Optimal Perfect Hash Family

Yirui He
School of Science and Engineering
the Chinese University of Hong Kong (Shenzhen)
Shenzhen, China
121090173@link.cuhk.edu.cn

***Abstract*—Perfect hash families (PHFs) are an essential tool in computer science, used for efficient data retrieval, storage, and organization. This paper explores the necessary lower bound and the sufficient lower bound for the size of a perfect hash family. Probabilistic methods like Lovász local lemma will be introduced. We finally use the discussed bounds to show the existence of the certain classes of optimal perfect hash family and provide explict construction.**

1. INTRODUCTION

Hash functions are fundamental tools in computer science, widely used for efficient data retrieval, storage, and organization. A hash function $\phi : V \to F$ maps elements from a large set V (the universe) to a smaller set F (the range), typically aiming to distribute elements uniformly across F. Hash functions are crucial in various applications, including hash tables, cryptographic protocols, and data integrity verification.

Despite their widespread use, hash functions can suffer from collisions, where two different elements in V are mapped to the same element in F. To mitigate this issue, hash families are introduced. A hash family is a set of hash functions from which one can randomly or deterministically choose a function for hashing.

Hash families have numerous applications. In compiler design, they are used to manage symbol tables efficiently. In cryptography, they provide mechanisms for secure data transmission and storage. In data structures, they support the implementation of hash tables and dictionaries, ensuring fast access and retrieval times.

Let n, a, t be positive integers. Let V be a set of size n and F be a set of size a. If a function $\phi : V \to F$ is injective when restricted to some subset X of V, we say that ϕ separates X; otherwise we say that ϕ reduces X.

Let s be a positive integer. An (n, q, t)-***perfect hash family*** *of size* s, denoted by $PHF(s; n, q, t)$, is a sequence $\phi_1, \phi_2, \cdots, \phi_s$ of functions such that for any subset $X \subseteq V$ $s.t.$ $|X| = t$, at least one of $\phi_1, \phi_2, \cdots, \phi_s$ separates X. A perfect hash family is **optimal** if s is the smallest positive integer such that $PHF(s; n, q, t)$ exists.

The concept of a perfect hash family (PHF) was introduced by Kurt Mehlhorn in the mid-1980s. Mehlhorn's work was pivotal in compiler design, where PHFs were used to prove lower bounds on the size of a program that constructs a hash function suitable for fast retrieval of fixed data, such as library function names [1]. This foundational work laid the groundwork for further developments in the field of hashing and data structures. For a comprehensive survey of this area, one can refer to the work of Czech, Havas, and Majewski [2].

Since its introduction, the theory of perfect hash families has developed significantly. Researchers like Newman and Wigderson have applied PHFs to circuit complexity problems, providing new insights and methodologies [3]. Their work demonstrated how PHFs could be used to simplify complex circuit designs by ensuring that specific subsets of inputs could be uniquely identified and processed without collisions, which is crucial in optimizing the performance of logical circuits.

Alon and Naor have constructed deterministic analogues of probabilistic algorithms using PHFs, thereby broadening the applications of PHFs in algorithm design [4]. They leveraged PHFs to create algorithms that, while deterministic, maintain the efficiency and robustness typically associated with probabilistic methods. This advancement was particularly impactful in fields such as network design and error correction, where reliability is paramount.

In the realm of cryptography, Blackburn, Burmester, Desmedt, and Wild have utilized PHFs for threshold cryptography, enhancing the security and efficiency of cryptographic protocols [5]. Their research focused on constructing PHFs that could efficiently manage keys and secrets in a distributed manner, ensuring that only authorized subsets of users could access the information. This application is vital for secure communications, multi-party computations, and distributed storage systems.

Further development in the field has been driven by Stinson, van Trung, and Wei, who leveraged PHFs and related structures like separating hash families to improve the construction of secure frameproof codes, key distribution patterns, and group testing algorithms [6]. Their work has shown how PHFs can be used to create robust coding schemes that are resistant to collusion, thereby enhancing data integrity and security in various applications.

They have also been used in the creation of cover-free families and separating systems. These structures are essential in scenarios where it is crucial to distinguish between different sets of elements based on their unique properties. For instance, in the design of optical networks, cover-free families help in ensuring that the signal paths are uniquely identifiable, preventing interference and improving overall network reliability.

