Tracebility of Reed Solomon Codes

Marcel Fernandez and Jorge J. Urroz, UPC, Barcelona

Valencia, July, 2019

The fingerprinting technique is used to determine the illegal redistribution of (digital) contents.

The fingerprinting technique is used to determine the illegal redistribution of (digital) contents.

Several hundred years ago logarithm tables were protected by fingerprinting them. The idea was to introduce tiny errors in the insignificant digits of $\log x$ (i.e., tenth digit right of the decimal point) of $\log x$ for a few random values of x. The set of x was different for each user.

The fingerprinting technique is used to determine the illegal redistribution of (digital) contents.

Several hundred years ago logarithm tables were protected by fingerprinting them. The idea was to introduce tiny errors in the insignificant digits of $\log x$ (i.e., tenth digit right of the decimal point) of $\log x$ for a few random values of x. The set of x was different for each user.

It consists of embedding imperceptible marks (fingerprints) in certain positions of the content which identifies the user.

When fingerprinting digital data one must address the problem of collusion.

When fingerprinting digital data one must address the problem of collusion.

Run diff on their two files to discover all the locations where the files differ.

When fingerprinting digital data one must address the problem of collusion.

Run diff on their two files to discover all the locations where the files differ. Change its value could involve an innocent user.

To avoid collusion we will use fingerprinting codes.

To avoid collusion we will use fingerprinting codes.

We will make two assumptions.

To avoid collusion we will use fingerprinting codes.

We will make two assumptions.

Marking assumption: Undetectable position remains unmodified.

To avoid collusion we will use fingerprinting codes.

We will make two assumptions.

Marking assumption: Undetectable position remains unmodified. Notice that two users could only detect those marks in which their copies differ.

To avoid collusion we will use fingerprinting codes.

We will make two assumptions.

Marking assumption: Undetectable position remains unmodified. Notice that two users could only detect those marks in which their copies differ.

In the detectable positions the traitors must insert one of their symbols. (narrow-sense envelope model) [Barg, Blakley, Kabatiansky 2003]

If there are L marks and each mark can have s different states, a fingerprint is a word of size L in an alphabet of size s.

If there are L marks and each mark can have s different states, a fingerprint is a word of size L in an alphabet of size s.

For us, the alphabet will be a Finite Field $\mathbb{F}_q$, and we consider (n, M, d) codes $C \in \mathbb{F}_q^n$ of size M and minimum distance d. When C is a subspace of dimension k we say that C is an (n, k, d) linear code.

If there are L marks and each mark can have s different states, a fingerprint is a word of size L in an alphabet of size s.

For us, the alphabet will be a Finite Field $\mathbb{F}_q$, and we consider (n, M, d) codes $C \in \mathbb{F}_q^n$ of size M and minimum distance d. When C is a subspace of dimension k we say that C is an (n, k, d) linear code.

Definition. A code is a c-frame proof code, c-FP, if no coallition can construct the word of an innocent user.

Example.

$$(2, 1, 1, 4, 3, 1, 2, 3, 1, 4)$$
$$(4, 1, 1, 4, 3, 4, 2, 3, 1, 1)$$
$$(2, 3, 1, 4, 3, 2, 2, 2, 1, 1)$$

$$(2, 3, 1, 4, 3, 4, 2, 3, 1, 4)$$

Example.

$$(2,1,1,4,3,1,2,3,1,4)$$
$$(4,1,1,4,3,4,2,3,1,1)$$
$$(2,3,1,4,3,2,2,2,1,1)$$

$$(2,3,1,4,3,4,2,3,1,4)$$

For $T \subset C$,

$$desc(T) = \{x \in \mathbb{F}_q^n : \text{ for each } 1 \leq i \leq n, x_i = t_i \text{ for some } t \in T\}$$

This is the set of pirate codewords that a coalition T can generate.

Example.

$$(2,1,1,4,3,1,2,3,1,4)$$
$$(4,1,1,4,3,4,2,3,1,1)$$
$$(2,3,1,4,3,2,2,2,1,1)$$

$$(2,3,1,4,3,4,2,3,1,4)$$

For $T \subset C$,

$desc(T) = \{x \in \mathbb{F}_q^n : \text{ for each } 1 \leq i \leq n, x_i = t_i \text{ for some } t \in T\}$

This is the set of pirate codewords that a coalition T can generate. C is c-FP if for $|T| \leq c$, desc(T)$\cap C = T$

Is desirable that a fingerprinting code is c-FP for c as big as possible. But is not enough. A coalition of c users could claim that a pirate copy was made by other completely disjoint coalition.

