A New Paradigm Hidden in Steganography*

Ira S. Moskowitz
Center for High Assurance Computer Systems
Naval Research Laboratory
Washington, DC 20375

Garth E. Longdon†
Center for High Assurance Computer Systems
Naval Research Laboratory
Washington, DC 20375

LiWu Chang
Center for High Assurance Computer Systems
Naval Research Laboratory
Washington, DC 20375

ABSTRACT

We discuss how steganography, in contrast to similar disciplines, requires a new paradigm based upon discontinuities and the absence of noise as a detection deterrent.

Keywords

information hiding, steganography

1. INTRODUCTION

Steganography, which is Greek for "covered writing," is a subset of the emerging discipline of *information hiding* [12, 1, 5, 18, 13]. It is the science of transmitting a message between two parties (Alice and Bob) in such a manner that an eavesdropper (Eve) will not be aware that the message exists. The terms "information hiding" and "steganography" are often, but incorrectly, used interchangeably. Information hiding is the broad term for the scientific discipline which studies topics such as covert and subliminal communication channels, detection of hidden information (e.g., steganography), watermarking of digital objects, and anonymity services. Unlike cryptography, which seeks to hide the *content* of the message, with steganography we seek to hide the *existence* of the message. Steganographically hidden messages are inserted into legitimate and obvious (with respect to Eve) communications between Alice and Bob. Eve's steganographic challenge, therefore, is to detect the message, not to understand it. Of course, steganography and cryptography can be used in conjunction, so that message content may be protected cryptographically, even if the steganographic "shield" fails and the existence of the message is discovered.

*US Government work.
Research supported by the Office of Naval Research.
NSPW2000, Ballycotton, Co. Cork, Republic of Ireland.

†ITT Industries

New Security Paradigm Workshop 9/00 Ballycotton, Co. Cork, Ireland
ACM ISBN 1-58113-260-3/01/0002

1.1 Paradigms old and new

The paradigm of cryptography (the "old" paradigm) is that cryptography can be modeled, measured, and utilized by the standards of information theory and noise. We have Shannon [21] to thank for this. Attempts have been made to extend this paradigm to steganography [6, 16, 25]. We find that these extensions, although useful, do not capture all of the essence of steganography. Note that the authors of [6, 16, 25] never claimed that their work did. We propose a "new paradigm" for steganography, based upon (1) discontinuous mathematical models, and (2) the lack of noise as a detection deterrent. This is not to say that the present steganographic models do not take, at least part of, this thinking into account. However, we feel that it is important to delineate these ideas as a new paradigm to force ourselves to think of steganography in a different light than that of cryptography. Perhaps by looking at steganography in light of our new paradigms, the present steganographic models can be "filled out" to capture more of the essence of steganography.

In this paper, we also discuss how (part of) the old paradigm applies to covert channels, but not to the steganographic equivalent—subliminal channels. Our ideas are preliminary and works-in-progress. We invited discussion, encouragement, and criticism from the workshop participants, and received it. Because much of this community's work is based upon ideas from Shannon, some may (especially the first author) find it hard to break away from the old paradigm of continuity and noise. We are quite respectful of the existing steganographic techniques. They are a useful assortment of engineering methods that seem to work, some better than others. The few existing formal models noted above are quite new and were developed to attempt to fill a void. They are a service to the community. It is our desire to continue to study the existing models, but with our new paradigm in mind. Our ultimate goal is a mathematical model of steganography that incorporates our new paradigm.

2. STEGANOGRAPHY— BACKGROUND MATERIAL

In this section we go over the standard terminology for steganography and include some simple examples.

2.1 Terminology

We will use the standard terminology for steganography as discussed at the First International Information Hiding

Workshop [19]. We assume that Alice wishes to send, via steganographic transmission, a message to Bob. Alice starts with a covermessage. The hidden message is called the embedded message. A steganographic algorithm combines the covermessage with the embedded message. The algorithm may or may not use a steganographic key (stegokey), which is similar to a cryptographic key in purpose and use — this is illustrated by using a dotted line in figure 1. The output of the steganographic algorithm is the stegomessage. The covermessage and stegomessage must be of the same datatype, but the embedded message may be of another datatype. We sometimes make the datatype explicit in our terminology, e.g., "coverimage." Figure 1 illustrates the embedding process. In steganography, we do not make the "strong" assumption that Eve has knowledge of the steganographic algorithm. This is why there may, or may not be, a stegokey involved in the embedding and extraction of a hidden message. Eve should not be able to determine from the stegomessage that there is an embedded message in it. Of course, in steganography we often make the assumption that Eve does not have access to the covermessage. Thus, Eve should not be able to tell if she is "observing" a legitimate covermessage or a stegomessage. Both Bob and Eve receive the stegomessage. Bob reverses the embedding process to extract the embedded message. In figure 2, we illustrate the extracting process.

We say that steganographic communication is *steganographically strong* if it is impossible for Eve to detect the steganography. It is the concept of "impossibility" that influences our new paradigm. Note that many authors refer to Eve as Wally, Wendy, Willy, etc. This is because the eavesdropper is often thought of as a warden due to the paper of Simmons [22]. We prefer to stick with eavesdropper since it is more general. Since the goal of this paper is to discuss the new paradigm associated with steganography, let us illustrate our thinking with some examples. There are certainly many more sophisticated and robust steganographic techniques than what we present here. We choose these two methods for (1) the historical significance of the first method, and (2) the simplicity and illustrative strength of both methods.

