Multiuser Authentication with Anonymity Constraints over Noisy Channels

Rémi A. Chou and Aylin Yener
Wireless Communications and Networking Laboratory
Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802

***Abstract*—We consider authentication of messages sent by L legitimate transmitters to a legitimate receiver over a noisy multiple access channel. We assume the presence of a computationally unbounded opponent who has access to noisy observations of the messages transmitted, and can perform impersonation or substitution attacks. In addition, we consider anonymity constraints where the legitimate receiver must be able to authenticate the messages he receives with respect to predetermined groups of transmitters, but must be kept ignorant of the transmitter's identity of a given message in a given group. Our main result is an authentication coding scheme for which asymptotically matching upper and lower bounds on the probability of successful attack are derived. Our result analytically quantifies the impact of a multiuser setting compared to a single-user setting, as well as the negative impact of anonymity constraints on the probability of successful attack.**

I. Introduction

Authentication aims at preventing the receiver of a message to be deceived by the claimed authorship of the message. Simmons' model on authentication [1] considers a transmitter who wishes to communicate a message M over a noiseless channel, such that the legitimate receiver, upon receiving M, can assess the identity of the emitter of M. It is assumed that the legitimate users share a secret key K. It is also assumed that an opponent, whose computational power is unbounded, can initiate one of the following attacks: (i) a substitution attack, for which the opponent intercepts the original message and sends a modified version to the legitimate receiver; or (ii) an impersonation attack, for which the opponent sends a fraudulent message to the receiver before the legitimate transmitter initiates any communication. P_I and P_S, the probabilities of successful impersonation and substitution attacks, respectively, are lower bounded by $P_I \geqslant 2^{I(K;M)}$ and $P_S \geqslant 2^{-H(K|M)}$ in [1]. See also [2], in which a simplified and unified proof of several related results via hypothesis testing is proposed.

The fact that practical communication systems are noisy, motivates the investigation of authentication over noisy channels. A first strategy in such a situation is to transform the noisy channel in a noiseless one via channel coding. For instance, [3] and [4] consider the effect of imperfect channel coding on authentication. Instead of decoupling the problem of authentication over noisy channel in channel coding and authentication over noiseless channel, [5] combines both tasks to take *advantage* of the channel noise. Specifically, [5] provides, in the case of a noisy communication channel, the lower bound $\max(P_I, P_S) \geqslant 2^{-H(K)}$, as well as an asymptotically matching upper bound for a coding scheme based on wiretap codes [6] under strong secrecy.

In this paper, we build upon and extend to the multiuser setting the approach put forward in [5]. In particular, we consider an authentication problem for *multiple legitimate transmitters* that communicate *strongly secure* messages to a legitimate receiver over a noisy multiple access channel in the presence of an opponent. As in [1], [2], [5], we consider impersonation and substitution attacks by the opponent. Additionally, we consider *anonymity constraints* that address, for instance, a setting in which anonymity of each transmitter is required, i.e., a legitimate receiver must identify the group of nodes that transmit the messages it decodes, and simultaneously should be kept uninformed of the identity of the transmitter of a particular message. This could correspond to a secret ballot, or an anonymous review. Another setting that we address is when anonymity among groups of transmitters is required, as, for example, in the case of several groups of people involved in a clinical trial. Specifically, each group is assigned a different drug. At the end of the trial, each participant submits a report of its experience to the principal investigator. Upon receiving all the reports, the latter must identify the group $\mathcal{G}$ associated with a particular report but the identity of the person who wrote the report should be kept anonymous among the group $\mathcal{G}$.

Our contributions are fourfold. (i) We quantify the impact of the multi-transmitter setting on the probability of successful attack compared to the single-user case in [5]. (ii) We provide an information-theoretic metric to assess anonymity. (iii) We quantify the negative impact of anonymity constraints on the probability of successful attack. (iv) We provide an achievable scheme relying on MAC-WT codes for strong secrecy, and prove its asymptotic optimality in terms of probability of successful attack.

The remainder of the paper is organized as follows. We describe the model under consideration in Section II. We propose a coding scheme in Section III, and analyze its probability of successful attack in Section IV. Finally, we

This work was supported in part by NSF grants CIF-1319338 and CNS-1314719.

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

provide concluding remarks in Section V.

II. Problem statement

We first introduce the following notation. We define $[\![x, y]\!]$ as the integers between $\lfloor x \rfloor$ and $\lceil y \rceil$. We denote the modulo-2 addition by $\oplus$. Finally, we denote the indicator function by $\mathbb{1}\{\omega\}$, which is equal to 1 if the predicate ω is true and 0 otherwise.

