

A Novel Text Steganography System in Cricket Match Scorecard

Md. Khairullah

Assistant Professor, Department of Computer Science and Engineering
Shahjalal University of Science and Technology, Sylhet
Bangladesh

ABSTRACT

Steganography can be defined as a method of hiding data within a cover media so that other individuals fail to realize their existence. Image, audio and video are some popular media for steganography. But text is ideal for steganography due to its ubiquity and smaller size compared to these media. However, text communication channels do not necessarily provide sufficient redundancy for covert communication. In this paper, a new approach for steganography in cricket match scorecard is proposed. The main idea is that additional zeroes can be added before a number without changing the value of a number in a scorecard.

General Terms

Steganography, Information Security

Keywords

Steganography, Scorecard, Cover media

1. INTRODUCTION

Cricket is a team sport played between two teams of eleven players. It is a bat-and-ball game played on a roughly elliptical grass field, in the centre of which is a flat strip of ground 22 yards (20.12 m) long, called a pitch [1]. The game of cricket is the second most popular game in the world, second only to soccer [2]. Cricket is first documented as being played in southern England in the 16th century. By the end of the 18th century, cricket had developed to the point where it had become the national sport of England. The expansion of the British Empire led to cricket being played overseas and by the mid-19th century the first international matches were being held. Today, the sport is played in more than 100 countries [3]. A scorecard is the instant source of information on the current status of the game of cricket. It is probably the easiest form of knowing the exact position of any match. The concept of scorecard is popularly used in every game but it has a major significance in cricket as the score changes with every ball [4].

Steganography is the art and science of writing hidden messages in such a way that no-one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity [5]. Cryptography — the science of writing in secret codes — addresses all of the elements necessary for secure communication over an insecure channel, namely privacy, confidentiality, key exchange, authentication, and non-repudiation. But cryptography does not always provide safe communication. Consider an environment where the very use of encrypted messages causes suspicion. If a nefarious government or Internet service provider (ISP) is looking for encrypted messages, they can easily find them. Consider the

following text file; what else is it likely to be if not encrypted? Steganography is the science of hiding information. Whereas the goal of cryptography is to make data unreadable by a third party, the goal of steganography is to hide the data from a third party [6].

2. PREVIOUS WORKS

Following is the list of works has been done on hiding information or text steganography carried out.

2.1 Using in Markup Languages

Some markup language feature can be used to hide information [7]. For instance feature of HTML document is their tags case insensitivity. For example, the tag
 can be also used as
 and
. As a result one can do text steganography in HTML documents by changing the small or large case of letters in document tags. We have to extract information by comparing these words with words in normal case. Then by using appropriate function in both hidden information is extracted. However these methods are not for those markup languages that have case sensitive tags. WML is an example of this category. In some cases the positions of tags are also used for text steganography. For example <U> </U> or like this <U> </U>. We extract information by comparing the tags positions. In WML the second text steganography method can be employed.

2.2 Using Specific Characters in Words

In this analytical and time-consuming method, some specific characters from certain words are selected [8]. For example the first words of each paragraph are selected in a manner that by placing the first characters of the selected words side by side, as a result it forms secret or hidden information is extracted. This method requires strong mental power and it also requires special text because not all type of texts can be used in this method.

2.3 Line Shifting Strategy

In printed document the lines of the text are vertically shifted to some degrees [9,10]. For example, some lines are shifted 1/300 inch up or down in the text and information are hidden by creating a hidden unique shape of the text. However, in this method, the distances can be observed by using special instruments of distance assessment and necessary changes can be introduced to destroy the hidden information. Also if the text is retyped or if character recognition programs (OCR) are used, the hidden information would get destroyed.

2.4 Word Shifting Strategy

Information is hidden in the text by shifting words horizontally and by changing distance between words, [9, 11]. This method is

acceptable for texts where the distance between words is varying. This method can be identified less, because change of distance between words to fill a line is quite common. But if somebody was aware of the algorithm of distances, he can compare the present text with the algorithm and extract the hidden information by using the difference. The text image can be also closely studied to identify the changed distances. Although this method is very time consuming, there is a high probability of finding information hidden in the text. The same as in the method described under 2-3, retyping of the text or using OCR programs destroys the hidden information.

2.5 Using Punctuation Signs

Proper placement of some punctuation signs such as full stop (.) and comma (,) appropriately can hide information in a text file [8]. This method requires identifying proper places for putting punctuation signs. The amount of information to hide in this method is trivial.

2.6 Using Synonyms

By using the synonym of words for certain words one can hide information in the text [10, 12]. A major advantage of this method is the protection of information in case of retyping or using OCR programs (contrary to methods listed under 2-3 and 2-4). However, this method may alter the meaning of the text.

