Two-level Fingerprinting Codes

Alexander Barg

University of Maryland, College Park

Joint work with

Prasanth Anthapadmanabhan

UMD

ISIT 2009, Seoul, Korea

Fingerprinting

Introduce *hidden marks* in every legal copy

Original: 011000010100011000111010100010101111 0010

$$\Downarrow$$

User 1: 011**1**0000101**0**0001100**0**0111**1**0101000101**1**0111101010

User 2: 011**0**0000101**1**0001100**0**0111**0**0101000101**1**0111101010

$$\vdots$$

User M: 011**1**0000101**1**0001100**0**0111**0**0101000101**0**1011110**0**010

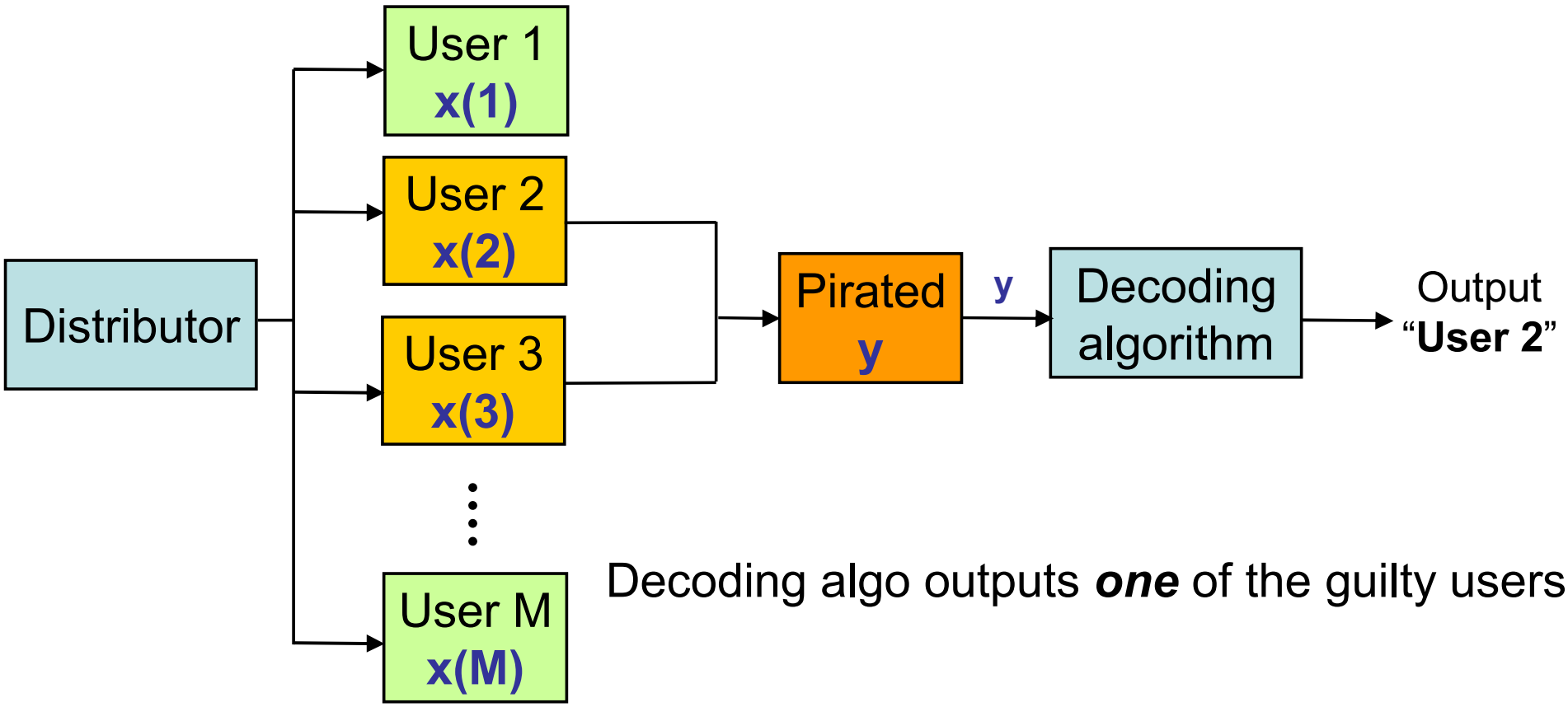
$$\left.\begin{array}{ll} \mathbf{x(1)}: & \mathbf{100111} \\ \mathbf{x(2)}: & \mathbf{010011} \\ & \vdots \\ \mathbf{x(M)}: & \mathbf{110000} \end{array}\right\}$$

"Fingerprints"

