Digital Steganography for Information Security

Anthony T.S. Ho, Siu-Chung Tam,
Siong-Chai Tan and Lian-Teck Yap

DataMark Technologies
Suite 106, Innovation Centre,
Nanyang Avenue,
Singapore 639798

Tel: (65) 793 7725
Fax: (65) 793 7790
Email: datamark@datamark-tech.com
Web: http://www.datamark-tech.com

Kok-Beng Neo and Sim-Peng Thia

CET Technologies
Singapore Technologies Building
100 Jurong East Street 21
Singapore 609602

Tel: (65) 567 6769
Fax: (65) 567 6300
Email: neokb@cet.st.com.sg
Web: http://www.cet.st.com.sg

ABSTRACT

The phenomenal growth in e-commerce applications through the Internet in the past few years has led to a genuine need, as well as a sense of urgency, for both small office and home office (SOHO) and corporate users to protect their data transactions through the Internet. These data transactions may include sensitive document transfer, digital signature authentication, digital watermarking for copyright protection, and digital data storage and linkage.

In this paper, the use of digital steganography for information security in various e-commerce applications through the Internet will be discussed in detail. These applications include digital watermarking for copyright protection of multimedia data, digital signature authentication and validation of electronic documents, digital data storage and linkage for binding digitized photographs with personal attribute information, as well as secure communication of multimedia data through the open channels. Enhanced information security will lead to wider e-commerce applications that involve e-communication, e-transactions, e-filing, and e-publications.

INTRODUCTION

The conventional way of securing data transactions is through the use of standard encryption key techniques such as RSA, DES, and 3DES. In the past three years, however; a new branch of data security techniques known as digital steganography has evolved and is continuing to receive a great deal of attention from both the academic and industrial communities [1,2]. Since then, a number of companies have established in the US and Europe to commercialize and market steganography products. Many of these steganography products were developed as plug-ins and OEM applications.

Instead of scrambling the data using either a standard symmetric or asymmetric key system as in the case of encryption, digital steganography exploits the use of a host data or message (also known as a container) to hide or embed another data or message into it. Unlike encryption, the host data or container used in steganography is not scrambled or hidden during the communication process. Only a hashed form of the hidden data derived from a mathematical combination of the host and hidden data is transmitted for decoding.

The word "steganography" actually stems from a Greek word meaning "covered writing". Some common analogies to the digital form can be found in nature. For example, a leaf insect (hidden message) exploits the natural surrounding of leaves (host message or container) to camouflage itself. Or an infantry soldier camouflages himself by painting and covered himself in colours that match its surrounding. One of the successful revolts of Han Chinese against the Mongols during the Yuan dynasty also exploited the use of steganography. In this scenario, the Chinese took advantage of the Mid-Autumn festival by distributing moon cakes (container) with a message inside (hidden data) to their members informing them of the planned revolt.

DIGITAL STEGANOGRAPHY

The applications of digital steganography in various e-commerce applications through the Internet will be discussed in detail. These applications include digital watermarking for copyright protection of multimedia data, digital signature authentication and validation of electronic documents, digital data storage and linkage for binding digitized photographs with personal attribute information, as well as secure communication of multimedia data. Targeting these applications, DataMark Technologies (DMT) have developed four digital steganography products based on their patent-pending algorithms [3,4,5], as follows:

1. Secure Communication (StegComm™)
2. Digital Signature Authentication (StegSign™)
3. Digital Watermarking (StegMark™)
4. Digital Storage and Linkage (StegSafe™)

1. SECURE COMMUNICATION

StegComm™ is a state-of-the-art digital steganography software package developed by DMT for confidential multimedia communication. The software allows the user to select a multimedia data file or "container" for embedding hidden text, audio sequence, video clip, or any form of data file. Figure 1 illustrates the basic concept of digital steganography when applied to text encoding. The contents of the text message are hashed with those of the container file to produce a key file. The key file is also known as a “Stegfile”.

Many conventional steganography techniques simply incorporate a combination of cryptography and steganography. The cryptography operation is used first to scramble the hidden text. For steganography operation, the scrambled data is then inserted or "hidden" into the least significant bits (LSB) of the container data. One of the common drawbacks of these techniques is that the container file has to be of certain size greater than the hidden file. Other limitations include the knowledge required on the exact location of the hidden text, the limited container data formats, and the export restriction of using encryption algorithms to certain countries. These difficulties are circumvented by the use of StegComm™. First, StegComm™ utilises a patent-pending lossless algorithm (the HTTY algorithm) that does not affect the data integrity of the container file. Second, the program is completely independent of the size of the container file relative to that of the hidden file. Third, as steganography is a relatively new field, there are currently no export restrictions on products that incorporate this technology.