Is desirable that a fingerprinting code is c-FP for c as big as possible. But is not enough. A coalition of c users could claim that a pirate copy was made by other completely disjoint coalition.

Definition A coordinate i separates A from B if $a_i \neq b_i$ for any $a \in A$ and $b \in B$.

Is desirable that a fingerprinting code is c-FP for c as big as possible. But is not enough. A coalition of c users could claim that a pirate copy was made by other completely disjoint coalition.

Definition A coordinate i separates A from B if $a_i \neq b_i$ for any $a \in A$ and $b \in B$.

$$(2, 1, 1, 4, 3, 1, 2, 3, 1, 4)$$
$$(4, 1, 1, 4, 3, 4, 2, 3, 1, 1)$$

$$(2, 3, 1, 4, 3, 2, 2, 2, 1, 1)$$
$$(2, 3, 1, 4, 3, 4, 2, 3, 1, 4)$$

A code is (c_1, c_2)-separating if for any two sets $A, B \in C$, $|A| = c_1, |B| = c_2$ there exist a coordinate that separates them

A code is (c_1, c_2)-separating if for any two sets $A, B \in C$, $|A| = c_1, |B| = c_2$ there exist a coordinate that separates them

In this context a $(c, 1)$-separating code is a c-FP code while the $(c, c)-$ separating codes are the c-secure frameproof codes, distinguishing coalitions.

Ok, but still not enough...

.

Ok, but still not enough...

The code $\{(0,0,0),(0,1,1),(1,1,0),(1,0,1)\}$ is a $(2,2)$-separating code, but it is impossible to identify the traitor coalition that generated $(0,1,0)$.

.

Ok, but still not enough...

The code $\{(0,0,0),(0,1,1),(1,1,0),(1,0,1)\}$ is a $(2,2)$-separating code, but it is impossible to identify the traitor coalition that generated $(0,1,0)$.

Definition. A code has the c-identifiable parent property, c-IPP, if for all $x \in \mathbb{F}_q^n$, $x \notin desc(T)$ or

$$\cap_{x \in desc(T)} T \neq \emptyset.$$

Always $|T| \leq c$.

Ok, but still not enough...

The code $\{(0,0,0),(0,1,1),(1,1,0),(1,0,1)\}$ is a $(2,2)$-separating code, but it is impossible to identify the traitor coalition that generated $(0,1,0)$.

Definition. A code has the c-identifiable parent property, c-IPP, if for all $x \in \mathbb{F}_q^n$, $x \notin desc(T)$ or

$$\cap_{x \in desc(T)} T \neq \emptyset.$$

Always $|T| \leq c$.

The running time for identification of traitors is $O(\binom{M}{c})$.

Definition. The code has the c-traceability property, c-TA, if whenever $x \in desc(T)$, then $d(x,t) < d(x,y)$ for some $x \in T$ and any $y \in C/T$.

Definition. The code has the c-traceability property, c-TA, if whenever $x \in desc(T)$, then $d(x,t) < d(x,y)$ for some $x \in T$ and any $y \in C/T$.

The running time becomes $O(M)$. Moreovoer, c-TA$\Rightarrow$ c-IPP, but also more restrictive.

Definition. The code has the c-traceability property, c-TA, if whenever $x \in desc(T)$, then $d(x,t) < d(x,y)$ for some $x \in T$ and any $y \in C/T$.

The running time becomes $O(M)$. Moreovoer, c-TA$\Rightarrow$ c-IPP, but also more restrictive....Right?

Definition. The code has the c-traceability property, c-TA, if whenever $x \in desc(T)$, then $d(x,t) < d(x,y)$ for some $x \in T$ and any $y \in C/T$.

The running time becomes $O(M)$. Moreovoer, c-TA$\Rightarrow$ c-IPP, but also more restrictive....Right?

The traceability property imposes a condition in the distance. In particular any code with $d > (1 - 1/c^2)n$ is TA and, in fact

$$d > n - n/c^2 \Rightarrow c - \text{TA}.$$

Proof. The descendent x coincide with at least n/c with an element of the coalition and the distance is smaller than $n - n/c$.