Collection is called a **code**,

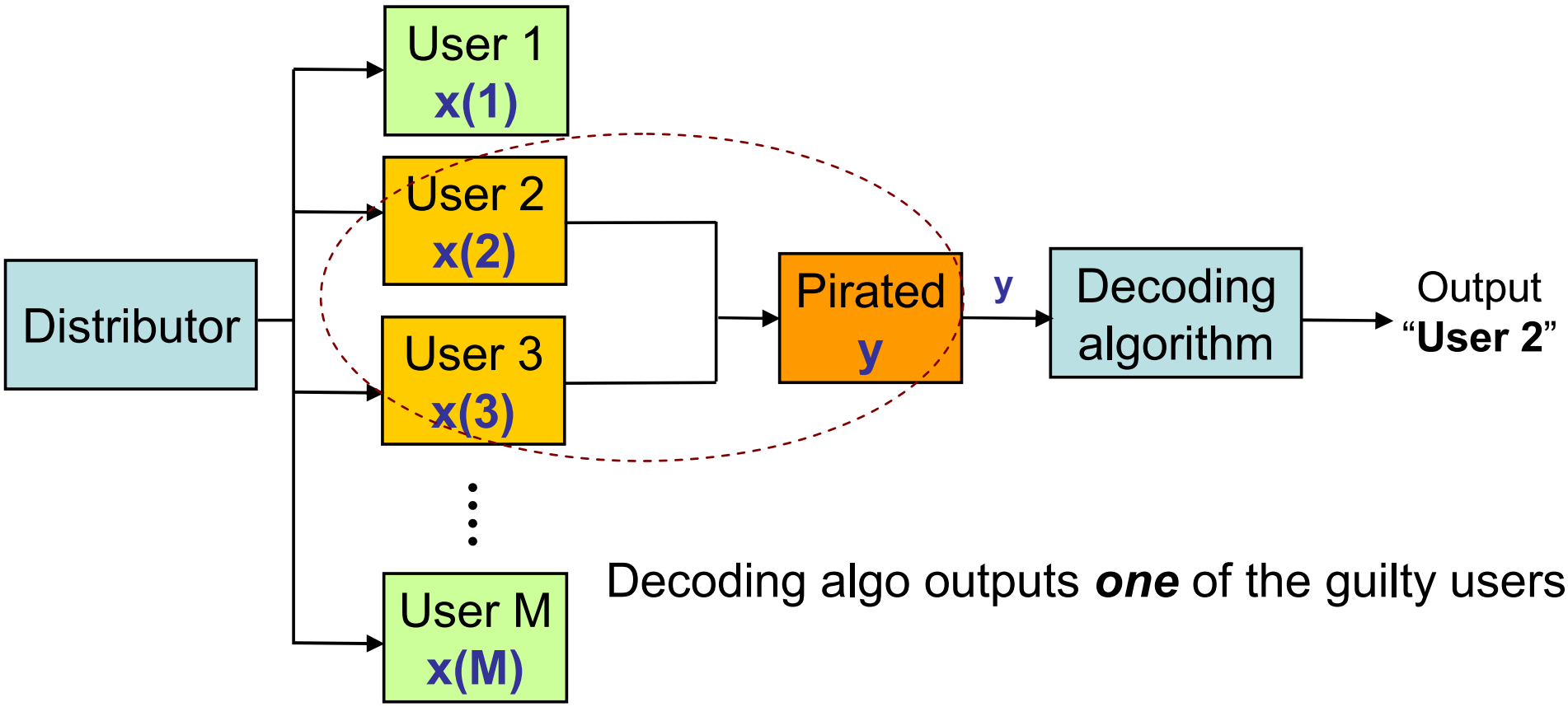
denoted by $C : \{1, \ldots, M\} \to \mathcal{Q}^n$

The fingerprinting problem



Decoding algo outputs ***one*** of the guilty users.

The fingerprinting problem



Decoding algo outputs ***one*** of the guilty users.

The Boneh-Shaw fingerprinting problem

Which forgeries are possible?

User 2: 011**1**0000101**0**0001100**0**0111**1**0101000101**1**0111101010

User 3: 011**0**0000101**1**0001100**0**0111**0**0101000101**1**0111101010

$$\mathbf{x}_1 : \quad \mathbf{100111}$$
$$\mathbf{x}_2 : \quad \mathbf{010011}$$
$$\mathbf{y} : \quad \mathbf{000111}$$

Envelope

Set of all possible forgeries a coalition can create.

$$\mathcal{E}(\mathbf{x}_1, \ldots, \mathbf{x}_t) = \{\mathbf{y} | y_i \in \{x_{1i}, \ldots, x_{ti}\}, \forall i\}$$

Randomized codes

Randomized code

Family of codes: $\{C_k, k \in \mathcal{K}\}, \quad C_k : \{1, \ldots, M\} \to \mathcal{Q}^n$
Decoding algos: $\{D_k, k \in \mathcal{K}\}, \quad D_k : \mathcal{Q}^n \to \{1, \ldots, M\} \cup \{0\}$

$$(\mathcal{C}, \mathcal{D}) \equiv \text{rv over } \{(C_k, D_k), k \in \mathcal{K}\}$$

$$\textbf{Rate } R = \frac{1}{n} \log_q M$$

Code family is public. Selection of k known only to distributor.

Coalition strategies

- Coalition $U \subseteq \{1, \ldots, M\}$ of size t observes $C_k(U) = \{\mathbf{x}_1, \ldots, \mathbf{x}_t\}$
- $V(\cdot, \ldots, \cdot)$: coalition strategy

$$V(\mathbf{x}_1, \ldots, \mathbf{x}_t) = \text{outputs a random forgery in } \mathcal{E}(\mathbf{x}_1, \ldots, \mathbf{x}_t)$$

$$\mathbf{Y}_{\mathcal{C},U,V} := V(\mathcal{C}(U))$$

- Probability of error for a given coalition U and strategy V:

$$e(\mathcal{C}, \mathcal{D}, U, V) = \mathbb{P}\left[\, \mathcal{D}(\mathbf{Y}_{\mathcal{C},U,V}) \notin U \,\right]$$

(One-level) Fingerprinting codes

Definition

$(\mathcal{C}, \mathcal{D})$ is ***t*-fingerprinting** with ε-error if

$$\max_V e(\mathcal{C}, \mathcal{D}, U, V) \leq \varepsilon, \quad \forall U \text{ with } |U| \leq t.$$

Informally: A randomized code is t-**fingerprinting** if for every coalition of size $\leq t$ and any coalition strategy, it is possible to identify <u>one</u> of the pirates with high probability.

