2010 Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering

Steganography and Visual Cryptography in Computer Forensics

George Abboud
Department of Computer Engineering and Computer Science
Speed School, University of Louisville
Louisville, KY USA
gtabbo01@louisville.edu

Jeffrey Marean
Department of Computer Engineering and Computer Science
Speed School, University of Louisville
Louisville, KY USA
jsmare01@louisville.edu

Roman V. Yampolskiy
Department of Computer Engineering and Computer Science
Speed School, University of Louisville
Louisville, KY USA
roman.yampolskiy@louisville.edu

Abstract— **Recently, numerous novel algorithms have been proposed in the fields of steganography and visual cryptography with the goals of improving security, reliability, and efficiency. This paper discusses and compares the two methodologies. Some similarities and differences are presented, in addition to discussing some of the best known algorithms for each. Lastly, an idea for a possible algorithm which combines the use of both steganography and visual cryptography is suggested. There are several ways of hiding data in files of different formats, leaving various signs of hidden data. Can data hidden in an original image be detected after it undergoes visual cryptography? Would that be a scenario which computer forensic investigators and forensic software developers have to account for?**

Keywords-Visual Cryptography, Steganography, Computer Forensics, Anti-forensics, Data Hiding, Secrecy, Novel Visual Cryptographic and Steganographic Methods, Forensic Investigation

I. INTRODUCTION

Steganography is the art, science, or practice in which messages, images, or files are hidden inside other messages, images, or files. The concept of steganography is not a new one; it dates back many millennia when messages used to be hidden on things of everyday use such as watermarks on letters, carvings on bottom sides of tables, and other objects. The more recent use of this concept emerged with the dawn of the digital world. Experiments have shown that data can be hidden in many ways inside different types of digital files. The main benefit of steganography is that the payload is not expected by the investigators who get to examine the computer data. The person sending the hidden data and the person meant to receive the data are the only ones who know about it; but to everyone else, the object containing the hidden data just seems like an everyday normal object.

Cryptography, on the other hand, is the enciphering and deciphering of data and information with secret code. Visual cryptography uses the same concept except that it is applied to images. Visual cryptography can also be somewhat deceiving to the inexperienced eye, in such a way that, if an image share were to fall into the wrong hands, it would look like an image of random noise or bad art depending on the individual's experience. In the world of forensics, such noise could represent important evidence in a criminal case, if it is recognized and decrypted successfully.

Steganography and visual cryptography are somewhat similar in concept. Ultimately they both are ways of hiding data from prying eyes and in many cases from forensic and security investigators. Some claim that visual cryptography is another type of steganography and some claim the inverse. Although in their basic purpose of hiding

978-0-7695-4052-8/10 $26.00 © 2010 IEEE
DOI 10.1109/SADFE.2010.14

IEEE computer society

information they are indeed similar, when it comes to the data transformation algorithms steganography and visual cryptography take advantage of different methodologies in order to protect their respective payload.

In steganography, only the sender and receiver are aware of the hidden data and typically if the loaded file falls into the hands of anyone else they wouldn't suspect the hidden data. Whereas in cryptography, when someone receives data that is encrypted the first thing that comes to their mind is the question of what is encrypted and how they can decrypt the hidden message.

II. IMAGE STEGANOGRAPHY

Steganography is an area in which many studies and intensive research have been carried out. There are several different methods and algorithms of hiding data in different types of files. One example of an advanced hiding technique in images is using image layers [1]. This method divides the original image into several blocks, and then creates layers for each block of the binary values of pixels as matrices. The second step to hide the secret bits is to search within these layers' rows and columns and try to find the best match between the binary value of the pixel that is being hidden and the binary value of the pixel where we want to hide it [1]. So for example, if the value of the pixel that we want to hide is '1001', but we did not find a '1001' in any rows or columns of the binary layers of the original image, but we did find a '1000' then this is selected as the closest match and that secret pixel is hidden there.

This method hides less data per block, it only hides 1 byte in an 8 x 8 pixels block whereas other methods like the LSB (Least Significant Bit) matching revisited method hides 1 bit in every pixel [2]. So this method hides less data per block which increases performance and sustains a better image quality. The significant thing about this method is that it doesn't rely on hiding data in the LSB of pixel values, but tries to find the best secret pixel – original image layer pixel binary value match in higher layers of the image thus preserving the quality of the image which makes it somewhat resistant to steganalysis.