The problem statement is as follows. Let $L \in \mathbb{N}^*$ and consider a set $\mathcal{L} \triangleq [\![1, L]\!]$ of L transmitters, a legitimate receiver, and an opponent. Let $Q \in \mathbb{N}^*$, the transmitters form groups according to a partition $\mathcal{P} \triangleq \{\mathcal{G}_q\}_{q\in\mathcal{Q}}$ of $\mathcal{L}$, where $\mathcal{Q} \triangleq [\![1, Q]\!]$. The transmitters wish to send messages to the legitimate receiver over B encoding blocks of length N. For $b \in \mathcal{B} \triangleq [\![1, B]\!]$, for $l \in \mathcal{L}$, we denote the message that Transmitter l wishes to transmit to the legitimate receiver over Block b by $M_{l,b}$. We define $M_{\mathcal{L},b} \triangleq (M_{l,b})_{l\in\mathcal{L}}$ and $M_{\mathcal{L}}^B = (M_{\mathcal{L},b})_{b\in\mathcal{B}}$. It is assumed that the messages are independent across transmitters, i.e., the L sequences $(M_{l,b})_{b\in\mathcal{B}}$, $l \in \mathcal{L}$, are mutually independent. Transmission is over a discrete memoryless multiple access channel $\left(\mathcal{X}_{\mathcal{L}}, W_{YZ|X_{\mathcal{L}}}, \mathcal{Y} \times \mathcal{Z}\right)$, where $\mathcal{X}_{\mathcal{L}}$, $\mathcal{Y}$, and $\mathcal{Z}$ are finite alphabets and $X_{\mathcal{L}} \triangleq (X_l)_{l\in\mathcal{L}}$. We assume that the legitimate receiver and the opponent are both unaware of how the L inputs of the channel are allocated to the different transmitters in $\mathcal{L}$. For $b \in \mathcal{B}$, we denote by $\mathbf{Y}_b$ and $\mathbf{Z}_b$ the outputs of the channel observed by the legitimate receiver and the opponent in Block b, respectively, when $\mathbf{X}_{\mathcal{L},b} \triangleq (\mathbf{X}_{l,b})_{l\in\mathcal{L}}$, is the input of the channel with $\mathbf{X}_{l,b}$ a codeword of length N emitted by Transmitter $l \in \mathcal{L}$. We also define $\mathbf{X}_{\mathcal{L}}^B \triangleq (\mathbf{X}_{\mathcal{L},b})_{b\in\mathcal{B}}$, $\mathbf{Y}^B \triangleq (\mathbf{Y}_b)_{b\in\mathcal{B}}$, and $\mathbf{Z}^B \triangleq (\mathbf{Z}_b)_{b\in\mathcal{B}}$.

For the purpose of authentication, we consider that each transmitter shares with the legitimate receiver a secret key, which is uniformly distributed over the alphabet $\mathcal{K} \triangleq \{0, 1\}^{\log|\mathcal{K}|}$. Let K_l denote the key shared between Transmitter $l \in \mathcal{L}$ and the legitimate receiver. The keys K_l's are assumed distinct, mutually independent, and private in the sense that Transmitter l is not willing to share K_l with any other transmitter. The objectives of the authentication scheme are as follows.

Reliability: $M_{\mathcal{L}}^B$ must be reconstructed by the legitimate receiver, i.e.,

$$\lim_{N\to\infty} \mathbb{P}\left[\widehat{M}_{\mathcal{L}}^B \neq M_{\mathcal{L}}^B\right] = 0, \tag{1}$$

where $\widehat{M}_{\mathcal{L}}^B$ denotes the decoded messages at the legitimate receiver.

Strong Secrecy: $M_{\mathcal{L}}^B$ must be kept secret from the opponent in the sense

$$\lim_{N\to\infty} I\left(M_{\mathcal{L}}^B; \mathbf{Z}^B\right) = 0. \tag{2}$$

Authentication: Upon forming $\widehat{M}_{\mathcal{L},b}$, $b \in \mathcal{B}$, the legitimate receiver must be able to uniquely identify the group of transmitters $\mathcal{G}_q$, from which the message $\widehat{M}_{l,b}$, $l \in \mathcal{L}$, has been transmitted. If at least one message cannot be identified, then all messages are rejected. An extension, not reported here due to space constraints, is to consider a threshold $\tau \in [\![2, Q]\!]$ on the number of groups of messages correctly identified by the receiver, to decide whether to authenticate or not the messages correctly identified; see also Section V.

Anonymity: For all $b \in \mathcal{B}$, for all $q \in \mathcal{Q}$, we require the anonymity of the transmitter in the group $\mathcal{G}_q$, i.e., for any $l \in \mathcal{L}$, the legitimate receiver can determine from which set of transmitters $\mathcal{G}_q$, the message $\widehat{M}_{l,b}$ has been sent but must not be able to determine from which transmitter in $\mathcal{G}_q$, $\widehat{M}_{l,b}$ has been sent. We thus require the observations of the receiver, $\mathbf{Y}^B$, to be independent of $K_{\mathcal{G}} \triangleq (K_l)_{l\in\mathcal{G}}$ for any $\mathcal{G} \subsetneq \mathcal{G}_q$ to ensure anonymity in the group of transmitters $\mathcal{G}_q$, i.e.,

$$\forall q \in \mathcal{Q}, \quad \max_{\substack{\mathcal{G}\subset\mathcal{G}_q \\ \text{s.t. } |\mathcal{G}|=|\mathcal{G}_q|-1}} I\left(\mathbf{Y}^B; K_{\mathcal{G}}\right) = 0. \tag{3}$$

