ON THE RELATIONSHIP BETWEEN THE TRACEABILITY PROPERTIES OF REED-SOLOMON CODES

JOSÉ MOREIRA AND MARCEL FERNÁNDEZ
Department of Telematics Engineering
Universitat Politècnica de Catalunya
C. Jordi Girona 1-3, 08034 Barcelona, Spain

MIGUEL SORIANO
Department of Telematics Engineering
Universitat Politècnica de Catalunya
C. Jordi Girona 1-3, 08034 Barcelona, Spain
and
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Av. Carl Friedrich Gauss 7, 08860 Castelldefels (Barcelona), Spain

(Communicated by Judy Walker)

Abstract. Fingerprinting codes are used to prevent dishonest users (traitors) from redistributing digital contents. In this context, codes with the traceability (TA) property and codes with the identifiable parent property (IPP) allow the unambiguous identification of traitors. The existence conditions for IPP codes are less strict than those for TA codes. In contrast, IPP codes do not have an efficient decoding algorithm in the general case. Other codes that have been widely studied but possess weaker identification capabilities are separating codes. It is a well-known result that a TA code is an IPP code, and an IPP code is a separating code. The converse is in general false. However, it has been conjectured that for Reed-Solomon codes all three properties are equivalent. In this paper we investigate this equivalence, providing a positive answer when the number of traitors divides the size of the ground field.

1. Introduction

Distributing digital contents is an activity which is prone to an undesirable attack: unauthorized redistributions performed by dishonest users. To fight against this, the distributor can apply the fingerprinting technique. This technique consists in making each copy of the content unique by embedding a mark before delivering it. The embedding process must satisfy several properties. First, the process must be robust. This means that it should not be possible to remove or degrade the mark once it is embedded without rendering the content unusable. Second, the marked content must not differ substantially from the original content and must retain the same functionality.

2000 Mathematics Subject Classification: Primary: 94B60; Secondary: 94B65.
Key words and phrases: Fingerprinting and traitor tracing, identifiable parent property, separating codes, MDS codes, Reed-Solomon codes.

This work has been supported in part by the Spanish Government through projects Consolider Ingenio 2010 CSD2007-00004 “ARES” and TEC2011-26491 “COPPI,” and by the Catalan Government under Grant 2009 SGR-1362. J. Moreira is the recipient of an FPU fellowship, AP2009-3854, from the Spanish Ministry of Education.

467 ©2012 AIMS
The distributor assigns a unique mark to each user, and delivers the marked content correspondingly. In other words, the users will receive different objects that will have identical appearance from their perspective. Due to the uniqueness of the objects, they are discouraged from redistributing their own copy. However, several dishonest users, called traitors, may form a coalition, compare their copies and create a pirated copy. This is known as a collusion attack. The underlying idea is that the mark contained in the pirated copy hides the identity of the traitors. What is worse, that mark could coincide with that of a user outside the coalition and frame that innocent user. A traitor tracing scheme should be prudently designed so that this situation never happens ideally, or it happens with arbitrarily small probability. A stronger traitor tracing scheme would provide, besides that, a mechanism to identify at least one of the traitors. This goal is achieved if the set of marks chosen by the distributor constitutes a code with tracing properties. Such codes are also known as fingerprinting codes and the marks are called fingerprints.

Codes with tracing properties can be classified according to their capabilities in terms of protecting innocent users and identifying traitors. These properties are not equivalent in the general case. However, in this paper we show the equivalence of some of these properties for numerous families of Reed-Solomon codes.

The paper is organized as follows. In the next section we introduce the topic and the notation, and present some previous results. In Section 3 we present the main results of the paper, showing the equivalence of some tracing properties for Reed-Solomon codes, when the coalition size divides the size of the ground field. Next, in Section 4 we provide an illustrative example and a table summarizing the results. Finally we present the conclusions.

2. Definitions and background

Let $\mathbb{F}_q$ be the finite field of $q$ elements, where $q = p^m$ for some prime number $p$ and some integer $m \geq 1$. We denote the vectors of $n$ coordinates over $\mathbb{F}_q$ in boldface, e.g. $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{F}_q^n$. The (Hamming) distance between two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^n$ is denoted $d(\mathbf{a}, \mathbf{b})$.

An $(n,M,d)$-code $C$ over $\mathbb{F}_q$ is a subset of $\mathbb{F}_q^n$ of size $M$, where $d$ is called the minimum distance of the code, defined as $d = \min\{d(\mathbf{a}, \mathbf{b}) : \mathbf{a}, \mathbf{b} \in C, \mathbf{a} \neq \mathbf{b}\}$. The elements of $C$ are called codewords. Moreover, if $C$ is a $k$-dimensional vector space over $\mathbb{F}_q$ (of size $M = q^k$), we say that $C$ is an $[n,k,d]$-linear code, or simply an $[n,k,d]$-code.