Another key advantage of the lossless algorithm is the option to select any digital data file from a webpage on the Internet. As the algorithm does not corrupt or overwrite the container file, multimedia data posted on any webspage, such as images (JPEG, GIF), video clips (AVI, MPEG) or audio files (WAV, MIDI), can be selected as the container file. Furthermore, customised container files, such as the voices and images of the sender captured via video conferencing, can be generated very easily. Therefore, the probability of knowing which container file used during encoding is infinitesimally small. It is almost like "finding a needle in a haystack."

The operations involved in using StegComm™ are illustrated in figure 2. A multimedia container file is first chosen from the PC hard disk or from a webpage on the Internet. The knowledge of this container file must be pre-determined and communicated securely between the sender and receiver. The algorithm generates a hash file or stegfile from the inputs of the container file and the hidden text. The stegfile contains random data based on a number of mathematical operations between the two input files.

The random data bears no data resemblance to either the container or the hidden file. For example, if a hacker were to intercept this stegfile and perform his normal decoding analysis on the data, without the knowledge of the container file, it is virtually impossible for him to decode the stegfile. The hidden file can therefore only be decoded if both the container and the stegfile are available at the receiver end. Figure 3 illustrates a graphical user interface (GUI) for StegComm™.

StegComm™ is currently being marketed in two product versions: Standard and Professional. For some corporate companies, such as banks and financial institutions, as well as government agencies, where data security is of paramount importance, the Professional version offers an additional layer of security by incorporating an encryption solution, such as DES or 3DES, to the stegfile prior to open channel communication. Passwords for both container file and stegfile are also available in the Professional version. However, for SOHO and home users, the Standard version is more than adequate for their day-to-day needs in secure data communication.

2. DIGITAL SIGNATURE AUTHENTICATION

StegSign™ is a software product specifically developed by DMT to prevent malicious tampering of private and confidential documents. These documents include company memos, Emails and letters. StegSign™ can provide a wide spectrum of applications in the e-commerce sector. Such e-commerce applications include business transactions between banks and customers, legal document exchanges between lawyers and clients, and scenarios involving non-repudiation issues. This product will detect any unwarranted tampering and alert the receiver side immediately.

StegSign™ incorporates patent-pending algorithms filed by DMT. A digital signature and a multimedia container password are embedded into the confidential document. The basic operation of StegSign™ is illustrated through a data flow diagram as shown in figure 4. The digital signature can be inputted as a handwritten signature or as a personal seal. The container password can either be a normal text string, an image, or a binary file. For the Professional version, encryption is available to provide another added layer of security for the "signed" document. More mathematical random lock combinations for data embedding are also included in the Professional version. Figure 5 illustrates a typical GUI sample from StegSign™.

3. DIGITAL WATERMARKING

StegMark™ is a digital watermarking software for copyright protection of digital images, music CDs, DVDs, and other forms of multimedia data. In the case of digital images, the files can come from a variety of sources, such as the Internet, digital still cameras, and video cameras. Many digital watermarking techniques in the market embed only a certain number of bits or characters into the image. However, StegMark™ can embed either a text or image watermark invisibly into an "unlabelled" image. The text watermark can be of many characters, for example, for a colour image of size 512 x 512, more than a few thousand characters may be embedded.

The image watermark technique of StegMark™ is currently the only digital watermarking product available in the market that offers the embedding of a company's logo/trademark into an image. For a 512 x 512 image, an image watermark of size up to 128 x 128 can be embedded entirely into the image, without the loss of image integrity. This unique "image-in-image" watermarking technique has already been filed for an international patent and is currently under pending status. The performance of any watermarking

technique is essentially a direct tradeoff between image integrity and robustness of the watermark. Figure 6 illustrates the data flow process of the StegMark™ product. A GUI of StegMark™ is shown in figure 7.

