

Data Retrieval and Extremal Set

M. Atici

*Department of Computer Science, Western Kentucky University,
Bowling Green, KY 42101,
E-mail: mustafa.atici@wku.edu*

Abstract

A hash table, or hash function (map), is a data structure that associates keys values with data. The main goal of hash table is to support fast data retrieval. That is when a key value is given to retrieve the corresponding data. In this study we will present a construction of hash functions from extremal set structure.

AMS Subject Classification:

Keywords:

1. Introduction to Perfect Hash Family and Extremal Set

1.1. Perfect Hash Family

Let T be a hash table with size m . Let W be a set of key values, where $|W| \leq m$ and U be a universal set, that is, set of all possible key values. Let f be a hash function from U to T . Then for any data the corresponding key value k is stored in slot $f(k)$ in T . It is said to be that data with key value k hashes to slot $f(k)$. $f(k)$ is hash value of key k .

It is very difficult to construct a hash function f for any possible key set W such that f is always one-to-one on W . That is, f is collision free. Therefore the idea of perfect hash family was introduced [1–7].

An (n, m, w) -perfect hash family is a set of functions \mathcal{F} such that

$$f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, m\}$$

for each $f \in \mathcal{F}$, and for any $W \subseteq \{1, 2, \dots, n\}$ such that $|W| = w$, there exist at least one $f \in \mathcal{F}$ such that $f|_W$ is one-to-one. We will use the notation $\text{PHF}(N; n, m, w)$ for (n, m, w) -perfect hash family with $|\mathcal{F}| = N$. Also, we will let $N(n, m, w)$ denote the minimum values N such that a $\text{PHF}(N; n, m, w)$ exists.

The motivation for the terminology “perfect hash family” is that we have a family of hash functions with the property that if at most w elements are to be hashed, then at least one function in the family yields no collisions when applied to the given w inputs. PHS’s can be constructed from combinatorial objects such as balanced incomplete block design and error correcting codes [8] and two recursive construction are given in [9].

We will typically depict a $PHF(N; n, m, w)$ in the form of an $N \times n$ array of m symbols, where each row of the array corresponds to one of the functions in the family. This array has the property that, for any subset of w columns, there exists at least one row such that the entries in the w given columns of that row are distinct.

Example 1.1. Perfect hash family \mathcal{F} , that is $PHF(3; 9, 4, 3)$ can be given:

1	2	3	4	5	6	7	8	9
1	2	3	4	3	3	1	2	4
1	1	2	1	3	4	2	4	3
2	3	3	4	2	4	1	1	1

Now, for any key set $W \subset \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ with $|W| = 3$, there will be a row such that the corresponding entries for W in that row are all distinct. For example $W = \{1, 4, 9\}$. The last row is one-to-one on W .

1.2. Extremal Set

Let the set $[n] = \{1, 2, \dots, n\}$ be given. Find the minimum cardinality of a collection \mathcal{S} of subsets of $[n]$ such that, for any two distinct elements $x, y \in [n]$, there exist disjoint subsets $A_x, A_y \in \mathcal{S}$ such that $x \in A_x, y \in A_y$, and $A_x \cap A_y = \phi$.

The extremal set problem was originally motivated by a problem in graph theory [10, 11]. In simple graph G , the *distance* between two vertices u, v of G is the least number of edges in a path joining u and v . Any such shortest path is called a *geodesic*. A set I of vertices of G is called *geodetic set* if the union of all the geodesics joining pairs of points of I is the whole graph G . Let $g(G)$ denote the minimum number of vertices in a geodetic set for G , and call $g(G)$ the *geodetic number* of G . Extremal set structure is used in [10] to give a lower bound on $g(G)$ of any given graph G .

Using Theorem 1.2 in [10] we can state the following theorem.

Theorem 1.2. For any given integer n , there exists an integer m such that

$$2 \times 3^{m-1} < n \leq 3^m$$

or

$$3^m < n \leq 4 \times 3^{m-1}$$

or

$$4 \times 3^{m-1} < n \leq 2 \times 3^m.$$

Example 1.3. Let $n = 8$. That is, $[n] = \{1, 2, 3, 4, 5, 6, 7, 8\}$. We can construct the extremal set \mathcal{S} as follows:

$\mathcal{S} = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{1, 2, 5, 6\}, \{3, 4, 7, 8\}, \{1, 3, 5, 7\}, \{2, 4, 6, 8\}\}$.

