Projective Systematic Authentication via Reed-Muller Codes

Hsuan-Po Liu* and Hessam Mahdavifar*†
*Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
†Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA
Emails: hsuanpo@umich.edu, h.mahdavifar@northeastern.edu

***Abstract*—In this paper, we study the problem of constructing projective systematic authentication schemes based on binary linear codes. In systematic authentication, a tag for authentication is generated and then appended to the information, also referred to as the source, to be sent from the sender. Existing approaches to leverage *projective* constructions focus primarily on codes over large alphabets, and the projection is simply into one single symbol of the codeword. In this work, we extend the projective construction and propose a general projection process in which the source, which is mapped to a higher dimensional codeword in a given code, is first projected to a lower dimensional vector. The resulting vector is then masked to generate the tag. To showcase the new method, we focus on leveraging binary linear codes and, in particular, Reed-Muller (RM) codes for the proposed projective construction. More specifically, we propose systematic authentication schemes based on RM codes, referred to as RM-A-codes. We provide analytical results for probabilities of deception, widely considered as the main metrics to evaluate the performance of authentication systems. Through our analysis, we discover and discuss explicit connections between the probabilities of deception and various properties of RM codes.**

I. Introduction

With the rapid expansion of wireless networks, the need for providing message integrity and authenticity has become increasingly crucial, and is widely regarded as one of the major goals of cryptography systems [1]. Authentication codes were first introduced in [2]. A theoretical framework for authentication was then introduced by Simmons [3], which considers an *unconditionally secure* authentication system, i.e., where the adversary may have unlimited computational power.

A conventional authentication system involves three parties: a *sender* who sends a message, a *receiver* who is the intended recipient of the message, and an *adversary* who attempts to attack by either impersonating the sender and inserting a message into the channel, or substituting an intercepted message with a fraudulent one. These two types of attacks are termed the *impersonation* attack and the *substitution* attack, respectively. The communication is assumed to take place over a public channel. To protect the system from the aforementioned attacks, the sender and the receiver utilize a shared secret key, known only to them, which is then used in the encoding rule of the underlying authentication code. The probabilities of successful impersonation and substitution attacks by the adversary are the probabilities of deception, which are considered to evaluate the performance of the authentication system.

A vast body of work is dedicated to designing authentication codes with various methodologies and under various constraints. Some of the major approaches to this problem include geometric codes [4], [5], nonlinear functions [6]–[8], algebraic constructions [9]–[11], and error-correcting codes [12]–[17]. In this work, we focus on authentication codes without secrecy, which are also known as *systematic* authentication codes, constructed via error-correcting codes using a *projective* construction [15]–[17]. In systematic authentication codes, a message is sent from the sender to the receiver through the public channel, including the source state (i.e., plaintext), appended with a tag. The tag is generated by an encoding rule from a shared secret key between the sender and the receiver. The *projective* constructions proposed in prior works [15]–[17], are specifically for codes over rather large underlying alphabets, and the projection is simply into one single symbol of the codeword.

In this paper, we extend upon the projective authentication methods and consider a general notion of projecting higher dimensional codewords generated by a specific error-correcting code to a lower dimensional vector, e.g., sub-blocks of codewords. This constitutes a major building block of the system. More specifically, the secret key is split into two sub-parts. The first part is used to indicate the subset of the codeword to be projected to the lower dimension, where it is masked by the second part of the key. In order to showcase the proposed scheme, we focus on designing new projective authentication codes based on binary linear codes, and in particular, Reed-Muller (RM) codes, referred to as RM-A-codes. The main motivation behind this choice is to demonstrate that the already existing physical layer blocks for binary error correction can be leveraged for authentication as well, leading to potential solutions for low-complexity low-cost communication devices in massive networks, such as in Internet-of-Things (IoT) networks. RM codes are one of the oldest families of codes, which have received renewed attention in recent years due to their capacity-achieving properties [18], [19] as well as their excellent performance in short blocklengths [20]–[22]. We demonstrate that RM codes are a perfect fit as a building block for the proposed systematic authentication based on binary linear codes, and present closed-

This work was supported in part by the Center for Ubiquitous Connectivity (CUbiC), sponsored by Semiconductor Research Corporation (SRC) and Defense Advanced Research Projects Agency (DARPA) under the JUMP 2.0 program, and in part by the Department of Energy under Grant DESC0022186.

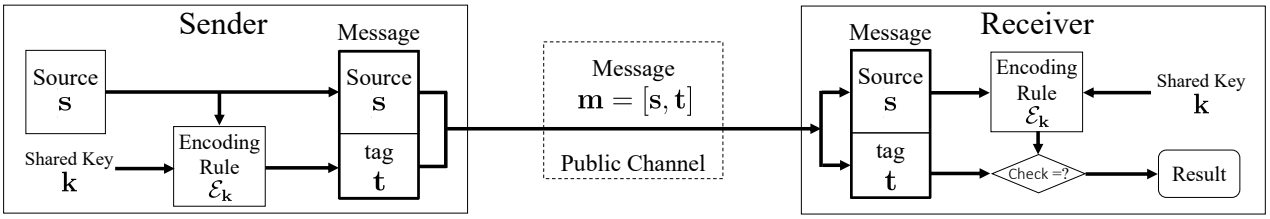
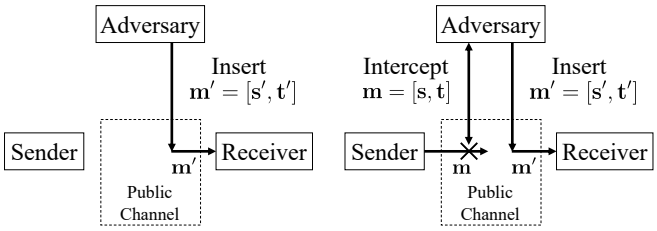


Fig. 1: Systematic authentication system

form expressions for the probability of success of attacks by an adversary in the considered authentication system. As for the probability of success for substitution attacks, we show that our construction reduces the computationally expensive calculations for characterizing the guarantees, and identifies the performance under different settings considering RM codes.