Let $C \subseteq \mathbb{F}_q^n$ be the $(n,M,d)$-fingerprinting code chosen by the distributor. Recall that every codeword in $C$ uniquely identifies a user. In a collusion attack several traitors compare their copies. Since fighting against arbitrary-size coalitions is a very hard problem, the maximum size of the coalition is usually upper bounded by a constant $c \geq 2$. Since the copies of the traitors have different fingerprints, say $\mathbf{t}_1, \ldots, \mathbf{t}_c \in C$, they will be able to detect a number of differences in several locations. Using this information, the traitors are free to generate a pirated copy which is different from their copies.

We remark that the locations of the content chosen by the distributor to embed the fingerprint are the same for every copy. Hence, the differences spotted by the traitors are where two of their assigned fingerprints differ. Also, no information about the fingerprint can be gained from locations where the fingerprint is not embedded. Hence, it becomes sufficient to study only the fingerprints assigned to the users.
Definition 1. Let $C$ be an $(n, M, d)$-code over $\mathbb{F}_q$ and let $T = \{t_1, \ldots, t_c\} \subseteq C$ be a subset (coalition) of $c$ codewords. We say that coordinate $i$ is undetectable for coalition $T$ if $t_{1,i} = t_{2,i} = \cdots = t_{c,i}$. If this condition is not satisfied, we say that coordinate $i$ is detectable.

It is customary to assume that undetectable coordinates remain unchanged in the pirated copy. This is known as the marking assumption [2, 3]. The motivation for this assumption arises from applications to protecting software, where modifying arbitrary locations may damage the content. For the case of the detectable coordinates, the coalition can set them to any arbitrary value of $\mathbb{F}_q$, or even make them unreadable.

Recall that the goal of the distributor consists in identifying one of the members of a coalition that generated a pirated copy. If the only restriction in the generation of the pirated word is the marking assumption, then identification with zero-error probability is not possible [2, 3].

Definition 2. Let $C$ be an $(n, M, d)$-code over $\mathbb{F}_q$ and let $T = \{t_1, \ldots, t_c\} \subseteq C$ be a subset of $c$ codewords. We say that $x \in \mathbb{F}_q^n$ is a descendant of $T$ if for each coordinate $1 \leq i \leq n$ there exists a $t_j \in T$ such that $x_i = t_{j,i}$. We call $t_j$ a parent of $x$. The set of all the descendants of $T$ is denoted $\text{desc}(T)$, that is

$$\text{desc}(T) = \{x \in \mathbb{F}_q^n : x_i \in \{t_{1,i}, \ldots, t_{c,i}\}\}.$$ 

Also, the $c$-descendant code of $C$, denoted $\text{desc}_c(C)$, is defined as

$$\text{desc}_c(C) = \bigcup_{T \subseteq C, |T| \leq c} \text{desc}(T).$$

The set of descendants that we have just introduced is also known as the narrow-sense envelope by some authors [1]. Assuming that the set of pirated words that a coalition $T$ can generate is $\text{desc}(T)$, then identification with zero-error probability is possible. Observe that $T' \subseteq \text{desc}(T)$, and that $T' \subseteq T$ implies $\text{desc}(T') \subseteq \text{desc}(T)$. Hence, solving the identification problem for coalitions of size exactly $c$ also solves the problem for coalitions of any size $c' \leq c$.

Definition 3. Let $C$ be a code, and let $A, B \subseteq C$ be two disjoint subsets of size $c_1$ and $c_2$ respectively: $A = \{a_1, \ldots, a_{c_1}\}$, $B = \{b_1, \ldots, b_{c_2}\}$. We say that coordinate $i$ is separating if $\{a_{1,i}, \ldots, a_{c_1,i}\} \cap \{b_{1,i}, \ldots, b_{c_2,i}\} = \emptyset$.

Similarly as in [17], we denote $\theta(A, B)$ the number of separating coordinates between the disjoint subsets $A$ and $B$. Whenever $\theta(A, B) = 0$, we say that the pair $(A, B)$ is a $(c_1, c_2)$-nonseparated configuration. Also, for a code $C$, we denote $\theta_{c_1, c_2}$ the smallest value $\theta(A, B)$ attained for disjoint subsets $A, B \subseteq C$ of size $c_1$ and $c_2$ respectively. Although in general $\theta(A, B)$ is not a metric in the mathematical sense of the term, clearly, $\theta_{1,1}$ is the minimum distance of the code.

The values $\theta_{c_1, c_2}$ will be useful in the characterization of codes with tracing properties.

Definition 4. A code $C$ is $(c_1, c_2)$-separating if it has $\theta_{c_1, c_2} > 0$. In other words, $C$ does not contain any $(c_1, c_2)$-nonseparated configuration.
by many authors [17, 14, 16, 12, 6, 7]. Relationships with similar concepts have also
been shown. See for example overviews [17, 6].

Recently, more attention has been paid to separating codes in connection with
digital fingerprinting. In the crypto literature, \((c, 1)\)- and \((c, c)\)-separating codes are
also known as \(c\)-frameproof codes and \(c\)-secure frameproof codes, respectively [21, 20].

