A\subseteq \mathcal{X}}\{P_{X_1}(A) - P_{X_2}(A)\}$ for any RVs X_1, X_2 over $\mathcal{X}$, $P_{K|V=v}(\mathcal{K}_v) - P_K(\mathcal{K}_v) \le \frac{1}{2}\mathsf{D}(P_{K|V=v}; P_K)$. Similarly, $-P_{K|V=v}(\mathcal{K}_v) + P_K(\mathcal{K}_v) \le \frac{1}{2}\mathsf{D}(P_{K|V=v}; P_K)$. Hence, the lemma follows. ■

The following lemma states that $T'T$ and K are almost independent. It will be used to specify a value for δ_1 in Lemma 1 (property 3-b).

Lemma 4: Let K_0, K_1, U, M', M be RVs over $\mathcal{K}_0, \mathcal{K}_1, \mathcal{U}, \mathcal{M}$ and $\mathcal{M}$ respectively with M', M being deterministic in U. Let $\{h_{k_0}\}_{k_0\in\mathcal{K}_0}$ be *any* family of functions from $\mathcal{M}$ to $\mathcal{T}$. Let $T' = h_{K_0}(M')$ and $T = h_{K_0}(M)$. If K_1 is independent of (K_0, M), then for $\Delta = \mathsf{D}(K_0K_1|U; K_0K_1)$,

$$\mathsf{D}(P_{T'T|K_1}; P_{T'T}) \le \sqrt{2\Delta \ln \frac{|\mathcal{K}_0||\mathcal{K}_1|}{\Delta}}. \quad (14)$$

Proof: On the one hand, we have

$$\begin{aligned}I(T'; K_1|T) &\le I(UK_0; K_1|T) \\ &= I(U; K_1|TK_0), \ (\text{as } I(K_0; K_1|T) = 0) \\ &= H(K_1|TK_0) - H(K_1|UTK_0) \\ &= H(K_1) - H(K_1|UK_0), \\ &\quad (K_1 \text{ is ind. of } TK_0;\ K_0U \text{ determines } T) \\ &= I(K_1; UK_0) = I(K_1; U|K_0) \le I(K_0K_1; U), \\ &\quad (\text{ as } K_0 \text{ and } K_1 \text{ are ind.})\end{aligned}$$

On the other hand, by [28, Lemma 1], we have

$$I(T'; K_1|T = t) \ge \frac{\left(\sum_{k_1} P_{K_1|T=t}(k_1)\mathsf{D}(P_{T'|T=t}; P_{T'|K_1T=k_1t})\right)^2}{2\ln 2}.$$

By the convexity of $f(x) = x^2$, we have

$$\begin{aligned}I(T'; K_1|T) &= \sum_t P_T(t) I(T'; K_1|T = t) \\ &\ge \frac{\left(\sum_{k_1,t} P_{K_1T}(k_1, t)\mathsf{D}(P_{T'|T=t}; P_{T'|K_1T=k_1t})\right)^2}{2\ln 2}.\end{aligned}$$

Thus, it follows that

$$\begin{aligned}&\sum_{k_1,t} P_{K_1T}(k_1, t)\mathsf{D}(P_{T'|T=t}; P_{T'|K_1T=k_1t}) \\ &\qquad\le \sqrt{2\ln 2 \cdot I(K_0K_1; U)}.\end{aligned}$$

After reformatting the left side of the inequality above, we can obtain the result from [28, Lemma 1] together with the independence between K_1 and T. ■

The following lemma will be used to specify a value for δ_2 in Lemma 1 (property 3-d) and to show that the third condition in Lemma 1 (property 3) can be satisfied.

Lemma 5: Let U, M', M be RVs over $\mathcal{U}, \mathcal{M}$ and $\mathcal{M}$ respectively s.t. M', M are deterministic in U. Let $\{h_{k_0} : \mathcal{M} \to \mathcal{T}\}_{k_0\in\mathcal{K}_0}$ be ϵ-ASU hash functions. Let K_0 be uniformly distributed over $\mathcal{K}_0$, $T' = h_{K_0}(M')$ and $T = h_{K_0}(M)$. If $P(M' = M) = 0$, then there exists function $d(t', t)$ satisfying $\sum_{t',t} d(t', t) \le \mathsf{D}(K_0|U; K_0)$ and

$$P_{T'T}(t', t) \le d(t', t) + \frac{\epsilon}{|\mathcal{T}|}. \quad (15)$$

Proof: Defining $\mathcal{K}_0(u, t', t) = \{k_0 : h_{k_0}(m') = t'; h_{k_0}(m) = t\}$, where m', m are the values of M' and M determined by $U = u$. Let $d(t', t) = \sum_u |P_{K_0U}(\mathcal{K}_0(u, t', t), u) - P_{K_0}(\mathcal{K}_0(u, t', t))P_U(u)|$. Then, we have

$$\begin{aligned}P_{T'T}(t', t) &= \sum_u P_{K_0U}(\mathcal{K}_0(u, t', t), u) \\ &\le d(t', t) + \sum_u P_{K_0}(\mathcal{K}_0(u, t', t))P_U(u) \\ &\le d(t', t) + \frac{\epsilon}{|\mathcal{T}|}, \quad (\text{as } |\mathcal{K}_0(u, t', t)| \le \tfrac{\epsilon|\mathcal{K}_0|}{|\mathcal{T}|}) \quad (16)\end{aligned}$$

For any u, $\{\mathcal{K}_0(u,t,t')\}_{t,t'}$ are disjoint. Hence,

$$\begin{aligned}\mathsf{D}(K_0|U;K_0) &= \sum_{k_0,u} |P_{K_0U}(k_0,u) - P_{K_0}(k_0)P_U(u)| \\ &\geq \sum_{t,t',u} |P_{K_0U}(\mathcal{K}_0(u,t',t),u) \\ &\quad - P_{K_0}(\mathcal{K}_0(u,t',t))P_U(u)| = \sum_{t,t'} d(t',t).\end{aligned}$$

This completes the proof. ■

B. Authentication Theorem

Based on the above lemmas, the following theorem is obtained, which indicates the conditions for the proposed MAC scheme to achieve ITS under polynomial MITM attacks.

Theorem 1: Let (W_1, W_2) be a wiretap channel satisfying $I(X;Y) \geq I(X;Z) + \tau$ for some $\tau > 0$ and P_X is a type P over $\mathcal{X}$ with $P(x) > 0$ $(\forall x \in \mathcal{X})$. Let $K = (K_0, K_1)$ be the pre-shared secure key between Alice and Bob. Consider that $\{h_{k_0} : \mathcal{M} \to \mathcal{T}\}_{k_0 \in \mathcal{K}_0}$ is an ϵ-ASU hash functions with $\epsilon = \min\{2^{-\Omega(\log n)}, \frac{2^{n^\omega}}{|\mathcal{T}|}\}$ for some $\omega \in (0,1)$ and $|\mathcal{K}_1| = 2^{\Omega(\log n)}$, where $\varphi(n) = \Omega(\log n)$ if $\lim_{n\to\infty} \frac{\varphi(n)}{\log n} = \infty$. Then, the proposed MAC scheme is secure.

In this theorem, it is assumed that $I(X;Y) \geq I(X;Z) + \tau$ for some constant $\tau > 0$. Such an assumption can be relaxed to the condition that W_2 is *not less noisy than* W_1. Actually, under this condition, there exists an RV U satisfying the following requirements: $U \to X \to YZ$ forms a Markov chain; and $I(U;Y) \geq I(U;Z) + \tau$ for a constant $\tau \in (0, I(U;Y) - I(U;Z))$. Let $(W_1' : \mathcal{U} \to \mathcal{Y}, W_2' : \mathcal{U} \to \mathcal{Z})$ be a virtual wiretap channel, in which $W_1' = P_{Y|U}$ and $W_2' = P_{Z|U}$. By using U and (W_1', W_2') to replace X and (W_1, W_2) in Th. 1 respectively, the proposed MAC scheme (denoted as Π') is secure for (W_1', W_2'). A secure MAC scheme Π for (W_1, W_2) can be induced by Π' as follows.

- *Encoding.* Alice first computes the codeword (M, U^n) of message M with Π', then obtains the output X^n by simulating the noisy channel $W_0, \mathcal{U} \to \mathcal{X}$ with input U^n, where $W_0 = P_{X|U}$, and finally, transmits M and X^n to Bob over the noiseless channel and (W_1, W_2), respectively;
- *Decoding.* After receiving the outputs of these two channels, Bob verifies M by using Π'.

In this following, the authentication theorem (i.e., Th. 1) will be proved. To this end, we aim to show that Alice can authenticate a polynomial number of messages with (K_0, K_1) under adaptively interleave two types of attacks from Oscar. In Type-I attack, Oscar revises M to $M'(\neq M)$ when Alice transmits (M, X^n); while in Type-II attack, Oscar selects $(\hat{M}, \hat{Y}^n)$ adaptively and transmits them to Bob over the noiseless channel. Oscar succeeds, when $g_{T'}(Y^n) \in \mathcal{C}_{K_1T'}$ after launching a Type-I attack (in which, $T' = h_{K_0}(M')$), or $g_{\hat{T}}(\hat{Y}^n) \in \mathcal{C}_{K_1\hat{T}}$ after launching a Type-II attack (in which, $\hat{T} = h_{K_0}(\hat{M})$).

The proof idea is as follows. $b_\ell = 1$ is used to denote Oscar's success in the ℓ-th attack (Type-I / Type-II). In a Type-I attack, two cases are considered: (1) $h_{K_0}(M') = h_{K_0}(M)$ (i.e., $T' = T$), where Oscar has high probability of success by the completeness of the coding scheme (f, g); and (2) $g_{T'}(Y^n) \in \mathcal{C}_{K_1T'}$ but $T' \neq T$. In case (1), suppose M' is independent of K_0, the success probability of Oscar is upper bounded by ϵ from the property of hash functions. Conceivably, if $\mathsf{D}(K_0|M';K_0)$ is small (i.e., M' is almost independent of K_0), Oscar still succeeds with a small probability. Note that M' depends on Oscar's view U_ℓ. Hence, it suffices to show that $\mathsf{D}(K_0|U_\ell;K_0)$ is small enough. In case (2), Lemma 1 (property 3) is used to prove that the probability of Oscar's success is small. In a Type-II attack, if $(\hat{M}, \hat{Y})$ (which depends on U_ℓ) is independent of K_1, then $g_{\hat{T}}(\hat{Y}^n) \in \mathcal{C}_{K_1\hat{T}}$ holds with probability $\frac{1}{|\mathbb{K}_1|}$. Conceivably, if $\mathsf{D}(K_1|U_\ell;K_1)$ is small, then $g_{\hat{T}}(\hat{Y}^n) \in \mathcal{C}_{K_1\hat{T}}$ should hold with a small change in success probability. As $\mathsf{D}(K_c|U_\ell;K_c) \leq \mathsf{D}(K_0K_1|U_\ell;K_0K_1)$ for $c = 0, 1$, we only need to prove $\mathsf{D}(K_0K_1|U_\ell;K_0K_1)$ is small, which can be obtained by combining Lemmas 2-5.

C. Proof of Th. 1

Proof: The completeness of the MAC scheme holds by Lemma 1 (property 1). In what follows, we mainly focus on the remaining two properties: authentication and key security.

Let $M^\nu = M_1 \cdots M_\nu$ be ν messages authenticated by Alice and X_i^n and Z_i^n be the input and output w.r.t. M_i over channel W_2, respectively. Note that M^ν is selected by Alice from distribution P_{M^ν} (especially independent of Oscar's random tape $Rand_o$); X_i^n is determined by (K_0K_1, M_i) and the randomness of sampling X_i^n from $\mathcal{C}_{K_1T}$; and Z_i^n depends on X_i^n and the channel noise of W_2. It follows that $(M^\nu, K_0K_1, X_1^nZ_1^n \cdots X_\nu^nZ_\nu^n)$ is independent of $Rand_o$ and hence has the same distribution when Oscar is not present. Hence, by Lemma 2, $I(K_0K_1; M^jZ_1^n \cdots Z_j^n) \leq j2^{-n\beta_2}$, for one constant $\beta_2 > 0$ and all $j \leq \nu$.

