Randomized Frameproof Codes For Content Protection

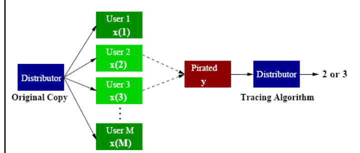
N. Prasanth Anthapadmanabhan and Alexander Barg

Problem Statement

Objective

To design a scheme to protect copyrighted content (esp. software) against piracy.

Traditional solution - **"Fingerprinting"**
(Boneh-Shaw 1998)

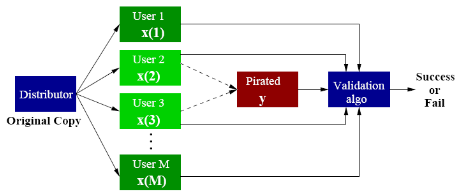


- Distributor embeds a distinct imperceptible fingerprint in each legal copy.
- Users may collude to create a pirated copy.
- Tracing algorithm identifies one of the guilty users as long as the coalition size does not exceed a certain threshold.

Drawbacks: Need large redundancy (i.e., low rates), complexity of tracing.

We will consider the following **modified** problem.

Fingerprinting with Validation



- Validation algorithm verifies fingerprint everytime a user tries to execute copy.
- Execution continues only if validation is successful.
- **Problem**: Assign fingerprints such that pirates can not frame an innocent user.

Pros

- Lower redundancy (increase in rates).
- Eliminates the need for tracing.

Cons

- Complexity increases. However, polynomial-time validation may help curb the effects.

Notation and Terminology

- Users: $[M] = \{1, \ldots, M\}$ Alphabet: binary
  Fingerprint length: n Rate: $R = \log_2 M/n$ (quantifies redundancy)
- Collection of fingerprints called a code. Distributor uses **randomization**, i.e., picks a code at random from a family of codes.

  Randomized code $\mathcal{C}$:
  Code family $\{C_k\}$, $|C_k| = M$; Probability of choosing "key" $k = \pi(k)$
- Validation algorithm:
  Checks whether fingerprint is present in current code.
  Preferably polynomial-time complexity.
- Code family known to all users. **Distributor keeps selection of k secret!**
- Coalition $U \subseteq [M]$ (pirates) of size t observes $C_k(U) = \{x_1, \ldots, x_t\}$.
- Goals: Distributor: maximize R
  Coalition: frame an innocent user, i.e., forge y s.t. $y \in C_k \backslash C_k(U)$.

Definition: Frameproof Code

What are the rules of the game?
Coalitions try to detect fingerprint positions by comparing their copies.

Marking assumption [Boneh-Shaw 1998]

The coalition can change only those positions of the fingerprint where they find a difference.

Example

x_1	:	110001000
x_2	:	100011010
y	:	110011000

Envelope: Set of all possible forgeries.

$$\mathcal{E}(x_1, \ldots, x_t) = \{y | y_i = x_{1i}, \forall i \text{ undetectable}\}$$

Definition

A randomized code $\mathcal{C}$ is *t-frameproof with ε-error* if: $\forall U \subseteq [M]$ s.t. $|U| \le t$,

$$\underbrace{\Pr\{\mathcal{E}(\mathcal{C}(U)) \cap (\mathcal{C} \backslash \mathcal{C}(U)) \neq \emptyset\}}_{\text{Prob. of framing}} \le \varepsilon.$$

Binary Frameproof Codes

Construction of $\mathcal{C}$:
Pick random $M \times n$ binary matrix with $P(1) = p, P(0) = 1 - p \quad (0 \le p \le 1)$.

Theorem

$\mathcal{C}$ is t-frameproof with error probability decaying exponentially in n for any rate

$$R < -p^t \log_2 p - (1-p)^t \log_2 (1-p).$$

	Comparison of Rates		
t	Randomized Frameproof	Deterministic Frameproof	Fingerprinting
2	0.5	0.2075	0.25
3	0.25	0.0693	0.0833
4	0.1392	0.04	0.0158
5	0.1066	0.026	0.0006

But validation has exponential complexity!

