International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 06 | Sep-2015 www.irjet.net p-ISSN: 2395-0072

A MELIORATED APPROACH TO TEXT STEGANOGRAPHY USING MARKUP LANGUAGES AND AES

K.Ramesh[1], P.Manivannan[2]

[1,2]Assistant Professor, Department of Information Technology, V.R.S. College of Engineering & Technology, Arasur, Villupuram District, Tamilnadu, India.

Abstract – Steganography is the art or practice of concealing a message, image or file within another message, image or file. It is simply hiding some data within some other file. It has been used over the course of history for a variety of purposes. But, web crawlers and advanced automated scripting systems may now be used to reveal hidden data and therefore entirely break the security provided by steganography. This paper proposes a new mechanism of steganography where the data is hidden in a webpage after being encrypted using AES, which is the global standard of encryption used in the present world and one which has not been broken even with advanced systems. A new method of key transfer is also suggested.

Keywords – SHA 256, AES, Bit message, Hashing.

I. INTRODUCTION

Security mechanisms and authentication systems are used by people of all origins to safeguard personal and sensitive data. Various security mechanisms have evolved over the course of the time. Each of these security mechanisms seem to be better than the previous one in some manner [1]. But, with the advancements in computing power, each new mechanism is bound to be broken at sometime or other. Scientists are at the verge of inventing quantum computers which if used for breaking encryption standards may break almost all the standards. It is therefore important to find new system that doesn't allow the data to be discovered either by eavesdropping or snooping.

II. LITERATURE REVIEW

Attacks on data are becoming more and more prevalent. The amount of time and money that people are willing to spend to steal data from others was directly proportional to the sensitivity and importance of the information. Passwords are the primary (and in most cases, the only) authentication system. However, computers produced for consumer usage can now break passwords and steganography relatively easily. This system is only made difficult by the usage of multi-core and multi-processor systems [6].

Some password and hash-breaking systems are now capable of using GPGPUs to perform computations. Many such possible attacks and one such implementation are described [3]. It is therefore evident that textual passwords were not the future of encryption systems. The data is to be safeguarded using some other system that doesn't reveal itself to the attacker or eavesdropper. The common technique that was used for generations for this purpose was steganography. Steganogrpahy of previous systems were used to hide data in images [4].

Usage of markup languages to hide data was first introduced by Mr.Susmita Mahato, Mr.Dilip Kumar Yadav, and Mr.Danish Ali Khan in their paper [2]. This paper introduced the method of hiding data in web pages by detecting the presence of tags along with their attributes. This paper however had the primary limitation of hiding data that were only a few bits long. This was not very helpful as almost all data in the present world will be atleast a few KBs. This was the major limiting factor and disadvantage of this paper.

Various types of cryptography and steganography were performed [4]. The relative positions of lines or words were altered to create grammatical or common mistakes [5]. The mistakes were not of great significance. This method was used to hide data. Water marks in images were also used to hide data [7]. The water mark was overlooked easily. This made data transfer easy. Common word spellings were altered in a seemingly long paragraph [8]. The altered paragraph when deciphered produced messages.

III. ATTACKS ON DATA

Attacks can be active or passive in its most basic form. Passive attacks can be considered as being less harmful than active attacks. Passive attacks monitor all the data being sent and received through the monitored network. Active attacks, on the other hand, monitor data and modify them before forwarding them to the destination. Various types of attacks are possible over the network on the data. Some of them are:

A. Eavesdropping

Since most of the data transmitted is in plain text format (unless it uses a SSL protocol).Eavesdropping is the most common method of attack on data. It is also next to impossible to identify this attack because no alteration is performed, whatsoever.

B. Compromised Key Attacks

Keys are used to encrypt data before being used for transmission. Keys are randomly generated or generated based on certain principles. Obtaining these key were very difficult, but not impossible. Once the hacker gets these keys, they are known as compromised keys. The compromised keys can be used to decrypt data, modify them or send completely irrelevant messages.

C. Man-in-the-middle Attacks

These attacks are used to modify data sent through the network or in other cases are used to completely sends a new message by the intermediate person. The receiver believes that the message is authentic and may perform actions such as updating databases based on the sent message.