(One-level) Fingerprinting codes

Definition

$(\mathcal{C}, \mathcal{D})$ is ***t*-fingerprinting** with ε-error if

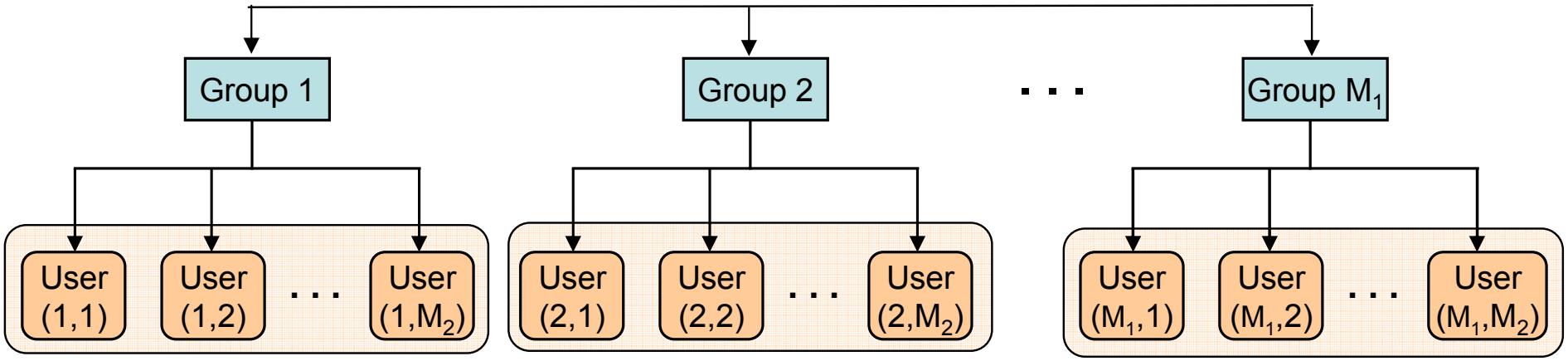
$$\max_V e(\mathcal{C}, \mathcal{D}, U, V) \leq \varepsilon, \quad \forall U \text{ with } |U| \leq t.$$

Informally: A randomized code is t-**fingerprinting** if for every coalition of size $\leq t$ and any coalition strategy, it is possible to identify <u>one</u> of the pirates with high probability.

- A. B., G.R. Blakley and G. Kabatiansky, *Digital fingerprinting*, IT, Apr. 2003
- G. Tardos, *Optimal probabilistic fingerprint codes*, 2003-2008 (Journ. ACM)
- N. Merhav and A. Somekh-Baruch, *Capacity of private fingerprinting*, IT, March 2005
- N. P. Anthapadmanabhan, A. B., and I. Dumer, *Fingerprinting capacity*, IT, July 2008
- E. Amiri and G. Tardos, *High-rate codes and FP capacity*, ACM SODA 2009
- P. Moulin, *Capacity and error exponents for FP*, 2008-09
- Y.-W. Huang and P. Moulin, previous talk.

Two-level codes

Users are organized in M_1 groups each containing M_2 users.



Accordingly,

$$C_k : [M_1] \times [M_2] \to \mathcal{Q}^n,$$
$$D_k : \mathcal{Q}^n \to ([M_1] \times [M_2]) \cup \{0\},$$
$$(R_1, R_2) = \left(\frac{1}{n} \log_q M_1, \frac{1}{n} \log_q M_2 \right).$$

Error probabilities

For user $\mathbf{u} \equiv (u_1, u_2)$, let $\mathcal{G}(\mathbf{u}) = u_1$, be its group index.

Error probabilities for a given coalition U and strategy V:

$$e(\mathcal{C}, \mathcal{D}, U, V) = \mathbb{P}\left[\, \mathcal{D}(\mathbf{Y}_{\mathcal{C},U,V}) \notin U \,\right]$$

$$\tilde{e}(\mathcal{C}, \mathcal{D}, U, V) = \mathbb{P}\left[\underbrace{\mathcal{G}(\mathcal{D}(\mathbf{Y}_{\mathcal{C},U,V}))}_{\text{group index of decoder output}} \notin \underbrace{\mathcal{G}(U)}_{\text{groups containing guilty users}} \right]$$

Two-level fingerprinting codes

Definition

$(\mathcal{C}, \mathcal{D})$ is (t_1, t_2)-**fingerprinting** with $(\varepsilon_1, \varepsilon_2)$-error (where $t_1 > t_2$) if

1. $\max_V e(\mathcal{C}, \mathcal{D}, U, V) \leq \varepsilon_2, \quad \forall U$ with $|U| \leq t_2$. $\Leftrightarrow$ t_2-fingerprinting
2. $\max_V \tilde{e}(\mathcal{C}, \mathcal{D}, U, V) \leq \varepsilon_1, \quad \forall U$ with $|U| \leq t_1$. $\Leftarrow$ t_1-fingerprinting

