

Available online at <http://www.mecspress.net/ijeme>

Text Steganography using Daily Emotions Monitoring

Sivabalan A/L Patiburn, Vahab Iranmanesh *, Phoey Lee Teh

Department of Computing and Information Systems Faculty of Science and Technology Sunway University
Bandar Sunway, Malaysia

Abstract

Early in 2008, it was emphasized that sending messages through public networks will draw the attention of third parties such as attackers, perhaps causing attempts to break and reveal messages. Thus, cryptography was initially used to send a text message by producing cipher text. Although cryptography provides confidentiality for sent message, but the cipher text that is generated appears nonsense for the attacker, which leads to being identified as the sensitive information. In order to overcome this problem, this paper introduces a novel mobile steganography technique, based on using emoticons to deliver sensitive information as a daily emotion monitoring application. Based on achieved result, 88 characters can be embedded and sent to the recipient successfully in a hidden way.

Index Terms: Text Steganography, Mobile Application, Emoticons.

© 2017 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

In the era of Information Technology (IT), cybercrime has always been a concern for online users. Phishing, social engineering, and third party attacks have made people reluctant to share their personal information. The necessity of security rises as soon as sensitive data's such as password, credit card number and personal details are accessible and can be captured by attackers [1]. Thus, cryptography was initially invented to send secret messages by producing cipher-text that is known as encrypted or encoded information. Since cipher text contains a form of the original message that does not make sense to a human or the computer without proper key to decrypt it, however, the main problem is that the generated cipher text appears gibberish and it can be detected by attackers easily [2-4]. On the other hand, the hidden data can be embedded within the other information that is called cover media in steganography. Steganography comes from the Greek words "stegano" and "graphy", meaning covered writing [5, 6]. In fact, steganography can provide non-significant changes on the cover media, which is difficult to be detected by human senses [7]. Fig. 1 shows two main processes, namely embedding and extraction that are used in steganography. In the embedding part, sender

* Corresponding author

E-mail address: vahab.iranmanesh@gmail.com

embeds hidden data within cover media in order to generate a stego media that contains a secret message. In the extraction side, since the recipient is aware of the embedding algorithm, he will use the same methodology, but in a reverse way to extract the secret message from the same stego media. Moreover, both sender and recipient can use an optional key as stego key in embedding and extraction processes, which can provide difficulties for the attacker to either, guess the steganography technique or extract the hidden data.

In general, the steganography can be classified into several approaches, based on which cover media such as text, image, audio and video is used to hide the data [7-10]. Additionally, since text requires less memory and time to for processing the data, it has been used as a cover media for several years [2, 3, 11]. However, text steganography is the most challenging method among other steganography techniques because it is believed to have deficiency of redundant information within the text file comparing to the image, audio and a video files, which can be exploited for hiding the data [7, 8, 9, 12]. This is due to the observable structure of the text, which means that any distortion within the text file as data hiding method can modify the text structure as well. Therefore, in order to overcome this issue, a mobile application is developed to use emoticons as the cover media in order to generate the stego media that it seems to be user's daily emotions (feelings) in different times. In fact, the main reason to use emoticons for hiding the data is quite a number of emoticons, which is about five billion emoticons per day by people through communication channels around the world [13]. As a result, a user is able to send maximum 88 characters per text to the recipient, which can be ignored to capture and understand by the attacker.

The rest of this paper is organized as follows. Section II describes the literature review in text steganography. Section III describes our proposed steganography technique. Section IV discuss our findings and presents conclusion and directions for future work.