The intuition behind (3) is as follows. Assume that Transmitter l is associated with the input l' of the multiple access channel, $l, l' \in \mathcal{L}$. Since the keys must be kept private among all transmitters, only Transmitter l has information about K_l. Hence, if $I\left(\mathbf{Y}^B; K_l\right) \neq 0$, then $I\left(\mathbf{X}_{\mathcal{L}}^B; K_l\right) \neq 0$ by the data processing inequality, and $I\left(\mathbf{X}_{l'}^B; K_l\right) \neq 0$ by privacy of the key K_l, which means that the legitimate receiver, who knows K_l, can potentially learn l if he tries to reconstruct $\mathbf{X}_{l'}^B$. Similarly, if $I\left(\mathbf{Y}^B; K_{\mathcal{G}}\right) \neq 0$ for $\mathcal{G} \subsetneq \mathcal{G}_q$, then the receiver can identify the set of messages sent by the group of transmitters $\mathcal{G}$ in $\mathcal{G}_q$.

In addition, we assume that the opponent can choose an arbitrary block $b \in \mathcal{B}$ and initiate one of the following attacks:

Impersonation attack: The opponent sends a fraudulent channel output $\widetilde{\mathbf{Y}}_b$ to the receiver when the transmitters are silent, where $\widetilde{\mathbf{Y}}_b$ is a function of all the observations of the opponent, i.e, $\mathbf{Z}^{b-1} \triangleq (\mathbf{Z}_i)_{i\in[\![1,b-1]\!]}$.

Substitution attack: The opponent blocks the transmission of the b-th block between the transmitters and the legitimate receiver while observing $\mathbf{Z}_b$, and sends a fraudulent channel output, $\widetilde{\mathbf{Y}}_b$, to the legitimate receiver, where $\widetilde{\mathbf{Y}}_b$ is a function of all the observations of the opponent, i.e, $\mathbf{Z}^b \triangleq (\mathbf{Z}_i)_{i\in[\![1,b]\!]}$.

Without loss of generality [5], we assume the channel between the opponent and the legitimate receiver noiseless. The impersonation or substitution attack is successful if the legitimate receiver accepts all the messages as authenticated. We denote probability of successful impersonation and substitution attacks in Block b by $P_{I,b}$ and $P_{S,b}$, respectively, and define the probability of successful attack by the opponent by

$$P_A \triangleq \max_{b\in\mathcal{B}} \max\left(P_{I,b}, P_{S,b}\right).$$

III. Proposed coding scheme

The basis of our coding scheme is a multiple access wiretap code for strong secrecy [7]. Although [7] only considers two inputs, the result generalizes to an arbitrary number of inputs. We refer to this coding scheme as $E^{\text{WT-MAC}}$.

Algorithm 1 describes the pre-authentication exchange among the transmitters that takes place before the communication to the legitimate receivers. In general, as explained in the

following, such cooperation among transmitters is necessary to allow authentication.

Algorithm 1 Cooperation among transmitters

for $q \in \mathcal{Q}$ **do**
  for $l \in \mathcal{L}$ **do**
    Transmitter l draws a binary sequence R_l uniformly distributed over $\mathcal{K}$ and sends $K_l \oplus R_l$ to all transmitters in $\mathcal{G}_q$
    Transmitter l computes $\overline{\Gamma}_q \triangleq \bigoplus_{l' \in \mathcal{G}_q} (K_{l'} \oplus R_{l'})$
  end for
  An arbitrary Transmitter in $\mathcal{G}_q$ sends to all transmitters in $\mathcal{G}_q$ a random permutation π_q over the set $[\![1, |\mathcal{G}_q|]\!]$
end for

Algorithm 2 Encoding at the transmitters

for Block $b \in \mathcal{B}$ **do**
  for $l \in \mathcal{L}$ **do**
    Transmitter l encodes $(M_{l,b}, R_l, \Gamma_l)$, with $E^{\text{WT-MAC}}$
  end for
end for