In fact, with $t_1 > t_2$,

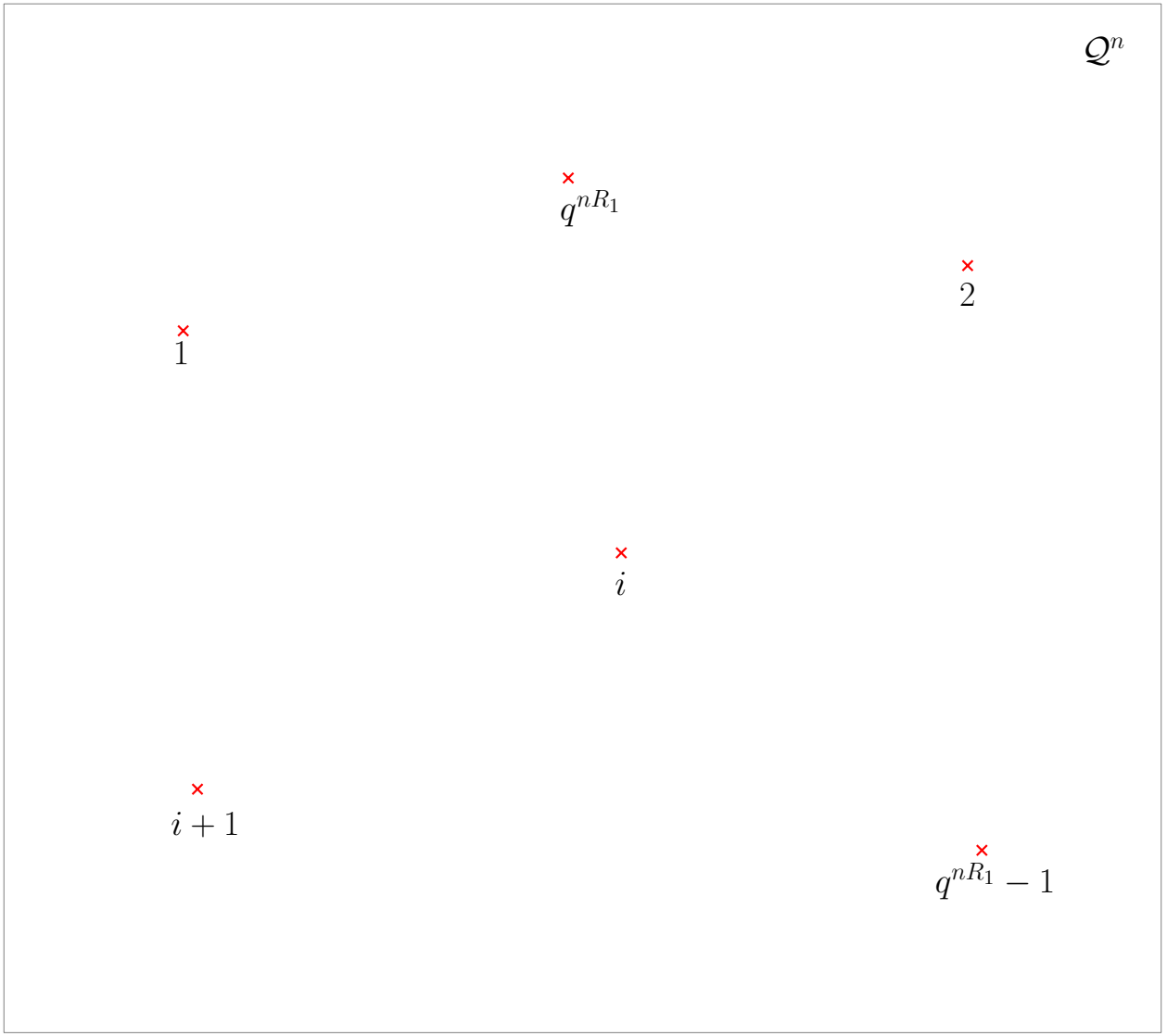
1. one-level t_1-fingerprinting $\Rightarrow$ (t_1, t_2)-fingerprinting

   Q: Can we achieve higher rates?

2. (t_1, t_2)-fingerprinting of size $M_1 M_2$ $\Rightarrow$
   one-level t_1-fingerprinting of size M_1.

