New Progress on Combinatorial Schemes for Broadcast Encryption and Codes for Multimedia Fingerprinting

New Progress on Combinatorial Schemes for Broadcast Encryption and Codes for Multimedia Fingerprinting

March　2018

GU　YUJIE

New Progress on Combinatorial Schemes for Broadcast Encryption and Codes for Multimedia Fingerprinting

Graduate School of Systems and Information Engineering

University of Tsukuba

March　2018

GU　YUJIE

New Progress on Combinatorial Schemes for Broadcast Encryption and Codes for Multimedia Fingerprinting

Yujie Gu

University of Tsukuba, 2018

Supervisor: Ying Miao

In this prosperous information age, to protect copyright of digital contents is an urgent problem to be solved. Protection against digital content copyright violation is an important but difficult challenge. The fingerprinting technique can be used to resist collusion attacks of illegal key redistribution for broadcast encryption and collusion attacks of illegal content redistribution for multimedia contents.

This dissertation is devoted to study anti-collusion schemes for broadcast encryption and anti-collusion codes for multimedia fingerprinting, which are based on the fingerprinting technique, by virtue of combinatorial methods, such as tools in extremal set theory, extremal graph theory, probabilistic methods and combinatorial design theory. We derive new upper and lower bounds for the anti-collusion combinatorial structures and also provide constructions to achieve our new bounds, which greatly improve the previously known results and further promote the development of broadcast encryption and multimedia fingerprinting.

Traceability schemes (TSs) were introduced by Stinson and Wei [85] as a generalization of traceability codes proposed by Chor, Fiat and Naor [25, 26]. Cover-free families (CFFs) were introduced by Kautz and Singleton [60] in the disguise of binary superimposed codes. In this dissertation, we find a very interesting phenomenon, that is, a t-TS is in fact a t^2-CFF. Based on this new relationship, we derive new upper bounds for TSs by exploiting a combinatorial structure called *own-subset*, investigated by Erdős, Frankl and Füredi in [39]. We construct many infinite families of optimal TSs, achieving the new upper bounds, by means of sunflowers and combinatorial designs. A constructive lower bound for TSs is also given, which has the same order of magnitude as the upper bounds.

We introduce a unified concept of parent-identifying schemes to include many combinatorial structures with the parent-identifying properties as special cases. An equivalent relationship between parent-identifying schemes and forbidden configurations is established. Moreover, we correspond the research problems of parent-identifying schemes to a kind of Turán-type problems. The original idea of parent-identifying codes was introduced by Hollmann *et al.* in [53]. Collins [29] studied the parent-identifying set systems (IPPS) for broadcast encryption. In this dissertation, we give an improvement on the known upper bound for IPPS by using techniques in extremal set theory. A probabilistic lower bound for IPPS is also provided, which has the same order of magnitude as the new upper bound. Furthermore, better upper bounds for 2-IPPS$(4, v)$ and 3-IPPS$(6, v)$ are respectively derived by means of the graph removal lemma in extremal graph theory.

We introduce a special kind of CFFs, called union-intersection-bounded families (UIBFs), due to its applications in broadcast encryption. The relationships among UIBF, TS, IPPS and CFF are investigated. The upper bounds and lower bounds for UIBF are provided, which have the same order of magnitude for certain cases.

Multimedia parent-identifying codes (MIPPCs) were introduced by Cheng *et al.* [22] for multimedia fingerprinting. In the literature, there is no lower bound for general MIPPCs. We show the first probabilistic existence result for MIPPCs, which asymptotically achieves the known upper bound in certain cases and is very close to the known upper bound in other cases.

Finally, we conclude the work presented in this dissertation. Some open problems related with the topics in this dissertation are also discussed.

ACKNOWLEDGMENTS

First, I would like to express my deep gratitude to my supervisor Prof. Ying Miao for his guidance, encouragement and patience to help me grow up from a research-infant. He introduced me many interesting topics upon which this dissertation is based. This dissertation would not have come into being without his endless support and helpful advices. I am also grateful to him for supporting me to attend many interesting conferences. It is fortunate for me to have him as my supervisor.

I would like to appreciate Prof. Maiko Shigeno and Prof. Yusuke Kobayashi who served as the members of my advisory group, for their time, support and friendly help during my whole Ph.D. program. I want to thank Prof. Masahiro Hachimori, Prof. Hiroyasu Ando and Prof. Masanori Sawa for serving as members of my degree evaluation committee. Their time, patience and valuable comments are also appreciated. Many thanks to Prof. Ryoh Fuji-Hara for many kind and useful suggestions for my study in Tsukuba.

I would like to appreciate Prof. Grigory Kabatiansky for his valuable comments and beneficial discussions. I am thankful to Prof. Gennian Ge, my master supervisor at Zhejiang University, who led me to the area of combinatorics and its applications. Many thanks to him for suggesting me coming to University of Tsukuba to pursue my Ph.D. degree.

I would like to thank Dr. Yuichiro Fujiwara at Chiba University, who taught me a lot on research in our joint seminars. I also want to thank Prof. Sanpei Kageyama, Prof. Masakazu Jimbo, Prof. Miwako Mishima, Prof. Nobuko Miyamoto, Prof. Koji Momihara, Dr. Kazuki Matsubara, Dr. Shoko Chisaki and Dr. Xiaonan Lu for their help and discussions in many workshops.

I am grateful to Dr. Sihuang Hu, Dr. Hengjia Wei, Dr. Shuxing Li, Dr. Yiwei Zhang and Dr. Chong Shangguan for their kind assistance. I am thankful to Prof. Minquan Cheng and Dr. Jing Jiang for many interesting discussions. I also thank all my laboratory members, Dr. Xun Wang, Dr. Ling Jiang, Satoshi Noguchi, Han Cai, Jinping Fan, Hua Shuai, Minfeng Shao, Wei Li and Yueting Li, for their discussions in weekly seminars and help in daily life.

Many thanks to Shohei for his encouragement, patience and help. Special thanks to my family for their boundless love and support.

Finally, I would like to quote Albert Einstein's words: "A hundred times every day I remind myself that my inner and outer life are based on the labors of other men, living and dead, and that I must exert myself in order to give in the same measure as I have received and am still receiving."

The future is coming.

Contents

Chapter 1
Introduction

With the development of communication networks, an urgent problem to be solved is to protect copyright of digital content owing to the ease of copying and manipulating multimedia data. Protection against digital content copyright violation is an important but difficult challenge. Digital content can be encrypted before it is transmitted to prevent attacks from unauthorized users. However, all content must eventually be decrypted before it will be intelligible to authorized users. Then the decrypted content may potentially be redistributed by some malicious authorized users. In order to personalize the copyrighted content assigned to the authorized user and rule out the unauthorized redistribution of digital content, digital fingerprinting was introduced to identify the authorized customers who redistribute their content for unintended purposes [19]. Its fundamental problem is to design efficient tracing algorithms based on fingerprinting codes which enable tracing to be carried out. The following are two typical types of copyright violation described in [83].

- Illegal key redistribution: Assuming that content is encrypted, there must be a mechanism for the content to be decrypted by an authorized user. The key used to decrypt the content may be copied and distributed to other users. Alternatively, these keys may be combined to create a new *pirate decoder*, which can subsequently be used to decrypt encrypted content illegally.

- Illegal content redistribution: Encrypted content is invariably decrypted once it gets to its authorized destination. Decrypted content can then be copied and transmitted to others, for example in an illegal *pirate broadcast*.

The main purpose of this dissertation is to investigate the anti-collusion key-distribution schemes for broadcast encryption and the anti-collusion codes for multimedia fingerprinting from the combinatorial viewpoint.

1.1 Broadcast encryption

Broadcast provides an efficient way to send information to many users simultaneously, that is one-to-many communication but not one-to-one communication. This is useful and profitable for the commercial broadcasting such as pay-TV, especially when a large number of long-lived users are designated. Naturally, the security problem of the broadcast network comes with it.

To protect the copyright of the broadcasted information, the distributor will first encrypt the data and then broadcast the encrypted information. Each authorized user, who paid for the copyright, can get a key from the distributor to decrypt the received encrypted information. Figure 1.1 provides a possible illustration of this situation. In the first stage, the distributor Bob may encrypt the data and upload to the public cloud. In the second stage, the authorized users Alice, Ted, Tim can download the encrypted file from the public platform and then decrypt it by using their assigned keys. If Alice, Ted and Tim are assigned the same key by Bob, then any one of them may pass his/her key to an unauthorized user without fear of being tracked down.

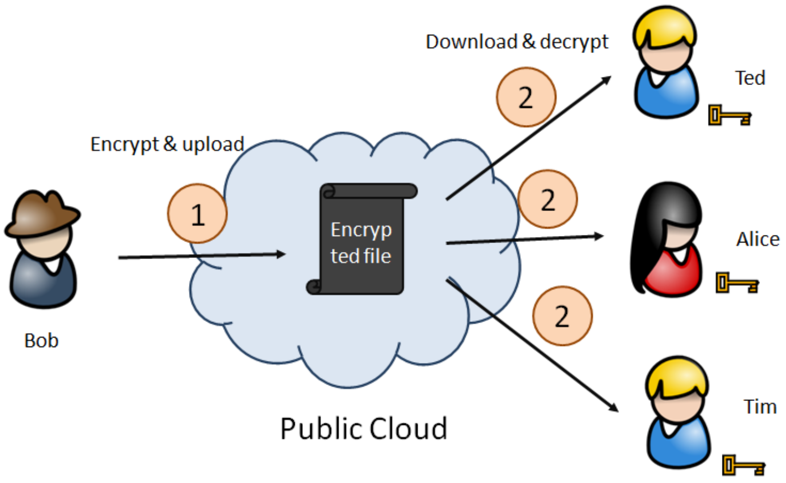

Figure 1.1: An illustration of broadcast encryption [94]

To hinder the illegal key redistribution, the distributor will personalize each authorized user by assigning a unique key as his/her *personal key*, which is essentially the same as that of digital fingerprinting in [19]. In this situation, the individual attack rarely occur. Since if one user, say Alice, illegally redistributes her personal key and once an illegal usage is captured, then Alice can be easily identified according to her personal key. As depicted in [83], the illegal key redistribution may mainly come from a collusion attack that some dishonest authorized users (*traitors*) work together to generate a pirate decoder and then redistribute it illegally. The key-distribution schemes for broadcast encryption are required to have some secure properties to resist the collusion attack. The following are two kinds of secure requirements which are widely investigated. Suppose that the number of traitors in the collusion does not exceed a predetermined threshold t.

- For any pirate decoder generated by the collusion attack, if it is confiscated, then at least one traitor will be traced back by virtue of some tracing algorithms.
- Any innocent user or group of a limited users cannot be framed by the collusion attack.

Here we remark that the traceability, in the static scenario, is only required to identify at least one traitor but not all traitors from a confiscated pirate. Since in a collusion, although a pirate is generated by several colluders, but it only contains partial information of each colluder. If one asks to trace back to all colluders from a pirate, then the desired security property is too strong and the corresponding schemes only can hold very few users.

In the literature, Fiat and Tassa [43] introduced the "dynamic traitor tracing" to identify all up to t traitors, where they assumed that the colluders do not construct a pirate decoder but decrypt the content and rebroadcast it [74]. The dynamic traitor tracing algorithms were also investigated by Berkman, Parnas and Sgall [12], Tassa [89], Roelse [71] and Laarhoven, Doumen, Roelse, Škorić and Weger [64]. Considering also in the dynamic scenario, Safavi-Naini and Wang [74] introduced the "sequential traitor tracing" to identify all up to t traitors, which improves the dynamic traitor tracing in the sense that it reduces computation by the broadcaster, and more importantly reduces the impact of delayed broadcasting by the pirate on the efficiency of the tracing process [14]. About the traceability in the dynamic scenario, the interested reader is referred to [12, 14, 43, 64, 71, 74, 89].

In this dissertation, we mainly focus on the key management schemes in the static scenario. In the following, we list four kinds of combinatorial structures for key-distribution schemes.

⧫ Traceability schemes

In 1994, Chor, Fiat and Naor introduced a traitor tracing scheme, the Chor-Fiat-Naor traceability scheme, applied to the broadcast encryption [25, 26]. To prevent unauthorized users from accessing the data, the data supplier encrypts the data blocks with session keys and gives the authorized users personal keys to decrypt them. Some unauthorized users (pirate users) might obtain some decryption keys from a group of traitors. Then the pirate users can decrypt data that they are not entitled to [85]. If a pirate decoder is confiscated, the Chor-Fiat-Naor traceability scheme can trace back to at least one traitor, by comparing the number of common base keys between the pirate decoder and each user's personal key, on the assumption that the number of traitors in the collusion does not exceed a predetermined threshold t.

In 1998, Stinson and Wei [85] generalized the Chor-Fiat-Naor traceability scheme to the Stinson-Wei traceability scheme. As stated in [85], in a broadcast encryption system, the data supplier generates a base set $\mathcal{X}$ of v keys and assigns w base keys to each authorized user, as the user's personal key. All authorized users can recover the session keys K, which are used to decrypt the data blocks, by using their personal keys. In the Chor-Fiat-Naor traceability scheme, the set $\mathcal{X}$ of base keys is partitioned into w subsets $\mathcal{S}_1, \ldots, \mathcal{S}_w$ (each of size v/w). Each personal key is a transversal of $(\mathcal{S}_1, \ldots, \mathcal{S}_w)$ (i.e., it contains exactly one base key from each $\mathcal{S}_i$). In this case, the pirate decoder generated by several traitors is also a transversal of $(\mathcal{S}_1, \ldots, \mathcal{S}_w)$, since otherwise the pirate decoder can not work. However, in the Stinson-Wei traceability scheme, each personal key is not necessarily a transversal. A personal key can be made up of any selection of w base keys from the set $\mathcal{X}$. The data supplier can use a w out of v threshold secret sharing scheme (such as the Shamir threshold scheme [75], for example) to construct v shares of the key K and then encrypt each share with a base key in $\mathcal{X}$. The pirate decoder can be made up of any w different base keys from the union of each traitor's personal key. If such a pirate decoder is captured and the size of the coalition does not exceed a predetermined threshold t, the Stinson-Wei traceability scheme also can reveal at least one traitor in the collusion by detecting the users who share the maximum base keys with the pirate decoder.

The Chor-Fiat-Naor traceability scheme is popular with the notion of "traceability code (TA code)". Upper bounds on the size of 2-TA codes and 3-TA codes were studied in [17] and [78], respectively. However, according to what we know, the upper bounds for general t-TA codes is still an open problem, which was conjectured in [17] by Blackburn, Etzion and Ng. The combinatorial properties for t-TA codes using error correcting codes were investigated in [42, 57, 82]. A construction for t-TA codes by virtue of constant weight codes was given in [65]. Infinite families of q-ary t-TA codes of constant rate bounded away from zero were proved to exist for small q in [17] and [59]. The Stinson-Wei traceability scheme has been studied as "traceability scheme (TS)". Upper bounds on the size of TS were studied in [29, 85], and constructions for TS were investigated in [68, 73, 85]. Objects related with traceability schemes, such as key distribution patterns, also have been studied by numerous researchers, see [81, 84, 93], for example.

⧫ Parent-identifying schemes

In 1998, Hollmann, van Lint, Linnartz and Tolhuizen proposed a digital fingerprinting scheme, based on codes with the identifiable parent property (IPP codes), to protect against piracy of software by embedding fingerprints into the copyrighted contents [53]. Given an IPP code, it is possible for every pirate copy (descendant) of digital content to identify at least one of its parents, that is, those authorized users each assigned with a fingerprint that contribute to the pirate copy, by computing the intersection of all groups of possible parents who can produce the pirate copy, again on the assumption that the number of parents in the collusion does not exceed a predetermined threshold t. IPP codes have been extensively studied in the literature, such as considering bounds on the maximum size of a t-IPP code [4, 6, 13], efficient traitor tracing algorithms for IPP codes [10], and relationships between IPP codes and separating hash families [2, 9, 77].

In 2009, Collins [29] suggested parent-identifying set systems (IPP set systems, or IPPSs) for broadcast encryption, which can be regarded as a generalization of IPP codes. The point of generalization from an IPP code to an IPPS is essentially the same as that from the Chor-Fiat-Naor traceability scheme to the Stinson-Wei traceability scheme, that is, instead of considering w-tuples, we consider w-subsets. Just as in the case of IPP codes, when a pirate copy is confiscated, the traitor tracing algorithm based on t-IPPS also needs to compute the intersection of all groups of possible parents with size at most t. Compared to the Stinson-Wei traceability scheme, the traitor tracing scheme based on IPPS can accommodate more users, but at the expenses of tracing efficiency.

⧫ Cover-free families

Cover-free families (CFFs) were introduced in 1964 by Kautz and Singleton [60] to study binary superimposed codes. Variants of this formulation have been investigated related to subjects such as information theory [60], group testing [55, 76], multiple access communication [41] and combinatorics [34, 35, 38, 46]. A t-CFF is a family of finite sets (blocks) in which no block is covered by the union of t others. From the viewpoint of broadcast system, a t-CFF is a kind of scheme in which any t traitors can not create another authorized user's personal key, which is closely related to a frameproof code used in digital copyright protection [19]. The relationship between frameproof codes and cover-free families was investigated in [82]. Frameproof codes were studied by numerous researchers, see [15, 20, 79, 85], for example.

In broadcast encryption, Chor, Fiat and Naor [25] studied the key-distribution schemes based on t-CFFs and mentioned that they can guarantee that any innocent user will not be framed by a coalition of size at most t. Although the CFFs may not have the traceability as that from TSs or IPPSs, but the key-distribution schemes based on CFFs can accommodate much more users than IPPSs.

⧫ Union-intersection-bounded families

As shown by Chor, Fiat and Naor [25], the schemes based on t-CFFs ensure that any innocent user can not be framed by a coalition of size at most t. However, t-CFFs cannot make sure that a group of $s \geq 2$ innocent users is not framed by any other coalition of size at most t. Inspired by this, we introduce a special kind of CFFs, called union-intersection-bounded families (UIBFs). The key-distribution schemes based on $(s, t; w)$-UIBFs can provide this desired security requirement. That is, the scheme based on $(s, t; w)$-UIBF guarantees that any group $\mathcal{G}$ of up to t traitors can not frame any other group $\mathcal{G}'$ of up to s users provided that $\mathcal{G} \cap \mathcal{G}' = \emptyset$. Obviously, the requirements of $(s, t; w)$-UIBFs are stronger than that of t-CFFs, but still weaker than that of t-IPPSs.

In summary, the security properties of schemes, based on the above four kinds of combinatorial structures, have a relationship as

$$t\text{-TSs} > \ t\text{-IPPSs} > (s, t; w)\text{-UIBFs} > t\text{-CFFs},$$

where "$>$" means "stronger". On the one hand, as can be seen, a t-TS has the strongest security property that for any t-collusion, at least one traitor can be traced back in time $O(M)$, where M is the number of users in this scheme. A t-CFF only guarantee that any innocent user cannot be framed by any t-collusion, but can not provide the ability of tracing back to traitors. On the other hand, the scheme based on a t-CFF can accommodate the most users among them, and a t-TS has a capacity much less than a t-CFF. Therefore, to find a tradeoff between the security property and the capacity of users for a key-distribution scheme is an interesting and meaningful problem. In some sense, t-IPPSs and $(s, t; w)$-UIBFs are two structures on the way of finding a balance between t-TSs and t-CFFs.

This dissertation is devoted to investigate these four kinds of schemes from the combinatorial viewpoint. We aim to determine the maximum capacity of each scheme, and to construct the schemes which could accommodate the maximum numbers of users.

1.2 Multimedia fingerprinting

In Section 1.1, the anti-collusion key-distribution schemes in broadcast encryption were required to resist the illegal key redistribution. In this section, we discuss the anti-collusion codes for multimedia fingerprinting, which are required to resist the illegal content redistribution.

Multimedia fingerprinting is a technology of protecting the copyright of multimedia content, such as audio, video, images, text, and various other modalities. Fingerprints for multimedia data can be embedded through a variety of watermarking techniques prior to their authorized distribution, among which spread-spectrum additive embedding is a widely employed robust embedding technique [27, 70]. As depicted in [24], in spread-spectrum embedding, a watermark signal, often represented by noise-like orthonormal basis signals, is added to the host signal. Usually, all signals are regarded as vectors in some signal spaces. Let $\mathbf{x}$ be the host multimedia signal, $\{\mathbf{e}_i : 1 \leq i \leq n\}$ be an orthonormal basis of noise-like signals, and let

$$\{\mathbf{f}_j = (\mathbf{f}_j(1), \mathbf{f}_j(2), \ldots, \mathbf{f}_j(n)) = \sum_{i=1}^{n} b_{ij}\mathbf{e}_i \; : \; 1 \leq j \leq M\}, \quad b_{ij} \in \{0,1\}$$

be a family of scaled watermarks to achieve the imperceptibility as well as to control the energy of the embedded watermark. The watermarked version of the content $\mathbf{y}_j = \mathbf{x}+\mathbf{f}_j$, $1 \leq j \leq M$, is then assigned to the authorized user U_j who has purchased the rights to access $\mathbf{x}$. The fingerprint $\mathbf{f}_j$ assigned to U_j can be represented uniquely by a vector $\mathbf{b}_j = (b_{1j}, b_{2j}, \ldots, b_{nj})^T \in \{0,1\}^n$ according to the linear independence of the basis $\{\mathbf{e}_i : 1 \leq i \leq n\}$. Clearly, an individual authorized user will not redistribute his/her received content without running the risk of being tracked down. However, several authorized users may collude to generate a pirate copy by virtue of some collusion attack models so that the venture traced by the pirate copy can be attenuated. For more details about the collusion attack models, the interested reader is referred to [67].

A *coalition* is a group $U = \{U_{j_1}, U_{j_2}, \ldots, U_{j_t}\}$ of authorized users who intend to work together to generate a pirate copy, where each authorized user U_{j_k} is dishonest and called a *traitor*. By the above assumptions, each authorized user U_{j_k}, $1 \leq k \leq t$, is assigned the content $\mathbf{y}_{j_k} = \mathbf{x}+\mathbf{f}_{j_k}$, where $\mathbf{f}_{j_k} = \sum_{i=1}^{n} b_{ij_k}\mathbf{e}_i$ and $b_{ij_k} \in \{0,1\}$. As in [24], we assume that they have no way of manipulating the individual orthonormal signals, that is, the underlying codeword needs to be taken and proceeded as a single entity. Normally, no colluder is willing to take more of a risk than any other colluder. As claimed in [67, 91], the averaging attack is the most fair choice for each colluder to take the risk of being identified. This attack also makes the pirate copy have better perceptional quality. Accordingly, a pirate copy $\mathbf{y}$ produced by an averaging attack of U is

$$\mathbf{y} = \sum_{k=1}^{t} \frac{1}{t}\mathbf{y}_{j_k} = \mathbf{x} + \frac{1}{t}\sum_{k=1}^{t} \mathbf{f}_{j_k} = \mathbf{x} + \sum_{k=1}^{t}\sum_{i=1}^{n} \frac{b_{ij_k}}{t}\mathbf{e}_i. \tag{1.1}$$

The pirate copy $\mathbf{y}$ is also called a *descendant* of U, and each user in U is called a *parent* of $\mathbf{y}$. One important problem in multimedia fingerprinting is how to detect colluders in U once a pirate copy $\mathbf{y}$ is confiscated.

In colluder detection phases, the distributor can extract the mixed fingerprints information from $\mathbf{y}$ by computing the correlation vector $\mathbf{T} = (\mathbf{T}(1), \mathbf{T}(2), \ldots, \mathbf{T}(n))^T$, where $\mathbf{T}(i) = \langle \mathbf{y}-\mathbf{x}, \mathbf{e}_i \rangle$, $1 \leq i \leq n$, and $\langle \mathbf{y}-\mathbf{x}, \mathbf{e}_i \rangle$ is the inner product of $\mathbf{y}-\mathbf{x}$ and $\mathbf{e}_i$. Obviously, $0 \leq \mathbf{T}(i) \leq 1$ for any $1 \leq i \leq n$. Now the traitor tracing problem is how to identify the colluders from the detection statistics of their mixed fingerprints information. The interested reader is referred to [24, 67] for more details.

A multimedia t-fingerprinting code is a set of authorized fingerprints, each of which is assigned to an authorized user, such that no matter which coalition of at most t traitors collude to produce a pirate copy, the distributor is always capable of tracing back to at least one or all colluders. Thus the main problems in multimedia fingerprinting are 1) to design a t-fingerprinting code with as large code size as possible, which corresponds to accommodate as many users as possible; and 2) to create as efficient as possible traitor tracing algorithms for the t-fingerprinting code to identify at least one or all colluders from a captured pirate copy.

In the following, we briefly recap two kinds of anti-collusion codes for multimedia fingerprinting.

⧫ Separable codes

Separable codes were introduced by Cheng and Miao in [24] and further studied in [16, 23, 21, 47] and so on. In the linear attack model, a $\bar{t}$-separable code guarantees that any two distinct groups, where each group consists of up to t users, can not generate the same pirate copy. Thus, once a pirate copy generated by at most t colluders is confiscated, the coalition set can be identified by checking all groups of up to t users. Obviously, the traitor tracing algorithm is not efficient, since its computational complexity is $O(M^t)$, where M is the number of all authorized users. Furthermore, Cheng and Miao [24] found that the binary frameproof codes, introduced in [19] for digital fingerprinting, can provide more efficient traitor tracing algorithm for multimedia contents with the computational complexity $O(M)$, whereas the frameproof codes do not have traceability in digital fingerprinting. However, the frameproof codes have a smaller capacity than separable codes. In [56], Jiang *et al.* introduced the concept of strongly separable codes, which is a special kind of separable codes. They showed that the strongly separable codes have the traitor tracing algorithm to identify all colluders with computational complexity $O(M)$, and also a larger capacity than frameproof codes. To some extent, the strongly separable codes lie between frameproof codes and separable codes.

⧫ Multimedia parent-identifying codes

Parent-identifying codes were introduced by Hollmann *et al.* [53] for digital fingerprinting to guarantee that at least one colluder can be traced back by virtue of some traitor tracing algorithms. As an analogue of that, Cheng *et al.* [22] proposed codes with the identifiable parent property for multimedia fingerprinting, which are also referred to as the multimedia parent-identifying codes (MIPPCs). They showed that, in a linear attack, t-MIPPCs ensure that at least one colluder can be identified by a tracing algorithm with computational complexity $O(M^t)$ by computing the intersection of all groups of up to t possible parents who can produce the pirate copy. Compared to separable codes, although MIPPCs can only identify at least one colluder, Cheng *et al.* [22] conjectured MIPPCs can have much more codewords than separable codes, and they proved this for several small parameters.

This dissertation is also devoted to investigate the maximum number of codewords in the multimedia anti-collusion codes.

1.3 Previous results

Among the known results on t-TS, Stinson and Wei [85] proved that a t-TS is a t-CFF, and derived an upper bound for the number of blocks in a t-TS by using this relationship. There is a huge gap between this upper bound and the lower bound determined by the size of t-TS constructed by using combinatorial structures in [85]. Collins [29] improved the upper bound for t-TS, and gave upper bounds for t-IPPS. Unfortunately, there is no construction which can achieve any of these known upper bounds. As a matter of fact, the known upper bounds for t-TS and t-IPPS are not tight.

Cheng *et al.* [22] derived upper bounds for MIPPCs by transferring the requirement of a t-MIPPC to a corresponding bipartite graph without cycles of length less than or equal to $2t$. They also gave some constructions for 3-MIPPCs of length 2 by means of generalized quadrangles. However, there are no general lower bounds or constructions for t-MIPPCs.

1.4 Contributions

This dissertation is devoted to study the anti-collusion schemes for broadcast encryption and the anti-collusion codes for multimedia fingerprinting by virtue of combinatorial methods, such as techniques in extremal set theory, extremal graph theory, probabilistic methods and combinatorial design theory. We derive new bounds for the anti-collusion combinatorial structures and also provide constructions to achieve our new bounds, which greatly improve the previously known results and further promote the development of broadcast encryption and multimedia fingerprinting. To be specific, we state our results in the following aspects.

As far as we know, in the literature, the relationship between TS and CFF has been studied for the same strength (i.e. a t-TS is a t-CFF), and this is also almost true for other relationships among various anti-collusion schemes. In this dissertation, we find a very interesting phenomenon, that is, a t-TS is in fact a t^2-CFF. This is the first relationship between two kinds of anti-collusion schemes which strengthens the strength from t to t^2. Based on this important discovery, new upper bounds for t-TS are derived. To obtain our new bounds, we use a combinatorial structure called *own-subset* by Erdős, Frankl and Füredi in [39]. Our new bounds show that some infinite families of t-TSs constructed by Stinson and

Wei [85] from combinatorial designs are in fact optimal. We generalize Stinson and Wei's construction to obtain more infinite families of optimal t-TS. We also describe a constructive lower bound for general t-TS, the size of which has the same order of magnitude as the new upper bounds. These results are exhibited in Chapter 4.

Collins [29] showed that the upper bound for t-IPPS(w, v) is $O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil})$. We give an improvement for this by showing that the upper bound for t-IPPS(w, v) is $O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil})$, which is realized by analyzing the minimum size of own-subsets possessed by some blocks in a t-IPPS. Next, by using the expurgation method, we prove that for fixed t, w and sufficiently large v, there exists a t-IPPS(w, v) with size $O(v^{\frac{w}{\lfloor t^2/4 \rfloor + t}})$, which has the same order of magnitude with the new upper bound when $\lfloor t^2/4 \rfloor + t$ is a divisor of w. We also study 2-IPPS(w, v) for small w. Better upper bounds for 2-IPPS$(4, v)$ and 3-IPPS$(6, v)$ are respectively derived by means of the well-known graph removal lemma in extremal graph theory. These results are presented in Chapter 5.

We introduce a special kind of CFF, called UIBF, due to its applications in broadcast encryption. The relationships among UIBF, TS, IPPS and CFF are investigated in Chapter 3. The upper bounds and lower bounds for UIBF are provided in Chapter 6, which have the same order of magnitude for certain cases.

The probabilistic existence result for MIPPCs is presented in Chapter 7, which is the first lower bound for general MIPPCs. We also compare the new lower bound with the known upper bound for MIPPCs, showing that the gap between them is very small.

Partial results of this dissertation have been preprinted for publication in [50, 51, 52].

1.5 Outline of this dissertation

In Chapter 2, we recap some basic notations and definitions. The definitions of TS, IPPS, UIBF, CFF, and their relationships are presented in Chapter 3. In Chapter 4, we provide the new upper bounds, new lower bounds and several constructions for TS. New bounds for IPPS and UIBF are exhibited in Chapter 5 and Chapter 6, respectively. In Chapter 7, we show the probabilistic existence result for MIPPC. Finally, we conclude this dissertation and list several related open problems in Chapter 8.

Chapter 2

Preliminary

In this chapter, we state some mathematical notations and concepts which will be used in this dissertation.

2.1 Set systems

Let v and w be positive integers such that $v \geq w$. Suppose $\mathcal{X}$ is a finite set of size v, then denote $2^{\mathcal{X}}$ as the *power set* of $\mathcal{X}$ which consists of all subsets of $\mathcal{X}$, and denote $\binom{\mathcal{X}}{w}$ as the collection of all w-subsets of $\mathcal{X}$, that is,

$$\binom{\mathcal{X}}{w} = \{B \in 2^{\mathcal{X}} : \ |B| = w\}.$$

Definition 2.1.1 *A set system is a pair $(\mathcal{X}, \mathcal{B})$, where $\mathcal{B} \subseteq 2^{\mathcal{X}}$. The cardinality of $\mathcal{B}$ is called the size of the set system $(\mathcal{X}, \mathcal{B})$. The elements of $\mathcal{X}$ are called points and the members of $\mathcal{B}$ are called blocks.*

Example 2.1.2 *Suppose $\mathcal{X} = \{1, 2, 3, 4\}$ and $\mathcal{B} = \{\{1\}, \{2\}, \{1, 2\}, \{2, 3, 4\}\}$. Then $(\mathcal{X}, \mathcal{B})$ is a set system with size $|\mathcal{B}| = 4$.*

Definition 2.1.3 *A set system $(\mathcal{X}, \mathcal{B})$ is w-uniform if $\mathcal{B} \subseteq \binom{\mathcal{X}}{w}$. A set system $(\mathcal{X}, \mathcal{B})$ is called uniform if it is w-uniform for some positive integer $w \leq v$.*

A w-uniform set system $(\mathcal{X}, \mathcal{B})$ is also denoted as a (w, v) *set system*. The following notion of sunflowers, or Δ-systems, was introduced by Erdős and Rado [40] in 1960.

Definition 2.1.4 *Let $(\mathcal{X}, \mathcal{B})$ be a set system. Then $(\mathcal{X}, \mathcal{B})$ is a sunflower or Δ-system if there exists a subset (core) $\Delta \subseteq \mathcal{X}$ such that*

(1) for any two distinct blocks $A, B \in \mathcal{B}$, we have $A \cap B = \Delta$;

(2) for any block $B \in \mathcal{B}$, the petal $B \setminus \Delta$ is nonempty.

Note that a family of pairwise disjoint sets is a sunflower (with an empty core).

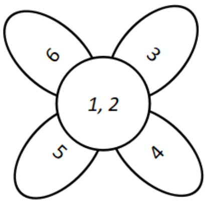

Figure 2.1: A depiction of sunflower $\mathcal{B}$ in Example 2.1.5

Example 2.1.5 *Suppose $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{B} = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 2, 6\}\}$. Let $\Delta = \{1, 2\} \subseteq \mathcal{X}$. Clearly, the intersection of any two blocks in $\mathcal{B}$ is Δ, and $B \backslash \Delta$ is nonempty for any $B \in \mathcal{B}$. Thus $\mathcal{B}$ is a sunflower. A direct depiction of this sunflower is as Figure 2.1.*

A notion, the *own-subset*, is extremely useful in deriving bounds for the size of set systems t-CFF in [39]. It is also an important tool we will exploit to investigate the combinatorial schemes in broadcast encryption.

