Lecture 1: From Digital Fingerprinting Security to Extremal Combinatorics: Frameproof Codes, Matchings, and Separating Hash Families

Lecturer: Professor Xiande Zhang

Notes written by Danni Li with assistance from ChatGPT

August 2026

Contents

1 Digital Fingerprints: Frameproof Codes

Frameproof codes come from the study of digital fingerprinting. In the model of Boneh and Shaw [1], each user receives a copy of the same digital content with a personal fingerprint $\boldsymbol{x} = (x_1, \ldots, x_\ell) \in [q]^\ell$. Thus a set of users can be represented by a code $\mathcal{C} \subseteq [q]^\ell$. Several users may work together and compare their copies. At coordinate i, they may output any symbol that appears at that coordinate among their fingerprints. This leads to the following definition.

Definition 1.1 (Descendant)**.** For $P \subseteq [q]^\ell$, the *descendant set* of P is

$$\mathrm{desc}(P) = \left\{ \boldsymbol{y} \in [q]^\ell : y_i \in \{x_i : \boldsymbol{x} \in P\} \text{ for every } i \in [\ell] \right\}.$$

Thus $\mathrm{desc}(P)$ is the set of all fingerprints that the users in P can produce by choosing symbols coordinate by coordinate. The security goal is simple: a group of at most c users should not be able to produce the fingerprint of another user.

Definition 1.2 (c-frameproof code)**.** A code $\mathcal{C} \subseteq [q]^\ell$ is a *c-frameproof code* if $\mathrm{desc}(P) \cap \mathcal{C} = P$ for every $P \subseteq \mathcal{C}$ with $|P| \le c$.

Example 1.3. The code $\mathcal{C} = \{000, 011, 101, 110\} \subseteq \{0,1\}^3$ is a 2-frameproof code. For example, if $P = \{000, 011\}$, then $\mathrm{desc}(P) = \{000, 001, 010, 011\}$ and hence $\mathrm{desc}(P) \cap \mathcal{C} = P$. The other pairs can be checked in the same way.

There is also a useful set form of the same condition. Let $V = [\ell] \times [q]$. For $\boldsymbol{x} = (x_1, \ldots, x_\ell)$, define the transversal set

$$z(\boldsymbol{x}) = \{(1, x_1), \ldots, (\ell, x_\ell)\}.$$

Then

$$\boldsymbol{x} \in \mathrm{desc}(P) \quad \Longleftrightarrow \quad z(\boldsymbol{x}) \subseteq \bigcup_{\boldsymbol{y} \in P} z(\boldsymbol{y}).$$

Definition 1.4 (c-cover-free family)**.** A family $\mathcal{F}$ is *c-cover-free* if $F \nsubseteq \bigcup_{A \in \mathcal{P}} A$ for every $F \in \mathcal{F}$ and every $\mathcal{P} \subseteq \mathcal{F} \setminus \{F\}$ with $|\mathcal{P}| \le c$.

The following are equivalent definitions of a c-frameproof code $\mathcal{C} \subseteq [q]^\ell$:

(i) For every $P \subseteq \mathcal{C}$ with $|P| \le c$, we have $\mathrm{desc}(P) \cap \mathcal{C} = P$.

(ii) For every $P \subseteq \mathcal{C}$ with $|P| \le c$ and every $\boldsymbol{x} \in \mathcal{C} \setminus P$, we have $\boldsymbol{x} \notin \mathrm{desc}(P)$.

(iii) For every $P \subseteq \mathcal{C}$ with $|P| \le c$ and every $\boldsymbol{x} \in \mathcal{C} \setminus P$, there is an $i \in [\ell]$ such that $x_i \notin \{y_i : \boldsymbol{y} \in P\}$.

(iv) The family $\{z(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{C}\}$ is c-cover-free.

Condition (iii) is the separating form of frameproofness. It says that a target codeword can always be separated from a group of at most c other codewords in some coordinate. Condition (iv) gives the same idea in set language. We write

$$M_{c,\ell}(q) = \max\left\{ |\mathcal{C}| : \mathcal{C} \subseteq [q]^\ell \text{ is a } c\text{-frameproof code} \right\}.$$

Our goal is to understand the size of $M_{c,\ell}(q)$ for fixed c and ℓ.

