DATA HIDING IN AUDIO CARRIER USING AES ALGORITHM IN DWT DOMAIN

Aarcha P.K[1], Lani Rachel Mathew[2]

[1]M.Tech Scholar, Dept. of Electronics and Communication Engineering,
[2]Assistant Professor, MarBaselios College of Engineering and Technology,
Trivandrum,India

ABSTRACT

In the communication field, more security is required due to the increase in data transmitted and received with less susceptibility to eavesdropping. It is achieved by several techniques like watermarking, steganography, cryptography etc. Steganography technique is highly secure because of the properties like transparency, capacity and robustness. Steganography technique can be done in different domains. Biorthogonal wavelet family is chosen for the proposed audio steganography method because it uses AES algorithm and hence provides high capacity and preserves the high quality of the stego signal. In the proposed technique, first the secret message is encrypted using AES algorithm. This encrypted message is then XORed with itself and embedded within the cover audio decomposed by biorthogonal wavelet family. The reverse operation is done at the receiver side to obtain back the original secret message. Performance evaluation is done based on the SNR, processing time and MSE.

Keywords: Audio Steganography, Advanced Encryption Standard, Discrete wavelet transfor, Wavelet families.

I. INTRODUCTION

Communication security is of utmost importance in the present scenario maintaining secrecy of data transmitted. Steganography technique is a kind of secure communication method that hides information within a cover medium like text, audio, video or image [1]. It is also called covered writing because it uses cover for hiding information to obtain better security [2] [3].Audio steganography is a technique that hides the secret information in a cover medium, which is an audio signal, and the embedded signal is transmitted using wireless communication. At the receiver side the secret message is retrieved from the stego signal. Stego signal is inaudible to the third party and hence they are unable to access the secret information. Steganography is the best method because it gives secrecy when compared to cryptography which gives only privacy. Both Steganography and Cryptography ensure data confidentiality [1].Several audio steganography techniques have been developed such as the LSB, parity coding, phase coding, spread spectrum and echo data hiding. [4] Djebbar et al 2012 have explained about different domain methods. In temporal domain, LSB method is mostly used, which is the easiest and earliest method used for data hiding. It has high embedding capacity but reduces the performance of robustness and hence is highly vulnerable to attack. Tanwar et al 2014 proposes a LSB substitution technique by raising the signals average power. This makes the system robust to various attacks that try to reveal the hidden message in addition to the unintentional attack like distortion [5]. The transform domain is another domain that exploits the frequency range of HAS directly by modifying only the mask regions or indirectly by altering the audio signal samples. [6] Bandyopadhay et al 2012 discusses two methods: modified LSB and Phase coding. The basic idea of phase coding is that it encodes the secret message bits in phase spectrum as phase shift, which is inaudible and with low transmission rate. Thus, it can be used only when a small amount of data needs to be concealed.Another method for audio steganography is in the Discrete Wavelet Transform (DWT) domain. Singh et al 2014 proposes a high capacity and high stego signal quality method based on samples comparison in DWT domain where selected coefficient of a segment value is altered by a threshold value depending upon the embedding cipher text bit [2]. Rekik et al 2012 introduces a new approach where DWT separates high frequency and low frequency components. In the next stage, low amplitude of the high frequency component is used to hide another audio signal. This increases the complexity hiding so that an eavesdropper cannot extract the hidden information even after suspecting the existence of a secret message [7]. Gupta et al 2014 proposes a new method that combines LSB and DWT. DWT is applied on audio files for extracting higher frequency components. Then LSB is used for hiding the secret message in those higher frequency components. This improves simple LSB algorithm so that persons with malicious intent cannot easily extract the message from audio file [8].

In the proposed system, wavelet families are used for embedding text in the cover audio file. For better security AES algorithm is used and after that XORing operation is done on the AES encrypted message bits. The rest of the paper is organized as follows. In the second section, the proposed system is explained, followed by the experimental results in the next section. The final section discusses about the conclusions arrived at.

II. PROPOSED METHOD

In the proposed method steganography technique is combined with cryptography for secure transmission of information. Here information is in the text format. For achieving high security, AES algorithm is used and after that, we do XORing on AES encrypted message bits with itself for embedding data in the audio file. Wavelet families have been used for decomposing the audio cover signal.