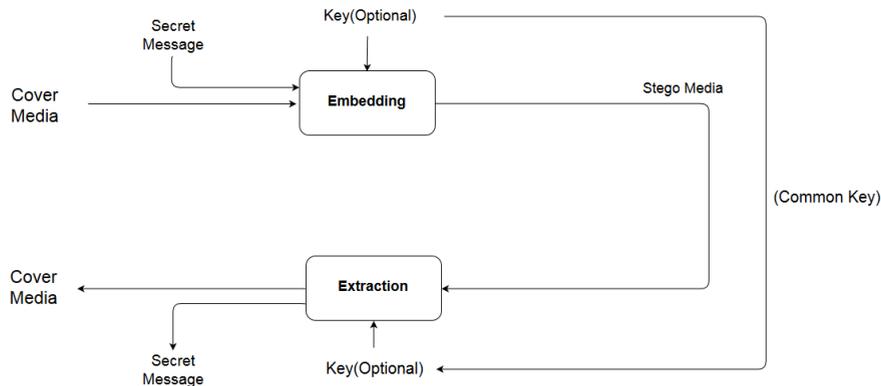


Fig.1. Steganography Process.

2. Related Work

Steganography is invented as soon as the need for privacy became important. In general, steganography techniques can be classified into different approaches such as image, text, audio and video steganography depending on the cover media that is used to embed the hidden data [7-12, 14]. Moreover, text steganography is one of the oldest information hiding techniques, which is considered as the hardest cover media to use for steganography purpose due to less redundant information available inside the text files for hiding the data [7, 8, 10, 11, 14-16]. Therefore, several researchers have developed some text techniques over the past decade, which can be categorized under three approaches, namely format-based, random and statistical generation and linguistic [11, 15, 17, 18], as fig, 2 shows. In fact, each text steganography approach manipulates the cover text (text file) in a different way to hide the secret message. In the following, several text steganography techniques that are categorized under each approach are explained in details.

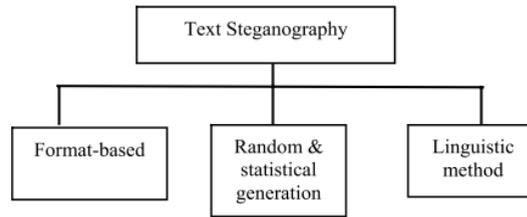


Fig.2. Text Steganography Approaches [10].

2.1. Format Based

Format-based text steganography modifies the existing formats of the text file in order to hide the secret message [1, 9, 10, 14, 15]. In general, the secret message that is used by any of techniques under this approach can be hidden in a way that it can be difficult to identify by human perception, whereas a computer as a machine can be utilized to detect the changes easily [1, 9]. In the following, several format based text steganography techniques in conjunction with examples are described.

2.1.1. White Spacing

White space text steganography is a common information hiding method, which utilizes the spaces within the text file that can be found in different locations such as the end of each sentence, each line, between words or after each paragraph [9, 12, 14]. As fig. 3 shows an example of adding white spaces at the end of each line, which can be used to embed any hidden data within the text file. However, in a case that the number of whitespace within the text is not sufficient, few characters as the secret message can be hidden respectively [7]. In addition, by comparing both modified and non-modified text file using any kind of text processor such as Microsoft word, the difference as whitespaces can be leaked as the hidden data [7].

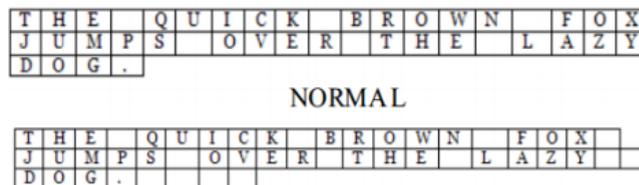


Fig.3. White Space Technique [16].

2.1.2 Word Shifting

World shifting is similar to white space approach in term of adding white space between the words, however, it is different since more than one white space can be added to the cover text horizontally left or right for encoding the hidden data [8, 14, 16]. Fig. 4 shows an example of word shifting by adding a white space between some words to hide the data. However, similar to white space approach, a comparison process between the original text file (cover text) and modified one (stego text) leads to find the differences between them as hidden data [3]. Moreover, retyping the stego text or using Character Recognition Programs (OCR) to scan printed stego text can destroy the hidden data [15].

