Zero-error coding for multiple-access channels as a new test bed for AG-codes

Elena Egorova, Grigory Kabatiansky

AGC^2T-16 Conference

Marseille, France,
2017

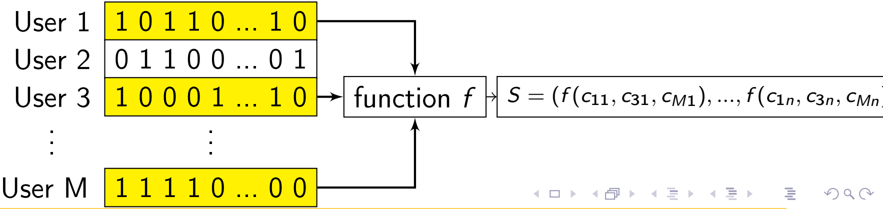
Outline

1. Introduction: when AG codes are better than random codes?
   - q-ary codes in Hamming distance $\Rightarrow q \geq 49$ (TVZ)
   - Authentication codes (Vladuts)
   - New areas of possible applications of AG codes: Multiple access channels (MAC) and Fingerprinting codes
2. Multiple access channels (MAC)
   - Problem statement & some previous results
   - Adder channel
   - Disjunctive channel
   - A& B channels
3. Malicious MAC or Digital fingerprinting codes
4. Weighted adder channel or Multimedia fingerprinting codes
5. Separating codes

Signature codes for MAC

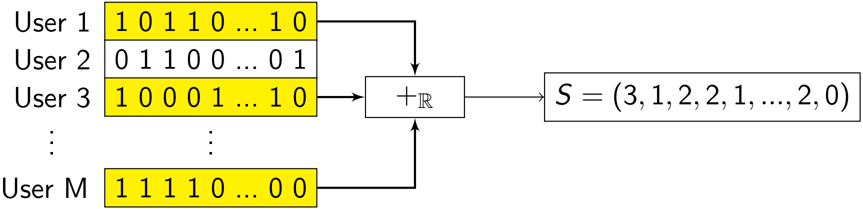
- Let M be the number of users, i-th user has its personal vector $c_i = (c_{i1}, \dots, c_{in})$ of length n, i.e. code $C = \{c_1, \dots, c_M\}$.
- **Input:** during each time slot t users or less are active (t might be equal to M), i.e. transmit their vectors.
  Let $I = \{i_1, \dots, i_k\}$, $k \leq t$ be a set of active users.
- **Output** is a vector S, its each position is some function of values at the corresponding position of transmitted vectors, i.e. $S = (\dots, f(c_{i_1 j}, \dots, c_{i_k j}), \dots)$, $i_l \in I$, $j \in [n]$

Signature code:= from S uniquely determine all active users



Adder channel

Definition. The input is binary vectors, the output is the sum of vectors (as vectors over $\mathbb{R}$).



Known results (based on random coding and entropy method):

$$\frac{\log t}{4t}(1 + o(1)) \leq R \leq \frac{\log t}{2t}(1 + o(1))$$

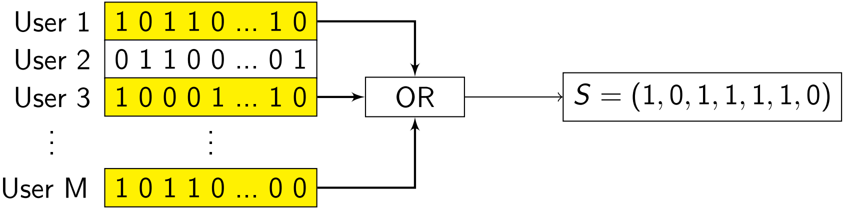
[D'yachkov A. G., Rykov V. V. On a Coding Model for a Multiple-Access Adder Channel 1981.]

Disjunctive channel

Definition. The input is binary vectors, the output is a bit-wise logical OR ($\vee$): $0 \vee 0 = 0$, $0 \vee 1 = 1 \vee 0 = 1 \vee 1 = 1$.

Corresponding codes called *superimposed codes* (Kautz, Singleton 1964).

In terms of sets: Erdos et al. 1982, Family of sets in which no set is covered by the union of two others.



