SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks. 2009; **2**:555–566
Published online 6 March 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sec.99

Chestega: chess steganography methodology

Abdelrahman Desoky and Mohamed Younis*,†

Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD, U.S.A.

Summary

Steganography is the science and art of avoiding the arousal of suspicion in covert communications. This paper presents *Che*ss *Stega*nography (Chestega), a novel methodology that exploits popular games like chess. Chestega conceals messages in chess related covers such as training documents, game analysis, news articles, etc. Unlike contemporary approaches, Chestega does not exploit noise to embed a message nor produce a detectable noise. Instead, authenticated data can be employed in the cover which makes it resilient to comparison attacks. Chestega is also a public approach that neither relies on the secrecy of its technique, nor need to employ a stega-key. The paper demonstrates the feasibility of employing authenticated Chess Cover that is generated by Chessmaster 8000. Chestega is further validated through steganalysis. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: steganography; information hiding; covert communication

1. Introduction

Information hiding and covert transmission of messages have been practiced since the time of ancient civilizations [1,2]. The techniques pursued in such area, which is often referred to as steganography, have grown increasingly sophisticated over the years, especially with the availability of digital media. The overall goal is not just to prevent an adversary from detecting and disrupting a message but to avoid the arousal of suspicion in covert communications. Steganography involves three steps; encoding a message, embedding the encoded message in a suitable cover and then delivering the cover to the recipient. Traditionally, the second step has been the differentiator among the various steganography techniques. Contemporary approaches are often categorized based on the steganographic cover type such as text, image, audio, or graph [3].

Most of the steganographical schemes found in the literature camouflage a message as noise in a cover that is assumed to look innocent. For example, the encoded message can be embedded as alteration of a digital image or an audio file without noticeable degradation [4]. Another example is concealing a message in a text-cover by altering the format and style of an existence text [5,6]. However, such alteration of authenticated covers can easily raise suspicion and the message is

*Correspondence to: Mohamed Younis, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD, U.S.A.
†E-mail: younis@umbc.edu

detectable regardless of whether or not a plaintext is reveled. Some other different approaches pursued in the linguistic path, such as null cipher [7], mimic functions [8,9], NICETEXT and SCRAMBLE [10], and translation-based [11–13]. The hidden message becomes to a great extent a foreign body in the cover and thus makes those schemes vulnerable to detection. In addition, contemporary steganography schemes rely on private or restricted access to the original unaltered cover in order to avoid the potential of comparison attacks, which is considered a major threat to the covert communication. Basically, an adversary can detect the presence of a hidden message when comparing to the original cover and finding out that some alterations have been made.

Unlike prior work, this paper promotes *Ches*s *Stega*nography (Chestega), a novel methodology that exploits popular games like chess. Chestega conceals messages in chess related covers such as training documents, game analysis, news articles, etc. It conceals a message in three steps. First, it determines the encoding parameters, meaning what aspect of the game would be used to hold steganographic code. Example encoding parameters include the chessboard, pieces, moves, etc. This is an initialization step that the communicating parties should agree on. Second, Chestega expresses the message using the selected encoding parameters. Third, it camouflages the message in a chess-cover. The main advantages of Chestega over other approaches are as follows. Since the message is not concealed as a noise in the chess-cover, the hidden message is anti-distortion. Chestega is a public approach that neither relies on the secrecy of its technique nor requires a stega-key. Chestega is also resilient to comparison attacks by employing untraceable or authenticated data. Finally, the popularity of Chess provides sufficient justification for the transmission of chess-covers among communication parties and would help in camouflages the communications process. It is worth noting that the presented approach is equally applicable to other games such as checkers, crosswords, domino, etc.

The remainder of this paper is organized as follows. The next section discusses prior work on steganography and highlights how Chestega is distinct. Section 3 describes Chestega in detail and highlights its advantages. The implementation of Chestega is discussed in Section 4. Section 5 presents the steganalysis validation. Finally, the paper is concluded in Section 6 with a summary and an outline of future research directions.

2. Related Work

Steganography techniques opt to hide a message in a cover so that no one would suspect any covert communication However, once suspicion is drawn, the approach is said to fail. In general, an adversary's attack is successful once he is able to detect or distort a noise, regardless of whether or not a plaintext is revealed. Steganography approaches are often categorized based on the cover type such as text, image, audio, or graph [3].

Textual steganography can be classified as textual format manipulation (TFM) and textual fabrication (TF) [4]. TFM modifies an original text by employing spaces, misspellings, fonts, font size, font style, colors, and non-color (as invisible ink) to embed an encoded message. However, comparing the original text with the modified text will trigger suspicion and enable adversary to pin down where the message is hidden in the text [5,6]. In addition, TFM can be distorted, discerned by human eye, or detected by a computer. On the other hand, textual fabrication techniques generate an entire text-cover for hiding a message rather than manipulating an existing text. Examples of these approaches are null cipher [7], mimic functions [8,9], NICETEXT and SCRAMBLE [10], and translation-based [11–13]. However, the text-cover that is generated by these approaches often has numerous linguistic flaws that can raise suspicion. In addition, revealing the hidden message may be feasible [4,11].

On the other hand, image steganography is based on manipulating digital images to conceal a message. Such manipulation often renders the message as noise. In general, image steganography suffers from several issues such as the potential of distortion, the significant size limitation of the messages that can be embedded, and the increased vulnerability to detection through digital image processing techniques [14]. Audio-covers have also been pursued. Example of audio steganography techniques include LSB [15,16], spread spectrum coding [17,18], phase coding [17,19], and echo hiding [20]. In general, these techniques are too complex, and like their image-based counterpart, are still subject to distortion and vulnerable to detection [5,17].