D. Sniffer Attacks

Sniffers are small software used to analyze networks and the data being sent through the affected computers. These sniffers can capture packets and forward them to specific destinations if programmed to do so. Sniffer attacks can recover any type of data being sent and may also be used to forward encrypted data which can then be decrypted and used (But, this still requires the process of decrypting the original key used to encrypt the data).

E. Cross Site Scripting

Cross Site Scripting (XSS) is a type of computer security vulnerability found in web applications. It allows attackers to inject client-side script into web pages viewed by others. It can be used to bypass security controls. It is mainly due to breaches in browser security.

F. Phishing Attack

In this type of attack, the user is made to enter passwords into fake page that imitates the original or genuine pages. These pages can be easily identified by looking at the URLs, since the URLs are all redirected URLs or shortened URLs.

IV. MECHANISM

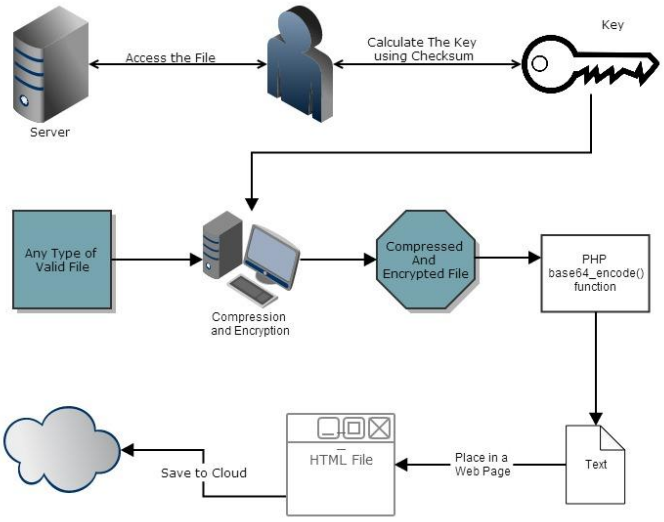

Figure 1 – Overview

The whole mechanism is divided into multiple steps to prevent any ambiguities in the process. The steps involved in the mechanism are explained as follows :

A. Key Transfer

Most of the encryption systems fail ultimately at this step. Any data that is encrypted has a key. This key transfer is done using some ordinary means (like email or IM) in most cases. Even in more complex transfer mechanisms, the attackers may capture the encrypted data and the key, thereby overcoming the defense provided by the encryption. The process that we use for data transfer is different from existing approaches. The receiver of the data chooses some random sequences of text or a common sentence as per his needs and transfers this to the sender of the data. The protocol or transfer method that is to be used for this transfer is Bit Message. The Bit Message protocol is used is a decentralised peer-to-peer communication protocol in which the sender cannot be spoofed.

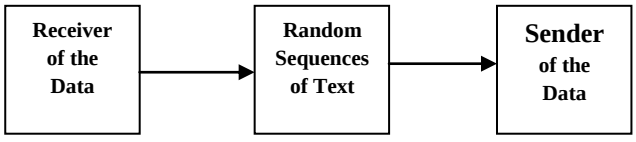

Figure 1 – Sending the Random Text

Now, this random piece of text is obtained by the sender of the encrypted data. He copies this piece of text and stores in a simple text file with an extension agreed upon previously by both the sender and the receiver. For this example, let us use the simple *.txt extension. Now, the

received text is in a text file. The SHA-256 of this file is calculated. SHA stands for Secure Hashing Algorithm and the output that is generated is always unique. This is used as the key for encryption.

B. Hash Conversion

This is a part were the actual key for the encryption has generated, the hashing is any algorithm that maps the arbitrary length data into a fixed length. Thus, the text sent by the receiver can have any information (Eg: Name, Address, etc) that will be of any length data. On calculating the hash value (i.e. SHA-256), the output of this will be 256 bits length hexadecimal data which is going to be used as a key for encryption process.