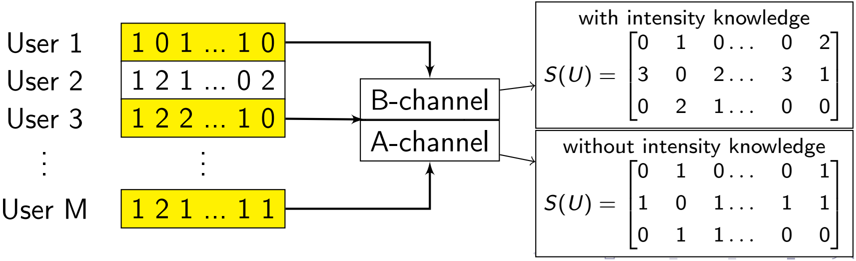
Result (random coding): [Erdos et al., D'yachkov & Rykov, 1982]

$$R \geq \frac{\ln 2}{t^2}(1 + o(1))$$

M-user q-frequency MAC with and without intensity knowledge [Chang& Wolf, 1981]

- $Q = \{0, 1, ..., q-1\}$, $C = \{c_1, ..., c_M\} \subseteq Q^n$
- Output of B-channel — composition of vectors from U, i.e. matrix $S(U) = ||w_{ij}||_{i=1..q, j=1..n}$, where w_{ij} equals the number of times when element $(i-1) \in Q$ appeared at $j-$th positions of vectors from U.
- Output of A-channel — matrix $S(U) = ||w_{ij}||_{i=1..q, j=1..n}$, element w_{ij} equals 1 if element $(i-1) \in Q$ appeared at j-th position of vectors from U and 0 otherwise.

U = {User1, User 3, User M}



B-channel and Adder channel

Note that B-channel with $q = 2$ is the same as the adder channel. Another name for the same problem is *Finding $\leq t$ counterfeit coins among M coins on exact (spring) scale.*
For $t = M$ random coding [Erdos & Renyi, 1964] proves that the minimal number of weightings is at most

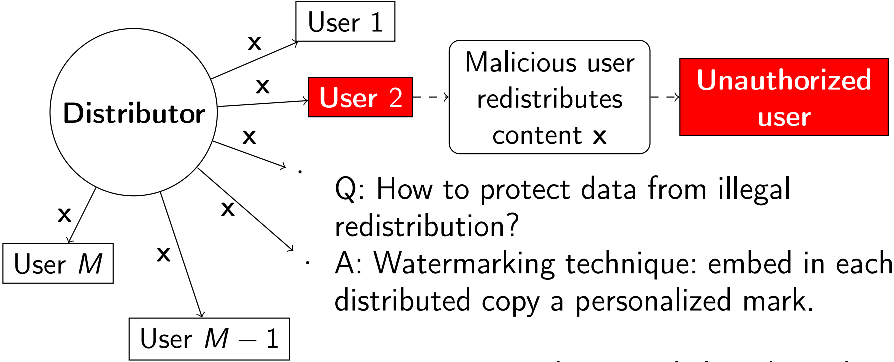
$$3M(\log_2 M)^{-1},$$

on the other hand, entropy bound says that the number of weightings is at least

$$2M(\log_2 M)^{-1}$$

Lindstrom, Counter and Mills provided exact construction with $2M(\log_2 M)^{-1}$.
If t is constant then random coding gives the best known lower bound except the case $t = 2$ when binary BCH codes give better bound.

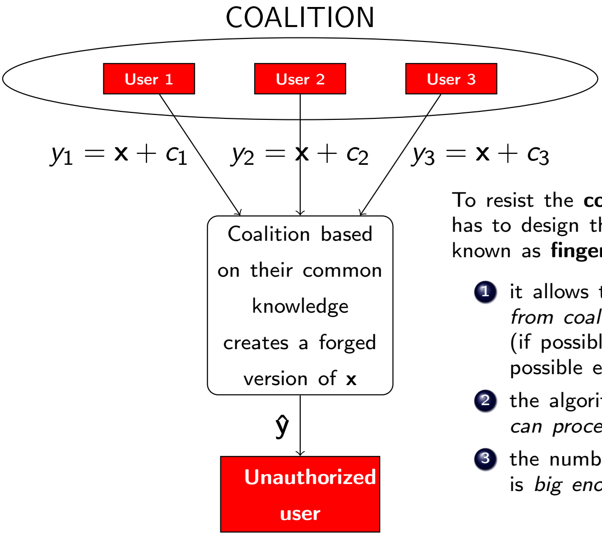
How to protect data from illegal redistribution or codes for Malicious MAC



Q: How to protect data from illegal redistribution?
A: Watermarking technique: embed in each distributed copy a personalized mark.

BUT: Watermarking can help only in the case of **single** traitor.