2.2 Kurak-McHugh Method

In 1992 C. Kurak and J. McHugh presented [14] detailing how one can hide an image inside of an image. The thrust for writing that paper was to show that one should not be too complacent about downgrading images from "private" to "public." The paper simply and graphically demonstrates that a public image that appears innocuous to a casual observer may, in fact, be hiding an embedded private image. We summarize the Kurak-McHugh method.

—— Start with a bitmapped version of a greyscale image that we wish to do the hiding in (the coverimage). Next, we consider a bitmap of the image that we wish to hide. The two images are merged into a bitmap (the stegoimage). The merging is done in the following manner. The bitmaps have one byte representing each pixel. Thus there are 256 levels of grey, ranging from 0 to 255 for each pixel. Replace the n least significant bits (LSB) of each pixel in the coverimage, with the n most significant bits (MSB) from the corresponding pixel of the image to be embedded. ——

For simplicity's sake, we assume that the coverimage and the embedded image are of the same size so that the pixels are in bijective correspondence. In [14], the authors vary n from 1 to 4 bits. We found that $n = 1$ is insufficient for preserving the quality of the original image (What we embed is often only an approximation to the original message that we wish to send. Questions of artistic quality and and what information we are actually trying to pass come into play here.) Values of $n > 2$ may cause Eve to notice that an image has been embedded. Therefore, we set $n = 2$ for discussion. Since the stegoimage differs from the coverimage by, at most, three grey levels (the two lowest bits affect the grey level anywhere from 0 (e.g., 2 LSB are (0,0)) to 3 (e.g., 2 LSB are (1,1)), it is visually impossible for Eve to detect the steganography. Of course, if Eve has knowledge of the algorithm, it is then trivial for Eve to detect the steganography.

Alice performs the embedding process as described above. The stegoimage can be passed to Bob in e-mail, or simply by posting the stegoimage on a web page. Pixel byte values must be unchanged through the storage and transmission processes. Thus, with this algorithm, a lossless method such as TIFF must be used. Note that some authors have steganographic methods that apply to methods such as JPEG, e.g., [9]. The web page approach may cause Eve the least suspicion, because Eve does not know the intended recipient of any surreptitious transmission from Alice. Bob receives the image (either through e-mail or from downloading it from the web) and then shifts every byte 6 bits to the left, thus uncovering the embedded image.

One can use the Kurak-McHugh method to deal with color images (they noted this trivial extension in their paper). Each pixel is represented by three bytes, one for each of the colors red (R), green (G), or blue (B). Every color byte is modified as for the greyscale byte. The conclusions are the same.

In terms of impact, the Kurak-McHugh paper was a huge success. If Alice is sending the stegoimage to Bob, the eavesdropper, Eve, cannot tell by looking that there is actually an embedded image hiding in the coverimage. However, is the Kurak-McHugh method steganographically strong? The answer is no.[1] Eve can determine and duplicate the stegoalgorithm and thus find the hidden picture. Can we modify the Kurak-McHugh method and make it steganographically strong? One would think that using the Kurak-McHugh method with cryptography would make the steganography impossible to detect. In fact, just the opposite is true, as we will discuss later. Accepting this causes us to rethink our paradigms about the use of noise—an important part of the new paradigm needed for steganography.

2.3 Our Text Hiding Method

There are many ways to hide text in an image. We present our own method which we feel is steganographically strong (but not necessarily robust). (Note that by using an image of text the Kurak-McHugh method would work.) We summarize the method in this paper. The full details of our

[1] Note that Kurak and McHugh never made, nor implied any such claims. This was not the purpose of their paper.

method and the underlying statistical analysis can be found in [17].

We start with the bitmap of an image. For the sake of simplicity, we will restrict ourselves in this paper to greyscale images with dimensions 500x500 pixels. Our text data is limited to 249 (ASCII) characters. Each character of the text is represented in binary form by a byte (eight bits) $(b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8)$. We use this representation to encode the text data. Each character is broken down into four sections of two bits each: (b_1, b_2), (b_3, b_4), (b_5, b_6), (b_7, b_8). We generate a list of 1000 unique random pixel coordinates and use that list as a stegokey. Each two bit section, from above, is then sequentially matched with a pixel from the stegokey. Now we mimic what Kurak-McHugh, along with many others, e.g., [13, sec. 3.2.1], have done with the popular, but non-robust LSB technique [7]. We replace the two LSB of the pixel in question with the matching two bit section. We do this for every character. We always end our text message with the null character, represented in binary as $(0, 0, 0, 0, 0, 0, 0, 0)$. This allows us to send a message shorter than 249 characters. To extract the embedded text, the algorithm is reversed. When the reverse algorithm reads the null character, it stops the extraction process. In general, the smaller the message, the harder it is for Eve to detect that there is an embedded message. This is why we change no more than $1000 = (249+1) \cdot 4$ out of the available 250,000 pixels.

3. DETECTING STEGO— PARADIGM SHIFT 1

Now that we have some simple examples to play with, let us examine the first part of our paradigm shift. In cryptography, Eve knows that there is an encrypted message. The job for Eve is to learn as much as possible about the encrypted message. In cryptography, it is not Alice or Bob's responsibility to hide their encrypted message. Rather, it is their job to make the message unintelligible to Eve, even if Eve may be able to bring large amounts of computational resources to bear upon the problem. Shannon modeled secrecy based upon probabilities and information theory. Perfect secrecy is achieved if the ciphertext and the plaintext are statistically independent. Mathematically, Shannon [21] expressed this as: Given finite messages M that are enciphered into possible cryptograms E we say that perfect secrecy is achieved iff $\forall E, \forall M, P(M|E) = P(M)$. This is a "yes or no" situation. However, in cryptography less than perfect secrecy is of great interest. This is very different than steganography (and this is the first part of our new paradigm).

