TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 18, Số 9 (2021): 1638-1648
ISSN: 2734-9918
HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 18, No. 9 (2021): 1638-1648
Website: *http://journal.hcmue.edu.vn*

Research Article

THE SMALLEST BASE OF $k-$SETS

***Le Phuc Lu*[1*], *Nguyen Dinh Song An*[2]**

[1]*University of Science, Vietnam National University Ho Chi Minh City, Vietnam*

[2]*Saint John's University, United States of America*

[*]*Corresponding author: Le Phuc Lu – Email: lephuclu@gmail.com*

Received: March 21, 2021; Revised: September 16, 2021; Accepted: September 20, 2021

ABSTRACT

In information theory, such as storage model, private sharing, or encryption sometimes we want to distribute a given database into many small parts, each of which is stored by a party in such a way that when there is a cooperation of a sufficient number of parties, we are able to recover the original information. For this purpose, this paper describes the way to work on a given finite set then construct a family of uniform subsets such that there exists only one permutation that maps one-to-one each subset. Of course, the optimality of construction will be considered through its size. By evaluating the number of occurrences of each element in the subsets, it is possible to establish the lower bound for that size and using the simple undirected graph to model. The construction step is only successful with relevant data and the general case is under further study.

Keywords: base; fixed point; graph; reconstruct permutation

1. Introduction

1.1. The related works

The reconstruction of permutation based on the given information was introduced by Rebecca Smith (2006) at a combinatorics conference. Bui et al. (2004) also discussed this in their book that "*There are a group of* 8 *people who stored an important document in the locked box. They required at least* 5 *people to unlock the box. What is the minimum number of locks and keys to satisfy that requirement?"*

The same question can be asked by replacing 8,5 by any positive integers n,k respectively.

These problems related to cryptography, information security, and the verified solution of the minimum number of keys is C_n^k. This is just one among many problems in this scope and the ways to prove the lower bound, also the construction step is quite hard,

Cite this article as: Le Phuc Lu, & Nguyen Dinh Song An (2021). The smallest base of k-sets. *Ho Chi Minh City University of Education Journal of Science, 18*(9), 1638-1648.

requires some estimation on the number of elements in the family of subsets. Before introducing the main problem, we discuss some basic theories.

1.2. Problem definition

For positive integer n, let $[n]=\{1,2,\ldots,n\}$ be the collection of all first n positive integers and k is some integer such that $1\le k\le n$. Define $k-$set as the subset of k elements of $[n]$. There are some studies that worked on the family of $k-$sets, such as Sean et al. (2016) or Diatonics et al (2008). Besides, the problems of finding the smallest size of the base of $k-$sets, of the symmetric group $Sym(n)$ play an important role in not only coding theory, but also computer science and other science scopes.

Namely, we need to find the minimum size of the family of $k-$sets such that there exists only one identity permutation acting on each $k-$set of that family and fixes them (the identity permutation maps each element to itself), to construct such family of $k-$set satisfying that condition.

For example, consider the special case $3-$sets with $n=5$, there are 10 subsets of size 3:

$$\{1,2,3\};\{1,2,4\};\{1,2,5\};\{1,3,4\};\{1,3,5\};\{1,4,5\};\{2,3,4\};\{2,3,5\},\{2,4,5\};\{3,4,5\}.$$

Consider some permutations of $S=[5]$; for example, $\sigma=(2,3,1),(4),(5)$ which maps

$1\to 2$, $2\to 3$, $3\to 1$, $4\to 4$, and $5\to 5$.

The permutation σ changes these $3-$sets, such as σ maps $\{1,2,5\}\to\{2,3,5\}$. Hence, σ will take some shuffle on these $3-$sets. On the other hand, not all $3-$set change, and we can see that σ maps $\{1,2,3\}$ to $\{2,3,1\}=\{1,2,3\}$ since in the set, the order is not important. In this work, we call $\{1,2,3\}$ as fixed point, and we formally define them as follows:

***Definition 1*. (fixed point).** Let n and k are positive integers such that $1\le k\le n$ and some permutation $\sigma\in Sym(n)$. The $k-$set A is a fixed point of σ if and only if σ is the bijection from A to A, namely $\sigma(A)=A$.