Recently, a Graph Steganography (Graphstega) methodology has been proposed [3]. Unlike all other schemes, the message is naturally embedded in the cover by simply generating the cover based on the message. Graphstega camouflages a message as data points in a graph and thus the message would not

DOI: 10.1002/sec

be detectable as noise. The approach is shown to be resilient to a wide range of attacks, including a comparison attack by untraceable or authenticated data. Graphstega represents a new paradigm in steganography research in which the message is hidden in the cover as data rather than noise. Chestega follows this new paradigm by exploiting popular domains that are explored by many individuals across the world, like Chess fans, in concealing messages.

Exploiting the use of games to hide messages has been considered by Hernandez-Castroa et al. [21]. Basically, game scripts are suggested as covers where messages are hidden in moves or comments. However, this work provides limited options for concealing messages and has considered neither the coding implication on the sequence of moves nor the use of authenticated data in terms of documented tournaments and games among known players. Only fabricated games are pursued as covers. In addition, the proposed approach is vulnerable to contrast attacks where the sequence of moves does not logically match the game flow, mostly caused by the message concealment process. Moreover, it is vulnerable to comparison and traffic attacks. Chestega overcomes these shortcomings by generating a noiseless cover through the use of an unaltered authenticated data or unsuspicious fabricated data (e.g., teaching chess).

3. Chestega Methodology

Chess is a very popular game that appeals to people of all ages world-wide. In addition to international competitions, there are numerous local, regional, and national chess tournaments almost everywhere. Chess games are reported and rated by an international chess federation and/or some local chapters. To standardize the storage and reporting chess games, they are represented using specific keywords and syntax, called the portable game notation (PGN) [22]. PGN is not the only chess notation that exists. However, in this paper PGN is used since it is the official and most popular chess notation.

Chestega averts suspicion in covert communication by concealing a message using chess data. Chess data in this context includes chessboard positions, pieces and their color, moves, tournament name, place and results, players, etc. These data can be exploited to conceal a message within a script of moves in a game, teaching sessions, game analysis, etc. The chess data can be authenticated, for example, citing actual games or tournaments, players, etc., or fabricated as part of teaching material or a made-up scenario. The fabricated data does not always have to look normal and legitimate, for example, reflecting only legal moves, instead it may be in a form of a natural noise, for example, an illegal move or a position, the use of illegal moves is often pursued by the chess community for teaching purposes.

The Chestega Cover can be in a form of a graph, for example, game statistics, image, for example, snap shot of the chessboard during a game, text, for example, teaching of tactics, audio, for example, game analysis, or a combination of these types. The tremendous volume of chess data in electronic and non-electronic format makes an adversary's job extremely difficult and renders Chestega an effective steganography methodology.

Chestega is composed of three modules whose ultimate goal is to define a configuration for the communicating parties to use. The first module mainly determines Chestega encoding parameters, meaning what aspect of the game would be used to hold steganographic code. These parameters are then used by the second and third modules to define a message encoder and a camouflage scheme, respectively. Figure 1 shows the interaction among the Chestega modules and how the generated configuration is used by the sender and recipient. The following subsections explain the Chestega modules in detail. Section 3.4 elaborates on how the communicating parties employ the Chestega configuration for covert message exchange.

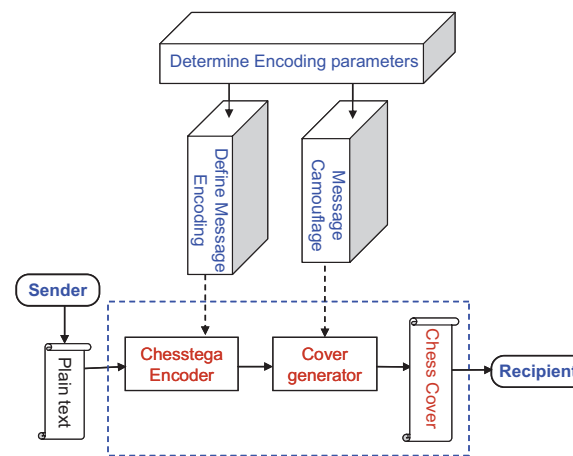

Fig. 1. An illustration of the interaction of the various Chestega modules and how the outputs of the individual modules are used for covert communication between two parties.

DOI: 10.1002/sec

3.1. Determining Encoding Parameters

There are numerous parameters in chess that can be exploited as a vehicle for concealing a message. A parameter in this context means some aspect about the game that is referred to with multiple values throughout a game. Examples of these parameters include the squares on the chessboard, pieces, moves, players, thinking time, etc. The encoding module of Chestega exploits these choices and determines the parameter(s) that will be employed for concealing a message. The selection criteria will be mostly driven by the message size, the style of the cover and the availability of authenticated data that would match the encoded message.

While the use of the chessboard is the most intuitive choice for message encoding, it puts a cap on the size of the message. For example, assume that the squares in the chessboard are serially numbered and the presence of a pawn indicates that the square conceal part of the message. Obviously, the maximum size of a message would be constrained by the number of pawns, that is, 16 and the number of squares in the board, that is, 64, and the message would not thus exceed 96 bits (16 pawn $\times$ 6 bits of 64 squares). On the other hand, the use of moves as an encoding vehicle would allow the concealment of long messages. Simply the message can be represented as a sequence of moves in a game. Given that most chess games, especially at the master level, last for an extended duration and involve many moves, it is feasible to conceal long messages subject to availability of authenticated data as we explain below. While concealing long messages is a challenge for all known steganography approaches, Chestega can hide relatively long messages as demonstrated in Section 4.

Although one would argue that the selection of a suitable cover is affected by the picked coding parameter and not the other way around, an imposed style for the cover would constrain the encoding of the message. For example, if chess teaching sessions are not an acceptable form of cover, for example, the communicating parties are known to play well, allowing illegal moves will raise suspicion and would not be acceptable. Thus, using moves as an encoding parameter will be restricted to only legal moves. The same applies to chessboard and pieces encoding, for example, having the two bishops tied to the same square color early in the game. In addition, the availability of authenticated data plays a major role in determining the encoding parameters. In general, being able to map each steganographic code to some realistic data that is publicly accessible would be a major advantage to any steganography approach. For example, concealing a message within a game analysis would need that every reference to a move in the game must match what happened in reality since the game is documented and anyone can examine the authenticity of the used data. While the use of fabricated chess data, for example, reporting on a fictitious game, is always possible, it requires more care for justifying the association between the communicating parties as discussed in Section 3.4.

