

King Fahd University for Petroleum and Minerals
College of Computer Science and Engineering

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

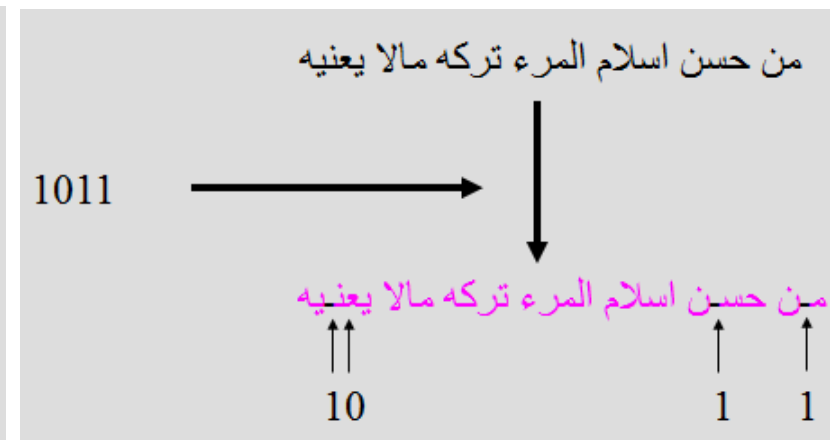
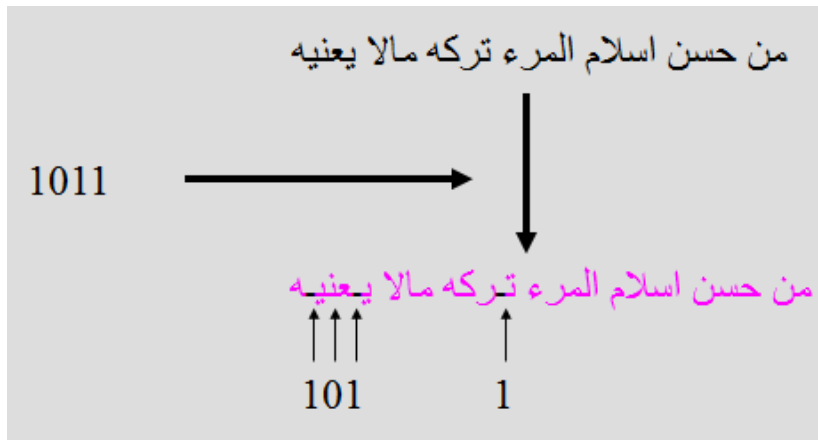
- Steganography is the ability to hide information in a cover media, e.g. pictures.
- Hiding information in text:
 - Challenging because of less un-used bits
 - Appreciated because of less size and the ease of transfer over the network.

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: م الله ال رح م ن ال رح ي م ب س

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' 'ز' 'ر' 'ذ' 'دا' 'خ' 'ح' 'ج' 'ث' 'ت' 'ة' 'ب' 'ا' 'أ' 'إ' 'ؤ' 'أ' 'آ' 'ي' 'ى' 'و' 'ه' 'ن' 'م' 'ل' 'ك' 'ق' 'ف' 'غ' 'ع' 'ظ' 'ط' 'ض' 'ص' 'ش'
 - 23 letters can come before Kashida
 - 'ع' 'ظ' 'ط' 'ض' 'ص' 'ش' 'س' 'خ' 'ح' 'ج' 'ث' 'ت' 'ب' 'ي' 'ه' 'ن' 'م' 'ل' 'ك' 'ق' 'ف' 'غ'
 - 4 special cases for letter (ل): (لآ، لإ، لآ، لآ) can't accept Kashida between.

Proposed Approach

MSCUKAT

Secret (string) الخطبة الأولى
الحث على العلم

Read Secret From File Secret File Name G:\Encryption\Encryption' Secret (string) Length 29

Secret (bits) 1110010001100000010001001100000111010001100000111011000110000000101000110000010010100011000000000010000000000111001000110000
00010001001100000110001000110000000100100110000000100010011000001001100000101100000000000010100000000000011100100011000
00001000100110000010110100011000001101010001100000000010000000000100111000110000000100010011000001001001001100000000010000000

Convert Secret To Bits Secret (bits) Length 464 No. of Ones (1s) 130 Perc. of Ones 28.02

Cover Media الحمد لله الذي خلق السماوات والأرض وما بينهما في ستة أيام وجعل في ذلك من المصالح العظيمة والحكم البالغة ما تتقاصر دونه فهوم ذوي الأفهام وأشهد أن لا إله إلا الله وحده لا شريك له الملك القدوس السلام وأشهد أن محمدا عبده ورسوله سيد الأنام ومصباح الظلام صلى الله عليه وعلى اله وأصحابه والتابعين لهم بإحسان ما تعاقبت الليلي والأيام وسلم تسليما .

