Towards Eliminating Steganographic Communication

Anthony Whitehead
School of Information Technology
Carleton University

Abstract

There have been a number of steganography embedding techniques proposed over the past few years. In turn, there has been great interest in steganalysis techniques as the embedding techniques improve. Specifically, universal steganalysis techniques have become more attractive since they work independently of the embedding technique. In this work, we examine the effectiveness of a basic universal technique that relies on some knowledge about the cover media, but not the embedding technique. We examine images as a cover media, and examine how a single technique that we call steganographic sanitization performs on 26 different steganography programs that are publicly available on the Internet. Our experiments are completed using a number of secret messages and a variety of different levels of sanitization. We benchmark the level of sanitizing versus the subsequent detection capability by trying to extract the hidden message after the sanitizing process However, since our intent is to remove covert communication, and not authentication information, we continue by examining how well the sanitization process preserves authentication information such as watermarks and digital fingerprints.

1. Introduction

Steganography literally means "covered writing" and dates back to the ancient Greeks where message runners would have messages tattooed to their shaven heads and dispatched once the hair grew back. Upon arrival at their destination, the head was once again shaven and the message could then be read. Other methods included carved messages on wooden tables that were later covered in wax. Today, steganography is the science of hiding information by embedding covert messages within other, seemingly harmless pieces of data. Steganography works by replacing bits of unused or imperceptible areas in regular computer files (such as graphics, sound, text, video data, etc) with bits of different information. This hidden information can be plain text, cipher text, or any other form of digital data such as images, documents and schematics.

Steganography is sometimes used when encryption is not permitted. Or, more commonly, steganography is used to supplement encryption. An encrypted file may still hide information using steganography, so even if the encrypted file is eventually deciphered, the hidden message is not seen. Conversely, encrypted messages can be hidden in non-encrypted data to circumvent the rules that disallow encryption. Moreover, should the hidden communication be found, depending on the encryption scheme it could be very difficult to determine what the message really is.

Steganalysis is the science of discovering such covert messages embedded in the media. The idea behind detecting steganographic communication is, that once discovered, it can be dealt with. However discovery techniques rely largely on statistical based methods [1] that are becoming less effective as the information hiding researchers begin to apply the basic principles of encryption: Randomness, large key spaces, uniform distribution, and the inability to allow sequence guessing [2]. As the steganography techniques become more sophisticated at covertly hiding the embedded messages, steganalysis will become a more computationally intensive process.

Given that information can be imperceptibly embedded into a cover medium with very little computational expense, we claim that it is also possible to scramble the information in the cover medium where steganographic communications might exist, also with little computational overhead. In this paper we experimentally examine the elimination of steganographic communication from cover media (in our tests we used images) on 26 different, publicly available Steganography programs that we were able to get from the Internet

In this paper we briefly review steganography, and steganalysis in Section 2. We continue by examining what threats steganography can induce and what can be done to help deter those risks in Section 3. Section 4 discusses methods for eliminating steganographic communications. In Section 5 we expose our experimental results and show how one technique that uses only the knowledge of the cover medium can effectively eliminate Steganography as a form of covert communication. Finally in Section 6 we sum up our results and draw some conclusions.

2. Overview of Steganography and Steganalysis

Steganography, as a process, can be simply explained as the embedding of one information source into another. As we see in Figure 1, we embed our hidden message which is often called the *payload*. The payload is embedded into a 2nd digital file often called the *carrier*. The result is the *stego-media* that is perceptually identical to the carrier.

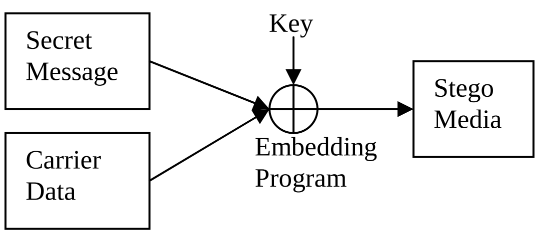

Figure 1: The steganographic embedding process

When we say *imperceptible,* we mean that the embedding program should produce no obvious artifacts in the resulting stego-media that would bring suspicion on the media. For example, in Figure 2, we present an image file to be used as carrier data. In Figure 3 we present the stego media presented after a 194 kilobyte message has been embedded into the carrier data. The message is actually the document you are reading right now. As you will notice, you cannot see, nor would you know by simple inspection that information is "hiding" within the image in Figure 3.