Papers by Fredman and Komlós [7], Alon [8], Körner and Marton [9], Atici, Magliveras, Stinson and Wei [10], Blackburn and Wild [11], and Stinson, Wei, and Zhu [12] have all contributed to the rich combinatorial understanding of PHFs, offering various perspectives and techniques for their analysis and application. These studies have explored the limits and capabilities of PHFs, examining how different constructions

and parameters affect their performance and applicability. This collective body of work has provided a deeper theoretical foundation for PHFs, enabling more sophisticated and efficient designs in practical applications.

For example, the work of Fredman and Komlós [7] on the size of separating systems and perfect hash families has provided key insights into the minimum size requirements for PHFs, helping to optimize their construction. Similarly, Körner and Marton's research [9] on new bounds for perfect hash families and separating systems has advanced our understanding of the trade-offs between size, efficiency, and functionality.

Overall, the development of perfect hash families has been marked by a series of significant theoretical advancements and practical applications. Researchers have continually pushed the boundaries of what PHFs can achieve, exploring new areas of application and refining existing methodologies to enhance their efficiency and reliability.

A key tool in the development of PHFs and other combinatorial structures is the probabilistic method. The probabilistic method, pioneered by Erdős and others, is a non-constructive approach used to demonstrate the existence of certain structures by showing that, if elements are chosen at random, the probability that a desired structure exists is non-zero. The spirit of the probabilistic method lies in its counterintuitive yet powerful ability to prove the existence of structures without necessarily providing an explicit example. This method has been instrumental in many areas of combinatorics, number theory, and computer science.

One of the most magical and widely used tools in the probabilistic method is the Lovász Local Lemma (LLL) [13]. The LLL is a powerful probabilistic tool used in combinatorics and theoretical computer science to show the existence of combinatorial structures meeting certain criteria, particularly in scenarios where events are mostly independent, but not entirely so. It provides a way to show that certain undesirable events do not all occur simultaneously, even if they are not completely independent.

For example, in graph theory, the LLL can be used to demonstrate the existence of a proper coloring of a graph under specific constraints. In this context, each event A_i represents the improper coloring of a vertex, and the dependency graph D outlines the interactions between these events. By applying the LLL, one can show that there exists a way to color the graph such that no two adjacent vertices share the same color, even when the number of colors available is relatively small. This lemma is also instrumental in the construction of perfect hash families, error-correcting codes, and the design of randomized algorithms, providing guarantees that certain configurations exist with non-zero probability. Its flexibility and broad applicability make the Lovász Local Lemma a cornerstone in the probabilistic method toolkit.

In this paper, the Lovász Local Lemma (LLL) will also be introduced and utilized to provide sufficient conditions and necessary conditions on the existence of a $PHF(s;n,q,t)$. This paper is mainly based on Blackburn's work [14] on these conditions, which will be discussed in Section 2. In Section 3, an explicit construction of good classes of perfect hash families is given. The constructed perfect hash family is then proved to be optimal using the necessary conditions given in Section 2.

2. PROBABILISTIC METHODS

We will first prove the Lovász local lemma [13], and use it to give a lower bound as a sufficient condition for a perfect hash family to exist.

Lemma 2.1 (The local lemma). *Suppose $D=(n,E)$ is a dependency graph for events $A_1,A_2,\cdots,A_n$ and there are real numbers $x_1,x_2,\cdots,x_n$ s.t.: for any $i\in[n]$, $x_i\in[0,1)$ and*

$$\Pr(A_i)\le x_i\prod_{j\in\tau(i)}(1-x_j), \tag{1}$$

then

$$\Pr\left(\bigcap_{i=1}^{n}\overline{A_i}\right)\ge\prod_{i=1}^{n}(1-x_i)>0.$$

Proof. The probability of $\bigcap_{i=1}^{n}\overline{A_i}$ can be reached by factorize it to the product of conditional probability. But we first need the following claim.

Claim: for any $S\subseteq[n]\quad s.t.\quad i\notin S$:

$$\Pr(A_i|\bigcap_{j\in S}\overline{A_j})\le x_i\prod_{j\in\tau(i)}(1-x_j)^{-1}. \tag{2}$$

We shall use induction on the size of S to prove the claim.

Note that (2) is true when $|S|=0$. Assume (2) is true when $|S|=k-1,(k\ge1)$, then for $|S|=k$:

Let $S_1=S\cap\tau(i)$, $S_2=S\setminus S_1$.