How robust a watermark is depends on whether it can survive various "attacks" that include contrast changes, cropping, scratches, and filtering. However, the image integrity of the "labelled" image must not degrade poorly from an increased level of robustness to these attacks. There are currently many exaggerating claims that some watermarking techniques can survive all kinds of image manipulation attacks. However, many of these attacks will destroy the watermark, simply because the labelled image values with the embedded watermark have now been significantly modified. Obviously, depending on the watermarking techniques, some attacks can be defended more successful than others.

StegMark™ has been tested repeatedly with a number of image attacks that included contrast stretching (reduction and sharpening), pixel defects, low pass and high pass filtering on the image-in-image watermarking technique. The image watermark is able to survive most of these attacks. Although some of the image watermark pixels were affected, the overall structure of the watermark remained intact and could still be recognized. The recognition of a watermark is an important issue as it can be used in the court of law to defend the true ownership of the intellectual property.

4. DIGITAL LINKAGE AND STORAGE

StegSafe™ is the latest of DMT steganography products that provides a secure data linkage between a digital image and attribute text information. The attribute information can be any personal records such as employee details, hospital patient medical records, or law enforcement records. Currently, many personal records with ID photographs are manually or electronically filed. Tampering to these records, such as changing the name, photograph, or medical conditions, can be performed if the hacker is able to gain security access to the database.

The main function of StegSafe™ is to securely link the personal record and digital photograph together and then create a hash file that can be safely stored in a database. This hash file is unique to and can only be decoded with the original photograph and associated personal record. Tampering with any one of these files will render the decoding process ineffective. The database administrator will be able to determine whether these files have been modified, by checking the original hash file with the digitized photograph. An optional password is also available to protect the hash file prior to data storage. The basic operation of StegSafe™ is illustrated in figure 8 and a GUI sample of this product is shown in figure 9.

CONCLUSIONS

The use of digital steganography for Infosecurity in various e-commerce applications through the Internet has been discussed in detail in this paper. These applications include digital watermarking for copyright protection of multimedia data, digital signature authentication and validation of electronic documents, digital data storage and linkage for binding digitized photographs with personal attribute information, as well as secure communication of multimedia data through open channels. Digital steganography can provide one of the safest and unrestricted information security tools in the market, and is poised to advance the pace of growth of e-commerce applications in Singapore and beyond.

REFERENCES

1. Memon, N. and Wong, P.W., "Protecting Digital Media Content", Communications of the ACM, Vol. 41, 7, July1998

2. "WWW References on Multimedia Watermarking and Data Hiding Research & Technology", http://www-nt.e-technik.uni-erlangen.de/%7Ehartung/watermarkinglinks.html
3. Ho, A.T.S., "Method and Apparatus for Camouflaging Data", PCT/SG98/00023, 18 March, 1998
4. Ho, A.T.S. and Tam, S.C., "Methods for Embedding Image, Audio and Video Watermarks in Digital Data", PCT/SG98/00039, 1 June, 1998
5. Ho, A.T.S., Tam, S.C., Tan, Siong Chai, and Yap, Lian Teck, "Methods of Digital Steganography for Multimedia Data", SG9803458-0, 28 October, 1998

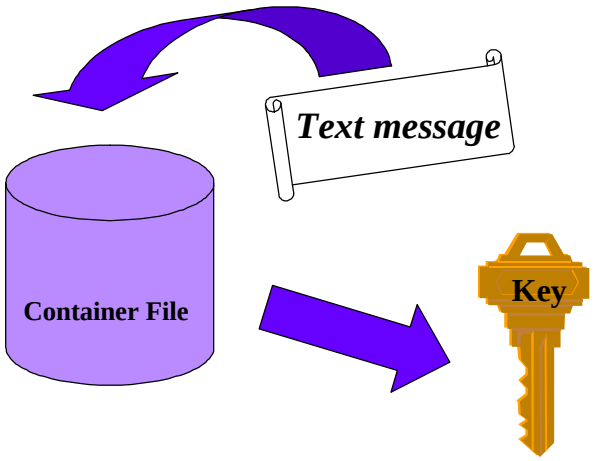

Figure 1 illustrates the basic concept of digital steganography when applied to text embedding