***Definition 2.* (base sets)**. Let n and k are positive integers such that $1\le k\le n$ and the collection of all $k-$sets S will be the base of $Sym(n)$ if and only if there is only the identity permutation fixing all of $k-$sets in S. Denote that base as $S=S(n,k)$.

In summary, we will try to answer the question: "*for the given n,k what is the smallest size of the base $S(n,k)$ such that there exists only one permutation (identity one) that fixes all of the subsets in the base? Construct some base like that.*"

This problem can be applied in the construction (erasure) combinatorial batch code mentioned before by Paterson et al. (2009), by Jung et al. (2018), or the distributed storage by Ishai et al. (2004).

2. Main results

2.1. Special case of the problem and some properties

Consider the problem in the small size with $n=5$ and $k=3$. We investigate some properties of the subset in the family satisfying the condition mentioned above.

Property 3. For all pairs $a,b\in\{1,2,3,4,5\}$, there exists a 3-set in the base S that has exactly a or b.

Proof. Indeed, suppose all 3-sets in S consist of both a,b or contain neither of them. Because the roles of a and b are equal in all 3-sets of S, therefore, the permutation (a,b) is different from the identity one since $a\neq b$, and it fixes all the subsets in S, contradiction.

Property 4. If in the base S, there exist two 3-sets that share exactly one element then that element must be fixed in all permutations that fix S.

Proof. Suppose that we have $\{a_1,b_1,c\}\in S$ and $\{a_2,b_2,c\}\in S$. Consider permutation σ that fixes all 3-sets of S. Then $\sigma(c)\in\{a_1,b_1,c\}$ and $\sigma(c)\in\{a_2,b_2,c\}$ which lead to $\sigma(c)\in\{a_1,b_1,c\}\cap\{a_2,b_2,c\}$. Therefore $\sigma(c)=c$.

2.2. Detailed solution

From the above simple example, we will prove the size of the base $S(5,3)=3$.

Choosing

$S=\big\{\{1,2,3\},\{1,2,4\},\{1,4,5\}\big\}$.

We will prove S is a base.

Consider some permutation σ that fixes all 3-sets above. According to **Property 4,** because $\{1,2,3\},\{1,4,5\}$ share same element 1, so $\sigma(1)=1$. Next, $\{1,2,3\}$ and $\{1,2,4\}$ share two elements $1,2$ so we have $\sigma(2)=2$. We also see that $\{1,2,4\}$ and $\{1,4,5\}$ share $1,4$ so $\sigma(4)=4$. From here considering set $\{1,2,3\}$, we see that $\sigma(3)=3$, therefore $\sigma(5)=5$. So S is base. Next, we suppose S is a base but $|S|\leq 2$.

If $|S|=1$, then there exists only one $3-$set, but there are six permutations fixing it (6 is the amount of permutation of 3 elements) so S is not the base.

If $|S|=2$, without the loss of generality, suppose that $\{1,2,3\}\in S$. Then the roles of each number in the following pairs $(1,2),(2,3),(3,1),(4,5)$ are equal and currently not satisfied with **Property 3.** Thus, there must exist a set that has exactly 1 of 2 elements from each of those pairs. Firstly, there exists a set A such that $1\in A$ and $2\notin A$ (similarly if $1\notin A$ and $2\in A$). We consider two following cases:

- If $3 \notin A$ then $A = \{1,4,5\}$. With pair $(2,3)$ and $(4,5)$, there must be at least one more set. Therefore $|S| > 2$ which contradicts the hypothesis that $|S| = 2$.
- If $3 \in A$ then $A = \{1,3,4\}$ or $A = \{1,3,5\}$. Then $(1,3)$ still have an equal role, implying that there must be one more 3-set. This also contradicts the hypothesis $|S| = 2$.

Hence, we get the smallest size of $S(5,3)$ is 3.

2.3. Remarks

We can see that there is more than one way to choose $S(5,3)$. For example besides the above chosen S, we could choose $S' = \{\{1,2,3\},\{1,2,4\},\{2,3,4\}\}$ to have a different base S'. We consider another small value of n.