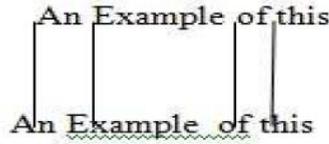


Fig.4. Word Shifting Technique [17].

2.1.3. Line Shifting

In line shifting approach, as it is cleared from its name, a degree of vertical shifting can be applied on each or selective lines within the text file to represent the hidden data [8, 12, 16]. For example, by assuming that the hidden data is converted into a binary format, in order to hide bit 0, a line is shifted up whereas shifting line to down can hide bit 1, as shown in fig. 5. Retyping the text file and using Character Recognition Program (OCR) are two limitations of using line shifting text steganography approach, which can destroy the hidden data [14, 15].

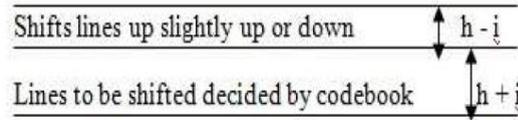


Fig.5. Line Shifting [17].

2.1.4. Language Feature Coding

In general, different languages are represented using different alphabets; this makes unique characteristics for each, which can be used for steganography purpose. For instance, since Persian and Arabic letters that contain a number of points/dots in some letters, around 18 letters, it allows this opportunity to hide secret messages by altering their positions [9, 14]. As fig. 6 shows, the location of a dot on top of the letter “F” in Persian/Arabic is changed slightly lower than actual location (right) to represent the hidden data. However, due to non-common alphabet characteristics among various languages, language feature coding text steganography approach cannot be generalized for the other languages [7, 12, 14].



Fig.6. Persian/Arabic Letter [12].

2.1.5. Emoticons and Lingo Embedding

Since people around the world have changed their behaviours to use several emoticons and lingo in their daily conversations instead of typing their feelings, several researchers have used as steganography features for hiding the data [8, 19]. In this case, a random text that contains several words that can be replaced with some emoticons and/or lingo is generated first. Later on, based on which character is mapped into which emoticon or lingo, they can be replaced with the words within the generated text file in order to represent the hidden data. For instance, fig. 7 shows an example of using emoticons and lingo to map the secret message “Attack”,

with each emoticon and lingo representing one character. Therefore, the characters “a”, “t”, “l”, “a”, “c” and “k” are replaced with “wink”, “tbh”, “tbh”, “wink”, “shy”, and “plox”, which are fitted within the generated random text. However, making the random text can be a challenging task since it must be generated in a way to contain several emoticons and/or lingoes for replacing with the hidden data.

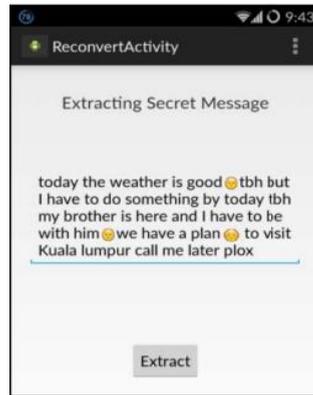


Fig.7. Emoticons Embedding [19].

2.1.6. Mathematics Equation

In different approach [14], a mathematical method is suggested, which an equation can be formed with an arbitrary number of variables and mathematical symbols, such as constants, operations, and functions. In fact, the idea behind suggested idea is to decrease the attention of the attacker by making some equations that are pretended to be a math quiz. In this way, the mapping technique is implemented by first changing the secret message into random values, to further complicate message identification of a third party. Then, each letter of the English alphabet, uppercase and lowercase, the numbers 0-9, and whitespace, have their own values that are pre-set. With the mapped values from the previous step, the next step now is to embed the values into an equation. In order to create an equation, a different key can be used. Fig. 8 shows the word “Attack Now”, which is converted into a mathematical equation that appears like a quiz in the recipient side.