Definition 2.1.6 *Let $(\mathcal{X}, \mathcal{B})$ be a set system and $B \in \mathcal{B}$ be a block. Then a subset $B_0 \subseteq B$ is called a $|B_0|$-own-subset of B if $B_0 \nsubseteq B'$ for any $B' \in \mathcal{B} \setminus \{B\}$.*

In the above Example 2.1.5, we can see that $\{3\}$ is a 1-own-subset of $\{1, 2, 3\}$, $\{1, 6\}$ is a 2-own-subset of $\{1, 2, 6\}$, and $\{2, 4, 5\}$ is a 3-own-subset of $\{1, 2, 4, 5\}$. But $\{1, 2\}$ is not a 2-own-subset of any block in $\mathcal{B}$.

2.2 Graphs

Definition 2.2.1 *A graph G is a pair (V, E) comprising a set V of vertices together with a set E of edges, where $E \subseteq \binom{V}{2}$.*

A graph defined by Definition 2.2.1 is usually described as undirected and simple. As can be seen, a graph $G = (V, E)$ corresponds to a 2-uniform set system $(\mathcal{X}, \mathcal{B})$, where $\mathcal{X} = V$ and $\mathcal{B} = E$. Generally, set systems are also investigated in the disguise of hypergraphs.

Definition 2.2.2 *A hypergraph $\mathcal{H}$ is a pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of elements called vertices and $\mathcal{E}$ is a set of non-empty subsets of $\mathcal{V}$ called edges or hyperedges. A hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is called k-uniform if for any $E \in \mathcal{E}$, $|E| = k$.*

A 2-uniform hypergraph is a *graph.* A (k-uniform) set system $(\mathcal{X}, \mathcal{B})$ such that $\emptyset \notin \mathcal{B}$ corresponds to a (k-uniform) hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{X}$ and $\mathcal{E} = \mathcal{B}$.

Definition 2.2.3 *A graph $G = (V, E)$ is complete if $E = \binom{V}{2}$. A complete graph on n vertices is denoted by K_n.*

Definition 2.2.4 *Let $G = (V, E)$ be a graph. A graph $G' = (V', E')$ is a subgraph of $G = (V, E)$ if $V' \subseteq V$, $E' \subseteq E$ and $E' \subseteq \binom{V'}{2}$.*

Definition 2.2.5 *A path is a graph $P = (V, E)$ of the form*

$$V = \{v_1, v_2, \dots, v_l\}, \quad E = \{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{l-1}, v_l\}\},$$

where v_i are all distinct. The cardinality of E is called the length of the path P.

Definition 2.2.6 *Adding an edge $\{v_l, v_1\}$ to a path $P = (V, E)$ forms a cycle. The number of edges in a cycle is called the length of the cycle. The length of a shortest cycle contained in a graph is called the girth of the graph.*

Example 2.2.7 *Figure 2.2 depicts a graph $G = (V, E)$, where $V = \{1, 2, 3, 4, 5\}$ and $E = \{\{1, 2\}, \{1, 4\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\}$. A subgraph $P = (\{1, 2, 4\}, \{\{1, 2\}, \{1, 4\}\})$ of G is a path of length 2. A subgraph $K = (\{3, 4, 5\}, \{\{3, 4\}, \{3, 5\}, \{4, 5\}\})$ of G is a cycle of length 3 and also a complete graph on $\{3, 4, 5\}$. The girth of graph G is 3.*

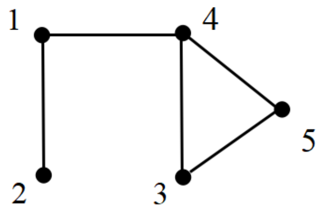

Figure 2.2: An example of graph in Example 2.2.7

Lemma 2.2.8 ([32]) *If a graph G on an n-vertex set contains more than $n - 1$ edges, then G contains a cycle.*

A well-known result in extremal graph theory, the graph removal lemma, was first stated by Alon, Duke, Lefmann, Rödl and Yuster in [3] and Füredi in [45], which is a generalization of the triangle removal lemma proved by Ruzsa and Szemerédi in [88]. A recent survey of graph removal lemma refers to [30].

Lemma 2.2.9 ([3, 30, 45]) *For any graph H and any $\varepsilon > 0$, there exists $\delta > 0$ such that any graph on n vertices which contains at most $\delta n^{v(H)}$ copies of H may be made H-free by removing at most εn^2 edges, where $v(H)$ is the number of vertices in graph H.*

2.3 Combinatorial designs

Definition 2.3.1 *A t-(v, k, λ) design, a t-design in short, is a k-uniform set system $(\mathcal{X}, \mathcal{B})$ with the property that every t-subset of $\mathcal{X}$ is contained in exactly λ blocks.*

Example 2.3.2 *A 2-$(7, 3, 1)$ design $(\mathcal{X}, \mathcal{B})$, also known as Fano plane, is shown in Figure 2.3, where $\mathcal{X} = \{1, 2, \ldots, 7\}$ and $\mathcal{B} = \{\{1, 2, 3\}, \{1, 4, 7\}, \{1, 5, 6\}, \{2, 4, 6\}, \{2, 5, 7\}, \{3, 4, 5\}, \{3, 6, 7\}\}$.*

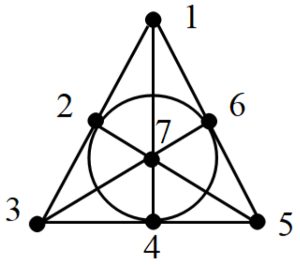

Figure 2.3: The Fano plane

Definition 2.3.3 *A t-(v, k, λ) packing, a t-packing in short, is a k-uniform set system $(\mathcal{X}, \mathcal{B})$ with the property that every t-subset of $\mathcal{X}$ is contained in at most λ blocks.*

In Example 2.3.2, one can get a 2-$(7, 3, 1)$ packing by deleting some blocks from $\mathcal{B}$.

Proposition 2.3.4 *Let $(\mathcal{X}, \mathcal{B})$ be a t-(v, k, λ) design. Then*

$$|\mathcal{B}| = \frac{\lambda\binom{v}{t}}{\binom{k}{t}}.$$

Corollary 2.3.5 *Let $(\mathcal{X}, \mathcal{B})$ be a t-(v, k, λ) packing. Then*

$$|\mathcal{B}| \leq \frac{\lambda\binom{v}{t}}{\binom{k}{t}}.$$

The following are two notable results on designs and packings.

In 2014, Keevash [62] proved the existence conjecture of τ-design by using the method of *randomised algebraic constructions*. More recently, Glock, Kühn, Lo and Osthus [49] provided a new proof for that via *iterative absorption*. The case $\tau = 2$ was also proved by Wilson four decades ago [28].

Theorem 2.3.6 ([62, 49]) *Given w, τ and λ, there exists a $v_0(w, \tau, \lambda)$ such that for any $v > v_0(w, \tau, \lambda)$, a τ-(v, w, λ) design exists if and only if for any $0 \leq i \leq \tau - 1$,*

$$\lambda\binom{v-i}{\tau-i} \equiv 0 \pmod{\binom{w-i}{\tau-i}}.$$

The asymptotic existence of τ-$(v, w, 1)$ packing was proved by Rödl by using the probabilistic method known as the *Rödl nibble* [72].

Theorem 2.3.7 ([72]) *Given w and τ, there exists a $v_0(w, \tau)$ such that for any $v > v_0(w, \tau)$, a τ-$(v, w, 1)$ packing $(\mathcal{X}, \mathcal{B})$ exists with size $|\mathcal{B}| = (1 - o(1))\binom{v}{\tau}/\binom{w}{\tau}$.*

2.4 Coding theory

Let n, q be positive integers and Q be a finite set with size q. Usually, $Q = \{0, 1, \dots, q-1\}$.

Definition 2.4.1 *A set $\mathcal{C} \subseteq Q^n$ is called a q-ary code of length n, denoted (n, q) code. Each $\mathbf{c} \in \mathcal{C}$ is called a codeword. The cardinality of $\mathcal{C}$ is called the code size.*

If $q = 2$, we also refer to $\mathcal{C}$ as a *binary* code.

Definition 2.4.2 *The Hamming distance between two codewords $\mathbf{u} = (\mathbf{u}(1), \dots, \mathbf{u}(n))$ and $\mathbf{v} = (\mathbf{v}(1), \dots, \mathbf{v}(n))$ is*

$$d(\mathbf{u}, \mathbf{v}) = |\{1 \le i \le n : \; \mathbf{u}(i) \neq \mathbf{v}(i)\}|.$$

Definition 2.4.3 *The Hamming weight of a codeword $\mathbf{c} = \{\mathbf{c}(1), \dots, \mathbf{c}(n)\} \in Q^n$ is*

$$wt(\mathbf{c}) = d(\mathbf{c}, \mathbf{0}) = |\{1 \le i \le n : \; \mathbf{c}(i) \neq 0\}|,$$

where $\mathbf{0} = (0, \dots, 0)$. A code $\mathcal{C}$ is called a constant weight code if there exists a constant $0 \le w \le n$ such that for any codeword $\mathbf{c} \in \mathcal{C}$, $wt(\mathbf{c}) = w$.

Example 2.4.4 *Let $Q = \{0, 1\}$ and $\mathcal{C} = \{1100, 1010, 1001, 0110, 0101, 0011\}$. Then $\mathcal{C}$ is a binary code of length 4 and size 6. It is also a constant weight code, since for any $\mathbf{c} \in \mathcal{C}$, we have $wt(\mathbf{c}) = 2$.*

An important parameter in coding theory is the *rate* of a code, which is defined as follows.

Definition 2.4.5 *The code rate of an (n, q) code $\mathcal{C}$ is*

$$R(\mathcal{C}) = \frac{\log_q |\mathcal{C}|}{n}.$$

The asymptotic code rate of $\mathcal{C}$ is often investigated for a certain class of codes. That is,

$$\lim_{n \to \infty} R(\mathcal{C}) = \lim_{n \to \infty} \frac{\log_q |\mathcal{C}|}{n}$$

or

$$\lim_{q \to \infty} R(\mathcal{C}) = \lim_{q \to \infty} \frac{\log_q |\mathcal{C}|}{n}.$$

2.5 Probabilistic methods

We first state some basic concepts in probability theory. Here and hereafter we consider Ω to be a discrete set.

Definition 2.5.1 *The sample space is a set Ω containing all possible outcomes.*

Definition 2.5.2 *An event is a subset of the sample space Ω. The event space $\mathcal{F}$ is the collection of all events in the sample space Ω.*

Definition 2.5.3 *A probability space is a three-tuple $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is the event space and P is a function from events to probabilities.*

Definition 2.5.4 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A discrete random variable $X : \Omega \to \mathbb{R}$ is a real-valued function from Ω to the measurable space $\mathbb{R}$.*

Denote R_X as the set of all possible values of the random variable X.

Definition 2.5.5 *Let X be a discrete random variable. The probability mass function of X is a function*

$$p(x) = P(X = x), \quad \forall\, x \in R_X,$$

such that

(1) $p(x) \geq 0, \ \forall\, x \in R_X$;

(2) $\sum_{x \in R_X} p(x) = 1$;

(3) $P(X \in A) = \sum_{x \in A} p(x), \ \forall\, A \subseteq R_X$.

Example 2.5.6 *When a single 6-sided die is tossed, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. The probability of each outcome is $P(i) = 1/6$, $\forall\, 1 \leq i \leq 6$.*

A discrete random variable $X : \Omega \to \mathbb{R}$ is defined as $X(i) = \lfloor i/2 \rfloor$, $\forall\, i \in \Omega$. The set of all possible values of X is $R_X = \{0, 1, 2, 3\}$. The probability mass function $p(x)$ of X is

$$\begin{aligned} p(0) &= P(X = 0) = P(1) = \frac{1}{6}, \\ p(1) &= P(X = 1) = P(2) + P(3) = \frac{1}{3}, \\ p(2) &= P(X = 2) = P(4) + P(5) = \frac{1}{3}, \\ p(3) &= P(X = 3) = P(6) = \frac{1}{6}. \end{aligned}$$

Definition 2.5.7 *Let X be a discrete random variable and $p(x)$ be the probability mass function of X. The expected value (or expectation) of X is*

$$E(X) = \sum_{x \in R_X} xp(x).$$

In Example 2.5.6, we have

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) = 0 + \frac{1}{3} + \frac{2}{3} + \frac{3}{6} = \frac{3}{2}.$$

The linearity of expectations is stated as follows.

Proposition 2.5.8 ([5]) *Let $X_1, \ldots, X_n$ be random variables, $X = c_1X_1 + \cdots + c_nX_n$. Linearity of expectation states that*

$$E(X) = c_1E(X_1) + \cdots + c_nE(X_n).$$

As depicted in [58], the power of this principle comes from there being no restrictions on the dependence or independence of the X_i's. In most arguments, we often use the following fact, known as the *Pigeonhole principle.*

Proposition 2.5.9 *There must be a point in the probability space for which $X \geq E(X)$ and a point for which $X \leq E(X)$.*

Probabilistic methods provide tools to argue that some structures with certain properties do exist. By the basic probabilistic methods, one defines an appropriate probability space of structures and then proves that the desired properties hold in this space with positive probability [5].

However, the *expurgation method* can work beyond this. Expurgation method, also called the alteration method or deletion method, deals with the situation that the random structure does not have all the desired properties but may have a few “blemishes”. And the desired structure can be shown to exist after removing the blemishes [5, 58].

2.6 Secret sharing schemes

Secret sharing schemes are important tools in cryptography. A secret-sharing scheme refers to a method by which a dealer distributes a secret amongst a group of players, each of whom is allocated a share of the secret, such that

(1) any subset in $\mathcal{A}$ can reconstruct the secret from its shares;

(2) any subset not in $\mathcal{A}$ can not reveal any partial information on the secret,

where $\mathcal{A}$ is a collection of subsets of shares, called the access structure. Usually, individual shares are of no use on their own.

Shamir [75] and Blakley [18] introduced the threshold secret sharing schemes, that is, the secret can be reconstructed only when a sufficient number of shares are combined together.

Definition 2.6.1 *Let n and k be positive integers such that $n \geq k$. A (k, n)-threshold scheme, also called a k out of n threshold scheme, is a method of sharing a secret amongst a set of n parties in such a way that*

(1) any k or more shares (parties) can reconstruct the secret;

(2) any $k-1$ or fewer shares (parties) can not reconstruct the secret.

The essential idea of the Shamir threshold scheme is that k points are sufficient to define a polynomial of degree $k-1$.

Example 2.6.2 *Suppose the dealer would like to use a (k, n)-threshold secret sharing scheme to share a secret S. First, the secret is assumed to be an element a_0 in a finite field $\mathbb{F}_q$, where q is a prime power such that $q > n$. The dealer chooses $k-1$ random elements $a_1, \ldots, a_{k-1} \in \mathbb{F}_q$ independently with uniform distribution, and builds a polynomial*

$$f(x) = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1}.$$

Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$ be n distinct nonzero elements. Each party i is assigned a point $(\alpha_i, f(\alpha_i))$, which is also called a share. Then any k shares can compute the constant term a_0 by using the polynomial interpolation.

For more details on the secret sharing schemes, the interested reader is referred to a survey [11] and the references therein.

2.7 Group testing

The field of group testing was introduced by Dorfman [33] in 1943. The motivation arose during the Second World War when the United States public health service and the selective service embarked upon a large scale project to weed out all syphilitic men called up for induction. However, at that time, testing every soldier individually was expensive and inefficient. Group testing provides an efficient way:

(a) pool soldiers into groups, and combine blood samples in each group together;

(b) test the combined sample to check if at least one soldier in the group has syphilis.

Accordingly, the two important problems in group testing are how to design the experiment groups and how to identify the syphilitic soldiers from the testing results. Recently, group testing is also found applications in molecular biology, drug discovery, information security and so on.

In the literature, many combinatorial structures were proposed for group testing. We list some of them as follows.

Definition 2.7.1 *Let n be a positive integer. A family $\mathcal{F} \subseteq 2^{[n]}$ is t-cover-free, if for any $t+1$ distinct blocks $A_0, A_1, \ldots, A_t \in \mathcal{F}$, we have*

$$A_0 \nsubseteq \bigcup_{1 \le i \le t} A_i.$$

The incidence matrix of t-cover-free family is also called *t-disjunct matrix.*

Definition 2.7.2 *Let n be a positive integer. A family $\mathcal{F} \subseteq 2^{[n]}$ is t-superimposed if for any distinct subfamilies $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ such that $\mathcal{F}_1 \neq \mathcal{F}_2$ and $1 \le |\mathcal{F}_i| \le t$, $i = 1, 2$, we have*

$$\bigcup_{A \in \mathcal{F}_1} A \neq \bigcup_{A \in \mathcal{F}_2} A$$

The incidence matrix of t-superimposed family is also called *$\bar{t}$-separable matrix.*

Definition 2.7.3 *Let n be a positive integer. A family $\mathcal{F} \subseteq 2^{[n]}$ is t-single-user tracing superimposed family, if for all choices of $\mathcal{F}_1, \ldots, \mathcal{F}_k$ with $1 \le |\mathcal{F}_i| \le t$,*

$$\bigcup_{A \in \mathcal{F}_1} A = \bigcup_{A \in \mathcal{F}_2} A = \cdots = \bigcup_{A \in \mathcal{F}_k} A$$

implies

$$\bigcap_{1 \le i \le k} \mathcal{F}_i \neq \emptyset.$$

Chapter 3

Combinatorial Schemes for Broadcast Encryption

In this chapter, we first present the combinatorial definitions of TS, IPPS, UIBF, CFF, and their corresponding applications in broadcast encryption. Their relationships are also described in Section 3.2.

3.1 Definitions

First, we state the definitions of TS, IPPS, UIBF and CFF from the combinatorial viewpoint as follows.

Definition 3.1.1 *Suppose $(\mathcal{X}, \mathcal{B})$ is a (w, v) set system with size $|\mathcal{B}| = M$. Let s, t, d be positive integers such that $s \leq t$ and $d \leq w$. Then*

(1) *$(\mathcal{X}, \mathcal{B})$ is a t-traceability scheme, denoted t-TS(w, M, v), provided that for every choice of $s \leq t$ blocks $B_1, B_2, \ldots, B_s \in \mathcal{B}$ and for any w-subset $T \subseteq \bigcup_{1 \leq j \leq s} B_j$, there does not exist a block $B \in \mathcal{B} \backslash \{B_1, B_2, \ldots, B_s\}$ such that $|T \cap B| \geq |T \cap B_j|$ for all $1 \leq j \leq s$.*

(2) *$(\mathcal{X}, \mathcal{B})$ is a t-parent-identifying set system, denoted t-IPPS(w, M, v), provided that for any w-subset $T \subseteq \mathcal{X}$, either $P_t(T)$ is empty, or*

$$\bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P} \neq \emptyset,$$

where

$$P_t(T) := \{\mathcal{P} \subseteq \mathcal{B} : |\mathcal{P}| \leq t, \ T \subseteq \bigcup_{B \in \mathcal{P}} B\}.$$

(3) *$(\mathcal{X}, \mathcal{B})$ is an $(s, t; d)$-union-intersection-bounded family, denoted $(s, t; d)$-UIBF(w, M, v), provided that for any $s + t$ distinct blocks $A_1, \ldots, A_s, B_1, \ldots, B_t \in \mathcal{B}$, we have*

$$|(\bigcup_{1 \leq i \leq s} A_i) \cap (\bigcup_{1 \leq j \leq t} B_j)| < d.$$

(4) *$(\mathcal{X},\mathcal{B})$ is a t-cover-free family, denoted t-CFF(w, M, v), provided that for any $t+1$ distinct blocks $B_0, B_1, \ldots, B_t \in \mathcal{B}$, we have*

$$B_0 \not\subseteq \bigcup_{1\le i\le t} B_i.$$

Example 3.1.2 *Let $\mathcal{X} = \{1,2,3,4,5\}$ and $\mathcal{B} = \{\{1,2,3\},\{1,2,4\},\{1,2,5\}\}$. By Definition 2.1.4, $(\mathcal{X},\mathcal{B})$ is a sunflower. It is easy to check that $(\mathcal{X},\mathcal{B})$ is a 2-TS$(3,5)$, a 2-IPPS$(3,5)$, a $(2,2;3)$-UIBF$(3,5)$ and a 2-CFF$(3,5)$.*

In the setting of applications of Definition 3.1.1 (1)–(4), $\mathcal{X}$ corresponds to the set of base keys possessed by the data supplier, and each block $B \in \mathcal{B}$ corresponds to a personal key issued to an authorized user. Accordingly, the cardinality of $\mathcal{B}$, i.e. the parameter M, is the number of users which the scheme can accommodate. A set of traitors holding $\mathcal{B}_0 \subseteq \mathcal{B}$, where $|\mathcal{B}_0| \le t$, as their personal keys can combine their personal keys to produce a pirate personal key, that is, a w-subset $T \subseteq \bigcup_{B\in\mathcal{B}_0} B$, to decode w shares of the key K, then use the w out of v threshold secret sharing scheme to recover K, and finally use K to decrypt the data block. The subset T is called a *descendant* generated by $\mathcal{B}_0$, while $\mathcal{B}_0$ is referred to as a *parent set* of T, and each $B \in \mathcal{B}_0$ is called a *parent* of T. Note that the parent sets are not in general unique [14]. Explicitly, the notation $P_t(T)$ in the above definition is the set of all parent sets of T with size at most t. We also refer to

$$\mathrm{Desc}_t(\mathcal{B}) := \{T \subseteq \mathcal{X} : P_t(T) \neq \emptyset,\ |T| = w\}$$

as the *set of descendants* of $\mathcal{B}$.

Example 3.1.3 *Let $\mathcal{X} = \{1,2,\ldots,6\}$ and $\mathcal{B} = \{B_1 = \{1,2,3\}, B_2 = \{2,4,5\}, B_3 = \{3,5,6\}\}$. If $T = \{3,4,5\}$, then*

$$T \subseteq B_1 \cup B_2 \quad \textit{and} \quad T \subseteq B_2 \cup B_3.$$

T is a descendant of $\{B_1, B_2\}$ (also $\{B_2, B_3\}$). Both $\{B_1, B_2\}$ and $\{B_2, B_3\}$ are parent sets of T with size 2, which implies

$$P_2(T) = \{\{B_1, B_2\}, \{B_2, B_3\}\}.$$

The set of descendants of $\mathcal{B}$ is

$$Desc_2(\mathcal{B}) = \binom{\mathcal{X}}{3} - \{\{1,4,6\}\} = \binom{\{1,\ldots,6\}}{3} - \{\{1,4,6\}\}.$$

When a pirate decoder $T \in \mathrm{Desc}_t(\mathcal{B})$ is confiscated, the data supplier wishes to identify at least one parent of T. The traitor tracing algorithm based on a t-TS, Definition 3.1.1 (1), allows a traitor to be identified in time $O(M)$ by exploiting the property that any user with the maximum number of base keys in common with the pirate T has to be a traitor [14]. t-IPPSs can be applied to trace back to at least one traitor in the intersection of all possible parents sets of T, which is not empty according to Definition 3.1.1 (2). Clearly, the traitor tracing algorithm based on a t-IPPS needs to check each subset of $\mathcal{B}$ with size at most t, resulting in time $O(M^t)$. As can be seen, both t-IPPS and t-TS can be used in traitor tracing schemes to trace back to at least one traitor, but the latter can accommodate more users, while the former is more efficient. Unfortunately, $(s,t;w)$-UIBFs and t-CFFs can not satisfy this requirement of tracing back to traitors. In some sense, they can provide some weaker security properties as follows. The scheme based on $(s,t;w)$-UIBF guarantees that any group $\mathcal{G} \subseteq \mathcal{B}$ of up to s traitors can not frame any other group $\mathcal{G}' \subseteq \mathcal{F}$ of up to t users provided that $\mathcal{G} \cap \mathcal{G}' = \emptyset$, since the intersection of $\mathcal{G}$ and $\mathcal{G}'$ contains less than w base keys in accordance with Definition 3.1.1 (3). Moreover, $(t,t;w)$-UIBFs can provide a sort of weak traceability. That is, once a pirate $T \in \mathrm{Desc}_t(\mathcal{B})$ is captured, a $(t,t;w)$-UIBF guarantees that each possible parent set of size at most t, which can generate the pirate copy T, must contain at least one traitor. Chor *et al.* [25, 26] applied t-CFFs in the traitor tracing scheme setting to ensure that any t users can not frame any innocent authorized user, for the reason that any innocent user has at least one base key not known to the traitors according to Definition 3.1.1 (4).

The parameter M is called the *size* of the set system. We also use t-TS(w,v) (t-IPPS(w,v), $(s,t;d)$-UIBF(w,v), t-CFF(w,v), resp.) to replace t-TS(w,M,v) (t-IPPS(w,M,v), $(s,t;d)$-UIBF(w,v), t-CFF(w,M,v), resp.) when M is unclear or not necessarily claimed.

- Denote $M_t(w,v)$ as the maximum size of a t-TS(w,v). A t-TS(w,v) is called *optimal* if it has size $M_t(w,v)$. Given parameters t,w and v, the goal is to explore the exact value of $M_t(w,v)$ and to construct optimal t-TS(w,v).

- Denote $I_t(w,v)$ as the maximum size of a t-IPPS(w,v). A t-IPPS(w,v) is called *optimal* if it has size $I_t(w,v)$. Given parameters t,w and v, the goal is to explore the exact value of $I_t(w,v)$ and to construct optimal t-IPPS(w,v).

- Denote $U_{s,t}(w,v;d)$ as the maximum size of an $(s,t;d)$-UIBF(w,v). An $(s,t;d)$-UIBF(w,v) is called *optimal* if it has size $U_{s,t}(w,v;d)$. Given parameters s,t,d,w and v, the goal is to explore the exact value of $U_{s,t}(w,v;d)$ and to construct optimal $(s,t;d)$-UIBF(w,v).

- Denote $f_t(w,v)$ as the maximum size of a t-CFF(w,v). A t-CFF(w,v) is called *optimal* if it has size $f_t(w,v)$. Given parameters t,w and v, the goal is to explore the exact value of $f_t(w,v)$ and to construct optimal t-CFF(w,v).

The research problems which are commonly considered for TS, IPPS and CFF are

- to derive (tight) upper and lower bounds on the maximum size of the desired set systems;
- to give constructions for the desired set systems with size as large as possible, especially with size achieving their upper bounds.

In this thesis, we are also concerned with the above types of problems for TS, IPPS, UIBF and CFF.

3.2 Relationships

The relationships among TS, IPPS and CFF have been studied in [85] and [29]. To be self-contained, we show the arguments of all relationships among TS, IPPS, UIBF and CFF in Lemma 3.2.1.

Lemma 3.2.1 *Let v, w, t be positive integers such that $v \geq w \geq 2$ and $t \geq 2$. Then*

(1) [29] A t-TS(w, v) is a t-IPPS(w, v).

(2) A t-IPPS(w, v) is a $(t, t; w)$-UIBF(w, v).

(3) A $(t, t; w)$-UIBF(w, v) is a t-CFF(w, v).

Proof: (1) Suppose $(\mathcal{X}, \mathcal{B})$ is a t-TS(w, v), then we would like to prove it is also a t-IPPS(w, v). Let $T \subseteq \mathcal{X}$ be a w-subset such that $P_t(T) \neq \emptyset$. Since $(\mathcal{X}, \mathcal{B})$ is a t-TS(w, v) and each $\mathcal{P} \in P_t(T)$ can generate T, we have

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \subseteq \mathcal{P}$$

for any $\mathcal{P} \in P_t(T)$. It implies that

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \subseteq \bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P} \neq \emptyset.$$

By Definition 3.1.1 (2), $(\mathcal{X}, \mathcal{B})$ is a t-IPPS(w, v).

(2) Suppose $(\mathcal{X}, \mathcal{B})$ is not a $(t, t; w)$-UIBF(w, v), then we would like to prove it is not a t-IPPS(w, v). By the definition of $(t, t; w)$-UIBF(w, v), there exist $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{B}$ such that $|\mathcal{B}_1| = t$, $|\mathcal{B}_2| = t$, $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ and

$$|(\bigcup_{B \in \mathcal{B}_1} B) \cap (\bigcup_{B \in \mathcal{B}_2} B)| \geq w.$$

Let T be a w-subset of $(\bigcup_{B\in\mathcal{B}_1} B)\cap(\bigcup_{B\in\mathcal{B}_2} B)$. Then $\mathcal{B}_1,\mathcal{B}_2 \in P_t(T)$. However, $\mathcal{B}_1\cap\mathcal{B}_2=\emptyset$ implies that $\bigcap_{\mathcal{P}\in P_t(T)}\mathcal{P}=\emptyset$. By Definition 3.1.1 (2), $(\mathcal{X},\mathcal{B})$ is not a t-IPPS(w,v).

(3) Suppose $(\mathcal{X},\mathcal{B})$ is not a t-CFF(w,v), then we would like to prove it is not a $(t,t;w)$-UIBF(w,v). By the definition of t-CFF(w,v), there exist $t+1$ blocks $B_1,\ldots,B_{t+1}\in\mathcal{B}$ such that $B_{t+1}\subseteq\bigcup_{1\le i\le t}B_i$. Arbitrarily choose other $t-1$ blocks $B_{t+2},\ldots,B_{2t}\in\mathcal{B}\setminus\{B_1,\ldots,B_{t+1}\}$, we have

$$|(\bigcup_{1\le i\le t}B_i)\cap(\bigcup_{t+1\le j\le 2t}B_j)|\ge|(\bigcup_{1\le i\le t}B_i)\cap B_{t+1}|=w.$$

By Definition 3.1.1 (3), $(\mathcal{X},\mathcal{B})$ is not a $(t,t;w)$-UIBF(w,v). □

The relationship between TS and CFF was investigated in [85]. That is, a t-TS is a t-CFF, which follows from Lemma 3.2.1 (1), (2) and (3). The relationship between IPPS and CFF was shown in [29]. That is, a t-IPPS is a t-CFF, which follows from Lemma 3.2.1 (2) and (3). In other words, t-TS is a special kind of t-IPPS, t-IPPS is a special kind of $(t,t;w)$-UIBF(w,v), and $(t,t;w)$-UIBF(w,v) is a special kind of t-CFF. In the following, we give some examples to certify this relationship.

Example 3.2.2 *(A t-IPPS may not be a t-TS.)*

Let $\mathcal{X}=\{1,\ldots,9\}$ *and* $\mathcal{B}=\{\{1,2,3,4\},\{1,2,3,5\},\{6,7,8,9\}\}$. *Then* $(\mathcal{X},\mathcal{B})$ *is a* 2*-IPPS*$(4,9)$ *but not a* 2*-TS*$(4,9)$.

Indeed, for any 4*-subset* $T\subseteq\mathcal{X}$ *which can be generated by two blocks in* $\mathcal{B}$, T *must contain at least one element of* $\{4,\ldots,9\}$. *Since each element of* $\{4,\ldots,9\}$ *determines one block, by Definition 3.1.1 (2),* $(\mathcal{X},\mathcal{B})$ *is a* 2*-IPPS*$(4,9)$.

However, $\{1,2,3,4\}$ *and* $\{6,7,8,9\}$ *can generate a* 4*-subset* $T=\{1,2,3,9\}$, *but*

$$|\{1,2,3,9\}\cap\{1,2,3,4\}|=3,$$
$$|\{1,2,3,9\}\cap\{1,2,3,5\}|=3,$$
$$|\{1,2,3,9\}\cap\{6,7,8,9\}|=1,$$

which contradicts Definition 3.1.1 (1) and implies $(\mathcal{X},\mathcal{B})$ *is not a* 2*-TS*$(4,9)$.

Example 3.2.3 *(A* $(t,t;w)$*-UIBF may not be a t-IPPS.)*

Let $\mathcal{X}=\{1,\ldots,9\}$ *and* $\mathcal{B}=\{\{1,2,3\},\{1,4,5\},\{3,4,6\},\{7,8,9\}\}$. *Then* $(\mathcal{X},\mathcal{B})$ *is a* $(2,2;3)$*-UIBF*$(3,9)$ *but not a* 2*-IPPS*$(3,9)$.

In fact, when we divide the four blocks in $\mathcal{B}$ *into two disjoint groups of size* 2, *it is equivalent to take two blocks from* $\{1,2,3\},\{1,4,5\},\{3,4,6\}$. *One case is*

$$|(\{1,2,3\}\cup\{1,4,5\})\cap(\{3,4,6\}\cup\{7,8,9\})|=|\{3,4\}|=2<3.$$

It is easy to check that in the two other cases, the intersection is also 2*. By Definition 3.1.1 (3),* $(\mathcal{X}, \mathcal{B})$ *is a* $(2,2;3)$*-UIBF*$(3,9)$*.*

However, any two of $\{1,2,3\}, \{1,4,5\}, \{3,4,6\}$ *can generate the same* 3*-subset* $T = \{1,3,4\}$*, which implies* $\bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P} = \emptyset$*. By Definition 3.1.1 (2),* $(\mathcal{X}, \mathcal{B})$ *is not a* 2*-IPPS*$(3,9)$*.*

Example 3.2.4 *(A* t*-CFF may not be a* $(t,t;w)$*-UIBF.)*

Let $\mathcal{X} = \{1,\ldots,9\}$ *and* $\mathcal{B} = \{\{1,2,3\},\{3,4,5\},\{5,6,7\},\{7,8,9\}\}$*. Then* $(\mathcal{X}, \mathcal{B})$ *is a* 2*-CFF*$(3,9)$ *but not a* $(2,2;3)$*-UIBF*$(3,9)$*.*

Indeed, since each block in $\mathcal{B}$ *contains at least one* 1*-own-subset, so any block can not be covered by other two blocks. By Definition 3.1.1 (4),* $(\mathcal{X}, \mathcal{B})$ *is a* 2*-CFF*$(3,9)$*.*

However,

$$|(\{1,2,3\} \cup \{5,6,7\}) \cap (\{3,4,5\} \cup \{7,8,9\})| = |\{3,5,7\}| = 3$$

implies that $(\mathcal{X}, \mathcal{B})$ *is not a* $(2,2;3)$*-UIBF*$(3,9)$*.*

Furthermore, by Lemma 3.2.1, we have the following corollary.

