Permutations resilient to deletions

Noga Alon* Steve Butler† Ron Graham‡ Utkrisht C. Rajkumar§

May 21, 2017

Abstract

Let σ be a permutation on $[n] = \{1, 2, \ldots, n\}$ which can be written in two-line notation, and let $\varphi : [n] \to S$ be a bijection. Construct τ (resp. β) by replacing the elements in σ as dictated by φ and then deleting up to d elements in the top (resp. bottom) line and contracting the result, making sure no symbol is deleted in both lines. The permutation σ is d-resilient if τ and β *always* uniquely determine φ (or equivalently, determine where the deletions in the top and bottom lines occurred).

Necessary and sufficient conditions for a permutation to be d-resilient are established in terms of whether a family of auxiliary graphs are acyclic. Also, constructions are given for d-resilient permutations which have size n exponential in d, this is best possible. It is further shown that for every fixed d and sufficiently large n a positive portion of all permutations of n elements are d-resilient.

1 Introduction

Let σ be a permutation on $[n] = \{1, 2, \ldots, n\}$ written in two-line notation. We consider the problem of whether a bijection $\varphi : [n] \to S$, where S is an arbitrary set of n elements, can be uniquely determined if we are given *partial* information about the permutation expressed using symbols from S.

More precisely we replace the entries in σ by the corresponding elements of S as dictated by φ, then for each of the top and bottom rows we delete up to d symbols (with no symbol being deleted in both rows), and contract the result (to remove indications of where the deletions occurred). This produces two lists: τ and β (for the top and bottom respectively). Given σ, τ, and β, can we uniquely determine φ? Equivalently, given σ, τ, and β, can we uniquely determine the location of the deletions in σ that produced τ and β?

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel. nogaa@tau.ac.il Research supported in part by a BSF grant, an ISF grant and a GIF grant.

†Dept. of Mathematics, Iowa State University, Ames, IA 50011 USA. butler@iastate.edu. Partially supported by a grant from the Simons Foundation (#427264, Steve Butler).

‡Dept. of Computer Science and Engineering, UC San Diego, La Jolla, CA 92093 USA. graham@ucsd.edu

§Dept. of Computer Science and Engineering, UC San Diego, La Jolla, CA 92093 USA. urajkuma@eng.ucsd.edu

Example 1. Let $S = \{A, B, C, \ldots, I\}$ and consider

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 7 & 4 & 1 & 8 & 5 & 2 & 9 & 6 & 3 \end{pmatrix},$$

$\tau = BDEFGHI$, $\beta = GDAHBFC$. Then there is a unique φ, and we have

$$\sigma_S = \begin{pmatrix} \boxed{A}\, B\, \boxed{C}\, D\ E\ F\, G\, H\, I \\ G\ D\ A\ H\, \boxed{E}\, B\, \boxed{I}\, F\, C \end{pmatrix},$$

where σ_S is the permutation σ using elements from S and the location of the deletions are boxed. Note that either line of σ_S can be used to give φ.

Example 2. Let $S = \{A, B, C, \ldots, I\}$ and consider

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 7 & 4 & 1 & 8 & 5 & 2 & 9 & 6 & 3 \end{pmatrix},$$

$\tau = ABCEGHI$, and $\beta = DAHEBIF$. Then there are two possible φ, and we have

$$\sigma_S = \begin{pmatrix} A\ B\, C\, \boxed{D}\, E\, \boxed{F}\, G\, H\ I \\ \boxed{G}\, D\ A\ H\ E\ B\ I\ F\, \boxed{C} \end{pmatrix}, \text{ or } \sigma_S = \begin{pmatrix} A\ B\ \boxed{F}\, C\, E\, G\, \boxed{D}\, H\ I \\ D\, \boxed{C}\ A\ H\, E\, B\ I\ \boxed{G}\, F \end{pmatrix}.$$

Definition 1. We say a permutation σ is *d-resilient* if for any choice of up to d deletions in both the top and bottom rows with no symbol being deleted in both rows, then τ and β are enough to uniquely determine φ.

Note that the permutation used in Examples 1 and 2 is not 2-resilient as there exists a τ and β which does not uniquely determine φ.