3.1 The Wire-Tap Channel

Wyner [24] first described a simplified eavesdropper scenario in cryptography in terms of a wire-tapper Eve, listening in on Alice and Bob [10, 8]. Alice's transmission to Bob may be noisy, and Eve's tapping also has noise in it. Alice wishes to send k source bits S^k which are encoded into n symbols though a noisy discrete memoryless channel channel X. Bob receives Y^n from the channel and Eve taps Z^n out of the channel. Both $X \rightarrow Y$ and $X \rightarrow Z$ have their noise characteristics modeled by the joint conditional probability $P_{Y,Z|X}$. Based upon what Bob receives, he "estimates" what S^k was. Alice wishes for this estimate to differ, in probability, from S^k as little as possible. This is the probability of error. However, Eve is learning information about what Alice transmitted. This is measured as the normalized conditional entropy as $\Delta = \frac{H(S^k|Z^n)}{H(S^k)}$. If Eve can determine without question what Alice sent, based upon what Eve received, then all probabilities are zero or one, and therefore $H(S^k \mid Z^n) = 0$, and $\Delta = 0$. This is the worst case in terms of secrecy. If Eve learns nothing about the distribution of S^k from knowing Z^n, then the two are statistically independent and Δ is maximized at the value 1. This is the best in terms of security. However, pragmatically secure communication can be done between Alice and Bob even when $\Delta = 1 - \epsilon$, ϵ small. In contrast, in steganography, there is no such thing as "almost does not know there is a hidden message." Therefore, the wire-tap model differs greatly for steganography. We must call our thinking into question when it comes to things like ϵ-security.

Of course Δ, is very similar to *unicity distance* [21], [15, secs. 7.2 & 7.3] which is expressed also as a normalized entropy. This measures how much plaintext can be revealed without enabling decryption of the entire ciphertext. This is not the case in steganography. The use of a normalized entropy must be called into question when it is an either/or situation, as it is in steganography.

3.2 Existing Steganographic Models

Consider the above scenario, but substitute steganography for cryptography. Let Δ again represent the amount of "information" that Eve can learn through eavesdropping.

- Should we still use an entropy-based measure? Entropy works well for cryptography. But is it the appropriate measure for steganography?
- How should one interpret Δ? Should anything other than boundary values for Δ be useful? Non-boundary values are useful for cryptography, where we are willing to live with less than perfect secrecy, but this is not the case for steganography.

To the best of our knowledge, all existing steganographic models are based upon a paradigm of entropy/information theory (which has continuous probability theory as its underlying core principle). Of course, the above wire-tapping scenario does not map exactly into a steganographic problem. Consider figure 1: Let C be a random variable representing the covermessage, E a random variable representing the embedded message, and S the random variable representing the stegomessage. The idea is that, statistically, the stegomessage should appear to be similar to a covermessage. Differences in statistical profiles, or conditional entropies, would alert Eve that there is an embedded message. What concerns us is that the prevailing paradigm assumes that probability distributions can be assigned to the set of legitimate cover messages. We would like to see more published work on how these distributions are actually assigned. Also, the existing paradigm does not include the idea of "spontaneous discovery." That is once Eve knows that there is hidden information, the game is over. Of course, we can get into a discussion (not in this paper) of what "knows" means. Obviously in the Kurak-McHugh method, Eve is definite in

her knowledge. The process of obtaining this knowledge might very well be a continuous process (such as hypothesis testing). What is not acceptable is the idea of a "little bit discovered." This of course is different than the acceptable idea (and what the existing models use) that if one knows that all messages under consideration have a given non-zero probability of containing a hidden message that it is then appropriate to discuss subtle differences in that probability. This distinction in approach must be drawn out.

In [6] Cachin uses the discrimination (relative entropy) $D(C \mid S)$ between the distributions C and S to define ϵ-security against a passive (just listening in) Eve; the stegosystem is ϵ-secure against a passive Eve iff $D(C \mid S) \leq \epsilon$. When the discrimination is zero, then the stegosystem is perfectly secure. We take issue with the concept of ϵ-security *in general* (not necessarily with how it was used in [6]). Is this the proper way to be thinking about steganography? Does ϵ-security mean that you have some knowledge that there is a hidden message, or does it mean that the odds have shifted by ϵ that there is a hidden message? Cachin nicely ties ϵ-security into hypothesis testing (detects a hidden message). However, we still feel that a continuous slide from perfection to detection is questionable. Perhaps there is a deeper concept describing this change that is not continuous. However, to defend [6], one must keep in mind that the purpose of this paper is to define a concept of steganographic security/insecurity when one has the ability to assign probabilities to what a legitimate cover might be. The author himself expresses the need for "caution."

In [11] Ettinger takes a game theoretic approach to detecting the steganography. A permitted "distortion" is allowed. This permitted distortion is allowed under the concept of "a distribution of locations." Is it possible for Eve to increase her computational efforts so that what was acceptable before is no longer acceptable? Is discovery not just a "yes/no" proposition? We must think about how and when to apply such a model. The formalism of all of the existing models seems to be correct only under the ability to assign distributions for what is a legitimate cover. (Note that the authors of those papers make no further claims.)