The rest of the paper is structured as follows. In Section II, we provide the preliminaries of systematic authentication codes. In Section III, we present the proposed projective constructions. We then analyze the theoretical results for our proposed construction in Section IV. Finally, we conclude the paper in Section V.



(a) Impersonation attack (b) Substitution attack

Fig. 2: Two types of attacks

II. Preliminaries

A. Systematic authentication codes

The systematic authentication code is defined as a four-tuple $(\mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{E}_{\mathbf{k}} : \mathbf{k} \in \mathcal{K}\})$, where $\mathcal{S}$ is the source space, $\mathcal{T}$ is the tag space, $\mathcal{K}$ is the key space, and $\mathcal{E}_{\mathbf{k}} : \mathcal{S} \mapsto \mathcal{T}$ is the encoding rule. During the authentication phase, which may happen only once at the beginning of the communication or at the beginning of every new round of communication, the sender first generates a source $\mathbf{s} \in \mathcal{S}$. Then given the secretly shared key $\mathbf{k}$ between the sender and the receiver, the encoding rule $\mathcal{E}_{\mathbf{k}} : \mathcal{S} \mapsto \mathcal{T}$ generates a tag $\mathbf{t} = \mathcal{E}_{\mathbf{k}}(\mathbf{s}) \in \mathcal{T}$. The message $\mathbf{m} \in \mathcal{M} = \mathcal{S} \times \mathcal{T}$ sent from the sender to the receiver is then denoted by the concatenation of $\mathbf{s}$ and $\mathbf{t}$ as $\mathbf{m} = [\mathbf{s}, \mathbf{t}]$. When the receiver receives a message (which includes a source vector and a tag vector), it checks the authenticity by verifying whether the received tag can be generated from the secretly shared key $\mathbf{k}$ through the encoding rule $\mathcal{E}_{\mathbf{k}} : \mathcal{S} \mapsto \mathcal{T}$ or not. If yes, the receiver accepts the received message; otherwise, the receiver discards it. We summarize the systematic authentication system in Fig. 1. Note that we will specify the dimension for each of the underlying vectors in the next section when we formally propose our construction.

The adversary is assumed to have the ability to insert messages into the public channel and/or to intercept messages that are sent over the public channel and to modify them. Two types of attacks are often considered in the authentication systems, referred to as the impersonation attack and the substitution attack. An impersonation attack occurs when the adversary inserts a new message $\mathbf{m}' = [\mathbf{s}', \mathbf{t}']$ into the public channel, see Fig. 2a. A substitution attack is when the adversary observes a message $\mathbf{m} = [\mathbf{s}, \mathbf{t}]$ that exists in the public channel, intercepts it, then inserts a new message $\mathbf{m}' = [\mathbf{s}', \mathbf{t}']$ into the channel, where $\mathbf{s}' \neq \mathbf{s}$, this attack is demonstrated in Fig. 2b. The security guarantees of the system are measured in terms of the adversary's probability of success with respect to each of the attacks. The probabilities of success for the impersonation and the substitution attacks by the adversary are denoted by P_{I} and P_{S}, respectively. These quantities are defined more explicitly in the next subsection.

B. Probability of deception

The probability of success of the impersonation attack is denoted by P_{I} and defined as

$$P_{\mathrm{I}} := \max_{\mathbf{s}', \mathbf{t}'} \mathbb{P}([\mathbf{s}', \mathbf{t}'] \text{ valid}), \tag{1}$$

where

$$\mathbb{P}([\mathbf{s}', \mathbf{t}'] \text{ valid}) = \frac{|\{\mathbf{k} \in \mathcal{K} : \mathbf{t}' = \mathcal{E}_{\mathbf{k}}(\mathbf{s}')\}|}{|\{\mathbf{k} \in \mathcal{K}\}|}.$$

Note that this probability is with respect to the space of all possible keys under a uniform distribution.

The probability of success of the substitution attack is denoted by P_{S} and defined as

$$P_{\mathrm{S}} := \max_{\mathbf{s}, \mathbf{t}} \max_{\mathbf{s}' \neq \mathbf{s}, \mathbf{t}'} \mathbb{P}([\mathbf{s}', \mathbf{t}'] \text{ valid} \,|\, [\mathbf{s}, \mathbf{t}] \text{ observed}), \tag{2}$$

where

$$\mathbb{P}([\mathbf{s}', \mathbf{t}'] \text{ valid} \,|\, [\mathbf{s}, \mathbf{t}] \text{ observed}) = \frac{\left|\left\{\mathbf{k} \in \mathcal{K} : \begin{matrix} \mathbf{t} = \mathcal{E}_{\mathbf{k}}(\mathbf{s}), \\ \mathbf{t}' = \mathcal{E}_{\mathbf{k}}(\mathbf{s}') \end{matrix}\right\}\right|}{|\{\mathbf{k} \in \mathcal{K} : \mathbf{t} = \mathcal{E}_{\mathbf{k}}(\mathbf{s})\}|}. \tag{3}$$

In this analysis, it is assumed that both the key and the source state are from uniform distributions on the key space and the source space, respectively.