Proof. The descendent x coincide with at least n/c with an element of the coalition and the distance is smaller than $n - n/c$.

Since the distance is $d > n - n/c^2$, an element of the code coincide with at most n/c^2 with any element of T, and less than n/c with $desc(T)$ so the distance to x is bigger than $n - n/c$

So, we have

$$d > n - n/c^2 \Rightarrow c - \mathrm{TA} \Rightarrow c - \mathrm{IPP} \Rightarrow (c, c) - \text{separating}.$$

So, we have

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

When dealing with linear MDS codes ($k = n - d + 1$), c-TA is the same as $d > n - n/c^2$. In fact, for many families of Reed Solomon codes, $d \leq n - n/c^2$ implies $(c, c) - \textit{inseparability}$.

So, we have

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

When dealing with linear MDS codes ($k = n - d + 1$), c-TA is the same as $d > n - n/c^2$. In fact, for many families of Reed Solomon codes, $d \leq n - n/c^2$ implies $(c, c) - \textit{inseparability}$.

Given $P = \{p_1, \ldots, p_n\} = \mathbb{F}_q^*$, or $P = \mathbb{F}_q$ the $[n, k, d]$ Reed Solomon code is definded as

$$RS(n, k) = \{(f(p_1), \ldots, f(p_n)) : \deg(f(x) < k\}.$$

So, we have

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

When dealing with linear MDS codes ($k = n - d + 1$), c-TA is the same as $d > n - n/c^2$. In fact, for many families of Reed Solomon codes, $d \leq n - n/c^2$ implies $(c, c) - \textit{inseparability}$.

Given $P = \{p_1, \ldots, p_n\} = \mathbb{F}_q^*$, or $P = \mathbb{F}_q$ the $[n, k, d]$ Reed Solomon code is definded as

$$RS(n, k) = \{(f(p_1), \ldots, f(p_n)) \, : \, \deg(f(x) < k\}.$$

Question. Is it true that a $(c, c)-$separating Reed Solomon code has $d > n - n/c^2$?

So, we have

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

When dealing with linear MDS codes ($k = n - d + 1$), c-TA is the same as $d > n - n/c^2$. In fact, for many families of Reed Solomon codes, $d \leq n - n/c^2$ implies $(c, c) - \textit{inseparability}$.

Given $P = \{p_1, \ldots, p_n\} = \mathbb{F}_q^*$, or $P = \mathbb{F}_q$ the $[n, k, d]$ Reed Solomon code is definded as

$$RS(n, k) = \{(f(p_1), \ldots, f(p_n)) : \deg(f(x) < k\}.$$

Question. Is it true that a $(c, c)-$separating Reed Solomon code has $d > n - n/c^2$?

It is enough to prove it for $k = \lceil n/c^2 \rceil + 1$.

So, we have

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

When dealing with linear MDS codes ($k = n - d + 1$), c-TA is the same as $d > n - n/c^2$. In fact, for many families of Reed Solomon codes, $d \leq n - n/c^2$ implies $(c, c) - \textit{inseparability}$.

Given $P = \{p_1, \ldots, p_n\} = \mathbb{F}_q^*$, or $P = \mathbb{F}_q$ the $[n, k, d]$ Reed Solomon code is definded as

$$RS(n, k) = \{(f(p_1), \ldots, f(p_n)) \, : \, \deg(f(x) < k\}.$$

Question. Is it true that a $(c, c)-$separating Reed Solomon code has $d > n - n/c^2$?

It is enough to prove it for $k = \lceil n/c^2 \rceil + 1$.

If it is true, then it is not possible to identify traitors when $k \geq n/c^2 + 1$.

Theorem. If $(n-d)|n$ and $d \leq n - n/c^2$, then $RS(n,k)$ is not $(c,c)-$separable.

Theorem. If $(n-d)|n$ and $d \leq n - n/c^2$, then $RS(n,k)$ is not $(c,c)-$separable.

Proof. Let $r = \frac{n}{k-1}$ and consider the polynomial $f(x) = \prod_0^{k-2}(\alpha^{-ir}x - 1)$. Since $f(\alpha^{-rh}x) = f(x)$,

$$\mathbf{v} = (f(p_1), \ldots, f(p_n)) = (b_1, \ldots b_r, b_1, \ldots, b_r, \ldots, b_1, \ldots, b_r).$$

Theorem. If $(n-d)|n$ and $d \leq n - n/c^2$, then $RS(n,k)$ is not $(c,c)-$separable.