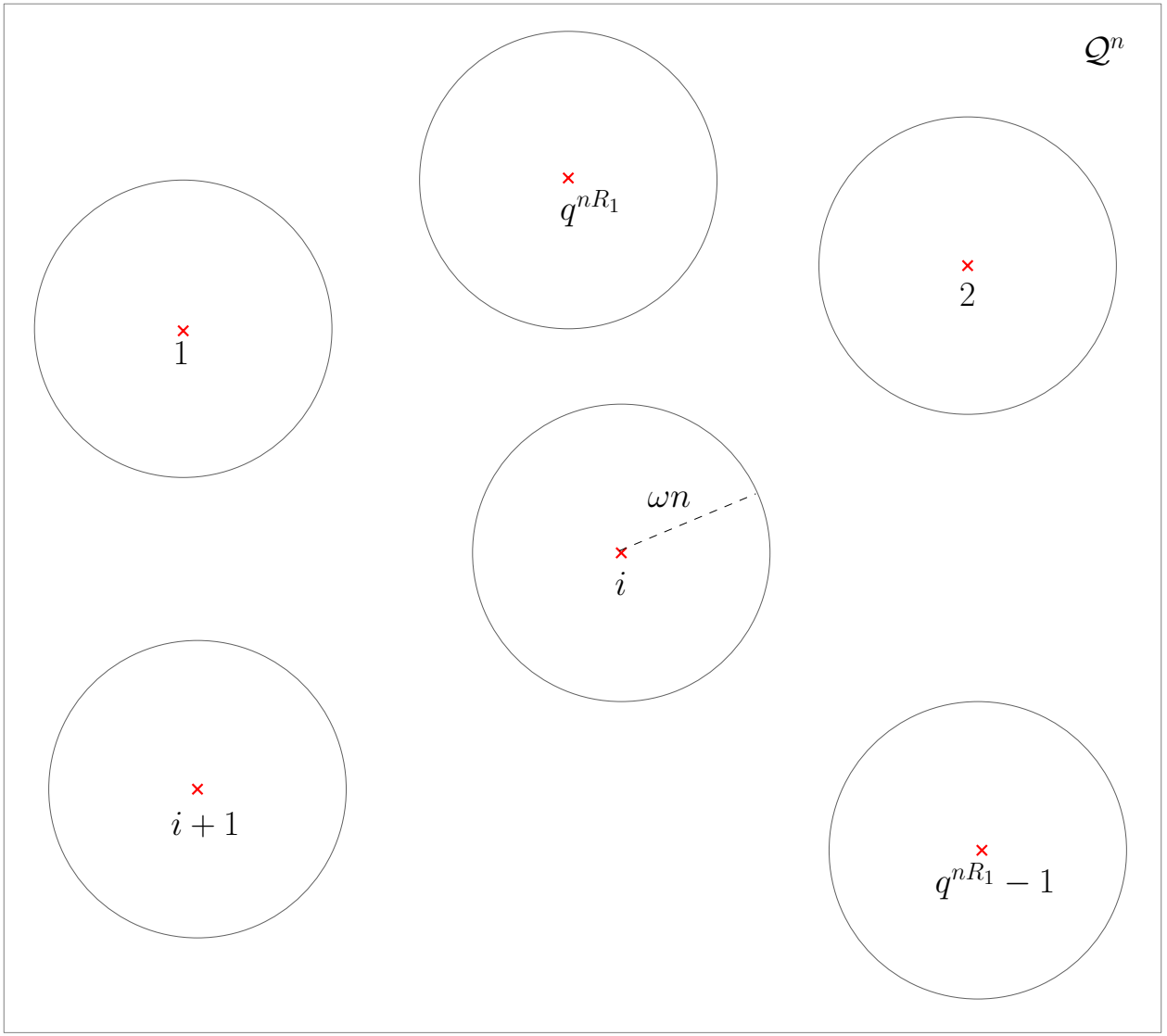
(2,1)-fingerprinting code

Want: $(\mathcal{C}_n, \mathcal{D}_n)$ which are (2,1)-fingerprinting with $(\varepsilon_{1,n}, \varepsilon_{2,n})$-error, such that $\varepsilon_{1,n} \to_n 0$ and $\varepsilon_{2,n} \to_n 0$



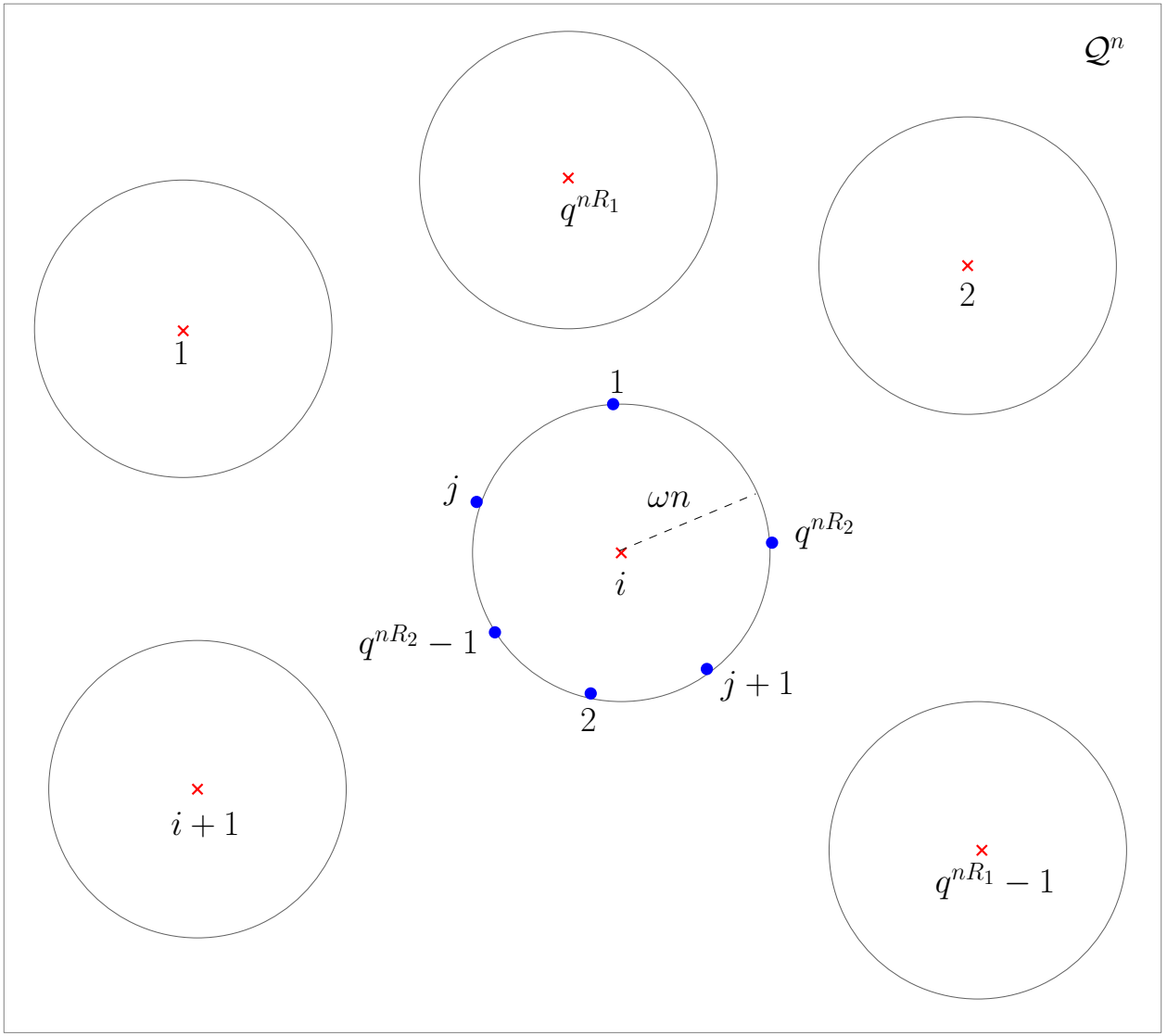
(2,1)-fingerprinting code

Want: $(\mathcal{C}_n, \mathcal{D}_n)$ which are (2,1)-fingerprinting with $(\varepsilon_{1,n}, \varepsilon_{2,n})$-error, such that $\varepsilon_{1,n} \to_n 0$ and $\varepsilon_{2,n} \to_n 0$