The connection between separating and fingerprinting codes is quite straightforward. Assume that a fingerprinting code has the \((c, 1)\)-separating property. Then
no coalition of size \(\leq c\) will be able to generate a pirated word that coincides with
the fingerprint of an innocent user. Moreover, using a \((c, c)\)-separating code, the
coalition can not even claim that the pirated word was generated by a disjoint
coalition of size \(\leq c\).

Still, the separating property is not enough to achieve unambiguous identification
of traitors. To see this, consider the case \(c = 2\) and the code
\[ C = \{a = (0, 0, 0), b = (0, 1, 1), c = (1, 1, 0), d = (1, 0, 1)\}, \]
which is \((2, 2)\)-separating. Since \((0, 1, 0) \in \text{desc}(\{a, b\}) \cap \text{desc}(\{a, c\}) \cap \text{desc}(\{b, c\})\),
one cannot decide which of the three possible pairs of codewords is the actual
coalition of traitors that generated the pirated word \((0, 1, 0)\).

Now, we present sufficient conditions to allow identification with zero-error prob-
ability.

**Definition 5.** A code \(C\) has the c-identifiable parent property (c-IPP) if for all \(x \in \mathbb{F}_q^n\), either \(x \notin \text{desc}_c(C)\) or
\[
\bigcap_{T \subseteq C, |T| \leq c, \text{ s.t. } x \in \text{desc}(T)} T \neq \emptyset.
\]

Note that for an IPP code the intersection of all coalitions of size \(\leq c\) that can
generate a given pirated word is nonempty. In particular, the codewords that lie
the intersection (1) belong to the coalition that generated the pirated word and can
be accused as traitors. Hence, for a code of size \(M\), the identification process runs
in time \(O(M^c)\) in the general case.

**Definition 6.** A code \(C\) has the c-traceability property (c-TA) if for all subsets
\(T \subseteq C\) of size at most \(c\), if \(x \in \text{desc}(T)\), then there exists \(t \in T\) such that \(d(x, t) < d(x, y)\)
for all \(y \in C \setminus T\).

That is, in a c-TA code the closest codeword to a descendant of a subset (coalition)
\(T\), in terms of Hamming distance, is in \(T\).

It is easy to see that every TA code is an IPP code. The main benefit of using
TA codes is that the identification process runs in time \(O(M)\). Nevertheless the
TA property imposes more restrictions to the code than the IPP property; see for
example [20].

The concepts of IPP and TA codes were originated in [4] (later in [5]), however
no specific name was given to such codes. IPP codes where further studied in [10].
There the authors coined the term “IPP,” that has been widely adopted in the
crypto literature. Also, IPP and TA codes have been investigated in [20] under the
names presented here.

A simple sufficient condition for an \((n, M, d)\)-code to posses the TA property was
presented in [4, 5]. Namely, if \(d > (1 - 1/c^2)n\), the code is c-TA. In addition, the
following chain of implications are also well-known results:

\[ d > (1 - 1/c^2)n \Rightarrow \theta_{1,c} > (1 - 1/c)n \Rightarrow c\text{-TA} \Rightarrow c\text{-IPP} \Rightarrow (c,c)\text{-separating}. \]

These results were presented later in the form of a theorem in [20].

2.1. MDS codes with tracing properties. Codes that achieve equality in the Singleton bound, \( M \leq q^{n-d+1} \), are called maximum distance separable (MDS) codes. Therefore, linear MDS codes satisfy \( k = n - d + 1 \).

Even though the results (2) are well-known and obvious, it took several years to prove the converse of the first and second implication for linear MDS codes. This result first appeared in [11].

**Theorem 1** ([11, Theorem 2.3]). Let \( C \) be an MDS \([n,k,d]\)-code over the finite field \( \mathbb{F}_q \) such that \( n \leq q + 1 \). Then, for \( c \geq 2 \), \( C \) is a \( c\text{-TA} \) code if and only if \( d > (1 - 1/c^2)n \).

Putting this together with (2), we have that for linear MDS codes

\[ d > (1 - 1/c^2)n \iff \theta_{1,c} > (1 - 1/c)n \iff c\text{-TA}. \]

A well-known family of linear MDS codes are Reed-Solomon codes [15]. We first give the following definition.

**Definition 7.** Let \( P = \{p_1, \ldots, p_n\} \) be a subset of \( n \) elements of \( \mathbb{F}_q \), called evaluation points. We define the \([n,k,d]\)-code \( G(n,k) \) over \( \mathbb{F}_q \) as

\[ G(n,k) = \{(f(p_1), \ldots, f(p_n)) : f(x) \in \mathbb{F}_q[x], \deg f(x) < k\}. \]

Note that \( G(n,k) \) is a linear MDS code, irrespective of the choice of the set of evaluation points. If \( P \) is the multiplicative group of the ground field, \( \mathbb{F}_q^\ast \), then \( G(n,k) \) is known as Reed-Solomon code, denoted RS\((n,k)\). If \( P = \mathbb{F}_q \), then it is known as extended Reed-Solomon code.