The Dynamic Compensation LSB Steganography method [3] provides an even higher resistance to steganalysis and histogram analysis. This method hides data in the LSB of the original image pixels, and then compensates dynamically on the resultant image. The experiments Xiangyang, Bin, and Fenlin did on this method showed that adding 1 to half the pixels of the image to hide data in resulted in a high sigma value, which means that the steganalysis is more likely to detect hidden data. So the dynamic compensation method proposed as an alternate method is to calculate sigma values based on different rows of pixels of blocks in the image. Then the lowest sigma value which is less than the threshold with which steganalysis detects the hidden information is taken as the threshold for adding 1 to the pixels to hide data in. So this dynamic compensation method picks and chooses blocks of rows of pixels in which to hide data in, as long as this alteration to the pixels maintains a sigma value lower than the chosen threshold to stay under the radar of steganalysis. Experimental studies on this method show that the embedding rate is close to 100% of pixels. Nevertheless, dynamic compensation causes RS (Regular Singular) Steganalysis sigma values to come closer to 0 implying a wrong judgment – as if saying that there is no hidden data in that image. The results of different steganalysis methods such as the conventional RS, conventional SPA (Sample Pair Analysis), and other improved RS and SPA steganalysis methods, show that the detection rate of data hidden using dynamic compensation is almost negligible, so this method proves successful in avoiding data hiding detection software even when embedding ratio is closer to 100%.

With advancement in methods of hiding data in images and the various new ways that one can hide data in images, we can foresee that it is a growing challenge for computer forensic investigators to detect hidden data. The fact that a computer forensics investigator is faced with thousands of image files when conducting analysis on a machine is challenging enough, not to mention the obstacles of detection software resistant hiding schemes.

III. DETECTION OF STEGANOGRAPHY

Niels Provos created a detection framework to research the claims of terrorists and criminals hiding data in images [4]. At first, he scanned eBay for two million images without any success in finding any hidden messages. Then he decided to widen the scope of the scan and tapped into the USENET archive where he scanned another million images. The scan resulted in 20,000 suspicious images using 'stegdetect'. Those images underwent a dictionary attack with a size of 1,800,000 words and phrases, but no hidden messages were found. These scans occurred a little after September of 2001. From this, we can conclude that both terrorists and criminals weren't using steganography, or that the available tools for detecting hidden messages weren't as reliable.

The detection of hidden data presents a big challenge to investigators and individuals looking for hidden data. For images only, there are hundreds of billions of images on the web and looking through all of them would be a very time consuming and computationally challenging task; let alone the other types of files that data can possibly be hidden in. Even if someone manages to go through all the current images on the web, what if some new algorithm

for hiding data in images emerges? Is the application used to scan the images for hidden data suitable for and capable of uncovering the hidden data? And is it feasible to go back and rescan all the images all over again with the same or other software updated to detect the hidden data by the new algorithm?

The answer to the above questions is that it's close to impossible to be able to accurately scan or attempt to detect hidden data on such a wide scope of suspect images. It is somewhat easier for investigators to scan for hidden data on a smaller scale such as an image of a hard drive, but they are still faced with the same software inaccuracy and the possibility of encountering unknown data-hiding algorithms.

IV. Other Types Of Steganography

Another interesting concept is one that is discussed in Steganography in MMS (Multimedia Messaging) by Mohammad Shirali-Shahreza [5]. With the expanding use of mobile communications, this becomes a very interesting area in which data hiding can be widely used. This method presents hidden communication using both text and image steganography. The author talks about hiding data in text messages or SMS by using the basic concept of abbreviation. He proposes the use of expressions like 'u' instead of 'you' or 'l8ter' instead of 'later'. While it is true that hidden data detection software designed to search for keywords in the regular form found in a language, it would require a simple modification to the software to have it also search for possible abbreviations. The method that he suggests hides data in both text and images. The data is first broken into two parts; each is proportional to the size of the text and the image. Then the size of the information is saved in the image for decoding purposes. Afterwards, the process of hiding data begins by looping through and hiding some bits in text and then some bits in the image. So some of the hidden data is in the text and some is in the image. This method doesn't require a sophisticated device or operating system on the mobile device as the author experimented using J2ME programming language which is compatible with most modern cell phones. So if a device is capable of sending MMS and SMS, this algorithm can be implemented on it.