Figure2: Carrier data

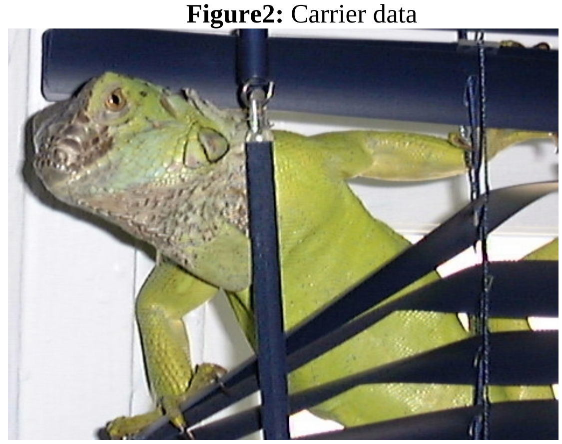

Figure 3: Stego-media with embedded document

Embedding without visual artifacts allows the observation of the communication without raising any suspicions. The ultimate intent of steganography is to maximize the communications bandwidth, minimize the perceptibility of the communication and ensure robustness of the embedding [3]. These three forces act in opposition to one another as shown in Figure 4.

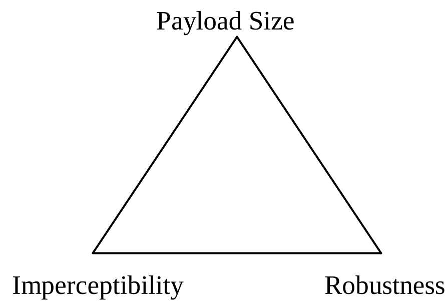

Figure 4: Tradeoffs in information embedding

The fundamental conclusions we can draw from the above are the following:

1. Imperceptible communication is not robust
2. Embedding large messages is not robust

Moreover, we can infer that multiple small messages are more likely to be noticed and therefore less desirable by those who are willing to attempt covert communication.

Steganalysis is the process of examining a message and looking for the existence of a covert message within the original message. Loosely classified into *passive* and *active* steganalysis, passive steganalysis simply tries to detect the presence of a message while active analysis attempts to extract the message length and location, estimate the secret key used in the embedding process, and ultimately extract the secret message itself. The methods to perform these tasks are largely statistical in nature [4,5,6,7]. However it is becoming much more difficult for statistical methods to detect the presence of steganographic embedding [8].

However, if our intent is to secure a network against covert communications, our efforts can be restricted to elimination of steganographic communication as a whole, rather than detection of specific instances. We next provide a detailed analysis of the threat of steganography to organizations.

3. Steganography Threat Analysis

While steganography may seem to be an excellent apparatus for the exchange of sensitive documents and information in a concealed manner, it can also be used in ways that are counter productive to our security measures. Besides the examples of hidden files within pictures, there is speculation terrorists may be using steganographic techniques to communicate via seemingly innocent Web sites. It seems to be inevitable that such techniques will grow in popularity among those who are trying to communicate and feel that secrecy is a very high priority. Since the mere fact that two people communicating can bring on suspicion by association, there exists extremely high utility value in keeping the communication itself hidden. It should come as little surprise that those who tend to engage in subversive activities will also utilize all of the tools available to keep their actions (and associations) private.

The core threat from steganography is lack of knowledge. Vital information that is leaked out can have extremely dire consequences. Since we may not be aware of the leak, there is no ability to counter or prepare for the subsequent attack. As well, information can be sent into the organization. Should there be a covert operative working on the inside, instructions and missions can be sent into the organization for the person in question to execute on. Furthermore, since any type of information (such as executable programs) can be hidden via steganography, it is feasible (with the aid of a covert operative) to infect the internal computer network. Perhaps even bring down a critical infrastructure system.