In [25] Zollner et. al. use conditional entropies to show that it is impossible to have any sort of steganographic security if Eve has knowledge of both the covermessage and the stegomessage. Without all of the fancy math, this boils down to the fact that Eve can compare covermessages and stegomessages and see that something is amiss. This is why all stegosystems are modeled with Eve only getting her hands on the stegomessage. The authors then go on to show that there must be uncertainty in the covermessage, or Eve could always tell if she had a stegomessage or a covermessage. Underlying this paper is, we feel, the *all or nothing* idea that we wish to pursue as part of our new paradigm. However, the emphasis of [25] is the need for *indeterminacy* in the set of covermessages in order to obfuscate Eve, a point that they make well!

In [2, 3] the authors discuss the appropriateness of using an information theoretical approach for modeling steganography. They discuss how Eve's computational power could influence such a model, and also consider some upper bounds for hidden information. A parallel to a one-time pad is discussed, as it also is in [6].

In [16] Mittelholzer discusses a perfect steganography scenario in light of issues of steganographic robustness — an important topic in digital watermarking. Mittelholzer also includes watermarking in his model. Even though watermarking is part of the larger field of information hiding, it is not identical to steganography. For example, in watermarking the fact that a digital watermark has been embedded in a covermessage is often a public fact. This is orthogonal to steganography. Therefore, *we* find it difficult to follow a model that attempts to incorporate both steganography and watermarking.

In cryptography, a small amount of discovery is allowed. In steganography, a small amout of discovery is not allowed. It is our desire to find/design a formal model that explicitly shows that partial discovery is not allowed. Of course, uncertainty in discovery is allowed (e.g., indeterminacy). This uncertainty in discovery can be expressed probabilistically, provided that one can show that distributions can be assigned.

3.3 Covert Channels

We note that the existing paradigm for covert channels is not appropriate for steganography. Steganography can be thought of as a subliminal channel. Simmons was the first to use the term *subliminal channel* in a general sense in [22]. A subliminal channel is a secondary communication between two parties Alice and Bob, such that the primary communication is publicly known, but the secondary communication is meant to be hidden. A covert channel differs in that there is communication between Alice and Bob that exists outside of the system design. A covert channel is allowed to exist if its information theoretical capacity is below an agreed-upon upper bound. This does not, and should not, work for steganography. Once Eve knows that there is hidden communication, the subliminal channel has been discovered. There is no such thing as partially subliminal, which is similar to the concept of being a little bit pregnant. The paradigm of covert channels, the old paradigm, is similar to that of cryptography, also the old paradigm. Steganography (subliminal channels) must have a new paradigm that does not include such distinctions as a little bit discovered (non-hidden)! However, steganographic models do rely upon the fact that one can be a little bit confused—through the indeterminacy of what is a legitimate cover.

3.4 Comments

All of the above models are important and of interest. They have their various strengths and weaknesses, depending upon what aspect of steganography one is attempting to model. At present, the community has yet to agree upon one model or approach as the definitive one. We wish to discuss how a system transitions from successful steganography to unsuccessful steganography. This transition is very different from that of cryptosystems or of "safe" covert channels. This is the first part of our new paradigm (noise being the other). Our ideas are raw and in need of refinement. We enjoyed the workshop participants' feedback.

3.5 Lack of Steganography— New Paradigm shift 1

In our view, steganographic communication exists when and only when Eve is not cognizant of the hidden message. Acceptable regions of indecision should only be allowed under the cloud of indeterminacy. The fact that one does not have the proper tools to detect the steganography should not be part of a formal model. Once Eve has any evidence that there is hidden information, the steganography has failed. This is a discontinuity. This is not to say that the underlying process may not, in fact be continuous. As in [6], it might be some sort of hypothesis that is accepted that causes Eve to detect the hidden communication. However, it is not, as in the wire-tap channel, a case where some amount of information is allowed to be leaked. This may not happen in steganography! The first part of our *new paradigm* is:

In steganography, the discovery of hidden information is not modeled in a continuous manner. We must readdress our old paradigms for secure systems to deal with discontinuities. Standard information theoretical models do not deal with "jumps."

The idea of a discontinuity arising from a (perhaps) continuous process had disturbed us for quite a while. It was when we started investigating the much-maligned field of mathematics called "Catastrophe Theory" [23] that we started to get a feel for how to approach modeling our new paradigm. A successful and complete model of steganography should deal with jump discontinuities.

Consider the polynomial $y = (x-3)^2 + \delta = x^2 - 6x + (9+\delta)$. This is a simple quadratic whose graph is a parabola with the minimum value of δ achieved when $x = 3$. In figure 3, we show the plots for three values of δ: $\delta = -1, 0, 2$. Note that the quadratic has two roots when $\delta = -1$, one root when $\delta = 0$, and no real roots when $\delta = 2$. This phenomenon is expressed in general in figure 4; here we plot the number of real roots against δ. Note that even though δ increases in a continuous manner, the number of real roots (intersections with the x-axis) has a discontinuity at zero (non-removable singularity). This simple example shows that a continuous natural event might have some features acting in a discontinuous manner, and any attempts to model those features in a continuous manner are contrary to the will of nature— this relates quite strongly to our new paradigm. We must call the old ways of thinking into question and look for new methods with which to model steganographic systems