The goal of this paper is to give a necessary and sufficient condition for a permutation σ to be d-resilient expressed in terms of a family of auxiliary graphs being acyclic (see Section 2). We also give a construction of d-resilient permutations which have size n exponential in d, and show that this is best possible. Moreover we show that for every fixed d and large n a positive portion of the permutations of n elements are d-resilient (see Section 3).

Comment. This problem can also be phrased in terms of a deletion channel where sent messages have portions deleted and then contracted before delivery (see [4]). A d-resilient permutation can be used to help detect and correct errors in the case when a message M of distinct symbols and $\sigma(M)$ (the message M with entries permuted as dictated by σ) are sent through a deletion channel and at most d deletions occur in each of M and $\sigma(M)$ before delivery and each entry occurs in at least one of M and $\sigma(M)$ (after the deletion). Due to the size of d-resilient permutations they are unlikely to find direct applications in the deletion channel problem.

Comment. This solution was inspired in part by the oral transmission protocols for Sanskrit literature in the Vedic period. This relied on interleaving patterns of words to combat transpositions, substitutions, insertions, and deletions of words.

2 A necessary and sufficient condition to be d-resilient

We will find it informative to work through the details for Example 2. We see that the symbols C, D, F, and G each only occur once and so were deleted when either τ or β was produced. The symbols A, B, E, H, and I are doubly-occurring, i.e., are both in τ and β. It is an easy exercise to show that for this permutation the identification of the doubly-occurring symbols can be determined (i.e., φ is known for the subset restricted to the doubly occurring symbols).

Given that we know the doubly occurring symbols, we have the situation illustrated in Figure 1 where we have marked with a line how σ connects the entries involved in a deletion.

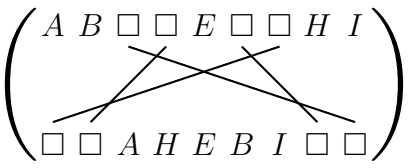

Figure 1: Intermediate step in working to recover σ_S.

The entries marked as □ in Figure 1 indicate that this is a location of a symbol that occurs only once. Examining the location of these we see that there are four blocks of contiguous □'s (two in the top line and two in the bottom line). By examining τ and β we can conclude that each one of these blocks has one entry which was deleted and the other was kept.

If we could uniquely determine which entries in all blocks were deleted and which were kept we could recover our message (i.e., we simply use the connections between the two lines to fill any gaps). However, it might be that there is more than one possibility to which entries in the blocks were deleted and which were kept.

We are in the latter case in that there are two ways in which entries could be deleted or kept, as shown in Figure 2. Here we have oriented the edges from where a symbol was deleted (marked with a "$*$") in one of the top or bottom lines to where it was kept in one of the bottom or top lines. This allows us to quickly determine the two possible σ_S.

Figure 2: Ambiguity found in working to recover σ_S.

This example highlights the key idea in why we might be unable to uniquely recover φ. Namely there is ambiguity in large blocks of □'s as to which specific entries were deleted and which were kept. When there is an ambiguity we *might* be able to propagate a change in σ_S through the blocks in a consistent manner to produce multiple possible σ_S.

To generalize what happens, we introduce an auxiliary (multi-)graph to the problem. Given a permutation σ on $[n] = \{1, 2, \ldots, n\}$ and $D \subseteq [n]$ (representing the location of indices involved in deletions), let $G(\sigma, D)$ be a bipartite (multi-)graph defined in the following way. Let $t_1, t_2, \ldots, t_i$ be the maximal contiguous blocks of elements of D in the top line of σ and let $b_1, b_2, \ldots, b_j$ be the maximal contiguous blocks of elements of D in the bottom line of σ. We now let

$$V\big(G(\sigma, D)\big) = \{t_1, t_2, \ldots, t_i, b_1, b_2, \ldots, b_j\}$$

and we add $|t_k \cap b_\ell|$ edges joining t_k and b_ℓ for all k and ℓ. Note that $G(\sigma, D)$ will have $|D|$ edges (i.e., one edge for each element in D).[1]

Since every edge in $G(\sigma, D)$ can be identified with an element of D, we can indicate whether a symbol is deleted in the top or bottom line by orienting the edge *away* from where the deletion occurs.