Corollary 3.2.5 *Let* v, w, t *be positive integers such that* $v \geq w \geq 2$ *and* $t \geq 2$*. Then*

(1) $M_t(w,v) \leq I_t(w,v)$*.*

(2) $I_t(w,v) \leq U_{t,t}(w,v;w)$*.*

(3) $U_{t,t}(w,v;w) \leq f_t(w,v)$*.*

$(\mathcal{X}, \mathcal{B})$	$\lim\limits_{v\to\infty} \log_v \lvert\mathcal{B}\rvert \geq$	$\lim\limits_{v\to\infty} \log_v \lvert\mathcal{B}\rvert \leq$	shown in
t-TS(w,v)	$\lceil w/t^2 \rceil$	$\lceil w/t^2 \rceil$	Chapter 4
t-IPPS(w,v)	$w/(\lfloor t^2/4 \rfloor + t)$	$\lceil w/(\lfloor t^2/4 \rfloor + t) \rceil$	Chapter 5
$(t,t;w)$-UIBF(w,v)	$w/(2t-1)$	$\lceil w/(2t-1) \rceil$	Chapter 6
t-CFF(w,v)	$\lceil w/t \rceil$	$\lceil w/t \rceil$	Reference [39]

Table 3.1: Bounds for TS, IPPS, UIBF and CFF

Table 3.1 presents a general overview of the bounds on the size of t-TS(w,v), t-IPPS(w,v), $(t,t;w)$-UIBF(w,v) and t-CFF(w,v). We draw Figure 3.1 to express the corresponding tools used to derive bounds for TS, IPPS and UIBF.

It is noteworthy that although several similar techniques are used to derive bounds for TS, IPPS and UIBF, their arguments are distinctive. For instance, to improve the known upper bounds of t-TS and t-IPPS, later we will use two different characteristics of own-subsets: (1) the largest possible number of own-subsets (with a fixed size) possessed by each block in $\mathcal{B}$ (Lemma 4.2.6), which improves the coefficient of the upper bound of t-TS; (2) the least possible size of an own-subset possessed by some block in $\mathcal{B}$ (Lemma 5.3.2), which leads to the improvement on the order of magnitude of the upper bound of t-IPPS.

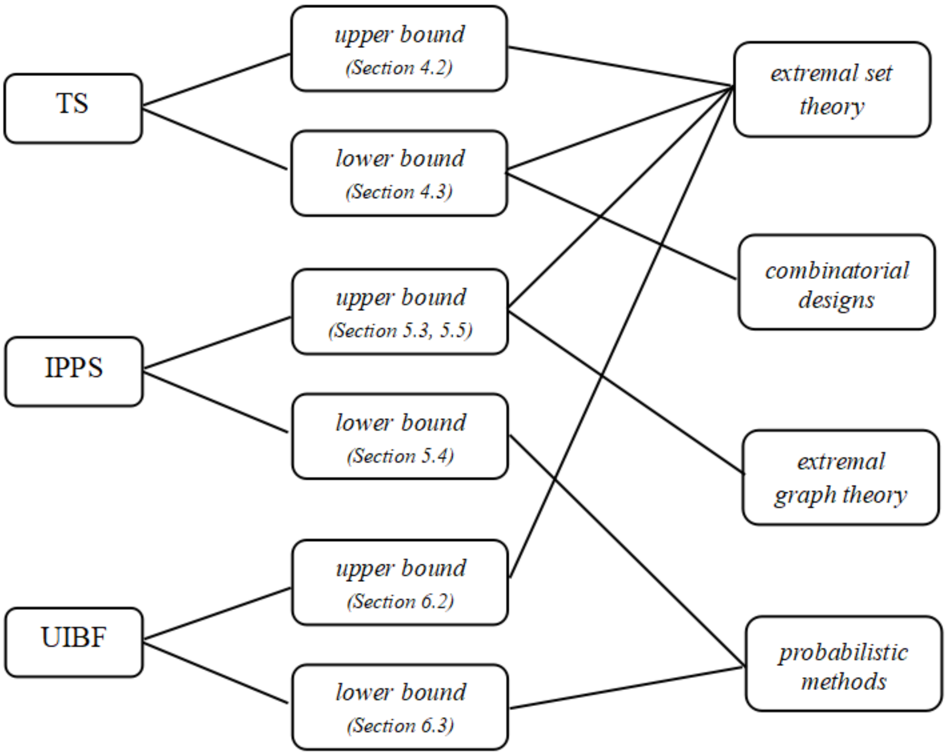
TS
IPPS
UIBF
upper bound (Section 4.2)
lower bound (Section 4.3)
upper bound (Section 5.3, 5.5)
lower bound (Section 5.4)
upper bound (Section 6.2)
lower bound (Section 6.3)
extremal set theory
combinatorial designs
extremal graph theory
probabilistic methods

Figure 3.1: Tools for TS, IPPS and UIBF

In the following Chapter 4, Chapter 5 and Chapter 6, we will derive bounds on the size of t-TS, t-IPPS and $(s, t; d)$-UIBF, respectively.

Chapter 4

Traceability Schemes

In this chapter, we focus on the traceability schemes (TSs). First, some remarks on the definition of TS and the related known results are stated in Section 4.1. Second, a new and interesting relationship between TS and CFF is proved in Section 4.2.1, New upper bounds on the size of TS are shown in Section 4.2.2 for general cases and Section 4.2.2 for several special cases, which greatly improve the previously known results. Third, based on the new upper bounds, we obtain optimal t-TS(w, v) from sunflowers for the case $w \leq t^2$ in Section 4.3.1. Constructions of TS by means of combinatorial designs are provided in Section 4.3.2, which also can achieve our new upper bounds in many cases. Moreover, a constructive lower bound for general TS is given in Section 4.3.3, which has the same order of magnitude with the new upper bounds. Finally, we conclude this chapter in Section 4.4. This chapter is based on results in [51].

4.1 Definition and known results

4.1.1 Definition

In Chapter 3.1, the definition of TS given by Stinson and Wei [85] was exposed in Definition 3.1.1 (1). Equivalently, the definition of TS can be described as follows.

Definition 4.1.1 *Let $(\mathcal{X}, \mathcal{B})$ be a (w, v) set system. Then $(\mathcal{X}, \mathcal{B})$ is a t-traceability scheme, denoted t-TS(w, v), if for any s blocks $B_1, \ldots, B_s \in \mathcal{B}$, $s \leq t$, and any w-subset $T \subseteq \bigcup_{1 \leq i \leq s} B_i$, we have*

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \subseteq \{B_1, B_2, \ldots, B_s\}.$$

The equivalence of Definition 3.1.1 (1) and Definition 4.1.1 can be easily derived from their descriptions. Noting that the TS defined in both Definition 3.1.1(1) and Definition 4.1.1 requires that $|T| = w$. Modifying it to the case $|T| \geq w$ and $T \subseteq \bigcup_{1 \leq i \leq s} B_i$, we have the following definition of t-TS*.

Definition 4.1.2 *Let $(\mathcal{X}, \mathcal{B})$ be a (w, v) set system. Then $(\mathcal{X}, \mathcal{B})$ is a t-traceability scheme*, denoted t-TS$^\star(w, v)$, if for any s blocks $B_1, \ldots, B_s \in \mathcal{B}$, $s \leq t$, and any subset $T \subseteq \bigcup_{1 \leq i \leq s} B_i$ such that $|T| \geq w$, we have*

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \subseteq \{B_1, B_2, \ldots, B_s\}.$$

In the following lemma, we show that Definition 4.1.1 (also Definition 3.1.1 (1)) and Definition 4.1.2 are in fact equivalent.

Lemma 4.1.3 *A set system $(\mathcal{X}, \mathcal{B})$ is a t-TS(w, v) if and only if it is a t-TS$^\star(w, v)$.*

Proof: If $(\mathcal{X}, \mathcal{B})$ is a t-TS*, then we immediately have that $(\mathcal{X}, \mathcal{B})$ is a t-TS.

Now we prove the converse.

If $(\mathcal{X}, \mathcal{B})$ is not a t-TS$^\star(w, v)$, then there exist $s \leq t$ blocks $B_1, \ldots, B_s \in \mathcal{B}$, a subset $T \subseteq \bigcup_{1 \leq i \leq s} B_i$ with $|T| \geq w$, and another block $B' \in \mathcal{B} \setminus \{B_1, \ldots, B_s\}$ such that

$$|T \cap B'| \geq \max_{1 \leq i \leq s} |T \cap B_i|. \tag{4.1}$$

Now we take a w-subset $T' \subseteq T$ such that $(T \cap B') \subseteq T'$. This is doable since $|T \cap B'| \leq |B'| = w$. Hence

$$\begin{aligned} |T' \cap B'| &= |T \cap B'| \\ &\geq \max_{1 \leq i \leq s} |T \cap B_i| \\ &\geq \max_{1 \leq i \leq s} |T' \cap B_i|, \end{aligned}$$

where the first inequality follows from (4.1) and the second inequality follows from $T' \subseteq T$. This implies that $(\mathcal{X}, \mathcal{B})$ is not a t-TS, completing the proof. □

The above lemma illustrates that considering the condition $T = w$ and $T \geq w$ is basically the same. Here we also remark that in the definition of t-TS, considering "for any $s \leq t$ blocks" is almost the same as that considering "for any t blocks", except for the trivial case that $|\mathcal{B}| \leq t$. Since for any set system $(\mathcal{X}, \mathcal{B})$ such that $|\mathcal{B}| > t$, if there exist $s < t$ blocks $B_1, \ldots, B_s \in \mathcal{B}$ and a w-subset T such that

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \nsubseteq \{B_1, B_2, \ldots, B_s\},$$

then we always can choose $t - s$ blocks $B_{s+1}, \ldots, B_t \in \mathcal{B} \setminus \{B_1, \ldots, B_s\}$ such that

$$\{B' \in \mathcal{B} : \ |T \cap B'| = \max_{B \in \mathcal{B}} |T \cap B|\} \nsubseteq \{B_{s+1}, \ldots, B_t\}.$$

This is doable since $|\mathcal{B}| \geq t+1$. Correspondingly, there also exist t blocks $B_1, \ldots, B_t$ contradicting the requirement of t-TS(w, v).

4.1.2 Known results

In [85], Stinson and Wei provided an upper bound for t-TS(w, v) as follows.

Theorem 4.1.4 ([85]) *For any $v \geq w \geq 2$, $t \geq 2$, we have*

$$M_t(w, v) \leq \binom{v}{\lceil w/t \rceil} \Big/ \binom{w-1}{\lceil w/t \rceil - 1}.$$

By investigating $\lceil w/t^2 \rceil$-own-subsets, Collins [29] improved the above upper bound.

Theorem 4.1.5 ([29]) *For any $v \geq w \geq 2,\ t \geq 2$, we have*

$$M_t(w, v) \leq \binom{v}{\lceil w/t^2 \rceil}.$$

However, in the literature, no construction can produce t-TS(w, v) achieving the above upper bound. As a matter of fact, this bound is still not tight. In the next section, we show an interesting discovery of the relationship between TS and CFF. Based on this important discovery, new upper bounds for t-TS(w, v) are derived, which are great improvements on the previously known upper bounds. Moreover, we describe several constructions in Section 4.3, which can produce infinite families of t-TS(w, v) achieving our new upper bounds.

4.2 New upper bounds

To show our new upper bounds, we first state an interesting discovery on the relationship between TS and CFF. As far as we know, this is the first relationship between two kinds of anti-collusion schemes which strengthens the strength from t to t^2.

4.2.1 A new relationship between TS and CFF

Stinson and Wei [85] showed that a t-TS(w, v) is a t-CFF(w, v), as described in Lemma 3.2.1. Here, we have a new relationship between TS and CFF as follows.

Lemma 4.2.1 *A t-TS(w, v) is a t^2-CFF(w, v).*

Proof: Assume that $(\mathcal{X}, \mathcal{B})$ is a t-TS(w, v). We would like to show that $(\mathcal{X}, \mathcal{B})$ is also a t^2-CFF(w, v). Suppose, on the contrary, that $(\mathcal{X}, \mathcal{B})$ is not a t^2-CFF(w, v). Then we try to prove $(\mathcal{X}, \mathcal{B})$ is not a t-TS. That is, we would like to look for t distinct blocks $B_1, \ldots, B_t \in \mathcal{B}$ and a descendant T generated by $B_1, \ldots, B_t$, such that there exists another block in $\mathcal{B} \backslash \{B_1, \ldots, B_t\}$ which has at least $\max\{|T \cap B_i| : 1 \leq i \leq t\}$ points in common with T.

By the definition of t^2-CFF, if it is not a t^2-CFF, there exists $B_0 \in \mathcal{B}$ which can be covered by the union of some other t^2 blocks from $\mathcal{B} \setminus \{B_0\}$. Denote $\mathcal{B}_0$ as the collection of such t^2 blocks.

By the pigeonhole principle, there exists at least one block $B \in \mathcal{B}_0$ such that $|B_0 \cap B| \geq \lceil \frac{w}{t^2} \rceil$. Denote

$$\max\{|B_0 \cap B| : B \in \mathcal{B}_0\} = \lceil \frac{w}{t^2} \rceil + \sigma_1,$$

where $0 \leq \sigma_1 < w - \lceil \frac{w}{t^2} \rceil$. Assume that $B_1 \in \mathcal{B}_0$ satisfies $|B_0 \cap B_1| = \lceil \frac{w}{t^2} \rceil + \sigma_1$.

For $2 \leq i \leq t$, choose $B_i \in \mathcal{B}_0 \setminus \{B_j : 1 \leq j \leq i-1\}$ such that

$$|B_i \cap (B_0 \setminus \cup_{1 \leq j \leq i-1} B_j)| = \sigma_i > 0,$$

and

$$|(\bigcup_{2 \leq i \leq t} B_i \cap B_0) \setminus B_1| = \sum_{2 \leq i \leq t} \sigma_i \geq \frac{1}{t+1}(w - \lceil \frac{w}{t^2} \rceil - \sigma_1). \tag{4.2}$$

The above $t-1$ blocks, satisfying $\sigma_i > 0$ for $2 \leq i \leq t$, are available, since if not, then B_0 would be covered by at most t distinct blocks, which contradicts Lemma 3.2.1. Inequality (4.2) can be realized by using the pigeonhole principle to $t^2 - 1$ blocks in $\mathcal{B}_0 \setminus \{B_1\}$, that is, there exist $t-1$ blocks $B_2, \ldots, B_t \in \mathcal{B}_0 \setminus \{B_1\}$ such that

$$|(\bigcup_{2 \leq i \leq t} B_i \cap B_0) \setminus B_1| \geq \frac{t-1}{t^2-1}(|B_0| - |B_1 \cap B_0|) = \frac{1}{t+1}(w - \lceil \frac{w}{t^2} \rceil - \sigma_1).$$

By the assumption, for $2 \leq i \leq t$, we have

$$\sigma_i \leq |B_i \cap B_0| \leq |B_1 \cap B_0| = \lceil \frac{w}{t^2} \rceil + \sigma_1.$$

Now we are going to show that each block B_i, $1 \leq i \leq t$, has more than $\sum_{2 \leq i \leq t} \sigma_i$ points which are not contained in any other B_j for $0 \leq j \leq t$ and $j \neq i$. To achieve this goal, we want to show

$$0 \leq |(\bigcup_{1 \leq i < j \leq t} B_i \cap B_j) \setminus B_0| < w - (\lceil \frac{w}{t^2} \rceil + \sum_{1 \leq i \leq t} \sigma_i). \tag{4.3}$$

Indeed, if

$$|(\bigcup_{1 \leq i < j \leq t} B_i \cap B_j) \setminus B_0| \geq w - (\lceil \frac{w}{t^2} \rceil + \sum_{1 \leq i \leq t} \sigma_i),$$

then any t of $B_0, B_1, \ldots, B_t$ can generate the same descendant, that is, a w-subset of

$$\bigcup_{0 \leq i < j \leq t} (B_i \cap B_j),$$

where $|\bigcup_{0 \leq i < j \leq t}(B_i \cap B_j)| \geq w$, which contradicts Lemma 3.2.1. Thus inequality (4.3) follows.

Thus, for each $1 \leq i \leq t$, we have

$$\begin{aligned}
|B_i \setminus \bigcup_{\substack{0\leq j\leq t,\\ j\neq i}} B_j| &= |B_i| - |B_i \cap (\bigcup_{\substack{0\leq j\leq t,\\ j\neq i}} B_j)| \\
&= |B_i| - |B_i \cap B_0| - |(B_i \cap (\bigcup_{\substack{1\leq j\leq t,\\ j\neq i}} B_j)) \setminus B_0| \\
&= w - |B_i \cap B_0| - |(\bigcup_{\substack{1\leq j\leq t,\\ j\neq i}} B_i \cap B_j) \setminus B_0| \qquad (4.4) \\
&> w - (\lceil \frac{w}{t^2} \rceil + \sigma_1) - [w - (\lceil \frac{w}{t^2} \rceil + \sum_{1\leq i\leq t} \sigma_i)] \\
&= \sum_{2\leq i\leq t} \sigma_i,
\end{aligned}$$

where the inequality follows from (4.3). That is, each block B_i, $1 \leq i \leq t$, has more than $\sum_{2\leq i\leq t} \sigma_i$ points which are not contained in any other B_j for $0 \leq j \leq t$ and $j \neq i$.

For $1 \leq i \leq t$, let $\hat{B}_i \subseteq B_i \setminus \bigcup_{\substack{0\leq j\leq t,\\ j\neq i}} B_j$ be a subset of B_i such that $|\hat{B}_i| = \sum_{2\leq i\leq t} \sigma_i$, which is available by (4.4). Clearly, $\hat{B}_i \cap \hat{B}_j = \emptyset$ for any $1 \leq i < j \leq t$.

To find a suitable descendant T of $B_1, \ldots, B_t$, we construct a set $A \subseteq \bigcup_{1\leq i\leq t} B_i$ by including all points in $B_0 \cap B_i$ and $\hat{B}_i$ for all $1 \leq i \leq t$. That is,

$$A = \bigcup_{1\leq i\leq t} ((B_0 \cap B_i) \cup \hat{B}_i).$$

Then the size of A is

$$\begin{aligned}
|A| &= | \bigcup_{1\leq i\leq t} (B_i \cap B_0)| + | \bigcup_{1\leq i\leq t} \hat{B}_i| \\
&= |B_1 \cap B_0| + |(\bigcup_{2\leq i\leq t} B_i \cap B_0) \setminus B_1| + \sum_{1\leq i\leq t} |\hat{B}_i| \\
&= (\lceil \frac{w}{t^2} \rceil + \sigma_1) + \sum_{2\leq i\leq t} \sigma_i + t \sum_{2\leq i\leq t} \sigma_i \\
&= \lceil \frac{w}{t^2} \rceil + \sigma_1 + (t+1) \sum_{2\leq i\leq t} \sigma_i \\
&\geq \lceil \frac{w}{t^2} \rceil + \sigma_1 + (t+1) \frac{1}{t+1} (w - \lceil \frac{w}{t^2} \rceil - \sigma_1) \\
&= w,
\end{aligned}$$

where the first equality follows from that $\hat{B}_i \cap B_0 = \emptyset$ for $1 \leq i \leq t$, the second equality follows from that $\hat{B}_i \cap \hat{B}_j = \emptyset$ for any $1 \leq i < j \leq t$, and the inequality follows from (4.2).

Let T be a w-subset of A such that $\bigcup_{1\leq i\leq t}(B_i \cap B_0) \subseteq T$. Such T exists since $T \subseteq A$ and $|\bigcup_{1\leq i\leq t}(B_i \cap B_0)| \leq |B_0| = w$. Furthermore, $T \subseteq A \subseteq \bigcup_{1\leq i\leq t} B_i$. Thst is, T is a descendant of $B_1, \ldots, B_t$. However,

$$|B_i \cap T| \leq |B_0 \cap B_i| + \sum_{2\leq i\leq t} \sigma_i \leq \lceil \frac{w}{t^2} \rceil + \sum_{1\leq i\leq t} \sigma_i, \; 1 \leq i \leq t,$$

and

$$|B_0 \cap T| = \lceil \frac{w}{t^2} \rceil + \sum_{1\leq i\leq t} \sigma_i,$$

which contradicts the definition of t-TS(w, v).

Therefore, for any $B \in \mathcal{B}$, it can not be covered by the union of any other t^2 blocks from $\mathcal{B} \setminus \{B\}$. The lemma follows. □

Here we remark that the converse of Lemma 4.2.1 is not correct. That is, a t^2-CFF(w, v) may not be a t-TS(w, v). We give evidence to this by virtue of the following example.

Example 4.2.2 *(A t^2-CFF may not be a t-TS.)*

Let $\mathcal{X} = \{1, 2, \ldots, 11\}$ and $\mathcal{B} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 2, 7, 8\}, \{1, 2, 9, 10\}, \{3, 5, 7, 11\}\}$. Then $(\mathcal{X}, \mathcal{B})$ is a 4-CFF$(4, 11)$, but not a 2-TS$(4, 11)$.

Indeed, each block has one private point which is not contained in any other blocks. Thus each block can not be covered by the union of any other four blocks. It is a 4-CFF$(4, 11)$.

However, $\{1, 2, 3, 4\}$ and $\{1, 2, 5, 6\}$ can generate a new 4-subset $\{3, 4, 5, 6\}$, and the fact

$$\begin{aligned}
|\{3, 4, 5, 6\} \cap \{1, 2, 3, 4\}| &= |\{3, 4\}| = 2, \\
|\{3, 4, 5, 6\} \cap \{1, 2, 5, 6\}| &= |\{5, 6\}| = 2, \\
|\{3, 4, 5, 6\} \cap \{3, 5, 7, 11\}| &= |\{3, 5\}| = 2
\end{aligned}$$

implies that $(\mathcal{X}, \mathcal{B})$ is not a 2-TS$(4, 11)$.

With regards to Lemma 4.2.1, another natural question may be whether t^2 is the critical value for the relationship between TS and CFF. That is, is it possible that a t-TS(w, v) is a t'-CFF(w, v), where t' is an integer such that $t' > t^2$? We give an example to show that a t-TS may not be a $(t^2 + 1)$-CFF.

Example 4.2.3 *(A t-TS may not be a $(t^2 + 1)$-CFF.)*

Let $\mathcal{X} = \{1, 2, \ldots, 15\}$ *and*

$$\begin{aligned} \mathcal{B} = \{B_1 &= \{1,2,3,4,5\}, \\ B_2 &= \{1,6,7,8,9\}, \\ B_3 &= \{2,6,10,11,12\}, \\ B_4 &= \{3,7,10,13,14\}, \\ B_5 &= \{4,8,11,13,15\}, \\ B_6 &= \{5,9,12,14,15\}\}. \end{aligned}$$

It is obvious that for any two distinct blocks B_i *and* B_j*, where* $1 \leq i < j \leq 6$*, we have*

$$|B_i \cap B_j| = 1.$$

By Definition 3.1.1 (1), it is easy to check that $(\mathcal{X}, \mathcal{B})$ *is a* 2*-TS*$(5, 15)$ *with size* 6*. In the meanwhile,* $(\mathcal{X}, \mathcal{B})$ *is also a* 4*-CFF*$(5, 15)$ *since no block can be covered by the union of four others.*

However, $(\mathcal{X}, \mathcal{B})$ *is not a* 5*-CFF*$(5, 15)$*. In fact,*

$$B_1 \subseteq B_2 \cup B_3 \cup B_4 \cup B_5 \cup B_6,$$

which implies that B_1 *can be covered by the union of five other blocks, contradicting Definition 3.1.1 (4).*

In conclusion, $(\mathcal{X}, \mathcal{B})$ *is a* 2*-TS*$(5, 15)$ *and also a* 4*-CFF*$(5, 15)$*, but not a* 5*-CFF*$(5, 15)$*.*

Furthermore, the above example can be generalized to the following.

Example 4.2.4 *(A* t*-TS may not be a* (t^2+1)*-CFF.)*

Let $\mathcal{X} = \{1, 2, \ldots, \binom{t^2+2}{2}\}$ *and*

$$\begin{aligned} \mathcal{B} = \{B_0 &= \{1,\ 2,\ 3,\ 4,\ \ldots,\ t^2,\ t^2+1\}, \\ B_1 &= \{1,\ t^2+2,\ t^2+3,\ t^2+4,\ \ldots,\ 2t^2,\ 2t^2+1\}, \\ B_2 &= \{2,\ t^2+2,\ 2t^2+2,\ 2t^2+3,\ \ldots,\ 3t^2-1,\ 3t^2\}, \\ B_3 &= \{3,\ t^2+3,\ 2t^2+2,\ 3t^2+1,\ \ldots,\ 4t^2-3,\ 4t^2-2\}, \\ B_4 &= \{4,\ t^2+4,\ 2t^2+3,\ 3t^2+1,\ \ldots,\ 5t^2-6,\ 5t^2-5\}, \\ &\cdots \\ B_{t^2} &= \{t^2,\ 2t^2,\ 3t^2-1,\ 4t^2-3,\ \ldots,\ \binom{t^2+2}{2}-2,\ \binom{t^2+2}{2}\}, \\ B_{t^2+1} &= \{t^2+1,\ 2t^2+1,\ 3t^2,\ 4t^2-2,\ \ldots,\ \binom{t^2+2}{2}-1,\ \binom{t^2+2}{2}\}\}. \end{aligned}$$

Clearly, $|B_i \cap B_j| = 1$ *for any* $0 \leq i < j \leq t^2+1$*.*

- *$(\mathcal{X},\mathcal{B})$ is a t-TS$(t^2+1,\binom{t^2+2}{2})$.*

  Indeed, for any $\mathcal{B}_0 \subseteq \mathcal{B}$ such that $|\mathcal{B}_0| \leq t$ and any (t^2+1)-subset $F \subseteq \bigcup_{B\in\mathcal{B}_0} B$, we have

  $$\max_{B\in\mathcal{B}} |B\cap F| \geq \max_{B\in\mathcal{B}_0} |B\cap F| \geq \frac{t^2+1}{|\mathcal{B}_0|} \geq \frac{t^2+1}{t} > t.$$

  For any $B_j \in \mathcal{B}\setminus\mathcal{B}_0$, we have

  $$|B_j\cap F| \leq |B_j \cap (\bigcup_{B\in\mathcal{B}_0} B)| = |\mathcal{B}_0| \leq t < \max_{B\in\mathcal{B}} |B\cap F|,$$

  as desired.

- *$(\mathcal{X},\mathcal{B})$ is a t^2-CFF$(t^2+1,\binom{t^2+2}{2})$.*

  In fact, for any t^2+1 distinct blocks $B_i, B_{j_1},\ldots,B_{j_{t^2}} \in \mathcal{B}$, we have

  $$|B_i \cap (\bigcup_{1\leq k\leq t^2} B_{j_k})| = t^2 < |B_i| = t^2+1,$$

  which implies that B_i cannot be covered by other t^2 blocks.

- *However, $(\mathcal{X},\mathcal{B})$ is not a (t^2+1)-CFF$(t^2+1,\binom{t^2+2}{2})$ since*

  $$B_0 \subseteq \bigcup_{1\leq i\leq t^2+1} B_i.$$

4.2.2 A new upper bound for general t-TS

In [39], Erdős, Frankl and Füredi proved the following result.

Lemma 4.2.5 ([39]) *Let $(\mathcal{X},\mathcal{B})$ be a t-CFF(w,v), and $B\in\mathcal{B}$. The number of $\lceil w/t\rceil$-own-subsets in B is at least $\binom{w-1}{\lceil w/t\rceil-1}$.*

Combining Lemma 4.2.1 and Lemma 4.2.5, we have the following lemma.

Lemma 4.2.6 *Let $(\mathcal{X},\mathcal{B})$ be a t-TS(w,v), and $B\in\mathcal{B}$. The number of $\lceil w/t^2\rceil$-own-subsets in B is at least $\binom{w-1}{\lceil w/t^2\rceil-1}$.*

In the following, we prove a new upper bound for t-TS by means of the double-counting technique. In combinatorics, *double-counting*, also called *counting in two ways*, is a combinatorial argument technique of describing a finite set from two perspectives, leading to two distinct expressions for the size of one set. In [66], van Lint and Wilson also regard it as “one of the most important tools in combinatorics”.

Theorem 4.2.7 *For any $v \geq w \geq 2,\ t \geq 2$, we have*

$$M_t(w,v) \leq \frac{\binom{v}{\lceil w/t^2 \rceil} - \binom{w-1}{\lceil w/t^2 \rceil}}{\binom{w-1}{\lceil w/t^2 \rceil - 1}}.$$

Proof: Suppose $(\mathcal{X},\mathcal{B})$ is a t-TS(w,M,v). Denote $\binom{\mathcal{X}}{\lceil w/t^2 \rceil}$ as the collection of all $\lceil w/t^2 \rceil$-subsets of $\mathcal{X}$. Clearly, $|\binom{\mathcal{X}}{\lceil w/t^2 \rceil}| = \binom{v}{\lceil w/t^2 \rceil}$. The following is to double count the set $\{(T,B) : T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil},\ B \in \mathcal{B},\ T \subseteq B\}$. Denote

$$\Sigma := |\{(T,B) : T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil},\ B \in \mathcal{B},\ T \subseteq B\}|.$$

Then we have

$$\Sigma = \sum_{T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil}} \sum_{\substack{B \in \mathcal{B} \\ \text{s.t. } T \subseteq B}} 1 = \sum_{B \in \mathcal{B}} \sum_{\substack{T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil} \\ \text{s.t. } T \subseteq B}} 1. \tag{4.5}$$

On one hand, fixing $B \in \mathcal{B}$, we have $\sum_{\substack{T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil} \\ \text{s.t. } T \subseteq B}} 1 = \binom{w}{\lceil w/t^2 \rceil}$. Then

$$\Sigma = \sum_{B \in \mathcal{B}} \binom{w}{\lceil w/t^2 \rceil} = M \binom{w}{\lceil w/t^2 \rceil}. \tag{4.6}$$

On the other hand, fixing $T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil}$, there are the following two possible cases.

(a) If T is a $\lceil w/t^2 \rceil$-own-subset of some block $B \in \mathcal{B}$, then we have $\sum_{\substack{B \in \mathcal{B} \\ \text{s.t. } T \subseteq B}} 1 = 1$.

(b) If T is not a $\lceil w/t^2 \rceil$-own-subset of any block $B \in \mathcal{B}$, then we have $\sum_{\substack{B \in \mathcal{B} \\ \text{s.t. } T \subseteq B}} 1 \leq M$.

For any $B \in \mathcal{B}$, denote $\mathcal{O}(B)$ as the collection of all $\lceil w/t^2 \rceil$-own-subsets of the block B. By Lemma 4.2.6, we have $|\mathcal{O}(B)| \geq \binom{w-1}{\lceil w/t^2 \rceil - 1}$. Without loss of generality, we assume that

$$|\bigcup_{B \in \mathcal{B}} \mathcal{O}(B)| = M \binom{w-1}{\lceil w/t^2 \rceil - 1} + \sigma,\ \sigma \geq 0.$$

Then the number of $T \in \binom{\mathcal{X}}{\lceil w/t^2 \rceil}$, satisfying the condition of case (b), is at most

$$\binom{v}{\lceil w/t^2 \rceil} - M \binom{w-1}{\lceil w/t^2 \rceil - 1} - \sigma.$$

Thus by the first equality of (4.5),

$$\begin{aligned}\Sigma &\leq [M\binom{w-1}{\lceil w/t^2\rceil-1}+\sigma]+M[\binom{v}{\lceil w/t^2\rceil}-M\binom{w-1}{\lceil w/t^2\rceil-1}-\sigma]\\ &=M[\binom{v}{\lceil w/t^2\rceil}-M\binom{w-1}{\lceil w/t^2\rceil-1}+\binom{w-1}{\lceil w/t^2\rceil-1}]-(M-1)\sigma \qquad (4.7)\\ &\leq M[\binom{v}{\lceil w/t^2\rceil}-M\binom{w-1}{\lceil w/t^2\rceil-1}+\binom{w-1}{\lceil w/t^2\rceil-1}].\end{aligned}$$

From (4.6) and (4.7), we have

$$M\binom{w}{\lceil w/t^2\rceil}\leq M[\binom{v}{\lceil w/t^2\rceil}-M\binom{w-1}{\lceil w/t^2\rceil-1}+\binom{w-1}{\lceil w/t^2\rceil-1}],$$

which implies

$$M\leq\frac{\binom{v}{\lceil w/t^2\rceil}-\binom{w}{\lceil w/t^2\rceil}}{\binom{w-1}{\lceil w/t^2\rceil-1}}+1=\frac{\binom{v}{\lceil w/t^2\rceil}-\binom{w-1}{\lceil w/t^2\rceil}}{\binom{w-1}{\lceil w/t^2\rceil-1}},$$

as desired. □

In the following, we use several examples to compare the new upper bound with the previous results for t-TS.