Read Cover Media from File Cover File Name G:\Encryption\Encryption' Cover Media Length 3713 Export to Text File

Embed Secret

Using Kashida (MSCUKAT)

Output Text الحمد لله الذي خلق السماوات والأرض وما بينهما في ستة أيام وجعل في ذلك من المصالح العظيمة والحكم البالغة ما تتقاصر دونه فهوم ذوي الأفهام وأشهد أن لا إله إلا الله وحده لا شريك له الملك القدوس السلام وأشهد أن محمدا عبده ورسوله سيد الأنام ومصباح الظلام صلى الله عليه وعلى اله وأصحابه والتابعين لهم بإحسان ما تعاقبت الليلي والأيام وسلم تسليما .

Capacity 1431 Percentage 38.54 Export to Text File Get Secret Back

Secret (back) الخطبة الأولى
الحث على العلم

Using Dotted Letters Kashida

Output Text الحمد لله الذي خلق السماوات والأرض وما بينهما في ستة أيام وجعل في ذلك من المصالح العظيمة والحكم البالغة ما تتقاصر دونه فهوم ذوي الأفهام وأشهد أن لا إله إلا الله وحده لا شريك له الملك القدوس السلام وأشهد أن محمدا عبده ورسوله سيد الأنام ومصباح الظلام صلى الله عليه وعلى اله وأصحابه والتابعين لهم بإحسان ما تعاقبت الليلي والأيام وسلم تسليما . أما بعد أيها الناس : اتقوا ربكم واشكروه على ما هداكم وعلمكم ما لم تكونوا تعلمون علمكم ما فيه صلاح دينكم ونياكم وحبب عنكم من العلم ما لا تدركه عقولكم ولا

