Analysis of two tracing traitor schemes via coding theory

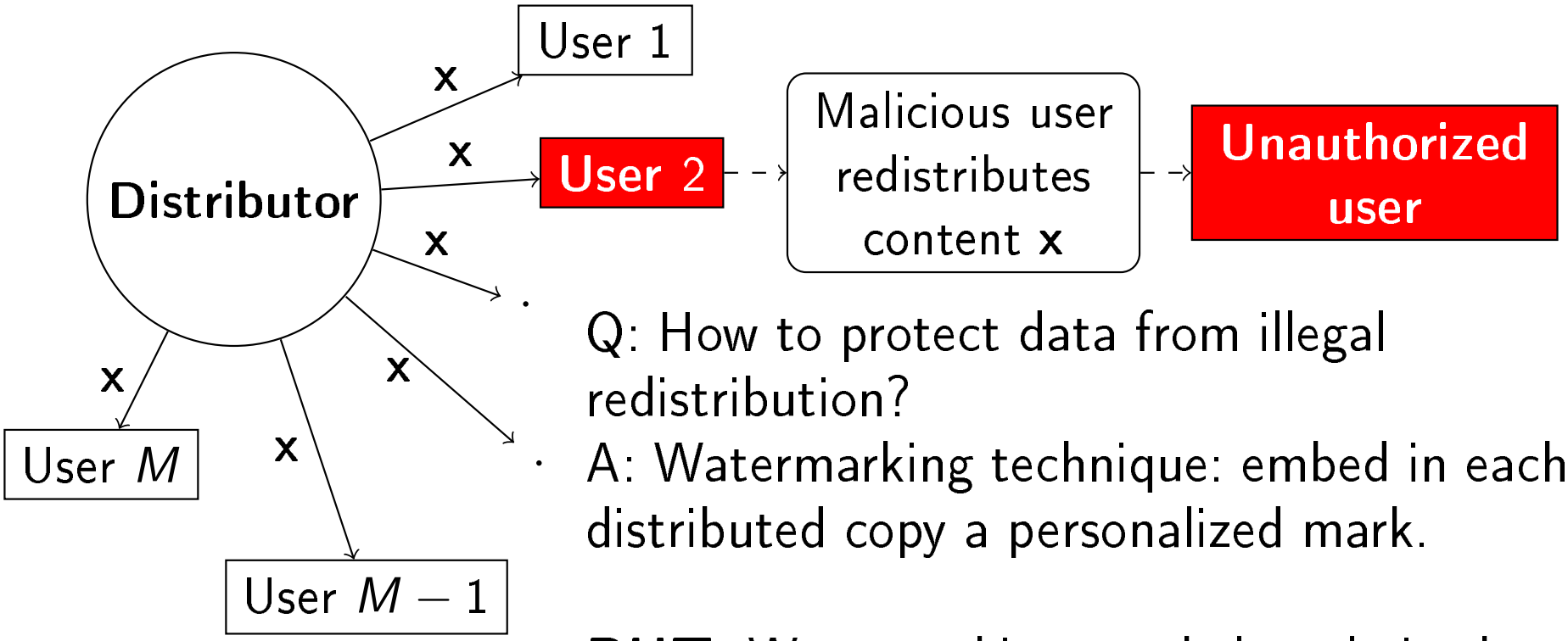
Elena Egorova, Grigory Kabatiansky

Skolkovo institute of science and technology (Skoltech)
Moscow, Russia

5th International Castle Meeting on Coding Theory and Applications