Example 4.2.8 *Let $t=2$ and $w=5$. The following Figure 4.1 and Figure 4.2 show our improvements on the known upper bounds for 2-TS$(5,v)$ and 4-TS$(25,v)$, respectively.*

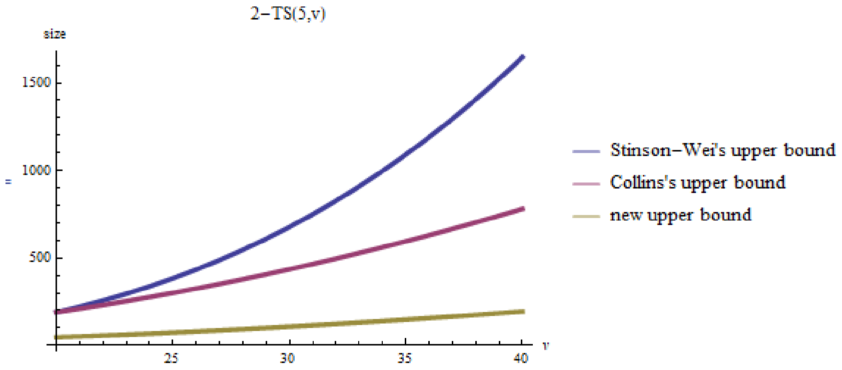

Figure 4.1: A comparison between known and new upper bounds for 2-TS$(5,v)$

Example 4.2.9 *Let $t=2$. We compare the known upper bound in Theorem 4.1.4 and our new upper bound in Theorem 4.2.7 for 2-TS(w,v), where $5\leq w\leq 10$, $2\leq v\leq 80$, in Figure 4.3. A comparison between Theorem 4.1.5 and Theorem 4.2.7 for 2-TS(w,v) is also shown in Figure 4.4.*

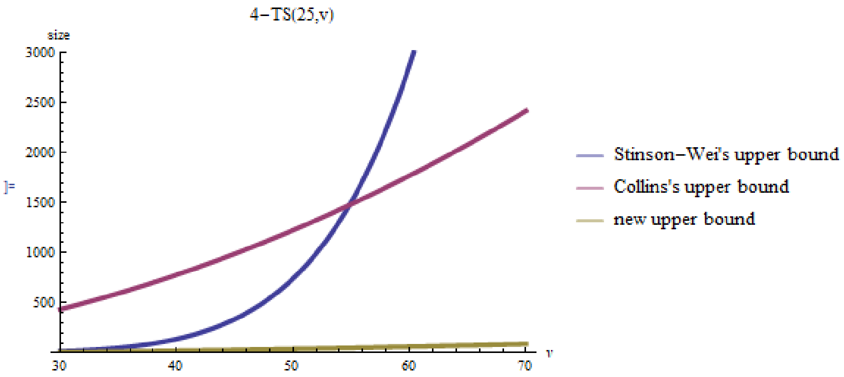

Figure 4.2: A comparison between known and new upper bounds for 4-TS$(25, v)$

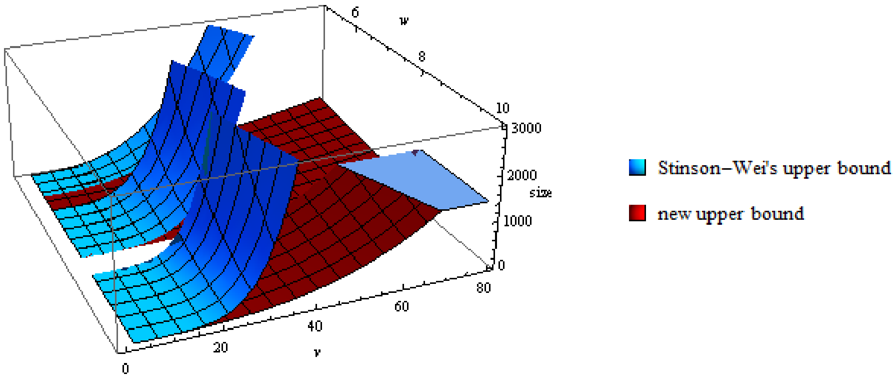

Figure 4.3: A comparison between Theorem 4.1.4 and Theorem 4.2.7 for 2-TS(w, v)

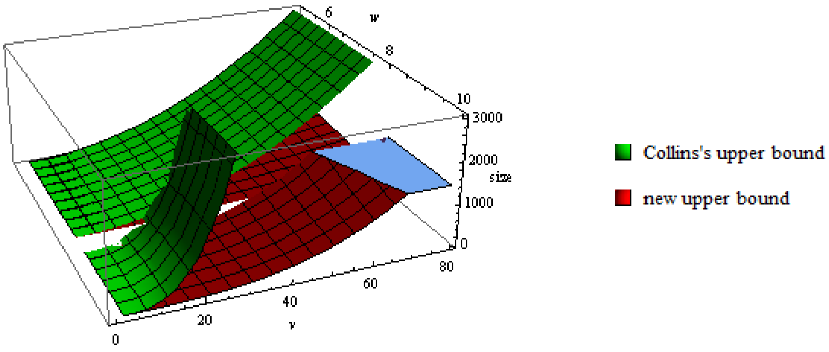

Figure 4.4: A comparison between Theorem 4.1.5 and Theorem 4.2.7 for 2-TS(w, v)

Example 4.2.10 *Let $t = 3$. In Figure 4.5 and Figure 4.6, we compare the known upper bounds in Theorem 4.1.4 and Theorem 4.1.5 with the new upper bound in Theorem 4.2.7 for 3-TS(w, v), where $5 \leq w \leq 25$, $20 \leq v \leq 80$, respectively.*

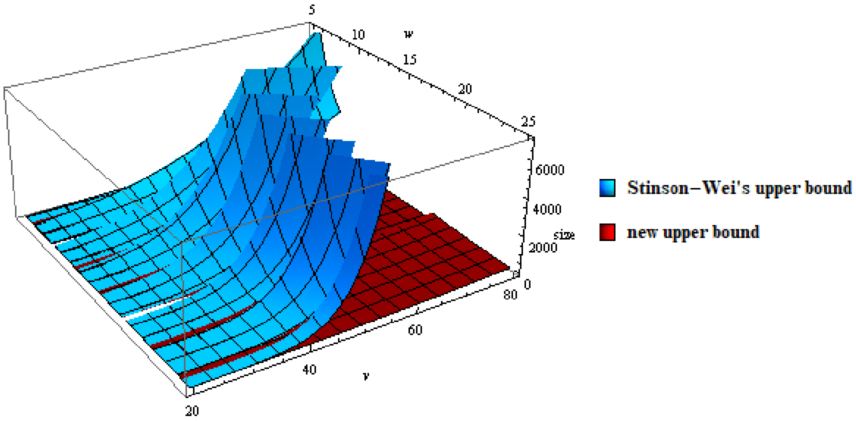

Figure 4.5: A comparison between Theorem 4.1.4 and Theorem 4.2.7 for 3-TS(w, v)

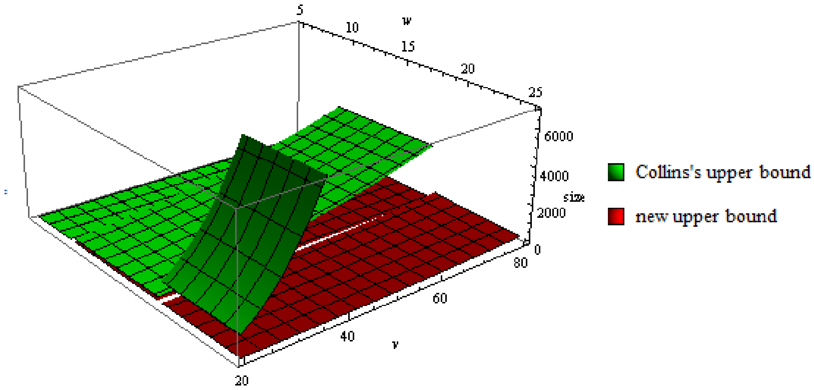

Figure 4.6: A comparison between Theorem 4.1.5 and Theorem 4.2.7 for 3-TS(w, v)

In the next section, we will give some constructions for t-TS which achieve the new upper bound in Theorem 4.2.7.

By using the double-counting technique, we also can slightly improve one known upper bound of t-CFF in [39]. Erdős, Frankl and Füredi [39] obtained the following upper bound for t-CFF(w, v).

Theorem 4.2.11 ([39]) *For any $v \geq w \geq 2$, $t \geq 2$, we have*

$$f_t(w, v) \leq \frac{\binom{v}{\lceil w/t \rceil}}{\binom{w-1}{\lceil w/t \rceil - 1}}.$$

Note that the double-counting technique used in the proof of Theorem 4.2.7 also can be applied to derive an upper bound for t-CFF(w,v). More precisely, we have the following upper bound for t-CFF(w,v), which is slightly better than that in Theorem 4.2.11.

Theorem 4.2.12 *For any $v \geq w \geq 2,\ t \geq 2$, we have*

$$f_t(w,v) \leq \frac{\binom{v}{\lceil w/t \rceil} - \binom{w-1}{\lceil w/t \rceil}}{\binom{w-1}{\lceil w/t \rceil - 1}}.$$

Proof: The proof is similar to that of Theorem 4.2.7 by using Lemma 4.2.5, and we omit it here. □

4.2.3 A better upper bound for t-TS in some special cases

Besides the upper bound for general t-TS(w,v) that we proved in the preceding subsection, a better upper bound for several special cases can be obtained. Erdős, Frankl and Füredi [39] provided the following bound for $f_r(w,v)$.

Theorem 4.2.13 ([39]) *Let $w = r(\lceil \frac{w}{r} \rceil - 1) + 1 + d$ where $0 \leq d \leq r-1$. Then for $v > 2d\lceil \frac{w}{r} \rceil \binom{w}{\lceil w/r \rceil}$,*

$$f_r(w,v) \leq \frac{\binom{v-d}{\lceil w/r \rceil}}{\binom{w-d}{\lceil w/r \rceil}}$$

holds in the following cases:

$$\text{(a) } d = 0, 1, \quad \text{(b) } d < r/(2\lceil \frac{w}{r} \rceil^2), \quad \text{(c) } \lceil \frac{w}{r} \rceil = 2 \text{ and } d < \lceil 2r/3 \rceil.$$

By using the relationship in Lemma 4.2.1, we have the following bound for t-TS.

Theorem 4.2.14 *Let $w = t^2(\lceil \frac{w}{t^2} \rceil - 1) + 1 + d$ where $0 \leq d \leq t^2 - 1$. Then for $v > 2d\lceil \frac{w}{t^2} \rceil \binom{w}{\lceil w/t^2 \rceil}$,*

$$M_t(w,v) \leq \frac{\binom{v-d}{\lceil w/t^2 \rceil}}{\binom{w-d}{\lceil w/t^2 \rceil}}$$

holds in the following cases:

$$\text{(a) } d = 0, 1, \quad \text{(b) } d < t^2/(2\lceil \frac{w}{t^2} \rceil^2), \quad \text{(c) } \lceil \frac{w}{t^2} \rceil = 2 \text{ and } d < \lceil 2t^2/3 \rceil.$$

Proof: This theorem follows from Lemma 4.2.1 and Theorem 4.2.13. □

Recalling the corresponding application in broadcast encryption, our new upper bound in Theorem 4.2.7 means that the traitor tracing scheme based on a t-TS(w,v) can accommodate at most $(\binom{v}{\lceil w/t^2 \rceil} - \binom{w-1}{\lceil w/t^2 \rceil})/\binom{w-1}{\lceil w/t^2 \rceil - 1}$ users, and futhermore, in several cases as described in Theorem 4.2.14, it can only accommodate at most $\binom{v-d}{\lceil w/t^2 \rceil}/\binom{w-d}{\lceil w/t^2 \rceil}$ users. In the next section, we will provide several constructions for t-TS, most of which are optimal, that is, the traitor tracing schemes based on such t-TS can accommodate the largest possible number of users.

4.3 New lower bounds

In this section, we give several constructions for t-TS(w,v). When $w \leq t^2$, in Section 4.3.1, we show optimal t-TS(w,v) from sunflowers based on the new upper bounds in the preceding section. When $w > t^2$, in Section 4.3.2, constructions of TS by means of combinatorial designs are provided, which can produce many infinite families of optimal TS achieving our new upper bounds. For the general TS, in Section 4.3.3, a constructive lower bound for general TS is given, which has the same order of magnitude with the new upper bounds.

4.3.1 When $w \leq t^2$: Constructions from sunflowers

Theorem 4.3.1 *For any $v \geq w \geq 2,\ t \geq 2$, we have*

$$M_t(w,v) \geq v - w + 1.$$

Proof: We prove this theorem by providing a construction by means of sunflowers. Suppose $\mathcal{X}$ is a v-set of points. Arbitrarily choose a $(w-1)$-subset $\Delta \subseteq \mathcal{X}$. Define

$$B_j := (\{j\} \cup \Delta) \subseteq \mathcal{X}, \text{ for } \forall j \in \mathcal{X} \setminus \Delta,$$

and denote $\mathcal{B} := \{B_j : j \in \mathcal{X} \setminus \Delta\}$. Clearly, $(\mathcal{X}, \mathcal{B})$ is a sunflower according to Definition 2.1.4. (The case $\Delta = \{v-w+2, \ldots, v\}$ is depicted in Figure 4.7.)

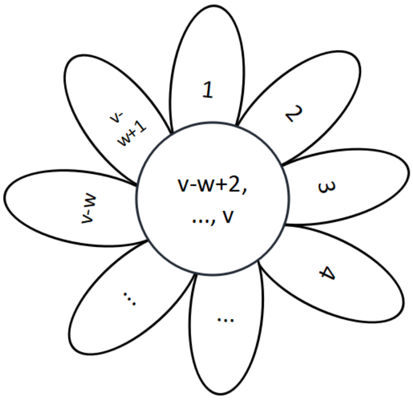

Figure 4.7: A construction of t-TS(w,v)

Then $(\mathcal{X}, \mathcal{B})$ is a t-TS(w,v) for any $t \geq 2$, since besides the common subset Δ, each block possesses a unique point. So

$$M_t(w,v) \geq |\mathcal{B}| = v - (w-1) = v - w + 1,$$

as desired. □

We have the following result for the case $w \leq t^2$.

Corollary 4.3.2 *For any $v \geq w \geq 2,\ t \geq 2$ with $w \leq t^2$, we have*

$$M_t(w,v) = v - w + 1.$$

Proof: From Theorem 4.2.7, we have $M_t(w,v) \leq v - w + 1$ for $w \leq t^2$. Then the corollary follows from Theorem 4.3.1. □

Note that the size of t-TS(w,v) we obtained in Theorem 4.3.1 is far from the upper bound in Theorem 4.2.7 for $w > t^2$. So we explore other constructions in the next subsection.

4.3.2 When $w > t^2$: Constructions from combinatorial designs

Combinatorial structures are often used to construct various configurations in coding theory. In this subsection, we use combinatorial designs to construct t-TS(w,v).

In [85], Stinson and Wei used τ-$(v,w,1)$ design to construct t-TS(w,v) as follows.

Theorem 4.3.3 ([85]) *If there exists a τ-$(v,w,1)$ design, then there exists a t-TS(w,v) with cardinality $\binom{v}{\tau}/\binom{w}{\tau}$, where $t = \lfloor\sqrt{(w-1)/(\tau-1)}\rfloor$.*

We provide a generalized construction as follows.

Theorem 4.3.4 *Let $w \equiv d+1 \pmod{t^2}$ where $0 \leq d \leq t^2-1$ and let $\tau = \lceil w/t^2 \rceil$. If there exists a τ-$(v-d,w-d,1)$ design, then there exists a t-TS(w,v) with cardinality $\binom{v-d}{\tau}/\binom{w-d}{\tau}$.*

Proof: Let $\mathcal{X}$ be a v-set of points. Suppose there exists a τ-$(v-d,w-d,1)$ design $(\mathcal{X}_0,\mathcal{B}_0)$, where $\mathcal{X}_0 \subseteq \mathcal{X}$, $|\mathcal{X}\setminus\mathcal{X}_0| = d$. Extending each block $B_0 \in \mathcal{B}_0$ to the d points of $\mathcal{X}\setminus\mathcal{X}_0$ to obtain

$$\mathcal{B} = \{B_0 \cup (\mathcal{X}\setminus\mathcal{X}_0) : B_0 \in \mathcal{B}_0\}.$$

Clearly, $|\mathcal{B}| = |\mathcal{B}_0| = \binom{v-d}{\tau}/\binom{w-d}{\tau}$, and $|B| = w$ for any $B \in \mathcal{B}$. We prove that $(\mathcal{X},\mathcal{B})$ is a t-TS(w,v).

Suppose $B_1,\ldots,B_s \in \mathcal{B}$, $2 \leq s \leq t$, are s distinct blocks and $B_{s+1} \in \mathcal{B}\setminus\{B_i : 1 \leq i \leq s\}$ is any other block. For any $B \subseteq \cup_{1\leq i\leq s} B_i$ such that $|B| = w$, we want to show that

$$|B \cap B_{s+1}| < \max\{|B \cap B_i| : 1 \leq i \leq s\}. \tag{4.8}$$

For $1 \leq i \leq s+1$, denote

$$B_i' = B_i \cap \mathcal{X}_0 \in \mathcal{B}_0.$$

Equivalently, B_i' is the corresponding original block in $\mathcal{B}_0$ which is extended to B_i. Denote $B' = B \cap \mathcal{X}_0$. Clearly,

$$B' \subseteq \bigcup_{1\leq i\leq s} B_i'. \tag{4.9}$$

Note that $|B'| \geq w - d$ rather than $|B'| = w - d$, so more detailed analyses are required. Assume that

$$|B \setminus B'| = |B \cap (\mathcal{X} \setminus \mathcal{X}_0)| = \delta,\ 0 \leq \delta \leq d.$$

Then $|B'| = w - \delta$. On the one hand, applying the pigeonhole principle to (4.9), we have

$$\max\{|B' \cap B'_i| : 1 \leq i \leq s\} \geq \lceil \frac{w-\delta}{s} \rceil \geq \lceil \frac{w-d}{t} \rceil = t(\lceil \frac{w}{t^2} \rceil - 1) + 1.$$

Since $\mathcal{X} \setminus \mathcal{X}_0$ is contained in each block $B \in \mathcal{B}$, we have $|B \cap B_i| = |B' \cap B'_i| + \delta$. Moreover,

$$\max\{|B \cap B_i| : 1 \leq i \leq s\} \geq t(\lceil \frac{w}{t^2} \rceil - 1) + \delta + 1. \tag{4.10}$$

On the other hand, since $(\mathcal{X}_0, \mathcal{B}_0)$ is a τ-$(v-d, w-d, 1)$ design, we have

$$\begin{aligned}|B' \cap B'_{s+1}| &\leq |(\bigcup_{1\leq i\leq s} B'_i) \cap B'_{s+1}| \\ &= |\bigcup_{1\leq i\leq s} (B'_i \cap B'_{s+1})| \\ &\leq \sum_{1\leq i\leq s} |B'_i \cap B'_{s+1}| \\ &\leq s(\tau - 1) \\ &\leq t(\lceil \frac{w}{t^2} \rceil - 1).\end{aligned}$$

Thus

$$|B \cap B_{s+1}| = |B' \cap B'_{s+1}| + \delta \leq t(\lceil \frac{w}{t^2} \rceil - 1) + \delta. \tag{4.11}$$

Hence, (4.8) can be obtained from (4.10) and (4.11). With the definition, $(\mathcal{X}, \mathcal{B})$ is a t-TS(w, v) and the theorem follows. □

As can be seen from Theorem 4.2.14 and Theorem 4.3.4, when all the parameters v, w, t, d, τ satisfy the conditions therein, the t-TS(w, v) constructed from τ-$(v-d, w-d, 1)$ design is optimal. From this point of view, the existence of τ-$(v, w, 1)$ design is crucial to the existence of optimal t-TS. We list several infinite families of optimal t-TS(w, v) as follows.

Theorem 4.3.5 *Let t be a prime power and d be an integer such that $0 \leq d < \lceil 2t^2/3 \rceil$. There exists an optimal t-TS$(t^2 + d + 1,\ t^{2n} + t^{2(n-1)} + \cdots + t^2 + d + 1)$ provided $n \geq 2 + \min\{2, d\}$.*

Proof: A 2-$(q^n + \cdots + q + 1,\ q + 1,\ 1)$ design exists whenever q is a prime power and $n \geq 2$ [28]. From Theorem 4.3.4, assuming $q = t^2$, for $0 \leq d \leq t^2 - 1$, there exists a t-TS$(t^2 + d + 1,\ t^{2n} + t^{2(n-1)} + \cdots + t^2 + d + 1)$. Noting that $\lceil \frac{t^2+d+1}{t^2} \rceil = 2$, if $0 \leq d < \lceil 2t^2/3 \rceil$, then this satisfies the condition of case (c) in Theorem 4.2.14. It follows that the above t-TS$(t^2+d+1,\ t^{2n}+t^{2(n-1)}+\cdots+t^2+d+1)$ is optimal when

$$t^{2n}+t^{2(n-1)}+\cdots+t^2+d+1 > 2d(t^2+d)(t^2+d+1),$$

which holds when $n \geq 2+\min\{2,d\}$. □

Theorem 4.3.6 *Let $t \geq 2$ be an integer such that t^2+1 is a prime power. Let d be an integer such that $0 \leq d < \lceil 2t^2/3 \rceil$. There exists an optimal t-TS$(t^2+d+1,\ (t^2+1)^n+d)$ provided $n \geq 2+\min\{2,d\}$.*

Proof: A 2-$(q^n,q,1)$ design exists whenever q is a prime power and $n \geq 2$ [28]. From Theorem 4.3.4, assuming that $q = t^2+1$ is a prime power, for $0 \leq d \leq t^2-1$, there exists a t-TS$(t^2+d+1,\ (t^2+1)^n+d)$. Noting that $\lceil \frac{t^2+d+1}{t^2} \rceil = 2$, if $0 \leq d < \lceil 2t^2/3 \rceil$, then this satisfies the condition of case (c) in Theorem 4.2.14. It follows that the above t-TS$(t^2+d+1,\ (t^2+1)^n+d)$ is optimal when

$$(t^2+1)^n+d > 2d(t^2+d)(t^2+d+1),$$

which holds when $n \geq 2+\min\{2,d\}$. □

Theorem 4.3.7 *Let t be a positive integer power of 2, and d be an integer such that $d \in \{0,1\}$ or $0 \leq d < t^2/18$. There exists an optimal t-TS$(2t^2+d+1,\ 2^nt^{2n}+d+1)$ provided $n \geq 2+2\min\{1,d\}$.*

Proof: A 3-$(q^n+1,q+1,1)$ design exists whenever q is a prime power and $n \geq 2$ [28]. From Theorem 4.3.4, assuming $q = 2t^2$ (a power of 2), for $0 \leq d \leq t^2-1$, there exists a t-TS$(2t^2+d+1,\ 2^nt^{2n}+d+1)$. Note that if $d \in \{0,1\}$ or $0 \leq d < t^2/18$, then this satisfies the condition of case (a) or (b) in Theorem 4.2.14. It follows that the above t-TS$(2t^2+d+1,\ 2^nt^{2n}+d+1)$ is optimal when

$$2^nt^{2n}+d+1 > d(2t^2+d+1)(2t^2+d)(2t^2+d-1),$$

which holds when $n \geq 2+2\min\{1,d\}$. □

Theorem 4.3.8 *Let t be a prime power and d be an integer such that $0 \leq d \leq t^2/4$. There exists an optimal t-TS$(t^2+d+1,\ t^6+d+1)$.*

Proof: A 2-$(q^3+1,q+1,1)$ design exists whenever q is a prime power [28]. From Theorem 4.3.4, assuming $q = t^2$, for $0 \leq d \leq t^2-1$, there exists a t-TS$(t^2+d+1,\ t^6+d+1)$. Noting that $\lceil \frac{t^2+d+1}{t^2} \rceil = 2$, if $0 \leq d < \lceil 2t^2/3 \rceil$, then this satisfies the condition of case (c) in Theorem 4.2.14. It follows that the above t-TS$(t^2+d+1,\ t^6+d+1)$ is optimal when

$$t^6+d+1 > 2d(t^2+d)(t^2+d+1),$$

which holds when $0 \leq d \leq t^2/4$. □

Theorem 4.3.9 *Let v be an integer such that $v \equiv 1, 5 \pmod{20}$, and $d \in \{0, 1, 2\}$. There exists an optimal 2-TS$(5 + d, v + d)$ with size $v(v - 1)/20$, provided $v > 2d^3 + 18d^2 + 39d$.*

Proof: A 2-$(v, 5, 1)$ design exists whenever $v \equiv 1, 5 \pmod{20}$ [28]. By Theorem 4.3.4, for $0 \leq d \leq 3$, there exists a 2-TS$(5 + d, v + d)$ with size $v(v - 1)/20$. Noting that $\lceil \frac{5+d}{2^2} \rceil = 2$, if $0 \leq d < \lceil 2t^2/3 \rceil = 3$, then this satisfies the condition of case (c) in Theorem 4.2.14. It follows that the above 2-TS$(5 + d, v + d)$ is optimal when $v > 2d^3 + 18d^2 + 39d$. □

In [85], Stinson and Wei gave constructions for t-TS via 2-$(q^2 + q + 1, q + 1, 1)$ design and 3-$(q^2 + 1, q + 1, 1)$ design, which are special cases of Theorem 4.3.5 and Theorem 4.3.7. Particularly, in the above theorems, we show that their constructions and our generalized constructions can produce infinite families of optimal t-TS.

By Theorem 2.3.6, we also have

Theorem 4.3.10 *Let $w \equiv d + 1 \pmod{t^2}$ where $0 \leq d \leq t^2 - 1$ and let $\tau = \lceil w/t^2 \rceil$. Then for any sufficiently large v such that*

$$\binom{v-d-i}{\tau-i} \equiv 0 \ (\mathrm{mod} \binom{w-d-i}{\tau-i}), \quad \forall\ 0 \leq i \leq \tau - 1, \tag{4.12}$$

there exists an optimal t-TS(w, v) in the following cases:

$$\text{(a) } d = 0, 1, \quad \text{(b) } d < t^2/(2\tau^2), \quad \text{(c) } \tau = 2 \text{ and } d < \lceil 2t^2/3 \rceil.$$

Proof: From Theorem 2.3.6, for sufficiently large v, there exists a τ-$(v - d, w - d, 1)$ design if the condition (4.12) holds. Then there exists a t-TS(w, v) with cardinality $\binom{v-d}{\tau}/\binom{w-d}{\tau}$, which follows from Theorem 4.3.4. It is optimal for the cases in Theorem 4.2.14. □

In the next subsection, we provide a constructive lower bound for general t-TS(w, v).

4.3.3 A general constructive lower bound

In [85], Stinson and Wei proposed to use another type of combinatorial designs, packings, to construct TS.

Lemma 4.3.11 ([85]) *If there exists a $\lceil w/t^2 \rceil$-$(v, w, 1)$ packing, then there exists a t-TS(w, v).*

A similar generalization with Theorem 4.3.4 is the following.

Theorem 4.3.12 *Let $w \equiv d + 1 \pmod{t^2}$ where $0 \leq d \leq t^2 - 1$. If there exists a $\lceil w/t^2 \rceil$-$(v - d, w - d, 1)$ packing, then there exists a t-TS(w, v).*

Considering the substance in the above theorem, we have the following constructive lower bound for general t-TS.

Theorem 4.3.13 *Let $w \equiv d + 1 \pmod{t^2}$ where $0 \le d \le t^2 - 1$. For any $v \ge w$, we have*

$$M_t(w, v) \ge \binom{v-d}{\lceil w/t^2 \rceil} \Big/ \binom{w-d}{\lceil w/t^2 \rceil}^2.$$

Proof: First, let $v' = v - d$, $w' = w - d$. Suppose $\mathcal{X}$ is a finite set of v points, and $\mathcal{X}_0$ is a d-subset of $\mathcal{X}$. Denote $\mathcal{X}' = \mathcal{X} \setminus \mathcal{X}_0$. We use the following Algorithm 1.

Clearly, for each i, we have

$$|\mathcal{D}_i \setminus \mathcal{D}_{i-1}| \le \binom{w'}{\lceil w/t^2 \rceil} \binom{v' - \lceil w/t^2 \rceil}{w' - \lceil w/t^2 \rceil}.$$

Hence

$$|\mathcal{B}| \ge \frac{\binom{v'}{w'}}{\binom{w'}{\lceil w/t^2 \rceil} \binom{v' - \lceil w/t^2 \rceil}{w' - \lceil w/t^2 \rceil}} = \binom{v'}{\lceil w/t^2 \rceil} \Big/ \binom{w'}{\lceil w/t^2 \rceil}^2.$$

Now, it is sufficient to prove that the collection of blocks $\mathcal{B}$ generated by Algorithm 1 is a t-TS(w, v). This follows from Theorem 4.3.12, since for the output $\mathcal{B}$ of Algorithm 1, the set system $(\mathcal{X}', \mathcal{B}')$, where $\mathcal{B}' = \{B \setminus \mathcal{X}_0 : B \in \mathcal{B}\}$, is a $\lceil w/t^2 \rceil$-$(v - d, w - d, 1)$ packing.

Consequently, we have

$$M_t(w, v) \ge \binom{v-d}{\lceil w/t^2 \rceil} \Big/ \binom{w-d}{\lceil w/t^2 \rceil}^2,$$

as desired. □

Moreover, we have

Theorem 4.3.14 *Let $w \equiv d + 1 \pmod{t^2}$ where $0 \le d \le t^2 - 1$. For any sufficiently large v, we have*

$$M_t(w, v) \ge (1 - o(1)) \binom{v-d}{\lceil w/t^2 \rceil} \Big/ \binom{w-d}{\lceil w/t^2 \rceil}.$$

Proof: This theorem follows from Theorem 4.3.12 and Theorem 2.3.7. □

We make remarks here that the constructive lower bound in Theorem 4.3.13 has the same order of magnitude as our general upper bound in Theorem 4.2.7, and the lower bound in Theorem 4.3.14 is very close to our upper bound in Theorem 4.2.14.

Our constructions of t-TS provide the way of distributing base keys for the data supplier to resist collusion attacks from traitors. The infinite families of optimal t-TS show that there exist TS-based traitor tracing schemes which can accommodate the largest possible number of users.

4.4 Summary

In this chapter, we investigated the traceability schemes (TS). We found a very interesting relationship between TS and CFF in Section 4.2.1, that is, a t-TS is a t^2-CFF. This relationship has a significant meaning for TS in both the combinatorial viewpoint and its applications in broadcast encryption. Based on this new relationship, we derived new upper bounds on the size of TS for general cases in Section 4.2.2 and for several special cases in Section 4.2.2, respectively. Our new bounds greatly improve the previously known upper bounds in [85] and [29]. By means of sunflowers, we constructed optimal t-TS(w, v) for the case $w \leq t^2$ in Section 4.3.1. By means of combinatorial designs, we constructed several infinite families of optimal t-TS(w, v) for the case $w > t^2$ in Section 4.3.2, A constructive lower bound for general TS, based on combinatorial packings, was given in Section 4.3.3, which has the same order of magnitude with the new upper bounds.

In Section 4.3, we have already shown that our new upper bounds in Theorem 4.2.7 and Theorem 4.2.14 can be achieved in many cases. It would be of interest to determine whether the new upper bounds are tight or not for other parameters. If they are tight, then constructions for t-TS achieving the upper bounds are also expected.

Chapter 5

Parent-Identifying Set Systems

In this chapter, we concentrate on the parent-identifying set systems (IPPSs). First, in Section 5.1, we propose a unified concept of parent-identifying schemes for many combinatorial structures with the parent-identifying property, such as parent-identifying codes for digital fingerprinting and single-user tracing superimposed families for group testing. An equivalent relationship between parent-identifying schemes and forbidden configurations is established, which is extremely useful to analyze and derive bounds for various parent-identifying schemes. Moreover, we correspond the research problems of parent-identifying schemes to a kind of Turán-type problems. In Section 5.2, some remarks on the definition of IPPS and the related known results of IPPS are stated. In Section 5.3, we prove a new upper bound for t-IPPS by virtue of techniques in extremal set theory, which greatly improve the previously known bound. In Section 5.4, we derive a lower bound for t-IPPS with the probabilistic methods, which has the same order of magnitude with the new upper bound in Section 5.3. In Section 5.5, we analyze the structure of 2-IPPS, and an equivalent combinatorial description of 2-IPPS is provided. We also derive better upper bounds for 2-IPPS$(4, v)$ and 3-IPPS$(6, v)$, respectively, by means of the well-known graph removal lemma in extremal graph theory. Finally, we summarize this chapter in Section 5.6. This chapter is based on results in [50, 51].

5.1 Forbidden configurations

In this section, we unify the combinatorial structures with the parent-identifying property by a concept of parent-identifying schemes (Definition 5.1.1). An equivalent relationship between parent-identifying schemes and forbidden configurations is established (Theorem 5.1.5), which is extremely useful to analyze and derive bounds for various parent-identifying schemes.

5.1.1 A unified concept of parent-identifying schemes

As Pearl wrote in the preface of his book [69], the central aim of many studies in the physical, behavioral, social, and biological sciences is the elucidation of cause-effect relationships among variables or events. Parent-identifying scheme provides a way to identify causes from an effect for some information systems such as digital fingerprinting and group testing. We use the "f-channel" to unify different functional mappings from causes to effects, see Figure 5.1.

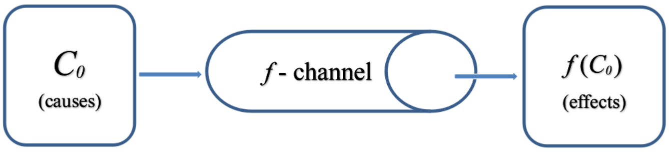

Figure 5.1: f-channel

Definition 5.1.1 *Let Q be a finite set of possible causes. A subset $\mathcal{C}$ of Q^n or 2^Q is a t-parent-identifying scheme under f-channel if for any $\mathcal{C}' \subseteq \mathcal{C}$ with $|\mathcal{C}'| \leq t$ and any $d \in f(\mathcal{C}')$, we have*

$$\bigcap_{\mathcal{P} \in P_t(d)} \mathcal{P} \neq \emptyset,$$

where

$$P_t(d) = \{\mathcal{P} \subseteq \mathcal{C} : \ |\mathcal{P}| \leq t, \, d \in f(\mathcal{P})\}.$$

The above definition exhibits the essential idea of cause clarification algorithms in parent-identifying schemes, that is, once an effect d is observed, one needs to check each subset $\mathcal{P} \subseteq \mathcal{C}$ with size at most t to see if d can be generated by $\mathcal{P}$. If d can be generated by $\mathcal{P}$, then we call d a *descendant* of $\mathcal{P}$ and call $\mathcal{P}$ a *possible parent set* of d. The intersection of all possible parent sets has to be a subset of true causes for d.

We remark that Anthapadmanabhan, Barg and Dumer [7] also investigated a t-input-single-output channel, which is similar to the above f-channel, under the marking assumption for digital fingerprinting from an information-theoretic viewpoint. The interested reader is referred to [7]. In this thesis, we are primarily concerned with the combinatorial properties of the f-channel in Definition 5.1.1.

In Definition 5.1.1, if $\mathcal{C} \subset Q^n$, then it can be regarded as a class of fingerprinting codes. Well-known examples include *codes with the t-identifiable parent property* (t-IPP codes) [4, 6, 8, 9, 10, 13, 53, 82, 90], where

$$f(\mathcal{P}) = \mathcal{P}(1) \times \mathcal{P}(2) \times \cdots \times \mathcal{P}(n),$$

and $\mathcal{P}(i) = \{\mathbf{c}(i) \in Q : \ \mathbf{c} = (\mathbf{c}(1), \ldots, \mathbf{c}(n)) \in \mathcal{P}\}$.

