International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)

Blind Data Extraction from Digital Media

Ch.Kedari Rao
Associate Professor
Sri Indu College of Engg&Tech
Ibrahimpatan, Hyderabad, TS

Oruganti Shylaja
M.Tech Scholar(CSE)
Sri Indu College of Engg&Tech
Ibrahimpatan, Hyderabad, TS

V. Prashanth
M.Tech Scholar(CSE)
Sri Indu College of Engg&Tech
Ibrahimpatan, Hyderabad, TS

Abstract: **We consider the problem of extracting blindly data embedded over a wide band in a spectrum (transform) domain of a digital medium (image, audio, video). We develop a novel multicarrier/ signature iterative generalized least-squares (M-IGLS) core procedure to seek unknown data hidden in hosts via multicarrier spread-spectrum embedding. Neither the original host nor the embedding carriers are assumed available. Experimental studies on images show that the developed algorithm can achieve recovery probability of error close to what may be attained with known embedding carriers and host autocorrelation matrix.**

Keywords: **annotation, blind detection, covert communications, data hiding, information hiding**

1. INTRODUCTION

"Steganography is the art of hiding information in ways that prevent the detection of the hidden message" [1]. The needed information is concealed and its' existence is known only to the sender and the intended recipient. Over the years, a variety of mediums have been used for steganography. One of the earliest recorded uses dates back to the time of Herodotus, some 2000 years ago. In one of his stories, Herodotus tells of a Persian noble man who shaved the head of one of his slaves and tattooed a secret message on his scalp. Once the slave's hair grew back, the nobleman sent him to his destination with instructions to shave his head thus revealing a plan to to instigate a revolt against the Persians Rapidly increasing power of personal mobile devices is providing much richer contents and social interactions to users on the move. In the existing system reversible data hiding technique the image is compressed and encrypted by using the encryption key and the data to hide is embedded in to the image by using the same encryption key. The user who knows the secret encryption key used can access the image and decrypt it after extracting or removing the data hidden in the image. After extracting the data hidden in the image then only can be the original image is retrieved.

This trend however is throttled by the limited battery lifetime of mobile devices and unstable wireless connectivity, making the highest possible quality of service experienced by mobile users not feasible. The recent cloud computing technology, with its rich resources to compensate for the limitations of mobile devices and connections, can potentially provide an ideal platform to support the desired mobile services. Tough challenges arise on how to effectively exploit cloud resources to facilitate mobile services, especially those with stringent interaction delay requirements. In this paper, we propose the design of a Cloud-based, novel Mobile social TV system.

Cloud computing is the provision of dynamically scalable and often virtualized resources as a services over the internet Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them. Cloud computing represents a major change in how we store information and run applications. A number of mobile TV systems have sprung up in recent years, driven by both hardware and software advances in mobile devices. Some early systems bring the living room experience to small screens on the move. But they focus more on barrier clearance in order to realize the convergence of the television network and the mobile network, than exploring the demand of "social" interactions among mobile users.

We propose the design of a Cloud-based, novel Mobile social TV system. The system effectively utilizes both PaaS (Platform-as-a-Service) and IaaS (Infrastructure-asa-Service) cloud services to offer the living-room experience of video watching to a group of disparate mobile users who can interact socially while sharing the video. To guarantee good streaming quality as experienced by the mobile users with time varying wireless connectivity, we employ a surrogate for each user in the IaaS cloud for video downloading and social exchanges on behalf of the user. Nowadays smart phones are shipped with multiple microprocessor cores and gigabyte RAMs; they possess more computation power than personal computers of a few years ago. On the other hand, the wide deployment of 3G broadband cellular infrastructures further fuels the trend. Apart from common productivity tasks like emails and web surfing, smart phones are flexing their strengths in more challenging scenarios such as real time video streaming and online gaming, as well as serving as a main tool for social exchanges.