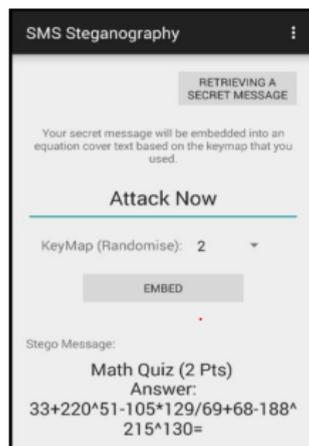


Fig.8. Math Equation Technique [14].

2.1.7. Sudoku Puzzle

Sudoku puzzle is the other technique, which hides a secret message in a 9x9 Sudoku puzzle [14, 20]. The secret message can be embedded into any rows and columns of the puzzle, which will be received by the recipient to extract the secret message from the puzzle. Fig. 10 shows a secret message hidden in one of the rows or columns of the Sudoku puzzle. Based on that, in a 9x9 Sudoku puzzle each of the numbers 1 through 9 appears in each row or column just once and there is no repetitious number in any row or column. As a result of this characteristic, we can arrange the numbers in each row or column in 9! Permutations, of course, when a row or column is set, we won't be able to set the other rows or columns as we wish, because they must observe the regulation of non-repetitiousness of the numbers lying in each row or column.

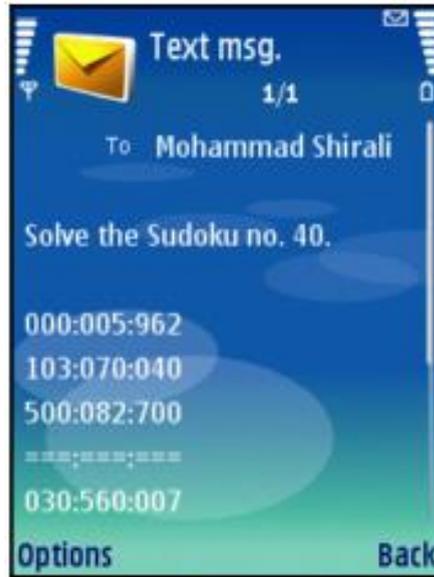


Fig.9. Sudoku Puzzle Technique [15].

2.2. Random and Statistical Generation

This approach is referred to generating a random text to play the role of the cover text, based on the character or word sequences [1, 8, 10, 12] at random method. In this way, random sequence of either character or words are utilized to hide the data. In addition, several sentences can be mimicked, which can have the same statistical properties with the original text file in order to encode the hidden data using statistical method. Although the new generated text file has similar statistical property with the original text file, which can be robust against statistical attacks, however in a case that the generated text doesn't provide meaning, this leads attacker to identify the hidden communication easily.

2.3. Linguistic Method

The linguistic text steganography is relied on two approaches, namely syntactic and semantic [10, 12]. In the syntactic method, finding the proper places to put several punctuations marks such as comma and full stop is the main challenge for encoding the hidden data [11, 12, 17, 2]. Thus, the number of the hidden characters (capacity) that can be hidden within the text would not be sufficient. For instance, as fig. 10 shows, a comma is

used after word “Albert” to represent a hidden data. On the other hand, the semantic method is referred to using alternative words as an indicators for representing the hidden data. Fig. 11 and 12 show two examples of semantic approach, which are relied on using abbreviation [7, 12, 16] and different word spelling [10, 11], based on UK and UK spelling, for hiding the data. In fact, both word abbreviation and spelling methods are able to store small information into the text since the text may not contain sufficient number of words that can be shorten as well as alternative words between UK and US languages [12]. As a result, in a case that the printed stego text will be re-typed, it may lead to destroy the hidden data.

Good morning Albert how are you
 Good morning Albert, how are you?

Fig.10. Syntactic Method [14].

1	0
218	Too late
ASAP	As Soon As Possible
C	See
CM	Call Me
F2F	Face to face

Fig.11. Abbreviation Technique [18].