However, it is not necessarily always the case. In some scenario, $\mathcal{C}$ may be a collection of subsets of Q, that is, $\mathcal{C} \subseteq 2^Q$, which forms a set system. A typical example is the *t-single-user tracing superimposed family*, introduced for applications in molecular biology [31, 1], where

$$f(\mathcal{P}) = \{\bigcup_{A \in \mathcal{P}} A\}.$$

In this dissertation, we investigate two parent-identifying schemes, according to their f-channels, in different applications, namely, IPPS for broadcast encryption (this chapter) and multimedia parent-identifying codes (MIPPC) for multimedia fingerprinting (Chapter 7). We focus on the size (or in other words, code rate), one of the most important parameters, of such parent-identifying schemes.

In the literature, Alon *et al.* [6] proved that the asymptotically optimal code rate of t-IPP code is $\frac{1}{\lfloor t^2/4 \rfloor + t}$. Csúrös *et al.* [31] and Alon *et al.* [1] showed that the code rate of a t-single-user tracing superimposed family is $\Theta(\frac{1}{t})$. However, there are no general lower bounds for t-IPPS and t-MIPPC in the literature. In this dissertation, we will use the probabilistic expurgation method to derive lower bounds for t-IPPS (Section 5.4) and t-MIPPC (Section 7.2), and compare them with the best known upper bounds.

To this end, in Section 5.1.2, we first establish an equivalent relationship between parent-identifying schemes and forbidden configurations.

5.1.2 A kind of Turán-type problems

As wrote by Füredi in [44], one of the most important problems in every branch of mathematics is the description of structure of its objects. In this subsection, we provide an equivalent description of parent-identifying schemes by using their forbidden configurations.

In extremal combinatorics, a class of extremal problems known as *Turán-type* problems, named for Pál Turán [92], consists of problems that aim to maximize the size of a certain object while avoiding a series of forbidden substructures. Turán-type problems are often very difficult and very little is known even about some simple cases. Recent surveys on Turán-type problems are referred to [44] and [61]. In the following, we correspond the research problems of parent-identifying schemes to a kind of Turán-type problems.

In [9], Barg *et al.* exploited the notion of *minimal forbidden configuration* to study IPP codes. We recap the related notations here.

Definition 5.1.2 *In a set $\mathcal{C}$, let $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ be a collection of subsets of $\mathcal{C}$ with $\mathcal{F}_i \subseteq \mathcal{C}$, $|\mathcal{F}_i| \leq t$, $i = 1, \ldots, m$. Then $\mathcal{F}$ is called a configuration if it has an empty intersection, $\cap_{1\leq i\leq m}\mathcal{F}_i = \emptyset$. Moreover, $\mathcal{F}$ is called a minimal configuration if it is minimal under inclusion, that is,*

$$\bigcap_{\substack{1\leq j\leq m,\\ j\neq i}} \mathcal{F}_j \neq \emptyset, \quad \forall 1 \leq i \leq m.$$

Denote $U(\mathcal{F}) = \cup_{1\leq i\leq m}\mathcal{F}_i$. The cardinality of $U(\mathcal{F})$ is called the *size* of the configuration $\mathcal{F}$. In [9], Barg *et al.* proved the following lemma for the size of a minimal configuration, which was also shown by Staddon *et al.* in [82] and Alon *et al.* in [6]. Denote $u := \lfloor (\frac{t}{2}+1)^2 \rfloor$. To be self-contained, we expose the proof in the following.

Lemma 5.1.3 ([6], [9], [82]) *Let $\mathcal{F}$ be a minimal configuration. Then $|U(\mathcal{F})| \leq u$.*

Proof: Suppose $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ be a minimal configuration. Then for each $1 \leq i \leq m$, there exists a codeword $\mathbf{x}_i$ such that

$$\mathbf{x}_i \notin \mathcal{F}_i \ \text{ and } \ \mathbf{x}_i \in \bigcap_{\substack{1 \leq j \leq m, \\ j \neq i}} \mathcal{F}_j.$$

Clearly, $\mathbf{x}_i \neq \mathbf{x}_j$ for all $1 \leq i < j \leq m$. Therefore

$$\begin{aligned} |U(\mathcal{F})| &= |\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}| + |\bigcup_{1 \leq i \leq m} \mathcal{F}_i \setminus \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}| \\ &\leq m + \sum_{1 \leq i \leq m} (|\mathcal{F}_i| - (m-1)) \\ &\leq m + m(t - m + 1) \\ &= -m^2 + (t+2)m \\ &\leq (\frac{t}{2}+1)^2. \end{aligned}$$

where the last inequality holds by taking $m = \frac{t}{2} + 1$. The lemma follows. □

Now we define the forbidden configuration.

Definition 5.1.4 *Let $\mathcal{C}$ be a t-parent-identifying scheme under f-channel. A (minimal) forbidden configuration in $\mathcal{C}$ is a (minimal) configuration $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ in $\mathcal{C}$ such that*

$$f(\mathcal{F}_1) \cap f(\mathcal{F}_2) \cap \cdots \cap f(\mathcal{F}_m) \neq \emptyset.$$

Then we have the following relationship.

Theorem 5.1.5 *A set $\mathcal{C}$ is not a t-parent-identifying scheme under f-channel if and only if there exists a minimal forbidden configuration in $\mathcal{C}$ with size at most u.*

Proof: The sufficiency follows from Definition 5.1.1 and Definition 5.1.4 directly. We focus on the necessity. If $\mathcal{C}$ is not a t-parent-identifying scheme under f-channel, then there exists a subset $\mathcal{C}' \subseteq \mathcal{C}$ and a descendant $d \in f(\mathcal{C}')$ such that

$$\bigcap_{\mathcal{P} \in P_t(d)} \mathcal{P} = \emptyset.$$

Then we can find a minimal forbidden configuration in

$$P_t(d) = \{\mathcal{P} \subseteq \mathcal{C} : \ |\mathcal{P}| \leq t, \, d \in f(\mathcal{P})\}.$$

This is doable since we can consecutively remove some $\mathcal{P}$ from $P_t(d)$, if the intersection of remaining subsets is still empty, until it forms a minimal configuration. The size of the minimal configuration, which is at most u, follows from Lemma 5.1.3. The proof is completed. $\square$

By Theorem 5.1.5, we have the following Turán-type problems.

Fact 5.1.6 *The research problem for parent-identifying schemes is to maximize the size $|\mathcal{C}|$ while avoiding any $\mathcal{C}_0 \subseteq \mathcal{C}$ such that $|\mathcal{C}_0| \leq u$ and $\mathcal{C}_0$ forms a minimal forbidden configuration as in Definition 5.1.2.*

Different f-channels lead to different parent-identifying schemes. By Theorem 5.1.5, different parent-identifying schemes correspond to different minimal forbidden configurations. To derive a lower bound for the size of a parent-identifying scheme, we can use the probabilistic expurgation method, that is, to estimate the expectation number of its corresponding minimal forbidden configurations, and then delete one element from each of them. We will derive probabilistic existence lower bounds for IPPS (Section 5.4) and MIPPC (Section 7.2) in this way.

5.2 Definition and known results

5.2.1 Definition

The definition of IPPS was provided in Definition 3.1.1 (2). Here we give some remarks for that definition. First, we note that the IPPS defined in Definition 3.1.1(2) requires that condition $|T| = w$. Modifying it to all the case $|T| \geq w$, we have the following definition of t-IPPS*.

Definition 5.2.1 *A t-parent-identifying* set system, denoted t-IPPS$^\star(w, v)$, is a set system $(\mathcal{X}, \mathcal{B})$ such that $\mathcal{B} \subseteq \binom{\mathcal{X}}{w}$ and $|\mathcal{X}| = v$, with the property that for any $T \subseteq \mathcal{X}$ such that $|T| \geq w$, either $P_t(T)$ is empty, or*

$$\bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P} \neq \emptyset,$$

where

$$P_t(T) := \{\mathcal{P} \subseteq \mathcal{B} : |\mathcal{P}| \leq t,\ T \subseteq \bigcup_{B \in \mathcal{P}} B\}.$$

Considering the relationship between IPPS and IPPS*, we have the following lemma.

Lemma 5.2.2 *A set system $(\mathcal{X}, \mathcal{B})$ is a t-IPPS(w, v) if and only if it is a t-IPPS$^\star(w, v)$.*

Proof: The sufficiency directly follows from their definitions. We focus on the necessity. Suppose $(\mathcal{X}, \mathcal{B})$ is a t-IPPS(w, v), we would like to show that it is also a t-IPPS$^\star(w, v)$. Consider any $T \subseteq \mathcal{X}$ with $|T| \geq w$ and $P_t(T) \neq \emptyset$. Choosing a w-subset $T_0 \subseteq T$, we have

$$P_t(T) \subseteq P_t(T_0),$$

since for any $\mathcal{P} \in P_t(T)$, we have $T \subseteq \bigcup_{B \in \mathcal{P}} B$ and then $T_0 \subseteq T \subseteq \bigcup_{B \in \mathcal{P}} B$, which implies that $\mathcal{P} \in P_t(T_0)$. By the definition of t-IPPS(w, v), we have

$$\bigcap_{\mathcal{P} \in P_t(T_0)} \mathcal{P} \neq \emptyset.$$

Hence we have

$$\emptyset \neq \bigcap_{\mathcal{P} \in P_t(T_0)} \mathcal{P} \subseteq \bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P}.$$

Thus $(\mathcal{X}, \mathcal{B})$ is a t-IPPS$^\star(w, v)$ by Definition 5.2.1, and the necessity follows. □

5.2.2 Known results

By investigating $\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil$-own-subsets, Collins gave an upper bound for t-IPPS(w, v) as follows.

Theorem 5.2.3 ([29]) *Let $v \geq w \geq 2,\ t \geq 2$ be integers. Then*

$$I_t(w, v) \leq \frac{\binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil}}{\binom{\lceil \frac{w}{\lfloor t/2 \rfloor + 1} \rceil - 1}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil - 1}} = O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil}).$$

In Chapter 4, to improve the known upper bounds for t-TS, we first proved that a t-TS is a t^2-CFF in Section 4.2 and then derived new upper bounds based on this new relationship. For t-IPPS, it was shown a t-IPPS is a t-CFF in Lemma 3.2.1. A natural question is that: is it possible that a t-IPPS is a t'-CFF, where $t' > t$? In the following, we give a negative answer by Example 5.2.4.

Example 5.2.4 *Let $v > w > t > 1$ be positive integers such that $v = w(t+1)$, $w = r(t+1)$. Let $\mathcal{X}$ be a finite set of v points. Considering a family of w-subsets $\mathcal{F} = \{F_0, F_1, \ldots, F_{t+1}\} \subseteq \binom{\mathcal{X}}{w}$ such that*

(c1) $F_i \cap F_j = \emptyset, \quad \forall\, i \neq j \in \{1, \ldots, t+1\}$,

(c2) $|F_i \cap F_0| = r, \quad \forall\, i \in \{1, \ldots, t+1\}$.

Now we claim that $\mathcal{F}$ is a t-IPPS but not a $(t+1)$-CFF.

On the one hand, $\mathcal{F}$ is a t-IPPS. That is, for any w-subset of $\mathcal{X}$ which belongs to the union of some t blocks in $\mathcal{F}$, at least one of these sets can be uniquely determined. Indeed, let T be a w-subset of $\mathcal{X}$ generated by some coalition $U \subseteq \mathcal{F}$.

- *If there exists an element $x_0 \in T$ such that $x_0 \notin F_0$, then by property (c1), this x_0 belongs to the unique defined set F_j, $1 \leq j \leq t+1$. This F_j must appear in each parent set of T.*
- *Otherwise, if there is no such $x_0 \in T$, then $T = F_0$. Therefore any coalition which can generate T has to contain F_0 since the union of any t blocks of $\{F_1, \ldots, F_{t+1}\}$ can not cover F_0 by the lack of cardinality, see property (c2).*

On the other hand, $\mathcal{F}$ is not a $(t+1)$-CFF, since F_0 can be covered by $F_1, \ldots, F_{t+1}$.

From the above example, we cannot improve the known upper bound for t-IPPS as that for t-TS, that is, by virtue of its relationship with t-CFF. In Section 5.3, we show our improvement on the upper bound for t-IPPS by investigating the own-subsets.

5.3 A new upper bound for t-IPPS

In this section, we prove a new upper bound for t-IPPS and compare it with the known results. We also analyze our new upper bound by virtue of Theorem 5.1.5.

Noting that the exponent of v in the upper bound of $N_t(w, v)$ in Theorem 5.2.3, that is $\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil$, is exactly the size of the own-subset investigated by Collins [29]. To improve the upper bound in Theorem 5.2.3, we should try to find an own-subset with smaller size possessed by some block.

In the following Theorem 5.3.1, we provide a better upper bound than Theorem 5.2.3, by showing that some block of a t-IPPS must contain at least one $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-own-subsets.

Theorem 5.3.1 *Let $v \geq w \geq 2,\ t \geq 2$ be integers. Then*

$$I_t(w, v) \leq \binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil} = O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil}).$$

Before proving this theorem, we need the following lemma.

Lemma 5.3.2 *Let $(\mathcal{X}, \mathcal{B})$ be a t-IPPS(w, v). There exists one block $B \in \mathcal{B}$ containing at least one $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-own-subset.*

Proof: Suppose on the contrary that each block $B \in \mathcal{B}$ does not contain any $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-own-subset. That is, for each block $B \in \mathcal{B}$ and each $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-subset $B_0 \subseteq B$, there exists another block $B' \in \mathcal{B} \setminus \{B\}$ such that $B_0 \subseteq B'$. Then we would like to derive a contradiction with the assumption that $(\mathcal{X}, \mathcal{B})$ is a t-IPPS(w, v). In other words, we try to find a collection of subsets $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_m$ of $\mathcal{B}$, each of which has size at most t, such that all $\mathcal{P}_i$ can generate the same descendant T, but their intersection is empty. To this end, we

- pick $m \leq \lfloor \frac{t}{2} \rfloor + 1$ blocks $B_1, \ldots, B_m$ from $\mathcal{B}$ to form $\mathcal{P}_0$, and construct a descendant T of $\mathcal{P}_0$;
- for each B_i, look for a subset $\mathcal{C}^{(i)} \subseteq \mathcal{B} \setminus \{B_i\}$ of at most $t - m + 1$ blocks to substitute it, forming $\mathcal{P}_i$ which has size at most t and is also a parent set of T.

In the following, we show the explicit process to realize this.

First, arbitrarily choose a block $B_1 \in \mathcal{B}$. Let $i = 1$ and $A_0 = B_0 = D_0 = \emptyset$. Next turn to execute a while loop.

While $1 \leq i \leq \lfloor \frac{t}{2} \rfloor$ and

$$|B_i \setminus (\cup_{0 \leq j \leq i-1}(A_j \cup D_j))| \geq \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil (\lceil \frac{t}{2} \rceil + 1),$$

take a subset $A_i \subseteq B_i \setminus (\cup_{0 \leq j \leq i-1}(A_j \cup D_j))$ such that $|A_i| = \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil \lceil \frac{t}{2} \rceil$. With the assumption, A_i can be covered by at most $\lceil \frac{t}{2} \rceil$ distinct blocks in $\mathcal{B}$ other than B_i. Denote $\mathcal{C}^{(i)} \subseteq \mathcal{B} \setminus \{B_i\}$ such that $|\mathcal{C}^{(i)}| \leq \lceil \frac{t}{2} \rceil$ and $A_i \subseteq \bigcup_{B \in \mathcal{C}^{(i)}} B$. Note that some B_j, $j \neq i$, may appear in $\mathcal{C}^{(i)}$. This is allowed since it does not increase the number of blocks that we are looking for to keep the size of $\mathcal{P}_i$ at most t. We have

$$B_i \setminus (\bigcup_{0 \leq j \leq i-1} (A_j \cup D_j) \cup A_i) \nsubseteq \bigcup_{0 \leq j \leq i-1} B_j,$$

since if not, B_i would be covered by at most t other blocks $\{B_j : 0 \leq j \leq i-1\} \cup \mathcal{C}^{(i)}$, which contradicts Lemma 3.2.1. This allows us to take another subset $D_i \subseteq B_i \setminus (\bigcup_{0 \leq j \leq i-1}(A_j \cup D_j) \cup A_i)$ such that $|D_i| = \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$ and

$$D_i \nsubseteq \bigcup_{0 \leq j \leq i-1} B_j.$$

With the assumption, there exists another block $B_{i+1} \in \mathcal{B} \setminus \{B_j : 1 \leq j \leq i\}$ such that $D_i \subseteq B_{i+1}$. Let $i = i + 1$ and continually execute the while loop of this paragraph.

The above while loop stops when one of the following two cases holds:

(a) $i = \lfloor \frac{t}{2} \rfloor + 1$;

(b) $i \leq \lfloor \frac{t}{2} \rfloor$ and $0 \leq |B_i \setminus (\cup_{0 \leq j \leq i-1}(A_j \cup D_j))| < \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil (\lceil \frac{t}{2} \rceil + 1)$.

Let $m = i$. Take $A_m \subseteq B_m \setminus (\cup_{0 \leq j \leq m-1}(A_j \cup D_j))$ such that

$$|A_m| = w - (m-1)\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil \lceil \frac{t}{2} \rceil - (m-1)\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil.$$

If $|A_m| = 0$, let $\mathcal{C}^{(m)} = \emptyset$. If not, we deal with the above two cases separately.

For case (a), $m = \lfloor \frac{t}{2} \rfloor + 1$, $0 < |A_m| \leq \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil \lceil \frac{t}{2} \rceil$. With the assumption, A_m can be covered by at most $\lceil \frac{t}{2} \rceil$ distinct blocks in $\mathcal{B}$ other than B_m. Denote $\mathcal{C}^{(m)} \subseteq \mathcal{B} \setminus \{B_m\}$ such that $|\mathcal{C}^{(m)}| \leq \lceil \frac{t}{2} \rceil$ and $A_m \subseteq \bigcup_{B \in \mathcal{C}^{(m)}} B$.

For case (b), $m \leq \lfloor \frac{t}{2} \rfloor$, $0 < |A_m| < \lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil (\lceil \frac{t}{2} \rceil + 1)$. With the assumption, A_m can be covered by at most $\lceil \frac{t}{2} \rceil + 1$ distinct blocks in $\mathcal{B}$ other than B_m. Denote $\mathcal{C}^{(m)} \subseteq \mathcal{B} \setminus \{B_m\}$ such that $|\mathcal{C}^{(m)}| \leq \lceil \frac{t}{2} \rceil + 1 \leq t - m + 1$ and $A_m \subseteq \bigcup_{B \in \mathcal{C}^{(m)}} B$.

Now, we have already taken $m \leq \lfloor \frac{t}{2} \rfloor + 1$ distinct blocks $B_1, \ldots, B_m \in \mathcal{B}$. Denote

$$T = \bigcup_{0 \leq j \leq m-1} (A_j \cup D_j) \cup A_m.$$

Clearly, $T \subseteq \mathcal{X}$ and $|T| = w$. Moreover,

$$T \subseteq \bigcup_{1 \leq j \leq m} B_j,$$

that is, $\mathcal{P}_0 := \{B_1, \ldots, B_m\} \in P_t(T)$.

On the other hand, for each $1 \leq i \leq m$, we have

$$T \subseteq (\bigcup_{B \in \mathcal{C}^{(i)}} B) \cup (\bigcup_{\substack{1 \leq j \leq m, \\ j \neq i}} B_j),$$

that is, each $\mathcal{P}_i := \mathcal{C}^{(i)} \cup \{B_j : 1 \leq j \leq m, j \neq i\} \subseteq \mathcal{B}$ is a parent set of T. Moreover, $|\mathcal{P}_i| \leq t$. That is,

$$\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_m\} \subseteq P_t(T).$$

However,

$$\bigcap_{0 \leq i \leq m} \mathcal{P}_i = \emptyset,$$

which implies

$$\bigcap_{\mathcal{P} \in P_t(T)} \mathcal{P} = \emptyset.$$

Hence $(\mathcal{X}, \mathcal{B})$ is not a t-IPPS(w, v), a contradiction to the assumption. Therefore the lemma follows. □

Proof of Theorem 5.3.1. Suppose $(\mathcal{X}, \mathcal{B})$ is a t-IPPS(w, v). By Lemma 5.3.2, there exists one block $B \in \mathcal{B}$ which contains at least one $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-own-subset. Delete this B from $\mathcal{B}$. The resulting $\mathcal{B} \setminus \{B\}$ is still a t-IPPS(w, v). Applying Lemma 5.3.2 repeatedly, we can successively delete blocks, which contain at least one $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-own-subset, from the newly obtained t-IPPS(w, v). Note that there are $\binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil}$ distinct $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-subsets from $\mathcal{X}$, and each $\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil$-subset, as own-subset of some block, can be deleted at most once. Hence

$$|\mathcal{B}| \leq \binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil},$$

and the theorem follows. □

In the following, we give examples to compare the new upper bound with the known result for t-IPPS.

Example 5.3.3 *Let $t = 2$ and $w = 7$. The following Figure 5.2 shows the improvements of the new bound for 2-IPPS$(7, v)$, where $20 \leq v \leq 50$.*

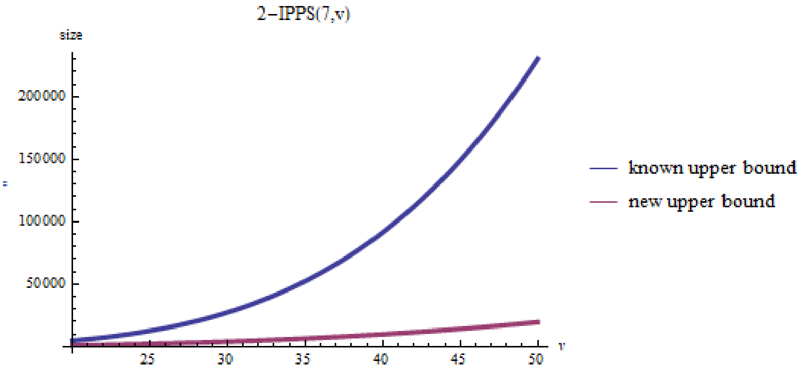

Figure 5.2: A comparison between the known and new upper bounds for 2-IPPS$(7, v)$

Example 5.3.4 *Let $t = 2$. A comparison between the known upper bound in Theorem 5.2.3 and our new upper bound in Theorem 5.3.1 for 2-IPPS(w, v) is shown in Figure 5.3, where $2 \leq w \leq 10$, $10 \leq v \leq 30$.*

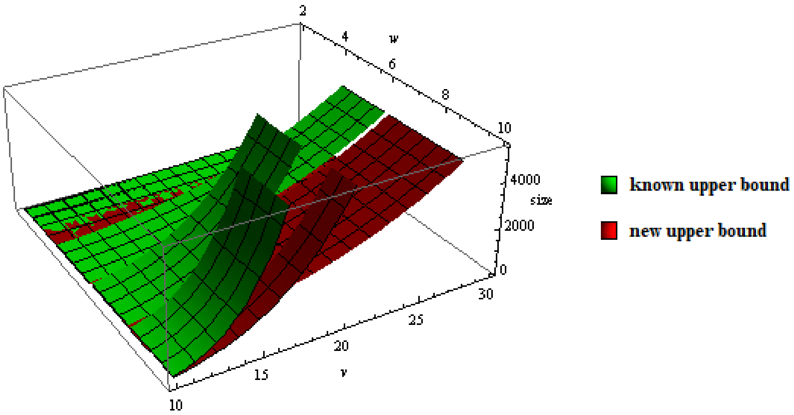

Figure 5.3: A comparison between the known and new upper bounds for 2-IPPS(w, v)

As can be seen, the known upper bound $O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + \lceil t/2 \rceil} \rceil})$ in Theorem 5.2.3 is improved to $O(v^{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil})$ in Theorem 5.3.1, which is a significant difference especially for fixed t, w and sufficiently large v.

Recall that in the proof of Lemma 5.3.2, we tried to construct a suitable T for which $\mathcal{P}_t(T)$ is a configuration. Note that the configuration $\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_m\}$, where $m \leq \lfloor \frac{t}{2} \rfloor + 1$, satisfies that

$$|\bigcup_{0 \leq i \leq m} \mathcal{P}_i| \leq m + \lceil \frac{t}{2} \rceil (m-1) + (t - m + 1) \leq \lfloor (\frac{t}{2} + 1)^2 \rfloor,$$

which implies that $\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_m\}$ is probably a minimal configuration by Lemma 5.1.3. Moreover, we have the following conjecture.

Conjecture 5.3.5 *Let $v \geq w \geq 2$, $t \geq 2$ be positive integers. The new upper bound $\binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil}$ is the exact upper bound for t-IPPS(w, v), up to a constant depending only on w and t.*

The above discussion shows that the anti-collusion key-distributing scheme based on a t-IPPS(w, v) can accommodate at most $\binom{v}{\lceil \frac{w}{\lfloor t^2/4 \rfloor + t} \rceil}$ users, which may be the exact maximum number of users it can hold, hopefully. In the next section, we provide a positive answer to this conjecture for certain cases.

5.4 A probabilistic lower bound for t-IPPS

In this section, we show a lower bound for t-IPPS with the probabilistic expurgation method.

First, according to Definition 5.1.4, we have the following description of the forbidden configuration in an IPPS. Let $(\mathcal{X}, \mathcal{B})$ be a t-IPPS(w, v). A *(minimal) forbidden configuration* $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ in $\mathcal{B}$ is a (minimal) configuration in $\mathcal{B}$ such that

$$|(\bigcup_{B \in \mathcal{F}_1} B) \cap (\bigcup_{B \in \mathcal{F}_2} B) \cap \cdots \cap (\bigcup_{B \in \mathcal{F}_m} B)| \geq w.$$

Therefore we have the following lemma.

Lemma 5.4.1 *In a t-IPPS(w, v), if a minimal forbidden configuration contains s blocks, where $2 \leq s \leq u$, then it is spanned by at most $(s-1)w$ points, that is,*

$$|\bigcup_{B \in U(\mathcal{F})} B| = |\bigcup_{1 \leq i \leq m} (\bigcup_{B \in \mathcal{F}_i} B)| \leq (s-1)w.$$

Proof: Suppose $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ is a minimal forbidden configuration with size $|U(\mathcal{F})| = s$. Then there exists a w-subset $W \subseteq \mathcal{X}$ such that

$$W \subseteq \bigcup_{B \in \mathcal{F}_i} B, \quad \forall\, 1 \leq i \leq m.$$

Thus each point in W appears in at least two distinct blocks of $U(\mathcal{F})$, since if not, it contradicts that $\mathcal{F}$ is a configuration. From $|U(\mathcal{F})| = s$, we know that $\mathcal{F}$ is spanned by at most sw points. Thus

$$|\bigcup_{B\in U(\mathcal{F})} B| = |W| + |\bigcup_{B\in U(\mathcal{F})} (B \setminus W)| \leq w + (s-2)w = (s-1)w,$$

as desired. □

Combining Theorem 5.1.5 and Lemma 5.4.1, we have the following corollary.

Corollary 5.4.2 *If a set system $(\mathcal{X}, \mathcal{B})$ is not a t-IPPS(w, v), then there exists an s-subset $\mathcal{U} \subseteq \mathcal{B}$, $2 \leq s \leq u$, such that $\mathcal{U}$ is spanned by at most $(s-1)w$ points.*

For convenience, we define the bad s-packet as follows.

Definition 5.4.3 *An s-subset $\mathcal{U} \subseteq \binom{\mathcal{X}}{w}$, $2 \leq s \leq u$, is called a bad s-packet if it is spanned by at most $(s-1)w$ points.*

Now we are going to prove the existence of good t-IPPS(w, v) for fixed w and sufficiently large v. From Lemma 5.4.1, we will show that bad s-packets are not typical and therefore their probability is small by using the technique similar to random coding. The process of deriving a lower bound for t-IPPS in Theorem 5.4.4 is to first randomly choose a family of blocks from $\binom{\mathcal{X}}{w}$, and then remove one block from each bad s-packet, $2 \leq s \leq u$, which may form a possible minimal forbidden configuration.

Theorem 5.4.4 *Let w and t be fixed positive integers such that $t \geq 2$. Then there exists a constant c, depending only on w and t, with the following property. For any sufficiently large integer v, there exists a t-IPPS(w, v) with size at least $cv^{\frac{w}{u-1}}$.*

Proof: Let $\mathcal{X}$ be a finite set of v points. Let $\binom{\mathcal{X}}{w}$ be the collection of all w-subsets (blocks) $B \subseteq \mathcal{X}$. Form a random subset $\mathcal{B} \subseteq \binom{\mathcal{X}}{w}$ of blocks by including each block independently with probability p, where $0 < p < 1$. We will determine the value of p later.

Let V denote the number of blocks in $\mathcal{B}$. Clearly, $E(V) = \binom{v}{w}p$.

Let X denote the number of all bad s-packets, $2 \leq s \leq u$, in $\mathcal{B}$.

For any s-subset $\mathcal{U} \subseteq \binom{\mathcal{X}}{w}$, $2 \leq s \leq u$, let $X(\mathcal{U})$ be the indicator random variable for the event $\mathcal{U} \subseteq \mathcal{B}$. Then

$$Pr(X(\mathcal{U})) = p^s$$

as all s blocks in $\mathcal{U}$ must be chosen to be in $\mathcal{B}$.

So, by the linearity of expectation,

$$\begin{aligned} E(X) &= \sum_{\substack{\mathcal{U} \text{ is a bad } s\text{-packet in } \binom{\mathcal{X}}{w}, \\ 2 \le s \le u}} Pr(X(\mathcal{U})) \\ &= \sum_{2 \le s \le u} N_s p^s, \end{aligned} \tag{5.1}$$

where N_s is the number of bad s-packets in $\binom{\mathcal{X}}{w}$.

For each $2 \le s \le u$, we have

$$N_s \le \binom{v}{(s-1)w} \binom{\binom{(s-1)w}{w}}{s}. \tag{5.2}$$

Indeed, since each bad s-packet is spanned by at most $(s-1)w$ points, so any $(s-1)w$ points in $\mathcal{X}$ may generate up to $\binom{\binom{(s-1)w}{w}}{s}$ bad s-packets in $\mathcal{X}$. There are $\binom{v}{(s-1)w}$ distinct subsets of size $(s-1)w$ in $\mathcal{X}$. The inequality (5.2) for N_s follows.

From (5.1) and (5.2),

$$E(X) \le \sum_{2 \le s \le u} \binom{v}{(s-1)w} \binom{\binom{(s-1)w}{w}}{s} p^s.$$

So, again by the linearity of expectation,

$$\begin{aligned} E(V-X) &= E(V) - E(X) \\ &\ge \binom{v}{w} p - \sum_{2 \le s \le u} \binom{v}{(s-1)w} \binom{\binom{(s-1)w}{w}}{s} p^s. \end{aligned} \tag{5.3}$$

Take $p = c_0 v^{\frac{(2-u)w}{u-1}}$, where c_0 is a constant chosen appropriately and depending only on w and t. Note that for fixed w, t and sufficiently large v, the value of p always can be chosen such that $0 < p < 1$. Then for (5.3) and sufficiently large v, we have

$$\begin{aligned} E(V-X) &\ge c_1 v^w p - c_2 v^{(u-1)w} p^u - c_3 \sum_{2 \le s \le u-1} v^{(s-1)w} p^s \\ &\ge c_0 c_1 v^w v^{\frac{(2-u)w}{u-1}} - c_0^u c_2 v^{(u-1)w} v^{\frac{(2-u)uw}{u-1}} - c_0^{u-1} c_3 \\ &\ge c v^{\frac{w}{u-1}}, \end{aligned}$$

where c_1, c_2, c_3 and c are constants depending only on w and t.

Thus, there exists at least one point in the probability space for which the difference $V - X$ is at least $cv^{\frac{w}{u-1}}$. That is, there is a family of blocks $\mathcal{B}$ which has at least $cv^{\frac{w}{u-1}}$ more blocks than bad s-packets, $2 \le s \le u$. Delete one block from each bad s-packets, $2 \le s \le u$, in $\mathcal{B}$, leaving a set $\mathcal{B}'$. This set $\mathcal{B}'$ does not contain any bad s-packets, $2 \le s \le u$, and has at least $cv^{\frac{w}{u-1}}$ blocks.

Thus the theorem follows by Corollary 5.4.2. □

We remark that the lower bound for t-IPPS(w,v) in Theorem 5.4.4 has order of magnitude $\frac{w}{u-1} = \frac{w}{\lfloor t^2/4\rfloor+t}$, which is extremely close to the order of magnitude $\lceil \frac{w}{\lfloor t^2/4\rfloor+t} \rceil$ of the upper bound in Theorem 5.3.1. Particularly, when $\lfloor t^2/4\rfloor + t$ is a divisor of w, the expurgation method provides asymptotically optimal t-IPPS(w,v) in the sense that

$$\lim_{v\to\infty} \frac{\log_v I_t(w,v)}{w} = \frac{1}{\lfloor t^2/4\rfloor + t}$$

meets the upper bound of Theorem 5.3.1. This is also a positive answer to Conjecture 5.3.5 for certain cases.

5.5 t-IPPS with small t and w

In the preceding section, we provided a probabilistic existence lower bound for t-IPPS(w,v) with general t and w. In this section, we pay attention to t-IPPS(w,v) for the case of $t = 2,3$, small positive integer w, and $v \geq w$. We aim to determine the exact value of $I_t(w,v)$ in these cases.

5.5.1 Bounds for 2-IPPS with $w = 2,3$

First, considering a 2-IPPS(w,v), we have the following lemma.