It is shown in [23] that we have $P_{\mathrm{S}} \geq P_{\mathrm{I}} \geq \frac{1}{|\mathcal{T}|}$. The core design criterion for constructing a good systematic

authentication code is minimizing the probabilities of deception P_{I} and P_{S}.

C. RM codes: A brief overview

An RM code [20] is denoted by $\mathrm{RM}(m,r)$, where m is a positive integer that defines the blocklength of the code as $n=2^m$, r is termed as the order of the code with $r\in\{0,1,\ldots,m\}$, which determines the dimension of the code as $\sum_{i=0}^{r}\binom{m}{i}$. Note that $\mathrm{RM}(m,r)$ generates a $(2^m,\sum_{i=0}^{r}\binom{m}{i})$-code. The resulting generator matrix $\mathbf{G}$ is constructed by

$$\mathbf{G}=\begin{bmatrix}\mathbf{G}_r\\ \vdots\\ \mathbf{G}_1\\ \mathbf{G}_0\end{bmatrix}_{\sum_{i=0}^{r}\binom{m}{i}\times 2^m}, \tag{4}$$

where $\mathbf{G}_0$ is an all-one vector with n entries, $\mathbf{G}_1$ is a matrix of dimension $m\times n$ that each column vector is a unique m-bit binary vector, and $\mathbf{G}_j$ is an $\binom{m}{j}\times n$ matrix that each row is constructed by an entry-wise product of a distinct set of j rows from $\mathbf{G}_1$, for $j\in\{2,3,\ldots,r\}$. Thus, $\mathbf{G}$ has $\sum_{i=0}^{r}\binom{m}{i}$ rows, with a minimum distance of 2^{m-r}.

Note also that the last row of each of the $\mathbf{G}_i$'s, for $i\in\{1,2,\ldots,r\}$, are in the form of $[\mathbf{1}_{2^{m-i}},\mathbf{0}_{n-2^{m-i}}]$. Such vectors, consisting of two separate sub-blocks of all-ones and all-zeros, play a critical role in the analysis of the probability of success for the substitution attacks. This is an advantage of employing RM-A-codes, where the set of such codewords is exactly known and can be characterized.

III. The Proposed RM-A-codes

First, recall a systematic authentication code, defined in Section II-A, as a four-tuple $(\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{E}_{\mathbf{k}}:\mathbf{k}\in\mathcal{K}\})$. We define the proposed projective construction based on RM codes, referred to as RM-A-codes, as follows:

Definition 1 (RM-A-codes)**.** Consider the source $\mathbf{s}\in\mathcal{S}=\{0,1\}^M$, tag $\mathbf{t}\in\mathcal{T}=\{0,1\}^l$, and let the key $\mathbf{k}\in\mathcal{K}$ to be a concatenation of two keys $\mathbf{k}_1\in\mathcal{K}_1\subset\{0,1\}^n$ and $\mathbf{k}_2\in\mathcal{K}_2=\{0,1\}^l$, such that $\mathbf{k}=[\mathbf{k}_1,\mathbf{k}_2]\in\mathcal{K}=\mathcal{K}_1\times\mathcal{K}_2$, for some positive integers n,M,l with $n>M\geq l$. More details on $\mathcal{K}_1$ are discussed later in Remark 1. The tag $\mathbf{t}$ is defined as $\mathbf{t}=\mathcal{E}_{\mathbf{k}}(\mathbf{s})=\mathbf{c}_{\mathbf{s},\mathbf{k}_1}+\mathbf{k}_2$, where $\mathbf{c}_{\mathbf{s},\mathbf{k}_1}$ is a vector selected as a subset of l entries from the entries of the codeword $\mathbf{c}_{\mathbf{s}}$ with the indices determined by $\mathbf{k}_1$, $\mathbf{c}_{\mathbf{s}}$ is a codeword encoded from the source $\mathbf{s}\in\{0,1\}^M$ as a subset of the information input $\mathbf{u}\in\{0,1\}^{\sum_{i=0}^{r}\binom{m}{i}}$, such that $\mathbf{u}=[\mathbf{0}_{\sum_{i=0}^{r}\binom{m}{i}-M-1},\mathbf{s},0]$, for a $(n,\sum_{i=0}^{r}\binom{m}{i})$-code from $\mathrm{RM}(m,r)$ with a generator matrix $\mathbf{G}$, where $M<\sum_{i=0}^{r}\binom{m}{i}$ and $n=2^m$, such that $\mathbf{c}_{\mathbf{s}}=\mathbf{u}\mathbf{G}$.

Remark 1. The key $\mathbf{k}_1$ is a length-n binary vector with a weight of l, i.e., containing l ones, where the indices of ones indicate the indices of the selected l entries from $\mathbf{c}_{\mathbf{s}}$ to construct $\mathbf{c}_{\mathbf{s},\mathbf{k}_1}$. Thus, given a systematic authentication code, we have $|\mathcal{K}_1|=\binom{n}{l}$. Note that we may reduce the length of $\mathbf{k}_1$ to $\lceil\log_2|\mathcal{K}_1|\rceil=\lceil\log_2\binom{n}{l}\rceil$.

To generate $\mathbf{c}_{\mathbf{s}}$ given $\mathbf{s}$ and $\mathbf{G}$ from $\mathrm{RM}(m,r)$ in (4), one needs to specify $\mathbf{u}$. To this end, we start by presenting the following lemmas.