V. Visual Cryptography

Visual cryptography is another way of sharing hidden data, except that it is limited to image formats. In its basic concepts, visual cryptography works in such a way that an image is split up into shares which look like white noise, but when those shares are overlaid they reveal the hidden image. Many studies have been performed in the area of visual cryptography and several algorithms have been developed.

One interesting visual cryptography method is the (t,n) Threshold Image Hiding Scheme [6]. This method hides a secret image into 'n' number of cover images, and can be recovered if 't' number of cover images are available. The hidden image can be up to 512 colors with a size as big as that of the cover images. This method uses Lagrange interpolating polynomial, MD5 hashing, and RSA signature to encrypt the image to be hidden [6]. The interesting thing about this algorithm is that during extraction of the hidden image from the cover images, it implements a cheat attack check where it checks whether these cover images are the same as the ones used to hide the data. If that check fails then the extraction of data is aborted. The authors of this method do not mention anything about the quality of the hidden image after extraction and how similar it is to the original image, although they do mention that the cover images used in their experiment are of relatively good quality with an average PSNR (Peaks of the signal-to-noise ratio) value of 31.34 [6].

Another visual cryptography algorithm is the Image Size Invariant Visual Cryptography [7]. This method hides two-tone secret image and splits it into binary transparencies which look like random noise images. Once those transparencies are stacked on top of each other, the secret image is revealed. The secret image can also be reconstructed by XOR computations of the transparencies. This algorithm is based on the conventional VSS (Visual Secret Sharing) method.

The JVW method is one that uses the concept of watermarking and visual cryptography jointly [8]. Since the DHCED (Data Hiding in Halftone Image by Conjugate Error Diffusion) method cannot prevent the secret image from being extracted with only one of the shares, JVW was proposed to overcome that issue [9]. JVW consists of two main steps; the first is to add some noise to the original multi-tone image. Introducing random noise to the original image breaks the direct correlation between it and the share images without affecting the perceptual quality, which means that when we overlay the shares we will still be able to identify the original image. The second step is to modify the DHCED algorithm to accommodate two halftone images instead of just one. An interesting point of this algorithm is that it does not reveal the secret image even if one has the original image and one of the shares; both shares have to be present to reveal the secret image [9].

Next the RIVC (Region Incrementing Visual Cryptography) method is discussed [10]. In RIVC, the original image is sectioned into 'n' number of secrets and then 'n+1' number of shares are then created. Any 'n' number of

shares stacked would reveal 'n-1' number of secrets [10]. The advantage to this method is that a user can pick which region of the secret image to assign to a secrecy level, and thus it makes it flexible and accommodating to user preferences. As this method may not seem to be as secure as other methods because of the fact that some levels of secrecy can still be revealed even if one doesn't have all the shares, it is hard for the person who is trying to reveal the secret data to know if the shares that they have are all the shares or if they're missing any. So if someone has 3 out of 5 shares and sees some data revealed, they may think that they've found the secret and stop looking for the other two. But if someone is using this method to hide a certain secret in a certain level, but decides to create other secrets as decoy, this doesn't guarantee the hider that others won't be able to reveal that secret if they happen to obtain the right shares. This is definitely an interesting method because it can be used in many ways and it is challenging to tell which shares reveal the real secret and which shares reveal decoy secrets.

The 'colour image secret sharing' is one of the newer proposed methods which are capable of encrypting a color image [11]. Its author claims that using the decryption module, perfect reconstruction can be achieved. Encryption of the image happens at the bit level of the blocks of the image. The result is a set of color-noise-like image shares. Because the encryption happens at the vector level, the shares have no correlation to the original image, which makes them resistant to brute force attacks that attempt to decrypt them. With this method, overlaying of shares doesn't reveal any data; the decryption module has to decrypt the shares for the data to be revealed. This is good for added security since only those with software which implements this algorithm are capable of revealing the secret image. Two advantages of this method are that it decrypts the image shares without altering the secret image or effecting its quality or dimensions, and that the decryption satisfies the perfect reconstruction property. This means that after decryption, one would obtain a revealed image that is identical in look and content to the original secret image.