Algorithm 3 Decoding at the legitimate receiver

for Block $b \in \mathcal{B}$ **do**
  The legitimate receiver forms $\left(\widehat{M}_{l,b}, \widehat{R}_{l,b}, \widehat{\Gamma}_{l,b}\right)_{l \in \mathcal{L}}$, an estimate of $(M_{l,b}, R_l, \Gamma_l)_{l \in \mathcal{L}}$, using the decoder associated with $E^{\text{WT-MAC}}$ in Algorithm 2.
  if for all $q \in \mathcal{Q}$, all the elements of $(\overline{K}_{q,i})_{i \in [\![1, |\mathcal{G}_q|-1]\!]}$ appears exactly once in $(\widehat{\Gamma}^{(2)}_{l,b})_{l \in \mathcal{L}}$ in positions indexed by $(l_{i,q})_{i \in [\![1,|\mathcal{G}_q|]\!]} \in \mathcal{L}^{|\mathcal{G}_q|}$, and

$$S_{q,b} \oplus \left(\underset{i \in [\![1,|\mathcal{G}_q|]\!]}{||} \widehat{\Gamma}^{(2)}_{l_{i,q},b} \right) || \left(\underset{i \in [\![1,|\mathcal{G}_q|]\!]}{||} \widehat{\Gamma}^{(1)}_{l_{i,q},b} \right) = F_q,$$

  where $||$ denotes concatenation, **then**
    The decoder accepts all the messages as authenticated
  else
    The decoder rejects all the messages
  end if
end for

Observe that after the protocol described in Algorithm 1 takes place, no transmitter has leaked any information about his key to the other transmitters. We then define for $q \in \mathcal{Q}$,

$$F_q \triangleq \bigoplus_{l \in \mathcal{S}_q} K_l. \tag{4}$$

F_q is understood as a super key meant to identify the group $\mathcal{G}_q$. We also define for $q \in \mathcal{Q}$, for $i \in [\![1, |\mathcal{G}_q|-1]\!]$

$$\overline{K}_{q,i} \triangleq F_q \left[\left[\!\left[1 + (i-1)\Delta^{(1)}_q, i\Delta^{(1)}_q \right]\!\right] \right],$$
$$\overline{K}_{q,|\mathcal{G}_q|} \triangleq F_q \left[\left[\!\left[1 + (|\mathcal{G}_q|-1)\Delta^{(1)}_q, n^* \right]\!\right] \right],$$

where for any $\mathcal{A} \subseteq [\![1, \log|\mathcal{K}|]\!]$, $F_q[\mathcal{A}]$ denote the bits of F_q in positions indexed by $\mathcal{A}$, $\Delta^{(1)}_q \triangleq \left\lfloor \frac{n^*}{|\mathcal{G}_q|} \right\rfloor$, and n^* is the smallest integer such that the elements of the sequence $(\overline{K}_{q,i})_{q \in \mathcal{Q}, i \in [\![1,|\mathcal{G}_q|-1]\!]}$ are distinct. We assume that the transmitters have access to $(\overline{K}_{q,i})_{q \in \mathcal{Q}, i \in [\![1,|\mathcal{G}_q|-1]\!]}$. It requires additional cooperation, the latter is, however, negligible since n^* is a constant that does not scale with $|\mathcal{K}|$. It also makes sacrifice, in terms of privacy with respect to other transmitters, n^* bits of the individual key of each transmitter. The number of sacrificed bits is, however, constant and does not scale with $|\mathcal{K}|$. Note also that $(\overline{K}_{q,i})_{q \in \mathcal{Q}, i \in [\![1,|\mathcal{G}_q|-1]\!]}$ can be determined by the legitimate receiver.

For any $q \in \mathcal{Q}$, we divide $\overline{\Gamma}_q$ in $|\mathcal{G}_q|$ parts, specifically, we define for $i \in [\![1, |\mathcal{G}_q|-1]\!]$

$$\overline{\Gamma}_{q,i} \triangleq \overline{\Gamma}_q \left[\left[\!\left[n^* + 1 + (i-1)\Delta^{(2)}_q, n^* + i\Delta^{(2)}_q \right]\!\right] \right],$$
$$\overline{\Gamma}_{q,|\mathcal{G}_q|} \triangleq \overline{\Gamma}_q \left[\left[\!\left[n^* + 1 + (|\mathcal{G}_q|-1)\Delta^{(2)}_q, \log|\mathcal{K}| \right]\!\right] \right],$$

where $\Delta^{(2)}_q \triangleq \left\lfloor \frac{\log|\mathcal{K}| - n^*}{|\mathcal{G}_q|} \right\rfloor$.

The transmitters then encode their messages as in Algorithm 2, where we have defined for $l \in \mathcal{L}$, $\Gamma_l \triangleq \left[\Gamma^{(1)}_l, \Gamma^{(2)}_l\right]$ with $\Gamma^{(1)}_l \triangleq \overline{\Gamma}_{q,\pi_q(l)}$, $\Gamma^{(2)}_l \triangleq \overline{K}_{q,\pi_q(l)}$, and with q such that $l \in \mathcal{G}_q$.

Finally, the legitimate receiver decodes its observations as in Algorithm 3, where for $l \in \mathcal{L}$, $\widehat{\Gamma}^{(1)}_l$, $\widehat{\Gamma}^{(2)}_l$ denote the estimate of $\Gamma^{(1)}_l$, $\Gamma^{(2)}_l$, respectively, and for $q \in \mathcal{Q}$, we have defined, provided that the indices $(l_{i,q})_{i \in [\![1,|\mathcal{G}_q|]\!]} \in \mathcal{L}^{|\mathcal{G}_q|}$ exist, $S_{q,b}$ as the sum

$$\bigoplus_{i \in [\![1,|\mathcal{G}_q|]\!]} \widehat{R}_{l_{i,q},b},$$

where the first n^* bits have replaced by zeros.