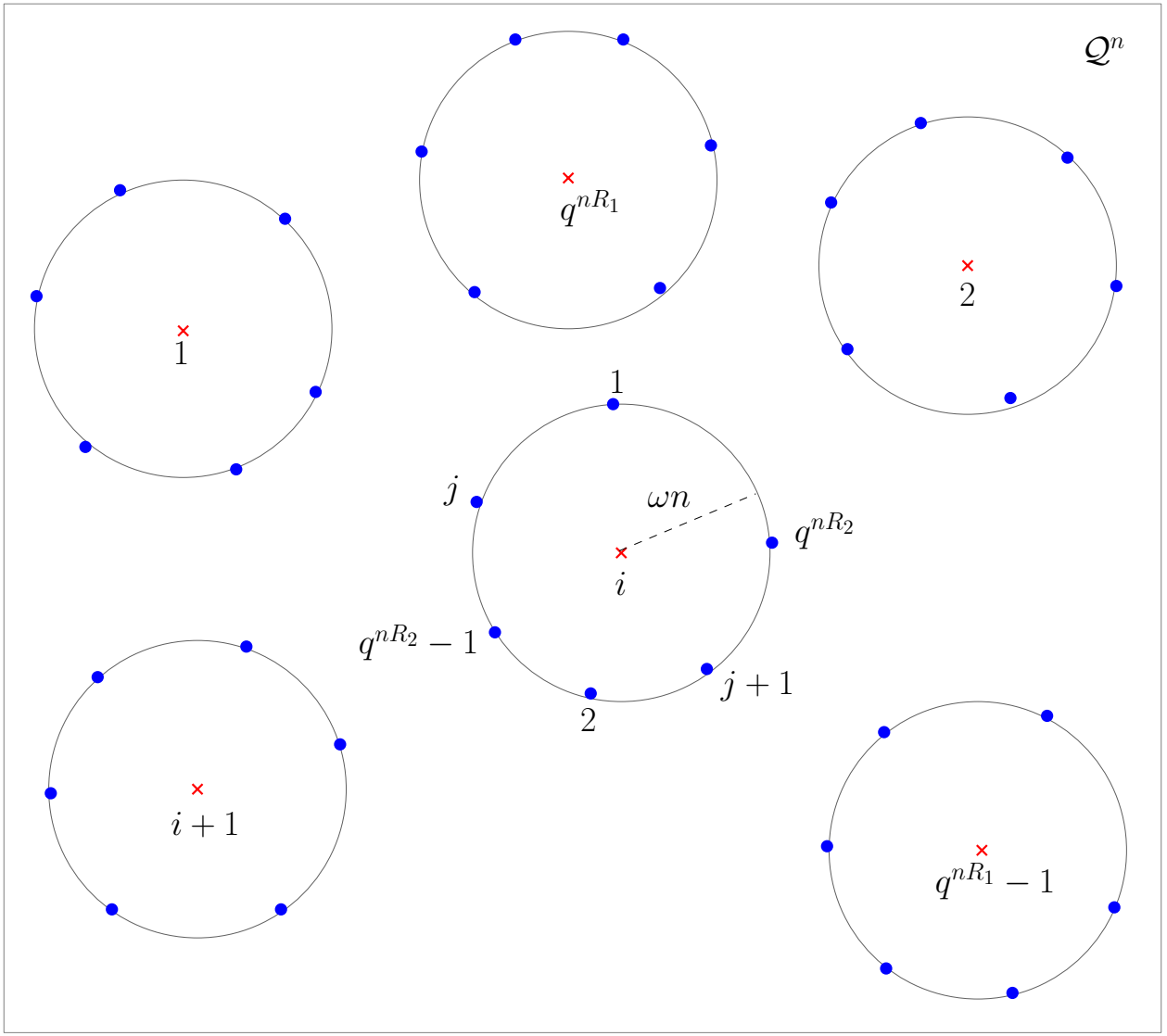
(2,1)-fingerprinting code

Want: $(\mathcal{C}_n, \mathcal{D}_n)$ which are (2,1)-fingerprinting with $(\varepsilon_{1,n}, \varepsilon_{2,n})$-error, such that $\varepsilon_{1,n} \to_n 0$ and $\varepsilon_{2,n} \to_n 0$



(2,1)-fingerprinting code

Want: $(\mathcal{C}_n, \mathcal{D}_n)$ which are (2,1)-fingerprinting with $(\varepsilon_{1,n}, \varepsilon_{2,n})$-error, such that $\varepsilon_{1,n} \to_n 0$ and $\varepsilon_{2,n} \to_n 0$



(2,1)-fingerprinting code

Decoding: minimum distance

$$D_{k,n}(\mathbf{y}) = \arg\min_{\mathbf{u}} d_H(C_{k,n}(\mathbf{u}), \mathbf{y})$$

