Perfect Hash Families: Constructions and Applications

by

Kyung-Mi Kim

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2003

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Let $\mathbf{A}$ and $\mathbf{B}$ be finite sets with $|\mathbf{A}| = n$ and $|\mathbf{B}| = m$. An (n, m, w)-*perfect hash family* is a collection $\mathcal{F}$ of functions from $\mathbf{A}$ to $\mathbf{B}$ such that for any $\mathbf{X} \subseteq \mathbf{A}$ with $|\mathbf{X}| = w$, there exists at least one $f \in \mathcal{F}$ such that f is one-to-one when restricted to $\mathbf{X}$. Perfect hash families are basic combinatorial structures and they have played important roles in Computer Science in areas such as database management, operating systems, and compiler constructions. Such hash families are used for memory efficient storage and fast retrieval of items such as reserved words in programming languages, command names in interactive systems, or commonly used words in natural languages. More recently, perfect hash families have found numerous applications to cryptography, for example, to broadcast encryption schemes, secret sharing, key distribution patterns, visual cryptography, cover-free families and secure frameproof codes.

In this thesis, we survey constructions and applications of perfect hash families. For constructions, we divided the results into three parts, depending on underlying structure and properties of the constructions: combinatorial structures, linear functionals, and algebraic structures. For applications, we focus on those related to cryptography.

Acknowledgements

I would like to thank my supervisor, Dr. Douglas Stinson, for his suggestion on a thesis topic, his guidance, support inspiration and patience. I would also like to thank my readers, Dr. Alfred Menezes and Dr. David McKinnon, for their valuable, helpful comments.

I would like to thank Dr. Simon R. Blackburn and Dr. Mustafa Atici for sending their papers at the right times, Dr. Michael J. Best for his encoragement in my graduate studies, and James Muir for answering all my questions about the practical computer matters of writing this thesis.

I would like to thank CACR, the Centre for Applied Cryptogrphic Research and all its members, MIC, the Ministry of Information and Communication, Korea IT Industry Promotion Agency, the Department of Combinatorics and Optimization and all its faculty members, many of my friends, and fellow graduate students for their various support and attention during writing this thesis.

Most of all, I would like to thank my family for their unending love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Perfect Hash Families

In this chapter, definitions and examples of perfect hash families are given, and then we will discuss fundamental properties of perfect hash families such as the bounds based on the results by Mehlhorn [22]. In addition, we will deal with the reasons that perfect hash families are studied with a focus on history and direction. Finally, we will introduce the organization of this thesis.

1.1 What is a Perfect Hash Family?

Let n, m and w be integers such that $n \geq m \geq w \geq 2$. Let $\mathbf{A}$ and $\mathbf{B}$ be finite sets with $|\mathbf{A}| = n$ and $|\mathbf{B}| = m$. Throughout the whole thesis, we will use these notations. An (n, m) *hash function* is any function $f : \mathbf{A} \to \mathbf{B}$. A hash function with finite domain is also known as a compression function. An (n, m) *hash family* is a collection $\mathcal{F}$ of functions such that each $f \in \mathcal{F}$ is an (n, m) hash function.

Definition 1.1 *An (n, m, w)-perfect hash family $\mathcal{F}$ is a collection of functions such that*

$$f : \mathbf{A} \to \mathbf{B}$$

for each $f \in \mathcal{F}$, where $|\mathbf{A}| = n$, $|\mathbf{B}| = m$, and for any $\mathbf{X} \subseteq \mathbf{A}$ such that $|\mathbf{X}| = w$, there exists at least one $f \in \mathcal{F}$ such that $f|_{\mathbf{X}}$ is injective.

In the above definition, the notation '$|_{\mathbf{X}}$' is used to denote the restriction to the set $\mathbf{X}$. We say that a function $f : \mathbf{A} \to \mathbf{B}$ *separates* $\mathbf{X} \subseteq \mathbf{A}$ if f is injective when restricted to $\mathbf{X}$. From now on, we will use the notation PHF$(N; n, m, w)$ for an (n, m, w)-perfect hash family with $|\mathcal{F}| = N$. Also, we will let $N(n, m, w)$ denote the minimum value N such that a PHF$(N; n, m, w)$ exists. If $N = N(n, m, w)$, then PHF$(N; n, m, w)$ is called *optimal*.

Example 1.1 We have a PHF$(4; 9, 3, 3)$. Consider the matrix:

$$\mathcal{M} = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 3 & 1 & 2 & 2 & 3 & 1 \\ 1 & 2 & 3 & 2 & 3 & 1 & 3 & 1 & 2 \end{pmatrix}$$

Let $\mathbf{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $\mathbf{B} = \{1, 2, 3\}$. Hence $n = 9, m = 3$.
For any $i = 1, 2, 3, 4$, Define $f_i(x) =$ the value of entry (i, x) of $\mathcal{M}$. Then we have a $(9, 3, 3)$-perfect hash family, $\mathcal{F} = \{f_i : 1 \leq i \leq 4\}$, as shown in Figure 1.1. In order to check that $\mathcal{F}$ is a PHF$(4; 9, 3, 3)$, for any subset $\mathbf{X} \subseteq \mathbf{A}$ with $|\mathbf{X}| = 3$, we have to find at least one function of $\mathcal{F}$ which separates $\mathbf{X}$. These verifications for all $\binom{9}{3}$ subsets of $\mathbf{A}$ of size 3 are shown in Table 1.1. As shown in Table 1.1, each 3-subset is separated by a function of $\mathcal{F}$. □

x	1	2	3	4	5	6	7	8	9
$f_1(x)$	1	1	1	2	2	2	3	3	3
$f_2(x)$	1	2	3	1	2	3	1	2	3
$f_3(x)$	1	2	3	3	1	2	2	3	1
$f_4(x)$	1	2	3	2	3	1	3	1	2

FIGURE 1.1: A PHF$(4; 9, 3, 3)$

Table 1.1: Verification that $\mathcal{F}$ is a PHF$(4; 9, 3, 3)$

X	i	**X**	i	**X**	i	**X**	i	**X**	i	**X**	i
123	2,3,4	124	3	125	4	126	2	127	4	128	3
129	2	134	4	135	2	136	3	137	3	138	2
139	4	145	4	146	3	147	1,3,4	148	1	149	1
156	2	157	1	158	1	159	1,2,4	167	1	168	1,2,3
169	1	178	3	179	4	189	2	234	2	235	3
236	4	237	2	238	4	239	3	245	3	246	4
247	1	248	1	249	1,2,3	256	1	257	1	258	1,3,4
259	1	267	1,2,4	268	1	269	1	278	4	279	2
289	3	345	2	346	4	347	1	348	1,2,4	349	1
356	3	357	1,2,3	358	2	359	1	367	1	368	1
369	1,3,4	378	2	379	3	389	4	456	2,3,4	457	3
458	4	459	2	467	4	468	2	469	3	478	4
479	3	489	2	567	2	568	3	569	4	578	3
579	2	589	4	678	2	679	4	689	3	789	2,3,4

Note: **X** is a 3-subset of **A** and i represents a function $f_i \in \mathcal{F}$. To save space, we write a subset **X** in the form 123, instead of $\{1, 2, 3\}$.

Let π be a partition of a set $\mathbf{A}$. When the elements of $\mathbf{X} \subseteq \mathbf{A}$ are in distinct parts of π, we say that a set $\mathbf{X}$ is *separated* by a partition π. Naturally, perfect hash families may also be thought of as sets of partitions with some special properties.

Proposition 1.1 *Suppose that* Π *is a set of partitions of a set* $\mathbf{A}$ *with* $|\Pi| = N$ *and for all* $\mathbf{X} \subseteq \mathbf{A}$ *with* $|\mathbf{X}| = w$*, there exists at least one* $\pi \in \Pi$ *that separates* $\mathbf{X}$*. Then there exists a PHF*$(N; n, m, w)$*. Conversely, a PHF*$(N; n, m, w)$ *gives rise to such a set* Π *of partitions of* $\mathbf{A}$*.*

Proof Let $\Pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$ be a set of partitions of a set $\mathbf{A}$. We can construct a collection $\mathcal{F}$ of functions by labeling the parts of each partition π_i with distinct elements of $\mathbf{B}$, and then defining f_i to map each $x \in \mathbf{A}$ to the label of the part of π_i containing x. Then resulting set of functions, say $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$, is an (n, m, w)-perfect hash family.

Conversely, suppose that $\mathcal{F} = \{f_1, \ldots, f_N\}$ is a PHF$(N; n, m, w)$. We can construct a set of partitions of $\mathbf{A}$, say $\Pi = \{\pi_1, \ldots, \pi_N\}$, by setting π_i to f_i for all $i = 1, 2, \ldots, N$. And then for any π_i, $x, y \in \mathbf{A}$ in the same part of π_i whenever $f_i(x) = f_i(y)$. Hence Π is the desired set of partitions of $\mathbf{A}$. □

Example 1.2 Let $\mathbf{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Applying the process described in Proposition 1.1 to the PHF$(4; 9, 3, 3)$ constructed in Example 1.1, we get the following;

$$\pi_1 = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}, \qquad \pi_2 = \{\{1, 4, 7\}, \{2, 5, 8\}, \{3, 6, 9\}\},$$
$$\pi_3 = \{\{1, 5, 9\}, \{2, 6, 7\}, \{3, 4, 8\}\}, \qquad \pi_4 = \{\{1, 6, 8\}, \{2, 4, 9\}, \{3, 5, 7\}\}.$$

Thus, $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4\}$ is the desired set of partitions of $\mathbf{A}$.

Conversely, we can construct, for $i = 1, 2, 3, 4$, by regarding $f_i(x)$ as π_i and for each

$x \in \mathbf{A}$, labeling the part for each partition π_i according to the given order in the above and, mapping $f_i(x)$ to the value j for jth part of π_i. Then $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$ is a $(9, 3, 3)$-perfect hash family. □

Let $\mathcal{M}$ be an $N \times n$ array of m symbols. We say that the ith row of $\mathcal{M}$ *separates* a subset $\mathbf{X}$ of columns of $\mathcal{M}$ if the ith rows of the columns in $\mathbf{X}$ are all distinct. A relationship between a perfect hash family and a certain type of array is given in the following proposition.

Proposition 1.2 *Suppose that there exists a PHF*$(N; n, m, w)$*. Then there exists an array* $\mathcal{M}$*, where size is* $N \times n$ *and which has entries in a set* $\mathbf{B}$ *of size* m*, such that for any subset* $\mathbf{X}$ *of columns of* $\mathcal{M}$ *with* $|\mathbf{X}| = w$*, there is at least one row of* $\mathcal{M}$ *that separates the subset* $\mathbf{X}$ *of columns of* $\mathcal{M}$*. Conversely, such an array gives rise to a PHF*$(N; n, m, w)$*.*

Proof For given an (n, m, w)-perfect hash family $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$, we can produce an array $\mathcal{M}$ of size $N \times n$ with entries in $\mathbf{B}$ as follows: Index the columns of $\mathcal{M}$ by the elements $x \in \mathbf{A}$, and index the rows of $\mathcal{M}$ by the set $\{1, 2, \ldots, N\}$, i.e., each row of the array corresponds to one of the functions in the family $\mathcal{F}$. Setting the value of the entry (i, x) in $\mathcal{M}$ to be $f_i(x)$, the resulting array satisfies the desired conditions.

In the reverse direction, suppose that $\mathcal{M}$ is an array of size $N \times n$, having entries in $\mathbf{B}$. For $i = 1, 2, \ldots, N$ and $x \in \mathbf{A}$, we define $f_i(x)$ to be the value of the entry (i, x) of $\mathcal{M}$. Hence $f_i(x) = f_i(y)$ for $f_i \in \mathcal{F}$ whenever the (i, x)th and (i, y)th entries of $\mathcal{M}$ are equal. Then we have a desired set $\mathcal{F} = \{f_i : 1 \leq i \leq N\}$, which is a PHF$(N; n, m, w)$. □

1.2 Classical Results for the Bounds on the Parameters of a Perfect Hash Family

There are some known bounds for the values $N(n,m,w)$. We can consider the behavior of $(N;n,m,w)$ as a function of n when m and w are fixed. In particular, to obtain the upper bound on the values $N(n,m,w)$, it suffices to present some constructions of PHF$(N;n,m,w)$. And for the lower bound, we will use the non-existence results for the PHF$(N;n,m,w)$. Actually, there are two types of the bounds; those bounds which are good when w is small compared with m and those which are good when w is close to m. In this section, we will introduce the classical results for them based on [22]. We will study further results in Chapter 3. Here and in the sequel all logs are to the base 2 unless otherwise stated.

Theorem 1.1

1. $N(n,m,w) \geq \dfrac{\binom{n}{w} m^w}{\binom{m}{w} n^w}$.
2. $N(n,m,w) \geq \left\lceil \dfrac{\log n}{\log m} \right\rceil$.

Proof

1. Let $\mathcal{F} \subseteq \{f : \mathbf{A} \to \mathbf{B}\}$ be an (n,m,w)-perfect hash family with $|\mathcal{F}| = N$. Note that there are $\binom{n}{w}$ subsets of $\mathbf{A}$ having size w. Suppose that $f \in \mathcal{F}$ separates $\mathbf{X} \subseteq \mathbf{A}$, $|\mathbf{X}| = w$. Then $|f^{-1}(i) \cap \mathbf{X}| \leq 1$ for all $i \in \mathbf{B}$. Hence the number of subsets $\mathbf{X}$ separated by f is

$$\sum_{\mathbf{X} \subseteq \mathbf{A}, |\mathbf{X}| = w} |f^{-1}(i_1)| \times \cdots \times |f^{-1}(i_w)| \quad \text{where} \quad \mathbf{X} = \{i_1, \ldots, i_w\}.$$

This expression is maximized when $|f^{-1}(i_j)| = \frac{n}{m}$ for $j = 1, \ldots, w$. So the maximal value of the sum is equal to $\binom{m}{w}(\frac{n}{m})^w$. Thus the number of distinct hash functions must be at least $\binom{n}{w}m^w / \binom{m}{w}n^w$.

2. Let $\mathcal{F} = \{f_1, \ldots, f_N\} \subseteq \{f : \mathbf{A} \mapsto \mathbf{B}, \text{ and } |\mathbf{A}| = n, |\mathbf{B}| = m\}$. $\mathcal{F}$ induces the following assignment of N dimensional m-ary vectors to the elements of $\mathbf{A}$: the value of the jth component of the vector v_t is $f_j(t)$. In fact, each vector corresponds to a column of the $N \times n$ array $\mathcal{M}$ described in Proposition 1.2. Observe that $v_s \neq v_t$ for $s \neq t$ since s is separated from t by at least one function in perfect hash family $\mathcal{F}$. Thus, $\{v_t : t \in \mathbf{A}\}$ is a set of n distinct m-ary vectors, implying that $N(n, m, w) \geq \left\lceil \frac{\log n}{\log m} \right\rceil$.

□

In the case $w = 2$, an $N \times n$ array $\mathcal{M}$ of m symbols is a PHF$(N; n, m, 2)$ if and only if no two columns of $\mathcal{M}$ are identical. We can easily derive the following result:

Theorem 1.2 *There exists a PHF$(N; n, m, 2)$ if and only if $n \leq m^N$.*

From this theorem, the fact that $N \geq \frac{\log n}{\log m}$ gives the result that N is $\Omega(\log n)$ for $w = 2$ and fixed m.

Theorem 1.3 $N(n, m, w) \leq \left\lceil \frac{\log \binom{n}{w}}{\log(m^w) - \log(m^w - w!\binom{m}{w})} \right\rceil$.

Proof As mentioned above, we can represent $\mathcal{F} = \{f_1, \ldots, f_N\}$ as an $N \times n$ array, say $\mathcal{M}(\mathcal{F}) = (f_i(x))_{x \in \mathbf{A}, 1 \leq i \leq N}$ with m symbols, where each row of the array corresponds the one of the functions in $\mathcal{F}$. Naturally, there are m^{nN} matrices of dimension $N \times n$ with m symbols.

We derive an upper bound on the number of non-perfect arrays. If $\mathcal{F}$ does not contain a perfect hash function for $\mathbf{X} = \{x_1, x_2, \ldots, x_w\}$, then the submatrix of $\mathcal{M}(\mathcal{F})$

given by columns $x_1, \ldots, x_w$ cannot have a row with w different values. So, there are $m^w - m(m-1)\cdots(m-w+1)$ possible rows of that submatrix, and hence the number of such submatrices is bounded above by $[m^w - m(m-1)\cdots(m-w+1)]^N$. Since there are $\binom{n}{w}$ subsets $\mathbf{X}$, the number of non-perfect matrices is bounded above by

$$\binom{n}{w}\left[m^w - m(m-1)\cdots(m-w+1)\right]^N m^{(n-w)N}.$$

There will be a perfect hash family $\mathcal{F}$, $|\mathcal{F}| = N$ provided that

$$\begin{aligned}
& m^{nN} > \binom{n}{w}\left[m^w - w!\binom{m}{w}\right]^N m^{(n-w)N} \\
\iff \qquad & m^{wN} > \binom{n}{w}\left[m^w - w!\binom{m}{w}\right]^N \\
\iff \qquad & N\log m^w \geq \log\binom{n}{w} + N\left[\log(m^w - w!\binom{m}{w})\right] \\
\iff \qquad & N\left[\log m^w - \log(m^w - w!\binom{m}{w})\right] \geq \log\binom{n}{w} \\
\iff \qquad & N \geq \frac{\log\binom{n}{w}}{\log m^w - \log(m^w - w!\binom{m}{w})}.
\end{aligned}$$

$N(n,m,w)$ is an integer, thus we must have

$$N(n,m,w) \leq \left\lceil \frac{\log\binom{n}{w}}{\log(m^w) - \log(m^w - w!\binom{m}{w})} \right\rceil.$$

□

From the above result, we have the following sufficient condition for the existence of a PHF$(N; n, m, w)$ provided that $N \geq we^{w^2/m}\ln n$, which says that N is $O(\log n)$ for fixed w and m.

Corollary 1.1 *If $N \geq we^{w^2/m} \ln n$, then there is an (n, m, w)-perfect hash family $\mathcal{F}$ with $|\mathcal{F}| = N$.*

Proof From the above proof, we observe that there will be a perfect hash family $\mathcal{F}$, $|\mathcal{F}| = N$, provided that

$$m^{nN} > \binom{n}{w} \left[m^w - m(m-1)\cdots(m-w+1) \right]^N m^{(n-w)N}.$$

However

$$\binom{n}{w} \left[m^w - m(m-1)\cdots(m-w+1) \right]^N m^{(n-w)N} < m^{nN}$$

$$\iff \qquad \binom{n}{w} \left[1 - \frac{m(m-1)\cdots(m-w+1)}{m^w} \right]^N < 1$$

$$\iff \qquad N \geq \ln \binom{n}{w} \Big/ -\ln \left[1 - \prod_0^{w-1} (1 - \frac{i}{m}) \right].$$

We can check the following:

$$\ln \binom{n}{w} = \ln \frac{n(n-1)\cdots(n-w+1)}{w(w-1)\cdots 21} \leq w \ln n.$$

By elementary calculus, $-\ln(1-x) \geq x$ for $0 < x < 1$, so

$$-\ln \left[1 - \prod_{i=0}^{w-1} \left(1 - \frac{i}{m} \right) \right] \geq \prod_{i=0}^{w-1} \left(1 - \frac{i}{m} \right),$$

and

$$\prod_{i=0}^{w-1} (1 - \frac{i}{m}) = e^{\sum_{i=0}^{w-1} \ln (1 - \frac{i}{m})}.$$

Now, we have

$$\begin{aligned}
\sum_{i=0}^{w-1} \ln\left(1-\frac{i}{m}\right) &= \int_0^w -\frac{1}{m}\ln\left(1-\frac{x}{m}\right)\mathrm{d}x \\
&= -m\int_0^w -\frac{1}{m}\ln\left(1-\frac{x}{m}\right)\mathrm{d}x \\
&= -m\left[\left(1-\frac{w}{m}\right)-\ln\left(1-\frac{w}{m}\right)+\frac{w}{m}\right] \\
&= -m\left(1-\frac{w}{m}\right)\ln\left(1-\frac{w}{m}\right)-w \\
&\leq -m\left(1-\frac{w}{m}\right)\left(-\frac{w}{m}\right)-w \\
&= -\frac{w^2}{m}.
\end{aligned}$$

By the above observation,

$$\begin{aligned}
-\ln\left[1-\prod_0^{w-1}\left(1-\frac{i}{m}\right)\right] &\geq \prod_{i=0}^{w-1}\left(1-\frac{i}{m}\right) \\
&= e^{\sum_{i=0}^{w-1}\ln\left(1-\frac{i}{m}\right)} \\
&\geq e^{-\frac{w^2}{m}}.
\end{aligned}$$

So,

$$1/-\ln\left[1-\prod_0^{w-1}\left(1-\frac{i}{m}\right)\right] \leq e^{w^2/m}.$$

As a result, we have,

$$\ln\binom{n}{w}\Big/-\ln\left[1-\prod_0^{w-1}\left(1-\frac{i}{m}\right)\right] \leq we^{w^2/m}\ln n.$$

Due to our assumption, $N \geq we^{w^2/m}\ln n$, which completes the proof. □

From those classical results, we know that $N(n,m,w)$ is $\Theta(\log n)$ for fixed m and w.

However, these results are non-constructive. We will study further results related to the bounds on the values $N(n, m, w)$ in Chapter 3.

1.3 Why PHF? History and Direction

Perfect hash families are basic combinatorial structures and they have played important roles in Computer Science, such as database management, operating systems, and compiler constructions. Such hash families are used for memory efficient storage and fast retrieval of items such as reserved words in programming languages, command names in interactive systems, or commonly used words in natural languages. These hash functions should be easily computable and minimize the amount of memory required. More recently, perfect hash families have found numerous applications to cryptography, for example, to broadcast encryption schemes, secret sharing, key distribution patterns, visual cryptography, cover-free families and secure frameproof codes.

We observed that there exists a PHF$(N; n, m, w)$ for fixed m and w in which N is $\Theta(\log n)$, but this result is non-constructive. Since there are several applications of hash families, much attention has been given to finding efficient methods to construct perfect hash families. We will proceed to describe various constructions and applications of perfect hash families.

1.4 Summary of Results

This thesis is divided into two separate parts on perfect hash families-constructions and applications.

In Chapter 2, we will review results on combinatorial structures. First we present

direct constructions using Design theory, Error-correcting codes, and other structures. Then we provide recursive methods to understand PHFs.

In Chapter 3, we will discuss further the bounds on the values $N(n, m, w)$, including the classic results we observed in Chapter 1. And then, by restricting the value m, we will present constructions for linear perfect hash families and some examples. Here we will consider the case when the value m is a large prime power, and greater than the value w. Moreover, we present some other ways to construct PHFs which are due to S. Blackburn.

In Chapter 4, we present theoretical constructions using function fields and algebraic curves over a finite field. The resulting families are linear PHF and we need some assumptions in order to explicitly construct such families. Under some assumptions, we can obtain some examples of perfect hash families. From those examples, we will analyze this method.

In the last Chapter, we consider the various applications of perfect hash families. Especially, we will focus on the cryptographic applications. We will give examples for those applications using the perfect hash families which we constructed or observed throughout the previous chapters.

Chapter 2

Constructions using Combinatorial Structures

In Chapter 1, we observed that there is a close relation between a perfect hash family and a certain combinatorial array. Besides this array, there are many other objects which are closely related to a perfect hash family. In this chapter, we will present further connections between perfect hash families and certain combinatorial structures. As well, we will explain various methods to construct perfect hash families using combinatorial structures.

2.1 Design Theory

In this section, we will introduce constructions using design theory such as, w-separating resolvable block design. For any prime power q, there is a corresponding resolvable block design. Fortunately, there is a connection between those two block designs under certain conditions. This section is based on the paper [4]. We need to introduce some concepts from design theory before we proceed further. Most

basic concepts and fundamental properties of design theory are based on [12] and [33].

Definition 2.1 *A* design *or* set system *is a pair* $(X, \mathcal{A})$ *such that the following properties are satisfied.*

1. X *is a set of elements called* points*, and*
2. $\mathcal{A}$ *is a collection of subsets of* X *called* blocks*.*

Definition 2.2 *Let* $v, k,$ *and* λ *be positive integers such that* $v \geq k \geq 2$*. A* (v, k, λ)-balanced incomplete block design *(which we abbreviate to* (v, k, λ)*-BIBD) is a set system* $(X, \mathcal{A})$ *such that the following properties are satisfied:*

1. $|X| = v$*,*
2. *each block contains exactly* k *points, and*
3. *every pair of distinct points is contained in exactly* λ *blocks.*

We give a few examples of BIBDs now. For simplicity, blocks will be written in the form abc, rather than $\{a, b, c\}$.

Example 2.1

- A $(7, 3, 1)$-BIBD.

$$\begin{aligned} X &= \{1, 2, 3, 4, 5, 6, 7\}, \text{ and} \\ \mathcal{A} &= \{124, 136, 157, 235, 267, 347, 456\}. \end{aligned}$$

- A $(9, 3, 1)$-BIBD.

$$X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \text{ and}$$

$$\mathcal{A} = \{123, 456, 789, 147, 258, 369, 159, 267, 348, 168, 249, 357\}.$$

- A $(10, 4, 2)$-BIBD.

$$X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \}, \text{ and}$$
$$\mathcal{A} = \{0123, 0145, 0246, 0378, 0579, 0689, 1278, 1369, 1479, 1568,$$
$$2359, 2489, 2567, 3458, 3467\}.$$

□

The following are basic necessary conditions for existence of a (v, k, λ)-BIBD.

Theorem 2.1

1. *In a (v, k, λ)-BIBD, every point occurs in exactly $r = \lambda(v-1)/(k-1)$ blocks.*
2. *A (v, k, λ)-BIBD has exactly $b = vr/k = \lambda(v^2 - v)/(k^2 - k)$ blocks.*
3. *(Fisher's inequality) If a (v, k, λ)-BIBD exists, then $b \geq v$ (or, equivalently, $r \geq k$ or $\lambda(v-1) \geq k^2 - k$).*

The value r is often called the *replication number* of the BIBD. Suppose $(X, \mathcal{A})$ is a (v, k, λ)-BIBD. A *parallel class* or *resolution* in $(X, \mathcal{A})$ is a subset of disjoint blocks from $\mathcal{A}$ whose union is X. That is, it is a partition of the point set X. Obviously, a parallel class contains v/k disjoint blocks, and a BIBD can have a parallel class only if $v \equiv 0 \bmod k$. A *resolvable BIBD* is a design whose blocks can be partitioned into parallel classes. In this case, there is a partition of $\mathcal{A}$ into r parallel classes.

Example 2.2 A resolvable $(6, 2, 1)$-BIBD is shown in Figure 2.1.
Let $X = \{0, 1, 2, 3, 4, 5\}$, and $r = \frac{6-1}{2-1} = 5$. Hence there are 5 parallel classes, each consists 3 blocks. In Figure 2.1, each row represents a parallel class. □

$\{0, 1\}$	$\{2, 5\}$	$\{3, 4\}$
$\{0, 2\}$	$\{1, 3\}$	$\{4, 5\}$
$\{0, 3\}$	$\{2, 4\}$	$\{1, 5\}$
$\{0, 4\}$	$\{3, 5\}$	$\{1, 2\}$
$\{0, 5\}$	$\{1, 4\}$	$\{2, 3\}$

FIGURE 2.1: A resolvable $(6, 2, 1)$-BIBD

Definition 2.3 [4] *A w-separating resolvable block design is a pair (X, Π), where the following properties are satisfied:*

1. *X is a finite set of elements which are called points.*
2. *Π is a finite set of parallel classes, each of which is a partition of X(the members of the parallel classes are called blocks).*
3. *For any subset Y of w points, there exists a parallel class $\pi \in \Pi$ such that the w points in Y occur in w different blocks in π.*

Let $|X| = v$, $|\Pi| = r$, $b = \sum_{\pi \in \Pi} |\pi|$, and $m = max\{|\pi| : \pi \in \Pi\}$. Then we denote such a design that satisfies the properties in Definition 2.3 as a w-SRBD (v, b, r, m). PHFs are related to SRBD by the following theorem.

Theorem 2.2 [15] *Suppose that there exists a PHF$(N; n, m, w)$. Then there exists a w-SRBD (n, b, N, m) for some $b \leq Nm$. Conversely, a w-SRBD (v, b, r, m) gives rise to a PHF$(r; v, m, w)$.*

Proof Let $\mathcal{F} = \{f_i : \mathbf{A} \to \mathbf{B}$, and $1 \leq i \leq N\}$ be an (n, m, w)-perfect hash family. We can construct the desired design (X, Π) as follows: Let $X = \mathbf{A}$. If we let $\pi_i = \{f_i^{-1}(j) : j \in \mathbf{B}\}$, then π_i is a partition of X. Hence $\Pi = \{\pi_i : 1 \leq i \leq N\}$ is a set of parallel classes and $|\pi| \leq m$ for all $\pi \in \Pi$. We have $|X| = n$, $|\Pi| = N$, and $b = \sum_{\pi \in \Pi} |\pi| \leq Nm$, where $m = max\{|\pi| : \pi \in \Pi\}$. By the definition of $\mathcal{F}$, for any subset Y of w points, there exists a parallel class $\pi \in \Pi$ such that the w points in Y occur in w different blocks in π. Thus we have a w-SRBD (n, b, N, m) for some $b \leq Nm$.

Conversely, suppose that (X, Π) is a w-SRBD (v, b, r, m). Let $\Pi = \{\pi_i : 1 \leq i \leq r\}$. Define a family $\mathcal{F}$ as follows: For any $1 \leq i \leq r$, Give the index to the blocks in π_i, and then define $f_i(x) = j$ whenever $x \in \mathbf{A}$ is in the jth block in π_i. Clearly, $j \leq m$ and f_i is an (n, m) hash function. Thus the resulting set $\mathcal{F} = \{f_i : 1 \leq i \leq r\}$ is a PHF$(r; v, m, w)$ since for any $x, y \in \mathbf{A}$, x and y are in the same block of π_i if and only if $f_i(x) = f_i(y)$. □

Example 2.3 A PHF$(3; 5, 3, 3)$ is shown in Figure 2.2, which is from [5]. Each row of the array corresponds to one of three functions. From this PHF$(3; 5, 3, 3)$,

1	2	3	3	3
1	1	2	3	3
1	1	1	2	3

FIGURE 2.2: A PHF$(3; 5, 3, 3)$

we can find 3-SRBD$(5, 9, 3, 3)$ as follows: Let $X = \{1, 2, 3, 4, 5\}$ be a set of points and $\Pi = \{\pi_1, \pi_2, \pi_3\}$ a set of parallel classes, where $\pi_i = \{f_i^{-1}(j) : 1 \leq j \leq 3\}$. We

have

$$f_1^{-1}(1) = \{1\}, \quad f_1^{-1}(2) = \{2\}, \quad f_1^{-1}(3) = \{3,4,5\},$$
$$f_2^{-1}(1) = \{1,2\}, \quad f_2^{-1}(2) = \{3\}, \quad f_2^{-1}(3) = \{4,5\},$$
$$f_3^{-1}(1) = \{1,2,3\}, \quad f_3^{-1}(2) = \{4\}, \quad f_3^{-1}(3) = \{5\}.$$

Thus,

$$\begin{aligned} \pi_1 &= \{\{1\},\{2\},\{3,4,5\}\}, \\ \pi_2 &= \{\{1,2\},\{3\},\{4,5\}\}, \\ \pi_3 &= \{\{1,2,3\},\{4\},\{5\}\}. \end{aligned}$$

Then the resulting (X, Π) is a 3-SRBD$(5, 9, 3, 3)$ by the above theorem. □

Proposition 2.1 *A resolvable (v, b, r, k, λ)-BIBD is a w-SRBD $(v, b, r, v/k)$ if $r > \lambda\binom{w}{2}$.*

Proof Let $(X, \mathcal{A})$ be a resolvable (v, b, r, k, λ)-BIBD and Π a set of parallel classes. To prove that (X, Π) is a w-SRBD $(v, b, r, v/k)$, it suffices to show that $(X, \mathcal{A})$ satisfies the property of the Definition 2.3. i.e., for any subset Y of w points, there exists a parallel class $\pi \in \Pi$ such that the w points in Y occur in w different blocks in π. We use a counting argument. Let $\mathbf{Y}$ be a set of w points of X. Suppose that there exists no parallel class $\pi \in \Pi$ separating $\mathbf{Y}$. Then each parallel class can not separate some pair of elements in $\mathbf{Y}$. By the definition of a resolvable (v, b, r, k, λ)-BIBD, we note that any pair of points in X occurs in exactly λ blocks. Thus there are at most λ parallel classes in Π that do not separate a fixed pair of

elements. Hence, there are at most $\lambda\binom{w}{2}$ parallel classes in Π that do not separate $\mathbf{Y}$. By assumption, the number of classes in Π that do not separate all the pairs in $\mathbf{Y}$ is at most $\lambda\binom{w}{2} < r$. Thus there exists at least one parallel class in Π that separates $\mathbf{Y}$, and hence we have a w-SRBD $(v, b, r, v/k)$. □

Corollary 2.1 *Suppose that there exists a resolvable (v, b, r, k, λ)-BIBD with $r > \lambda\binom{w}{2}$. Then there is a PHF$(r; v, v/k, w)$.*

Proof By Theorem 2.1, any resolvable (v, b, r, k, λ)-BIBD with $r > \lambda\binom{w}{2}$ is a w-SRBD $(v, b, r, v/k)$. Then, by Theorem 2.2, we have a PHF$(r; v, v/k, w)$. □

Example 2.4 We presented a resolvable $(6, 15, 5, 2, 1)$-BIBD in Example 2.2. By Proposition 2.1, there is a PHF$(5; 6, 3, 3)$ as shown in Figure 2.3.