American English	British English
Favorite	Favourite
Criticize	Criticise
Fulfill	Filfil
Center	Centre
Dialog	Dialogue
Medieval	Mediaeval
Check	Cheque
Defense	Defence
Tire	Tyre

Fig.12. Word Spelling Technique [15].

3. Proposed Method

In this study, an android mobile application is developed to utilize emoticons as features for encoding and decoding the hidden data in a hidden way. In fact, it is assumed that the developed text steganography technique is kind of mobile application, which allows users to share their daily emotions (feelings) in different times using emoticons. The main reason to use emoticons for steganography purpose in this study is that since many people have used them to deliver their expressions in texts based communication systems such as SMS and social apps; they are considered to be exploited as the cover text for hiding the data. Thus, it provides this opportunity for the sender to encode the hidden data in a format that is less suspicious for the attacker. As a result, this will enable the integrity and confidentiality of the hidden data to be preserved.

Similar to the text steganography techniques that have been discussed before, the methodology that is used to develop suggested text steganography is divided into two parts, as shown in fig. 13. In the sender side, a process that is called embedding is used to encode the hidden data using emoticons in order to send it through public channels such as SMS or social apps. On the other hand, the recipient uses the extraction process, which is similar to encoding process but in a reverse way, in conjunction with a common key (optional) to decode the hidden data from generated emoticons. In the following, both embedding and extraction processes are explained in details.

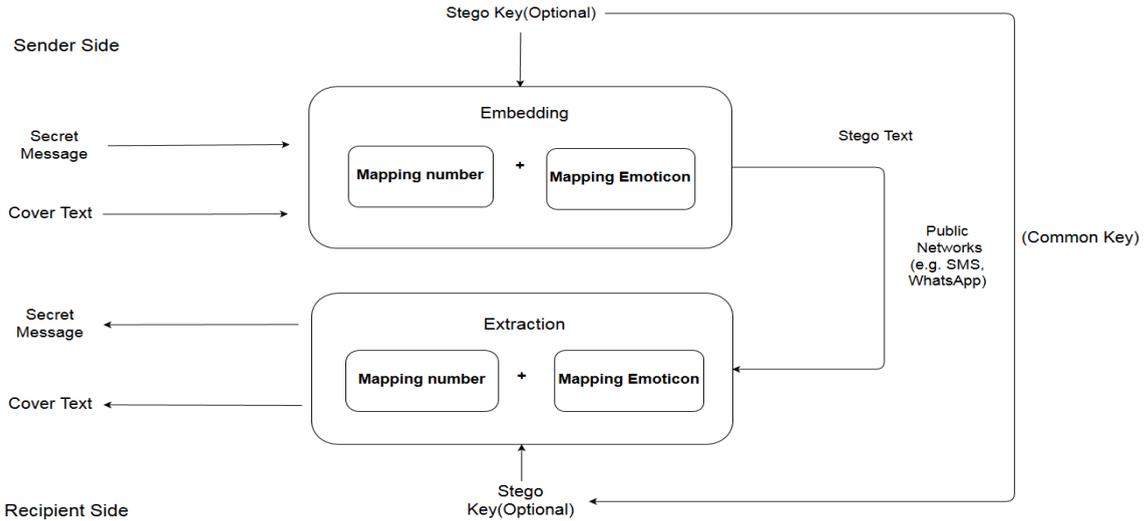


Fig.13. Proposed Steganography Technique.

3.1. Embedding Process

Fig. 14 shows the GUI of the developed mobile application, which can be used for hiding the data. In this way, a user can type his/her secret message within the pink box, which will be sent to the recipient. Furthermore, the user must select a random key among several pre-defined keys in the grey box, which plays the main role in the embedding process. Table 1 shows an example of keys (e.g. 1), which can convert each character of hidden data into proper emotions. For example, assuming the character “h” and “i” are the hidden data that must be sent to the recipient. In this way, a fixed value, 7 and 8 respectively, will be assigned to each character in order to map into proper emoticons. In addition, in a case that different key will be chosen by the sender, different numbering as well as emoticons will be generated.