Lemma 5.5.1 *A (w,v) set system $(\mathcal{X},\mathcal{B})$ is a 2-IPPS(w,v) if and only if the following cases hold:*

(IPPSa) *For any three distinct blocks $A,B,C \in \mathcal{B}$, we have*

$$|(A\cup B)\cap(A\cup C)\cap(B\cup C)| < w.$$

(IPPSb) *For any four distinct blocks $A,B,C,D \in \mathcal{B}$, we have*

$$|(A\cup B)\cap(C\cup D)| < w.$$

Proof: From Theorem 5.1.5, a (w,v) set system $(\mathcal{X},\mathcal{B})$ is not a 2-IPPS(w,v) if and only if there exists a minimal forbidden configuration in $\mathcal{C}$ with size at most 4. Since any two blocks in $\mathcal{B}$ are different, we only need to consider the minimal forbidden configuration in $\mathcal{B}$ with size 3 and 4. If there exists a minimal forbidden configuration in $\mathcal{B}$ with size 3, say A,B,C, then the configuration $\mathcal{F} = \{\{A,B\},\{A,C\},\{B,C\}\}$ should be forbidden, which is the opposite side of (IPPSa). Similarly, the minimal forbidden configuration in $\mathcal{B}$ with size 4 is the opposite side of (IPPSb). The lemma follows. □

By virtue of a construction for traceability schemes from sunflowers in Theorem 4.3.1, we have

Lemma 5.5.2 *For any $v \geq w \geq 2,\ t \geq 2,$ we have*

$$I_t(w, v) \geq v - w + 1.$$

Proof: By Corollary 3.2.5 (1) and Theorem 4.3.1, we have

$$I_t(w, v) \geq M_t(w, v) \geq v - w + 1,$$

as desired. □

The following corollary follows from Lemma 5.5.2 and Theorem 5.3.1.

Corollary 5.5.3 *For any $v \geq w \geq 2,\ t \geq 2$ such that $w \leq \lfloor t^2/4 \rfloor + t$, we have*

$$v - w + 1 \leq I_t(w, v) \leq v.$$

We start from the first non-trivial case, that is, $t = 2$ and $w = 2$.

Theorem 5.5.4 *For any $v \geq 2$, we have $I_2(2, v) = v - 1$.*

Proof: From Corollary 5.5.3, we have $v - 1 \leq I_2(2, v) \leq v$. It suffices to prove $I_2(2, v) < v$. Suppose not, and $(\mathcal{X}, \mathcal{B})$ is a 2-IPPS$(2, v)$ with size v. We aim to find a contradiction to the definition of IPPS.

First, there exists one point $x \in \mathcal{X}$ such that x appears in at least two distinct blocks, since $\lceil \frac{2v}{v} \rceil = 2$. That is,

$$2 \leq |\{B \in \mathcal{B} :\ x \in B\}| := b(x) \leq v - 1.$$

Then there exists another point $y \in \mathcal{X} \backslash \{x\}$ such that y appears in at least two blocks, since $\lceil \frac{2v - b(x) - 1}{v - 1} \rceil = 2$. Without loss of generality, we assume that $A, B \in \mathcal{B}$ are two distinct blocks containing x, and $C, D \in \mathcal{B}$ are two distinct blocks containing y. Note that $\{A, B\} \neq \{C, D\}$. If $|\{A, B, C, D\}| = 3$, then, without loss of generality, assume $A = C = \{x, y\}$. It implies $A \subseteq B \cup D$, a contradiction to Lemma 5.5.1 (IPPSa). If $|\{A, B, C, D\}| = 4$, then $\{x, y\} \subseteq (A \cup C) \cap (B \cup D)$, a contradiction to Lemma 5.5.1 (IPPSb). Thus $I_2(2, v) < v$ and the lemma follows.

□

A direct corollary from Theorem 5.5.4 is as follows.

Corollary 5.5.5 *For any $v \geq 2,\ t \geq 2$, we have $I_t(2, v) = v - 1$.*

For $w \geq 3$, we have the following observation.

Proposition 5.5.6 *Let $v \geq w \geq 3$ and $(\mathcal{X}, \mathcal{B})$ be a 2-IPPS(w, v). If there exist two distinct blocks $A, B \in \mathcal{B}$ such that $|A \cap B| = w - 1$, then $|\mathcal{B}| \leq v - w + 1$.*

Proof: With the assumption, we first claim that for any point $x \in \mathcal{X} \setminus (A \cap B)$, x is contained in at most one block in $\mathcal{B}$. Suppose not, then there exists one point $x_0 \in \mathcal{X} \setminus (A \cap B)$ contained in two blocks $C, D \in \mathcal{B}$. Clearly, $\{A, B\} \neq \{C, D\}$. If $|\{A, B, C, D\}| = 3$, then, without loss of generality, assume $A = C$. It implies that any two blocks of A, B, D can generate a w-subset $(A \cap B) \cup \{x_0\}$, a contradiction to Lemma 5.5.1 (IPPSa). If $|\{A, B, C, D\}| = 4$, then $(A \cap B) \cup \{x_0\} \subseteq (A \cup C) \cap (B \cup D)$, a contradiction to Lemma 5.5.1 (IPPSb). Thus the claim follows.

Based on the above claim, $\mathcal{B}$ with the maximum number of blocks is from the construction in Lemma 5.5.2 (actually Theorem 4.3.1), where $\Delta = A \cap B$. It implies $|\mathcal{B}| \leq v - w + 1$, as desired. □

Here we remark that if one would like to explore the exact value of $I_2(w, v)$, $v \geq w \geq 3$, the first step may need to analyze the set system with block size w and

$$\max\{|B_1 \cap B_2| : B_1, B_2 \in \mathcal{B}, B_1 \neq B_2\} \leq w - 2.$$

5.5.2 An upper bound for 2-IPPS$(4, v)$: A graph theoretic approach

In this subsection, we consider the case $t = 2$ and $w = 4$. By Theorem 5.4.4 and Theorem 5.3.1, for sufficiently large v, we have

$$cv^{4/3} \leq I_2(4, v) \leq \frac{1}{2}v^2,$$

where c is a positive constant. One interesting problem is to determine the order of magnitude of the size of 2-IPPS$(4, v)$. In the following, we prove

Theorem 5.5.7 *$I_2(4, v) = o(v^2)$.*

Before we prove Theorem 5.5.7, we do some preparations. First, by Corollary 5.5.6, we have

Corollary 5.5.8 *Let $(\mathcal{X}, \mathcal{B})$ be a 2-IPPS$(4, v)$. If there exist two distinct blocks $B_1, B_2 \in \mathcal{B}$ such that $|B_1 \cap B_2| = 3$, then $|\mathcal{B}| \leq v - 3$.*

To prove Theorem 5.5.7, we have to consider the 2-IPPS$(4, v)$ $(\mathcal{X}, \mathcal{B})$ where

$$\max\{|B_1 \cap B_2| : B_1, B_2 \in \mathcal{B}, B_1 \neq B_2\} \leq 2,$$

since the case in Corollary 5.5.8 has already satisfied Theorem 5.5.7.

Moreover, we have the following obvious observation.

Proposition 5.5.9 *In a set system $(\mathcal{X}, \mathcal{B})$, the number of blocks in $\mathcal{B}$ that contains at least one 1-own-subset is at most v.*

The above proposition shows that for any set system $(\mathcal{X}, \mathcal{B})$, one can remove at most v blocks from $\mathcal{B}$ to satisfy that each of the remaining blocks in $\mathcal{B}$ does not contain any 1-own-subset. In the following discussion, we may assume that a set system $(\mathcal{X}, \mathcal{B})$ such that

- $\max\{|B_1 \cap B_2| : B_1, B_2 \in \mathcal{B}, B_1 \neq B_2\} \leq 2$,
- for any $B \in \mathcal{B}$ and any $x \in B$, there exists $B' \in \mathcal{B} \setminus \{B\}$ such that $x \in B'$.

(5.4)

We also have the following observation.

Proposition 5.5.10 *Let $(\mathcal{X}, \mathcal{B})$ be a 2-IPPS$(4, v)$. If there exist two distinct blocks $B_1, B_2 \in \mathcal{B}$ such that $|B_1 \cap B_2| = 2$, then there does not exist $B' \in \mathcal{B} \setminus \{B_1, B_2\}$ such that $B_i \setminus (B_1 \cap B_2) \subseteq B'$, where $i = 1, 2$.*

Proof: Suppose not, that is there exists $B' \in \mathcal{B} \setminus \{B_1, B_2\}$ such that $B_1 \setminus (B_1 \cap B_2) \subseteq B'$. Then $B_1 \subseteq (B_2 \cup B')$, a contradiction to Lemma 5.5.1 (IPPSa). □

What is more, we have

Lemma 5.5.11 *Let $(\mathcal{X}, \mathcal{B})$ be a 2-IPPS$(4, v)$ such that (5.4). Then*

$$|\{B \in \mathcal{B} : \exists B' \in \mathcal{B} \setminus \{B\} \text{ such that } |B \cap B'| = 2\}| \leq v - 1. \tag{5.5}$$

Proof: Suppose on the contrary that the left-hand side of (5.5) is no less than v. Then we would like to find a contradiction to that $\{\mathcal{X}, \mathcal{B}\}$ is a 2-IPPS$(4, v)$.

To this end, we construct a graph $G = (\mathcal{X}, \mathcal{E})$, where $\mathcal{X}$ is the vertex-set and $\mathcal{E}$ is the edge-set. For each $B \in \{B \in \mathcal{B} : \exists B' \in \mathcal{B} \setminus \{B\}$ s.t. $|B \cap B'| = 2\}$, the 2-subset $B \setminus (B \cap B')$ forms an edge in $\mathcal{E}$, where B' is a block in $\mathcal{B} \setminus \{B\}$ such that $|B \cap B'| = 2$. Note that a block B may contribute more than one edge to $\mathcal{E}$. From Proposition 5.5.10, we have that any edge in $\mathcal{E}$ only belongs to one block in $\mathcal{B}$ and any two edges in $\mathcal{E}$ arising from the same block are adjacent. According to the assumption, we have $|\mathcal{E}| \geq v$. That is, G is a graph on a v-vertex-set containing more than $v - 1$ edges. By Lemma 2.2.8, there exists a cycle in G. Hence there exists a path of length 3 as Figure 5.4. If $a = d$, the path is a cycle of length 3.

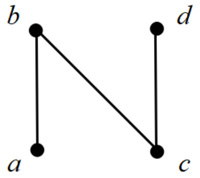

Figure 5.4: A path of length 3

Case 1. If a, b, c belong to the same block $B \in \mathcal{B}$, then let $B = \{a, b, c, v_1\}$, where $v_1 \in \mathcal{X} \setminus \{a, b, c\}$. By the first assumption of (5.4) and the way of constructing G, there exist $B_1 \in \mathcal{B} \setminus \{B\}$ such that $B \cap B_1 = \{c, v_1\}$, and $B_2 \in \mathcal{B} \setminus \{B, B_1\}$ such that $B \cap B_2 = \{a, v_1\}$. Since any two blocks in $\mathcal{B}$ intersect at at most two points, there exists a point $x \in B_1$ and $x \notin B$, $x \notin B_2$. By the second assumption of (5.4), there exists a block $B_3 \in \mathcal{B} \setminus \{B, B_1, B_2\}$ such that $x \in B_3$. Now we have $|\{a, c, v_1, x\}| = 4$ and

$$\{a, c, v_1, x\} \subseteq B \cup B_3 \ \text{ and } \ \{a, c, v_1, x\} \subseteq B_1 \cup B_2.$$

However, $\{B, B_3\}$ and $\{B_1, B_2\}$ are disjoint, which implies a contradiction to Lemma 5.5.1 (IPPSb).

Similarly, we can derive a contradiction to Lemma 5.5.1 when b, c, d belong to the same block of $\mathcal{B}$.

Case 2. Now we consider the case that there does not exist any block $B \in \mathcal{B}$ such that $\{a, b, c\} \subseteq B$ or $\{b, c, d\} \subseteq B$. Let $B_1 = \{a, b, v_1, v_2\} \in \mathcal{B}$, where $v_1 \neq v_2$, $v_1 \neq a, b, c$ and $v_2 \neq a, b, c$. Let B_2, B_3 be the blocks such that $\{b, c\} \subseteq B_2$ and $\{c, d\} \subseteq B_3$, respectively. Since $\{a, b, c\}$ is not contained in any block of $\mathcal{B}$, we have $B_1 \neq B_2$ and $B_1 \neq B_3$. Since $\{b, c, d\}$ is not contained in any block of $\mathcal{B}$, we have $B_2 \neq B_3$. By the way of constructing G, there exists $B'_1 \in \mathcal{B} \setminus \{B_1\}$ such that $B'_1 \cap B_1 = \{v_1, v_2\}$. Then $B'_1 \neq B_2$ follows from the first assumption of (5.4). Now we have $|\{b, c, v_1, v_2\}| = 4$ and

$$\{b, c, v_1, v_2\} \subseteq B_1 \cup B_2, \ \ \{b, c, v_1, v_2\} \subseteq B_1 \cup B_3 \ \text{ and } \ \{b, c, v_1, v_2\} \subseteq B'_1 \cup B_2.$$

If $B'_1 = B_3$, then $\{b, c, v_1, v_2\}$ can be generated by any two of B_1, B_2, B_3, a contradiction to Lemma 5.5.1 (IPPSa). If $B'_1 \neq B_3$, then $\{B_1, B_3\}$ and $\{B'_1, B_2\}$ are disjoint but both of them can generate $\{b, c, v_1, v_2\}$, a contradiction to Lemma 5.5.1 (IPPSb).

This completes the proof. □

Now we are going to prove Theorem 5.5.7 by virtue of the well-known graph removal lemma. A special case of Lemma 2.2.9, where $H = K_k$, was also shown by Alon, Duke, Lefmann, Rödl and Yuster in [3].

Lemma 5.5.12 ([3]) *For every $\gamma > 0$ and every positive integer k, there exists a constant $\delta = \delta(k, \gamma) > 0$ such that every graph G on n vertices, containing less than δn^k copies of the complete graph K_k on k vertices, contains a set of less than γn^2 edges whose deletion destroys all copies of K_k in G.*

Now we prove Theorem 5.5.7 by means of Lemma 5.5.12.

Proof of Theorem 5.5.7. Proving Theorem 5.5.7 is equivalent to proving that for any $\varepsilon > 0$, there exists $v_0 = v_0(\varepsilon)$ such that for any $v > v_0$, we have $I_2(4, v) < \varepsilon v^2$.

Suppose $(\mathcal{X}, \mathcal{B})$ is a 2-IPPS$(4, v)$ of size M. If there exist two distinct blocks $B_1, B_2 \in \mathcal{B}$ such that $|B_1 \cap B_2| = 3$, then Corollary 5.5.8 ensures that $M \leq v - 3$, which implies $I_2(4, v) = o(v^2)$. So we need only consider the case that $\max\{|B_1 \cap B_2| : B_1, B_2 \in \mathcal{B}, B_1 \neq B_2\} \leq 2$.

First, by Proposition 5.5.9 and Lemma 5.5.11, we can remove at most $2v$ blocks from $\mathcal{B}$ to make the remaining $\mathcal{B}'$ satisfy that any block does not contain any 1-own-subset and any two blocks intersect at at most one point. Clearly, $|\mathcal{B}'| \leq |\mathcal{B}| = M$ and $|\mathcal{B}'| \geq |\mathcal{B}| - 2v = M - 2v$.

Now we construct a graph $G = (\mathcal{X}, \mathcal{E})$ by the following way: for each $B \in \mathcal{B}'$, any 2-subset of B forms an edge in $\mathcal{E}$. Obviously, any block contributes $\binom{4}{2} = 6$ edges, which actually form a copy of the complete graph K_4. Since any two distinct blocks in $\mathcal{B}'$ intersect at at most one point, thus any two copies of K_4 in G that arise from two distinct blocks are edge-disjoint. Hence, $|\mathcal{E}| \geq 6(M - 2v)$. Accordingly, one needs to delete at least $M - 2v$ edges from $\mathcal{E}$ to destroy all copies of complete graph K_4 in G.

Assume that for sufficiently large v, we have $M \geq \varepsilon v^2$ for some $\varepsilon > 0$. Then we need to delete at least $M - 2v \geq \frac{\varepsilon}{2} v^2$ edges from $\mathcal{E}$ to destroy all copies of K_4 in G. By Lemma 5.5.12, let $k = 4$ and $\gamma = \frac{\varepsilon}{2}$, we should have that G contains at least δv^4 copies of K_4, where $\delta = \delta(\varepsilon)$ is a positive constant.

For these copies of K_4 in G, the number of copies of K_4 which contain at least two edges arising from the same block in $\mathcal{B}'$ is $O(v^3)$. Indeed, by an upper bound of 2-IPPS$(4, v)$ in Theorem 5.3.1 that $M \leq \binom{v}{2}$, there are at most $\binom{v}{2}$ ways to choose a block in $\mathcal{B}'$, and $\binom{6}{2} = 15$ ways to choose two edges from that block. The above process decides at least three vertices and there are at most $v - 3$ ways to choose another vertex to form a copy of K_4. Thus $O(v^3)$ follows.

Since G contains at least δv^4 copies of K_4, there exists a copy of K_4 in which any two edges come from two different blocks. Denote one such copy of K_4 as Figure 5.5. We can suppose that $B_1, B_2, B_3, B_4 \in \mathcal{B}'$ are four distinct blocks such that

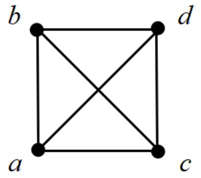

Figure 5.5: A copy of K_4

$\{a, b\} \subseteq B_1$, $\{a, c\} \subseteq B_2$, $\{b, d\} \subseteq B_3$ and $\{c, d\} \subseteq B_4$. Now we have $|\{a, b, c, d\}| = 4$ and

$$\{a, b, c, d\} \subseteq B_1 \cup B_4 \quad \text{and} \quad \{a, b, c, d\} \subseteq B_2 \cup B_3.$$

However, $\{B_1, B_4\}$ and $\{B_2, B_3\}$ are disjoint, which implies a contradiction to Lemma 5.5.1 (IPPSb).

Thus for any $\varepsilon > 0$ and sufficiently large v, we have $M < \varepsilon v^2$, as desired. □

In the next subsection, we will prove a similar upper bound for 3-IPPS$(6, v)$ by using the graph removal lemma.

5.5.3 An upper bound for 3-IPPS$(6, v)$

By Theorem 5.3.1 and Theorem 5.4.4, we have

$$cv^{6/5} \leq I_3(6, v) \leq \frac{1}{6}v^2,$$

where c is a positive constant. In this subsection, by a similar way as the argument of Theorem 5.5.7, we prove a better upper bound for 3-IPPS$(6, v)$ as follows.

Theorem 5.5.13 *$I_3(6, v) = o(v^2)$.*

Before we go to the proof of Theorem 5.5.13, we first prove the following lemma.

Lemma 5.5.14 *Let $(\mathcal{X}, \mathcal{B})$ be a 3-IPPS$(6, v)$. If there exist two distinct blocks $A, B \in \mathcal{B}$ such that $|A \cap B| \geq 2$, then $|\mathcal{B}| = o(v^2)$.*

Proof: Suppose on the contrary that, under the assumption, there exists $\varepsilon > 0$ such that for sufficiently large v, $|\mathcal{B}| \geq \varepsilon v^2$.

By the assumption, there are two points $\{a, b\} \subseteq A \cap B$. Now we would like to remove some points from each block of $\mathcal{B}$ to obtain a truncated set system. For each block $B \in \mathcal{B} \setminus \{A, B\}$,

- if $\{a, b\} \subseteq B$, then we remove a, b from B, obtaining B';
- if a (or b) exists in B, then we first remove a (or b) and another arbitrary point from B, obtaining B';
- if $\{a, b\} \nsubseteq B$, then we remove any two points from B, obtaining B'.

From the above process, we obtain a set system $(\mathcal{X}, \mathcal{B}')$ such that for any $B' \in \mathcal{B}'$, $|B'| = 4$.

We first claim that $\mathcal{B}'$ is not a multi-set. Suppose not, then there exist two distinct blocks $C, D \in \mathcal{B}$ such that $|(C \cap D) \setminus \{a, b\}| \geq 4$. Then $(C \cap D) \cup \{a, b\} \subseteq (A \cup C) \cap (B \cup D)$, a contradiction to the assumption that $(\mathcal{X}, \mathcal{B})$ is a 3-IPPS$(6, v)$. Thus any two blocks in $\mathcal{B}'$ are distinct, which implies that each $B' \in \mathcal{B}'$ corresponds to a block $B \in \mathcal{B}$.

By the assumption, we have $|\mathcal{B}'| \geq \varepsilon v^2 - 2$. Then from Theorem 5.5.7, we know that $(\mathcal{X}, \mathcal{B}')$ is not a 2-IPPS$(4, v)$. Then we have the following two possible cases.

- If there exist three distinct blocks $C', D', E' \in \mathcal{B}'$ such that any two of them can generate the same 4-subset $T' \subseteq \mathcal{X} \setminus \{a, b\}$, denote $C, D, E \in \mathcal{B}$ as the corresponding blocks of C', D' and E'. Then any triple in
$$\{A, C, D\},\ \{A, C, E\},\ \{A, D, E\},\ \{B, C, D\},\ \{B, C, E\},\ \{B, D, E\}$$
can generate the same 6-subset $T' \cup \{a, b\}$. But their intersection is empty, which contradicts the definition of 3-IPPS$(6, v)$.

- If there exist four distinct blocks $C', D', E', F' \in \mathcal{B}$ such that $\{C, D\}$ and $\{E, F\}$ can generate the same 4-subset $T' \subseteq \mathcal{X} \setminus \{a, b\}$, denote $C, D, E, F \in \mathcal{B}$ as the corresponding blocks of C', D', E' and F'. Then $\{A, C, D\}$ and $\{B, E, F\}$ can generate the same 6-subset $T' \cup \{a, b\}$. But $\{A, C, D\} \cap \{B, E, F\} = \emptyset$, which contradicts the definition of 3-IPPS$(6, v)$.

Thus for any $\varepsilon > 0$ and sufficiently large v, $|\mathcal{B}| < \varepsilon v^2$, establishing the lemma. □

Now we are going to prove Theorem 5.5.13.

Proof of Theorem 5.5.13. Let $(\mathcal{X}, \mathcal{B})$ be a 3-IPPS$(6, v)$. If there exist two blocks in $\mathcal{B}$ intersecting at at least two points, then by Lemma 5.5.14, we have $|\mathcal{B}| = o(v^2)$. Hence in the following, we assume that for any two distinct blocks $A, B \in \mathcal{B}$, $|A \cap B| \leq 1$.

We aim to prove $|\mathcal{B}| = o(v^2)$. Suppose on the contrary that there exists a constant $\varepsilon > 0$ such that for sufficiently large v, $|\mathcal{B}| \geq \varepsilon v^2$. We would like to find a contradiction to the assumption that $(\mathcal{X}, \mathcal{B})$ is a 3-IPPS$(6, v)$.

First, we construct a graph $G = (\mathcal{X}, \mathcal{E})$. For any block $B \in \mathcal{B}$, any 2-subset of B forms an edge in $\mathcal{E}$. Thus a block $B \in \mathcal{B}$ contributes to a copy of K_6 in G. By the assumption that any two distinct blocks intersect at at most one point, we have that any two copies of K_6 generated from two distinct blocks are edge-disjoint.

Then we need to delete at least εv^2 edges from $\mathcal{E}$ to destroy all copies of K_6 in G. By Lemma 5.5.12, let $k = 6$ and $\gamma = \frac{\varepsilon}{2}$, we should have that G contains at least δv^6 copies of K_6, where $\delta = \delta(\varepsilon)$ is a positive constant.

For these copies of K_6 in G, the number of copies of K_6 which contain at least two edges arising from the same block in $\mathcal{B}$ is $O(v^5)$. Indeed, by an upper bound of 2-IPPS$(6, v)$ in Theorem 5.3.1 that $|\mathcal{B}| \leq \binom{v}{2}$, there are at most $\binom{v}{2}$ ways to choose a block in $\mathcal{B}$, and $\binom{15}{2}$ ways to choose two edges from that block. The above process decides at least three vertices and there are at most $(v-3)^3$ ways to choose another three vertices to form a copy of K_6. Thus $O(v^5)$ follows.

Since G contains at least δv^6 copies of K_6, there exists a copy of K_6 in which any two edges come from two different blocks. Denote one such copy of K_6 as Figure 5.6.

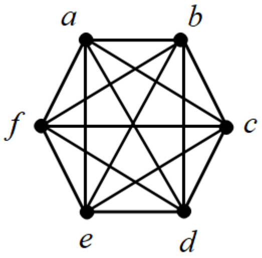

Figure 5.6: A copy of K_6

We can suppose that $A, B, C, D, E, F \in \mathcal{B}$ are six distinct blocks such that $\{a,b\} \subseteq A$, $\{c,d\} \subseteq B$, $\{e,f\} \subseteq C$, $\{a,f\} \subseteq D$, $\{b,c\} \subseteq E$ and $\{d,e\} \subseteq F$. Now we have $|\{a,b,c,d,e,f\}| = 6$ and

$$\{a,b,c,d,e,f\} \subseteq A \cup B \cup C \quad \text{and} \quad \{a,b,c,d,e,f\} \subseteq D \cup E \cup F.$$

However, $\{A,B,C\} \cap \{D,E,F\} = \emptyset$ implies a contradiction to the definition of 3-IPPS$(6, v)$.

Thus for any $\varepsilon > 0$ and sufficiently large v, we have $|\mathcal{B}| < \varepsilon v^2$, as desired. □

We remark that for $t \geq 4$ and $w = \lfloor (\frac{t}{2}+1)^2 \rfloor$, we may cannot have a similar argument as that of Theorem 5.5.7. Since in a graph, we can only get $2t$ vertices from t distinct edges, and the fact $w = \lfloor (\frac{t}{2}+1)^2 \rfloor > 2t$ for any $t \geq 4$ implies that $2t$ vertices are not enough to form a w-subset. But we believe that this obstacle can be removed by virtue of hypergraphs or some elaborate analyses. Precisely, we have the following conjecture.

Conjecture 5.5.15 *Suppose $t \geq 4$ is a positive integer, then*

$$I_t(w, v) = o(v^2),$$

where $w = \lfloor (\frac{t}{2}+1)^2 \rfloor$.

5.6 Summary

In this chapter, we investigated IPPSs. First, to unify many combinatorial structures with the parent-identifying property in such as digital fingerprinting and group testing, we proposed a unified concept of parent-identifying schemes in Section 5.1. An equivalent relationship between parent-identifying schemes and forbidden configurations was established. We also corresponded the problems for parent-identifying schemes to a kind of Turán-type problems.

Next, we showed a new upper bound for t-IPPS by virtue of techniques in extremal set theory in Section 5.3, which greatly improve the previously known bound.

Moreover, we proved a lower bound for t-IPPS with the probabilistic methods in Section 5.4, which has the same order of magnitude with the new upper bound in Section 5.3 for certain cases. Furthermore, in Section 5.5, we gave some bounds for 2-IPPS with small block size w. Particularly, better upper bounds for 2-IPPS$(4, v)$ and 3-IPPS$(6, v)$ were provided by means of the famous graph removal lemma in extremal graph theory, respectively.

Although we improved the known results by presenting new bounds on the size of IPPS in this chapter, to determine the exact maximum size of a t-IPPS(w, v) for the specific parameters t, w and v is still far from being solved.

Chapter 6

Union-Intersection-Bounded Families

In this chapter, we investigate a special kind of CFF due to its applications in broadcast encryption, called union-intersection-bounded family (UIBF). In Section 6.1, we show some properties of UIBF. An upper bound for general $(s, t; d)$-UIBF is proved in Section 6.2.1 by means of techniques in extremal set theory. With a result of CFF by Erdős, Frankl and Füredi in [39], we also provide a better upper bound for $(1, t; d)$-UIBF in Section 6.2.2. A probabilistic lower bound for $(s, t; d)$-UIBF is exhibited in Section 6.3. We make a summary of this chapter in Section 6.4. This chapter is based on results in [52].

6.1 Properties

The definition of UIBF was given in Definition 3.1.1. In this section, we provide some properties of UIBF. First, we have

Proposition 6.1.1 *Let $v \geq w \geq 2$, $t \geq s \geq 1$ and $d \leq w$ be positive integers. Then we have the following relationships:*

(1) An $(s, t; d)$-UIBF(w, v) is an $(s, t-1; d)$-UIBF(w, v).

(2) An $(s, t; d)$-UIBF(w, v) is an $(s-1, t; d)$-UIBF(w, v).

(3) An $(s, t; d-1)$-UIBF(w, v) is an $(s, t; d)$-UIBF(w, v).

(4) An $(s, t; d)$-UIBF(w, v) is an s-CFF(w, v).

(5) An $(s, t; d)$-UIBF(w, v) is a t-CFF(w, v).

Proof: For (1), we prove by assuming the opposite. Assume $(\mathcal{X}, \mathcal{B})$ is not an $(s, t-1; d)$-UIBF(w, v). Then by the definition, there exist $s + (t-1)$ distinct blocks $A_1, \ldots, A_s, B_1, \ldots, B_{t-1} \in \mathcal{B}$ such that

$$|(\bigcup_{1 \leq i \leq s} A_i) \cap (\bigcup_{1 \leq j \leq t-1} B_j)| \geq d.$$

By arbitrarily choosing a block $B_t \in \mathcal{B}\backslash\{A_1, \ldots, A_s, B_1, \ldots, B_{t-1}\}$, we directly have

$$|(\bigcup_{1\le i\le s} A_i) \cap (\bigcup_{1\le j\le t} B_j)| \ge d,$$

which implies that $(\mathcal{X}, \mathcal{B})$ is not an $(s, t; d)$-UIBF(w, v). Thus (1) follows.

The relationship (2) can be proved in a similar way, and we omit its proof. The relationship (3) is obvious. The relationship (4) follows from (1) and (3). The relationship (5) follows from (2) and (3). □

Consequently, we have the following corollary.

Corollary 6.1.2 *Let $v \ge w \ge 2$, $t \ge s \ge 1$ and $d \le w$ be positive integers. Then*

(1) $U_{s,t}(w, v; d) \le U_{s,t-1}(w, v; d)$;

(2) $U_{s,t}(w, v; d) \le U_{s-1,t}(w, v; d)$;

(3) $U_{s,t}(w, v; d-1) \le U_{s,t}(w, v; d)$.

6.2 Upper bounds for $(s, t; d)$-UIBF

In this section, we provide an upper bound for general $U_{s,t}(w, v; d)$ and a better upper bound for $U_{1,t}(w, v; d)$.

6.2.1 An upper bound for general UIBF

In this subsection, we prove the following theorem.

Theorem 6.2.1 *Let $v \ge w \ge 2$, $t \ge s \ge 1$ and $d \le w$ be positive integers. We have*

$$U_{s,t}(w, v; d) \le \binom{v}{\lceil \frac{d}{s+t-1} \rceil}.$$

To prove this upper bound, we use the notion own-subset. For any $(s, t; d)$-UIBF, we have the following lemma.

Lemma 6.2.2 *In an $(s, t; d)$-UIBF(w, v), there exists one block containing at least one $\lceil \frac{d}{s+t-1} \rceil$-own-subset.*

Proof: Suppose $(\mathcal{X}, \mathcal{B})$ is an $(s, t; d)$-UIBF(w, v) over a finite set $\mathcal{X}$, where $|\mathcal{X}| = v$. Assume on the contrary that for any $B \in \mathcal{B}$, B does not contain any $\lceil \frac{d}{s+t-1} \rceil$-own-subset. We would like to derive a contraction with that $(\mathcal{X}, \mathcal{B})$ is an $(s, t; d)$-UIBF(w, v). Explicitly, we aim to find out $s+t$ distinct blocks $A_1, \ldots, A_s, B_1, \ldots, B_t \in \mathcal{B}$ such that

$$|(\bigcup_{1\le i\le s} A_i) \cap (\bigcup_{1\le j\le t} B_j)| \ge d. \tag{6.1}$$

First, arbitrarily choose one block $A_1 \in \mathcal{B}$. Let $i = 1$ and $A_0 = B_0 = D_0 = E_0 = \emptyset$. Next turn to execute a while loop.

While $i < s$, we do the following three steps.

Step 1. Take a $\lceil \frac{d}{s+t-1} \rceil$-subset $D_i \subseteq A_i \setminus \bigcup_{0 \le j < i}(D_j \cup E_j)$ such that

$$D_i \nsubseteq \bigcup_{0 \le j < i} A_j \quad \text{and} \quad D_i \nsubseteq \bigcup_{0 \le j < i} B_j.$$

We claim that D_i is available. Indeed, if this kind of D_i is not available, it would imply that there does not exist any $\lceil \frac{d}{s+t-1} \rceil$-subset in $(A_i \setminus \bigcup_{0 \le j < i}(D_j \cup E_j)) \setminus \bigcup_{0 \le j < i} A_j$, or that there does not exist any $\lceil \frac{d}{s+t-1} \rceil$-subset in $(A_i \setminus \bigcup_{0 \le j < i}(D_j \cup E_j)) \setminus \bigcup_{0 \le j < i} B_j$. Without loss of generality, we may assume that there does not exist any $\lceil \frac{d}{s+t-1} \rceil$-subset in $(A_i \setminus \bigcup_{0 \le j < i}(D_j \cup E_j)) \setminus \bigcup_{0 \le j < i} A_j$. Equivalently,

$$0 \le |(A_i \setminus \bigcup_{0 \le j < i} (D_j \cup E_j)) \setminus \bigcup_{0 \le j < i} A_j| < \lceil \frac{d}{s+t-1} \rceil.$$

Then with the assumption that A_i does not contain any $\lceil \frac{d}{s+t-1} \rceil$-own-subset, we could take a block $A^\star \in \mathcal{B} \setminus \{A_1, \ldots, A_i\}$ such that

$$((A_i \setminus \bigcup_{0 \le j < i} (D_j \cup E_j)) \setminus \bigcup_{0 \le j < i} A_j) \subseteq A^\star.$$

This would imply that A_i can be covered by $i+1 \le s$ blocks $A_1, \ldots, A_{i-1}, B_{i-1}, A^\star$, contradicting Proposition 6.1.1 (4). Thus the required D_i is available.

Next, with the assumption that A_i does not contain any $\lceil \frac{d}{s+t-1} \rceil$-own-subset, we take a block $B_i \in \mathcal{B} \backslash \{A_1, \ldots, A_i, B_1, \ldots, B_{i-1}\}$ such that $D_i \subseteq B_i$. Then turn to Step 2.