3.2. Defining Message Encoder

Chestega creates an encoded representation of the plaintext message and then camouflages it in a chess cover. The obvious constraint that Chestega imposes on the message encoder is to generate steganographic code that can be embedded in the cover. For example, when a chessboard is employed as an encoding parameter, the message encoder should not refer to a non-existing square. In addition, the variations in the data values may have to be considered. For example, when moves are pursued as an encoding vehicle the target square has to abide with the chess rules for a piece move. Encoding the message as numerical values is not the only option. As discussed in the previous section, the use of characters can be a feasible choice. Messages can be concealed using the name of players, tournament locations, opening techniques, etc. Character based encoding will be mostly challenged by finding appropriate authenticated data that can be referenced in the same chess cover.

Given the availability of numerous encoding techniques in the literature that fit [23,24], the balance of this section will focus on an example that illustrates how to meet the message encoding constraint. This example will be used in Section 4 to demonstrate the applicability of Chestega. A message is encoded as follows. First, the message is converted to a binary string which is then partitioned into groups of a particular number of bits that is agreed upon among communicating parties and such that all constraints on the range of steganographic code values can take are met. Finally, a decimal representation is generated for the individual groups. For example, a 7-bit based grouping, the value of each group would range between 0 and 127 (0000000 to 1111111 in binary). The following describes the encoding of a sample message:

- The plaintext of the message is "*he doesn't love you*."
- The concatenated binary string of the ASCII representation of this message is: "01101000011001

DOI: 10.1002/sec

01001000000110010001101111011001010111001101101110100100100111010000100000011011000101011110111011011001100100 100 10000001111001011011110111010100100000."

- Slicing this string (from the previous step) into 7 bits each, as set and agreed upon by communicating parties, will result in: 0110100 0011001 0100100 0000110 0100011 0111101 1001010 1110011 0110111 0100100 1001110 1000010 0000011 0110001 1011110 1110110 0110010 1001000 0001111 0010110 1111011 1010100 100000.
- Converting the individual slices (from the previous step) into decimals results in: "52 25 36 6 35 61 74 115 55 36 78 66 3 49 94 118 50 72 15 22 123 84 32."

It is worth noting that the range of the resulting decimal values can be easily narrowed or widened by partitioning the binary string into groups of less or more than 7 bits. Again, this encoding scheme is just for illustration and many alternative schemes can be employed.

3.3. Message Camouflaging Scheme

As mentioned before, Chestega camouflages a message by concealing the encoded message as data in a chess-cover. In other words, the encoded message will be the data that is referenced in the cover. The message may constitute a subset or a full set of the data in the cover. The latter case makes the message's decoding a straightforward exercise; basically by applying the reverse process using all data items. However, the use of a partial data set would require a pre-agreement among the communicating parties on how to pick the data items that are relevant to the decoding of the message. For example, the sender may agree with the recipient on considering only every other data item according to the order of appearance in the cover. The use of a subset of the chess data can make Chestega more resilient to attacks. Basically, an adversary would have to try all possible subsets of the data in order to identify the relevant items assuming that he will suspect the presence of a hidden message in a particular set of chess-related documents and attempt to guess the encoding scheme. The same applies when multiple encoding parameters are pursued. It would be very difficult for an adversary to identify what parameter to investigate and what not; especially when an explicit interaction between the sender and the receiver does not take place, for example, by posting the cover on a publicly accessible website. Note that in this paper a plaintext is concealed for simplicity. In realty the ciphertext is concealed rather than plaintext, which is common practice in steganography.

Chestega supports multiple cover styles and types. A style in this context means how and why chess data is presented. Example styles include teaching documents, puzzles, game reports, news articles, etc. A cover can be focused on a single game or discuss multiple games. While it is a common practice for a chess player to read and analyze unrelated games, for example, checking various posting on the Internet, it is feasible nonetheless to relate various chess games appearing in a cover. Example themes for relating a collection of games in a cover include:

1. The opening strategies of a chess game.
2. Similar positions of some pieces or applying similar concepts such as scarifying peace(s), controlling open files, short castle, opposite castle, etc.
3. The names of chess player, tournaments, events, etc.
4. The date and place of games, for example, country, city, etc.
5. Political or rivalry aspects of the played games, for example, US versus the former Soviet Union.

It is worth noting that identifying the theme and generating some text to legitimize the appearance of unrelated games in a chess cover can be automated through the use of a natural language generation (NLG) systems [25]. Many of the computer based chess tools such as Chessmaster employs NLG systems to generate analysis and comments.

Meanwhile, the type of a cover indicates its format. The most intuitive cover type is the use of images when using the chessboard for encoding. Basically, the relevant pieces will be placed in the right squares and a capture of the chessboard will then constitute the cover. Alternatively, text covers can be employed in the form of detailed description, game analysis, teaching sessions, etc. The use of the PNG notation would be appropriate in that case. In addition, a graph-cover [3] can be employed when game or tournament statistics are used to conceal a message. Moreover, an audio cover may be pursued in the form of expert commentary, live update of a game, etc. The encoding parameters, picked by module 1, guide the process of selecting the most suitable cover type and style. As mentioned earlier, long messages would make some encoding parameters such as moves an appropriate choice for concealing a message and would also make detailed description and analysis of games a favorable cover. Also, an encoding that causes illegal moves

DOI: 10.1002/sec

would mandate the use of educational chess documents as cover. It is worth noting that multiple cover types may be involved. For example, the game analysis can include a number of images of the chessboard that summaries the status of the game at different instants.