Lemma 1. *Given a generator matrix $\mathbf{G}$ from $\mathrm{RM}(m,r)$, we have $\mathbf{c}_{\mathbf{s}}+\mathbf{c}_{\mathbf{s}'}=\mathbf{1}_n$ when $\mathbf{u}=\mathbf{s}$ and $\mathbf{u}'=\mathbf{s}'$, where $\mathbf{c}_{\mathbf{s}}=\mathbf{u}\mathbf{G}$ and $\mathbf{c}_{\mathbf{s}'}=\mathbf{u}'\mathbf{G}$, if $\mathbf{s}$ and $\mathbf{s}'$ are binary vectors that differ only in the last entry (referred to as neighboring vectors).*

Proof. Let $\mathbf{u}=[u_1,u_2,\ldots,u_{\sum_{i=0}^{r}\binom{m}{i}-1},0]$ and $\mathbf{u}'=[u_1,u_2,\ldots,u_{\sum_{i=0}^{r}\binom{m}{i}-1},1]$ to be a pair of binary neighboring vectors. Then we have

$$\mathbf{c}_{\mathbf{s}}+\mathbf{c}_{\mathbf{s}'}=\mathbf{u}\mathbf{G}+\mathbf{u}'\mathbf{G}=(\mathbf{u}+\mathbf{u}')\mathbf{G}=[\mathbf{0}_{\sum_{i=0}^{r}\binom{m}{i}-1},1]\mathbf{G}. \tag{5}$$

Equivalently, $\mathbf{c}_{\mathbf{s}}+\mathbf{c}_{\mathbf{s}'}$ is equal to the last row in $\mathbf{G}$, which is the all-one vector according to (4). □

The next lemma demonstrates that using plain RM codes results in the probability of success for the substitution attack being one. Hence, we will modify the structure by considering sub-codes of RM codes, i.e., by letting $\mathbf{s}$ be a sub-vector of $\mathbf{u}$ while the remaining entries of $\mathbf{u}$ are fixed to zeros. This will be clarified later.

Lemma 2. *For any codeword $\mathbf{c}_{\mathbf{s}}$, the adversary can pick $\mathbf{c}_{\mathbf{s}'}$, where $\mathbf{s},\mathbf{s}'\in\mathcal{S}$, with $\mathbf{c}_{\mathbf{s}}+\mathbf{c}_{\mathbf{s}'}=\mathbf{1}_n$, resulting in $P_{\mathrm{S}}=1$.*

With Lemmas 1 and 2, we have the following proposition.

Proposition 3. *RM-A-codes ensure $P_{\mathrm{S}}<1$ if we set the last entry in $\mathbf{u}$ frozen to be 0.*

Proof. Lemma 1 indicates that we will have $\mathbf{c}_{\mathbf{s}}+\mathbf{c}_{\mathbf{s}'}=\mathbf{1}_n$ if we do not set the last entry of $\mathbf{u}$ to be 0. Consequently, Lemma 2 implies that it will lead to an authentication code with $P_{\mathrm{S}}=1$, which is undesirable. Except $\mathbf{c}_{\mathbf{s}'}$, there is no other codeword (together with a tag), that form a valid pair for all choices of the key. Hence, removing $\mathbf{c}_{\mathbf{s}'}$ results in $P_{\mathrm{S}}<1$. □

Proposition 3 implies that we should set the last entry of $\mathbf{u}$ frozen to be 0, since the core design criterion for constructing a good authentication code is to minimize the probabilities of deception and one should naturally avoid $P_{\mathrm{S}}=1$.

As specified by Proposition 3 with $M<\sum_{i=0}^{r}\binom{m}{i}$, we obtain $\mathbf{u}=[\mathbf{0}_{\sum_{i=0}^{r}\binom{m}{i}-M-1},\mathbf{s},0]$ such that $\mathbf{c}_{\mathbf{s}}=\mathbf{u}\mathbf{G}=[\mathbf{0}_{\sum_{i=0}^{r}\binom{m}{i}-M-1},\mathbf{s},0]\mathbf{G}$. To construct an RM-A-code as defined in Definition 1, we need to determine the length of the source M, the length of the tag l, and $\mathrm{RM}(m,r)$ to generate a $(2^m,\sum_{i=0}^{r}\binom{m}{i})$-code which has a blocklength $n=2^m$. We end this section by illustrating a toy example of RM-A-codes.

Example 1. Let $M=2$, $l=1$, and choose $m=2,r=1$ to generate a $(4,3)$-code by $\mathrm{RM}(2,1)$. In such code, we obtain the generator matrix as

$$\mathbf{G}=\begin{bmatrix}1&0&1&0\\1&1&0&0\\1&1&1&1\end{bmatrix}. \tag{6}$$

	$\mathbf{c}_{[0,0]}$	$\mathbf{c}_{[0,1]}$	$\mathbf{c}_{[1,0]}$	$\mathbf{c}_{[1,1]}$
$\mathbf{c}_{\mathbf{s},[1,0,0,0]} + [0]$	0	1	1	0
$\mathbf{c}_{\mathbf{s},[1,0,0,0]} + [1]$	1	0	0	1
$\mathbf{c}_{\mathbf{s},[0,1,0,0]} + [0]$	0	0	1	1
$\mathbf{c}_{\mathbf{s},[0,1,0,0]} + [1]$	1	1	0	0
$\mathbf{c}_{\mathbf{s},[0,0,1,0]} + [0]$	0	1	0	1
$\mathbf{c}_{\mathbf{s},[0,0,1,0]} + [1]$	1	0	1	0
$\mathbf{c}_{\mathbf{s},[0,0,0,1]} + [0]$	0	0	0	0
$\mathbf{c}_{\mathbf{s},[0,0,0,1]} + [1]$	1	1	1	1