Step 2. Take a $\lceil \frac{d}{s+t-1} \rceil$-subset $E_i \subseteq B_i \setminus \bigcup_{0 \le j < i}(D_j \cup E_j \cup D_i)$ such that

$$E_i \nsubseteq \bigcup_{0 \le j \le i} A_j \quad \text{and} \quad E_i \nsubseteq \bigcup_{0 \le j < i} B_j.$$

This E_i is available, which can be argued in a similar way as that D_i in Step 1 is available.

By the assumption that B_i does not contain any $\lceil \frac{d}{s+t-1} \rceil$-own-subset, we take a block $A_{i+1} \in \mathcal{B} \backslash \{A_1, \ldots, A_i, B_1, \ldots, B_i\}$ such that $E_i \subseteq A_{i+1}$. Then turn to Step 3.

Step 3. Let $i = i + 1$. Then continue to execute this while loop.

After the above loop, we have taken $A_1, \dots, A_s, B_1, \dots, B_{s-1} \in \mathcal{B}$. Now we explain how to choose $B_s, B_{s+1}, \dots, B_{t-1}$. We also execute a while loop. Now $i = s$. Let $\widehat{E} := \bigcup_{0 \le j < s} E_j$.

While $i < t$, take a $\lceil \frac{d}{s+t-1} \rceil$-subset $D_i \subseteq A_s \setminus \bigcup_{0 \le j < i}(D_j \cup \widehat{E})$ such that

$$D_i \nsubseteq \bigcup_{0 \le j < s} A_j \quad \text{and} \quad D_i \nsubseteq \bigcup_{0 \le j < i} B_j.$$

This D_i is available, which can be argued in a similar way, by virtue of Proposition 6.1.1 (5), as that the above $D_1, \dots, D_{s-1}$ are available. By the assumption that A_s does not contain any $\lceil \frac{d}{s+t-1} \rceil$-own-subset, we can take a block $B_i \in \mathcal{B} \backslash \{A_1, \dots, A_s, B_1, \dots, B_{i-1}\}$ such that $D_i \subseteq B_i$. Then let $i = i + 1$ and continue the while loop.

After this while loop, we have taken $A_1, \dots, A_s, B_1, \dots, B_{t-1} \in \mathcal{B}$. Next we choose the last block B_t. First, by Proposition 6.1.1 (5), we know that A_s can not be covered by any other t blocks. Thus

$$\begin{aligned}
A_s \setminus \bigcup_{0 \le j < t} (D_j \cup \widehat{E}) &\nsubseteq \bigcup_{0 \le j < s} A_j, \\
A_s \setminus \bigcup_{0 \le j < t} (D_j \cup \widehat{E}) &\nsubseteq \bigcup_{0 \le j < t} B_j, \\
|A_s \setminus \bigcup_{0 \le j < t} (D_j \cup \widehat{E})| &= k - (s + t - 2) \lceil \frac{d}{s+t-1} \rceil.
\end{aligned}$$

Then we take a subset $D_t \subseteq A_s \setminus \bigcup_{0 \le j < t}(D_j \cup \widehat{E})$ such that

$$\begin{aligned}
D_t \cap (A_s \setminus \bigcup_{0 \le j < s} A_j) &\neq \emptyset, \\
D_t \cap (A_s \setminus \bigcup_{0 \le j < t} B_j) &\neq \emptyset, \\
d - (s + t - 2) \lceil \frac{d}{s+t-1} \rceil \le &|D_t| \le \lceil \frac{d}{s+t-1} \rceil.
\end{aligned}$$

With the assumption, there exists a block $B_t \in \mathcal{B} \setminus \{A_1, \dots, A_s, B_1, \dots, B_{t-1}\}$ such that $D_t \subseteq B_t$.

Until now, we have finally taken $s + t$ distinct blocks $A_1, \dots, A_s, B_1, \dots, B_t \in \mathcal{B}$ such that

$$\begin{aligned}
&D_i \subseteq A_i \cap B_i, \quad &&\forall\, 1 \le i < s, \\
&D_i \subseteq A_s \cap B_i, \quad &&\forall\, s \le i \le t, \\
&E_j \subseteq B_j \cap A_{j+1}, \quad &&\forall\, 1 \le j < s,
\end{aligned}$$

where $D_1, \ldots, D_t, E_1, \ldots, E_{s-1}$ are pairwise disjoint. Hence we have

$$(\bigcup_{1\le i\le t} D_i) \cup (\bigcup_{1\le j<s} E_j) \subseteq (\bigcup_{1\le i\le s} A_i) \cap (\bigcup_{1\le j\le t} B_j)$$

and then

$$\begin{aligned}|(\bigcup_{1\le i\le s} A_i) \cap (\bigcup_{1\le j\le t} B_j)| &\ge |(\bigcup_{1\le i\le t} D_i) \cup (\bigcup_{1\le j<s} E_j)| \\ &= \sum_{1\le i<t} |D_i| + \sum_{1\le j<s} |E_j| + |D_t| \\ &\ge (s+t-2)\lceil \frac{d}{s+t-1} \rceil + (d-(s+t-2)\lceil \frac{d}{s+t-1} \rceil) \\ &= d,\end{aligned}$$

which implies (6.1) and contradicts that $(\mathcal{X}, \mathcal{B})$ is an $(s,t;d)$-UIBF(w,v).

This completes the proof. □

Now we prove Theorem 6.2.1.

Proof of Theorem 6.2.1. Suppose $(\mathcal{X}, \mathcal{B})$ is an $(s,t;d)$-UIBF(w,v). By Lemma 6.2.2, there exists one block $B \in \mathcal{B}$ which contains at least one $\lceil \frac{d}{t+s-1} \rceil$-own-subset. Delete this B from $\mathcal{B}$. The resulting $\mathcal{B} \setminus \{B\}$ is still an $(s,t;d)$-UIBF(w,v). Applying Lemma 6.2.2 repeatedly, we can successively delete blocks to obtain a series of $(s,t;d)$-UIBF(w,v)'s, each of which contains at least one $\lceil \frac{d}{t+s-1} \rceil$-own-subset. Note that there are $\binom{v}{\lceil \frac{d}{t+s-1} \rceil}$ distinct $\lceil \frac{d}{t+s-1} \rceil$-subsets from $\mathcal{X}$, and each $\lceil \frac{d}{t+s-1} \rceil$-subset, as an own-subset of some block, can be deleted at most once. Hence

$$|\mathcal{B}| \le \binom{v}{\lceil \frac{d}{t+s-1} \rceil},$$

and the theorem follows. □

We remark here that for the special case $s = 1$, we can improve this upper bound by exploring the number of $\lceil \frac{d}{t+s-1} \rceil$-own-subsets possessed by each block, which will be discussed in the next subsection.

6.2.2 A better upper bound for $(1,t;d)$-UIBF

The $(1,t;d)$-UIBF has been studied under the guise of a $(t;k-d)$-cover-free family [86] and a superimposed distance code [36]. Its applications to broadcast encryption were investigated in [48, 63] and to group testing for correcting errors in [54]. We do not expose the detailed applications here, and the interested reader is referred to [48, 54, 63].

In this subsection, we show a new upper bound for $(1,t;d)$-UIBF(w,v), which is better than that from the general upper bound in Theorem 6.2.1. The idea to derive a better upper bound for $(1,t;d)$-UIBF is to explore the number of $\lceil d/t\rceil$-own-subsets possessed by each block. The following lemma by Erdős, Frankl and Füredi [39] is useful.

Lemma 6.2.3 ([39]) *Let $\mathcal{F}$ be a family of d-subsets of a ground set, and $F \in \mathcal{F}$. If F can not be covered by the union of any other t distinct members of $\mathcal{F} \setminus \{F\}$, then F has at least $\binom{d-1}{\lceil d/t\rceil-1}$ $\lceil d/t\rceil$-own-subsets.*

Hence we have the following lemma.

Lemma 6.2.4 *Let $(\mathcal{X},\mathcal{B})$ be a $(1,t;d)$-UIBF(w,v). For any $B \in \mathcal{B}$, the number of $\lceil d/t\rceil$-own-subsets in B is at least $\lfloor \frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}$.*

Proof: From the definition of $(1,t;d)$-UIBF(w,v), we have that any d-subset of B can not be covered by the union of any other t distinct blocks in $\mathcal{B}$. There are $\lfloor \frac{w}{d}\rfloor$ disjoint d-subsets in B. By Lemma 6.2.3, the number of $\lceil d/t\rceil$-own-subsets in B is at least

$$\lfloor \frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1},$$

as desired. □

By virtue of Lemma 6.2.4, we have

Theorem 6.2.5 *Let $v \geq w \geq 2$, $t \geq 2$ and $d \leq w$ be positive integers. We have*

$$U_{1,t}(w,v;d) \leq \frac{\binom{v}{\lceil \frac{d}{t}\rceil} - \binom{w}{\lceil \frac{d}{t}\rceil}}{\lfloor \frac{w}{d}\rfloor\binom{d-1}{\lceil \frac{d}{t}\rceil-1}} + 1.$$

Proof: Suppose $\mathcal{B}$ is a $(1,t;d)$-UIBF(w,v) over $\mathcal{X}$ with size M. Denote $\binom{\mathcal{X}}{\lceil d/t\rceil}$ as the collection of all $\lceil d/t\rceil$-subsets of $\mathcal{X}$. Clearly, $|\binom{\mathcal{X}}{\lceil d/t\rceil}| = \binom{v}{\lceil d/t\rceil}$. The following is to double count the set $\{(T,B) : T \in \binom{\mathcal{X}}{\lceil d/t\rceil}, B \in \mathcal{B}, T \subseteq B\}$. Denote

$$\Sigma := |\{(T,B) : T \in \binom{\mathcal{X}}{\lceil d/t\rceil}, B \in \mathcal{B}, T \subseteq B\}|.$$

Then we have

$$\Sigma = \sum_{T\in\binom{\mathcal{X}}{\lceil d/t\rceil}} \sum_{\substack{B\in\mathcal{B}\\ \text{s.t. } T\subseteq B}} 1 = \sum_{B\in\mathcal{B}} \sum_{\substack{T\in\binom{\mathcal{X}}{\lceil d/t\rceil}\\ \text{s.t. } T\subseteq B}} 1. \tag{6.2}$$

On the one hand, fixing $B \in \mathcal{B}$, we have $\sum_{\substack{T\in\binom{\mathcal{X}}{\lceil d/t\rceil}\\ \text{s.t. } T\subseteq B}} 1 = \binom{k}{\lceil d/t\rceil}$. Then

$$\Sigma = \sum_{B\in\mathcal{B}} \binom{w}{\lceil d/t\rceil} = M\binom{w}{\lceil d/t\rceil}. \tag{6.3}$$

On the other hand, fixing $T \in \binom{\mathcal{X}}{\lceil d/t\rceil}$, there are the following two possible cases.

(a) If T is a $\lceil d/t\rceil$-own-subset of some block $B \in \mathcal{B}$, then we have $\sum_{\substack{B\in\mathcal{B}\\ \text{s.t. } T\subseteq B}} 1 = 1.$

(b) If T is not a $\lceil d/t\rceil$-own-subset of any block $B \in \mathcal{B}$, then we have $\sum_{\substack{B\in\mathcal{B}\\ \text{s.t. } T\subseteq B}} 1 \leq M.$

For any $B \in \mathcal{B}$, denote $\mathcal{O}(B)$ as the collection of all $\lceil d/t\rceil$-own-subsets of the block B. By Lemma 6.2.4, we have $|\mathcal{O}(B)| \geq \lfloor \frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}$. Without loss of generality, we may assume that

$$|\bigcup_{B\in\mathcal{B}} \mathcal{O}(B)| = M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} + \sigma,\ \sigma \geq 0,$$

for the reason that any $\lceil d/t\rceil$-own-subset is contained in exactly one block. Then the number of $T \in \binom{\mathcal{X}}{\lceil d/t\rceil}$, satisfying the condition of case (b), is at most

$$\binom{v}{\lceil d/t\rceil} - M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} - \sigma.$$

Thus by the first equality of (6.2),

$$\begin{aligned}\Sigma &\leq [M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} + \sigma] + M[\binom{v}{\lceil d/t\rceil} - M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} - \sigma]\\ &= M[\binom{v}{\lceil d/t\rceil} - M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} + \lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}] - (M-1)\sigma \\ &\leq M[\binom{v}{\lceil d/t\rceil} - M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} + \lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}].\end{aligned} \tag{6.4}$$

From (6.3) and (6.4), we have

$$M\binom{w}{\lceil d/t\rceil} \leq M[\binom{v}{\lceil d/t\rceil} - M\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1} + \lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}],$$

which implies

$$M \leq \frac{\binom{v}{\lceil d/t\rceil} - \binom{w}{\lceil d/t\rceil}}{\lfloor\frac{w}{d}\rfloor\binom{d-1}{\lceil d/t\rceil-1}} + 1,$$

as desired. □

We remark here that when $d = w$, a $(1, t; w)$-UIBF(w, v) is exactly a t-CFF(w, v). Its upper bound in Theorem 6.2.5 is consistent with those in [39] and Theorem 4.2.7.

6.3 A lower bound for UIBF

In this section, we show a lower bound for $(s, t; d)$-UIBF(w, v) by virtue of the probabilistic expurgation method.

Theorem 6.3.1 *Let $w \geq d$ and $t \geq s \geq 1$ be fixed positive integers. Then there exists a constant c, depending only on w, s and t, with the following property. For any sufficiently large integer v, there exists an $(s,t;d)$-UIBF(w,v) with size at least $cv^{\frac{d}{t+s-1}}$.*

Proof: Let $\mathcal{X}$ be a finite set of v points. Let $\binom{\mathcal{X}}{w}$ be the collection of all w-subsets (blocks) $F \subseteq \mathcal{X}$. Form a random subset $\mathcal{F} \subseteq \binom{\mathcal{X}}{w}$ of blocks by including each block independently at random with probability p, where $0 < p < 1$. We will determine the value of p later.

Let V denote the number of blocks in $\mathcal{F}$. Clearly, $E(V) = \binom{v}{w}p$.

To form an $(s,t;d)$-UIBF, we need to destroy all the subfamilies of cardinality $s'+t'$ in $\mathcal{F}$, $1 \leq s' \leq s,\ 1 \leq t' \leq t$, which contradict the definition. For convenience, we define the forbidden configuration as follows. Let $s' \leq s,\ t' \leq t$ be positive integers. A collection of $t'+s'$ distinct blocks $\{F_1, F_2, \ldots, F_{t'+s'}\} \subseteq \binom{\mathcal{X}}{w}$ is called an (s',t')-*forbidden configuration* if there exist s' distinct blocks, say $\mathcal{F}_1 \subseteq \{F_1, F_2, \ldots, F_{t'+s'}\}$ with $|\mathcal{F}_1| = s'$, such that

$$|(\bigcup_{F \in \mathcal{F}_1} F) \cap (\bigcup_{\substack{1 \leq i \leq t'+s', \\ F_i \notin \mathcal{F}_1}} F_i)| \geq d.$$

Now we are going to count the number of all forbidden configurations in $\mathcal{F}$. Let X denote the total number of (s',t')-forbidden configurations in $\mathcal{F}$ for all $s' \leq s$ and $t' \leq t$. Let $X_{(s',t')}$ denote the number of (s',t')-forbidden configurations in $\mathcal{F}$. Clearly,

$$E(X) = \sum_{\substack{1 \leq s' \leq s, \\ 1 \leq t' \leq t}} E(X_{(s',t')}).$$

Note that each (s',t')-forbidden configuration is spanned by at most

$$(t'+s')w - |(\bigcup_{F \in \mathcal{F}_1} F) \cap (\bigcup_{\substack{1 \leq i \leq t'+s', \\ F_i \notin \mathcal{F}_1}} F_i)| \leq (t'+s')w - d$$

points in $\mathcal{X}$. Thus we have

$$E(X_{(s',t')}) \leq \binom{v}{(t'+s')w-d} \binom{\binom{(t'+s')w-d}{w}}{t'+s'} p^{t'+s'}.$$

Indeed, there are $\binom{v}{(t'+s')w-d}$ subsets of $\mathcal{X}$ containing $(t'+s')w-d$ points. Each subset with $(t'+s')w-d$ points can generate $\binom{(t'+s')w-d}{w}$ blocks of size w, yielding at most $\binom{\binom{(t'+s')w-d}{w}}{t'+s'}$ distinct forbidden configurations. Since each block in $\binom{\mathcal{X}}{w}$ is included into $\mathcal{F}$ independently at random with probability p, any $t'+s'$ distinct blocks in $\binom{\mathcal{X}}{w}$ are included in $\mathcal{F}$ simultaneously with probability $p^{t'+s'}$. The above inequality follows.

By the linearity of expectation, we have

$$\begin{aligned} E(V-X) &= E(V) - E(X) \\ &= E(V) - \sum_{\substack{1\le s'\le s,\\ 1\le t'\le t}} E(X_{(s',t')}) \\ &\ge \binom{v}{k} p - \sum_{\substack{1\le s'\le s,\\ 1\le t'\le t}} \binom{v}{(t'+s')w-d} \binom{\binom{(t'+s')w-d}{w}}{t'+s'} p^{t'+s'}. \end{aligned} \tag{6.5}$$

Take $p = \kappa' v^{\frac{d}{t+s-1}-w}$, where κ' is a constant chosen appropriately depending only on w, s and t. Note that for fixed w, s, t and sufficiently large v, the value of p always can be chosen such that $0 < p < 1$. Then for (6.5) and sufficiently large v, we have

$$\begin{aligned} E(V-X) &\ge \kappa_1 v^w p - \sum_{\substack{1\le s'\le s,\\ 1\le t'\le t}} \kappa_2 v^{(t'+s')w-d} p^{t'+s'} \\ &= \kappa_1 \kappa' v^w v^{\frac{d}{t+s-1}-w} - \sum_{\substack{1\le s'\le s,\\ 1\le t'\le t}} \kappa_2 (\kappa')^{t'+s'} v^{(t'+s')w-d} v^{(t'+s')(\frac{d}{t+s-1}-w)} \\ &\ge \kappa_1' v^w v^{\frac{d}{t+s-1}-w} - \kappa_2' v^{(t+s)w-d} v^{(t+s)(\frac{d}{t+s-1}-w)} \\ &\ge \kappa v^{\frac{d}{t+s-1}}, \end{aligned}$$

where κ_1, κ_2, κ_1', κ_2' and κ are constants depending only on w, s and t.

Thus, there exists at least one point in the probability space for which the difference $V - X$ is at least $\kappa v^{\frac{d}{t+s-1}}$. That is, there is a family of blocks $\mathcal{F}$ which has at least $\kappa v^{\frac{d}{t+s-1}}$ more blocks than forbidden configurations inside. Delete one block from each forbidden configuration in $\mathcal{F}$, leaving a set $\mathcal{B}$. This set $\mathcal{B}$ does not contain any forbidden configuration, and has at least $\kappa v^{\frac{d}{t+s-1}}$ blocks.

This completes the proof. □

We remark here that the order of magnitude of the size of $(s,t;d)$-UIBF(w,v) in Theorem 6.3.1 is $\frac{d}{t+s-1}$, which is extremely close to that of the upper bound, $\lceil \frac{d}{t+s-1} \rceil$, in Theorem 6.2.1. Especially, when $t+s-1$ is a divisor of d, they are equal.

6.4 Summary

In this chapter, we focused on UIBF, which is a special kind of CFF. We provided upper bounds for UIBF in Section 6.2 and a lower bound for UIBF in Section 6.3, respectively. However, there is still a gap between the upper and lower bounds. It would be of interest to close the gap for UIBF.

Chapter 7

Multimedia Parent-Identifying Codes

Fingerprinting used in multimedia scenario was investigated recently by a number of authors, see [91, 24] for example. In this chapter, we focus on a kind of anti-collusion codes for multimedia fingerprinting, namely, multimedia parent-identifying codes (MIPPCs). The upper bound for MIPPC was given in [22] by establishing a connection with bipartite graphs without cycles of small length. However, there is no general lower bounds for MIPPC in the literature. In Section 7.1, we recap some definitions and the previously related results. Next, we provide the first probabilistic lower bound for MIPPC in Section 7.2. We analyze the new lower bound for MIPPC and show that the new lower bound is very close to the known upper bound. Finally, we summarize this chapter in Section 7.3. This chapter is based on results in [50].

7.1 Definition and known results

7.1.1 Definition

In the unified Definition 5.1.1 of parent-identifying schemes, the set of all authorized users' fingerprints for multimedia fingerprinting is a code $\mathcal{C} \subseteq Q^n$. Since the most considered attack model in multimedia fingerprinting is the averaging attack as formula (1.1), a pirate copy, generated by at most t dishonest users $\mathcal{C}' \subseteq \mathcal{C}$, can reflect the information of all traitors in $\mathcal{C}'$, that is,

$$f_m(\mathcal{C}') = \{\mathrm{desc}(\mathcal{C}')\} = \{\mathcal{C}'(1) \times \mathcal{C}'(2) \times \cdots \times \mathcal{C}'(n)\},$$

where

$$\mathcal{C}'(i) = \{\mathbf{c}(i) \in Q : \ \mathbf{c} = (\mathbf{c}(1), \dots, \mathbf{c}(n)) \in \mathcal{C}'\}$$

and $\mathrm{desc}(\mathcal{C}')$ is called the *descendant set* of $\mathcal{C}'$.

Example 7.1.1 *Let* $\mathcal{C} = \{000, 012, 020, 112\}$ *and* $\mathcal{C}' = \{000, 012, 020\} \subseteq \mathcal{C}$. *Then*

$$\mathcal{C}'(1) = \{0\}, \quad \mathcal{C}'(2) = \{0, 1, 2\}, \quad \mathcal{C}'(3) = \{0, 2\},$$

and

$$desc(\mathcal{C}') = \{0\} \times \{0, 1, 2\} \times \{0, 2\}.$$

The definition of codes with the identifiable parent property for multimedia fingerprinting was proposed by Cheng *et al.* in [22].

Definition 7.1.2 *An (n, q) code $\mathcal{C}$ has the t-identifiable parent property for multimedia fingerprinting, denoted t-MIPPC(n, q), if for any subcode $\mathcal{C}' \subseteq \mathcal{C}$ such that $|\mathcal{C}'| \leq t$, we have*

$$\bigcap_{\mathcal{S} \in S_t(\mathcal{C}')} \mathcal{S} \neq \emptyset,$$

where

$$S_t(\mathcal{C}') = \{\mathcal{S} \subseteq \mathcal{C} : \ |\mathcal{S}| \leq t, \ desc(\mathcal{S}) = desc(\mathcal{C}')\}.$$

The cardinality of $\mathcal{C}$ is called the *size* of this t-MIPPC(n, q). Denote $N_t(n, q)$ as the maximum size of a t-MIPPC(n, q). A t-MIPPC(n, q) $\mathcal{C}$ is called *optimal* if it has size $N_t(n, q)$. When q tends to infinity, the *asymptotic code rate* of a t-MIPPC(n, q) with size N is denoted as

$$R_m(n, t) = \lim_{q \to \infty} \frac{\log_q N}{n}.$$

A t-MIPPS(n, q) has the *asymptotically optimal code rate* if its asymptotic code rate is

$$\lim_{q \to \infty} \frac{\log_q N_t(n, q)}{n}.$$

Example 7.1.3 *Let*

$$\mathcal{C} = \begin{pmatrix} 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 & 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 & 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 & 1 & 2 & 3 & 4 \end{pmatrix},$$

where the column vectors are codewords. By Definition 7.1.2, $\mathcal{C}$ is a 3-MIPPC$(4, 4)$ of size $|\mathcal{C}| = 8$. Indeed, the descendant sets of all $\mathcal{C}' \subseteq \mathcal{C}$ such that $|\mathcal{C}'| \leq 3$ are distinct.

However, $\mathcal{C}$ is not a 4-MIPPC$(4, 4)$, since the first four codewords have the same descendant set as that of the last four codewords.

7.1.2 Known results

In the literature, Cheng *et al.* [22] transferred the requirement of a t-MIPPC to a corresponding bipartite graph without cycles of length less than or equal to $2t$, obtaining the following upper bound.

Theorem 7.1.4 ([22]) *Let n, t, q be positive integers. Then*

$$N_t(n,q) \le \begin{cases} q^{\frac{n}{2}}(q^{\frac{n}{2t}} + 2c) & \text{if } n \text{ is even} \\ q^{\frac{n}{2}}(q^{\frac{n+1}{2t}} + c(q^{\frac{1}{2}} + q^{-\frac{1}{2}})) & \text{if } n \text{ is odd and } t \text{ is even} \\ q^{\frac{n}{2}}(q^{\frac{n}{2t}} + c(q^{\frac{1}{2}} + q^{-\frac{1}{2}})) & \text{if } n \text{ is odd and } t \text{ is odd,} \end{cases}$$

where c is a constant depending only on t.

From Theorem 7.1.4, we have

Corollary 7.1.5 *Let n and t be positive integers. Then*

$$\begin{aligned} R_m(n,t) &= \lim_{q\to\infty} \frac{\log_q N_t(n,q)}{n} \\ &\le \begin{cases} \frac{1}{2} + \frac{1}{2t} & \text{if } n \text{ is even} \\ \frac{1}{2} + \max\{\frac{n+1}{2tn}, \frac{1}{2n}\} & \text{if } n \text{ is odd and } t \text{ is even} \\ \frac{1}{2} + \max\{\frac{1}{2t}, \frac{1}{2n}\} & \text{if } n \text{ is odd and } t \text{ is odd.} \end{cases} \end{aligned} \tag{7.1}$$

In the next section, we will prove a probabilistic lower bound for MIPPC.

7.2 Probabilistic existence result for t-MIPPC

7.2.1 A lower bound for t-MIPPC

By Definition 5.1.4, we have the following description of forbidden configurations in a t-MIPPC. Let $\mathcal{C}$ be a t-MIPPC, a (minimal) *forbidden* configuration in $\mathcal{C}$ is a (minimal) configuration $\mathcal{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_m\}$, $\mathcal{F}_i \subseteq \mathcal{C}$, $|\mathcal{F}_i| \le t$, $1 \le i \le m$, such that

$$\mathrm{desc}(\mathcal{F}_1) = \mathrm{desc}(\mathcal{F}_2) = \cdots = \mathrm{desc}(\mathcal{F}_m),$$

that is, for each $1 \le i \le n$,

$$\mathcal{F}_1(i) = \mathcal{F}_2(i) = \cdots = \mathcal{F}_m(i).$$

From Theorem 5.1.5, we have the following corollary.

Corollary 7.2.1 *An (n,q) code $\mathcal{C}$ is not a t-MIPPC(n,q) if and only if there exists a minimal forbidden configuration in $\mathcal{C}$ with size at most u.*

Now we are going to prove the existence of good t-MIPPC(n,q) for fixed n, t and sufficiently large q. From Corollary 7.2.1, we will show that the minimal forbidden configurations with size at most u are not typical and therefore their probability is small by using the random coding technique. The process of deriving a lower bound for t-MIPPC in Theorem 7.2.2 is to first randomly choose a collection of words from Q^n, and then destroy all the possible minimal forbidden configurations with size at most u by removing one word from each of them.

Theorem 7.2.2 *Let n and t be fixed positive integers such that $n \geq 2, t \geq 2$. Then there exists a constant c, depending only on n and t, with the following property. For all sufficiently large integers q, there exists a t-MIPPC(n, q) with size $cq^{\frac{tn}{2t-1}}$.*

Proof: Let $Q = \{0, 1, \ldots, q-1\}$ be a set of cardinality q. Choose words $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M \in Q^n$ uniformly and independently at random, where M is an integer to be decided later. Denote $\mathcal{C} := \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}$.

Now we would like to remove some words from $\mathcal{C}$ to avoid the forbidden configurations of size at most u. To this end, we first define the *bad subfamily* as follows. A subfamily $\mathcal{C}_0 \subseteq \mathcal{C}$ is called *bad* if there exist m subsets of $\mathcal{C}_0$, say $\mathcal{F}_1, \ldots, \mathcal{F}_m$, such that

(a) $\mathcal{C}_0 = \bigcup_{1 \leq i \leq m} \mathcal{F}_i$ and $|\mathcal{F}_i| \leq t$ for any $1 \leq i \leq m$;

(b) $\bigcap_{1 \leq i \leq m} \mathcal{F}_i = \emptyset$;

(c) for any $1 \leq i \leq n$, $\mathcal{F}_1(i) = \mathcal{F}_2(i) = \cdots = \mathcal{F}_m(i)$.

Now we are going to find out all bad subfamilies in $\mathcal{C}$ with size at most u and remove one word from each bad subfamily. It is clear that $u \geq 2t$. The following process will be divided into two cases according to the size of a bad subfamily: (1) the size is less than $2t$; (2) the size is not less than $2t$.

Case 1. We first consider the case when the size of a bad subfamily is less than $2t$. Suppose $\mathcal{C}_0 = \{\mathbf{f}_1, \ldots, \mathbf{f}_\delta\} \subseteq \mathcal{C}$ is a bad subfamily, where $2 \leq \delta < 2t$. Then for each coordinate $1 \leq i \leq n$, we have a useful observation:

$$|\mathcal{C}_0(i)| = |\{\mathbf{f}_1(i), \ldots, \mathbf{f}_\delta(i)\}| \leq \frac{\delta}{2}. \tag{7.2}$$

Indeed, for each $\mathbf{f}_j(i)$, $1 \leq j \leq \delta$, there exists another word $\mathbf{f}_k \in \mathcal{C}_0 \setminus \{\mathbf{f}_j\}$ such that $\mathbf{f}_k(i) = \mathbf{f}_j(i)$. If not, without loss of generality, we may assume that $\mathbf{f}_1(i) \neq \mathbf{f}_k(i)$ for any $2 \leq k \leq \delta$. Since $\mathcal{C}_0$ is a bad subfamily, there exist m subsets $\mathcal{F}_1, \ldots, \mathcal{F}_m$ of $\mathcal{C}_0$ satisfying conditions (a), (b) and (c). By condition (a), there exists a subset $\mathcal{F}_s$, $1 \leq s \leq m$, such that $\mathbf{f}_1 \in \mathcal{F}_s$. Accordingly, $\mathbf{f}_1(i) \in \mathcal{F}_s(i)$. By condition (c), for any $\mathcal{F}_j$, $1 \leq j \leq m$, we have $\mathcal{F}_j(i) = \mathcal{F}_s(i)$. Thus $\mathbf{f}_1(i) \in \mathcal{F}_j(i)$, which implies $\mathbf{f}_1 \in \mathcal{F}_j$. Accordingly,

$$\mathbf{f}_1 \in \bigcap_{1 \leq j \leq m} \mathcal{F}_j \neq \emptyset,$$

which does not satisfy condition (b) and thus contradicts that $\mathcal{C}_0$ is a bad subfamily. Thus (7.2) follows. Actually, the right-hand in inequality (7.2) should be $\lfloor \frac{\delta}{2} \rfloor$, but we omit the floor-function symbol for convenience.

Based on the above observation, we estimate the probability of the event that a given δ-subfamily of $\mathcal{C}$ forms a bad subfamily. In the following estimation, we always consider the case that q is much larger than n and t. Let $\mathcal{C}_0 \subseteq \mathcal{C}$ such that $|\mathcal{C}_0| = \delta$, $2 \leq \delta < 2t$. Clearly, there are $\binom{M}{\delta}$ distinct such $\mathcal{C}_0$ in $\mathcal{C}$. Let $X(\mathcal{C}_0)$ be the event that $\mathcal{C}_0$ forms a bad subfamily. For each $1 \leq i \leq n$, the i-th coordinates $\mathcal{C}_0(i)$ contributes to the event $X(\mathcal{C}_0)$ with the probability at most

$$\binom{q}{\delta/2}(\delta/2)^{\delta}/q^{\delta}.$$

Since each coordinate contributes to the event $X(\mathcal{C}_0)$ independently, hence we have

$$Pr(X(\mathcal{C}_0)) \leq \binom{q}{\delta/2}^{n}(\delta/2)^{\delta n}/q^{\delta n}.$$

Therefore, the expectation number of bad subfamilies in $\mathcal{C}$ with size less than $2t$ is at most

$$\sum_{2\leq\delta<2t}\binom{M}{\delta}\binom{q}{\delta/2}^{n}(\delta/2)^{\delta n}/q^{\delta n}. \tag{7.3}$$

Case 2. Next we consider the case when the size of a bad subfamily is not less than $2t$. Suppose $\mathcal{C}_0 = \{\mathbf{f}_1, \ldots, \mathbf{f}_\gamma\} \subseteq \mathcal{C}$ is a bad subfamily, where $2t \leq \gamma \leq u$. Similarly, for each $1 \leq i \leq n$, we have an observation

$$|\mathcal{C}_0(i)| = |\{\mathbf{f}_1(i), \ldots, \mathbf{f}_\gamma(i)\}| \leq t. \tag{7.4}$$

If not, then $|\mathcal{C}_0(i)| \geq t+1$. Without loss of generality, we may assume that $|\mathcal{C}_0(i)| = |\{\mathbf{f}_1(i), \ldots, \mathbf{f}_{t+1}(i)\}| = t+1$. Since $\mathcal{C}_0$ is a bad subfamily, there exist m subsets $\mathcal{F}_1, \ldots, \mathcal{F}_m$ of $\mathcal{C}_0$ satisfying conditions (a), (b) and (c). By condition (a), we know that $\mathcal{F}_1(i)$ contains at most t elements in $\mathcal{C}_0(i)$, and there exists another $\mathcal{F}_j$, $j \neq 1$, such that

$$\mathcal{F}_j(i) \cap (\mathcal{C}_0(i) \setminus \mathcal{F}_1(i)) \neq \emptyset.$$

However, this implies $\mathcal{F}_j(i) \neq \mathcal{F}_1(i)$, which contradicts condition (c) and implies that $\mathcal{C}_0$ is not a bad subfamily. Thus (7.4) follows.