□

1	1	2	3	3	2
1	2	1	2	3	3
1	3	2	1	2	3
1	3	3	2	1	2
1	2	3	3	2	1

FIGURE 2.3: A PHF$(5; 6, 3, 3)$

An *affine plane of order n* is an $(n^2, n, 1)$-BIBD. By Theorem 2.1, an affine plane of order n which is a resolvable BIBD has the replication number

$$r = \lambda(v-1)/(k-1) = 1(n^2-1)/(n-1) = n+1$$

and b blocks where

$$b = vr/k = n^2(n+1)/n = n(n+1).$$

And it is well-known that, for any prime power q, there exists an affine plane of order q, i.e., a $(q^2, q, 1)$-BIBD.

Example 2.5 Consider the case $q = 3$, i.e, an affine plane of order 3. We have this $(9, 3, 1)$-BIBD represented in Example 2.1. The parallel classes $\{\Pi_1, \Pi_2, \Pi_3, \Pi_4\}$ of a $(9, 3, 1)$-BIBD are as follows: □

Π_1	$\{1,2,3\}$	$\{4,5,6\}$	$\{7,8,9\}$
Π_2	$\{1,4,7\}$	$\{2,5,8\}$	$\{3,6,9\}$
Π_3	$\{1,5,9\}$	$\{2,6,7\}$	$\{3,4,8\}$
Π_4	$\{1,6,8\}$	$\{2,4,9\}$	$\{3,5,7\}$

FIGURE 2.4: A $(9, 3, 1)$-BIBD

Theorem 2.3 (Bose's Inequality) *If $(X, \mathcal{A})$ is a resolvable BIBD, then $b \geq v + r - 1$.*

When $b = r + v - 1$ (or, equivalently, $r = k + \lambda$), the resolvable BIBD is called an affine resolvable design. Clearly, an affine plane is an affine resolvable design.

Theorem 2.4 *Any two blocks from different parallel classes of an affine resolvable (v, k, λ)-BIBD intersect in exactly k^2/v points.*

Corollary 2.2 *Let w be an integer such that $w \geq 2$, and let q be a prime power such that $q + 1 > \binom{w}{2}$, then there exists a PHF$(q + 1; q^2, q, w)$.*

Proof An affine plane of order q is a resolvable $(q^2, q^2+q, q+1, q, 1)$-BIBD. In this BIBD, $r = q+1 > \binom{w}{2}$, and hence there exists a PHF$(q+1; q^2, q, w)$. □

If we substitute $q = 3$ and $w = 3$ to the above result, then using Example 2.5, we have a PHF$(4; 9, 3, 3)$, which was shown in Figure 1.1. With the above construction, if we consider the relations between two values, say $N(n, m, w)$ and n, then $N \approx n^{\frac{1}{2}} = 2^{\frac{1}{2}\log n}$.

From Chapter 1, we observed that $N(n, m, w)$ is $O(\log n)$ for fixed m and w. For any PHF$(N; n, m, w)$ constructed using the design theory such as w-SRBD, resolvable BIBD and so on, N is $\Omega(n)$. (Any resolvable (v, b, r, k, λ)-BIBD gives to an $(r; v, \frac{v}{k}, w)$-perfect hash family provided that $r > \lambda\binom{w}{2}$. We know that $b = vr/k$ and $b \geq v + r - 1$. Hence we have

$$\frac{vr}{k} \geq v + r - 1 \iff \frac{vr}{k} - r \geq v - 1 \iff r(\frac{v-k}{k}) \geq v - 1$$

$$\iff r \geq \frac{(v-1)k}{v-k} \iff r \geq \frac{v-1}{\frac{v}{k}-1}.$$

If we regard N, n, and m as r, v, and $\frac{v}{k}$, respectively, then, for a PHF$(N; n, m, w)$ derived from a resolvable (v, b, r, k, λ)-BIBD, we have that $N \geq \frac{n-1}{m-1}$.)

The method described in this section gives a simple construction, but it is limited in that it cannot be applied to obtain PHF with arbitrary $m \geq w$. (From the previous observations that $r \geq k + \lambda$, $vr = bk$, and $b = \lambda v(v-1)/k(k-1)$, we have that

$$\frac{r}{\lambda} > \binom{w}{2} \iff \frac{v-1}{k-1} > \frac{w(w-1)}{2} \iff \frac{n-1}{n/m-1} > \frac{w(w-1)}{2}.$$

Thus, for $n \geq m \geq w \geq 2$, as $n \to \infty$, the left side of above last inequality goes to m and the right side approaches to w^2. So we cannot obtain PHF in which m is $O(w)$.)

2.2 Error-Correcting Codes

Error-correcting codes are designed to correct errors in the transmission of data over noisy communication channels. They are widely used in applications such as transmitting pictures from deep space, design of registration numbers, and storage of data on magnetic tape and CDs. In this section, we will present a construction of perfect hash families from error-correcting codes. Then we can obtain other results using some combinatorial structures, because those structures and an error-correcting code are very closely related. Before stating our construction, we need to introduce some concepts and notations for the error-correcting codes.

Definition 2.4

1. *Let l, K, d and q be positive integers and let Q be a set of size q. An (l, K, d, q)-code is a set $\mathcal{C} \subseteq Q^l$ such that $|\mathcal{C}| = K$ and the minimum distance of $\mathcal{C}$ is d.*

2. *The elements of the q-ary code $\mathcal{C}$ are called* codewords.

3. *The (Hamming) distance between two codewords of $\mathcal{C}$, say, $x = x_1x_2\ldots x_l$, and $y = y_1y_2\ldots y_l$ indicates the number of places where they differ. The Hamming distance is denoted by $d(x, y)$. The Hamming distance of $\mathcal{C}$ is $d = min\{d(x, y) : x, y \in \mathcal{C}, x \neq y\}$.*

Example 2.6

1. We have an example of $(5, 4, 3, 2)$ code $\mathcal{C}_1$ as follows: Let $Q = \{0, 1\}$ and $\mathcal{C}_1 = \{x_1, x_2, x_3, x_4\}$, where

$$x_1 = 00000, \quad x_2 = 01101, \quad x_3 = 10110, \quad x_4 = 11011.$$

Then $\min\{d(x_i, x_j) : 1 \leq i < j \leq 4\} = 3$.

2. $\mathcal{C}_2 = \{000, 120, 301, 321, 012, 230, 103, 111, 023, 310, 132, 222, 031, 201, 213, 333\}$ is a $(3, 16, 2, 4)$ code.

□

Because most structures such as Mutually Orthogonal Latin Rectangles, Mutually Orthogonal Latin Squares, Orthogonal Arrays, Transversal Designs we will introduce at the sequential sections can be thought as the suitable error correcting codes, we can derive many applications from the below theorem.

Theorem 2.5 [4] *Let $w \geq 2$ be an integer and suppose $d > l(1 - 1/\binom{w}{2})$. Suppose that there exists an (l, K, d, q) code. Then there exists a PHF$(l; K, q, w)$.*

Proof Let $\mathcal{C}$ be the set of codewords of an (l, K, d, q) code. We write each codeword of $\mathcal{C}$ as $(c_{i1}, c_{i2}, \ldots, c_{iK})$ with $c_{ij} \in \{1, 2, \ldots, q\}$, where $1 \leq i \leq K, 1 \leq j \leq l$. For each j, we define a function f_j from $\{1, 2, \ldots, K\}$ to $\{1, 2, \ldots, q\}$ by the rule, $f_j(i) = c_{ij}$, and, let $\mathcal{F} = \{f_1, \ldots, f_l\}$.
Now let $\mathbf{X}$ be a subset of $\{1, 2, \ldots, K\}$ with $|\mathbf{X}| = w$. Since the minimum distance of the code is d, it follows that given any pair of elements x and y of $\mathbf{X}$, there are at most $l - d$ functions from $\mathcal{F}$ such that the values of these $l - d$ functions evaluated on x and y are the same. Since there are $\binom{w}{2}$ possible pairs of distinct elements from $\mathbf{X}$, it follows that there is at least one function $f \in \mathcal{F}$ such that the values of f on $\mathbf{X}$ are all distinct, provided that $l > \binom{w}{2}(l - d)$. This shows that an (l, K, d, q) code gives rise to a PHF$(l; K, q, w)$ if $d\binom{w}{2} > l\left(\binom{w}{2} - 1\right)$. □

Example 2.7 To illustrate Theorem 2.5,

1. If we let $w = 2$ in $(5, 4, 3, 2)$ code $\mathcal{C}_1$, then $3 > 5(1 - \frac{1}{\binom{w}{2}})$. Thus we can obtain a PHF$(5; 4, 2, 2)$ as shown in Figure 2.5.

0	0	1	1
0	1	0	1
0	1	1	0
0	0	1	1
0	1	0	1

FIGURE 2.5: A PHF$(5; 4, 2, 2)$

2. Similarly, if we let $w = 2$ and consider a $(3, 16, 4, 2)$ code described in Example 2.6, then we can obtain a PHF$(3; 16, 4, 2)$ as shown in Figure 2.6.

0	0	0	0	1	2	3	2	3	1	1	2	3	1	2	3
0	1	2	3	2	3	1	0	0	0	3	1	2	1	2	3
0	2	3	1	0	0	0	1	2	3	2	3	1	1	2	3

FIGURE 2.6: A PHF$(3; 16, 4, 2)$

□

Now we present a method to construct an error-correcting code using polynomials. Let q be a prime, $l \leq q$, and $k < q$. Let $\mathcal{P}(q, l, k)$ be the set of polynomials $a(x) \in \mathbb{Z}_q[x]$ having degree at most $k - 1$. Define the set

$$\mathcal{C}(q, l, k) = \{(a(0), a(1), \ldots, a(l-1)) : a(x) \in \mathcal{P}(q, l, k)\},$$

then $\mathcal{C}(q, l, k)$ is a q-ary code of length l having distance $d = l - k + 1$ and $|\mathcal{C}(q, l, k)| = q^k$.

Corollary 2.3 *Suppose we take p to be prime, $p \geq \binom{w}{2}$, $q = p^j$, and $k = p^{j-1}$, where $j \geq 1$. Then there exists a PHF$(p^{j-1}; p^{jp^{j-1}}, p^j, w)$.*

Proof Using the above construction, we have a $(q-1, q^k, q-k, q)$ code if we let $l = q-1$, for q a prime and $k < q-1$, which is known as *Reed-Solomon Code*, briefly, RS code. Now, if we take p, q, and k as given above, then

$$\frac{d}{l} = \frac{p^j - p^{j-1}}{p^j - 1} = \frac{p^{j-1}(p-1)}{p^j - 1} > \frac{p^{j-1}(p-1)}{p^j} = 1 - \frac{1}{p} \geq 1 - \frac{1}{\binom{w}{2}}.$$

By Theorem 2.5, we can obtain the desired PHF. □

In the above result, if we consider the relation between two values, $N(n, m, w) = N$ and n, then $N = C \log n$ where $C = 1/(j \log p)$.

For any PHF$(N; n, m, w)$ constructed using the error-correction codes, N is $O(n)$. (We use the Plotkin Bound, which states that for any (l, K, d, q) code $\mathcal{C}$ for which $l < 2d$, it holds that $K \leq 2[\frac{d}{2d-l}]$. This implies that

$$\frac{d}{n} \leq \frac{K(q-1)}{(K-1)q}.$$

Combining this with our condition

$$\frac{d}{n} > 1 - \frac{1}{\binom{w}{2}},$$

we see that

$$\begin{aligned} & \frac{k(q-1)}{(k-1)q} > 1 - \frac{1}{\binom{w}{2}} \\ \iff & \frac{1}{\binom{w}{2}} > \frac{k-q}{k(q-1)} \\ \iff & \binom{w}{2} < \frac{(k-1)q}{k-q}. \end{aligned}$$

In the resulting PHF, we have

$$\binom{w}{2} < \frac{(n-1)m}{n-m}.$$

Now, as $n \to \infty$, the right side of this inequality approaches m. So, we cannot construct PHF in which m is $O(w)$, as with the construction using design theory.

2.3 Combinatorial Structures

As mentioned above, most combinatorial structures which we will introduce later can be suitable error-correcting codes. Thus we can construct perfect hash families using the construction with the error-correcting codes.

2.3.1 Latin Rectangle and Latin Squares

Definition 2.5 *For $m \leq n$, let B be a set of size n.*

1. *An $m \times n$ Latin rectangle is an $m \times n$ array M consisting of elements of B, with the property that each row of the array is a permutation of B, and no element of B occurs twice in any column of the array. If $n = m$, then such a rectangle is a Latin square of order n.*

2. *Suppose that M_1 and M_2 are $m \times n$ Latin rectangles with entries from sets B_1 and B_2, respectively, of size n. We say that M_1 and M_2 are orthogonal provided that, for every $x \in B_1$ and for every $y \in B_2$, there is at most one cell (i, j) such that $M_1(i, j) = x$ and $M_2(i, j) = y$.*

3. *A set of $t \geq 2$ Latin rectangles is said to be mutually orthogonal if any two of the t rectangles are orthogonal. We will abbreviate the term 'Mutually*

Orthogonal Latin Rectangles' to 'MOLR', and 'Mutually Orthogonal Latin Squares' to 'MOLS'.

Theorem 2.6 [37] *Suppose there are t MOLR of size $m \times n$. Then there is a PHF$(t+2; mn, n, w)$ provided that $t \geq \binom{w}{2} - 1$. For $m = n$, there are t MOLS of order n, which implies that there exists a PHF$(t+2; n^2, n, w)$ provided that $t \geq \binom{w}{2} - 1$.*

Proof To use Theorem 2.5, it suffices to make an error-correcting code satisfying suitable conditions from the given structure. Let $M_1, M_2, \ldots, M_t$ be t MOLR of size $m \times n$. Consider the set $\mathcal{C} = \{(i, j, M_i(i,j), \ldots, M_t(i,j)) : 1 \leq i \leq n, 1 \leq i \leq m\}$. Then $|\mathcal{C}| = mn$, each element in $\mathcal{C}$ is of length $t+2$ and has components a n-set. Moreover, by the definition of MOLR, the minimum distance is $t+1$. Thus $\mathcal{C}$ is a $(t+2, mn, t+1, n)$ code. By the given assumptions, we have that

$$\frac{t+1}{t+2} = 1 - \frac{1}{t+2} > 1 - \frac{1}{t+1} \geq 1 - \frac{1}{\binom{w}{2}}.$$

By Theorem 2.5, we can obtain a PHF$(t+2; mn, n, w)$. Obviously, in the special case $m = n$, we can derive a PHF$(t+2; n^2, n, w)$ from t MOLS of order n, provided that $t \geq \binom{w}{2} - 1$. □

Example 2.8 The four MOLR of size 5×10 given by [29] are

0123456789	0123456789	0123456789	0123456789
1234567890	3456789012	7890123456	9012345678
2345678901	6789012345	4567890123	8901234567
3456789012	9012345678	1234567890	7890123456
4567890123	2345678901	8901234567	6789012345

From this, we obtain a $(6, 50, 5, 10)$ code $\mathcal{C}$ as follows:

110000	121111	132222	143333	154444
165555	176666	187777	198888	109999
211379	222480	233591	244602	255713
266824	277935	288046	299157	200268
312648	323759	334860	345971	356082
367193	378204	389315	390426	301537
413917	424028	435139	446240	457351
468462	479573	480684	491795	402806
514286	525397	536408	547519	558620
569731	570842	581953	592064	503175

If we put $w = 3$, then $t = 4 \geq \binom{3}{2} - 1$ holds. From this code, we obtain a PHF$(6; 50, 10, 3)$. □

Let $N(m, n)$ be the maximum integer t such that there exists a set of t MOLR of size $m \times n$. Shiue and Mullen [29] present the following construction of orthogonal Latin rectangles:
For $1 \leq k \leq m$, Let $R(k) = (r_{ij})$ be an $m \times n$ Latin rectangle where

$$r_{ij} = ki + j \pmod{n}, \quad 0 \leq i < m,\ 0 \leq j < n.$$

They obtain the following results and provide a list for the values $N(m, n)$.

Theorem 2.7 [29] *For* $2 \leq m \leq n$,

1. $N(2, n) = n - 1$ *for all* n.

2.

$$N(3,n) = \begin{cases} n-1 & \text{if } n \text{ is odd,} \\ \frac{n}{2}-1 & \text{if } n \text{ is even.} \end{cases}$$

3. If p is prime, $N(p,n) = n-1$ if all prime divisors of n are at least p.

4. $N(m,n) = n-1$ if $2 \leq m \leq p$ where p is the smallest prime divisor of n.

5. If $n = {p_1}^{e_1} \cdots {p_r}^{e_r}$ with p_i distinct primes, then

$$N(n,n) = min_{1 \leq i \leq r}\{p_i - 1\}.$$

6. If p is prime and $a \leq b$, $N(p^a, p^b) = p^{b-a+1} - 1$.

7. If $n = pq$ with p and q prime, then $N(q+1, pq) = p-1$.

Now, if we find values t and w satisfying $N(m,n) \geq t \geq \binom{w}{2} - 1$ and $2 \leq w \leq n$ for given values m and n, then we can obtain a PHF $(t+2; mn, n, w)$ by Theorem 2.6.

Example 2.9 By the above theorem, $N(3,9) = 8$ since 3 is the smallest prime divisor of 9. By assigning some suitable values for t and w, we can construct a PHF$(t+2; 27, 9, w)$. First, we can find a value w which satisfies $8 \geq t \geq \binom{w}{2} - 1$. Since $t \leq 8$, we must have $2 \leq w \leq 4$. If we take $w = 4$, then $t \geq \binom{4}{2} - 1$, i.e., $t \geq 5$. Thus, we have a PHF$(5+2; 3 \times 9, 9, 4)$, i.e., we can obtain a PHF$(7; 27, 9, 4)$ using 5 MOLR of order 3×9. □

Using similar processes as with the above example, we can obtain the following Table 2.1 for $m \leq n \leq 30$ by referring to the results in [29]. Table 2.1 shows the possible values w and t in order to obtain a PHF$(t+2; mn, n, w)$ for fixed m and n with $m \leq n \leq 30$.

Table 2.1: Possible Values for a PHF$(t+2; mn, n, w)$

m	n	$N(m,n)$	w	t
2	$3,4,5$	$2,3,4$	3	2
·	$6,7,8,9$	$5,6,7,8$	4	5
·	$10,11,12,13,14$	$9,10,11,12,13$	5	9
·	$15,16,17,18,19,20$	$14,15,16,17,18,19$	6	14
·	$21,22,23,24,25,26,27$	$20,21,22,23,24,25,26$	7	20
·	$28,29,30$	$27,28,29$	8	27
3	3	2	2	2
·	$5,6,8,9,10$	$4,2,4,8,4$	3	2
·	$7,12,14,16,18,21,24,27,30$	$6,5,6,7,8,6,7,8,5$	4	5
·	$11,13,20,22,26,28$	$10,12,9,10,12,13$	5	9
·	$15,17,19$	$14,16,18$	6	14
·	$23,25$	$22,24$	7	20
$3,\cdots,29$	29	28	8	27
4	9	2	2	2
·	$5,8,10,12,15,18$	$4,3,4,3,4,3$	3	2
·	$14,16,21,24,27,30$	$6,7,6,7,8,5$	4	5
·	$11,13,20,22,26,28$	$10,12,9,10,12,13$	5	9
·	$17,19$	$16,18$	6	14
·	$23,25$	$22,24$	7	20
5	$9,12$	$2,2$	2	2
·	$5,10,15,16,18,20,24$	$4,4,4,3,3,4,4$	3	2

continued on next page

continued from previous page

m	n	$N(m,n)$	w	t
·	$7,14,21,27,28,30$	$6,6,6,8,6,5$	4	5
·	$11,13,22$	$10,12,10$	5	9
·	$17,19$	$16,18$	6	14
·	$23,25$	$22,24$	7	20
6	$9,12,15,20,30$	$2,2,2,2,2$	2	2
·	$16,18,24,25$	$3,3,4,4$	3	2
·	$7,14,21,27,28$	$6,6,6,8,6$	4	5
·	$11,13,22,26$	$10,12,10,12$	5	9
·	$17,19$	$16,18$	6	14
·	$23,25$	$22,24$	7	20
7	$9,15,18,20,24$	$2,2,2,2,2$	2	2
·	$16,25$	$3,4$	3	2
·	$7,14,21,28$	$6,6,6,6$	4	5
·	$11,13,20,22,26,28$	$10,12,9,10,12,13$	5	9
·	$17,19$	$16,18$	6	14
·	$23,25$	$22,24$	7	20
8	$9,15,18,20,21,24,28$	$2,2,2,2,2,2,2$	2	2
·	$16,25$	$3,4$	3	2
·	27	8	4	5
·	$11,13,22,26$	$10,12,10,12$	5	9
·	$17,19$	$16,18$	6	14
$8,\cdots,23$	23	22	7	20
9	$9,15,18,20,21,24,28$	$2,2,2,2,2,2$	2	2

continued on next page

continued from previous page

m	n	$N(m,n)$	w	t
·	25	4	3	2
·	27	8	4	5
·	11, 13, 22, 26	10, 12, 10, 12	5	9
·	17, 19	16, 18	6	14
10	15, 20, 23, 24, 27, 28	2, 2, 2, 2, 2, 2	2	2
10, ··· , 25	25	4	3	2
·	27	8	4	5
·	11, 13, 22, 26	10, 12, 10, 12	5	9
·	17, 19	16, 18	6	14
11	15, 21, 24, 27, 28	2, 2, 2, 2	2	2
11	11, 13, 22, 26	10, 12, 10, 12	5	9
·	17, 19	16, 18	6	14
12	15, 21, 24, 27, 28	2, 2, 2, 2, 2	2	2
·	13, 22, 26	12, 10, 12	5	9
·	17, 19	16, 18	6	14
13	15, 21, 27, 28	2, 2, 2, 2	2	2
·	13, 26	10, 12, 12	5	9
·	17, 19	16, 18	6	14
14	15, 21, 27, 28	2, 2, 2, 2	2	2
·	17, 19	16, 18	6	14
15	15, 21, 27	2, 2, 2	2	2
·	17, 19	16, 18	6	14

continued on next page

continued from previous page

m	n	$N(m,n)$	w	t
16	$21, 27$	$2, 2$	2	2
·	$17, 19$	$16, 18$	6	14
17	$21, 27$	$2, 2$	2	2
·	$17, 19$	$16, 18$	6	14
·	23	22	7	20
·	29	28	8	27
18	$21, 27$	$2, 2$	2	2
·	19	18	6	14
19	$21, 27$	$2, 2$	2	2
·	19	18	6	14
20	$21, 27$	$2, 2$	2	2
21	$21, 27$	$2, 2$	2	2
22	$21, 27$	$2, 2$	2	2
$23, \cdots, 27$	27	2	2	2

Note : The case $N(m,n) = 1$ is excluded from the results given in [29] since $t \geq 2$.

It is well-known that $1 \leq N(n) \leq n-1$, where $N(n)$ is the maximum number of MOLS of order n. Specifically, if q is a prime power, then $N(q) = q-1$. Thus, we have a PHF$(q+1; q^2, q, w)$ provided that a prime power q with $q > \binom{w}{2}$ by Theorem 2.6.

Table 2.1 provides many explicit perfect hash families. For example, if we want to find a PHF$(N; n, m, w)$ for fixed $N = 4$ and $w = 3$, then we can refer to the above table to find suitable values n and m. Besides, we know that two MOLS of order m exist for any integer $m \geq 3$, $m \neq 6$. Specifically, for $t = 2$, if w satisfies $t \geq \binom{w}{2} - 1$,

i.e., if $2 \leq w \leq 3$, then we have a PHF$(4; m^2, m, 3)$ for any integer $m \geq 3$, $m \neq 6$.

Example 2.10 Let us illustrate this with an example. Let $m = 3$. We have two MOLS of order 3, say M_1, M_2, as follows:

$$M_1 = \begin{array}{|ccc|}\hline 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \\ \hline \end{array} \qquad M_2 = \begin{array}{|ccc|}\hline 1 & 3 & 2 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \\ \hline \end{array}$$

Then, we can construct a $(4, 9, 3, 3)$ code, say $\mathcal{C}$, as described in Theorem 2.6:

$$\mathcal{C} = \{1111, 1233, 1322, 2123, 2212, 2331, 3132, 3221, 3313\}.$$

This gives rise to a PHF$(4; 9, 3, 3)$ as shown in Figure 2.7, which is isomorphic to the one given in Figure 1.1. □

1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3
1	3	2	2	1	3	3	2	1
1	3	2	3	2	1	2	1	3

FIGURE 2.7: A PHF$(4; 9, 3, 3)$-II

Naturally, we have the following question: How about the case $m = 6$? Is there any method to construct a perfect hash family from the Latin square of order 6? To answer these questions, we need to introduce weaker concept for two Latin squares. Two Latin squares on the same symbols are *r-orthogonal* if, when superimposed, there are exactly r distinct ordered pairs. By the definition of MOLS, two MOLS of order n yields n^2-orthogonal Latin squares of order n.

Theorem 2.8 *Suppose there are r-orthogonal Latin squares of order n. Then there is a PHF$(4; r, n, 3)$.*

Proof Suppose that $M = (m_{ij})_{1 \le i,j \le n}$ and $N = (n_{ij})_{1 \le i,j \le n}$ are r-orthogonal Latin squares of order n. Let T be a set of r ordered pairs (i, j) such that r ordered pairs (m_{ij}, n_{ij}) are all distinct. Then we can obtain a $(4, r, 3, n)$ code as in Theorem 2.6. Since $3 > 4\left(1 - \frac{1}{\binom{3}{2}}\right)$, there is a PHF$(4; r, n, 3)$. □

Theorem 2.9 [12] *For n a positive integer, two r-orthogonal Latin squares of order n exist if and only if $r \in \{n, n^2\}$ or $n + 2 \le r \le n^2 - 2$, except when*

1. *$n = 2$ and $r = 4$;*
2. *$n = 3$ and $r \in \{5, 6, 7\}$;*
3. *$n = 4$ and $r \in \{7, 10, 11, 13, 14\}$;*
4. *$n = 5$ and $r \in \{8, 9, 20, 22, 23\}$;*
5. *$n = 6$ and $r = 36$;*
   and possibly when $r = n^2 - 3$, and
   $n \in \{6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 19, 20, 22, 23, 25, 26\}$.

Although there do not exist two MOLS of order 6, By Theorem 2.9 we know there are two r-orthogonal latin squares of order 6 for $r = 6, 8 \le r \le 34$ and $r \ne 33$. The following example shows the construction of a PHF with two Latin squares of order 6 using Theorem 2.8.

Example 2.11 Here are two Latin squares of order 6, denoted M_1, M_2.

$$M_1 = \begin{array}{|cccccc|}\hline 1&2&3&4&5&6\\ 2&1&4&3&6&5\\ 3&4&5&6&1&2\\ 4&3&6&5&2&1\\ 5&6&1&2&3&4\\ 6&5&2&1&4&3\\ \hline\end{array} \qquad M_2 = \begin{array}{|cccccc|}\hline 1&2&3&4&5&6\\ 2&1&4&5&6&3\\ 3&4&1&6&2&5\\ 4&5&6&1&3&2\\ 5&6&2&3&1&4\\ 6&3&5&2&4&1\\ \hline\end{array}$$

Even though they are not mutually orthogonal, there are exactly 13 distinct ordered pairs, say T, as follows:

$$\begin{aligned} T &= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,4),\\ &\quad (2,6),(3,3),(3,5),(3,6),(4,5),(5,5)\} \end{aligned}$$

that is, M_1 and M_2 yield a 13-orthogonal Latin squares of order 6. Hence, we obtain a $(4,13,5,6)$ code $\mathcal{C}$,

$$\begin{aligned} \mathcal{C} &= \{1111,1222,1333,1444,1555,1666,2435,2653,\\ &\quad 3351,3512,3625,4523,5531\} \end{aligned}$$

Finally, A PHF$(4;13,6,3)$ is shown in Figure 2.8. □

1	1	1	1	1	1	2	2	3	3	3	4	5
1	2	3	4	5	6	4	6	3	5	6	5	5
1	2	3	4	5	6	3	5	5	1	2	2	3
1	2	3	4	5	6	5	3	1	2	5	3	1

FIGURE 2.8: A PHF$(4;13,6,3)$

2.3.2 Orthogonal Arrays

Definition 2.6 *An* orthogonal array*, denoted $OA_\lambda(t,k,v)$, is a $\lambda v^t \times k$ array A on a set of v symbols, such that within any t columns of A every possible t-tuple (not necessarily distinct) of symbols occurs in exactly λ rows of A. For any two rows r_1 and r_2 of A, define*

$$\mu(r_1, r_2) = |\{j : A(r_1, j) = A(r_2, j)\}|$$

and define

$$\mu(A) = max\{\mu(r_1, r_2) : r_1 \neq r_2\}.$$

When $\lambda = 1$ and $t = 2$, we denote such an orthogonal array by $OA(k, v)$.

Example 2.12 An $OA(4,3)$ is shown in Figure 2.9. For this array A, the value of $\mu(A)$ is 1 since no two rows are identical. □

1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	1	2
3	1	3	2
3	2	1	3
3	3	2	1

FIGURE 2.9: An $OA(4,3)$

Theorem 2.10 [37, Theorem 2.12] *Suppose there is an $OA_\lambda(t,k,v)$, say A. Then there exists a PHF$(k; \lambda n^t, v, w)$ provided $k > \mu(A)\binom{w}{2}$.*

Proof Let A be an $OA_\lambda(t,k,v)$. We can construct a v-ary code $\mathcal{C}$ of length k having λv^t codewords by regarding each row of an array as a codeword of a code. Then the minimum distance d is $k-\mu(A)$. Moreover, the parameters of resulting code satisfies the following:

$$\frac{k-\mu(A)}{k} < 1 - \frac{1}{\binom{w}{2}},$$

provided that

$$k > \mu(A)\binom{w}{2} \quad \Longleftrightarrow \quad \frac{k}{\mu(A)} > \binom{w}{2}.$$

Thus, by Theorem 2.5, we have a PHF$(k;\lambda v^t, v, w)$. □

Corollary 2.4 [37, Theorem 2.13] *Suppose there exists an $OA_1(t,k,v)$. Then there exists a PHF $(k;v^t,v,w)$ provided $k > (t-1)\binom{w}{2}$.*

Proof It suffices to show $\mu(A) = t-1$ if $\lambda = 1$. Then, we can use above theorem. If $\mu(A) \geq t$, i.e., there exist two distinct rows r_1 and r_2 such that they have more than t columns having the same values, then this contradicts that condition $\lambda = 1$. Thus $\mu(A) = max\{\mu(r_1,r_2) : r_1 \neq r_2\} = t-1$. □

Let q be a prime power, and $\lambda = 1$. An $OA_\lambda(t,q,q)$ gives rise to a PHF$(q;q^t,q,w)$ provided $q > (t-1)\binom{w}{2}$. In fact, an $OA_\lambda(t,q,q)$ is equivalent to Reed-Solomon code.