A. AES Algorithm:

AES design is based on the principle known as a substitution-permutation network, i.e. it is a combination of both substitution and permutation which gives fast result in both software and hardware. In this method, 16 bit key is used while the data to be encrypted is of length 16 and 32. The length of the cipher key is also 16. AES operates on a 4 × 4 matrix of bytes, known as state. If there are 16 bytes b1, b2…b16, can be represented as matrix:

$$\begin{bmatrix} b1 & b5 & b9 & b13 \\ b2 & b6 & b10 & b14 \\ b3 & b7 & b11 & b15 \\ b4 & b8 & b12 & b16 \end{bmatrix}$$

The key size used for an AES cipher specifies the number of cycles of transformation rounds that convert the input, called the plaintext, into the final output, called the cipher text. The number of cycles needed for corresponding bit keys is as follows:

- 10 cycles for 128-bit keys.
- 12 cycles for 192-bit keys.
- 14 cycles for 256-bit keys.

Each round has 4 processing steps that depends on the encryption key itself and set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key [10].

1. Key Expansions—round keys are derived from the cipher key using Rijndael's key schedule. AES requires a separate 128-bit round key block for each round plus one more.
2. Initial Round is Add Round Key—each byte of the state is combined with a block of the round key using bitwise XOR.
3. Rounds

- SubBytes—A non-linear substitution step where each byte is replaced with another according to a lookup table.

In the SubBytes step, each $a_{i,j}$ byte in the state matrix is replaced with a SubByte $S(a_{i,j})$ using an 8-bit substitution box(S box). This operation will provides the non-linearity in the cipher. S-box is constructed by combining the inverse function with an invertible affine transformation that will avoid attacks that based on algebraic properties. While performing the decryption process, inverse SubBytes step is performed, this requires first taking the affine transformation and then finding the multiplicative inverse.

- Shift Rows—A transposition step where the last three rows of the state are shifted cyclically a certain number of steps.The Shift Rows step operates on the rows of the state that cyclically shifts the bytes in every row by a certain offset. For AES, the primary row is kept unmoved. Each byte of the second row is shifted one to the left. Similarly, the third and fourth rows are shifted by offsets of two and three respectively. For blocks of sizes 128 bits and 192 bits, the shifting model is that the same. For a 256-bit block, the primary row is unchanged and also the shifting for the second, third and fourth row is 1 byte, 3 bytes and 4 bytes respectively. This alter only applies for the Rijndael cipher once used with a 256-bit block, normally AES doesn’t use 256-bit blocks.
- Mix Columns—A mixing operation that operates on the columns of the state, combining the four bytes in each column. The Mix Columns function takes four bytes as input and outputs, where each input byte affects all four output bytes. Along with ShiftRows, Mixcolumns provides diffusion within the cipher. During this operation, each column is transformed using a fixed matrix,

$$\begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{bmatrix}$$

Matrix multiplication is followed by multiplication, which will be a modulo irreducible polynomial $x^8+x^4+x^3+x+1$ and addition operation is simply XOR of the entries. Entries are 8 bit treated as coefficients of polynomial of order x^7. Addition is simply XOR.

- Add Round Key

4. Final Round (no Mix Columns)

- SubBytes
- Shift Rows
- Add Round Key.

B. Wavelet Transform

The wavelet transform is just like the Fourier transform, in which the main difference is that the Fourier transform decomposes the signal into sines and cosines, i.e. the functions localized in Fourier space, whereas within the wavelet transform, it makes use of functions that are localized in each real space and Fourier space. In standard, the wavelet transform can be expressed via the following equation:

$$F(a,b) = \int_{-\infty}^{\infty} f(x)\varphi^{*}_{(a,b)}(x)dx \qquad (1)$$

where the * is the complex conjugate symbol and φ is the scaling function. Wavelet transform will segment the signal into shifted and scaled versions of the original wavelet [9]. There are two types of wavelet transforms: discrete wavelet transform (DWT) and continuous wavelet transform (CWT).