2. SPREAD SPECTRUM

Spread Spectrum is not a new technology it dates back to the 1930's when it was first used with wideband Frequency Modulation (FM) systems by Major E. H. Armstrong [7]. However, it was not until the early 1940's, that the potential for Military Communications was realized. George Antheil, a composer, and the actress Hedy Lamarr invented a secret communication system. The system, which received a patent, manipulated radio frequencies between transmission and reception to develop an unbreakable code. This meant that top-secret messages could not be intercepted [8]. This was the beginning of today's Spread Spectrum. From the 1940's to the present day significant research has been carried out in the

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)

area of Spread Spectrum for covert communications. Most of this research was carried out by, or on behalf of the military, and it is only in recent years that some of the findings have been released to the public. This has subsequently resulted in the development of commercial applications. Whether we are aware of it or not, most of us have had some interaction with systems using Spread Spectrum Technology. Current applications include Global Positioning Systems (GPS), integrated bar code scanners, digital cellular telephone communications, and others.

Spread Spectrum Technology forms the basis of Spread Spectrum Steganography. As such, it is essential to understand exactly what it is and how it works, in order to use it for steganographic purposes. The theoretical background explaining the basis of Spread Spectrum technology came with the publication of a paper by Claude Shannon on the mathematical theory of communication [11]. Shannon's theorem is as follows:

$$C = W \log2(1 + S/N) \qquad \text{Equation 1}$$

where C = data rate in bits per second, W = bandwidth (Hz), S = average signal power (W), N = mean white gaussian noise power (W).

It can be seen from the equation that the only options available to increase a channel's capacity are to increase either the bandwidth (W) or the signal to noise ratio (S/N). We can see that there is a "relationship between the ability of a channel to transfer error free information, compared with the signal to noise ratio existing in the channel, and the bandwidth used to transmit the information" [12]. Equation 1 can be manipulated to:

$$W = (NC)/S \qquad \text{Equation 2}$$

From equation 2, it can be seen that for any given noise to signal ratio, a low information error rate can be achieved by increasing the bandwidth used to transfer the information.
In order to be considered as a Spread Spectrum system, the system must meet the following criteria [12]:

1. The transmitted signal bandwidth is much greater than the information bandwidth.
2. Some function other than the information being transmitted is employed to determine the resultant transmitted bandwidth.

To accomplish the spreading required, the data transmitted is modulated together with a wideband encoding signal. A direct consequence is that the energy used to transmit the signal appears as noise.

The main advantage of spreading the signal and making it appear as noise is that it makes the communication very difficult to find in the frequency spectrum, thus making it more difficult to track and more difficult to jam. The spreading of the data across the frequency spectrum also makes the signal resistant to noise and interference, thus increasing the likelihood that the signal will be received correctly as sent. Another advantage is that signals generated using Spread Spectrum techniques are unlikely to interfere with other signals even if they are transmitted on the same frequency.

3. SPREAD SPECTRUM TECHNIQUES

There are several techniques currently in use for generating Spread Spectrums. These include Direct Sequence Spread Spectrum (DSSS), Frequency Hopping Spread Spectrum (FHSS), and Time Hopping Spread Spectrum. Each technique differs in its implementation and has certain advantages/disadvantages. In addition to the above, there are a number of hybrid techniques which offer certain advantages over, or extend the usefulness of the other techniques. These hybrids are Frequency Hopped/Direct Sequence Modulation and Time-Frequency Hopping.

The various Spread Spectrum techniques make use of a digital code sequence for the modulation process. The code sequence is called pseudo-random noise (PN). Pseudo-random noise is a signal similar to noise which satisfies one or more of the standard tests for statistical randomness. The signal appears to lack a definite pattern but in fact it is comprised of a deterministic sequence of pulse that repeat after a long period of time. In Spread Spectrum systems, the transmissions of the pseudo-random noise appear as noise to any receiver that is not locked on the transmitter frequencies or is incapable of correlating a locally generated pseudorandom sequence with the received signal.

3.1. Direct Sequence Spread Spectrum (DSSS)

The basic principle behind the Direct Sequence Spread Spectrum (DSSS) technique is the modulation of the RF carrier with a digital code sequence. The code sequence utilizes a chip rate, which is much higher than the bandwidth of the data signal and is used directly to modulate the carrier, thus directly setting the transmitted bandwidth.

A two-stage process is used to produce the DSSS. During the first stage, data is spread across the spectrum. This is achieved by dividing the data stream into a symbol stream (small pieces of one bit or more) and then allocating each part of the divided data to a frequency channel across the spectrum.