3.5.1 Catastrophe Theory:

Catastrophe theory was developed in the 1970's by the great French mathematician Rene Thom [23]. In some sense, catastrophe theory was the unsuccessful precursor to chaotic dynamical systems. As the name implies, catastrophe theory models discontinuities in a system's behavior, e.g., when does a dog decide to bark, what is the difference between genius and insanity, when does the bubble burst on Internet stocks, etc.? In short, it shows how discontinuities can describe certain aspects of continuous natural systems, which is a scenario quite like what we have described with steganography. In figure 5, we see the plot of the parametrized surface

$$(r, \theta) \rightarrow (r\cos(\theta), r\sin(\theta), \frac{r}{2\pi}\theta), \qquad r \in [0,1],\ \theta \in [0, 2\pi]$$

(Note: This is similar to the Riemann surface of $\log(\zeta)$) The mathematics describing figure 5 are not important. Rather, the importance lies in its interpretation. Our example is motivated by Arnol'd's example [4, p. 7-8] of the "technical proficiency-enthusiasm-achievement" scientist. Note that standard illustrations of catastrophe theory often use "folded" surfaces — for simplicity we just stay with "cut" surfaces. Our interpretation of figure 5 is of the skill of a mathematician. The upper most regions of the surface represents *genius*, the middle *normal*, and the bottom *pre-algebraist*. Think of 3-dimensional space R^3 with coordinates (r, θ, z). The coordinate r is ability, θ is effort, and z is mental state (we do not intend for this example to be an exact representation of what makes up a mathematician's skill—it is for illustrative purposes only). When we project down to the polar plane, we arrive at figure 6. In other words, when we only have a partial view of the mathematician's skill, it seems that there is a discontinuous jump from pre-algebraist to genius, which is a view that many have of mathematicians. This is the same behavior when we looked at the roots in the previous example. It is not an exact match, but the ideas are very similar. What we see from this example is that depending upon how one views a physical system or phenomenon, it may appear discontinuous.

We are presently investigating catastrophe theory to see if it can be used as a model for steganography. One must move carefully when using catastrophe theory. Many think of it as the cold fusion of modern mathematics. However, the underlying mathematics are sound, it is the applications that must be carefully examined. It is our opinion that the new paradigm that started with catastrophe theory laid the foundations for the 1980's rage in chaos and fractals. Steganography must use a new paradigm that includes discontinuous jumps. Reliance upon the old paradigms of entropy must be examined. Now we will discuss the second part of our new paradigm.

4. NOISE IS BAD FOR STEGANOGRAPHY— PARADIGM SHIFT 2

One can achieve perfect secrecy in cryptography by using a one-time pad. If Eve intercepts the encrypted transmission, what she gets is total noise (of course this is only true if the random number generator behaves properly). This is the best that one could hope for with respect to cryptography. This old paradigm must be examined when it comes to steganography. We thought that we could use (white) noise to assist in steganography. We found that we were wrong. This was the paradigm that we took from cryptography, and the paradigm that must be changed. In retrospect, we see that the old paradigm is obviously wrong when it comes to steganography. However, we had to learn our lesson. Note that we know of no existing models of steganography that advocate "white noise." We bring up the issue to show how different steganography is from cryptography. Hopefully, expressing the second part of our new paradigm will cause others not to erroneously think the same way that we un-

fortunately did (at first). In retrospect, it seems obvious. However, one can use noise, but in a controlled manner. The noise must imitate what noise a legitimate coverimage would have. Thus, we get back to the idea of some sort of indeterminacy which is a linchpin of the existing steganographic models.

In steganography, the use of noise may make things worse, not better. One can use the inherent noise in a covermessage, but adding additional noise may cause the steganography to be discovered.

4.1 Kurak-McHugh—again?

One can easily adjust the Kurak-McHugh method to not let Eve know what the embedded image is, even if Eve has determined that there is an embedded image. Thus, we can achieve cryptographic security when the steganography has failed. Simply encrypt the embedded bits so that the 2 LSB in the stegoimage appear as white noise. By white noise, we mean that the 2 LSB are statistically equivalent to having each pixel's 2 LSB randomly and independently generated from a uniform draw of the (decimal) values 0,1,2,3. We use Blowfish [20] to do this as follows.

The 2 MSB of each pixel of the embedded image are saved into an array which is encrypted using Blowfish in Cipher Block Chaining (CBC) mode. The encryption key is a 16 byte MD5 hash of a passphrase. The encrypted array is then stored, two bits at a time, replacing the 2 LSB of each pixel in the coverimage, thus forming the stegoimage. The embedded hidden image is recovered by a reversal of this process. The 2 LSB of each pixel are saved to an array, which is decrypted using Blowfish CBC with the decryption key being equal to the encryption key. The decrypted array is then used, two bits at a time, to form the 2 MSB of each pixel in the recovered hidden image.

Even though the above approach keeps the hidden image (ignoring the 6 LSB) cryptographically secure it does not keep the hidden image steganographically secure. This is extremely important! Our experiments have shown that there are "artifacts" residing in the 2 LSB. This is independent of what image type (JPEG, TIFF, PNG, etc.) the original image was before we realized its bitmap. We discuss this below. Not all images that we used had these artifacts, but most did.