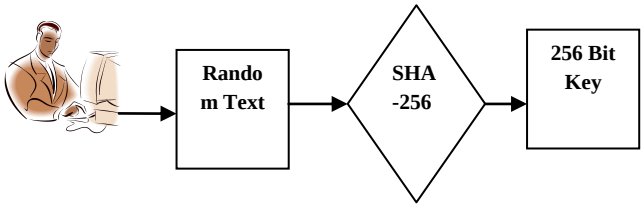

Figure 2 – Generation of Hash Value

C. Encryption

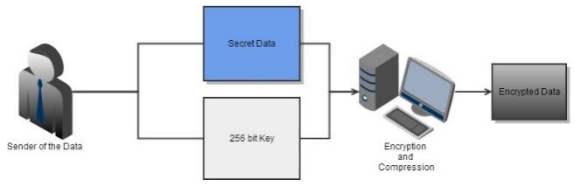

Figure 3 – Encryption of the Original Data

Encrypting is a process of encoding messages or information in such a way that only authorized parties can read it. Encryption doesn't prevent hacking but it reduces the likelihood that the hacker will be able to read the data that is encrypted. It is a scheme of converting the normal media or plain text into an unreadable form, thus it allows only the authorized person to access the encrypted data who has the key. Here, we use AES (Advanced Encryption Standard) algorithm which was developed by two Belgian cryptographers. It belongs to Rijndael cipher which belongs to a family of ciphers with different key and block sizes. For AES, NIST selected three members of the Rijndael family, each with a block size of 128 bits, but three different key lengths: 128, 192 and 256 bits. What we use for our project is 256 bits length key since it is unbreakable, also provides the higher level of security to our data from the hackers and other unwanted users.After the process of encrypting the file (any type of media files), the encrypted file is compressed with a 7zip. 7zip is a compression program that is used to reduce the original size of the file. Also, for an additional security the encrypted file can be put into a password protected zip file which will act as an additional security.

D. Converting Text using PHP Base64 Algorithm

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. In the PHP language, a Base64_Encode function which converts any type of input file into a scrambled text file. We can achieve this by simply using a function in any PHP environments. Thus, here the original source media is converted into a normal text file which will be not in a user readable form. The output text data will be higher than the size of the original file size. In order to reduce the sizing problem here, the source file was compressed in the previous module so that the size after this conversion will look equal to the size of the original source file before the compression.The above function can be reversed using the PHP Base64_Decode function, which transfers the scrambled text file into the original source file. The only problem here is that the scrambled text should not be damaged or there should be no loss made to it since it can affect the original source file.

E. Embedding on HTML Web Page

The output text after the conversion of PHP Base64_Encode function is embedded into a HTML web-page, this embedding process is done only at the back-end of the code. Hence, the data embedded in the HTML web-page will be secured enough, also if someone sees the source code of the web page; they will not get any idea only seeing the scrambled text data. The data is secure also it will be not known that there is some hidden data other than the one who is receiving/awaiting for the source file. The HTML page in which the code is embedded can be published in web or also it can be posted to any cloud service by which the receiver sends a key to the sender. By this method if the receiver will be notified with the web address that has the source file or the cloud service name to which the receiver must have access so that they can access the file to get the data.

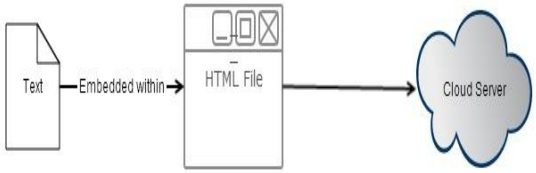

Figure 4 – Saving the Text to a Cloud Server

F. Decryption

Decryption is a reverse process of what we done until now which unsurprisingly gives back the original source data at the end of this process. This process can be initiated only when the receivers know about the logic of what is done from the beginning and also they should have the appropriate key for reversing the entire process.