Villa Manor, Estonia,
2017

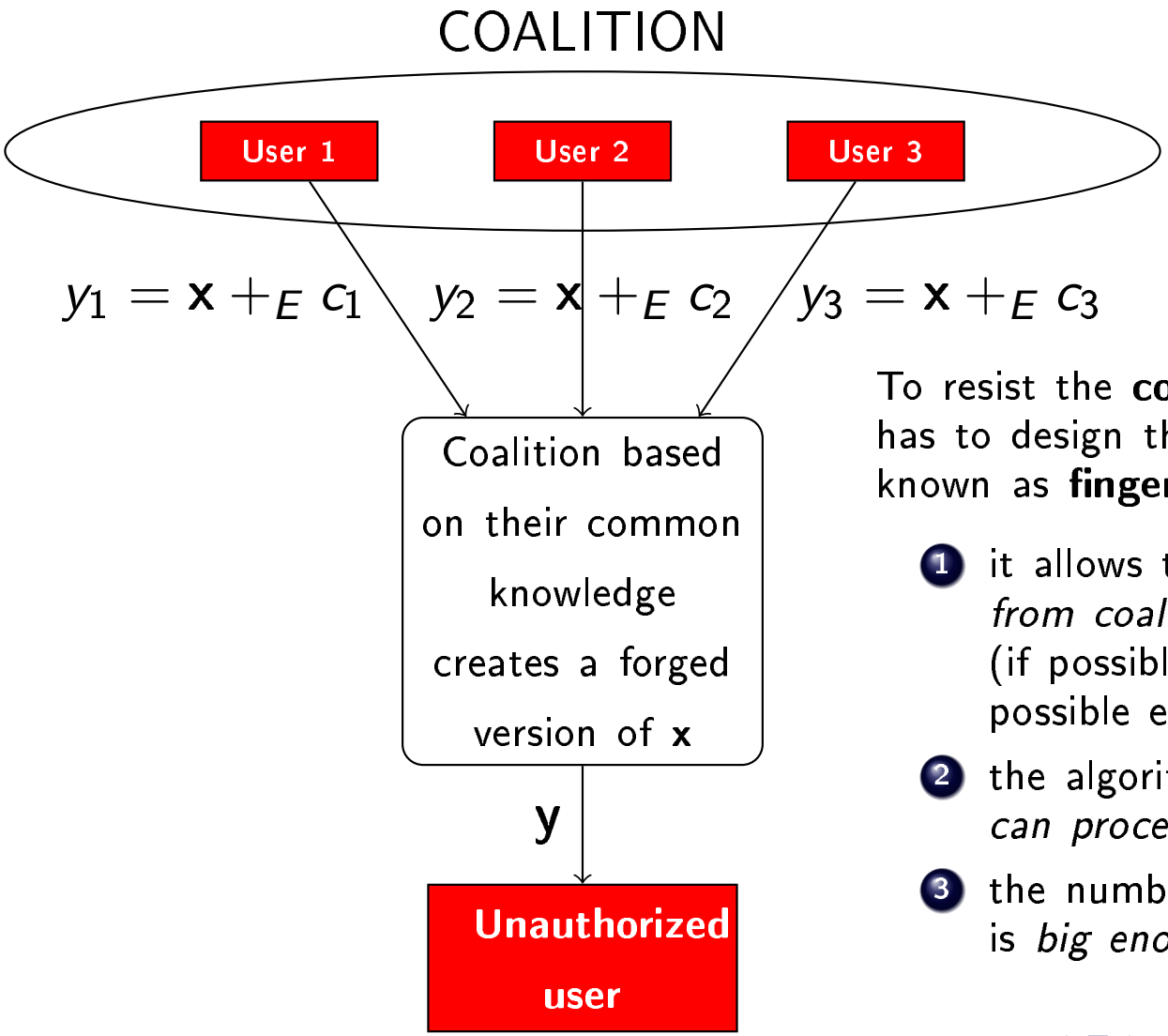
Watermarking vs Collusion attack



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.

Fingerprinting codes against 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*.