The selected cover style also has to suit the desired frequency of communications. Through the use of some styles it may be legitimate to generate a new cover every day or so. For example, it is customary for a chess website to report on recent games on daily basis, or even more frequently. On the other hand, some covers may not justify more than one message per month, season or even a year. For example, the statistics of chess activities in a local region are not something that gets reported very often. Finally, some covers enable broadcasts or non-private announcements to a set of interested parties and would thus make the association between the sender and the recipient unsuspected. Posting an opinion about a chess game on a website of chess fans is a perfect example in this category. Finally, the cover style may depend on whether the use of authenticated data is required. Chess-covers that involve verbose documents make the use of authenticated data less favored since many constraints may be imposed to ensure consistency. For example, using moves of a publicly watched game to conceal a message requires the moves and possibly their order to match what happened in the game. That is obviously harder than reporting the moves of a private or fictitious game.

3.4. Chestega Configuration

A sender and a recipient who communicate covertly using Chestega must agree on the following, which constitutes the Chestega configuration:

(a) The particular specifications of message encoding/decoding scheme including the parameters that are employed for concealing the message.
(b) The style and type of the chess-cover so that the recipient would know what to decode.
(c) How to establish a covert channel for them to communicating, that is, delivering a Chestega Cover to the recipient.

The first and second items are addressed by the three modules discussed in the previous subsections. The third item, which is referred to as the Chestega communication protocol, mainly defines how the cover will be delivered to the recipient without raising suspicion. Contemporary steganography approaches in the literature have focused on how to conceal a message and not on how to camouflage its transmittal. It is however argued that covert transmittal of the steganographic cover is very crucial to the success of steganography. At the core of the cover transmittal issue is how to prevent the association between the sender and recipient from drawing suspicion. For example, exchanging e-mail messages would automatically imply a relationship between the communicating parties. Similarly, downloading files from a website indicates an interest in the accessed material. With advances in monitoring tools for network and Internet traffic, profiles of user's access pattern can be easily established. An adversary most probably will suspect the presence of a hidden message, even if the content does not look suspicious, because of the observed traffic pattern and the lack of a justification for the interest in the contents of such message traffic. Therefore, it is very important to rationalize the receiving of the steganographic cover in order to avoid attracting any attention that may trigger an attack

Chestega enables an effective solution to the issue of cover's transmittal to recipients. The use of chess allows a legitimate association among the communicating parties and would thus make sharing a chess-cover an ordinary practice. Chess is a very popular game and has many fans and players all over the world. Such popularity makes the transmission of the chess-covers via e-mail, posting them on web pages or even downloading chess-related articles is a natural matter. In addition, casual message exchange that includes no hidden messages can be pursued in order to avoid the formation of a communications pattern that may draw attention. Explicit message transmission is not the only means for sharing the cover. Web posting in public discussion forums and mailing magazines via postal services are sample of other means that can be pursued. In summary, the way of delivering the hidden message can raise suspicion even when using a resilient steganographic technique. Chestega averts the suspicion that may arise during covert communications not only by camouflaging a message but also its transmittal.

4. Chestega Implementation

The following scenario illustrates how Chestega can be used. Bob and Alice are undercover agents and they communicate covertly using chess. They agree on a date and time using a specific online chess

DOI: 10.1002/sec

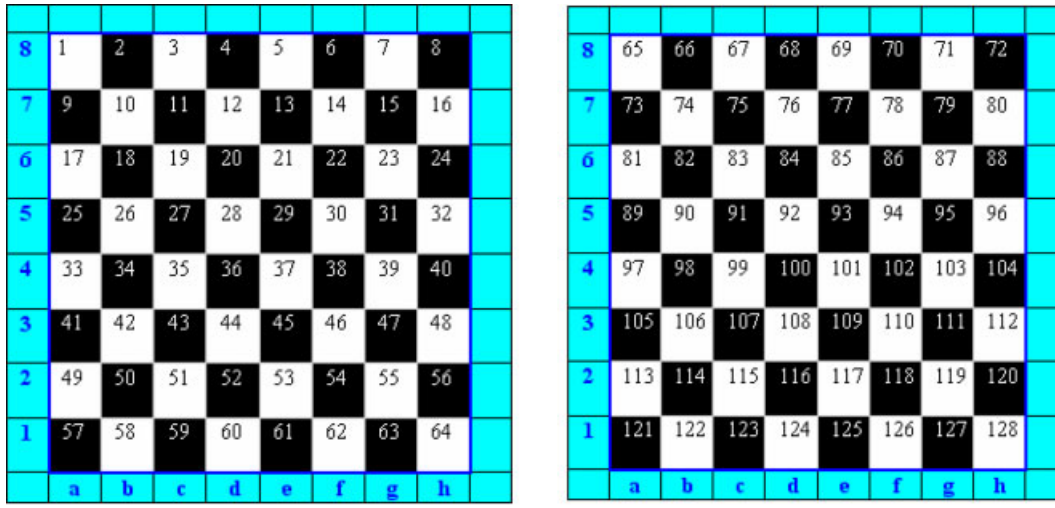

Fig. 2. The steganographic code when encoding the While side (left) and the Black side (right) of a chessboard.

provider to play a game, examine games, or teach chess for the purpose of communicating covertly. This section demonstrates the applicability of Chestega and validates the feasibility of the concealment process through two examples. In the first example only the chessboard is used to encode a message. The second example employs the first letter of a real chess player as an encoding scheme. These examples are also indended to show how one can define his Chestega configuration. It should be noted this section shows just few examples of possible implementations. In addition, the goal of this section is to show Chestega capabilities in concealing data rather than making the adversary's task hard to decode a message. Using cryptosystem to protect a message is straightforward and is not the focus of this paper.