Proof. Let $r = \frac{n}{k-1}$ and consider the polynomial $f(x) = \prod_0^{k-2}(\alpha^{-ir}x - 1)$. Since $f(\alpha^{-rh}x) = f(x)$,

$$\mathbf{v} = (f(p_1), \ldots, f(p_n)) = (b_1, \ldots b_r, b_1, \ldots, b_r, \ldots, b_1, \ldots, b_r).$$

The coalitions $T_1 = \{\mathbf{v}, \mathbf{v_c}, \mathbf{v_{2c}}, \ldots, \mathbf{v_{r-c}}\}$ and $T_2 = \{\mathbf{b_1}, \ldots, \mathbf{b_c}\}$ are inseparable.

Theorem. If $(n-d)|n$ and $d \leq n - n/c^2$, then $RS(n,k)$ is not $(c,c)-$separable.

Proof. Let $r = \frac{n}{k-1}$ and consider the polynomial $f(x) = \prod_0^{k-2}(\alpha^{-ir}x - 1)$. Since $f(\alpha^{-rh}x) = f(x)$,

$$\mathbf{v} = (f(p_1), \ldots, f(p_n)) = (b_1, \ldots b_r, b_1, \ldots, b_r, \ldots, b_1, \ldots, b_r).$$

The coalitions $T_1 = \{\mathbf{v}, \mathbf{v_c}, \mathbf{v_{2c}}, \ldots, \mathbf{v_{r-c}}\}$ and $T_2 = \{\mathbf{b_1}, \ldots, \mathbf{b_c}\}$ are inseparable.

Corollary The $RS(n,k)$ with distance $d \leq n - n/c^2$ for $c \geq \sqrt{q-1}$ is $(c,c)-$ inseparable

Theorem. If $c|q$ and $d \leq n - n/c^2$ then $RS(n,k)$ is $(c,c)-$inseparable.

Theorem. If $c|q$ and $d \leq n - n/c^2$ then $RS(n,k)$ is $(c,c)-$inseparable.

Definition. A polynomial L is linearized if $L(a\alpha + b\beta) = aL(\alpha) + bL(\beta)$ for any $a, b \in \mathbb{F}_q$ and $\alpha, \beta \in \mathbb{F}_{q^m}$

Theorem. If $c|q$ and $d \leq n - n/c^2$ then $RS(n,k)$ is $(c,c)-$inseparable.

Definition. A polynomial L is linearized if $L(a\alpha + b\beta) = aL(\alpha) + bL(\beta)$ for any $a, b \in \mathbb{F}_q$ and $\alpha, \beta \in \mathbb{F}_{q^m}$

Lemma. $L_s(x) = \prod_{u \in \mathbb{F}_{q'}} (x-u)^{q^s}$ is a linearized polynomial.

Theorem. If $c|q$ and $d \leq n - n/c^2$ then $RS(n,k)$ is $(c,c)-$inseparable.

Definition. A polynomial L is linearized if $L(a\alpha + b\beta) = aL(\alpha) + bL(\beta)$ for any $a, b \in \mathbb{F}_q$ and $\alpha, \beta \in \mathbb{F}_{q^m}$

Lemma. $L_s(x) = \prod_{u \in \mathbb{F}_{q'}} (x-u)^{q^s}$ is a linearized polynomial.

Proof. It is enought to prove it for $s = 0$ since $(a+b)^p = a^p + b^p$ and

$$\begin{aligned} L_s(a\alpha + b\beta) &= L^q_{s-1}(a\alpha + b\beta) = a^q L^q_{s-1}(\alpha) + b^q L_s - 1^q(\beta) \\ &= aL_s(\alpha) + bL_s(\beta) \end{aligned}$$

Proof of the theorem. We assume $c^2 \leq q = p^m$. Since $c|q$, $c = p^l$. Consider $L(x)$ with $q' = p^{m-2l}$. $|Ker(L)| = q/c^2$, $|Img(L)| = c^2$. Take $B \subset Img(L)$ with $|B| = c$. Then $Img(L) = U_i(B + \beta_i)$.

Proof of the theorem. We assume $c^2 \leq q = p^m$. Since $c|q$, $c = p^l$. Consider $L(x)$ with $q' = p^{m-2l}$. $|Ker(L)| = q/c^2$, $|Img(L)| = c^2$. Take $B \subset Img(L)$ with $|B| = c$. Then $Img(L) = U_i(B + \beta_i)$.