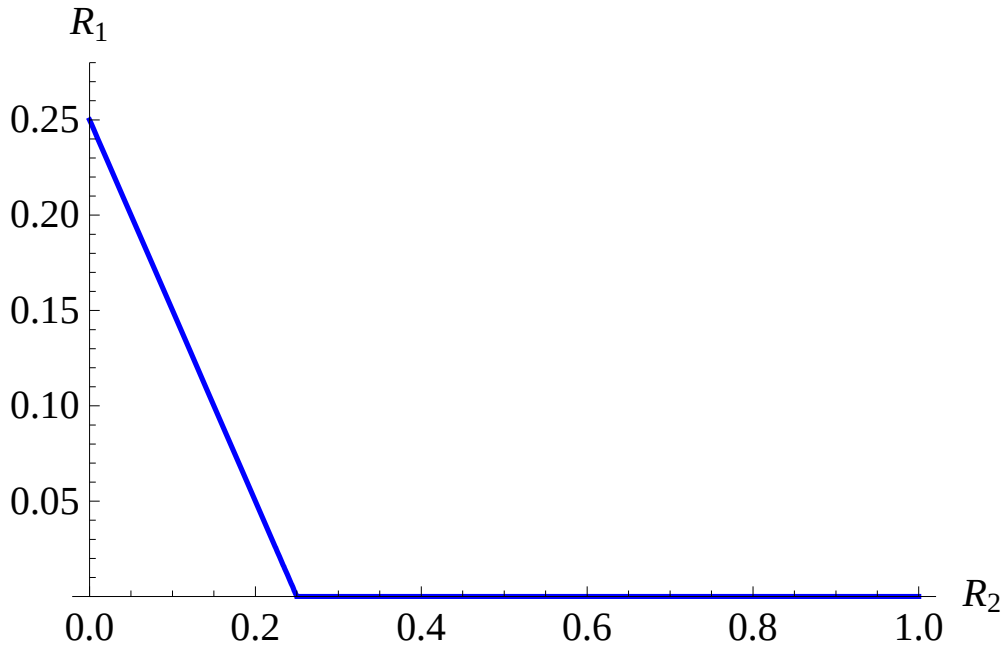
Theorem

$(\mathcal{C}_n, \mathcal{D}_n)$ *is* $(2,1)$-**fingerprinting** *if there exists an* $\omega \in [0, \frac{q-1}{2q}]$ *such that*

$$R_1 < 1 - h\left(\frac{1}{2}\left(1 - \frac{1}{q}\right) + \omega\right),$$
$$R_2 < h(\omega).$$

Remark: For such ω's to exist we need large alphabet, e.g., $q \geq t + 1$ (recall IPP codes)

(2,1)-fingerprinting: Achievable rates

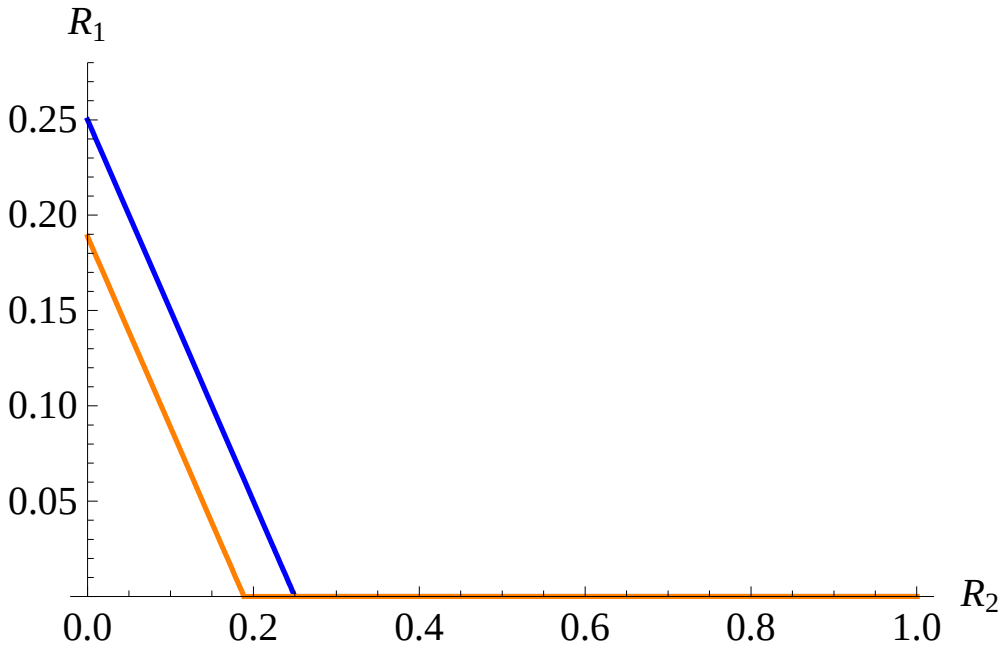


Inner bound

2-fingerprinting

[ABD, 2008]