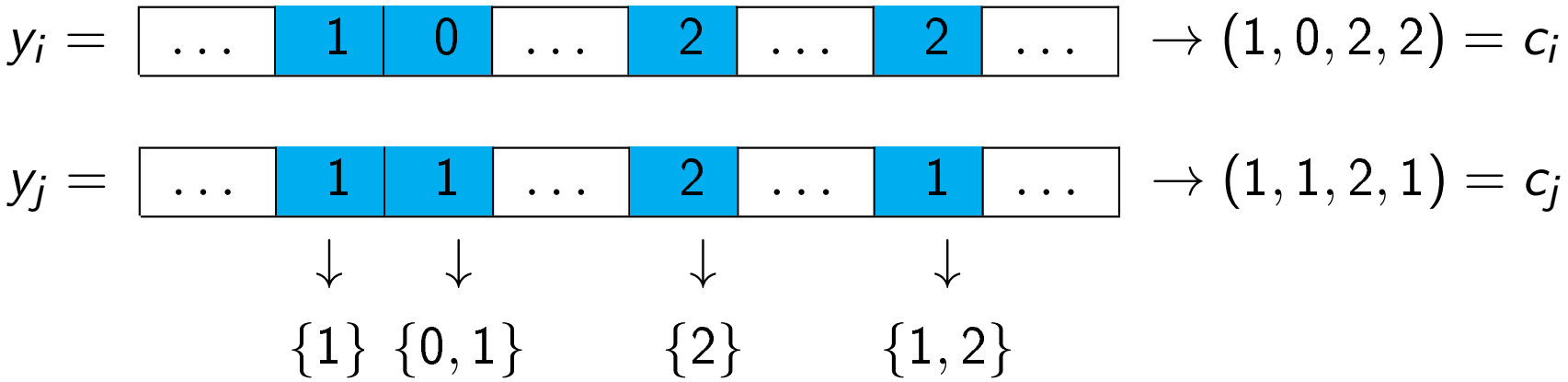
Collusion attack paradigm

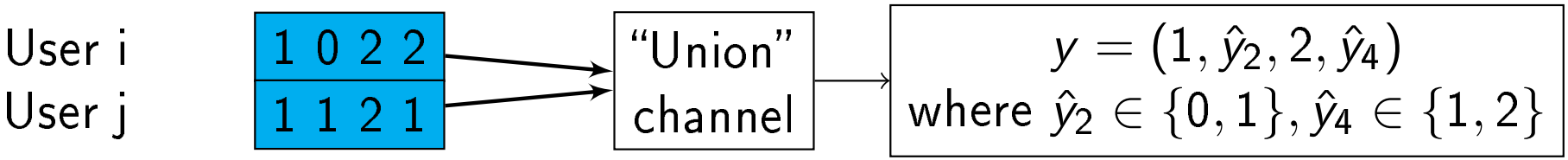
Let $C = \{c_1, ..., c_M\} \subset A^n$ be a fingerprinting code. All users know the code.
We assume that no more than t users can form a coalition $U = \{i_1, ..., i_k\}, k \leq t$.
Let $< U >$ be a set of all forged vectors that coalition U can generate according to specificity of the considered model (examples are on the next slide).
Goal of the distributor: construct a code C that satisfies the t-*Identifiable parent property* (t-IPP for short).
Definition. A code C has t-IPP property if for any vector $y \in A^n$

$$\bigcap_{U:\, |U| \leq t,\, y \in \langle U \rangle} U \neq \emptyset$$

or no one coalition U of cardinality t can create y.

Examples 1: t-IPP codes





Goal: for any t-coalition and any given y generated by the coalition the distributor can correctly identify at least one member of the coalition.

t-IPP codes

Formally, let $P_j(U) = \{u_j \mid u = (u_1, ..., u_n) \in U\}$, then

$$y \in \langle U \rangle_t := \{(y_1, ..., y_n) \in Q^n \mid y_j \in P_j(U)\}$$

Definition [Hollmann et al, 1998] A code C called t-**IPP code** if for any vector $y \in Q^n$

$$\bigcap_{U:\, |U| \le t,\, y \in \langle U \rangle_t} U \neq \emptyset$$

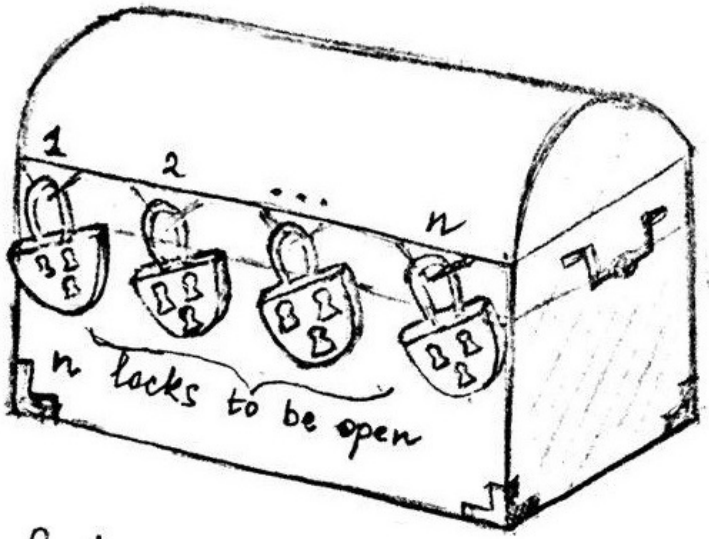
or no one coalition U of cardinality t can create y.