2.7 Feature Coding Method

Some of the features of the text can be altered to hide some information [13]. For example, the end part of some characters such as h, d, b or so on, is elongated or shortened a little thereby hiding information in the text. In this method, a large volume of information can be hidden in the text without making the reader aware of the existence of such information in the text. By placing characters in a fixed shape, the information is lost. Retyping the text or using OCR program (as in methods 2-3 and 2-4) destroys the hidden information.

2.8 Using Abbreviation

Use of abbreviations can hide some information, though with very less capacity [8]. For example, only a few bits can be hidden in a file of several kilobytes.

2.9 White Space Strategy

We can add some extra white spaces in the text [8, 14] to hide some information. These white spaces can be placed at the end of each line, at the end of each paragraph or between the words. This method can be implemented on any arbitrary text and does not raise attention of the reader. However, the volume of information hidden under this method is very little. Also, some text editor programs automatically delete extra white spaces and thus destroy the hidden information.

2.10 Displacing Letter Points

Arabic, Persian and Urdu texts are used for this technique [15, 16, 20]. One of the characteristics of these languages is abundance of points in its letter. In Persian 18 letters out of 32 letters have points. From these 18, 3 letters have 2 points each, 5 letters have 3 points each and the other 10 letters have 1 point each. These 1 point letters are used to hide the information by shifting position of point a little bit vertically high with respect to the standard point position in the text.

2.11 Displacing the Diacritics

Urdu and Arabic texts can be used for this type of text steganography [17, 18]. One of the characteristics of Arabic languages is that the vowel symbols are located either on the upper or lower side of the character to be modified by the vowel. This characteristic is used to hide the information by shifting position of the vowel sign (diacritics) a little bit vertically high or low with respect to the standard diacritics position in the text.

2.12 Using Extension Letter

Text steganography is applied on Arabic text [19] for this algorithm. Arabic language has a special extension character, which can be arbitrarily inserted between characters for formatting purposes.

3. THE PROPOSED TECHNIQUE

Padding arbitrary number of zeros before a number or after the fractional part of a number does not alter the value of the number. So we can intentionally pad zeroes before a number to represent some information. A cricket scorecard contains huge number of numerical figures to represent the overall condition of the match. A typical score card contains the run scored by a batsman and the number of balls he spent to score that run. Recent score cards contain number of 4s and 6s scored by batsman also. In some cases a scorecard can contain number of minutes the batsman was present in the field to represent his run-scoring rate with respect to time. Besides this the performance of the bowling team is also recorded as the bowlers name, number of over bowled by a bowler, number of maiden over, runs spent in his over, number of wickets and economy rate. The batting team has 11 bats man though all of them may not required to bat. On the other hand the bowling team must use minimum of 5 bowlers. So there are six numbers per batsman and five numbers per bowler that can be optionally zero padded to represent some information. So if the batting team requires 9 batsman and the bowling team use 6 bowlers, there are 84 numerical figures in an innings in a scorecard. One day and T-20 matches have 2 innings per match and in test matches there are up to 4 innings. So in one day and T-20 matches score card of a match contains around 160 numerical figures. And for test matches the number should be double.

Now if we want to hide a '1' bit we can pad a meaningless zero before the number and if want to hide a '0' bit we may leave the number unpadding. By these 160 numbers can hide 160 bits, i.e. 20 bytes.

Algorithm-1: Hiding Secret Bits

Input: Cover scorecard, Secret bit stream

Output: Stego scorecard

Step-1: For each batsman of the batting team repeat steps 2 to 4.

Step-2: Consider the run, the minutes, the balls, the 4s, the 6s and the strike rate sequentially.

Step-3: Consider next six consecutive bits in the secret bit stream. Consider the first bit for the run, the second bit for the minutes and so on.

Step-4: If the secret bit is 1 add an extra zero before the corresponding numerical figure in the scorecard.

Step-5: For each bowler of the bowling team repeat steps 6 to 8.
 Step-6: Consider the over, the maidens, the run, the wicket, and the economy rate sequentially.
 Step-7: Consider next five consecutive bits in the secret bit stream. Consider the first bit for the over, the second bit for the maidens and so on.
 Step-8: If the secret bit is 1 add an extra zero before the corresponding figure in the scorecard.
Algorithm-2: Extracting Bits from the Received Score Card
 Input: Stego scorecard
 Output: Extracted secret bit stream
 Step-1: for each batsman of the batting team repeat steps 2 and 3.
 Step-2: Consider the run, the minutes, the balls, the 4s, the 6s and the strike rate sequentially.

Step-3: If there is an unnecessary 0 before a figure in the scorecard add a '1' bit in the output bit stream and if not add a '0' bit in the output bit stream.
 Step-4: for each bowler of the bowling team repeat steps 5 and 6.
 Step-5: Consider the over, the maidens, the run, the wicket, and the economy rate sequentially.
 Step-6: If there is an unnecessary 0 before a figure in the scorecard add a '1' bit in the output bit stream and if not add a '0' bit in the output bit stream.