Once the hidden data characters are converted into proper emoticons, in order to decrease the level of suspicion on the generated emoticons as the stego text, a fixed sentence “Today I Felt: “ is located as a pre-fix as well as a random time as the timestamp after each emoticon to pretend that the user is sending his/her emotions (feeling) for the mentioned times. Additionally, generating timestamps is based on one random initial timestamps that is started from 1.00 a.m. and it continued till 11.59 p.m. by incremental step that changes previous timestamp minute and hour randomly. As a result, the generated stego text can be shown in the blue box, which will be sent to the recipient for decoding.

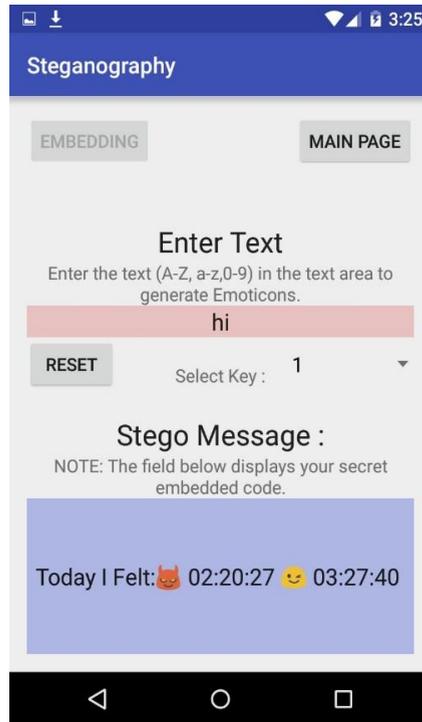


Fig.14. Embedding GUI.

Table 1. Character and Number Conversion using Key 1

Key 1		
User Input	Mapped Number	Emoticons
A	0	😄
B	1	😁
C	2	😂
D	3	😊
E	4	😇
F	5	😏
G	6	😌
H	7	😈
I	8	😊
J	9	😋
K	10	😍

L	11	
M	12	
N	13	
O	14	
P	15	
Q	16	
R	17	
S	18	
T	19	
U	20	
V	21	
W	22	
X	23	
Y	24	
Z	25	
0	26	
1	27	
2	28	
3	29	
4	30	
5	31	
6	32	
7	33	
8	34	
9	35	
Space	36	

3.2. Extraction Process

Similar to embedding process, two sub-steps mapping number and mapping emoticon are utilized in the

stego text, but in a reverse way to extract the hidden data. Fig. 15 shows the extraction GUI, which uses the same stego key (grey box) that is used by the sender in conjunction with the sent stego text (green box) to extract the hidden data (yellow box). In fact, as fig. 16 shows, selecting wrong key by the recipient leads to produce gibberish words, which doesn't provide any meaning for the recipient. In addition, before processing the attained stego text, a filtering is implemented to remove unnecessary data, such as generated timestamps and pre-fix sentence "Today I Felt:" Thus, only emoticons, which are the main features for hiding the data, are passed to the next step. Therefore, mapping to emoticon sub-step is applied on the filtered cover text to convert each emoticon into its respective number. As a result, the attained numbers are used by mapping number sub-step to map each number to define character, based on the chosen stego key (e.g. 1).

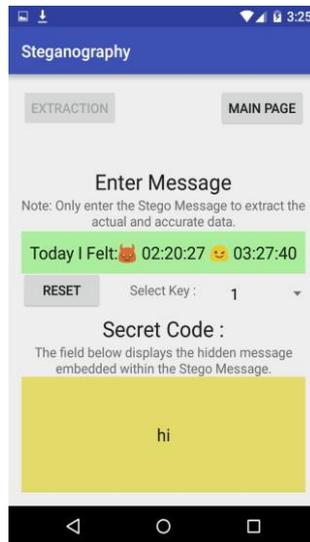


Fig.15. Extraction GUI.

