Reconstruction of Sequences Over Non-Identical Channels

Michal Horovitz and Eitan Yaakobi, *Senior Member, IEEE*

***Abstract*— Motivated by the error behavior in the DNA storage channel, in this paper, we extend the previously studied *sequence reconstruction problem* by Levenshtein. The reconstruction problem studies the model in which the information is read through multiple noisy channels, and the decoder, which receives all channel estimations, is required to decode the information. For the combinatorial setup, the assumption is that all the channels cause at most some t errors. Levenshtein considered the case in which all the channels have the same behavior, and we generalize this model and assume that the channels are not identical. Thus, different channels may cause different maximum numbers of errors. For example, we assume that there are N channels, which cause at most t_1 or t_2 errors, where $t_1 < t_2$, and the number of channels with at most t_1 errors is at least $\lceil pN \rceil$, for some fixed $0 < p < 1$. If the information codeword belongs to a code with minimum distance d, the problem is then to find the minimum number of channels that guarantees successful decoding in the worst case. A different problem we study in this paper is where the number of channels is fixed, and the question is finding the minimum distance d that provides exact reconstruction. We study these problems and show how to apply them for the cases of substitutions and transpositions.**

***Index Terms*— The sequence reconstruction problem, DNA storage, Hamming errors, the Johnson graph.**

I. Introduction

The *sequence reconstruction problem* was first proposed and studied by Levenshtein in [9] and [10]. In this model, a codeword is transmitted over multiple channels and a decoder, which receives all channel outputs, decodes the transmitted word. The assumption is that all channels are the same and are uncorrelated, with the only exception that all channel outputs are different from each other. This model was originally motivated by chemical and biological processes where the information is replicated and can be read from different noisy sources. However, it was also shown to be relevant in storage technologies, where the stored information has multiple copies or where a single copy is read by several different read heads. Specifically, the applicability of this model is most relevant to *DNA storage* [1], [2], [17]–[19]. Both for *in vitro* and *in vivo* storage systems, the information has a large number of copies stored in DNA strands and the goal is to read these strands and reconstruct the data, while every estimation of the data is erroneous.

The reconstruction model studied by Levenshtein and later by others was combinatorial. Suppose all words belong to some space V with a distance function ρ. It is assumed that the information codeword x belongs to a code with minimum distance d and the number of errors in every channel is at most t. Then, the goal is to find the minimum number of channels that guarantees unique decoding in the worst case. Clearly, if $t < \lfloor (d-1)/2 \rfloor$, then a single channel is sufficient. Otherwise, it was shown in [9] that this number has to be greater than the largest intersection of two balls with radius t and minimum distance d between their centers, that is, greater than

$$\max\{|B_t(x) \cap B_t(z)| : x, z \in V, \ \ \rho(x, z) \geq d\},$$

where $B_t(x) = \{y \in V \ : \ \rho(x, y) \leq t\}$. Later, this combinatorial problem was studied for several channels. Levenshtein [9] studied the cases of substitution errors, the Johnson graphs, and several more general metric distances. More results for other general error graphs were given in [11] and [12], and in [6]–[8], it was studied for permutations. The case of permutations with the Kendall's τ distance was investigated in [16] as well as the Grassmann graph case. Levenshtein's results for deletions and insertions in [10] were extended in [14] for insertions and in [3] for deletions. In [15], the connection between the reconstruction problem and associative memories was studied, and in [5] it was analyzed for the purpose of asymptotically improving the Gilbert-Varshamov bound.

The motivation for the paradigm studied in this paper comes from the error behavior in DNA storage. We generalize Levenshtein's model and assume a combinatorial model where the channels are not identical. When reading the data stored in DNA strands, it may happen that some estimations are more noisy than the others [17]. In the reconstruction model this is translated to channels that cause a different maximum number of errors. For example, it is known that for substitution errors, if the transmitted word belongs to a code with minimum Hamming distance 3 and there are at most 2 errors in every channel, then 7 channels are necessary and sufficient for successful decoding. However, if at most 2 channels cause two errors (and the rest 1 error), then we show that 5 channels are necessary and sufficient for successful decoding. In [13],

Manuscript received September 7, 2017; revised March 13, 2018; accepted June 11, 2018. Date of publication July 25, 2018; date of current version January 18, 2019. This work was supported in part by the Israel Science Foundation under Grant 1624/14 and in part by the United States–Israel Binational Science Foundation (BSF) under Grant 2015816. This paper was presented in the 2017 IEEE International Symposium on Information Theory [4].

M. Horovitz is with the Department of Computer Science, Tel-Hai College, Upper Galilee 1220800, Israel, and also with the Galilee Research Institute - Migal, Kiryat Shmona 11016, Israel (e-mail: horovitzmic@telhai.ac.il).

E. Yaakobi is with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yaakobi@cs.technion.ac.il).

Communicated by A. Jiang, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2018.2859813

a similar problem was studied for the setup in which every channel can cause a different number of insertions.

In this paper, we mainly focus on studying the following two problems. In the first problem we assume that the input set, i.e. the code and its minimum distance, is given, and we seek to find the minimum number of channels required for exact reconstruction. In the second problem, the number of channels is given, and then we study the minimum distance of a code which is required for exact reconstruction. In these two problems, we consider three cases which depend upon whether the information about the number of errors in each channel is given, only the distribution of the number of errors in the channels is given, or only the average number of errors is given.

Formally, we define the first model as follows. Let ℓ be the number of possible types of channels. For $T = (t_1, \ldots, t_\ell)$ and $P = (p_1, \ldots, p_\ell)$, where $t_1 < \cdots < t_\ell \in \mathbb{N}$ and $0 < p_1 < \cdots < p_{\ell-1} < p_\ell = 1$, we say that a system with N channels is a (T, P)-*channel system* if for all i, $1 \leq i \leq \ell$, $\lceil p_i N \rceil$ of the channels cause at most t_i errors. For example, Levenshtein's model is a special case with $\ell = 1$ and $p_1 = 1$. We study the minimum number of channels required for a (T, P)-channel system for successful decoding when the information is a codeword which belongs to a code with minimum distance d. The first two cases we consider here take into account whether the decoder may or may not know the type of each channel. We also study the case where only the average number of errors is known. We solve the general cases, and focus on substitution and transposition errors for $\ell = 2$.

In the second problem the number of channels N is fixed, and the goal is to find the minimum distance d of a code which guarantees exact reconstruction. As in the first problem, we consider here also three parallel questions according to the knowledge about the error behavior of the channels. That is, whether the decoder knows the number of errors in each channel, the distribution of the number of errors, or only the average number of errors.

The rest of the paper is organized as follows. In Section II, we formally define the models and the main problems studied in the paper. In Section III we solve the first problem, i.e., we find the minimum number of channels required for exact reconstruction, where the input set is given, and in Section IV we demonstrate this solution for substitution and transposition errors. Then, in Section V we solve the second question, where the number of channels is given, and the goal is to find the minimum distance of the code required for exact reconstruction. Later on, in Section VI, we consider two special systems. Finally, we conclude in Section VII.

II. Definitions and Problems Setup

For a positive integer h, we denote by $[h]$ the set $\{1, 2, \ldots, h\}$. Let V be a finite set with a distance function $\rho : V \times V \to \mathbb{N}$, when $\mathbb{N}$ is the set of all non-negative integer numbers. For $x \in V$, the ball of radius t centered at x is the set $B_t(x) = \{y \in V : \rho(x, y) \leq t\}$. A combinatorial channel C is called a t-*error channel*, if for any input $x \in V$ the output of C is in $B_t(x)$. Note that for $t < t'$, a t-error channel is also a t'-error channel.

A *channel system* is a system consisting of some N combinatorial channels $C_1, C_2, \ldots, C_N$. The *size* of the channel system is the number of channels N comprised in it. We say that a word $x \in V$ is *transmitted over the channel system* if x is transmitted over C_i for all $i \in [N]$, and y_i is the output of the ith channel. The sequence $(y_1, \ldots, y_N)$ is called the *outputs sequence* of the system. The receiver applies a decoding function $\mathcal{D}(y_1, \ldots, y_N)$ in order to reconstruct the transmitted word x, and *exact reconstruction* happens when $x = \mathcal{D}(y_1, \ldots, y_N)$. In this paper we only refer to the exact reconstruction problem and we assume that all channel outputs are different from each other. We say that a channel system *supports exact reconstruction for* U, $U \subseteq V$, if there exists a decoding function $\mathcal{D}$ such that for each $x \in U$, $x = \mathcal{D}(y_1, \ldots, y_N)$ where $(y_1, \ldots, y_N)$ is a possible outputs sequence when x is transmitted over the system.

Let $T = (t_1, \ldots, t_\ell)$ and $P = (p_1, \ldots, p_\ell)$ be such that $t_1 < t_2 < \cdots < t_\ell \in \mathbb{N}$ and $0 < p_1 < p_2 < \cdots < p_{\ell-1} < p_\ell = 1$. A channel system with N combinatorial channels is called a (T, P)-*channel system* if for each $i \in [\ell]$, $\lceil p_i N \rceil$ of the channels are t_i-error channels.

We consider two models, which depend upon whether the behavior of each specific channel is known or unknown to the decoder. In the first channel system, called the *sequenced-channel system*, the decoder *knows* the maximum number of errors in every channel. In this model we assume, without loss of generality, that for each $i \in [\ell]$ the first $\lceil p_i N \rceil$ channels are t_i-error. However, in the second channel system, called the *non-sequenced-channel system*, only the distribution of the errors in the channels is known to the decoder, but the number of errors in each individual channel is unknown. For example, in the non-sequenced model, the decoder knows that half of the channels are t_1-error channels, and the rest are t_2-error channels, but it does not know for each specific channel, if it is a t_1-error or t_2-error channel. But, in the sequenced model, the decoder knows also that the first half of the channels are t_1-error. We will also consider a channel system where only t, the average number of errors, is known, when t is not necessarily an integer number. Such a system will be called a t-*channel system*.

For $U \subseteq V$, we denote by $N^u(T, P, U, V)$ the minimum size of a (T, P)-non-sequenced-channel system such that every $x \in U$ has exact reconstruction. Similarly, $N^k(T, P, U, V)$ and $N^a(t, U, V)$ are defined for the sequenced- and t-channel systems, respectively. Note that $N^k(T, P, U, V) \leq N^u(T, P, U, V)$, and clearly the values of $N^k(T, P, U, V)$, $N^u(T, P, U, V)$, and $N(t, U, V)$ depend also on the distance function ρ. Yet, the distance function will be clear from the context, so to simplify, we omit ρ in these notations. In the rest of the paper, whenever we write g, we refer to $g \in \{k, u\}$. The first problem we study in this paper is formulated as follows.

Problem 1: Let V be a finite set with distance function $\rho : V \times V \to \mathbb{N}$, $T = (t_1, \ldots, t_\ell)$, $t \geq 0$, and $P = (p_1, \ldots, p_\ell)$. For all $U \subseteq V$, find the values of $N^k(T, P, U, V)$, $N^u(T, P, U, V)$, and $N^a(t, U, V)$.

Problem 1 is a generalized version for the setup presented and solved by Levenshtein for $\ell = 1$ [9]. The solution for a general ℓ, which will be described in Subsection III-B, extends the result by Levenshtein, where in this case $T = (t)$, $P = (1)$ and $N^k(T, P, U, V) = N^u(T, P, U, V)$.

Problem 1 is studied in Section III. In Subsection III-A, we find the values $N^k(T, P, U, V)$ and $N^u(T, P, U, V)$ for $\ell = 2$, that is, the channel system is given by $T = (t_1, t_2)$ and $P = (p, 1)$. Then, in Subsection III-B, we generalize this combinatorial solution for all ℓ, and in Subsection III-C we present the solution for $N^a(t, U, V)$. In Section IV, we apply the solution for $\ell = 2$ for two types of errors; for substitutions in Subsection IV-A and for transpositions in Subsection IV-B, both over the binary alphabet. In Section VI we slightly modify the model of a (T, P)-channel system and study some special cases of $\ell = 2$ when there exists a fixed number of t_1-error channels and the rest of the channels are t_2-error channels, and vice versa. These problems are solved for both cases, the sequenced- and non-sequenced-channel systems.

In the sequel, when writing x and z, we assume that $x \neq z$, and we use a bold notation $(\boldsymbol{x}, \boldsymbol{z})$ when $\boldsymbol{x}$ and $\boldsymbol{z}$ are vectors. It is straightforward to verify that

$$N^g(T, P, U, V) = \max\{N^g(T, P, \{x, z\}, V) : x, z \in U\}, \text{ and}$$
$$N^a(t, U, V) = \max\{N^a(t, \{x, z\}, V) : x, z \in U\}.$$

We also define $N^g(T, P, d, V)$ as the maximum value of $N^g(T, P, U, V)$, for all $U \subseteq V$ of minimum distance at least d, that is, $d(U) \geq d$, where $d(U) = \min\{\rho(x, z) : x, z \in U\}$. As before, we get that

$$N^g(T, P, d, V) = \max\{N^g(T, P, \{x, z\}, V) : x, z \in V,\ \rho(x, z) \geq d\},$$

and $N^a(t, d, V)$ is defined similarly. Therefore, we focus on finding the values of $N^g(T, P, \{x, z\}, V)$ and $N^a(t, \{x, z\}, V)$ for all $x, z \in U$. For $V = \{0, 1\}^n$, we denote the values of $N^g(T, P, d, V)$ and $N^a(t, d, V)$ by $N^g(T, P, d, n)$ and $N^a(t, d, n)$, respectively.

The second problem we study in this paper is formulated as follows. Assume that N, the size of the channel system, is given. We define three models of channel systems, each one is of size N. Denote by t_i the number of maximum errors in the i-th channel, and by t the average number of errors, i.e., $t = \frac{\sum_{i=1}^N t_i}{N}$. Without loss of generality $t_1 \leq t_2 \leq \cdots \leq t_N$. We define the three models as follows,

(1) a (T, N)-sequenced-channel system - where the N-tuple $T = (t_1, t_2, \ldots, t_N)$ is given,
(2) a (T, N)-non-sequenced-channel system - where the multiset $T = \{t_1, t_2, \ldots, t_N\}$ is given, and
(3) a (t, N)-channel system - where t, the average number of errors is given.

Then, we study the minimum distance d required for exact reconstruction in each one of these three models. We denote by $D^k(T, N, V)$, $D^u(T, N, V)$, and $D^a(t, N, V)$, the minimum distance of the codes required for exact reconstruction for (T, N)-sequenced-, (T, N)-non-sequenced-, and (t, N)- channel systems, respectively. Note that $D^k(T, N, V) \leq D^u(T, N, V) \leq D^a(t, N, V)$, where we denote these values by ∞ if there does not exist such a minimum distance. In the sequel, we use the notation $T = (t_1, \ldots, t_N)$ to denote the multiset $\{t_1, \ldots, t_N\}$ or the N-tuple $(t_1, \ldots, t_N)$, where the meaning will be clear from the context, according to the model. Note that the parameter N is redundant in the non-sequenced and in the sequenced models, nevertheless, it is left for clarification. The second problem is defined as follows.

Problem 2: Let V be a finite set with distance function $\rho : V \times V \to \mathbb{N}$. For all N, $T = (t_1, \ldots, t_N)$, and $t \geq 0$, find the values of $D^k(T, N, V)$, $D^u(T, N, V)$, and $D^a(t, N, V)$.

In Section V we solve Problem 2, using reduction to Problem 1. We also provide some examples for substitution errors.

III. Problem 1 - Minimum Size of a Channel-System

In this section we consider Problem 1 for every $U \subseteq V$. In Subsection III-A we find the minimum number of channels required for exact reconstruction in a (T, P)-channel system where $\ell = 2$, that is, the values of $N^u(T, P, U, V)$ and $N^k(T, P, U, V)$. Then, in Subsection III-B, we extend this analysis for arbitrary ℓ. Finally, in Subsection III-C we solve the problem of $N^a(t, U, V)$, where only the average number of errors is known. In the rest of this section, and unless stated otherwise, we assume that $U = \{x, z\} \subseteq V$.

A. The Case $\ell = 2$

In this subsection we study Problem 1 for $\ell = 2$. This result extends the case studied by Levenshtein when all the channels are identical [9]. For $\ell = 2$ we have, $T = (t_1, t_2)$, and $P = (p, 1)$, where $t_1 < t_2 \in \mathbb{N}$ and $0 < p < 1$. Recall that a (T, P)-channel system of size N is a set of N combinatorial channels, where $\lceil pN \rceil$ of the channels are t_1-error channels and the others are t_2-error channels.

For the rest of this subsection we assume that $T = (t_1, t_2)$ and $P = (p, 1)$. We define

$$I(x, z, t_1, t_2) = B_{t_1}(x) \cap B_{t_2}(z), \quad I(x, z, t_1) = B_{t_1}(x) \cap B_{t_1}(z),$$

and

$$N(x, z, t_1, t_2) = |I(x, z, t_1, t_2)|, N(x, z, t_1) = |I(x, z, t_1)|.$$

Note that the values of $B_t(x)$, $I(x, z, t_1, t_2)$, and $N(x, z, t_1, t_2)$ depend also on V which is omitted to simplify the notations and it will be clear from the context.

We start with a simple proposition that will be used in the paper.

Proposition 1: For $r_1, r_2 \in \mathbb{N}$ and $0 < p \leq 1$ the followings hold.

