Exploit Kashida Adding to Arabic e-Text for High Capacity Steganography

Ahmed A. AlNazer & Adnan A. Gutub

International Workshop on Frontiers of Information Assurance & Security (FIAS 2009)
in conjunction with the IEEE 3rd International Conference on Network & System Security (NSS 2009)
Gold Coast, Queensland, AUSTRALIA, 19-21 October 2009.
Outline

- Objective
- Introduction
- Background
- Proposed Approach
- Improvement and Comparison
- Summary and Conclusion
- Future Work
Objective

- To build a steganography schema and tool that maximize the capacity of Arabic Text cover media by maximizing the use of Kashida (Arabic extension letter) in all possible location to hide a secret.
Introduction

- Steganography is the ability to hide information in a cover media, e.g. pictures.

- Hiding information in text:
  - Challenging because of less un-used bits
  - Appreciated because of less size and the ease of transfer over the network.
Background

- Different languages :- different properties.
- Arabic language:
  - 28 characters
  - Joined characters when writing words
  - Extension character (Kashida) may be embedded between two Arabic characters *
- Example: بـسـ م اهـل رحـم الن ال رحـيـم
Background

- Pointed Letters and Kashida (by Dr. Gutub)
  - Adding extensions after pointed letters.
  - Adding extensions before pointed letters.

- Low security, increase capacity *
Proposed Approach

- Studied Arabic letters to see their applicability to add Kashida.
- Built a steganography schema and tool to put Kashida whenever possible.
- Compare proposed approach with a previous approach in terms of capacity.
Proposed Approach

- Arabic letters applicability with Kashida
  - 35 keyboard letters can come after Kashida
    - س, ر, ز, ث, ح, ج, ث, ت, ك, ب, ئ, ئ, أ, أ, أ
    - 4 special cases for letter (لا, لاأ, ل، ل، ل، لا، ل) can’t accept Kashida between.
  - 23 letters can come before Kashida
    - غ, ط, ض, ص, ش, س, ح, خ, ج, ث, ت, ب, ب
    - لا, لاأ, ل، ل، ل، لا، ل

- 23 letters can come before Kashida
  - غ, ط, ض, ص, ش, س, ح, خ, ج, ث, ت, ب, ب
    - لا, لاأ, ل، ل، ل، لا، ل
Proposed Approach
**Improvement and Comparison**

- **Experiment:** data taken from 15 Khotbas (written religious speeches) in the literature with different length.

- **Capacity comparison:**
  - Using proposed method gives an average of 39% capacity
    - i.e. utilize 39% of the cover media to hide a secret
  - Using - old method - Kashida with dotted letters gives an average of 16% capacity.
## Improvement and Comparison

<table>
<thead>
<tr>
<th>#</th>
<th>Cover Media Length</th>
<th>MSCUKAT Capacity</th>
<th>MSCUKAT Per %</th>
<th>Dotted Letters Capacity</th>
<th>Per %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,357</td>
<td>861</td>
<td>40.88</td>
<td>1,653</td>
<td>21.29</td>
</tr>
<tr>
<td>2</td>
<td>2,503</td>
<td>845</td>
<td>41.66</td>
<td>1,785</td>
<td>19.72</td>
</tr>
<tr>
<td>3</td>
<td>2,905</td>
<td>977</td>
<td>36.03</td>
<td>1,649</td>
<td>21.35</td>
</tr>
<tr>
<td>4</td>
<td>2,990</td>
<td>909</td>
<td>38.72</td>
<td>1,741</td>
<td>20.22</td>
</tr>
<tr>
<td>5</td>
<td>3,137</td>
<td>962</td>
<td>36.59</td>
<td>1,681</td>
<td>20.94</td>
</tr>
<tr>
<td>6</td>
<td>3,337</td>
<td>997</td>
<td>35.31</td>
<td>1,883</td>
<td>18.69</td>
</tr>
<tr>
<td>7</td>
<td>3,591</td>
<td>924</td>
<td>38.10</td>
<td>1,677</td>
<td>20.99</td>
</tr>
<tr>
<td>8</td>
<td>3,656</td>
<td>933</td>
<td>37.73</td>
<td>1,622</td>
<td>21.70</td>
</tr>
<tr>
<td>9</td>
<td>3,689</td>
<td>873</td>
<td>40.32</td>
<td>1,639</td>
<td>21.48</td>
</tr>
<tr>
<td>10</td>
<td>3,713</td>
<td>930</td>
<td>37.85</td>
<td>1,751</td>
<td>20.10</td>
</tr>
<tr>
<td>11</td>
<td>3,747</td>
<td>894</td>
<td>39.37</td>
<td>1,784</td>
<td>19.73</td>
</tr>
<tr>
<td>12</td>
<td>3,794</td>
<td>921</td>
<td>38.22</td>
<td>1,606</td>
<td>21.92</td>
</tr>
<tr>
<td>13</td>
<td>3,893</td>
<td>855</td>
<td>41.17</td>
<td>1,603</td>
<td>21.96</td>
</tr>
<tr>
<td>14</td>
<td>4,040</td>
<td>932</td>
<td>37.77</td>
<td>1,728</td>
<td>20.37</td>
</tr>
<tr>
<td>15</td>
<td>5,567</td>
<td>880</td>
<td>40.00</td>
<td>1,623</td>
<td>21.69</td>
</tr>
</tbody>
</table>

| Average | 39.00 | 21.00 |
Improvement and Comparison

![Graph showing comparison between Kadhida with Dotted and MSCUKAT]
Improvement and Comparison

- Analysis:
  - Using proposed technique is giving 244% better than using old Kashida with dotted letters.
  - Limitation of the capacity of using Kashida with dotted letters affect the ability to hide a long secret in a limited size cover media.
  - This implies an advantage of using this proposed idea that it gives us more possibility to hide longer secrets.
Improvement and Comparison

- Analysis of the secret:
  - Studied the secret with different file sizes
  - Analyzed the number of 1s in the secret and its percentage compared to secret size.
  - Opened a future work to better utilize the cover media to have more capacity.

27% average number of ones in the secret
Summary and Conclusion

- Study of characteristics of Arabic letters and how Kashida can be embedded to answer the questions:
  - Is it proper to use Kashida whenever possible?
  - How many places in Arabic text can put Kashida?

- Steganography schema and tool to embed secret with maximum utilization of Kashida between Arabic letters.

- Using proposed method is giving 244% better than using Kashida with dotted letters.
Future Work

- Enhance the way of embedding Kashida
- Benefit of the less number of ones in the secret (27%)
- Encrypt the secret
- Use other file formats as secret
Q & A