TABLE I: Authentication matrix for the toy example

The information input $\mathbf{u}$ is a length-3 vector as $\mathbf{u} = [\mathbf{s}, 0] = [s_1, s_2, 0]$, where $\mathbf{s} = [s_1, s_2] \in \mathcal{S} = \{[0,0], [0,1], [1,0], [1,1]\}$. We obtain the codeword as $\mathbf{c_s} = \mathbf{uG} = [s_1 + s_2, s_2, s_1, 0]$. The keys $\mathbf{k}_1$ and $\mathbf{k}_2$ are constructed as $\mathbf{k}_1 \in \mathcal{K}_1 = \{[1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1]\}$ and $\mathbf{k}_2 \in \mathcal{K}_2 = \{[0], [1]\}$. Then, the tags can be generated from the given key, recall that $\mathbf{t} = \mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{k}_2$. Then, for instance, given $\mathbf{k}_1 = [1,0,0,0]$ and $\mathbf{k}_2 = [0]$, we have $\mathbf{t} = \mathbf{c}_{\mathbf{s},[1,0,0,0]} + [0] = [s_1 + s_2]$, where $\mathbf{c}_{\mathbf{s},[1,0,0,0]}$ denotes selecting the first entry in the codeword $\mathbf{c_s}$. By calculating the tags $\mathbf{t}$'s over all sources in $\mathcal{S}$ and all keys $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$, we have the authentication matrix shown in TABLE I. Applying the definitions of the probabilities of deception yields $P_{\mathrm{I}} = 0.5$, $P_{\mathrm{S}} = 0.5$.

This example shows a simple case for our construction. In the next section, we analyze theoretical closed-form expressions for P_{I} and P_{S}.

IV. Analysis

A. The probability of success of the impersonation attack P_{I}

The following theorem shows that the proposed scheme has the lowest possible probability of success of the impersonation attack, i.e., the adversary cannot do better than a random assignment of the tag.

Theorem 4. *Given an RM-A-code as defined in Definition 1, we have*

$$P_{\mathrm{I}} = \max_{\mathbf{s}',\mathbf{t}'} \frac{|\{\mathbf{k} \in \mathcal{K} : \mathbf{t}' = \mathcal{E}_{\mathbf{k}}(\mathbf{s}') = \mathbf{c}_{\mathbf{s}',\mathbf{k}_1} + \mathbf{k}_2\}|}{|\{\mathbf{k} \in \mathcal{K}\}|} = \frac{1}{2^l}. \quad (7)$$

Proof. Since $|\mathcal{K}_1| = \binom{n}{l}$ and $|\mathcal{K}_2| = 2^l$, we have

$$|\{\mathbf{k} \in \mathcal{K}\}| = |\mathcal{K}| = |\mathcal{K}_1 \times \mathcal{K}_2| = \binom{n}{l} \cdot 2^l. \quad (8)$$

Furthermore, with $\mathcal{T} = \mathcal{K}_2$, we have

$$|\{\mathbf{k} \in \mathcal{K} : \mathbf{t}' = \mathcal{E}_{\mathbf{k}}(\mathbf{s}') = \mathbf{c}_{\mathbf{s}',\mathbf{k}_1} + \mathbf{k}_2\}| = |\mathcal{K}_1| = \binom{n}{l}. \quad (9)$$

Thus, with (8) and (9), we have $P_{\mathrm{I}} = \frac{1}{2^l}$. □

B. Analysis for P_{S}: The probability of success of the substitution attack

The following theorem presents a simplified form for the calculation of P_{S}, enabling a more efficient method for calculating and characterizing the probability of success of the substitution attack, with the assistance of a straightforward linearity property that given $\mathbf{s}, \mathbf{s}' \in \mathcal{S} = \{0,1\}^M$ and a generator matrix $\mathbf{G}$ from $\mathrm{RM}(m, r)$, we have $\mathbf{c_s} + \mathbf{c_{s'}} = \mathbf{c_{s+s'}}$.

Theorem 5. *The quantity P_{S}, defined in* (2)*, can be equivalently computed as*

$$P_{\mathrm{S}} = \max_{\tilde{\mathbf{s}} \neq \mathbf{0}_M} \max_{\tilde{\mathbf{t}}} \frac{|\{\mathbf{k}_1 \in \mathcal{K}_1 : \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1} = \tilde{\mathbf{t}}\}|}{\binom{n}{l}}, \quad (10)$$

where $\tilde{\mathbf{s}} \in \mathcal{S}$, excluding the all-zero source vector, and $\tilde{\mathbf{t}} \in \mathcal{T}$.