As $(M^j, K_0K_1, X_1^nZ_1^n \cdots X_j^nZ_j^n)$ is independent of $Rand_o$, $I(K_0K_1; Rand_oM^jZ_1^n \cdots Z_j^n) \leq j2^{-n\beta_2}$. Let $K \stackrel{def}{=} K_0K_1$ and $V_j \stackrel{def}{=} Rand_oM^jZ_1^n \cdots Z_j^n$. By [28, Lemma 1],

$$\mathsf{D}(K|V_j;K) \leq \sqrt{2j\ln 2} \cdot 2^{-n\beta_2/2}. \tag{17}$$

According to adversary model, Oscar can adaptively switch between the two attacks. (I) When Alice sends out (M_j, X_j^n), Oscar can can adaptively select a message $M_j'(\neq M_j)$ and modify M_j to M_j'; (II) at any time, Oscar can adaptively select and transmit a pair $(\hat{M}, \hat{Y}^n)$ to Bob noiselessly.

The result of the ℓ-th attack (either type I or type II above) is denoted by a binary variable b_ℓ, in which, $b_\ell = 1$ iff the attack is successful.

Assume that Alice has authenticated $M^{j_\ell - 1}$ to Bob before the ℓ-th attack launched by Oscar. Then, the *view* of Oscar is $U_\ell := (V_{j_\ell - 1}, b_1, \ldots, b_{\ell-1})$, where a party's view includes his random tape and the received data externally.

Suppose that the ℓ-th attack is Type-I. Then $b_\ell = 1$ if and only if $g_{T'_{j_\ell}}(Y^n_{j_\ell}) \in \mathcal{C}_{K_1T'_{j_\ell}}$ for $T'_{j_\ell} = h_{K_0}(M'_{j_\ell})$. Define event $T'_{j_\ell} = T_{j_\ell} (:= h_{K_0}(M_{j_\ell}))$ by col_ℓ, and event $g_{T'_{j_\ell}}(Y^n_{j_\ell}) \in \mathcal{C}_{K_1T'_{j_\ell}}$ with $T_{j_\ell} \neq T'_{j_\ell}$, by mis_ℓ. Then, we have $P(b_\ell = 1) = P(\mathsf{col}_\ell) + P(\mathsf{mis}_\ell)$.

Suppose that the ℓ-th attack is Type-II. Then $b_\ell = 1$ iff $g_{\hat{T}_\ell}(\hat{Y}_\ell^n) \in \mathcal{C}_{K_1\hat{T}_\ell}$ for $\hat{T}_\ell = h_{K_0}(\hat{M}_\ell)$, where $(\hat{M}_\ell, \hat{Y}_\ell^n)$ is Oscar's output in this attack.

Let L be the upper bound on the attack number of Oscar. then the probability of Oscar success can be denoted by $\Pr\left(\vee_{\ell=1}^{L} b_\ell = 1\right)$.

Since each successful adversary has to experience the first successful attack, we restrict to an adversary who will stop after launching a successful attack. Accordingly, $b_\ell = 1$ means that $b_1 = \cdots = b_{\ell-1} = 0$.

Denote the original authentication game by Γ. Now we modify Γ to Γ' such that in Type-I attack, $b_\ell \stackrel{def}{=} \mathsf{col}_\ell$ (instead of $b_\ell \stackrel{def}{=} \mathsf{col}_\ell \vee \mathsf{mis}_\ell$). Consider an adversary Oscar$'$ for Γ' who simply follows Oscar's actions by setting each (unknown) mis_ℓ as 0 (even if it is 1). The view of Oscar$'$ in Γ' differs from that of Oscar in Γ only if $\mathsf{mis}_\ell = 1$ in Γ' for some ℓ. Thus,

$$P(succ(\Gamma)) \leq P(succ(\Gamma')) + \sum_\ell P(\mathsf{mis}_\ell(\Gamma')). \tag{18}$$

As $P(succ(\Gamma')) \leq \sum_{\ell=1}^{L} P(b_\ell(\Gamma') = 1)$, we only need to bound $P(b_\ell(\Gamma') = 1)$ and $P(\mathsf{mis}_\ell(\Gamma'))$.

Bounding $P(\mathsf{mis}_\ell(\Gamma'))$.

Lemma 6: $P(\mathsf{mis}_\ell(\Gamma')) \leq 2^{-\varsigma n^{\omega}} + \sqrt{2\Delta \ln \frac{|\mathcal{K}_0||\mathcal{K}_1|}{\Delta}} + \Delta$ for a constant $\varsigma > 0$, where $\Delta = \mathsf{D}(K|U_\ell; K)$.

Proof: We first show that $U_\ell M_{j_\ell} \to K_1 T_{j_\ell} \to X^n_{j_\ell} \to Y^n_{j_\ell}$ forms a Markov chain. Then we have the following two facts:

(a) Given $X^n_{j_\ell}$, $Y^n_{j_\ell}$ is completely determined by the noise in channel W_1 while this noise occurs after fixing $(X^n_{j_\ell}, K_1 T_{j_\ell} U_\ell M_{j_\ell})$ and hence is independent of the latter;

(b) Given $K_1 T_{j_\ell}$, $X^n_{j_\ell}$ is determined by the randomness for sampling it from $\mathcal{C}_{K_1 T_{j_\ell}}$, which is independent of $U_\ell M_{j_\ell}$.

By Lemma 1 (property 3) with δ_1 from Lemma 5 and δ_2 from Lemma 4, as well as the fact that $\mathsf{D}(K_0|U_\ell; K_0) \leq \mathsf{D}(K_0K_1|U_\ell; K_0K_1)$ (from triangle inequality), the lemma holds. □

Bounding $P(b_\ell(\Gamma') = 1)$.

Let $\bar{U}_\ell = (V, b_1, b_2, \ldots, b_{\ell-1})$, where $V = Rand_0 M^\nu Z_1^n \cdots Z_\nu^n$. Denote by $\bar{\mathcal{U}}_\ell^0$ the set of possible values for $\bar{U}_\ell$ with $b_1, b_2, \ldots, b_{\ell-1} = 0^{\ell-1}$. Then, $P(b_\ell = 1) = \sum_{u_\ell \in \bar{\mathcal{U}}_\ell^0} P(b_\ell = 1, \bar{U}_\ell = u_\ell)$. For given $V = v$, we take $u_\ell = v\|0^{\ell-1}$ $(\ell = 1, 2, \ldots, L)$.

1) Type-I Attack Case: In this case, $b_\ell = \mathsf{col}_\ell$. As M'_{j_ℓ}, M_{j_ℓ} depend on U_ℓ (part of $\bar{U}_\ell$),

$$\mathcal{E}_{u_\ell} \stackrel{def}{=} \{(k_0, k_1) \in \mathcal{K} : h_{k_0}(M'_j) \neq h_{k_0}(M_j)\} \tag{19}$$

completely depends on $\bar{U}_\ell = u_\ell$. So, from Lemma 3, we have

$$\begin{aligned}\Pr(b_\ell = 1|\bar{U}_\ell = u_\ell) &= P_{K|\bar{U}_\ell = u_\ell}(\mathcal{E}^c_{u_\ell}) \\ &\leq P_K(\mathcal{E}^c_{u_\ell}) + \frac{1}{2}\mathsf{D}(P_{K|\bar{U}_\ell = u_\ell}; P_K) \leq \epsilon + \frac{1}{2}\mathsf{D}(P_{K|\bar{U}_\ell = u_\ell}; P_K).\end{aligned} \tag{20}$$

Averaging over $\bar{U}_\ell$, $\Pr(b_\ell = 1) \leq \epsilon + \frac{1}{2}\mathsf{D}(P_{K|\bar{U}_\ell}; P_K)$.

2) Type-II Attack Case: In this case, given $\bar{U}_\ell = u_\ell$, $(\hat{M}_\ell, \hat{Y}^n_\ell)$ is deterministic in u_ℓ as the view U_ℓ of Oscar$'$ is part of $\bar{U}_\ell$. Since $\mathcal{C}_t$ is a codebook with decoder $g_t(\cdot)$, $g_{\hat{T}_\ell}(\hat{Y}^n_\ell) \in \mathcal{C}_{K_1 \hat{T}_\ell}$ holds for at most one K_1 when K_0 and u_ℓ are fixed. Thus, given $\bar{U}_\ell = u_\ell$, $b_\ell = 1$ holds for at most $|\mathcal{K}_0|$ choices of (K_0, K_1). Let $\widehat{\mathcal{E}}_{u_\ell} = \{(k_0, k_1) : g_{\hat{T}_\ell}(\hat{Y}^n_\ell) = \perp\}$. Then, by Lemma 3, we have

$$\Pr(b_\ell = 1|\bar{U}_\ell = u_\ell) \leq P_{K|\bar{U}_\ell = u_\ell}(\widehat{\mathcal{E}}^c_{u_\ell}) \leq \frac{1}{|\mathcal{K}_1|} + \frac{1}{2}\mathsf{D}(P_{K|\bar{U}_\ell = u_\ell}; P_K). \tag{21}$$

Averaging over $\bar{U}_\ell$, $\Pr(b_\ell = 1) \leq \frac{1}{|\mathcal{K}_1|} + \frac{1}{2}\mathsf{D}(P_{K|\bar{U}_\ell}; P_K)$.

Bounding $\mathsf{D}(P_{K|\bar{U}_\ell}; P_K)$. Given $\bar{U}_\ell = u_\ell = v0^{\ell-1}$, we have $K \in \mathcal{E}_{u_i}$ for any $i < \ell$. Let $\mathcal{K}^\ell_v \stackrel{def}{=} \cap_{i=1}^{\ell-1} \mathcal{E}_{u_i}$. In Type-I attack, b_ℓ in Γ' depends on $(K_0, M'_{j_\ell}, M_{j_\ell})$, which further depends on $(K_0, V_{j_\ell}, b_1, \ldots, b_{\ell-1})$. In Type-II attack, b_ℓ in Γ' depends on $(K, \hat{M}_\ell, \hat{Y}_\ell)$, which is further depends on $(K, V_{j_\ell}, b_1, \ldots, b_{\ell-1})$. It follows that $(b_1, \ldots, b_\ell)$ is deterministic in (K, V). As $\bar{U}_\ell = (V, b_1, b_2, \ldots, b_{\ell-1})$, according to the Bayes rule $P_{AB} = P_A P_{B|A}$, it holds that $P_{K\bar{U}_\ell}(k, u_\ell) = P_{KV}(k, v)$ if $(b_1, b_2, \ldots, b_{\ell-1})$ determined by (k, v) is $0^{\ell-1}$; and 0, otherwise. Note $\mathcal{K}^\ell_v$ is the set of all possible k such that $(b_1, \ldots, b_{\ell-1})$ determined by (k, v) is $0^{\ell-1}$. Thus,

$$P_{\bar{U}_\ell}(u_\ell) = \sum_{k \in \mathcal{K}^\ell_v} P_{KV}(k, v) = P_{KV}(\mathcal{K}^\ell_v, v). \tag{22}$$