If $S_1=\emptyset$, then A_i is independent of A_j, $\forall j\in S$. Then (2) is trivial. Otherwise, we define the notation $B(S)=\bigcap_{i\in S}\overline{A_i}$. Assume $S_1=\{j_1,j_2,\cdots,j_r\}$. Then we have the following lower bound:

$$\begin{aligned}
\Pr(B(S_2)|B(S_2)) &= \Pr\left(\bigcap_{t=1}^{r}\overline{A_{j_t}}|B(S_2)\right)\\
&=\prod_{t=1}^{r}\Pr\left(\overline{A_{j_t}}|B(S_2)\cap\bigcap_{s=1}^{t-1}\overline{A_{j_s}}\right)\\
&=\prod_{t=1}^{r}\left(1-\Pr\left(A_{j_t}|B(S_2)\cap\bigcap_{s=1}^{t-1}\overline{A_{j_s}}\right)\right)\\
&\ge\prod_{t=1}^{r}\left(1-\Pr(A_{j_t})\prod_{j\in\tau(i)}(1-x_j)^{-1}\right)\\
&\ge\prod_{t=1}^{r}(1-x_i),
\end{aligned} \tag{3}$$

where the first inequality is a result of the induction hypothesis (2), and the second inequality holds because of (1).

$$\begin{aligned}
\Pr(A_i|B(S)) &= \frac{\Pr(A_i\cap B(S_1)|B(S_2))}{\Pr(B(S_1)|B(S_2))}\\
&\le\frac{\Pr(A_i|B(S_2))}{\Pr(B(S_1)|B(S_2))}=\frac{\Pr(A_i)}{\Pr(B(S_1)|B(S_2))}\\
&\le\Pr(A_i)\prod_{t=1}^{r}(1-x_i)^{-1},\quad\text{by (3)}\\
&\le x_i,\quad\text{by (1).}
\end{aligned}$$

We have the claim prove at the second inequality, and we can use the last result to prove our original statement.

$$\begin{aligned}\Pr\left(\bigcap_{i=1}^{n}\overline{A_i}\right) &= (1-\Pr(A_i))(1-\Pr(A_2|\overline{A_1}))\cdots \\ &\quad \cdots(1-\Pr(A_n|\bigcap_{j=1}^{n-1}\overline{A_j})) \\ &\geq \prod_{i=1}^{n}(1-x_i)\end{aligned}$$

Proved. □

Using Lemma 2.1 we can easily show the most known form of Lovász local lemma.

Corollary 2.1.1 (Lovász local lemma). *Suppose $D=(n,E)$ is a dependency graph for events $A_1, A_2, \cdots, A_n$ s.t. : $\Pr(A_i) \leq p$ and $|\tau(i)| \leq d$, for any $i \in [n]$. Then there is a positive probability that none of the events occurs if*

$$ep(d+1) \leq 1 \tag{4}$$

Proof. Let $x_i = \frac{1}{d+1}$, $\forall i \in [n]$. Then

$$\begin{aligned}x_i \prod_{t\in\tau(i)}(1-x_t) &\geq \frac{1}{d+1}(1-\frac{1}{d+1})^d \\ &\geq \frac{1}{e(d+1)} \\ &\geq p \\ &\geq \Pr(A_i)\end{aligned} \tag{5}$$

We can conclude from Lemma 2.1 that $\Pr\left(\bigcap_{i=1}^{n}\overline{A_i}\right) > 0$. □

Remark. In the following we replace the condition (4) by $4pd \leq 1$, since if p, d satisfy $4pd \leq 1$ then they must satisfy $ep(d+1) \leq 1$.

The following theorem is a result of LLL. We can connect LLL to the existence of a perfect hash family if we see $Pr(A) > 0$ as the existence of some event that satisfies A.

Theorem 2.2. *A $PHF(s;n,q,t)$ exists whenever*

$$s > \frac{\ln 4(\binom{n}{t}-\binom{n-t}{t})}{\ln q^t - \ln q^t - t!\binom{q}{t}} \tag{6}$$

Proof. Let V be a set of size n and F be a set of size q.

Let $\phi_1, \cdots, \phi_s$ be functions chosen uniformly and independently at random from V to F.

For any $X \subseteq V$ *s.t.* $|X| = t$, let A_X be the event that X is reduced by all of $\phi_1, \cdots, \phi_s$. Then $\phi_1, \cdots, \phi_s$ is a perfect hash family if none of A_X occurs, *i.e.* $PHF(s;n,q,t)$ exists if $\bigcap \overline{A_X} \neq \emptyset$.