Corollary 2.5 [37, Corollary 2.14] *For any prime power q and for any integer t such that $2 \leq t < q$, there exists a PHF$(q;q^t,q,w)$ provided that $(t-1)\binom{w}{2} < q$.*

Proof We observed that an $OA_1(t,q,q)$ gives rise to a PHF$(q;q^t,q,w)$ provided $q > (t-1)\binom{w}{2}$. We have that $1 \leq \binom{w}{2} < \frac{q}{t-1}$, since $w \geq 2$. Thus we have a

PHF$(q; q^t, q, w)$, for $2 \leq t < q$, provided that $(t-1)\binom{w}{2} < q$. □

Corollary 2.6 [37, Corollary 2.15] *For any prime power q, let n, m, and t be any positive integers satisfying $n \geq m$ and $2 \leq t \leq q^n$. Then there exists a PHF$(q^n; q^{m+(t-1)n}, q^m, w)$ provided that $(t-1)\binom{w}{2} < q^m$.*

Proof Suppose that $(t-1)\binom{w}{2} < q^m$. We obtain a PHF$(q^n; q^{m+(t-1)n}, q^m, w)$ using Theorem 2.10. Let $v = q^m$ and let $k = q^n$. Then we want to find the value λ satisfying the following equation:

$$\lambda q^{mt} = q^{m+(t-1)n},$$

i.e.,

$$\lambda = q^{m+(t-1)n-mt} = q^{-m(t-1)+(t-1)n} = q^{(n-m)(t-1)}.$$

Now it suffices to show that $q^n > \mu(A)\binom{w}{2}$. In [6], we have such an orthogonal array A that satisfies $\mu(A) \leq (t-1)q^{n-m}$, that is, $\frac{q^n}{\binom{w}{2}} \geq \frac{q^m}{t-1}$. Then, combining the condition $(t-1)\binom{w}{2} < q^m$ gives $\binom{w}{2} < \frac{q^m}{t-1} \leq \frac{q^n}{\mu(A)}$ which completes the proof. □

2.3.3 Transversal Design

Definition 2.7 *A* transversal design $TD_\lambda(k, m)$ *is triple $(X, \mathcal{G}, \mathcal{A})$ such that following properties are satisfied :*

1. *X is a set of km points,*
2. *$\mathcal{G}$ is a partition of X into k groups of size m,*
3. *$\mathcal{A}$ is a set of λm^2 blocks, each of size k,*

4. *given any two points x, y from different groups, there exist λ blocks containing x and y.*

A *super-simple transversal design* is one in which the intersection of any two blocks has at most two elements.

Example 2.13 A $TD_1(4,3)$:

$$(X, \mathcal{G} = \{G_1, G_2, G_3, G_4\}, \mathcal{A} = \{A_i : 1 \leq i \leq 9\}),$$

where

$$\begin{aligned}
X &= \{a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3, d_1, d_2, d_3\}, \\
G_1 &= \{a_1, a_2, a_3\}, \quad G_2 = \{b_1, b_2, b_3\}, \\
G_3 &= \{c_1, c_2, c_3\}, \quad G_4 = \{d_1, d_2, d_3\}, \\
A_1 &= \{a_1, b_1, c_1, d_1\}, \quad A_2 = \{a_2, b_2, c_2, d_2\}, \quad A_3 = \{a_3, b_3, c_3, d_1\}, \\
A_4 &= \{a_1, b_2, c_3, d_2\}, \quad A_5 = \{a_2, b_3, c_1, d_2\}, \quad A_6 = \{a_3, b_1, c_2, d_2\}, \\
A_7 &= \{a_1, b_3, c_2, d_3\}, \quad A_8 = \{a_2, b_1, c_3, d_3\}, \quad A_9 = \{a_3, b_2, c_1, d_3\}.
\end{aligned}$$

□

Theorem 2.11 [37, Theorem 2.11] *Suppose there exists a super-simple $TD_\lambda(k, m)$. Then there exists a PHF$(k; \lambda m^2, m, w)$ provided $k > w(w-1)$.*

Proof Let $(X, \mathcal{G}, \mathcal{A})$ be a super-simple transversal design, $TD_\lambda(k, m)$. We can construct an m-ary code $\mathcal{C}$ of length k consisting of λm^2 codewords by regarding each block in $\mathcal{A}$ as a codeword in $\mathcal{C}$. And then, the minimum distance d of resulting code $\mathcal{C}$ is at most 2 since the intersection of any two blocks has at most two

elements. Moreover, the condition, $k < w(w-1) = 2\binom{w}{2} \leq d\binom{w}{2}$ gives rise to a PHF$(k; \lambda m^2, m, w)$ by Theorem 2.5. □

Example 2.14 A $TD_1(4,3)$ given in the above example is a super-simple transversal design. From this, we can construct a PHF$(4;9,3,2)$ as follows: First, for simplicity, we replace a_i, b_i, c_i, d_i with i for $i = 1,2,3$. Then we obtain a $(4,9,2,3)$ code $\mathcal{C}$ as shown in Figure 2.10.

$$\mathcal{C} = \{1111, 2222, 3331, 1232, 2312, 3122, 1323, 2133, 3213\}.$$

And then we obtain the desired a PHF$(4;9,3,2)$ as follows: □

1	2	3	1	2	3	1	2	3
1	2	3	2	3	1	3	1	2
1	2	3	3	1	2	2	3	1
1	2	1	2	2	2	3	3	3

FIGURE 2.10: A PHF$(4;9,3,2)$

2.3.4 Difference Matrices

Definition 2.8 [12] *Let $(G, \odot)$ be a group of order n. An $(n,k;\lambda)$-*difference matrix *is a $k \times n\lambda$ matrix $D = (d_{ij})$ with entries from G, so that for each $1 \leq i < j \leq k$, the multiset*

$$\{d_{il} \odot {d_{jl}}^{-1} : 1 \leq l \leq n\lambda\}$$

contains every element of G exactly λ times. When G is abelian, typically additive notation is used, so that differences $d_{il} - d_{jl}$ are employed.

Example 2.15 [12]

1. A $(3, 6; 2)$-difference matrix is shown below:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 2 & 1 & 0 \\ 0 & 2 & 2 & 0 & 1 & 1 \\ 2 & 2 & 0 & 1 & 1 & 0 \\ 2 & 0 & 2 & 1 & 0 & 1 \end{pmatrix}$$

2. A $(15, 5; 1)$- difference matrix is shown below:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 0 \\ 2 & 5 & 7 & 9 & 12 & 4 & 1 & 13 & 10 & 8 & 6 & 3 & 11 & 14 & 0 \\ 6 & 3 & 14 & 10 & 7 & 13 & 4 & 9 & 12 & 1 & 5 & 8 & 2 & 11 & 0 \\ 10 & 6 & 1 & 11 & 2 & 7 & 12 & 5 & 9 & 14 & 4 & 13 & 8 & 3 & 0 \end{pmatrix}$$

□

The following shows the necessary conditions for the existence of $(n, k; \lambda)$ difference matrix.

Theorem 2.12 [12]

1. *An $(n, k; \lambda)$ difference matrix does not exist if $k > \lambda n$.*

2. *An $(n, k; \lambda)$ difference matrix does not exist if $n \equiv 2 \pmod 4$ and λ is odd.*

The following lemma provides a method to construct a difference matrix.

Lemma 2.1 *Let n_0 and w be a positive integers. If* $\gcd(n_0, \binom{w}{2}!) = 1$, *then there exists an* $(n_0, \binom{w}{2} + 1; 1)$*-difference matrix.*

Proof Define the matrix $D = (d_{i,j})$ by

$$d_{i,j} = ij \pmod{n_0}, 0 \leq i \leq n_0 - 1, 0 \leq j \leq \binom{w}{2}.$$

If there exist j_1, j_2 such that $d_{i,j_1} - d_{i,j_2} \equiv d_{h,j_1} - d_{h,j_2} \pmod{n_0}$ for $0 \leq i < h \leq \binom{w}{2}$, then

$$\begin{aligned} ij_1 - ij_2 &\equiv hj_1 - hj_2 \pmod{n_0}, \text{ i.e.,} \\ i(j_1 - j_2) &\equiv h(j_1 - j_2) \pmod{n_0}, \text{ i.e.,} \\ (h - i)(j_1 - j_2) &\equiv 0 \pmod{n_0}. \end{aligned}$$

Since $h - i \in \mathbb{Z}_{\binom{w}{2}!}$ and $\gcd(n_0, \binom{w}{2}!) = 1$, there exists an inverse of $(h - i)$. Hence $j_1 = j_2$. Thus D is an $(n_0, \binom{w}{2} + 1; 1)$-difference matrix. □

Example 2.16 If we substitute $n_0 = 5$ and $w = 3$ in the above lemma, then $\gcd(5, \binom{3}{2}!) = 1$. We obtain $(5, 4; 1)$ difference matrix as shown in Figure 2.11. □

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 2 & 4 & 1 & 3 \\ 0 & 3 & 1 & 4 & 2 \end{pmatrix}$$

FIGURE 2.11: A $(5, 4, 1)$ difference matrix

Theorem 2.13 [7, Theorem 2.2.2] *Suppose there exists an* $(n, k; 1)$*-difference matrix* $D = (d_{i,j})$ *where* $k > \binom{w}{2}$. *Then there exists a PHF*$(k; n^2, n, w)$.

Proof Let $\mathbf{A} = \mathbb{Z}_n \times \mathbb{Z}_n$, $|\mathbf{A}| = n^2$. For $1 \leq i \leq k$, each row $(d_{i,1}, d_{i,2}, \ldots, d_{i,n})$ of D gives rise to a partition of $\mathbf{A}$ into n parts; $(a, b) \in \mathbf{A}$ is in the jth part of the partition if and only if $d_{i,a} + b \equiv j$.

Let $\mathcal{F}$ be any set of these partitions of order $\binom{w}{2} + 1$ as described in Proposition 1.1. Let $\mathbf{X} \subseteq \mathbf{A}$ be a set of w points. Now a partition $\pi \in \mathcal{F}$ fails to separate $\mathbf{X}$ if and only if π fails to separate some pair of elements in $\mathbf{X}$.

Let (a_1, b_1) and (a_2, b_2) be distinct elements in $\mathbf{A}$. We prove that there is at most one partition induced from D that does not separate $\{(a_1, b_1), (a_2, b_2)\}$. For suppose the ith and jth rows of D give rise to partitions of $\mathbf{A}$ that does not separate $\{(a_1, b_1), (a_2, b_2)\}$. Then $d_{i,a_1} + b_1 \equiv d_{i,a_2} + b_2$ and $d_{j,a_1} + b_2 \equiv d_{j,a_2} + b_1$.

But then, $d_{i,a_1} - d_{j,a_1} \equiv (d_{i,a_2} + b_2 - b_1) - (d_{j,a_2} + b_2 - b_1) \equiv d_{i,a_2} - d_{j,a_2} \pmod{n}$.

Since every element of $\mathbb{Z}_n$ occurs exactly once in the vector which is the difference of rows i and j of D, we have that $a_1 = a_2$. But now the equality $d_{i,a_1} + b_1 \equiv d_{i,a_2} + b_2$ implies that $b_1 = b_2$. This is a contradiction, since we are assuming that (a_1, b_1) and (a_2, b_2) are distinct. Hence, there are at most $\binom{w}{2}$ partitions in $\mathcal{F}$ that fail to separate $\mathbf{X}$. Since $|\mathcal{F}| = \binom{w}{2} + 1$, there is a partition in $\mathcal{F}$ that separates $\mathbf{X}$.

Hence, we have an (n^2, n, w)-perfect hash family $\mathcal{F}$ such that $\mathcal{F} = \binom{w}{2}$. Thus, for $k > \binom{w}{2}$, there exists a PHF$(k; n^2, n, w)$. □

Example 2.17 Recalling Example 2.16, we can obtain a PHF$(4; 25, 5, 3)$ as follows: Let

$$A = \{(00), (01), (02), (03), (04), (10), (11), (12), (13), (14),$$

$$(20),(21),(22),(23),(24),(30),(31),(32),(33),$$
$$(34),(40),(41),(42),(43),(44)\}.$$

Each row i, $1 \leq i \leq 4$, is corresponds the partition π_i as follows:

$$\begin{aligned}
\pi_1 &= \{(00),(10),(20),(30),(40)\},\{(01),(11),(21),(31),(41)\},\\
&\quad \{(02),(12),(22),(32),(42)\},\{(03),(13),(23),(33),(43)\},\\
&\quad \{(04),(14),(24),(34),(44)\},\\
\pi_2 &= \{(00),(14),(23),(32),(41)\},\{(01),(10),(24),(33),(42)\},\\
&\quad \{(02),(11),(20),(34),(43)\},\{(03),(12),(21),(30),(44)\},\\
&\quad \{(04),(13),(22),(31),(40)\},\\
\pi_3 &= \{(00),(13),(21),(34),(42)\},\{(01),(14),(22),(30),(43)\},\\
&\quad \{(02),(10),(23),(31),(44)\},\{(03),(11),(24),(32),(40)\},\\
&\quad \{(04),(12),(20),(33),(41)\},\\
\pi_4 &= \{(00),(12),(24),(31),(43)\},\{(01),(13),(20),(32),(44)\},\\
&\quad \{(02),(14),(21),(33),(40)\},\{(03),(10),(22),(34),(41)\},\\
&\quad \{(04),(11),(23),(30),(42)\}.
\end{aligned}$$

□

By Lemma 2.16 and Theorem 2.13, we have following result.

Corollary 2.7 *Let n_0 and w be a positive integers. If $\gcd(n_0, \binom{w}{2}!) = 1$, then there exists a PHF$(\binom{w}{2}+1; {n_0}^2, n_0, w)$.*

Example 2.18 If we substitute $w = 3$ in the above corollary, then there exists a PHF$(4; n^2, n, 3)$ for all n such that $\gcd(n, 3!) = 1$, namely, whenever n is odd and $n \not\equiv 0 \pmod 3$. Of course, we have a PHF$(4; n^2, n, 3)$ for $n \geq 3$ and $n \neq 6$, from the existence of two MOLS of order n. □

In summary, we consider the connections between the combinatorial structures,

Theorem 2.14 [12]

1. *An $OA_\lambda(k, n)$ is equivalent to a $TD_\lambda(k, n)$.*
2. *t MOLS of order n is equivalent to an $OA(t + 2, n)$.*
3. *An $(n, k; \lambda)$-difference matrix over $\mathbb{Z}_n$ gives rise to a resolvable $OA_\lambda(k, n)$ and hence to an $OA_\lambda(k + 1, n)$ and a $TD_\lambda(k + 1, m)$.*

2.4 Recursive Constructions

2.4.1 Recursive Construction I

In this section, first we introduce the Product Theorem which can be used to construct a new perfect hash family from smaller known families. Then we present various results obtained from the Product Theorem.

Theorem 2.15 [7, Theorem 2.3.1] *Suppose the following exist:*

- *a PHF$(N_1; n_0, n_1, w)$,*
- *a PHF$(N_2; n_1, m, w)$.*

Then there is a PHF$(N_1 N_2; n_0, m, w)$.

Proof Let $\mathcal{F}_1 \subseteq \{g : \mathbf{A}_0 \to \mathbf{A}_1\}$ be a PHF$(N_1; n_0, n_1, w)$ and $\mathcal{F}_2 \subseteq \{h : \mathbf{A}_1 \to \mathbf{B}\}$ be a PHF$(N_2; n_1, m, w)$. Define a family of hash functions, say $\mathcal{F} \subseteq \{f : \mathbf{A}_0 \to \mathbf{B}\}$ by $f = h \circ g, \quad g \in \mathcal{F}_1, h \in \mathcal{F}_2$. We show that the resulting family is a PHF$(N_1 N_2; n_0, m, w)$.

Let $\mathbf{X} \subseteq \mathbf{A}_0$ be a set of w points. Because $\mathcal{F}_1$ is an (n_0, n_1, w) perfect hash family, there exists at least one $g \in \mathcal{F}$ such that $g|_{\mathbf{X}}$ is injective. And hence, $\mathbf{Y} = \{g(x) : x \in \mathbf{X}\}$ is a subset of $\mathbf{A}_1$ with $|\mathbf{Y}| = w$. Because $\mathcal{F}_2$ is an (n_1, m, w) perfect hash family, there exists at least one $h \in \mathcal{F}_2$ such that $h|_{\mathbf{Y}}$ is injective. If we let $f = h \circ g$, then f is an one-to-one when restricted to $\mathbf{X}$.

Thus, for any $\mathbf{X} \subseteq \mathbf{A}_0$, $|\mathbf{X}| = w$, there exists at least one $f \in \mathcal{F}$ such that $f|_{\mathbf{X}}$ is injective, and we have a PHF$(N_1 N_2; n_0, m, w)$. □

Example 2.19 To illustrate the above theorem, let $\mathcal{F}_1$ be a PHF$(4, 13, 6, 3)$ as shown in Figure 2.8 and let $\mathcal{F}_2$ be a PHF$(3; 6, 3, 3)$ as indicated in Figure 2.12 given in [3].

1	1	2	2	3	3
1	3	1	2	2	3
1	3	2	3	1	2

FIGURE 2.12: A PHF$(3; 6, 3, 3)$

Then we obtain a PHF$(12; 13, 3, 3)$ by replacing the symbol i of $\mathcal{F}_1$ with the ith column of $\mathcal{F}_2$, which is given in Figure 2.13. □

With Theorem 2.15 and the previous results from this chapter, we are able to obtain many explicit examples of perfect hash families as follows:

1	1	1	1	1	1	1	1	2	2	2	2	3
1	1	1	1	1	1	3	3	1	1	1	2	2
1	1	1	1	1	1	3	3	2	2	2	3	1
1	1	2	2	3	3	2	3	2	3	3	3	3
1	3	1	2	2	3	2	3	1	2	3	2	2
1	3	1	3	1	2	3	2	2	1	2	1	1
1	1	2	2	3	3	2	3	3	1	1	1	2
1	3	1	2	2	3	1	2	2	1	3	3	1
1	3	2	3	1	2	2	1	1	1	3	3	2
1	1	2	2	3	3	3	3	1	1	3	2	1
1	3	1	2	2	3	2	3	1	3	2	1	1
1	3	2	3	1	2	1	2	1	3	1	2	1

FIGURE 2.13: A PHF$(12; 13, 3, 3)$

Theorem 2.16 *Suppose there exists a PHF*$(N_0; n_1, m, w)$,

1. *Suppose there exists a* w*-SRBD*(n_0, b, r, n_1) *with* $r > \lambda\binom{w}{2}$. *Then there exists a PHF*$(rN_0; n_0, m, w)$. *Moreover, suppose that there exists a resolvable* $(n_0, b, r, n_0/n_1, \lambda)$*-BIBD, then there exists a PHF*$(rN_0; n_0, m, w)$.

2. *Suppose there exist* $\binom{w}{2} - 1$ *MOLR of size* $n_0 \times n_1$. *Then there exists a PHF* $((\binom{w}{2} + 1)N_0; n_0 n_1, m, w)$. *Indeed, it holds for the case* $n_0 = n_1$.

3. *Suppose there exists an* $OA_\lambda(t, k, n_0)$, *where* $k > \mu(A)\binom{w}{2}$. *Then there exists a PHF*$(kN_0; \lambda(n_0)^t, m, w)$.

4. *Suppose there exists a super-simple* $TD_\lambda(k, n_0)$, *where* $k > w(w-1)$. *Then there exists a PHF*$(kN_0; \lambda {n_0}^2, m, w)$.

5. *Suppose there is an* $(n_0, \binom{w}{2}+1; 1)$*-difference matrix. Then there is a PHF*$((\binom{w}{2}+1)N_0; {n_0}^2, m, w)$.

6. *Suppose* $\gcd(n_0, \binom{w}{2}!) = 1$. *Then there is a PHF*$((\binom{w}{2} + 1)N_0; {n_0}^2, m, w)$.

The following theorem shows the construction of infinite classes of PHF.

Theorem 2.17 [4] *Suppose there exists a PHF*$(N_0; n_0, m, w)$*, and suppose that* $\gcd(n_0, \binom{w}{2}!) = 1$*. Then there is a PHF*$((\binom{w}{2}+1)^j N_0; {n_0}^{2^j}, m, w)$ *for any integer* $j \geq 1$*.*

Proof By induction on $j \geq 1$. For $j = 1$, the result holds by Theorem 2.16. Suppose the result is true for all $j \leq k$, i.e., there exists a PHF$((\binom{w}{2}+1)^k N_0; {n_0}^{2^k}, m, w)$. There exists a PHF$(\binom{w}{2} + 1; {n_0}^{2^k 2}, {n_0}^{2^k}, w)$ since $\gcd({n_0}^{2^k}, \binom{w}{2}!) = 1$. Thus, by the induction hypothesis and Theorem 2.15, there exists a PHF$((\binom{w}{2} + 1)^{k+1} N_0; {n_0}^{2^{k+1}}, m, w)$. Therefore, there is a PHF$((\binom{w}{2}+1)^j N_0; {n_0}^{2^j}, m, w)$ for any integer $j \geq 1$. □

For some parameters in Theorem 2.17, if we let $N_0 = 3, w = 3, n_0 = 5, m = 3$ as given in [4], then we obtain $N \approx .556(\log n)^2$. Generally speaking, for fixed values m and w, we obtain the relationship

$$N = \frac{N_0}{(\log n_0)^2}(\log n)^2.$$

Corollary 2.8 *There exists a PHF*$(2^{j+3}; 7^{2^{j+1}}, 7, 3)$ *for all* $j \geq 1$*.*

Proof We have an affine plane of order 7, and if we let $w = 3$, then by Corollary 2.2, we obtain a PHF$(8; 49, 7, 3)$. Moreover, $\gcd(49, \binom{3}{2}!) = 1$ implies a PHF$(8 \times 4^j; 49^{2^j}, 7, 3)$ for all $j \geq 1$ by the above theorem, that is, there exists a PHF$(2^{j+3}; 7^{2^{j+1}}, 7, 3)$ for all $j \geq 1$. □

With the above parameters, we have

$$N = 2^{2j+3} = 2\left(\frac{\log n}{\log 7}\right)^2 \approx .253(\log n)^2.$$

In [37], for any given m and w, infinite classes of perfect hash families are constructed in which N is $O(C^{log^*(n)})$, where C is a constant depending on the value w as follows:

Theorem 2.18 [37, Theorem 3.6] *Suppose there exists a PHF* $(N_0; q^{l_0}, m, w)$, *where* q *is a prime power and* $q^{l_0} \geq w^2 - w$. *Then for all* $h \geq 0$, *there exists a PHF* $(N_0 P_h; q^{l_h}, m, w)$, *where* $P_0 = 1$,

$$P_h = q^{l_{h-1}} P_{h-1}, \text{ and}$$

$$l_h = l_{h-1} \left\lfloor q^{l_{h-1}} / \binom{w}{2} \right\rfloor$$

for all $h \geq 1$.

Example 2.20 The following shows the behavior of the parameters of perfect hash families, PHF$(N; n, m, w)$, produced by the above theorem. The results are obtained using Maple.

1. Put $N_0 = 2$, $q = 3$, $l_0 = 1$, and $m = w = 2$, since there exists a PHF$(2; 3, 2, 2)$ by Theorem 1.2. Then

   - When $h = 1$, there exists a PHF$(6; 27, 2, 2)$.
   - When $h = 2$, there exists a PHF$(162, 44342648824303776994824963061914989280 3, 2, 2)$.
   - For $h > 3$, n is a very large integer.

2. Put $N_0 = 4$, $q = 9$, $l_0 = 1$, and $m = w = 3$, since there exists a PHF$(4; 9, 3, 3)$ by Example 1.1. Then

   - When $h = 1$, there exists a PHF$(36; 729, 3, 3)$.

- When $h = 2$, there exists a PHF$(26244; n, 3, 3)$, where n is a very large integer.

3. Put $N_0 = 2$, $q = 5$, $l_0 = 1$, $m = 3$, and $w = 2$, since there exists a PHF$(2; 5, 3, 2)$ by Theorem 1.2. Then

   - When $h = 1$, there exists a PHF$(50; 3125, 3, 2)$.
   - When $h = 2$, there exists a PHF$(156250; n, 3, 2)$, where n is a very large integer.

□

2.4.2 Recursive Construction II

In this section, we first introduce the Kronecker-product type Theorem, and then we apply it to other previous results.

Theorem 2.19 [7, Theorem 2.3.2] *Suppose the following exist:*

- *a PHF*$(N_1; n_0 n_1, m, w)$,
- *a PHF*$(N_2; n_2, n_1, w - 1)$,
- *a PHF*$(N_3; n_2, m, w)$.

Then there is a PHF$(N_1 N_2 + N_3; n_0 n_2, m, w)$.

Proof Let $\mathcal{F}_1 \subseteq \{f : \mathbf{A}_0 \times \mathbf{A}_1 \to \mathbf{B}\}$ be a PHF$(N_1; n_0 n_1, m, w)$, let $\mathcal{F}_2 \subseteq \{g : \mathbf{A}_2 \to \mathbf{A}_1\}$ be a PHF$(N_2; n_2, n_1, w - 1)$, and let $\mathcal{F}_3 \subseteq \{h : \mathbf{A}_2 \to \mathbf{B}\}$ be a

PHF$(N_3; n_2, m, w)$.

For any $(x, y) \in \mathbf{A}_0 \times \mathbf{A}_1$, define $\mathcal{F} = \{\varphi : \mathbf{A}_0 \times \mathbf{A}_2 \to \mathbf{B}\}$ by

$$\begin{aligned} \varphi_h(x, y) : &= h(y) \quad h \in \mathcal{F}_3, \quad \text{and} \\ \varphi_{f,g}(x, y) : &= f(x, g(y)) \quad f \in \mathcal{F}_1, g \in \mathcal{F}_2. \end{aligned}$$

Let $\mathbf{X} = \{(x, y) : x \in \mathbf{A}_0, y \in \mathbf{A}_2\}$ be any subset of $\mathbf{A}_0 \times \mathbf{A}_2$, $|\mathbf{X}| = w$. To prove that $\mathcal{F}$ is a PHF$(N_1N_2 + N_3; n_0n_2, m, w)$, we have to show that there exists at least one $\varphi \in \mathcal{F}$ such that φ is one-to-one when restricted to $\mathbf{X}$. We consider two cases for $\mathbf{X}$.

- **Case I:** Suppose the y-coordinates of $\mathbf{X}$ are all different. Then there must exist φ on $\mathbf{X}$ in which the w district values because of the fact that $\mathcal{F}_3$ is a PHF(N_3, n_2, m, w).

- **Case II:** Suppose that the y-coordinates of $\mathbf{X}$ are chosen from at most $w - 1$ different values, say $y_1, y_2, \ldots, y_v$, $v \leq w - 1$. Using the fact that $\mathcal{F}_2$ is a PHF$(N_2; n_2, n_1, w - 1)$, $g(y_1), g(y_2), \ldots, g(y_v)$ are distinct points of $\mathbf{A}_1$. Now consider the function of $\mathcal{F}_1$. In $\mathcal{F}_1$, the w points under consideration comprise w distinct points of $\mathbf{A}_0 \times \mathbf{A}_1$. Using the fact that $\mathcal{F}_1$ is a PHF$(N_1; n_0n_1, m, w)$, there must exist a function of $\mathcal{F}$ in which the w points have distinct values.

□

Example 2.21 To illustrate the above theorem, let $\mathcal{F}_1$ be a PHF$(3; 6, 3, 3)$ as follows:

	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
f_1	1	1	2	2	3	3
f_2	1	3	1	2	2	3
f_3	1	3	2	3	1	2

Let $\mathcal{F}_2$ and $\mathcal{F}_3$ be a PHF$(4;9,3,2)$ described in 2.10 and $\mathcal{F}_3$ a PHF$(4;9,3,3)$ described in 1.1, respectively. Then we have a PHF$(16;18,3,3)$ as shown in Figure 2.14. □

1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	2	3	3	1	2	1	2	3	1	2	3	3	1	2	1	2	3
1	2	3	2	3	1	3	1	2	1	2	3	2	3	1	3	1	2
1	1	2	1	1	1	1	1	2	2	3	3	2	3	2	2	3	3
1	1	2	1	2	1	2	1	1	2	3	3	3	3	2	3	3	2
1	1	2	2	1	1	1	2	1	2	3	3	3	2	3	3	3	2
1	1	1	1	1	1	2	2	2	2	3	2	3	3	3	3	3	3
1	3	1	1	3	1	1	3	1	2	2	3	2	2	3	2	2	3
1	3	1	3	1	1	1	1	3	2	2	3	2	3	2	3	2	2
1	3	1	1	1	3	3	1	1	2	2	3	3	2	2	2	3	2
1	3	1	3	3	3	1	1	1	2	2	2	2	2	2	3	3	3
1	3	2	1	3	2	1	3	2	3	1	2	3	1	2	3	1	2
1	3	2	3	2	1	2	1	3	3	1	2	1	2	3	2	3	1
1	3	2	3	1	2	3	2	1	3	1	2	2	3	1	1	2	3
1	3	1	3	3	3	2	2	2	3	1	3	1	1	1	2	2	2

FIGURE 2.14: A PHF$(16;18,3,3)$

Corollary 2.9 [4] *Suppose there is a PHF* $(\alpha;m^2,m,3)$. *Then, for any* $k \geq 2$, *there is a PHF*$(\alpha\binom{k}{2};m^k,m,3)$.

Proof By induction on k. For $k=2$, it is clearly true since a $(\alpha;m^2,m,3)$ is given.

By Theorem 1.2, there exists a PHF$(N; n, m, 2)$ if and only if $n \leq m^N$. If we consider the case $m = 3$, then there exists a PHF$(k-1; 3^{k-1}, 3, 2)$. Suppose that there exists a PHF$(\alpha\binom{k-1}{2}; m^{k-1}, m, 3)$. Define the parameters to apply Theorem 2.19 as follows:

$$\begin{aligned} N_1 = 4, \qquad & N_2 = k-1, \quad N_3 = \alpha\binom{k-1}{2}, \\ n_0 = m, \qquad & n_1 = m, \quad n_2 = m^{k-1}. \end{aligned}$$

Then, we have a PHF$(2k^2 - 2k; 3^k, 3, 3)$, since

$$\begin{aligned} N_1N_2 + N_3 &= \alpha(k-1) + \alpha\binom{k-1}{2} \\ &= \alpha(k-1)(1 + \frac{k-2}{2}) \\ &= \frac{1}{2}\alpha k(k-1) \\ &= \alpha\binom{k}{2}. \end{aligned}$$

Thus, there is a PHF $(\alpha\binom{k}{2}; m^k, m, 3)$, for all $k \geq 2$. □

If we substitute $\alpha = 4$ in the above theorem, then there exists a PHF$(4; m^2, m, 3)$ for $m \geq 3$, and $m \neq 6$ as observed before. Moreover, by Corollary 2.9, there exists a PHF$(4\binom{k}{2}; m^k, m, 3)$ for all $k \geq 2$. Thus we can obtain a PHF$(2k^2 - 2k; m^k, m, 3)$, for $m \geq 3, m \neq 6$ and $k \geq 2$.

From Corollary 2.9, we have a PHF$(N; n, m, 3)$ with $N = \alpha\binom{k}{2}$ and $n = m^k$ for fixed m, i.e.,

$$N = \frac{\alpha(k^2 - k)}{2} \approx \frac{\alpha}{2(\log m)^2}(\log n)^2.$$

Corollary 2.10 [4] *Suppose there is a PHF$(\alpha; m^2, m, 3)$ and a PHF $(\beta; m^2, m, 4)$. Then, for any $k \geq 3$, there is a PHF $\left(\beta(\alpha\binom{k}{3} + 1)\right); m^k, m, 4)$.*

Proof By induction on k. For $k = 3$, clearly, there exists a PHF$(\alpha + 1; m^2, m, 3)$. By Theorem 2.9, i.e., there exists a PHF$(\alpha\binom{k-1}{2}; m^k, m, 3)$, for $k \geq 3$. Suppose that there exists a PHF$(\alpha\binom{k-1}{2}; m^{k-1}, m, 3)$. Thus we have following PHFs:

- a PHF$(\beta; m^2, m, 4)$,
- a PHF$\left(\alpha\binom{k-1}{2}; m^{k-1}, m, 3\right)$,
- a PHF$\left(\beta(\alpha\binom{k-1}{3} + 1)\right); m^{k-1}, m, 4)$.

Then,

$$\beta\alpha\binom{k-1}{2} + \beta\left(\alpha\binom{k-1}{3} + 1\right) = \beta\left(\alpha\binom{k}{3} + 1\right).$$

Thus, by Theorem 2.19, there exists a PHF$\left(\beta(\alpha\binom{k}{3} + 1)\right); m^k, m, 4)$. Therefore, there is a PHF $\left(\beta(\alpha\binom{k}{3} + 1)\right); m^k, m, 4)$, for all $k \geq 3$. □

Example 2.22 To illustrate the above results, let p be a prime such that $p \geq 17$ and $p = 11$. And we substitute $\alpha = 4$, $\beta = 6$, and $m = p$, respectively. Then we have a PHF$(6; p^2, p, 4)$ given in [8] and a PHF$(4; p^2, p, 3)$. Thus we can obtain a PHF$(4k(k-1)(k-2) + 6; p^k, p, 4)$ for all $k \geq 3$. With these parameters,

$$N \approx \frac{4}{(\log p)^3}(\log n)^3.$$

□

Chapter 3

Linear Perfect Hash Families

This chapter is organized in three parts. First, we will study some more bounds on the values $N(n, m, w)$ which come from [15], [22], [21], and [10]. Secondly, we will deal with the special classes of linear perfect hash families which S. Blackburn and P. Wild introduced and studied in [10]. Finally, we will summarize some other constructions based on the results by S. Blackburn.