One can show optimality of the choice of $(F_q)_{q \in \mathcal{Q}}$, as explained in the following, and that (1)–(3) hold. We omit the proofs due to space constraints.

Proposition 1. *To minimize the probability of a successful attack, one should have for all $b \in \mathcal{B}$, for all $q \in \mathcal{Q}$,*

$$\lim_{N \to \infty} I\left(\mathbf{Y}_b; F_q\right) = H\left(F_q\right). \tag{5}$$

A Proof follows from an application of [2, Section IV]. A first consequence that can be shown from (5) and (3) is that $H(F_q) \leqslant \log|\mathcal{K}|$, $q \in \mathcal{Q}$, which shows optimality of our choice of $(F_q)_{q \in \mathcal{Q}}$. A second consequence that can be shown from (5) and (3) is that, in general, the communication rates in Algorithm 1, must be strictly positive, i.e., the transmitters must cooperate. A formal counter-example includes a class of switch channels. Nevertheless, as seen in Remark 1, the total communication rate associated to Algorithm 1 can be chosen vanishing to zero. Note also that cooperation at the transmitters is not needed when $Q = L$.

IV. Characterization of the probability of a successful attack

For $b \in \mathcal{B}$, denote by $\mathcal{C} \triangleq \{c_d\}_{d \in \mathcal{D}}$, where $\mathcal{D} \triangleq [\![1, D]\!]$, the set of lengths taken by the parts of $\mathcal{P}$, i.e., $\mathcal{C} \triangleq \{|\mathcal{G}_q|\}_{q \in \mathcal{Q}}$. For

$b \in \mathcal{B}$, for $d \in \mathcal{D}$, define n_d as the number of parts of $\mathcal{P}$ with length c_d. We thus have $\sum_{d\in\mathcal{D}} n_d c_d = L$ and $\sum_{d\in\mathcal{D}} n_d = Q$.

A. Lower bound on the probability of a successful attack

To lower bound the probability of a successful attack by an opponent, one can consider any strategy and study its probability of success. Assume that for each $d \in \mathcal{D}$, the opponent successively guesses at random with n_d tries the keys associated to the groups of size c_d. We assume that for a given d, the opponent can redraw sequences that he has already drawn for previous d, i.e., the opponent draws with replacement for different d. One can show that the probability of a successful attack, as $|\mathcal{K}|$ goes to infinity, is $\frac{\prod_{d\in\mathcal{D}} n_d!}{|\mathcal{K}|^Q}$.

B. Upper bound on the probability of a successful attack

Consider an arbitrary attack strategy, denoted by e, performed by the opponent. Let $\widetilde{m}_{\mathcal{L},b}(e) \triangleq (\widetilde{m}_{l,b})_{l\in\mathcal{L}}$, $\widetilde{r}_{\mathcal{L},b}(e) \triangleq (\widetilde{r}_{l,b})_{l\in\mathcal{L}}$, $\widetilde{\gamma}_{\mathcal{L},b}(e) \triangleq (\widetilde{\gamma}_{l,b})_{l\in\mathcal{L}} = \left([\widetilde{\gamma}^{(1)}_{l,b}, \widetilde{\gamma}^{(2)}_{l,b}]\right)_{l\in\mathcal{L}}$, be the messages decoded with Algorithm 3 in block b by the legitimate receiver upon receiving $\widetilde{\mathbf{Y}}_b$, and where $\widetilde{\gamma}^{(1)}_{l,b}$, $\widetilde{\gamma}^{(2)}_{l,b}$ corresponds to the estimate of $\gamma^{(1)}_{l,b}$, $\gamma^{(2)}_{l,b}$, respectively.

For $q \in \mathcal{Q}$, if all the elements of $(\overline{k}_{q,i})_{i\in\mathcal{G}_q}$ appears exactly once in $(\widetilde{\gamma}^{(2)}_{l,b})_{l\in\mathcal{L}}$ in positions indexed by $(\widetilde{l}_{i,q})_{i\in[\![1,\mathcal{G}_q]\!]} \in \mathcal{L}^{|\mathcal{G}_q|}$, then define

$$\widetilde{\sigma}_{\widetilde{l}_{i,q},b} \triangleq \widetilde{s}_{q,b} \oplus \left(\underset{i\in[\![1,|\mathcal{G}_q|]\!]}{||} \widetilde{\gamma}^{(2)}_{\widetilde{l}_{i,q},b}\right) || \left(\underset{i\in[\![1,|\mathcal{G}_q|]\!]}{||} \widetilde{\gamma}^{(1)}_{\widetilde{l}_{i,q},b}\right),$$

where $\widetilde{s}_{q,b}$ is defined as the sum

$$\bigoplus_{i\in[\![1,|\mathcal{G}_q|]\!]} \widetilde{r}_{\widetilde{l}_{i,q},b},$$

where the first n^* bits have been replaced by zeros.