Therefore, denoting $\epsilon' = \max(\epsilon, \frac{1}{|\mathcal{K}_1|})$, we have

$$\begin{aligned}&\mathsf{D}(P_{K|\bar{U}_\ell}; P_K) \\ &\quad= \sum_v \sum_{k \in \mathcal{K}^\ell_v} |P_{KV}(k, v) - P_{KV}(\mathcal{K}^\ell_v, v) P_K(k)| \\ &\qquad+ \sum_v \sum_{k \notin \mathcal{K}^\ell_v} |P_{KV}(\mathcal{K}^\ell_v, v) P_K(k)| \\ &\quad\leq \mathsf{D}(K|V; K) + 2\sum_v P_{KV}(\mathcal{K} \backslash \mathcal{K}^\ell_v, v) \\ &\quad\leq 2\mathsf{D}(K|V; K) + 2\sum_v P_K(\mathcal{K} \backslash \mathcal{K}^\ell_v) P_V(v) \text{ (Lemma 3)} \\ &\quad\leq 2\mathsf{D}(K|V; K) + 2(\ell - 1)\epsilon'.\end{aligned} \tag{23}$$

Finalizing the bound on $P(\mathbf{Succ}(\Gamma))$. As U_ℓ is part of $\bar{U}_\ell$, we have $\mathsf{D}(K|U_\ell; K) \leq \mathsf{D}(K|\bar{U}_\ell; K)$. Note that $\mathsf{D}(K|V; K) \leq \sqrt{2\nu \ln 2} \cdot 2^{-n\beta_2/2}$. By Lemma 6 and calculus analysis, there exist $\varsigma' > 0$ and $\omega' < \omega$ such that $P(\mathsf{mis}_\ell(\Gamma'))$ is bounded by

$$2^{-\varsigma' n^{\omega'}} + \sqrt{4(\ell - 1)\epsilon' \ln \frac{|\mathcal{K}_0||\mathcal{K}_1|}{2(\ell - 1)\epsilon'}} + 2(\ell - 1)\epsilon'. \tag{24}$$

Summarizing the bound on $P(b_\ell = 1)$, we have $P(b_\ell = 1) \leq \sqrt{2\nu \ln 2} \cdot 2^{-n\beta_2/2} + \ell\epsilon'$.

As $P(\mathbf{Succ}(\Gamma')) \leq \sum_\ell P(b_\ell(\Gamma') = 1)$ and ν is polynomially bounded, Eq. (18) gives

$$\begin{aligned}P(\mathbf{Succ}(\Gamma)) &\leq \sum_\ell P(\mathsf{mis}_\ell(\Gamma')) + \sum_\ell P(b_\ell(\Gamma') = 1) \\ &\leq 2^{-\varsigma'' n^{\omega'}} + \sum_{\ell=1}^{L} \left(\sqrt{4(\ell - 1)\epsilon' \ln \frac{|\mathcal{K}_0||\mathcal{K}_1|}{2(\ell - 1)\epsilon'}} + 3\ell\epsilon'\right) \\ &\leq 2^{-\varsigma'' n^{\omega'}} + 2L\sqrt{L\epsilon' \ln \frac{|\mathcal{K}_0||\mathcal{K}_1|}{\epsilon'}} + 3L^2\epsilon',\end{aligned}$$

for some $\varsigma'' > 0$. Since ϵ' is negligible and L is polynomial in n, $P(\mathbf{Succ}(\Gamma))$ is negligible. Hence the authentication property is satisfied.

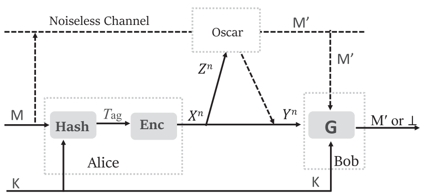

Fig. 3. The specific authentication model.

Finally, we prove that the key security property is satisfied. From Eq. (17) and (23), we have

$$\mathsf{D}(P_{K|\bar{U}_L}; P_K) \leq 2\sqrt{2j\ln 2} \cdot 2^{-n\beta_2/2} + 2(L-1)\epsilon',$$

in which, $\epsilon' = \max(\frac{1}{|\mathcal{K}_1|}, \epsilon)$, L is the upper bound on the attack number of Oscar, and $\bar{U}_L = View(Oscar)$. Hence, $\mathsf{D}(P_{K|\bar{U}_L}; P_K)$ is negligible. By [28, Lemma 1], we have

$$I(K; \bar{U}_L) \leq \mathsf{D}(P_{K|\bar{U}_L}; P_K)\frac{|\mathcal{X}|}{\log[\mathsf{D}(P_{K|\bar{U}_L}; P_K)]}.$$

Thus, $I(K; View(Oscar))$ is arbitrarily small. This completes the proof for the theorem. ■

VII. Message Authentication Channel Capacity

In this section, we discuss the efficiency analysis of a secure MAC scheme. We first focus on minimizing the usage of wiretap channel in a specific authentication syntax model. Then, we compare the authentication rate of the proposed secure MAC scheme with that of a natural scheme.

A. Authentication Channel Capacity

Since the information transmission over a wiretap channel is every expensive, we aim to minimize the usage of it. To this end, a specific authentication syntax model (F, G) with $F = (Id_{\mathcal{M}}, Enc \circ Hash)$ is studied, where $F(M, K) = (M, Enc(Hash(M, K)))$, $Id_{\mathcal{M}} : \mathcal{M} \to \mathcal{M}$ is a identity map, $Hash : \mathcal{M} \times \mathcal{K} \to \mathcal{T}_{ag}$ is a keying hash function, and $Enc : \mathcal{T}_{ag} \to \mathcal{X}^n$ is an encoder. Note that, the hash function $Hash$ can be regarded as hashing family $\{Hash_k : \mathcal{M} \to \mathcal{T}_{ag}\}_{k\in\mathcal{K}}$. In order to analyze the efficiency of authentication, it is assumed that hashing family $\{Hash_k\}_{k\in\mathcal{K}}$ is 1-*universal*.

The specific model is shown in Fig. 3. Concretely, if Alice wants to authenticate M to Bob with a pre-shared key K, she computes the hash value $T_{ag} = Hash(M, K)$, and then encodes T_{ag} to X^n with encoder Enc. Finally, Alice transmits M and X^n to Bob over a noiseless channel and (W_1, W_2), respectively.

Definition 5: For a secure MAC scheme (F, G) with $F = (Id_{\mathcal{M}}, Enc \circ Hash)$, the authentication rate $\rho_{\mathbf{auth}}$ can be rewritten as $\rho_{\mathbf{auth}} = \rho_{\mathbf{tag}} \cdot \rho_{\mathbf{chan}}$, where $\rho_{\mathbf{tag}} = \frac{\log|\mathcal{M}|}{\log|\mathcal{T}_{ag}|}$ and $\rho_{\mathbf{chan}} = \frac{\log|\mathcal{T}_{ag}|}{n}$. $\rho_{\mathbf{tag}}$ is called *tag rate* and $\rho_{\mathbf{chan}}$ is called *authentication channel coding rate*. The *authentication channel capacity* of (W_1, W_2) is defined as $C_{\mathbf{auth}} \overset{def}{=} sup\{\rho_{\mathbf{chan}} : \rho_{\mathbf{chan}}$ is an achievable authentication channel coding rate$\}$.

In order to achieve a higher authentication rate, we need to improve both the authentication channel rate and the tag rate. As tag rate $\rho_{\mathbf{tag}}$ mainly depends on traditional cryptographic techniques (which is out of the scope of this work), we mainly focus on authentication channel coding rate $\rho_{\mathbf{chan}}$. The single-letter representation of the authentication channel capacity is presented in the following theorem.

Theorem 2: If W_1 is less noisy than W_2, then the authentication channel capacity of (W_1, W_2)

$$C_{\mathbf{auth}} = \max_{U\to X\to YZ} H(U|Z). \tag{25}$$

Proof: We first prove, for any $\delta > 0$, if $U \to X \to YZ$, the authentication channel coding rate $H(U|Z) - \delta$ is achievable.

Suppose that (W_1, W_2) is a wiretap channel with $I(X; Y) \geq I(X; Z)$. Let $K = (K_0, K_1)$ be the secret key, and $h : \mathcal{M} \times \mathcal{K}_0 \to \mathcal{T}$ be the ϵ-ASU hash function in the proposed secure MAC scheme (i.e., Scheme 1 satisfies the security requirements in Th. 1). Note that Scheme 1 belongs to the specific authentication model by taking $Hash(K, M) = (h_{K_0}(M), K_1)$ (i.e., $\mathcal{T}_{ag} = \mathcal{T} \times \mathcal{K}_1$).

In our secure MAC scheme construction, $\mathcal{T} \subset [\mathbb{J}]$. The constraint for $\mathbb{J}$ is $\frac{\log\mathbb{J}}{n} < H(X|Y) + \tau$ (Lemma 1), in which, τ only need to meet the following constraint: $H(X|Z) > H(X|Y) + \tau$ (by Lemma 1 and Th. 1). Therefore, for $\forall\delta \in (0, H(X|Z) - H(X|Y))$, defining $\tau = H(X|Z) - H(X|Y) - \delta/2$ and let $|\mathcal{T}| = \mathbb{J} = 2^{n(H(X|Z) - \frac{2}{3}\delta)}$, we have $\rho_{\mathbf{auth}} = \frac{1}{n}[\log|\mathcal{T}| + \log|\mathcal{K}_1|] = H(X|Z) - \frac{2}{3}\delta + \frac{1}{n}\Omega(\log n) \geq H(X|Y) - \delta$, $\forall\delta \in (0, H(X|Z) - H(X|Y))$, when n is large enough. Here $|\mathcal{K}_1| = 2^{\Omega(\log n)}$ (please refer to Th. 1). Thus, the authentication channel coding rate $H(X|Y) - \delta$ is achievable when $I(X; Y) \geq I(X; Z)$.

Let $U \to X \to YZ$ be a Markov chain with $I(U; Z) > I(U; Y)$. Similar to the discussion after Th. 1, we can design a secure MAC scheme Π for (W_1, W_2), such that the authentication channel coding rate of Π is $H(U|Z) - \delta$ (for any $\delta > 0$). Thus, if W_1 is *less noisy than* W_2, then the authentication channel coding rate $H(U|Z) - \delta$ is achievable for any RV U with $U \to X \to YZ$ forming a Markov chain.

Then, we prove that for any achievable rate $\rho_{\mathbf{auth}}$ it must be satisfied that $\rho_{\mathbf{auth}} \leq \max_{U\to X\to YZ} H(U|Z)$. Let $(W_1' : \mathcal{U} \to \mathcal{Y}, W_2' : \mathcal{U} \to \mathcal{Z})$ be a virtual wiretap channel, where $W_1' = P_{Y|U}$, $W_2' = P_{Z|U}$ and $U \to X \to YZ$ forms a Markov chain. Let (F', G') be a secure MAC scheme over (W_1', W_2') with $F' = (Id_{\mathcal{M}}, Enc\circ Hash)$. Then, for any $\varepsilon > 0$, the authentication channel coding rate $\rho_{\mathbf{auth}}(F', G')$ can be bounded as follows.

$$\begin{aligned}
\rho_{\mathbf{auth}}(F', G') &= \frac{1}{n}\log|\mathcal{T}_{ag}| = \frac{1}{n}H(T_{ag}) \\
&= \frac{1}{n}\big[H(T_{ag}|Z^n, M) + I(T_{ag}; Z^n, M)\big] \\
&\leq \frac{1}{n}\big[H(T_{ag}|Z^n) + I(T_{ag}; Z^n, M)\big] \\
&\leq \frac{1}{n}\big[H(U^n|Z^n) + I(T_{ag}; Z^n, M)\big] \\
&= \frac{1}{n}\big[H(U^n|Z^n) + I(T_{ag}; Z^n|M)\big] \\
&\quad \text{(as } \{Hash_k\}_k \text{ is 1-}universal) \\
&= H(U|Z) + \frac{1}{n}[H(Z^n|M) - H(Z^n|M, T_{ag})\big] \\
&\leq H(U|Z) + \frac{1}{n}I(K; Z^n|M) \leq H(U|Z)
\end{aligned}$$

$$+\frac{1}{n}I(K;Z^n,M)$$
$$\leq H(U|Z)+\frac{1}{n}I(K;View(Oscar))$$
$$\leq H(U|Z)+\varepsilon,$$

when n is large enough.