2 The Upper Bound for Frameproof Codes: Double Counting

The upper bound uses *own subsequences*. They tell us when a small set of coordinates is enough to identify one codeword.

2.1 Own Subsequences

Definition 2.1 (Own subsequence)**.** Let $\mathcal{C} \subseteq [q]^\ell$, $\boldsymbol{x} \in \mathcal{C}$, and $S \subseteq [\ell]$. Write $\boldsymbol{x}_S = (x_i : i \in S)$. We call $\boldsymbol{x}_S$ an *own subsequence* of $\boldsymbol{x}$ if $\boldsymbol{y}_S \neq \boldsymbol{x}_S$ for every $\boldsymbol{y} \in \mathcal{C} \setminus \{\boldsymbol{x}\}$.

Thus the coordinates in S already identify $\boldsymbol{x}$ inside $\mathcal{C}$. For $S \subseteq [\ell]$, let

$$U_S = \{\boldsymbol{x} \in \mathcal{C} : \boldsymbol{x}_S \text{ is an own subsequence of } \boldsymbol{x}\}.$$

Each word in $[q]^{|S|}$ can correspond to at most one member of U_S. Hence $|U_S| \le q^{|S|}$.

Lemma 2.2 (Covering property). *Let $\mathcal{C} \subseteq [q]^\ell$ be a c-frameproof code. If $S_1, \dots, S_c \subseteq [\ell]$ and $S_1 \cup \cdots \cup S_c = [\ell]$, then $\mathcal{C} = U_{S_1} \cup \cdots \cup U_{S_c}$.*

Proof. Suppose that some $\boldsymbol{x} \in \mathcal{C}$ does not belong to $U_{S_1} \cup \cdots \cup U_{S_c}$. For each $j \in [c]$, there is a codeword $\boldsymbol{y}_j \in \mathcal{C} \setminus \{\boldsymbol{x}\}$ with $(\boldsymbol{y}_j)_{S_j} = \boldsymbol{x}_{S_j}$. Since $S_1, \dots, S_c$ cover $[\ell]$, each symbol x_i appears at coordinate i in at least one of $\boldsymbol{y}_1, \dots, \boldsymbol{y}_c$. Thus $\boldsymbol{x} \in \mathrm{desc}(\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_c\})$, which contradicts frameproofness. □

This lemma is the main coding step in the upper bound. We now combine it with a matching argument.

2.2 Blackburn's Upper Bound

Fix c and ℓ. Let $d = \lceil \ell/c \rceil$, and let $\lambda \in \{0, \dots, c-1\}$ satisfy $\ell \equiv \lambda + 1 \pmod{c}$. Then

$$\ell = (\lambda + 1)d + (c - \lambda - 1)(d - 1). \tag{2.1}$$

For a family $\mathcal{A}$, let $\nu(\mathcal{A})$ be its matching number. Define

$$m(\ell, d, \lambda) = \max\left\{|\mathcal{A}| : \mathcal{A} \subseteq \binom{[\ell]}{d},\ \nu(\mathcal{A}) \le \lambda\right\}.$$

Theorem 2.3 (Blackburn's upper bound [2]). *For fixed c and ℓ,*

$$M_{c,\ell}(q) \le \frac{\binom{\ell}{d}}{\binom{\ell}{d} - m(\ell, d, \lambda)} q^d + O(q^{d-1}).$$

Proof. Let $\mathcal{C} \subseteq [q]^\ell$ be a c-frameproof code.

Step 1. Remove codewords with short own subsequences. Define

$$D = \mathcal{C} \setminus \bigcup_{S \in \binom{[\ell]}{d-1}} U_S.$$

Since $|U_S| \le q^{d-1}$ for each $S \in \binom{[\ell]}{d-1}$, we have $|\mathcal{C} \setminus D| \le \binom{\ell}{d-1} q^{d-1}$. Hence

$$|D| \ge |\mathcal{C}| - \binom{\ell}{d-1} q^{d-1}. \tag{2.2}$$

For each $T \in \binom{[\ell]}{d}$, set $V_T = U_T \cap D$. Then $|V_T| \le q^d$.

