Discrete Mathematics 64 (1987) 239–252
North-Holland

SHORTEST PREFIX STRINGS CONTAINING ALL SUBSET PERMUTATIONS

Alejandro A. SCHÄFFER*
AT&T Bell Laboratories, Murray Hill, NJ 07974, U.S.A.

Received 22 October 1985
Revised 8 November 1985

What is the length of the shortest string consisting of elements of $\{1, \ldots, n\}$ that contains as subsequences all permutations of any k-element subset? Many authors have considered the special case where $k = n$. We instead consider an incremental variation on this problem first proposed by Koutas and Hu. For a fixed value of n they ask for a string such that for all values of $k \leq n$, the prefix containing all permutations of any k-element subset as subsequences is as short as possible. The problem can also be viewed as follows:

For $k = 1$ one needs n distinct digits to find each of the n possible permutations. In going from k to $k + 1$, one starts with a string containing all k-element permutations as subsequences, and one adds as few digits as possible to the end of the string so that the new string contains all $(k + 1)$-element permutations.

We give a new construction that gives shorter strings than the best previous construction. We then prove a weak form of lower bound for the number of digits added in successive suffixes. The lower bound proof leads to a construction that matches the bound exactly. The length of a shortest prefix string is

$$k(n-2) + \lfloor \tfrac{1}{3}(k+1) \rfloor + 3, \quad \text{for } k > 2.$$

The lengths for $k = 1, 2$ are n and $2n - 1$. This proves the natural conjecture that requiring the strings to be prefixes strictly increases the length of the strings required for all but the smallest values of k.

1. Introduction and problem history

Knuth [3] poses the following problem which he attributes to Karp: What is the length of the shortest string consisting of elements of $\{1, \ldots, n\}$ that contains all permutations of the set as subsequences? Several published papers [1], [4], [6] give constructions for strings of length

$$n^2 - 2n + 4, \quad \text{for } n \geq 3.$$

For example the string 1231231 of length 7 contains the permutations 123, 132, 213, 231, 312, and 321 as subsequences. Savage [8] generalizes Adleman's construction to the problem of finding a shortest string containing any permutation of any k-element subset, obtaining an upper bound of

* This work was done while the author was a summer employee of AT&T Bell Laboratories. During the academic year he is a student in the Department of Computer Science, Stanford University, Stanford, CA 94305 where any correspondence should be sent.

$$k(n-2)+4, \quad \text{for } 3 \leqslant k \leqslant n.$$

Notice that the bounds are identical for $k=n$. Both of these upper bounds also appear in a technical report by Newey [7]. The best lower bound known for $k=n$ is

$$n^2 - cn^{7/4+\varepsilon}, \quad \text{where } \varepsilon > 0 \text{ and } c \text{ depends on } \varepsilon,$$

due to Kleitman and Kwiatkowski [5].

In all of the constructions cited the relationship between the string for n and the string for $n+1$ is nontrivial. Koutas and Hu, therefore, also consider the following related problem [6]. Let $F(n, 1)=n$ be the length of a shortest string containing all the elements of $\{1, \ldots, n\}$. Now add as few numbers as possible to the end of the string until it contains all permutations of any two-element subset of $\{1, \ldots, n\}$. The length of this string is $F(n, 2)$; $F(n, 2)=2n-1$ because the last number among the first n does not need to appear again. Keep adding digits on the right, trying to minimize the length, so that prefixes of the string contain all permutations of k-element subsets, $k=3, 4\ldots$ It is essential to observe that very few strings that achieve $F(n, k)$ can be extended minimally to strings that achieve $F(n, k+1)$. This means that in defining $F(n, k+1)$ one should minimize over all prefixes satisfying the subsequence condition that are of length $F(n, k)$. Koutas and Hu show by exhaustive case analysis that

$$F(n, 3)=3n-2, \qquad F(n, 4)=4n-4, \qquad F(n, 5)=5n-5.$$

They conjecture that

$$F(n, k)=k(n-1), \quad \text{for } 4 \leqslant k \leqslant n-1.$$

The notation of Koutas and Hu is a bit deceptive for the case of arbitrary k. The reason is that the definition for the string for $k+1$ presumes that its shortest prefix which contains all k-element permutations is as short as possible. Therefore we define $L(n, k)$ to be a sequence of k numbers where the ith number is the fewest number of digits one has to add to an optimal prefix for $i-1$ to get a shortest prefix for i. We can restate the earlier results as:

$$L(n, 1)=n, \qquad L(n, 2)=n, n-1, \qquad L(n, 3)=n, n-1, n-1,$$
$$L(n, 4)=n, n-1, n-1, n-2, \qquad L(n, 5)=n, n-1, n-1, n-2, n-1.$$

Cai [2] disproves the conjecture of Koutas and Hu by constructing a string where for every fourth value of k starting with $4, 8, \ldots$ one adds only $n-2$ numbers instead of $n-1$. Thus the length of Cai's string is

$$k(n-1)+1-\lfloor \tfrac{1}{4}k \rfloor, \quad \text{for } 1 \leqslant k \leqslant n.$$