Note that, any secure MAC scheme (F,G) over (W_1,W_2) with $F=(Id_{\mathcal{M}},Enc\circ Hash)$ can be regarded as a secure MAC scheme (F',G') over (W_1',W_2') with $U=X$, $F'=(Id_{\mathcal{M}},Enc'\circ Hash)$ and $G'=G$, where $Enc'=Id_{\mathcal{X}^n}\circ Enc$ and $Id_{\mathcal{X}^n}:\mathcal{X}^n\to\mathcal{X}^n$ is a identity map. Thus, $\rho_{\mathbf{auth}}(F,G)\leq\max_{U\to X\to YZ}H(U|Z)$. This completes the proof. ■

Similar to the discussion after Th. 1, the assumption that $I(X;Y)\geq I(X;Z)$ can be relaxed to the condition that W_2 is *not less noisy than* W_1.

Theorem 3: If W_2 is not less noisy than W_1, then the authentication channel capacity of (W_1,W_2)

$$C_{\mathbf{auth}}=\max_{U\to X\to YZ}H(U|Z). \tag{26}$$

B. Comparison With a Natural Scheme

A special ϵ-ASU hashing family with a high efficiency will be utilized to design a secure MAC scheme as follows.

Lemma 7: [35] For a prime power q and a integer $s(\geq 1)$, there exists an $\frac{s}{q}$-ASU hashing family from $\mathcal{M}$ to $\mathcal{T}$ with secure key space $\mathcal{K}_0$, in which, $|\mathcal{K}_0|=q^s$, $|\mathcal{M}|=q^{2^s}$, and $|\mathcal{T}|=q$.

To realize the proposed MAC scheme, we only need to specify h_k, $\mathcal{K}_0$, $\mathcal{K}_1$ and τ. By taking $\tau=H(X|Z)-H(X|Y)-\delta/2$, from Th. 2, we have $\rho_{\mathbf{chan}}=H(X|Z)-\delta$. We utilize $\frac{s}{q}$-ASU in Lemma 7 to realize hashing family $\{h_k\}_{k|in\mathcal{K}_0}$, where $q=|\mathcal{T}|=2^{n\cdot\rho_{\mathbf{chan}}}=2^{n(H(X|Z)-\delta)}$, $|\mathcal{K}_0|=q^s$, and $|\mathcal{M}|=q^{2^s}$. Let $|\mathcal{K}_1|=2^{\log^2 n}$. In this case, the requirements in the authentication theorem (i.e., Th. 1) are satisfied when $s<2^{n^{\omega}}$ for some $\omega\in(0,1)$. Accordingly, $\rho_{\mathbf{tag}}=2^s$ and hence $\rho_{\mathbf{auth}}=[H(X|Z)-\delta]\cdot\rho_{\mathbf{tag}}=[H(X|Z)-\delta]2^s$, where $s<2^{n^{\omega}}$ for some $\omega\in(0,1)$. Specifically, if $s=\log n$, then $|\mathcal{M}|=2^{n^2[H(X|Z)-\delta]}$, $|\mathcal{K}_0|=2^{n\log n[H(X|Z)-\delta]}$, and $\rho_{\mathbf{auth}}=n[H(X|Z)–\delta]$. Here, the length of authentication key (K_0,K_1) is $n\log n[H(X|Z)-\delta]+\log^2 n$.

In our scheme, tag T is first calculated; and then is encoded to X^n with the coding scheme in Lemma 1. Similarly, it is easy to construct a natural variant scheme, with the change that: T is encoded at the sender to X^n utilizing the encoder of the secrecy coding scheme presented in [14]; and the receiver just need to obtain T with secure decoder in [14], and then check the consistency between T and M'. since T is fully protected by secure coding, the natural scheme is secure. Denoting the secrecy capacity of (W_1,W_2) ais C_s, the authentication rate $\rho_{\mathbf{auth}}$ of the natural one can be expressed as $\rho_{\mathbf{auth}}=\rho_{\mathbf{tag}}C_s$. From [14], if W_1 is more capable[6] than W_2, then $C_s=H(X|Z)-H(X|Y)$ for some probability distribution P_X on $\mathcal{X}$.

Fig. 4 compares the proposed MAC scheme with the natural scheme in terms of authentication rate. It shows that the authentication rate of the proposed MAC scheme is larger than that of the natural scheme when $0<\delta<H(X|Y)$ (note that

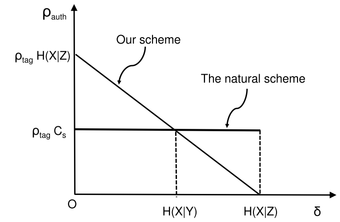

Fig. 4. Comparison between our scheme and the natural scheme in terms of authentication rate, where $C_s=H(X|Z)-H(X|Y)$ and $H(X|Y)>0$.

δ can be arbitrarily small). From that fact that the capacity achieving code T is used in the natural MAC scheme, and the proposed MAC scheme achieve a higher authentication channel rate, we can deduce that it is unnecessary to obtain tag T from X^n, as well as to protect secrecy of T. Actually, the secrecy capacity achievable channel coding of a wiretap channel has two main purposes: (1) the secret information cannot be leaked to the attacker Oscar; and (2) the secret information can be recovered by the receiver Bob. In the proposed MAC scheme, it just need to realize the first purpose but not the second one. This is because: if $M'=M$, then T can be recovered by Bob from M' in the noiseless channel and part of secret key K_0; otherwise, Bob only needs to reject this message. Therefore, in the proposed MAC scheme, it is not necessary for Bob to recover T from Y^n. One exception is that, the rate of these two schemes are identical when W_1 is noiseless (i.e., when $H(X|Y)=0$).

Example 2: Let W_1 and W_2 be the BECs with erasure probability e_0 and e_1, respectively, where $e_1>e_0\geq 0$. Then, from Lemma 1, for any $\mathbb{I}$ and $\mathbb{J}$ with $1\leq|\mathbb{I}|<2^{n(e_1-e_0-\tau)}$ and $1\leq|\mathbb{J}|<2^{n(e_0+\tau)}$, there exists disjoint subsets $\mathcal{C}_{ij}\subset\mathsf{T}_P^n$ $(i\in[\mathbb{I}],\ j\in[\mathbb{J}])$ such that Properties (1)-(3) in Lemma 1 are satisfied, where $0<\tau<e_1-e_0$. By taking $\tau=e_1-e_0-\delta/2$, from the discussion in Section VII-B, the authentication channel coding rate $\rho_{\mathbf{chan}}=\frac{\log|\mathbb{J}|}{n}=e_1-\delta\to e_1$ when $\delta\to 0$. Note that, $C_s=e_1-e_0$. Thus, $\rho_{\mathbf{chan}}>C_s$ if $e_0>0$; and $\rho_{\mathbf{chan}}=C_s$ if $e_0=0$ (i.e., the main channel is noiseless).

VIII. Conclusion

In this paper, we have proposed a novel authentication framework over a wiretap channel for authenticating multiple messages, where Alice sends insecure information S over the noiseless channel and an encoded tag T over the wiretap channel. Traditional authentication framework can be considered as a special case of the proposed framework. Based on this framework, an efficient MAC scheme has been devised for authenticating multiple messages, to realize ITS with the same secret key K. To this end, a novel channel coding has been developed based on random coding techniques. The proposed MAC scheme has been proved to provide information-theoretic security by detailed analysis of adversary model and rigorous mathematical derivation. Moreover, we have also demonstrated that *authentication rate* $\rho_{\mathbf{auth}}=\rho_{\mathbf{tag}}\cdot(H(X|Z)-\delta)$ for any fixed tag rate $\rho_{\mathbf{tag}}$ and any small constant $\delta>0$. Therefore, the proposed MAC scheme can achieve higher efficiency, compared with the state-of-the-art schemes. Furthermore, the authentication channel capacity, which equivalents to minimizing the usage of wiretap channel, has been discussed in a specific authentication syntax model.

[6]Channel W_1 is more capable than channel W_2 means when for every input X, $I(X;Y)>I(X;Z)$.

Theoretical result demonstrate that, if W_1 is less noisy than W_2 then the authentication channel capacity of (W_1, W_2) is $\max_{U\to X\to YZ} H(U|Z)$.

Appendix

In this Appendix, we first give some useful definitions and lemmas. Then, we prove Lemma 1.

A. Preparation

Let X and Y be RVs over sets $\mathcal{X}$ and $\mathcal{Y}$ respectively with a joint distribution P_{XY}. Let (X^n, Y^n) be n independent outputs according to P_{XY}. In this case, (X^n, Y^n) is called a *discrete memoryless multiple source* (DMMS) with variables X and Y. For $\mathcal{A} \subseteq \mathcal{X}^n$, let $\widetilde{P}_{X^nY^n}$ be the joint distribution of (X^n, Y^n) conditional on $\mathcal{A}$. That is, $\widetilde{P}_{X^nY^n}(x^n, y^n) \stackrel{def}{=} P^n_{XY}(x^n, y^n)/P^n_X(\mathcal{A})$ for any $x^n \in \mathcal{A}, y^n \in \mathcal{Y}^n$. Marginal distributions $\widetilde{P}_{X^n}(x^n) = \sum_{y^n\in\mathcal{Y}^n} P^n_{XY}(x^n, y^n)/P^n_X(\mathcal{A}) = P^n_X(x^n)/P^n_X(\mathcal{A})$ and $\widetilde{P}_{Y^n}(y^n) = \sum_{x^n\in\mathcal{A}} P^n_{XY}(x^n, y^n)/P^n_X(\mathcal{A})$.

For any index set $\mathcal{B}$, any collection of disjoint subsets $\{\mathcal{A}_b\}_{b\in\mathcal{B}}$ with $\cup_{b\in\mathcal{B}}\mathcal{A}_b = \mathcal{A}$ forms a partition of $\mathcal{A}$. Certainly, a partition of $\mathcal{A}$ does not depend on the index set $\mathcal{B}$. The generality of $\mathcal{B}$ is only for ease of presentation. For a partition $\{\mathcal{A}_b\}_{b\in\mathcal{B}}$ of $\mathcal{A}$, let $\widetilde{P}_{Y^n|b}(y^n) \stackrel{def}{=} \widetilde{P}(Y^n = y^n|X^n \in \mathcal{A}_b)$. That is,

$$\begin{aligned}\widetilde{P}_{Y^n|b}(y^n) &= \sum_{x^n\in\mathcal{A}_b} \widetilde{P}_{X^nY^n}(x^n, y^n)/\widetilde{P}_{X^n}(\mathcal{A}_b)\\ &= \sum_{x^n\in\mathcal{A}_b} P^n_{XY}(x^n, y^n)/P^n_X(\mathcal{A}_b). \end{aligned} \tag{27}$$

In other words, $\widetilde{P}_{Y^n|b}$ equals the marginal distribution of Y^n in P^n_{XY} conditional on $X^n \in \mathcal{A}_b$.

A partition can also be characterized through a mapping. Specifically, for mapping $\sigma : \mathcal{A} \to \mathcal{B}$, let $\mathcal{A}_b \stackrel{def}{=} \sigma^{-1}(b)$ for $b \in \mathcal{B}$. Then $\{\mathcal{A}_b\}_{b\in\mathcal{B}}$ forms a partition of $\mathcal{A}$. On Moreover, given one partition $\{\mathcal{A}\}_{b\in\mathcal{B}}$, one can define $\sigma : \mathcal{A} \to \mathcal{B}$ by $\sigma(x) = b$ for all $x \in \mathcal{A}_b$. Thus, without further specification, we will simply call a mapping σ *a partition of size* $|\mathcal{B}|$ *for* $\mathcal{A}$.