4.1. Chessboard-Based Example

Intuitively, the chessboard is the most basic encoding venue. The encoding scheme of this example is similar to the one discussed in Section 3. A chessboard is an 8×8 square, which renders 64 squares. Since the chess pieces are two colors, white and black, the encoding of the chessboard will be the double of 64 squares which is 128 squares. The squares are encoded from 0 (in binary 0000000) to 127 (in binary 1111111). Employing an index that starts at 1 referring to 0 in decimal (in binary 0000000) up to 128 referring to 127 in decimal (in binary 1111111), as shown in Figure 2. Each move will be thus represented by 7 binary digits, referring to the index of the target square by a piece of a particular color. The message to be concealed is "*he doesn't love you*," which will be encoded as detailed in this section and also explained in Section 3.2 to "*52 25 36 6 35 61 74 115 55 36 78 66 3 49 94 118 50 72 15 22 123 84 32*." Table 1 shows each steganographic code and its corresponding move in the PNG notation. The PNG moves are then used to conceal the message in a chess-cover (Figure 3). In this cover, particular chess moves from unaltered authenticated games are used to camouflage the message in a chess training lesson.

Table 1. Encoded message using steganographic code in Figure 2.

Binary	Decimal	Color	PGN Square
0110100	52	W	d2
0011001	25	W	a5
0100100	36	W	d4
0000110	6	W	f6
0100011	35	W	c4
0111101	61	W	e1
1001010	74	B	b7
1110011	115	B	c2
0110111	55	W	g2
0100100	36	W	d4
1001110	78	B	f7
1000010	66	B	b8
0000011	3	W	c8
0110001	49	W	a2
1011110	94	B	f5
1110110	118	B	f2
0110010	50	W	b2
1001000	72	B	h8
0001111	15	W	g7
0010110	22	W	f6
1111011	123	B	c1
1010100	84	B	d6
100000	32	W	h5

DOI: 10.1002/sec

This lesson is about trading off piece(s) in order to gain a superior position. The following games demonstrate that having less material and good position can lead for winning.

Anderssen defeated Dufresne by sacrificing a piece to open the central files against the uncastled Black King, and despite his seemingly adequate development and ;counterattacking chances, Black comes out a tempo short in one of the ;finest combinations on record, justly known as the "Evergreen Game."

1. e4	e5
2. Nf3	Nc6
3. Bc4	Bc5
4. b4	Bxb4
5. c3	Ba5
6. d4	exd4
7. O-O	d3
8. Qb3	Qf6
9. e5	Qg6
10. Re1	Nge7
11. Ba3	b5
12. Qxb5	Rb8
13. Qa4	Bb6

The Chessmaster recommends: Knight at b1 to d2.
Analysis: You move your knight at b1 to d2, which blocks Black's pawn at d3. Black answers with a castle. You move your knight to e4, which threatens Black's pawn at d3. Black responds with the pawn to d5, which disengages the pin on Black's pawn at f7 and forks your bishop at c4 and your knight at e4. Your pawn captures pawn en passant, which pins Black's pawn at f7, protects your bishop at c4 and your knight at e4, and attacks Black's knight at e7.

Black counters with pawn takes pawn, which removes the threat on Black's knight at e7 and isolates your pawn at c3. Your bishop at a3 takes pawn, which pins Black's knight at e7, attacks Black's rook at b8, and creates a passed pawn on c3. Black responds with the bishop to h3, which threatens checkmate (queen takes pawn), pins your pawn at g2 with a partial pin, and blocks your pawn at h2. You move your knight at f3 to g5, which frees your pawn at g2 from the pin. As a result of this line of play, you win two pawns for a pawn. Additionally, your mobility is greatly increased. Also, Black's pawn structure is somewhat weakened. Finally, the pressure on Black's King is slightly increased.

14. Nbd2	Bb7
15. Ne4	Qf5
16. Bxd3	Qh5
17. Nf6+	gxf6
18. exf6	Rg8
19. Rad1	Qxf3
20. Rxe7+	Nxe7
21. Qxd7+	Kxd7
22. Bf5+	Ke8
23. Bd7+	Kd8
24. Bxe7#	1-0

==========

This brilliancy-prize game by Henry Edward Bird, one of England's premier players for half a century, features a speculative queen sacrifice with the unusual combination of two rooks and knights against queen, rook and ;knight. A delight!

1. e4	e6
2. d4	d5
3. Nc3	Nf6
4. exd5	exd5
5. Nf3	Bd6
6. Bd3	O-O
7. O-O	h6
8. Re1	Nc6
9. Nb5	Bb4
10. c3	Ba5
11. Na3	Bg4
12. Nc2	Qd7
13. b4	Bb6
14. h3	Bh5
15. Ne3	Rfe8
16. b5	Ne7
17. g4	Bg6
18. Ne5	Qc8
19. a4	c6
20. bxc6	bxc6
21. Ba3	Ne4
22. Qc2	Ng5
23. Bxe7	Rxe7
24. Bxg6	fxg6
25. Qxg6	Nxh3+
26. Kh2	Nf4
27. Qf5	Ne6
28. Ng2	Qc7

The Chessmaster recommends: Queen to d3.
Analysis: You move your queen to d3. Black counters by moving the rook to f8, which attacks your pawn at f2. You move your king to g1, which frees your knight at e5 from the pin and protects your pawn at f2. Black responds by moving knight to c5, which attacks your queen. You move your queen to e2, which moves it to safety. Black replies by moving the rook at f8 to e8. You move your queen to a2, which frees your knight at e5 from the pin. Black responds with rook captures knight. Your pawn captures rook, which pins Black's pawn at d5 and creates a passed pawn on e5. Black answers with rook captures pawn. As a result of this sequence of moves, you win a rook for a knight and a pawn.

29. a5	Bxa5
30. Rxa5	Rf8
31. Ra6	Rxf5

...

==========

.....

Fig. 3. The chess cover that conceal "*he doesn't love you*" using a chessboard-based encoding.