In [18, 19], the authors posed the following question, which has motivated our work in this paper.

**Question 1.** Is it the case that \( d > (1 - 1/c^2)n \) for all \( c\text{-IPP} \) Reed-Solomon codes of length \( n \) and minimum distance \( d \)?

In fact, we will see below that, for many families of Reed-Solomon codes, the condition \( d \leq (1 - 1/c^2)n \) implies not only losing the \( c\text{-IPP} \) property, but also losing the \((c,c)\)-separating property. Hence, the converse of all the implications in (2) holds for such families.

The observation that we have just made suggests a generalization of the previous question as follows.

**Question 2.** Is it the case that \( d > (1 - 1/c^2)n \) for all \((c,c)\)-separating \( G(n,k) \) codes of Definition 7 of length \( n \) and minimum distance \( d \)?

Observe that for the same subset of evaluation points and any \( k' \leq k \), we have \( G(n,k') \subseteq G(n,k) \subseteq \mathbb{F}_q^n \). Therefore, to provide a positive answer to the question above, we only need to show that \( G(n,k = \lceil n/c^2 \rceil + 1) \) has \( \theta_{c,c} = 0 \), for every possible pair of values \( q \) and \( c \), and every possible choice of evaluation points.

The motivation of these questions arises from the fact that the amount of information (fingerprint) that we can embed in a digital document is limited. Assume that we can embed no more than \( n \) symbols from \( \mathbb{F}_q \). Then, there exists a \( c\text{-TA} \) Reed-Solomon code that can allocate \( q^k \) users, for any \( k < n/c^2 + 1 \). For the same
value of \( n \), if the distributor needs to allocate more users, then by Theorem 1 the code will not be \( c \)-TA. In this situation, is there a chance that we can still identify traitors? The remark made above suggests that for \( k \geq n/c^2 + 1 \) there are neither \( c \)-IPP nor \((c, c)\)-separating codes, hence identification with zero-error probability would not be possible.

In this paper we are mainly concerned with giving an answer to these questions when \( G(n, k) \) is a Reed-Solomon code. However, we will also exploit the constructions presented to give some answers for other \( G(n, k) \) codes.

2.2. Previous results. An answer to the questions above for the case \( c = 2 \) can be found in [17]. It is written there that in 1986 G. D. Katsman and S. N. Litsyn applied Mattson-Solomon polynomials and linearized polynomials to Reed-Solomon codes obtaining

\[
\theta_{2,2} = n - 4(k - 1).
\]

Taking \( k \geq n/4 + 1 \), we have \( \theta_{2,2} = 0 \). Therefore \((2, 2)\)-separating \( \Rightarrow d > (1 - 1/4)n \), which means that the converse of every implication in (2) holds for Reed-Solomon codes and the particular case \( c = 2 \). Unfortunately, the proof of this nice result has not been published.

Also, in [18, 19] a custom-made construction of \( G(n, k) \) codes is presented, defined over sufficiently large alphabets. They have minimum distance \( d = (1 - 1/c^2)n \) and they are not \((c, c)\)-separating. Nevertheless, no specific relation is given between the code parameters.

In [8] a related result is presented for Reed-Solomon codes \( RS(n, k) \) such that their ground field contains the \((k - 1)\)th roots of unity. The idea there was to restate the separating condition algebraically, as a system of equations. From [8, Theorem 7], and from the proof provided by the authors, the following corollary is immediate.

**Corollary 1.** Let \( RS(n, k) \) be a Reed-Solomon code over \( \mathbb{F}_q \) with minimum distance \( d \). If \( n - d \) divides \( q - 1 \), then the code is \((c, c)\)-separating if and only if \( d > (1 - 1/c^2)n \).

3. Equivalence of the tracing properties of Reed-Solomon codes

We begin by showing some upper and lower bounds of \( \theta_{c_1, c_2} \) for linear and MDS codes. These bounds were presented for the particular cases \( c_1 = c_2 = 2 \) in [17], and \( c_1 = 1, c_2 \) arbitrary in [11].

**Lemma 1.** Let \( C \) be an \([n, k, d]\)-code and let \( c_1, c_2 \) be two positive integers. Then,

\[
\max\{0, d - (c_1c_2 - 1)(n - d)\} \leq \theta_{c_1, c_2} \leq \max\{0, d - (c_1 + c_2 - 2)(k - 1)\}.
\]

If \( C \) is additionally an MDS code and \( c_1, c_2 \geq 2 \), then

\[
\max\{0, d - (c_1c_2 - 1)(n - d)\} \leq \theta_{c_1, c_2} \leq \max\{0, d - (c_1 + c_2 - 2)(k - 1) - c_1 - c_2 + 3\}.
\]