Although the main threat at the moment would seem to be of direct concern to national security, it is well within the realm of imagination that the technology can be used for purposes in the financial and commercial markets. Information regarding money laundering, insider trading, the illegal drug trade, the distribution of child pornography and trafficking in humans can all be concealed using steganography. Although steganography is not yet the sort of threat that IT auditors and emergency preparedness officials will come up against on a regular basis, it is one that needs a thorough understanding. Furthermore, a willingness to address methods to secure their organizations now will go a long way in preventing a disaster in the future.

4. Eliminating Steganographic Communication

Since detection and possible decryption of a covert message is computationally intense, it is not feasible to analyze every possible data item that passes through an organizations gateway. However, steganographic sanitization is an extremely fast process that allows for the communication channel to be effectively scrubbed of covert communications. While the solution itself does not lend any further information to find culprits of espionage and mal-intent, it does help to secure the organization overall.

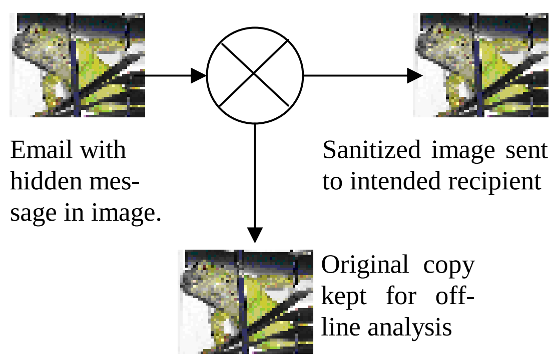

Figure5: Strategy for removing covert messages.

The Elimination strategy as outlined in Figure 5 has a three tier effect:

1. Removes the network as a channel for covert communication.
2. Allows further analysis of items in an off-line situation and allows targeted analysis.

3. Forces the two communicating parties into further communication to determine what happened to the covert message.

These effects provide the following benefits for the organizations security:

1. Important information will not escape the organization forcing the covert operatives to use another more traditional (hopefully, easier to detect) form of communication.
2. A more detailed analysis of the original media is possible, allowing for the potential detection of covert communication and subsequent target for further investigation.
3. An aura of confusion and mistrust may be created between the sending and receiving operatives. This may force further communication between the two parties using methods that are easier to detect.

Finally, as we are sanitizing the entire channel, all communication that is not covert will also be scrubbed. Since the sanitized data is imperceptibly changed, the recipients of "clean" data should not notice our interference. We continue with a theoretical analysis of the elimination strategy followed by a detailed explanation of our implemented strategy.

4.1 Game Theory Analysis of the Elimination Strategy

Steganography is the study of information hidden within a carrier host data set. A message is hidden in the carrier data is sent in a covert way designed to be imperceptible to a casual analysis. An attacker (warden) may seek to detect the presence of the message, and may introduce additional distortion in an attempt to destroy the hidden message. Ultimately, the resulting data set is analyzed using information shared with the information hider (such as a decryption key) to extract the hidden message. Thus, there is a game between the information hider and the attacker (warden).

The attacker (warden) seeks to minimize information communicated while the information hider seeks to maximize it. The value of the mutual information for the pair of hiding and attack strategies bounds the communication rate for the information hider. Since the warden determines the channel subject to a distortion constraint, the optimal attack strategy is the solution of a rate-distortion problem. Hence, the optimal attack is equivalent to optimal data compression subject to a distortion constraint, with the information hider determining the distribution on the carrier data. Since the information hider determines the input distribution to the channel, the optimal hiding strategy corresponds to the solution of a constrained channel capacity problem. Hence, the optimal hiding strategy is equivalent to optimal channel coding, with the warden determining the channel transition probabilities. The optimal pair of strategies is then a saddle point for the conditional mutual information function [5].