During the second stage, the modulation phase, the DSSS transmitter utilizes a phase varying modulation technique (QPKS – Quadrature Phase Shift Key or BPSK – Binary Phase Shift Key) to modulate each piece of data with a higher data rate bit sequence (chipping code), a code called pseudo-random noise (PN). This increases the bandwidth according to a spread ratio which is based on the length of the chip

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)

sequence.

Figure Time and frequency domains of the original data, the result of the stage 1 spreading and the final modulated data.
By modulating the carrier with the digital code sequence, the signal produced is centered at the carrier frequency. The resulting spectrum has a (sinx/x)2 form as can be seen in which shows a DSSS Spectrum produced using a Binary Phase Shift Key (BPSK) to modulate the data with the code sequence.
Although DSSS has very good noise and anti-jamming performance and is very difficult to intercept, it does have some disadvantages. The circuitry required to produce the spectrum is complex, it requires a large bandwidth channel with relatively small phase distortions and requires a long acquisition time since the PN codes are long. Also, DSSS suffers from what is known as a "Near-Far" effect. This effect occurs when an interfering transmitter is much closer to the receiver than the intended transmitter. It is possible that the interference caused by the closer interfering transmitter will result in the receiver only receiving a signal from the interfering transmitter. Consequently, proper data detection is not possible .

3.2. Frequency Hopping Spread Spectrum (FHSS)

Another Spread Spectrum technique is the Frequency Hopping Spread Spectrum (FHSS). FHSS has an advantage over DSSS in that it is not as affected by the "Near-Far" effect. The basic principle behind the Frequency Hopping Spread Spectrum (FHSS) technique is that the carrier frequency is periodically modified (hopped) across a specific range of frequencies. The frequencies, across which the carrier jumps is the spreading code. The shifting pattern is determined by the chosen code sequence (frequency shift key – FSK). The amount of time spent on each hop is known as the dwell time and is in the range of 3ms-100ms.

Two types of Frequency Hopping signals may be used, slow hopping and fast hopping. With slow hopping, the hopping rate is smaller than the message bit rate, meaning that in one hop, one or more data bits are transmitted. While in fast hopping, one data bit is divided over more than one hop (the hopping rate is greater than the message bit rate) [17, 19].

There are advantages and disadvantages with both slow and fast hopping. With slow hopping, coherent data detection is possible, but data can be lost if a single frequency hop channel is jammed. To overcome this, it is necessary to use error correcting codes. Fast hopping disposes of the need for error codes since one bit of data is spread over a number of hops. However, fast hopping has the disadvantage that due to phase discontinuities, coherent data detection is not possible [17].

Like DSSS, FHSS has advantages and disadvantages. In comparison to DSSS, FHSS provides a greater amount of spreading and has a relatively short acquisition time since the chip rate is considerably less than in DSSS. It is also less susceptible to the "Near-Far" effect which can cause problems with DSSS. On the negative side, FHSS requires a complex frequency synthesizer to generate the hops and it requires error correction.

4. MODULE DESCRIPTION

The most crucial phase of any project is the implementation. This includes all those activities that take place to convert from the old system to the new system. It involves setting up of the system for use by the concerned end user. A successful implementation involves a high level of interaction between the analyst, programmers and the end user. The most common method of implementation is the phased approach, which involves installation of the system concurrently with the existing system. This has its advantage in that the normal activity carried out, as part of the existing system is anyway hampered. The end users are provided with sufficient documentation and adequate training in the form of demonstration/presentation in order to familiarize with the system.

MODULES

1. Steganography
2. Multi-Carrier Spread Spectrum Embedding
3. Image encryption and watermarking
4. Image decryption and extraction

Steganography: Steganography includes the concealment of information within computer files. In digital steganography, electronic communications may include steganographic coding inside of a transport layer, such as a document file, image file, program or protocol. Digital steganography can hide confidential data (i.e. secret files) very securely by embedding them into some media data called "vessel data." The vessel data is also referred to as "carrier, cover, or dummy data". In Steganography images used for vessel data. The embedding operation in practice is to replace the "complex areas" on the bit planes of the vessel image with the confidential data. The most important aspect of Steganography is that the embedding capacity is very large. For a 'normal' image, roughly 50% of the data might be replaceable with secret data before image degradation becomes apparent.

