

Non-Uniqueness of Minimal Superpermutations

Nathaniel Johnston^{a,b}

^a*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

^b*Department of Mathematics & Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada*

Abstract

We examine the open problem of finding the shortest string that contains each of the $n!$ permutations of n symbols as contiguous substrings (i.e., the shortest superpermutation on n symbols). It has been conjectured that the shortest superpermutation has length $\sum_{k=1}^n k!$ and that this string is unique up to relabelling of the symbols. We provide a construction of short superpermutations that shows that, if the conjectured minimal length is true, then uniqueness fails for all $n \geq 5$. Furthermore, uniqueness fails spectacularly; we construct more than doubly-exponentially many distinct superpermutations of the conjectured minimal length.

Keywords: superpermutation, permutation, shortest superstring

1. Introduction

The shortest superstring problem asks for the shortest string containing each string s_i in a given set $S := \{s_1, s_2, \dots, s_m\}$ as a contiguous substring. The shortest superstring problem is NP-hard [1] in general, and finding algorithms for computing or approximating shortest superstrings is an active area of research [2, 3].

In certain special cases, the shortest superstring problem simplifies considerably. Perhaps most well-known of these cases arises when the set of substrings S consists of all words of length k on n symbols, where k and n are given positive integers. There are n^k such words, so there is a trivial lower bound of length $n^k + k - 1$ on any string containing them all as substrings. More interestingly, however, it has been shown [4] that this length is always attainable. A string containing all such words that attains this minimal length is called a *De Bruijn sequence*.

In the present paper, we consider the related problem that arises when the set of substrings S contains all permutations of n symbols. That is, we are interested in the problem that asks for the shortest string on the n symbols $[n] := \{1, 2, 3, \dots, n\}$ that contains all $n!$ permutations of those symbols as contiguous substrings. Following the terminology of [5], we call a string containing all permutations in this way a *superpermutation*, and we call a superpermutation of minimum length a *minimal superpermutation*.

Email address: nathaniel.johnston@uwaterloo.ca (Nathaniel Johnston)

Similar to the De Bruijn sequence case, there are $n!$ permutations of n symbols, so a trivial lower bound on the length of a minimal superpermutation is $n! + n - 1$. Unlike in the De Bruijn sequence case, however, this bound is not tight when $n \geq 3$. For $n = 2$ a minimal superpermutation is trivially 121, which contains both permutations 12 and 21 as substrings. For $n = 3$ it is not difficult to show that a shortest string containing each of 123, 132, 213, 231, 312, and 321 as contiguous substrings is 123121321. For small values of n , the optimal solutions have been computed via brute-force computer search [6] — Table 1 summarizes the shortest strings for $1 \leq n \leq 4$.

n	Minimal Superpermutation	Length
1	1	1
2	121	3
3	123121321	9
4	123412314231243121342132413214321	33

Table 1: The shortest strings containing every permutation of $\{1, 2, \dots, n\}$ for $1 \leq n \leq 4$.

Computing minimal superpermutations for $n \geq 5$ isn't feasible via brute-force, but the solutions for $n \leq 4$ provide some structure that is easily generalized. For example, there is a natural recursive construction that can be used to create small superpermutations on n symbols for any n (more specifically, superpermutations of length $\sum_{k=1}^n k!$), and this procedure produces the superpermutations provided in Table 1 when $n \leq 4$. This method of constructing small superpermutations has been rediscovered independently in several different forms [5, 7, 8, 9, 10] and it gives, for all n , what are currently the only known superpermutations of length at most $\sum_{k=1}^n k!$. Furthermore, the superpermutations provided in Table 1 have been shown by brute force to be unique up to relabelling the symbols (e.g., the only minimal superpermutations in the $n = 2$ case are 121 and 212, which are related to each other by swapping the roles of 1 and 2).