Step 2. Reduce the problem to matchings. We claim that if $T_1, \dots, T_{\lambda+1} \in \binom{[\ell]}{d}$ are pairwise disjoint, then

$$D = V_{T_1} \cup \cdots \cup V_{T_{\lambda+1}}. \tag{2.3}$$

By (2.1), the remaining coordinates can be split into $c - \lambda - 1$ sets $S_{\lambda+2}, \dots, S_c$, each of size $d - 1$. Thus

$$T_1 \cup \cdots \cup T_{\lambda+1} \cup S_{\lambda+2} \cup \cdots \cup S_c = [\ell].$$

Lemma 2.2 gives

$$\mathcal{C} = U_{T_1} \cup \cdots \cup U_{T_{\lambda+1}} \cup U_{S_{\lambda+2}} \cup \cdots \cup U_{S_c}.$$

After we intersect both sides with D, the last $c - \lambda - 1$ terms vanish by the definition of D. This proves (2.3).

Now fix $\boldsymbol{x} \in D$ and define

$$\mathcal{A}_{\boldsymbol{x}} = \left\{ T \in \binom{[\ell]}{d} : \boldsymbol{x} \notin V_T \right\}.$$

If $\mathcal{A}_{\boldsymbol{x}}$ contained $\lambda + 1$ pairwise disjoint sets, then (2.3) would fail. Hence $\nu(\mathcal{A}_{\boldsymbol{x}}) \leq \lambda$, so $|\mathcal{A}_{\boldsymbol{x}}| \leq m(\ell, d, \lambda)$. Therefore $\boldsymbol{x}$ belongs to at least $\binom{\ell}{d} - m(\ell, d, \lambda)$ sets V_T.

Step 3. Double count. Let

$$K = \{(\boldsymbol{x}, T) : \boldsymbol{x} \in V_T,\ T \in \tbinom{[\ell]}{d}\}.$$

Since $|V_T| \leq q^d$ for every T,

$$|K| \leq \binom{\ell}{d} q^d.$$

On the other hand, each $\boldsymbol{x} \in D$ is contained in at least $\binom{\ell}{d} - m(\ell, d, \lambda)$ sets V_T. Hence

$$|K| \geq |D| \left(\binom{\ell}{d} - m(\ell, d, \lambda) \right).$$

It follows that

$$|D| \leq \frac{\binom{\ell}{d}}{\binom{\ell}{d} - m(\ell, d, \lambda)} q^d.$$

Together with (2.2), this gives the desired bound. □

The code argument is now finished. The only term that remains is $m(\ell, d, \lambda)$, which is a matching problem.

2.3 The Erdős Matching Problem

For $\mathcal{A} \subseteq \binom{[\ell]}{d}$, the matching number $\nu(\mathcal{A})$ is the largest number of pairwise disjoint sets in $\mathcal{A}$. The number $m(\ell, d, \lambda)$ is the largest size of such a family with $\nu(\mathcal{A}) \leq \lambda$.

Conjecture 2.4 (Erdős Matching Conjecture [3])**.**

$$m(\ell, d, \lambda) = \max \left\{ \binom{\ell}{d} - \binom{\ell - \lambda}{d}, \binom{(\lambda+1)d - 1}{d} \right\}.$$

3 The Lower Bound for Frameproof Codes: Faithful Induced Packings

The lower bound reverses the main idea of the upper bound. We start with a largest family of d-sets with matching number at most λ. Its complement will be used as a template. We then pack many copies of this template into a complete multipartite hypergraph.