The discrete wavelet transform (DWT) is a part of the wavelet transform using a discrete set of the wavelet scales and translations. DWT provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies. It is much like the human ear which reveals similar time-frequency resolution characteristics. The Discrete Wavelet Transform (DWT) is a special case of the WT that provides a compact representation of a signal in time and frequency that may be computed correctly. The DWT is defined by the following equation:

$$W(j,k) = \sum_{j}\sum_{k} x(k)2^{-\frac{j}{2}}\varphi(2^{-j}n - k) \qquad (2)$$

where ψ(t) is a time function with finite energy and fast decay called the mother wavelet. The DWT analysis can be performed using a fast, pyramidal algorithm related to multirate filter banks [10]. On applying wavelet transformations to the audio signal, an approximation components (low frequency) and detail components (high frequency) of the signal can be obtained.

$$y_{high}(k) = \sum_{n} x[n]g[2k - n] \qquad (3)$$

$$y_{low}(k) = \sum_{n} x[n]h[2k - n] \qquad (4)$$

where y_{high} [k], y_{low} [k] are the outputs of the high pass (g) and lowpass (h) filters by sub sampling by 2. Because of the down sampling the number of resultant wavelet coefficients is accurately the same as the number of input points. Biorthogonal wavelet are used for audio signal data embedding and extraction. By applying biorthogonal wavelets into the audio signal, it will split components into numerous frequency bands called sub bands known as,

- Approximation LL: Horizontally and vertically low pass.
- Vertical LH: Horizontally low pass and vertically high pass.
- Horizontal HL: Horizontally high pass and vertically low pass.
- Diagonal HH: Horizontally and vertically high pass.

Those subbands are used for in addition processing. HAS contains fairly small differential range i.e. noisy sounds normally have a tendency to reduce out weaker sounds. Thus HAS has a range of 20 to 20000 Hz that is sensitive to high frequency parts compared to its low frequency parts. For this reason, hiding secret message in low frequency sub band will be greater suitable. This doesn't affect the sound quality of resulting audio signal.

C. Implementation of Data Embedding

The embedding process is summarized in Fig.1, and the implementation process of data embedding is explained as follows

1. Input secret message and cover signal : Secret message is a text file, which is encrypted and embedded in to a cover signal. This cover signal should be large enough to hide the message. After selecting the input secret message and cover signal, length of the audio file as well as length of the text files needs to be calculated.

2. Encryption : The secret message should be converted into another form, normally to binary, so that the intruder will not be able to attack the secret information. Next step is to find the length of message which is assumed as N bits long. A random number is used to generate the private key of the same length as that of the message. This is performed to match the length of encrypting message to its original message.XOR operator is then applied on the AES encrypted message bits with itself to generate corresponding N bit cipher message.

3. The audio signal will be read and decomposed in to coeffients using biorthogonal wavelet.

4. Each encrypted text bits is embedded in to the coeffients of the decomposed audio signal by biorthogonal wavelet. This process will be continued until all the bits are embedded into the cover audio file Finally we can get the audio with hidden data, i.e. audio stego signal.

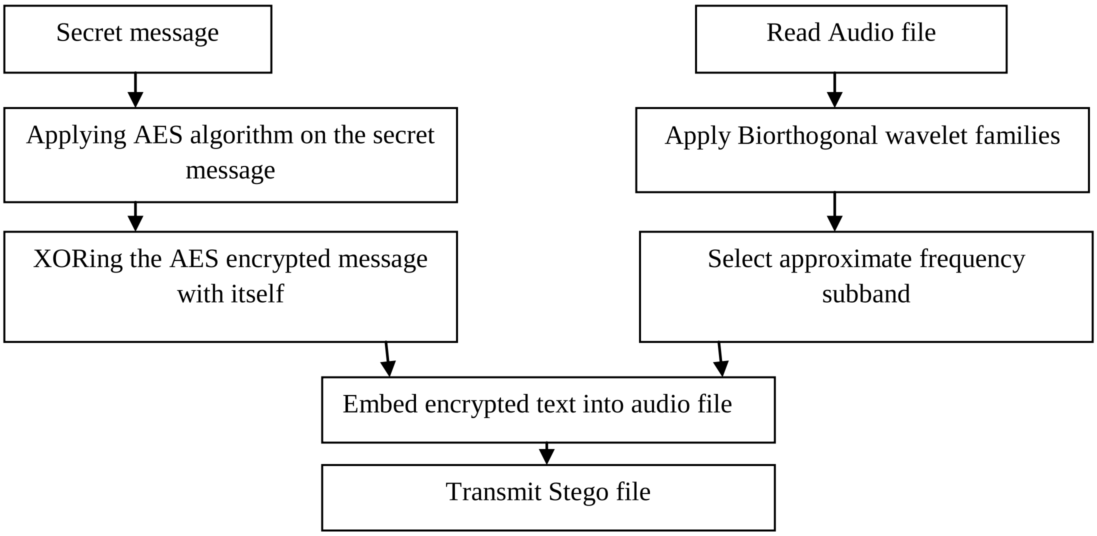
Secret message
Read Audio file
Applying AES algorithm on the secret message
Apply Biorthogonal wavelet families
XORing the AES encrypted message with itself
Select approximate frequency subband
Embed encrypted text into audio file
Transmit Stego file