(a) If $r_1 > \left\lfloor \frac{r_2}{p} \right\rfloor$, then $\lceil p \cdot r_1 \rceil > r_2$.

(b) If $r_1 \leq \left\lfloor \frac{r_2}{p} \right\rfloor$, then $\lceil p \cdot r_1 \rceil \leq r_2$.

Proof:

(a) For an integer r_1, the equality $r_1 > \left\lfloor \frac{r_2}{p} \right\rfloor$ is equivalent to $r_1 \geq \left\lfloor \frac{r_2}{p} \right\rfloor + 1$. Thus, we have
$$\lceil p \cdot r_1 \rceil \geq p \cdot r_1 \geq p \cdot (\lfloor r_2/p \rfloor + 1) > p \cdot r_2/p = r_2.$$

(b) If $\lceil p \cdot r_1 \rceil = p \cdot r_1$, then
$$\lceil p \cdot r_1 \rceil = p \cdot r_1 \leq p \cdot \lfloor r_2/p \rfloor \leq p \cdot r_2/p = b.$$

Otherwise, $\lceil p \cdot r_1 \rceil = \lfloor p \cdot r_1 \rfloor + 1$, and

$$\lceil p \cdot r_1 \rceil = \lfloor p \cdot r_1 \rfloor + 1 < p \cdot \lfloor r_2/p \rfloor + 1 \leq p \cdot (r_2/p) + 1 = r_2 + 1.$$

□

The following theorem solves Problem 1 for the sequenced model.

Theorem 2: $N^k(T, P, U, V) = N + 1$, *where*

$$N = \min\{\lfloor N(x, z, t_1)/p \rfloor, N(x, z, t_2)\}.$$

Proof: The above conditions are symmetric for x and z. Thus, without loss of generality, let x be the transmitted word.

For the first direction we prove that a (T, P)-sequenced-channel system of size J, where $J > \min\{\lfloor N(x, z, t_1)/p \rfloor, N(x, z, t_2)\}$, supports exact reconstruction for $U = \{x, z\}$. If $J > \min\{\lfloor N(x, z, t_1)/p \rfloor, N(x, z, t_2)\}$ then at least one of the following conditions holds:

(1) $J > \lfloor N(x, z, t_1)/p \rfloor$ which implies $\lceil pJ \rceil > N(x, z, t_1)$ by Proposition 1(a), or
(2) $J > N(x, z, t_2)$.

Assume the word x was transmitted over the channel system. Therefore, the first $\lceil pJ \rceil$ outputs are in $B_{t_1}(x)$ and all the J outputs are in $B_{t_2}(x)$. Thus, if the first condition holds, since the first $\lceil pJ \rceil$ of the outputs are in $B_{t_1}(x)$ and $\lceil pJ \rceil > N(x, z, t_1)$, at least one of the first $\lceil pJ \rceil$ outputs is not in $B_{t_1}(z)$. However, for z to be decoded, all of these outputs must be in $B_{t_1}(z)$. Thus, z cannot be decoded. Otherwise, if the second condition holds, since all J outputs are in $B_{t_2}(x)$ and $J > N(x, z, t_2)$, not all the outputs are in $B_{t_2}(z)$. But, for z to be decoded, all the outputs must be in $B_{t_2}(z)$. Hence, for both cases, the word z will not be decoded.

For the second direction we prove that if the following two conditions hold simultaneously:

(1) $J \leq \lfloor N(x, z, t_1)/p \rfloor$ which implies $\lceil pJ \rceil \leq N(x, z, t_1)$ by Proposition 1(b), and
(2) $J \leq N(z, x, t_2)$,

then a channel system of size J does not support exact reconstruction for $U = \{x, z\}$. We present a sequence of J outputs which can be an outputs sequence when transmitting either x or z. We chose the first $m = \min\{J, N(x, z, t_1)\}$ outputs from $I(x, z, t_1)$, where $m \geq \lceil pJ \rceil$ by the first condition. The rest of the outputs are in $I(x, z, t_2)$, which is possible by the second condition. Thus, the first $\lceil pJ \rceil$ outputs are in $B_{t_1}(x)$, and all the J outputs are in $B_{t_2}(x)$, and the same holds for z. □

The rest of this subsection is dedicated for presenting the solution for the non-sequenced model. Note that if x is transmitted over a (T, P)-non-sequenced-channel system with N channels, then at least $\lceil pN \rceil$ of the outputs are in $B_{t_1}(x)$, and all the N outputs are in $B_{t_2}(x)$. Thus, to support exact reconstruction for x we require that for every $z \in U$, there are no N outputs such that all the following three conditions hold simultaneously:

(1) at least $\lceil pN \rceil$ of the outputs are in $B_{t_1}(x)$,
(2) at least $\lceil pN \rceil$ of the outputs are in $B_{t_1}(z)$,
(3) all N outputs are in $I(x, z, t_2) = B_{t_2}(x) \cap B_{t_2}(z)$.

Thus, we conclude

Lemma 3: A (T, P)-non-sequenced-channel system of size J supports exact reconstruction for U, if and only if at least one of the following four conditions holds:

(1) $\lceil pJ \rceil > N(x, z, t_1, t_2)$,
(2) $\lceil pJ \rceil > N(z, x, t_1, t_2)$,
(3) $J > N(x, z, t_2)$, *or*
(4) $2\lceil pJ \rceil - N(x, z, t_1) > J$.

Proof: The above conditions are symmetric for x and z. Thus, without loss of generality, let x be the transmitted word.

For the first direction we prove that if at least one of the above conditions holds, then the system supports exact reconstruction. If Condition (1) or Condition (3) holds, since at least $\lceil pJ \rceil$ of the outputs are in $B_{t_1}(x)$ and all the J outputs are in $B_{t_2}(x)$, then not all the outputs can be in $B_{t_2}(z)$. If Condition (2) holds, there are no $\lceil pJ \rceil$ outputs in $B_{t_1}(z)$. Thus, if one of conditions (1), (2), or (3) holds, then z will not be decoded. For Condition (4), assume that we have m outputs in $I(x, z, t_1)$, $m \leq N(x, z, t_1)$. In order for z to be a possible output for the decoder, we must have at least $\lceil pJ \rceil - m$ outputs in $I(z, x, t_1, t_2) \setminus I(x, z, t_1)$. Furthermore, since x was transmitted at least $\lceil pJ \rceil - m$ outputs are in $I(x, z, t_1, t_2) \setminus I(x, z, t_1)$. Thus, we must have that $2\lceil pJ \rceil - m \leq J$ in contradiction to Condition (4).

For the second direction we prove that if the following four conditions hold simultaneously:

(1) $\lceil pJ \rceil \leq N(x, z, t_1, t_2)$,
(2) $\lceil pJ \rceil \leq N(z, x, t_1, t_2)$,
(3) $J \leq N(x, z, t_2)$, and
(4) $2\lceil pJ \rceil - N(x, z, t_1) \leq J$,

then the channel system of size J does not support exact reconstruction for $U = \{x, z\}$. For this part, we present a set of J outputs, which any order of them may be an outputs sequence when transmitting either x or z. Let $m = N(x, z, t_1)$. If $m < \lceil pJ \rceil$ then m outputs are chosen from $I(x, z, t_1)$, at least $\lceil pJ \rceil - m$ outputs are in $I(x, z, t_1, t_2) \setminus I(x, z, t_1)$ (by Conditions (1) and (4)), at least $\lceil pJ \rceil - m$ in $I(z, x, t_1, t_2) \setminus I(x, z, t_1)$ (by Conditions (2) and (4)), and all the outputs are in $I(x, z, t_2)$ (by Condition (3)). Otherwise, $m \geq \lceil pJ \rceil$, and then at least $\lceil pJ \rceil$ outputs are in $I(x, z, t_1)$, and all the rest are in $I(x, z, t_2)$ (by Condition (3)). Thus, at least $\lceil pJ \rceil$ of the outputs are in $B_{t_1}(x)$, and all the J outputs are in $B_{t_2}(x)$, and the same holds for z. □

We are now ready to find the value of $N^u(T, P, U, V)$. Let us define

$$N'(x, z, t_1, p) = \min\{L : 2\lceil pL \rceil - L > N(x, z, t_1), L \geq 1\} - 1, \quad (1)$$

where here and in the rest of this paper $\min \emptyset = \infty$. The following theorem establishes our result in calculating the value of $N^u(T, P, U, V)$ by using Lemma 3 and the definition of $N'(x, z, t_1, p)$.

Theorem 4: $N^u(T, P, U, V) = N + 1$, *where*

$$N = \min\{\lfloor N(x, z, t_1, t_2)/p \rfloor, N(x, z, t_2), \lfloor N(z, x, t_1, t_2)/p \rfloor, N'(x, z, t_1, p)\}.$$

Proof: If a (T, P)-channel system consists of $J = N + 1$ channels, then, by the definition of N, at least one of the following conditions holds:

(1) $J > \lfloor N(x, z, t_1, t_2)/p \rfloor$ which implies by Proposition 1(a) $\lceil pJ \rceil > N(x, z, t_1, t_2)$,
(2) $J > \lfloor N(z, x, t_1, t_2)/p \rfloor$ which implies by Proposition 1(a) $\lceil pJ \rceil > N(z, x, t_1, t_2)$,
(3) $J > N(x, z, t_2)$, or
(4) $2\lceil pJ \rceil - N(x, z, t_1) > J$.

Thus, by Lemma 3, a channel system of size J supports exact reconstruction.

For the second direction we have to prove that if $J \leq N$, then J channels are not sufficient for exact reconstruction where $U = \{x, z\}$. The following four conditions hold simultaneously:

(1) $\lceil pJ \rceil \leq N(x, z, t_1, t_2)$ which is derived by Proposition 1(b) from $J \leq \lfloor N(x, z, t_1, t_2)/p \rfloor$,
(2) $\lceil pJ \rceil \leq N(z, x, t_1, t_2)$ which is derived by Proposition 1(b) from $J \leq \lfloor N(x, z, t_1, t_2)/p \rfloor$,
(3) $J \leq N(x, z, t_2)$, and
(4) $2\lceil pJ \rceil - N(x, z, t_1) \leq J$.

Then, we apply again Lemma 3 to conclude that exact reconstruction is not supported. □

Remark 5: We note that a (T, P)-non-sequenced-channel system of size J, where $J > N^u(T, P, \{x, z\}, V)$, may not support exact reconstruction for $U = \{x, z\}$. That could happen only if $J \leq \min\{\lfloor N(x, z, t_1, t_2)/p \rfloor, \lfloor N(z, x, t_1, t_2)/p \rfloor, N(x, z, t_2)\}$. The reason for this undesirable phenomena is that $N^u(T, P, U, V)$ may be determined by $N'(x, z, t_1, p)+1$, that is, $N^u(T, P, U, V) = N'(x, z, t_1, p) + 1$ which means that $2\lceil pN^u(T, P, U, V)\rceil - N^u(T, P, U, V) > N(x, z, t_1)$. However, $2\lceil pJ \rceil - J \leq N(x, z, t_1)$. For example, assume $V = \{0, 1\}^n$, ρ is the Hamming distance function, $d = 3$, $T = (1, 4)$, and $P = (1/3, 1)$. Thus, we have $N^u(T, P, d, n) = 1$, but 2 channels (and even more) are not sufficient (the vectors $\boldsymbol{y}_1 = (1, 0, 0, \ldots, 0)$ and $\boldsymbol{y}_2 = (1, 1, 1, 1, 0, \ldots, 0)$ might be the outputs of either $\boldsymbol{x} = (0, 0, \ldots, 0)$ or $\boldsymbol{z} = (1, 1, 1, 0, \cdots, 0)$).

Thus, another interesting problem is to find a threshold $N_2^u(T, P, U, V)$ such that for all $N \geq N_2^u(T, P, U, V)$, a (T, P)-non-sequenced-channel system of size N supports exact reconstruction, but a (T, P)-non-sequenced-channel system of size $N_2^u(T, P, U, V) - 1$ does not support exact reconstruction. Unfortunately, we can not answer this question completely. However, we know that $N^u(T, P, U, V) \leq N_2^u(T, P, U, V) \leq N'$, where $N' = \min\{\lfloor N(x, z, t_1, t_2)/p \rfloor, \lfloor N(z, x, t_1, t_2)/p \rfloor, N(x, z, t_2)\} + 1$.

Let J be the size of a (T, P)-non-sequenced channel system. We can conclude the following.

- *If $J < N^u(T, P, \{x, z\}, V)$ then exact reconstruction is not supported.*
- *If $J = N^u(T, P, \{x, z\}, V)$ or $J \geq N'$ then exact reconstruction is supported.*
- *Otherwise, $N^u(T, P, \{x, z\}, V) < J < N'$, and we do not know whether exact reconstruction is supported or not.*

This phenomena happens also for general ℓ in the non-sequenced model (see Remark 11).

In order to complete the study of the value $N^u(T, P, U, V)$ we are only left with studying the value of $N'(x, z, t_1, p)$.

Proposition 6: For $0 < p \leq 1/2$:

$$N'(x, z, t_1, p) = \begin{cases} 0 & \text{if } N(x, z, t_1) = 0, \\ \infty & \text{otherwise.} \end{cases}$$

For $1/2 < p < 1$:

$$\left\lfloor \frac{N(x, z, t_1) - 2}{2p - 1} \right\rfloor \leq N'(x, z, t_1, p) \leq \left\lfloor \frac{N(x, z, t_1)}{2p - 1} \right\rfloor.$$

Proof: For $0 < p \leq 1/2$, $2\lceil pL \rceil - L \leq 1$ for all L. Thus, in this case the value of $N'(x, z, t_1, p)$ is an immediate result by Equation (1).

For $1/2 < p < 1$, denote $a = N(x, z, t_1)$. For each δ, $0 \leq \delta < 1$, let $A_\delta = \{J : \lceil pJ \rceil - pJ = \delta, J \geq 1\}$, and $J_\delta = \min\{L : 2\lceil pL \rceil - L > a, L \geq 1, L \in A_\delta\}$. For $A_\delta \neq \emptyset$, let L_δ be an element in A_δ, i.e., $\lceil pL_\delta \rceil - pL_\delta = \delta$. Thus, $2\lceil pL_\delta \rceil - L_\delta > a$ if and only if $2(pL_\delta + \delta) - L_\delta > a$, which holds if and only if $L_\delta > \frac{a-2\delta}{2p-1}$, that is equivalent to $L_\delta \geq \left\lfloor \frac{a-2\delta}{2p-1} \right\rfloor + 1$. Thus, if $A_\delta \neq \emptyset$ then $J_\delta = \left\lfloor \frac{a-2\delta}{2p-1} \right\rfloor + 1$. We can conclude that $N'(x, z, t_1, p) = \min\{J_\delta - 1 : 0 \leq \delta < 1\}$. Therefore,

$$\left\lfloor \frac{N(x, z, t_1) - 2}{2p - 1} \right\rfloor \leq N'(x, z, t_1, p) \leq \left\lfloor \frac{N(x, z, t_1)}{2p - 1} \right\rfloor$$

□

The following corollary is deduced immediately by Proposition 6 and Theorem 4.

Corollary 7: $N^u(T, P, U, V) = N + 1$ where N is defined as follows. For $0 < p \leq 1/2$:

$$N = \begin{cases} 0 & \text{if } N(x, z, t_1) = 0 \\ \min\{\lfloor N(x, z, t_1, t_2)/p \rfloor, & \text{otherwise.} \\ \quad \lfloor N(z, x, t_1, t_2)/p \rfloor, & \\ \quad N(x, z, t_2)\}. & \end{cases}$$

For $1/2 < p < 1$:

$$N = \min\{\lfloor N(x, z, t_1, t_2)/p \rfloor, N(x, z, t_2), \lfloor N(z, x, t_1, t_2)/p \rfloor, N'(x, z, t_1, p)\}.$$

In Section IV we show how to apply the result from Corollary 7 in order to explicitly solve Problem 1 with $\ell = 2$ for substitution and transposition errors over the binary alphabet.

B. The General Case

In this section, we extend the solution from Subsection III-A. We provide a combinatorial translation for the general case of Problem 1, where $T = (t_1, \ldots, t_\ell)$ and $P = (p_1, \ldots, p_{\ell-1}, p_\ell)$, $t_1 < t_2 < \cdots < t_\ell \in \mathbb{N}$, and $0 < p_1 < p_2 < \cdots < p_{\ell-1} < p_\ell = 1$. Remember that a (T, P)-channel system of size N consists of N channels, where for each $i \in [\ell]$, $\lceil p_i N \rceil$ channels are t_i-error channels.

Theorem 8 and Theorem 10 generalize Theorem 2 and Theorem 4 for arbitrary ℓ, respectively.

Theorem 8: $N^k(T, P, U, V) = N + 1$, where

$$N = \min\{\lfloor N(x, z, t_i)/p_i \rfloor : i \in [\ell]\}.$$

Furthermore, for all $J \geq N^k(T, P, U, V)$, a (T, P)-sequenced-channel system of size J supports exact reconstruction for U.