Let $\mathcal{A} \subseteq \binom{[\ell]}{d}$ be a largest family with $\nu(\mathcal{A}) \leq \lambda$. Thus $|\mathcal{A}| = m(\ell, d, \lambda)$. Define the template $\mathcal{F}$ to be the d-uniform hypergraph on $[\ell]$ with

$$E(\mathcal{F}) = \binom{[\ell]}{d} \setminus \mathcal{A}. \tag{3.1}$$

Hence

$$|E(\mathcal{F})| = \binom{\ell}{d} - m(\ell, d, \lambda).$$

We will construct many copies of $\mathcal{F}$ in a complete multipartite hypergraph.

Definition 3.1 (Complete multipartite d-graph)**.** Let $H_\ell^{(d)}(q)$ be the ℓ-partite, d-uniform hypergraph with parts $V_i = \{i\} \times [q]$ for $i \in [\ell]$. Its edges are all d-sets that meet each part in at most one vertex.

Thus $|E(H_\ell^{(d)}(q))| = \binom{\ell}{d} q^d$.

3.1 Faithful Induced Packings

We first recall the notion of a packing.

Definition 3.2 ($\mathcal{F}$-packing)**.** An $\mathcal{F}$-packing in $H_\ell^{(d)}(q)$ is a collection of pairwise edge-disjoint copies of $\mathcal{F}$.

Each copy of $\mathcal{F}$ contains $|E(\mathcal{F})|$ edges. Therefore, if an $\mathcal{F}$-packing contains M copies, then

$$M|E(\mathcal{F})| \le |E(H_\ell^{(d)}(q))|.$$

Hence

$$M \le \frac{\binom{\ell}{d} q^d}{|E(\mathcal{F})|}. \tag{3.2}$$

This is the natural upper bound for the size of an $\mathcal{F}$-packing.

For frameproof codes, we need two extra properties.

Definition 3.3 (Faithful induced packing)**.** An $\mathcal{F}$-packing in $H_\ell^{(d)}(q)$ is called

- *faithful* if vertex i of the template $\mathcal{F}$ is always mapped into the part V_i;
- *induced* if distinct copies with vertex sets U_i and U_j satisfy $|U_i \cap U_j| \le d$, and if $|U_i \cap U_j| = d$, then this common d-set is not an edge of either copy.

The faithful condition gives the direct link between packings and codes. Indeed, a faithful copy has exactly one vertex in each part. If

$$U = \{(1, x_1), \ldots, (\ell, x_\ell)\}$$

is its vertex set, then it gives the codeword $(x_1, \ldots, x_\ell) \in [q]^\ell$.

The induced condition is needed to show that the resulting code is frameproof.

Proposition 3.4 (From packings to frameproof codes)**.** *Let $U_1, \ldots, U_M$ be the vertex sets of a faithful induced $\mathcal{F}$-packing in $H_\ell^{(d)}(q)$. Then the corresponding M words in $[q]^\ell$ form a c-frameproof code.*

Proof. Suppose that the code is not c-frameproof. By the cover-free form of frameproofness, there exist $1 \le r \le c$ and distinct members $U_0, U_1, \ldots, U_r$ such that

$$U_0 \subseteq U_1 \cup \cdots \cup U_r.$$

For each vertex of U_0, choose one U_i that contains it. This gives pairwise disjoint sets $T_1, \ldots, T_r$ such that $T_i \subseteq U_0 \cap U_i$ and $U_0 = T_1 \cup \cdots \cup T_r$. Set $T_{r+1} = \cdots = T_c = \varnothing$. Since the packing is induced, $|U_0 \cap U_i| \le d$, and hence $|T_i| \le d$. Recall that

$$\ell = (\lambda + 1)d + (c - \lambda - 1)(d - 1).$$

If at most λ of the sets T_i had size d, then

$$|U_0| = \sum_{i=1}^{c} |T_i| \le \lambda d + (c - \lambda)(d - 1) = \ell - 1,$$

a contradiction. Hence at least $\lambda + 1$ of the sets T_i have size d.