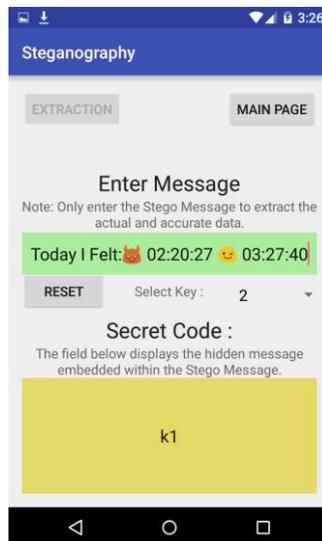


Fig.16. Wrong Hidden Data.

4. Discussion & Conclusion

Text steganography is one of data hiding techniques, which is utilized the text file in order to deliver the data in a hidden way. However, due to the limitation of redundancy of data within the text file, it is considered as the most challenging steganography approach among the other steganography techniques. In order to overcome to this issue, a novel steganography technique, based on developing a mobile application suggested to embed the hidden data into emoticons, which can be considered as a daily emotions monitoring tool. This is due to the pre-fix sentence "Today I Felt:" and timestamps that are used along the emoticons, which provide a better appearance against being detected by the attacker.

Furthermore, the suggested steganography provides more capacity since each byte (character) of hidden data can be encoded using an emoticon rather than bit encoding [9]. In addition, due to the generating fixed text as pre-fix and timestamp, compare to any random sentence that is generated in [19], more characters can be hidden. Furthermore, using a key to reshuffle the number and emotions that are assigned to each character is similar to the technique that is used in [14], which can provide more security for the suggested steganography technique. However, the maximum number of characters (capacity) that can be hidden using proposed text steganography is 88 characters, due to the limited timestamps that are generated by incremental step randomly. In the future, we are planning to add more features such as supporting special characters. This will enable people to send out data like email address and password. Secondly, we are planning to add a math function to the embedding and extraction process, in order to provide a degree of complexity for the attacker.

References

- [1] I. Banerjee, S. Bhattacharyya and G. Sanyal, "Study and Analysis of Text Steganography Tools", *International Journal of Computer Network and Information Security*, vol. 5, no. 12, pp. 45-52, 2013.
- [2] M. Agarwal, "Text Steganographic Approaches: A Comparison", *International Journal of Network Security & Its Applications*, vol. 5, no. 1, pp. 91-106, 2013.
- [3] N. Rani and J. Chaudhary, "Text Steganography Techniques: A Review", *International Journal of Engineering Trends and Technology (IJETT)*, vol. 4, no. 5, pp. 3013-3015, 2013.
- [4] A. Majumder and S. Changder, "A Novel Approach for Text Steganography: Generating Text Summary Using Reflection Symmetry", *Procedia Technology*, vol. 10, pp. 112-120, 2013.
- [5] A. AminAli and A. Seddik Saad, "New Text Steganography Technique by using Mixed-Case Font", *International Journal of Computer Applications*, vol. 62, no. 3, pp. 6-9, 2013.
- [6] S. Gupta and R. Jain, "An Innovative Method of Text Steganography", in *Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP)*, Solan, India, 2015, pp. 60-64.
- [7] M. Shirali-Shahreza and M. Shirali-Shahreza, "Text Steganography in SMS", *International Conference on Convergence Information Technology (ICCIT)*, Gyeongju, Republic of Korea, 2007, pp. 2260-2265.
- [8] T. P. Nagarhalli, "A New Approach to SMS Text Steganography using Emoticons", *National Conference on Role of Engineers in Nation Building (NCRENB)*, 2014, pp. 1-3.
- [9] S. Dulera, D. Jinwala and A. Dasgupta, "Experimenting with the Novel Approaches in Text Steganography", *International Journal of Network Security & Its Applications*, vol. 3, no. 6, pp. 213-225, 2011.
- [10] S. Kingslin and N. Kavitha, "Evaluative Approach towards Text Steganographic Techniques", *Indian Journal of Science and Technology*, vol. 8, no. 29, 2015.
- [11] M. Agarwal, "Text Steganographic Approaches: A Comparison", *International Journal of Network Security & Its Applications*, vol. 5, no. 1, pp. 91-106, 2013.
- [12] A. Abdul-Aziz Gutub and M. Fattani, "A Novel Arabic Text Steganography Method Using Letter Points and Extensions", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 1, no. 3, pp. 502-505, 2007.