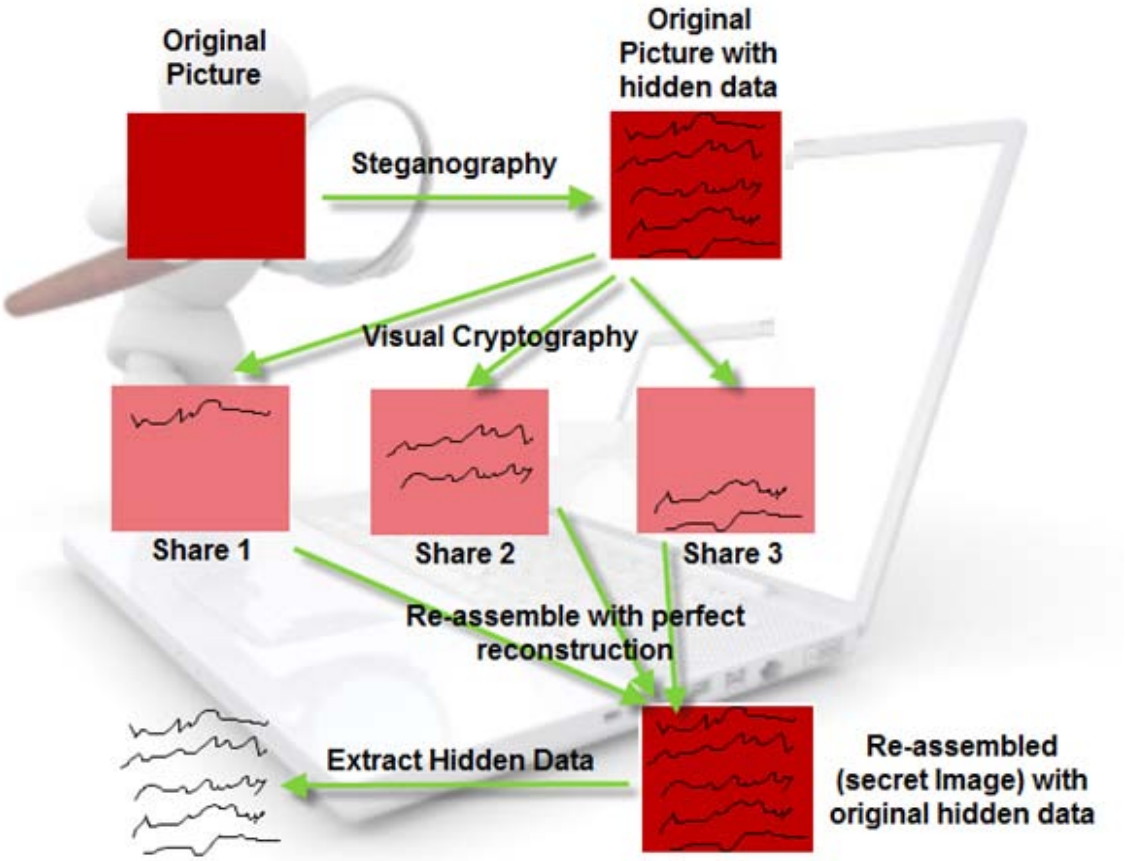

Figure 1: Proposed algorithm using both steganography and visual cryptography with perfect reconstruction.

VI. Proposed Algorithm for Further Research

Steganography and visual cryptography have so far been dealt with as two separate entities as far as possibility of use. A few algorithms touch on the concept of using steganography and visual cryptography together, such as the

JVW method mentioned above. JVW mentions the use of watermarking, embedding another image inside an image, and then using it as a secret image. The secret image would get split into shares which would need to be overlaid to reveal that secret image. The use of steganography alongside visual cryptography is a strong concept and adds a lot of challenges to detecting such hidden and encrypted data. To expand on this concept, research can be done on more ways where steganography can be used in conjunction with visual cryptography (See Figure 1). For example, imagine an algorithm which uses one of the strong algorithms of steganography to hide data (not necessarily another image) inside an image, and then uses that image as a secret image with a strong visual cryptography method. Basically we would then have a secret image with hidden data which would be split up into shares. These shares can also be innocent images, not necessarily noise images. Then when these shares are re-assembled or decoded to reconstruct the original image we would then have a revealed image which still contains the hidden data. So the receiver would be able to extract the hidden data from the revealed image. This algorithm cannot exist without having a perfect reconstruction property in the visual cryptography method. The reason for that is that if our reconstruction process or even the encryption process alters the image data, then it would consequently alter our hidden data which would make it impossible to extract the hidden data from the revealed image.