We can directly compute that

$$\begin{aligned}p &= \Pr(A_X) \\ &= (1-\frac{q(q-1)\cdots(q-t+1)}{q^t})^s \\ &= (\frac{q^t - t!\binom{q}{t}}{q^t})^s\end{aligned} \tag{7}$$

To compute $d = \sup_{X\subseteq V}\tau(X)$:

Note that $\forall X_1, X_2 \subseteq V$, A_{X_1} and A_{X_2} are independent if $X_1 \cap X_2 = \emptyset$. Hence $\tau(X) \leq \binom{n}{t} - \binom{n-t}{t}$, since there are $\binom{n}{t}$ subsets of size t of V and there are $\binom{n-t}{t}$ subsets X' *s.t.* $X \cap X' = \emptyset$ for fixed X.

Hence $d \leq \binom{n}{t} - \binom{n-t}{t}$.

Then we can conclude from Corollary 2.1.1 that $\bigcap \overline{A_X} \neq \emptyset$, and thus $PHF(s;n,q,t)$ exists. □

In the followings, we will give out two lower bounds as necessary conditions for a perfect hash family to exist. We will later use these to show our construction for a perfect hash family is optimal. A new definition will be introduced which is crucial in proving the mentioned lower bounds.

Definition 2.1. Let V be a set of size n and F be a set of size q. For any subset $R \subseteq \{\phi : V \to F\}$ and any subset $W \subseteq V$, let W_R denote the subset of W consisting of those elements $w \in W$ *s.t.* $\forall\ v \in W \setminus \{w\}$, the subset $\{w, v\}$ is separated by some $\phi \in R$.

When we want to argue that $PHF(s;n,q,t)$ does not exist under certain condition, we often need to give a counter-example and prove by contradiction. To reach contradiction, we often need to find a subset of V that cannot be separated by any $\phi \in R$. V_R, or W_R in general, then plays a very important role in finding such subset. We are interested in the size of W_R, and we have the following lemma.

Lemma 2.3. *Let $R \subseteq \{\phi : V \to F\}$ and let $W \subseteq V$. If $|W| > q^{|R|}$, then $|W_R| \leq q^{|R|} - 1$.*

Proof. Let $l = |R|$, $R = \{\phi_1, \cdots, \phi_l\}$.

Define $\sigma : W \to F^l$ by

$$\sigma(v) := (\phi_1(v), \cdots, \phi_l(v)),\ \forall\ v \in W.$$

$\forall\ v \neq w \in W_R$, $\sigma(v) \neq \sigma(w)$, by the definition of W_R. Hence $\sigma|_{W_R}$ is injective.

Since $|W| > q^l = |F^l|$, then σ is not surjective. Hence $\sigma|_{W_R}$ is not surjective.

Hence $|W_R| < |F^l| = q^l$, $|W_R| \leq q^l - 1$. □

The above lemma gives a upper bound for W_R and we can always conclude that $W_R < W$ which means W/W_R is not empty. The set W/W_R has great property in giving the counter-example we have mentioned before. It will be shown the following theorems.

Theorem 2.4. *Let $S \subseteq \{\phi : V \to F\}$ be a (n,q,t)-PHF. Let e be a positive integer. If*

$$\begin{aligned}&t = 2 \text{ and } n > q^e \text{ or} \\ &t \geq 3 \text{ and } n > (t-1)(q^e-1),\end{aligned} \tag{8}$$

then $|S| > (t-1)e$.

Proof. To show that $|S| > (t-1)e$, we only need to show that it is impossible for a (n,q,t)-PHF R to be of size $(t-1)e$.

Suppose not, *i.e.* $\exists\ R \in PHF(n,q,t)$, *s.t.* $|R| = (t-1)e$, then

Let $R = \{\phi_1, \cdots, \phi_{(t-1)e}\}$. Let $R_i = \{\phi_{(i-1)e+1}, \cdots, \phi_{ie}\}$, $i = 1, \cdots, t-1$.

Since $|V| = n > q^e$, (by (8)), then according to lemma 2.3:

$$\begin{aligned} &|V_{R_i}| \le q^e - 1, \\ &|\bigcup_{i=1}^{t-1} V_{R_i}| \le (t-1)(q^e-1) < n. \end{aligned} \tag{9}$$

Hence there exists $v_t \in V/\bigcup_{i=1}^{t-1} V_{R_i}$.

For any $i \in \{1, \cdots, t-1\}$, there exists $v_i \in V$ *s.t.* for any $\phi \in R_i$, ϕ does not separate $\{v_t, v_i\}$, since $v_t \notin V_{R_i}$.