- $n = 3$, there is not any base S because there exists only one set $\{1,2,3\}$, which contradicts **Property 3**.
- $n = 4$, we can prove that min $|S| = 3$ similarly to **2.2.**
- $n = 6$, min $|S| = 3$ with an instant of $S(6,3)$: $S = \{\{1,2,3\},\{1,2,4\},\{1,4,5\}\}$.

We already proved that these sets fix all five elements in $\{1,2,3,4,5\}$ so even if there is another element 6, we still have $f(6) = 6$. Notice that if we choose below 3-sets

$S' = \{\{1,2,3\},\{1,2,4\},\{2,3,4\}\}$,

then it does not work because 5 and 6 are not present.

2.4. The general problem and lower boundary

Now we consider the general problem, in which 5 elements will become n elements, and $3-$sets become $k-$sets. Denote $S = S(n,k)$ as the base set that satisfies the given condition. Then there will have some similar observations as in case $k = 3$:

- For every two elements $a,b \in [n]$, there exist some subsets in S with exactly one of them (similarly to **Property 3**).
- There is one element that does not appear in any subset in S (the corollary of the condition above).
- If in S, there are two subsets that share one common element then that element must be fixed for all permutations that fix S (similarly to **Property 4**).

Back to the problem, let x be the number of elements that appear in at least 2 sets. Let y be the number of elements that appear in exactly one set. Since there is no more than one absent element, we count the number of appearances of each element in the subsets in S to get

$x + y \geq n - 1.$

Let m be the number of k-sets for building base S. Because each k-set $s \in S$, the set s contains exactly k element so by counting the relationship (subset, element), we have

$$km \geq 2x + y.$$

It is clear that $y \leq m$, thus $x \geq n - m - 1$. From here, we can conclude that

$$km \geq 2x + y = x + (x + y) \geq (n - m - 1) + (n - 1) \text{ or}$$

$$m \geq \frac{2(n-1)}{k+1}.$$

Therefore, we have the lower bound of m is $\left\lceil \frac{2(n-1)}{k+1} \right\rceil$

(*by* $\lceil a \rceil$ *we denote the smallest integer that is not smaller than the real number* a).

2.5. The structures

2.5.1. Additional conditions

We will build the set S satisfies the condition $m = \left\lceil \frac{2(n-1)}{k+1} \right\rceil$ and add more condition that no element exists in more than two subsets. Then we have an important estimation: "the number of elements that appear in two k-sets will not exceed m chooses 2".

Indeed, suppose $x > C_m^2$ then by the pigeonhole principle, two elements will appear in two k-sets (and they will not appear in any other subset).

Next, consider $2(n-1) = a(k+1) + r$ with $a \in \mathbb{Z}^+$ and $r \in \{0,1,\ldots,k\}$. If $r = 0$ then $m = a$, so all evaluation must result in the system of equations

$$\begin{cases} x + y = n - 1 \\ 2x + y = km \\ y = m \end{cases}$$

By solving this system of equations, we have

$$y = m = a \text{ and } x = n - m - 1 = \frac{a(k+1)}{2} - a = \frac{a(k-1)}{2}.$$

We have also

$$x \leq C_m^2 = C_a^2 \Leftrightarrow \frac{a(k-1)}{2} \leq \frac{m(m-1)}{2}$$

or

$$k \leq a = \frac{2(n-1)}{k+1} \Leftrightarrow n - 1 \geq \frac{k(k+1)}{2}.$$

If $r > 0$ then $m = a + 1$. Then we will still have $2x + y = km$ but $x + y \in \{n-1, n\}$ so we get $x \in \{km - n + 1, km - n\}$. Similarly, we have

$$n \geq \frac{k(k+1)}{2} \text{ or } n - 1 \geq \frac{k(k+1)}{2}.$$

(depending on if there is or there is no element that does not appear in any subset).