Now when we pick any two different numbers, say x, y from the set $[n] = \{1, 2, 3, 4, 5, 6, 7, 8\}$, there will be two disjoint set $A, B \in \mathcal{S}$ such that $x \in A$ and $y \in B$ and $A \cap B = \emptyset$. The inimality of $|\mathcal{S}|$ is shown in [10].

2. Algorithm to Construct Extremal Set

In this section we give the algorithm to compute the extremal set \mathcal{S} for any given integer n and corresponding set $[n] = \{1, 2, \dots, n\}$.

Extremal set ($[n]$)

1. Using Theorem 1.2 determine m .
2. Set N to 3^m or $4 \cdot 3^{m-1}$ or $2 \cdot 3^m$.
That is, $L = q_1 q_2 \cdots q_k$, where q_i is 2 or 3 for $i = 1, 2, \dots, k$.
3. Construct a matrix $M = (m_{ij})_{L \times k}$. Each row of M is $(\alpha_1, \alpha_2, \dots, \alpha_k)$ where $1 \leq \alpha_i \leq q_i$.
4. Define $A_{ij} = \{r : m_{ri} = j, 1 \leq i \leq k, 1 \leq j \leq q_i, 1 \leq r \leq L\}$. That is, A_{ij} is a set of rows such that the i th columns are j and $\mathcal{S}(L) = \{A_{ij} | 1 \leq j \leq q_i, 1 \leq i \leq k\}$.
5. Define $\mathcal{S} = \{B | B = A - \{x : n + 1 \leq x \leq L\} \text{ where } A \in \mathcal{S}(L)\}$.

Example 2.1. Let $n = 17$. Then $L = 2 \cdot 3^2 = 18$. Hence $L = 2 \times 3 \times 3$ (Here $q_1 = 2, q_2 = 3$, and $q_3 = 3$). Therefore the matrix M is:

	1	2	3	
1	1	1	1	α_1
2	1	1	2	α_2
3	1	1	3	α_3
4	1	2	1	α_4
5	1	2	2	α_5
6	1	2	3	α_6
7	1	3	1	α_7
8	1	3	2	α_8
9	1	3	3	α_9
10	2	1	1	α_{10}
11	2	1	2	α_{11}

	1	2	3	
12	2	1	3	α_{12}
13	2	2	1	α_{13}
14	2	2	2	α_{14}
15	2	2	3	α_{15}
16	2	3	1	α_{16}
17	2	3	2	α_{17}
18	2	3	3	α_{18}

Hence $\mathcal{S}(18) = \{A_{ij} | 1 \leq j \leq q_i, 1 \leq i \leq k\}$, where the A_{ij} 's are:

$$\begin{aligned}
A_{11} &= \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \\
A_{12} &= \{10, 11, 12, 13, 14, 15, 16, 17, 18\}, \\
A_{21} &= \{1, 2, 3, 10, 11, 12\}, \\
A_{22} &= \{4, 5, 6, 13, 14, 15\}, \\
A_{23} &= \{7, 8, 9, 16, 17, 18\}, \\
A_{31} &= \{1, 4, 7, 10, 14, 16\}, \\
A_{32} &= \{2, 5, 8, 11, 15, 17\}, \\
A_{33} &= \{3, 6, 9, 12, 16, 18\}.
\end{aligned}$$

Therefore $\mathcal{S} = \{B_{ij}\}$, where the B_{ij} 's are:

$$\begin{aligned}
B_1 &= \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \\
B_2 &= \{10, 11, 12, 13, 14, 15, 16, 17\}, \\
B_3 &= \{1, 2, 3, 10, 11, 12\}, \\
B_4 &= \{4, 5, 6, 13, 14, 16\}, \\
B_5 &= \{7, 8, 9, 16, 17\}, \\
B_6 &= \{1, 4, 7, 10, 14, 16\}, \\
B_7 &= \{2, 5, 8, 11, 15, 17\}, \\
B_8 &= \{3, 6, 9, 12, 16\}.
\end{aligned}$$

Theorem 2.2. The above algorithm computes the extremal set \mathcal{S} for any integer n and corresponding set $[n] = \{1, 2, \dots, n\}$.