Let $P = \{v_1, \cdots, v_{t-1}, v_t\}$. Then P cannot be separated by any function ϕ in $R = \bigcup_{i=1}^{t-1} R_i$.

Contradiction! □

Theorem 2.4 is a very strict lower bound for the size of a perfect hash family. However, under certain condition, the lower bound can be more strict. But it is a lot more difficult to prove. It will be shown in Theorem 2.6. But we shall first reveal a nice theorem of optimal perfect hash family using the above sufficient condition (Theorem 2.2) and necessary condition (Theorem 2.4).

Theorem 2.5. *Let e, t be integers such that $e \ge 2$. Let d be a real number such that $0 < d - e < 1/(t-1)$. Then, for sufficiently large q, an optimal $PHF(\lfloor q^d \rfloor, q, t)$ has size $(t-1)e+1$.*

Proof. Let $s = (t-1)e+1$, $n = \lfloor q^d \rfloor$.

Theorem 2.2 states that a $PHF(s; n, q, t)$ exists, whenever

$$s > \frac{\ln 4(\binom{n}{t} - \binom{n-t}{t})}{\ln q^t - \ln q^t - t!\binom{q}{t}} \tag{10}$$

The right hand side of the inequality tends to $d(t-1)$ with $n = \lfloor q^d \rfloor$, as $q \to \infty$.

Since $0 < d-e < 1/(t-1)$, then $d(t-1) < (t-1)e+1 = s$. Hence s satisfies the inequality (10) for sufficiently large q.

Therefore, a $PHF(s; n, q, t)$ exists.

To show that a $PHF(s; n, q, t)$ is optimal, we observe that:

Since $d > e$, then $n = \lfloor q^d \rfloor > (t-1)(q^e - 1)$ for sufficient large q.

We can they use Theorem 2.4 to conclude that any $PHF(n, q, t)$ should have size larger or equal to $s = (t-1)e+1$.

Therefore $PHF(s; n, q, t)$ is an optimal $PHF(n, q, t)$.

□

Next theorem is also necessary lower bound for the size $PHF(n, q, t)$. The lower bound it provide is larger than the lower bound in Theorem 2.4 by only 1. But is cost a lot more work and analysis to reach the bound.

Theorem 2.6. *Let $S \subseteq \{\phi : V \to F\}$ be a (n, q, t)-PHF. Let e be a positive integer. If $n > q^{e+1}/(t-1) + t(q^e - 1) + q - 1$, then $|S| > (t-1)e + 1$.*

Proof. Suppose there exists $R \subseteq \{\phi : V \to F\}$ s.t. $|R| = (t-1)e+1$ and $R \in PHF(n, q, t)$.

Assume $R = \{\phi_1, \cdots, \phi_{(t-1)e+1}\}$. We define the followings:

$$\begin{aligned} &R_0 = \{\phi_0\}; \\ &R_i = \{\phi_{(i-1)e+1}, \cdots, \phi_{ie}\},\ i = 1, \cdots, t-1. \end{aligned} \tag{11}$$

Since $|V| = n > q^e$, then $|V_{R_i}| \le q^e - 1$, $\forall\ i = 1, \cdots, t-1$.

Let $W = V/\bigcup_{i=1}^{t-1} V_{R_i}$, then we have

$$\begin{aligned} |W| &\ge n - \sum_{i=1}^{t-1} |V_{R_i}| \\ &> q^{e+1}/(t-1) + q^e + q - 2. \end{aligned} \tag{12}$$

For any $v \in W$, there exists $\phi \in R_i$, and $v_i \in V$, *s.t.* ϕ does not separate $\{v, v_i\}$.

For any $v \in W$, define $\underline{v}^i = (\phi_{(i-1)e+1}(v), \cdots, \phi_{ie}(v)) \in F^e$, $i = 1, \cdots, t-1$.

We say $v \equiv_i u$ if $\underline{v}^i = \underline{u}^i$, $\forall\ u, v \in V$.

We say $v \equiv_0 u$ if $\phi_0(u) = \phi_0(v)$, $\forall\ u, v \in W$.

Let $C_1, \cdots, C_l$ denote the equivalence classes of equivalence relation $\equiv_0$ that contains more than one element.

Let $C = \bigcup_{i=1}^{l} C_i$, $S = W/C$.

Since there are at most q equivalence classes in total, then $|S|$ = number of equivalence classes that contain only one element $\le q - l$.