A collection of games are included, each starts with a move that correspond to a steganographic code in the encoded message. The theme of the cover is that sacrificing a piece in chess may be the gate to winning the game. The order of the games corresponds to the moves in Table 1. The selection of games was done by querying the Chessmaster database. There are multiple databases for chess games and distinct moves, which enable the automation of identifying the contents of a chess-cover. The text in the cover, other than the first paragraph, is auto-generated using Chessmaster. Given the size of the full cover, only the first two games are included. It is worth noting that in practice links to the individual games can be used in order to avoid lengthy text and make it easy to browse the contents.

DOI: 10.1002/sec

Table 2. The letter to binary mapping for the encoding scheme.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Binary	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
Letters	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

4.2. Non-Game Example

This example demonstrates how a message can be concealed without using contemporary chess parameters such as pieces, chessboard and moves. The idea is to use data that is not related to how the game is played, for example, components, rules, etc., but rather related to the players, tournaments, place, date, etc. That is why the example is called "non-game." Again the message "*he doesn't love you*" is to be concealed. The bit string for this message shown in Section 3.2 is again used here. However, the bit string is sliced into 4-bit sets. The encoding scheme for this example is to pick any 16 letters in the English alphabet and map each to a distinct combination of 4 bits. Table 2 shows the mapping used in this section. It is worth noting that other mappings will work as well. In fact unordered mapping may even be better from a security point of view. In addition, it is important to mention that slicing the message into 5 bits or more is still feasible. However, some combinations would then need to be mapped to a sequence of two letters and would slightly restrict the cover generation as explained.

Now, the corresponding letter for every 4-bit slice of the message's bit string is determined and a name of a chess player that starts with the same letter is identified. Table 3 shows the results of this step. Again multiple databases of chess players do exist with large number of names on record. The names listed in Table 3 are found by searching the Chessmaster database. If 5-bit slices of the bit string are used, names that match 2 consecutive letters will be needed for some combinations and the search becomes somewhat constrained.

The chess cover, Figure 4, is simply generated by looking for games for the identified players. Again Chessmaster is used to collect the description of the individual games. In other words, unaltered authenticated game information is used for camouflaging the message. It should be noted that some details are omitted from the games in Figure 4 in order to simplify the presentation and to highlight the key features. The cover basically lists the games such that the name of the player with white pieces matches that in Table 3. The cover includes many other names that are not related to the messages. The appearance of other player names in the cover creates a huge fugue that confuses the adversary and convinces him that nothing is hidden.

Table 3. Encoded message using player names.

Binary	*Decimal*	Letter	Player name
0110	6	G	Grunfeld, E.
1000	8	I	Ivanchuk, V.
0110	6	G	Grunfeld, E
0101	5	F	Hoffmann, F. A.
0010	2	C	Chigorin, M.
0000	0	A	Anderssen, A.
0110	6	G	Geller, E.
0100	4	E	Edinburgh
0110	6	G	Gheorghiu, F.
1111	15	P	Pierre de Saint-Amant
0110	6	G	Gligoric, S.
0101	5	F	Frederic Lazard
0111	7	H	Henry Bird
0011	3	D	David Bronstein
0110	6	G	Glucksberg
1110	14	O	Ossip Bernstein
1001	9	J	Johannes Zukertort
0010	2	C	Chigorin, M.
0111	7	H	Hebden, M.
0100	4	E	Emanuel Lasker
0010	2	C	Captain Smith
0000	0	A	Anderssen, A
0110	6	G	Gaspariantz
1100	12	M	MacDonnell, A.
0110	6	G	Gligoric, S.
1111	15	P	Paulsen, L.
0111	7	H	Hennings, A.
0110	6	G	Geller, Y.
0110	6	G	Garcia, G.
0101	5	F	Frederic Lazard
0010	2	C	Carl Hamppe
0000	0	A	Alexander Halprin
0111	7	H	Hjartarson, J.
1001	9	J	Janowski, D.
0110	6	G	Gurgenidze, B.
1111	15	P	Pillsbury, H.
0111	7	H	Horowitz, A.
0101	5	F	Flohr, S.
0010	2	C	Capablanca, J.
0000	0	A	Alekhine, A.

5. Steganalysis Validation

The aim of this section is to show the resilience of Chestega to possible attacks. Again the success of

DOI: 10.1002/sec

The following is a list of good games for beginners to check and enrich their tactics.

[White "Grunfeld, E."]
[Black "Bogoljubow, E."]
[Result "1-0"]
Austria's Ernst Grunfeld was a great theoretician who possessed an encylopedic knowledge of the openings. A prominent star in the 1920s, he later became too content with colorless draws. Here is one his finest early efforts.
1. d4 Nf6
....
19. Rd8# 1-0
==========
[White "Ivanchuk, V."]
[Black "Angelov, K."]
[Result "1-0"]
After a weirdly violent opening exchange, Black finds his Knight difficult to extract.
1. e4 d5
....
29. Rc1 1-0
==========
[White "Grunfeld, E."]
[Black "Alekhine, A."]
[Result "0-1"]
Another superb Alekhine combination, as he outplays opening expert Grunfeld in the middle game.
1. d4 Nf6
....
34. Qf1 Bd4+
==========
[White "F. A. Hoffmann"]
[Black "A. D. Petrov"]
[Result "0-1"]
The main feature of this ancient game is the simultaneous assault by White on f7 and Black on f2, the weakest square on each side. Black's maneuvers culminate in a magnificent queen sacrifice and a relentless king hunt.
1. e4 e5
....
22. gxh4 Be3#
==========
.....
==========
[White "Alekhine, A."]
[Black "Yates, F."]
[Result "0-1"]
Frederick Yates was England's outstanding representative after Blackburne was no longer on the scene. Here is his most celebrated victory which earned him a brilliancy prize.
1. d4 Nf6
....
50. Kf3 Bg1+
==========

Fig. 4. A chess-cover that conceals a message using the first letter in the name of the player with the White pieces.

steganography is qualified with its ability for avoiding an adversary's suspicion of the presence of a hidden message. It is assumed that an adversary will perform all possible investigations. In addition, the adversary is also aware of Chestega, as a public methodology, but he does not know the detailed of Chestega configuration that the sender and recipient employ for their covert communication.