2.5.2. Building and proving by using a graph model

We will continue by using graphs. The specific steps are as follows:

Let A, B be the set that contains elements appearing twice and once, respectively. Let $|A| = x$, $|B| = y$, we always have $2x + y = km$. Based on whether we choose $x + y = n - 1$ or $x + y = n$ (corresponds to whether we have an element that does not appear), we can calculate the values of x, y such that $x \leq C_m^2$ and $y \leq m$. For simplicity purposes, we choose A is the set $[x]$.

We let $k-1$ the first positions of each set equal to the elements in A. The elements in B will fill in the k^{th} position of each subset of S. The distribution of the elements in B into sets must satisfy the following conditions:

- Each element appears exactly twice at two different k-sets.
- There are two k-sets that share exactly one element.

In order to achieve this, we consider the completed graph $G = (V, E)$ in which V is a set of vertices representing m subsets in a base $S = S(k, n)$, and E is a set of edges representing the elements in A. If two k-sets share an element then they will be connected by an edge, and because of the aforementioned condition, it is a simple undirected graph. Hence, we can enumerate the edges of a graph G by using the elements from A, each number is used once.

Because $|E| \geq |A|$, this can always be done (and there could be some edges that are not used).

Then, the number x that on an edge connecting two vertices representing k-sets V_1, V_2 then $x \in V_1, V_2$. Notice that $(k-1)m < 2x$ could happen, so some elements in A will be chosen to be the k^{th} element for two k-sets of S (here we care about the order of elements in the subset because of its simplicity, and it does not take effect on the original problem).

Lastly, for the k-sets in S missing the k^{th} position, we fill in that position with elements from B such that there are no two elements in the same k-set. Because $|B| < m$,

this is always true. We could see that the structure built above is adequate. Consider permutation f fixes all the k-sets in S being built with the aforementioned steps.

- For all $a \in A$, there exists two k-sets V_1, V_2 that contain it $(a \in V_1 \cap V_2)$ so permutation f fixes V_1, V_2 also satisfies $f(a) = a$. Hence f fixes all elements in A.
- For each element $b \in B$, there exists some k-sets that only contain b and $k-1$ elements of A (already fixed) so we also have $f(b) = b$.
- Lastly, if there is an element that does not appear, that element will also be fixed because the other $n-1$ elements have already been fixed.

2.6. Examples

We consider the following examples:

Example 1. For $n = 22, k = 6,$ then we calculate the number of $k-$sets in the base S is $m = \left\lceil \frac{2(n-1)}{k+1} \right\rceil = 6.$ Because this is a divisible case, so we must discard an element, suppose it is $22.$ At the same time, we also have

$$\begin{cases} x + y = 21 \\ 2x + y = 6 \times 6 \end{cases} \Leftrightarrow \begin{cases} x = 15 \\ y = 6 \end{cases}.$$

So 15 elements appear twice, and six elements appear once. Consider the following graph, the edges are enumerated based on the alphabetical order of the name of the vertices that it connects:

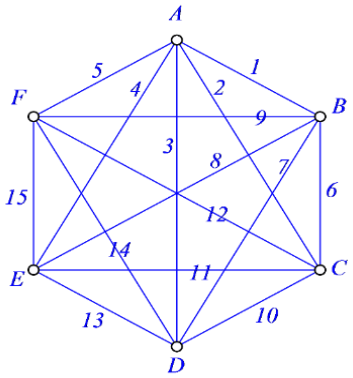

Figure 1. *Constructed graph for* $m = 6$ *and* 15 *elements*

From here we can build a complete model with all six sets:

$A = \{1, 2, 3, 4, 5\}, B = \{1, 6, 7, 8, 9\}, C = \{2, 6, 10, 11, 12\},$

$D = \{3, 7, 10, 13, 14\}, E = \{4, 8, 11, 13, 15\}, F = \{5, 9, 12, 14, 15\}.$

Table 1. *The distribution of elements into subsets in base*

Set	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	×	×	×	×	×											×					
2	×					×	×	×	×	×							×				
3		×				×				×	×	×						×			
4			×				×			×			×	×					×		
5				×				×			×		×		×					×	
6					×				×			×		×	×						×

Example 2. Consider example $n=99, k=10$ then the number of subsets of the base S is $m=\left\lceil \frac{2(n-1)}{k+1} \right\rceil =18.$ This is not a divisible case, so building the model becomes flexible.