Since $|W| > q$, then there exists $u, v \in W$, *s.t.*, $\phi_0(u) = \phi_0(v)$. Hence $l \ge 1$, $|S| \le q - 1$. We thus obtain a lower bound for $|C|$:

$$\begin{aligned} |C| &= |W/S| \\ &\ge |W| - |S| \\ &> q^{e+1}/(t-1) + q^e - 1. \end{aligned}$$

Let $[u]_i = \{v \in V | v \equiv_i u\}$ denote the equivalence class of the equivalence relation $\equiv_i$.

Let $\{u_j \in C | j \in J\}$ be a set of representation of the equivalence classes of $\equiv_i$.

Since $\underline{u}_j^i \in F^e$, hence $J \le |F^e| = q^e$.

Let $n_i^u = |[u]_i \cap C|$. Since $C = \biguplus_{j \in J} [u]_j^i \cap C$, then $|C| = \sum_{j \in J} |[u]_j^i \cap C| = \sum_{j \in J} n_i^{u_j}$. And we can have the following observation:

$$\begin{aligned} \sum_{u \in C} n_i^u &= \sum_{j \in J} \sum_{u \in [u]_j^i \cap C} n_i^u \\ &= \sum_{j \in J} \sum_{u \in [u]_j^i \cap C} n_i^{u_j} \\ &= \sum_{j \in J} (n_i^{u_j})^2 \\ &\ge \frac{\left(\sum_{j \in J} n_i^{u_j}\right)^2}{|J|} \\ &= \frac{|C|^2}{q^e}, \end{aligned}$$

where the first inequality is induced by the Cauchy–Schwarz inequality. Hence we have

$$\begin{aligned} \sum_{u \in C} \sum_{i=1}^{t-1} n_i^u &= \sum_{i=1}^{t-1} \sum_{u \in C} n_i^u \\ &\ge \frac{(t-1)|C|^2}{q^e}. \end{aligned}$$

As the maximal should be greater or equal to the average, hence there exists $\overline{u} \in C$, $s.t.$

$$\begin{aligned}\sum_{i=1}^{t-1} n_i^{\overline{u}} &\geq \frac{(t-1)|C|}{q^e} \\ &> (t-1)\left(\frac{q^{e+1}}{(t-1)} + q^e - 1\right) q^e \\ &> q + t + 2.\end{aligned} \tag{13}$$

Case 1: If there exists $1 \leq i \neq j \leq t-1$, $s.t.$ there exists $v \neq \overline{u}$, $s.t.$ $v \in [\overline{u}]_i \cap [\overline{u}]_j$, then,

let $u_i = \overline{u}$, $u_j = v$.

For any $k = 1, \cdots, t-1$, $k \neq i, j$, there exists $u_k \in V$, $s.t.$ ϕ does not separate $\{u_k, \overline{u} = u_i\}$, $\forall\ \phi \in V_{R_k}$.

Since $\overline{u} \in C$, then $\exists\ u_t \in V$, $s.t.$ $u_t \equiv_0 \overline{u} = u_i$, ϕ_0 does not separate $\{u_t, \overline{u} = u_i\}$.

Case 2: Otherwise, for any $1 \leq i \neq j \leq t-1$, $[\overline{u}]_i \cap [\overline{u}]_j = \{u\}$, then,

Claim: there exists $w_1 \neq w_2 \in [\overline{u}] \cap C$ for some i, $s.t.$ $w_1 \equiv_0 w_2$.

Suppose not, $i.e.$ no two elements of $\bigcup_{k=1}^{t-1} [\overline{u}]_k \cap C$ belong to the same equivalence class of the equivalence relation $\equiv_0$, then,

$$\begin{aligned}q &\geq l \\ &\geq |\bigcup_{k=1}^{t-1} [\overline{u}]_k \cap C| \\ &= |\{\overline{u}\} \cap \biguplus_{k=1}^{t-1} ([\overline{u}]_k \cap C/\{\overline{u}\})| \\ &= 1 + \sum_{i=1}^{t-1} (n_i^{\overline{u}} - 1) \\ &= \sum_{i=1}^{t-1} n_i^{\overline{u}} - t + 2 \\ &> q, \text{ by (13).}\end{aligned}$$

Thus we reach a contradiction and the claim is proved.

Let $u_i = w_1$, $u_t = w_2$. $\forall\ k = 1, \cdots, t-1$, $k \neq i$, $\exists\ u_k \in V$, $s.t.$ ϕ does not separate $\{u_k, u_i\}$, $\forall\ \phi \in V_{R_k}$.