Fig 1: Block Diagram for Data Embedding

D. Implementation of Data Extraction

Data extraction is the reverse process of data embedding which is summarized in Fig 2. The following steps explains this process:

1. Input Audio stego signal : Select the Audio stego signal from which data is to be extracted. This signal must be same as the one stored using the stego signal in message hiding process.

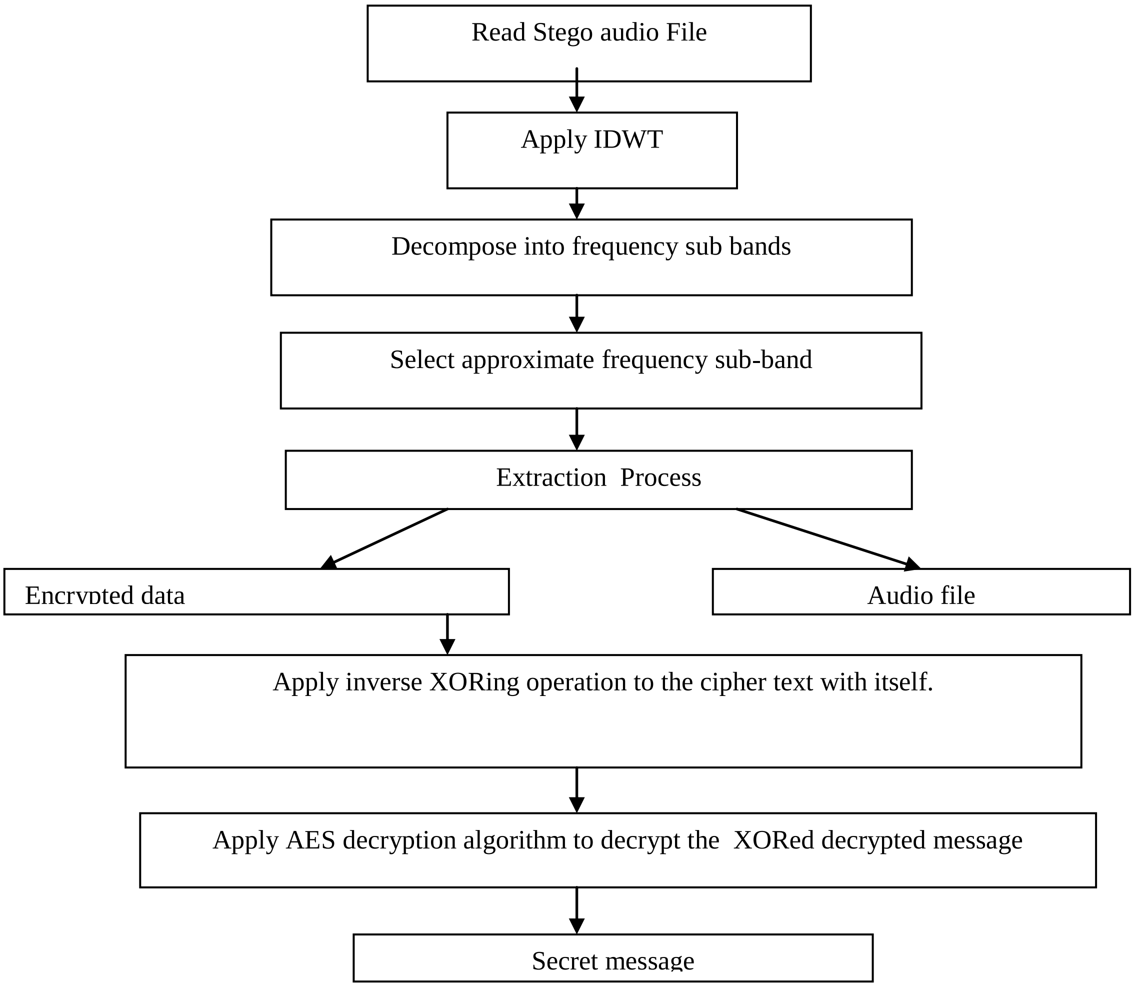
Read Stego audio File
Apply IDWT
Decompose into frequency sub bands
Select approximate frequency sub-band
Extraction Process
Encrvpted data
Audio file
Apply inverse XORing operation to the cipher text with itself.
Apply AES decryption algorithm to decrypt the XORed decrypted message
Secret message