5.1. Traffic Analysis

One of the possible attacks an adversary may pursue is to analyze the communications traffic and the access patterns to publicly available or exchanged documents, images, graphs, files, etc. For example, the intelligence community has a number of tools at their disposal for analyzing traffic on the Internet, tracking access to websites, monitoring checked out literature from public libraries, etc. The main goal of a traffic attack is to detect unusual or questionable association between a sender and a recipient. Traffic analysis intuitively can identify who communicates with whom. The relationship between the communicating parties will be then qualified based on the contents of the message. Traffic attacks can be applied to any contemporary steganographic techniques regardless of the steganographic cover type (e.g., image, graph, audio file, text, etc) and can achieve successful results with relatively low costs. In the context of Chestega, the subject of the cover is checked rather than its validity and consistency. If someone sends, receives, accesses some materials without a legitimate reason for doing so, suspicion can be raised and further investigation may be warranted. The additional investigations will involve a thorough analysis of a steganographic cover, as detailed in the next subsection.

Traffic analysis is deemed ineffective with Chestega. Chestega camouflages the transmittal of a hidden message to appear legitimate and thus suspicion is averted. Basically, Chestega ensures that the involved parties establish a covert communication channel by having a well-defined relationship with each other. The popularity of Chess makes the association between a sender and recipient to appear legitimate. Analyzing the traffic between them will not reveal any questionable association and will not trigger any further investigation. Finally, it is noted that if further investigations on a Chestega Cover are triggered by traffic analysis, they would not be successful, as elaborated next. In Chestega, differentiating between a chess-cover that has or does not have a hidden message is extremely difficult.

5.2. Contrast and Comparison Attacks

One of the intuitive sources of noise that may alert an adversary is the presence of contradictions in the

DOI: 10.1002/sec

chess-cover, such as finding inaccurate details about a game or some naïve move made by a professional player. Contradictions can also be spotted when using data that indicates a clear violation of the rules in a game. The use of authenticated or untraceable data will definitely counter such an attack. Untraceable data means data that is based on a private context, for example, a game between two unknown amateurs, and thus cannot be contrasted or compared. Meanwhile, noise in the context of comparison attacks reflects alteration of authenticated or previously used data. The goal is to find any incorrect and inconsistent data that may imply the manipulation of chess-cover contents to include a hidden message. The vulnerability of Chestega to comparison attacks depends on how the cover is generated. Automating the generation of chess-cover through the use of contemporary chess databases and analysis tools makes the cover very resilient to this type of attacks. As demonstrated in Section 4, the use of a tool like Chessmaster has allowed the selection of appropriate games that match the encoded messages, and facilitated the generation of the game description and analysis. An adversary cannot detect any discrepancy in a chess-cover when examining the authenticity of the data and the consistency of the text with respect to the style of what Chessmaster usually generates.

It is worth noting that the traffic analysis, discussed earlier, can also be pursued as a base for launching comparison attacks in case the data is not publicly accessible. In that case, current data is compared to a record of old data in order to search for any inconsistency over some period of time. Countering such an attack is always a challenge because it requires consistency with data that was previously used over an extended period of time. Contradictions would surely raise suspicion about the existence of a hidden message. Chestega, as demonstrated through examples, is simply made contrast-aware. The flexibility in messages encoding and the ability of employing more than one cover type enable Chestega to avert such attack. Moreover, there are multiple styles of chess-cover that would even allow erroneous chess moves to appear in the cover. In addition, the description and analysis of an untraceable game do not imply the same game will be referenced in future communications and make consistency across multiple communication sessions an unexpected property. Finally, the popularity of chess ensures the availability of sufficient sources of correct and consistent data for embedding all sorts of messages and would not justify the need to repeatedly refer to the same untraceable or authenticated data.

5.3. Linguistics-Based Attacks

Linguistics examination distinguishes the text that is under attack from normal human language. Distinguishing the text from normal human language can be done through the examination of meaning, syntax, lexicon, rhetoric, semantic, coherence, and any other issues that can help to detect or suspect the existence of a hidden message. These examinations are used to determine whether or not the text that is under attack is abnormal. Generally, the text produced by natural language generation (NLG) systems usually meets the expected properties of a normal human language. Since a chess-cover is generated by contemporary tools like Chessmaster, which employs NLG techniques, a chess cover is most likely free of linguistic errors. Furthermore, if there are errors in the NLG engine, it should not be a concern for two reasons; first, it applies to all the generated text with and without a hidden message; second, nothing is concealed in errors. Therefore, it is obvious that Chestega is capable of passing any linguistic attack by both human and machine examinations.

On the other hand, a statistical attack refers to tracking the profile of the used text. A statistical signature (profile) of a text refers to the frequency of words and characters used. An adversary may use the statistical profile of chess text that contains no hidden message and compare it to a statistical profile of the suspected chess cover to detect any differences. An alteration in the statistical signature of a chess text can be a possible way of detecting a noise that an adversary would watch for. Tracking statistical signatures may be an effective means for attack since it can be easily automated and combined with traffic analysis. Nonetheless, Chestega is resistant to statistical attacks. As demonstrated, the chess cover is generated using public tools such as Chessmaster without any alteration. Any added sentences or even paragraphs to highlight a theme and legitimatize the reference to multiple games, are very small in size compared to the Chessmaster generated text. Obviously, it is expected that the statistical profile of chess-cover would match that of Chessmaster's text, deeming statistical attacks on Chestega to be very ineffective.

