Beyond Single-Deletion Correcting Codes:
Substitutions and Transpositions

Ryan Gabrys, Venkatesan Guruswami, João Ribeiro, Ke Wu
Three types of common errors

Substitution

Deletion (insertion)

Transposition
Substitution

AGCGCT

AGCGTT
Deletion(insertion)
Transposition
How about the interplay?

- Transposition
- Deletion
- Substitution

One del or one transposition

One del and one sub

One del or one sub
## Our Results

<table>
<thead>
<tr>
<th>Alphabet</th>
<th>Error type</th>
<th>Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>One del or one sub (edit error)</td>
<td>$\log n + O_q (\log \log n)$</td>
</tr>
<tr>
<td>2</td>
<td>One del or one adjacent trans</td>
<td>$\log n + O (\log \log n)$</td>
</tr>
<tr>
<td>1</td>
<td>One del AND one sub</td>
<td>$4\log n + O (\log \log n)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>List of size 2</td>
</tr>
</tbody>
</table>
## Our Results

<table>
<thead>
<tr>
<th>Alphabet</th>
<th>Error type</th>
<th>Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>q</td>
<td>One del or one sub (edit error)</td>
<td>$\log n + O_q(\log \log n)$</td>
</tr>
</tbody>
</table>
Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary
Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary
Correcting one edit error: binary

VT sketch: \( f(x) = \sum_{i=1}^{n} i \cdot x_i \)

If substitution: \( x = 0 \ 0 \ 1 \ 0 \ 1 \) \( \quad f(x) = 8 \)

[Lev 65, VT65]
VT sketch:  \[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(x) = 8 \]

\[ y = 00111 \quad f(y) = 12 \]

\[ f(y) - f(x) = \sum_{i=1}^{n} i \cdot (y_i - x_i) = e(y_i - x_i) = 4 \]

Position of substitution
VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$x = 00101 \quad f(x) = 8$

$y = 00111 \quad f(y) = 12$
Correcting one edit error: deletion

VT sketch: \[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(x) = 8 \]

\[ y = 0 \ 0 \ 0 \ 0 \ 1 \]

\[ f(y) = 4 \]

\[ f(x) - f(y) = \sum_{d} y_i + dx_d = 4 \]
VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$f(x) = 8$

$y = \begin{bmatrix} ? & 0 & 0 & 0 & 1 \end{bmatrix}$

$f(y) = 4$

$f(x) - f(y) = \sum_{d}^{n-1} y_i + dx_d = 4$
VT sketch: \[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ x = 0 0 1 0 1 \quad f(x) = 8 \]

\[ y = 0 0 0 ? 1 \quad f(y) = 4 \]

\[ f(x) - f(y) = \sum_{d}^{n-1} y_i + dx_d = 4 \]
VT sketch

VT sketch: \( f(x) = \sum_{i=1}^{n} i \cdot x_i \)

\( \log n + 2 \) bits!
Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary
VT code for ternary?

\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ x = 0 2 1 0 1 2 \]

\[ f(x) = 24 \]
VT code for ternary: substitution

\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(x) = 24 \]

\[ y = 0 \ 1 \ 1 \ 0 \ 1 \ 2 \]

\[ f(y) = 22 \]

\[ f(y) - f(x) = e(y_i - x_i) = 2 \]
\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(\text{#0}) = 2 \]
\[ f(\text{#1}) = 2 \]
\[ f(\text{#2}) = 2 \]

\[ x = 2 1 1 0 1 2 \]
\[ x = 0 2 1 0 1 2 \]
\[ f(x) = 24 \]

\[ y = 0 1 1 0 1 2 \]
\[ f(y) = 22 \]

\[ f(y) - f(x) = e(y_i - x_i) = 2 \]
$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

\[
\begin{align*}
  x &= 0 \, 2 \, 1 \, 0 \, 1 \, 2 \\
  f(x) &= 24 \\
  y &= 0 \, 1 \, 1 \, 0 \, 1 \, 2 \\
  f(y) &= 22
\end{align*}
\]
VT code for ternary: deletion of 0

\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[
\begin{align*}
  f(x) &= 24 \\
  y &= 0 \ 2 \ 1 \ 1 \ 2 \\
  f(x) - f(y) &= \sum_{d} y_i = 3
\end{align*}
\]
\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ x = 0 \ 2 \ 1 \ 0 \ 1 \ 2 \]
\[ f(x) = 24 \]

\[ y = 0 \ 2 \ 1 \ ? \ 1 \ 2 \]
\[ f(y) = 21 \]

\[ f(x) - f(y) = \sum_{d} y_i = 3 \]

Position of deletion
VT code for ternary: deletion of 1

\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(x) = 24 \]

\[ y = 0 \ 2 \ 0 \ 1 \ 2 \]

\[ f(y) = 18 \]

\[ f(x) - f(y) = \sum_{d}^{n-1} y_i + d = 6 \]
\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ x = 102012 \quad \quad x = 021012 \quad \quad f(x) = 24 \]

\[ y = 02012 \quad \quad f(y) = 18 \]

\[ f(x) - f(y) = \sum_{d} y_i + d = 6 \]
\[ x = 102012 \]
\[ \sum_{d}^{n-1} y_i + d \]

\[ x = 021012 \]
\[ \sum_{d}^{n-1} y_i + d \]
Fails when there is a chunk with an average 1
Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary
Bias the weight!

\[ f(x) = \sum_{i=1}^{n} i \cdot x_i \]

\[ f(x) = \sum_{i=1}^{n} i \cdot w(x_i) \]

\[ w(0) = 0 \]
\[ w(1) = 1 \]
\[ w(2) = 2 \log n \]
\[ f(x) = \sum_{i=1}^{n} i \cdot w(x_i) \]

\[ w(0) = 0 \]
\[ w(1) = 1 \]
\[ w(2) = 2 \log n \]

\[ x = 102 \ldots \]
\[ f(x) = 1 + 6 \log n \]

\[ x = 021 \ldots \]
\[ f(x) = 4 \log n + 3 \]
As long as the chunk of avg 1 has length $< 2 \log n$

$x = 1 \overbrace{02020202020202}^{< 2 \log n} 2 \ldots$

Run of 0 after deleting all 2’s $< \log n$

$x = 1 \overbrace{0000000}^{< \log n} 2 \ldots$

Runlength replacement [SWGY17]
Weighted VT: encoding

$x \rightarrow \text{Regular } x' \rightarrow \text{Weighted VT of } x' \rightarrow \log n + \log \log n \text{ redundancy!}$
1. Extends to general $q$

2. Binary code correcting one del and one sub

3. Binary code correcting one del and one adjacent trans