However, this does not prove an upper bound on $F(n, k)$ for general k. For this reason our sequence notation captures the incremental nature of the problem much better than the original measure of length, F. Cai's construction first saves an extra number at $k = 8$. However, we will show that it is already possible to save an extra number at $k = 6$. It is conceivable that the cost of saving a number earlier in the prefix, as demanded by the problem statement, is that later in the string it is not possible to save as many numbers as Cai's construction saves. If this were the case, the optimal string would actually be longer than Cai's string. Using the sequence notation two strings are compared *lexicographically*, and the one that saves an extra number *sooner* is better, regardless of which string is longer. More concretely, Cai's construction does show that $L(n, k)$ is lexicographically less than or equal to

$$n, \overline{n-1, n-1, n-2, n-1}, \quad \text{for } 1 \leqslant k \leqslant n,$$

where we use the horizontal bar to mean the sequence consisting of the first k elements of the infinite sequence in which the digits under the bar are repeated.

We first improve the bound to

$$n, n-1, \overline{n-1, n-2}, \quad \text{for } 1 \leqslant k \leqslant n.$$

Then we prove a lower bound of

$$n, n-1, n-1, \overline{n-2, n-1, n-2}, \quad \text{for } 1 \leqslant k \leqslant n,$$

and finally, we exhibit a string matching this lower bound exactly. That is, the second sequence is $L(n, k)$. This shows that the restriction to prefix strings does indeed increase the length, since the upper bound for the general problem is strictly less than the length for the prefix problem provided $k \geqslant 5$.

The prefix restriction is natural because it limits the need for inductive reasoning in the proofs to the rightmost pieces, i.e., the suffixes, of the string. The fundamental difficulty in improving the upper bound for this problem is that only a few strings that achieve $L(n, k)$ can be extended (minimally) to achieve $L(n, k+1)$. The fundamental difficulting in proving the lower bound is that the pattern of suffixes added for successive values of k is complicated, and as a result the case analysis for what can and cannot be found at the right end of the string is messy.

2. Improved upper bound

In this section we constructively prove the following upper bound:

Theorem 1. *$L(n, k)$ is lexicographically less than or equal to*

$$n, n-1, \overline{n-1, n-2}, \quad \textit{for } 1 \leqslant k \leqslant n.$$

Before beginning the formal construction we try to give some intuition for how Koutas and Hu arrived at their conjecture, and why their bound as well as Cai's can be improved. We call the suffix added to an optimal prefix in going from

$k-1$ to k, *block* k. The first block must contain all n numbers. Thereafter the kth block need not include the last number of block $k-1$. To see this suppose that we are looking for the permutation $x_1 \ldots x_k$ as a subequence where x_k is the last number in block $k-1$. We can assume as an inductive hypothesis that $x_1 \ldots x_{k-1}$ occurs as a subsequence of the first $k-1$ blocks. Furthermore $x_{k-1} \neq x_k$, so we don't need the last number to find $x_1 \ldots x_{k-1}$ as a subsequence. This implies that $x_1 \ldots x_k$ is actually a subsequence of the first $k-1$ blocks, and hence also a subsequence of the first k blocks even if block k does not contain x_k. This shows that $L(n, k)$ is bounded by $n, \overline{n-1}$.

Koutas and Hu show that in the fourth block one can save an extra number; their string has only $n-2$ numbers in that block [6]. Cai shows that one can save an extra number for every four blocks [2]. If the fourth block is missing x_1 and x_2, and if they occur in that order in the third block then one has to make sure that all the permutations that end with x_2x_1 occur as subsequences. Cai constructs a nice inductive argument that restricts this difficulty to the last 4 blocks even if earlier blocks also save an extra number.

However, it turns out that one can save an extra number more often than every fourth block, at the cost of refining the argument. This construction then guides us to various observations about how much more improvement one might expect. The lemmata and proofs of Section 3 in turn guide the construction of a string in Section 4 that matches the lower bound of Section 3. In our first construction that achieves the weaker bounds of Theorem 1, blocks $5, 7, 9, \ldots$ have $n-1$ numbers each, while blocks $4, 6, 8, \ldots$ have $n-2$ numbers each.

Construction. Set

$$A_1 = 1, \ldots, n, \qquad A_4 = 1, \ldots, n-5, n-1, n-3, n-4,$$
$$A_2 = 1, \ldots, n-1, \qquad A_5 = 1, \ldots, n-5, n-2, n-1, n, n-3,$$
$$A_3 = 1, \ldots, n-3, n, n-2, \qquad A_6 = 1, \ldots, n-4, n-2, n-1.$$

The string for $k=1$ is A_1; the string for $k=2$ is A_1A_2; the string for $k>2$ is the first k blocks of $A_1A_2\overline{A_3A_4A_5A_6}$. Notice that A_3 and A_5 have $n-1$ numbers each and A_4 and A_6 have $n-2$ numbers each. Also, notice that the odd blocks after block 1 are missing the last number of the previous block. Thus in the inductive proof of correctness we only need to check the cases where k is even.