ϕ does not separate $\{u_k, u_t\}$, $\forall\ \phi \in V_{R_i} \cup V_0$.

In either case, let $P = \{\phi_1, \cdots, \phi_t\}$. Then P cannot be separated by any function ϕ in $R = \bigcup_{i=0}^{t-1} R_i$. Proved. □

Corollary 2.6.1. *Let $S \subseteq \{\phi : V \to F\}$ be an $(n, q, 3) - PHF$. If $n > 2q^{\lceil e \rceil - 1}$ then $|S| \geq 2\lceil e \rceil - 1$ and if $n > q^{\lceil e \rceil}/2 + 2(q^{\lceil e \rceil - 1} - 1) + q - 1$, then $|S| \geq 2\lceil e \rceil$.*

Proof. The two bounds can be easily verified by letting $t = 3$ and $e = \lceil d \rceil - 1$ in Theorem 2.4 and Theorem 2.6, respectively. □

3. EXPLICIT CONSTRUCTION

This section provides a simple construction of an optimal perfect hash family when $t = 3$.

Let r be a fixed integer such that $r \geq 2$. We may construct an optimal $PHF(3; r^3, r^2, 3)$ as follows. Let $V = \mathbb{F}_r^3$ and $F = \mathbb{F}_r^2$. Define functions $\phi_1, \phi_2, \phi_3 : V \to F$ by

$$\begin{aligned}\phi_1((a,b,c)) &= (a,b), \\ \phi_2((a,b,c)) &= (b,c), \\ \phi_3((a,b,c)) &= (a,c),\end{aligned}$$

for all $(a, b, c) \in V$.

Theorem 3.1. *The set $S = \{\phi_1, \phi_2, \phi_3\}$ defined above is an optimal $PHF(3; r^3, r^2, 3)$.*

Proof. We first need to check whether S is a perfect hash family with respect to $(r^3, r^2, 3)$.

Suppose not, $i.e.$, $\exists\ l = \{x_1, x_2, x_3\}$, $s.t.$, $x_i \neq x_j$, $1 \leq i < j \leq 3$, and none of the functions in S separates S.

Assume $x_i = (a_i, b_i, c_i)$, $i = 1, 2, 3$.

Since ϕ_1 does not separate l, then, without loss of generality, assume $\phi_1(x_1) = \phi_1(x_2)$, $i.e.$:

$$a_1 = a_2,\ b_1 = b_2.$$

In this case, ϕ_2 and ϕ_3 must separate x_1 and x_2; otherwise, $c_1 = c_2$ and thus $x_1 = x_2$, which is a contradiction.

Since ϕ_2 does not separate l, then, without loss of generality, assume $\phi_2(x1) = \phi_2(x_3)$, then:

$$b_2 = b_1 = b_3,\ c_1 = c_3.$$

Since ϕ_3 does not separate l, then, without loss of generality, assume $\phi_3(x2) = \phi_3(x1)$, then:

$$a_2 = a_3,\ c_2 = c_3,$$

thus $x_2 = x_3$, while x_2 should be equal to x_3 by its construction.

By this contradiction, we conclude that S is indeed a $PHF(r^3, r^2, 3)$.

Corollary 2.6.1 states that a $PHF(s; r^3, r^2, 3)$ has the property that $s \geq 3$ provided $r \geq 2$. Hence $S = \{\phi_1, \phi_2, \phi_3\}$ is indeed an optimal $PHF(3; r^3, r^2, 3)$.

□

We may also construction the following $PHF(6; q^2, q, t)$ for all prime number p satisfying $p = 11$ or $p \geq 17$.

Let $F = \mathbb{F}_p$ and $V = \mathbb{F}_p^2$. Define $\phi_1, \cdots, \phi_6 : V \to F$ by:

$$\begin{aligned}\phi_1((a,b)) &= a, \\ \phi_2((a,b)) &= b, \\ \phi_3((a,b)) &= b - a, \\ \phi_4((a,b)) &= b - 2a, \\ \phi_5((a,b)) &= b - 3a, \\ \phi_6((a,b)) &= b - 5a,\end{aligned}$$

for all $a, b \in F$.

Theorem 3.2. *The functions ϕ_i defined above form a $PHF(6; p^2, p, 4)$ for all prime numbers p such that $p = 11$ or $p \geq 17$.*

Proof. To show that $\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6\}$ form a $PHF(6; p^2, p, 4)$, we need to prove that for any 4-subset of V, at least one of these functions separates the subset.