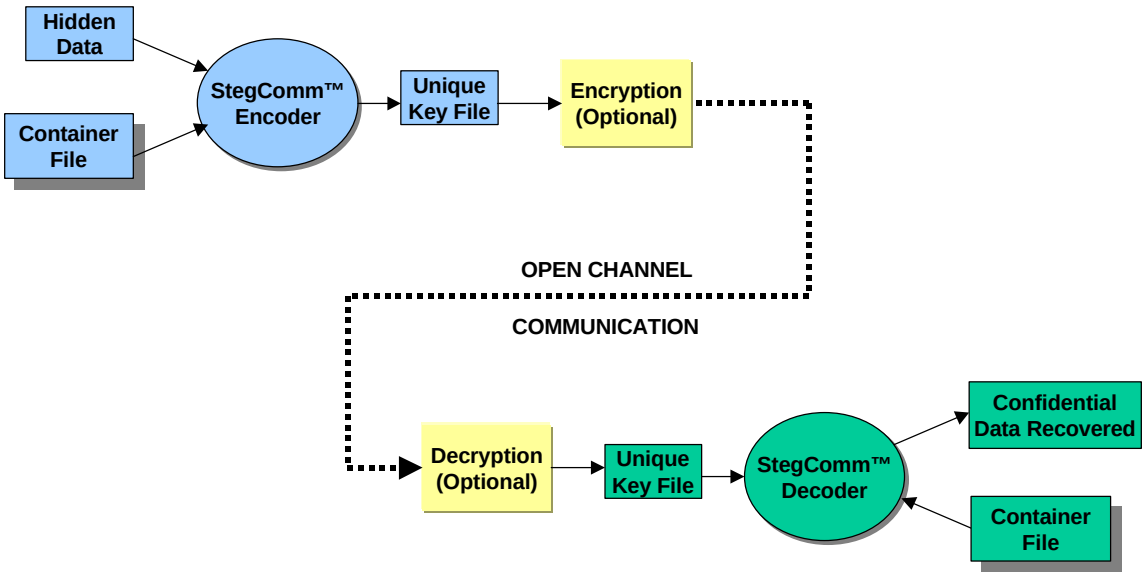

Figure 2 illustrates the operations of StegComm™ through a data flow diagram

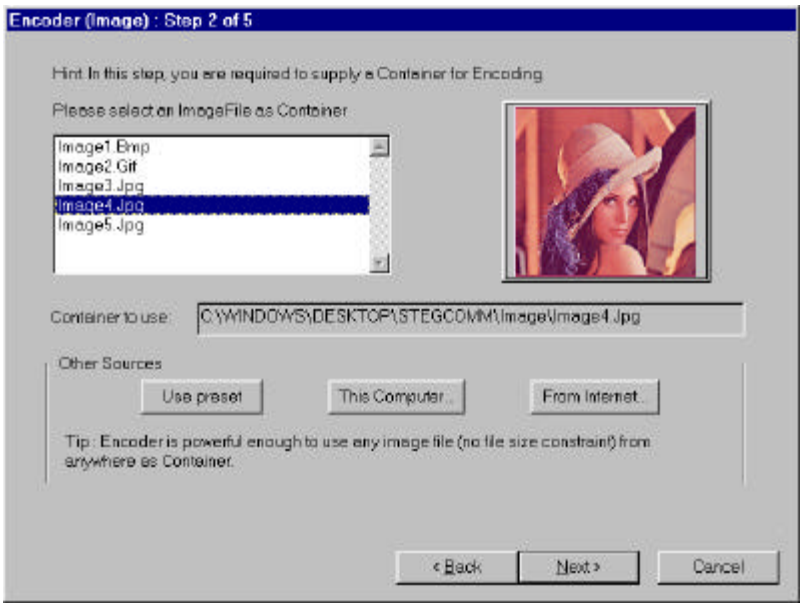

Figure 3 illustrates a GUI of StegComm™