Suppose $k=4$. Block 4 is missing n and $n-2$, but block 3 ends in $n, n-2$; therefore, the only 4-element permutations that could cause difficulty are those ending in $n-2, n$. Let x_1x_2n-2, n be such a permutation. The last chance to find n is at the end of block 3 and the $n-2$ immediately before it occurs in block 2. If $x_2 \neq n-1$ then we can certainly find x_1x_2 among the first $2n-3$ numbers. If $x_2 = n-1$ then we use the $n-1$ of block 1, but we now know that $x_1 < n-1$, therefore x_1 occurs earlier in the string.

Backwards reconstruction of the "worst possible permutation" also works for larger values of k. Suppose now $k = 6$ and the worst possible permutation is $x_1 \ldots x_6$. Block 6 is missing $n-3$ and n. Block 5 ends in $n, n-3$, so $x_6 = n$ and $x_5 = n-3$. This means we need to use the $n-3$ of block 4. The only number after it in block 4 is $n-4$, so $x_4 = n-4$, otherwise we can certainly find the permutation as a subsequence. We use the $n-4$ of block 3; the only number after it in block 3 that does not already occupy a place in the permutation is $n-2$, so $x_3 = n-2$. We use the $n-2$ of block 2 and $x_2 = n-1$. All the numbers that might appear earlier in the permutation come before $n-1$ in the first block so we can find x_1 there.

For the inductive step assume that any permutation $x_1 \ldots x_j$ for $j < k$, can be found as a subsequence in the first j blocks. We consider two cases depending on whether k is divisible by 4 or not. If $k \equiv 0 \pmod 4$, then the worst possible permutation ends in $n-4, n-3, n-1, n-2, n$. The $n-4$ used is in block $k-5$ which is of the form A_3. In that block the only numbers after $n-4$ have already been used, so we can find x_{k-5} in that block. By the induction hypothesis we can find $x_1 \ldots x_{k-6}$ in the first $k-6$ blocks. If $k \equiv 2 \pmod 4$ then the worst possible permutation ends in $n-1, n-2, n-4, n-3, n$. We use the $n-1$ in block $k-5$ which is of the form A_5. In that block the only numbers after $n-1$ already occur in the permutation, so we can find x_{k-5} in that block, and once again we can find $x_1 \ldots x_{k-6}$ in the first $k-6$ blocks. □

This construction suggests many questions about the detailed structure of a shortest string. Two particularly obvious ones are:
- Is it possible to have fewer than $n-2$ numbers in a block?
- Is it possible to have $n-2$ numbers in 2 consecutive blocks?

One might think that as k gets large the class of "worst possible permutations" becomes very restricted and therefore it would seem likely that many numbers could be omitted. Because of the incremental nature of the problem, however, one must know the exact sequence $L(n, k)$ before making any claims about how few numbers need appear in block $k+1$. For example one could have a string that has $n-1$ numbers in the penultimate block and $n-3$ numbers in the last block, but the problem statement requires that one prefer a string with a block of size $n-2$ followed by a block of size $n-1$, if such a string exists.

3. A lexicographic lower bound

In this section we consider by how much one can improve on the construction in the previous section. We prove a series of lemmata that when put together give a very precise characterization of how many numbers one might be able to save in constructing the block sequence. Because the lower bound proof assumes greediness, so that if it is possible to save numbers, one saves them immediately, it cannot address the question of length of the string. It is conceivable that one cannot be as greedy as Theorem 2 suggests, but nevertheless the length of the shortest prefix string is less than the length given by the block length sequence in Theorem 2. However, in Section 4 we show the significance of the lower bound by exhibiting a string that matches it exactly. For the first two lemmata we assume that S is an optimal string, so that all n of its prefixes are as short as possible.

Lemma 1. *No number occurs twice in a block of S.*

Proof. Let $j \leqslant n$ be given. By definition every $(j-1)$-element permutation can be found in the first $j-1$ blocks. Thus to find any j-element permutation we need at most one element of the jth block. S is a shortest string, and hence the jth block need not contain two copies of the same number. Since j was arbitrary, no block contains two copies of the same number. □

Lemma 2. *Let $j \leqslant n$ and a subset, P, of $\{1, \ldots, n\}$ of size $n-j =: p$ be given. There is a permutation of the elements of P that cannot be found as a subsequence in the first $p-1$ blocks of S. That is, some element of block p is needed to find that permutation.*

Proof. We build such a permutation, $x_1 \ldots x_p$ explicitly. Let x_1 be the element of P that occurs last in block 1. Let $P_i = P \setminus \{x_1, \ldots, x_i\}$. For $1 \leqslant i < p$, let x_{i+1} be the element of P_i that occurs latest in block $i+1$. By Lemma 1, it is impossible to use two elements of the same block to find the p-element permutation, $x_1 \ldots x_p$ as a subsequence. □

It is essential that the set specified in Lemma 2 is arbitrary.

Lemma 3. *No number can be missing from two consecutive blocks. Furthermore, if block j is missing two distinct numbers x_1 and x_2, then the numbers cannot both be missing from block $j+2$.*

Proof. Suppose x_1 is missing from both block j and block $j+1$. By Lemma 2 we can construct a j-element permutation not using x_1 that requires some element of block j in order to be found as a subsequence. If we now append x_1 to the permutation (string), the resulting $(j+1)$-element permutation cannot be found in the first $j+1$ blocks.