6. Conclusion

In this paper, Chestega, a novel methodology for steganography, has been presented. Chestega promotes

DOI: 10.1002/sec

the use of popular games like Chess as effective venues for covert communication. Chestega averts suspicion in covert communication by concealing a message using chess data. Chess data in this context includes chessboard positions, pieces and their color, moves, tournament name, place and results, players, etc. These data are exploited to conceal a message within a script of moves in a game, teaching sessions, game analysis, etc. Unlike most contemporary approaches, Chestega does not exploit noise to embed a message nor introduce a detectable noise. Instead, authenticated data can be employed in the cover which makes Chestega resilient to comparison attacks. Chestega also legitimizes the interactions between a sender and a recipient based on their interest in chess and thus makes traffic analysis ineffective. Numerous types of Chestega cover, for example, textual, image, graph, audio, can be pursued. In addition, the cover can be auto-generated by contemporary tools like Chessmaster, which employs natural language generation systems, and is thus resilient to linguistic and statistical profile attacks. Chestega is applicable to other games such as checkers, crosswords, domino, etc. and is worth investigating in the future.

References

1. Kipper G. Investigator's Guide to Steganography. CRS Press LLC: Boca Raton, Florida, USA, 2004; 15–16.
2. Davern P, Scott M. "Steganography its history and its application to computer based data files," *Internal Report Working Paper: CA-0795*, School of Computing, Dublin City University 1995. http://computing.dcu.ie/research/papers/1995/0795.pdf.
3. Desoky A, Younis M. Graphstega: Graph Steganography Methodology. *Journal of Digital Forensic Practice* 2008; **2**(1): 27–36.
4. Johnson NF, Katzenbeisser S. A survey of steganographic techniques. In Information Hiding, Katzenbeisser S, Petitcolas F (eds). Artech House: Norwood, MA, 2000; 43–78.
5. Bennett K. "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text," *CERIAS Tech Report 2004–13*, Purdue University, 2004.
6. Shirali-Shahreza MH, Shirali-Shahreza M. "A New Approach to Persian/Arabic Text Steganography" in the *Proceedings of 5th IEEE/ACIS International Conference on Computer and Information Science (ICIS-COMSAR 2006)* pp. 310–315, 10–12, Honolulu, Hawaii, July 2006.
7. Kahn D. The Codebreakers: The Story of Secret Writing (revised ed). Scribner: USA, 1996.
8. Wayner P. Mimic Functions. *Cryptologia* 1992; **XVI/3**: 193–214.
9. Wayner P. Disappearing Cryptography (2nd edn). Morgan Kaufmann: San Francisco, California, USA, 2002; 81–128.
10. Chapman M, Davida G. "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," in the *Proceedings of the International Conference on Information and Communications Security*, Vol. 1334 of Lecture Notes in Computer Science, Springer, pp. 335–345, Beijing, P.R. China, November 1997.
11. Grothoff C, Grothoff K, Alkhutova L, Stutsman R, Atallah M. "Translation-based steganography," Technical Report CSD TR# 05–009, Purdue University, 2005. (CERIAS Tech Report 2005–39) http://grothoff.org/christian/ lit-tech.ps.
12. Grothoff C, Grothoff K, Alkhutova L, Stuts Man R, Atallah M. "Translation-based steganography." In the *Proceedings of Information Hiding Workshop (IH 2005)*, pp. 213–233. Springer-Verlag, Barcelona, Spain, June 2005.
13. Stutsman R, Grothoff C, Atallah M, Grothoff K. "Lost in Just the Translation" in the *Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC'06)*, Dijon, France, April 2006.
14. Martin A, Sapiro G, Seroussi G. Is image steganography natural? *IEEE Transactions on Image Processing* 2005; **14**(12): 2040–2050.
15. Cvejic N, Seppanen T. "Increasing robustness of LSB audio steganography using a novel embedding method," in the *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04)*, pp. 533–537, Las Vegas, Nevada, April 2004.
16. Cvejic N, Seppanen T. "Reduced distortion bit-modification for LSB audio steganography" '04. 2004 in the *Proceedings of the 7th International Conference on Signal Processing (ICSP 04)*, Vol. 3 pp. 2318–2321, Beijing, China, August 2004.
17. Bender W, Gruhl D, Morimoto N, Lu A. Techniques for Data Hiding. *IBM Systems Journal* 1996; **35**(3 and 4): 313–336.
18. Kirovski D, Malvar H. "Spread-spectrum audio watermarking: requirements, applications, and limitations" in the *Proceedings of the 4th IEEE Workshop on Multimedia Signal Processing*, pp. 219–224 Cannes, France, October 2001.
19. Ansari R, Malik H, Khokhar A. "Data-hiding in audio using frequency-selective phase alteration," in the *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04)*, Vol. 5, 17–21 pp. 389–92 May 2004.
20. Gruhl D, Lu A, Bender W. "Echo Hiding," in the *Proceedings of First International Workshop on Information Hiding*, Lecture Notes in Computer Science, Vol. 1174 Springer, pp. 295–316, Cambridge, UK, May 1996.
21. Castroa J, Blasco-Lopez I, Estévez-Tapiador JM, Garnacho AR. Steganography in games: A general methodology and its application to the game of Go. *Computers and Security* 2006; **25**(1): 64–71. http://www.very-best.de/pgn-spec.htm.
22. Koblitz N. A Course in Number Theory and Cryptography (2nd edn). Springer: Germany, 1994; 54–76.
23. Kessler GC. "An Overview of Steganography for the Computer Forensics Examiner" An edited version, issue of Forensic Science Communications (Technical Report), Vol. 6, No. 3, July 2004. http://www.fbi.gov/hq/lab/fsc/backissu/july2004/research/2004_03_research01.htm. http://www.garykessler.net/library/fsc_stego.html.
24. Reiter E, Dale R. Building Natural Language Generation Systems. Cambridge University Press: Cambridge, UK, 2000.
25. Grune D. "Two-level grammars are more expressive than Type 0 grammars or are they?" SIGPLAN Notices, USA, Vol. 28, No.8, pp. 43–45, August 1993.

DOI: 10.1002/sec