Proof. We have

$$\begin{aligned}
P_{\mathrm{S}} &= \max_{\mathbf{s},\mathbf{t}} \max_{\mathbf{s}' \neq \mathbf{s},\mathbf{t}'} \mathbb{P}([\mathbf{s}', \mathbf{t}'] \text{ valid} \,|\, [\mathbf{s}, \mathbf{t}] \text{ observed}) \\
&= \max_{\mathbf{s},\mathbf{t}} \max_{\mathbf{s}' \neq \mathbf{s},\mathbf{t}'} \frac{\left|\left\{\mathbf{k} \in \mathcal{K} : \begin{matrix} \mathbf{t} = \mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{k}_2, \\ \mathbf{t}' = \mathbf{c}_{\mathbf{s}',\mathbf{k}_1} + \mathbf{k}_2 \end{matrix}\right\}\right|}{|\{\mathbf{k} \in \mathcal{K} : \mathbf{t} = \mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{k}_2\}|} \\
&\overset{(b)}{=} \max_{\mathbf{s},\mathbf{t}} \max_{\mathbf{s}' \neq \mathbf{s},\mathbf{t}'} \frac{\left|\left\{\mathbf{k} \in \mathcal{K} : \begin{matrix} \mathbf{t} = \mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{k}_2, \\ \mathbf{t} + \mathbf{t}' = \mathbf{c}_{\mathbf{s}+\mathbf{s}',\mathbf{k}_1} \end{matrix}\right\}\right|}{\binom{n}{l}} \\
&= \max_{\mathbf{s},\mathbf{t}} \max_{\mathbf{s}' \neq \mathbf{s},\mathbf{t}'} \frac{|\{\mathbf{k}_1 \in \mathcal{K}_1 : \mathbf{t} + \mathbf{t}' = \mathbf{c}_{\mathbf{s}+\mathbf{s}',\mathbf{k}_1}\}|}{\binom{n}{l}} \\
&\overset{(c)}{=} \max_{\tilde{\mathbf{s}} \neq \mathbf{0}_M} \max_{\tilde{\mathbf{t}}} \frac{\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \tilde{\mathbf{t}} = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1}\right\}\right|}{\binom{n}{l}},
\end{aligned} \quad (11)$$

where (b) is by the linearity of the codes that $\mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{c}_{\mathbf{s}',\mathbf{k}_1} = \mathbf{c}_{\mathbf{s}+\mathbf{s}',\mathbf{k}_1}$, together with noting that $|\{\mathbf{k} \in \mathcal{K} : \mathbf{t} = \mathbf{c}_{\mathbf{s},\mathbf{k}_1} + \mathbf{k}_2\}| = \binom{n}{l}$. In (c), we let $\tilde{\mathbf{s}} = \mathbf{s} + \mathbf{s}'$ and $\tilde{\mathbf{t}} = \mathbf{t} + \mathbf{t}'$. Note that since $\mathbf{s} \neq \mathbf{s}'$, we must have $\tilde{\mathbf{s}} \neq \mathbf{0}_M$, which implies $\mathbf{c}_{\tilde{\mathbf{s}}} \neq \mathbf{0}_n$. □

Remark 2. Theorem 5 implies that, in RM-A-codes, the expression for P_{S} in (2) involving two maximizations over all codewords can be simplified to (10) which involves only a maximization over all nonzero codewords with calculating the maximum number of appearance of the valid tags corresponding to each codeword. Later, we show this can be even more simplified to a search only over the values of the weight of the codewords.

For further analysis, we define the maximum probability of appearance of a valid tag $\tilde{\mathbf{t}}$ in a nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}}$ given the key $\mathbf{k}_1$ as

$$P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}}) := \max_{\tilde{\mathbf{t}}} \frac{\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \tilde{\mathbf{t}} = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1}\right\}\right|}{\binom{n}{l}}, \quad (12)$$

where $\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \tilde{\mathbf{t}} = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1}\right\}\right|$ can be regarded as the number of appearance of $\tilde{\mathbf{t}}$ in the coordinates of $\mathbf{c}_{\tilde{\mathbf{s}}}$, such that, according to Theorem 5, we have

$$P_{\mathrm{S}} = \max_{\mathbf{c}_{\tilde{\mathbf{s}}}} P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}}).$$

For instance, considering Example 1 in the previous section, a valid nonzero codeword would be $\mathbf{c}_{\tilde{\mathbf{s}}} = [1,1,0,0]$ (setting $\tilde{\mathbf{s}} = [0,1]$). To generate a valid tag $\tilde{\mathbf{t}} = [1]$, we have either $\mathbf{k}_1 = [1,0,0,0]$ or $\mathbf{k}_1 = [0,1,0,0]$. Thus, the number of

appearance of $\tilde{\mathbf{t}} = [1]$ is $\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \tilde{\mathbf{t}} = [1] = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1}\right\}\right| = |\{[1,0,0,0],[0,1,0,0]\}| = 2$.

The following lemma presents an expression for the number of appearances of the tag $\tilde{\mathbf{t}}$ given certain weights for both the nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}}$ and the tag $\tilde{\mathbf{t}}$. We denote $\mathrm{wt}(\cdot)$ as the weight of a vector.

Lemma 6. *For a given nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}}$ with $\mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = w$, the number of tags $\tilde{\mathbf{t}}$ with $\mathrm{wt}(\tilde{\mathbf{t}}) = w_t$ is given by*

$$\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \begin{matrix}\tilde{\mathbf{t}} = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1}, \\ \mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = w, \mathrm{wt}(\tilde{\mathbf{t}}) = w_t\end{matrix}\right\}\right| = \binom{w}{w_t}\binom{n-w}{l-w_t}. \tag{13}$$

There exists a special case in which the nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}}$ with $\mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = w$ can be divided into two sub-blocks where one vector is an all-one vector as $\mathbf{1}_w$, while the other one is an all-zero vector as $\mathbf{0}_{n-w}$, i.e., $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$. We analyze the number of appearances of the tags in such special cases in the following lemma.