Proof: The above conditions are symmetric for x and z. Thus, without loss of generality, let x be the transmitted word. For the first direction we prove that a (T, P)-sequenced-channel system of size J, where $J > \min\{\lfloor N(x,z,t_i)/p_i \rfloor : i \in [\ell]\}$, supports exact reconstruction for $U = \{x, z\}$. If $J > \min\{\lfloor N(x,z,t_i)/p_i \rfloor : i \in [\ell]\}$ then there exists $i \in [\ell]$, such that $J > \lfloor N(x,z,t_i)/p_i \rfloor$, which implies by Proposition 1(a) that $\lceil p_i J \rceil > N(x,z,t_i)$. The word x was transmitted over the channel system. Therefore, the first $\lceil p_i J \rceil$ outputs are in $B_{t_i}(x)$. But, since $\lceil p_i J \rceil > N(x,z,t_i)$, at least one output from the first $\lceil p_i J \rceil$ outputs is not in $B_{t_i}(z)$. However, for z to be decoded, all the first $\lceil p_i J \rceil$ outputs must be in $B_{t_i}(z)$. Hence, z cannot be decoded.

For the second direction we prove that if for all $i \in [\ell]$, $J \leq \lfloor N(x,z,t_i)/p_i \rfloor$ then a channel system of size J does not support exact reconstruction for $U = \{x, z\}$.

By Proposition 1(b), $J \leq \lfloor N(x,z,t_i)/p_i \rfloor$ implies $\lceil p_i J \rceil \leq N(x,z,t_i)$. Now, we present a sequence of J outputs, which is an outputs sequence when transmitting either x or z. We place the first $\lceil p_1 J \rceil$ outputs in $I(x,z,t_1)$, then we add the next $(\lceil p_2 J \rceil - \lceil p_1 J \rceil)$ outputs in $I(x,z,t_2)$, then $(\lceil p_3 J \rceil - \lceil p_2 J \rceil)$ additional outputs in $I(x,z,t_3)$, and so on. This can be applied by the fact that $\lceil p_i J \rceil \leq N(x,z,t_i)$ for all $i \in [\ell]$. Thus, the first $\lceil p_i J \rceil$ of the outputs are in $B_{t_i}(x)$, for all $i \in [\ell]$, and the same holds for z. Hence, z might be decoded when x was transmitted. □

Next, we consider the non-sequenced case. Recall that if x is transmitted over a (T, P)-channel system of size N, then for all $i \in [\ell]$ at least $\lceil p_i N \rceil$ of the outputs are in $B_{t_i}(x)$ (if $i = \ell$ we have that all the N outputs are in $B_{t_\ell}(x)$). Then, x does not have exact reconstruction if there exists a different word z, where for all $i \in [\ell]$ at least $\lceil p_i N \rceil$ of the outputs are in $B_{t_i}(z)$.

Lemma 9: A non-sequenced-channel system of size J does not support exact reconstruction for U if and only if for all $i, j \in [\ell]$ the following inequality holds

$$N(x,z,t_i,t_j) \geq \lceil p_i J \rceil + \lceil p_j J \rceil - J.$$

Proof: For the first direction, we assume that the system does not support exact reconstruction. For all $i \in [\ell]$, we denote by A_i, B_i the sets of outputs in $B_{t_i}(x), B_{t_i}(z)$, respectively. Since the system does not support exact reconstruction for $U = \{x, z\}$, we conclude that for all i, j, $|A_i| \geq \lceil p_i J \rceil$, $|B_j| \geq \lceil p_j J \rceil$, and $J \geq |A_i \cup B_j|$ Thus, we have:

$$\begin{aligned} J &\geq |A_i \cup B_j| = |A_i| + |B_j| - |A_i \cap B_j| \\ &\geq \lceil p_i J \rceil + \lceil p_j J \rceil - N(x,z,t_i,t_j), \end{aligned}$$

as required.

In the second direction we are given that for all $i, j \in [\ell]$

$$N(x,z,t_i,t_j) \geq \lceil p_i J \rceil + \lceil p_j J \rceil - J,$$

and we prove that a channel system of size J does not support exact reconstruction for $U = \{x, z\}$. For this part, we present a set of J outputs, which any order of them may be an outputs sequence when transmitting either x or z. Let us assume that x is transmitted. The J outputs can be divided as follows. First, we place $\lceil p_1 J \rceil$ outputs in $I(x,z,t_1,t_\ell)$ by choosing the outputs according to their distance from z, preferring the closest. Then, in the second step we do the same for t_2 to have additional $(\lceil p_2 J \rceil - \lceil p_1 J \rceil)$ outputs in $I(x,z,t_2,t_\ell)$. Then, in the third step we do the same for t_3, and so on. Thus, we have J outputs for x which were chosen by ℓ steps.

We prove now, that this sequence of outputs, may be an outputs sequence when transmitting z. We have to prove that for each $i \in [\ell]$, at least $\lceil p_i J \rceil$ outputs are in $B_{t_i}(z)$. Let $i \in [\ell]$. Then, for each $j \in [\ell]$, we define

$$a_{i,j} = \min\{\lceil p_j J \rceil - \lceil p_{j-1} J \rceil, N(x,z,t_j,t_i) - a_{i,j-1}\},$$

where $p_0 = 0, a_{i,0} = 0$. By induction on j, it is readily proved, that for all $0 \leq j \leq \ell$, at the j-th step of the algorithm, we choose exactly $a_{i,j}$ additional words from $B_{t_i}(z)$ (actually, from $I(x,z,t_j,t_i)$). Therefore, $a_{i,j} \geq 0$ for all i, j, and the number of outputs in $B_{t_i}(z)$ is $a_i = \sum_{k=1}^{\ell} a_{i,k}$. Thus, we have to prove that $a_i \geq \lceil p_i J \rceil$. If for all j, $a_{i,j} = \lceil p_j J \rceil - \lceil p_{j-1} J \rceil$, then $a_i = J - \lceil p_0 J \rceil \geq \lceil p_i J \rceil$. Otherwise, let j_i be the largest index such that $a_{i,j_i} = N(x,z,t_{j_i},t_i) - a_{i,j_i-1}$. Then,

$$a_i \geq \sum_{k=j_i-1}^{\ell} a_{i,k} = J - \lceil p_{j_i} J \rceil + N(x,z,t_{j_i},t_i) \geq \lceil p_i J \rceil.$$

□

In order to continue with the analysis to study the value of $N^u(T, P, U, V)$, we define the following term:

$$\begin{aligned} N'(x,z,t_i,t_j,p_i,p_j) = \min\{L : \lceil p_i L \rceil + \lceil p_j L \rceil - L \\ > N(x,z,t_i,t_j), L \geq 1\} - 1. \end{aligned} \quad (2)$$

The following theorem establishes this result in calculating the value of $N^u(T, P, U, V)$ by using Lemma 9 and $N'(x,z,t_i,t_j,p_i,p_j)$.

Theorem 10: $N^u(T, P, U, V) = N + 1$, where

$$N = \min\{N'(x,z,t_i,t_j,p_i,p_j) : i, j \in [\ell]\},$$

that is,

$$\begin{aligned} N = \min\{&\lfloor N(x,z,t_i,t_\ell)/p_i \rfloor : i \in [\ell-1]\} \\ &\cup \{\lfloor N(z,x,t_i,t_\ell)/p_i \rfloor : i \in [\ell-1]\} \\ &\cup \{N(x,z,t_\ell)\} \\ &\cup \{N'(x,z,t_i,t_j,p_i,p_j) : i, j \in [\ell-1]\}. \end{aligned}$$

Proof: The proof where $N = \min\{N'(x,z,t_i,t_j,p_i,p_j) : i, j \in [\ell]\}$ is essentially similar to the proof of Theorem 4 by using Lemma 9. Therefore we omit this part. For the other version of N, note that by the definition of $N'(x,z,t_i,t_j,p_i,p_j)$ in Equation (2), the followings hold. For all $i \in [\ell-1]$, we have that $N'(x,z,t_i,t_\ell,p_i,p_\ell) = \lfloor N(x,z,t_i,t_\ell)/p_i \rfloor$, $N'(x,z,t_\ell,t_i,p_\ell,p_i) = \lfloor N(z,x,t_i,t_\ell)/p_i \rfloor$, and $N'(x,z,t_\ell,t_\ell,p_\ell,p_\ell) = N(z,x,t_\ell)$. □

The following remark generalizes Remark 5 for general ℓ.

Remark 11: We note that a (T, P)-non-sequenced-channel system of size J, where $J > N^u(T, P, \{x, z\}, V)$, may not *support exact reconstruction for $U = \{x, z\}$. Let J be the size of a (T, P)-non-sequenced-channel system, and*

$$N' = 1 + \min\{\lfloor N(x, z, t_i, t_\ell)/p_i \rfloor : i \in [\ell - 1]\} \cup \{\lfloor N(z, x, t_i, t_\ell/p_i) \rfloor : i \in [\ell - 1]\} \cup \{N(x, z, t_\ell)\}.$$

We can conclude the following.

- *If $J < N^u(T, P, \{x, z\}, V)$ then exact reconstruction is not supported.*
- *If $J = N^u(T, P, \{x, z\}, V)$ or $J \geq N'$ then exact reconstruction is supported.*
- *Otherwise, $N^u(T, P, \{x, z\}, V) < J < N'$, and we do not know whether exact reconstruction is supported or not.*

C. The t-Channel System

In this subsection we solve Problem 1 for the case where only t, the average number of errors in all the channels, is known. That is, we find the value of $N^a(t, U, V)$ for all $t \geq 0$ and $U \subseteq V$. We extend the distance function ρ to the domain $V^N \times V^N$ such that if $x, z \in V$ then $\rho(x^N, z^N) = N \cdot \rho(x, z)$ We also assume that if $\boldsymbol{x}, \boldsymbol{z} \in V^N \times V^N$ where $\rho(\boldsymbol{x}, \boldsymbol{y}) = d'$ then there exists $\boldsymbol{y} \in V^N$ such that $\rho(\boldsymbol{x}, \boldsymbol{y}) \leq \lfloor (d' + 1)/2 \rfloor$ and $\rho(\boldsymbol{z}, \boldsymbol{y}) \leq \lfloor (d' + 1)/2 \rfloor$. This assumption is reasonable in many types of error, for example, substitutions and transpositions, which are rigorously solved in Section IV for Problem 1. In this model we do not assume that the N outputs are different.

Lemma 12: Let $U \subseteq V$, such that $d(U) = d$. Then,

$$N^a(t, U, V) = \min\{N : \lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor\}.$$

Furthermore, a t-channel system of size N supports exact reconstruction for U if and only if $\lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor$.

Proof: The outputs of the N channels can be translated to one channel which causes at most $\lfloor tN \rfloor$ errors, and the transmitted word x is translated to transmitting x^N. Thus, if $U \subseteq V$ is a code with minimum distance d, then U^N is a code with minimum distance dN, consisting of N repetitions of each word from U. Therefore, this code can correct at most $\lfloor (dN - 1)/2 \rfloor$ errors. A t-channel system of size N with input U corresponds to one channel with the code U^N.

Thus, if $\lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor$ then we can use the minimum distance decoding algorithm in order to find out x^N from the code U^N. Otherwise, $\lfloor tN \rfloor > \lfloor (dN - 1)/2 \rfloor$ and let $x, z \in U$ such that $\rho(x, z) = d$. Then $\rho(x^N, z^N) = dN$, where ρ is the extended version of the distance function. By our assumption, there exists $\boldsymbol{y} \in V^N$ such that $\boldsymbol{y} \in B_{t'}(x^N) \cap B_{t'}(z^N)$ where $t' = \lfloor (dN + 1)/2 \rfloor = \lfloor (dN - 1)/2 \rfloor + 1 \leq \lfloor tN \rfloor$. We can present $\boldsymbol{y}$ as an output sequence of $(y_1, y_2, \ldots, y_N)$ of either x or z in a t-channel system of size N.

Hence, we conclude that a t-channel system of size N supports exact reconstruction for U if and only if $\lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor$, and

$$N^a(t, U, V) = \min\{N : \lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor\}.$$

□

Recall that $N^a(t, d, V)$ was defined as the maximum value of $N^a(t, U, V)$, for all $U \subseteq V$ such that $d(U) \geq d$. That is,

$$N^a(t, d, V) = \max\{N(t, U, V) : d(U) \geq d\}.$$

The next proposition proves that $N^a(t, d, V) = N^a(t, U, V)$ for all $U = \{x, z\} \subseteq V$ such that $d = \rho(x, z)$, that is $N^a(t, U, V) \geq N^a(t, U', V)$ where $d(U) = d$ and $d(U') = d' > d$. This desirable property was defined by Levenshtein in [9] as the *monotonicity by intersection*.

Proposition 13: Let $U = \{x, z\} \subseteq V$ such that $\rho(x, z) = d$. Then, $N^a(t, d, V) = N^a(t, U, V)$.

Proof: The first direction $N^a(t, d, V) \geq N^a(t, U, V)$ is trivial by the definitions of $N^a(t, d, V)$ and U.

For the second direction, assume to the contrary that $N^a(t, d, V) < N^a(t, U, V)$. By Lemma 12 we conclude that $N^a(t, U, V)$ depends on the distance between x and z, $d = \rho(x, z)$, and does not depend on the specific values of x or z. Furthermore, a t-channel system of size N supports exact reconstruction for U' of distance d' if and only if $\lfloor tN \rfloor \leq \lfloor (d'N - 1)/2 \rfloor$. Thus, $N^a(t, d, V) < N^a(t, U, V)$ implies $\lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor$, but, $\lfloor tN \rfloor > \lfloor (d'N - 1)/2 \rfloor$ for some $d' > d$. We get $\lfloor (d'N - 1)/2 \rfloor < \lfloor (dN - 1)/2 \rfloor$ for $d' > d$, which is a contradiction. Thus, $N^a(t, d, V) = N^a(t, U, V)$ for all $U = \{x, z\} \subseteq V$ of distance d. □

Following, in this subsection, we search for an explicit solution for $N^a(t, d, V)$, which, by Proposition 13 equals to $N^a(t, U, V)$ for all V of distance d. We prove some properties of this value and find the exact solution in almost all the cases.

Note that by applying Proposition 13, we can get the following result.

Lemma 14: For all $t \geq 0$ and a positive integer d, $N^a(t, d, V) = \infty$ or $N^a(t, d, V) = 1$.

Proof: Assume to the contrary that $1 < N = N^a(t, d, V)$ for $N \in \mathbb{N}$. Thus, by Lemma 12 we have $\lfloor t \rfloor > \lfloor (d - 1)/2 \rfloor$ and $\lfloor tN \rfloor \leq \lfloor (dN - 1)/2 \rfloor$. We prove by induction that if $\lfloor t \rfloor > \lfloor (d - 1)/2 \rfloor$, then for all $N' \in \mathbb{N}$, $\lfloor tN' \rfloor > \lfloor (dN' - 1)/2 \rfloor$. The basis of the induction is $N' = 1$. For the step, we assume that $\lfloor tN' \rfloor > \lfloor (dN' - 1)/2 \rfloor$, and we prove that $\lfloor t(N' + 1) \rfloor > \lfloor (d(N' + 1) - 1)/2 \rfloor$. The following inequality holds

$$\lfloor t(N' + 1) \rfloor \geq \lfloor tN' \rfloor + \lfloor t \rfloor > \lfloor (dN' - 1)/2 \rfloor + \lfloor (d - 1)/2 \rfloor + 1.$$

For even d we can continue

$$\begin{aligned} &\lfloor (dN' - 1)/2 \rfloor + \lfloor (d - 1)/2 \rfloor + 1 \\ &\quad = (dN' - 2)/2 + (d - 2)/2 + 1 \\ &\quad = (d(N' + 1) - 2)/2 \\ &\quad = \lfloor (d(N' + 1) - 1)/2 \rfloor, \end{aligned}$$

and, for odd d we continue as follows:

$$\begin{aligned} &\lfloor (dN' - 1)/2 \rfloor + \lfloor (d - 1)/2 \rfloor + 1 \\ &\quad \geq (dN' - 2)/2 + (d - 1)/2 + 1 \\ &\quad = (d(N' + 1) - 1)/2 \\ &\quad = \lfloor (d(N' + 1) - 1)/2 \rfloor. \end{aligned}$$

Thus, we conclude that $N^a(t, d, V) = \infty$ or $N^a(t, d, V) = 1$. □

Remark 15: We note that a t-channel system of size J, where $J > N^a(t, U, V)$, may not support exact reconstruction for U of distance d. That could happen since $\lfloor t \cdot N^a(t, U, V)\rfloor \le \lfloor (d \cdot N^a(t, U, V) - 1)/2\rfloor$ does not imply $\lfloor tJ\rfloor \le \lfloor (dJ - 1)/2\rfloor$ for $J > N^a(t, U, V)$, and by Lemma 12, a t-channel system of size J supports exact reconstruction for U if and only if $\lfloor tJ\rfloor \le \lfloor (dJ - 1)/2\rfloor$. For example, if $t = 7/4$ and $d = 3$ then $N^a(t, U, V) = 1$, but a t-channel system of size 2 (or any even size) does not support exact reconstruction for U of distance d.

As described in Remark 15, in the general case, a t-channel system of size J where $J > N^a(t, U, V)$ may not support exact reconstruction for U. Yet, in the rest of this subsection, we prove that in most of the cases, including for an integer t, this undesirable phenomena does not hold. Furthermore, in this section we prove several properties on $N^a(t, d, V)$, and we find an explicit solution for almost all the parameters t and d. The proofs of all the claims in the rest of this subsection are presented in Appendix A.

Lemma 16: If a t-channel system of size J, for some even J, supports exact reconstruction for U, then for all even positive integer, N, a t-channel system of size N supports exact reconstruction for U.