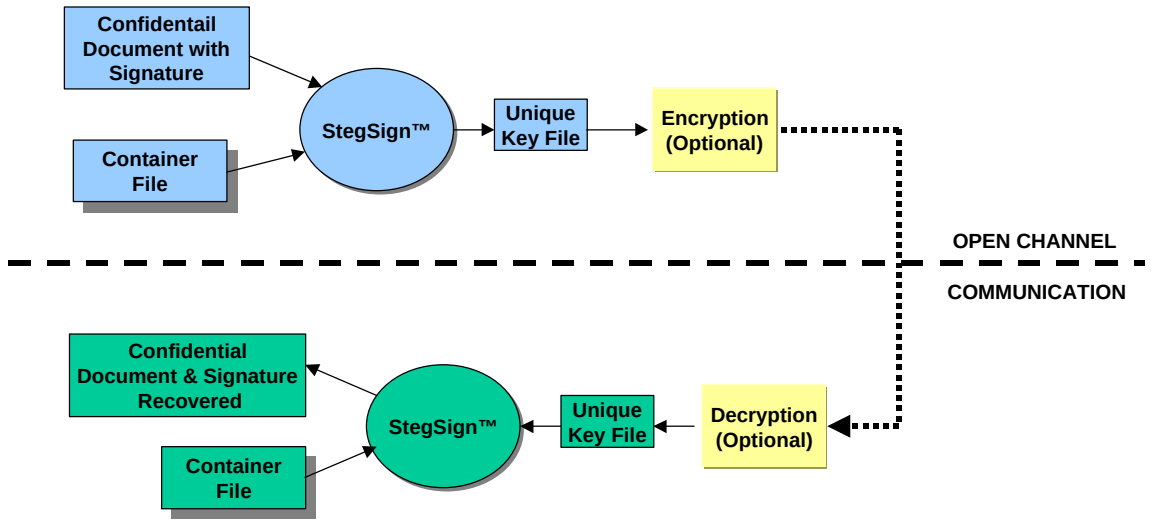

Figure 4 illustrates the operation of StegSign™ through a data flow diagram

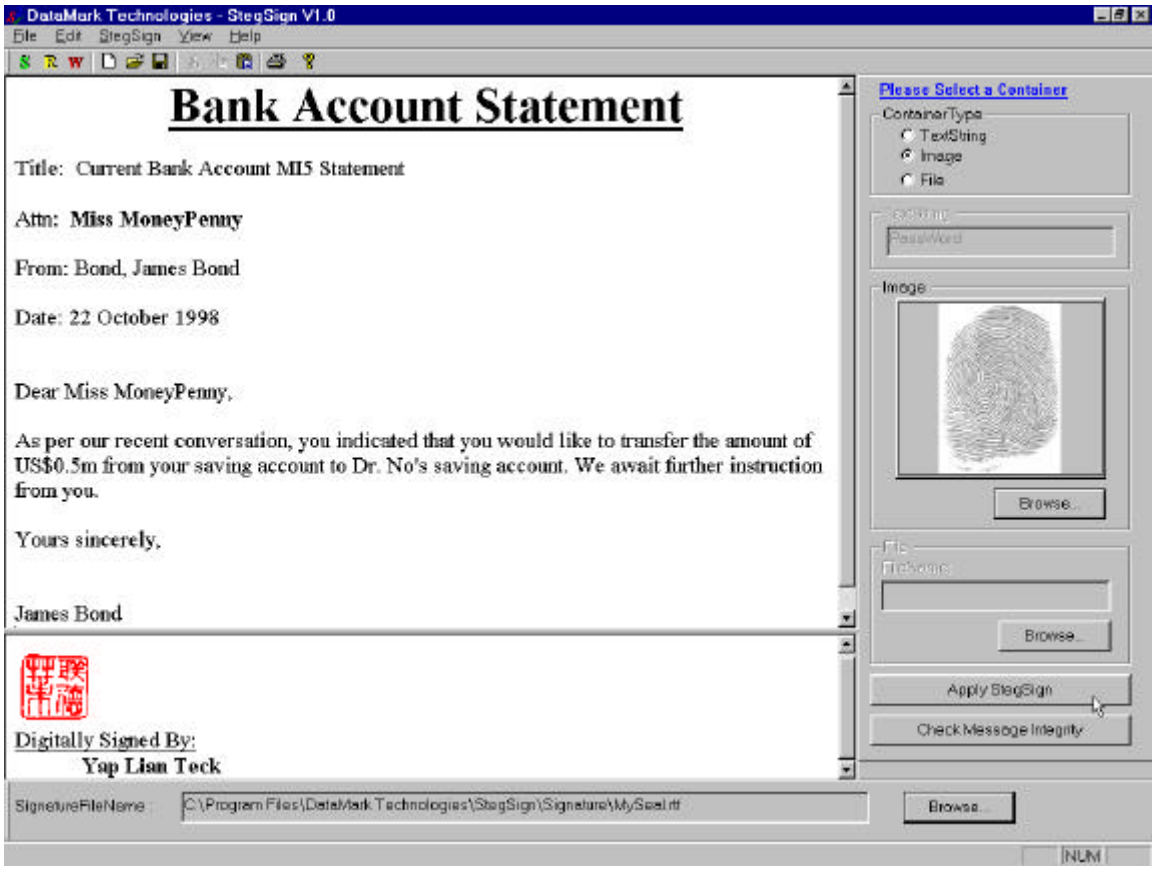

Figure 5 illustrates a GUI of StegSign™