Let $m_{\mathcal{L},b} \triangleq (m_{l,b})_{l\in\mathcal{L}}$, $r_{\mathcal{L}} \triangleq (r_l)_{l\in\mathcal{L}}$, $\gamma_{\mathcal{L}} \triangleq (\gamma_l)_{l\in\mathcal{L}}$, be the messages encoded by the legitimate transmitters.

For $q \in \mathcal{Q}$, let $(l_{i,q})_{i\in[\![1,|\mathcal{G}_q|]\!]} \in \mathcal{L}^{|\mathcal{G}_q|}$ be such that $||_{i\in[\![1,|\mathcal{G}_q|]\!]} \gamma^{(2)}_{l_{i,q}} = f_q[[\![1, n^*]\!]]$. We define for $q \in \mathcal{Q}$,

$$\overline{\gamma}_q \triangleq \left(\underset{i\in[\![1,|\mathcal{G}_q|]\!]}{||} \gamma^{(2)}_{l_{i,q}}\right) || \left(\underset{i\in[\![1,|\mathcal{G}_q|]\!]}{||} \gamma^{(1)}_{l_{i,q}}\right), \tag{6}$$

$$\begin{aligned}\overline{\sigma}_q &\triangleq \overline{\gamma}_q \oplus \bigoplus_{l\in\mathcal{G}_q} r_l \\ &= f_q.\end{aligned} \tag{7}$$

The opponent chooses his strategy to maximize his success, given its observations $\mathbf{Z}^{b-1}$ for an impersonation attack, or given its observations $\mathbf{Z}^b$ for a substitution attack. Hence, averaging over the opponent's observations, the probabilities of successful impersonation and substitution attacks are

$$\begin{aligned}P_{I,b} &= \mathbb{E}_{\mathbf{Z}^{b-1}}\left[\sup_e \left\{g(\mathcal{A}(e), \mathbf{z}^{b-1})\right\}\right], \\ P_{S,b} &= \mathbb{E}_{\mathbf{Z}^{b}}\left[\sup_e \left\{g(\mathcal{A}(e), \mathbf{z}^{b})\right\}\right],\end{aligned} \tag{8}$$

where we have defined for any $\gamma_{\mathcal{L}}$, for any $r_{\mathcal{L}}$, for any $m_{\mathcal{L},b}$, for any opponent's attack e, the set

$$\mathcal{I}_{\sigma,\widetilde{\sigma}} \triangleq \{q \in \mathcal{Q} : \overline{\sigma}_q \text{ appears exactly } |\mathcal{G}_q| \text{ times in } (\widetilde{\sigma}_{l,b})_{l\in\mathcal{L}}\},$$

the event $\mathcal{A}(e) \triangleq \{|\mathcal{I}_{\sigma,\widetilde{\sigma}}| = Q \text{ and } \exists l \in \mathcal{L}, \widetilde{m}_{l,b} \neq m_{l,b}\}$, and for any $\mathbf{z}^{b-1}$

$$\begin{aligned}&g(\mathcal{A}(e), \mathbf{z}^{b-1}) \\ &\triangleq \sum_{\gamma_{\mathcal{L}}} \sum_{r_{\mathcal{L}}} \sum_{m_{\mathcal{L},b}} \mathbb{1}\{\mathcal{A}(e)\}\, p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}}, m_{\mathcal{L},b} \middle| \mathbf{z}^{b-1}\right).\end{aligned}$$

The realization of the event $\mathcal{A}(e)$ means that all Q groups of messages in $\mathcal{P}$, are accepted as authenticated and at least one message has been modified.

We now define for any opponent's attack e, the event,

$$\mathcal{A}'(e) \triangleq \{|\mathcal{I}_{\sigma,\widetilde{\sigma}}| = Q\}.$$

Hence, by (8) and since $\forall\gamma_{\mathcal{L}}, \forall r_{\mathcal{L}}, \forall m_{\mathcal{L},b}, \mathbb{1}\{\mathcal{A}'(e)\} \geqslant \mathbb{1}\{\mathcal{A}(e)\}$, we have

$$P_{I,b} \leqslant \mathbb{E}_{\mathbf{Z}^{b-1}}\left[\sup_e \left\{g(\mathcal{A}'(e), \mathbf{z}^{b-1})\right\}\right]. \tag{9}$$