Returning to our example shown in Figure 1 we have $D = \{3, 4, 6, 7\}$, $t_1 = (3, 4)$, $t_2 = (6, 7)$, $b_1 = (7, 4)$, $b_2 = (6, 3)$, and $G(\sigma, D)$ is a (simple) 4-cycle. We can interpret the situation shown in Figure 2 as coming from two different cyclic orientations of the 4-cycle.

Theorem 1. *If $|D| \leq 2d$ and $G(\sigma, D)$ has a cycle, then σ is not d-resilient.*

Proof. To show σ is not d-resilient we only need to find a τ and β that can produce more than one valid σ_S. Start by fixing a cycle in $G(\sigma, D)$ and orient the edges $G(\sigma, D)$ in the following manner.

- First orient the edges of the cycle to produce in-degree and out-degree one at each vertex of the cycle.
- Orient the *remaining* edges so that there are at most d edges directed into the $\{t_k\}$ collectively and at most d edges directed into the $\{b_\ell\}$ collectively. (This can be done, for example, by always orienting towards the $\{t_k\}$ until there are d edges directed into them, and then orienting all the remaining edges towards the $\{b_\ell\}$.)

Call this orientation H_1. Let H_2 be the orientation found by starting with H_1 and reversing all the edges of the cycle. The orientations will indicate how to delete entries in forming the τ and β, i.e., we delete the entries which correspond to the tails of the directed arcs (here remembering that each arc is associated with a unique entry in the top row and unique entry in the bottom row). By our assumptions on the size of D and the choice of orientations, we will delete at most d symbols in each of the rows.

By construction we note that H_1 and H_2 have the same in- and out-degree at each vertex, i.e., we will consistently delete/keep the same number of symbols in each block. However for any vertex where the cycle passed through the location of the deletions will be slightly different between H_1 and H_2 (this is what leads to the ambiguity!).

Let φ be a bijection and σ_S the permutation according to φ. We now construct a second valid way to write σ_S by doing the following.

[1] This process is similar to what is done to construct random graphs with prescribed degrees, namely we have a matching (the connections joining indices in the top and bottom lines), and then we group a cluster of endpoints together to form a vertex. Here our clusters are defined by the contiguous blocks.

1. Place the edges of the orientation of H_1 between the two lines and for any element that has an edge oriented out replace the symbol with a $*$.

2. Replace the orientation H_1 with the orientation H_2.

3. For each block t_k and b_ℓ, move the non-$*$ entries to correspond to the vertices with an edge directed in; while preserving the relative order of the entries. The entries with edges directed out will now obtain a $*$.

4. Replace any symbol with a $*$ by using the edges of the orientation, i.e., with what it connects with.

The key step is the third step, because we have guaranteed two things to happen. First we have changed the orientation of at least one edge (from the cycle) and thus the location of at least one entry has changed (i.e., this is a different bijection, φ'). Second if we delete the original representation by H_1 and the new representation by H_2 then they will produce the same τ and β, giving us more than one valid σ_S for the same τ and β. □

So the existence of a cycle in $G(\sigma, D)$ with $|D| \leq 2d$ can lead to ambiguity. We next show that this is essentially the only possible way to have an ambiguity.

Theorem 2. *If for all $|D| \leq 2d$ the graph $G(\sigma, D)$ is acyclic, then σ is d-resilient.*

Proof. First we demonstrate that from τ and β we can determine the location of all doubly occurring symbols in σ.

Suppose that the symbol x occurs in τ in position q. Then x must be in one of positions q, $q+1$, $\ldots$, $q+d$ in the top line of σ_S (i.e., it could move down by at most d entries); which in turn gives that the location of x is in one of $d+1$ possible positions in the bottom line of σ_S. It now suffices to show that these positions are pairwise distance more than d apart, since then the positions are associated with non-overlapping portions of β from which we can determine the location of x.

So suppose that some pair of positions are pairwise distance d or less apart. Then there are a pair of symbols y and z so that the distance between them in both the top and bottom lines is at most distance d. Now form the set D by taking y, z, and all elements between y and z in both the top and bottom lines. This has size $|D| \leq 2d$ and the vertices y and z are both in the same block on the top and bottom and thus $G(\sigma, D)$ would have a two-cycle, which contradicts our assumption.