Classical game theory suggests that that this may indicate a model of a zero sum matrix game that can have optimal strategies. Mark Ettinger, of Los Alamos National Labs had performed an analysis of a game of steganography with a 2 players [5]. The first player intends to pass messages using steganography while the second intends to remove the message. This game exactly models our method of steganographic sanitization via an email server. His analysis shows that the attackers (the person removing hidden information) optimal strategy includes the flipping of all of the bits, which distorts the image beyond recognition. Not very useful when we intend to cleanse all data going through our communication channel.

By finding an equilibria point for minimizing channel distortion, we found that an optimal strategy in this case is for the warden was to flip some of the least perceptual bits in the image. Our methods meet this requirement and, in fact, exceed the equilibria requirements by randomizing the areas where perception of the distortion would be minimized. Remember that equilibria means that no player benefits from changing their strategy and this does not necessarily mean that the strategy is optimal. The optimal strategy does, in fact, include flipping all bits that allow the distortion to be minimized; our method performs a variant on this strategy since the optimal strategy is easily reversed.

4.2 Our Elimination Strategy

Our elimination strategy takes into account that we know the type of carrier media and takes advantage of this knowledge. While we focus on images for our experimental evaluations, the concepts behind the method could be applied to other carrier types as well. We rely on the basic idea that if one can imperceptibly embed information into an image, one should be able to imperceptibly remove the hidden information by altering the areas where extra information can be hidden. Effectively if the hiding places are so unnecessary to the visual quality of an image, then there is no harm in randomizing these areas. This method while conceptually simple has several implementation complexities that we will discuss next.

As we discussed earlier, the optimal strategy would be to flip all of the bits in the image, however, this would distort the image content beyond recognition. Initially one may assume that you want to apply the optimal strategy to only the hiding areas within the carrier data file. This would, in essences, make the data as different as possible from the original, but it is much too simple to reverse the process by simply inverting all the bits of the extracted data. This forces us into a less than optimal strategy of randomizing the bits that comprise the hiding places in the carrier media.

Our approach to randomization is to employ encryption techniques with randomly generated key sets. Recall that one of the core capabilities of a good encryption scheme is to make the output bit stream appear random. The other important capability is that it is difficult to revert to the original bit string from the resulting cyphertext. To this end, we concatenate all of the bits that comprise all of the possible hiding areas in image data and consider this our source string *H*. We then use the source string *H* as plaintext and proceed to encrypt the bits to produce the ciphertext string of bits *C*. The final step is to replace the bits that comprise *H* in the image data with the bits from *C* of the ciphertext.

Figure 6: Stego-media with embedded document

Our approach will cause no more distortion than that of the steganography techniques themselves. As Figure 6 shows, there is no visual distortion that is apparent to the image after the steganographic sanitization process has been employed. By exploiting the knowledge of where information *can* be hidden in an image we are able to effectively *remove* hidden information. However, the technique of sanitization is subject to the typical hacker/security arms race. Should there be a discovery of new areas in the image data, the construction of the plaintext string *H* needs to be adjusted to take that into account. For example, some common hiding places for information in image data include LSB (Least Significant Bits), DCT (Discrete Cosine Transform) coefficients, and DWT (Discrete Wavelet Transform) coefficients for which our method considers.

5. Experimental Evidence

We have conducted a large number of experiments using steganography programs that are publicly available on the Internet. In recent years, the number of publicly available steganography programs has reduced due to global events and suspected terrorist use of these programs. None-the-less we were able to find and successfully use a total of twenty-six different steganography programs when testing our elimination strategy. Our tests involved embedding multiple different types of information into the images that were used as the carrier data, followed by a steganographic sanitization, followed by an attempt to retrieve the initially embedded data. The result of the data retrieval was judged using a binary decision: Was the data successfully retrieved and usable? Furthermore, we performed a qualitative visual analysis of the images to determine whether or not a perceptible change had occurred as a result of our sanitization process. Finally, we also conducted a “load and save” test that loaded the stego-media and simply resaved the image, applying appropriate compression algorithms where necessary. The time required to sanitize a typical image (2 mega pixels) is less than 500 milliseconds which has little impact in a practical implementation of an SMTP filter, WWW proxy server or firewall.