**Proof.** Let \( T_1, T_2 \) be any two disjoint subsets of \( C \) of size \( c_1 \) and \( c_2 \) respectively. Note that two different codewords of \( C \) agree in at most \( n - d \) coordinates. Also, the maximum number of coordinates where the codewords of two disjoint subsets of \( C \) have a common element (nonseparating coordinates) is \( n - \theta_{c_1, c_2} \). Hence, for every codeword \( t \in T_1 \), the codewords in \( T_2 \) can agree together in at most \( c_2(n - d) \) coordinates of \( t \). Since \( T_1 \) has \( c_1 \) elements, we have that \( n - \theta_{c_1, c_2} \leq c_2(n - d) \), which proves the lower bounds.
To prove the upper bounds, construct two subsets $T_1$ and $T_2$ in the following way. First, take any $t_1, t_2 \in C$ such that $d(t_1, t_2) = d$. Such codewords exist, by definition of minimum distance. Put $t_1$ into $T_1$ and $t_2$ into $T_2$. For the remaining $c_1 - 1$ codewords of $T_1$, choose codewords such that agree with $t_2$ in $k - 1$ disjoint coordinates, in the positions where $t_1$ and $t_2$ differ. This can be done by virtue of [11, Lemma 2.2]. Operate in the same way for the codewords of the set $T_2$. Therefore, the number of coordinates where the elements of $T_1$ and $T_2$ have a common element is $n - d + (c_1 + c_2 - 2)(k - 1)$, which proves the upper bound in (3). Moreover, if the code is MDS and $c_1, c_2 \geq 2$, we can set an additional coordinate of every codeword of $T_2 \setminus \{t_2\}$ to agree with a coordinate of a given codeword $t'_1 \in T_1 \setminus \{t_1\}$. Similarly, we can set an additional coordinate of each codeword of $T_1 \setminus \{t_1, t'_1\}$ to agree with a coordinate of any other codeword in $T_2 \setminus \{t_2\}$. This reduces the number of nonseparating coordinates in $c_1 + c_2 - 3$, and proves the upper bound in (4).

Consider the case $c_1 = c_2 = c$. For linear MDS codes, and from the previous lemma, it is clear that whenever $d - 2(c - 1)(k - 1) - 2c + 3 \leq 0$, we have $\theta_{c,c} = 0$. Therefore the code is not $(c,c)$-separating. Also, when $d - (c^2 - 1)(n - d) > 0$, then $\theta_{c,c} > 0$ and the code is $(c,c)$-separating. In fact, the latter condition implies that the code is $c$-TA. In conclusion, there is an “uncertainty interval,” in terms of $d$, in which the $(c,c)$-separating property remains to be characterized, namely

$$\frac{2(c - 1)n + 2c - 3}{2c - 1} < d \leq (1 - 1/c^2)n.$$

3.1. Cyclic codes with multiplicative subgroups in the ground field. Whenever the set of evaluation points $P$ is a multiplicative subgroup with generator element $\alpha$, the code $G(n,k)$ is (linearly equivalent to) a cyclic code. We denote by $t^{(i)}$ the cyclic rotation of $t \in \mathbb{F}_q^n$ in $i$ coordinates to the right. In this case, it is easy to see that if the polynomial $f(x)$ generates the codeword $t \in G(n,k)$, the polynomial $f(\alpha^{-i}x)$ generates the codeword $t^{(i)}$.

The following result, together with (2), generalizes Corollary 1 for any $G(n,k)$ code generated with a multiplicative subgroup of evaluation points.

**Proposition 1.** Let $P$ be a multiplicative subgroup of $\mathbb{F}_q^*$. Also, let $G(n,k)$ be the code from Definition 7 generated with the set of evaluation points $P$, of minimum distance $d$. If $n - d$ divides $n$ and $d \leq (1 - 1/c^2)n$, then the code is not $(c,c)$-separating.

**Proof.** We need to show that under the conditions stated the code contains a $(c,c)$-nonseparated configuration.

Denote $r = n/(k - 1)$, and consider the polynomial

$$f(x) = \prod_{i=0}^{k-2} (\alpha^{ir} x - 1),$$

where $\alpha$ is a generator of $P$. Note that $f(x)$ is a polynomial of degree $k - 1$. Hence, the codeword generated from $f(x)$, say $t$, is in $G(n,k)$. It is easy to see that $f(\alpha^{-ih}x) = f(x)$ for any integer $h$. Hence, $t^{(rh)} = t$. This, together with the fact that the polynomial has degree $k - 1$, means that the codeword $t$ consists of $k - 1$ concatenations of a vector of $r$ distinct elements, say $b = (b_1, \ldots, b_r)$. Now take $c' = \min\{c, r\} \leq c$ and construct the following set of codewords:

$$T_1 = \{t(\alpha^{ic'}) : 0 \leq i < [r/c']\}.$$
From the starting assumptions, $n - d = k - 1 \geq n/c^2$, which implies that $\lvert T_1 \rvert = \lceil r/c' \rceil \leq c' \leq c$. Since $t$ is the repeated concatenation of the vector $b$, of length $r$, and $c'[r/c'] \geq r$, it is clear that for all $1 \leq j \leq n$, there exists a codeword $t^{(ic')}$ in $T_1$ such that $t^{(ic')}_{j} \in \{b_1, \ldots, b_{c'}\}$.