Capacity 613 Percentage 16.51 Export to Text File Get Secret Back

Secret (back) الخطبة الأولى
الحث على العلم



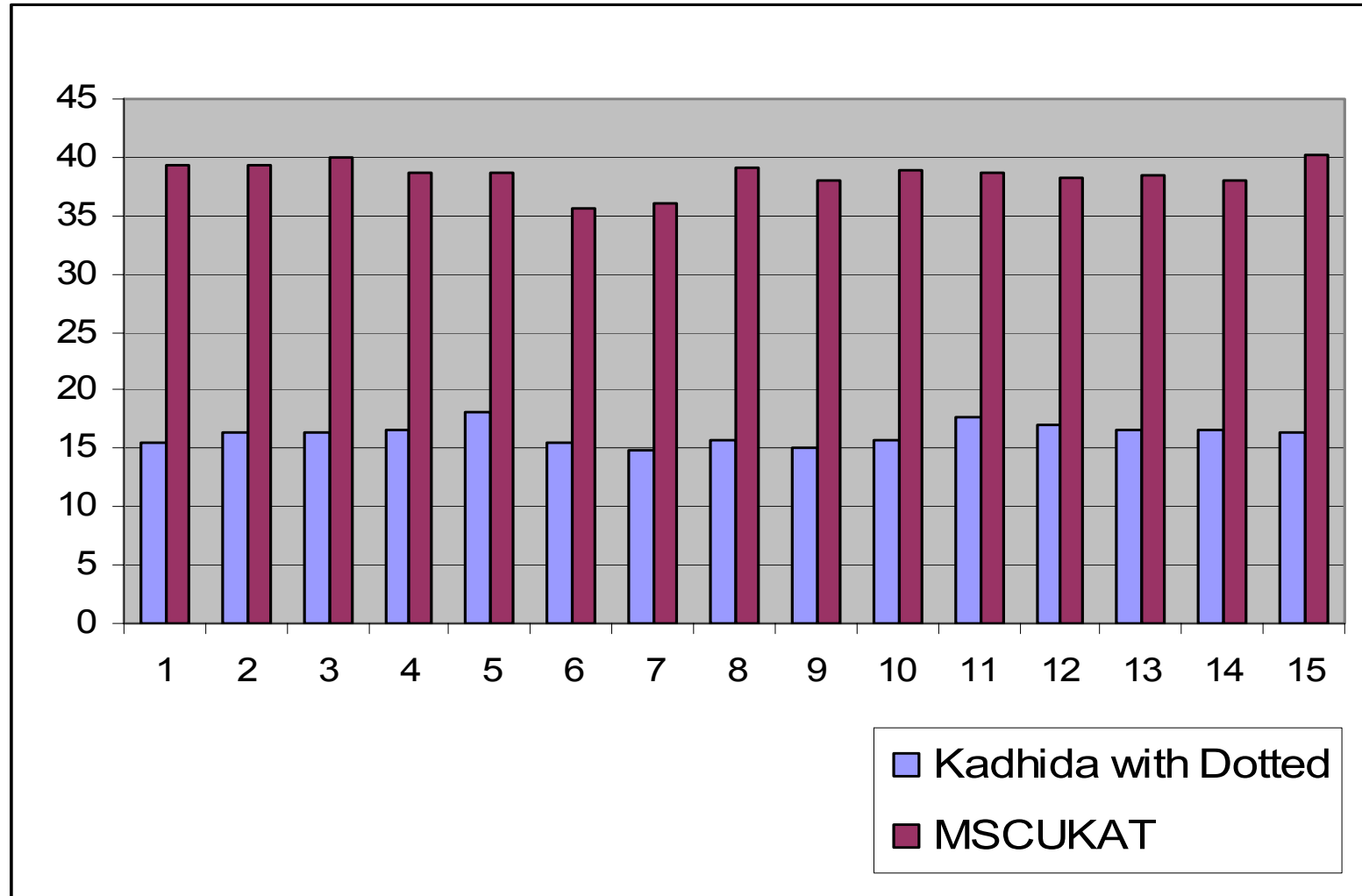
Improvement and Comparison

- 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

#	Cover Media Length	MSCUKAT		Dotted Letters	
		Capacity	Per %	Capacity	Per %
1	2,357	861	40.88	1,653	21.29
2	2,503	845	41.66	1,785	19.72
3	2,905	977	36.03	1,649	21.35
4	2,990	909	38.72	1,741	20.22
5	3,137	962	36.59	1,681	20.94
6	3,337	997	35.31	1,883	18.69
7	3,591	924	38.10	1,677	20.99
8	3,656	933	37.73	1,622	21.70
9	3,689	873	40.32	1,639	21.48
10	3,713	930	37.85	1,751	20.10
11	3,747	894	39.37	1,784	19.73
12	3,794	921	38.22	1,606	21.92
13	3,893	855	41.17	1,603	21.96
14	4,040	932	37.77	1,728	20.37
15	5,567	880	40.00	1,623	21.69
Average			39.00		21.00

Improvement and Comparison



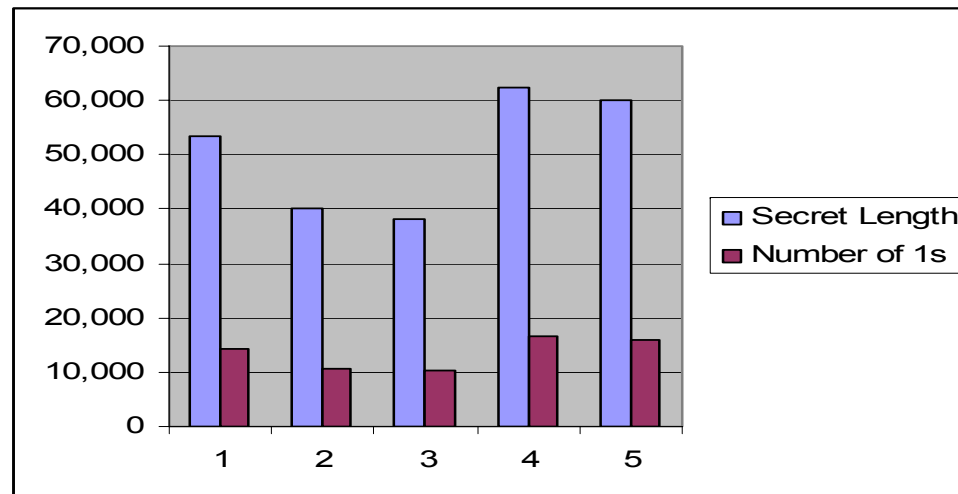


Improvement and Comparison

- 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