Our results, outlined in Table 1, indicate that clearly some embedding techniques are susceptible to lossy image compression such as JPEG, and that non-compressed images are not affected by a save and load situation. This result is not surprising since the compression algorithms can be considered a form of channel distortion by removing elements of the image data that are not visually important. For example, the JPEG compression reduces the colour range of an image and blurs the edge boundaries slightly. In essence, by re-saving an image with JPEG compression, we are altering the DCT table where information was hidden.

Our tests were also run on 3 levels of additive distortion for each steganographic embedding software. Outlined in Table 1, we were able to successfully sanitize images from all 26 methods using level 2 or higher, and most were sanitized with only level one distortion. But in all 3 levels of distortion, the qualitative analysis was unable to notice the distortion.

Since we are looking to remove covert communication and leave typical communicators unaware of our interference, it is fundamentally important that we also do not remove watermarking information that is used to authenticate images and protect against copyright violators. In our next experiment we examine the effectiveness of the sanitization process in maintaining watermark information. As Table 2 outlines, most watermark embeddings tested were not affected by our sanitization process at the lowest level of distortion. However as we increased the distortion level, the watermarks were being removed more often. Recall, from Figure 4, that one of the tradeoffs in the feature triangle was robustness. Since the watermark data is small, watermarking systems are able to apply redundant embeddings to increase the robustness. As we increase the distortion level of our proposed method, we increase coverage of the triangle from the top down.

Table 1: Test results on 26 different steganography programs. ✗ indicates failure, hidden information WAS NOT removed , ✓ - indicates success, hidden information WAS removed, ⊠ indicates that the steganography software does not support this carrier file type, or test scenario.

Software	Carrier File	Message File	Visual Inspection 1-5 poor-good	Scrambling steganographic areas for 3 different scramble levels and the load and save test results			
				1	2	3	Load & Save
1. Stash-It! V 1.1	BMP	Txt	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Exe	5	✓	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠
2. Invisible Secrets 3	BMP	Txt	5	✗	✓	✓	✗
		Zip	5	✗	✓	✓	✗
		Exe	5	✓	✓	✓	✗
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Exe	5	✓	✓	✓	✓
3. Jstep Shell 2.0	BMP			⊠	⊠	⊠	⊠
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓
		Exe	5	✓	✓	✓	✓
4. S Tools 4	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
		Exe	5	✓	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠
5. Steganos 3	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠
6. Hide in Picture	BMP	Txt	5	✗	✓	✓	✗
		Zip	5	✗	✓	✓	✗
		Doc	5	✗	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠
7. Contraband Hell	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠

8. Blindside	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
9. PGE10	BMP JPG	Txt Zip Doc Txt Zip Doc	5 5 5 5 5 5	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓	✗ ✗ ✗ ✓ ✓ ✓
10. Stego 4	BMP JPG	Txt Zip Doc	5 5 5	✗ ✗ ✗ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
11. Camouflage v1.2.1	BMP JPG	Txt Zip Doc Txt Zip Doc	5 5 5 5 5 5	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓	✗ ✗ ✗ ✓ ✓ ✓
12. Hide4PGP**	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
13. BMP Secrets	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
14. wbStego4	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
15. Steganos v 4.11	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
16. In Plain View	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
17. JP Hide and Seek	BMP JPG	 Txt Zip	 5 5	☒ ✓ ✓	☒ ✓ ✓	☒ ✓ ✓	☒ ✓ ✓
18. Courier V1.0a	BMP JPG	Txt Zip Doc Exe	5 5 5	✓ ✓ ✓ ✓ ☒	✓ ✓ ✓ ✓ ☒	✓ ✓ ✓ ✓ ☒	✗ ✗ ✗ ✗ ☒
19. Digital Picture Envelop	BMP JPG	Txt Zip Doc	5 5 5	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✓ ✓ ✓ ☒	✗ ✗ ✗ ☒
20. Hide and	BMP	Txt	5	✓	✓	✓	✗