The effect that we demonstrate seems to hold, irrespective of the file type the image is. Figure 7 is the bitmapped version of a TIFF file. Figure 8 is the bitmap when we move every byte (R byte, G byte, B byte for each pixel) from figure 7, six places to the left. This forces the 2 LSB from figure 7 to become the 2 MSB, and all of the other bits making up the byte to become zero. One can easily see that the bright spots from figure 7 leave very visible artifacts upon the lower bit planes. Thus, to use cryptography to enforce steganographic robustness would force the encryption to mimic the artifact pattern both visually and at the more complex statistical level. However, when we attempt to embed the 2 MSB of figure 9 into figure 7 by encrypting as above and resulting in figure 10, and then shift the bits left 6, we are left with figure 11, which is white noise. Thus, it is obvious that something is "wrong" with figure 10. Therefore, using cryptography without mimicking the artifact pattern of the coverimage lets Eve know that there is am embedded image in the coverimage. We do not know how to force the encryption to mimic the artifact pattern. This seems to be quite complex. Note, of course, that after decrypting the 2 MSB as given in figure 10, we have the 2 MSB representation of figure 9 as shown in figure 12.

4.2 Discussion — New Paradigm shift 2

From the above we note that adding totally random white noise is exactly the *wrong* thing to do with respect to steganography. In the example given above, Eve can easily, through trivial statistical tests, determine that there is something "fishy" with respect to the 2 LSB. Most legitimate images would not have the 2 LSB appear as white noise. Therefore adding noise to increase security — the old paradigm from cryptography — fails miserably here. The noise must be added in a manner consistent with the coverimage. This is not to say that all present models and techniques of steganography ignore this thinking. Our goal, rather, is to emphasize the difference between the paradigms of cryptography with those of steganography. This is non-trivial and is part of our current research.

5. CONCLUSION

We have shown how two staples of cryptography: a continuous information theoretic-based foundation, and the use of noise, should not be staples for steganographic modeling. Steganographic models must contain some way of dealing with (catastrophic) jumps from not knowing, to knowing, that there is hidden information. We have shown that this type behavior is possible in other continuous physical/mathematical systems. Therefore, we feel it is imperative to incorporate it into steganographic models. Adding noise during the steganographic embedding phase can cause the steganography to fail. The transition from a covermessage to a stegomessage must be carefully done so that Eve does not know that the covermessage has been tampered with. In cryptography, one need not hide the fact that a message has been encrypted. However, in steganography one must hide the fact that a message has been embedded. Since the philosophies of the two are so different, so should the guiding paradigms be different.

6. ACKNOWLEDGEMENTS

We appreciate the helpful comments from the reviewers, R. Heilizer, A. Pfitzmann, and the workshop participants.