Collusion attacks



To resist the **collusion attack** distributor has to design the set of user marks, known as **fingerprinting code**, such that

1. it allows to identify *at least one traitor from coalition* with zero error (if possible) or with the minimum possible error rate,
2. the algorithm of identification *can proceed in real time*,
3. the number of authorized users is *big enough*.

Discrete model

$$x_i = \boxed{\dots}\boxed{1}\boxed{0}\boxed{\dots}\boxed{2}\boxed{\dots}\boxed{2}\boxed{\dots} \to (1, 0, 2, 2) = c_i$$

$$x_j = \boxed{\dots}\boxed{1}\boxed{1}\boxed{\dots}\boxed{2}\boxed{\dots}\boxed{1}\boxed{\dots} \to (1, 1, 2, 1) = c_j$$

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$\{1\} \quad \{0, 1\} \quad \{2\} \quad \{1, 2\}$$

User i $\boxed{1\ 0\ 2\ 2}$ $\to$ $\boxed{\text{Union}}$ $\to$ $\hat{x} = (1, \hat{x}_2, 2, \hat{x}_4)$
User j $\boxed{1\ 1\ 2\ 1}$ $\to$ where $\hat{x}_2 \in \{0, 1\}, \hat{x}_4 \in \{1, 2\}$

Main problem: for any t-coalition and any given $\hat{x}$ generated by the coalition the distributor can correctly identify at least one member of the coalition.

Codes with Identifiable parent property (IPP)

Definition. A code C called t-**IPP code** if for any vector $\hat{\mathbf{x}} \in Q^n$ the intersection of all coalitions that can create $\hat{\mathbf{x}}$ is not empty, i.e.

$$\bigcap_{U:\ |U| \leq t,\ \hat{\mathbf{x}} \in \langle V \rangle_t} U \neq \emptyset$$

or no one coalition of cardinality t can create $\hat{\mathbf{x}}$.

IPP codes as codes for malicious MAC[Barg A. et al., 2003]: users from coalition can be considered as active users, but the output of MAC is under control of a coalition.

As a results the code (distributor) cannot recover the entire set of active users, and the distributor's goal is to find for sure at least one user from the coalition.

Good t-IPP code exists, i.e. $R \geq c(t) > 0 \Leftrightarrow t < q = |Q|$.
[Barg A. et al., 2001, based on random coding]

Multimedia digital fingerprinting codes = continuous model

Digital content: $\mathbf{x} \in \mathbb{R}^L$ — host multimedia signal.
Multimedia digital fingerprinting code: let $\mathbf{f}_1, \ldots, \mathbf{f}_n \in \mathbb{R}^L$ be noise-like orthonormal signals, then for $i = 1, ..., M$

$$\boxed{\mathbf{w}_i = \sum_{j=1}^{n} b_{i,j}\mathbf{f}_j, \ \text{, where } b_{i,j} \in \{1, -1\} \text{ or } \{0, 1\}}$$

— fingerprint for the i-th user.

Embedding of fingerprints: watermarked version of the content for the $i-$th user

$$\boxed{\mathbf{y}_i = \mathbf{x} + \sum_{j=1}^{n} b_{i,j}\mathbf{f}_j = \mathbf{x} + \mathbf{w}_i}.$$

Assumption: members of a coalition $U \subset \{1, \dots, M\}$ have no information about signals $\mathbf{f}_j$ and, therefore, they have no way of manipulating them, except for linear attack.
Linear attack:

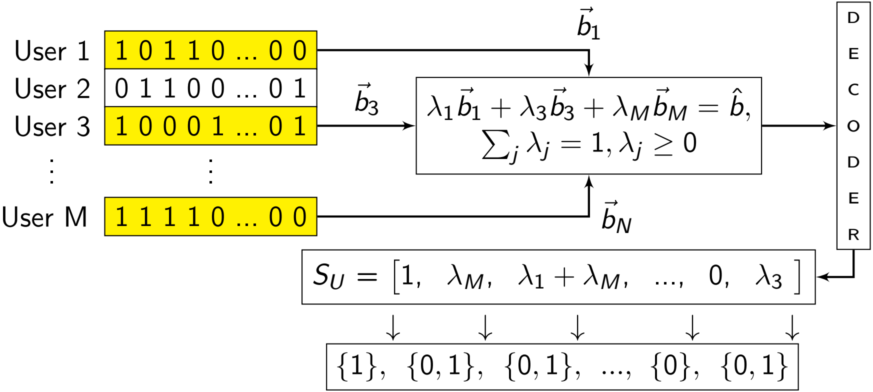
$$\widehat{\mathbf{y}} = \sum_{i \in U} \lambda_i \mathbf{y}_i, \text{ where } \sum \lambda_i = 1 \text{ and } \lambda_i \geq 0.$$