Let $R(C) = \frac{\log_q N}{n}$ — rate of a code C.

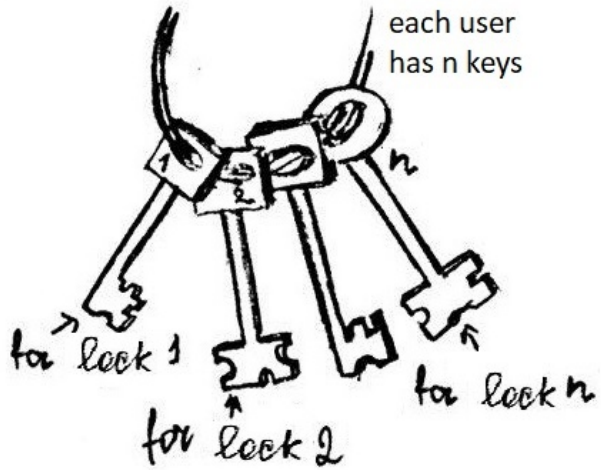
Theorem [Barg A. et al., 2001]
Good t-IPP code exists, i.e. $R \geq c(t) > 0 \Leftrightarrow t < q = |Q|$.

$\Rightarrow$ for the most interesting case of binary codes the IPP property does not hold even in the case of only two traitors.

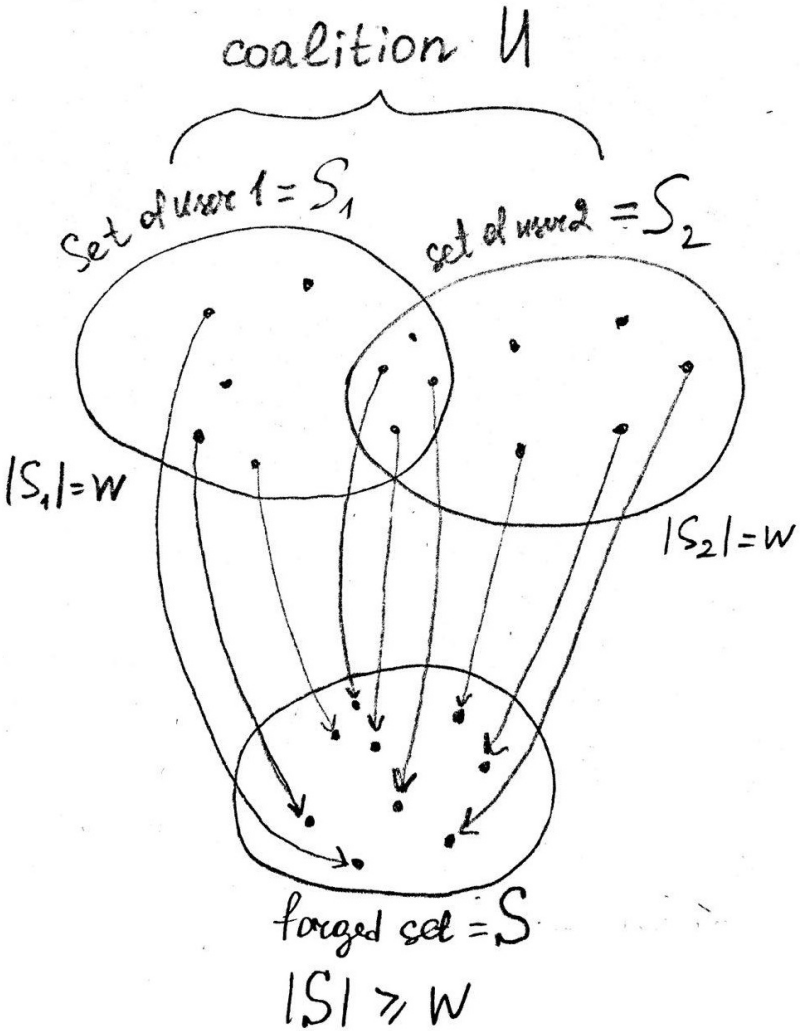
Toy example of IPP codes



each lock has q keyholes



Example 2: t-IPP set systems



Distributors chooses: a set of base keys X, $|X| = N$ and a parameter $w < N$.

i-th user has: the corresponding set of shares $S_i \subset X$, $|S_i| = w$.