Now consider $f_i(x) = L(x) - \beta_i$ The coalitions $T_1 = \{\mathbf{v_1}, \ldots, \mathbf{v_c}\}$ and $T_2 = \{\mathbf{b} : b \in B\}$ are inseparable.

Theorem. There exist a Reed Solomon code with distance $d = 8$ over F_{11} which is $(2, 2)-$ inseparable.

Theorem. There exist a Reed Solomon code with distance $d = 8$ over F_{11} which is $(2, 2)-$ inseparable.

Proof. Consider the polynomial $f_1 = 0$, and take $g_1(x) = C_0 \prod_{i=1}^{3}(x - \alpha_i)$, for some constant C_0 and any three elements $\alpha_i \in \mathbb{F}_{11}$. Now let

$$f_2(x) = \sum_{i=1}^{3} g_1(\alpha_{3+i}) \frac{\prod_{j \neq i}(x - \alpha_{3+j})}{\prod_{j \neq i}(\alpha_{3+i} - \alpha_{3+j})} + C_1 \prod_{i=1}^{3}(x - \alpha_{3+i}).$$

Finally consider

$$g_2(x) = C_2 \prod_{i=1}^{3}(x - \alpha_{6+i}).$$

By construction $\{f_1(\alpha_i), f_2(\alpha_i)\} \cap \{g_1(\alpha_i), g_2(\alpha_i)\} \neq \emptyset$ for $i = 1, \ldots, 9$.

Now, selecting C_1, C_2 such that

$$\begin{aligned}
&\sum_{i=1}^{3} g_1(\alpha_{3+i}) \frac{\prod_{j\neq i}(\alpha_{10}-\alpha_{3+j})}{\prod_{j\neq i}(\alpha_{3+i}-\alpha_{3+j})} \\
= \quad & C_2 \prod_{i=1}^{3}(\alpha_{10}-\alpha_{6+i}) - C_1 \prod_{i=1}^{3}(\alpha_{10}-\alpha_{3+i}) \\
&\sum_{i=1}^{3} g_1(\alpha_{3+i}) \frac{\prod_{j\neq i}(\alpha_{11}-\alpha_{3+j})}{\prod_{j\neq i}(\alpha_{3+i}-\alpha_{3+j})} \\
= \quad & C_2 \prod_{i=1}^{3}(\alpha_{11}-\alpha_{6+i}) - C_1 \prod_{i=1}^{3}(\alpha_{11}-\alpha_{3+i}),
\end{aligned}$$

which is possible whenever

$$\prod_{i=1}^{3}(\alpha_{10}-\alpha_{6+i})(\alpha_{11}-\alpha_{3+i}) \neq \prod_{i=1}^{3}(\alpha_{11}-\alpha_{6+i})(\alpha_{10}-\alpha_{3+i}),$$

we get

$$\begin{aligned} f_2(\alpha_{10}) &= g_2(\alpha_{10}) \\ f_2(\alpha_{11}) &= g_2(\alpha_{11}) \end{aligned}$$

and hence we find the $(2, 2)$-inseparable code that we wanted.

An example is by taking $\alpha_i = i$. Then

$$\prod_{i=1}^{3}(\alpha_{10} - \alpha_{6+i})(\alpha_{11} - \alpha_{3+i}) = \prod_{i=1}^{3}(4-i)(8-i) = 7!/4,$$

while

$$\prod_{i=1}^{3}(\alpha_{11} - \alpha_{6+i})(\alpha_{10} - \alpha_{3+i}) = \prod_{i=1}^{3}(5-i)(7-i) = 4 \cdot 6!.$$

In this case

$$\begin{aligned} g_1(x) &= x^3 - 6x^2 + 11x - 6 \\ f_2(x) &= 5x^3 + 10x + 9 \\ g_2(x) &= 10x^3 + x^2 + x + 9 \end{aligned}$$

and we have the inseparable words

$$\begin{aligned} f_1 &= (0,0,0,0,0,0,0,0,0,0,0) \\ f_2 &= (2,3,9,6,2,5,1,9,4,5,9) \\ g_1 &= (0,0,0,6,2,5,10,1,6,9,5) \\ g_2 &= (6,1,10,5,2,6,0,0,0,5,9) \end{aligned}$$