Lemma 17: Let $N^a(t, d, V) = 1$. If d is even or if a t-channel system of size 2 supports exact reconstruction for U of distance d, then for all $N \ge 1$ a t-channel system of size N supports exact reconstruction for U.

Theorem 18: Let $t \ge 0$, d be a positive integer, and $U \subseteq V$ where $d(U) = d$. Then

(1) If $d < \lfloor 2t\rfloor$ then $N^a(t, d, V) = \infty$,

(2) If $d > \lceil 2t\rceil$ then $N^a(t, d, V) = 1$, and exact reconstruction is supported for U for any size of the system.

(3) If $d = \lfloor 2t\rfloor$ and d is even then $N^a(t, d, V) = \infty$.

(4) If $d = \lfloor 2t\rfloor$ and d is odd then for all even N, a t-channel system of size N does not support exact reconstruction.

Thus, for an integer t, if $d > 2t$ then $N^a(t, d, V) = 1$, and exact reconstruction is supported for U for any size of the system. Otherwise, $N^a(t, d, V) = \infty$.

IV. Problem 1 - Examples

In this section we apply the solution for Problem 1 with $\ell = 2$ for two types of errors, for substitution errors in Subsection IV-A, and for transposition errors in Subsection IV-B. In this section we use the notations which were defined in Section II, where in some of the notations we add the subscript H or J, to denote the Hamming or the Johnson distance function for the case of substitution or transposition errors, respectively. We choose to focus here on the cases of substitutions and transpositions as these are examples for the errors that can be encountered when using DNA strands for storage. Substitutions correspond to the case where a base symbol is synthesized or sequenced with another symbol, and similarly a transposition is the case where two base symbols change their positions within the DNA strand [17].

A. Substitution Errors

Let $V_H = \{0, 1\}^n$ be the set of all length n words over the binary alphabet. The Hamming distance function $\rho_H : V_H \times V_H \to \mathbb{N}$ is defined by $\rho_H(\boldsymbol{x}, \boldsymbol{z}) = |\{i : x_i \ne z_i\}|$.

Note, that for all $\boldsymbol{x}, \boldsymbol{z} \in V_H$, $N_H(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ and $N_H(\boldsymbol{x}, \boldsymbol{z}, t)$ depend only on $d = \rho_H(\boldsymbol{x}, \boldsymbol{z})$. Thus, for $\boldsymbol{x}, \boldsymbol{z} \in V_H$ such that $d = \rho_H(\boldsymbol{x}, \boldsymbol{z})$, we denote by $N_H(d, n, t_1, t_2)$ and $N_H(d, n, t)$ the values $N_H(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ and $N_H(\boldsymbol{x}, \boldsymbol{z}, t)$, respectively. The next proposition proves that for all $d \ge 1$, $N^g_H(T, P, U, n) \ge N^g_H(T, P, U', n)$ where $d(U) = d$ and $d(U') = d' > d$. This desirable property is known as the monotonicity by intersection [9]. Recall that $N^g(T, P, d, V)$ was defined as the maximum value of $N^g(T, P, U, V)$, for all $U \subseteq V$ such that $d(U) \ge d$. That is,

$$N^g(T, P, d, V) = \max\{N^g(T, P, U, V) : d(U) \ge d\}.$$

Thus we prove in Proposition 19 that $N^g_H(T, P, d, n) = N^g_H(T, P, \{x, z\}, V)$ for all $U = \{x, z\} \subseteq V_H$ such that $d = \rho_H(x, z)$.

Proposition 19: Let $U = \{\boldsymbol{x}, \boldsymbol{z}\} \subseteq V_H$ such that $\rho_H(\boldsymbol{x}, \boldsymbol{z}) = d$. Then, $N^g_H(T, P, d, V) = N^g_H(T, P, U, V)$.

Proof: The first direction $N^g_H(T, P, d, V) \ge N^g_H(T, P, U, V)$ is trivial by the definitions of $N^g_H(T, P, d, V)$ and U. For the second direction, we note that in the Hamming case, $N^g(T, p, U, V)$ depends on the distance between x and z, $d = \rho(x, z)$, and does not depend on the specific values of x or z. It can be readily verified that $N_H(d, n, t)$, $N_H(d, n, t_1, t_2)$, and $N'_H(d, t_1, p)$ are non-increasing functions of d, Therefore, the function $N^g_H(T, P, U, V)$ for $U = \{\boldsymbol{x}, \boldsymbol{z}\} \subseteq \{0, 1\}^n$ of distance d is defined as the minimum between some of these non-increasing functions. Thus, for all U' of distance d', where $d' > d$, $N^g_H(T, P, U, n) \ge N^g_H(T, P, U', n)$, Hence $N^g_H(T, P, U, n) = N^g_H(T, P, d, n)$. □

According to Proposition 19, in order to calculate the value of $N^g_H(T, P, d, n)$ it is enough to find the value of $N^g_H(T, P, \{\boldsymbol{x}, \boldsymbol{z}\}, V_H)$, where $\rho_H(\boldsymbol{x}, \boldsymbol{z}) = d$ and $\boldsymbol{x}, \boldsymbol{z} \in V_H$. Therefore, according to Theorem 2 and Theorem 4, for $T = (t_1, t_2)$ and $P = (p, 1)$, we conclude that

$$N^k_H(T, P, d, n) = \min\{\lfloor N_H(d, n, t_1)/p\rfloor, N_H(d, n, t_2)\} + 1, \tag{3}$$

and

$$N^u_H(T, P, d, n) = \min\{\lfloor N_H(d, n, t_1, t_2)/p\rfloor, N_H(d, n, t_2), N'_H(d, t_1, p)\} + 1, \tag{4}$$

where

$$N'_H(d, t_1, p) = \min\{L : 2\lceil pL\rceil - L > N_H(d, n, t_1, L \ge 1\} - 1.$$

In the sequel, we find the values of $N^k_H(T, P, d, n)$ and $N^u_H(T, P, d, n)$. We start by computing the values of $N_H(d, n, t)$ and $N_H(d, n, t_1, t_2)$.

The following lemma was shown in [9], where we use the equality $t - \left\lceil \frac{d}{2}\right\rceil = \left\lfloor t - \frac{d}{2}\right\rfloor$.

Lemma 20: For $t, d \geq 1$,

$$N_H(d,n,t) = \sum_{i=0}^{\lfloor t-\frac{d}{2}\rfloor} \binom{n-d}{i} \cdot \sum_{k=d-t+i}^{t-i} \binom{d}{k},$$

where $\binom{a}{b} = 0$ if $a < b$ or $b < 0$.

We conclude that in the sequenced model the solution is simply derived by the expression in (3) while we use the result from Lemma 20. In particular, for a fixed p and n sufficiently large, we get that $N_H^k(T, P, d, n) = \lfloor N_H(d, n, t_1)/p \rfloor$. Furthermore, at the end of this subsection, we provide some examples.

For the non-sequenced model, we compute the value of $N_H(d, n, t_1, t_2)$ for all t_1, t_2. This value is presented in the following lemma, which generalizes Lemma 20, and is proved by similar combinatorial computation. The value $N_H(d, n, t)$, which is presented in Lemma 20, can be obtained from the Lemma 21, by substituting $t = t_1 = t_2$. For the completeness and readability of the results in the paper we prove the three following lemmas in Appendix B.

Lemma 21: For $1 \leq t_1 \leq t_2$ and $d \geq 1$,

$$N_H(d,n,t_1,t_2) = \sum_{i=0}^{\lfloor \frac{t_1+t_2-d}{2}\rfloor} \binom{n-d}{i} \cdot \sum_{k=d-t_2+i}^{t_1-i} \binom{d}{k}.$$

The following two lemmas compare between the three components which determine the value of $N_H^u(T, P, d, n)$, for $d \geq 1$, $t_1 < t_2 \in \mathbb{N}$, and fixed $0 < p < 1$. Lemma 22 compares between $\lfloor N_H(d, n, t_1, t_2)/p \rfloor$ and $N_H(d, n, t_2)$.

Lemma 22: For any fixed p and n sufficiently large the following holds. If d is odd, $p \leq 1/2$, and $t_2 = t_1 + 1$, then

$$N_H(d, n, t_2) < \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Otherwise,

$$N_H(d, n, t_2) \geq \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

The following lemma compares between the values of $N'_H(d, t_1, p)$ and $\min\{\lfloor N_H(d, n, t_1, t_2)/p \rfloor, N(d, t_2)\}$. Recall that according to Proposition 6, for $0 < p \leq 1/2$, $N'_H(d, t_1, p) \in \{0, \infty\}$, and by Lemma 22 if $1/2 < p < 1$ then $N_H(d, n, t_1, t_2)/p \leq N_H(d, n, t_2)$. Thus, in Lemma 23 we compare only between $\lfloor N_H(d, n, t_1, t_2)/p \rfloor$ and $\left\lfloor \frac{N_H(d,n,t_1)}{2p-1} \right\rfloor$ for $1/2 < p < 1$.

Lemma 23: For any fixed p and n sufficiently large the following holds. If d is even, $t_2 = t_1 + 1$, and ($(1/2 < p \leq 2/3)$ or $(2/3 < p < 3/4$ and $d < \frac{2-2p}{3p-2})$), then

$$\left\lfloor \frac{N_H(d, n, t_1)}{2p - 1} \right\rfloor > \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Otherwise,

$$\left\lfloor \frac{N_H(d, n, t_1)}{2p - 1} \right\rfloor \leq \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

According to Corollary 7, Lemma 22, and Lemma 23, we can now summarize the results for the case of binary substitution errors.

Corollary 24: For any fixed p and n sufficiently large the following holds.

- *For $0 < p \leq 1/2$:*

$$N_H^u(T, P, d, n) = \begin{cases} 1 & \text{if } d > 2t_1, \\ \Theta(n^{\lfloor \frac{t_1+t_2-d}{2} \rfloor}) & \text{otherwise.} \end{cases}$$

- *For $1/2 < p < 1$:*

$$N_H^u(T, P, d, n) = \Theta(n^{\lfloor \frac{2t_1-d}{2} \rfloor}).$$

More specifically,

- *For $0 < p \leq 1/2$:*

$$N_H^u(T, P, d, n) = \begin{cases} 1 & \text{if } d > 2t_1, \\ N_H(d, n, t_2) + 1 & \text{otherwise, if } d \text{ is odd and } t_2 = t_1 + 1, \\ \left\lfloor \frac{N_H(d,n,t_1,t_2)}{p} \right\rfloor s + 1 & \text{otherwise.} \end{cases}$$

- *For $1/2 < p < 1$:*

$$N_H^u(T, P, d, n) = \begin{cases} \left\lfloor \frac{N_H(d,n,t_1,t_2)}{p} \right\rfloor + 1 & \text{if } d \text{ is even}, t_2 = t_1 + 1, \text{ and } \left(\left(\frac{1}{2} < p \leq \frac{2}{3}\right) \vee \left(\frac{2}{3} < p < \frac{3}{4} \wedge d < \frac{2-2p}{3p-2}\right) \right), \\ N'_H(d, t_1, p) + 1 & \text{otherwise.} \end{cases}$$

To understand the results for substitution errors errors better, we demonstrate some of them. Let $L_1 = N_H(d, n, t_1) + 1$ and $L_2 = N_H(d, n, t_2) + 1$ be the solutions for the cases where all the channels are identical and cause at most t_1 and t_2 errors, respectively. In addition, for $T = (t_1, t_2)$ and $P = (p, 1)$, we denote $L^u = N_H^u(T, Pd, n)$ and $L^k = N_H^k(T, P, d, n)$. That is, L^u and L^k are the solutions to Problem 1 with $\ell = 2$ for the non-sequenced and for the sequenced models, respectively. Clearly, $L_1 \leq L^k \leq L^u \leq L_2$.

Note that by Levenshtein's result [9] and Lemma 20, if $d \leq 2t_1$ then $L_1 = \Theta(n^{\lfloor t_1 - d/2 \rfloor})$ and $L_2 = \Theta(n^{\lfloor t_2 - d/2 \rfloor})$. Using Theorem 2 and the Equation (3), we have $L^k = \Theta(n^{\lfloor t_1 - d/2 \rfloor})$, and similarly the value of L^u is derived by Corollary 24.

The following examples compare between the four values, L_1, L_2, L^k, and L^u. These examples emphasize the benefit produced from knowing that a fraction p of the channels are t_1-error, and not t_2-error, where $t_1 < t_2$. Additionally, these examples highlight the advantage of the sequenced model on the non-sequenced one.

- For fixed p, $0 < p \leq 1/2$, $d = 1$, and $T = (2, 4)$, we have $L_1 = L^k = \Theta(n)$ and $L^u = \Theta(n^2)$, while $L_2 = \Theta(n^3)$,
- For fixed p, $0 < p \leq 1/2$, $d = 1$, and $T = (2, 8)$, we have $L_1 = L^k = \Theta(n)$ and $L^u = \Theta(n^4)$, while $L_2 = \Theta(n^7)$,
- For fixed p, $1/2 < p \leq 2/3$, $d = 2$, and $T = (4, 5)$, we have $L_1 = L^k = L^u = \Theta(n^3)$, while $L_2 = \Theta(n^4)$.

For $T = (t_1, t_2)$, where $t_2 > t_1$, and a fixed p, $0 < p < 1$, the cases of $d = 2t_1 - 1$ or $d = 2t_1$ are interesting, since by Lemma 20, in both cases $N_H(d,n,t_1) = \binom{2t_1}{t_1}$, and thus L_1 is independent on n. For these parameters, we have $L_1 = \binom{2t_1}{t_1} + 1$, $L^k = \left\lfloor \binom{2t_1}{t_1}/p \right\rfloor + 1$, $L^u = \Theta(1)$ for $p > 1/2$, $L^u = \Theta(n^{\lfloor \frac{t_1+t_2-d}{2} \rfloor})$ for $p \le 1/2$, and $L_2 = \Theta(n^{\lfloor \frac{2t_2-d}{2} \rfloor})$.

B. Transposition Errors

For $\boldsymbol{x} \in \{0,1\}^n$, a transposition error transposes the symbols x_i and x_j for some $i, j \in [N]$, $i \neq j$. Note that transpositions do not change the Hamming weight of a word. Therefore we consider $V_J = J_w^n$, the set of all length n words over the binary alphabet with Hamming weight w. The Johnson distance function $\rho_J : V_J \times V_J \to \mathbb{N}$ is defined by $\rho_J(x,z) = \frac{|\{i: x_i \neq z_i\}|}{2}$.

Note, that for all $\boldsymbol{x}, \boldsymbol{z} \in V_J$, $N_J(x,z,t_1,t_2)$ and $N_J(x,z,t)$ depend only on $d = \rho_J(x,z)$. Thus, for $\boldsymbol{x}, \boldsymbol{z} \in V_J$ such that $d = \rho_J(x,z)$, we denote by $N_J(d,n,t_1,t_2)$ and $N_J(d,n,t)$ the values $N_J(\boldsymbol{x},\boldsymbol{z},t_1,t_2)$ and $N_J(\boldsymbol{x},\boldsymbol{z},t)$, respectively. Note that all these values depend also on w which is omitted to simplify the notations. The monotonicity by intersection property holds also for this case, as described in the following proposition. The proof is omitted since it essentially the same as the proof of Proposition 19 for substitution errors.

Proposition 25: Let $U = \{\boldsymbol{x}, \boldsymbol{z}\} \subseteq V_H$ such that $\rho_J(\boldsymbol{x},\boldsymbol{z}) = d$. Then, $N_J^g(T,P,d,V) = N_J^g(T,P,U,V)$.

According to Proposition 25, in order to calculate the value of $N_J^g(T,P,d,n)$ it is enough to find the value of $N_J^g(T,P,\{\boldsymbol{x},\boldsymbol{z}\},V_J)$, for some $\boldsymbol{x},\boldsymbol{z}$ such that $\rho_J(\boldsymbol{x},\boldsymbol{z}) = d$. Therefore, according to Theorem 2 and Theorem 4, for $T = (t_1,t_2)$ and $P = (p,1)$, we conclude that

$$N_J^k(T,P,d,n) = \min\{\lfloor N_J(d,n,t_1)/p \rfloor, N_J(d,n,t_2)\} + 1, \tag{5}$$

and

$$N_J^u(T,P,d,n) = \min\{\lfloor N_J(d,n,t_1,t_2)/p \rfloor, N_J(d,n,t_2), N_J'(d,t_1,p)\} + 1, \tag{6}$$

where

$$N_J'(d,t_1,p) = \min\{L : 2\lceil pL \rceil - L > N_J(d,n,t_1), L \ge 1\} - 1.$$

The following lemma was shown in [9].

Lemma 26: For $t, d \ge 1$,

$$N_J(d,n,t) = \sum_{i=0}^{t} \binom{n-w-d}{i} \cdot \sum_{a=0}^{t-i}\sum_{b=0}^{t-i} \binom{d}{a}\binom{d}{b}\binom{w-d}{a+b+i-d},$$

where $\binom{a}{b} = 0$ if $a < b$ or $b < 0$.

Thus, as in the substitution errors case, we conclude that the solution for the sequenced model is directly deduced from (5) together with the result from Lemma 26. We also note that for a fixed p and n sufficiently large, $N_J^k(T,P,d,n) = \lfloor N_J(d,n,t_1)/p \rfloor$. At the end of this subsection, we provide some examples for this value.