These observations have led to the following conjectures:

Conjecture 1 ([5, 6, 7, 8, 9, 11]). *The length of the minimal superpermutation on n symbols is $\sum_{k=1}^n k!$.*

Conjecture 2 ([5]). *The minimal superpermutation on n symbols is unique up to interchanging the roles of the symbols.*

In this work we generalize the well-known method of construction of small superpermutations in a new way. Our construction shows that, for all $n \geq 5$, at least one of these two conjectures fails. It is known that Conjecture 1 has to at least be “almost” true in the sense that the conjectured length can't be too far from the true minimal length [5, Theorem 18]. On the other hand, our construction shows that if Conjecture 1 is true, then Conjecture 2 fails spectacularly as n increases. For example, our construction in the $n = 8$ case provides more than 3×10^{50} distinct superpermutations of the conjectured minimal length.

Before proceeding, we present the notational conventions and representations of permutations that we will use throughout this paper. In order to avoid double-counting superpermutations that are related to each other by interchanging the roles of the symbols, we only consider superpermutations that begin with the substring $12 \cdots n$. We typically represent a permutation $\sigma \in S_n$ via its one-line representation $\sigma(1)\sigma(2) \cdots \sigma(n)$, which we think of as a string (for example, the identity permutation corresponds to the string $12 \cdots n$).

It is also sometimes useful for us to represent a permutation σ by another encoding that we now describe. Define $p_k \in S_n$ to be the permutation that cyclically increases the first k positive integers by one and acts as the identity otherwise: $p_k(i) = i + 1$ if $1 \leq i < k$, $p_k(i) = i$ if $k < i \leq n$, and $p_k(k) = 1$. Then p_k^j cyclically increases the first k positive integers by j and acts as the identity otherwise, and for every permutation $\sigma \in S_n$, there is a unique tuple of exponents (j_2, j_3, \dots, j_n) with $0 \leq j_i < i$ for all i so that $\sigma = p_2^{j_2} \circ p_3^{j_3} \circ \cdots \circ p_n^{j_n}$. To write σ in this way, first choose j_n so that the correct integer is mapped to n , then choose j_{n-1} so that the correct integer is mapped to $n - 1$, and so on. To see uniqueness, we simply note that there are exactly $n!$ tuples (j_2, j_3, \dots, j_n) satisfying $0 \leq j_i < i$ for all i , which is the same as the size of S_n .

Rather than explicitly writing a permutation in the form $p_2^{j_2} \circ p_3^{j_3} \circ \cdots \circ p_n^{j_n}$, we simply write the exponents in order within square brackets and a subscript c , as in $[j_2 j_3 \cdots j_n]_c$. We call this representation the *circular shift representation*. For example, both 42351 (one-line notation) and $[0121]_c$ (circular shift representation) represent the same permutation.

As perhaps a slightly more intuitive way of converting a permutation's circular shift representation to its one-line representation, start with the string $12 \cdots n$. Cycle the first 2 characters of the string to the left the number of times indicated by j_2 , then cycle the first 3 characters of the string to the left the number of times indicated by j_3 , and so on. So to obtain the one-line representation of $[0121]_c$, we start with 12345 and transform it as follows:

$$12345 \xrightarrow{0} 12345 \xrightarrow{1} 23145 \xrightarrow{2} 14235 \xrightarrow{1} 42351.$$

2. Construction of a Small Superpermutation

In this section we briefly present a well-known method for constructing a family of strings $\{M_1, M_2, M_3, \dots\}$ with the property that M_n is a superpermutation on $[n]$ of length $|M_n| = \sum_{k=1}^n k!$.

First, let $M_1 = 1$, the string with just one symbol. Then we construct M_{n+1} from M_n via the following procedure:

- (a) Let P_j ($0 \leq j < n!$) be the j th permutation of $[n]$ that appears as a substring of M_n .
- (b) Let Q_j ($0 \leq j < n!$) be the string of length $2n + 1$ obtained by concatenating the string P_j , the symbol $n + 1$, and again the string P_j , in that order.
- (c) Concatenate $Q_0, Q_1, \dots, Q_{n!-1}$, in that order, overlapping as much as possible each consecutive pair of strings.

To demonstrate this procedure in the $n = 2$ case, start with $M_2 = 121$. Then $P_0 = 12$ and $P_1 = 21$, so $Q_0 = 12312$ and $Q_1 = 21321$. If we concatenate Q_0 and Q_1 , overlapping as much as possible, we get $M_3 = 123121321$, where we overlapped the 2 in the middle of the string. The procedure is illustrated in the $n = 3$ case in Figure 1.