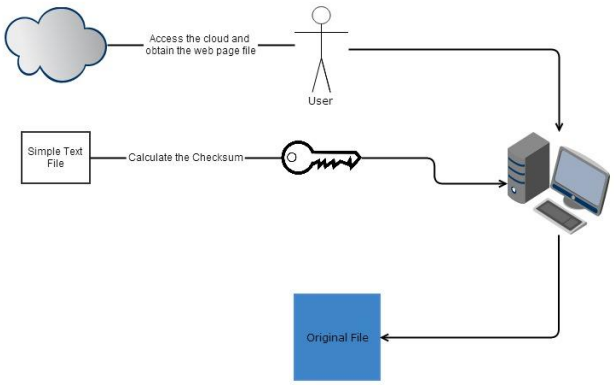

Figure 5 – The Decryption Process

V. THE PROCESS

Consider the situation where Anamika and Santhosh are to transfer messages privately. The whole procedure is known between them. Santhosh first uses the Bit Message client and sends a message to Anamika. Now, the message has been received by Anamika. The subject of the message is used as the title and content is the content of the text file that she creates.

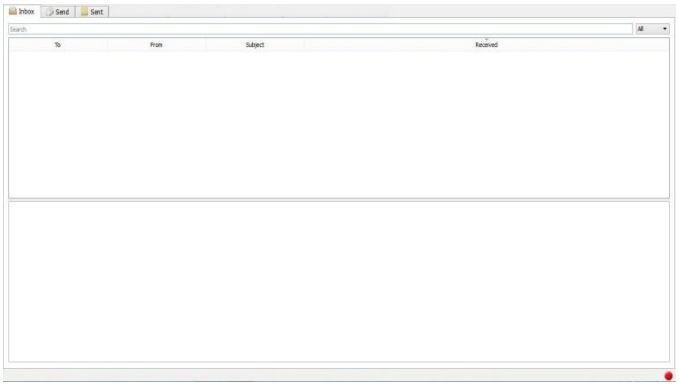

Figure 6 – The Bit Message Client

Now, she calculates the SHA256 of this text file. The calculated SHA256 for this method is E7EB01F539C3E39DBFA53BB9FE7EA2843D285DF55B2 B6BE543475D4B3E6F3E0C. 256 bits was selected because it is the maximum key size allowed by AES. Now, Anamika selects a file to be sent and encrypts it using the SHA256 key.

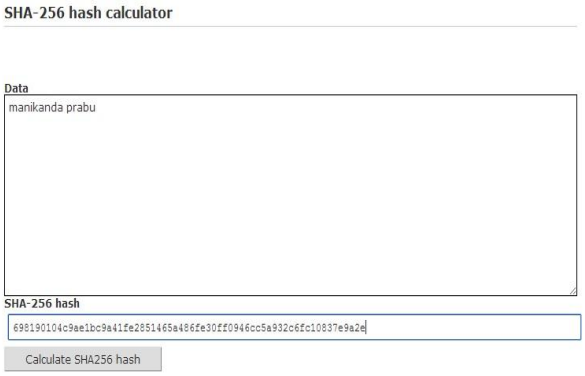

Figure 7 – Hash Conversion Process

The encrypted file is now converted to textual format using the PHP base64_encode function. The page where it is converted is shown in the following diagram.

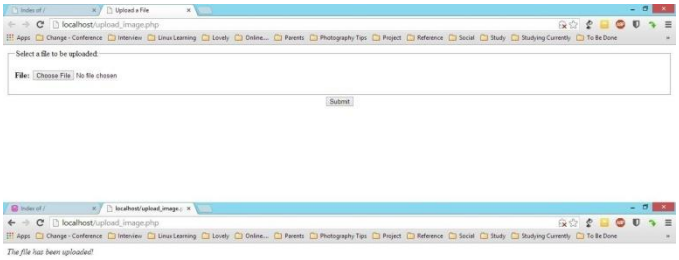

Figure 8 – Uploading the File

This produces an output of text that can be passed without any problem through the network.