3.1 Further Bounds on the Values $N(n, m, w)$

In Chapter 1, we studied bounds on the values $N(n, m, w)$ based on the results by Mehlhorn [22]. Theorem 1.1 gave a very rough bound on the value $N(n, m, w)$. For many years, improved bounds have been addressed, considering many sets of parameters. Specifically, in this section, we discuss the various lower bounds which give necessary conditions for the existence of PHF.

The following theorem is stated and proved by Fredman-Komlos in [15] by using techniques from information theory and graph theory:

Theorem 3.1 $N(n,m,w) \geq \dfrac{\binom{n-1}{w-2} m^{w-2} \log(n-w+2)}{\binom{m-1}{w-2} n^{w-2} \log(m-w+2)}$.

For brevity, we call this bound FK's bound. Consider the following computation:

$$\frac{\binom{n-1}{w-2} m^{w-2} \log(n-w+2)}{\binom{m-1}{w-2} n^{w-2} \log(m-w+2)}$$

$$= \frac{n(n-1)\cdots(n-w+2)}{m(m-1)\cdots(m-w+2)} \frac{\log(n-w+2)}{\log(m-w+2)}$$

$$\leq \frac{m^{w-1}}{m(m-1)\cdots(m-w+2)} \frac{\log n}{\log(m-w+2)},$$

for fixed $m \geq w \geq 2$. The above observation says that FK's bound is asymptotically equal to

$$\frac{m^{w-1}}{m(m-1)(m-2)\cdots(m-w+2)} \frac{\log n}{\log(m-w+2)},$$

as $n \to \infty$ with w and m fixed. In [21], Korner and Martin improved FK's bound as follows:

Theorem 3.2

$$N(n,m,w) \geq N_j(n,m,w) = \frac{\binom{n}{j+1} m^{j+1}}{\binom{m}{j+1} n^{j+1}} \frac{\log\left(\frac{n-j}{w-j-1}\right)}{\log\left(\frac{m-j}{w-j-1}\right)}.$$

As $n \to \infty$ with w and m fixed, we have that $N(n,m,w)$ is bounded below by a function which is approximately

$$max_{\ 0 \leq j \leq w-2}\ \{B_j \log n\},$$

where

$$B_j = \frac{m^{j+1}}{m(m-1)\cdots(m-j-1)} \frac{1}{\log\left(\frac{m-j}{w-j-1}\right)} \qquad \textit{for} \quad 0 \le j \le w-2.$$

For simplicity, this bound is called KM's bound. Note that FK's bound is equal to KM's bound when $j = w-2$. Both bounds, KM's bound and FK's bound, are obtained using information-theoretic techniques. We observed the upper bound on the values $N(n,m,w)$ in Theorem 1.3,

$$N(n,m,w) \ge \left\lceil \frac{\log \binom{n}{w}}{\log(m^w) - \log(m^w - w!\binom{m}{w})} \right\rceil.$$

S. Blackburn improved it using graph theory as follows:

Theorem 3.3 [8, Theroem 1] *A PHF*$(N; n, q, w)$ *exists whenever*

$$N > \frac{\log 4\left(\binom{n}{w} - \binom{n-w}{w}\right)}{\log q^w - \log\left(q^w - w!\binom{q}{w}\right)}.$$

More precisely, the above bound is better than the bound given in Theorem 1.3 whenever $\binom{n}{w} < \frac{4}{3}\binom{n-w}{w}$, because

$$\log 4\left(\binom{n}{w} - \binom{n-w}{w}\right) < \log \binom{n}{w}$$

$$\iff \quad \log \frac{4\left(\binom{n}{w} - \binom{n-w}{w}\right)}{\binom{n}{w}} < 0$$

$$\iff \quad \frac{4\left(\binom{n}{w} - \binom{n-w}{w}\right)}{\binom{n}{w}} < 1$$

$$\iff \quad 4\left(\binom{n}{w} - \binom{n-w}{w}\right) < \binom{n}{w}$$

$$\iff \quad 3\binom{n}{w} > 4\binom{n-w}{w}.$$

Example 3.1 To illustrate the application of the previous bounds, we will compute the bounds on $N(n, m, w)$ numerically, for $3 \leq n \leq 100$ and for $m = w = 3$, as follows:

Table 3.1: Comparison with the bounds on $N(n, 3, 3)$

n	I	II	III	IV	V	VI	Interval of $N(n, 3, 3)$
3	1	1	1	1	·	6	[1, 6]
4	2	2	2	2	6	12	[2, 6]
5	3	2	3	3	10	15	[3, 10]
6	3	2	3	3	12	18	[3, 12]
7	3	2	4	4	15	20	[4, 15]
8	3	2	4	4	17	21	[4, 17]
9	4	2	4	4	18	23	[4, 18]
10	4	3	5	5	20	24	[5, 20]
11	4	3	5	5	21	25	[5, 21]
12	4	3	5	5	22	26	[5, 22]
13	4	3	5	5	23	26	[5, 23]
14	4	3	6	6	24	27	[6, 24]
15	4	3	6	6	25	28	[6, 25]
16	4	3	6	6	26	28	[6, 26]
17	4	3	6	6	26	29	[6, 26]
18	4	3	6	6	27	29	[6, 27]
19	4	3	6	6	28	30	[6, 28]

continued on next page

continued from previous page

n	I	II	III	IV	V	VI	Interval of $N(n,3,3)$
20	4	3	7	7	29	30	[7, 29]
21	4	3	7	7	29	31	[7, 29]
22	4	3	7	7	30	32	[7, 30]
23	4	3	7	7	30	32	[7, 30]
24	4	3	7	7	31	32	[7, 31]
25	4	3	7	7	31	32	[7, 31]
26	4	3	7	7	32	33	[7, 32]
27	5	3	7	7	32	33	[7, 32]
28	5	4	7	7	33	33	[7, 33]
29	5	4	7	7	33	34	[7, 33]
30	5	4	8	8	34	34	[8, 34]
31	5	4	8	8	34	34	[8, 34]
32	5	4	8	8	34	35	[8, 35]
33	5	4	8	8	35	35	[8, 35]
34,35	5	4	8	8	35	35	[8, 35]
36,37	5	4	8	8	36	36	[8, 36]
38	5	4	8	8	36	36	[8, 36]
39,40	5	4	8	8	37	36	[8, 36]
41	5	4	8	8	37	37	[8, 37]
42,43	5	4	8	8	38	37	[8, 37]
44	5	4	8	8	38	37	[8, 37]
45	5	4	9	9	39	37	[9, 37]
46	5	4	9	9	39	38	[9, 38]

continued on next page

continued from previous page

n	I	II	III	IV	V	VI	Interval of $N(n,3,3)$
47	5	4	9	9	39	38	[9, 38]
48	5	4	9	9	39	38	[9, 38]
49,50	5	4	9	9	40	38	[9, 38]
51,52	5	4	9	9	40	39	[9, 39]
53,54,55	5	4	9	9	41	39	[9, 39]
56,57	5	4	9	9	41	39	[9, 39]
58,59,60,61,62	5	4	9	9	42	40	[9, 40]
63,64	5	4	9	9	43	40	[9, 40]
65,66,67	5	4	9	9	43	41	[9, 41]
68,69	5	4	9	9	44	41	[9, 41]
70,71,72,73	5	4	10	10	44	41	[10, 41]
74,75,76,77,78,79	5	4	10	10	45	42	[10, 42]
80,81,82	5	4	10	10	46	42	[10, 42]
83,84,85,86	5	5	10	10	46	43	[10, 43]
87,88,89,90,91,92,93	5	5	10	10	47	43	[10, 43]
94	5	5	10	10	47	44	[10, 44]
95,96,97,98,99,100	5	5	10	10	48	44	[10, 44]

The bounds used are as follows:

- I : From Theorem 1.1-1,

$$\left\lceil \frac{\binom{n}{3}3^3}{\binom{3}{3}n^3} \right\rceil = \left\lceil \frac{9(n-1)(n-2)}{2n^2} \right\rceil .$$

- II : From Theorem 1.1-2,

$$\left\lceil \frac{\log n}{\log 3} \right\rceil .$$

- III : From Theorem 3.1,

$$\left\lceil \frac{\binom{n-1}{3-2} 3^{3-2} \log(n-3+2)}{\binom{3-1}{3-2} n^{3-2} \log(3-3+2)} \right\rceil = \left\lceil \frac{3(n-1)\log(n-1)}{2n} \right\rceil .$$

- IV : From Theorem 3.2,

$$\max\{N_0(n,3,3), N_1(n,3,3)\} = N_1(n,3,3),$$

where,

$$N_0 = \left\lceil \frac{\log n - 1}{\log 3 - 1} \right\rceil , \quad N_1 = \left\lceil \frac{3(n-1)\log(n-1)}{2n} \right\rceil .$$

- V : From Theorem 1.3,

$$\left\lceil \frac{\log \binom{n}{3}}{\log(3^3) - \log(3^3 - 3!\binom{3}{3})} \right\rceil = \left\lceil \frac{\log \frac{n(n-1)(n-2)}{6}}{\log \frac{27}{21}} \right\rceil .$$

- VI: From Theorem 3.3,

$$\left\lceil \frac{\log 4 \left(\binom{n}{3} - \binom{n-3}{3} \right)}{\log 3^3 - \log \left(3^3 - 3!\binom{3}{3} \right)} \right\rceil = \left\lceil \frac{\log(6n^2 - 30n + 40)}{\log \frac{27}{21}} \right\rceil .$$

The bounds I, II, III, and IV represent lower bounds on the values $N(n,3,3)$; the bounds V and VI are upper bounds. For $m = w = 3$ and $3 \leq n \leq 100$, since $N_0 \leq N_1$, FK's bound is the same as KM's bound, as seen in the above table. Actually, KM's bound is strictly better that FK's bound only when w is small when compared with m. And for the upper bounds V and VI, VI is better than V when

$3\binom{n}{3} < 4\binom{n-3}{3}$, i.e., $n > 41$.

Atici found the value $N(n,3,3)$ for $4 \leq n \leq 11$ in [3], as shown in Table 3.2. □

Table 3.2: PHF$(N;n,3,3)$ for $4 \leq n \leq 11$

n	$N(n,3,3)$	n	$N(n,3,3)$
4	2	5	3
6	3	7	4
8	4	9	4
10	5	11	6

In Chapter 1, elementary probabilistic and non-constructive arguments were used to show that $N(n,m,w)$ is $O(\log n)$ for fixed m and w. So far, we have focused on the situation when the value m is close to the value w. Now we will consider the case when the value m is a sufficiently large prime power, and w is small compared with m. For these types of parameters, S. Blackburn and P. Wild introduced and studied the lower bound on $N(n,m,w)$ in [10]. We will present their results without proof.

Theorem 3.4 [10] *Let $\mathcal{F}$ be an (n,m,w)-perfect hash family. Let d be a positive integer and let m be a prime power. Thus the following hold:*

1. *If $w = 2$ and $n > m^d$, then $|\mathcal{F}| > d$.*
2. *If $w \geq 3$ and $n > (w-1)(m^d - 1)$, then $|\mathcal{F}| > (w-1)d$.*

We note that the first part of the above theorem follows from the result discussed earlier which states that a PHF$(N;n,m,2)$ exists if and only if $N \geq \frac{\log n}{\log m} > d$.

Example 3.2 We will compare the above result with the previous lower bounds for the prime power m and fixed $w = 3$.

1. $m = 4$. By Theorem 3.4, we have

   - If $n \geq 7$, then $N(n, 4, 3) \geq 3$,
   - If $n \geq 30$, then $N(n, 4, 3) \geq 5$,
   - If $n \geq 127$, then $N(n, 4, 3) \geq 7$.

Table 3.3: Comparison on the lower bounds on $N(n, 4, 3)$

	I	II	III	IV
7	2	2	2	3
8	2	2	2	3
9	2	2	3	3
10	2	2	3	3
11	3	2	4	3
16	3	2	4	3
17	3	3	4	3
30	3	3	4	5
32	3	3	5	5
65	3	4	5	5
68	3	4	6	5
127	3	4	6	7
147	3	4	7	7

The bounds used are:

- I : $= \left\lceil \frac{8(n-1)(n-2)}{3n^2} \right\rceil$.
- II : $= \left\lceil \frac{\log n}{2} \right\rceil$.
- III $:= \left\lceil \frac{4(n-1)\log(n-1)}{3n \log 3} \right\rceil$.
- IV : From the previous observation.

For the case when $m = 4$ and $w = 3$, Atici [3] determined the exact value of $N(n, 4, 3)$ for $5 \leq n \leq 10$, as indicated in Table 3.4.

Table 3.4: A PHF$(N; n, 4, 3)$ for $5 \leq n \leq 10$

n	$N(n,4,3)$	n	$N(n,4,3)$
5	2	6	2
7	3	8	3
9	3	10	4

2. $m = 5$. By Theorem 3.4, we have

   - If $n \geq 9$, then $N(n,5,3) \geq 3$,
   - If $n \geq 48$, then $N(n,5,3) \geq 5$,
   - If $n \geq 249$, then $N(n,5,3) \geq 7$.

Table 3.5: Comparison on the lower bounds on $N(n,5,3)$

	I	II	III	IV
9	2	2	2	3
13	2	2	3	3
26	2	3	3	3
33	2	2	4	3
48	2	3	4	5
75	3	3	4	5
90	3	3	5	5
126	3	4	5	5
249	3	4	5	7

The bounds used are:

- I : $= \left\lceil \frac{25(n-1)(n-2)}{12n^2} \right\rceil$.
- II : $= \left\lceil \frac{\log n}{\log 5} \right\rceil$.
- III $:= \left\lceil \frac{5(n-1)\log(n-1)}{4n} \right\rceil$.

- IV : From the previous observation.

Recall we had the explicit examples from Table 1.1, such as a PHF$(4; n, 5, 3)$ for $n = 10, 15, 20$.

3. $m = 9$. By Theorem 3.4, we have

   - If $n \geq 17$, then $N(n, 9, 3) \geq 3$,
   - If $n \geq 161$, then $N(n, , 3) \geq 5$,
   - If $n \geq 1456$, then $N(n, 9, 3) \geq 7$.

Table 3.6: Comparison on the lower bounds on $N(n, 9, 3)$

	I	II	III	IV
17	2	2	2	3
45	2	2	3	3
82	2	3	3	3
161	2	3	3	5

The bounds used are:

- I : $= \left\lceil \frac{81(n-1)(n-2)}{56n^2} \right\rceil$.
- II : $= \left\lceil \frac{\log n}{\log 9} \right\rceil$.
- III $:= \left\lceil \frac{3(n-1)\log(n-1)}{8n} \right\rceil$.
- IV : From the previous observation.

We had a PHF$(4; 27, 9, 3)$ from Table 1.1. Moreover, we are able to construct a PHF$(3, 27, 9, 3)$, which we deal with in the next section.

As observed, the lower bound given by Blackburn and Wild is better when m is large compared with the value w.

In summary, if n is sufficiently large, as indicated in the above theorem, then the cardinality of a perfect hash family is at least $(w-1)e+1$:

Theorem 3.5 [8] *Let e be an integer such that $e \geq 2$. Suppose a PHF$(s; n, q, w)$ exists with $n > m^e(w-1)$. Then $s \geq (w-1)e+1$.*

3.2 Linear PHF

Let q be a prime power and let $n = q^d$ for some positive integer d. Let $\mathcal{F} \subseteq \{f : \mathbf{A} \to \mathbf{B}\}$ be an (n, m, w)-perfect hash family. $\mathcal{F}$ is a *linear perfect hash family* if $\mathbf{B}$ can be identified with a finite field $\mathbb{F}_q$ and $\mathbf{A}$ can be identified with a d-dimensional vector space over $\mathbb{F}_q$, and $\mathcal{F}$ is a set of linear functionals under this identification. A linear (q^d, q, w)-perfect hash family is called *optimal* if $|\mathcal{F}| = d(w-1)$.
In [8], Blackburn presents explicit examples of the linear PHF and the optimal PHF, respectively, as follows:

Example 3.3 [8, Theorem 5] There exists a PHF$(6; p^2, p, 4)$ for all primes p such that $p = 11$ or $p \geq 17$.

- Let p be a prime such that $p = 11$ or $p \geq 17$.
- Let $F = \mathbb{F}_p$ and let $V = F^2$.
- Define functions $f_i : V \to F$ for all $1 \leq i \leq 6$ by

$$f_1((a,b)) = a, \qquad f_2((a,b)) = b,$$
$$f_3((a,b)) = b-a, \qquad f_4((a,b)) = b-2a,$$
$$f_5((a,b)) = b-3a, \qquad f_6((a,b)) = b-5a,$$

for all $a, b \in \mathbb{F}_p$.

Now we show that $\mathcal{F} = \{f_i : 1 \leq i \leq 6\}$ is a PHF$(6; p^2, p, 4)$. Suppose not, For any 4-subset $W = \{x_1, x_2, x_3, x_4\}$, where $x_i = (a_i, b_i)$, there are 6 possible pairs and each pair cannot agree on more than two functions. Then each pair provides each gradient, i.e., slope of $\{\infty, 0, 1, 2, 3, 5\}$. Here we consider the absolute value of each slope modulo some primes p, i.e., $p > 7$. Especially, if we construct five gradients, then the last one is uniquely determined. There are several cases to consider. Here we look at two typical cases:

Case I:

- if the gradient $g_\infty = \frac{b_2 - b_1}{a_2 - a_1}$ from x_1 and x_2 is ∞, then we obtain

$$a_1 = a_2 \pmod{p}.$$

- if the gradient $g_0 = \frac{b_3 - b_1}{a_3 - a_1}$ from x_1 and x_3 is 0, then we obtain

$$b_3 = b_1 \pmod{p}.$$

- if the gradient $g_1 = \frac{b_4 - b_1}{a_4 - a_1}$ from x_1 and x_4 is 1, then we obtain

$$b_4 - b_1 = a_4 - a_1 \pmod{p}.$$

- if the gradient $g_2 = \frac{b_3 - b_2}{a_3 - a_2}$ from x_2 and x_3 is 2, then we obtain

$$b_3 - b_2 = 2(a_3 - a_2) \pmod{p}.$$

- if the gradient $g_3 = \frac{b_4 - b_2}{a_4 - a_2}$ from x_2 and x_4 is 3, then we obtain

$$b_4 - b_2 = 3(a_4 - a_2) \pmod{p}.$$

Then, by computing those equations:

$$\begin{aligned} a_1 &= a_2 \pmod p, \\ b_1 &= b_3 \pmod p, \\ b_4 - b_1 &= a_4 - a_1 \pmod p, \\ b_3 - b_2 &= 2(a_3 - a_2) \pmod p, \\ b_4 - b_2 &= 3(a_4 - a_2) \pmod p. \end{aligned}$$

Then the equation results in $a_3 = a_4$. Thus in this case, for any prime $p \geq 11$, x_3 and x_4 cannot make the gradient 5.

Case II:

- if the gradient $g_\infty = \frac{b_2 - b_1}{a_2 - a_1}$ from x_1 and x_2 is ∞, then we obtain

$$a_1 = a_2 \pmod p.$$

- if the gradient $g_0 = \frac{b_3 - b_4}{a_3 - a_4}$ from x_3 and x_4 is 0, then we obtain

$$b_3 = b_4 \pmod p.$$

- if the gradient $g_1 = \frac{b_3 - b_1}{a_3 - a_1}$ from x_1 and x_3 is 1, then we obtain

$$b_3 - b_1 = a_3 - a_1 \pmod p.$$

- if the gradient $g_2 = \frac{b_4 - b_2}{a_4 - a_2}$ from x_2 and x_4 is 2, then we obtain

$$b_4 - b_2 = 2(a_4 - a_2) \pmod p.$$

- if the gradient $g_3 = \frac{b_4 - b_1}{a_4 - a_1}$ from x_1 and x_4 is 3, then we obtain

$$b_4 - b_1 = 3(a_4 - a_1) \pmod{p}.$$

Then, by computing those equations:

$$\begin{aligned} a_1 &= a_2 \pmod{p}, \\ b_3 &= b_4 \pmod{p}, \\ b_3 - b_1 &= a_3 - a_1 \pmod{p}, \\ b_4 - b_2 &= 2(a_4 - a_2) \pmod{p}, \\ b_4 - b_1 &= 3(a_4 - a_1) \pmod{p}. \end{aligned}$$

Then the equation results in $\frac{b_3 - b_2}{a_3 - a_2} \equiv \frac{2}{3} \pmod{p}$. Thus in this case, last gradient from x_2 and x_3 cannot be 5 if $\frac{2}{3} \not\equiv 5 \pmod{p}$, namely, $p \neq 13$.

By consideration of all possible cases, when $p = 11$ or $p \geq 17$, no set of four points is associated with the set $\{\infty, 0, 1, 2, 3, 5\}$ of gradients.

Suppose that there exists a PHF$(N; p^2, p, 4)$. Then by letting $d = 1$ and $m = p$, we have $N \geq 5$ by Theorem 3.5. But there is no known PHF$(5; p^2, p, 4)$. And we know that there exists a PHF$(7; 7^2, 7, 4)$ and a PHF$(7; 13^2, 13, 6)$ using the fact that there exist 5 MOLS of order 7 and 13. But we cannot apply the above construction to verify the existence of a PHF$(6; 13^2, 13, 4)$. By finding the solution $\{(a_i, b_i) : 1 \leq i \leq 4\}$ of the following system:

$$\begin{aligned} a_1 = a_2, \qquad & b_3 = b_4, \\ b_3 - b_1 = a_3 - a_1, \qquad & b_4 - b_2 = 2(a_4 - a_2), \\ b_4 - b_1 = 3(b_4 - b_1), \qquad & 3(b_4 - b_2) = 2(a_3 - a_2), \end{aligned}$$

we can find the 4-set which fails to be separated by the family $\mathcal{F}$ defined above, for example, {(1,10),(1,1),(3,3),(4,3)}. □

Example 3.4 [8] There exists an optimal PHF$(3; r^3, r^2, 3)$ for all $r \geq 2$ as follows:

- Let r be a fixed integer such that $r \geq 2$.
- Let T be a set of size r.
- Let $V = T^3$ and $F = T^2$.
- Define functions $f_1, f_2, f_3 : V \to F$ by

$$\begin{aligned} f_1((a,b,c)) &= (a,b), \\ f_2((a,b,c)) &= (b,c), \\ f_3((a,b,c)) &= (a,c), \end{aligned}$$

  for all $a, b, c \in T$.

We show that $\mathcal{F} = \{f_1, f_2, f_3\}$ is a PHF$(3; r^3, r^2, 3)$. Suppose not, i.e., the set of functions constructed by the above is not a perfect hash family. For three distinct elements of V, say $W = \{x_i = (a_i, b_i, c_i) \in V : i = 1, 2, 3\}$, suppose that there exists no function in $\mathcal{F}$ which separates W. Then for any two distinct points from the 3-set W, there exists at least one function in $\mathcal{F}$ such that they have the same value. Note that there are three possible pairs in W, namely $\{x_1, x_2\}, \{x_1, x_3\}$, and $\{x_2, x_3\}$. It is impossible for each pair to agree on more than two functions, since $x_1 \neq x_2 \neq x_3$. Thus, without loss of generality, $f_1(x_1) = f_2(x_2)$, $f_2(x_1) = f_2(x_3)$, and $f_3(x_2) = f_3(x_1)$. Then, by definition of f_i, $x_1 = x_2 = x_3$, which

gives a contradiction. Thus $\{f_1, f_2, f_3\}$ is a PHF$(3; r^3, r^2, 3)$.
We now verify its optimality: Note that a PHF$(N; r^3, r^2, 3)$ has the property that $N \geq 3$, provided that $r > 2$, by previous results. There exists a PHF$(2; n, 4, 3)$ only if $n \leq 6$, i.e., there exists no PHF$(2; 8, 4, 3)$. □

To attempt to find an optimal linear PHF, Blackburn and Wild introduce the following result. Even though this theorem is non-constructive, it gives a tight bound on the minimum cardinality of a linear (q^d, q, w) perfect hash family for a sufficiently large prime power q.

Theorem 3.6 [10, Theorem 4] *Let d, w be integers such that $d, w \geq 2$ and let q be a prime power. Set $n = q^d$. Then there exists a linear PHF$(N; n, q, w)$ with $N \geq d(w-1)$. Specifically, if $q \geq \binom{w}{2}^{d(w-1)}$, then there exists a PHF$(N; q^d, q, w)$ with $N = d(w-1)$, that is, there exists an optimal linear PHF$(N; q^d, q, w)$.*

To construct an optimal linear PHF by using the above theorem, they let V be the vector space with $|V| = n = q^d$ which has dimension d over $\mathbb{F}_q$ and let $k = d(w-1)$. And then by using the dual space V^* of V, which consists of all linear functionals $\phi : V \mapsto \mathbb{F}_q$, they attempt to construct a suitable sequence $(\phi_1, \ldots, \phi_k) \in (V^*)^k$, say $\mathcal{F} = \{\phi_1, \ldots, \phi_k\}$, that is a perfect hash family. If we consider the techniques of the proof in [10], then we may construct explicit classes of optimal linear perfect hash families in certain fields as follows:

Theorem 3.7 [10, Theorem 5] *Let d, w be integers such that $d, w \geq 2$ and let q be a prime power. Suppose that there exist finite fields*

$$F_0 < F_1 < F_2 < \cdots < F_{d(w-2)}$$

such that $|F_{d(w-2)}| = q$ *and* $[F_i : F_{i-1}] \geq d$ *for any* $i \in \{1, 2, \ldots, d(w-2)\}$.
Define a sequence $(\alpha^1, \alpha^2, \ldots, \alpha^{d(w-1)})$ *of row vectors of length* d *as follows:*
For all integers i *such that* $1 \leq i \leq d$, *define* α^i *to be the* i*th standard basis vector.*
For all integers i *such that* $d+1 \leq i \leq d(w-1)$, *define*

$$\alpha^i = (\beta^i{}_1, \beta^i{}_2, \ldots, \beta^i{}_d),$$

where $\{\beta^i{}_1, \beta^i{}_2, \ldots, \beta^i{}_d\}$ *is any subset of* F_{i-d} *which is linearly independent over* F_{i-d-1}. *Set* $V = (F_{d(w-2)})^d$ *and define functionals* $\phi_1, \phi_2, \ldots, \phi_{d(w-1)}$ *by*

$$(v)\phi_i = v(\alpha^i)^T,$$

for all $v \in V$. *Then* $\mathcal{F} = \{\phi_1, \phi_2, \ldots, \phi_{d(w-1)}\}$ *is an optimal linear* (q^d, q, w) *perfect hash family.*

Remark: It is not easy to construct an optimal linear PHF using the method given in Theorem 3.7 because it is difficult to find a suitable sequence of finite fields satisfying the condition stated in Theorem 3.7.

3.3 Other Constructions

In [9], S. Blackburn introduced the notation of $n_{N,w}(q)$, which is the largest integer n such that there exists a PHF$(N; n, q, w)$ for fixed N, q and w. He focuses on the construction of new classes of PHF for which $\lim_{q\to\infty} \frac{n_{N,w}(q)}{q}$ exists. Here we will present one way to construct perfect hash families given in [9]. Throughout this section, perfect hash families are regarded as sets of partitions as mentioned in Proposition 1.1.

Theorem 3.8 [9] *There exists a PHF*$(k; a^k, a^{k-1}, k)$ *for* $a \geq 2$.

Proof Let $k \geq 2$ and $a \geq 2$ be integers. Let A be a set of size a and let $V = A^k$. Define a set of partitions $\mathcal{F} = \{\pi_1, \ldots, \pi_k\}$ as follows:
$(a_1, a_2, \ldots, a_k) \in V$ and $(b_1, b_2, \ldots, b_k) \in V$ are in the same part of π_i if and only if $a_j = b_j$ for all $j \in \{1, 2, \ldots, k\} \setminus \{i\}$. Then each partition π_i has a^{k-1} parts, each of size a. Then for $X \subseteq V$ with $|X| = w$, $w \leq k$, X can be separated by at least $k - (w-1)$ of the partitions in $\mathcal{F}$. If we prove this, then we naturally obtain that there exists a PHF$(k; a^k, a^{k-1}, k)$.
Suppose not, i.e., $X = \{x^1, x^2, \ldots, x^w\}$ cannot be separated by w partitions. Define a colored graph $G = (V, E)$ by

- Let $V := X$ be the vertex set and
- Define the edge set E as follows: for $x, y \in X$, there exists an edge joining x and y if and only if there exists j such that x and y are in the same part of π_j.
- Give the color j to the edge joining x and y, where x and y are in the same part of π_j.

Now the graph G has w vertices and consists of edges of all w colors (if not, i.e., if there is a missing color, say h, then π_h can separate X). Thus G contains a cycle. Let C be the shortest cycle consisting of $|C|$ distinct colors, say $C = x^1 x^2 \cdots x^c$, where $x^c = x^1$. Let the edge between x^1 and x^2 be colored j, i.e., ${x^1}_i = {x^2}_i$ for all $i \in \{1, 2, \ldots, k\} \setminus \{j\}$. Moreover, since each color occurs once in the cycle for all $i \in \{2, 3, \ldots, c-1\}$, we have that x^i and x^{i+1} agree in their jth position. But this implies that the jth position of x^1 differs from the jth position of x^c. This contradicts the fact that $x^1 = x^c$. $\square$

For $k = 2$, we already produced a PHF$(2; a^2, a, 2)$ in Theorem 1.2. For $k = 3$, we presented a PHF$(3; a^3, a^2, 3)$ in Example 3.4.

Example 3.5 To illustrate the method given in Theorem 3.8, we construct a PHF$(2; 9, 3, 2)$ as follows: Let $\mathbf{A} = \{(ab) : a, b \in \{1, 2, 3\}\}$ and $\mathbf{B} = \{1, 2, 3\}$. Then the class of partitions, $\mathcal{F} = \{\pi_1, \pi_2\}$ is shown in Figure 3.1. □

	(11)	(12)	(13)	(21)	(22)	(23)	(31)	(32)	(33)
π_1	1	2	3	1	2	3	1	2	3
π_2	1	1	1	2	2	2	3	3	3

FIGURE 3.1: A PHF$(2; 9, 3, 2)$

Blackburn shows the following method to construct some perfect hash families in [9].

Theorem 3.9 [9] *There exists a PHF*$(3; 3a^2, a^2 + 2a, 4)$ *for* $a \geq 2$.

Proof We translate the diagram of a PHF$(3; 3a^2, a^2 + 2a, 4)$ given in [9] to the following form:

Let $A = \{(h, i, j) : 1 \leq h \leq 3, \quad 1 \leq i, j \leq a\}$. Define the class of hash functions $\mathcal{F} = \{f_1, f_2, f_3\}$ by

A	$(1, i, j)$	$(2, i, j)$	$(3, i, j)$
f_1	$(1, i, j)$	$(2, i)$	$(3, j)$
f_2	$(3, j)$	$(1, i, j)$	$(2, i)$
f_3	$(2, i)$	$(3, j)$	$(1, i, j)$

We will show that any 4-subset X of A can be separated by the functions of $\mathcal{F}$. Note that any two points with the distinct first position are separated by all functions in

$\mathcal{F}$. Moreover, any two points with the same first position are separated by at least two functions in $\mathcal{F}$. Then we can classify the 4-sets X, into four cases, as follows:

- all points have the same value on the first position.
- 3 points have the same value on the first position.
- 2 points have the same value; the other two points have the different values on the first position, i.e., X have all 3 values on the first position.
- 2 points have the same value; 2 points have a different common value on the first position.

For the first three cases, X can be separated by at least one function of $\mathcal{F}$ by the previous observations. For the last case, without loss of generality, say $X = \{x_1 = (1, i_1, j_1), x_2 = (1, i_2, j_2), x_3 = (2, i_3, j_3), x_4 = (2, i_4, j_4)\}$. There exists a function in $\mathcal{F}$ separating the set X, that is, if $i_1 = i_2$ and $i_3 = i_4$, then f_2 separates the set X; if $i_1 = i_2$ and $i_3 \neq i_4$, then f_1 separates the set X; if $i_1 \neq i_2$ and $i_3 = i_4$, then f_3 separates the set X;if $i_1 \neq i_2$ and $i_3 \neq i_4$, then f_1 separates the set X. □

Remark: $(h, i, j) \in A$ means the cell (i, j) of the hth square, and (h, k) represents the kth position of the rectangle corresponding to the hth square of the diagram in [9].