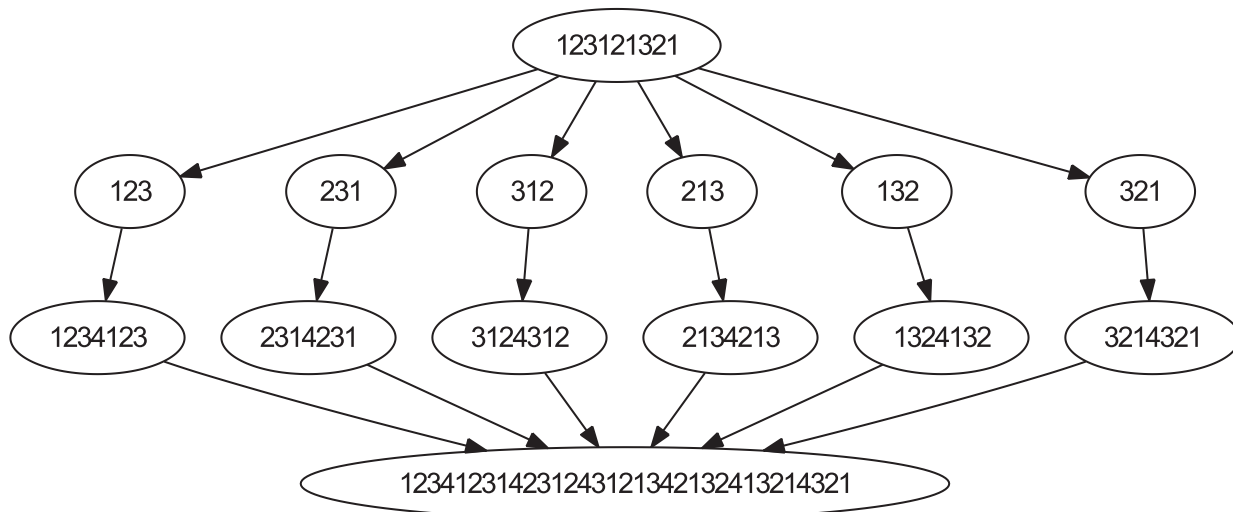


Figure 1: A tree that demonstrates how to construct M_4 from M_3 .

The fact that each M_n is a superpermutation on $[n]$ with $|M_n| = \sum_{k=1}^n k!$ follows easily via induction, so we don't present the argument here. The intuition behind why these superpermutations are so efficient is simple — they are the “greedy” superpermutations obtained by starting with the string $12 \cdots n$ and repeatedly appending as few symbols as possible so as to add a new permutation as a contiguous substring. However, minimality and uniqueness have only been proved when $n \leq 4$.

It is interesting to observe the circular shift representation of all permutations in the order that they appear in M_n (and indeed, this observation is our primary reason for introducing circular shift representations in the first place). For example, when $n = 3$, the permutations appear in M_n in the order 123, 231, 312, 213, 132, 321; these have circular shift representations $[00]_c$, $[01]_c$, $[02]_c$, $[10]_c$, $[11]_c$, and $[12]_c$, respectively. In other words, the order of the permutations in M_3 is the same as the order we get if we “count” in the circular shift representation. It turns out that this is true in general, as demonstrated by the following proposition.

Proposition 3. *Let n and j be positive integers with $0 \leq j < n!$. If we write*

$$j = \sum_{i=2}^n j_i \cdot n!/i!$$

with $0 \leq j_i < i$ for all i then the $(j + 1)$ th permutation to appear in M_n as a contiguous substring is $[j_2 j_3 \cdots j_n]_c$.

Before proving Proposition 3, we note that every integer $0 \leq j < n!$ can be written (uniquely) in the desired form — this is just a slight variant on the factorial number system [12] in which the digits are allowed to increase from left to right, rather than from right to left.

Proof of Proposition 3. We prove the result by induction. We already noted that the result is true when $n = 3$, and it is also easily verified when $n = 1$ or $n = 2$. We thus move directly to the inductive step.