For the second part assume that x_1 and x_2 are missing from blocks j and $j+2$. By Lemma 1, they can occur at most once each in block $j+1$. We can assume without loss of generality that they occur in the order x_1x_2. By Lemma 2, we can construct a j-element permutation that does not use x_1 or x_2, but requires some element of block j to be found as a subsequence in the first j blocks. We can then append x_2x_1 to the permutation, and the resulting $(j+2)$-element permutation cannot be found as a subsequence in the first $j+2$ blocks. □

Lemma 4. *If block j is missing the last two numbers of block $j-1$, and if the last or penultimate number of block $j-1$ is the last or penultimate number of block $j-3$ and is missing from block $j-2$, then block $j+1$ can be missing only one number (the last number of block j).*

Proof. Let x_1 be the last or penultimate number of block $j-3$ and the last or penultimate number of block $j-1$ that does not occur in either block $j-2$ or block j. Let x_2 be the last number of block j. Note that x_2 may occur in block $j-1$, but it cannot occur after x_1 because if there is a number after x_1 in block $j-1$ it is the last number of that block and must be missing from block j. Suppose we wanted to omit some number x_3 in addition to x_2 from block $j+1$. By Lemma 3, x_3 cannot be one of the numbers missing from block j. We can build a $(j-2)$-element permutation that does not use any member of $\{x_1, x_2, x_3\}$ and requires some element of block $j-2$ to be found as a subsequence. We append $x_1x_2x_3$ to that permutation making a $(j+1)$-element permutation.

By assumption x_1 does not occur in block $j-2$, so we need to go to the last two elements of block $j-1$ to find it. Also, x_2 does not occur after x_1 in block $j-1$, so the best we can do is use the x_2 that occurs as the last element of block j. We assumed that x_3 does not appear in block $j+1$, and thus we have a $(j+1)$-element permutation that cannot be found in the first $j+1$ blocks, a contradiction. □

Lemma 4 provides the first substantial piece of an induction step. Under its restrictive assumptions we can prove that after two blocks of size $n-2$ there cannot be a third such block. Lemma 5 fills in the remaining piece by showing that the number missing in that third block of size $n-1$ is also missing two blocks later, and by considering the contents of the block in between.

Lemma 5. *Suppose block j has $n-2$ numbers, block $j+1$ has $n-2$ numbers, and block $j+2$ has $n-1$ numbers. Suppose further that the last number of block $j-2$ is the penultimate number of block j and does not appear in block $j+1$. Then we can conclude that block $j+3$ has at least $n-2$ numbers, and if block $j+3$ has exactly $n-2$ numbers, then block $j+4$ has at least $n-2$ numbers. Moreover, block $j+4$ can be missing two numbers only if the last number of block $j+1$ is the penultimate number of block $j+3$ and is missing from block $j+4$.*

Proof. Let x_1 and x_2 be the last two numbers of block j; both are missing from block $j+1$, and x_1 is missing from block $j-1$. Let x_3 be the last number of block $j+1$; x_3 is missing from block $j+2$. Let x_4 be the last number of block $j+2$, and let $x_5 \neq x_4$ be any other number that does not occur in block $j+3$. Suppose for the moment that $x_4 = x_1$. Then we build a $(j+1)$-element permutation not using x_4 or x_5 that requires some element of block $j+1$ to be found as a subsequence. We append x_4x_5 to that permutation, and the resulting $(j+3)$-element permutation cannot be found in the first $j+3$ blocks.

Suppose instead that $x_1 \neq x_4$, and that x_5 is defined as before. We build a $(j-1)$-element permutation not using x_1, x_3, x_4, or x_5 that requires some element of block $j-1$. We try to append $x_1x_3x_4x_5$ to this permutation to build a $(j+3)$-element permutation that does not occur as a subsequence of the first $j+3$ blocks. The only circumstance under which we cannot build this permutation is if $x_1 = x_5$. This proves that block $j+3$ can be missing at most two numbers, and if two numbers are missing one must be x_1, but x_1 cannot be the last number of block $j+2$ because we dispensed with that case in the previous paragraph.

Consider what can be missing from block $j+4$. We built a $(j+2)$-element permutation that ended in $x_1x_3x_4$ where we needed the x_4 that occurs as the last element of block $j+2$. The elements x_1 and x_4 do not occur in block $j+3$ if that block has $n-2$ numbers, so they cannot be missing from block $j+4$. If block $j+4$ were missing two numbers other than x_3, we could make sure they did not occur in that $(j+2)$-element permutation, and then append them in reverse order to get a $(j+4)$-element permutation that could not be found as a subsequence of the first four blocks. Thus block $j+4$ can be missing at most x_3 and one other number x_6. In order to complete the proof we need to show that if both numbers are missing from block $j+4$, then block $j+3$ ends in x_3x_6.

First suppose that block $j+3$ has x_3 as its last number (instead of the penultimate as we claim). Then we can build a $(j+1)$-element permutation that does not use x_1, x_3, or x_6 and requires some element of block $j+1$ in order to be found as a subsequence. The number x_1 does not occur in block $j+1$, x_3 does not occur in block $j+2$, and, in fact, the next occurrence of x_3 is at the end of block $j+3$. Therefore if we append $x_1x_3x_6$ to our permutation, we get a $(j+4)$-element permutation that cannot be found in the first $j+4$ blocks.