For each such T_i, we have $T_i = U_0 \cap U_i$. Let $I_i = \{j \in [\ell] : T_i \cap V_j \neq \varnothing\}$. Since the copies are faithful, $|I_i| = d$. The induced condition says that T_i is not an edge of the copy of $\mathcal{F}$. Therefore $I_i \notin E(\mathcal{F})$, and hence $I_i \in \mathcal{A}$ by (3.1). The sets T_i are pairwise disjoint subsets of U_0. Since U_0 contains one vertex from each part, the corresponding sets I_i are also pairwise disjoint. Thus $\mathcal{A}$ contains $\lambda + 1$ pairwise disjoint members. This contradicts $\nu(\mathcal{A}) \leq \lambda$. □

Thus every faithful induced $\mathcal{F}$-packing gives a frameproof code of the same size. The remaining question is how large such a packing can be.

Liu, Ma, and Shangguan proved that the packing upper bound in (3.2) can be reached asymptotically.

Theorem 3.5 (Near-optimal faithful induced packing [4]). *For every fixed ℓ, d and every $\varepsilon > 0$, all sufficiently large q admit a faithful induced $\mathcal{F}$-packing in $H_\ell^{(d)}(q)$ with at least*

$$(1-\varepsilon)\frac{\binom{\ell}{d}q^d}{|E(\mathcal{F})|}$$

copies.

By Proposition 3.4, each copy in this packing gives one codeword. Since $|E(\mathcal{F})| = \binom{\ell}{d} - m(\ell, d, \lambda)$, we obtain the following lower bound.

Theorem 3.6 (Lower bound for frameproof codes [4]). *For fixed $c \geq 2$ and $\ell \geq 2$, every $\varepsilon > 0$, and all sufficiently large q,*

$$M_{c,\ell}(q) \geq (1-\varepsilon)\frac{\binom{\ell}{d}}{\binom{\ell}{d} - m(\ell, d, \lambda)}q^d.$$

Consequently, together with Theorem 2.3,

$$\lim_{q\to\infty}\frac{M_{c,\ell}(q)}{q^d} = \frac{\binom{\ell}{d}}{\binom{\ell}{d} - m(\ell, d, \lambda)}.$$

3.2 Pippenger–Spencer Theorem

We now explain the main idea behind Theorem 3.5. The goal is to find a faithful induced $\mathcal{F}$-packing whose size is close to the upper bound

$$\frac{|E(H_\ell^{(d)}(q))|}{|E(\mathcal{F})|}.$$

The main tool is a near-perfect matching theorem for hypergraphs.

Theorem 3.7 (Pippenger–Spencer [5]). *For every integer $t \geq 2$, $K \geq 1$, and $\varepsilon > 0$, there are $\gamma > 0$ and D_0 such that the following holds. Let H be a t-uniform hypergraph with $|V(H)| = n$ and $D \geq D_0$. Suppose*

(i) *all but at most γn vertices x satisfy $|d_H(x) - D| \leq \gamma D$;*

(ii) *$\Delta(H) < KD$;*

(iii) *$\Delta_2(H) < \gamma D$.*

Then H contains a matching that covers all but at most εn vertices.

The host H here is any t-uniform hypergraph that is almost regular, has controlled maximum degree, and has very small maximum codegree; it need not be complete.

Liu–Ma–Shangguan form an auxiliary hypergraph from candidate template copies and apply the theorem to that auxiliary hypergraph. The Liu–Ma–Shangguan argument can be summarized as follows:

(1) find a large faithful $H_\ell^{(d+1)}(1)$-packing $\mathcal{P}$ in $H_\ell^{(d+1)}(q)$;

(2) randomly sparsify the edge set of $H_\ell^{(d)}(q)$;

(3) construct an auxiliary $|E(\mathcal{F})|$-uniform hypergraph whose vertices are the retained d-edges and whose edges represent template copies;

(4) use concentration estimates to verify near-regularity, controlled maximum degree, and very small maximum codegree;

(5) apply a Pippenger-type near-perfect matching theorem; the resulting matching gives a faithful induced $\mathcal{F}$-packing.

4 Separating Hash Families

Frameproof codes ask us to separate one target codeword from one group. Separating hash families allow several groups and ask for one coordinate that separates all of them.