A few experiments were conducted using a hex editor (HxD) and visual cryptography software called 'Visual Cryptography Share Encryptor' [12]. Some plain text was hidden using HxD into an image file.

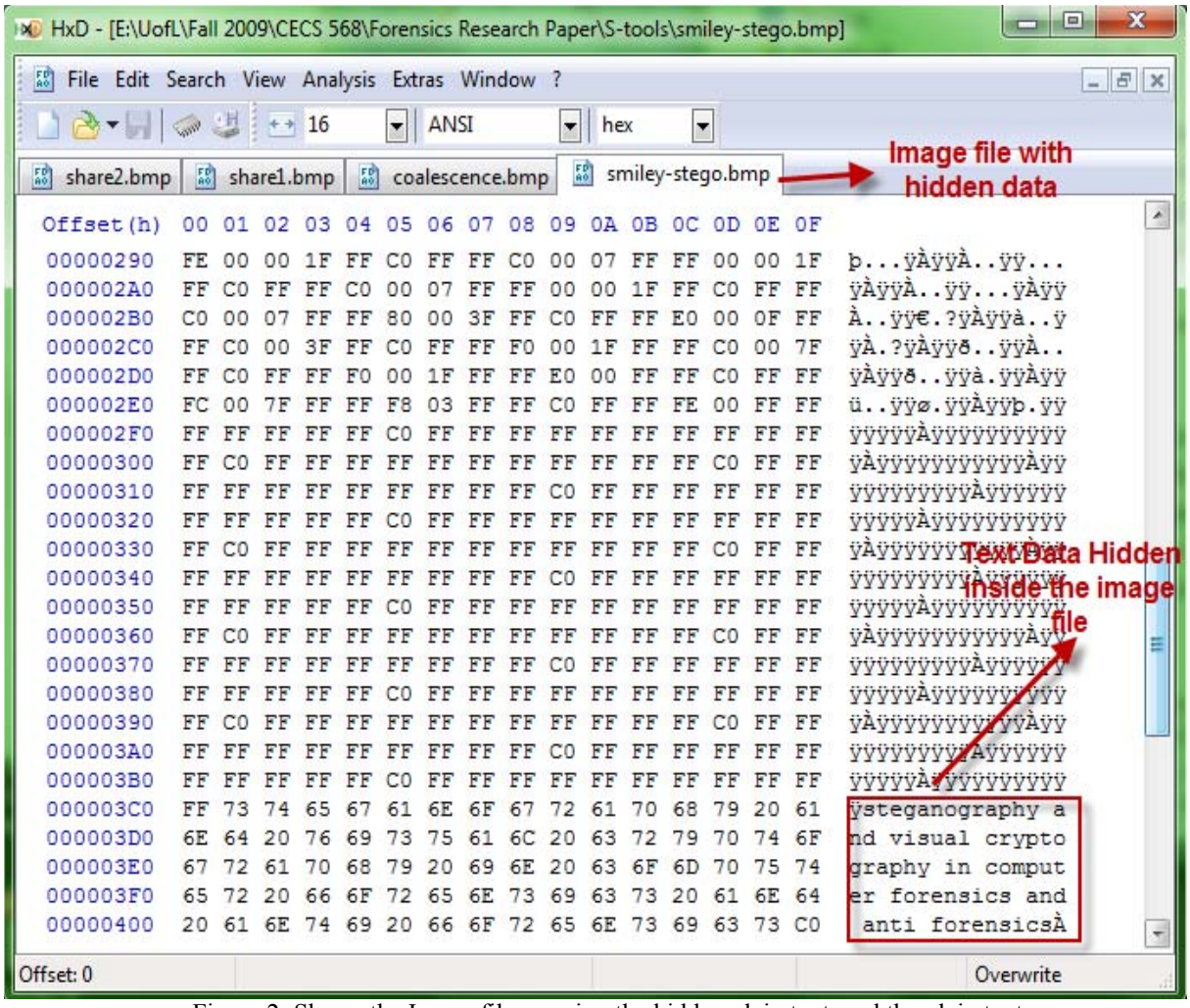

Figure 2: Shows the Image file carrying the hidden plain text, and the plain text.