Example 3.6 To illustrate the above method, let $a = 2$. Then a PHF$(3; 12, 8, 4)$ is shown in Figure 3.2. □

With a similar method, Blackburn also presents a construction of a PHF$(5; 3a^3, a^3 + 5a^2 + a, 7)$ for $a \geq 2$ in [9].

A	(111)	(112)	(121)	(122)	(211)	(212)
	(221)	(222)	(311)	(312)	(321)	(322)
f_1	(111)	(112)	(121)	(122)	(21)	(21)
	(22)	(22)	(31)	(32)	(31)	(32)
f_2	(31)	(32)	(31)	(32)	(111)	(112)
	(121)	(122)	(21)	(21)	(22)	(22)
f_3	(21)	(21)	(22)	(22)	(31)	(32)
	(31)	(32)	(111)	(112)	(121)	(122)

FIGURE 3.2: A PHF$(3; 12, 8, 4)$

Chapter 4

Constructions using Algebraic Structures

In this chapter, we will present some methods to construct perfect hash families using the algebraic structures like special global function fields and algebraic curves. These results are based on [24] and [41].

We will introduce a theorem for the construction of a PHF based on a suitable function field. We also discuss some basic concepts and properties which are needed in these constructions. There is a 1-1 correspondence between function fields and algebraic curves. That makes it possible to translate definitions and results from algebraic function fields to algebraic curves (and vice versa). Then, we will give an algorithm for the actual construction of a perfect hash family using these techniques. Even though such a construction needs assumptions like those in Chapter 3, the resulting family is a kind of a linear PHF. Moreover, the relation between algebraic function fields and algebraic curves shows how to construct a PHF using an algebraic curve. Finally, we will introduce some examples which apply the previous results.

Throughout this chapter, we assume that q is a prime power and $\mathbb{F}_q$ is the finite field of q elements.

4.1 Preliminaries

In this section, we present the main theorem for our construction. Then, we deal with relevant concepts and results. Let us consider the following theorem.

Theorem 4.1 [24] *For the global function field $F/\mathbb{F}_q$ with genus g, let $P_1, \ldots, P_N$ be distinct rational places of F. Let G be a divisor of F with $deg(G) \geq 2g+1$ and $supp(G) \cap \{P_1, \ldots, P_N\} = \emptyset$. Then there exists a perfect hash family PHF $(N; q^{deg(G)-2g+1}, q, w)$ whenever $2 \leq w \leq q$ and $N > \binom{w}{2} deg(G)$.*

First, we give a brief review of the necessary background for the above theorem.

Definition 4.1 *An* algebraic function field *$F/\mathbb{F}_q$ of* one variable over *$\mathbb{F}_q$ is an extension field $F \supseteq \mathbb{F}_q$ that is a finite algebraic extension of $\mathbb{F}_q(x)$ for some element $x \in F$ which is transcendental over $\mathbb{F}_q$.*

Definition 4.2 *A* place *P of F is the maximal ideal of some valuation ring of F. A place of degree* 1 *is called* rational. *Let $\mathbf{P}_F$ denote the set of rational places of F.*

Let $N(F)$ denote the number of rational places of a function field F. In Theorem 4.1, the cardinality of the perfect hash family depends on the values of $g(F)$ and $N(F)$. Unfortunately, we cannot always compute those values. Thus, for an explicit construction, we must use a suitable function field F such that $g(F)$ and $N(F)$ are known. The following is one of known results on the value $N(F)$.

Theorem 4.2 (Hasse-Weil Bound) *Let $F/\mathbb{F}_q$ be an algebraic function field of genus g. Then the number $N(F)$ of rational places of $F/\mathbb{F}_q$ satisfies*

$$|N(F) - (q+1)| \leq 2g\sqrt{q}.$$

Definition 4.3 *1. A divisor G is a formal sum*

$$G = \sum_{P \in \mathbf{P}_F} n_P \cdot P \qquad \text{with } n_P \in \mathbb{Z}, \text{ where all but finitely many } n_P = 0.$$

2. The support $supp(G)$ of a divisor G is defined by

$$supp(G) := \{P \in \mathbf{P}_F : n_P \neq 0\}.$$

Definition 4.4 *Let F be an algebraic function field. For a divisor G of F, we define the* Riemann-Roch space $\mathcal{L}(G)$ *by*

$$\mathcal{L}(G) = \{x \in F^* : div(x) + G \geq 0\} \cup \{0\}.$$

Then $\mathcal{L}(G)$ is a finite-dimensional vector space over a finite field $\mathbb{F}_q$ and we denote its dimension by $l(G)$.

The following is the most important theorem in the theory of algebraic function fields.

Theorem 4.3 (Riemann-Roch Theorem) *Let F be an algebraic function field of genus g. Then for any divisor G of F we have*

$$l(G) \geq deg(G) + 1 - g.$$

and equality holds whenever $deg(G) \geq 2g - 1$.

Let $F/\mathbb{F}_q$ be an algebraic function field of genus g with $N(F) \geq 1$ and let G be a divisor of F. To each rational place P of F with $P \notin supp(G)$ we associate the map $h_P : \mathcal{L}(G) \mapsto \mathbb{F}_q$ defined by

$$h_P(f) = f(P) \quad \text{for all} \quad f \in \mathcal{L}(G).$$

The map h_P defined in the above will play an important role in the construction of perfect hash families. Note that we will deal with a divisor G where $deg(G) \geq 2g+1$. By the Riemann-Roch theorem, we can obtain the following lemma.

Lemma 4.1 [41] *Let* $\mathcal{F} = \{h_P : P \in T\}$ *where* T *is a subset of all rational places of* F *satisfying* $T \cap supp(G) = \emptyset$. *If* $deg(G) \geq 2g + 1$, *then* $|\mathcal{F}| = |T|$.

Now, we discuss the connections between algebraic curves over finite fields and global function fields. From an algebraic function field of one variable $F/\mathbb{F}_q$, there exists a non-singular projective curve $\mathcal{X}$ (unique up to isomorphism) whose function field $\mathbb{F}_q(\mathcal{X})$ is F. In [31], Stichtenoth presents the construction of such a curve $\mathcal{X}$, from a given function field $F/\mathbb{F}_q$ as follows:

- Choose $x, y \in F$ such that $F = \mathbb{F}_q(x, y)$; Choose a separating element $x \in F/\mathbb{F}_q$. Then, since $F/\mathbb{F}_q(x)$ is a finite separable field extension, there is some $y \in F$ satisfying $F = \mathbb{F}_q(x, y)$.
- Let $G(X, Y) \in \mathbb{F}_q[X, Y]$ be an irreducible polynomial with $G(x, y) = 0$.
- Let $W = \{P \in \mathbf{A^2} : G(P) = 0\}$ and $\overline{W} \subseteq \mathbf{P^2}$, i.e., $\overline{W}$ is the projective closure of W, where $\mathbf{A^2}$ is the 2-dimensional affine space and $\mathbf{P^2}$ is the 2-dimensional projective space.

- Let $\mathcal{X}$ be the non-singular model of $\overline{W}$.
- Then $F \simeq \mathbb{F}_q(x, y)$.

More background including definitions can be found in other references such as [31] and [39]. As mentioned before, there is a 1-1 correspondence between algebraic curves and a function field, as follows:

Theorem 4.4 [24] *The map $\delta : \mathcal{X} \mapsto \mathbb{F}_q(\mathcal{X})$ yields a natural correspondence between smooth projective curves over $\mathbb{F}_q$ and global function fields (of one variable), up to isomorphism.*

The map δ in the above theorem induces a correspondence between the points $P \in \mathcal{X}$ and the places of $F/\mathbb{F}_q$, and it preserves the genus, i.e., the curve $\mathcal{X}$ and its function field $\mathbb{F}_q(\mathcal{X})$ have the same genus. Thus the correspondence δ makes it possible to translate definitions and results from algebraic function fields to algebraic curves (and vice versa) such as divisors G, degree of G, and Riemann-Roch space $\mathcal{L}(G)$. Those will play important factors in the explicit construction of perfect hash families. Therefore, we can express Theorem 4.1 in terms of an algebraic curves as follows:

Theorem 4.5 [41] *For an algebraic curve $\mathcal{X}/\mathbb{F}_q$ with genus g, let $T = \{P_1, \ldots, P_N\}$ be a subset of $\mathbb{F}_q$-rational points. Let G be a divisor of F with $deg(G) \geq 2g + 1$ and $supp(G) \cap T = \emptyset$. Then there exists a perfect hash family PHF $(N; q^{l(G)}, q, w)$ whenever $2 \leq w \leq q$ and $N > \binom{w}{2} deg(G)$.*

In fact, the resulting family $\mathcal{F}$ is a kind of a linear perfect hash family, as defined in Chapter 3. That is, $\mathcal{F}$ consists of linear functionals from the vector space $\mathcal{L}(G)$ to the underlying finite field $\mathbb{F}_q$.

4.2 Constructions and Examples

In this section, we will present a construction for perfect hash families, which is based on Theorem 4.1.

Algorithm 1 [Construction a PHF$(N; n, q, w)$]

Input: *A prime power q.*

Step 1. *Choose a global function field F over $\mathbb{F}_q$.*

Step 2. *Compute the genus $g(F)$ of F and $N(F)$.*

Step 3. *Choose positive integers t, w, and N satisfying the following conditions:*

- $2 \leq w \leq q$,
- $\binom{w}{2} t < N \leq N(F)$,
- $t \geq 2g + 1$.

Step 4. *Choose distinct rational places $P_1, P_2, \ldots, P_N$ of F, and put $T = \{P_1, P_2, \ldots, P_N\}$.*

Step 5. *Choose a divisor G of F with $deg(G) = t$ and $supp(G) \cap T = \emptyset$.*

Step 6. *For the vector space $\mathcal{L}(G)$, define the function $h_P : \mathcal{L}(G) \rightarrow \mathbb{F}_q$ by*

$$h_P(f) = f(P) \quad \text{for all } P \in T.$$

Step 7. *Set $n = q^{deg(G)-2g(F)+1}$.*

Output: *$\mathcal{F} = \{h_P : P \in T\}$ is a PHF$(N; n, q, w)$.*

Analysis

- In Step 1 and Step 2, we assume that, for the chosen function field F, the computation of the values $g(F)$ and $N(F)$ is easy and F has sufficiently many rational places. Here the existence of an efficient algorithm for the computation $g(F)$ and $N(F)$ is an important question.

- In Step 2, it is sufficient to find an upper bound $\overline{g}$ for $g(F)$. Then we can modify the algorithm, using the value $\overline{g}$ instead of $g(F)$.

- For the value $N(F)$, in Step 2, it is not necessary to find the exact value, but we have to know a lower bound $\underline{N}$ for $N(F)$ and then apply $\underline{N}$ to the algorithm instead of $N(F)$.

- Because of our assumptions that F has many rational places, Step 4 can be carried out efficiently.

- Step 6 is based on the previous observation, i.e., the existence of h_P on the Riemann-Roch space $\mathcal{L}(G)$. Moreover,by Lemma 4.1, the cardinality of the class $\{h_P : P \in T\}$ is the same value N that is chosen in Step 3.

- By Theorem 4.1, the output $\mathcal{F}$ is an (n, q, w) perfect hash family with $|\mathcal{F}| = N$.

Example 4.1 Let q be a prime power, and let $K = \mathbb{F}_q$.

1. Let F be the rational function field over K. Then $g(F) = 0$ and $N(F) = q+1$. We can construct a PHF$(N; q^{t+1}, q, w)$, for suitable values N, t, and w by following the above algorithm. More precisely, there exists a PHF$(t\binom{w}{2} + 1; q^{t+1}, q, w)$ for $w \geq 2$ and $t \geq 1$.

2. Elliptic function fields F, i.e., $g(F) = 1$, are classified as follows:

   - If $ch(K) \neq 2$, then there are x, y such that $F = K(x, y)$ and $y^2 = f(x) \in K[x]$ with a square free polynomial $f(x)$ of degree 3.
   - If $ch(K) = 2$, then there are $x, y \in F$ such that $F = K(x, y)$ and $y^2 + y = f(x) \in K[x]$ with degree 3 or $y^2 + y = x + \frac{1}{ax+b}$ with $a, b \in K$ and $a \neq 0$.

   By Theorem 4.2, $(\sqrt{q} - 1)^2 \leq N(F)$, i.e., $\underline{N} = \lceil (\sqrt{q} - 1)^2 \rceil$. Then by the above algorithm, we have a PHF$(N; q^t, q, w)$ for $\lceil (\sqrt{q} - 1)^2 \rceil \geq N \geq t\binom{w}{2} + 1$ and $t \geq 3$.

3. Suppose that there exists r such that $q = r^2$. Let F be the *Hermitian function field* $\mathbb{F}_q(x, y)$ defined by $y^r + y = x^{r+1}$. Then $g(F) = r(r-1)/2$ and $N(F) = r^3 + 1$ are known. We can construct a PHF$(N; q^{t+1-r(r-1)}, q, w)$, where positive integers t, w, and N satisfy the following:

$$t \geq 2g + 1 = r(r-1) + 1, \quad 2 \leq w \leq q, \text{ and } \binom{w}{2} t < N \leq r^3 + 1.$$

□

Suppose that there exists a function field with genus g. For w such that $2 \leq w \leq q$, if we let $t = 2g + 1$, then by Algorithm 1 and Theorem 4.2, we can obtain a PHF$((2g + 1)\binom{w}{2} + 1; q^{g+2}, q, w)$ provided that $q - 2g\sqrt{q} \geq (2g + 1)\binom{w}{2}$. If we apply the parameters of the resulting PHF to Theorem 3.5, i.e., if there exists a PHF$(s; q^{g+2}, q, w)$ with $q \gg w$, then $s \geq (w-1)(g+1)+1$ since $q^{g+2} > q^{g+1}(w-1)$.

With these parameters, we conclude that

$$N \approx \frac{w(w-1)}{\log q} \log n.$$

Of course, it is another problem to construct explicitly a PHF$((w-1)(g+1)+1; q^{g+2}, q, w)$.

Example 4.2 To illustrate the above observations, we can make a list of the possible pairs (g, q) where we can construct a PHF$((2g+1)\binom{w}{2}+1; q^{g+2}, q, w)$ for $2 \leq w \leq 4$. To do this, we refer to Table 4.1 which from [24].

Table 4.1: Bounds for $N_q(g)$

$g \setminus q$	2	3	4	5	8	9	16	27
1	5	7	9	10	14	16	25	27
2	6	8	10	12	18	20	33	48
3	7	10	14	16	24	28	38	56
4	8	12	15	18	25-27	30	45-46	64-66
5	9	12-13	17-18	20-22	29-32	32-35	49-54	55-76

where $N_q(g)$ means the number of rational places of a function field with genus g over $\mathbb{F}_q$.

- When $w = 2$, then there exists a PHF$(2g+2; q^{g+2}, q, 2)$ for the following pairs (g, q):

$$\begin{aligned} &(g,q) \quad \text{for } g \in \{1,2\}, \quad q \in \{2,3,4,5,8,9,16,27\}, \\ &(g,q) \quad \text{for } g \in \{3,4\}, \quad q \in \{3,4,5,8,9,16,27\}, \\ &(5,q) \quad \text{for } q \in \{4,5,8,9,16,27\}. \end{aligned}$$

- When $w = 3$, then there exists a PHF$(6g+4; q^{g+2}, q, 3)$ for the following pair

(g, q):

$$\begin{aligned}
&(1, q) && \text{for } q \in \{8, 9, 16, 27\},\\
&(g, q) && \text{for } g \in \{2, 3\}, \quad q \in \{3, 4, 5, 8, 9, 16, 27\},\\
&(4, q) && \text{for } q \in \{9, 16, 27\},\\
&(5, q) && \text{for } q \in \{16, 27\}.
\end{aligned}$$

- When $w = 4$, then there exists a PHF$(12g + 7; q^{g+2}, q, 4)$ for the following pairs (g, q):

$$\begin{aligned}
&(g, q) && \text{for } g \in \{1, 2\}, \quad q \in \{16, 27\},\\
&(g, q) && \text{for } g \in \{3, 4\}, \quad q \in \{16, 27\}.
\end{aligned}$$

□

Example 4.3 [24] Consider the Garcia-Stichtenoth tower of global function fields which is found in [31].

Let q be a square prime power, i.e., $q = r^2$. Let $K_1 \subseteq K_2 \subseteq \cdots$ be function fields over $\mathbb{F}_q$, which are defined to be $K_1 = \mathbb{F}_q(x_1)$ and $K_i = K_{i-1}(x_i)$ for $i \geq 2$, with

$$x_i{}^r + x_i = \frac{x_{i-1}{}^r}{x_{i-1}{}^{r-1} + 1}.$$

Then $N(K_i) > (r-1)r^i$ and $g(K_i) < r^i$ for all $i \geq 1$. Then an infinite class of examples of perfect hash families is obtained from the Garcia-Stichtenoth tower of global function fields.

1. Let $r \geq 4$ be a prime power and let c be a real number with $1 \leq c \leq (r-2)/2$.

2. Let $w = \lfloor (\frac{2r}{c+1})^{1/2} \rfloor$. Then by the definition of c and the observation below,

$$2 \leq c+1 \leq \frac{r}{2} \quad \Longleftrightarrow \quad \left(\frac{2r}{r}\right)^{1/2} \leq \left(\frac{2r}{c+1}\right)^{1/2} \leq \left(\frac{2r}{2}\right)^{1/2},$$

and the value w satisfies $2 \leq w \leq q^{1/4} < q$.

3. For $i \geq 1$, let $t_i = \lfloor (c+1)r^i \rceil$. For the following observation, we can choose N_i satisfying

$$t_i = \lfloor (c+1)r^i \rceil \geq \lfloor 2r^i \rfloor > 2g(K_i).$$

4. For $i \geq 1$, let $t_i = (r-1)r^i$. We verify the condition $\binom{w}{2} t_i < N_i \leq N(K_i)$ as follows:

$$\begin{aligned}
\binom{w}{2} t_i &\leq \frac{q^{1/4}(q^{1/4}-1)}{2} \lfloor (c+1)r^i \rceil \\
&\leq \frac{q^{1/2}-q^{1/4}}{2} \lfloor (\frac{r}{2})r^i \rceil \\
&= \frac{(r-r^{1/2})r^{i+1}}{4} \\
&\leq (r-r^{1/2})r^i \\
&< (r-1)r^i.
\end{aligned}$$

5. Choose distinct rational places $P_1, P_2, \ldots, P_{N_i}$ of K_i, say $T_i = \{P_1, P_2, \ldots, P_{N_i}\}$.

6. Choose a divisor G_i of K_i with $deg(G_i) = t_i \geq 2g(K_i)+1$ and $supp(G_i) \cap T_i = \emptyset$. Then

$$\begin{aligned}
l(G_i) &= deg(G_i) - g(K_i) + 1 \quad = \quad t_i - g(K_i) + 1 \\
&\leq g(K_i) + 2 \quad < \quad r^i + 2
\end{aligned}$$

$$< \quad cq^{i/2}.$$

Then there exists a PHF$(N_i; n_i, q, w)$, where for all $i \geq 1$,

$$N_i = (\sqrt{q} - 1)q^{i/2}, \qquad w = \left\lfloor \left(\frac{2q}{c+1} \right)^{1/4} \right\rfloor, \qquad \text{and } n_i = q^{\lfloor cq^{i/2} \rfloor}.$$

If we let $r = 5$ and $c = 1$, then by above the construction, we have $w = \lfloor 10/\sqrt{2} \rfloor = 2$ and $n_i = 25^{\lfloor 25^{i/2} \rfloor} = 5^{2 \cdot 5^i}$. Thus there exists a PHF$(4 \cdot 5^i; 5^{2 \cdot 5^i}, 25, 2)$. With these parameters, we obtain the infinite class of perfect hash families satisfying

$$N \approx \frac{2}{\log 5} \log n = .862 \log n.$$

□

Theorem 4.6 [24, Corollary 7.3.7] *For any integers $m \geq w \geq 2$, there exists a sequence of perfect hash families PHF $(N_i; n_i, m, w)$, $i = 1, 2, \ldots$, such that $n_i \to \infty$ as $i \to \infty$ and $N_i \leq C \log n_i$ for all $i \geq 1$, for a constant C depending only on m and w.*

Now, we will deal with algebraic curves over a finite field. More precisely, we will only consider smooth projective plane curves, which are closely related to the algebraic function fields described before. We have an explicit formula for the genus of such curves.

Theorem 4.7 *For a smooth projective plane curve $\mathcal{X}$ over $\mathbb{F}_q$ of degree d,*

$$g = \frac{1}{2}(d-1)(d-2).$$

From the connection between function fields and algebraic curves, we also present the following algorithm and examples.

Algorithm 2 [Construct a PHF$(N; n, q, w)$]

Input: *A prime power q.*

Step 1. *Choose a smooth projective plane curve $\mathcal{X}$ over $\mathbb{F}_q$, which has function field $F(\mathcal{X})$.*

Step 2. *Compute the genus $g(F)$ and $N(F)$.*

Step 3. *Choose positive integers t, w, and N satisfying the following conditions:*

- $2 \leq w \leq q$,
- $\binom{w}{2} t < N \leq N(F)$,
- $t \geq 2g + 1$.

Step 4. *Choose distinct $\mathbb{F}_q$-rational points $P_1, P_2, \ldots, P_N$ of $\mathcal{X}$, and put $T = \{P_1, P_2, \ldots, P_N\}$.*

Step 5. *Choose a divisor G of F with $deg(G) = t$ and $supp(G) \cap T = \emptyset$.*

Step 6. *For the vector space $\mathcal{L}(G)$, define the function $h_P : \mathcal{L}(G) \to \mathbb{F}_q$ by*

$$h_P(f) = f(P) \quad \text{for all } P \in T.$$

Step 7. *Set $n = q^{deg(G)-2g(F)+1}$.*

Output: *$\mathcal{F} = \{h_P : P \in T\}$ is a PHF$(N; n, q, w)$.*

Analysis: In step 1, we only consider smooth projective plane curves. So we can compute the genus of the chosen curve $\mathcal{X}$ using Theorem 4.7. Moreover, we already observed that there is a correspondence between rational places and rational points in the case of smooth projective curves. Naturally, we can apply most of the properties for the construction to those curves. Thus, by Theorem 4.5, we can obtain a PHF$(N; n, q, w)$ with this algorithm.

Example 4.4 Let q be a prime power.

1. Let $\mathcal{X}/\mathbb{F}_q$ be the projective line. Then $g(\mathcal{X}) = 0$ and the number of $\mathbb{F}_q$-rational points of $\mathcal{X}$ is $q+1$. Thus we can construct a PHF$(N; q^{t+1}, q, w)$, for suitable values N, t, and w. Especially, if let $N = q+1$ and $t = 1$, then we can obtain a PHF$(q+1; q^2, q, w)$ for suitable values w satisfying $w(w-1) < q+1$. This result was already derived by using the affine resolvable design and $q-1$ MOLS of order q in Chapter 2.

2. Suppose that $q = p^u$, p is a prime. There are elliptic curves with $N_q(1)$ rational points, where

$$N_q(1) = \begin{cases} q + \lfloor 2\sqrt{q} \rfloor & \text{if } p \text{ divides } \lfloor 2\sqrt{q} \rfloor \text{ and } u \geq 3, \\ q + \lfloor 2\sqrt{q} \rfloor + 1 & \text{otherwise.} \end{cases}$$

   Using such curves, we can construct a PHF$(N; q^t, q, w)$, for suitable values N, t, and w as follows:

   (a) Choose positive integers t, w, and N with $2 \leq w \leq q$, $t \geq 3$, and $\binom{w}{2} t < N \leq N_q(1)$.

   (b) Choose distinct rational points $P_1, P_2, \ldots, P_N$ of F, say $T = \{P_1, P_2, \ldots, P_N\}$.

(c) Choose a divisor G of F with $deg(G) = t \geq 2g+1 = 3$ and $supp(G) \cap T = \emptyset$.

3. Let $q = r^2$ be a prime power. Use the Hermitian Curve $\mathcal{X}$ over $\mathbb{F}_q$ defined by

$$y^r + y = x^{r+1}.$$

This is the same as the construction using the Hermitian function field, as observed in Example 4.1. Thus we have a PHF$(q\sqrt{q}+1; q^{t+1-\sqrt{q}(\sqrt{q}-1)/2}, q, w)$ if $q\sqrt{q}+1 > t\binom{w}{2}$ using the above algorithm.

□

Example 4.5 Let $q = r^2$ be a prime power. Using the Garcia-Stichtenoth curve, we have a sequence of algebraic curves $\mathcal{X}_i$ as follows: Let $\mathcal{X}_1$ be the projective line, with the function field $\mathbb{F}_q(\mathcal{X}_1) = \mathbb{F}_q(x_1)$. Let $\mathcal{X}_i$ be obtained by adjoining a new equation,

$$x_i{}^r + x_i = \frac{x_{i-1}{}^r}{x_{i-1}{}^{r-1} + 1}.$$

for $i \geq 2$. Then we know that the number of $\mathbb{F}_q$- rational points of $\mathcal{X}_i$ is more than $(r-1)r^i$, and genus of $\mathcal{X}_i$ is less than r^i for all $i \geq 1$. Applying the method given in Example 4.3, for every positive integer i, we can also obtain a perfect hash family

$$\text{PHF}\left((r-1)r^i; \lceil r^{2cr^i} \rceil, r^2, \left\lfloor \left(\frac{2r}{c+1}\right)^{1/2} \right\rfloor\right),$$

where $r \geq 4$ is a prime power and c is a real number with $1 \leq c \leq (r-2)/2$.

Theorem 4.8 [41, Theorem 3.3] *For any integers $m \geq w$, there exist a construction of PHF$(N; n, m, w)$ such that $N = C \log n$, where C is a constant independent*

of n, and n can go to ∞.

Remark: Even though the results in this chapter are constructive, the preceding construction is still theoretical, so it is difficult to illustrate even one simple example. Moreover, there are several questions which must be answered in order to implement the above algorithms:

1. What kinds of function fields or algebraic curves (explicitly) shall we choose for the construction?

2. For given function field F or algebraic curve $\mathcal{X}$, how can we compute $N(F)$ and $g(F)$?

3. How can we choose a set T described in the above algorithms? Is this easy to do, according to the choice of N in step 3?

4. How can we choose a divisor G satisfying some appropriate conditions, as shown in step 4?

Chapter 5

Applications

So far, we have discussed various methods to construct perfect hash families. As mentioned before, perfect hash families have many applications in computer science, including language translation systems, hypertext, hypermedia, and file management. Here we focus on cryptographic applications such as broadcast encryption, secret sharing schemes, key distribution patterns, visual cryptography, cover-free families, traceability schemes, and multicast re-keying schemes. In later sections, we will deal with the concepts required for those applications and most of all, we consider constructions using certain perfect hash families.

5.1 Secret Sharing Schemes

This section is based on [7]. In a secret sharing scheme, we need a trusted authority, denoted TA. The TA has a secret value $K \in \mathcal{K}$, called a *secret* or a *key*, where $\mathcal{K}$ is a specified finite set. The TA uses a share generation algorithm to split K into n shares, denoted $s_1, s_2, \ldots, s_n$, where each share $s_i \in \mathcal{S}$, and $\mathcal{S}$ is a specified finite set. A share generation algorithm has to satisfy two properties: any authorized subset

can compute K from the shares they jointly hold, but no unauthorized subset has any information about K. There is a special type of secret sharing scheme called a threshold scheme.

Definition 5.1 *Let w, n be positive integers, $w \leq n$. A* (w, n)-threshold scheme *is a method of sharing a key K among a set of w users, in such a way that any n users can compute the value K, but no group of $w-1$ users can do so.*

The well-known Shamir threshold scheme invented in 1979 [28] is one way to obtain a (w, n)-threshold scheme based on polynomial interpolation over $\mathbb{Z}_p$, where $p \geq n+1$ is prime. In the Shamir scheme, the TA constructs a random polynomial $a(x) \in \mathbb{Z}_p[x]$ of degree at most $w-1$ in which the constant term is the key, K. Every participant P_i obtains a point (x_i, y_i) on this polynomial. The following explains the method to construct the scheme:

- The TA chooses n distinct, nonzero elements of $\mathbb{Z}_p$, denoted x_i, $1 \leq i \leq n$. For $1 \leq i \leq n$, the TA gives the value x_i to the participant P_i, where the value x_i is public.

- Suppose the TA wants to share a secret key $K \in \mathbb{Z}_p$, he secretly chooses (independently at random) $w-1$ elements of $\mathbb{Z}_p$, $a_1, \ldots, a_{w-1}$.

- For $1 \leq i \leq n$, the TA computes the shares $y_i = a(x_i)$, where

$$a(x) = K + \sum_{j=1}^{w-1} a_j x^j \pmod{p}.$$

- For $1 \leq i \leq n$, the TA gives the share y_i to P_i.

In [7], S. Blackburn presents a method of share expansion: Suppose there is a sharing scheme for a key K and m users. Then we can expand the sharing for the

same key K to $n > m$ users, without changing the threshold w, using a certain perfect hash family, instead of a new sharing algorithm as follows:

Theorem 5.1 *Suppose there exists a PHF$(N; n, m, w)$ and there are N independent (w, m)-threshold schemes for sharing the given secret. Then we can construct a (w, n)-threshold scheme, in which each participant receives N shares from (w, m)-threshold schemes.*

Proof

Let $\mathcal{F} = \{f_1, \ldots, f_N\}$ be an (n, m, w) perfect hash family and let $\mathcal{S} = \{S^1, S^2, \ldots, S^N\}$ be a collection of N (w, m)-threshold schemes, for sharing a given secret K, where $S^i = \{s^i{}_1, \ldots, s^i{}_m\}$ for all $1 \leq i \leq N$. Then we generate a new secret sharing scheme $\mathcal{T} = \{t_1, \ldots, t_n\}$ as follows:

for any $1 \leq j \leq n$,

$$t_j = \{s^i{}_{f_i(j)} : 1 \leq i \leq N\}.$$

The resulting scheme is a (w, n) secret sharing scheme for the given secret K. Any $w - 1$ users possess at most $w - 1$ shares from each of the (w, m) schemes and so then can know nothing of the secret. Any w users possess w distinct shares from at least one of the schemes by the construction based on the perfect hash family $\mathcal{F}$ and hence these w participants are able to determine the secret. Furthermore, the share expansion of the (w, n) secret sharing scheme is equal to N times the share expansion of the (w, m) scheme. □

Blackburn [7] points out that it is desirable to find a perfect hash family whose size is as small as possible, in order to find a secret sharing scheme having a small share expansion.

Example 5.1 Let $K = 10$ be the secret key and suppose that we have constructed

shares from four $(3,3)$ threshold schemes in $\mathbb{Z}_{13}$, say

$$\begin{aligned} S^1 &= \{{s^1}_1 = 3, {s^1}_2 = 5, {s^1}_3 = 2\}, \\ S^2 &= \{{s^2}_1 = 2, {s^2}_2 = 2, {s^2}_3 = 6\}, \\ S^3 &= \{{s^3}_1 = 1, {s^3}_2 = 5, {s^3}_3 = 4\}, \\ S^4 &= \{{s^4}_1 = 2, {s^4}_2 = 11, {s^4}_3 = 10\}. \end{aligned}$$

Then by using a PHF$(4;9,3,3)$ described in Example 1.1, we construct a new sharing $\mathcal{T} = \{t_1, \ldots, t_9\}$ for 9 users, which is a $(3,9)$ threshold scheme, as follows:

$$\begin{aligned} t_1 &= ({s^1}_{f_1(1)}, {s^2}_{f_2(1)}, {s^3}_{f_3(1)}, {s^4}_{f_4(1)}) = ({s_1}^1, {s_1}^2, {s_1}^3, {s_1}^4) = (3,2,1,2) \\ t_2 &= ({s^1}_{f_1(2)}, {s^2}_{f_2(2)}, {s^3}_{f_3(2)}, {s^4}_{f_4(2)}) = ({s_1}^1, {s_2}^2, {s_2}^3, {s_2}^4) = (3,2,5,11) \\ t_3 &= ({s^1}_{f_1(3)}, {s^2}_{f_2(3)}, {s^3}_{f_3(3)}, {s^4}_{f_4(3)}) = ({s_1}^1, {s_3}^2, {s_3}^3, {s_3}^4) = (3,6,4,10) \\ t_4 &= ({s^1}_{f_1(4)}, {s^2}_{f_2(4)}, {s^3}_{f_3(4)}, {s^4}_{f_4(4)}) = ({s_2}^1, {s_1}^2, {s_3}^3, {s_2}^4) = (5,2,4,11) \\ t_5 &= ({s^1}_{f_1(5)}, {s^2}_{f_2(5)}, {s^3}_{f_3(5)}, {s^4}_{f_4(5)}) = ({s_2}^1, {s_2}^2, {s_1}^3, {s_3}^4) = (5,2,1,10) \\ t_6 &= ({s^1}_{f_1(6)}, {s^2}_{f_2(6)}, {s^3}_{f_3(6)}, {s^4}_{f_4(6)}) = ({s_2}^1, {s_3}^2, {s_2}^3, {s_1}^4) = (5,6,5,2) \\ t_7 &= ({s^1}_{f_1(7)}, {s^2}_{f_2(7)}, {s^3}_{f_3(7)}, {s^4}_{f_4(7)}) = ({s_3}^1, {s_1}^2, {s_2}^3, {s_3}^4) = (2,2,5,10) \\ t_8 &= ({s^1}_{f_1(8)}, {s^2}_{f_2(8)}, {s^3}_{f_3(8)}, {s^4}_{f_4(8)}) = ({s_3}^1, {s_2}^2, {s_3}^3, {s_1}^4) = (2,2,4,2) \\ t_9 &= ({s^1}_{f_1(9)}, {s^2}_{f_2(9)}, {s^3}_{f_3(9)}, {s^4}_{f_4(9)}) = ({s_3}^1, {s_3}^2, {s_1}^3, {s_2}^4) = (2,6,1,11). \end{aligned}$$

□

5.2 Visual Cryptography

A *Visual Cryptography Scheme*(VCS) for a set $\mathcal{P}$ of n participants is a method to encode a secret image (SI) into n shadow images called shares, where each participant in $\mathcal{P}$ receives one share. Certain qualified subsets of participants can *visually* recover the SI, but other, forbidden sets of participants have no information on SI. A 'visual recovery' of the qualified set X means that they can see the SI by xeroxing the shares given to the participants in X onto transparencies, and then stacking them. Thus the participants in a qualified set X will be able to see the SI without any knowledge of cryptography and without performing any cryptographic computation. Specifically, we focus on the construction of (w, n)-threshold VCS using perfect hash families. Before we introduce such a method, we will observe some notions and properties on the VCS, which are based on the results in [2] and [23].