For the non-sequenced model, we compute the value of $N_J(d,n,t_1,t_2)$ for all t_1, t_2. This value is presented in the following lemma, which generalizes Lemma 26, and is proved by similar combinatorial computation. The value $N_J(d,n,t)$, which is presented in Lemma 26, can be obtained from Lemma 27, by substituting $t = t_1 = t_2$. For the completeness of the results in the paper we prove the following lemma in Appendix B.

Lemma 27: For $t_1 \le t_2$:

$$N_J(d,n,t_1,t_2) = \sum_{i=0}^{t_1} \binom{n-w-d}{i} \cdot \sum_{a=0}^{t_1-i}\sum_{b=0}^{t_2-i} \binom{d}{a}\binom{d}{b}\binom{w-d}{a+b+i-d}.$$

The following two lemmas compare between the three components which determine the value of $N_J^u(T,P,d,n)$, for $d \ge 1$, $t_1 < t_2 \in \mathbb{N}$, and fixed $0 < p < 1$. Lemma 28 compares between $\lfloor N_J(d,n,t_1,t_2)/p \rfloor$ and $N_J(d,n,t_2)$.

Lemma 28: For any fixed p and n sufficiently large

$$N_J(d,n,t_2) > \lfloor N_J(d,n,t_1,t_2)/p \rfloor.$$

Proof: Note that

$$N_J(d,n,t_2) = \Theta(n^{t_2}) \text{ and } N_J(d,n,t_1,t_2) = \Theta(n^{t_1}).$$

Thus, for $t_2 > t_1$, $N_J(d,n,t_2) > \lfloor N_J(d,n,t_1,t_2)/p \rfloor$. □

The following lemma compares between the values of $N_J'(d,t_1,p)$ and $\min\{\lfloor N_J(d,n,t_1,t_2)/p \rfloor, N_J(d,n,t_2)\}$. Recall that according to Proposition 6, for $0 < p \le 1/2$, $N_J'(d,t_1,p) \in \{0,\infty\}$, and by Lemma 28 if $1/2 < p < 1$ then $\lfloor N_J(d,n,t_1,t_2)/p \rfloor \le N_J(d,n,t_2)$. Thus, in Lemma 29 we compare only between $\lfloor N_J(d,n,t_1,t_2)/p \rfloor$ and $\left\lfloor \frac{N_J(d,n,t_1)}{2p-1} \right\rfloor$ for $1/2 < p < 1$. The proof of this Lemma is presented in Appendix B.

Lemma 29: For any fixed p and n sufficiently large the following holds.

$$\left\lfloor \frac{N_J(d,n,t_1)}{2p-1} \right\rfloor < \lfloor N_J(d,n,t_1,t_2)/p \rfloor$$

if and only if

$$\frac{p}{2p-1} < \sum_{b=0}^{t_2-t_1} \binom{d}{b} \frac{1}{(t_1-d+b)^2 \cdots (t_1-d+1)^2}.$$

According to Corollary 7, Lemma 28, and Lemma 29, we can now summarize the results for transposition errors.

Corollary 30: For any fixed p and n sufficiently large the following holds. If $0 < p \le 1/2$ and $d > 2t_1$ then $N_J^u(T,P,d,n) = 1$. Otherwise, $N_J^u(T,P,d,n) = \Theta(n^{t_1})$ More specifically,

- *For $0 < p \le 1/2$:*

$$N_J^u(T,P,d,n) = \begin{cases} 1 & \text{if } d > 2t_1, \\ \lfloor N_J(d,n,t_1,t_2)/p \rfloor + 1 & \text{otherwise.} \end{cases}$$

- *For $1/2 < p < 1$:*

$$N_J^u(T,P,d,n) = \min\left\{\left\lfloor \frac{N_J(d,n,t_1,t_2)}{p} \right\rfloor, N_J'(d,t_1,p)\right\} + 1 = \Theta(n^{t_1}).$$

To understand the results for transposition errors better, we demonstrate some of them. Let $L_1 = N_J(d,n,t_1) + 1$ and $L_2 = N_J(d,n,t_2) + 1$ be the solutions for the cases where all the channels are identical and cause at most t_1 and t_2 errors, respectively. Additionally, for $T = (t_1, t_2)$ and $P = (p, 1)$, we denote $L^u = N^u_J(T, P, d, n)$ and $L^k = N^k_J(T, P, d, n)$. That is, L^u and L^k are the solutions to Problem 1 with $\ell = 2$ for the non-sequenced and for the sequenced models, respectively. Clearly, $L_1 \le L^k \le L^u \le L_2$.

Note that by Levenshtein [9] and Lemma 26, if $d \le 2t_1$ then $L_1 = \Theta(n^{\lfloor t_1 \rfloor})$ and $L_2 = \Theta(n^{\lfloor t_2 \rfloor})$. Using Theorem 2 and Equation (5) we have $L^k = \Theta(n^{\lfloor t_1 \rfloor})$, and the value of L^u is similarly derived by Corollary 30.

The following examples compare between the four values, L_1, L_2, L^k, and L^u. These examples emphasize the benefit produced from knowing that a fraction p of the channels are t_1-error, and not t_2-error, where $t_1 < t_2$. Note that for the Johnson distance function we have $\Theta(L^k) = \Theta(L^u)$.

- For fixed p, $d = 1$, and $T = (2, 4)$: $L_1 = L^k = L^u = \Theta(n^2)$, while $L_2 = \Theta(n^4)$,
- For fixed p, $d = 1$, and $T = (2, 8)$: $L_1 = L^k = L^u = \Theta(n^2)$, while $L_2 = \Theta(n^8)$,
- For fixed p, $d = 2$, and $T = (4, 5)$: $L_1 = L^k = L^u = \Theta(n^4)$, while $L_2 = \Theta(n^5)$.

V. Problem 2 - Minimum Distance

In this section we solve Problem 2. That is, we assume that the size of the system, N, is given, and we study the minimum distance of the code, d, that is required for exact reconstruction. Let $t_1 \le t_2 \le \cdots \le t_N$ be N positive integers that denote the maximum number of errors in the channels. In a similar way to Problem 1, we consider the following three cases:

(1) a (T, N)-sequenced-channel system, where the N-tuple $T = (t_1, t_2, \ldots, t_N)$ is known,
(2) a (T, N)-non-sequenced-channel system, where the multiset $T = \{t_1, t_2, \ldots, t_N\}$ is known, and
(3) a (t, N)-channel system, where $t = \frac{\sum_{i=1}^N t_i}{N}$ is known.

The solution to Problem 2 is given by reduction to Problem 1. Let $T = (t_1, \ldots, t_N)$ where $t_1 \le t_2 \le \cdots \le t_N \in \mathbb{N}$, $T = (t_1, \ldots, t_N)$ denotes the multiset $T = \{t_1, t_2, \ldots, t_N\}$ for the non-sequenced model, or the N-tuple, $T = (t_1, t_2, \ldots, t_N)$, in the sequenced case, where the exact meaning will be clear from the context. We define $T' = (t'_1, t'_2, \ldots, t'_\ell)$ which consists of the set of T, where $t_1 = t'_1 < t_2 < \cdots < t_\ell = t_N$. We define the function $f : [N] \to [\ell]$ such that $f(i) = i'$ if $t_i = t'_{i'}$, and the function $g : [\ell] \to [N]$ is defined as follows: $g(i') = \max\{i : f(i) = i'\}$. We also define $P = (p_1, p_2, \ldots, p_\ell)$, such that for $i' \in [\ell]$, $p_j = \frac{g(i')}{N}$.

The following theorem establishes the connection between a solution to Problem 1 and Problem 2, by using the definition of T' and P. We remind that $D^k(T, N, V)$, $D^u(T, N, V)$, and $D^a(t, N, V)$ are the minimum distances of the codes which are required for exact reconstruction in the sequenced, non-sequenced, and average models, respectively. Clearly, by the definitions of the models,

$$D^k(T, N, V) \le D^u(T, N, V) \le D^a(t, N, V).$$

We note that as in Problem 1, in the two first cases we assume that the N outputs are different, and in the third case we assume, as describe in Subsection III-C, that if $\boldsymbol{x}, \boldsymbol{z} \in V^N$ such that $\rho(\boldsymbol{x}, \boldsymbol{z}) = d'$, then there exists $\boldsymbol{y} \in V^N$ such that $\boldsymbol{y} \in B_{t'}(\boldsymbol{x}) \cap B_{t'}(\boldsymbol{z})$ for $t' = \lfloor (d' + 1)/2 \rfloor$.

Theorem 31: For all $N \ge 1$ the following properties hold:

- *for the average model,*
$$\lfloor 2t \rfloor \le D^a(t, N, V) \le \lceil 2t \rceil + 1,$$
and if t is integer then $D^a(t, N, V) = 2t + 1$,
- *for the non-sequenced model,*
$$D^u(T, N, V) = \min\{d : \text{A } (T', P)\text{-non-sequenced-channel system of size } N \text{ supports exact reconstruction for all } U \subseteq V \text{ such that } d(U) \ge d\},$$
- *and for the sequenced model,*
$$D^k(T, N, V) = \min\{d : N^k(T', P, d, V) \le N\}.$$

Proof: The solutions for all cases are derived immediately by reduction to Problem 1. For the non-sequenced case we could not find a more explicit solution, since $N > N^u(T', P, U, V)$ does not guarantee an exact reconstruction as explained in Remark 5. However for the average model we apply Theorem 18 to find an explicit solution, and for the sequenced model we could simplify the solution to Problem 2 by the property that a (T', P)-sequenced-channel system of size N supports exact reconstruction for $U \subseteq V$ if and only if $N \ge N^k(T', P, U, V)$.

In more details, for the average model, by the reduction to Problem 1, it can be readily verified that

$$D^a(t, N, V) = \min\{d : \text{a } t\text{-channel system of size } N \text{ supports exact reconstruction for all } x, z \text{ such that } \rho(x, z) \ge d\}.$$

Using Theorem 18 we conclude the value of $D^a(t, N, V)$.

The sequenced model is derived similarly. By the reduction to Problem 1 and by (T', P) definition, we have

$$D^k(T, N, V) = \min\{d : \text{A } (T', P)\text{-sequenced-channel system of size } N \text{ supports exact reconstruction for all } U \subseteq V \text{ such that } d(U) \ge d\},$$

and since a (T', P)-sequenced-channel system of size N supports exact reconstruction for U if and only if $N \ge N^k(T', P, U, V)$, we have

$$D^k(T, N, V) = \min\{d : N^k(T', P, d, V) \le N\}.$$

Yet, for the non-sequenced model, by Remark 5, a (T', P)-non-sequenced-channel system of size N, $N > N^k(T', P, U, V)$, might not support exact reconstruction for U. Therefore, in this model, unlike in the sequenced case, we could not simplify the result which obtained by the reduction to Problem 1. □

For the average model, Theorem 31 provides an explicit solution to Problem 2, that is, $D^a(t, N, V)$ has an explicit expression. But, for the other two models, even though Theorem 31 provides a solution to Problem 2, it does not give an explicit expression for the minimum distance. Thus, in the rest of this section we will show how to derive an explicit solution to Problem 2 in these two models for both substitution and transposition errors. The technique which will be presented can be used also for other distance functions. We first prove Lemma 32, which provides an insight about the reduction to Problem 1 for the general case. Then, we apply this lemma to find an explicit solution, for transpositions in Theorem 33, and for substitutions in Theorem 34.

Let T, N, T' and P be as defined earlier in this section, where the (T, N)-channel system is given, and (T', P) is defined by the reduction. In Subsection III-B, we presented conditions for supporting exact reconstruction for U in a (T', P)-channel system of size N. The following lemma provides equivalent conditions on the given parameters T and N. For the sequenced model, by Theorem 8 and Proposition 1, a (T', P)-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x}, \boldsymbol{z}\}$ if and only if there exists $i' \in [\ell]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) < \lceil p_{i'} N\rceil$. Similarly, in the non-sequenced model, by Lemma 9, a (T', P)-non-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x}, \boldsymbol{z}\}$ if and only if there exist $i', j' \in [\ell]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}, t'_{j'}) < \lceil p_{i'} N\rceil + \lceil p_{j'} N\rceil - N$. The following lemma, establishes an equivalent condition on T and N.

Lemma 32: Let $T = (t_1, \dots, t_N)$ such that $t_1 \le t_2 \le \cdots, t_N \in \mathbb{N}$, and let $T' = (t'_1, \dots, t'_\ell)$, f, g, and $P = (p_1, \dots, p_\ell)$ as defined earlier in this section. Then, the followings hold:

- *there exists $i' \in [\ell]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) < \lceil p_{i'} N\rceil$ if and only if there exists $i \in [N]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t_i) < i$, and $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) = N(\boldsymbol{x}, \boldsymbol{z}, t_i)$,*
- *there exist $i', j' \in [\ell]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}, t'_{j'}) < \lceil p_{i'} N\rceil + \lceil p_{j'} N\rceil - N$ if and only if there exist $i, j \in [N]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t_i, t_j) < i + j - N$, and $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}, t'_{j'}) = N(\boldsymbol{x}, \boldsymbol{z}, t_i, t_j)$.*

Proof: For the easier direction, assume that there exists $i' \in [\ell]$ such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) < \lceil p_{i'} N\rceil$. Let $i = g(i') \in [N]$, then, by the definitions of f, g, and P we have: $N(\boldsymbol{x}, \boldsymbol{z}, t_{g(i')}) < \left\lceil \frac{g(i')}{N} \cdot N \right\rceil$, that is, $N(\boldsymbol{x}, \boldsymbol{z}, t_i) < i$. Similarly, if there exist $i', j' \in [\ell]$, such that $N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}, t'_{j'}) < \lceil p_{i'} N\rceil + \lceil p_{j'} N\rceil - N$, then we can conclude that for $i = g(i')$ and $j = g(j')$ $N(\boldsymbol{x}, \boldsymbol{z}, t_i, t_j) < i + j - N$.

For the second direction, assume that $N(\boldsymbol{x}, \boldsymbol{z}, t_i) < i$, and let $r = \max\{k : f(k) = f(i)\}$, and $i' = f(i) = f(r)$. That is, $t_r = t_i = t'_{i'}$, $i \le r$, $r = g(i')$, and $p_{i'} = \frac{r}{N}$. Thus we conclude,

$$N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) = N(\boldsymbol{x}, \boldsymbol{z}, t_i) < i \le r = p_{i'} N = \lceil p_{i'} N\rceil .$$

The second part in the direction is proved similarly. We assume that $N(\boldsymbol{x}, \boldsymbol{z}, t_i, t_j) < i + j - N$, and we define $r_i = \max\{k : f(k) = f(i)\}$, $r_j = \max\{k : f(k) = f(j)\}$, $i' = f(i) = f(r_i)$, and $j' = f(i) = f(r_j)$. That is, $t_{r_i} = t_i = t'_{i'}$, $i \le r_i$, $r_i = g(i')$, and $p_{i'} = \frac{r_i}{N}$, and the same holds for j. Thus we conclude,

$$\begin{aligned} N(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}, t'_{j'}) &= N(\boldsymbol{x}, \boldsymbol{z}, t_i, t_j) < i + j - N \\ &\le r_i + r_j - N \\ &= \lceil p_{i'} N\rceil + \lceil p_{j'} N\rceil - N. \end{aligned}$$

□

Now, we focus on substitution and transposition errors. As mentioned in Section IV, for each $n \in \mathbb{N}$ and $\boldsymbol{x}, \boldsymbol{z} \in \{0, 1\}^n$, $N_H(\boldsymbol{x}, \boldsymbol{z}, t, t')$ and $N_J(\boldsymbol{x}, \boldsymbol{z}, t, t')$ depend on $d = \rho(\boldsymbol{x}, \boldsymbol{z})$ and n, and do not depend on the specific words $\boldsymbol{x}$ and $\boldsymbol{z}$. Thus, in Section IV we denoted the values of $N_H(\boldsymbol{x}, \boldsymbol{z}, t, t')$ and $N_J(\boldsymbol{x}, \boldsymbol{z}, t, t')$ by $N_H(d, n, t, t')$ and $N_J(d, n, t, t')$ for the Hamming and for the Johnson cases, respectively. We also use Propositions 19 and 25 for substitutions and transpositions, respectively. These propositions allow us to replace between $N_H(\boldsymbol{x}, \boldsymbol{z}, t, t')$, $N_J(\boldsymbol{x}, \boldsymbol{z}, t, t')$ and $N(d, n, t, t')$, $N(d, n, t, t')$ where $\boldsymbol{x}, \boldsymbol{z} \in \{0, 1\}^n$ and $\rho(\boldsymbol{x}, \boldsymbol{z}) = d$.

In addition, by Lemma 27, for all d, t, and t', the following holds $N_J(d, n, t, t') \to \infty$ as $n \to \infty$, or $N_J(d, n, t, t') \equiv 0$ (i.e., $N_J(d, n, t, t') = 0$ for all $n \in \mathbb{N}$). For the Hamming distance function, by Lemma 21, we have the same property for almost all the parameters d, t, and t'. In general we use the notation $N(d, n, t, t')$ (without the subscript J or H), and if $N(d, n, t, t')$ holds the property described in this paragraph, we say that $N(d, n, t, t')$ *increases on n* for d, t, and t'.

In the sequel, we apply Lemma 32 to find an explicit solution to Problem 2 for the sequenced and for the non-sequenced models, for both substitution and transposition errors over the binary alphabet. We use the property of $N(d, n, t, t')$ to simplify the solution. The same technique can be applied to find an explicit solution for other distance functions in which $N(d, n, t, t')$ increases on n.

