ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(29), DOI: 10.17485/ijst/2015/v8i29/84227, November 2015* ISSN (Online) : 0974-5645

Text Steganography in SMS Using Similarity of Glyphs in Unicode Characters

Ananthi Sheshasaayee* and D. Sumathy

[1]PG and Research Department of Computer Science, Quaid E Millath Government College for Women, Chennai - 600001, Tamil Nadu, India; ananthi.research@gmail.com, sumathy.research@gmail.com

Abstract

Background: Online bank transactions are widely carried out through the mobiles and PDAs there by reducing much time and energy involved for any bank operation. All m-banking transactions are carried out through SMS communications between user mobile and bank server through wireless media exposed to lot of malware vulnerabilities. The problem of securing SMS using text steganography, within limited length 70/160 characters is investigated in this paper. One of the main objectives of this paper is selection of a text steganography technique for protecting the OTP SMS from the bank server to user mobile. **Methods:** In this paper a novel technique of hiding secret six-digit OTP under an innocent looking cover text is proposed. The property of the similarity of glyphs of Unicode characters is combined with frequency of occurrence of English letters is used to hide OTP. The similarities of glyphs provide invisibility and letter frequency improves the overall hiding ratio. **Result:** This method will help to secure OTP SMS generated by the bank server from malware Trojans along the communication channel and also hide the very presence of the secret information within the SMS. By analysing the result, there is no change in the way the cover text appears after embedding OTP. The extraction of the secret bits is also done effectively. In conclusion, this method will efficiently hide the secret bits and provides a medium level of security. It can be used to a maximum of 10-12 digits of OTP. Due to the limited size of 70 Unicode characters per SMS, the payload capacity gets affected if more.

Keywords: Letter Frequency, M-Banking, OTP SMS, Text Steganography, Unicode Glyphs

1. Introduction

Electronic banking (e-banking) is the highly preferred method for performing bank transactions. With time getting dearer and technology becoming cheaper, the user is free to perform any business ignoring spatial and temporal barriers. Smart mobile phones are considered as one's personal tool to conduct mobile banking (m-banking).

1.1 One-Time Password Authentication

One-Time Password is an authentication factor used by banks to authorize the user initiated bank transaction. Figure 1 explains the sequence of actions carried out during m-banking transactions. OTP is sent to the user's registered mobile number via SMS[1].

The OTP is valid only for a transaction and a specific amount of time. There is a sequence of operations that is undergone when an e-banking transaction is initiated. The user completes the bank transaction by entering the OTP as requested by the bank server. The OTP once generated automatically becomes invalid when any one of the following happens.

- Transaction time exceeds the time limit.
- New transaction session begins.
- Login through different IP address.

1.2 Text Steganography in M-Banking

Text steganography is a method of hiding the secret text under an innocuous cover text with the help of stego-

* *Author for correspondence*

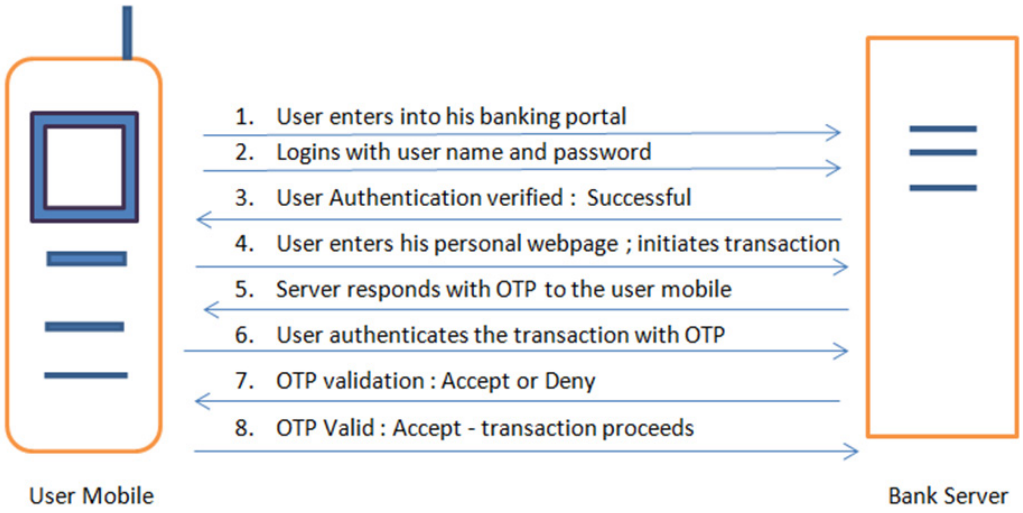


Figure 1. OTP generation and user authentication[1].