Suppose that the result holds for a given value of n . Then in the construction of M_{n+1} given earlier, P_j is the $(j + 1)$ th permutation of n symbols that appears in M_n , so $P_j = [j_2 j_3 \cdots j_n]_c$ by the inductive hypothesis. Then Q_j contains $n + 1$ permutations of $[n + 1]$ as substrings, the first of which is $[j_2 j_3 \cdots j_n 0]_c$. The second permutation contained as a substring of Q_j is just a cyclic permutation of the first: $[j_2 j_3 \cdots j_n 0]_c \circ p_{n+1} = [j_2 j_3 \cdots j_n 1]_c$. Similarly, the third permutation that appears as a substring of Q_j is $[j_2 j_3 \cdots j_n 1]_c \circ p_{n+1} = [j_2 j_3 \cdots j_n 2]_c$, and so on up to $[j_2 j_3 \cdots j_n n]_c$. The result follows. \square

3. Non-Uniqueness When $n = 5$

In Section 4 we present our generalized construction of small superpermutations for arbitrary n , but for now we simply present and motivate it in the first non-trivial case (i.e., when $n = 5$) for clarity.

To start, notice that the first 60 permutations to appear as substrings of M_5 are those of the form $[0xyz]_c$ for some $0 \leq x < 3$, $0 \leq y < 4$, and $0 \leq z < 5$, while the last 60 permutations to appear as substrings of M_5 are those of the form $[1xyz]_c$. Phrased differently, the first 60 permutations are those in which the symbols 1 and 2 have not been transposed. In other words, when written in one-line notation, if the 4 and 5 are erased, we are left with 123, 231, or 312. Similarly, the last 60 permutations to appear in M_5 are those that result in 213, 132, or 321 when the 4 and 5 are removed from their one-line representation. Since this property has nothing to do with the order of 4 and 5 in the permutation's one-line representation, we conclude that a permutation is one of the first (last) 60 permutations to appear as a substring of M_5 if and only if interchanging the roles of 4 and 5 in its one-line representation results again in one of the first (last) 60 permutations to appear in M_5 .

This leads immediately to the key idea: we can just replace all 4's by 5's (and vice-versa) in the last half of M_5 to generate a new string M'_5 that is of the same length and is still a superpermutation. These two strings are written in full in Table 2, and it is easily verified by computer that they are indeed distinct superpermutations of length $\sum_{k=1}^5 k! = 153$.

4. Generalized Construction

We now show how the idea from Section 3 extends to the $n \geq 6$ case. While the argument only provides two distinct superpermutations of the conjectured minimal length in the $n = 5$ case, our main result shows that the number of small superpermutations produced grows extremely quickly as n increases.

M_5	1234512341523412534123541231452314253142351423154231· ·245312435124315243125431 2 134521342513421534213542· ·1324513241532413524132541321453214352143251432154321
M'_5	1234512341523412534123541231452314253142351423154231· ·245312435124315243125431 2 135421352413521435213452· ·1325413251432513425132451321543215342153241532145321

Table 2: Two distinct superpermutations on $[5]$ of length $\sum_{k=1}^5 k! = 153$. The second string is obtained from the first simply by interchanging the roles of the symbols 4 and 5 in the second half of the string (i.e., after the bold **2**).

Theorem 4. *There are at least*

$$\prod_{k=1}^{n-4} (n-k-2)!^{k \cdot k!}. \quad (1)$$

distinct (up to relabelling of the symbols) superpermutations on $[n]$ of length $\sum_{k=1}^n k!$.

Before proving the theorem, note that the formula (1) gives the empty product when $n \leq 4$, which agrees with the previously-mentioned fact that the minimal superpermutation is unique in these cases. For $n = 5, 6, 7, 8$ it gives the values 2, 96, 8153726976, and approximately 3×10^{50} respectively. A text file containing all 96 small superpermutations described by the theorem in the $n = 6$ case can be downloaded from [13]. A text file containing all such superpermutations in the $n = 7$ case would be larger than 43 terabytes in size and is thus omitted.

The following lemma is our main tool in the proof of the theorem. Before presenting the lemma, we note that each permutation appears in M_n exactly once [5, Lemma 4]. It follows that the string $T_{j,k}$ defined in the statement of the lemma is unique.