Then the image with the hidden text is split into shares, each time using various schemes, resulting in image shares that look like noise. Notice the plain text cannot be spotted anywhere in the image data shown via the hex editor.

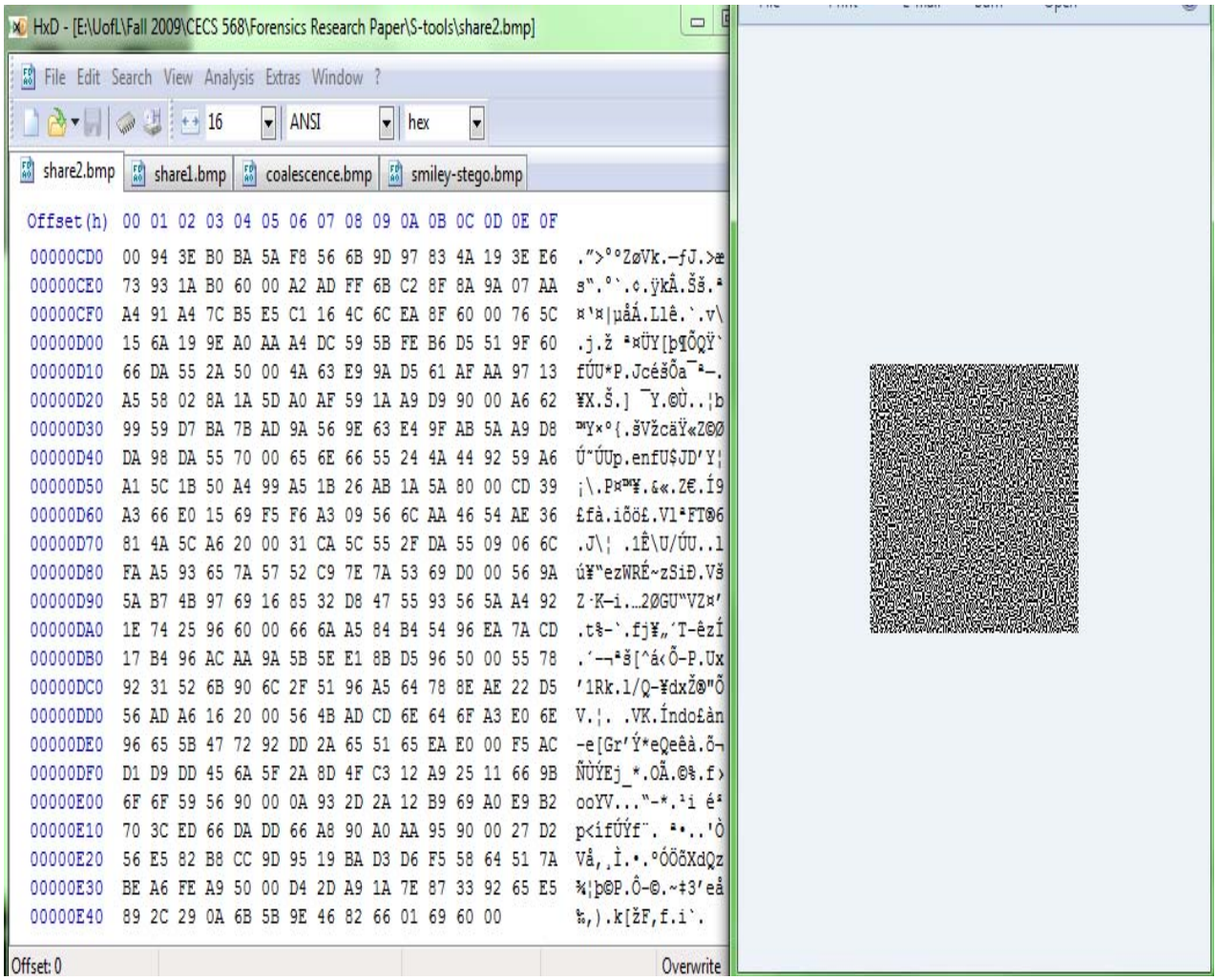

Figure 3: Showing one of the shares after applying a visual cryptography scheme on it.

Now, using the shares from the previous step, the image is reconstructed (See Figure 4). Again, notice that the data is now lost because of the absence of perfect reconstruction. Both, getting the shares and using them to reconstruct the hidden image, was done using the 'Visual Cryptography Share Encryptor' software.