key known only to both the sender and the receiver. This method of steganography is trickiest as it uses the text to conceal which has very little redundant information unlike image, video or audio that can be used for hiding as in Figure 2.

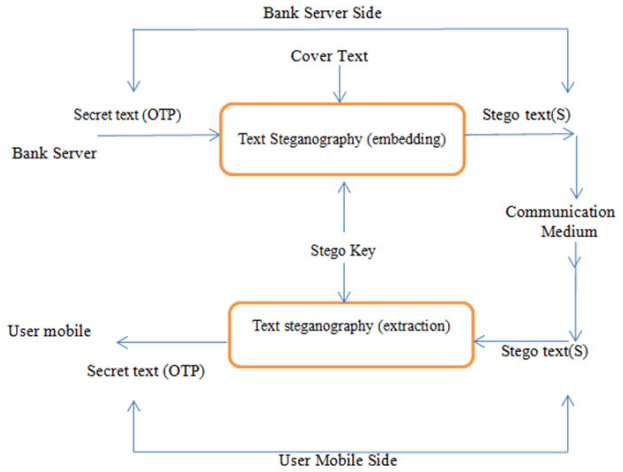


Figure 2. OTP (One-Time Password) from Bank to User mobile.

Text steganography can be used to hide the OTP that is sent to user mobile for authentication. An innocent cover text with the OTP embedded, can be sent from the bank server as SMS to the user mobile. The user can extract the OTP with the stego-text and authenticate himself[2].

1.3 Unicode Standard

Unicode standard is the universal coding of characters internationally that takes a written forms of characters and text. It includes all the multilingual characters that are used in different parts of the globe. It defines a consistent way of encoding multilingual text that enables the exchange of information internationally. Unicode can be implemented by different character encoding with the most commonly used are UTF-8, UTF-16. UTF-8 is similar to ASCII coding that uses the same 8 bit code. UTF-16 or UCS-2 uses two byte for each character[3].

1.4 Frequency of Occurrence of English Letters

The use of letter frequencies and frequency analysis plays a fundamental role in cryptograms and several word puzzle games, including Hangman, Scrabble and the television game show Wheel of Fortune. The analysis of the entries in the Concise Oxford dictionary gives the list of letters that occur very frequently in any piece of English text[4]. This feature of English language is used to hide more amounts of secret bits inside the cover text thereby improving the payload capacity of the stego-text drastically.

2. Threats to OTP

The OTP SMS during m-banking transactions is vulnerable to various attacks along the communication channel as it travels from Short Message Entity (SME), SMSC (Short Message Service Centre), SMS GMSC (SMS gateway MSC), HLR (Home Location Register), MSC (Mobile Switching Centre), VLR (Visitor Location Register), BSS (Base Station System)[5].

Table 1. Selected English alphabets for hiding process based on frequency of occurrence

S.No	Letter frequency %	Letter symbol	Secret pair- 00	Secret pair-01	Secret pair-10	Secret pair-11
			ASCII code	Unicode	Unicode	Unicode
1.	8.167	a	0061	0251	0430	FF41
2.	12.7	e	0065	0435	212F	FF45
3.	6.094	h	0068	04BB	13C2	FF48
4.	6.966	i	0069	0456	2170	FF49
5.	7.507	o	006F	03BF	043E	1D0F
6.	1 per line	- (hyphen)	2014	2500	2012	2013
7.	1 per line	. (full Stop)	002E	00B7	0323	02D1
8.	19.18182	(space)	00A0	2000	2003	2002

2.1 Wireless Interception

The research shows that the communication path between mobile phones and base stations can be eaves dropped and decrypted due to loopholes in protocols[6].