The code $G(n, k)$ contains all the constant codewords in $\mathbb{F}_q^n$, hence one can construct the set

$$T_2 = \{(b_i, \ldots, b_1) : 1 \leq i \leq c'\},$$

of size $c' \leq c$, which is disjoint from $T_2$. Since for $T_1$ and $T_2$ every coordinate is not separating, then $\theta(T_1, T_2) = 0$. It follows that the code is not $(c, c)$-separating. \qed

**Corollary 2.** Let $RS(n, k)$ be a Reed-Solomon code over $\mathbb{F}_q$ with minimum distance $d$. If $c \geq \sqrt{q - 1}$ and $d \leq (1 - 1/c^2)n$, then the code is not $(c, c)$-separating.

**Proof.** From Proposition 1, we have that if $k \geq \lceil n/c^2 \rceil + 1$ and $\lceil n/c^2 \rceil$ divides $n$, then $G(n, k)$ is not $(c, c)$-separating. Reed-Solomon codes have $n = q - 1$. Taking $c \geq \sqrt{q - 1}$, we have $\lceil n/c^2 \rceil = 1$, and the proof follows. \qed

It is well-known \cite{18, 19} that $c$-IPP codes over $\mathbb{F}_q$ do not exist for $c \geq q$. The previous corollary gives a tighter bound for the case of Reed-Solomon codes.

### 3.2. Coalition Size Dividing the Ground Field Size

This section contains the main result of our paper, which comes in the form of the following theorem.

**Theorem 2.** Let $RS(n, k)$ be a Reed-Solomon code over $\mathbb{F}_q$ with minimum distance $d$. If $c$ divides $q$ and $d \leq (1 - 1/c^2)n$, then the code is not $(c, c)$-separating.

In fact, from the proof of the theorem, one can easily see that it is valid for any code $G(n, k)$ with an arbitrary set of evaluation points $P$ of size $q - c^2 < |P| \leq q$.

The proof is based on a special class of polynomials known as linearized polynomials.

**Definition 8.** A polynomial of the form

$$L(x) = \sum_{i=0}^{h} l_{i} x^{q^i},$$

with coefficients $l_{i}$ in an extension field $\mathbb{F}_{q^m}$ of $\mathbb{F}_q$ is called a linearized polynomial over $\mathbb{F}_{q^m}$.

Let us present some important, well-known facts \cite{13} about linearized polynomials. First, if $L(x)$ is a linearized polynomial over $\mathbb{F}_{q^m}$, then

$$L(\alpha \xi + b \beta) = aL(\alpha) + bL(\beta),$$

for all $\alpha, \beta \in \mathbb{F}_{q^m}$ and all $a, b \in \mathbb{F}_q$. Thus, the polynomial function $L : \mathbb{F}_{q^m} \rightarrow \mathbb{F}_{q^m}$, defined as $x \mapsto L(x)$, is a linear operator on $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. Also, the following result will be useful in our proof below.

**Theorem 3** (\cite{13, Theorem 3.52}). Let $U$ be a vector subspace of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. Then for any nonnegative integer $s$, the polynomial

$$L(x) = \prod_{\mu \in U} (x - \mu)^{q^s}$$

is a linearized polynomial over $\mathbb{F}_{q^m}$.

For our purposes, we will deal with linearized polynomials over $\mathbb{F}_q = \mathbb{F}_{p^m}$ such that their roots also lie in $\mathbb{F}_q$. We are now in position to prove Theorem 2.
Proof of Theorem 2. We prove the theorem by finding a \((c, c)-\)nonseparated configuration again.

If \(c^2 > q\), the code is not \((c, c)-\)separating by Corollary 2. Henceforth, we shall assume that \(c^2 \leq q = p^m\). This, together with the fact that \(c\) divides \(q\), implies that \(c^2\) also divides \(q\), that is \(c = p^r\) for some \(r \leq m/2\). For any \(n\) such that \(q - c^2 < n \leq q\), and from the fact that \(d \leq (1 - 1/c^2)n\), we have that the code contains, at least, all the codewords generated from polynomials of any degree up to \(q/c^2 = p^{m-2r}\).

Now, consider the polynomial
\[
L(x) = \prod_{\mu \in U} (x - \mu),
\]
where \(U\) is a vector subspace of \(\mathbb{F}_q\) over \(\mathbb{F}_p\) of dimension \(m - 2r\) and size \(q/c^2\). Note that \(L(x)\) is a linearized polynomial by Theorem 3. Also, from (5), the polynomial function \(L : \mathbb{F}_q \to \mathbb{F}_q\) is an homomorphism with \(|\ker L| = q/c^2\) and \(|\text{im } L| = c^2\).