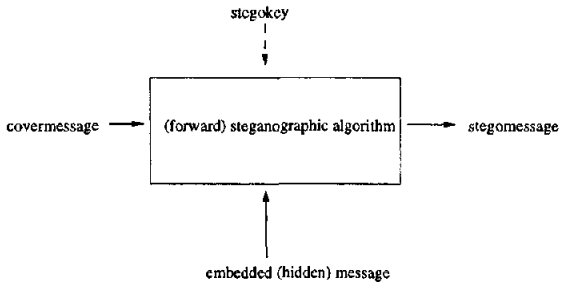

Figure 1: Embedding the hidden message

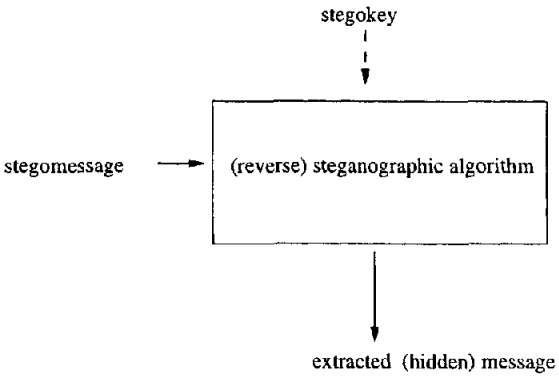

Figure 2: Extracting the hidden message

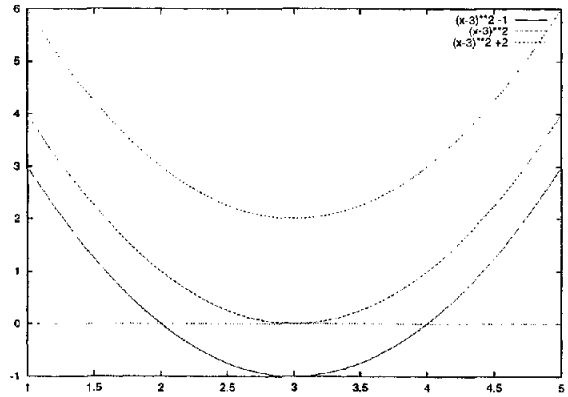

Figure 3: Real roots

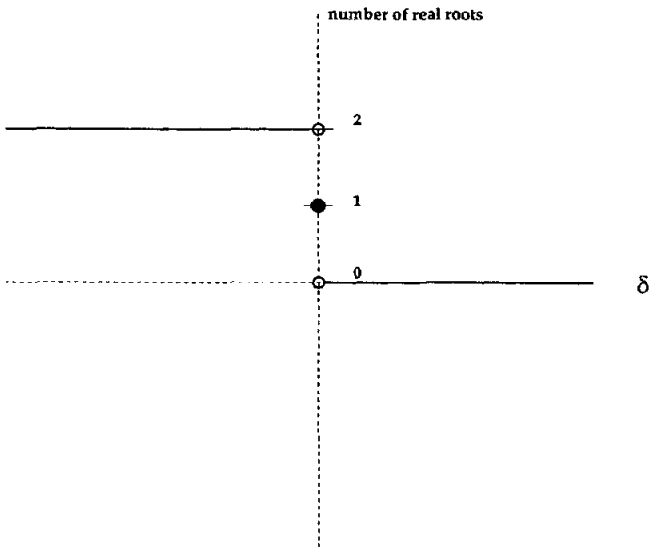

Figure 4:Discontinuity

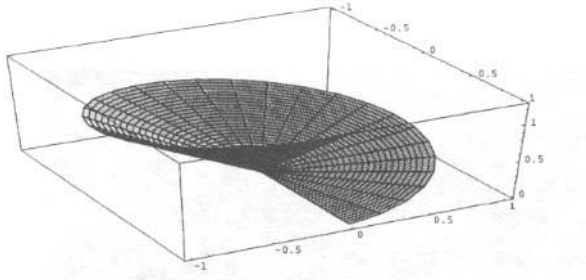

Figure 5: Parametrized surface in R^3 of mathematician's skill

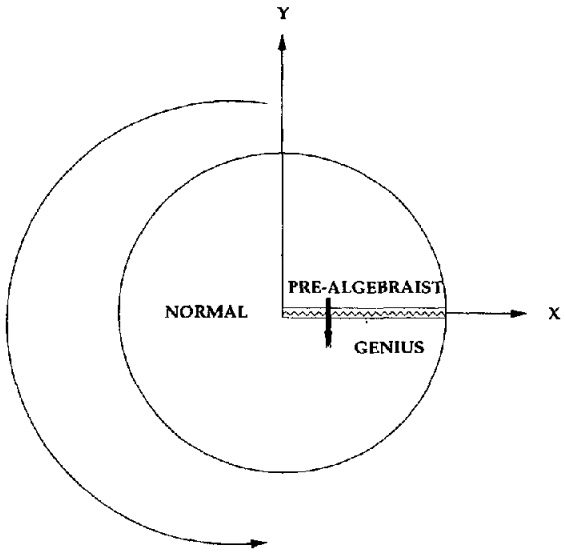

Figure 6: Mathematician's skill with hidden variable

Figure 7: Coverimage

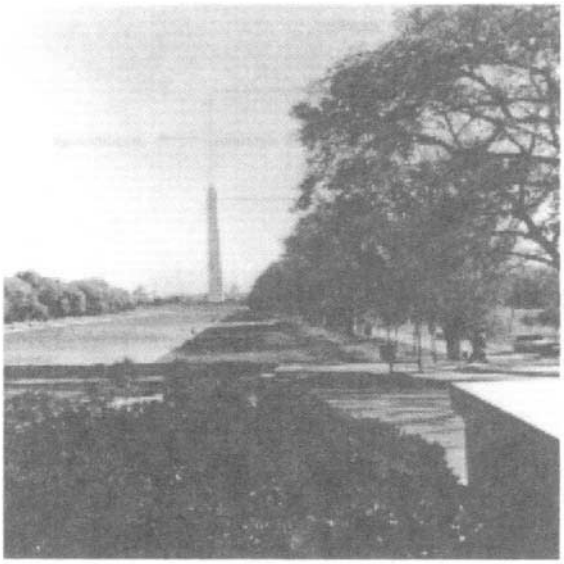

Figure 9: Image to be embedded

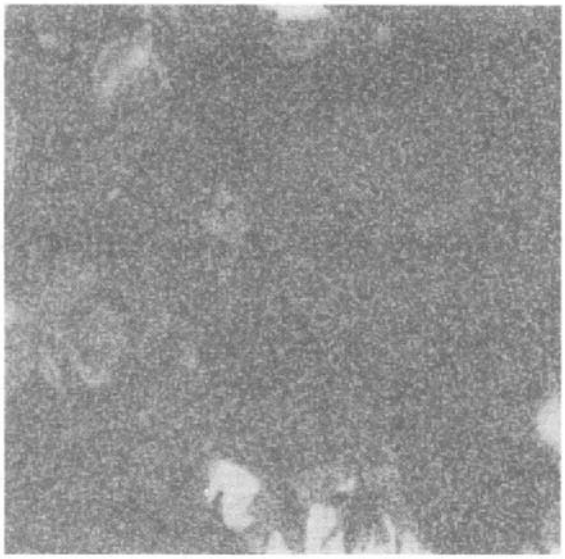

Figure 8: Coverimage (shifted 6 bits to the left)

Figure 10: Stegoimage

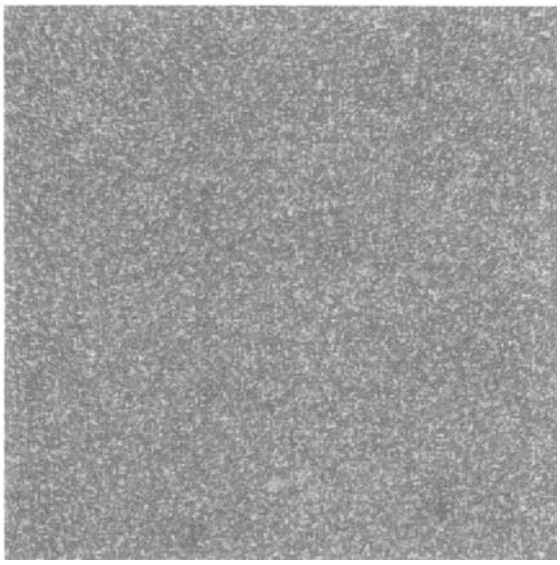

Figure 11: White noise

Figure 12: Recovered extracted image

7. REFERENCES

[1] R.J. Anderson, editor: *Information Hiding: First International Workshop*, vol. 1174 of *Lecture Notes in Computer Science*. Springer-Verlag, 1996.