2.2 Mobile Phone Trojans

The mobile phone Trojans are innocent looking malware designed by the criminals to steal OTP with a prime motive of making money. ZITMO (Zeus-In-The-MObile) is used to steal credentials and mSpy is a mobileapp installed to monitor all the activities performed in the mobile under siege are some of known malware Trojans[6].

2.3 Phishing

Phishing attack is a form social engineering that gains user's trust and confidence to gather user's credentials of economic value using deceptive looking websites, e-mails and many more[7].

2.4 Man-in-the-Browser (MITB) Attacks

In MITB attack, the attacker gains control of the online banking interface through compromised credentials and perform the customer's intended banking transaction. The attacker takes charge of the channel between user and the browser[8].

2.5 Passwords Stealing and Identity Theft

The user name and password that are used to authenticate a genuine customer are stolen using shoulder-surfing and guessing by person of close association[9].

3. Proposed Technique: Text Steganography Using Unicode and Frequency of Occurrence in English Letters

When an m-banking transaction is initiated, as a second level of authentication, OTP is generated. The OTP is need to be hidden using text steganography to prevent the various threats along the, communication medium.

The proposed idea is to use the multilingual character glyphs of Unicode that are look-alikes with the English characters of the cover text to improve invisibility as listed in Table 1. The frequency of occurrence of English letters gives a maximum embedding capacity thereby improves the payload capacity. The letters that have frequency more than six percent are alone considered for hiding. The basic criteria that measure the goodness of a good steganographic algorithm are addressed.

These 8 characters are chosen from Unicode and used to hide the 2 binary bit of secret information. So, if the OTP is, say 6 digits, then coded in BCD makes it 24 bits. The first 12 characters of the cover text are used for hiding the OTP (in BCD form) and the stego-text is sent to the user mobile as SMS.

3.1 Embedding Process

The secret text is M is to be embedded in the cover text based in the procedure given below.

Say the random cover message is taken. It should have at least 24 characters in order to code 24 bits of the cipher

text. Assume, 'Bank on the go with our mobile banking service. Anytime, anywhere bank', be the random cover text. The characters are selected for hiding based on the Table 1 is given in italics. The capitalized letters are ignored.

B*a*nk *o*n th*e* g*o* w*i*th *o*ur m*o*b*i*l*e* b*a*nk*i*ng s*e*rv*i*c*e*. Anyt*i*m*e*, *a*nywh*e*r*e* b*a*nk.

There are 20 characters + 11 spaces + 2 periods = 33 letters can be used for hiding.

Let the OTP be 439767. The OTP coded in BCD form is

M = 0100 0011 1001 0111 0110 0111

It requires only 12 characters for hiding. So the first 12 characters of the cover, present in the table are used.

Embedding algorithm:

Input : Cover text , Stegokey and M

Output : Stego-text, S

(i) Open a random cover text. Convert OTP to BCD form (M).

(ii) Scan the cover text to pick out the selected characters used for hiding as given in the table.

(iii) Compute the number of characters required for embedding.

(iv) For each two digits of the OTP in BCD, compute

if bits = 00 ***then*** no change (ASCII code)

else replace it with the multilingual glyphs as in table 1.

Repeat step (iv) until all the bits of M is embedded.

(v) Return the stego text.(S)

This is the message i.e. stego-text, S sent by the bank server as SMS via the communication medium.

3.2 Extraction Process

The SMS is received at the user mobile. The extraction of the OTP happens at the user mobile using the same table used for embedding. Only the first 12 characters (as given in the table 1) of stego-text are exacted and checked for the Unicode values. Based on the values secret OTP, M is obtained.

Extracting algorithm:

Input : Stego text, S and Stegokey

Output : Secret message M

(i) Open the stego text S.

(ii) Scan the stego text to find the first 12 ligatures characters that hides the secret.

(iii) Check the code of the selected characters.

(iv) ***If*** the code is within U+0061 to U+007A

then the secret message is 00

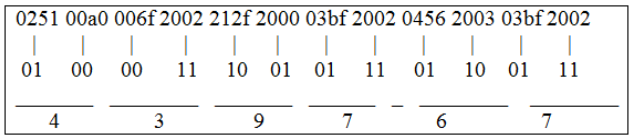
else secret message is { 01,10,11 } based on table 1

(v) Repeat steps (iii) and (iv) for first 12 characters to obtain the binary OTP in BCD (M).