Figure-1 represents a typical score card of an innings of a one day international and figure-2 represents the same score card after hiding the following bit stream in the score card.

Secret Bit Stream:
 1010010101101010101010111101000111001111000111100011
 11111111100001010101000001

| | | | | | | |
|--|-----|-----|----|----|----|----------------------|
| ODI no. 2839 | | | | | | |
| 2008/09 season | | | | | | |
| Australia in South Africa ODI Series – 3rd ODI | | | | | | |
| South Africa v Australia | | | | | | |
| Played at Newlands, Cape Town, on 9 April 2009 – day/night (50-over match) | | | | | | |
| Result South Africa won by 25 runs | | | | | | |
| South Africa innings (50 overs maximum) | R | M | B | 4s | 6s | SR |
| HH Gibbs c Hopes b Bracken | 26 | 61 | 47 | 3 | 0 | 55.31 |
| *GC Smith c Haddin b Johnson | 8 | 26 | 14 | 1 | 0 | 57.14 |
| JH Kallis c Hopes b Geeves | 70 | 115 | 86 | 5 | 0 | 81.39 |
| AB de Villiers c Hauritz b Johnson | 80 | 104 | 87 | 7 | 0 | 91.95 |
| JP Duminy c MEK Hussey b Johnson | 32 | 40 | 26 | 2 | 1 | 123.07 |
| JA Morkel b Johnson | 29 | 41 | 25 | 2 | 1 | 116.00 |
| MV Boucher not out | 28 | 26 | 15 | 1 | 2 | 186.66 |
| RE van der Merwe not out | 6 | 3 | 4 | 0 | 1 | 150.00 |
| Extras (lb 4, w 2, nb 4) | 10 | | | | | |
| Total (6 wickets; 50 overs; 211 mins) | 289 | | | | | (5.78 runs per over) |

Did not bat J Botha, DW Steyn, WD Parnell
 Fall of wickets 1-28 (Smith, 6.2 ov), 2-50 (Gibbs, 14.2 ov), 3-164 (Kallis, 35.6 ov), 4-198 (de Villiers, 40.4 ov), 5-230 (Duminy, 44.2 ov), 6-283 (Morkel, 49.2 ov)

| | | | | | | |
|------------|----|---|----|---|------|-----------|
| Bowling | O | M | R | W | Econ | |
| MG Johnson | 10 | 2 | 34 | 4 | 3.40 | |
| NW Bracken | 10 | 0 | 58 | 1 | 5.80 | (1w) |
| B Geeves | 10 | 0 | 67 | 1 | 6.70 | (4nb, 1w) |
| NM Hauritz | 8 | 0 | 37 | 0 | 4.62 | |
| JR Hopes | 8 | 0 | 64 | 0 | 8.00 | |
| DJ Hussey | 4 | 0 | 25 | 0 | 6.25 | |

Fig 1: Scorecard of an innings of a one-day international match

ODI no. 2839
2008/09 season
Australia in South Africa ODI Series - 3rd ODI
South Africa v Australia

Played at Newlands, Cape Town, on 9 April 2009 - day/night (50-over match)
Result South Africa won by 25 runs

| South Africa innings (50 overs maximum) | R | M | B | 4s | 6s | SR |
|---|-----|-----|----------------------|----|----|--------|
| HH Gibbs c Hopes b Bracken | 026 | 61 | 047 | 3 | 0 | 055.31 |
| *GC Smith c Haddin b Johnson | 8 | 026 | 14 | 01 | 00 | 57.14 |
| JH Kallis c Hopes b Geeves | 070 | 115 | 086 | 5 | 00 | 81.39 |
| AB de Villiers c Hauritz b Johnson | 080 | 104 | 087 | 7 | 00 | 091.95 |
| JP Duminy c MEK Hussey b Johnson | 032 | 040 | 26 | 02 | 1 | 123.07 |
| JA Morkel b Johnson | 29 | 041 | 025 | 02 | 1 | 116.00 |
| MV Boucher not out | 028 | 026 | 015 | 01 | 2 | 186.66 |
| RE van der Merwe not out | 6 | 03 | 04 | 00 | 01 | 150.00 |
| Extras (1b 4, w 2, nb 4) | 10 | | | | | |
| Total (6 wickets; 50 overs; 211 mins) | 289 | | (5.78 runs per over) | | | |