We start with Theorem 33 for transposition errors, since this distance function holds the property of $N(d, n, t, t')$ for all parameters d, t, and t', that is for the Johnson distance function, $N_J(d, n, t, t')$ increases on n for all parameters d, t, and t'.

Theorem 33: Let $V_J = J^n_w$, ρ_J is the Johnson distance function, N is a positive constant number, and $T = (t_1, \dots, t_N)$ where $t_1 \le \cdots \le t_N \in \mathbb{N}$. Then, for n sufficiently large,

$$D^k_J(T, N, V_J) = 2t_1 + 1$$

and

$$D^u_J(T, N, V_J) = \min\{t_i + t_j + 1 : i, j \in [N], i + j = N + 1\}.$$

Proof: For the Johnson distance function, by Lemma 27, $N_J(d, n, t, t')$ increases on n, for all t, t' and d. That is, $N_J(d, n, t, t') = 0$ for all n, or $N_J(d, n, t, t') \to \infty$ as $n \to \infty$. Let $T' = (t'_1, t'_2, \dots, t'_\ell)$ and $P = (p_1, p_2, \dots, p_\ell)$ be defined as described earlier in this section.

Let $U = \{\boldsymbol{x}, \boldsymbol{z}\}$ be of distance d. Thus, by Theorem 8 and Proposition 1, a (T', P)-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x}, \boldsymbol{z}\}$ of distance d if and only if there exists $i' \in [\ell]$ such that $N_J(d, n, t'_{i'}) = N_J(\boldsymbol{x}, \boldsymbol{z}, t'_{i'}) < \lceil p_{i'} N\rceil$, and by Lemma 32 this holds if and only if there exists $i \in [N]$ such that $N_J(d, n, t_i) < i$. But, $N_J(d, n, t_i) = N_J(d, n, t_i, t_i)$ increases on n, and therefore, for n sufficiently large, $N_J(d,n,t_i) < i$ implies $N_J(d,n,t_i) = 0$ which is equivalent to $d > 2t_i$. Recall that $D_J^k(T,N,V_H)$ equals to the minimum between all these possible d. Thus, we can conclude that in the sequenced model, for n sufficiently large,

$$D_J^k(T,N,V_H) = \min\{2t_i+1 : i \in [N]\} = 2t_1+1.$$

Similarly, in the non-sequenced model, by Lemma 9, a (T',P)-non-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x},\boldsymbol{z}\}$ of distance d if and only if there exist $i',j' \in [\ell]$ such that $N_J(d,n,t'_{i'},t'_{j'}) = N_J(\boldsymbol{x},\boldsymbol{z},t'_{i'},t'_{j'}) < \lceil p_{i'}N\rceil + \lceil p_{j'}N\rceil - N$, and by Lemma 32 this holds if and only if there exist $i,j \in [N]$ such that $N_J(d,n,t_i,t_j) < i+j-N$. But, $N_J(d,n,t_i,t_j)$ increases on n, and therefore, for n sufficiently large, $N_J(d,n,t_i,t_j) < i+j-N$ implies $N_J(d,n,t_i,t_j)=0$ and $i+j-N>0$, where the condition $N_J(d,n,t_i,t_j)=0$ is equivalent to $d > t_i+t_j$. Recall that $D_J^u(T,N,V_H)$ equals to the minimum d required for exact reconstruction. Thus, we can conclude that in the non-sequenced model, for n sufficiently large,

$$\begin{aligned} D_J^u(T,N,V_H) &= \min\{t_i+t_j+1 : i,j\in[N], i+j-N>0\} \\ &= \min\{t_i+t_j+1 : i,j\in[N], i+j=N+1\}. \end{aligned}$$

□

The following theorem provides an explicit solution to Problem 2 for the sequenced and for the non-sequenced models for substitution errors, by applying Theorem 31, and Lemma 32. Note that in this case $N_H(d,n,t,t')$ increases on n for almost all the parameters d, t, and t', and we use this property in the proof.

Theorem 34: Let $V_H = \{0,1\}^n$, ρ_H is the Hamming distance function, N is a positive constant number, and $T = (t_1,\ldots,t_N)$ where $t_1 \le \cdots \le t_N \in \mathbb{N}$. Then, for n sufficiently large,

$$D_H^k(T,N,V_H) = \min\{D_1,D_2,D_3\}$$

where

$$\begin{aligned} D_1 &= 2t_1+1, \\ D_2 &= \min\left\{2t_i : i\in[N], \binom{2t_i}{t_i} < i\right\}, \text{ and} \\ D_3 &= \min\left\{2t_i-1 : i\in[N], \binom{2t_i}{t_i} + \binom{2t_i}{t_i-1} < i\right\}, \end{aligned}$$

and

$$D_H^u(T,N,V_H) = \min\{D_1,D_2,D_3\},$$

where

$$\begin{aligned} D_1 &= \min\{t_i+t_j+1 : i,j\in[N], i+j=N+1\}, \\ D_2 &= \min\left\{t_i+t_j : i\le j\in[N], \binom{t_i+t_j}{t_i} < i\right\}, \text{ and} \\ D_3 &= \min\left\{t_i+t_j-1 : i\le j\in[N], \binom{t_i+t_j}{t_i} + \binom{t_i+t_j}{t_i-1} < i\right\}. \end{aligned}$$

Proof: By Lemma 21 $N_H(d,n,t,t')$ increases on n for all t, t' and d, except for two cases: $d=t+t'$ and $d=t+t'-1$. If $d=t+t'$ then $N_H(d,n,t,t') = \binom{t+t'}{t}$, and for $d=t+t'-1$ we have $N_H(d,n,t,t') = \binom{t+t'}{t} + \binom{t+t'}{t-1}$.

Let $T' = (t'_1,t'_2,\ldots,t'_\ell)$ and $P = (p_1,p_2,\ldots,p_\ell)$ be defined as described earlier in this section.

Let $U = \{\boldsymbol{x},\boldsymbol{z}\}$ be of distance d. Thus, by Theorem 8 and Proposition 1, a (T',P)-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x},\boldsymbol{z}\}$ of distance d if and only if there exists $i' \in [\ell]$ such that $N_H(d,n,t'_{i'}) = N_H(\boldsymbol{x},\boldsymbol{z},t'_{i'}) < \lceil p_{i'}N\rceil$, and by Lemma 32 this holds if and only if there exists $i \in [N]$ such that $N_H(d,n,t_i) < i$. Next we present equivalent conditions to $N_H(d,n,t_i) < i$ for all i. If $d = 2t_i-1$ then we substitute $N_H(d,n,t_i) = \binom{2t_i}{t_i}$, and if $d=2t_i$ then $N_H(d,n,t_i) = \binom{2t_i}{t_i} + \binom{2t_i}{t_i-1}$. Otherwise, $N_H(d,n,t_i) = N_H(d,n,t_i,t_i)$ increases on n, and therefore, for n sufficiently large, $N_H(d,n,t_i) < i$ implies $N_H(d,n,t_i) = 0$ which is equivalent to $d > 2t_i$. Recall that $D_H^k(T,N,V_H)$ equals to the minimum between all these possible d. Thus, we can conclude that in the sequenced model, for n sufficiently large,

$$D_H^k(T,N,V_H) = \min\{D_1,D_2,D_3\}$$

where

$$\begin{aligned} D_1 &= \min\{2t_i+1 : i\in[N]\} = 2t_1+1, \\ D_2 &= \min\left\{2t_i : i\in[N], \binom{2t_i}{t_i} < i\right\}, \text{ and} \\ D_3 &= \min\left\{2t_i-1 : i\in[N], \binom{2t_i}{t_i} + \binom{2t_i}{t_i-1} < i\right\}. \end{aligned}$$

Similarly, in the non-sequenced model, by Lemma 9, a (T',P)-non-sequenced-channel system of size N supports exact reconstruction for $U = \{\boldsymbol{x},\boldsymbol{z}\}$ of distance d if and only if there exist $i',j' \in [\ell]$ such that $N_H(d,n,t'_{i'},t'_{j'}) = N_H(\boldsymbol{x},\boldsymbol{z},t'_{i'},t'_{j'}) < \lceil p_{i'}N\rceil + \lceil p_{j'}N\rceil - N$, and by Lemma 32 this holds if and only if there exist $i,j \in [N]$ such that $N_H(d,n,t_i,t_j) < i+j-N$. Following we present equivalent conditions to $N_H(d,n,t_i,t_j) < i+j-1$ for all $i \le j$. If $d = t_i+t_j$ then we substitute $N_H(d,n,t_i,t_j) = \binom{t_i+t_j}{t_i}$, and if $d=t_i+t_j$ then $N_H(d,n,t_i,t_j) = \binom{t_i+t_j}{t_i} + \binom{t_i+t_j}{t_i-1}$. Otherwise, $N(d,n,t_i,t_j)$ increases on n, and therefore, for n sufficiently large, $N_H(d,n,t_i,t_j) < i+j-N$ implies $N(d,n,t_i,t_j)=0$ and $i+j-N>0$, were $N_H(d,n,t_i,t_j)=0$ is equivalent to $d > t_i+t_j$. Recall that $D_H^u(T,N,V_H)$ equals to the minimum d required for exact reconstruction. Thus, we can conclude that in the non-sequenced model, for n sufficiently large,

$$D_H^u(T,N,V_H) = \min\{D_1,D_2,D_3\}$$

where

$$\begin{aligned} D_1 &= \min\{t_i+t_j+1 : i,j\in[N], i+j-N>0\} \\ &= \min\{t_i+t_j+1 : i,j\in[N], i+j=N+1\}, \\ D_2 &= \min\left\{t_i+t_j : i\le j\in[N], \binom{t_i+t_j}{t_i} < i\right\}, \text{ and} \\ D_3 &= \min\left\{t_i+t_j-1 : i\le j\in[N], \binom{t_i+t_j}{t_i} + \binom{t_i+t_j}{t_i-1} < i\right\}. \end{aligned}$$

□

TABLE I
EXAMPLES FOR THE SOLUTION TO PROBLEM 2

T	$D^k(T,N,V)$	$D^u(T,N,V)$
(t_1,t_2) $J,\ H$	$2t_1+1$	t_1+t_2+1
(t_1,t_2,t_3) $J,\ H(t_3>1)$	$2t_1+1$	$\min\{t_1+t_3,2t_2\}+1$
$(1,1,1)$ H	2	2
(t_1,t_2,t_3,t_4) $J,\ H(t_3>1)$	$2t_1+1$	$\min\{t_1+t_4,t_2+t_3\}+1$
$(1,1,1,1)$ H	1	1
$(1,1,1,t_4)$ H	2	2
(t_1,t_2,t_3,t_4,t_5) $J,\ H(t_3>1)$	$2t_1+1$	$\min\{t_1+t_5,t_2+t_4,2t_3\}+1$

In the following, we apply Theorems 33 and 34 and present explicit solutions to Problem 2 for some parameters of T, for both substitution and transposition errors in the two models, the sequenced and the non-sequenced. The examples are presented in Table I, where J and H in the left column are abbreviations for Johnson and Hamming, and represent the transposition and substitution cases, respectively. In the left column we write the parameter T and some conditions on it. For example, (t_1,t_2,t_3), J, $H(t_3>1)$ means that this row presents the solution for transpositions (J) for all (t_1,t_2,t_3), and for substitutions (H) for all (t_1,t_2,t_3) such that $t_3>1$.

VI. SPECIAL SYSTEMS FOR $T=(t_1,t_2)$

In this section we study special cases of two types of channels. First, we define a new problem, and then present its solution. For $T=(t_1,t_2)$, $t_1<t_2$, and a constant positive integer a, a channel system with N combinatorial channels is called a (T,i,a)-*channel system*, $i\in\{1,2\}$, if a of the channels are t_i-error channels, while the rest are t_{3-i}-error channels. If the size of a system is smaller than a, then all the channels are t_i-error.

Under this model, we consider both cases, sequenced- and non-sequenced- channel systems. For $U\subseteq V$, we denote by $N^u(T,i,a,U,V)$ and $N^k(T,i,a,U,V)$ the value of the minimum size of a (T,i,a)-non-sequenced- and (T,i,a)-sequenced- channel system such that each $x\in U$ has exact reconstruction, respectively. This problem is formulated as follows.

Problem 3: Let V be a finite set with some distance function $\rho: V\times V\to\mathbb{N}$, for all $U\subseteq V$ and $i\in\{1,2\}$, find the values of $N^u(T,i,a,U,V)$ and $N^k(T,i,a,U,V)$.

As before, we focus on sets of the form $U=\{x,z\}$ since $N^g(T,i,a,U,V)=\max\{N^g(T,i,a,\{x,z\},V): x,z\in U\}$.

The solution for this problem is presented in the next three theorems. The first theorem solves the problem for constant number of t_1-error channels. In this case, the minimum number of channels which are required for exact reconstruction does not depend on knowing the behavior of each channel. The last two theorems present the solutions for constant number of t_2-error channels; Theorem 38 for the non-sequenced-channel system, and Theorem 37 for the sequenced one.

The proofs in this section apply similar techniques to those presented in the proofs for Problem 1 (see, for example, Theorems 2 and 4). For completeness and readability of this paper we preset these proofs in Appendix C.

In the rest of this section we denote U to be $U=\{x,z\}\subseteq V$, $T=(t_1,t_2)$ where $t_1<t_2$ and a is a constant positive integer.

Theorem 35: $N^k(T,1,a,U,V) = N^u(T,1,a,U,V) = N+1$, where

$$N=\begin{cases} N(x,z,t_2) & \text{if } N(x,z,t_1)\ge a,\\ N(x,z,t_1) & \text{otherwise.}\end{cases}$$

Furthermore,

- *a $(T,1,a)$-sequenced-channel system of size J supports exact reconstruction for U for all $J\ge N^k(T,1,a,U,V)$, and*
- *a $(T,1,a)$-non-sequenced-channel system of size J supports exact reconstruction for U, for all $N(x,z,t_1)<J\le a$. and $N(x,z,t_2)<J$.*

We note that according to Theorem 35 in almost all cases

$$N^k(T,1,a,U,V)=N^u(T,1,a,U,V)=N(x,z,t_2)+1.$$

The following remark explains the undesirable phenomena in which a $(T,1,a)$-non-sequenced-channel system of size $J>N^u(T,1,a,U,V)$ does not support exact reconstruction.

Remark 36: A $(T,1,a)$-non-sequenced-channel system of size $J>N^u(T,1,a,U,V)$ may not support exact reconstruction for U. By Theorem 35 it could happen only if $a<J\le N(x,z,t_2)$. The reason is that in this model it is not guaranteed which of the a outputs are the outputs of t_1-error channels. For example, $U=\{\boldsymbol{x},\boldsymbol{z}\}\subseteq\{0,1\}^n$ where $\rho_H(\boldsymbol{x},\boldsymbol{z})=3$, and $t_1=1$, $t_2=2$ and $a=2$. For these parameters we have $N(\boldsymbol{x},\boldsymbol{z},t_1)=0<2=a$ and $N(\boldsymbol{x},\boldsymbol{z},t_2)=6$. Thus, $N^u(T,1,a,U,V)=1$, but a $(T,1,a)$-non-sequenced-channel system of size $J=4$, does not support exact reconstruction for U. To prove it, we present a set of J outputs, which any order of them may be an outputs sequence when transmitting either x or z. We choose $a=2$ outputs from $I(\boldsymbol{x},\boldsymbol{z},t_1,t_2)$ and another $a=2$ outputs from $I(\boldsymbol{z},\boldsymbol{x},t_1,t_2)$. Thus, the decoder cannot distinguish between transmitting x or z. One can readily verify that in this case, exact reconstruction is not supported also for $J\in\{5,6\}$.

In the second case, we have that $i=2$ and a is the number of channels with at most t_2 errors. First, we state the solution to Problem 3 for the sequenced model.

Theorem 37: $N^k(T,2,a,U,V)=N+1$, where

$$N=\min\{N(x,z,t_1)+a, N(x,z,t_2)\}.$$

Furthermore, for all $J\ge N^k(T,2,a,U,V)$, a $(T,2,a)$-sequenced-channel system of size J supports exact reconstruction for U.

Here, we note again that in almost all cases

$$N^k(T,2,a,U,V)=N(x,z,t_1)+a+1.$$

Lastly, we solve Problem 3 for the non-sequenced model.

Theorem 38: $N^u(T, 2, a, U, V) = N + 1$, *where*

$$N = \min\{N(x, z, t_1, t_2) + a, N(x, z, t_2), N(z, x, t_1, t_2) + a, N(x, z, t_1) + 2a\}.$$

Furthermore, for all $J \geq N^u(T, 2, a, U, V)$, a $(T, 2, a)$-non-sequenced-channel system of size J supports exact reconstruction for U.

Once again, we note that in almost all the cases

$$N^u(T, 2, a, U, V) = N(x, z, t_1) + 2a + 1.$$

According to the previous theorem, one can verify that for the Hamming case with $a = 2$, $t_1 = 1, t_2 = 2$, and $\rho(\boldsymbol{x}, \boldsymbol{z}) = 3$, we get that $N^u(T, 2, a, U, V) = 5$, while if all channels cause at most 2 errors, then the number of channels for exact reconstruction is 7 [9].