Clearly, \(\text{im } L\) is a vector subspace of \(\mathbb{F}_q\) of dimension \(2r\).

Now, take a vector subspace \(B \subseteq \text{im } L\) of dimension \(r\) and size \(c\). Regard \(B\) as an additive subgroup of \(\mathbb{F}_q\) and consider its \(c\) cosets, which partition \(\text{im } L\):
\[
B_i = \beta_i + B, \quad 1 \leq i \leq c.
\]
We can assume without loss of generality that \(\beta_1 = 0\). Now consider the following \(c\) polynomials
\[
f_i(x) = L(x) - \beta_i, \quad 1 \leq i \leq c.
\]
Observe that for every \(\gamma \in \mathbb{F}_q\) there is exactly one polynomial \(f_i(x)\) such that \(f_i(\gamma) \in B\). To see this, note that if \(L(\gamma)\) lies in the coset \(B_i\) of \(\text{im } L\), that is \(L(\gamma) = \beta_i + b\) for some \(b \in B\), then the polynomial \(f_i(\gamma) = L(\gamma) - \beta_i = \beta_i + b - \beta_i\) evaluates to \(b \in B\). The fact that the \(c\) cosets \(B_i\) partition \(\text{im } L\) into disjoint subsets implies that there is only one \(f_i(x)\) satisfying this condition.

Now, consider the set of codewords
\[
T_1 = \{t_1, \ldots, t_c\},
\]
where \(t_i\) is the codeword generated from the polynomial \(f_i(x)\), and the set of \(c\) constant codewords
\[
T_2 = \{(b, \ldots, b) : b \in B\},
\]
Obviously, \(T_1\) and \(T_2\) are disjoint, because \(\text{deg } f_i(x) \geq 1\). Also, \(\theta(T_1, T_2) = 0\), which proves that the code is not \((c, c)-\)separating.

This construction applies whenever the code contains, at least, all the codewords generated from polynomials of degree up to \(q/c^2\). Since \(k - 1 \geq (q - 1)/c^2\), this happens in particular for the Reed-Solomon code. Finally, we remark that one can choose an arbitrary coset \(\beta_i + B\) for the generation of the constant codewords of the set \(T_2\).

However, there are other families of Reed-Solomon codes that can benefit from the constructions of the nonseparated configuration presented in the previous proof.

Proposition 2. Let \(\text{RS}(n, k)\) be a Reed-Solomon code over \(\mathbb{F}_q\) with minimum distance \(d\). If
\[
c' = \sqrt[\frac{q}{c^2}]\]
is an integer and \(d \leq (1 - 1/c^2)n\), then the code is not \((c, c)-\)separating.
Proof. Note that the code contains codewords generated from polynomials of degree at least \( \left\lceil \frac{q}{c^2} \right\rceil = q/c^2 \). Also, \( c' \) and \( c^2 \) must divide \( q \), which is implied by (6). Using the construction from the proof of Theorem 2, one can easily see that the code is not \((c', c')\)-separating. The proof follows by noting that \( c' \leq c \).

3.3. Summary of Results for Reed-Solomon Codes. We summarize here the results shown in the paper for the case of Reed-Solomon codes, \( RS(n, k) \).

1. For any

\[
d \leq \frac{2(c-1)n + 2c - 3}{2c - 1}
\]

the code is not \((c, c)\)-separating.

2. The implication

\[
d > (1 - 1/c^2)n \iff (c, c)\text{-separating}
\]

is true for families of Reed-Solomon codes when any of the following situations occurs: (a) \( c = 2 \); (b) \( c^2 > q \); (c) \( k - 1 \) divides \( n \); and (d) \( \sqrt{q/\left\lceil q/c^2 \right\rceil} \) is an integer value.

4. Example

Let us illustrate the proof of Theorem 2 with the following example. Consider the finite field \( \mathbb{F}_{27} = \mathbb{F}_3[x]/(x^3 + 2x + 1) \) with primitive element \( \alpha = \bar{x} \). Let \( c = 3 \) and take the Reed-Solomon code \( RS(n = 26, k = 4) \). First, we take the subgroup (or vector space over \( \mathbb{F}_3 \)) \( U = \{0, 1, \alpha^{13}\} \) and construct the linearized polynomial

\[
L(x) = (x - 0)(x - 1)(x - \alpha^{13}) = x^3 + \alpha^{13}x.
\]