Proof. The minimality of such $\mathcal{S}(L)$ is coming from Theorem 1.2 in [10]. We need to show that \mathcal{S} that is computed in the above algorithm is indeed an extremal set. Let x, y be two different number in $[n]$. Then there exist two α_i, α_j that corresponds to x , and y respectively. Hence x will be in some $A_{i_1 j_1}$ and y will be in some other $A_{i_2 j_2}$ such that $A_{i_1 j_1} \cap A_{i_2 j_2} = \emptyset$. ■

3. Construction of PHF from Extremal Set

In this section we will give the construction of perfect hash family of $(n, 3, 2)$. We first state the following theorem from [10].

Theorem 3.1. Let $h(n)$ denote the minimum cardinality of extremal set $S \subset 2^{[n]}$. Then

$$h(n) = \begin{cases} 3m & \text{if } 2 \times 3^{m-1} < n \leq 3^m, \\ 3m + 1 & \text{if } 3^m < n \leq 4 \times 3^{m-1}, \\ 3m + 2 & \text{if } 4 \times 3^{m-1} < n \leq 2 \times 3^m. \end{cases}$$

Theorem 3.2. Let n be a given integer. Then there exist PHF($m : n, 3, 2$) if $2 \times 3^{m-1} < n \leq 3^m$ or PHF($m + 1 : n, 3, 2$) if $3^m < n \leq 4 \times 3^{m-1}$ or $4 \times 3^{m-1} < n \leq 2 \times 3^m$.

Proof. We first construct the matrix M in the algorithm **Extremal Set** ($[n]$). Then take transpose of that matrix. Finally chop off the column(s) if $n < N$. ■

Example 3.3. Let us construct the perfect hash family PHF($2 : 9, 3, 2$). First $9 = 3^2$. Hence $m = 2$ and the matrix M is as follows:

	1	2
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

The transpose of the matrix M is PHF($2 : 9, 3, 2$).

	1	2	3	4	5	6	7	8	9
1	1	1	1	2	2	2	3	3	3
2	1	2	3	1	2	3	1	2	3

References

- [1] N. Alon and M. Naor. Derandomization, Witnesses for Boolean Matrix Multiplication and Constructions of Perfect Hash Functions, 1994.
- [2] M. Atici, S. S. Magliveras, D. R. Stinson, and W. D. Wei. Some Recursive Constructions for Perfect Hash Families, *Journal of Combinatorial Design*, 4(5):353–363, 1995.
- [3] M. L. Fredman and J. Komlos. On the Size of Separating System and Families of Perfect Hash Functions, *SIAM J. Alg. Disc. Meth.*, 5(1):61–68, 1984.
- [4] M. Korner and K. Marton. New Bounds for Perfect Hashing via Information Theory, *Europ. J. Comp.*, 9:523–530, 1988.

- [5] K. Mehlhorn. On the Program Size of Perfect and Universal Hash Functions, *Proc. 23rd Annual IEEE Symposium on Foundations of Computer Science*, pp. 170–175, 1982.
- [6] K. Melhorn. *Data Structures and Algorithms*, Vol. 1, Springer-Verlag, 1984.
- [7] H. C. Thomas, E. L. Charles, L. R. Ronalt, and S. Clifford. *Introduction to Algorithms*, Second Edition, MIT Press and McGraw Hill, 2001.
- [8] E. F. Brickell. A Problem in Broadcast Encryption, Fifth Vermont Summer Workshop on Combinatorics and Graph Theory, 1991.
- [9] M. Atici. PhD Disertation, 1996.
- [10] M. Atici and A. Vince. Geodetics in Graphs, an Extremal Set Problem, and Perfect Hash Families, *Graphs and Combinatorics*, 18:403–413, 2002.
- [11] M. Atici. On the Edge Geodetic Number of a Graph, *Intern. J. Computer Math.*, 80(7):853–861, 2003.