The Unicode equivalent of the 12 ligatures used for hiding after step (iii)

0251 00a0 006f 2002 212f 2000 03bf 2002 0456 2003 03bf 2002

The Output of the extracting algorithm:



The extracted binary pairs are converted to BCD form and are used for user authentication to complete the initiated m-banking transaction.

4. Analysis of the Results

In this paper, the proposed method uses Unicode and letter frequency of English alphabets to increase payload. The characters are encoded in UTF-16 that occupies two bytes per character. As the stego-text with the secret OTP is sent via mobile SMS, the maximum characters are restricted to 70 Unicode characters/SMS unlike other methods where there is no restriction on the size of the cover text[10-12]. This method has the highest payload within a limited size of 70 characters. The appearance is also invisible. The Figure 3 gives the plain text and Figure 4 shows the stegotext with OTP embedded using the proposed technique of replacing with Unicode characters.

Bank on the go with our mobile banking service. Anytime, anywhere bank.

Figure 3. The plain text.

Bankonth gowithourmobile banking service. Anytime, anywhere bank.

Figure 4. The Stego-text with OTP5.

4.1 Similarity measure using Jaro-Winkler Distance

For comparing the similarity between cover text and stego-text, Jaro Winkler distance[13] method for comparing two strings was computed using the formula.

$$d_w = d_j + (l * p(1 - d_j))$$

where d_j is the Jaro distance for strings s_1 and s_2, l is the length of common prefix at the start of the string up to a maximum of 4 characters, p is a constant scaling factor for how much the score is adjusted upwards for having common prefixes. The standard value is $p = 0.1$.

The Jaro-Winkler distance is a measure of similarity between two strings. It is a refinement of the Jaro distance metric ($\boldsymbol{d_j}$)[13]. The higher the Jaro-Winkler distance score the strings are said to be more similar. The score is within the range of 0 to 1, where 0 implies no similarity and 1 implies exact match. Table 2 shows the list of various similarity coefficients between the 24 bits OTP and the cover text.

Table 2.

Cover text size(in chars)	Matching Characters	Jaro-Winkler Score (d_w)
70	12	0.8971
70	10	0.9142
70	8	0.9313
69	12	0.8956
69	10	0.9130
69	8	0.9304
68	12	0.8940
68	10	0.9117
68	8	0.9293
67	12	0.8925
67	10	0.8925
67	8	0.9283
66	12	0.8908
66	10	0.9090
66	8	0.9272
Average score :		**0.9105**

The average Jaro-Winkler score of comparing 15 samples of cover text and the stego-text is 0.9105, which means they closely similar within a restriction of 70 characters as cover text size.

4.2 Similarity m easure using Histogram

For visually comparing the similarity between the plain text and stegotext histogram technique has been done. These histograms as shown in Figure 6 and Figure 7 visually explain that both plain text and stegotext with the embedded OTP are almost identical and look alike. This proves the invisibility of the secret message and thereby hides its existence.

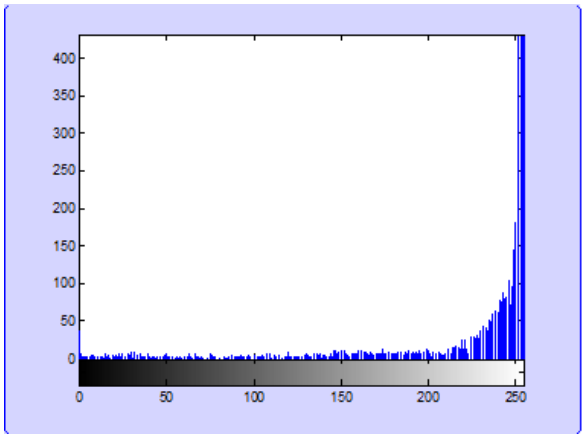


Figure 5. Histogram of plain text.