Now suppose that x_3 is neither the last nor the penultimate number of block $j+3$. This means that x_6 is the last element of block $j+3$, and that some other number, y_1, is the penultimate number in block $j+3$. Once again construct a $(j-1)$-element permutation that requires some element of block $j-1$; we shall see momentarily which numbers should not be used in this permutation, so that they can be appended to it. Append x_1 to the permutation. The next occurrence of x_1 is as the penultimate number in block j. Append the penultimate number of block $j+1$, call it y_2, which is known to be distinct from x_1 because x_1 does not occur in block $j+1$. Next append the last number of block $j+2$ other than y_2; call this number y_3. We can assume that x_1 is not among the last two numbers of block $j+2$ because if it were, then we could apply Lemma 4 with j in that Lemma corresponding to $j+3$ in this one to conclude that block $j+4$ would have $n-1$ numbers. Therefore, we do not need the explicit restriction that $y_3 \neq x_1$, and know that y_3 is among the last two numbers of block $j+2$. Observe that neither the last element of block $j+2$ nor x_1 occur in block $j+3$, therefore $\{x_1, y_2, y_3\} \not\subseteq \{x_6, y_1\}$. Choose $y_4 \in \{x_6, y_1\} \setminus \{x_1, y_2, y_3\}$, and append y_4x_3 to the permutation to get a $(j+4)$-element permutation that cannot be found in the first $j+4$ blocks provided that the original $(j-1)$-element permutation contains no member of $\{x_1, x_3, y_2, y_3, y_4\}$.

Since Lemma 2 guarantees that such a $(j-1)$-element permutation exists, the final assertion that the last number of block $j+1$ is the penultimate number of block $j+3$ holds. $\square$

We can actually get even more information from the last construction in the proof. If $x_6 \neq y_3$, then we could just append $x_6 x_3$ to the permutation at the last step and get an even broader contradiction implying that block $j+4$ could not be missing anything other than x_6. If there is any hope of having only $n-2$ elements in block $j+4$, then $x_6 = y_3$. That is, the last element of block $j+3$ must be among the last two numbers of block $j+2$. Since it cannot be the last number, and yet occur in block $j+3$, x_6 must be the penultimate number of block $j+2$. The reader can check that the string exhibited in Section 4 has this property.

It is important to see how Lemmata 4 and 5 can be used in tandem. First one establishes the conditions of Lemma 5 in order to show that two consecutive blocks have at least $n-2$ numbers and that if they both have exactly that many then the last two numbers of the first are missing from the second, and so on. The conclusions of Lemma 5 are precisely the assumptions needed to apply Lemma 4 to show that block after the two of size $n-2$ has $n-1$ elements. After applying Lemma 4 one can again apply Lemma 5 for the next two blocks, repeating the process until j reaches n.

Theorem 2. *$L(n, k)$ is lexicographically greater than or equal to*

$$n, n-1, n-1, \overline{n-2, n-1, n-2}, \quad \text{for } 1 \leqslant k \leqslant n.$$

Proof. We mentioned before that Koutas and Hu show that the first 5 elements of this sequence must be the same as for $L(n, k)$. We prove that if both blocks 6 and 7 have exactly $n-2$ numbers then the last number of block 4 is the penultimate number of block 6 and is missing from block 7. We also prove that blocks 6 and 7 can be missing at most two numbers each. This establishes the hypotheses necessary to apply Lemma 5 for the first time with $j=6$. Lemma 5 shows that if blocks 6 and 7 have no more than $n-2$ numbers each, we can then apply Lemma 4, to show that block 8 has $n-1$ numbers. The conclusion of Lemmata 5 and 4 now enable us to again apply Lemma 5 with $j=9$, and we can repeat the cycle as often as necessary to obtain the sequence given in the statement of the Theorem. Of course, there may be a block where Lemmata 4 and 5 would permit the omission of 2 numbers, but for other reasons we can omit only 1; however, this does not violate Theorem 2 since we compare the sequences of block lengths lexicographically.

Let y_1 be the element missing from block 2. We show that y_1 is also missing from blocks 4 and 6 and occurs as the penultimate element of block 3. Suppose first that two other elements v_1, v_2 were missing from block 4. Notice that both v_1 and v_2 must occur in block 3, so they are distinct from the last element of block 2, call it v_3. Thus if v_1, v_2 occur in that order in block 3, the 4-element permutation $y_1v_3v_2v_1$ cannot be found as a subsequence in the first four blocks. This shows that y_1 is missing from block 4.

Next suppose that y_1 is the last element of block 3 and the other number missing from block 4 is v_4. Then we begin with any pair not using y_1 or v_4 that requires some element of block 2 and append y_1v_4 to get a 4-element permutation that cannot be found in the first four blocks.

Next suppose that y_1 is not the penultimate element of block 3, so that block 3 ends v_5v_4 where v_5 is some other number that occurs in blocks 1, 2, 3, and 4. Let v_6 be the last element of block 2, and use the technique of Lemma 2 to build a 3-element permutation out of $\{v_4, v_5, v_6\}$ that requires some element of block 3 to be found as a subsequence. Since v_6 is missing from block 3, that element must be v_4 or v_5, both of which occur after y_1 in block 3. If we append y_1 to the 3-element permutation, we get a 4-element permutation that cannot be found in the first four blocks. Thus y_1 must be the penultimate number of block 3.