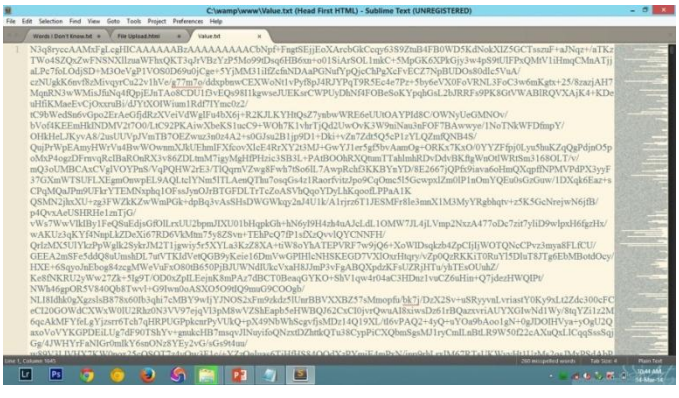

Figure 9 – The Converted Text

This is pasted in any HTML page and is published to the web by Anamika. This page is accessed by Santhosh and the contents are retrieved. When he uses the PHP base64_decode function, he gets the original encrypted

file. When he enters the SHA256 of the original text file, the file that was to be passed is received successfully.

VI. APPLICATIONS

This method can be used in all areas where data security and privacy is of prime importance. The most important application is usage in military applications. Data regarding missiles and particular attack or mission details can be safely transmitted using this method. Agents working undercover can safely transmit data and the data transmitted will be safe as long as the method is concealed.

VII. CONCLUSION

It is common for data to be stolen intercepted or hijacked, everything is possible in this digital world safety and security was defined by the limits of the user's knowledge. It is therefore sensible to use proper security measures and enjoy security and privacy.

REFERENCES

[1] Nadeem, A. Javed, M.Y. "A Performance Comparison of Data Encryption Algorithms", Information and Communication Technologies, 2005. ICICT 2005. 27-28 Aug. 2005, pg : 84-89, Print ISBN: 0-7803-9421-6.

[2] P.Vignesh, K.Ramesh Kumar, "Usage of GPGPU For Password Cracking Using Infiband Switched Fabric Link And a Prevention Protocol", International Journal of Advanced Computer Science and Information Technology, Paper ID - CNIACSIT-103-206,Volume – 1, Issue – 1, Mar-2013.

[3] Susmita Mahato, Dilip Kumar Yadav, Danish Ali Khan, "A Modified Approach to Text Steganography using HyperText Markup Language", International Conference on Advanced Computing and Communication Technologies. ISSN – 2327 -0632, pg : 40-44.

[4] A.Joseph Raphael, Dr. V. Sundaram, "Cryptography and Steganography A Survey", International Journal on Computer Technology and Applications, vol. 2 (3), ISSN:2229-6093, pp.626 630, 2010.

[5] S. H. Low, N. F. Maxcmchuk, J.T. Brassil, and L.O'Gorman, "Document marking and identification using both line and word shifting", Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95), Vol.2, pp.853860.

[6] K.Ramesh ,"Test Suite Generation using Genetic Algorithm and Evolutionary Techniques with Dynamically Evolving Test Cases" International Journal of Innovation and Scientific Research 2 (2), 296-300.

[7] D. Huang, and H. Yan, "Later word distance changes represented by Sinc Waves for watermarking text images", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11(12), pp.1237 1245, December 2001.

[8] M. Shirali-Shahreza, "Text Steganography by Changing words spelling," Proceedings of the 10 th International Conference on Advanced Communication technology, vol.3, pp1912 1913, 17-20 February 2008.

BIOGRAPHY

Mr.K.Ramesh, received his Bachelor degree in Computer Science and Engineering from Annamalai University, Chidambaram in 2008 and his Master degree in Software Engineering from Anna University of Technology, Coimbatore in 2010. At present he is working as Assistant Professor in Department of Information Technology at V.R.S. College of Engineering and Technology, Arasur. His areas of interest are Computer Networks, Software Engineering and Quality Assurance, Software Testing.

Mr.P.Manivannan received his Bachelor degree in Computer Science and Engineering from Anna University, Chennai in 2009 and his Master degree in Network Engineering from Anna University of Technology, Coimbatore in 2011. At present he is working as Assistant Professor in Department of Information Technology at V.R.S. College of Engineering and Technology, Arasur. His areas of interest are Computer Networks, Cryptography and Network Security, Mobile Communication.