Definition 4.1 (Separating hash family)**.** Let X be a set of size n, and let $\mathcal{H} = \{h_1, \dots, h_N\}$ be a family of functions from X to $[q]$. Let $w_1, \dots, w_t$ be positive integers. We say that $\mathcal{H}$ is a *separating hash family of type* $\{w_1, \dots, w_t\}$, denoted by $\mathrm{SHF}(N; n, q, \{w_1, \dots, w_t\})$, if the following holds.

For every collection of pairwise disjoint sets $C_1, \dots, C_t \subseteq X$ with $|C_i| = w_i$, there is a function $h \in \mathcal{H}$ such that $h(C_i) \cap h(C_j) = \varnothing$ for all $i \neq j$. In this case, we say that h *separates* $C_1, \dots, C_t$.

For fixed N, q and type $\{w_1, \dots, w_t\}$, let

$$C(N, q, \{w_1, \dots, w_t\}) = \max\{n : \text{there exists an } \mathrm{SHF}(N; n, q, \{w_1, \dots, w_t\})\}.$$

We also write $u = w_1 + \cdots + w_t$.

For fixed N and type $\{w_1, \dots, w_t\}$, the main question is the growth of $C(N, q, \{w_1, \dots, w_t\})$ as $q \to \infty$.

The results discussed can be summarized as follows.

Setting	**Result or question**	**Reference**
General bounds	$\Omega_{u,N}\left(q^{N/(u-1)}\right) \le C(N, q, \{w_1, \dots, w_t\}) \le O\left(q^{\lceil N/(u-1)\rceil}\right)$.	Blackburn [8]; Blackburn et al. [6]
$(u-1) \mid N$	The two bounds have the same exponent. Hence $C(N, q, \{w_1, \dots, w_t\}) = \Theta\left(q^{N/(u-1)}\right)$.	From the general bounds
$(u-1) \nmid N$, $u \le N \le 2u-3$	The exponent in the general upper bound is attainable: $C(N, q, \{w_1, \dots, w_t\}) \ge q^{2-o(1)}$.	Wei–Zhang–Ge [7]
$(u-1) \nmid N$	The exponent is now known for all N: $C(N, q, \{w_1, \dots, w_t\}) \ge q^{\lceil N/(u-1)\rceil - o(1)}$.	Personal communication
$(u-1) \nmid N$	Determine whether $C(N, q, \{w_1, \dots, w_t\}) = o\left(q^{\lceil N/(u-1)\rceil}\right)$ or $C(N, q, \{w_1, \dots, w_t\}) = \Theta\left(q^{\lceil N/(u-1)\rceil}\right)$.	Open
$N = 6$, type $\{2, 3\}$	$C(6, q, \{2, 3\})$ is the first unresolved case with $t = 2$ in the finer problem above.	Open

References

[1] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, *IEEE Trans. Inform. Theory* **44** (1998), 1897–1905.

[2] S. R. Blackburn, Frameproof codes, *SIAM J. Discrete Math.* **16** (2003), 499–510.

[3] P. Erdős, A problem on independent r-tuples, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **8** (1965), 93–95.

[4] M. Liu, Z. Ma, and C. Shangguan, Near optimal probabilistic constructions of frameproof codes, *IEEE Trans. Inform. Theory* **71** (2025), 4137–4144, DOI: 10.1109/TIT.2025.3558360.

[5] N. Pippenger and J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs, *J. Combin. Theory Ser. A* **51** (1989), 24–42.

[6] S. R. Blackburn, T. Etzion, D. R. Stinson, and G. M. Zaverucha, A bound on the size of separating hash families, *J. Combin. Theory Ser. A* **115** (2008), 1246–1256.

[7] X. Wei, X. Zhang, and G. Ge, Separating hash families with large universe, *J. Combin. Theory Ser. A* **216** (2025), 106075.

[8] S. R. Blackburn, Perfect hash families: probabilistic methods and explicit constructions, *J. Combin. Theory Ser. A* **92** (2000), 54–60.