Did not bat J Botha, DW Steyn, WD Parnell

Fall of wickets 1-28 (Smith, 6.2 ov), 2-50 (Gibbs, 14.2 ov), 3-164 (Kallis, 35.6 ov), 4-198 (de Villiers, 40.4 ov), 5-230 (Duminy, 44.2 ov), 6-283 (Morkel, 49.2 ov)

| Bowling | O | M | R | W | Econ |
|------------|-----|----|-----|----|----------------|
| MG Johnson | 10 | 2 | 034 | 04 | 03.40 |
| NW Bracken | 010 | 00 | 058 | 01 | 05.80 (1w) |
| B Geeves | 010 | 00 | 067 | 1 | 6.70 (4nb, 1w) |
| NM Hauritz | 8 | 0 | 037 | 0 | 04.62 |
| JR Hopes | 8 | 00 | 64 | 00 | 8.00 |
| DJ Hussey | 4 | 0 | 25 | 0 | 06.25 |

Fig 2: The scorecard of figure-1 after hiding secret information in it

One important thing we must remember that the capacity of a test match score card is roughly 1.5 times of a one day or a T-20 as there are at least three innings in general.

4. CONCLUSION

For text steganography various methods have been proposed. We represent a novel technique of hiding information in a cricket scorecard. Scorecards are very much common for cricket personals as well as fans. The capacity of this method is also in fair margin. We hide roughly 10 bytes of data in the sample innings scorecard of figure-1. So, our proposed method can be a good choice for text steganography.

5. REFERENCES

- | | |
|---|--|
| <p>[1] Cricket Game, http://cricket.deepthi.com</p> <p>[2] The history of cricket, http://www.essortment.com</p> <p>[3] Cricket Archive: full list of ICC members, http://www.cricketarchive.co.uk/Archive/Countries/index.html</p> <p>[4] Cricket Score Card is the Summary of the Match, http://www.articlesbase.com</p> <p>[5] Steganography, http://en.wikipedia.org</p> <p>[6] Steganography: Hiding Data Within Data, http://sover.net</p> | <p>[7] K. Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", Purdue University, CERIAS Tech. Report 2004-13.</p> <p>[8] T. Moerland, "Steganography and Steganalysis", May 15, 2003, www.liacs.nl/home/tmoerlan/privtech.pdf</p> <p>[9] S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting", <i>Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95)</i>, 2-6 April 1995, vol.2, pp. 853 - 860.</p> <p>[10] A.M. Alattar and O.M. Alattar, "Watermarking electronic text documents containing justified paragraphs and irregular line spacing", <i>Proceedings of SPIE -- Volume 5306, Security, Steganography, and Watermarking of Multimedia Contents VI</i>, June 2004, pp. 685-695.</p> <p>[11] Y. Kim, K. Moon, and I. Oh, "A Text Watermarking Algorithm based on Word Classification and Inter word Space Statistics", <i>Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR '03)</i>, 2003, pp. 775-779</p> <p>[12] M. Niimi, S. Minewaki, H. Noda, and E. Kawaguchi, "A Framework of Text-based Steganography Using</p> |
|---|--|

- SD Form Semantics Model", *Pacific Rim Workshop on Digital steganography 2003*, Kyushu Institute of Technology, Kitakyushu, Japan, July 3-4, 2003.
- [13] K. Rabah, "Steganography-The Art of Hiding Data", *Information Technology Journal*, vol. 3, Issue 3, pp.245-269, 2004.
- [14] D. Huang, and H. Yan, "Inter word Distance Changes Represented by Sine Waves for Watermarking Text Images", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, December 2001, pp. 1237-1245
- [15] M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A New Approach to Persian/Arabic Text Steganography", *Proceedings of 5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS*, June 2006.
- [16] M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A Robust Page Segmentation Method for Persian/Arabic Document", *WSEAS Transactions on Computers*, vol. 4, Issue 11, Nov. 2005, pp. 1692-1698.
- [17] Mohammed Aabed, Sameh Awaideh, Abdul-Rahman Elshafei, and Adnan Gutub, Arabic Diacritics Based Steganography", *IEEE International Conference on Signal Processing and Communications (ICSPC 2007)*, Pages: 756-759, Dubai, UAE, 24-27 November 2007
- [18] Adnan Gutub, Yousef Elarian, Sameh Awaideh, and Aleem Alvi, "Arabic Text Steganography Using Multiple Diacritics", *WoSPA 2008 – 5th IEEE International Workshop on Signal Processing and its Applications*, University of Sharjah, Sharjah, U.A.E. 18 – 20 MARCH 2008.
- [19] Adnan Abdul-Aziz Gutub, and Manal Mohammad Fattani, "A Novel Arabic Text Steganography Method Using Letter Points and Extensions", *PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY VOLUME 21 MAY 2007 ISSN 1307-6884 pp 28-31*
- [20] Jibrán Ahmed Memon, Kamran Khawaja, Hameedullah Kazi, "EVALUATION OF STEGANOGRAPHY FOR URDU /ARABIC TEXT", *Journal of Theoretical and Applied Information Technology*, pp 232-237.