Multi-Carrier Spread Spectrum Embedding: The technique of spread spectrum may allow partly fulfilling the above requirements. Advantages of spread spectrum techniques are widely known: Immunity against multi-path distortion, no need for frequency planning, high flexibility and variable data rate transmission. The capability of minimising multiple access interference in direct-sequence code- division-multiple-access system is given by the cross-correlation properties of spreading codes. In the case of multi-path propagation the capability of distinguishing one component fiom thers in the

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)

composite received signal is offered by the auto-correlation properties of the spreading codes.

Image encryption and watermarking: The host image is an 8-bit or higher grey level image which must ideally be the same size as the plaintext image or else resized accordingly using the same proportions. Pre-conditioning the cipher and the convolution processes are undertaken using a Discrete Fourier Transform (DFT). The output will include negative floating point numbers upon taking the real component of a complex array. The array must be rectified by adding the largest negative value in the output array to the same array before normalization. For color host images, the binary cipher text can be inserted into one or all of the RGB components.The binary plaintext image should have homogeneous margins to minimize the effects of ringing due to 'edge effects' when processing the data using Fourier transform.

5. FEASIBILITY REPORT & SYSTEM TESTING

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.
Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY: This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY: This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY: The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

SYSTEM TESTING: The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that theSoftware system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTING

Unit testing: Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing: Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test: Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)

System Test: System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing: White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing: Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

Test Cases

Unit Testing: Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach: Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing: Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

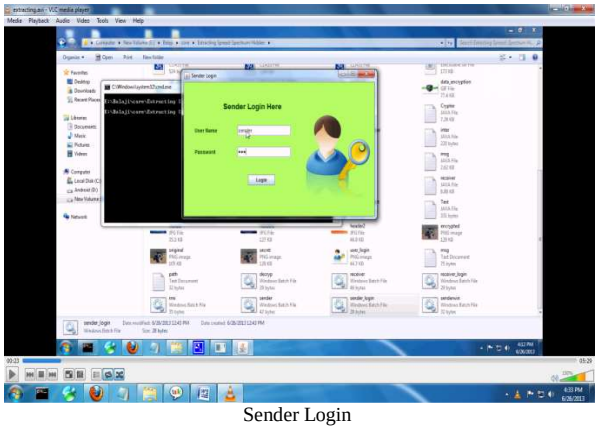
Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing: User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

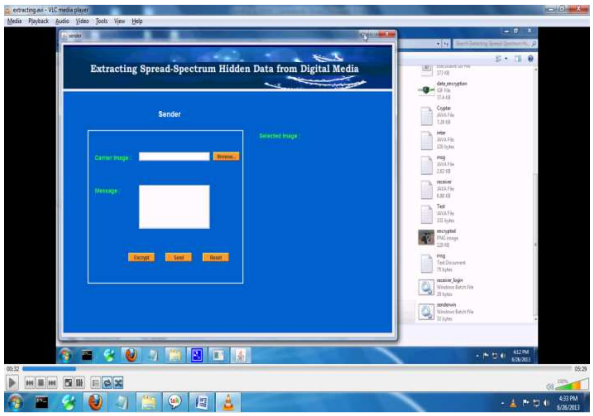
6. CONCLUSION

This paper presents our view of what might become a trend for mobile TV, i.e., mobile social TV based on agile resource supports and rich functionalities of cloud computing services. We introduce a generic and portable mobile social TV framework, CloudMoV, which makes use of both an IaaS cloud and a PaaS cloud. The framework provides efficient transcoding services for most platforms under various network conditions and supports for co-viewing experiences through timely chat exchanges among the viewing users. By employing one surrogate VM for each mobile user, we achieve ultimate scalability of the system. Through an in-depth investigation of the power states in commercial 3G cellular networks, we then propose an energy-efficient burst transmission mechanism that can effectively increase the battery lifetime of user devices.