Define $\widetilde{\Sigma}_b \triangleq \bigtimes_{d\in\mathcal{D}} \widetilde{\Sigma}_{d,b}$, where $\bigtimes$ denotes the Cartesian product and for $d \in \mathcal{D}$, $\widetilde{\Sigma}_{d,b}$ is the set of sequences consisting of n_d distinct elements that appear exactly c_d times in $(\widetilde{\sigma}_{l,b})_{l\in\mathcal{L}}$. We arbitrarily order the sequence $(\overline{\sigma}_q)_{q\in\mathcal{Q}}$ so that the first n_1 elements are such that $|\mathcal{G}_q| = c_1$, the following n_2 elements are such that $|\mathcal{G}_q| = c_2$, and so forth. Then, we have

$$\begin{aligned}\mathbb{1}\{\mathcal{A}'(e)\} &= \mathbb{1}\left\{\bigcup_{a_b\in\widetilde{\Sigma}_b} \{(\overline{\sigma}_q)_{q\in\mathcal{Q}} = a_b\}\right\} \\ &\leqslant \sum_{a_b\in\widetilde{\Sigma}_b} \mathbb{1}\{(\overline{\sigma}_q)_{q\in\mathcal{Q}} = a_b\}.\end{aligned} \tag{10}$$

Then, for any $a_b \in \widetilde{\Sigma}_b$, we have

$$\begin{aligned}&\sum_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} \mathbb{1}\{(\overline{\sigma}_q)_{q\in\mathcal{Q}} = a_b\}\, p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\ &\overset{(a)}{=} \sum_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} \mathbb{1}\left\{\left(\overline{\gamma}_q\right)_{q\in\mathcal{Q}} = a_b \oplus \left(\bigoplus_{l\in\mathcal{G}_q} r_l\right)_{q\in\mathcal{Q}}\right\} \\ &\qquad \times p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\ &\overset{(b)}{\leqslant} \sum_{r_{\mathcal{L}}} \max_{\gamma_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\ &\leqslant \sum_{r_{\mathcal{L}}} \max_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\ &= |\mathcal{K}|^L \max_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right),\end{aligned} \tag{11}$$

where (a) holds by (6) and (7), (b) holds because only one term is nonzero in the sum $\sum_{\gamma_{\mathcal{L}}}$ for a fixed $r_{\mathcal{L}}$. Hence, for any $b \in \mathcal{B}$, for any $\mathbf{z}^{b-1}$, for any opponent's strategy e, we have

$$\sum_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}, m_{\mathcal{L},b}} \mathbb{1}\{\mathcal{A}'(e)\}\, p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}}, m_{\mathcal{L},b} \middle| \mathbf{z}^{b-1}\right)$$

$$\begin{aligned}
&\overset{(a)}{=} \sum_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} \mathbb{1}\left\{\mathcal{A}'(e)\right\} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\
&\overset{(b)}{\leqslant} \sum_{a_b \in \widetilde{\Sigma}_b} |\mathcal{K}|^L \max_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right) \\
&\overset{(c)}{\leqslant} \prod_{d \in \mathcal{D}} n_d! |\mathcal{K}|^L \max_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right), \qquad (12)
\end{aligned}$$

where (a) holds by marginalization over $m_{\mathcal{L},b}$, (b) holds by combining (10) and (11), (c) holds because the sum $\sum_{a_b \in \widetilde{\Sigma}_b}$, which depends on the opponent's strategy e and the values taken by $(\widetilde{\sigma}_{l,b})_{l \in \mathcal{L}}$, has at most $\prod_{d \in \mathcal{D}} n_d!$ terms.

We now use the following lemma, whose proof is omitted due to space constraints.

Lemma 1. *Let A and B be two correlated randoms variables over finite alphabets. For any $\epsilon \geqslant 0$, if $I(A;B) \leqslant \epsilon$, then*

$$0 \leqslant 2^{-H_\infty(A|B)} - 2^{-H_\infty(A)} \leqslant 2\left(2\ln 2\right)^{1/4} \epsilon^{1/4},$$

where $H_\infty(A|B)$ is the average min-entropy of A given B [8].

We thus have

$$\begin{aligned}
P_{I,b} &\overset{(a)}{\leqslant} \mathbb{E}_{\mathbf{Z}^{b-1}}\left[\sup_e \left\{g(\mathcal{A}'(e), \mathbf{z}^{b-1})\right\}\right] \\
&\overset{(b)}{\leqslant} \mathbb{E}_{\mathbf{Z}^{b-1}}\left[\prod_{d \in \mathcal{D}} n_d! |\mathcal{K}|^L \max_{\gamma_{\mathcal{L}}, r_{\mathcal{L}}} p\left(\gamma_{\mathcal{L}}, r_{\mathcal{L}} \middle| \mathbf{z}^{b-1}\right)\right] \\
&= \prod_{d \in \mathcal{D}} n_d! |\mathcal{K}|^L 2^{-H_\infty\left(\Gamma_{\mathcal{L}} R_{\mathcal{L}} \middle| \mathbf{Z}^{b-1}\right)} \\
&\overset{(c)}{\leqslant} \prod_{d \in \mathcal{D}} n_d! |\mathcal{K}|^L \left(2^{-H_\infty(\Gamma_{\mathcal{L}} R_{\mathcal{L}})} + 2\left(2\ln 2\right)^{1/4} \delta(N)^{1/4}\right) \\
&\overset{(d)}{=} \prod_{d \in \mathcal{D}} n_d! \left(\frac{1}{|\mathcal{K}|^Q} + 2\left(2\ln 2\right)^{1/4} \delta(N)^{1/4} |\mathcal{K}|^L\right) \\
&\xrightarrow{N \to \infty} \frac{\prod_{d \in \mathcal{D}} n_d!}{|\mathcal{K}|^Q}, \qquad (13)
\end{aligned}$$