Fig 2: Block Diagram for Data Extraction

2. Inverse DWT is applied on the stego audio signal to decompose the audio signal into low frequency and high frequency components.

3. Select the frequency sub band corresponding to the approximate components.

4. The stego signal is decomposed to obtain encrypted text and audio signal from the reverse operation of the embedding stage..

5. The encrypted text is converted back to XORed message bits using inverse XORing operation with itself. Later, AES decryption algorithm is performed to achieve the original secret message.

III. EXPERIMENTAL RESULTS

Steganographic approaches are evaluated in subjective and objective evaluation. Subjective Evaluation is performed by human’s auditory perception where the listeners are asked to identify the good quality signal from the stego audio &original audio. Objective Evaluation is mainly done by measuring the distortion level in steganographic object. The evaluation of the stego audio is done by objective measure. The main parameters used for objective evaluation are SNR, speed and MSE.

Table 1 Comparison of different cover signal with secret message “STEGANOGRAPHY123”.

Cover audio signal (.wav)	Stego Audio signal (.wav)	SNR (dB)	Processing time		MSE
			Encryption (sec)	Decryption (sec)	
1	01	79.20	0.350	2.395	0.0006
Vega	02	115.80	0.328	3.4138	0.0001
Lathe	03	75.32	1.430	5.6139	0.0041
Godzila	04	118.46	0.334	2.8498	0.0001
Violin	05	148.91	0.340	2.6254	.00011

We have considered five different audio files in this experiment to embed digital data. The embedded data is a text file namely “STEGANOGRAPHY123”. We applied Biorthogonal wavelets on cover signal and the coefficient is taken where the data is to be hidden using a pre determined threshold value T. SNRseg is defined as the average of the SNR values over short segments:

$$\mathbf{SNR}_{\mathbf{seg}} = \frac{10}{M}\sum_{m=0}^{M-1} log_{10} \sum_{i=Nm}^{Nm+N-1}\left(\frac{x^2(i)}{(x(i)-y(i))^2}\right) \qquad \mathbf{(5)}$$

where x(i) is the original audio signal, y(i) is the stego audio signal. The length of segments is typically15 to 20 ms for audio. The SNRseg is applied for frames which have energy above a specified threshold in order to avoid silence regions. Signal-to-Noise Ratio (SNR), is a special case of SNRseg, when M=1 and one segment encompasses the whole record [9].

The SNR is measured as

$$\mathbf{SNR=10log_{10}}\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2} \qquad \mathbf{(6)}$$

Table.1 shows a comparison of different cover signal with respect to SNR of Threshold T equal to zero. In these tests use 1.wav, vega.wav, lathe.wav, godzila.wav, violin.wav files as a cover signals with text file as a secret message. The results in table show the output with threshold value T equal to 0.

Figure.3 shows comparison waveforms of different output stego-signal on different input cover signal at threshold value 0.