Definition 5.2 *Let* $\mathcal{P} = \{1, 2, \ldots, n\}$ *be a set of elements called* participants *and let* $2^{\mathcal{P}}$ *denote the collection of all subsets of* $\mathcal{P}$. *Let* $\Gamma_Q \subseteq 2^{\mathcal{P}}$ *and* $\Gamma_F \subseteq 2^{\mathcal{P}}$, *where* $\Gamma_Q \cap \Gamma_F = \emptyset$. *We call the members of* Γ_Q qualified sets *; and members of* Γ_F *are called* forbidden sets. *The pair* (Γ_Q, Γ_F) *is called the* access structure *of the scheme.*

Define Γ_0 *to consist of all minimal qualified sets:*

$$\Gamma_0 = \{A \in \Gamma_Q : A' \notin \Gamma_Q \quad \textit{for all } A' \subset A\}.$$

Example 5.2 Let $\mathcal{P} = \{1, 2, 3, 4\}$. Define the qualified sets to be

$$\Gamma_Q = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3\}\}.$$

Then any subset of $2^{\mathcal{P}} \setminus \Gamma_Q$ can be a forbidden set Γ_F and the pair (Γ_Q, Γ_F) is an access structure. Note that $\{1,2\},\{2,3\} \subseteq \{1,2,3\}$; $\{1,2\},\{2,3\} \in \Gamma_Q$, and hence, we have

$$\Gamma_0 = \{\{1,2\},\{2,3\},\{3,4\}\}.$$

□

In the VCS, it is assumed that the message consists of a collection of black and white pixels. Throughout this section, a boolean matrix $S = (s_{ij})$ represents the scheme as follows:

$$(s_{ij}) = 1 \quad \Longleftrightarrow \quad \text{the } j\text{th subpixel in the } i\text{th share is black.}$$

Definition 5.3 [2] *Let (Γ_Q, Γ_F) be an access structures on a set of n participants. A (Γ_Q, Γ_F, l)-VCS with the relative difference $\alpha(l)$ and set of thresholds $\{(X, t_X)\}_{X \in \Gamma_Q}$ is realized using the two $n \times l$ basis matrices S^0 and S^1 if the following two conditions hold.*

1. ***Contrast condition:*** *If $X = \{i_1, i_2, \ldots, i_P\} \in \Gamma_Q$, then the "or" V of rows $i_1, i_2, \ldots, i_P$ of S^0 satisfies $w(V) \leq t_X - \alpha(l) \cdot l$; whereas, for S^1 it results that $w(V) \geq t_X$.*

2. ***Security condition:*** *If $X = \{i_1, i_2, \ldots, i_P\} \in \Gamma_F$, then the two $p \times l$ matrices obtained by restricting S^0 and S^1 to rows $i_1, i_2, \ldots, i_P$ are identical up to a column permutation.*

Example 5.3 By using a random column permutation, we can construct suitable shares, which is used as basis matrices for VCS. For example, if we construct a $(2,2)$-scheme, then we can construct basis matrices by random column permutation from

the following set of columns:

$$S^0 \Longleftarrow \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad S^1 \Longleftarrow \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

. □

Actually, a VCS can be regarded as a visual variant of the w out of n secret sharing problem. If we consider the scheme with the strong access structure having basis

$$\Gamma_0 = \{B \subseteq \mathcal{P} : |B| = w\},$$

then, for n participants, any w (or more than w) participants of them can see the secret image by stacking their transparency (by the contrast condition), but any set of at most $w-1$ of them gain no information about it (by the security condition). Hence, we can obtain a (w, n)-threshold VCS. In the previous section, we discussed a way to expand the secret sharing scheme using perfect hash families. Here we will introduce a similar construction for a (w, n)-threshold VCS using a PHF$(N; n, m, w)$. To do this, we need two basis matrices S^0 and S^1 for a small (w, m)-threshold VCS for $m \geq w$. Naor and Shamir first presented a method to construct (w, m)-threshold VCS for $w \geq m$ in [23]. We first deal with this base constructions, and then we generalize it.

Algorithm 3 [Construction of (w, w)-threshold VCS] We construct two basis matrices S^0 and S^1 for a (w, w)-threshold VCS with 2^{w-1} subpixels as follows:

Step 1. Consider the ground set $G = \{g_1, g_2, \ldots, g_w\}$ of w elements.

Step 2. Let $E = \{\pi_1, \pi_2, \ldots, \pi_{2^{w-1}}\}$ be the collection of all subsets of W with even cardinality, and let $O = \{\sigma_1, \sigma_2, \ldots, \sigma_{2^{w-1}}\}$ be the collection of all subsets of W with odd cardinality.

Step 3. For $1 \leq i \leq w$ and $1 \leq j \leq 2^{w-1}$, define matrices $S^0 = ({s_{ij}}^0)$ and $S^1 = ({s_{ij}}^1)$ as follows:

$$\begin{aligned} {s_{ij}}^0 &= 1 \iff g_i \in \pi_j, \\ {s_{ij}}^1 &= 1 \iff g_i \in \sigma_j. \end{aligned}$$

To illustrate the above algorithm, we provide an example.

Example 5.4 Let $w = 3$. Then $G = \{1, 2, 3\}$,

$$\begin{aligned} E &= \{\emptyset, \{1,2\}, \{1,3\}, \{2,3\}\}, \quad \textit{and} \\ O &= \{\{1\}, \{2\}, \{3\}, \{1,2,3\}\}. \end{aligned}$$

Then we have $(3,3)$-threshold VCS with 4 subpixels, which has two basis matrices, S^0 and S^1 as follows:

$$S^0 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad S^1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Note that the "or" V of all rows of S^0 is (0111), i.e., $w(V) = 3$; and the "or" V of all rows of S^1 is (1111), i.e., $w(V) = 4$, which satisfies the contrast condition. Moreover, all 2×3 submatrices obtained by restricting to any two rows of S^0 and S^1 are equal up to column permutation, which is the security condition. $\square$

Now, we describe the method to construct (w, m)-threshold VCS using an (m, w, w) perfect hash family for $w < m$.

Algorithm 4 [Construction of (w, m)-threshold VCS using perfect hash families]

We can construct two basis matrices T^0 and T^1 for a (w, m)-threshold VCS as follows:

Step 1. Construct matrices S^0 and S^1 for a (w, w)-threshold VCS using the above algorithm.

Step 2. Construct an (m, w, w) perfect hash family, say $\mathcal{F} = \{f_1, f_2, \ldots, f_s\}$. Clearly, $s \geq N(m, w, w)$.

Step 3. For $1 \leq i \leq s$, define matrices ${B_i}^0$ and ${B_i}^1$, respectively, by

$$\begin{aligned} j\text{th row of } {B_i}^0 &= f_i(j)\text{th row of } S^0, \\ j\text{th row of } {B_i}^1 &= f_i(j)\text{th row of } S^1, \end{aligned}$$

for $1 \leq j \leq m$.

Step 4. Define basis matrices T^0 and T^1 for (w, m)-threshold VCS by

$$\begin{aligned} T^0 &= ({B_1}^0 : {B_2}^0 : \cdots : {B_s}^0), \\ T^1 &= ({B_1}^1 : {B_2}^1 : \cdots : {B_s}^1). \end{aligned}$$

Example 5.5 To illustrate the above algorithm, let $w = 3$ and $m = 4$. There are two 3×4 matrices S^0 and S^1 for $(3,3)$-threshold VCS given in Example 5.4. And there is a PHF$(2; 4, 3, 3)$, given in [3], as follows:

	1	2	3	4
f_1	1	2	3	3
f_2	1	1	2	3

Then

$$B_1{}^0 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad B_2{}^0 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

and

$$B_1{}^1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad B_2{}^1 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Thus, we have $(3,4)$-threshold VCS with 8 subpixels, which has two basis matrices, as follows;

$$T^0 = \begin{pmatrix} 0 & 1 & 1 & 0 & : & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & : & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & : & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & : & 0 & 0 & 1 & 1 \end{pmatrix},$$

and

$$T^1 = \begin{pmatrix} 1 & 0 & 0 & 1 & : & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & : & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & : & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & : & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Note that the "or" V of any 3 rows of T^0 is (01110111), i.e., $w(V) = 6$ and the "or" V of any 3 rows of T^1 has Hamming weight $w(V) = 7$, which satisfies the contrast condition. Moreover, all 3×8 submatrices, obtained by restricting to any three rows of T^0 and T^1, are equal up to a column permutation, which is the security condition. □

Theorem 5.2 *Suppose there exists a PHF*$(N; n, m, w)$. *Then we can construct a* (w, n)*-threshold VCS.*

Proof It suffices to construct two basis matrices M_0 and M_1 for (w, n)-threshold VCS. Given a perfect hash family, we apply Algorithm 4 to construct a (w, n)-threshold VCS replacing m with n. More precisely, if $w = m$, then the basis matrices T^0 and T^1 in Algorithm 4 are the same as the basis matrices M_0 and M_1 for (w, n)-threshold VCS. If $w < m$, then we can use Algorithm 4 twice, that is, after constructing two basis matrices T^0 and T^1 for a (w, m)-threshold VCS, we replace T^0 and T^1 with S_0 and S_1. respectively, in Step 1 of Algorithm 4. And then, in Step 3 s and m are regarded as N and n, respectively. Finally, we obtain basis matrices M_0 and M_1 for (w, n)-threshold VCS. □

Remark: Given a PHF$(N; n, m, w)$, we can choose the method to construct S_0 and S_1 for (w, m)-threshold VCS, i.e., if $w = m$, then the size of S_0 and S_1 is $w \times 2^{w-1}$; otherwise, $w \times s \cdot 2^{w-1}$, which is at least $N(m, w, w) \cdot 2^{w-1}$. And then by

applying the similar method of step 3 of Algorithm 4 using a PHF$(N; n, m, w)$ to obtain basis matrices. From those blocks, we construct basis matrices for M_0 and M_1 for (w, n)-threshold VCS. The resulting basis matrices M_0 and M_1 have the size $n \times N \cdot l'$, where l' represents the number of columns of the initial basis matrices. In the above process, for the value l, a number of subpixels of a (w, n)-threshold VCS, depends on the value N and the number of subpixels l' of a (w, m)-threshold VCS.

Actually, Naor and Shamir provided a (w, n)-threshold VCS with $l = 2^{O(w \log w)} \cdot \log n$ in [23]. Since it is attempted to find the scheme with l as small as possible, it is important to obtain a PHF$(N; n, m, w)$ such that the value N as small as possible. More exactly, we have a (w, n)-threshold VCS with $l = N \cdot 2^{w-1}$ provided that a PHF$(N; n, w, w)$ exists. Generally, we have a (w, n)-threshold VCS with $l = N \cdot N_0 \cdot 2^{w-1}$ provided that a PHF$(N; n, w, w)$ and a PHF$(N_0; m, w, w)$, namely, $l = O(\log n)$ for fixed w and m.

To illustrate the above theorem, we provide an example.

Example 5.6 We have a PHF$(2; 5, 4, 3)$ described in Figure 5.1. From Example

1	2	3	4	3
1	1	2	1	3

FIGURE 5.1: A PHF$(2; 5, 4, 3)$

5.5, we have basis matrices S^0 and S^1 for a $(3, 4)$-threshold VCS. Then we can obtain a $(3, 5)$-threshold VCS with 16 pixels, having the basis matrices M_0 and M_1

as follows:

$$M_0 = \begin{pmatrix} 01010110 & \vdots & 01100110 \\ 01010110 & \vdots & 01100110 \\ 00110101 & \vdots & 01010110 \\ 00110011 & \vdots & 01100110 \\ 00110101 & \vdots & 00110101 \end{pmatrix},$$

and

$$M_1 = \begin{pmatrix} 10010101 & \vdots & 10010101 \\ 01011001 & \vdots & 10010101 \\ 00110101 & \vdots & 01011001 \\ 00110011 & \vdots & 10010101 \\ 00110101 & \vdots & 00110101 \end{pmatrix}.$$

□

A (w, n)-threshold VCS can be constructed with different basis matrices, using different perfect hash families. For example, we can construct a $(3, 5)$-threshold VCS which is different with that given in Exmple 5.6 as follows:

Example 5.7 We have a $(3, 5)$-threshold VCS with 12 subpixels using a PHF$(3; 5, 3, 3)$. Form two basis matrix S^0 and S^1 of the $(3, 3)$-threshold VCS given in Example 5.4 and a PHF$(3; 5, 3, 3)$ given in Table 3.2. Then we can obtain a $(3, 5)$-threshold VCS with 12 pixels having basis matrices M_0 and M_1, as follows:

$$M_0 = \begin{pmatrix} 0110 & \vdots & 0110 & \vdots & 0110 \\ 0101 & \vdots & 0110 & \vdots & 0110 \\ 0011 & \vdots & 0101 & \vdots & 0110 \\ 0011 & \vdots & 0011 & \vdots & 0101 \\ 0011 & \vdots & 0011 & \vdots & 0011 \end{pmatrix},$$

and

$$M_1 = \begin{pmatrix} 1001 & \vdots & 1001 & \vdots & 1001 \\ 0101 & \vdots & 1001 & \vdots & 1001 \\ 0011 & \vdots & 0101 & \vdots & 1001 \\ 0011 & \vdots & 0011 & \vdots & 0101 \\ 0011 & \vdots & 0011 & \vdots & 0011 \end{pmatrix}.$$

□

5.3 Cover-free Families

There are many applications of (t, w)-CFF to multicast security, including group key predistribution and group session key distribution. Specifically, in the case $t = 1$, $(1, w)$-CFF are used to construct some schemes such as blacklisting (broadcast exclusion) and anti-jamming schemes, and to provide network source authentication. This section is based on [30].

Definition 5.4 *Let $(X, \mathcal{A})$ be a set system with $X = \{x_1, \ldots, x_v\}$ and let $\mathcal{A} = \{A_i \subseteq X : i = 1, \ldots, n\}$. $(X, \mathcal{A})$ is a* a (t, w)-cover free family *provided that, for any two disjoint subsets of blocks $P, F \subseteq \mathcal{A}$, where $|P| = t$ and $|F| = w$, it holds*

that

$$\bigcap_{A_i \in P} A_i \not\subseteq \bigcup_{A_j \in F} A_j.$$

Such a (t, w)-cover-free family is denoted as a (t, w)-CFF(v, n). In particular, we denote the case $t = 1$ by w-CFF(v, n), instead of $(1, w)$-CFF, i.e., if for any $A \subseteq \mathcal{A}$ with $|A| \leq w$, and for any $B \in \mathcal{A} \backslash A$, it holds that

$$B \not\subseteq \bigcup_{A_0 \in \mathcal{A}} A_0.$$

In other words, in a w-CFF(v, n), the union of any w blocks in $\mathcal{A}$ cannot cover any other one.

Let us look at a method to use a perfect hash family to construct a cover-free family.

Theorem 5.3 *Suppose there exists a PHF*$(N; n, m, w)$*. Then there exists a* $(w-1)$*-CFF*(Nm, n)*.*

Proof Let $\mathcal{F} \subseteq \{f : \mathbf{A} \to \mathbf{B}\}$ be an $(N; n, m, w)$ perfect hash family.

We define

$$X = \mathcal{F} \times \mathbf{B} = \{(f, j) : f \in \mathcal{F},\ j \in \mathbf{B}\}.$$

For each $1 \leq i \leq n$, we define a block A_i of X by

$$A_i = \{(f, f(i)) : f \in \mathcal{F}\},$$

and $\mathcal{A} = \{A_i : 1 \leq i \leq n\}$. Then $(X, \mathcal{A})$ is a $(w-1)$-CFF(Nm, n). Clearly, $|X| = Nm$ and $|\mathcal{A}| = n$. For any w blocks $A_{i_1}, A_{i_2}, \ldots, A_{i_w}$, since $\mathcal{F}$ is a PHF$(N; n, m, w)$, there exists a function $f \in \mathcal{F}$ such that f restricted on $\{i_1, i_2, \ldots, i_w\}$ is one-to-one. It follows that $f(i_1), \ldots, f(i_w)$ are w distinct elements in $\mathbf{B}$, which also implies that $(f, f(i_1)), \ldots, (f, f(i_w))$ are w distinct elements in $A_{i_1}, \ldots, A_{i_w}$, respectively. So the

union of any $w-1$ blocks in $\mathcal{A}$ cannot cover the remaining one. □

Actually, we can define a suitable incidence matrix for a cover-free family using a PHF$(N;n,m,w)$. First, for $1 \leq i \leq N$, we define an $m \times n$ matrix M_i, in which each column and row represents an element of $\mathbf{A} = \{1,2,\ldots,n\}$ and $\mathbf{B} = \{1,2,\ldots,m\}$, respectively as follows:

$$\mathcal{M}_i = (m_{k,j})_{1\leq k\leq m,\ 1\leq j\leq n} = \begin{cases} 1 & \text{if } \ f_i(j) = k \\ 0 & \text{otherwise.} \end{cases}$$

Now, we define an incidence matrix $\mathcal{M}$ of size $Nm \times n$ from N matrices M_i as follows:

$$\mathcal{M} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix}$$

Then, in $\mathcal{M}$, each row and column represents a point and a block of the CFF, respectively.

Example 5.8 We can construct a 2-CFF$(12,9)$ from a PHF$(4;9,3,3)$ described in Figure 1.1 as follows: First, we construct the incidence matrix $\mathcal{M}$, as described above. Then

$$\mathcal{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Finally

$$\begin{aligned} X &= \{1, 2, \ldots, 12\}, \text{ and} \\ \mathcal{A} &= \{\{1,4,7,10\}, \{1,5,8,11\}, \{1,6,9,12\}, \{2,4,9,11\}, \{2,5,7,12\}, \\ &\quad \{2,6,8,10\}, \{3,4,7,12\}, \{3,5,8,10\}, \{3,6,9,12\}\} \end{aligned}$$

is a set system which is 2-CFF$(12, 9)$. □

A (w, n)-threshold VCS, which we considered in the previous section, can also be used to construct $(w-1)$-CFF(l, n), where $l = N \cdot 2^{w-1}$, as follows ([2]): Let S^0 and S^1 be basis matrices of a (w, n)-threshold VCS with l subpixels on the set $\mathcal{P}$ of n participants. Let $G = \{g_1, g_2, \ldots, g_l\}$ be a ground set of l elements. For $i = 1, \ldots, n$, row i of S^1 represents the set $A_i = \{g_p : M(i,p) = 1\}$. Because of

the contrast condition, for any set $Y = \{j_1, j_2, \ldots, j_w\} \subseteq \{1, 2, \ldots, n\}$, the matrix $M[Y]$, for each row $i \in \{1, \ldots, w\}$, has at least one column with a "1" in the ith row and "0"s in the other rows. This implies that the sets $A_{j_1}, A_{j_2}, \ldots, A_{j_w}$ are such that the union of any $w-1$ of them does not contain the remaining one. Hence, any matrix $M \in C_1$ represents a family $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ of subsets over the ground set G having the property that the union of any $w-1$ of them does not cover any of the remaining sets. Thus $\mathcal{A}$ is a $(w-1)$-CFF(l, n).

Example 5.9 To illustrate the above observation, we have 2-CFF$(16, 9)$ from $(3, 9)$-threshold VCS with 16 subpixels which is constructed by a PHF$(4; 9, 3, 3)$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A_1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
A_2	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
A_3	1	0	0	1	0	0	1	1	0	0	1	1	0	0	1	1
A_4	0	1	0	1	1	0	0	1	0	0	1	1	0	1	0	1
A_5	0	1	0	1	1	0	0	1	1	0	0	1	0	0	1	1
A_6	0	1	0	1	0	0	1	1	0	1	0	1	1	0	0	1
A_7	0	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1
A_8	0	0	1	1	0	1	0	1	0	0	1	1	1	0	0	1
A_9	0	0	1	1	0	0	1	1	1	0	0	1	0	1	0	1

□

5.4 Broadcast Encryption

Broadcast encryption was first introduced by Fiat and Naor [14]. Let K be secret information, which is to be broadcast. Let P be a privileged subset of the users $\mathcal{U}$.

Generally, a broadcast encryption scheme (BES) consists of the following phases:

1. Split K into shares $s_1, s_2, \ldots, s_v$ using a (t, v)-threshold scheme.

2. Encrypt every share s_i with a key k_i in such a way that

   (a) every user $U_j \in P$ can compute at least t of the keys $k_1, k_2, \ldots, k_v$, that is, they can decrypt t shares of K, and then reconstruct K.

   (b) any coalition F, such that $F \cap P = \emptyset$ and $|F| \leq w$, can compute at most $t-1$ of the keys, that is, they can decrypt at most $t-1$ shares of K, and therefore they cannot obtain any information about K.

The keys k_i are obtained from suitable key predistribution scheme(s). For example, Fiat-Naor Key distribution patterns (KDPs) can be used to construct an efficient BES. Here we consider a method to construct broadcast encryption scheme using perfect hash families.

Algorithm 5 [Broadcast Encryption Scheme]
Let $\mathcal{F} = \{f_1, f_2, \ldots, f_N\}$ be a PFH$(N; n.m.w)$.

Step 1. Construct an $Nm \times n$ incidence matrix $\mathcal{M}$ as described in the previous section, where each row represents Nm schemes for key distribution and each column is indexed by one of the n users. Give an index (i, j), for $1 \leq j \leq m$ and $1 \leq i \leq N$ to row of $\mathcal{M}$, and k for $1 \leq k \leq n$ to column of $\mathcal{M}$.

Step 2. For $\mathcal{M} = (m_{(i,j),k})$, we define

$$\begin{aligned} users(\mathcal{F}_{i,j}) &:= \{u_k : m_{(i,j),k} = 1\} = f_i^{-1}(j), \quad \text{and} \\ schemes(k) &:= \{\mathcal{F}_{i,j} : m_{(i,j),k} = 1\} = \{\mathcal{F}_{i,f_i(k)} : 1 \leq i \leq N\}. \end{aligned}$$

Step 3. Construct an initial key for each user in $\mathcal{U}$ as follows:

- A key $l_{i,j}$ is given to every user in $users(\mathcal{F}_{i,j})$.
- For every user $u_h \in users(\mathcal{F}_{i,j})$, a key ${l^h}_{i,j}$ is given to every user in the set $users(\mathcal{F}_{i,j}) \backslash \{u_h\}$.

Step 4. Split the secret $K \in \mathbb{Z}_p$ into N shares, using the (N, N)-threshold scheme. We denote these shares as $s_1, \ldots, s_N$, and $K = \sum_{i=1}^{N} s_i \pmod{p}$.

Step 5. Let $P \subseteq \mathcal{U}$ be a set of users which K is being broadcast. For all i, j, construct the group keys $k_{i,j}$ for $P \cap users(\mathcal{F}_{i,j})$ as follows:

$$k_{i,j} = l_{i,j} + \sum_{\{u_h \in users(\mathcal{F}_{i,j}) \backslash P\}} {l^h}_{i,j}$$

Step 6. Use $k_{i,j}$ to encrypt s_i for all i, j, obtaining for $1 \leq i \leq N$ and $1 \leq j \leq m$, the ciphertext $E_{k_{i,j}}(s_i)$.

Step 7. Broadcast the Nm encryption of the shares. If we let

$$b_i = \{E_{k_{i,1}}(s_i), E_{k_{i,2}}(s_i), \ldots, E_{k_{i,m}}(s_i)\},$$

then the broadcast information b_P can be denoted $\bigcup_{i=1}^{N} b_i$.

Step 8. Each user $u_k \in P$ carries out the following process:

- For all $\mathcal{F}_{i,j} \in \text{schemes}(k)$, u_k construct the group keys $k_{i,j}$ for $P \cap users(\mathcal{F}_{i,j})$.
- u_k decrypt b_i using the key $k_{i,j}$ and obtain share s_i for all $1 \leq i \leq N$.
- u_k compute $\sum_{i=1}^{N} s_i \pmod{p}$ to obtain the secret K.

Let us illustrate this algorithm with an example.

Example 5.10 We can construct a broadcast encryption scheme using a PHF$(4; 9, 3, 3)$ in Figure 1.1.

1. We use the incidence matrix $\mathcal{M}$ constructed in Example 5.8.

2. From the matrix $\mathcal{M}$, we have

$$\begin{aligned}
users(\mathcal{F}_{1,1}) &= \{u_1, u_2, u_3\}, & users(\mathcal{F}_{1,2}) &= \{u_4, u_5, u_6\},\\
users(\mathcal{F}_{1,3}) &= \{u_7, u_8, u_9\}, & users(\mathcal{F}_{2,1}) &= \{u_1, u_4, u_7\},\\
users(\mathcal{F}_{2,2}) &= \{u_2, u_5, u_8\}, & users(\mathcal{F}_{2,3}) &= \{u_3, u_6, u_9\},\\
users(\mathcal{F}_{3,1}) &= \{u_1, u_5, u_7\}, & users(\mathcal{F}_{3,2}) &= \{u_2, u_6, u_8\},\\
users(\mathcal{F}_{3,3}) &= \{u_3, u_4, u_9\}, & users(\mathcal{F}_{4,1}) &= \{u_1, u_6, u_8\},\\
users(\mathcal{F}_{4,2}) &= \{u_2, u_4, u_9\}, & users(\mathcal{F}_{4,3}) &= \{u_3, u_5, u_7\}.
\end{aligned}$$

3. A total of 12 keys will be given to each user as indicated below:

u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9
$l_{1,1}$	$l_{1,1}$	$l_{1,1}$	$l_{1,2}$	$l_{1,2}$	$l_{1,2}$	$l_{1,3}$	$l_{1,3}$	$l_{1,3}$
${l^2}_{1,1}$	${l^1}_{1,1}$	${l^1}_{1,1}$	${l^5}_{1,2}$	${l^4}_{1,2}$	${l^4}_{1,2}$	${l^8}_{1,3}$	${l^7}_{1,3}$	${l^7}_{1,3}$
${l^3}_{1,1}$	${l^3}_{1,1}$	${l^2}_{1,1}$	${l^6}_{1,2}$	${l^6}_{1,2}$	${l^5}_{1,2}$	${l^9}_{1,3}$	${l^9}_{1,3}$	${l^8}_{1,3}$
$l_{2,1}$	$l_{2,2}$	$l_{2,3}$	$l_{2,1}$	$l_{2,2}$	$l_{2,3}$	$l_{2,1}$	$l_{2,2}$	$l_{2,3}$
${l^4}_{2,1}$	${l^5}_{2,2}$	${l^6}_{2,3}$	${l^1}_{2,1}$	${l^2}_{2,2}$	${l^3}_{2,3}$	${l^1}_{2,1}$	${l^2}_{2,2}$	${l^3}_{2,3}$
${l^7}_{2,1}$	${l^8}_{2,2}$	${l^9}_{2,3}$	${l^7}_{2,1}$	${l^8}_{2,2}$	${l^9}_{2,3}$	${l^4}_{2,1}$	${l^5}_{2,2}$	${l^6}_{2,3}$
$l_{3,1}$	$l_{3,2}$	$l_{3,3}$	$l_{3,3}$	$l_{3,1}$	$l_{3,2}$	$l_{3,1}$	$l_{3,2}$	$l_{3,3}$
${l^5}_{3,1}$	${l^6}_{3,2}$	${l^4}_{3,3}$	${l^1}_{3,1}$	${l^2}_{3,2}$	${l^3}_{3,3}$	${l^1}_{3,1}$	${l^2}_{3,2}$	${l^3}_{3,3}$
${l^7}_{3,1}$	${l^8}_{3,2}$	${l^9}_{3,3}$	${l^7}_{3,1}$	${l^8}_{3,2}$	${l^9}_{3,3}$	${l^5}_{3,1}$	${l^6}_{3,2}$	${l^4}_{3,3}$
$l_{4,1}$	$l_{4,2}$	$l_{4,3}$	$l_{4,2}$	$l_{4,3}$	$l_{4,1}$	$l_{4,3}$	$l_{4,1}$	$l_{4,2}$
${l^6}_{4,1}$	${l^4}_{4,2}$	${l^5}_{4,3}$	${l^2}_{4,2}$	${l^3}_{4,3}$	${l^1}_{4,1}$	${l^3}_{4,3}$	${l^1}_{4,1}$	${l^2}_{4,2}$
${l^8}_{4,1}$	${l^9}_{4,2}$	${l^7}_{4,3}$	${l^9}_{4,2}$	${l^7}_{2,3}$	${l^8}_{4,1}$	${l^5}_{4,3}$	${l^6}_{4,1}$	${l^4}_{4,2}$

4. TA sets up a $(4, 4)$ threshold scheme in $\mathbb{Z}_p$. Let s_1, s_2, s_3, s_4 be the four shares of the secret $K = \sum_{i=1}^{4} s_i \pmod{p}$.

5. Suppose that the TA wants to broadcast a message to the set $P = \{u_1, u_2, u_3, u_4\}$. The following are the keys used in this scheme:

(i,j)	$users(\mathcal{F}_{i,j}) \cap P$	$k_{i,j}$
$(1,1)$	$\{1,2,3\}$	$l_{1,1}$
$(1,2)$	$\{4\}$	$l_{1,2} + l^5{}_{1,2} + l^6{}_{1,2}$
$(1,3)$	$\emptyset$	$l_{1,3} + l^7{}_{1,3} + l^8{}_{1,3} + l^9{}_{1,3}$
$(2,1)$	$\{1,4\}$	$l_{2,1} + l^7{}_{2,1}$
$(2,2)$	$\{2\}$	$l_{2,2} + l^5{}_{2,2} + l^8{}_{2,2}$
$(2,3)$	$\{3\}$	$l_{2,3} + l^6{}_{2,3} + l^9{}_{2,3}$
$(3,1)$	$\{1\}$	$l_{3,1} + l^5{}_{3,1} + l^7{}_{3,1}$
$(3,2)$	$\{2\}$	$l_{3,2} + l^6{}_{3,2} + l^8{}_{3,2}$
$(3,3)$	$\{3,4\}$	$l_{3,3} + l^9{}_{3,3}$
$(4,1)$	$\{1\}$	$l_{4,1} + l^6{}_{4,1} + l^8{}_{4,1}$
$(4,2)$	$\{2,4\}$	$l_{4,2} + l^9{}_{4,2}$
$(4,3)$	$\{3\}$	$l_{4,3} + l^5{}_{4,3} + l^7{}_{4,3}$

6. The broadcast b_P consists of the following values:

$$
\begin{aligned}
b_{1,1} &= E_{l_{1,1}}(s_1) \\
b_{1,2} &= E_{l_{1,2}+l^5{}_{1,2}+l^6{}_{1,2}}(s_1) \\
b_{1,3} &= E_{l_{1,3}+l^7{}_{1,3}+l^8{}_{1,3}+l^9{}_{1,3}}(s_1) \\
b_{2,1} &= E_{l_{2,1}+l^7{}_{2,1}}(s_2) \\
b_{2,2} &= E_{l_{2,2}+l^5{}_{2,2}+l^8{}_{2,2}}(s_2) \\
b_{2,3} &= E_{l_{2,3}+l^6{}_{2,3}+l^9{}_{2,3}}(s_2) \\
b_{3,1} &= E_{l_{3,1}+l^5{}_{3,1}+l^7{}_{3,1}}(s_3) \\
b_{3,2} &= E_{l_{3,2}+l^6{}_{3,2}+l^8{}_{3,2}}(s_3) \\
b_{3,3} &= E_{l_{3,3}+l^9{}_{3,3}}(s_3)
\end{aligned}
$$

$$
\begin{aligned}
b_{4,1} &= E_{l_{4,1}+l^{6}{}_{4,1}+l^{8}{}_{4,1}}(s_4) \\
b_{4,2} &= E_{l_{4,2}+l^{9}{}_{4,2}}(s_4) \\
b_{4,3} &= E_{l_{4,3}+l^{5}{}_{4,3}+l^{7}{}_{4,3}}(s_4)
\end{aligned}
$$

7. Any user in P can compute the group key $k_{i,j}$ and then obtain all shares s_1, s_2, s_3, s_4 and secret K. However, any coalition of size 3 cannot obtain at least one share and hence they cannot reconstruct K. For example, if $F = \{u_5, u_6, u_7\}$ is a coalition, then they can obtain s_1, s_3, s_4 from decrypting $b_{1,2}, b_{3,1}, b_{4,3}$, but not s_2. Thus they cannot obtain the secret K.