Collusion attack: coalition $U = \{i_1, ..., i_k\}$ generates any element of

$$\langle U \rangle = \{S \mid S \subset \bigcup_{u \in U} S_u, |S| \geq w\}$$

t-IPP set systems

The goal: is, given any forged $\geq w-$subset of X, to be able to identify at least one user from the malicious coalition.

Definition

The family $F = \{S_1, ..., S_M\}$ of w-subsets of X, $|X| = N$ is called a t-IPP set system (t-IPPSS) if for every $S \subseteq X$, $|S| \geq w$ the intersection of all coalitions of size $\leq t$ that can generate S is either not empty, i.e.

$$\bigcap_{U:|U|\leq t, S\in<U>} U \neq \emptyset$$

or there is no $\leq t$-coalition that can generate S.

Remark: q-ary IPP codes of length n can be considered as a "weak" IPP set systems with $w = n$ and $N = qn$.

Traceability

Definition

Family $F = \{S_1, ..., S_M\}$ of w-subsets of X, $|X| = N$ is called t-**traceability set system** (t-TSS) if for any coalition $U = \{i_1, ..., i_k\} \subset [M]$, $k \leq t$ and any $S \in < U >$ holds

$$|S \cap S_j| < \max_{u \in U} |S \cap S_u| \text{ for any j} \notin U$$

It means that the search of malicious users reduces to the search of "closest" sets. For the corresponding binary codes, it's equivalent to the minimal distance decoding.
Moreover, any t-TSS is also t-IPPSS, so the lower bounds on the size of t-TSS are also valid for t-IPPSS.

Sufficient condition of t-traceability

Lemma

If $|S_i \cap S_j| < w/t^2$ for any $S_i, S_j \in F, i \neq j$, then F is a t-traceability set system.

Proof

Consider any coalition $U = \{i_1, ..., i_k\}$, $k \leq t$ and any $S \in < U >$.
Then, $\max_{i_l \in U} |S_{i_l} \cap S| \geq w/t$ since $|S| \geq w$.
On the other hand, for any $j \notin U$,

$$|S_j \cap S| \leq |S_j \cap (S_{i_1} \cup ... \cup S_{i_k})| \leq |S_j \cap S_{i_1}| + ... + |S_j \cap S_{i_k}| < t \cdot \frac{w}{t^2} = \frac{w}{t}, \text{ Q.E.D.}$$

The system of sets given in this lemma is the same as a constant-weight code of length N, weight w and distance $d \geq 2\tau n$ where $\tau = \omega(1 - 1/t^2)$ and $\omega := w/N$ (relative weight).

Let $M(N, \omega)$ be the maximum possible size of t-TSS with $|X| = N$ and weight ωN. Denote by

$$R_t^{TSS}(N, \omega) := \frac{\log_2 M(N, \omega)}{N}$$

the maximum possible rate and by

$$R_t(\omega) := \liminf_{N \to \infty} R_t^{TSS}(N, \omega)$$

.

Theorem

There exists a t-TSS code with rate

$$R_t(\omega) \geq H(\omega) - \omega H\left(\frac{\tau}{\omega}\right) - (1 - \omega) H\left(\frac{\tau}{1 - \omega}\right)$$

where H is binary entropy function and $\tau = \omega(1 - 1/t^2)$.

Proof of the theorem

Proof: Let $m_w(N, 2v)$ be the maximal cardinality of constant-weight code of length N, weight w and minimal distance $2v$. The corresponding Gilbert bound has the following form

$$m_w(N, 2v) \geq \frac{\binom{N}{w}}{\sum_{i=0}^{v-1} \binom{w}{i}\binom{N-w}{i}}$$

Asymptotically:

$$R = \frac{\log_2 m_w(N, 2v)}{N} \geq H(\omega) - \omega H\left(\frac{\tau}{\omega}\right) - (1-\omega) H\left(\frac{\tau}{1-\omega}\right) \quad (1)$$

where $\tau = vN^{-1}$.
Hence, there are constant-weight codes which rate satisfies (1). The corresponding family of sets is, according to lemma, t-TSS, Q.E.D.