If we include $99,$ then we have

$$\begin{cases} x+y=99 \\ 2x+y=18\times 10 \end{cases} \Leftrightarrow \begin{cases} x=81 \\ y=18 \end{cases}$$

Since we have $y=m$ so building the model will be similar to **Example 1**.

If we discard $99,$ then we have

$$\begin{cases} x+y=98 \\ 2x+y=18\times 10 \end{cases} \Leftrightarrow \begin{cases} x=82 \\ y=16 \end{cases}.$$

Here we have $2x=64>162=(k-1)m$ so when we let element 82 appears twice in order to build $k-1$ the first elements of each subset, then there will be an element appearing at the position k^{th} in those two subsets. Lastly, we fill in the k^{th} element of each subset because 16 elements appear once.

2.7. Extended analysis

In the previous part, we just consider the case that each element contained in at most 2 subsets so we get the condition $n-1\geq \frac{k(k+1)}{2}$ or $n\geq \frac{k(k+1)}{2}$. This condition implies that the construction cannot be applied for all values of (n,k).

For example, in case $k=n,$ the problem will not be solved since each subset must take all the elements of the original set, then can make the difference among elements. And about the case $k=n-1,$ we consider all subsets of size k of $[n]$ then it is easy to check that this base satisfies the condition. Next, we consider pairs (n,k) satisfying

$$k<n<\frac{k(k+1)}{2}.$$

Thus, for each pair (n,k) that satisfies the above conditions (it is clear that there exists some such base S) then to find the smallest size of S, we can conclude that there exist some element appears more than three times in the finding model. So we will have

$$m \geq \left\lceil \frac{2(n-1)}{k+1} \right\rceil + 1$$

(since the previous lower bound cannot be used anymore).

Example 3. Consider $(n,k) = (14,5)$ then $m = 5$. Denote x, y as the number of elements that appear in $2,1$ subsets and suppose that there just only two such kinds of elements then

$$\begin{cases} x + y = 14 \\ 2x + y = 5 \times 5 \end{cases} \Leftrightarrow \begin{cases} x = 11 \\ y = 3 \end{cases}$$

not satisfied since $x = 11 > C_5^2 = 10$. Now we discard 14 and let 1 appears 3 then we get

$$\begin{cases} x + y = 12 \\ 2x + y + 3 = 25 \end{cases} \Leftrightarrow \begin{cases} x = 10 \\ y = 2 \end{cases}.$$

We construct the graph by fill the numbers from $2 \to 11$ on each edge as below.

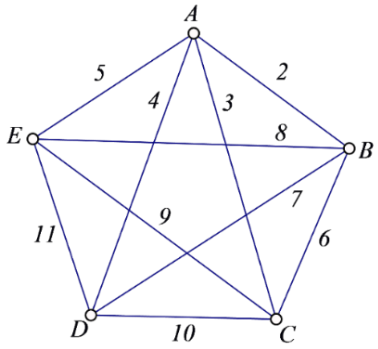

Figure 2. *Construction the graph for* $m = 5$ *subsets and* 11 *elements*

From here, the base S contains $m = 5$ subsets as follow

$$\begin{cases} A = \{1,2,3,4,5\}, \\ B = \{1,2,6,7,8\}, \\ C = \{1,3,6,9,10\}, \\ D = \{4,7,10,11,12\}, \\ E = \{5,8,9,11,13\}. \end{cases}$$

3. Conclusion and future works

On evaluating the number of occurrences of each element in the sets, we have established the lower bound for the base and built the basis using a graph model. The construction step is only successful with the appropriate data, the general case is being studied further. With the characteristics of the problem, its applicability to data storage, security problems, private information retrieval is completely feasible.