Forged content: $\widehat{\mathbf{y}} = \mathbf{x} + \sum_{i \in U} \lambda_i \mathbf{w}_i = \mathbf{x} + \sum_{i \in U} \sum_{j=1}^{n} \lambda_i b_{i,j} \mathbf{f}_j$.
Identification: the dealer evaluates

$$T = (\tau_1, \dots, \tau_n), \text{ where } \tau_j = (\widehat{\mathbf{y}} - \mathbf{x}, \mathbf{f}_j) = \sum_{i=1}^{t} \lambda_i b_{i,j}$$

and wants to find at least one member of a coalition or the whole coalition.

i-th user corresponds to the vector of coefficients $\vec{b}_i = (b_{i1}, \ldots, b_{in})$.



Distributor's goal: construct a code C such that any coalition $U, |U| \leq t$ can be uniquely recovered from its *signature* $\mathbf{S}_U$.

Weighted adder channel

How to find a coalition by its signature?
Let vectors $\vec{b}_i$, $i = 1, ..., M$ form a parity-check matrix B of the **binary BCH code** correcting t errors. Then different coalitions have different signatures. Indeed, if they coincide then we have linear dependency of $2t$ or less columns of matrix B — contradiction.
The rate of the corresponding code is

$$R \geq \frac{1}{t}$$

Unfortunately, it doesn't give a decoding algorithm.
Moreover, this construction fully relies on the assumption of exact evaluation of signatures.
What AG codes can do for this problem?

Separation and Hashing

A sequence $(A_1, \dots, A_t)$ of pairwise disjoint sets of codevectors called a $(s_1, \dots, s_t)$-configuration if $|A_j| = s_j$ for all j. Such a configuration is separated if there is a position i, such that for all $l \neq l'$ every vector of A_l is different from every vector of $A_{l'}$ on position i.

Definition. A code is $(s_1, \dots, s_t)$-separating if every $(s_1, \dots, s_t)$-configuration is separated.

Definition. A code is t-hash if for any t different code vectors there is a position which separates them.

Note that t-hash is $(1, \dots, 1)$-separating.

Remark: If the minimal code distance d satisfies

$$\binom{t}{2}(n-d) > n \text{ then code is } t\text{-hash.}$$

Open problem: can we replace for AG codes this condition for a somewhat weaker one?

Conclusion

It's known that AG codes sometimes can be very useful and perform better than random coding

- Signature codes for different models of *multiple access channels via AG codes* :
  -improve lower bounds
  -provide explicit constructions
- The same question for different types of separating codes.

References

- Chang S. C., Wolf J. K., "On the T-user M-frequency noiseless multiple-access channel with and without intensity information", *IEEE Trans. Inform. Theory* vol. 27, no. 1, pp. 41-48, 1981.
- Kautz W.H. and Singleton R.R. "Nonrandom binary superimposed codes", *IEEE Trans. Inform. Theory*, vol. 10, no. 4, pp. 363-377, 1964.
- A.G. Dyachkov, I.V. Vorob'ev, N.A. Polyansky and V.Yu. Shchuin, "Bounds for the rate of disjunctive codes", *Problems Information Transmission*, vol. 50, no. 1, pp. 31-63, 2014.
- A. Barg, G. Cohen, S. Encheva, G, Kabatiansky, and G. Zémor, "A hypergraph approach to the identifying parent property: the case of multiple parents," *SIAM J. Disc. Math*, vol. 14, pp. 423–431, 2001.
- D. Boneh and J. Shaw.
  Collusion-secure fingerprinting for digital data.
  IEEE Trans. Inform. Theory, 44(5):1897–1905, 1998.
- A. Barg, G. R. Blakley, and G. Kabatiansky.
  Digital fingerprinting codes: Problem statements, constructions, identification of traitors.
  IEEE Trans. Inform. Theory, 49(4):852–865, 2003.
- M. Fernandez, G.Kabatiansky, and J. Moreira.
  Almost IPP-codes or provably secure digital fingerprinting codes.
  In *Proc. IEEE International Symp. Information Theory (ISIT 2015)*, pages 1595–1599. IEEE Computer Society, 2015.
- K.J.R. Liu, W. Trappe, Z.J.Wang, M. Wu and H.Zhao. Multimedia fingerprinting forensics for traitor tracing. Vol. 4. Hindawi Publishing Corporation, 2005.

THANK YOU FOR YOUR ATTENTION!