Next suppose that block 6 is missing two numbers, w_1 and w_2 neither of which is y_1, and that the numbers occur in that order in block 5. Note also that they must be distinct from the last element of block 4, z_1, because z_1 does not occur in block 5. Build a 2-element permutation requiring some element of block 2 that does not use any element of $\{w_1, w_2, y_1, z_1\}$. Append $y_1z_1w_2w_1$ to get a 6-element permutation that cannot be found in the first 6 blocks. It is necessary to know that y_1 is the penultimate element of block 3 to ensure that we have to go to the end of block 4 to find z_1 after it. This also shows that block 6 can be missing at most two numbers. Let w_1 be the number missing from block 6 other than y_1, if there is a second number missing. If y_1 were the last element of block 5, then we could not find any 6-element permutation ending in y_1w_1 where the fourth element in the permutation required some member of block 4. The reason is that y_1 is also missing from block 4. Therefore, w_1 must be the last number of block 5, if block 6 is missing two numbers.

Next we show that if block 6 is missing two numbers, which must be y_1 and w_1, then block 7 can be missing at most two numbers one of which must be z_1, the last number of block 4. Suppose instead that block 7 is missing two other numbers, z_2 and z_3, and that they occur in that order in block 6. Note that both z_2 and z_3 are distinct from both w_1 and y_1 because neither w_1 nor y_1 occurs in block 6. We begin with a 2-element permutation not using any member of $\{w_1, y_1, z_1, z_2, z_3\}$ that requires some element of block 2. Then we append $y_1z_1w_1z_3z_2$ to get a 7-element permutation. To find y_1 we have to go to the penultimate number of block 3; the next occurrence of z_1 is at the end of block 4; the next occurrence of w_1 is at the end of block 5; finally, we cannot find z_3z_2 in blocks 6 and 7.

It remains only to show that z_1, the last number of block 4, is the penultimate element of block 6. We know from Lemma 3 that z_1 must reappear somewhere in block 6, since it is missing from block 5. We show first that if block 7 is missing two numbers then z_1 cannot be the last number of block 6. Suppose y_2 were the other number missing from block 7. Build a 5-element permutation not using y_2 or z_1 that requires some element of block 5. Then append $z_1 y_2$, to get a 7-element permutation that cannot be found in the first 7 blocks.

The remaining possibility is that z_1 occurs before the penultimate number in block 6. Let z_4 be the last element of block 6; it does not occur in block 7. We just proved that $z_1 \neq z_4$. In order to complete the proof we show first that no member of block 4 and block 6 can reappear in block 5 after z_4. For this argument it does not matter where y_1 and the last number of block 3 which are both missing from block 4 reappear in block 5. Of course, the last number of block 5, w_1, does occur after z_4 in that block. Suppose that some other number, u, did occur after z_4 in block 5. Observe that u is distinct from w_1 and y_1 because they do not occur in block 6; $u \neq z_1$ because z_1 is missing from block 5. We start with a two-element permutation that requires some element of block 2, and append y_1 for which we must use the penultimate member of block 3. If u occurs after w_1 in block 4, we append $uw_1z_4z_1$ to get a 7-element permutation that cannot be found in the first 7 blocks. To see this observe we assumed that u occurs after w_1 in block 4, so we must use the w_1 at the end of block 5, and z_1 and z_4 are missing from block 7. The other case has u before w_1 in block 4. Then we make a slight change by instead appending $w_1uz_4z_1$ to the permutation to get a bad 7-element permutation. We need to use the u of block 5 and it occurs after z_4 in that block, so we cannot find the permutation. A similar argument shows that the last number of block 3 cannot occur after z_4 in block 5, since we can just substitute that number for u in the last bad permutation we constructed resulting in similar problems.

Now we return to the objective of showing that z_1 is the penultimate element of block 6. Suppose there is a z_5 distinct from z_4 that comes after z_1 in block 6. Again z_5 must be distinct from y_1 and w_1 since they are missing from block 6. We start with a 2-element permutation using no member of $\{w_1, y_1, z_1, z_4, z_5\}$ that requires some number of block 2. We append y_1 to force us to the penultimate number of block 3. Next we append z_4 and w_1 in the order opposite to their order in block 4. This forces us to use either the last element of block 5 (w_1) or the z_4 near the end of block 5. We just proved in the previous paragraph that z_5 cannot occur after z_4 in block 5. Therefore if we make z_5 the sixth number in our permutation, we have to go to block 6 to find it. We then append z_1 which we assumed does not occur after z_5 in block 6, and does not occur in block 7. This gives a 7-element permutation that cannot be found in the first 7 blocks. Therefore z_1 must be the penultimate number of block 6, as claimed. □

4. Optimal Construction

In this section we prove constructively that the lower bound of the previous section is also an upper bound.

Theorem 3. $L(n, k) = n, n-1, n-1, \overline{n-2, n-1, n-2}$, *for* $1 \leqslant k \leqslant n$.