Lemma 7. *Given a nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$ with $\mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = w$, there exists exactly one valid tag $\tilde{\mathbf{t}} = [\mathbf{1}_{w_t}, \mathbf{0}_{l-w_t}]$ with $\mathrm{wt}(\tilde{\mathbf{t}}) = w_t$. Furthermore, for this $\tilde{\mathbf{t}}$ we have*

$$\left|\left\{\mathbf{k}_1 \in \mathcal{K}_1 : \begin{matrix}\tilde{\mathbf{t}} = \mathbf{c}_{\tilde{\mathbf{s}},\mathbf{k}_1} = [\mathbf{1}_w, \mathbf{0}_{n-w}], \\ \mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}], \\ \mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = w, \mathrm{wt}(\tilde{\mathbf{t}}) = w_t\end{matrix}\right\}\right| = \binom{w}{w_t}\binom{n-w}{l-w_t}. \tag{14}$$

The following lemma states that the maximum $P_{\mathrm{t}}(\cdot)$ over weight-w nonzero codewords occurs for a nonzero codeword of the type $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$, if such a codeword exists.

Lemma 8. *Given a nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$ and an arbitrary valid nonzero codeword $\mathbf{c}'_{\tilde{\mathbf{s}}}$ which $\mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = \mathrm{wt}(\mathbf{c}'_{\tilde{\mathbf{s}}}) = w$, we have $P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]) \geq P_{\mathrm{t}}(\mathbf{c}'_{\tilde{\mathbf{s}}})$.*

Furthermore, the following lemma helps us to further reduce the complexity of calculating the probability of success of the substitution attack P_{S}.

Lemma 9. *Given $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$ and $\mathbf{c}'_{\tilde{\mathbf{s}}} = [\mathbf{1}_{w'}, \mathbf{0}_{n-w'}]$, when $w' = n - w$, we have $P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]) = P_{\mathrm{t}}(\mathbf{c}'_{\tilde{\mathbf{s}}} = [\mathbf{1}_{w'}, \mathbf{0}_{n-w'}])$.*

With Lemmas 8 and 9, we end the theoretical analysis by presenting the theorem which demonstrates the closed-form expression for P_{S}.

Theorem 10. *Given an RM-A-code as defined in Definition 1, we have*

$$P_{\mathrm{S}} = \max_{w} \max_{w_t} \frac{\binom{w}{w_t}\binom{n-w}{l-w_t}}{\binom{n}{l}}, \tag{15}$$

where the maximization is over all w, with $\frac{n}{2^r} \leq w \leq \frac{n}{2}$, for which there exists a codeword of the form $[\mathbf{1}_w, \mathbf{0}_{n-w}]$.

Proof. Based on Lemma 8, we know that the nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}}$ with the largest $P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}})$ among all valid nonzero codewords $\mathbf{c}'_{\tilde{\mathbf{s}}}$ with identical weights $\mathrm{wt}(\mathbf{c}_{\tilde{\mathbf{s}}}) = \mathrm{wt}(\mathbf{c}'_{\tilde{\mathbf{s}}}) = w$ is always the nonzero codeword $\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]$, such that $P_{\mathrm{S}} = \max_w P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}])$. Thus, by maximizing $P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}])$ from choosing the weight w from all valid nonzero codewords with $[\mathbf{1}_w, \mathbf{0}_{n-w}]$'s naturally meets P_{S}. Lemma 9 indicates that, when $w' = n - w$, we have $P_{\mathrm{t}}(\mathbf{c}_{\tilde{\mathbf{s}}} = [\mathbf{1}_w, \mathbf{0}_{n-w}]) = P_{\mathrm{t}}(\mathbf{c}'_{\tilde{\mathbf{s}}} = [\mathbf{1}_{w'}, \mathbf{0}_{n-w'}])$. Then, when $w = w' = \frac{n}{2}$, we have $\mathbf{c}_{\tilde{\mathbf{s}}} = \mathbf{c}'_{\tilde{\mathbf{s}}}$. Therefore, the optimization problem only has to consider $w \leq \frac{n}{2}$ as the upper bound. Combining the lower bound which depends on the chosen $\mathrm{RM}(m, r)$ and M, as $\frac{n}{2^r} \leq w$, yields $\frac{n}{2^r} \leq w \leq \frac{n}{2}$. □

Remark 3. The proposed projective construction can be, in principle, generalized to all binary linear codes. RM-A-code, in fact, is a special case for constructing such a projective construction for systematic authentication codes. The structure of RM codes allows us to express the range of w's explicitly in the statement of Theorem 10, which reduces the computational overhead significantly compare to the expression for P_{S} in (2). For general codes, finding the range of w's for which codewords of the form $[\mathbf{1}_w, \mathbf{0}_{n-w}]$ exist is difficult and can be exponentially complex.