Note that Theorem 38 can also be derived by a slight modification in Theorem 4. We denote $m = N(x, z, t_1)$ and we define here

$$N'(x, z, t_1, p) = \min\{L : 2\lceil pL\rceil - L > m, \lceil pL\rceil > m, L \geq 1\} - 1,$$

instead of the previous definition, where

$$N'(x, z, t_1, p) = \min\{L : 2\lceil pL\rceil - L > m, L \geq 1\} - 1.$$

This change has no affect on Theorem 4, since for fixed p, $0 < p < 1, 2\lceil pL\rceil - L \leq \lceil pL\rceil$. Then, by substituting $\lceil pL\rceil = L - a$ in Theorem 4 we can conclude Theorem 38.

VII. Conclusion

In this paper we study a generalization of the reconstruction problem studied by Levenshtein. We assume here that all channels do not behave the same and the number of errors in different channels can vary. In the first problem we assume that the behavior of the channel system is known and for a given code we study the required number of channels for exact reconstruction. Under this problem we considered the cases in which the decoder knows the number of errors in every channel, only the distribution for the number of errors, or the average number of errors. In the second problem the number of channels is given and then we followed the same study in order to find the minimum distance of the code that guarantees exact reconstruction.

Appendix A

In this part we present the omitted proofs in Subsection III-C. The following proofs apply Lemmas 12 and 14, which were already proved in Subsection III-C.

Lemma 16: If a t-channel system of size J, for some even J, supports exact reconstruction for U, then for all even positive integer, N, a t-channel system of size N supports exact reconstruction for U.

Proof: Denote $d = d(U)$, and suppose that a t-channel system of size J, for some even J, supports exact reconstruction for U. First, we prove that a t-channel system of size 2 supports exact reconstruction for U. Eventually, we prove an equivalent claim; if a t-channel system of size 2 does not support exact reconstruction for U, then for every even N, a t-channel system of size N does not support exact reconstruction for U. We prove it by induction on N.

For the second part, we prove by induction on N, that if a t-channel system of size 2 supports exact reconstruction for U, then for each even N, a t-channel system of size N supports exact reconstruction for U.

For the first part, if a t-channel system of size 2 does not support exact reconstruction for U, then, by Lemma 12,

$$\lfloor 2t\rfloor > \lfloor (2d-1)/2\rfloor,$$

and we have to prove that for each even N

$$\lfloor tN\rfloor > \lfloor (dN-1)/2\rfloor.$$

The basis of the induction is $N = 2$. For the step, we assume correctness for an even N and for 2, and we prove for $N + 2$. This holds since

$$\begin{aligned}\lfloor t(N+2)\rfloor &\geq \lfloor tN\rfloor + \lfloor 2t\rfloor \\ &> \lfloor (dN-1)/2\rfloor + \lfloor (2d-1)/2\rfloor + 1 \\ &= (dN-2)/2 + (2d-2)/2 + 1 \\ &= (d(N+2)-2)/2 \\ &= \lfloor (d(N+2)-1)/2\rfloor.\end{aligned}$$

For the second part we assume that a t-channel system of size 2 supports exact reconstruction for U. Then, by Lemma 12,

$$\lfloor 2t\rfloor \leq \lfloor (2d-1)/2\rfloor,$$

and we have to prove that for each even N

$$\lfloor tN\rfloor \leq \lfloor (dN-1)/2\rfloor.$$

The basis of the induction is $N = 2$. For the step, we assume correctness for an even N and for 2, and we prove for $N + 2$. Again, this holds since

$$\begin{aligned}\lfloor t(N+2)\rfloor &\leq \lfloor tN\rfloor + \lfloor 2t\rfloor + 1 \\ &\leq \upsilon \lfloor (dN-1)/2\rfloor + \lfloor (2d-1)/2\rfloor + 1 \\ &= (dN-2)/2 + (2d-2)/2 + 1 \\ &= \upsilon(d(N+2)-2)/2 \\ &= \lfloor (d(N+2)-1)/2\rfloor.\end{aligned}$$

□

Lemma 17: Let $N^a(t, d, V) = 1$. If d is even or if a t-channel system of size 2 supports exact reconstruction for U of distance d, then for all $N \geq 1$ a t-channel system of size N supports exact reconstruction for U.

Proof: The proof is by induction on N. For even d, the basis is $N = 1$, where for odd d the basis is $N = 1$ and $N = 2$. Now, let $N \in \mathbb{N}$. We assume that for $L \leq N$, a t-channel system of size L, supports exact reconstruction for U, and we prove that a t-channel system of size $N + 1$ also supports exact reconstruction for U. By Lemma 12, $\lfloor tL\rfloor \leq \lfloor (dL-1)/2\rfloor$ for all $L \leq N$, and we have to prove that $\lfloor t(N+1)\rfloor \leq \lfloor (d(N+1)-1)/2\rfloor$.

If d is odd and N is odd, then it holds by Lemma 16. Otherwise, d is even or $N \geq 2$ is even, and then

$$\begin{aligned}\lfloor t(N+1)\rfloor &\leq \lfloor tN\rfloor + \lfloor t\rfloor + 1\\ &\leq \lfloor (dN-1)/2\rfloor + \lfloor (d-1)/2\rfloor + 1\\ &\leq (dN-2)/2 + (d-1)/2 + 1\\ &= (d(N+1)-1)/2,\end{aligned}$$

But, $\lfloor t(N+1)\rfloor$ is an integer value, and therefore $\lfloor t(N+1)\rfloor \leq \lfloor (d(N+1)-1)/2\rfloor$. □

Theorem 18: Let $t \geq 0$, d a positive integer, and $U \subseteq V$ where $d(U) = d$. Then

(1) If $d < \lfloor 2t\rfloor$ then $N^a(t,d,V) = \infty$,

(2) If $d > \lceil 2t\rceil$ then $N^a(t,d,V) = 1$, and exact reconstruction is supported for U for any size of the system.

(3) If $d = \lfloor 2t\rfloor$ and d is even then $N^a(t,d,V) = \infty$.

(4) If $d = \lfloor 2t\rfloor$ and d is odd then for all even N, a t-channel system of size N does not support exact reconstruction.

Thus, for an integer t, if $d > 2t$ then $N^a(t,d,V) = 1$, and exact reconstruction is supported for U for any size of the system. Otherwise, $N^a(t,d,V) = \infty$.

Proof: In all the cases we use the claim from Lemma 12, a t-channel system of size N supports exact reconstruction for U of distance d if and only if $\lfloor tN\rfloor \leq \lfloor (dN-1)/2\rfloor$.

(1) By Lemma 14 we have to prove that a t-channel system of size $N = 1$ does not support exact reconstruction for U. This holds since

$$\begin{aligned}\lfloor (dN-1)/2\rfloor = \lfloor (d-1)/2\rfloor &\leq \lfloor (2t-1-1)/2\rfloor\\ &\leq \lfloor t-1\rfloor < \lfloor t\rfloor = \lfloor tN\rfloor.\end{aligned}$$

(2) By Lemma 17 we have to prove that a system of size $N = 1$ and a system of size $N = 2$ support exact reconstruction for U. For $N = 1$ we have,

$$\begin{aligned}\lfloor (dN-1)/2\rfloor = \lfloor (d-1)/2\rfloor &\geq \lfloor ((2t+1)-1)/2\rfloor\\ &= \lfloor t\rfloor = \lfloor tN\rfloor.\end{aligned}$$

and for $N = 2$

$$\begin{aligned}\lfloor (dN-1)/2\rfloor = \lfloor (2d-1)/2\rfloor &\geq \lfloor (2(2t+1)-1)/2\rfloor\\ &= \lfloor 2t + 1/2\rfloor\\ &\geq \lfloor 2t\rfloor = \lfloor tN\rfloor.\end{aligned}$$

(3) By Lemma 14 we have to prove that a t-channel system of size $N = 1$ does not support exact reconstruction for U. This holds since

$$\begin{aligned}\lfloor (dN-1)/2\rfloor = \lfloor (d-1)/2\rfloor &= \lfloor (\lfloor 2t\rfloor - 1)/2\rfloor\\ &= v(\lfloor 2t\rfloor - 2)/2\\ &= \lfloor 2t\rfloor/2 - 1\\ &\leq \lfloor t\rfloor - 1 < \lfloor t\rfloor = \lfloor tN\rfloor.\end{aligned}$$

(4) By Lemma 16, it is sufficient to prove that a t-channel system of size $N = 2$ does not support exact reconstruction for U. This holds since

$$\begin{aligned}\lfloor (dN-1)/2\rfloor = \lfloor (2d-1)/2\rfloor &= \lfloor (2\lfloor 2t\rfloor - 1)/2\rfloor\\ &= (2\lfloor 2t\rfloor - 2)/2\\ &= \lfloor 2t\rfloor - 1\\ &< \lfloor 2t\rfloor = \lfloor tN\rfloor.\end{aligned}$$

Thus, for an integer t, if $d > 2t$ then by (2) we conclude that $N^a(t,d,V) = 1$ and exact reconstruction is supported for U for any size of the system. Otherwise, by (1) and (3), $N^a(t,d,V) = \infty$. □

Appendix B

For the completeness of the results in the paper, we present in this section some omitted proofs of lemmas which are used in this paper.

Let $V = \{0,1\}^n$ be the set of all length n words over the binary alphabet, $x, z \in V$, and $\rho : V \times V \to \mathbb{N}$ is a distance function. Recall that $N(x,z,t_1,t_2)$ is the number of elements from V which are of distance at most t_1 from x, and at most t_2 from z. If $N(x,z,t_1,t_2)$ depends only on $d = \rho(x,z)$ we denoted this value by $N(d,t_1,t_2)$.

In this section we prove the value of $N(d,t_1,t_2)$ for two types of errors: the value for substitution errors (Hamming distance) which is presented in Lemma 21, is proved in Appendix B-A, and for transposition errors (Johnson distance), which is presented in Lemma 27, is proved in Appendix B-B.

Appendix B-A and Appendix B-B contains also proofs of the comparisons which determine $N_H(d,n,t_1,t_2)$ and $N_J(d,n,t_1,t_2)$ (Hamming and Johnson distance), respectively.

A. Proofs for Lemmas 21,22,23

In this subsection we prove Lemma 21, that is, we prove the value of $N_H(d,n,t_1,t_2)$, and we also prove Lemmas 22 and 23 which contribute to determine $N_H^u(T,P,d,n)$ for substitution errors.

Recall that the Hamming distance function $\rho_H : V_H \times V_H \to \mathbb{N}$, where $V_H = \{0,1\}^n$, is defined by $\rho_H(x,z) = |\{i : x_i \neq z_i\}|$, and in this case, $N_H(x,z,t_1,t_2)$ depends only on $d = \rho_H(x,z)$. Thus, we denote by $N_H(d,n,t_1,t_2)$ the value $N_H(x,z,t_1,t_2)$ where $d = \rho_H(x,z)$.

Lemma 21: For $1 \leq t_1 \leq t_2$ and $d \geq 1$,

$$N_H(d,n,t_1,t_2) = \sum_{i=0}^{\left\lfloor \frac{t_1+t_2-d}{2}\right\rfloor} \binom{n-d}{i} \cdot \sum_{k=d-t_2+i}^{t_1-i} \binom{d}{k}.$$

Proof: Let $\boldsymbol{x} = (x_1,\ldots,x_n)$ and $\boldsymbol{z} = (z_1,\ldots,z_n)$ such that $\rho_H(\boldsymbol{x},\boldsymbol{z}) = d$. Let A be the set of indices where $x_i \neq z_i$, and $B = [n] \setminus A$, i.e., $A = \{j : x_j \neq z_j\}$, and $B = \{j : x_j = z_j\}$.

Let $\boldsymbol{y} \in I(\boldsymbol{x},\boldsymbol{z},t_1,t_2)$, that is, $m = \rho_H(\boldsymbol{x},\boldsymbol{y}) \leq t_1$ and $\rho_H(\boldsymbol{y},\boldsymbol{z}) \leq t_2$. The m positions in which $\boldsymbol{x}$ and $\boldsymbol{y}$ differ can be partitioned into two disjoint sets, i positions in B, and $k = m - i$ indices in A. That is, $i = \left|\{j : x_j = z_j \text{ and } x_j \neq y_j\}\right|$ and $k = \left|\{j : x_j = y_j \text{ and } x_j \neq z_j\}\right|$.

By $\rho_H(\boldsymbol{x}, \boldsymbol{y}) \leq t_1$ we have $i + k \leq t_1$, and from $\rho_H(\boldsymbol{z}, \boldsymbol{y}) \leq t_2$ we get $i + d - k \leq t_2$. Thus, $d - t_2 + i \leq k \leq t_1 - i$, which implies $d - t_2 + i \leq t_1 - i$, and hence $i \leq \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$.

For the second direction, we note that two different choices of i indices from A, and k elements from B, where $0 \leq i \leq \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$, and $d - t_2 + i \leq k \leq t_1 - i$, yield two different elements in $I(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$. □

Lemma 22: For any fixed p and n sufficiently large the following holds. If d is odd, $p \leq 1/2$, and $t_2 = t_1 + 1$, then

$$N_H(d, n, t_2) < \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Otherwise,

$$N_H(d, n, t_2) \geq \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Proof: Note that

$$N_H(d, n, t_2) = \Theta(n^{\left\lfloor \frac{2t_2-d}{2} \right\rfloor})$$

and

$$N_H(d, n, t_1, t_2) = \Theta(n^{\left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor}).$$

Thus, we compare between the powers $\left\lfloor \frac{2t_2-d}{2} \right\rfloor$ and $\left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$. If $t_2 = t_1 + 1$ and d is odd then $\left\lfloor \frac{2t_2-d}{2} \right\rfloor = \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$. In all other cases, $\left\lfloor \frac{2t_2-d}{2} \right\rfloor > \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$, and hence $N_H(d, n, t_2) > \lfloor N_H(d, n, t_1, t_2)/p \rfloor$.

For the case of $t_2 = t_1 + 1$ and odd d, we compare the coefficients of the dominant powers. Denote $d = 2m + 1$.

$$\begin{aligned} N_H(d, n, t_2) = &\left(\binom{d}{m} + \binom{d}{m+1}\right) \cdot \binom{n-d}{t_1 - m} \\ &+ \sum_{k=m-1}^{m+2} \binom{d}{k} \cdot \binom{n-d}{t_1-m-1} \\ &+ \Theta(n^{t_1-m-2}), \\ N_H(d, n, t_1, t_2) = &\binom{d}{m} \cdot \binom{n-d}{t_1 - m} \\ &+ \sum_{k=m-1}^{m+1} \binom{d}{k} \cdot \binom{n-d}{t_1-m-1} \\ &+ \Theta(n^{t_1-m-2}). \end{aligned}$$

Thus, the coefficient of the dominant powers in $N(d, t_2)$ is twice the coefficient of the corresponding term in $N(d, t_1, t_2)$. But, $N_H(d, n, t_1, t_2)$ is multiplied by $1/p$. Thus, for $p > 1/2$ we have $\lfloor N_H(d, n, t_1, t_2)/p \rfloor \leq N_H(d, n, t_2)$, and for $p < 1/2$, $\lfloor N_H(d, n, t_1, t_2)/p \rfloor > N_H(d, n, t_2)$.

For $p = 1/2$, we compare the coefficient of the second dominant powers in these two terms and get that $\sum_{k=m-1}^{m+2} \binom{d}{k} < 2 \cdot \sum_{k=m-1}^{m+1} \binom{d}{k}$. Thus, we conclude that for this case $\lfloor N_H(d, n, t_1, t_2)/p \rfloor > N_H(d, n, t_2)$. □

Lemma 23: For any fixed p and n sufficiently large the following holds. If d is even, $t_2 = t_1 + 1$, and ($(1/2 < p \leq 2/3)$ or $(2/3 < p < 3/4$ and $d < \frac{2-2p}{3p-2})$), then

$$\left\lfloor \frac{N_H(d, n, t_1)}{2p - 1} \right\rfloor > \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Otherwise,

$$\left\lfloor \frac{N_H(d, n, t_1)}{2p - 1} \right\rfloor \leq \lfloor N_H(d, n, t_1, t_2)/p \rfloor .$$

Proof: Note that $N_H(d, n, t_1) = \Theta(n^{\left\lfloor \frac{2t_1-d}{2} \right\rfloor})$ and $N_H(d, n, t_1, t_2) = \Theta(n^{\left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor})$. Thus, we compare the powers $\left\lfloor \frac{2t_1-d}{2} \right\rfloor$ and $\left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$. If $t_2 = t_1 + 1$ and d is even then $\left\lfloor \frac{2t_1-d}{2} \right\rfloor = \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$. In all other cases, $\left\lfloor \frac{2t_1-d}{2} \right\rfloor < \left\lfloor \frac{t_1+t_2-d}{2} \right\rfloor$, and hence, $N_H(d, n, t_1) < \lfloor N_H(d, n, t_1, t_2)/p \rfloor$.