This indicates that the algorithms used in this software lack the perfect reconstruction property since they do alter the data either in the process of obtaining the shares, or in the process of reconstructing the hidden image. So if we can establish a perfect reconstruction property in our visual cryptography method to where we are able to encrypt the image containing data into shares and then decrypt those shares back into an image and not alter the data, then this would potentially be an even more secure algorithm to communicate data. Perfect reconstruction can also be used for other purposes, such as being able to receive secret financial document shares and being able to reconstruct them into the exact financial document that was originally hidden [13]. So this is potentially a good area to research and explore where both steganography and visual cryptography can be used in conjunction.

On the other hand, this experiment presents a good way to fight steganography by altering the data but not completely destroying the image. So if an image is suspected to have some hidden data, this process of visual cryptography and then decryption would alter the data so it is corrupt but at the same time the image would still make sense to a human viewer.

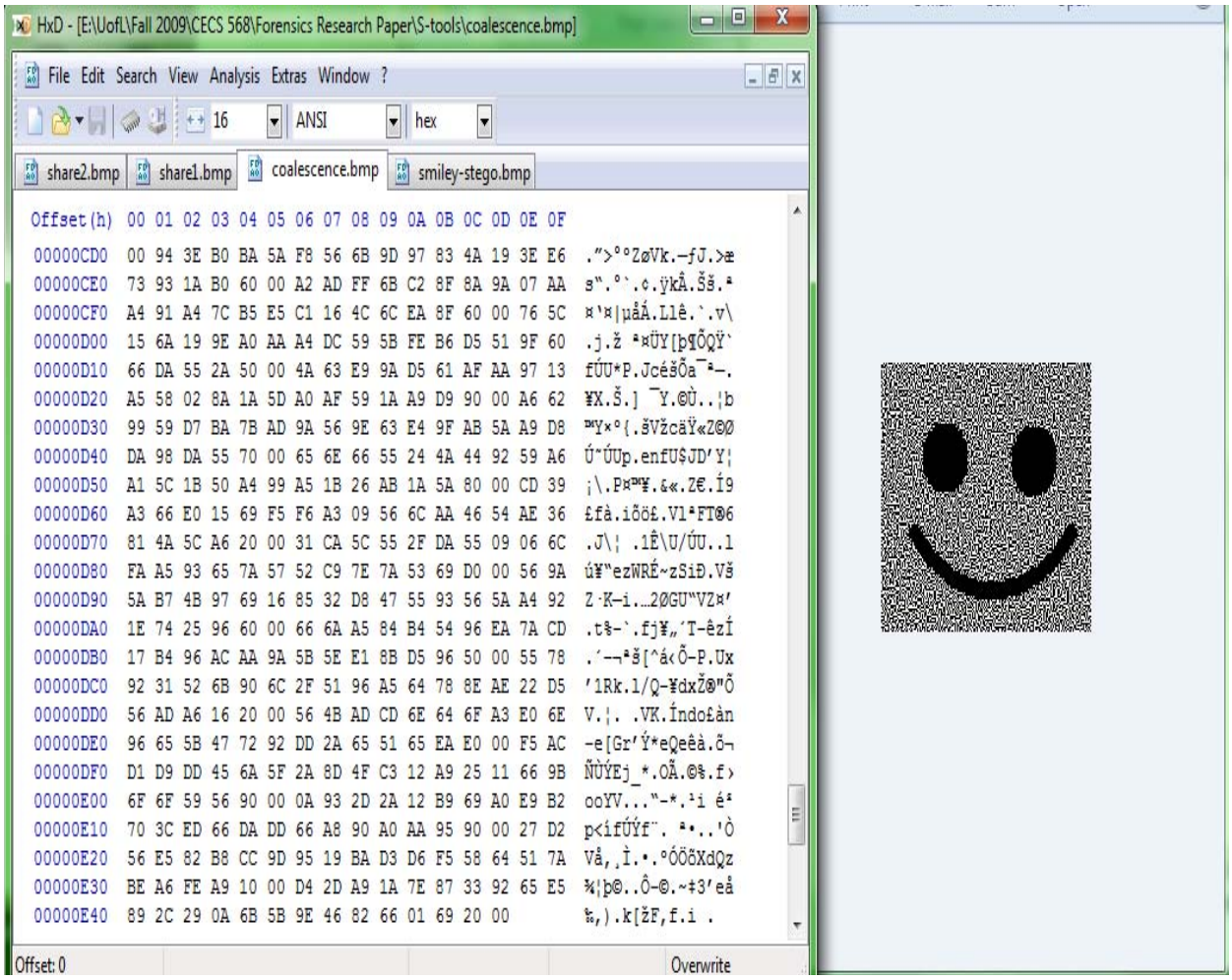

Figure 4: Shows the revealed image.