Lemma 5. *Let $2 \leq k < n$ be integers. For each $0 \leq j < k!$, let $T_{j,k}$ be the shortest substring of M_n that contains each of the $(j \cdot n!/k! + 1)$ th through $((j+1) \cdot n!/k!)$ th permutations to appear in M_n as substrings. Then for all j ,*

- (a) *there exists $1 \leq \ell < k$ such that the last ℓ characters of $T_{j,k}$, in order, are the same as the first ℓ characters of $T_{j+1,k}$;*
- (b) *the first and last $k+1$ characters of $T_{j,k}$ are $1, 2, \dots, k$, and $k+1$ (not necessarily in that order); and*
- (c) *the set of permutations that are substrings of $T_{j,k}$ remains unchanged under any interchange of the roles of the symbols $k+2, k+3, \dots, n$ in $T_{j,k}$.*

Proof. We first prove properties (a) and (b) by induction. Both of these properties are clearly true by inspection in the $n = 3$ case. For the inductive step, assume that the result holds for M_n . If we fix k then properties (a) and (b) clearly still hold for M_{n+1} by the

construction given in Section 2. To see that properties (a) and (b) also hold for M_{n+1} when $k = n$, simply note that we have $T_{j,k} = Q_j$ in this case, and Q_j clearly has both properties.

We now prove property (c) (directly; not via induction). It is a direct consequence of Proposition 3 that there exist fixed j_2, j_3, \dots, j_k such that $T_{j,k}$ contains as substrings exactly the permutations whose cyclic shift representations are of the form $[j_2 j_3 \cdots j_n]_c$, where j_i is free to vary from 0 to $i - 1$ when $k + 1 \leq i \leq n$. It follows that there exists a fixed permutation τ_j on $[k + 1]$ such that $\sigma \in S_n$ is a substring of $T_{j,k}$ if and only if deleting the symbols $k + 2, k + 3, \dots, n$ from its one-line representation results in a cyclic permutation of τ_j . Since this criterion for being a substring of $T_{j,k}$ is clearly independent of the locations of the symbols $k + 2, k + 3, \dots, n$ in the one-line representation of σ , condition (c) follows. \square

Proof of Theorem 4. Our procedure is to recursively construct the given number of small superpermutations from the well-known superpermutation M_n . Observe that the procedure that we are about to describe reduces exactly to the procedure described in Section 3 when $n = 5$.

- (1) Let $k = n - 3$ and define $SP := \{M_n\}$ to be the set of all superpermutations of the desired length that we have constructed so far.
- (2) For each $T \in SP$ and each $0 \leq j < k!$, let $T_{j,k}$ be as in Lemma 5. By the lemma, we can interchange the roles of $k + 2, k + 3, \dots, n$ within the section of T that equals $T_{j,k}$ without affecting whether or not T is a superpermutation. We could do this for each value of j from 0 to $k! - 1$, however we only do it for the $k! - k$ values of j that are not multiples of $(k - 1)!$ (we will explain the reason for this shortly). Add all of the superpermutations constructed in this way to SP . The cardinality of SP is thus multiplied by $(n - k - 1)!^{k! - k}$ during this step.
- (3) Decrease k by 1.
- (4) If $k = 1$, stop. Otherwise, return to step (2).

It is perhaps not clear in step (2) why we only consider the values of j that are not divisible by $(k - 1)!$. The reason why we avoid $j = 0$ is straightforward: we want all of our superpermutations to start with the substring $12 \cdots n$. The reason why we avoid all other values of j that are divisible by $(k - 1)!$ is that if we swapped the roles of $k + 2, k + 3, \dots, n$ in the same way for each j from $i \cdot (k - 1)!$ to $(i + 1) \cdot (k - 1)! - 1$ for some i , then we would end up with a superpermutation that is the same as one produced in the next iteration of the algorithm (i.e., after we decreased k by 1). We would thus end up double-counting some superpermutations. As is, in each iteration of the algorithm we alter part of each string T that has never been altered before, so we can be sure that each step in the procedure really does produce superpermutations that are not already in the set SP .