Now we are going to estimate the probability of the event that a given γ-subfamily of $\mathcal{C}$ forms a bad subfamily. Let $\mathcal{C}_0 \subseteq \mathcal{C}$ such that $|\mathcal{C}_0| = \gamma$, $2t \leq \gamma \leq u$. Clearly, there are $\binom{M}{\gamma}$ distinct such $\mathcal{C}_0$ in $\mathcal{C}$. Let $X(\mathcal{C}_0)$ be the event that $\mathcal{C}_0$ forms a bad subfamily. For each $1 \leq i \leq n$, the i-th coordinates $\mathcal{C}_0(i)$ contributes to the event $X(\mathcal{C}_0)$ with the probability at most

$$\binom{q}{t}t^{\gamma}/q^{\gamma}.$$

Since each coordinate contributes to the event $X(\mathcal{C}_0)$ independently, hence we have

$$Pr(X(\mathcal{C}_0)) \leq \binom{q}{t}^{n}t^{\gamma n}/q^{\gamma n}.$$

Therefore, the expectation number of bad subfamilies in $\mathcal{C}$ with size $\geq 2t$ and $\leq u$ is at most

$$\sum_{2t \leq \gamma \leq u} \binom{M}{\gamma} \binom{q}{t}^n t^{\gamma n}/q^{\gamma n}. \tag{7.5}$$

Now, we form a set $\mathcal{B}$ by choosing one word from each bad subfamily in $\mathcal{C}$ of size ≥ 2 and $\leq u$. Then from (7.3) and (7.5), we have

$$|\mathcal{B}| \leq \sum_{2 \leq \delta < 2t} \binom{M}{\delta} \binom{q}{\delta/2}^n (\delta/2)^{\delta n}/q^{\delta n} + \sum_{2t \leq \gamma \leq u} \binom{M}{\gamma} \binom{q}{t}^n t^{\gamma n}/q^{\gamma n}.$$

Define $\hat{\mathcal{C}} = \mathcal{C} \setminus \mathcal{B}$. Clearly, any two words in $\hat{\mathcal{C}}$ are distinct, since we removed one word from each bad subfamily of size 2. Moreover,

$$\begin{aligned} |\hat{\mathcal{C}}| &= |\mathcal{C}| - |\mathcal{B}| \\ &\geq M - \sum_{2 \leq \delta < 2t} \binom{M}{\delta} \binom{q}{\delta/2}^n (\delta/2)^{\delta n}/q^{\delta n} - \sum_{2t \leq \gamma \leq u} \binom{M}{\gamma} \binom{q}{t}^n t^{\gamma n}/q^{\gamma n}. \end{aligned} \tag{7.6}$$

We claim that $\hat{\mathcal{C}}$ is a t-MIPPC. Since if not, by Corollary 7.2.1, there exists a minimal forbidden configuration with size at most u. Correspondingly, there exists a bad subfamily of size at most u. But we have already destroyed all bad subfamilies in $\mathcal{C}$ by removing one word from it. Thus there does not exist any forbidden configuration with size at most u in $\hat{\mathcal{C}}$. Hence $\hat{\mathcal{C}}$ is a t-MIPPC.

For sufficiently large q, let $M = \varepsilon q^{\frac{tn}{2t-1}}$, where ε is a constant chosen appropriately and depending only on n and t. Substituting it into (7.6), we have

$$\begin{aligned} |\hat{\mathcal{C}}| &\geq M - \sum_{2 \leq \delta < 2t} \binom{M}{\delta} \binom{q}{\delta/2}^n (\delta/2)^{\delta n}/q^{\delta n} - \sum_{2t \leq \gamma \leq u} \binom{M}{\gamma} \binom{q}{t}^n t^{\gamma n}/q^{\gamma n} \\ &\geq M - \kappa_1 \sum_{2 \leq \delta < 2t} M^\delta q^{-\delta n/2} - \kappa_2 \sum_{2t \leq \gamma \leq u} M^\gamma q^{(t-\gamma)n} \\ &= M - \kappa_2 M^{2t} q^{-tn} - \kappa_1 \sum_{2 \leq \delta \leq 2t-1} M^\delta q^{-\delta n/2} - \kappa_2 \sum_{\substack{2t \leq \gamma \leq u, \\ \gamma \neq 2t}} M^\gamma q^{(t-\gamma)n} \\ &\geq \varepsilon q^{\frac{tn}{2t-1}} - \kappa_2 \varepsilon^{2t} q^{\frac{tn}{2t-1}} - \kappa_1' q^{\frac{n}{2}} - \kappa_2' q^{\frac{n}{2t-1}} \\ &\geq c q^{\frac{tn}{2t-1}}, \end{aligned}$$

where $\kappa_1, \kappa_2, \kappa_1', \kappa_2'$ and c are constants depending only on n and t. The theorem follows. □

7.2.2 Analysis of the new lower bound for t-MIPPC

We compare the asymptotic code rate of a t-MIPPC between the new lower bound in Theorem 7.2.2 and the known upper bound in Corollary 7.1.5 for several cases,

(n,t)	$(2,3)$	$(4,4)$	$(6,5)$	$(8,6)$	$(9,7)$	$(13,10)$	$(15,11)$	$(17,12)$
$R_m(n,t) \geq$	0.6	0.571	0.556	0.545	0.538	0.526	0.524	0.522
$R_m(n,t) \leq$	0.667	0.625	0.583	0.563	0.571	0.554	0.545	0.544

Table 7.1: A comparison of Theorem 7.2.2 and Corollary 7.1.5

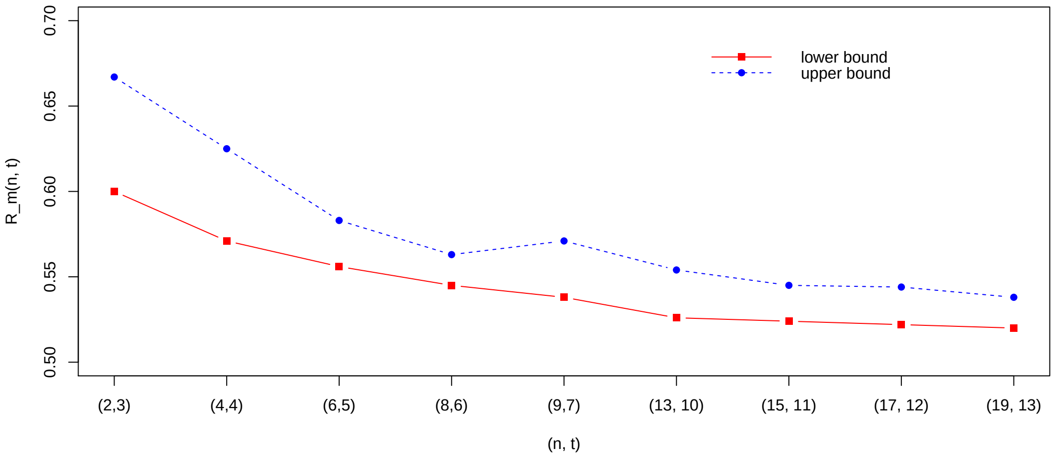

Figure 7.1: A comparison of Theorem 7.2.2 and Corollary 7.1.5

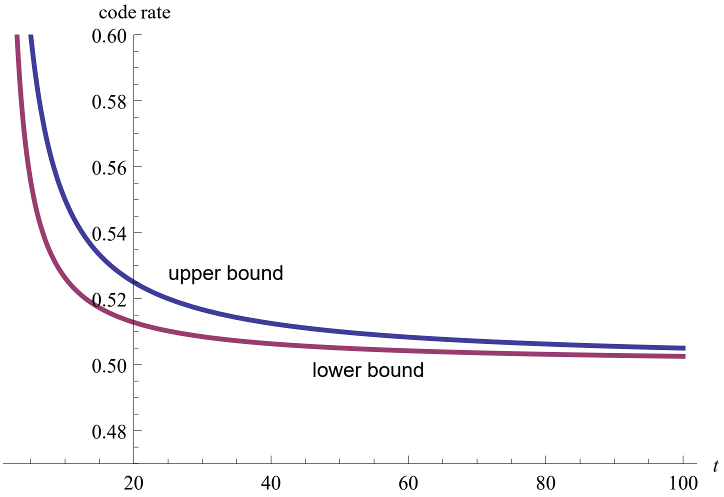

Figure 7.2: A comparison of upper bound and lower bound for t-MIPPCs

see Table 7.1 and Figure 7.1. Figure 7.2 also shows that the gap between the known upper bound and the new lower bound is very small, especially for large t.

Furthermore, we compare the code rate of a t-MIPPC with another kind of multimedia fingerprinting code, namely, $\bar{t}$-separable code. Separable code was introduced in [24] and was studied by several authors, see [16, 21, 23, 24, 47] for example.

Definition 7.2.3 *An (n, q) code $\mathcal{C}$ is called a $\bar{t}$-separable code, denoted $\bar{t}$-SC(n, q), if for any two subcodes $\mathcal{C}_1 \subseteq \mathcal{C}$, $\mathcal{C}_2 \subseteq \mathcal{C}$ such that $|\mathcal{C}_1| \leq t$, $|\mathcal{C}_2| \leq t$ and $\mathcal{C}_1 \neq \mathcal{C}_2$, we have*

$$desc(\mathcal{C}_1) \neq desc(\mathcal{C}_2),$$

that is, there is at least one coordinate i, $1 \leq i \leq n$, such that $\mathcal{C}_1(i) \neq \mathcal{C}_2(i)$.

The relationship between $\bar{t}$-SC and t-MIPPC is as follows.

Proposition 7.2.4 ([22]) *A $\bar{t}$-SC(n, q) is a t-MIPPC(n, q).*

Proof: The relationship is directly from Definition 7.1.2 and Definition 7.2.3. □

By Definition 7.2.3, the 3-MIPPC$(4, 4)$ shown in Example 7.1.3 is also a $\bar{3}$-SC$(4, 4)$. However, the converse of Proposition 7.2.4 is not always true. That is, a t-MIPPC(n, q) may not be a $\bar{t}$-SC(n, q), which can be seen in the following example.

Example 7.2.5 *(A t-MIPPC may not be a $\bar{t}$-SC.)*

Let

$$\mathcal{C} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

where the column vectors are codewords. By Definition 7.1.2, $\mathcal{C}$ is a 3-MIPPC$(3, 2)$ of size $|\mathcal{C}| = 4$.

However, $\mathcal{C}$ is not a $\bar{3}$-SC$(3, 2)$. In fact, let $\mathcal{C}_1 = \{100, 010\} \subseteq \mathcal{C}$, $\mathcal{C}_2 = \{100, 010, 110\} \subseteq \mathcal{C}$, then $|\mathcal{C}_1| = 2 \leq 3$, $|\mathcal{C}_2| = 3$, $\mathcal{C}_1 \neq \mathcal{C}_2$ but

$$desc(\mathcal{C}_1) = desc(\mathcal{C}_2) = \{0, 1\} \times \{0, 1\} \times \{0\},$$

which contradicts Definition 7.2.3.

When $t = 2$, Cheng *et al.* [22] proved that a 2-MIPPC is exactly a $\bar{2}$-separable code. In Theorem 7.2.2, for fixed n such that $n \equiv 0 \pmod 3$ and sufficiently large q, the asymptotic code rate of a 2-MIPPC can be at least $2/3$, which matches the asymptotically optimal code rate of $\bar{2}$-separable code in [47] and [16]. This implies that, in Theorem 7.2.2, the probabilistic expurgation method provides the 2-MIPPC with asymptotically optimal code rate when n is a multiple of 3.

For fixed $n \geq 2, t \geq 3$ and sufficiently large q, Blackburn [16] proved that, when $(t-1)|n$, the asymptotically optimal code rate of a $\bar{t}$-separable code is $1/(t-1)$. However, the fact

$$\frac{t}{2t-1} > \frac{1}{t-1}$$

implies that a t-MIPPC may provide much more codewords than a $\bar{t}$-separable code, which is illustrated in Figure 7.3.

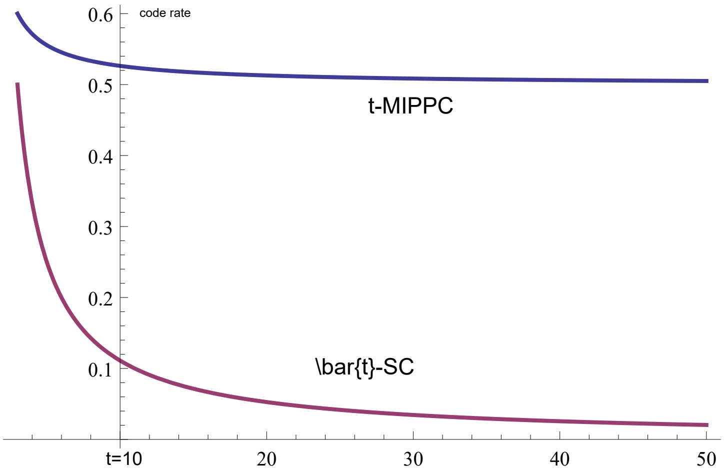

Figure 7.3: A comparison of lower bounds for t-MIPPC and $\bar{t}$-SC

7.3 Summary

In this chapter, we concentrated on a kind of anti-collusion codes for multimedia fingerprinting, that is, MIPPCs. We proved the first probabilistic lower bound for MIPPCs in Section 7.2. A comparison between the new lower bound and the known upper bound for MIPPCs was also established. However, there is still a gap between the lower bound and the known upper bound for MIPPCs in many cases. Clearly, the possible way to close the gap is to improve either the lower bound or the upper bound, or both.

Chapter 8

Conclusions and Open Problems

In this chapter, we draw a brief conclusion of new results obtained in this dissertation, and also propose several interesting open problems.

8.1 Conclusions

In this dissertation, we investigated the anti-collusion schemes for broadcast encryption and the anti-collusion codes for multimedia fingerprinting. We derived new upper and lower bounds for the anti-collusion combinatorial structures and also provided constructions to achieve the new bounds, which greatly improve the previously known results. In the following, we briefly list the new results in this dissertation.

▶ **Traceability schemes**

- We found a very interesting relationship between traceability schemes and cover-free families.
- We derived new upper bounds for traceability schemes, which greatly improve the known results. As a side benefit, we improved the upper bound of cover-free family in [39] by Erdős, Frankl and Füredi.
- We constructed the optimal t-traceability schemes(w, v) by means of sunflowers for the case $w \leq t^2$.
- We affirmed that some constructions of traceability schemes by Stinson and Wei in [85] are in fact optimal. Moreover, we generalized Stinson-Wei's construction to produce more infinite families of optimal traceability schemes achieving our new upper bounds by virtue of combinatorial designs.
- We provided a constructive lower bound for traceability schemes, which has the same order of magnitude with the new upper bounds.

▶ **Parent-identifying set systems**

- We introduced a unified concept of parent-identifying schemes to include many combinatorial structures with the parent-identifying property as special cases.

- We established an equivalent relationship between parent-identifying schemes and forbidden configurations. Furthermore, we corresponded the research problems of parent-identifying schemes to a kind of Turán-type problems.
- We derived a new upper bound for parent-identifying set systems, which is much better than the known bound.
- We provided a probabilistic lower bound for parent-identifying set systems, which has the same order of magnitude with the new upper bound for certain cases.
- We analyzed 2-parent-identifying set systems(w, v) and determined the maximum size of them for small w.
- We proved better upper bounds for 2-parent-identifying set systems$(4, v)$ and 3-parent-identifying set systems$(6, v)$ by virtue of the well-known graph removal lemma, respectively.

▶ **Union-intersection-bounded families**

- We proposed a special kind of cover-free families, called union-intersection-bounded families, due to its applications in broadcast encryption.
- We showed upper bounds for union-intersection-bounded families by virtue of techniques in extremal set theory.
- We derived a lower bound for union-intersection-bounded families by means of probabilistic expurgation method.

▶ **Multimedia parent-identifying codes**

- We proved the first lower bound for multimedia parent-identifying codes, which is much larger than the other kinds of multimedia anti-collusion codes.

8.2 Open problems

In this section, we list several interesting open problems related with the topics in this dissertation.

- Are there constructions of traceability schemes not from combinatorial designs? Recently, Egorova and Kabatiansky [37] exploited the constant weight codes to construct 2-traceability schemes. Are there other possible tools?

- We have already determined the exact maximum size of a traceability scheme for many cases in Chapter 4. What is the exact maximum size of a traceability scheme for the other cases?

- How to explicitly construct optimal parent-identifying set systems?

- Is the Conjecture 5.3.5 correct when $\lfloor t^2/4 \rfloor + t$ is not a divisor of w?

- When w is relatively large, such as w is linear with v, how about the size of a t-parent-identifying set system(w, v)?

- Can the Conjecture 5.5.15 be proved positively?

- For 2-parent-identifying set system$(4, v)$, we proved that $I_2(4, v) = o(v^2)$ in Theorem 5.5.7. Is it the case that there exists a 2-parent-identifying set system$(4, v)$ with size $\kappa v^{2-o(1)}$, where κ is a constant?

- How to explicitly construct optimal union-intersection-bounded families?

- Is it possible to determine the exact maximum size of a union-intersection-bounded family when $s+t-1$ is a divisor of d? We have shown it is $\Theta(v^{\frac{d}{s+t-1}})$. How about the coefficient?

- What is the exact order of magnitude of the maximum size of a union-intersection-bounded family when $s + t - 1$ is not a divisor of d? It should be some value between $\frac{d}{s+t-1}$ and $\lceil \frac{d}{s+t-1} \rceil$.

- Is it possible to improve our lower bound for multimedia parent-identifying codes (Theorem 7.2.2) significantly?

- Can the probabilistic lower bounds for parent-identifying set systems, union-intersection-bounded families and multimedia parent-identifying codes be explicitly constructed?

- How about the code rate of a q-ary t-multimedia parent-identifying code for small q?

Bibliography

[1] N. Alon and V. Asodi, Tracing a single user, *Europ. J. Comb.*, vol. 27, pp. 1227–1234, 2006.

[2] N. Alon, G. Cohen, M. Krivelevich, and S. Litsyn, Generalized hashing and parent-identifying codes, *J. Combin. Theory Ser. A*, vol. 104, pp. 207–215, 2003.

[3] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, The algorithmic aspects of the regularity lemma, *J. Algorithms*, vol. 16, pp. 80–109, 1994.

[4] N. Alon, E. Fischer, and M. Szegedy, Parent-identifying codes, *J. Combin. Theory Ser. A*, vol. 95, pp. 349–359, 2001.

[5] N. Alon and J. H. Spencer, The Probabilistic Method, Fourth Edition, Wiley Series in Discrete Mathematics and Optimization, Wiley & Sons, Hoboken, New Jersey, 2016.

[6] N. Alon and U. Stav, New bounds on parent-identifying codes: The case of multiple parents, *Combin., Probab. Comput.,* vol. 13, no. 6, pp. 795–807, 2004.

[7] N. P. Anthapadmanabhan, A. Barg, and I. Dumer, On the fingerprinting capacity under the marking assumption, *IEEE Trans. Inf. Theory,* vol. 54, no. 6, pp. 2678–2689, June 2008.

[8] A. Barg, G. R. Blakley, and G. Kabatiansky, Digital fingerprinting codes: Problem statements, constructions, identification of traitors, *IEEE Trans. Inf. Theory,* vol. 49, pp. 852–865, 2003.

[9] 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. Discrete Math.*, vol. 14, pp. 423–431, 2001.

[10] A. Barg and G. Kabatiansky, A class of I.P.P. codes with efficient identification, *J. Complexity*, vol. 20, pp. 137–147, 2004.

[11] A. Beimel, Secret sharing schemes: A survey, USA, NY, New York: Springer-Verlag, 2011.

[12] O. Berkman, M. Parnas, and J. Sgall, Efficient dynamic traitor tracing, *SIAM J. Computing*, vol. 30, pp. 1802–1828, 2001.

[13] S. R. Blackburn, An upper bound on the size of a code with the k-identifiable parent property, *J. Combin. Theory Ser. A*, vol. 102, pp. 179–185, 2003.

[14] S. R. Blackburn, Combinatorial schemes for protecting digital content, in *Surveys in Combinatorics, 2003 (Bangor),* vol. 307 of *London Math. Soc. Lecture Note Ser.*, pp. 43–78. Cambridge Univ. Press, Cambridge, 2003.

[15] S. R. Blackburn, Frameproof codes, *SIAM J. Discrete Math.,* vol. 16, pp. 499–510, 2003.

[16] S. R. Blackburn, Probabilistic existence results for separable codes, *IEEE Trans. Inf. Theory*, vol. 61, pp. 5822–5827, 2015.

[17] S. R. Blackburn, T. Etzion, and S.-L. Ng, Traceability codes, *J. Combin. Theory Ser. A*, vol. 117, pp. 1049–1057, 2010.

[18] G. R. Blakley, Safeguarding cryptographic keys, in *Proc. of the 1979 AFIPS National Computer Conference*, vol. 48, pp. 313–317, AFIPS Press, 1979.

[19] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, *IEEE Trans. Inf. Theory,* vol. 44, pp. 1897–1905, 1998.

[20] Y. M. Chee and X. Zhang, Improved constructions of frameproof codes, *IEEE Trans. Inf. Theory,* vol. 58, pp. 5449–5453, 2012.

[21] M. Cheng, H.-L. Fu, J. Jiang, Y.-H. Lo, and Y. Miao, New bounds on $\overline{2}$-separable codes of length 2, *Des. Codes Cryptogr.*, vol. 74, pp. 31–40, 2015.

[22] M. Cheng, H.-L. Fu, J. Jiang, Y.-H. Lo, and Y. Miao, Codes with the identifiable parent property for multimedia fingerprinting, *Des. Codes Cryptogr.*, vol. 83, pp. 71–82, 2017.

[23] M. Cheng, L. Ji, and Y. Miao, Separable codes, *IEEE Trans. Inf. Theory*, vol. 58, pp. 1791–1803, 2012.

[24] M. Cheng and Y. Miao, On anti-collusion codes and detection algorithms for multimedia fingerprinting, *IEEE Trans. Inf. Theory*, vol. 57, pp. 4843–4851, 2011.

[25] B. Chor, A. Fiat, and M. Naor, Tracing traitors, in *Cryto'94 (Lecture Notes in Computer Science),* Berlin, Heidelberg, New York: Springer-Verlag, vol. 839, pp. 480–491, 1994.

[26] B. Chor, A. Fiat, M. Naor, and B. Pinkas, Tracing traitors, *IEEE Trans. Inf. Theory*, vol. 46, pp. 893–910, May 2000.

[27] I. J. Cox, J. Kilian, F. T. Leighton, and T. G. Shamoon, Secure spread spectrum watermarking for multimedia, *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[28] C. J. Colbourn and J. H. Dinitz (eds.), The CRC Handbook of Combinatorial Designs, Second Edition, Chapman & Hall/CRC, Boca Raton, Florida, 2007.

[29] M. J. Collins, Upper bounds for parent-identifying set systems, *Des. Codes Cryptogr.*, vol. 51, pp. 167–173, 2009.

[30] D. Conlon and J. Fox, Graph removal lemmas, in *Surveys in Combinatorics 2013*, vol. 409 of *London Math. Soc. Lecture Note Ser.*, pp. 1–49. Cambridge Univ. Press, Cambridge, 2013.

[31] M. Csúrös and M. Ruszinkó, Single-user tracing and disjointly superimposed codes, *IEEE Trans. Inf. Theory,* vol. 51, pp. 1606–1611, 2005.

[32] R. Diestel, Graph Theory, Fifth Edition, Graduate Texts in Mathematics 173, Springer-Verlag, Berlin, Heidelberg, 2016.

[33] R. Dorfman, The detection of defective members of large populations, *Annals of Math. Statistics*, vol. 14, pp. 436–440, 1943.

[34] A. G. D'yachkov and V. V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii,* vol. 18, pp. 7–13, 1982.

[35] A. G. D'yachkov and V. V. Rykov, A survey of superimposed code theory, *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.,* vol. 12, pp. 229–242, 1983.

[36] A. G. D'yachkov, V. V. Rykov, and A. M. Rashad, Superimposed distance codes, *Problems of Control and Inf. Theory*, vol. 18, pp. 237–250, 1989.

[37] E. Egorova ans G. Kabatiansky, Analysis of two tracing traitor schemes via coding theory, in *Proceding of International Castle Meeting on Coding Theory and Applications 2017*, *Coding Theory and Applications*, pp. 84–92, 2017.

[38] P. Erdős, P. Frankl, and Z. Füredi, Families of finite sets in which no set is covered by the union of two others, *J. Combin. Theory Ser. A*, vol. 33, pp. 158–166, 1982.

[39] P. Erdős, P. Frankl, and Z. Füredi, Families of finite sets in which no set is covered by the union of r others, *Israel J. Math.*, vol. 51, pp. 79–89, 1985.

[40] P. Erdős and R. Rado, Intersection theorems for systems of sets, *J. London Math. Soc.*, vol. 35, pp. 85–90, 1960.

[41] P. Z. Fan, M. Darnell, and B. Honary, Superimposed codes for the multiaccess binary adder channel, *IEEE Trans. Inf. Theory*, vol. 41, pp. 1178–1182, 1995.

[42] M. Fernandez, J. Cotrina, M. Soriano, and N. Domingo, A note about the identifier parent property in Reed-Solomon codes, *Computers and Security*, vol. 29, pp. 628–635, 2010.

[43] A. Fiat and T. Tassa, Dynamic traitor tracing, *J. Cryptol.*, vol. 14, pp. 211–223, 2001.

[44] Z. Füredi, Turán type problems, in *Surveys in Combinatorics, 1991 (Guildford),* vol. 166 of *London Math. Soc. Lecture Note Ser.*, pp. 253–300. Cambridge Univ. Press, Cambridge, 1991.

[45] Z. Füredi, Extremal hypergraphs and combinatorial geometry, in *Proceedings of the International Congress of Mathematicians*, Vol. 1, 2 (Zürich, 1994), pp. 1343–1352, Birkhäuser, Basel, 1995.

[46] Z. Füredi, On r-cover-free families, *J. Combin. Theory Ser. A*, vol. 73, pp. 172–173, 1996.

[47] F. Gao and G. Ge, New bounds on separable codes for multimedia fingerpringting, *IEEE Trans. Inf. Theory,* vol. 60, pp. 5257–5262, 2014.

[48] J. A. Garay, J. Staddon, and A. Wool, Long-lived broadcast encryption, in *Advances in Cryptology – Crypto'00, Lecture Notes in Computer Science*, vol. 1880, pp. 335–353, 2000.

[49] S. Glock, D. Kühn, A. Lo, and D. Osthus, The existence of designs via iterative absorption, 2016. [online] Available: https://arxiv.org/abs/1611.06827

[50] Y. Gu, M. Cheng, G. Kabatiansky, and Y. Miao, Probabilistic existence results for parent-identifying schemes, preprint.

[51] Y. Gu and Y. Miao, Bounds on traceability schemes, *IEEE Transactions on Information Theory*, to appear.

[52] Y. Gu and Y. Miao, Union-intersection-bounded families and their applications to broadcast encryption, submitted.

[53] H. D. L. Hollmann, J. H. van Lint, J.-P. Linnartz, and L. M. G. M. Tolhuizen, On codes with the identifiable parent property, *J. Combin. Theory Ser. A*, vol. 82, pp. 121–133, 1998.

[54] T. Huang and C. Weng, A note on decoding of superimposed codes, *J. Comb. Optim.,* vol. 7, pp. 381–384, 2003.

[55] F. K. Hwang and V. T. Sós, Non-adaptive hypergeometric group testing, *Studia Sci. Math. Hungar.,* vol. 22, pp. 257–263, 1987.

[56] J. Jiang, M. Cheng, and Y. Miao, Strongly separable codes, *Des. Codes Cryptogr.*, vol. 72, pp. 303–318, 2016.

[57] H. Jin and M. Blaum, Combinatorial properties for traceability codes using error correcting codes, *IEEE Trans. Inf. Theory,* vol. 53, pp. 804–808, 2007.

[58] S. Jukna, Extremal Combinatorics with Applications in Computer Science, Second Edition, Texts in Theoretical Computer Science, an EATCS series, Springer, 2011.

[59] G. Kabatiansky, Good ternary 2-traceability codes exist, in *Proc. IEEE Symp. Inf. Theory,* Chicago, IL, pp. 203, 2004.

[60] W. H. Kautz and R. C. Singleton, Nonrandom binary superimposed codes, *IEEE Trans. Inf. Theory*, vol. 10, pp. 363–377, 1964.

[61] P. Keevash, Hypergraph Turán problems, in *Surveys in Combinatorics 2011*, Cambridge University Press, pp. 83–140, 2011.

[62] P. Keevash, The existence of designs, 2014. [online] Available: http://arxiv.org/abs/1401.3665

[63] R. Kumar, S. Rajagopalan, and A. Sahai, Coding constructions for blacklisting problems without computational assumptions, in *Advances in Cryptology – Crypto'99, Lecture Notes in Computer Science*, vol. 1666, pp. 609–623, 1999.

[64] T. Laarhoven, J. Doumen, P. Roelse, B. Škorić, and B. de Weger, Dynamic Tardos traitor tracing schemes, *IEEE Trans. Inf. Theory*, vol. 59, pp. 4230–4242, 2013.

[65] T. Lindkvist, J. Lofvenberg, and M. Svanstrom, A class of traceability codes, *IEEE Trans. Inf. Theory*, vol. 48, pp. 2094–2096, Jul. 2002.

[66] J. H. van Lint and R. M. Wilson, A Course in Combinatorics, Cambridge University Press, 2001.

[67] K. J. R. Liu, W. Trappe, Z. J. Wang, M. Wu, and H. Zhao, Multimedia Fingerprinting Forensics for Traitor Tracing, New York: Hindawi, 2005.

[68] J. Löfvenberg and J. Larsson, Comments on 'New results on frameproof codes and traceability schemes', *IEEE Trans. Inf. Theory,* vol. 56, pp. 5888–5889, Nov. 2010.

[69] J. Pearl, Causality: Models, Reasoning, and Inference, Second Edition, Cambridge Univ. Press, New York, 2009.

[70] C. I. Podilchuk and W. Zeng, Image-adaptive watermarking using visual models, *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 525–539, Apr. 1998.

[71] P. Roelse, Dynamic subtree tracing and its application in pay-TV systems, *Int. J. Inf. Security*, vol. 10, pp. 173–187, 2011.

[72] V. Rödl, On a packing and covering problem, *European J. Combin.,* vol. 6, pp. 69–78, 1985.

[73] R. Safavi-Naini and Y. Wang, New results on frame-proof codes and traceability schemes, *IEEE Trans. Inf. Theory,* vol. 47, pp. 3029–3033, Nov. 2001.

[74] R. Safavi-Naini and Y. Wang, Sequential traitor tracing, *IEEE Trans. Inf. Theory,* vol. 49, pp. 1319–1326, May 2003.

[75] A. Shamir, How to share a secret, *Comm. ACM*, vol. 22, pp. 612–613, 1979.

[76] C. Shangguan and G. Ge, New bounds on the number of tests for disjunct matrices, *IEEE Trans. Inf. Theory,* vol. 62, pp. 7518–7521, 2016.

[77] C. Shangguan and G. Ge, Separating hash families: A Johnson-type bound and new constructions, *SIAM J. Discrete Math.,* vol. 30, pp. 2243–2264, 2016.

[78] C. Shangguan, J. Ma, and G. Ge, New upper bounds for parent-identifying codes and traceability codes, *Des. Codes Cryptogr.*, to appear.

[79] C. Shangguan, X. Wang, G. Ge, and Y. Miao, New bounds for frameproof codes, *IEEE Trans. Inf. Theory,* vol. 63, no. 11, pp. 7247–7252, Nov. 2017.

[80] A. Silverberg, J. Staddon, and J. L. Walker, Applications of list decoding to tracing traitors, *IEEE Trans. Inf. Theory,* vol. 49, pp. 1312–1318, 2003.

[81] J. N. Staddon, A combinatorial study of communication, storage and traceability in broadcast encryption systems, *Ph.D. Thesis, University of California at Berkeley,* 1997.

[82] J. N. Staddon, D. R. Stinson, and R. Wei, Combinatorial properties of frameproof and traceability codes, *IEEE Trans. Inf. Theory,* vol. 47, pp. 1042–1049, 2001.

[83] D. R. Stinson, Cryptography: Theory and Practice, Third Edition, Chapman & Hall/CRC, 2006.

[84] D. R. Stinson, T. van Trung, and R. Wei, Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *J. Statist. Planning Inference,* vol. 86, pp. 595–617, 2000.

[85] D. R. Stinson and R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM J. Discrete Math.,* vol. 11, pp. 41–53, 1998.

[86] D. R. Stinson and R. Wei, Generalized cover-free families, *Discrete Math.,* vol. 279, pp. 463–477, 2004.

[87] D. R. Stinson, R. Wei, and L. Zhu, Some new bounds for cover-free families, *J. Combin. Theory Ser. A*, vol. 90, pp. 224–234, 2000.

[88] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, vol. 18, pp. 939–945, 1978.

[89] T. Tassa, Low bandwidth dynamic traitor tracing schemes, *J. Cryptol.*, vol. 18, pp. 167–183, 2005.

[90] V. D. To and R. Safavi-Naini, On the maximal codes of length 3 with the 2-identifiable parent property, *SIAM J. Discrete Math.*, vol. 17, pp. 548–570, 2004.

[91] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu, Anti-collusion fingerprinting for multimedia, *IEEE Trans. Signal Processing*, vol. 51, pp. 1069–1087, 2003.

[92] P. Turán, On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok*, vol. 48, pp. 436–452, 1941.

[93] R. Wei, Traceability schemes, frameproof codes, key distribution patterns and related topics: A combinatorial approach, *Ph.D. Thesis, University of Nebraska,* 1998.

[94] N. Yoosuf, How to re-encrypt without decrypting in the cloud, 2013. [online] http://mohamednabeel.blogspot.jp/2013/03/abe-how-to-re-encrypt-without.html

List of Publications

1. **Y. Gu** and Y. Miao, Bounds on traceability schemes, *IEEE Transactions on Information Theory*, to appear.
2. **Y. Gu** and Y. Miao, Union-intersection-bounded families and their applications to broadcast encryption, submitted.
3. **Y. Gu**, M. Cheng, G. Kabatiansky, and Y. Miao, Probabilistic existence results for parent-identifying schemes, preprint.
4. J. Jiang, **Y. Gu**, M. Cheng, Y. Miao, and D. Wu, Sequential traitor tracing for multimedia fingerprinting, in preparation.