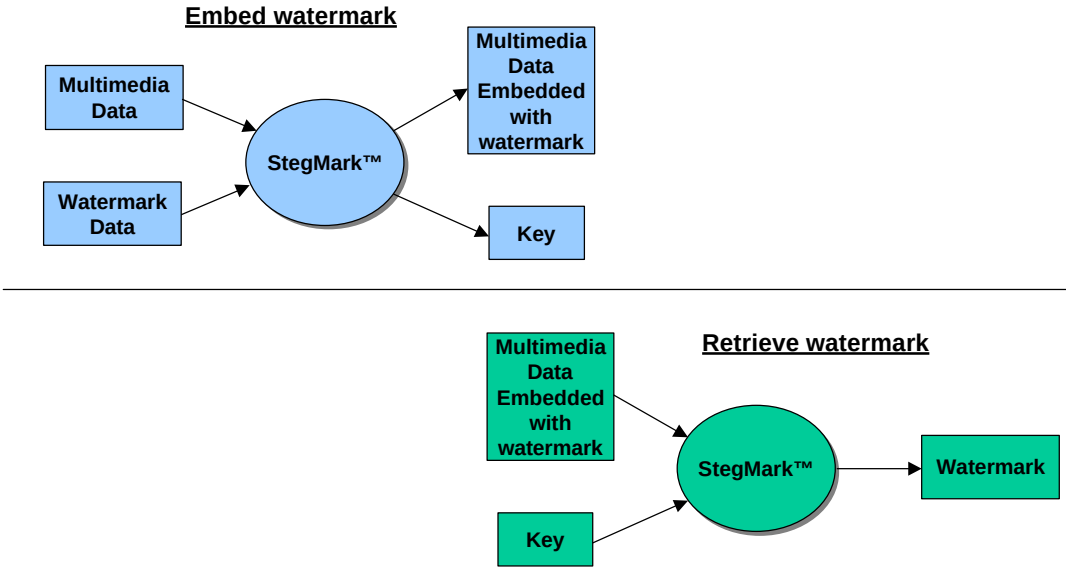

Figure 6 illustrates the operation of StegMark™ through a data flow diagram

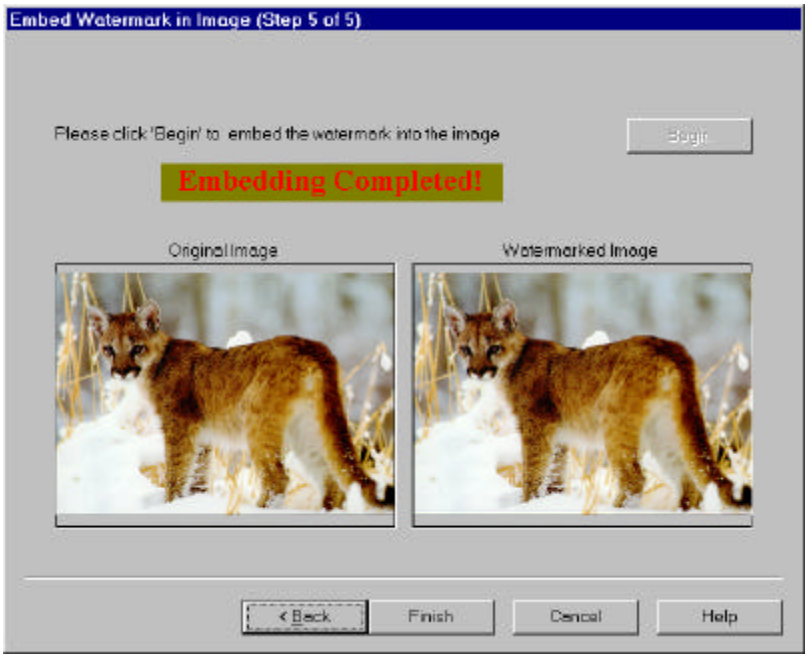

Figure 7 illustrates a GUI of StegMark™