The codeword generated from \( L(x) \) is

\[
(0, \alpha^{13}, \alpha^9, \alpha^{13}, \alpha^3, \alpha^{16}, \alpha, \alpha^3, \alpha^{22}, \alpha^{13}, \alpha, \alpha, \alpha^9, 0, 1, \alpha^{22}, 1, \alpha^{16}, \alpha^3, \alpha^{14}, \alpha^9, 1, \alpha^{14}, \alpha^{14}, \alpha^{14}, \alpha{22}),
\]

where it can be read that \( \text{im } L = \{0, 1, \alpha, \alpha^3, \alpha^9, \alpha^{13}, \alpha^{14}, \alpha^{16}, \alpha^{22}\} \). Since \( c^2 = |\text{im } L| \), we take for example the subgroup \( B = \{0, 1, \alpha^{13}\} \leq \text{im } L \) of \( c \) elements and its \( c \) cosets:

\[
B_1 = \beta_1 + S = \{0, 1, \alpha^{13}\} \\
B_2 = \beta_2 + S = \{\alpha, \alpha^3, \alpha^{9}\} \\
B_3 = \beta_3 + S = \{\alpha^{14}, \alpha^{16}, \alpha^{22}\},
\]

where \( \beta_1 = 0, \beta_2 = \alpha \) and \( \beta_3 = \alpha^{14} \). Now consider the polynomials \( f_i(x) = L(x) - \beta_i \), for \( 1 \leq i \leq c \). Due to space constraints, we will only show the first 16 coordinates of their corresponding codewords, which are

\[
(0, \alpha^{13}, \alpha^9, \alpha^{13}, \alpha^3, \alpha^{16}, \alpha, \alpha^3, \alpha^{22}, \alpha^{13}, \alpha, \alpha, \alpha^9, 0, 1, \alpha^{22}, \ldots) \\
(\alpha^{14}, \alpha^{22}, 1, \alpha^{22}, \alpha^{13}, \alpha^9, 0, \alpha^{13}, \alpha^{3}, \alpha^{22}, 0, 0, 1, \alpha^{14}, \alpha^{16}, \alpha^{3}, \ldots) \\
(\alpha, \alpha^3, \alpha^{16}, \alpha^3, \alpha^{22}, 1, \alpha^{14}, \alpha^{22}, \alpha^{13}, \alpha^{3}, \alpha^{14}, \alpha^{16}, \alpha, \alpha^9, \alpha^{13}, \ldots),
\]

where the elements of the coset \( B_2 \) have been highlighted. These codewords constitute the set \( T_1 \) from the proof of Theorem 2. Now, call \( T_2 \) the set formed by the constant codewords

\[
(\alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \ldots) \\
(\alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \alpha^3, \ldots)
\]

Advances in Mathematics of Communications Volume 6, No. 4 (2012), 467–478
that is, the codewords with constant elements in the coset $B_2$. As we have just shown, for every coordinate, one of the three codewords in $T_1$ has an element from $B_2$. Hence, $\theta(T_1, T_2) = 0$, and both coalitions can generate the same descendant: 

$$(\alpha, \alpha^3, \alpha^9, \alpha^3, \alpha^3, \alpha^3, \alpha, \alpha^3, \alpha^9, \alpha^9, \alpha^3, \ldots).$$

Note that if we had considered the extended Reed-Solomon code instead, it would also have had $k = 4$, and the additional coordinate involved would also have satisfied the same property.

**Table 1.** Some known families of Reed-Solomon codes $RS(n = q - 1, k = \lceil n/c^2 + 1\rceil)$ where $d > (1 - 1/c^2)n \Leftrightarrow (c, c)$-separating:

(a) $c = 2$; (b) $c^2 > q$; (c) $k - 1$ divides $n$; and (d) $\sqrt{q}/[q/c^2]$ is an integer value.

| $c = 2$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14-15 | 16 | 17-18 | 19 | 20-22 | 23-24 | 25 | 26 | 27 | 28-31 | 32 | 33 | 34-46 | $\geq 47$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$q = 64$</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
</tr>
<tr>
<td>81</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>(c)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>(c)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>625</td>
<td>(c)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>729</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>2187</td>
<td>(d)</td>
<td>(c)</td>
<td>(c)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td>(d)</td>
<td></td>
</tr>
</tbody>
</table>

Illustratively, in Table 1 we show some families of Reed-Solomon codes, for certain values of $c$ and $q$, satisfying $d > (1 - 1/c^2)n \Leftrightarrow (c, c)$-separating, i.e., of dimension $k = \lceil (q-1)/c^2 + 1\rceil$. This, together with several computer-assisted searches, suggests a positive answer to Question 1.

5. Conclusion

In this paper we have discussed the tracing properties of Reed-Solomon codes. Our main goal was to give an answer to the question posed by Silverberg et al. in [18, 19]: *Is it the case that $d > (1 - 1/c^2)n$ for all $c$-IPP Reed-Solomon codes of length $n$ and minimum distance $d$?*

We have given a positive answer for some families of Reed-Solomon, when $c$ divides the field size. Also, we have benefited from the proposed constructions to extend the results to other families of punctured Reed-Solomon codes. Obviously this does not provide a full answer to the question but hopefully it gives some hints that may be useful in finding the final response.
ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their thorough and valuable comments, which motivated major improvements on this manuscript.

REFERENCES


Received October 2011; revised June 2012.

E-mail address: jose.moreira@entel.upc.edu
E-mail address: marcel@entel.upc.edu
E-mail address: soriano@entel.upc.edu