For any partition $\sigma : \mathcal{A} \to \mathcal{B}$, $\sigma(X^n)$ has a distribution induced by random variable X^n. As $\sigma(x^n) = b$ iff $x^n \in \mathcal{A}_b$, we have $\Pr(\sigma(X^n) = b) = \widetilde{P}_{X^n}(\mathcal{A}_b) = P^n_X(\mathcal{A}_b)/P^n_X(\mathcal{A})$. Thus,

$$\begin{aligned}\mathsf{D}(Y^n|\sigma(X^n); Y^n) &= \sum_{b\in\mathcal{B}} \widetilde{P}_{X^n}(\mathcal{A}_b) \sum_{y^n\in\mathcal{Y}^n} |\widetilde{P}_{Y^n|b}(y^n) - \widetilde{P}_{Y^n}(y^n)|\\ &= \sum_{b\in\mathcal{B}} \widetilde{P}_{X^n}(\mathcal{A}_b)\mathsf{D}(\widetilde{P}_{Y^n|b}; \widetilde{P}_{Y^n}). \end{aligned} \tag{28}$$

Assume that $\mathcal{A} = \mathsf{T}^n_P$, where T^n_P is a typical set with type $P = P_X$ over $\mathcal{X}$. In the following lemma, Csiszár [28] shows that there exists a partition σ that partitions T^n_P into k subsets of almost equal size so that $\sigma(X^n)$ is almost independent of Y^n, when k is not sufficiently large.

Lemma 8: [28] Let $W : \mathcal{X} \to \mathcal{Y}$ be a DMC with input RV X and output RV Y. If X is according to a type P with $P(x) > 0$ $(\forall x \in \mathcal{X})$. Then, $\forall\tau > 0$, there exists a constant $\beta > 0$ such that, when n is large enough and $k \le |\mathsf{T}^n_P|2^{-n(I(X;Y)+\tau)}$, T^n_P has a partition $\sigma : \mathsf{T}^n_P \to \{1, \ldots, k\}$ satisfying

$$|\mathcal{A}_i| = \frac{|\mathsf{T}^n_P|}{k}(1 + \epsilon_i); \quad \mathsf{D}(Y^n|\sigma(X^n); Y^n) < 2^{-n\beta} \tag{29}$$

where $\mathcal{A}_i = \sigma^{-1}(i)$ and $|\epsilon_i| \le 2^{-n\beta}$. Moreover, if σ is uniformly distributed in all possible partitions, then Eq. (29) holds, except for an exponentially small (in n) probability.

Remark: This lemma can be trivially generalized to the setting $\sigma' : \mathsf{T}^n_P \to \mathcal{B}$ with $|\mathcal{B}| = k$ as it does not depend on the choice of $\mathcal{B}$. Specifically, for any σ in the lemma and $\mathcal{B} = \{b_1, \ldots, b_k\}$, define $\sigma' = \sigma \circ \pi$, where mapping $\pi : \mathcal{B} \to \{1, \ldots, k\}$ with $\pi(b_i) \mapsto i$ is one-to-one mapping. With $\mathcal{A}_{b_i} = \mathcal{A}_i$, σ satisfies Eq. (29) if and only if $\sigma \circ \pi$ satisfies the corresponding properties with index set $\mathcal{B}$.

For any set $\mathcal{A}$, there are $k^{|\mathcal{A}|}$ partitions of size k. One can sample a uniformly random partition $\sigma : \mathcal{A} \to \mathcal{B}$ by assigning $\sigma(x)$ to a uniformly random element $b \in \mathcal{B}$ for any $x \in \mathcal{A}$.

B. Useful Lemmas

Now, we present some lemmas that will be used to prove Lemma 1 later. First of all, some basic properties of typical sequences are introduced as follow.

Lemma 9: [32, Chap 1.2] Let X_1, X_2, and X be RVs over $\mathcal{X}$ and Y be an RV over $\mathcal{Y}$. Then,

1. For each type Q of $\mathcal{X}^n$,
$$(n+1)^{-|\mathcal{X}|} \cdot 2^{nH(Q)} \le |\mathsf{T}^n_Q| \le 2^{nH(Q)}. \tag{30}$$
2. There exists a constant $c > 0$ s.t. for $\forall\epsilon > 0$, $\forall x^n \in \mathsf{T}^n_{[X]_\epsilon}$,
$$\begin{aligned}2^{-n[H(X)+c\epsilon]} &\le P^n_X(x^n) \le 2^{-n[H(X)-c\epsilon]},\\ (1-\epsilon)2^{n[H(X)-c\epsilon]} &\le |\mathsf{T}^n_{[X]_\epsilon}| \le 2^{n[H(X)+c\epsilon]}, \end{aligned} \tag{31}$$
in which, inequality (31) holds for a sufficiently large n.
3. There exists a constant $c > 0$ s.t. for $\forall\epsilon > 0$, $\forall x^n \in \mathsf{T}^n_{[X]_\epsilon}$, and $\forall y^n \in \mathsf{T}_{[Y|X]_\epsilon}(x^n)$, we have
$$\begin{aligned}2^{-n[H(Y|X)+c\epsilon]} &\le P^n_{Y|X}(y^n|x^n) \le 2^{-n[H(Y|X)-c\epsilon]},\\ (1-\epsilon)2^{n[H(Y|X)-c\epsilon]} &\le |\mathsf{T}^n_{[Y|X]_\epsilon}(x^n)| \le 2^{n[H(Y|X)+c\epsilon]}, \end{aligned} \tag{32}$$
where inequality (32) holds when n is large enough.
4. There exists constants λ_1 and $\lambda_2 > 0$ such that when n is large enough, for any $x^n \in \mathsf{T}^n_{[X]_\epsilon}$
$$P^n_Y(\mathsf{T}^n_{[Y]_\epsilon}) \ge 1 - 2^{-n\lambda_1\epsilon^2}, \tag{33}$$
$$P^n_{Y|X}(\mathsf{T}^n_{[Y|X]_\epsilon}(x^n)|x^n) \ge 1 - 2^{-n\lambda_2\epsilon^2}. \tag{34}$$

Lemma 10 bounds $E[|\mathsf{T}_{[W]_\epsilon}(Z^n_1) \cap \mathsf{T}_{[W]_\epsilon}(Z^n_1)|]$ for Z^n_1, Z^n_2 chosen from typical set T^n_P uniformly at random. The main idea is that for a random subset B of S, $E(|B|) = \sum_{y\in S} P(y \in B)$. Therefore, we only need to bound

$$\sum_{y^n} P(y^n \in \mathsf{T}_{[W]_\epsilon}(Z^n_1) \cap \mathsf{T}_{[W]_\epsilon}(Z^n_2)). \tag{35}$$

It is easy to find that $y^n \in \mathsf{T}_{[W]_\epsilon}(Z^n)$ for a typical Z^n (Z^n with type P_X satisfies this) implies $Z^n \in \mathsf{T}^n_{[X|Y]_\epsilon}(y^n)$. Thus, Eq. (35) is bounded by $\sum_{y^n} P(Z^n_1, Z^n_2 \in \mathsf{T}^n_{[X|Y]_\epsilon}(y^n))$. Notice that Z^n_1, Z^n_2 are independent and $P(Z^n \in \mathsf{T}^n_{[X|Y]_\epsilon}(y^n)) \approx 2^{-nI(X;Y)}$. Then, the desired bound for Eq. (35) can be obtained.

Lemma 10: Assume RVs X and Y are connected by DMC $W : \mathcal{X} \to \mathcal{Y}$ where $P_X = P$ for some type P. If (Z^n_1, Z^n_2)

are chosen from T_P^n uniformly at random, then there exists a constant $c>0$ s.t. for $\forall\epsilon>0$, when n is large enough,

$$E\Big(|\mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n)|\Big)\le 2^{n[H(Y|X)-I(X;Y)+c\epsilon]}. \quad (36)$$

Proof: For a fixed set S and its random subset $B\subseteq S$, $E(|B|)=E(\sum_{y\in S}\mathbf{1}_B(y))=\sum_{y\in S}P(y\in B)$, where $\mathbf{1}_B(y)=1$ if $y\in B$; and 0 otherwise. Thus,

$$E\Big(|\mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n)|\Big) = \sum_{y^n\in\mathcal{Y}^n} P(y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n)). \quad (37)$$

Since $P_X=P$, $y^n\in \mathsf{T}_{[W]_\epsilon}(x^n)$ $(\forall x^n\in \mathsf{T}^n_P)$ implies $|P_{x^ny^n}(a,b)-P_X(a)P_{Y|X}(b|a)|\le\frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}$ for all a,b. Summation over a yields

$$|P_{y^n}(b)-P_Y(b)|\le\frac{\epsilon}{|\mathcal{Y}|}. \quad (38)$$

This further implies that $|P_{x^ny^n}(a,b)-P_{y^n}(b)P_{X|Y}(a|b)|\le\frac{c'\epsilon}{|\mathcal{X}||\mathcal{Y}|}$ for some constant $c'>0$. So for $x^n\in \mathsf{T}^n_P$, $y^n\in \mathsf{T}_{[W]_\epsilon}(x^n)$ implies $x^n\in \mathsf{T}^n_{[X|Y]_{c'\epsilon}}(y^n)$. It follows that $\{x^n\in \mathsf{T}^n_P: y^n\in \mathsf{T}^n_{[W]_\epsilon}(x^n)\}\subseteq\{x^n\in \mathsf{T}^n_P: x^n\in \mathsf{T}^n_{[X|Y]_{c'\epsilon}}(y^n)\}\subseteq \mathsf{T}^n_{[X|Y]_{c'\epsilon}}(y^n)$, which has a size of at most $2^{n[H(X|Y)+c''\epsilon]}$ for some constant $c''>0$ by Lemma 9 (3). Therefore, Eq. (37) gives

$$\begin{aligned}
&\sum_{y^n\in\mathcal{Y}^n} P(y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n))\\
&\quad=\sum_{y^n\in \mathsf{T}^n_{[Y]_\epsilon}} P(y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n))\text{(by Eq. (38))}\\
&\quad\le\sum_{y^n\in \mathsf{T}^n_{[Y]_\epsilon}} P(Z_1^n,Z_2^n\in \mathsf{T}_{[X|Y]^n_{c'\epsilon}}(y^n))\\
&\quad\overset{*}{\le}\sum_{y^n\in \mathsf{T}^n_{[Y]_\epsilon}} \frac{2^{n[H(X|Y)+c''\epsilon]}}{|\mathsf{T}^n_P|}\times\frac{2^{n[H(X|Y)+c''\epsilon]}}{|\mathsf{T}^n_P|-1}\\
&\quad\le 2(n+1)^{2|\mathcal{X}|}\sum_{y^n\in \mathsf{T}^n_{[Y]_\epsilon}} 2^{-2n[I(X;Y)-c''\epsilon]}\text{(Lemma 9(1))}\\
&\quad\le 2^{-n[I(X;Y)-H(Y|X)-(2c''+c^*+1)\epsilon]}\ \text{(Lemma 9(2))},
\end{aligned}$$

for some $c^*>0$. Ineq (*) holds as Z_1^n,Z_2^n is a uniformly random pair in T^n_P. The lemma holds with $c=2c''+c^*+1$. ■

Lemma 11 essentially states that if we sample a subset A of size at most $2^{n(I(X;Y)-\tau)}$ from T^n_P uniformly at random for some $\tau>0$, then most likely A is an error-correcting code with an exponentially small error. The basic idea is as follows. By the previous lemma, if the sampled set is $\{Z_1^n,\ldots,Z_\ell^n\}$, then $\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)\cap\cup_{j\ne i}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)$ has a size of *roughly* $2^{n(H(Y|X)-\tau)}$, which is an exponentially small part of $\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)$. Thus, A is a codebook under a typical decoding with an exponentially small error.