[2] R.J. Anderson: *Stretching the Limits of Steganography*, In R. Anderson, editor, *Information Hiding: First International Workshop*, vol. 1174 of *LNCS*, pp. 39-48. Springer-Verlag, 1996.

[3] R.J. Anderson and F.A.P. Petitcolas: *On The Limits of Steganography*, IEEE Journal of Selected Areas in Communications, 16(4), pp. 474-481, May 1998.

[4] V.I. Arnol'd: *Catastrophe Theory*, Third, Revised and Expanded Ed., Springer-Verlag, Berlin, 1992.

[5] D. Aucsmith, editor: *Information Hiding: Second International Workshop*, vol. 1525 of *Lecture Notes in Computer Science*. Springer-Verlag, 1998.

[6] C. Cachin: *An Information-Theoretic Model for Steganography* In D. Aucsmith, editor, *Information Hiding: Second International Workshop*, vol. 1525 of *LNCS*, pp. 306-318. Springer-Verlag, 1998.

[7] L. Chang and I.S. Moskowitz *Critical Analysis of Security in Voice Hiding Techniques* In Y. Han, T. Okamoto, and S. Qing, editors, *Information and Communications Security: First International Conference*, vol. 1334 of *Lecture Notes in Computer Science*, pp. 203-216, Springer-Verlag, 1997.

[8] I. Csiszar: *Broadcast Channels with Confidential Messages* IEEE Transaction on Information Theory, V. IT-24, No. 3, pp. 339-349, May 1978.

[9] D.L. Currie, III and C.E. Irvine: *Surmounting the Effects of Lossy Compression on Steganography*, In *National Information System Security Conference*, Baltimore, MD, pp. 194-201, October 1996

[10] M. van Dijk: *On a Special Class of Broadcast Channels with Confidential Messages*, IEEE Transactions on Information Theory, V. 43, No. 2, pp. 712-714, March 1997.

[11] J.M. Ettinger: *Steganalysis and Game Equilibria* In D. Aucsmith, editor, *Information Hiding: Second International Workshop*, vol. 1525 of *LNCS*, pp. 319-328. Springer-Verlag, 1998.

[12] D. Kahn: *The History of Steganography* In R. Anderson, editor, *Information Hiding: First International Workshop*, vol. 1174 of *LNCS*, pp. 1-6, Springer-Verlag, 1996.

[13] S. Katzenbeisser and F. Petitcolas, editors: *Information Hiding Techniques for Steganography and Digital Watermarking*. Artech House, 2000.

[14] C Kurak & J. McHugh: *A Cautionary Note on Image Downgrading* In *Computer Security Applications Conference*, San Antonio, TX, USA, pp. 153-159, Dec. 1992.

[15] A.J. Menezes, P.C. van Oorschot, & S.A. Vanstone: *Handbook of Applied Cryptography* CRC Press, Florida, 1997.

[16] T. Mittelholzer: *An Information-Theoretic Approach to Steganography and Watermarking* In A. Pfitzmann, editor, *Information Hiding: Third International Workshop*, vol. 1768 of *LNCS*, pp. 1-16. Springer-Verlag, 2000.

[17] I.S. Moskowitz, G.E. Longdon, & L. Chang: *A Method of Steganographic Communication* In Preparation, 2000.

[18] A. Pfitzmann, editor: *Information Hiding: Third International Workshop*, vol. 1768 of *Lecture Notes in Computer Science*. Springer-Verlag, 1999.

[19] B. Pfitzmann: *Information Hiding Terminology* In R. Anderson, editor, *Information Hiding: First International Workshop*, vol. 1174 of *LNCS*, pp. 347-350. Springer-Verlag, 1996.

[20] B. Schneier: *Description of a New Variable-Length Key, 64-Bit Block Cipher (Blowfish)*, In R. Anderson, editor, *Fast Software Encryption, Cambridge Security Workshop Proceedings*, vol. 809 of *LNCS*, pp. 191-204. Springer-Verlag, 1994 (Blowfish implementation written by Eric Young) .

[21] C.E Shannon: *Communication theory of Secrecy Systems* Bell System Technical Journal, Vol. 28, pp. 656-715, 1949.

[22] G. Simmons: *The Prisoners' Problem and the Subliminal Channel* In D. Chaum, editor, *Advances in Cryptology: Proceedings of Crypto 83*, pp. 51-67. Plenum Press, 1984.

[23] R. Thom: *Structural Stability and Morphogenesis*, W.A. Benjamin, Reading, MA, (French Ed. 1972) 1975.

[24] A.D. Wyner: *The Wire-Tap Channel* The Bell System Technical Journal, V. 54, No. 8, pp. 1355-1387, October 1975.

[25] J. Zollner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, & G. Wolf: *Modeling the Security of Steganographic Systems* In D. Aucsmith, editor, *Information Hiding: Second International Workshop*, vol. 1525 of *LNCS*, pp. 344-354. Springer-Verlag, 1998