- [13] T. Dua, "Emojis by the numbers: A Digiday data dump - Digiday", Digiday, 2015. [Online]. Available: <http://digiday.com/brands/digiday-guide-things-emoji>. [Accessed: 16- Oct- 2016].
- [14] L. Min Yang, V. Iranmanesh and J. C. Quiroz, "A New Approach to SMS Steganography using Mathematical Equations", *International Conference on Computer Applications & Technology (ICCAT)*, Rome, Italy, 2015.
- [15] L. POR and B. Delina, "Information Hiding: A New Approach in Text Steganography", *7th International Conference on Applied Computer and Applied Computational Science (ACACOS)*, Hangzhou, China, 2008, pp. 689-695.
- [16] H. Singh, P. Singh and K. Saroha, "A Survey on Text Based Steganography", *Proceedings of the 3rd National Conference*, New Delhi, 2009, pp. 26-27.
- [17] V. Saraswathi and S. Kingslin, "Different Approaches to Text Steganography: A Comparison", *International Journal of Emerging Research in Management & Technology (ERMT)*, vol. 3, no. 11, pp. 124-127, 2014.
- [18] S. Samanta, S. Dutta and G. Sanyal, "A Novel Approach of Text Steganography using Nonlinear Character Positions (NCP)", *International Journal of Computer Network and Information Security*, vol. 6, no. 1, pp. 55-60, 2013.
- [19] V. Iranmanesh, H. Jing Wei, S. Lee Dao-Ming and O. Ayodeji Arigbabu, "On using Emoticons and Lingoes for Hiding Data in SMS", *Technology Management and Emerging Technologies (ISTMET)*, Kuching, Malaysia, 2015, pp. 103-107.
- [20] M. Shirali-Shahreza and M. Shirali-Shahreza., "Steganography in SMS by Sudoku Puzzle", *International Conference on Computer Systems and Applications (ACS)*, Doha, Qatar, 2008, pp. 844-84.

Authors' Profiles



Sivabalan A/L Patiburn completed diploma in Information Technology at Sunway University Malaysia. Currently, he is pursuing his Degree in Computer Security & Networking.



Vahab Iranmanesh, graduated from University Technology Malaysia (UTM) in information security in 2010. Currently, he is pursuing his Ph.D in communications and computer networks engineering at University Putra Malaysia (UPM). Vahab has been working as a lecturer at Sunway University Malaysia since 2014. He has done several researches in the field of steganography, which has led to publish several research papers in various conferences. His research interests include steganography, information security, network security and handwritten signature verification.



Teh Phoey Lee, graduated from University Putra Malaysia (UPM) in management information systems. Currently, she is associate professor of information systems at Sunway University, Malaysia since 2013. Her research interests include visualization, mobile application, data mining and data analysis.

How to cite this paper: Sivabalan A. L Patiburn, Vahab Iranmanesh, Phoey Lee Teh, "Text Steganography using Daily Emotions Monitoring", *International Journal of Education and Management Engineering (IJEME)*, Vol.7, No.3, pp.1-14, 2017.DOI: 10.5815/ijeme.2017.03.01