Let consider the following situation, originating from an idea that,**"***There are* n *users and each of them stores a unique file on the server. System admin does not know who is the owner of each file so he performs a list of queries that choosing some* k *files and asking the users who are the owners of those files. So the smallest number of queries that need to find the exact owner is also the smallest base of* $S(n,k)$ *discussed in this study"*

Through analyzing the above solutions, we have a general remark that, if an element $x \in \{1,2,\ldots,n\}$ appears in a group of at least 2 subsets then its image $\sigma(x)$ will belong to the intersections of these sets. In case the size of that intersection is 1 then $\sigma(x) = x$.

This allows us to construct some elements that appear in more than 2 subsets. For this idea, we may arrange elements appropriately in m subsets to expand the bound:

$$n \le C_m^0 + C_m^1 + C_m^2 + \cdots + C_m^m = 2^m.$$

The idea of the element connecting a group of subsets rather than just two subsets related to the edge in the hypergraph, so further research on that theory is likely to help thoroughly solve the given problem.

❖ ***Conflict of Interest:*** *Authors have no conflict of interest to declare.*

❖ ***Acknowledgement:*** *This research was supported by Vietnam National University Ho Chi Minh City (VNU-HCM) under the grant number NCM2019-18-01.*

REFERENCES

Bui, D. K., Nguyen, D. T., & Hoang, H. D. (2004). *Giao trinh ma hoa thong tin – Li thuyet va ung dung* [*Course book of cryptography – Theory and Application*]. Labour and Social publisher company limited, 90-100.

Diatonics, P., Fulman, J., & Guralnick, R. (2008). On fixed points of permutations. *J. Algebraic Combi., 28*(1), 189-218.

Jung, J., Mummert, C., Niese, E., & Schroeder, M. (2018). On erasure combinatorial batch codes. *Designs, Codes and Cryptography, 12*(1), 49.

Ishai, Y., Kushilevitz, E., & Ostrovsky, R. (2004). Batch codes and their applications. *Proceedings of STOC 2004, ACM Press,* 262-271.

Paterson, M. B., Stinson, D. R., & Wei, R. (2009). Cominatorial Batch Codes. *Communications in Advanced Mathematical, 3*(1), 13.

Smith, R. (2006). Permutation Reconstruction. *The Electronic journal of combinatorics, 13*(11).

Sean, E., Kevin, F., & Ben, G. (2016). Permutations fixing a k-set. *International Mathematics Research Notices,* (21), 6713-6731.

CƠ SỞ NHỎ NHẤT CỦA CÁC TẬP CON k – SETS

Lê Phúc Lữ[1*], Nguyễn Đình Song Ân[2]

[1]Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam

[2]Đại học Saint John's, Mĩ

[*]Tác giả liên hệ: Lê Phúc Lữ – Email: lephuclu@gmail.com

Ngày nhận bài: 21-3-2021; ngày nhận bài sửa: 16-9-2021; ngày duyệt đăng: 20-9-2021

TÓM TẮT

Trong các lĩnh vực về lí thuyết thông tin nhưdựng mô hình lưu trữ, chia sẻ riêng tư, mã hóa... đôi khi ta muốn phân tán một mẫu dữ liệu cho trước thành nhiều phần nhỏ, mỗi phần được lưu giữ bởi một party mà khi một số lượng đủ nhiều các party phối hợp với nhau thì sẽ có cách khôi phục lại được thông tin gốc. Hướng tới mục tiêu đó, bài viết này mô tả việc xuất phát từ một tập hợp hữu hạn,xây dựng một họ các tập con cùng số phần tử sao cho tồn tại duy nhất một hoán vị là ánh xạ 1-1 vào mỗi tập con. Tất nhiên, tính tối ưu sẽ được xét thông qua kích cỡ nhỏ nhất của họ các tập con đó. Bằng cách đánh giá số lượt xuất hiện của mỗi phần tử trong các tập con, ta có thể thiết lập được thành công chặn dưới cho số tập con, đồng thời xây dựng được bằng mô hình graph đơn vô hướng. Bước xây dựng chỉ thành công với những dữ liệu thích hợp và trường hợp tổng quát đang được nghiên cứu thêm.

Từ khóa: tập cơ sở; điểm bất động; đồ thị; khôi phục hoán vị