Lemma 11: Let P be a type over $\mathcal{X}$. Assume integer $\ell\le 2^{n(I(X;Y)-\tau)}$ for a constant $\tau>0$ and RVs X and Y are connected by DMC $W:\mathcal{X}\to\mathcal{Y}$ with $P_X=P$. Let $A:=\{Z_1^n,\ldots,Z_\ell^n\}$ (indexed uniformly at random) be a purely random subset of T^n_P of size ℓ. Let (f,g) be a code with codebook A, where encoding $f:[\ell]\to A$, $f(i)\mapsto Z_i^n$, and decoding $g(Y^n)=i$ if there exists a unique i s.t. $Y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_i^n)$ and $g(Y^n)=\perp$ otherwise. Then, there exist two constants $\lambda>0$ and $\epsilon_0>0$ (not depending on ℓ) s.t., with probability at least $1-2^{-n\tau/2}$ (over the choice of A), we have $e(A)\le 2^{-n\lambda\epsilon^2}$, for any $\epsilon<\epsilon_0$ and when n is large enough.

Proof: We first compute

$$\begin{aligned}
&E(|\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)\cap\cup_{j\ne i}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)|)\\
&\quad\le\sum_{j\in[\ell]\backslash\{i\}} E\Big(|\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_j^n)|\Big)\\
&\quad\le \ell\cdot E\Big(|\mathsf{T}^n_{[W]_\epsilon}(Z_1^n)\cap \mathsf{T}^n_{[W]_\epsilon}(Z_2^n)|\Big)\\
&\qquad\times(Z_1^n,\ldots,Z_\ell^n\text{ are symmetric})\\
&\quad\le \ell\cdot 2^{-n[I(X;Y)-H(Y|X)-c\epsilon]}\quad\text{(by Lemma 36)}\\
&\quad\le 2^{n[H(Y|X)-\tau+c\epsilon]},\quad (n\text{ is large enough})
\end{aligned}$$

for some constant $c>0$. Hence,

$$\sum_{i=1}^n\frac{1}{\ell}E(|\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)\cap\cup_{j\ne i}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)|)\le 2^{n[H(Y|X)-\tau+c\epsilon]}. \quad (39)$$

By Markov inequality, with probability $1-2^{-n\tau/2}$ (over A),

$$\sum_{i=1}^n\frac{1}{\ell}|\mathsf{T}^n_{[W]_\epsilon}(Z_i^n)\cap\cup_{j\ne i}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)|\le 2^{n[H(Y|X)-\frac{\tau}{2}+c\epsilon]}. \quad (40)$$

Denote the collection of such A by $\mathcal{A}$.

By Lemma 9 (3), there exists a constant $\hat{c}>0$ s.t. $P^n_{Y|X}(y^n|x^n)\le 2^{-n[H(Y|X)-\hat{c}\epsilon]}$, $\forall\epsilon>0,\forall x^n\in \mathsf{T}^n_P,\forall y^n\in \mathsf{T}_{[W]_\epsilon}(x^n)$. Thus, there exists a constant $c'>0$ s.t. for any $A\in\mathcal{A}$, when I is uniformly distributed in $[\ell]$,

$$P\Big(Y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_I^n)\cap\cup_{j\ne I}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)\Big)\le 2^{-n(\tau/2-c'\epsilon)}. \quad (41)$$

Note that an error occurs only if $Y^n\in \mathsf{T}^n_{[W]_\epsilon}(Z_I^n)\cap\cup_{j\ne I}\mathsf{T}^n_{[W]_\epsilon}(Z_j^n)$ or if $Y^n\notin \mathsf{T}^n_{[W]_\epsilon}(Z_I^n)$. Thus, by Lemma 9 (4), there exists a constant $\lambda_0>0$ s.t. $e(A)\le 2^{-n(\tau/2-c'\epsilon)}+2^{-n\lambda_0\epsilon^2}$. Lemma follows by the fact that $\lambda<\lambda_0$ and ϵ_0 is small enough (dependent on τ,c',λ_0). ■

The lemma below states that a random subset A of T^n_P with $|A|=\ell$ is uniformly distributed over all subsets with size ℓ.

Lemma 12: For a type P and integer s, a subset $A\subseteq \mathsf{T}^n_P$ is sampled by including each $x^n\in \mathsf{T}^n_P$ with probability $1/s$. Then given $|A|=\ell$, A is uniformly distributed over all possible subsets of T^n_P of size ℓ.

Proof: Let $N=|\mathsf{T}^n_P|$. Then, a particular set A of size ℓ is sampled with probability $s^{-\ell}(1-1/s)^{N-\ell}$, which does not depend on the specific element of A. Therefore, given $|A|=\ell$, A occurs with probability $1/\binom{N}{\ell}$. ■

In the following lemma, we aim to claim that, for a random partition $\mathcal{A}_1,\ldots,\mathcal{A}_s$ of T^n_P, most of $\mathcal{A}_j$'s are codebooks with high probability, when $s<|\mathsf{T}^n_P|2^{-n(I(X;Y)-\theta)}$. The main idea for proof is based on the following fact: if $E(X)\le L$ for $L>0$ and RV X, then $P(X>uL)\le 1/u$ for any $u>0$. Actually, this fact is a simple consequence of Markov inequality.

Lemma 13: Let RVs X,Y be connected by DMC $W:X\to Y$. For a type P and $s=|\mathsf{T}^n_P|2^{-n(I(X;Y)-\theta)}$, $\mathcal{A}_1,\ldots,\mathcal{A}_s$ is a random partition of T^n_P: for each $x^n\in \mathsf{T}^n_P$, take a uniformly random $i\in[s]$ and put x^n into $\mathcal{A}_i$. Regard $\mathcal{A}_j$

with $|\mathcal{A}_j| \le 2^{I(X;Y)-\theta/2}$ as a codebook in Lemma 11 and $\mathcal{A}_j$ with $|\mathcal{A}_j| > 2^{I(X;Y)-\theta/2}$ as a codebook with decoding error $e(\mathcal{A}_j) > 2^{-n\lambda\epsilon^2}$. Then, there exist constants $\lambda > 0, \epsilon_0 > 0$ such that, with probability $1-2^{-n\theta/8+1}$ (over the randomness of partition), there are at most $2^{-n\theta/8}s$ possible j's with $e(\mathcal{A}_j) > 2^{-n\lambda\epsilon^2}$, for any $\epsilon < \epsilon_0$.

Proof: By Lemma 12, given $|\mathcal{A}_j| = \ell$, $\mathcal{A}_j$ is uniformly distributed over all possible subsets of T_P^n of size ℓ. Then, by Lemma 11, given $|\mathcal{A}_j| = \ell \le 2^{I(X;Y)-\theta/2}$, there exist constants $\lambda > 0$ and $\epsilon_0 > 0$ (not depending on ℓ) such that, with probability $1-2^{-n\theta/4}$, $\mathcal{A}_j$ is a codebook with

$$e(\mathcal{A}_j) \le 2^{-n\lambda\epsilon^2}, \tag{42}$$

for any $\epsilon < \epsilon_0$. By symmetry of $\mathcal{A}_1, \ldots, \mathcal{A}_s$, we have that ϵ_0 and λ are invariant with j. On the other hand, as $E(|\mathcal{A}_j|) = |\mathsf{T}_P^n|/s = 2^{n(I(X;Y)-\theta)}$, from Markov inequality, $P(|\mathcal{A}_j| > 2^{n(I(X;Y)-\frac{\theta}{2})}) \le 2^{-n\theta/2}$.

Define Boolean function $F(\mathcal{A}_j) = 1$ if and only if either $\mathcal{A}_j$ violates Eq. (42) or $|\mathcal{A}_j| > 2^{n(I(X;Y)-\frac{\theta}{2})}$. In other words, $F(\mathcal{A}_j) = 1$ if and only if $e(\mathcal{A}_j) > 2^{-n\lambda\epsilon^2}$. Then, $P(F(\mathcal{A}_j) = 1) < 2^{-n\theta/4+1}$. Thus, $E\left(\frac{1}{s}\sum_{j=1}^{s} F(\mathcal{A}_j)\right) \le 2^{-n\theta/4+1}$.

Thus, by Markov inequality, $P\left(\frac{1}{s}\sum_{j=1}^{s} F(\mathcal{A}_j) > 2^{-n\theta/8}\right) \le 2^{-n\theta/8+1}$. That is, with probability $1-2^{-n\theta/8+1}$ (over the randomness of a partition), $\frac{1}{s}\sum_{j=1}^{s} F(\mathcal{A}_j) \le 2^{-n\theta/8}$. In other words, with probability $1-2^{-n\theta/8+1}$, there are at most $2^{-n\theta/8}s$ possible j's with $e(\mathcal{A}_j) > 2^{-n\lambda\epsilon^2}$ (i.e., $F(\mathcal{A}_j) = 1$). ■

C. Proof of Lemma 1

Proof: Part I (for properties 1-2). Since $P_X = P$, we have $P_{XY}(x,y) = P(x)W_1(y|x)$ and $P_{XZ}(x,z) = P(x)W_2(z|x)$. Let $\widetilde{P}_{X^nZ^n}(x^n,z^n) \overset{def}{=} P_{XZ}^n(x^n,z^n)/P_X^n(\mathsf{T}_P^n)$ for $x^n \in \mathsf{T}_P^n$ and $z^n \in \mathcal{Z}^n$. Then, its marginal distribution $\widetilde{P}_{X^n}(x^n) = \frac{1}{|\mathsf{T}_P^n|}$ for $x^n \in \mathsf{T}_P^n$.

For any $\theta \in (0,\tau)$, let s_1, s_2 be any integers with $1 \le s_1 \le 2^{n[I(X;Y)-I(X;Z)-\tau]}$ and $s_2 = |\mathsf{T}_P^n| 2^{-n[I(X;Y)-\theta]}$ Consider independent and uniformly random partitions of T_P^n, $\sigma_1 : \mathsf{T}_P^n \to \{1,\ldots,s_1\}$ and $\sigma_2 : \mathsf{T}_P^n \to \{1,\ldots,s_2\}$. Then, $\sigma = (\sigma_1,\sigma_2)$ is a partition of size s_1s_2 for T_P^n. Let $\mathcal{A} = \mathsf{T}_P^n$. By Lemma 8 with Z in the role of Y and $\sigma = (\sigma_1,\sigma_2)$ (hence $\mathcal{B} = [s_1]\times[s_2]$ in the remark after Lemma 8 and $k = s_1s_2 \le |\mathsf{T}_P^n| 2^{-n(I(X;Z)+(\tau-\theta))}$), there exists $n_1 > 0, \alpha_1 > 0$ and $\beta_1 > 0$ so that following results hold with probability $1–2^{-n\alpha_1}$ over σ,

$$|\mathcal{A}_{ij}| = \frac{|\mathsf{T}_P^n|}{s_1s_2}(1+\epsilon_{ij}) \tag{43}$$

$$\mathsf{D}(Z^n|\sigma(X^n); Z^n) < 2^{-n\beta_1} \tag{44}$$

for $n \ge n_1$, where $\mathcal{A}_{ij} = \sigma_1^{-1}(i) \cap \sigma_2^{-1}(j)$ and $|\epsilon_{ij}| \le 2^{-n\beta_1}$.