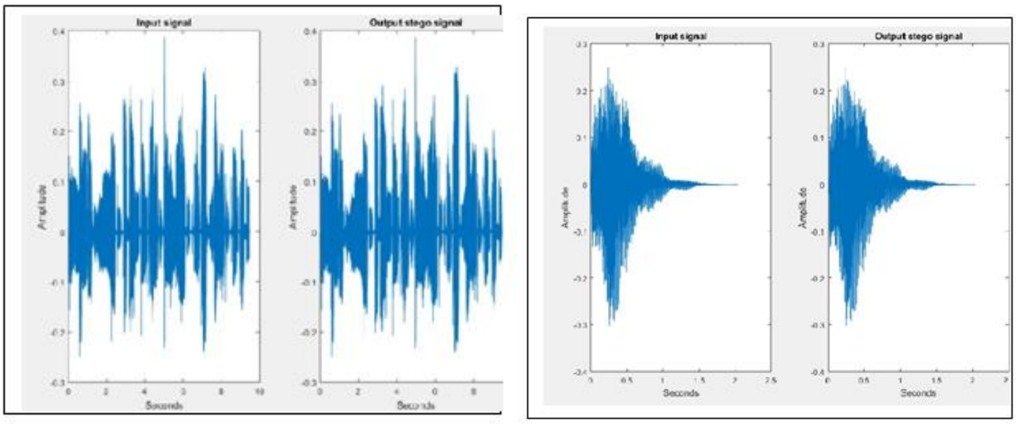

Fig3. Comparison of different audio signal & its output stego signal.

IV.CONCLUSION

This paper illustrates wavelet method used for digital audio steganography. Here, a high security and high stego-signal quality audio steganography scheme is presented. A high SNR which indicates better perceptual quality is observed. A perfect reproduction of original audio file is observed to have an MSE equal to zero. A stego audio file greatly differs from the original audio when MSE has large value. Biorthogonal wavelet is the apt choice because the wavelet coefficients obtained is even numbered for embedding as we are using AES algorithm which produces the data in even(16,32) i.e. embedding process of data should be done in lower number of rows extension for the best result using wavelet family.

REFERENCES

[1] Mirza Tabinda and Vijaya Ahire, “Analysis of wavelet families on Audio steganography using AES”, International Journal of Advances in Computer Science and Technology (IJACST), Vol.3, No.11, pp no : 26-31, 2014.

[2] Satish Singh Verma, Ravindra Gupta and Gaurav Shrivastava, “A Novel Technique for Data Hiding in Audio Carrier by Using Sample Comparison in DWT Domain”, 2014 Fourth International Conference on Communication Systems and Network Technologies, pp no: 639-643.

[3] Ashwini Mane, Gajanan Galshetwar and Amutha Jeyakumar, “Data Hiding Technique : Audio Steganography using LSB Technique”, International Journal of Engineering Research and Applications (IJERA),Vol. 2, Issue 3, pp no:1123-1125, May-Jun 2012.

[4] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim and Habib Hamam, “Comparative study of digital audio steganography techniques”, EURASIP Journal on Audio, Speech, and Music Processing, pp no: 2-5, 2012.

[5] Rohit Tanwar, Bhasker Sharma and Sona Malhotra, “A Robust Substitution Technique to implement Audio Steganography”, International Conference on Reliability, Optimization and Information Technology - ICROIT, pp no: 6-8, February 2014.

[6] Prof. Samir Kumar Bandyopadhyay and Barnali Gupta Banik, "LSB Modification and Phase Encoding Technique of Audio Steganography Revisited", International Journal of Advanced Research in Computer and Communication Engineering Vol. I, Issue 4, June 2012.

[7] Siwar Rekik, Driss Guerchi, Sid-Ahmed Selouani and Habib Hamam, "Speech steganography using wavelet and Fourier transforms", EURASIP Journal on Audio, Speech, and Music Processing, pp no:1to20, 2012.

[8] Neha Gupta and Ms. Nidhi Sharma, "Dwt and LSB Based Audio Steganography", International Conference on Reliability, Optimization and Information Technology - ICROIT, pp no: 6-8, 2014.

[9] Michael Weeks, "Digital Signal Processing Using MATLAB and Wavelets", Pearson publications, ISBN, pp no:81-297.

[10] S.G Mallat, "A Theory for Multi resolution Signal Decomposition: The Wavelet Representation" IEEE. Transactions on Pattern Analysis and Machine Intelligence,Vol.11, 674-693, 1989.