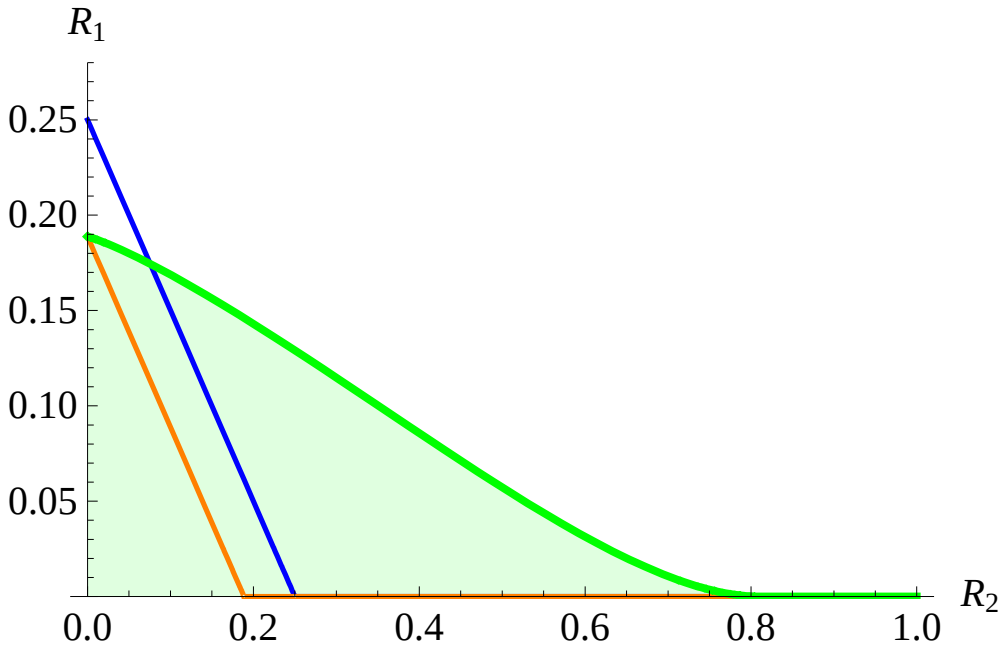
(2,1)-fingerprinting: Achievable rates



Inner bound
2-fingerprinting
[ABD, 2008]

Inner bound
2-FP with MD decoder
[Lin, Shahmohammadi, ElGamal, 2007]

(2,1)-fingerprinting: Achievable rates

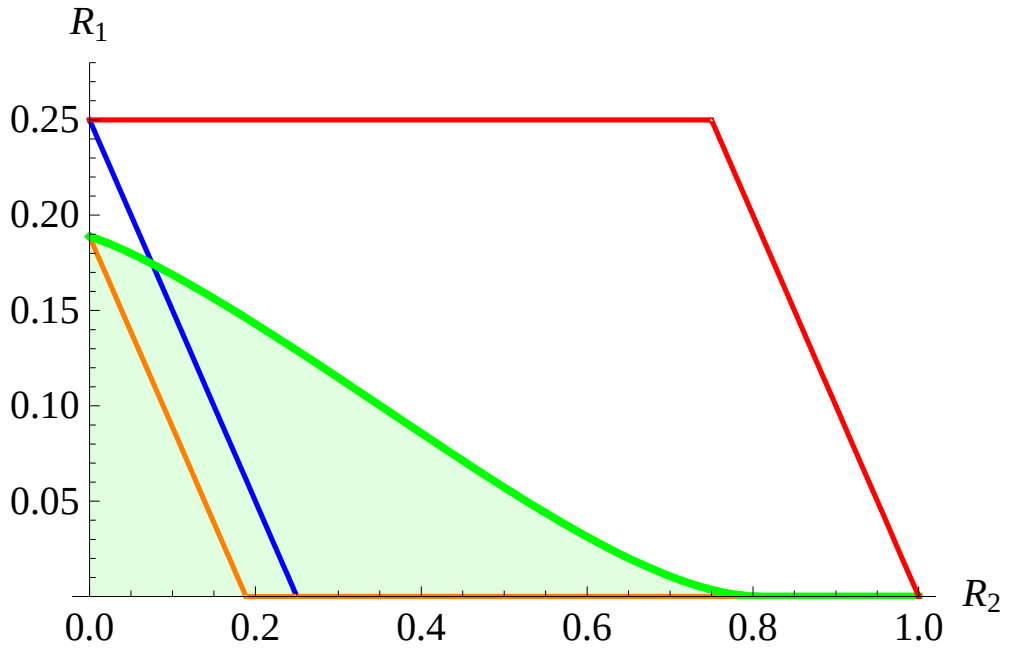


Inner bound
2-fingerprinting
[ABD, 2008]

Inner bound
2-FP with MD decoder
[Lin, Shahmohammadi, ElGamal, 2007]

This work

(2,1)-fingerprinting: Achievable rates



Inner bound
2-fingerprinting
[ABD, 2008]

Inner bound
2-FP with MD decoder
[Lin, Shahmohammadi, ElGamal, 2007]

This work

Outer bound
From one-level codes
[Amiri and Tardos '09]
[Huang and Moulin '09]

(t,1)-fingerprinting code

Extending the idea...

Theorem

$(\mathcal{C}_n, \mathcal{D}_n)$ *is* $(t,1)$-**fingerprinting** *if there exists an* ω *with*

$$\frac{t-1}{t}\left(1-\frac{1}{q^{t-1}}\right)+\omega \leq \frac{q-1}{q}$$

such that

$$R_1 < 1 - h\left(\frac{t-1}{t}\left(1-\frac{1}{q^{t-1}}\right)+\omega\right),$$

$$R_2 < h(\omega).$$

(t,2)-fingerprinting code

Similarly...

Theorem

$(\mathcal{C}_n, \mathcal{D}_n)$ *is* $(t, 2)$-**fingerprinting** *if there exists an* ω *with*

$$\frac{t-1}{t}\left(1 - \frac{1}{q^{t-1}}\right) + \omega \leq \frac{q-1}{q}$$

such that

$$R_1 < 1 - h\left(\frac{t-1}{t}\left(1 - \frac{1}{q^{t-1}}\right) + \omega\right),$$

$$R_2 < h(\omega) - (\dots).$$

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Open

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Open

- Two-level codes with polynomial-time decoding

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Open

- Construction for (t_1, t_2) FP
- Two-level codes with polynomial-time decoding

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Open

- Construction for $(t_1, t_2, \ldots, t_m)$ FP
- Construction for (t_1, t_2) FP
- Two-level codes with polynomial-time decoding

Conclusion

- Introduced *two-level codes* for fingerprinting
  $\approx$ unequal error protection
- A bound on achievable rates under minimum-distance decoding for $(t, 1)$ fingerprinting, $(t, 2)$ fingerprinting

Open

- Upper bounds
- Construction for $(t_1, t_2, \ldots, t_m)$ FP
- Construction for (t_1, t_2) FP
- Two-level codes with polynomial-time decoding