We shall first introduce the concept of gradients: The gradients (slopes) of the lines through pairs of points (a_i, b_i) are given by:

$$\text{Gradient} = \frac{b_j - b_i}{a_j - a_i} \quad \text{for} \quad i \neq j.$$

A set of 4 points (a_i, b_i) in the plane $\mathbb{F}_p^2$ can be represented by the gradients of the lines passing through each pair of points. There are $\binom{4}{2} = 6$ pairs, and hence 6 possible gradients.

The set of gradients must cover at most 6 distinct values in $\mathbb{F}_p \cup \{\infty\}$. These gradients correspond to the lines determined by the pairs of points.

A 4-set of points is reduced by all of $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6$ if and only if the set of gradients is $\{\infty, 0, 1, 2, 3, 5\}$. If 5 of these gradients are specified, the sixth one is uniquely determined.

For $p = 11$ or $p \geq 17$, no set of 4 points in $\mathbb{F}_p^2$ can have gradients covering all 6 values simultaneously. This implies that in any subset of size 4, at least one pair of points will have a gradient different from $\{\infty, 0, 1, 2, 3, 5\}$.

Each function ϕ_i maps (a, b) to a, b, or a linear combination thereof. If $\phi_i((a, b)) = \phi_i((c, d))$ for some points, then these points are not separated by ϕ_i. By the property of gradients, at least one function ϕ_i will separate any subset of size 4.

Thus, it is established that for any prime $p = 11$ or $p \geq 17$, the functions $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6$ indeed form a perfect hash family $PHF(6; p^2, p, 4)$. This completes the proof of Theorem 3.2. □

Let $e = 1$, Theorem 2.6 states that for a $PHF(s; q, n, t)$, if

$$n > q^{e+1}/(t-1) + t(q^e - 1) + q - 1$$

, then

$$s > (t-1)e + 1.$$

In this case, $n = q^2$ satisfies the above inequality, hence our construction of $PHF(6; q^2, q, t)$ is indeed an optimal $PHF(q^2, q, t)$ for prime number $p = 11$ or $p \geq 17$.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Kenneth Shum for his guidance throughout this semester. His tireless suggestions and support whenever I had questions have been invaluable.

REFERENCES

[1] K. Mehlhorn, "Data structures and algorithms. 1. sorting and searching," *Springer*, 1984.
[2] P. Czech, G. Havas, and B. Majewski, *Perfect Hashing*. Elsevier, 1997.
[3] I. Newman and A. Wigderson, "On the complexity of decodable error-correcting codes and private information retrieval," *IEEE Transactions on Information Theory*, 1996.
[4] N. Alon and M. Naor, "Derandomization, witnesses, and the chromatic number," *Journal of Combinatorial Theory, Series A*, 1996.
[5] S. R. Blackburn, M. Burmester, Y. Desmedt, and J. Wild, "Efficient multiparty computations secure against dishonest minority," in *Springer*, 1996.
[6] D. R. Stinson, N. van Trung, and R. Wei, "Secure frameproof codes, key distribution patterns, and cover-free families," *Journal of Cryptology*, 2000.
[7] M. L. Fredman and J. Komlós, "On the size of separating systems and perfect hash families," *SIAM Journal on Algebraic and Discrete Methods*, 1984.
[8] N. Alon, "Probabilistic methods in combinatorial optimization," *SIAM Journal on Discrete Mathematics*, 1992.
[9] J. Körner and K. Marton, "New bounds for perfect hash families and separating systems," *Journal of Combinatorial Theory, Series A*, 1988.
[10] A. Atici, S. S. Magliveras, D. R. Stinson, and R. Wei, "On perfect hash families and group testing," *Designs, Codes and Cryptography*, 1995.
[11] S. R. Blackburn and J. Wild, "Perfect hash families and bipartite ramanujan graphs," *Journal of Combinatorial Theory, Series A*, 1999.
[12] D. R. Stinson, R. Wei, and L. Zhu, "Combinatorial properties of perfect hash families," *Discrete Mathematics*, 1997.
[13] P. Erdős and L. Lovász, "Problems and results on 3-chromatic hypergraphs and some related questions," in *Infinite and Finite Sets*, ser. Colloquium Math. Soc. Janos Bolyai, A. Hajnal, R. Rado, and V. Sós, Eds. Amsterdam: North-Holland, 1975, vol. 11, pp. 609–627.
[14] S. R. Blackburn, "Perfect hash families: Probabilistic methods and explicit constructions," *Journal of Combinatorial Theory, Series A*, 2000.