□

Theorem 5.4 *If there exists a PHF$(N; n, m, w)$, then there exists a broadcast encryption scheme which is a secure against coalitions of size w.*

Proof It suffices to show that the scheme described above is secure against coalition of size w. Note that the group keys $k_{i,j}$ can be computed by all members of $users(\mathcal{F}_{i,j}) \cap P$. and no individual user not in $users(\mathcal{F}_{i,j})$ can compute $k_{i,j}$, but a subset of two users in $users(\mathcal{F}_{i,j}) \backslash P$ can compute $k_{i,j}$. For any coalition F, where $F \cap P = \emptyset$ and $|F| \leq w$, say $F = \{u_{h_1}, \ldots, u_{h_w}\}$, there exists a function $f_i \in \mathcal{F}$ such that $f_i(h_1), f_i(h_2), \ldots, f_i(h_w)$ are distinct. Thus there exists i such that no subset of two users is in $(users(\mathcal{F}_{i,j}) \backslash P) \cap F$ for all $1 \leq j \leq m$. Thus the users in F cannot obtain $k_{i,j}$ for all $1 \leq j \leq m$ to decrypt b_i. Thus they can obtain at most $N - 1$ shares, and hence they cannot obtain any information about K. □

Remark: We can apply any w-CFF(v, n) to a broadcast encryption scheme for n users and v schemes, which is a secure against coalitions of size w.

5.5 A Multicast Re-Keying Scheme

In a multicast network, the group key is the initial shared key between all the users, and it is used for encrypting the broadcasted information. Here we will consider the situation when the users in the network change, i.e., a new session key for the new group of users is needed.

In this section, a re-keying scheme is regarded as a group eviction problem. The concepts and ideas are mostly based on [26]. That is, re-keying the multicast group $\mathcal{U}\backslash\mathcal{M}$ involves evicting the users in $\mathcal{M}$. A re-keying scheme is *w-resilient* if a coalition of up to w users from $\mathcal{M}$ is unable to compute any keys in the set of new session keys. Although there are many methods to solve the re-keying problem, we will deal with two efficient schemes, the OR scheme and the AND scheme, based on the existence of a perfect hash family. The OR scheme can be only used to remove a small number of users; the AND scheme can be used to remove a large number of users.

We need some assumptions and notations as follows:

- the group controller(GC) knows all the system keys.
- $\mathcal{U}$: the set of users, say $\mathcal{U} = \{u_1, \ldots, u_n\}$.
- $\mathcal{M}$: the set of evicted members.
- $k^{\mathcal{U}}$: a session key for users is $\mathcal{U}$.
- $\mathcal{K}$: a set of auxiliary keys
- $\mathcal{K}(u_i) \subseteq \mathcal{K}$: the set of auxiliary keys of the user u_i.
- $M(\mathcal{F})$: an incidence matrix of size $Nm \times n$ based on a function family $\mathcal{F}$.

5.5.1 First Version of a Re-Keying Scheme

Assume that the users in $\mathcal{U}$ know an initial session key $k^{\mathcal{U}}$. First, we consider the case when GC is not a group member. A basic re-keying scheme consists of the following three phases:

1. Key initialization: the GC generates $\mathcal{K}$ and then sends $\mathcal{K}(u_i)$ to users u_i, for $1 \leq i \leq n$.

2. Broadcast: the GC broadcasts an encrypted (new) session key which is only decryptable by a specified group.

3. Decryption: the authorized users are able to decrypt the encrypted session key while unauthorized users are not.

Let $\mathcal{F} \subseteq \{f : \mathbf{A} \to \mathbf{B}\}$ be a PHF$(N; n, m, w+1)$.
First, we present the OR scheme using the perfect hash family, $\mathcal{F}$. The scheme works as follows.

1. Key initialization:

   - GC generates $\mathcal{K} = \{k_{(f,m)} : f \in \mathcal{F}, m \in \mathbf{B}\}$ and then sends $\mathcal{K}(u_i) = \{k_{f,f(i)} : f \in \mathcal{F}\}$ to users u_i, for $1 \leq i \leq n$.

2. Broadcast: Assume that $\mathcal{M}$ is the set of users to be removed and assume $|\mathcal{M}| \leq w$,

   - GC randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$
   - GC encrypts it with keys $k \in \mathcal{K}\setminus\mathcal{K}(\mathcal{M})$.
   - GC broadcasts $\{E_k(k^{\mathcal{U}\setminus\mathcal{M}}) : k \in \mathcal{K}, k \notin \mathcal{K}(\mathcal{M})\}$

3. Decryption:

   - Each user u_i in $\mathcal{U}\backslash\mathcal{M}$ uses one of his auxiliary keys $k \in \mathcal{K}(u_i)$ to decrypt $E_k(k^{\mathcal{U}\backslash\mathcal{M}})$ and obtain the key $k^{\mathcal{U}\backslash\mathcal{M}}$.

Theorem 5.5 *If there exists a PHF$(N; n, m, w+1)$, then there exists a w-resilient re-keying scheme in which the number of auxiliary keys for each user and the GC are N and Nm, respectively, and the number of broadcast transmissions to remove up to w users is less than $N(m-1)$.*

Proof It suffices to show that the OR scheme described above is w-resilient. Each encryption key is derived from $\mathcal{K}\backslash\mathcal{K}(\mathcal{M})$, and users in $\mathcal{M}$ who do not possess keys that are used to encrypt $k^{\mathcal{U}\backslash\mathcal{M}}$ are unable to decrypt the broadcasted information. On the other hand, any user in $\mathcal{U}\backslash\mathcal{M}$ has at least one key to decrypt. We assume that $\mathcal{M} = \{u_{i_1}, \ldots, u_{i_l}\}$ and $l \leq w$, and hence for any user $u_i \notin \mathcal{M}$, there exists a function $f_j \in \mathcal{F}$ which separates $\{i_1, \ldots, i_l, i\}$. It follows that $k_{(f_j, f_j(i))}$ is in $\mathcal{K}(u_i) \subseteq \mathcal{K}\backslash\mathcal{K}(\mathcal{M})$, and hence u_i can decrypt the broadcasted information, obtaining a new session key.

Moreover, the number of keys for each user in $\mathcal{U}$ and the GC are $|\mathcal{K}(u_i)| = N$ and $|\mathcal{K}| = Nm$, respectively. Since the number of the broadcast ciphertexts depends on the number of encryption keys and $|\mathcal{K}(\mathcal{M})| \geq N$, we have

$$|\mathcal{K}\backslash\mathcal{K}(\mathcal{M})| \leq Nm - N = N(m-1),$$

and thus the number of the broadcast transmission is at most $N(m-1)$. □

Example 5.11 Recall a PHF$(4; 9, 3, 3)$ in $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$, described in Figure 1.1. We assume that $\mathcal{U} = \{u_1, \ldots, u_9\}$, and $\mathbf{B} = \{1, 2, 3\}$.

1. Key initialization:

   - GC generates $\mathcal{K}$ as follows:

$\mathcal{K}(u_1)$	$\mathcal{K}(u_2)$	$\mathcal{K}(u_3)$	$\mathcal{K}(u_4)$	$\mathcal{K}(u_5)$	$\mathcal{K}(u_6)$	$\mathcal{K}(u_7)$	$\mathcal{K}(u_8)$	$\mathcal{K}(u_9)$
$k_{f_1,1}$	$k_{f_1,1}$	$k_{f_1,1}$	$k_{f_1,2}$	$k_{f_1,2}$	$k_{f_1,2}$	$k_{f_1,3}$	$k_{f_1,3}$	$k_{f_1,3}$
$k_{f_2,1}$	$k_{f_2,2}$	$k_{f_2,3}$	$k_{f_2,1}$	$k_{f_2,2}$	$k_{f_2,3}$	$k_{f_2,1}$	$k_{f_2,2}$	$k_{f_2,3}$
$k_{f_3,1}$	$k_{f_3,2}$	$k_{f_3,3}$	$k_{f_3,3}$	$k_{f_3,1}$	$k_{f_3,2}$	$k_{f_3,2}$	$k_{f_3,3}$	$k_{f_3,1}$
$k_{f_4,1}$	$k_{f_4,2}$	$k_{f_4,3}$	$k_{f_4,2}$	$k_{f_4,3}$	$k_{f_4,1}$	$k_{f_4,3}$	$k_{f_4,1}$	$k_{f_4,2}$

   And then GC sends keys in $\mathcal{K}(u_i)$ to each user u_i for $1 \leq i \leq 9$.

   Thus, the GC generates and stores 12 auxiliary keys, and then sends 4 auxiliary keys to each user.

2. Broadcast: Assume that $\mathcal{M} = \{u_8, u_9\}$ is the set of users to be removed. Then

   - GC randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$
   - GC encrypts it with keys not belonging to $\mathcal{K}(\mathcal{M})$, i.e.,

     $$\mathcal{K}' := \mathcal{K}\setminus\mathcal{K}(\mathcal{M}) = \{k_{f_1,1}, k_{f_1,2}, k_{f_1,4}, k_{f_2,1}, k_{f_2,2}, k_{f_2,4}, k_{f_3,2}, k_{f_3,3}, k_{f_3,4}, k_{f_4,3}, k_{f_4,4},\}.$$

   - GC broadcasts $\{E_{k'}(k^{\mathcal{U}\setminus\mathcal{M}}) : k' \in \mathcal{K}'\}$.

3. Decryption:

   - Each user u_i in $\mathcal{U} \setminus \mathcal{M}$ uses one of his auxiliary keys $k' \in \mathcal{K}(u_i) \cap \mathcal{K}'$ to decrypt $E_{k'}(k^{\mathcal{U}\setminus\mathcal{M}})$ and obtain a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$. For example, user $u_7 \in \mathcal{U} \setminus \mathcal{M}$ has two keys, $k_{f_2,1}$ and $k_{f_4,3}$ in $\mathcal{K}'$ and he can obtain $k^{\mathcal{U}\setminus\mathcal{M}}$.

□

Next, we present the AND scheme using the perfect hash family, $\mathcal{F}$. This scheme works as follows.

1. Key initialization:
   - The same as with the OR scheme.
2. Broadcast: Assume that $\mathcal{M}$ is the set of users to be removed and assume $|\mathcal{U} \setminus \mathcal{M}| \leq w$. Let $\mathcal{U} \setminus \mathcal{M} = \{u_{i,1}, \ldots, u_{i,l}\}$ for $l \leq w$.
   - For each $u_{i,j} \in \mathcal{U} \setminus \mathcal{M}$, define

     $$k(u_{i,j}) = \bigoplus_{f \in \mathcal{F}} k_{f,f(u_{i,j})},$$

     where $\bigoplus$ is the exclusive-or (assuming all the auxiliary keys are strings with the same length.)
   - GC randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$.
   - GC encrypts it with $k(u_{i,j})$, $u_{i,j} \in \mathcal{U} \setminus \mathcal{M}$.
   - GC broadcasts $\{E_{k(u_{i,j})}(k^{\mathcal{U}\setminus\mathcal{M}}) : u_{i,j} \in \mathcal{U}\setminus\mathcal{M}\}$.
3. Decryption:
   - Each user $u_{i,j}$ in $\mathcal{U} \setminus \mathcal{M}$ computes $k(u_{i,j})$ and decrypts $E_{k(u_{i,j})}(k^{\mathcal{U}\setminus\mathcal{M}})$ and obtain the key $k^{\mathcal{U}\setminus\mathcal{M}}$.

Theorem 5.6 *If there exists a PHF$(N; n, m, w+1)$, then there exists a w-resilient re-keying AND scheme. To remove l users from $\mathcal{U}$, $n - w \leq l \leq n - 1$, each user*

has to store N auxiliary keys, and the GC has to store Nm auxiliary keys, and the number of the transmissions is $n - |\mathcal{M}| \leq w$.

Proof It suffices to show that the AND scheme described above is w-resilient. Let $\mathcal{U} \setminus \mathcal{M} = \{u_{i,1}, \ldots, u_{i,w}\}$. Each user $u_{i,j} \in \mathcal{U} \setminus \mathcal{M}$ knows the keys $k_{f,f(u_{i,j})}$ for all $f \in \mathcal{F}$, and so can compute $k(u_{i,j})$ and decrypt $E_{k(u_{i,j})}(k^{\mathcal{U}\setminus\mathcal{M}})$ to obtain $k^{\mathcal{U}\setminus\mathcal{M}}$. By the definition of $\mathcal{F}$, there exists at least a function f such that $k_{f,f(u_{i,j})}$ is unknown to the users in $\mathcal{M}$ and so they are unable to find $k(u_{i,j})$, and hence they can not decrypt the broadcasted information to obtain a new session key.
Moreover, the number of keys for each user in $\mathcal{U}$ and the GC are the same values, i.e.,$|\mathcal{K}(u_i)| = N$ and $|\mathcal{K}| = Nm$, respectively. The number of the broadcast transmissions is $n - |\mathcal{M}|$ since the number of the broadcast ciphertexts depends on the encryption keys $k(u_{i,j})$. □

Example 5.12 Let $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$ be a PHF$(4; 9, 3, 3)$ described in Figure 1.1. We assume that $\mathcal{U} = \{u_1, \ldots, u_9\}$, and $\mathbf{B} = \{1, 2, 3\}$. Then the AND scheme works as follows:

1. Key initialization:

   - The same as with the OR scheme constructed in Example 5.11. Thus the number of keys stored by the GC and each user is the same as with Example 5.11.

2. Broadcast: Assume that $\mathcal{M} = \{u_1, \ldots, u_7\}$ is the set of users to be removed, i.e., $|\mathcal{U} \setminus \mathcal{M}| = 2$.

- For each user $u_i \in \mathcal{U}\backslash\mathcal{M} = \{u_8, u_9\}$, we have

$$\begin{aligned} k(u_8) &= k_{f_{1,3}} \oplus k_{f_{2,2}} \oplus k_{f_{3,3}} \oplus k_{f_{4,1}}, \quad \text{and} \\ k(u_9) &= k_{f_{1,3}} \oplus k_{f_{2,3}} \oplus k_{f_{3,1}} \oplus k_{f_{4,2}}, \end{aligned}$$

  where $\bigoplus$ is the exclusive-or (assuming all the auxiliary keys are strings with the same length.)

- GC randomly chooses a new session key $k^{\mathcal{U}\backslash\mathcal{M}}$.
- GC encrypts it with $k(u_8)$ and $k(u_9)$.
- GC broadcasts $\{E_{k(u_{i,j})}(k^{\mathcal{U}\backslash\mathcal{M}}) : u_{i,j} \in \mathcal{U}\backslash\mathcal{M}\}$.
- GC broadcasts $\{E_{k(u_8)}(k^{\mathcal{U}\backslash\mathcal{M}}),\ E_{k(u_9)}(k^{\mathcal{U}\backslash\mathcal{M}})\}$.

3. Decryption:

- u_8 and u_9 can compute $k(u_8), k(u_9)$ and decrypt $E_{k(u_8)}(k^{\mathcal{U}\backslash\mathcal{M}})$ and $E_{k(u_9)}(k^{\mathcal{U}\backslash\mathcal{M}})$, respectively. Hence users u_8 and u_9 can obtain a new session key $k^{\mathcal{U}\backslash\mathcal{M}}$. However, if $F = \{u_5, u_6\}$ is a coalition, then they don't know $k_{f_3,3}$ for $k(u_8)$ and $k_{f_4,2}$ for $k(u_9)$. Thus they cannot obtain a new session key $k^{\mathcal{U}\backslash\mathcal{M}}$.

□

5.5.2 Second Version of a Re-Keying Scheme

From now on, we assume that the GC is dynamic, i.e., after the system initialization, each user can establish a new subgroup and the user who wants first to do it plays a role as the GC. In this situation, we need other component, a TA(trusted

authority) who initializes the system and assigns auxiliary keys to the group. After the initialization phase, the only means of communication among the group users is through a multicast channel, on which users in the group may broadcast messages that will be received by all users in the group.

There are three phases in a dynamic controller re-keying scheme.

1. Key initialization: the TA assigns auxiliary keys $\mathcal{K}(u_i)$ to each user u_i of the group.

2. Broadcast : a user broadcasts an encrypted session key which is only decryptable by a specified target group.

3. Decryption : the authorized users are able to decrypt the encrypted session key while unauthorized users are not.

Let $\mathcal{F} = \{f_1, \ldots, f_N\}$ be a PHF$(N; n, m, w+2)$.

First we present the OR re-keying scheme with dynamic controller using the perfect hash family, $\mathcal{F}$. This scheme works as follows.

1. Key initialization:

   - TA generates a group key $k^{\mathcal{U}}$.
   - TA constructs N symmetric $m \times m$ matrices, say $G^l = ({k^l}_{u,v})_{1 \le u,v \le m}$, for $1 \le l \le N$, each entry of which consists of a set of auxiliary keys $\mathcal{K}$.
   - For any $1 \le i \le n$, TA generates the set of auxiliary keys of the user u_i to be
   $$\mathcal{K}(u_i) = \bigcup_{l=1}^{N} \mathcal{K}(u_i)^l,$$
   where $\mathcal{K}(u_i)^l$ is the $f_l(i)$th row of matrix G^l for all $1 \le l \le N$.

- TA secretly sends $\mathcal{K}(u_i)$ and $k^{\mathcal{U}}$ to each u_i, for $1 \leq i \leq n$.

2. Broadcast: Suppose that a user u_i wants to establish a session key for a group $\mathcal{U}\backslash\mathcal{M}$, i.e., u_i will play a role of as the GC. We assume that $\mathcal{M}$ is the set of users to be removed and assume $|\mathcal{M}| \leq w$,

   - u_i randomly chooses a new session key $k^{\mathcal{U}\backslash\mathcal{M}}$
   - u_i encrypts it with all his auxiliary keys not belonging to $\mathcal{M}$
   - u_i broadcasts $\{E_k(k^{\mathcal{U}\backslash\mathcal{M}}) : k \in \mathcal{K}(u_i), k \notin \mathcal{K}(\mathcal{M})\}$.

3. Decryption:

   - Each user u_j in $\mathcal{U}\backslash\mathcal{M}$ uses one of his auxiliary keys $k \in \mathcal{K}(u_j)$ to decrypt $E_k(k^{\mathcal{U}\backslash\mathcal{M}})$ and obtain the key $k^{\mathcal{U}\backslash\mathcal{M}}$.

Theorem 5.7 *If there exists a PHF$(N; n, m, w+2)$, then there exists a w-resilient re-keying OR scheme with dynamic controller in which the number of auxiliary keys for each user is Nm, and the number of transmission to update a session key $k^{\mathcal{U}\backslash\mathcal{M}}$, where $|\mathcal{M}| \leq w$, is less than $m(N-1)$.*

Proof It suffices to show that the OR scheme described above is w-resilient. Suppose that u_i wants to establish a session key for a group $\mathcal{U}\setminus\mathcal{M}$, where $|\mathcal{M}| \leq w$. The encryption key k is derived from $\mathcal{K}(u_i)\backslash\mathcal{K}(\mathcal{M})$, and users in $\mathcal{M}$ who do not possess keys that are used to encrypt $k^{\mathcal{U}\backslash\mathcal{M}}$ are unable to decrypt the broadcasted information. On the other hand, any user u_j in $\mathcal{U}\backslash\mathcal{M}$ has at least one key to decrypt one of the broadcasts ciphertexts and obtain $k^{\mathcal{U}\backslash\mathcal{M}}$.
We assume that $\mathcal{M} = \{u_{i_1}, \ldots, u_{i_w}\}$ and $X = \{i_1, i_2, \ldots, i_w, i, j\}$. Since $\mathcal{F}$ is a PHF$(N; n, m, w+2)$, there exists a function $f_\alpha \in \mathcal{F}$ such that f_α separates X, that is, $k^{\alpha}{}_{f_\alpha(i),f_\alpha(j)} \in \mathcal{K}(u_i)\backslash\mathcal{K}(\mathcal{M})$. Moreover user u_i has the keys of the $f_\alpha(j)$th row of

the symmetric matrix G^α, and hence u_i and u_j share a common key which is not in $\mathcal{K}(\mathcal{U}\backslash\mathcal{M})$ since $k^\alpha{}_{f_\alpha(i),f_\alpha(j)} = k^\alpha{}_{f_\alpha(j),f_\alpha(i)}$. Thus the common key $k^\alpha{}_{f_\alpha(i),f_\alpha(j)}$ can be used by u_j to decrypt the ciphertexts in the broadcast.

Moreover, the number of auxiliary keys for each user is Nm, since each auxiliary key is from each row of N matrices. The number of broadcast ciphertexts depends on the number of encryption keys. There exists at least one function which separates the sets, whose size is at most $w+1$. Thus, we have

$$|\mathcal{K}(u_i)\backslash\mathcal{K}(\mathcal{M})| \leq Nm - m = m(N-1),$$

i.e., the number of the broadcast transmissions is at most $m(N-1)$. □

Example 5.13 Recall the PHF$(4;9,3,3)$, $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$, described in Figure 1.1. We assume that $\mathcal{U} = \{u_1, \ldots, u_9\}$, and $\mathbf{B} = \{1,2,3\}$.

1. Key initialization:

   - TA generates a group key $k^{\mathcal{U}}$.
   - TA constructs 4 symmetric 3×3 matrices as follows:

$$G^1 = \begin{pmatrix} k^1{}_{1,1} & k^1{}_{1,2} & k^1{}_{1,3} \\ k^1{}_{2,1} & k^1{}_{2,2} & k^1{}_{2,3} \\ k^1{}_{3,1} & k^1{}_{3,2} & k^1{}_{3,3} \end{pmatrix} = \begin{pmatrix} k^1{}_{1,1} & k^1{}_{1,2} & k^1{}_{1,3} \\ k^1{}_{1,2} & k^1{}_{2,2} & k^1{}_{2,3} \\ k^1{}_{1,3} & k^1{}_{2,3} & k^1{}_{3,3} \end{pmatrix}$$

$$G^2 = \begin{pmatrix} k^2{}_{1,1} & k^2{}_{1,2} & k^2{}_{1,3} \\ k^2{}_{2,1} & k^2{}_{2,2} & k^2{}_{2,3} \\ k^2{}_{3,1} & k^2{}_{3,2} & k^2{}_{3,3} \end{pmatrix} = \begin{pmatrix} k^2{}_{1,1} & k^2{}_{1,2} & k^2{}_{1,3} \\ k^2{}_{1,2} & k^2{}_{2,2} & k^2{}_{2,3} \\ k^2{}_{1,3} & k^2{}_{2,3} & k^2{}_{3,3} \end{pmatrix}$$

$$G^3 = \begin{pmatrix} {k^3}_{1,1} & {k^3}_{1,2} & {k^3}_{1,3} \\ {k^3}_{2,1} & {k^3}_{2,2} & {k^3}_{2,3} \\ {k^3}_{3,1} & {k^3}_{3,2} & {k^3}_{3,3} \end{pmatrix} = \begin{pmatrix} {k^3}_{1,1} & {k^3}_{1,2} & {k^3}_{1,3} \\ {k^3}_{2,1} & {k^3}_{2,2} & {k^3}_{2,3} \\ {k^3}_{1,3} & {k^3}_{2,3} & {k^3}_{3,3} \end{pmatrix}$$

$$G^4 = \begin{pmatrix} {k^4}_{1,1} & {k^4}_{1,2} & {k^4}_{1,3} \\ {k^4}_{2,1} & {k^4}_{2,2} & {k^4}_{2,3} \\ {k^4}_{3,1} & {k^4}_{3,2} & {k^4}_{3,3} \end{pmatrix} = \begin{pmatrix} {k^4}_{1,1} & {k^4}_{1,2} & {k^4}_{1,3} \\ {k^4}_{2,1} & {k^4}_{2,2} & {k^4}_{2,3} \\ {k^4}_{1,3} & {k^4}_{2,3} & {k^4}_{3,3} \end{pmatrix}$$

- For each user $u_i, 1 \leq i \leq 9$, TA generates $\mathcal{K}(u_i)$ as follows:

$\mathcal{K}(u_1)$	$\mathcal{K}(u_2)$	$\mathcal{K}(u_3)$
$\{{k^1}_{1,1}, {k^1}_{1,2}, {k^1}_{1,3}\}$	$\{{k^1}_{1,1}, {k^1}_{1,2}, {k^1}_{1,3}\}$	$\{{k^1}_{1,1}, {k^1}_{1,2}, {k^1}_{1,3}\}$
$\{{k^2}_{1,1}, {k^2}_{1,2}, {k^2}_{1,3}\}$	$\{{k^2}_{1,2}, {k^2}_{2,2}, {k^2}_{2,3}\}$	$\{{k^2}_{1,3}, {k^2}_{2,3}, {k^2}_{3,3}\}$
$\{{k^3}_{1,1}, {k^3}_{1,2}, {k^3}_{1,3}\}$	$\{{k^3}_{1,2}, {k^3}_{2,2}, {k^3}_{2,3}\}$	$\{{k^3}_{1,3}, {k^3}_{2,3}, {k^3}_{3,3}\}$
$\{{k^4}_{1,1}, {k^4}_{1,2}, {k^4}_{1,3}\}$	$\{{k^4}_{1,2}, {k^4}_{2,2}, {k^4}_{2,3}\}$	$\{{k^4}_{1,3}, {k^4}_{2,3}, {k^4}_{3,3}\}$

$\mathcal{K}(u_4)$	$\mathcal{K}(u_5)$	$\mathcal{K}(u_6)$
$\{{k^1}_{1,2}, {k^1}_{2,2}, {k^1}_{2,3}\}$	$\{{k^1}_{1,2}, {k^1}_{2,2}, {k^1}_{2,3}\}$	$\{{k^1}_{1,2}, {k^1}_{2,2}, {k^1}_{2,3}\}$
$\{{k^2}_{1,1}, {k^2}_{1,2}, {k^2}_{1,3}\}$	$\{{k^2}_{1,2}, {k^2}_{2,2}, {k^2}_{2,3}\}$	$\{{k^2}_{1,3}, {k^2}_{2,3}, {k^2}_{3,3}\}$
$\{{k^3}_{1,3}, {k^3}_{2,3}, {k^3}_{3,3}\}$	$\{{k^3}_{1,1}, {k^3}_{1,2}, {k^3}_{1,3}\}$	$\{{k^3}_{1,2}, {k^3}_{2,2}, {k^3}_{2,3}\}$
$\{{k^4}_{1,2}, {k^4}_{2,2}, {k^4}_{2,3}\}$	$\{{k^4}_{1,3}, {k^4}_{2,3}, {k^4}_{3,3}\}$	$\{{k^4}_{1,1}, {k^4}_{1,2}, {k^4}_{1,3}\}$

$\mathcal{K}(u_7)$	$\mathcal{K}(u_8)$	$\mathcal{K}(u_9)$
$\{{k^1}_{3,1}, {k^1}_{3,2}, {k^1}_{3,3}\}$	$\{{k^1}_{1,3}, {k^3}_{2,3}, {k^1}_{3,3}\}$	$\{{k^1}_{1,3}, {k^1}_{2,3}, {k^1}_{3,3}\}$
$\{{k^2}_{1,1}, {k^2}_{1,2}, {k^2}_{1,3}\}$	$\{{k^2}_{1,2}, {k^2}_{2,2}, {k^2}_{2,3}\}$	$\{{k^2}_{1,1}, {k^2}_{1,2}, {k^2}_{1,3}\}$
$\{{k^3}_{1,2}, {k^3}_{2,2}, {k^3}_{2,3}\}$	$\{{k^3}_{1,2}, {k^3}_{2,2}, {k^3}_{2,3}\}$	$\{{k^3}_{1,1}, {k^3}_{1,2}, {k^3}_{1,3}\}$
$\{{k^4}_{1,3}, {k^4}_{2,3}, {k^4}_{3,3}\}$	$\{{k^4}_{1,1}, {k^4}_{1,2}, {k^4}_{1,3}\}$	$\{{k^4}_{1,2}, {k^4}_{2,2}, {k^4}_{2,3}\}$

- TA sends keys in $\mathcal{K}(u_i)$ and $k^{\mathcal{U}}$ to each user u_i for $1 \leq i \leq 9$.

2. Broadcast: Assume that $\mathcal{M} = \{u_9\}$ is the set of users to be removed, and the user u_1 wants to establish a session key for a group $\mathcal{U} \setminus \mathcal{M}$.

   - u_1 randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$.
   - u_1 encrypts it with keys in $\mathcal{K}(u_1) \setminus \mathcal{K}(\mathcal{M})$;

   $$\mathcal{K}(u_1) \setminus \mathcal{K}(\mathcal{M}) = \{k^1{}_{1,1}, k^1{}_{1,2}, k^2{}_{1,1}, k^2{}_{1,2}, k^4{}_{1,1}\}.$$

   - u_1 broadcasts $\{E_k(k^{\mathcal{U}\setminus\mathcal{M}}) : k \in \mathcal{K}(u_1)\setminus\mathcal{K}(\mathcal{M})\}$.

3. Decryption:

   - Each user u_j in $\mathcal{U}\setminus\mathcal{M} = \{u_2, u_3, u_4, u_5, u_6, u_7, u_8\}$ has the key for decrypting $E_k(k^{\mathcal{U}\setminus\mathcal{M}})$ as follows:

u_2	$k^1{}_{1,1}, k^1{}_{1,2}, k^2{}_{1,2}$	u_3	$k^1{}_{1,1}, k^1{}_{1,2}$
u_4	$k^1{}_{1,2}, k^2{}_{1,1}, k^2{}_{1,2}$	u_5	$k^1{}_{1,2}, k^2{}_{1,2}$
u_6	$k^1{}_{1,2}, k^4{}_{1,1}$	u_7	$k^2{}_{1,1}, k^2{}_{1,2}$
u_8	$k^2{}_{1,2}, k^4{}_{1,1}$		

   Thus, u_9 cannot obtain a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$.

□

Next, we present the AND Scheme with dynamic controller using the perfect hash family, $\mathcal{F}$. This works as follows.

1. Key initialization:

   - The same as with the OR scheme with dynamic controller.

2. Broadcast: Assume that $\mathcal{M}$ is the set of users to be removed and assume $|\mathcal{U} \setminus \mathcal{M}| \leq w+1$. Assume that a user u_i wishes to establish a session key for the group $\mathcal{U} \setminus \mathcal{M} = \{u_{i,1}, \ldots, u_{i,l}\}$ for $l \leq w$.

   - For each $u_j \in \mathcal{U} \setminus \mathcal{M}$, $1 \leq j \leq w$, define

$$k(u_i \to u_j) = \bigoplus_{f_\alpha \in \mathcal{F}} k^\alpha{}_{f_\alpha(i), f_\alpha(j)},$$

     where $\bigoplus$ is the exclusive-or (assuming all the auxiliary keys are strings with the same length.)

   - To update a session key for the group $\mathcal{U} \setminus \mathcal{M}$, user u_i randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$
   - u_i encrypts it with his keys $k(u_i \to u_j)$ for all $u_j \in \mathcal{U} \setminus \mathcal{M}$.
   - u_i broadcasts $\{E_{k(u_i \to u_j)}(k^{\mathcal{U}\setminus\mathcal{M}}) : u_j \in \mathcal{U}\setminus\mathcal{M}\}$.

3. Decryption:

   - Each user u_j in $\mathcal{U}\setminus\mathcal{M}$ computes $k(u_i \to u_j)$ and decrypts $E_{k(u_i \to u_j)}(k^{\mathcal{U}\setminus\mathcal{M}})$ and obtain the key $k^{\mathcal{U}\setminus\mathcal{M}}$.