Since we now know the locations of all the doubly occurring symbols, we also know the locations of entries involved in deletion, i.e., there is a unique D associated with τ and β. We also know that $G(\sigma, D)$ does not contain a cycle. We now observe that if there were two (or more) possible σ_S, then they would have to correspond to two distinct orientations, say H_1 and H_2, of the edges of $G(\sigma, D)$, and moreover that the orientations would have the same in- and out-degrees at each vertex. (This last statement follows from noting that we know how many symbols were in these blocks from τ and β and also the length of the blocks.)

So suppose there were two possible σ_S and let e_1 be any edge of H_2 which has a different orientation from H_1. Now going into the vertex it is oriented towards there must be some other edge that initially was oriented into the vertex in H_1 but is now oriented out, call that

edge e_2. Now we can repeat this procedure finding e_1, e_2, e_3, ..., until we eventually come across an edge which goes into a previously seen vertex by this procedure. But at such a point we have a directed cycle in H_2, and more importantly a cycle in $G(\sigma, D)$, which is impossible. So there can only be one σ_S. Since this is true for any τ and β we have σ is d-resilient. □

3 Construction of d-resilient permutations

By Theorems 1 and 2 it is easy to show that σ is 1-resilient if and only if the permutation does not map adjacent entries to adjacent entries. This first happens with $n = 4$, for example

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

The smallest possible 2-resilient permutations have length 18, for example

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 6 & 16 & 9 & 1 & 5 & 11 & 15 & 2 & 7 & 12 & 17 & 4 & 8 & 14 & 18 & 10 & 3 & 13 \end{pmatrix}. \tag{1}$$

It is not immediately obvious that d-resilient permutations exist for larger d, however the following result shows that not only do they exist, there is also an efficient procedure to produce a d-resilient permutation.

Theorem 3. *For any n and d satisfying $n > 3^{2d}$ there is a d-resilient permutation σ of $[n]$. Such a σ can be found by a polynomial time algorithm (in n).*

Call a graph H an (n, d)-double path graph if it has n vertices, its edge set is a union of two Hamiltonian paths, and its girth is at least $2d + 1$. Given such a graph, number its vertices by the integers $1, 2, 3, \ldots, n$ according to the order of the first Hamiltonian path (corresponding to the top row of the two-line representation of σ), and the ordering of the second Hamiltonian path will then correspond with the bottom row of the two-line representation of σ. As an example the permutation in (1) produces the graph shown in Figure 3.

Lemma 4. *For any (n, d)-double path graph H the corresponding permutation is d-resilient.*

Proof. If there is a $D \subset [n]$ so that $G(\sigma, D)$ contains a cycle, then so does the induced subgraph of H on D. Since H has girth $2d + 1$ then $G(\sigma, D)$ is acyclic for all $|D| \leq 2d$ and so by Theorem 2 we have σ is d-resilient. □

Lemma 5. *If $n > 3^{2d}$ then there is an (n, d)-double path graph H.*

Proof. We apply a variant of the method of Erdős and Sachs [2]. Starting with a graph H on the set of vertices $[n]$ with the edge set being the union of the Hamiltonian path $1, 2, \ldots, n$ (in this order) and another Hamiltonian path P, we keep modifying P as long as there is a cycle of length at most $2d$ in H. We show how to perform these modifications in order to get rid of all cycles of length at most $2d$ keeping the first Hamiltonian path and maintaining the property that the second one, P, also stays a Hamiltonian path. In each modification

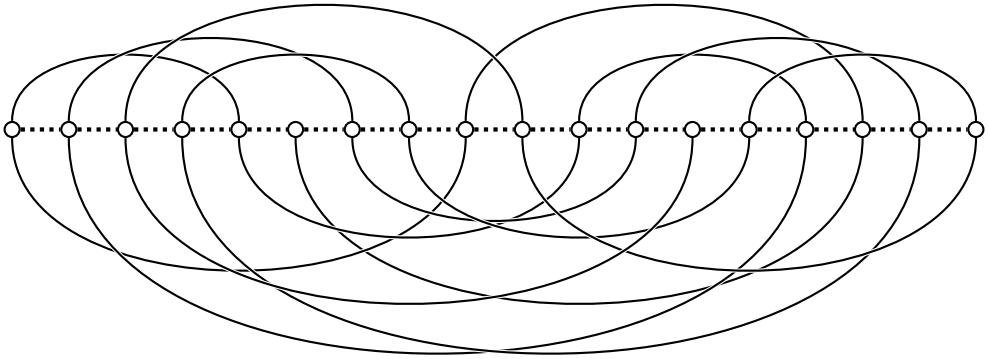

Figure 3: A graph corresponding to the permutation in (1); the dashed line being the top row and the solid line being the bottom row. Note the graph has girth 5.