Linear Frameproof Codes

Linear codes can be validated by verifying parity-checks in $O(n^2)$!!

Construction of $\mathcal{C}$: Suppose we have $M = 2^{nR}$ users.

1. Pick random $n(1-R) \times n$ parity-check matrix with $P(0) = P(1) = 1/2$.
2. Binary vectors satisfying the parity-check matrix form a linear code of size $\ge 2^{nR}$.
3. Assign to each a user a unique codeword selected uniformly at random.

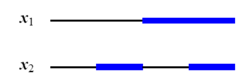
Theorem

$\mathcal{C}$ is 2-frameproof with error probability decaying exponentially in n for any rate $R < 0.5$.

Matches the rate obtained with exponential complexity!

Proof idea:

For any two pirates, with high prob. the cross-section of their fingerprints contain each of $(0,0), (0,1), (1,0), (1,1)$ in $\approx n/4$ coordinates.



Given that fingerprints (x_1, x_2) satisfy above condition

$$\Pr\{\mathcal{E}(\mathcal{C}(U)) \cap (\mathcal{C} \backslash \mathcal{C}(U)) \neq \emptyset | \mathcal{C}(U) = (x_1, x_2)\}$$
$$= \Pr\{\exists y \in \mathcal{C} : y \in \mathcal{E}(x_1, x_2) \backslash \{x_1, x_2\} | \mathcal{C}(U) = (x_1, x_2)\}$$
$$\lessapprox \underbrace{2^{n/2}}_{\text{envelope size}} \times \underbrace{2^{-n(1-R)}}_{\Pr\{y \in \mathcal{C}\}} \longrightarrow_n 0 \text{ if } R < 0.5. \qquad \blacksquare$$

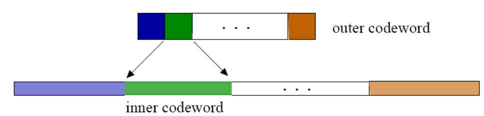
Do there exist linear t-frameproof codes for $t > 2$? **NO!**

Proposition

There do NOT exist binary linear t-frameproof codes with ε-error, $0 \le \varepsilon < 1$, if $t > 2$.

Polynomial-time validation for larger t

We use **code concatenation** for larger t.



- Outer code C_{out}: (deterministic) q-ary linear $[N, K, \Delta]$ code.
- Inner code C_{in}: randomized binary (m, q) code, t-frameproof with ε-error.
- For each of the N coordinates of the outer code, generate an independent instance of the randomized binary code.
- $\mathcal{C}$: Concatenated, randomized binary $(n = Nm, q^K)$ code

Theorem

If $\frac{\Delta}{N} \ge 1 - \frac{1}{t}(1 - \xi)$ for C_{out} and the error probability $\varepsilon < \xi$ for $\mathcal{C}_{in}$, then $\mathcal{C}$ is t-frameproof with error probability $2^{-ND(\xi||\varepsilon)}$ and has a poly(n) validation algorithm.

Validation: Exhaustive search at inner level. Parity-checks at outer level.
Choose appropriate scaling, for e.g., $m \sim \log_2(n)$, to obtain poly(n) complexity.

Use explicit codes in the above construction:

- C_{out}: $[q-1, K]$ Reed-Solomon code with rate $\le (1-\xi)/t$
- C_{in}: randomized binary t-frameproof with error probability $\varepsilon = 2^{-m\beta}$ for some $\beta > 0$, and rate

$$R_t = \max_{p \in [0,1]} \left[-p^t \log_2 p - (1-p)^t \log_2 (1-p)\right].$$

Taking ξ arbitrarily small and m sufficiently large to satisfy $\varepsilon < \xi$, we obtain:

Corollary

The concatenated code is t-frameproof with error prob. $\exp(-\Omega(n))$, validation complexity $O(n^2)$ and rate $\approx R_t/t$.