Construction. Set

$$\begin{aligned}
A_1 &= 1, \ldots, n,\\
A_2 &= 1, \ldots, n-1,\\
A_3 &= 1, \ldots, n-3, n, n-2,\\
A_4 &= 1, \ldots, n-5, n-1, n-4, n-3,\\
A_5 &= 1, \ldots, n-6, n-2, n-5, n, n-4, n-1,\\
A_6 &= 1, \ldots, n-6, n-2, n-5, n-3, n-4,\\
A_7 &= 1, \ldots, n-6, n-1, n, n-2, n-5,\\
A_8 &= 1, \ldots, n-6, n-4, n-1, n-3, n, n-2,\\
A_9 &= 1, \ldots, n-6, n-4, n-1, n-5, n,\\
A_{10} &= 1, \ldots, n-6, n-2, n-3, n-4, n-1,\\
A_{11} &= 1, \ldots, n-6, n, n-2, n-5, n-3, n-4,\\
A_{12} &= 1, \ldots, n-6, n, n-2, n-1, n-3,\\
A_{13} &= 1, \ldots, n-6, n-4, n-5, n, n-2,\\
A_{14} &= 1, \ldots, n-6, n-3, n-4, n-1, n-5, n,\\
A_{15} &= 1, \ldots, n-6, n-3, n-4, n-2, n-5,\\
A_{16} &= 1, \ldots, n-6, n, n-1, n-3, n-4,\\
A_{17} &= 1, \ldots, n-6, n-5, n, n-2, n-1, n-3,\\
A_{18} &= 1, \ldots, n-6, n-5, n, n-4, n-1,\\
A_{19} &= 1, \ldots, n-6, n-3, n-2, n-5, n,\\
A_{20} &= 1, \ldots, n-1, n-3, n-4, \quad n-2, n-5,\\
A_{21} &= 1, \ldots, n-1, n-3, n, n-2,\\
A_{22} &= 1, \ldots, n-5, n-4, n-1, n-3.
\end{aligned}$$

The string for k consists of the first k blocks of

$$A_1A_2A_3A_4\overline{A_5A_6A_7A_8A_9A_{10}A_{11}A_{12}A_{13}A_{14}A_{15}A_{16}A_{17}A_{18}A_{19}A_{20}A_{21}A_{22}},$$

where each A is one block. If $k \leqslant 5$ all of the elements such as $n-6$ which are less than 1 are deleted. We can check that the string contains all k-element permutations by the same "worst possible permutation" technique that we used to prove Theorem 1. Once again we need only consider those values of k for which block k has only $n-2$ elements. Notice that every block begins with $1, \ldots, n-6$, so these elements can have no role at the end of the worst possible permutation. Therefore, for each value of k that we want to check we will only exhibit the longest suffix of the worst possible permutation consisting of elements from $\{n-5, n-4, n-3, n-2, n-1, n\}$.

Before exhibiting the worst possible suffixes, we observe that the suffixes of the repeating blocks were not constructed haphazardly even if that appears to be the case. For example consider the blocks starting with A_{15}. That block has $n-2$ elements but its predecessor has $n-1$. In the previous section we showed that in that case block A_{15} must be missing the last element and another element of its predecessor, A_{14} which is not the penultimate element; it is easiest to choose the other missing element as the antepenultimate element of A_{14}. In this way we ensure that all the bad permutations for $k \equiv 15 \pmod{18}$ end in $n-5, n, n-1$, since $n-5$ is the only number after $n-1$ in block A_{14} other than n which must appear later in the permutations. We also showed that A_{16} must be missing the last two numbers of A_{15}, and that the penultimate number of block A_{15} must be the last number of block A_{13}. Notice also that the relative order of numbers that appear in consecutive blocks (after block 5) is never changed.

For $k = 4$ the worst possible suffix is $n-2, n$. For $k = 6$ it is $n-2, n-3, n-4, n-1, n$, and one also needs to check permutations ending in $n-2, n-3, n-$

Table 1

k (mod 18)	suffix(es)
0	$n, n-5, n-4, n-3, n-2$ or $n, n-5, n-4, n-1, n-2$ or $n, n-5, n-3, n-1, n-2$ or $n, n-4, n-3, n-1, n-2$ or $n, n-5, n-4, n-3, n-1, n-2$
1	$n-3, n-1, n-4$
3	$n-3, n-1, n, n-5, n-4$ or $n-3, n-1, n, n-2, n-4$ or $n-3, n-1, n-5, n-2, n-4$ or $n-3, n, n-5, n-2, n-4$ or $n-3, n-1, n, n-5, n-2, n-4$
4	$n-5, n-2, n$
6	$n-5, n-2, n-3, n-1, n$ or $n-5, n-2, n-3, n-4, n$ or $n-5, n-2, n-1, n-4, n$ or $n-5, n-3, n-1, n-4, n$ or $n-5, n-2, n-3, n-1, n-4, n$
7	$n-1, n-4, n-3$
9	$n-1, n-4, n-5, n-2, n-3$ or $n-1, n-4, n-5, n, n-3$ or $n-1, n-4, n-2, n, n-3$ or $n-1, n-5, n-2, n, n-3$ or $n-1, n-4, n-5, n-2, n, n-3$
10	$n-2, n, n-5$
12	$n-2, n, n-1, n-4, n-5$ or $n-2, n, n-1, n-3, n-5$ or $n-2, n, n-4, n-3, n-5$ or $n-2, n-1, n-4, n-3, n-5$, or $n-2, n, n-1, n-4, n-3, n-5$
13	$n-4, n-3, n-1$
15	$n-4, n-3, n-2, n, n-1$ or $n-4, n-3, n-2, n-5, n-1$ or $n-4, n-3, n, n-5, n-1$ or $n-4, n-2, n, n-5, n-1$ or $n-4, n-3, n-2, n, n-5, n-1$
16	$n, n-5, n-2$.