where (a) holds by (9), (b) holds since (12) is valid for any opponent's attack e, (c) holds by Lemma 1 and strong secrecy with $\delta(N)$ such that $\lim_{N\to\infty} \delta(N) = 0$, (d) holds since $\Gamma_{\mathcal{L}}$ contains Q independent sequences uniformly distributed over $\mathcal{K}$, that are independent of $R_{\mathcal{L}}$, which in turn, is a sequence of L independent sequences uniformly distributed over $\mathcal{K}$.

Replacing $\mathbf{Z}^{b-1}$ by $\mathbf{Z}^b$, we obtain the same upper bound (13) for $\lim_{N\to\infty} P_{S,b}$. Finally, one can show

$$\lim_{N\to\infty} \max_{b \in \mathcal{B}} \max\left(P_{I,b}, P_{S,b}\right) \leqslant \frac{\prod_{d \in \mathcal{D}} n_d!}{|\mathcal{K}|^Q}.$$

C. Probability of a succesful attack

By combining the lower bound of Section IV-A and the upper bound of Section IV-B we obtain the following result.

Theorem 1. *For any $L, B \in \mathbb{N}\backslash\{0\}$, for any partition $\mathcal{P}$ of $\mathcal{L}$, the probability of a successful attack of the authentication scheme of Section III satisfies*

$$\lim_{N\to\infty} P_A \overset{|\mathcal{K}|\to\infty}{\sim} \frac{\prod_{d \in \mathcal{D}} n_d!}{|\mathcal{K}|^Q}. \qquad (14)$$

Observe that the decay of the probability of successful attack with respect to $|\mathcal{K}|$ only depends on Q, the number of parts of $\mathcal{P}$. Consequently, anonymity benefits to the opponent since Q is maximal and equal to L when no anonymity constraint holds. Observe also that, except for the case $Q = 1$, for which $\lim_{N\to\infty} P_A \overset{|\mathcal{K}|\to\infty}{\sim} \frac{1}{|\mathcal{K}|}$, all the transmitters benefit from a multiuser setting compared to a single-user setting in terms of probability of successful attack. This observation is further discussed in Section V.

Remark 1. *If one chooses $\log|\mathcal{K}| = \omega_N$, where $\omega_N = o(N)$ and $\lim_{N\to\infty} \omega_N = +\infty$, then the communication rates in Algorithm 1 vanish as $N \to \infty$.*

V. Concluding remarks

As mentioned in Section II, a threshold, τ, on the number of correctly identified groups of messages can be introduced for the authentication decision. The choice of τ influences the numerator of the r.h.s. in (14) in a non-trivial way, and induces the following trade-off. If τ is large, the receiver might refuse up to $\tau - 1$ correctly authenticated groups of messages, which might not be desirable since it represents wasted transmissions. On the other hand, if τ is small, the probability of a successful attack increases by a factor $|\mathcal{K}|$ each time τ is decreased by one. This extension can be found in an upcoming journal version.

Our proposed authentication scheme relies on codes for the multiple access wiretap channel (MAC-WT) under strong secrecy. While [7] provides the existence of such codes, providing a constructive and low-complexity counterpart of [7] remains challenging. In the case of two transmitters, we have recently proposed a polar coding scheme for MAC-WT in [9], which can be used to implement the proposed scheme in Section III.

References

[1] G. J. Simmons, "Authentication theory/coding theory," in *Advances in Cryptology*. Springer, 1985, pp. 411–431.

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

[3] C. Boncelet, "The NTMAC for authentication of noisy messages," *IEEE Trans. Inf. Forensics and Security*, vol. 1, no. 1, pp. 35–42, 2006.

[4] Y. Liu and C. Boncelet, "The CRC–NTMAC for noisy message authentication," *IEEE Trans. Inf. Forensics and Security*, vol. 1, no. 4, pp. 517–523, 2006.

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

[6] I. Csiszár, "Almost independence and secrecy capacity," *Problems of Information Transmission*, vol. 32, no. 1, pp. 40–47, 1996.

[7] M. Yassaee and M. Aref, "Multiple access wiretap channels with strong secrecy," in *Proc. IEEE Inf. Theory Workshop*, 2010, pp. 1–5.

[8] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM journal on computing*, vol. 38, no. 1, pp. 97–139, 2008.

[9] R. Chou and A. Yener, "Polar coding for the multiple access wiretap channel via rate-splitting and cooperative jamming," in *Proc. of IEEE Int. Symp. Info. Theory*, 2016.