To get the cardinality of the set SP after the algorithm terminates (i.e., the number of distinct superpermutations that we have constructed), we simply recall that its cardinality

is increased by a factor of $(n - k - 1)!^{k! - k}$ for each value of k from $n - 3$ down to 2. That is, the number of distinct superpermutations that we have constructed is

$$\prod_{k=2}^{n-3} (n - k - 1)!^{k! - k},$$

which can be seen to equal the value given in the statement of the theorem by shifting the value of k by 1. \square

5. Conclusions

We have shown that the standard construction of small superpermutations generalizes in a natural way that disproves at least one of two long-standing conjectures. While it is possible that Conjecture 1 is false and Conjecture 2 is true, we consider this possibility rather unlikely, as Conjecture 1 being false would seem to indicate that minimal superpermutations in general do not have as much structure as suggested by the $n \leq 4$ cases. We thus expect that either both conjectures fail, or Conjecture 1 is true and Conjecture 2 is false.

Our results also shed some light on approaches to proving Conjecture 1 that will *not* work. Upon seeing the abundance of symmetry in the superpermutations in the $1 \leq n \leq 4$ cases, it is tempting to try to prove that minimal superpermutations must have certain properties that the M_n have. For example, one could try to prove that a minimal superpermutation on $[n]$ must be a palindrome or must contain exactly $(n - 1)!$ occurrences of the symbol n . However, our generalized construction shows that minimal superpermutations need not have either of these properties.

While computing the length of minimal superpermutations is a fundamental combinatorial problem, and one that can be understood by anyone with even a modest mathematical background, it seems to have received relatively little attention. We hope that our results spark further interest.

Acknowledgements. Thanks are extended to Dan Ashlock for helpful conversations. The author was supported by the University of Guelph Brock Scholarship, the Natural Sciences and Engineering Research Council of Canada, and the Mprime Network.

References

- [1] J. Gallant, D. Maier, and J. A. Storer. On finding minimal length superstrings. *Journal of Computer and System Sciences*, 20:50–58, 1980.
- [2] M. K. Goldberg and D. T. Lim. A learning algorithm for the shortest superstring problem. In *Proceedings of the Atlantic Symposium on Computational Biology and Genome Information Systems and Technology*, pages 171–175, 2001.
- [3] H. Kaplan and N. Shafir. The greedy algorithm for shortest superstrings. *Information Processing Letters*, 93(1):13–17, 2005.
- [4] T. van Aardenne-Ehrenfest and N. G. de Bruijn. Circuits and trees in oriented linear graphs. *Simon Stevin*, 28:203–217, 1951.
- [5] D. Ashlock and J. Tillotson. Construction of small superpermutations and minimal injective superstrings. *Congressus Numerantium*, 93:91–98, 1993.

- [6] M. Hamm. Sequence A180632. *The On-Line Encyclopedia of Integer Sequences*, Sep. 13, 2010, retrieved Oct. 21, 2012. <http://oeis.org/A180632>
- [7] J. A. Barnett. *Permutation Strings*, retrieved Oct. 21, 2012. <http://www.notatt.com/permutations.pdf>
- [8] J. Orendorff. Generate sequence with all permutations. *StackOverflow*, Feb. 16, 2010, retrieved Oct. 21, 2012. <http://stackoverflow.com/a/2274978>
- [9] Y. Filmus. What is the shortest string that contains all permutations of an alphabet? *Mathematics - Stack Exchange*, Dec. 25, 2010, retrieved Oct. 21, 2012. <http://math.stackexchange.com/a/15512>
- [10] Kolistivra. Turkish Informatics Olympiad problem. *TopCoder Forums*, Nov. 20, 2007, retrieved Oct. 21, 2012. <http://apps.topcoder.com/forums/?module=Thread&threadID=594178>
- [11] J. A. Barnett. Re: Integer Strings Containing All Possible Length 4 Sub-Strings. *The Math Forum*, Aug. 30, 2001, retrieved Oct. 21, 2012. <http://mathforum.org/kb/message.jspa?messageID=1689849>
- [12] D. E. Knuth. Volume 3: Sorting and Searching. *The Art of Computer Programming* (3rd ed.), Addison–Wesley, pp. 209, 1973.
- [13] N. Johnston. *96 Small Superpermutations*. Published electronically at <http://www.njohnston.ca/superperm6.txt>, 2012.