For the case of $t_2 = t_1 + 1$ and even d, we compare the coefficients of the dominant powers.

$$\begin{aligned} N_H(d, n, t_1) &= \binom{d}{d/2} \cdot \binom{n-d}{t_1 - d/2} + \Theta(n^{t_1-d/2-1}), \\ N_H(d, n, t_1, t_2) &= \left(\binom{d}{d/2 - 1} + \binom{d}{d/2}\right) \cdot \binom{n-d}{t_1 - d/2} \\ &\quad + \Theta(n^{t_1-d/2-1}). \end{aligned}$$

Thus, the coefficient of the dominant term in $\left\lfloor \frac{N_H(d,n,t_1)}{2p-1} \right\rfloor$ is

$$\frac{1}{2p - 1}\binom{d}{d/2},$$

while the corresponding coefficient in $\lfloor N_H(d, n, t_1, t_2)/p \rfloor$ is

$$\frac{1}{p}\left(\binom{d}{d/2} + \binom{d}{d/2 - 1}\right) = \frac{2d + 2}{(d + 2)p}\binom{d}{d/2}.$$

The inequality

$$\frac{2d + 2}{(d + 2)p} < \frac{1}{2p - 1}$$

holds if and only if

$$(p \leq 2/3) \text{ or } (2/3 < p < 3/4 \text{ and } d < \frac{2 - 2p}{3p - 2}).$$

Therefore, we conclude that

$$\left\lfloor \frac{N_H(d, n, t_1, t_2)}{p} \right\rfloor < \left\lfloor \frac{N_H(d, n, t_1)}{2p - 1} \right\rfloor$$

if and only if d is even, $t_2 = t_1 + 1$, and $((1/2 < p \leq 2/3)$ or $(2/3 < p < 3/4$ and $d < \frac{2-2p}{3p-2}))$. □

B. Proofs for Lemmas 27 and 29

In this subsection we prove Lemma 27, that is we prove the value of $N_J(d, n, t_1, t_2)$, and we also prove Lemma 29 which contributes in determining $N_J^u(T, P, d, n)$ for transposition errors.

For $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$, a transposition error transposes the symbols x_i and x_j. Note that transpositions do not change the Hamming weight of a word. Therefore we consider $V_J = J_w^n$, the set of all length n words over the binary alphabet with Hamming weight w. The Johnson distance function $\rho_J : V_J \times V_J \to \mathbb{N}$ is defined by $\rho_J(x, z) = \frac{|\{i : x_i \neq z_i\}|}{2}$.

Note, that $N_J(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ depends only on $d = \rho_J(\boldsymbol{x}, \boldsymbol{z})$. Thus, in this case, we denote by $N_J(d, n, t_1, t_2)$ the value $N_J(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ where $d = \rho_J(\boldsymbol{x}, \boldsymbol{z})$.

Lemma 27: For $t_1 \leq t_2$:

$$N_J(d, n, t_1, t_2) = \sum_{i=0}^{t_1} \binom{n-w-d}{i} \cdot \sum_{a=0}^{t_1-i} \sum_{b=0}^{t_2-i} \binom{d}{a} \binom{d}{b} \binom{w-d}{a+b+i-d}.$$

Proof: Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{z} = (z_1, \ldots, z_n)$, two binary words of Hamming weighht w, where $\rho_J(\boldsymbol{x}, \boldsymbol{z}) = d$. Let us partition the n indices into four disjoint sets:

- $I = \{j : x_j = z_j = 1\}$, $|I| = w - d$,
- $A = \{j : x_j = 1,\ z_j = 0\}$, $|A| = d$,
- $B = \{j : x_j = 0,\ z_j = 1\}$, $|B| = d$, and
- $C = \{j : x_j = z_j = 0\}$, $|C| = n - w - d$.

Let $\boldsymbol{y} \in I(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ where $\boldsymbol{y}$ is of Hamming weight w. The w positions in which $y_j = 1$ can be partitioned into four sets as follows: $w - d - i$ positions in I, $d - a$ in A, $d - b$ in B, and the rest, $a + b + i - d$, in C.

By the definition of the sets,

$$\begin{aligned}\rho_J(\boldsymbol{x}, \boldsymbol{y}) &= \frac{\big|\{j : j \in I \cup A \wedge y_j = 0\}\big| + \big|\{j : j \in B \cup C \wedge y_j = 1\}\big|}{2} \\ &= \frac{i + a + (d-b) + (a+b+i-d)}{2} \\ &= i + a,\end{aligned}$$

and

$$\begin{aligned}\rho_J(\boldsymbol{z}, \boldsymbol{y}) &= \frac{\big|\{j : j \in I \cup B \wedge y_j = 0\}\big| + \big|\{j : j \in A \cup C \wedge y_j = 1\}\big|}{2} \\ &= \frac{i + b + (d-a) + (a+b+i-d)}{2} \\ &= i + b.\end{aligned}$$

Recall that $\boldsymbol{y} \in I(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$, i.e., $\rho_J(\boldsymbol{x}, \boldsymbol{y}) \leq t_1$ and $\rho_J(\boldsymbol{y}, \boldsymbol{z}) \leq t_2$. Therefore, $i + a \leq t_1$ and $i + b \leq t_2$.

For the second direction, we note that two different choices of i indices from I (to be the 0s in I), a indices from A (to be the 0s in A), b indices from B (to be the 0s in B), and $a + b + i - d$ indices from C (to be the 1s in C), where $0 \leq i + a \leq t_1$ and $0 \leq i + b \leq t_2$, yield two different elements in $I(\boldsymbol{x}, \boldsymbol{z}, t_1, t_2)$ of Hamming weight w. □

Lemma 29: For any fixed p and n sufficiently large the following holds.

$$\left\lfloor \frac{N_J(d, n, t_1)}{2p-1} \right\rfloor < \lfloor N_J(d, n, t_1, t_2)/p \rfloor$$

if and only if

$$\frac{p}{2p-1} < \sum_{b=0}^{t_2-t_1} \binom{d}{b} \frac{1}{(t_1-d+b)^2 \cdots (t_1-d+1)^2}.$$

Proof: Note that

$$\begin{aligned}\frac{N_J(d, n, t_1)}{2p-1} &= \frac{1}{2p-1} \cdot \binom{w-d}{t_1-d} \cdot \binom{n-w-d}{t_1} + \Theta(n^{t_1-1}), \\ \frac{N(d, t_1, t_2)}{p} &= \frac{1}{p} \cdot \left(\sum\nolimits_{b=0}^{t_2-t_1} \binom{d}{b} \binom{w-d}{t_1+b-d} \right) \\ &\quad \cdot \binom{n-w-d}{t_1} + \Theta(n^{t_1-1}).\end{aligned}$$

Thus, we compare between

$$\frac{1}{2p-1} \cdot \binom{w-d}{t_1-d}$$

and

$$\frac{1}{p} \cdot \left(\sum_{b=0}^{t_2-t_1} \binom{d}{b} \binom{w-d}{t_1+b-d} \right).$$

The following holds

$$\frac{1}{2p-1} \cdot \binom{w-d}{t_1-d} < \frac{1}{p} \cdot \left(\sum_{b=0}^{t_2-t_1} \binom{d}{b} \binom{w-d}{t_1+b-d} \right)$$

if and only if

$$\begin{aligned}\frac{p}{2p-1} &< \sum\nolimits_{b=0}^{t_2-t_1} \binom{d}{b} \frac{\binom{w-d}{t_1+b-d}}{\binom{w-d}{t_1-d}} \\ &= \sum\nolimits_{b=0}^{t_2-t_1} \binom{d}{b} \frac{1}{(t_1-d+b)^2 \cdots (t_1-d+1)^2}.\end{aligned}$$

□

Appendix C

We present in this section two omitted proofs from Section VI.

Theorem 35: $N^k(T, 1, a, U, V) = N^u(T, 1, a, U, V) = N + 1$, where

$$N = \begin{cases} N(x, z, t_2) & \text{if } N(x, z, t_1) \geq a, \\ N(x, z, t_1) & \text{otherwise.} \end{cases}$$

Furthermore,

- *a $(T, 1, a)$-sequenced-channel system of size J supports exact reconstruction for U for all $J \geq N^k(T, 1, a, U, V)$, and*
- *a $(T, 1, a)$-non-sequenced-channel system of size J supports exact reconstruction for U, for all $N(x, z, t_1) < J \leq a$. and $N(x, z, t_2) < J$.*

Proof: If $N(x, z, t_1) < a$, then a $(T, 1, a)$-channel system of size at most $N(x, z, t_1) + 1$ contains only t_1-channels. Thus, according to Levenshtein [9], $N^k(T, 1, a, U, V) = N^u(T, 1, a, U, V) = N(x, z, t_1) + 1$. Note that in the sequenced model, for all $J \geq N^k(T, 1, a, U, V) = N(x, z, t_1) + 1$, the first $N^k(T, 1, a, U, V)$ channels are t_1-error channels, and only their outputs are sufficient to decode correctly. Yet, in the non-sequenced model, if $N(x, z, t_1) + 1 \leq J \leq a$ then all the channels are t_1-error, and exact reconstruction is supported. In addition, if $J > N(x, z, t_2)$ exact reconstruction is supported by Levenshtein result for a system where all the channels are t_2-error. But, for

$N(x, z, t_1) + 1 \leq a < J \leq N(x, z, t_2)$ exact reconstruction may not be supported as noted in Remark 36.

In the second case, $N(x, z, t_1) \geq a$, and then by Levenshtein's result a $(T, 1, a)$-channel system of size J, $J \geq N(x, z, t_2) + 1$, supports exact reconstruction for the two models. For the second direction we have $a \leq J \leq N(x, z, t_2)$. Let us assume that x is transmitted over the system, the first a outputs are in $I(x, z, t_1)$ and all the J outputs are in $I(x, z, t_2)$. Then z can also be a possible output of the decoder in both the sequenced and the non-sequenced models. □

Theorem 37: $N^k(T, 2, a, U, V) = N + 1$*, where*

$$N = \min\{N(x, z, t_1) + a, N(x, z, t_2)\}.$$

Furthermore, for all $J \geq N^k(T, 2, a, U, V)$*, a* $(T, 2, a)$*-sequenced-channel system of size* J *supports exact reconstruction for* U*.*

Proof: For the first direction, if $J > \min\{N(x, z, t_1) + a, N(x, z, t_2)\}$ then either $J > N(x, z, t_1) + a$ or $J > N(x, z, t_2)$. Since $t_1 < t_2$, a t_1-channel system is also a t_2-channel system. Thus, if $J > N(x, z, t_2)$ we can apply the solution by Levenshtain for a system where all the channels are t_2-error. Otherwise, $J > N(x, z, t_1) + a$, and we can apply Levenshtein's solution for the subsystem consists of only the first $N(x, z, t_1) + 1$ channels.

For the second direction, we present a sequence of J outputs, which can be an outputs sequence when transmitting either x or z. The first $(J - a)$ outputs will be in $I(x, z, t_1)$, which is possible by the condition $J \leq N(x, z, t_1) + a$, and the other outputs will be chosen from $I(x, z, t_2)$, which is possible by $J \leq N(x, z, t_2)$. Thus, for this outputs sequence, the decoder cannot distinguish between transmitting x or z. □

Theorem 38: $N^u(T, 2, a, U, V) = N + 1$*, where*

$$N = \min\{N(x, z, t_1, t_2) + a, N(x, z, t_2), N(z, x, t_1, t_2) + a, N(x, z, t_1) + 2a\}.$$

Furthermore, for all $J \geq N^u(T, 2, a, U, V)$*, a* $(T, 2, a)$*-non-sequenced-channel system of size* J *supports exact reconstruction for* U*.*

Proof: The proof is similar to the one of Theorem 4. If a $(T, 2, a)$-channel system consists of $J \geq N+1$ channels, then, by the definition of N, at least one of the following conditions holds:

(1) $J - a > N(x, z, t_1, t_2)$,
(2) $J - a > N(z, x, t_1, t_2)$,
(3) $J > N(x, z, t_2)$, or
(4) $2(J - a) - N(x, z, t_1) > J$.

Without loss of generality, let x be the transmitted word. If Condition (1) or (3) holds, then not all the outputs are in $B_{t_2}(z)$. If Condition (2) holds, there are no $J - a$ outputs in $B_{t_1}(z)$. Thus, if one of the conditions (1), (2), or (3) holds, then z will not be decoded. Regarding Condition (4), assume that we have m outputs in $I(x, z, t_1)$, $m \leq N(x, z, t_1)$. In order for z to be a possible output of the decoder, we must have at least $J - a - m$ outputs in $I(z, x, t_1, t_2) \setminus I(x, z, t_1)$. Furthermore, since x was transmitted at least $J - a - m$ outputs are in $I(x, z, t_1, t_2) \setminus I(x, z, t_1)$. Thus, we must have that $2(J - a) - m \leq J$ in contradiction to Condition (4).

For the second direction we present a set of N outputs which any order of them can be an outputs sequence of transmitting either x or z. The following four conditions hold simultaneously:

(1) $N - a \leq N(x, z, t_1, t_2)$,
(2) $N - a \leq N(z, x, t_1, t_2)$,
(3) $N \leq N(x, z, t_2)$, and
(4) $2(N - a) - N(x, z, t_1) \leq N$.

Let $m = N(x, z, t_1)$. If $m < N - a$, then m outputs are in $I(x, z, t_1)$, at least $N - a - m$ in $I(x, z, t_1, t_2) \setminus I(x, z, t_1)$ (by Conditions (1) and (4)), at least $N - a - m$ in $I(z, x, t_1, t_2) \setminus I(x, z, t_1)$ (by Conditions (2) and (4)), and all the others in $I(x, z, t_2)$ (by Condition (3)). Otherwise, $m \geq N - a$, and then at least $N - a$ outputs are in $I(x, z, t_1)$ and a in $I(x, z, t_2)$ (by Condition (3)). Thus, at least $N - a$ of the outputs are in $B_{t_1}(x)$, and all the N outputs are in $B_{t_2}(x)$, and the same holds for z. □

Acknowledgment

The authors thank the anonymous reviewer and the Associate Editor Prof. Andrew Jiang for their valuable comments and suggestions, which have contributed for the clarity of the paper and its presentation.

References

[1] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "A DNA-based archival storage system," in *Proc. ASPLOS*, Atlanta, GA, USA, 2016, pp. 637–649.

[2] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, p. 1628, 2012.

[3] R. Gabrys and E. Yaakobi, "Sequence reconstruction over the deletion channel," in *Proc. Int. Symp. Inf. Theory*, Jul. 2016, pp. 1596–1600.

[4] M. Horovitz and E. Yaakobi, "Reconstruction of sequences over non-identical channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 1510–1514.

[5] T. Jiang and A. Vardy, "Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1655–1664, Aug. 2004.

[6] E. Konstantinova, "Reconstruction of permutations distorted by single reversal errors," *Discrete Appl. Math.*, vol. 155, no. 18, pp. 2426–2434, 2007.

[7] E. Konstantinova, V. Levenshtein, and J. Siemons. (Feb. 2007). "Reconstruction of permutations distorted by single transposition errors." [Online]. Available: https://arxiv.org/abs/math/0702191v1

[8] E. Konstantinova, "On reconstruction of signed permutations distorted by reversal errors," *Discrete Math.*, vol. 308, pp. 974–984, Mar. 2008.

[9] V. I. Levenshtein, "Efficient reconstruction of sequences," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 2–22, Jan. 2001.

[10] V. I. Levenshtein, "Efficient reconstruction of sequences from their subsequences or supersequences," *J. Combinat. Theory A*, vol. 93, no. 2, pp. 310–332, 2001.

[11] V. I. Levenshtein, E. Konstantinova, E. Konstantinov, and S. Molodtsov, "Reconstruction of a graph from 2-vicinities of its vertices," *Discrete Appl. Math.*, vol. 156, pp. 1399–1406, May 2008.

[12] V. I. Levenshtein and J. Siemons, "Error graphs and the reconstruction of elements in groups," *J. Combinat. Theory, A*, vol. 116, no. 4, pp. 795–815, 2009.

[13] K. Mazooji, personal communication, 2017.

[14] F. Sala, R. Gabrys, C. Schoeny, and L. Dolecek, "Exact reconstruction from insertions in synchronization codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2428–2445, Apr. 2017.

[15] E. Yaakobi and J. Bruck, "On the uncertainty of information retrieval in associative memories," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 106–110.

[16] E. Yaakobi, M. Schwartz, M. Langberg, and J. Bruck, "Sequence reconstruction for Grassmann graphs and permutations," in *Proc. Int. Symp. Inf. Theory*, Jul. 2013, pp. 874–878.

[17] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," *Sci. Rep.*, vol. 7, Jul. 2017, Art. no. 5011.
[18] S. M. H. T. Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao, and O. Milenkovic, "DNA-based storage: Trends and methods," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 3, pp. 230–248, Sep. 2015.
[19] S. M. H. T. Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," *Sci. Rep.*, vol. 5, Aug. 2015, Art. no. 14138.

Michal Horovitz was born in Israel in 1987. She received the B.Sc. degree from the departments of Mathematics and Computer Science at the Open University, Ra'anana, Israel, in 2009, and the Ph.D. degree from the Computer Science Department at the Technion - Israel Institute of Technology, Haifa, Israel, in 2017. She is currently a Lecturer in the Department of Computer Science, Tel-Hai College, Israel and she is also a researcher in The Galilee Research Institute - Migal, Upper Galilee, Israel. Her research interests include coding theory with applications to non-volatile memories, information theory, combinatorics, machine learning, and optimization.

Eitan Yaakobi (S'07–M'12–SM'17) is an Assistant Professor at the Computer Science Department at the Technion—Israel Institute of Technology. He received the B.A. degrees in computer science and mathematics, and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California, San Diego, in 2011. Between 2011-2013, he was a postdoctoral researcher in the department of Electrical Engineering at the California Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, data storage and retrieval, and voting theory. He received the Marconi Society Young Scholar in 2009 and the Intel Ph.D. Fellowship in 2010-2011.