Theorem 5.8 *If there exists a PHF$(N; n, m, w+2)$, then there exists a w-resilient re-keying AND scheme with dynamic controller, such that the number of auxiliary keys for each user is Nm. To update a session key for group $\mathcal{U} \setminus \mathcal{M}$, the number of transmissions is $n - |\mathcal{M}| - 1 \leq w$.*

Proof It suffices to show that the AND scheme with dynamic controller described above is w-resilient. Each user $u_j \in \mathcal{U}\setminus\mathcal{M}$ knows the key $k^\alpha{}_{f_\alpha(j), f_\alpha(i)}$ since the matrix G^α is symmetric. Hence u_j can compute $k(u_i \to u_j)$ and decrypt $E_{k(u_i \to u_j)}(k^{\mathcal{U}\setminus\mathcal{M}})$

to obtain $k^{\mathcal{U}\setminus\mathcal{M}}$. Assume that w users $u_{i_1}, \ldots, u_{i_w}$ from $\mathcal{M}$ collude to find $k^{\mathcal{U}\setminus\mathcal{M}}$. They succeed only if they can calculate $k(u_i \to u_j)$ for some j. By definition of a PHF$(N; n, m, w+2)$, there exists at least a function $f_\alpha \in \mathcal{F}$ such that $k^\alpha{}_{f_\alpha(i),f_\alpha(j)}$ is unknown to the users in $\mathcal{M}$ and so they are unable to find $k(u_i \to u_j)$. So they cannot decrypt the broadcast information to obtain a new session key.
Moreover, the number of auxiliary keys for each user is Nm, since each auxiliary key is from a row of the N matrices. The number of the broadcast transmissions is $n - 1 - |\mathcal{M}| \leq w$, since the number of the broadcast ciphertexts depends on the number of the encryption keys $k(u_i \to u_j)$. □

Example 5.14 Recall the PHF$(4; 9, 3, 3)$, $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$, described in Figure 1.1. We assume that $\mathcal{U} = \{u_1, \ldots, u_9\}$, and $\mathbf{B} = \{1, 2, 3\}$. Then the AND scheme works as follows:

1. Key initialization:

   - The same as with the OR scheme constructed in the Example 5.13.

2. Broadcast: Assume that $\mathcal{M} = \{u_1, \ldots, u_7\}$ is the set of users to be removed,i.e., $|\mathcal{M}| = 7$. Suppose that u_8 wants to establish the session key for $\{u_8, u_9\}$.

   - Define

$$\begin{aligned} k(u_8 \to u_9) &= k^1_{f_1(8),f_1(9)} \oplus k^2_{f_2(8),f_2(9)} \oplus k^3_{f_3(8),f_3(9)} \oplus k^3_{f_3(8),f_3(9)} \\ &= k^1_{3,3} \oplus k^2_{3,2} \oplus k^3_{2,3} \oplus k^4_{2,1}, \end{aligned}$$

     where $\oplus$ is the exclusive-or (assuming all the auxiliary keys are strings with the same length.)

- To update a session key for the group $\mathcal{U} \setminus \mathcal{M}$, user u_8 randomly chooses a new session key $k^{\mathcal{U}\setminus\mathcal{M}}$
- u_8 encrypts it with his keys $k(u_8 \to u_9)$.
- u_8 broadcasts $E_{k(u_8\to u_9)}(k^{\mathcal{U}\setminus\mathcal{M}})$.

3. Decryption:

- User u_9 in $\mathcal{U}\setminus\mathcal{M}$ computes $k(u_8 \to u_9)$ since the matrix $G^\alpha, \alpha \in \{1,2,3,4\}$ is symmetric. And hence user u_9 can decrypt $E_{k(u_8\to u_9)}(k^{\mathcal{U}\setminus\mathcal{M}})$ and obtain the key $k^{\mathcal{U}\setminus\mathcal{M}}$.

□

Remark: From the above observation, we can compute the key storage and communication complexity of the OR scheme and the AND scheme described above as follows: Let $\mathcal{U} = \{u_1, \ldots, u_n\}$ and $\mathcal{M} \subseteq \mathcal{U}$ be the set of users to be evicted. Let w be given an integer.

- For $|\mathcal{M}| \leq w$, there exists a w-resilient OR re-keying scheme from a PHF$(N; n, m, w+1)$ such that the numbers of keys of each user and the GC are N and Nm, respectively. Moreover, the number of transmissions is at most $N(m-1)$.
- There exists a w-resilient OR re-keying scheme with dynamic controller from PHF$(N; n, m, w+2)$ in which the numbers of auxiliary keys for each user Nm, and the number of transmissions to update a session key and exclude up to w users from the group, is at most $m(N-1)$.
- For $n - w \leq |\mathcal{M}| \leq n-1$, there exists a w-resilient AND scheme from PHF$(N; n., m, w+1)$ such that the number of keys of each user and the GC

are N and Nm, respectively. And the number of transmissions is at most w.

- For $n - w - 1 \leq |\mathcal{M}| \leq n - 1$, there exists a w-resilient AND scheme from PHF$(N; n, m, w + 2)$ in which the numbers of auxiliary keys for each user Nm, and the number of transmissions to update a session key and exclude more than $n - w - 1$ users from the group, is at most w.

In Chapter 1, we observed that a PHF$(N; n, m, w)$ with N is $O(\log n)$ exists for fixed m and w.

5.6 The Traceability Scheme

Codes are used to protect copyrighted materials by providing some forms of traceability for pirated data. In [11], Boneh and Shaw first introduced the methods for assigning codewords for the purpose of fingerprinting digital data. Roughly speaking, there are "strong" versions of traceability that allow at least one member of a coalition that constructed a pirate decoder to be traced, and there are "weaker" versions of this concept which ensure that no coalition can frame a disjoint user or group of users. All these concepts can be formulated as codes having certain properties. There have been many results on these schemes, such as codes with the identifiable parent property (IPP), frameproof (FP) codes, and secure frameproof (SFP) codes. We will focus on the relationships between various notions of traceability and perfect hash families. We consider a code $\mathcal{C}$ of length l on an alphabet Q with $|Q| = q$, and we call it an (l, n, q)-code if $|\mathcal{C}| = n$. Moreover, an (l, n)-code denotes a binary code of length l and size n. Before proceeding further, we need some definitions which are used throughout this section.

Definition 5.5 *Let $\mathcal{C}$ be an (l, n, q)-code and let $w \geq 2$ be an integer.*

1. *For any subset of codewords* $\mathcal{C}_0 \subseteq \mathcal{C}$, *define the set of* descendants *of* $\mathcal{C}_0$, *denoted* $\boldsymbol{desc}(\mathcal{C}_0)$, *by*

$$\boldsymbol{desc}(\mathcal{C}_0) = \{x \in Q^l : x_i \in \{a_i : a \in \mathcal{C}_0\}, 1 \leq i \leq l\}.$$

This set consist of the l*-tuples that could be produced by a coalition holding the codewords in the set* $\mathcal{C}_0$.

2. *Let* w *be a positive integer. The* w-descendant code *of* $\mathcal{C}$, *denoted* $\boldsymbol{desc}_w(\mathcal{C})$, *is defined as follows:*

$$\boldsymbol{desc}_w(\mathcal{C}) = \bigcup_{\mathcal{C}_0 \subseteq \mathcal{C}, |\mathcal{C}_0| \leq w} \boldsymbol{desc}(\mathcal{C}_0).$$

This set consists of the l*-tuples that could be produced by some coalition of size at most* w.

3. *Suppose that* $x \in \boldsymbol{desc}_w(\mathcal{C})$. *We define the set of* suspect coalition *as follows:*

$$\boldsymbol{susp}_w(x) = \{\mathcal{C}_0 \subseteq \mathcal{C} : |\mathcal{C}_0| \leq w,\ x \in \boldsymbol{desc}(\mathcal{C}_0)\}.$$

Example 5.15 Let $\mathcal{C} = \{(112), (232), (131)\}$ be a code of length 3 and let $w = 2$. Then we can find $\mathbf{desc}_2(C_0)$ for all $C \subseteq \mathcal{C},\ |C| \leq 2$ as indicated below:

$C_0 \subseteq \mathcal{C}$ with $\lvert C_0\rvert \leq 2$	$\mathbf{desc}_2(C_0)$
$\{112\}$	$\{112\}$
$\{232\}$	$\{232\}$
$\{131\}$	$\{131\}$
$\{(112),(232)\}$	$\{(112),(232),(132),(212)\}$
$\{(112),(131)\}$	$\{(111),(131),(112),(132)\}$
$\{(232),(131)\}$	$\{(232),(131),(231),(132)\}$

And hence,

$$\mathbf{desc}_2(\mathcal{C}) = \{(112),(232),(131),(132),(212),(111),(231)\}.$$

Let $x = (232) \in \mathbf{desc}_2(\mathcal{C})$, then

$$\mathbf{susp}_2(x) = \{\{(232),(112)\},\{(232),(131)\}\}.$$

□

Definition 5.6 [30] *Let $\mathcal{C}$ be an (l,n,q)-code and let $w \geq 2$ be an integer. Let $\mathcal{C}_i \subseteq \mathcal{C}$, $i = 1,2,\ldots,t$, be all the subsets of $\mathcal{C}$ such that $\lvert\mathcal{C}_i\rvert \leq w$, where $t = \sum_{j=1}^{w} \binom{n}{j}$.*

1. *$\mathcal{C}$ is a* w-frameproof code (w-FPC) *provided that for all $x \in \boldsymbol{desc}_w(\mathcal{C})$, $x \in \boldsymbol{desc}(\mathcal{C}_i) \cap \mathcal{C}$ implies $x \in \mathcal{C}_i$.*

2. *$\mathcal{C}$ is a* w-secure frameproof code (w-SFPC) *provided that for all $x \in \boldsymbol{desc}_w(\mathcal{C})$, $x \in \boldsymbol{desc}(\mathcal{C}_i) \cap \boldsymbol{desc}(\mathcal{C}_j)$ implies that $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, where $i \neq j$.*

3. *$\mathcal{C}$ has the* w identifiable parent property (w-IPP) *provided that for all $x \in$*

desc$_w(\mathcal{C})$, *it holds that*

$$\bigcap_{\{i:x\in \boldsymbol{desc}(\mathcal{C}_i)\}} \mathcal{C}_i \neq \emptyset.$$

The meaning of the above definition is as follows:

- A code is w-*frameproof* if no coalition of size at most w can frame another user not in the coalition by producing the codeword held by that user.
- A code is w-*secure frameproof* if no coalition of size at most w can frame a disjoint coalition of size at most w by producing an l-tuple that could have been produced by the second coalition.
- A w-*IPP code*, given any fingerprint created by a coalition of size at most w, at least one of the members of the coalition can be identified.

For simplicity, we will focus on the binary codes with frameproof and secure frameproof. Let $\mathcal{C} = \{c^1, c^2, \ldots, c^n\} \subseteq \{0,1\}^l$ be a binary code where each c^i indicates a codeword. We call a binary l-tuple $x \in \{0,1\}^l \backslash \mathcal{C}$ an *unregistered word*. For a given code $\mathcal{C}$, we define the incidence matrix $\mathcal{M}(\mathcal{C})$ to be an $n \times l$ matrix, having entries from $\{0,1\}$, in which the rows are the n codewords in Γ. We consider a subset of codewords of $\mathcal{C}$, say $C_0 = \{c^{u_1}, \ldots, c^{u_d}\} \subseteq \mathcal{C}$. For $i \in \{1, 2, \ldots, l\}$, a bit position i is *undetectable* for C if

$$w_i{}^{u_1} = \cdots = w_i{}^{u_d}.$$

Let $U(C_0)$ be the set of undetectable bit positions for C_0. Then

$$F(C_0) = \{x \in \{0,1\}^l : x|_{U(C_0)} = c^{u_i}|_{U(C_0)} \text{ for all } c^{u_i} \in C_0\}$$

is called the *feasible set* of C_0. This represents the set of all possible l-tuples that could be produced by the coalition C_0 by comparing the d codewords they jointly

hold, which is clearly equivalent to **desc**(C_0).

5.6.1 Frameproof Code

Here we only focus on binary codes. In this case, if we use the notation $F(C)$, the relevant proof and illustration is more convenient. That is to say, in order to prove that Γ is w-FP, it suffices to show that $F(C) \cap \Gamma \subseteq C$ for any $C \subseteq \Gamma$.

Example 5.16 1. For any integer k, if we consider the code $\mathcal{C}$ which is obtained from a $k \times k$ identity matrix. Then this $\mathcal{C}$ is a k-FPC(k, k).

2. Let $\mathcal{C} = \{100, 010, 001, 111\}$ and let $w = 2$. Then $\mathcal{C}$ is a 2-FPC$(3, 4)$ code, since for any $C_i \subseteq \mathcal{C}$ with $|C_i| \leq 2$, we can verify that $F(C_i) \cap \mathcal{C} = C_i$ as follows:

$C_i \subseteq \mathcal{C}$ with $\lvert C_i \rvert \leq 2$	$F(C_i)$	$F(C_i) \cap \mathcal{C}$
$\{100\}$	$\{100\}$	$\{100\}$
$\{010\}$	$\{010\}$	$\{010\}$
$\{001\}$	$\{001\}$	$\{001\}$
$\{111\}$	$\{111\}$	$\{111\}$
$\{100, 010\}$	$\{100, 010, 110, 000\}$	$\{100, 010\}$
$\{100, 001\}$	$\{100, 001, 101, 000\}$	$\{100, 001\}$
$\{100, 111\}$	$\{100, 111, 101, 110\}$	$\{100, 111\}$
$\{010, 001\}$	$\{010, 001, 011, 000\}$	$\{010, 001\}$
$\{010, 111\}$	$\{010, 111, 110, 011\}$	$\{010, 111\}$
$\{001, 111\}$	$\{001, 111, 011, 101\}$	$\{001, 111\}$

□

There is a method to enlarge a frameproof code, using a perfect hash family, as follows:

Theorem 5.9 [36] *Suppose that there exists a PHF*$(N; n, m, w+1)$ *and there exists a* w*-FPC*(l, m)*. Then there exists a* w*-FPC*(Nl, n)*.*

Proof Let $\mathcal{C} = \{c^1, \ldots, c^m\}$ be a w-FPC(l, m) and let $\mathcal{F} = \{f_1, \ldots, f_N\}$ be an $(n, m, w+1)$ perfect hash family. We can construct a new code $\mathcal{C}' = \{x^1, \ldots, x^n\} \subseteq \{0,1\}^{Nl}$ as follows:

$$x^j = (c^{f_1(j)}||c^{f_2(j)}||\cdots||c^{f_N(j)}) \quad \text{for} \quad 1 \leq j \leq n.$$

Now we will show that $\mathcal{C}'$ is a w-FPC(Nl, n). Let $W \subseteq \mathcal{C}'$, $W = \{x^{i_1}, \ldots, x^{i_w}\}$. Assume that $\Gamma' \cap F(W) \neq W$. Then there exists a codeword $x^{i_{w+1}} \in (\mathcal{C}' \cap F(W)) \backslash W$, i.e., $x^{i_{w+1}} \in \mathcal{C}' \backslash W$ and

$$x^{i_{w+1}}|_{U(W)} = x^{i_j}|_{U(W)} \quad \text{for } 1 \leq j \leq w.$$

If $I = \{i_1, \ldots, i_w, i_{w+1}\}$, then since $\mathcal{F}$ is a PHF$(N; n, m, w+1)$, there exists an $f \in \mathcal{F}$ such that $f|_I$ is one-to-one. Thus we have $w+1$ different codewords $c^{f(i_j)} \in \mathcal{C}$, for $1 \leq j \leq w+1$, such that $c^{f(i_{w+1})}$ is in $F(C_0)$, where $C_0 = \{c^{f(i_j)} : 1 \leq j \leq w\}$. Thus $\mathcal{C} \cap F(C_0) \neq C_0$, which contradicts the fact that $\mathcal{C}$ is a w-frameproof code. □

Example 5.17 There exists a PHF$(7; 7, 4, 3)$ as shown in Figure 5.2. We have 2-FPC $(3, 4)$ code, $\mathcal{C}$, from Example 5.16. We obtain a 2-FPC $(21, 7)$ code, $\mathcal{C}'$, as

1	2	3	4	1	2	3
1	2	3	4	2	1	4
1	2	3	4	2	1	4
1	2	3	4	4	3	2
2	3	2	3	1	1	4
2	4	1	2	3	4	3
1	1	2	2	3	4	3

FIGURE 5.2: A PHF$(7; 7, 4, 3)$

described in Theorem 5.9. The following matrix represents the new code, $\mathcal{C}'$:

$$\begin{pmatrix}
1&0&0&0&1&0&0&0&1&1&1&1&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&1&1&1&0&1&0&1&0&0&1&1&1\\
1&0&0&0&1&0&0&0&1&1&1&0&1&0&1&0&0&0&1&1&1\\
1&0&0&0&1&0&0&0&1&1&1&1&1&1&1&0&0&1&0&1&0\\
0&1&0&0&0&1&0&1&0&0&0&1&1&0&0&1&0&0&1&1&1\\
0&1&0&1&1&1&1&0&0&0&1&0&0&0&1&1&1&1&0&0&1\\
1&0&0&1&0&0&0&1&0&0&1&0&0&0&1&1&1&1&0&0&1
\end{pmatrix}$$

□

Corollary 5.1 [36] *For any integer $j \geq 1$, there exists a 2-FPC$(6 \times 4^j, 5^{2^j})$.*

Proof There exists a PHF$(2; 5, 4, 3)$, as shown in Figure 5.1. And since $\gcd(5, \binom{3}{2}!) = 1$, by Theorem 2.17 and $\mathcal{F}$ described above, we have a PHF$(2 \times 4^j; 5^{2^j}, 4, 3)$ for any integer $j \geq 1$. Then, by Example 5.16 and Theorem 5.9, we can obtain a 2-FPC$(6 \times 4^j, 5^{2^j})$. □

Corollary 5.2 *Suppose that there exists a w-FPC(v,q) $\mathcal{C}_1$, and an (N,n,d,q)-code $\mathcal{C}_2$. Then there exists a w-FPC(vN,n), provided that $d > N\left(1-\frac{1}{\binom{w+1}{2}}\right)$.*

Proof Let $\mathcal{C}_1$ be a w-FPC (v,q) code and let $\mathcal{C}_2$ be a (N,n,d,q) code. If $d > N\left(1-\frac{1}{\binom{w+1}{2}}\right)$, then, by Theorem 2.5, we have a PHF$(N;n,q,w+1)$ from the given (N,n,d,q)-code $\mathcal{C}_2$. Thus, by Theorem 5.9, there exists a w-FPC(vN,n). □

Similarly, D. Stinson and R. Wei presented and proved the following theorem:

Theorem 5.10 [36] *If there exists a w-FPC(v,q) and an (N,n,d,q) code with minimum Hamming distance $d > N\left(1-\frac{1}{w}\right)$, then there exists a w-FPC(vN,n).*

5.6.2 Secure Frameproof Code

We also consider the case for the binary codes in this section. First we consider the following example to illustrate a (binary) secure frameproof code as follows:

Example 5.18 Let $\mathcal{C}$ be the binary $(3,4)$-code described in Example 5.16. By computing $F(C_i)$ for all C_i such that $|C_i| = 2$,we can verify that $\mathcal{C}$ is a 2-SFPC, i.e.,

$$\begin{aligned} F(\{100,010\}) \cap F(\{001,111\}) &= \emptyset, \\ F(\{100,010\}) \cap F(\{010,111\}) &= \emptyset, \\ F(\{100,111\}) \cap F(\{010,001\}) &= \emptyset. \end{aligned}$$

□

Theorem 5.11 *Any w-SFPC (l,n) is a w-FPC(l,n).*

Proof Let $\mathcal{C}$ be a w-SFPC(l,n). Suppose that $\mathcal{C}$ is not a w-FPC(l,n). Then there exists a set $C \subseteq \mathcal{C}$ and a codeword c^j such that $|C| \leq w$ and $c^j \in F(C) \backslash C$. Suppose we define $C_1 = C$ and $C_2 = \{c^j\}$. Then, we have $|C_1| \leq w$, $|C_2| \leq w$, $C_1 \cap C_2 = \emptyset$ and $F(C_1) \cap F(C_2) = \{c^j\} \neq \emptyset$. This contradicts the fact that $\mathcal{C}$ is a w-SFPC(l,n). □

Now we present the method described in Theorem 5.9 to expand the w-SFPC using a perfect hash family.

Theorem 5.12 [35] *Suppose that there exists a w-SFPC(l,m) and there exists a PHF$(N;n,m,2w)$. Then there exists a w-SFPC(lN,n).*

Proof Let $\mathcal{C} = \{c^1, \ldots, c^m\}$ be a w-SFPC(l,m) and let $\mathcal{F} = \{f_1, \ldots, f_N\}$ be a PHF$(N;n,m,2w)$. We will construct a new code $\mathcal{C}' = \{x^1, \ldots, x^n\} \subseteq \{0,1\}^{lN}$, which consists of n codewords of length lN, as follows:

$$x^j = (c^{f_1(j)} || c^{f_2(j)} || \cdots || c^{f_N(j)}) \quad \text{for} \quad 1 \leq j \leq n.$$

We will show that $\mathcal{C}'$ is a w-SFPC(lN,n). Let $C_1, C_2 \subseteq \mathcal{C}'$, $C_1 = \{x^{i_1}, \ldots, x^{i_w}\}$, $C_2 = \{x^{i_w+1}, \ldots, u^{i_{2w}}\}$ such that $C_1 \cap C_2 = \emptyset$. Assume that $F(C_1) \cap F(C_2) \neq \emptyset$. Then there exists a codeword $x^l \in F(C_1) \cap F(C_2)$, i.e., there is common undetectable position for C_1 and C_2. Since $\mathcal{F}$ is a PHF$(N;n,m,2w)$, there exists an $f \in \mathcal{F}$ such that $f|_I$ is one-to-one, where $I = \{i_1, \ldots, i_{2w}\}$. Thus $W_1 = \{c^{f(i_j)} : 1 \leq j \leq w\}$ and $W_2 = \{c^{f(i_j)} : w+1 \leq j \leq 2w\}$ are disjoint subset of $\mathcal{C}$. But $F(W_1) \cap F(W_2) \neq \emptyset$, which contradicts the fact that $\mathcal{C}$ is a w-secure frameproof code. □

Corollary 5.3 *There exists a 2-SFPC$(3 \cdot 7^{j+1}, 7^{2^j})$ for all $j \geq 0$.*

Proof We have a PHF$(7;7,4,3)$, $\mathcal{F}$, in Figure 5.2. And since $\gcd(7, \binom{3}{2}!) = 1$, by Theorem 2.17 and the perfect hash family $\mathcal{F}$, we have a PHF$(7 \times 7^j; 7^{2^j}, 4, 4)$ for any integer $j \geq 1$. And then by Example 5.18 and Theorem 5.9, we can obtain a 2-SFPC$(3 \cdot 7^{j+1}, 7^{2^j})$. □

5.6.3 Identifiable Parent Property

An (l, n, q)-code is a w-IPP code if and only if

$$\bigcap_{\mathcal{C}_0 \in \mathbf{susp}_w(c)} \mathcal{C}_0 \neq \emptyset,$$

for all $c \in \mathbf{desc}_w(\mathcal{C})$. To illustrate this definition, we consider the following example.

Example 5.19 We present a $(3, 6, 3)$ code, $\mathcal{C}$, and consider coalition of size at most 2:

$$\mathcal{C} = \{011, 101, 110, 202, 102, 210\}.$$

For $x = (111) \in \mathbf{desc}_2(\mathcal{C})$,

$$\mathbf{susp}_2(x) = \{\{011, 101\}, \{011, 110\}, \{101, 110\}, \{011, 102\}, \{101, 210\}\}.$$

Then

$$\bigcap_{\mathcal{C}_0 \in \mathbf{susp}_2(x)} \mathcal{C}_0 = \emptyset.$$

Thus this code $\mathcal{C}$ is not a 2-IPP code. □

Example 5.20 We present a $(3, 7, 5)$ 2-IPP code $\mathcal{C}$ as follows:

$$\mathcal{C} = \{000, 011, 023, 103, 204, 330, 440\}.$$

If $c = (x_1x_2x_3) \in \mathbf{desc}_w(\mathcal{C})$ and any coordinate of c is non-zero then at least one parent of c can be identified as follows:

x_1	i.p.	x_2	i.p.	x_3	i.p.
1	103	1	011	1	011
2	204	2	023	2	023
3	330	3	330	3	103
4	440	4	440	4	204

where i.p. means an identifiable parent. Finally, if $c = (0, 0, 0)$, then c_1 must be a parent. □

Now we discuss the relationships between 2-IPP codes and perfect hash families.

Theorem 5.13 [19] *Let $\mathcal{C}$ be a (N, n, q) code which is 2-IPP. Let $M(\mathcal{C})$ be an $(0, 1)$-array from a code $\mathcal{C}$, where each codeword corresponds to one of columns. Then $M(\mathcal{C})$ is a PHF$(N; n, q, 3)$.*

Proof Suppose that $M(\mathcal{C})$ is not a PHF$(N; n, q, 3)$, then there exist three columns r_1, r_2, and r_3 of $M(\mathcal{C})$ violate the PHF property. For any row c, let x_c be an element that is repeated, i.e., it occurs in at least two of the three given columns in row c. Then for $x = (x_1x_2 \cdots x_N)$, we have $\{r_1, r_2\}, \{r_1, r_3\}, \{r_2.r_3\} \in \mathbf{susp}_2(x)$. Thus $\mathcal{C}$ is not a 2-IPP code. □

For the general case, we have the following theorem, which can be proved with a similar argument to the above theorem.

Theorem 5.14 [30] *Suppose that there exists a w-IPP(N, n, q) code. Then there exists a PHF$(N; n, q, w+1)$, provided that $n \geq w+1$.*

Instead of proving the above theorem, we illustrate with an example, as follows:

Example 5.21 From a $(3, 7, 5)$ 2-IPP code $\mathcal{C}$ given in Example 5.20, we can construct a PHF$(3; 7, 5, 3)$, as described in Figure 5.3 □

0	0	0	1	2	3	4
0	1	2	0	0	3	4
0	1	3	3	4	0	0

FIGURE 5.3: A PHF$(3; 7, 5, 3)$

Now we observe a necessary condition on the existence of a w-IPP code as follows:

Theorem 5.15 [30] *Suppose $\mathcal{C}$ is any (l, n, q) code, and $n - 1 \geq w \geq q$. Then $\mathcal{C}$ is not a w-IPP code.*

At last, we will present the following theorem without proof.

Theorem 5.16 [30] *Suppose that there exists a PHF$(N; n, q, \left\lfloor \frac{(w+2)^2}{4} \right\rfloor)$. Then there exists a w-IPP (N, n, q) code.*

Bibliography

[1] N. Alon. Explicit construction of exponential sized families of k-independent sets. *Discrete Mathematics*, 58:191–193, 1986.

[2] G. Ateniese, C. Blundo, A. De Santis, and D. Stinson. Visual cryptography for general access structures. *Information and Computation*, 129:86–106, 1996.

[3] Mustafa Atici. *Hash Families: Resursive Constructions and Applications to Cryptography*. PhD thesis, University of Nebraska, 1996.

[4] M. Atici, S. S. Magliveras, D. R. Stinson, and R. Wei. Some recursive constructions for perfect hash families. *Journal of Combinatorial Designs*, 4:353–363, 1996.

[5] M. Atici, D. R. Stinson, and R. Wei. A new practical algorithm for the construction of a perfect hash function. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 35:127–145, 2000.

[6] Jurgen Bierbrauer. Maximal orthogonal latin rectangles. *Journal of Statistical Planning and Inference*, 56:39–47, 1996.

[7] S. R. Blackburn. Combinatorics and Threshold Cryptography. In Combinatorial Designs and Their Applications. CRC Press, 49–70, 1999.

[8] S. R. Blackburn. Perfect hash families: Probablistic methods and explicit constructions. *Journal of Combinatorial Theory, Series A*, 92:54–60, 2000.

[9] S. R. Blackburn. Perfect hash families with few functions. *Preprint*, 2002.

[10] S. R. Blackburn and P. R. Wild. Optimal linear perfect hash families. *Journal of Combinatorial Theory, Series A*, 83:233–250, 1998.

[11] Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. *Lecture Notes in Computer Science (CRYTO '95)*, 963:452–465, 1995.

[12] C. J. Colbourn and J. H. Dinitz. *The CRC Handbook of Combinatorial Designs*. CRC Press, 1996.

[13] Z. J. Czech, G. Havas, and B. S. Majewski. Perfect hashing. *Theoretical Computer Science*, 182:1–143, 1997.

[14] A. Fiat and M. Naor. Broadcast encryption. *Lecture Notes in Computer Science(CRYPTO '93)*, 773:480–491, 1994.

[15] Michael L. Fredman and Janos Komlos. On the size of separating systems and families of perfect hash functions. *SIAM J. Alg. Disc. Meth.*, 5:61–68, 1984.

[16] Michel L. Fredman, Janos Komlos, and Endre Szemeredi. Storing a sparse table with o(1) worst case access time. *Journal of the Association for Computing Machinery*, 31:538–544, 1984.

[17] Soren Have Hansen. *Rational Points on Curves over Finite Fields*. MATEMATISK INSTITUT, 1995.

[18] A. S. Hedayat, N. J. A. Sloane, and John Stufken. *Orthogonal Arrays.* Springer, 1999.

[19] H. D. L. Hollmann, J. H. van Lint, J-P. Linnartz, and L. M. G. M. Tolhuizen. On codes with the identifiable parent property. *Journal of Combinatorical Theory A*, 82:121–133, 1998.

[20] P. Horak, A. Rosa, and J. Siran. Maximal orthogonal latin rectangles. *Ars Combinatoria*, 47:129–145, 1997.

[21] J. Korner and K. Martin. New bounds for perfect hashing via information theory. *Europ. J. Combinatorics*, 9:523–530, 1988.

[22] Kurt Mehlhorn. *Data Structures and Algorithms 1:Sorting and Searching.* Springer-Verlag, 1984.

[23] M. Naor and A. Shamir. *Visual Cryptography (Eurocrypt'94)*, volume 950. Springer-Verlag, 1995.

[24] Harald Niederreiter and Chaoping Xing. *Rational Points on Curves over Finite Fields: Theory and Applications.* Cambridge University Press, 2001.

[25] H. J. Ryser. Combinatorial theorem with an application to latin rectangles. *American Mathematical Society, Proceedings*, 2:550–552, 1951.

[26] Rei Safavi-Naini and Huaxiong Wang. New constructions for multicast rekeying schemes using perfect hash families. *Precedings of the 7th ACM Conference on Computer and Communications Security*, pages 228–234, 2000.

[27] P. Sarkar and D. R. Stinson. Frameproof and IPP codes. *Lecture Notes in Computer Science (INDOCRYPT 2001)*, 2247:117–126, 2000.

[28] A. Shamir. How to share a secret. *Communications of the ACM*, 22:612–613, 1979.

[29] Jau-Shyong Shiue and Gary L. Mullen. A simple construction for orthogonal latin rectangles. *Journal of Combinatorial Mathemtics and Combinatorial Computing*, 9:161–166, 1991.

[30] J. N. Staddon, D. R. Stinson, and R. Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Transactions on Information Theory*, 47:1042–1049, 2001.

[31] Henning Stichtenoth. *Algebraic Function Fields and Codes*. Springer-Verlag, 1993.

[32] Douglas R. Stinson. *Cryptography : Theory and Practice*. CRC Press, 1995.

[33] Doug Stinson. *Lecture Notes: CO634*. University of Waterloo, 2002.

[34] Doug Stinson. *Lecture Notes: CS798*. University of Waterloo, 2002.

[35] D. R. Stinson, Tran van Trung, and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *Journal of Statistical Planning and Inference*, 86:595–617, 2000.

[36] D. R. Stinson and R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes. *SIAM J. Discrete Math.*, 11:41–53, 1998.

[37] D. R. Stinson, R.Wei, and L.Zhu. New constructions for perfect hash families and related structures using combinatorial designs and codes. *Journal of Combinatorial Designs*, 8:189–200, 2000.

[38] Xiaodong Sun. Explicit interpolation sests using perfect hash families. Technical Report DIMACS Technical Report 2000-28, University, 2000.

[39] M. A. Tsfasman and S. G. Vladut. *Algebraic-Geometric Codes*. Kluwer Academic Publisher, 1991.

[40] J. H. Van Lint and R. M. Wilson. *A Couse in Combinatorics*. Cambridge University Press, 1992.

[41] H. Wang and C. Xing. Explicit constructions of perfect hash families from algebraic curves over finite fields. *Journal of Combinatorial Theory, Series A*, 93:112–124, 2001.