7. OUTPUT SCREENS

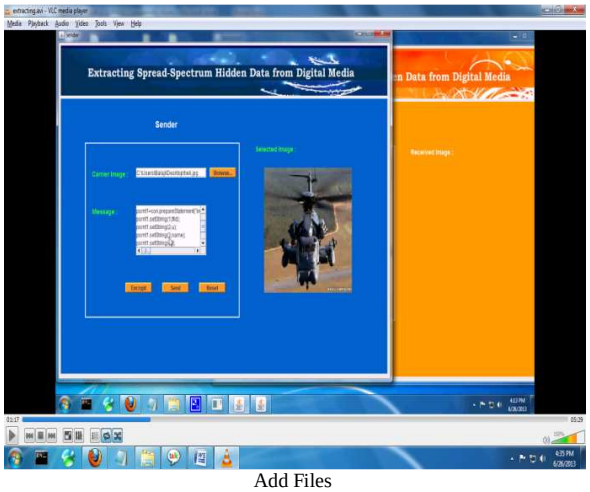


Sender Login

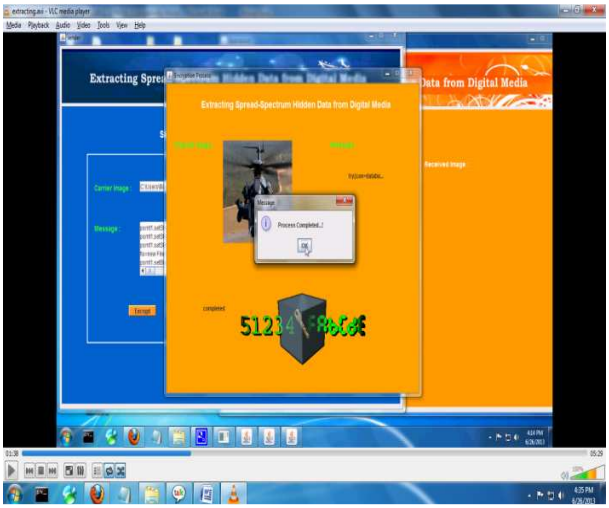


Sender Homepage

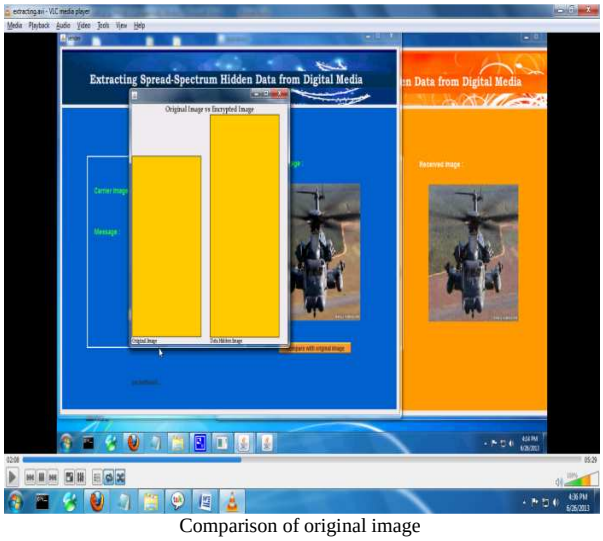
International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 2, Issue 2, February 2015)