we switch some pair of edges of P which are far from each other, that is, omit them and connect their endpoints by new edges in the unique way ensuring that the modified P will stay a Hamiltonian path. Here are the details.

As long as H contains a cycle of length at most $2d$, let C be a shortest cycle in H, and let e be an arbitrary edge of P that belongs to C (there must be such an edge, as the other Hamiltonian path contains no cycle at all). By assumption

$$n - 1 > 1 + 2 \cdot 3 + 2 \cdot 3^2 + 2 \cdot 3^3 + \cdots + 2 \cdot 3^{2d-1}$$

implying that P contains an edge e' whose distance (in H) from e is at least $2d$. We now switch at e, e' (that is, delete them and add the required edges to keep P a Hamiltonian path). This way we get rid of the cycle C. Any new cycle created this way either contains only one of the new edges, and then its length is at least $2d + 1$, or contains both and then its length is at least twice the length of the shortest cycle deleted. Proceeding in this way we increase the length of the shortest cycle after a finite number of steps, and when the process terminates we get the required graph. Note that since the number of cycles of length t in a graph of maximum degree 4 and n vertices is smaller than $n \cdot 3^t$ the process terminates after at most

$$O(n(3^3 + \cdots + 3^{2d})) = O(n^2)$$

steps, and each step is efficient as finding a shortest cycle in a graph is efficient. □

The assertions of Theorem 3 follow from the two preceding lemmas. This shows that we can find d-resilient permutations which have size n exponential in d, we now note that this is best possible.

Theorem 6. *If there is a permutation σ of $[n]$ that is d-resilient, then $d \leq O(\log n)$.*

Proof. By the known results about cycles in graphs with n vertices and $2n - 2$ edges (see [1]), the graph constructed from the permutation σ as in the discussion above contains a short cycle on a set of vertices S, $|S| \leq 2 \log_3 n + O(1)$. This in turn implies that $G(\sigma, D)$ has a cycle of length $2 \log_3 n + O(1)$ and hence we have that $d \leq 2 \log_3 n + O(1)$. □

Finally we note that asymptotically a positive portion of permutations are d-resilient.

Proposition 7. *For any fixed d there is a positive real $\epsilon(d)$ and $n_0 = n_0(d)$ so that the probability that a random permutation σ of $[n]$ is d-resilient is at least $\epsilon(d)$.*

Proof. It is known that for every fixed integer d a random 4-regular graph on n vertices, for large n, has girth bigger than $2d$ with probability at least some $\delta(d) > 0$. By the known results about contiguity (see [3]) this random graph is the edge disjoint union of two Hamiltonian cycles with probability that tends to 1 as n tends to infinity. This implies the required assertion, by Lemma 4. □

Acknowledgements

Utkrisht Rajkumar thanks Young-Han Kim for support and guidance. Noga Alon and Ron Graham thank the Simons Institute for the Theory of Computing at UC Berkeley, where part of this work was done. The authors thank the referees for feedback on an earlier version of this paper.

References

[1] N. Alon, S. Hoory and N. Linial, The Moore bound for irregular graphs, Graphs and Combinatorics 18 (2002), 53-57.

[2] P. Erdős and H. Sachs, Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl, Wiss. Z. Matin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe 12(1963), 251–257.

[3] J. H. Kim and N. Wormald, Random matchings which induce Hamiltonian cycles, and hamiltonian decompositions of random regular graphs, J. Combinatorial Theory, Series B 81 (2001), 20–44.

[4] N.J.A. Sloane, On single-deletion-correcting codes, in *Codes and Designs: Proceedings of a Conference Honoring Professor Dijen K. Ray-Chaudhuri on the Occasion of His 65th Birthday, Ohio State University, May 18-21, 2000*, 2002.