Let $\mathcal{A}_{\cdot j} = \cup_i \mathcal{A}_{ij}$. Then, $\mathcal{A}_{\cdot j} = \sigma_2^{-1}(j)$ and hence $\{\mathcal{A}_{\cdot j}\}_{j=1}^{s_2}$ is the explicit representation of partition σ_2. By Lemma 13, there exist constants $\lambda > 0$ and $\epsilon_0 > 0$ such that with probability $1-2^{-n\theta/8+1}$ (over σ), there are at most $2^{-n\theta/8}s_2$ possible j's with $e(\mathcal{A}_{\cdot j}) > 2^{-n\lambda\epsilon^2}$, for any $\epsilon < \epsilon_0$.

Define $\mathsf{Bad}(\sigma)$ as the event: under σ, either Eqs. (43) and (44) fails, or $|\{j : e(\mathcal{A}_{\cdot j}) > 2^{-n\lambda\epsilon^2}\}| > s_2 2^{-n\theta/8}$. Then, $\Pr[\mathsf{Bad}(\sigma)] \le 2^{-nc+2}$ for $c = \min(\alpha_1, \theta/8)$. From Eqs. (28), (44) and $\mathcal{B} = [s_1]\times[s_2]$, since

$$\widetilde{P}_{X^n}\left(\sigma_1^{-1}(i) \cap \sigma_2^{-1}(j)\right) = \widetilde{P}_{X^n}\left(\sigma_1^{-1}(i)\right) \cdot \widetilde{P}_{X^n}\left(\sigma_2^{-1}(j)\right) \tag{45}$$

(as σ_1, σ_2 are independent), we have

$$\begin{aligned} &\mathsf{D}(Z^n|\sigma(X^n); Z^n) \\ &= \sum_{i,j} \widetilde{P}_{X^n}(\mathcal{A}_{ij}) \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n}) \\ &= \sum_{j=1}^{s_2} \widetilde{P}_{X^n}(\mathcal{A}_{\cdot j}) \left(\sum_{i=1}^{s_1} \widetilde{P}_{X^n}(\mathcal{A}_{i\cdot}) \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n}) \right) \le 2^{-n\beta_1} \end{aligned} \tag{46}$$

where $\widetilde{P}_{X^n}(\mathcal{A}_{ij}) = \frac{|\mathcal{A}_{ij}|}{|\mathsf{T}_P^n|} = \frac{1}{s_1s_2} + \frac{\epsilon_{ij}}{s_1s_2}$ for $|\epsilon_{ij}| \le 2^{-n\beta_1}$. Let $\bar{\epsilon}_{\cdot j} = \sum_{i=1}^{s_1} \epsilon_{ij}/s_1$. We have $\widetilde{P}_{X^n}(\mathcal{A}_{\cdot j}) = \frac{1}{s_2} + \frac{\bar{\epsilon}_{\cdot j}}{s_2}$ with $|\bar{\epsilon}_{\cdot j}| \le 2^{-n\beta_1}$. Thus, as $\mathsf{D}(Q_1; Q_2) \le 2$ for any distribution Q_1, Q_2, Eq. (46) implies $\frac{1}{s_2}\sum_{j=1}^{s_2}\left(\sum_{i=1}^{s_1} \widetilde{P}_{X^n}(\mathcal{A}_{i\cdot}) \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n})\right) < 2^{-n\beta_1+2}$. Similarly, we obtain

$$\frac{1}{s_1s_2}\sum_{j=1}^{s_2}\left(\sum_{i=1}^{s_1} \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n})\right) < 2^{-n\beta_1+4}. \tag{47}$$

When Eq. (47) holds, Markov inequality implies that the number of j's with

$$\frac{1}{s_1}\sum_{i=1}^{s_1} \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n}) < 2^{-n\beta_1/2+4} \tag{48}$$

is at least $s_2(1-2^{-n\beta_1/2})$.

For any σ with $\neg\mathsf{Bad}$, we already know that Eq. (47) holds and the number of j's with $e(\mathcal{A}_{\cdot j}) > 2^{-n\lambda\epsilon^2}$ is bounded by $s_2 2^{-n\theta/8}$. Hence, if we let $\mathcal{J}'$ be the set of j such that Eq. (48) holds and $e(\mathcal{A}_{\cdot j}) \le 2^{-n\lambda\epsilon^2}$, then for any σ with $\neg\mathsf{Bad}$,

$$|\mathcal{J}'| \ge s_2(1-2^{-n\theta/8}-2^{-n\beta_1/2}). \tag{49}$$

For each $j \in \mathcal{J}'$, make $|\mathcal{A}_{ij}| = \frac{|\mathsf{T}_P^n|}{s_1s_2}(1-2^{-n\beta_1})$ by cutting a *uniformly random* subset of a proper size from $\mathcal{A}_{ij}$. Then, for $j \in \mathcal{J}'$ and $i \in [s_1]$, denote $\mathcal{A}_{ij}$, $\mathcal{A}_{i\cdot}$, and $\mathcal{A}_{\cdot j}$ by $\mathcal{C}_{ij}$, $\mathcal{C}_{i\cdot}$, and $\mathcal{C}_{\cdot j}$, respectively. Let $\mathcal{C} = \cup_{i\in[s_1], j\in\mathcal{J}'} \mathcal{C}_{ij}$. Update $\widetilde{P}_{X^nZ^n}(x^n,z^n) = P_{XZ}^n(x^n,z^n)/P_X^n(\mathcal{A})$ to $\widetilde{P}_{X^nZ^n}(x^n,z^n) = P_{XZ}^n(x^n,z^n)/P_X^n(\mathcal{C})$. and update $\widetilde{P}_{X^n}(x^n)$, $\widetilde{P}_{Z^n}$ correspondingly. Then, we have $\widetilde{P}_{X^n}(\mathcal{C}_{i\cdot}) = 1/s_1$. Note that $\mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n})$ is updated by a multiplicative factor $P_X^n(\mathcal{A})/P_X^n(\mathcal{C})$. Then, Eq. (48) is now updated to

$$\frac{1}{s_1}\sum_{i=1}^{s_1} \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n}) < 2^{-n\beta_1+5} \tag{50}$$

for every $j \in \mathcal{J}'$. Let $\mathcal{J}$ be a *uniformly random* subset of $\mathcal{J}'$ of size $\mathbb{J}$ and let $\mathbb{I} = s_1$. Then, with probability at least $1-2^{-nc+2}$ over σ (i.e., when $\neg\mathsf{Bad}$ occurs), we get $\mathcal{J}$ s.t. (1) for any $j \in \mathcal{J}$, $\mathcal{C}_{\cdot j}$ is the codebook of a channel coding (f_j, g_j) with average error probability at most $2^{-n\lambda\epsilon^2+1}$ as the cutting operation on $\mathcal{A}_{ij}$ can increase the average error probability by at most $\frac{1+2^{-n\beta_1}}{1-2^{-n\beta_1}} < 2$; (2) for any $j \in \mathcal{J}$, $\frac{1}{s_1}\sum_{i=1}^{s_1} \mathsf{D}(\widetilde{P}_{Z^n|(i,j)}; \widetilde{P}_{Z^n}) < 2^{-n\beta_1+5}$. Therefore, for any $P_{IJ} = P_J/s_1$, $\mathsf{D}(\widetilde{P}_{Z^n|(J,I)}; \widetilde{P}_{Z^n}) < 2^{-n\beta}$ for $\beta < \beta_1$ (not depending on P_J) and n large enough.

As $\lim_{n\to\infty}\frac{1}{n}\log(s_2(1-2^{-n\theta/8}-2^{-n\beta_1/2}))=H(X|Y)+\theta$ and θ is arbitrary in $(0,\tau)$, we can define $\mathbb{J}$ to be any value as long as $\frac{1}{n}\log\mathbb{J}<H(X|Y)+\tau$. Thus, $\mathbb{J}$ and $\mathbb{I}$ can take any value in the required condition.

So far we have proved that for $1-2^{-nc+2}$ fraction of σ (denoted by set Good), uniform RV $\mathcal{J}$ from $\mathcal{J}'$ and uniform RV $\mathcal{C}_{ij}$ from $\mathcal{A}_{ij}$ will satisfy properties 1-2. Note the uniformity of $\mathcal{J}$ and $\mathcal{C}_{ij}$ is unnecessary for property 1-2 and it is for the proof of property 3 in the following.

Part II (continue for property 3). We continue to prove property 3, based on set Good, the uniformity of $\mathcal{C}_{ij}, \mathcal{J}$ above and properties 1-2. We will show that for a large fraction of Good, there exists some choice of $\mathcal{J}$ and $\mathcal{C}_{ij}$ (in properties 1-2) that also satisfies property 3.

Let $r=\frac{|\mathsf{T}_P^n|}{s_1s_2}(1-2^{-n\beta_1})$, $\mathcal{C}_{ij}=\{u_1,\ldots,u_r\}$ and $\mathcal{C}_{ij'}=\{v_1,\ldots,v_r\}$ where elements are ordered uniformly at random. Since g_j uses typicality decoding (Lemma 11), for any σ,

$$\begin{aligned}
&P\Big(g_{J'}(\hat{Y}^n)\in\mathcal{C}_{IJ'}\Big)\\
&\le P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)\\
&=\sum_{i,j',j,t}P_{IJJ'\hat{X}^n}(i,j,j',u_t)\\
&\quad\times P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|IJJ'\hat{X}^n=ijj'u_t\Big)\\
&=\sum_{i,j',j,t}P_{IJJ'}(i,j,j')\frac{1}{r}\times P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|\hat{X}^n=u_t\Big)\\
&\quad(J'\to IJ\to\hat{X}^n\to\hat{Y}^n\text{ Markovian assumption})\\
&\le\sum_{i,j',j,t}\frac{P_{JJ'}(j,j')P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|\hat{X}^n=u_t\Big)}{r\mathbb{I}}+\delta_1.
\end{aligned}$$

where the last inequality is from condition (a) in property 3. Furthermore, by condition (c) in property 3, we have

$$\begin{aligned}
&P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)-\delta_1-\delta_2\\
&\qquad\le\sum_{i,j',j,t}\frac{2^{n^\omega}P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|\hat{X}^n=u_t\Big)}{r\mathbb{J}(\mathbb{J}-1)\mathbb{I}}.
\end{aligned}\tag{51}$$

Notice $\sum_{i,j',j,t}\frac{1}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r}P\Big(\tilde{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|\hat{X}^n=u_t\Big)$ equals $P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)$ but with $P_{IJJ'\hat{X}^n}=\frac{1}{r\mathbb{I}\mathbb{J}(\mathbb{J}-1)}$ (i.e., $I,(J,J')$ are uniform RVs and independent of each other, and $\hat{X}^n$ uniformly distributed over $\mathcal{C}_{IJ}$). We now bound $P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)$ under this setting. By Lemma 9 (4), for some $\lambda_1>0$,

$$\begin{aligned}
&P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)-2^{-n\lambda_1\epsilon^2}\\
&\le P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(\hat{X}^n)\cap\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{IJ'})\Big)\\
&=\sum_{i,j',j,t}\frac{P\Big(\hat{Y}^n\in\mathsf{T}_{[W]_\epsilon}(u_t)\cap\mathsf{T}_{[W]_\epsilon}(\mathcal{C}_{ij'})|\hat{X}^n=u_t\Big)}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r}\\
&\le\sum_{i,j',j,t,t'}\frac{\Big|\mathsf{T}_{[W]_\epsilon}(u_t)\cap\mathsf{T}_{[W]_\epsilon}(v_{t'})\Big|}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r2^{n(H(Y|X)-\epsilon)}},
\end{aligned}\tag{52}$$

where $\mathcal{C}_{ij}=\{u_1,\ldots,u_r\}$ and $\mathcal{C}_{ij'}=\{v_1,\ldots,v_r\}$.