1, n. Table 1 lists the suffixes for larger values of k by congruence class. Note that the suffix for $k=6$ is different from the suffix for $k=24, 42, \ldots$ because in the first case the last three blocks are $A_4A_5A_6$ whereas in the other cases the last three blocks are $A_{22}A_5A_6$. For $k=9$ the only change from the general case $k \equiv 9 \pmod{18}$ is that $n-1, n-5, n-2, n, n-3$ and $n-1, n-4, n-5, n-2, n, n-3$ are replaced by $n-4, n-5, n-2, n, n-3$ and $n-4, n-1, n-5, n-2, n, n-3$ because $n-1$ and $n-4$ occur in different orders in A_4 and A_{22}. If $k \pmod{18} \in \{2, 5, 8, 11, 14, 17\}$, then the string ends with a block of size $n-1$, so we don't need to check anything.

These can be checked by the same tedious "worst possible permutation" method we used to prove Theorem 1. As an example we work out the case $k \equiv 12 \pmod{18}$. This corresponds to a string ending with a block of the form A_{12}. That block is missing $n-5$ and $n-4$ which occur in that order in A_{11}. In A_{11}, both $n-3$ and $n-4$ occur after $n-5$, so we need to check permutations ending in $n-3, n-5$ and permutations ending in $n-4, n-5$. Note that there will be multiple bad suffixes only in those cases where block k is missing two numbers which are not the last two numbers of block $k-1$.

First consider the suffix $n-3$, $n-5$. There are two unused numbers in A_{10} after $n-3$, namely $n-1$ and $n-4$. If the permutation ends in $n-1, n-3, n-5$, then the only unused number in A_9 after $n-1$ is n; the only unused number in A_8 after n is $n-2$. There are no unused numbers in A_7 after $n-2$. This corresponds to the entry $n-2, n, n-1, n-3, n-5$ in Table 1. If instead the permutation ends in $n-4, n-3, n-5$, then in A_9 there are two unused numbers after $n-4$, namely $n-1$ and n. If we choose n, then we must use $n-2$ before it, to get the suffix $n-2, n, n-4, n-3, n-5$. If we choose $n-1$ instead of n, we get the suffix $n-2, n-1, n-4, n-3, n-5$ which can be found in A_7 through A_{11} or $n-2, n, n-1, n-4, n-3, n-5$ which can be found in A_6 through A_{11}.

The other bad permutations end in $n-4, n-5$. The only unused number after the $n-4$ in A_{10} is $n-1$, so we can find any permutation that does not end in $n-1, n-4, n-5$. The only unused number in A_9 after the $n-1$ is n. The only unused number in A_8 after the n is $n-2$. Therefore these permutations end in $n-2, n, n-1, n-4, n-5$, the first entry in the corresponding row of the table. There are no unused numbers in A_7 after the $n-2$. This implies that if we can find all permutations lf length $k-i$ in the first $k-i$ blocks for $1 \leqslant i \leqslant 6$, then we can find all permutations of length k in the first k blocks. $\square$

Corollary. *$F(n, k) = k(n-2) + \lfloor \frac{1}{3}(k+1) \rfloor + 3$, for $k > 2$.*

Acknowledgments

Drs. Al Aho, Brian Kernighan, Chris Van Wyk and Professors Donald Knuth and Andrew Yao made helpful comments on earlier drafts.

References

[1] L. Adleman, Short Permutation Strings, Discrete Math. 10 (1974) 197–200.

[2] M.-C. Cai, A new bound on the length of the shortest string containing all r-permutations, Discrete Math. 39 (1982) 329–330.

[3] V. Chvátal, D.A. Klarner and D.E. Knuth, Selected combinatorial research problems, Technical report CS-72-292, Stanford University.

[4] G. Galbiati and F.P. Preparata, On permutation embedding sequences, SIAM J. Appl. Math. 30 (1976) 421–423.

[5] D.J. Kleitman and D.J. Kwiatkowski, A lower bound on the length of a sequence containing all permutations as subsequences, J. of Combin. Theory Ser. A 21 (1976) 129–136.

[6] P.J. Koutas and T.C. Hu, Shortest string containing all permutations, Discrete Math. 11 (1975) 125–132.

[7] M. Newey, Notes on a problem involving permutations as subsequences, Technical Report CS-73-340, Stanford University.

[8] Carla Savage, Short strings containing all k-element permutations, Discrete Math. 42 (1982) 281–285.