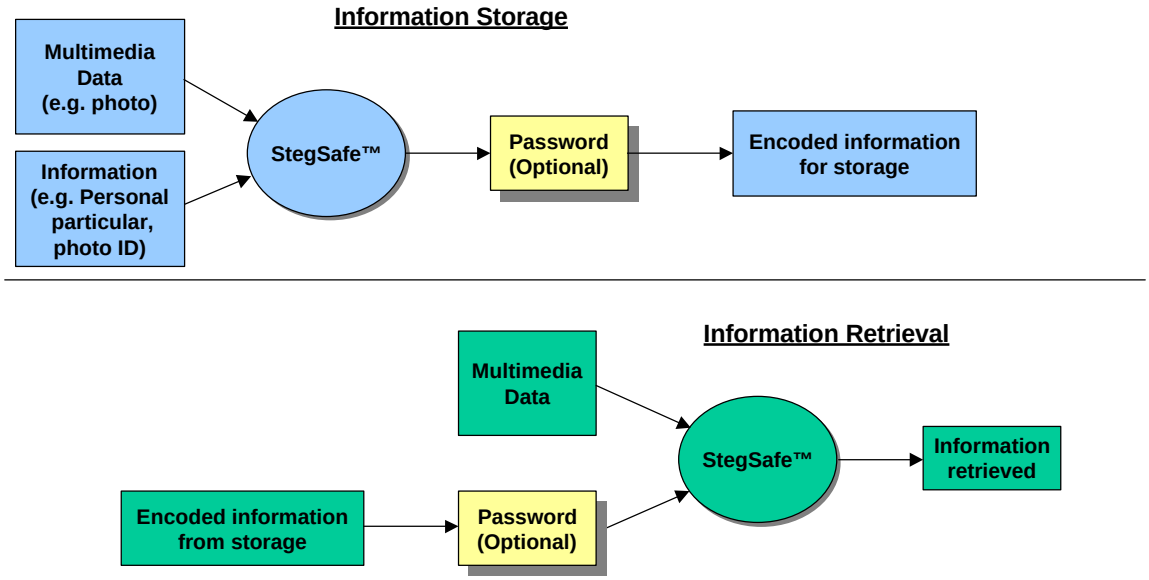

Figure 8 illustrates the operation of StegSafe™ through a data flow diagram

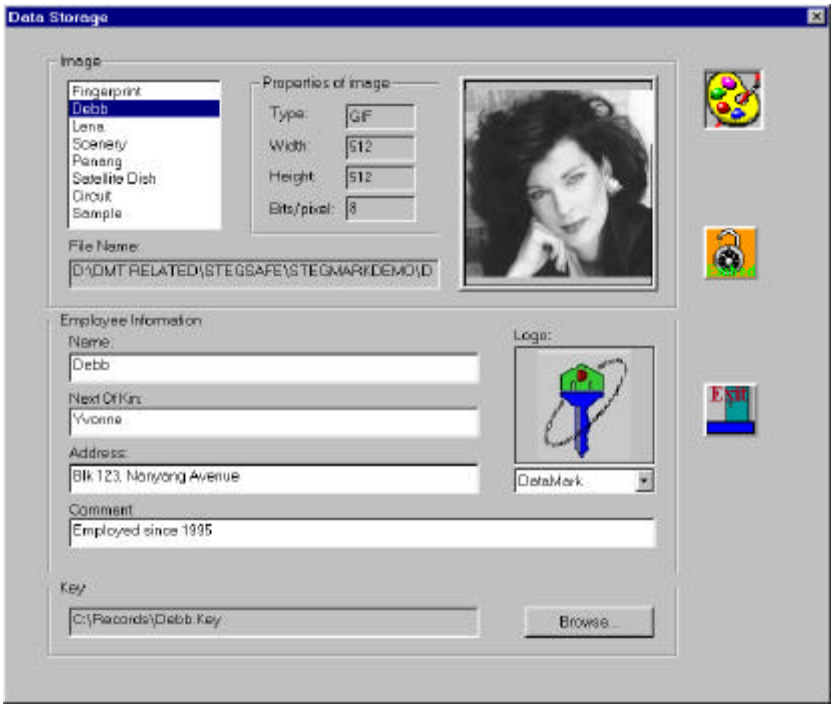

Figure 9 illustrates a GUI of StegSafe™