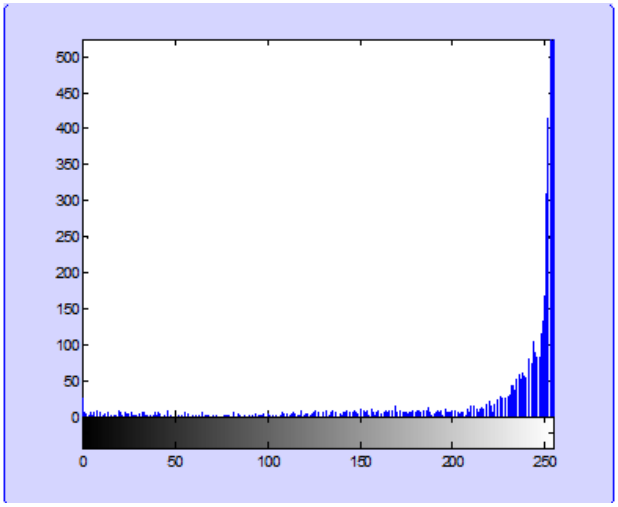


Figure 6. Histogram of the stego text with OTP.

5. Conclusion

The main criteria for any data hiding method are high payload capacity, robustness and invisibility. The proposed method has high level of invisibility because the alphabets of the cover text are replaced with look-alike glyphs of Unicode characters based on the BCD equivalent of the secret text. The frequency of occurrences property of English letters is used to increase the payload capacity. The stego-text is robust and preserves the hidden information during all kind of transformation like compression/ decompression, change of font style, copy-paste and many more. On the other side, as all characters used here, are encoded using Unicode that occupies two bytes, the length of the cover text can be no more than 70 characters

per SMS. The proposed method provides an efficient Text Steganography technique to hide 6 digits OTP with high payload within the restriction of 70 Unicode characters per SMS. This method will also work even if there is an increase in the number of digits of OTP; say from 6 to 8 or even more.

6. References

1. Ananthi S, Sumathy D. OTP Encryption Techniques in Mobiles for Authentication and Transaction Security. International Journal of Innovative Research in Computer and Communication Engineering. 2014 Oct; 2(10):6192–201 ISSN (Online): 2320–9801 ISSN (Print): 2320–9798.
2. Ananthi S, Sumathy K. A Systematic Approach towards the techniques of Text Steganography. Proceedings of International Conference on Research Trends in Computer Science. Chennai: 2013 Aug. p. 88–96.
3. Prof. Abdul SRAMS, et. al. Text Steganography Based on Unicode of Characters in multilingual. 2013 Jul-Aug; 3(4):1153–65. ISSN: 2248–9622.
4. Frequency of English letters. 12/12/2014; Available from: https://www.wikipedia/letter_frequency.
5. Ananthi S, Sumathy D. A Framework of security issues and standards for efficient SMS. International Journal of Computer Technology and Applications. 2014 Mar-Apr; 5(2):469–77. ISSN 2229–6093.
6. Collin M, et. al. SMS-based one-time passwords: attacks and defense. DIMVA 2013. LNCS. 2013; 7967; Springer-Verlag: Berlin Heidelberg. p. 150–9.
7. Safa H, et. al. Securing SMS based one time password technique from man in the middle attack. International Journal of Engineering Trends and Technology. 2014 May; 11(3):154–58. ISSN:2231–5381.
8. Dauda S, CISA. Man in the Browser – A Threat to Online Banking. ISACA Journal. 2013; (4):1–3.
9. Attacks on SMS. 15/05/2014; Available from: http://securityblog.s21sec.com/2010/09/zeus-mitmo-man-in-mobile-i.html.
10. Text to Unicode convertor. 15/12/2015 Available from: https://www.branah.com/unicode-converter.
11. Indradip B, et. al. Novel text steganography through special code generation. Proceedings of the International Conference on Systemics, Cybernetics and Informatics. 2011. p. 298–303.
12. Souvik B et. al. Hiding Data in text using ASCII Mapping Technology (AMT). International Journal of Computer Applications. 2013 May; 70(18). (0975 – 8887)
13. Similarity measure for strings. 05/01/2015; Available from: http://en.wikipedia.org/wiki/Jaro-Winkler_distance.