VII. CONCLUSION

In this paper, the definitions of steganography and visual cryptography have been discussed along with several studies done on various algorithms of each type. Steganography and visual cryptography have many similarities and differences, and thus have various uses in the digital and real worlds. Different algorithms for steganography and visual cryptography have different advantages and power, as well as disadvantages and weaknesses. So we notice that certain methods are easier to detect than others. But generally, the job of forensic and security investigators is not easy. When steganography and visual cryptography detection tools are used exclusively, it is almost impossible for investigators to uncover hidden or encrypted data. On the other hand, if these detection tools are used in conjunction with other tools and factors that narrow down the search to a somewhat smaller data set, then it makes the lives of investigators much easier and gives them a better chance of detecting suspicious data.

We notice that using an algorithm with a solid reconstruction method will allows us to reconstruct shares back into the original, unaltered image. This algorithm would present a great area for further exploration which would open up some more venues in the world of forensics and anti-forensics. It would be very interesting to learn how detectable data is after applying visual cryptography with perfect reconstruction to an image with hidden data.

Also, an interesting detection question is whether we can reconstruct a set of 'n' shares into a meaningful image that is different than the image used to create those shares by omitting some of the n original shares and by including an additional share specifically constructed for such purpose. Basically this is a question about the uniqueness of the shares created by different visual cryptography algorithms. So if we obtain a set of shares and attempt to reconstruct them, could they construct an image with illegal content although they might not have come from an image with illegal content? How unique are those shares that we obtain from these different visual cryptographic algorithms and how much influence can be exerted by an unethical investigator during the decryption process?

REFERENCES

[1] O. Kurtuldu and N. Arica, "A new steganography method using image layers," in *Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on*, 2008, pp. 1-4.

[2] J. Mielikainen, "LSB matching revisited," *Signal Processing Letters, IEEE,* vol. 13, pp. 285-287, 2006.

[3] L. Xiangyang, L. Bin, L. Fenlin, "A Dynamic Compensation LSB Steganography Resisting RS Steganalysis," in *SoutheastCon, 2006. Proceedings of the IEEE*, 2006, pp. 244-249.

[4] Provos, N. (2001). Scanning USENET for Steganography. from http://niels.xtdnet.nl/stego/usenet.php

[5] M. Shirali-Shahreza, "Steganography in MMS," in *Multitopic Conference, 2007. INMIC 2007. IEEE International*, 2007, pp. 1-4.

[6] C. Chin-Chen and L. Iuon-Chang, "A new (t, n) threshold image hiding scheme for sharing a secret color image," in *Communication Technology Proceedings, 2003. ICCT 2003. International Conference on*, 2003, pp. 196-202 vol.1.

[7] L. Hao and Y. Faxin, "Data Hiding in Image Size Invariant Visual Cryptography," in *Innovative Computing Information and Control, 2008. ICICIC '08. 3rd International Conference on*, 2008, pp. 25-25.

[8] F. Ming Sun and O. C. Au, "Data hiding in halftone images by conjugate error diffusion," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, 2003, pp. II-920-II-923 vol.2.

[9] F. Ming Sun and O. C. Au, "Joint visual cryptography and watermarking," in *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, 2004, pp. 975-978 Vol.2.

[10] W. Ran-Zan, "Region Incrementing Visual Cryptography," *Signal Processing Letters, IEEE,* vol. 16, pp. 659-662, 2009.

[11] R. Lukac and K. N. Plataniotis, "Colour image secret sharing," *Electronics Letters,* vol. 40, pp. 529-531, 2004.

[12] P. R. Busse. (2003, *Visual Encryptor*. Available: http://compsci.snc.edu/cs460_archive/2003/busspr/VisualEncryptor.html

[13] J. Cai, "A Short Survey on Visual Cryptography Schemes." 2004, http://www.cs.toronto.edu/~jcai/paper.pdf