C. Numerical Analysis

Next, we provide numerical results for the proposed RM-A-codes with different blocklengths. The results are shown in TABLE II. In this setting, the source length is set as $M = 4$, tag length as $l = 3$, and the order of the RM code is $r = 1$, together with $m = \{4, 5, 6, 7, 8\}$. In TABLE II, it can be observed that P_{S} decreases while the blocklength increases. Furthermore, since l is fixed as $l = 3$, we have P_{I} as a constant $P_{\mathrm{I}} = \frac{1}{2^l} = 0.125$, which meets the lower bound. We leave a more thorough numerical analysis for our future works.

m	4	5	6	7	8
Size	$(16, 5)$	$(32, 6)$	$(64, 7)$	$(128, 8)$	$(256, 9)$
P_{I}	0.125	0.125	0.125	0.125	0.125
P_{S}	0.4	0.3817	0.3810	0.3780	0.3765

TABLE II: $M = 4$, $l = 3$, and $r = 1$, varies m

V. Conclusion

In this paper, we proposed a projective construction of systematic authentications based on binary linear codes, and studied a particular case based on RM codes, referred to as the RM-A-codes. The theoretical results are provided for the probabilities of deception. Furthermore, we have discussed explicit connections between the probability of success for the substitution attack and the RM code structure, which captures certain properties in the structures of error-correcting codes that are not very well understood. A potential direction for future work is to extend the projective construction for systematic authentications to more general classes of binary linear codes.

References

[1] S. A. A. Hakeem, H. H. Hussein, and H.-C. Kim, "Security requirements and challenges of 6g technologies and applications," *Sensors (Basel, Switzerland)*, vol. 22, 2022.

[2] E. N. Gilbert, F. J. MacWilliams, and N. J. Sloane, "Codes which detect deception," *bell system technical journal*, vol. 53, no. 3, pp. 405–424, 1974.

[3] G. J. Simmons, "Authentication theory/coding theory," in *Annual International Cryptology Conference*, 1985.

[4] J. Bierbrauer, T. Johansson, G. Kabatianskii, and B. Smeets, "On families of hash functions via geometric codes and concatenation," in *Advances in Cryptology—CRYPTO'93: 13th Annual International Cryptology Conference Santa Barbara, California, USA August 22–26, 1993 Proceedings 13*. Springer, 1994, pp. 331–342.

[5] J. Bierbrauer, "Universal hashing and geometric codes," *Designs, Codes and Cryptography*, vol. 11, pp. 207–221, 1997.

[6] S. Chanson, C. Ding, and A. Salomaa, "Cartesian authentication codes from functions with optimal nonlinearity," *Theoretical Computer Science*, vol. 290, no. 3, pp. 1737–1752, 2003.

[7] C. Ding and H. Niederreiter, "Systematic authentication codes from highly nonlinear functions," *IEEE transactions on information theory*, vol. 50, no. 10, pp. 2421–2428, 2004.

[8] C. Carlet, C. Ding, and H. Niederreiter, "Authentication schemes from highly nonlinear functions," in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 739–743.

[9] T. Helleseth and T. Johansson, "Universal hash functions from exponential sums over finite fields and galois rings," in *Advances in Cryptology—CRYPTO'96: 16th Annual International Cryptology Conference Santa Barbara, California, USA August 18–22, 1996 Proceedings 16*. Springer, 1996, pp. 31–44.

[10] C. Xing, H. Wang, and K. Y. Lam, "Constructions of authentication codes from algebraic curves over finite fields," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 886–892, 2000.

[11] F. Özbudak and Z. Saygi, "Some constructions of systematic authentication codes using galois rings," *Designs, Codes and Cryptography*, vol. 41, no. 3, pp. 343–357, 2006.

[12] T. Johansson, G. Kabatianskii, and B. Smeets, "On the relation between a-codes and codes correcting independent errors," in *Advances in Cryptology—EUROCRYPT'93: Workshop on the Theory and Application of Cryptographic Techniques Lofthus, Norway, May 23–27, 1993 Proceedings 12*. Springer, 1994, pp. 1–11.

[13] G. A. Kabatianskii, B. Smeets, and T. Johansson, "On the cardinality of systematic authentication codes via error-correcting codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 566–578, 1996.

[14] H. Wang, C. Xing, and R. Safavi-Naini, "Linear authentication codes: bounds and constructions," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 866–872, 2003.

[15] C. Ding and X. Wang, "A coding theory construction of new systematic authentication codes," *Theoretical computer science*, vol. 330, no. 1, pp. 81–99, 2005.

[16] C. Ding, T. Helleseth, T. Klove, and X. Wang, "A generic construction of cartesian authentication codes," *IEEE transactions on information theory*, vol. 53, no. 6, pp. 2229–2235, 2007.

[17] Y. Liu, L. Qu, and C. Li, "New constructions of systematic authentication codes from three classes of cyclic codes." *Adv. Math. Commun.*, vol. 12, no. 1, pp. 1–16, 2018.

[18] E. Abbe and C. Sandon, "Reed-muller codes have vanishing bit-error probability below capacity: a simple tighter proof via camellia boosting," *arXiv preprint arXiv:2312.04329*, 2023.

[19] G. Reeves and H. D. Pfister, "Reed–muller codes on bms channels achieve vanishing bit-error probability for all rates below capacity," *IEEE Transactions on Information Theory*, 2023.

[20] E. Abbe, A. Shpilka, and M. Ye, "Reed–muller codes: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 67, no. 6, pp. 3251–3277, 2020.

[21] M. V. Jamali, X. Liu, A. V. Makkuva, H. Mahdavifar, S. Oh, and P. Viswanath, "Reed-Muller subcodes: Machine learning-aided design of efficient soft recursive decoding," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 1088–1093.

[22] ——, "Machine learning-aided efficient decoding of Reed-Muller subcodes," *IEEE Journal on Selected Areas in Information Theory*, pp. 260–275, 2023.

[23] D. R. Stinson, "Combinatorial characterizations of authentication codes," *Designs, Codes and Cryptography*, vol. 2, no. 2, pp. 175–187, 1992.