Let ξ be the randomness to select $\mathcal{J}$ from $\mathcal{J}'$ and to select $\mathcal{C}_{sd}$ from $\mathcal{A}_{sd}$ for all s,d. Let η be the randomness to order elements in $\mathcal{C}_{sd}$ for all s,d. So far we have assumed ξ, η, and σ are fixed. As $J\neq J'$ (so $u_t\neq v_{t'}$), it is not hard to see that, over the randomness of (ξ,η,σ), RV $(u_t,v_{t'})$ for fixed (t,t') has a probability distance $2^{-n\gamma}$ from a uniformly random pair (U,V) in T_P^n for some constant $\gamma>0$. Therefore,

$$\begin{aligned}
&E\Big(\sum_{i,j',j,t,t'}\frac{\Big|\mathsf{T}_{[W]_\epsilon}(u_t)\cap\mathsf{T}_{[W]_\epsilon}(v_{t'})\Big|}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r2^{n(H(Y|X)-\epsilon)}}\Big)\\
&\le 2^{-n(\gamma-2\epsilon)}+E\Big(\sum_{i,j,t,j',t'}\frac{\Big|\mathsf{T}_{[W]_\epsilon}(U)\cap\mathsf{T}_{[W]_\epsilon}(V)\Big|}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r2^{n(H(Y|X)-\epsilon)}}\Big)\\
&\le 2^{-n\gamma/2}+r2^{-n(I(X;Y)-c'\epsilon)},\quad(c'\text{ constant, Lemma 36})\\
&=2^{-n\gamma/2}+2^{-n(\theta-c'\epsilon)}\le 2^{-n\gamma''+1},
\end{aligned}$$

for $\gamma''<\min\{\gamma/2,\theta/2\}$. So for $1-2^{-n\gamma''/2+1}$ fraction of σ, there exist ξ and η so that

$$\sum_{i,j',j,t,t'}\frac{|\mathsf{T}_{[W]_\epsilon}(u_t)\cap\mathsf{T}_{[W]_\epsilon}(v_{t'})|}{\mathbb{I}\mathbb{J}(\mathbb{J}-1)r2^{n(H(Y|X)-\epsilon)}}\le 2^{-n\gamma''/2}.\tag{53}$$

Denote the set of σ by Good'. For $\sigma\in\mathsf{Good}\cap\mathsf{Good}'$, from Eq. (52) and (53), we have Eq. (51) is bounded by $2^{-n\lambda\epsilon^2+n^\omega}+2^{-n\gamma''/2+n^\omega}$. Hence, property 3 is satisfied if we take $\epsilon=\sqrt[3]{\frac{1}{n^{1-\omega}}}$, as $2^{n^\omega-n\gamma''/4}+2^{-n\lambda_1\epsilon^2+n^\omega}<2^{-n^\omega}$ when n is large enough.

As a summary, for $P(\mathsf{Good}\cap\mathsf{Good}')>1-2^{-nc+2}-2^{-n\gamma''/2+1}$ fraction of σ, properties 1-3 are satisfied. ■

REFERENCES

[1] G. J. Simmons, "Authentication theory/coding theory," in *Proc. CRYPTO* (Lecture Notes in Computer Science), vol. 196. Berlin, Germany: Springer, 1985, pp. 411–431.

[2] C. Lin, D. He, X. Huang, and K. K. R. Choo, "OBFP: Optimized blockchain-based fair payment for outsourcing computations in cloud computing," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3241–3253, 2021.

[3] K. Phalak and A. A. Saki, "Quantum puf for security and trust in quantum computing," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 11, no. 2, pp. 333–342, Jun. 2021.

[4] Z. Sheng, H. D. Tuan, A. A. Nasir, H. V. Poor, and E. Dutkiewicz, "Physical layer security aided wireless interference networks in the presence of strong eavesdropper channels," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3228–3240, 2021.

[5] J. Zhang, H. Du, Q. Sun, B. Ai, and D. W. K. Ng, "Physical layer security enhancement with reconfigurable intelligent surface-aided networks," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3480–3495, 2021.

[6] U. M. Maurer, "Authentication theory and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1350–1356, Jul. 2000.

[7] N. Aldaghri and H. Mahdavifar, "Physical layer secret key generation in static environments," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2692–2705, 2020.

[8] S. Ribouh, K. Phan, A. V. Malawade, Y. Elhillali, A. Rivenq, and M. A. A. Faruque, "Channel state information-based cryptographic key generation for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7496–7507, Dec. 2021.

[9] G. Li, Y. Xu, W. Xu, E. Jorswieck, and A. Hu, "Robust key generation with hardware mismatch for secure MIMO communications," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 5264–5278, 2021.

[10] V. Korzhik *et al.*, "Performance evaluation of keyless authentication based on noisy channel," in *Proc. MMM-ACNS* (Communications in Computer and Information Science), V. Gorodetsky, I. Kotenko, and V. A. Skormin, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 115–126.

[11] D. Chen, S. Jiang, and Z. Qin, "Message authentication code over a wiretap channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2301–2305.

[12] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[13] L. Lai, H. El Gamal, and H. V. Poor, "Authentication over noisy channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 906–916, Feb. 2009.

[14] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[15] A. Bunin, Z. Goldfeld, H. Permuter, S. S. Shitz, P. Cuff, and P. Piantanida, "Key and message semantic-security over state-dependent channels," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1541–1556, 2020.

[16] J. Tang, L. Jiao, K. Zeng, H. Wen, and K.-Y. Qin, "Physical layer secure MIMO communications against eavesdroppers with arbitrary number of antennas," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 466–481, 2021.

[17] Q. Li, C. Li, and J. Lin, "Constant modulus secure beamforming for multicast massive MIMO wiretap channels," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 264–275, 2020.

[18] A. Khisti, "Secret-key agreement over non-coherent block-fading channels with public discussion," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7164–7178, Dec. 2016.

[19] W. K. Harrison, T. Fernandes, M. A. C. Gomes, and J. P. Vilela, "Generating a binary symmetric channel for wiretap codes," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 8, pp. 2128–2138, Aug. 2019.

[20] L. Li, Z. Chen, D. Zhang, and J. Fang, "A full-duplex Bob in the MIMO Gaussian wiretap channel: Scheme and performance," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 107–111, Jan. 2016.

[21] A. Hyadi, Z. Rezki, A. Khisti, and M. S. Alouini, "Secure broadcasting with imperfect channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2215–2230, Mar. 2016.

[22] S.-H. Lee, W. Zhao, and A. Khisti, "Secure degrees of freedom of the Gaussian diamond-wiretap channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 496–508, Jan. 2017.

[23] A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin, "Strong secrecy on the binary erasure wiretap channel using large-girth LDPC codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 585–594, Sep. 2011.

[24] M. R. Cribbs, R. A. Romero, and T. T. Ha, "Alternative codes and phase rotation extensions for alternating space-time coding-based physical layer security," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1123–1143, 2021.

[25] O. Gunlu, P. Trifonov, M. Kim, R. F. Schaefer, and V. Sidorenko, "Privacy, secrecy, and storage with nested randomized polar subcode constructions," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 514–525, Jan. 2022.

[26] N. Xie, J. Chen, and L. Huang, "Physical-layer authentication using multiple channel-based features," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2356–2366, 2021.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[28] I. Csiszár, "Almost independence and secrecy capacity," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 40–47, Jan. 1996.

[29] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

[30] D. Chen, N. Zhang, N. Cheng, K. Zhang, Z. Qin, and X. Shen, "Physical layer based message authentication with secure channel codes," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 5, pp. 1079–1093, Sep. 2020.

[31] D. Chen *et al.*, "An LDPC code based physical layer message authentication scheme with prefect security," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 748–761, Apr. 2018.

[32] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless System*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[33] P. Baracca, N. Laurenti, and S. Tomasin, "Physical layer authentication over MIMO fading wiretap channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2564–2573, Jul. 2012.

[34] A. Ferrante, N. Laurenti, C. Masiero, M. Pavon, and S. Tomasin, "On the achievable error region of physical layer authentication techniques over Rayleigh fading channels," 2013, *arXiv:1303.0707*.

[35] D. R. Stinson, "Universal hashing and authentication codes," *Des., Codes Cryptogr.*, vol. 4, no. 3, pp. 369–380, Jul. 1994.

[36] R. Safavi-Naini and P. R. Wild, "Information theoretic bounds on authentication systems in query model," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2426–2436, Jun. 2008.

[37] S. Jiang, "Keyless authentication in a noisy model," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 1024–1033, Jun. 2014.

[38] 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.

[39] W. Tu and L. Lai, "Keyless authentication and authenticated capacity," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3696–3714, May 2018.

[40] Z. Gu, H. Chen, P. Xu, Y. Li, and B. Vucetic, "Physical layer authentication for non-coherent massive SIMO-enabled industrial IoT communications," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3722–3733, 2020.

Dajiang Chen (Member, IEEE) received the Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC) in 2014. He was a Post-Doctoral Fellow with the Broadband Communications Research (BBCR) Group, Department of Electrical and Computer Engineering, University of Waterloo, Canada, from 2015 to 2017. He is currently an Associate Professor with the School of Information and Software Engineering, UESTC. His current research interests include physical layer security, secure channel coding, and machine learning and its applications in wireless network security and wireless communications.

Shaoquan Jiang received the B.S. and M.S. degrees in mathematics from the University of Science and Technology of China, Hefei, China, in 1996 and 1999, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005. He was a Research Assistant with the Institute of Software, Chinese Academy of Sciences, Beijing, China, from 1999 to 2000, a Faculty Member with the UESTC, Chengdu, China, from 2005 to 2013, and has been a Faculty Member with Mianyang Normal University, Mianyang, China, since 2013. He was a Post-Doctoral Researcher with the University of Calgary, Calgary, AB, Canada, from 2006 to 2008, and a Visiting Research Fellow with Nanyang Technological University, Singapore, from 2008 to 2009. His research interests include public-key-based secure protocols.

Ning Zhang (Senior Member, IEEE) received the B.E. degree from Beijing Jiaotong University in 2007, the M.S. degree from the Beijing University of Posts and Telecommunications in 2010, and the Ph.D. degree from the University of Waterloo, Canada, in 2015. From 2015 to 2017, he was a Post-Doctoral Research Fellow at the University of Waterloo and the University of Toronto, Canada, respectively. Since 2017, he has been an Assistant Professor at Texas A&M University-Corpus Christi, USA. His current research interests include next generation mobile networks, physical layer security, machine learning, and mobile edge computing. He serves/served as an Associate Editor for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE ACCESS, and *IET Communications*, and an Area Editor for *Encyclopedia of Wireless Networks* (Springer) and Cambridge Scholars.

Lei Liu (Member, IEEE) received the B.Eng. degree in electronic information engineering from Zhengzhou University, Zhengzhou, China, in 2010, and the M.Sc. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2013 and 2019, respectively. He is currently a Lecturer with the State Key Laboratory of Integrated Service Networks, Xidian University, and the Xidian Guangzhou Institute of Technology. His research interests include vehicular ad hoc networks, intelligent transportation, mobile-edge computing, and the IoT.

Kim-Kwang Raymond Choo (Senior Member, IEEE) received the Ph.D. degree in information security from the Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship with The University of Texas at San Antonio. He was a recipient of the 2019 IEEE Technical Committee on Scalable Computing Award for Excellence in Scalable Computing (Middle Career Researcher), and Best Paper Awards from IEEE SYSTEMS JOURNAL in 2021, IEEE Computer Society's Bio-Inspired Computing STC Outstanding Paper Award in 2021, IEEE DSC 2021, *IEEE Consumer Electronics Magazine* in 2020, *Journal of Network and Computer Applications* in 2020, *EURASIP Journal on Wireless Communications and Networking* in 2019, IEEE TrustCom 2018, and ESORICS 2015. He is the Founding Co-Editor-in-Chief of *ACM Distributed Ledger Technologies: Research and Practice*, and the Founding Chair of the IEEE TEMS TC on Blockchain and Distributed Ledger Technologies.