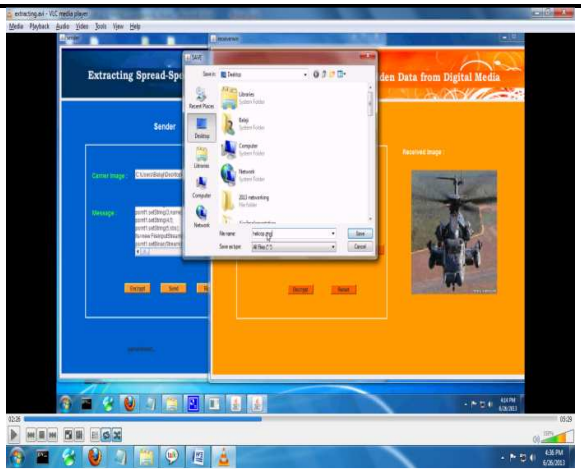
Add Files



Complete Encrypt Process



Comparison of original image


Save Image

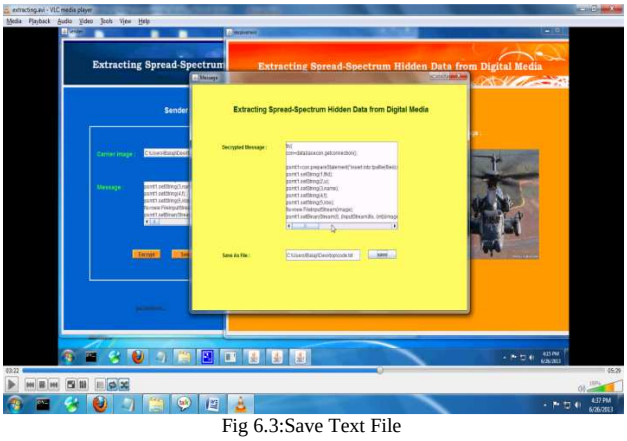


Fig 6.3:Save Text File

REFERENCES

[1] Ming Li, Michel Kulhandjian, Dimitris A. Pados, Extracting Spread-Spectrum Hidden Data from Digital Media: IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. X, NO. X, MONTH YEAR X
[2] Neil F. Johnson and Sushil Jajodia, Steganograph: Seeing the Unseen. IEEE Computer, February 1998: 26-34
[3] Duncan Sellars, An Introduction to Steganography
[4] Gary C. Kessler, An Overview of Steganography for the Computer Forensics Examiner, Forensic Science Communications, July 2004, Volume 6, Number 3
[5] William Stalling. Network and Internetwork Security. Addison-Wesley professional computing series Addison-Wesley, 1996. ISBN 0-201-63337-X.
[6] W. Bender, D. Gruhl, N. Morimoto and A. Lu, Techniques for Data Hiding, IBM Systems Journal, Vol. 35, Nos 3 & 4, 1996, pp 313-336
[7] S.K.Pal, P.K.Saxena and S.K.Muttoo, The Future of Audio Steganography, Pacific Rim Workshop on Digital Steganography 2002 (STEG'02) July 11-12, 2002
[8] E. Armstrong, A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation, Proc. IRE, Vol. 24, pp. 689-740, May1936
[9] A. Arasu and H. Garcia-Molina, "Extracting Structured Data from Web Pages," Proc. SIGMOD Int'l Conf. Management of Data, 2003.
[10] L. Arlotta, V. Crescenzi, G. Mecca, and P. Merialdo, "Automatic Annotation of Data Extracted from Large Web Sites," Proc. Sixth Int'l Workshop the Web and Databases (WebDB), 2003.

[11] P. Chan and S. Stolfo, "Experiments on Multistrategy Learning by Meta-Learning," Proc. Second Int'l Conf. Information and Knowledge Management (CIKM), 1993.
[12] W. Bruce Croft, "Combining Approaches for Information Retrieval," Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval, Kluwer Academic, 2000.
[13] V. Crescenzi, G. Mecca, and P. Merialdo, "RoadRUNNER: Towards Automatic Data Extraction from Large Web Sites," Proc. Very Large Data Bases (VLDB) Conf., 2001.
[14] S. Dill et al., "SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation," Proc. 12th Int'l Conf./ World Wide Web (WWW) Conf., 2003.
[15] H. Elmeleegy, J. Madhavan, and A. Halevy, "Harvesting Relational Tables from Lists on the Web," Proc. Very Large Databases (VLDB) Conf., 2009.
[16] D. Embley, D. Campbell, Y. Jiang, S. Liddle, D. Lonsdale, Y. Ng, and R. Smith, "Conceptual-Model-Based Data Extraction from Multiple-Record Web Pages," Data and Knowledge Eng., vol. 31,no. 3, pp. 227-251, 1999.
[17] D. Freitag, "Multistrategy Learning for Information Extraction," Proc. 15th Int'l Conf. Machine Learning (ICML), 1998.
[18] D. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, 1989.

About the authors:

Ch.Kedari Rao, Presently Pursuing Ph.D in CSE from JNTUH., Hyderabad. He has eight years of teaching experience and currently working as an associate professor in the Dept. of CSE of Sri Indu College of Engineering & Technology, Hyderabad.

Oruganti Shylaja completed B.Tech. Currently Pursuing M.Tech 2nd Year in Sri Indu College of Engineering and Technology. Areas of Interest are Cloud Computing, Web Technologies, Computer Networks and Data Base Management Systems.

V. Prashanth completed B.Tech from Jaya Prakash Narayana College of Engineering and Technology, Mahaboobnagar, JNTUH with Distinction. Currently Pursuing M.Tech 2nd Year in Sri Indu College of Engineering and Technology. Areas of Interest are Cloud Computing, Web Technologies, Computer Networks and Data Base Management Systems.