Seek 95 v1.1		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG			⊠	⊠	⊠	⊠
21. Jsteg Shell v2.0	BMP			⊠	⊠	⊠	⊠
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓
		Exe	5	✓	✓	✓	✓
22. MASKER v5.0	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓
23. F5	BMP	Txt	5	✓	✓	✓	⊠
		Zip	5	✓	✓	✓	⊠
		Doc	5	✓	✓	✓	⊠
24. WinStegano	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓
25. Gif it Up	GIF	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓
26. Stella	BMP	Txt	5	✓	✓	✓	✗
		Zip	5	✓	✓	✓	✗
		Doc	5	✓	✓	✓	✗
	JPG	Txt	5	✓	✓	✓	✓
		Zip	5	✓	✓	✓	✓
		Doc	5	✓	✓	✓	✓

Table 2: Results of sanitization process watermarked images from various programs.

Software	Watermark intact?		
	L1	L2	L3
1. Picture Shark	YES	YES	NO
2. Alpvision	YES	NO	NO
3. ASI Watermark	NO	NO	NO
4. UniDream	YES	YES	YES
5. Trans Watermark	YES	YES	YES

5. Conclusions and Review

National security is one example where steganographic communication, if left unchecked, can have extremely dire consequences. As steganographic communication techniques become more sophisticated and appear to be statistically random, the detection of steganography will become more and more difficult, if not impossible to complete in a reasonable time frame. The alternative is to ensure that steganographic communication is not able to occur in the first place. In this paper we have proposed a method that we call steganographic sanitization that uses knowledge of the carrier medium to cleanse all communication of possible steganographic content as it passes through a single point be it an SMTP server, a Web proxy server or a firewall.

We have proposed a novel alternative method for securing a communication channel against steganographic use by eliminating steganography rather than trying to detect it. Moreover, we have shown that the method is an extremely successful method against publicly available steganography programs. Because detection is going to become even more computationally intensive, we believe that detection will not be possible in real time and our proposed sanitization process will allow the communication channel to remain open and subsist at a higher security level.

Our experiments show that our proposed method is extremely vigorous and capable of removing steganographic messages from images without distorting the carrier data too greatly. The sanitization process, while conceptually very simple, has many implementation complications that require domain expertise in each individual carrier data type that one hopes to secure. Furthermore, the sanitization process does allow for subsequent detailed examination of carrier data without seriously degrading the communication channel bandwidth.

Finally, it is worth noting that the strategy we have applied to image data is directly applicable to other forms of steganographic embedding. Embeddings in audio files can be sanitized using the same process of enumerating the bits used by the information hiding processes and randomizing these bits subsequently in the sanitization process.

6. References

[1] N. F. Johnson and S. Jajodia, “Workshop on Information Hiding Proceedings”, Portland, Oregon, USA, 15 - 17 April 1998. Lecture Notes in Computer Science, Vol. 1525, Springer-Verlag: 273-289.

[2] W. Thompson, A. Yasinsac, T. McDonald, Semantic “Encryption Transformation Scheme”, International Workshop on Security in Parallel and Distributed Systems, San Francisco, 2004.

[3] N.F. Johnson, Z. Duric, S. Jajodia *Information Hiding, Steganography and Watermarking - Attacks and Countermeasures* Kluwer Academic Pub Books, November 2000.

[4] R. Chandramouli, “Mathematical approach to steganalysis”, Proceedings of Security and Watermarking of Multimedia Contents IV, SPIE Photonics West, Calif. 2002.

[5] J. Mark Ettinger. Steganalysis and Game Equilibria, IHW'98 - Proceedings of the International Information hiding Workshop, April. 1998, 319-328.

[6] N. F. Johnson, S Jajodia. “Steganalysis: The Investigation of Hidden Information”, IEEE Information Technology Conference (IT98), Sept. 1998.

[7] N. Provos, “Probabilistic Methods for Improving Information Hiding” , CITI Technical report, Jan, 2001.

[8] N. Provos, “Defending Against Statistical Steganalysis”, 10th USENIX Security Symposium. Washington, DC, August 2001.