Previous results:

1. Bound from **[Tracing traitors, 1994]** $R_t(\omega_0) \geq \frac{1}{8t^4}$ with $\omega_0 = \frac{1}{2t^2}$.

2. Bound from **[Gu, Y., Miao, Y. 2016]** $|M(N,\omega)| \geq \frac{\binom{N}{\lceil w/t^2 \rceil}}{\binom{w}{\lceil w/t^2 \rceil}}$

what asymptotically means $R_t(\omega) \geq H(\frac{\omega}{t^2}) - 2\omega H(\frac{1}{t^2})$.

For $t = 2$: theorem proves that $R_2 = \max_\omega R_t(\omega) \geq 0.0181$ and the best known bound is $R_{2,M} = 0.0159$ [Gu, Miao, 2016].

For $t = 3$: the new bound is $R_3 = 0.00316$ and the best known bound is $R_{3,M} = 0.0027$ [Gu, Miao, 2016].

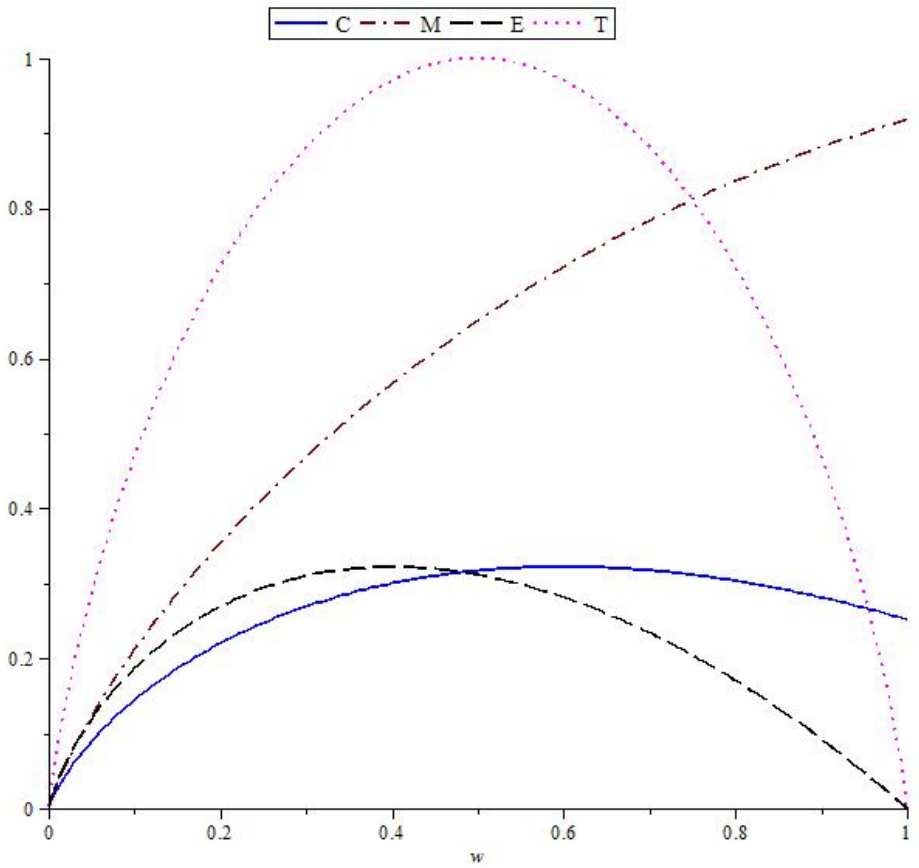
For arbitrary t: since

$$H(\omega) + H(t^{-2}) \geq (1-\omega)H\left(t^{-2}\omega/(1-\omega)\right) + H(t^{-2}\omega),$$

it can be seen that the new bound is always better than previously known ones.

Conclusion

- We proved the new lower bound on the rate of t-TSS codes, which, of course, is the lower bound for t-IPPSS. But we don't know BETTER lower bounds for t-IPPSS.
- It was proved by Gu and Miao in 2016 that t-TSS is t^2-superimposed (cover-free) codes. That gives rather good upper bound for t-TSS. But known upper bounds for IPPSS are rather weak, we think.



Trivial bound: $R_T(\omega) \leq H(\omega)$.

Bound of Collins: $R_C(\omega) \leq H(\omega/3) - 2\omega/3$.

Bound of Miao: $R_M(\omega) \leq H(\omega/3)$

Bound of Erdös: $R_E(\omega) \leq H(\omega/2) - \omega$.

Thank you for your attention!