SPEECH ENCRYPTION USING FAST FOURIER TRANSFORM TECHNIQUES

S.Sridharan* , E.Dawson** , and J.O'Sullivan#

*School of Electrical& Electronic Systems Engineering
Queensland Institute of Technology

**Department of Mathematics
Queensland Institute of Technology

#Division of Radio Physics
CSIRO, Sydney

ABSTRACT - A speech encryption system based on permutation of FFT coefficients is described. Results of simulation and cryptanalysis of the system are presented.

INTRODUCTION

Scramblers are needed to ensure privacy in speech transmission in radio communication, the telephone network and in the emerging cellular mobile radio systems. In a typical speech scrambler, speech is first digitized and the digital sequence is scrambled. The scrambled sequence is converted back to the analog form for transmission. If it is ensured that the scrambled signal occupies the same bandwidth as the original speech signal, then it can be used with existing telephone, satellite and mobile communication systems.

The above approach to scrambling is referred to as analog scrambling even though digital techniques are used in the implementation. A totally digital encryption is also possible but is not attractive at present due to the inability of such a system to retain naturalness of speech and the individual voice characteristics (Lee & Chou, 1986).

Some of the speech scramblers currently available use time domain techniques. A survey of the options that are available for time domain encryption are described in Mitchell & Piper (1985). An alternative technique of speech scrambling which has recently received considerable attention makes use of the Fast Fourier Transform (FFT) technique and performs scrambling in the frequency domain (Sakurai, Koga & Muratani, 1984). In this method the DFT coefficients of frames of speech are permuted. The interest in this method stems from the availability of fast FFT processors such as the AUSTEK A4162 which can perform a 256 point complex FFT in 0.2 ms. The aim of this paper is to present the findings of the research that has been conducted on a speech encryption scheme in which the DFT coefficients are permuted. Theoretical and simulation results of the following aspects of the system are presented :

(1). Choice and generation of the permutation matrices
(2). Synchronization and effects of channel distortion
(3). Residual intelligibility and recovered speech quality
(4). Cryptanalysis of the system

FORMULATION OF THE FREQUENCY DOMAIN SCRAMBLING PROCESS

Let x represent a sampled time speech vector of length N and F represent the N by N Discrete Fourier Transform (DFT) matrix. Consider the DFT of x given by

$$u = Fx$$

The scrambling is performed by a permutation matrix P applied on the speech DFT vector u to produce a vector v

$$v = Pu$$

The scrambled speech for transmission in the time domain y is obtained by applying the inverse transformation F^{-1} on v

$$y = F^{-1}v$$

CHOICE AND GENERATION OF THE PERMUTATION MATRICES

In order that the bandwidth of the system does not increase due to the encryption process, the permutation is restricted to M DFT coefficients lying within the speech band 300-3000Hz. The number of possible permutations is M!. Not all of these permutations can be used since some of them leave considerable residual intelligibility in the scrambled signal. The proposed system derives its cryptanalytic strength from frequent variation of the permutation matrix in addition to the nominal key cardinality expressed by M!.

An efficient method of generating the permutation matrices is to use the method proposed by Sloane (1983). This method enables all M! permutations to be generated from a random number seed lying between 0 and 1. We use these random numbers as the keys for the encryption process. The random numbers are generated using a nonlinear combination of shift register sequences similar to that described in Asenstorfer, Gray & Dawson (1987).

Since a large amount of memory space is needed to store the permutations it is economical to generate the permutations in real time. It is also important to test the selected permutations in order to screen those which do not sufficiently destroy the intelligibility of speech. Even though intelligibility is a substantially subjective matter it is possible to use objective tests which are useful (if not ideal) indicators of intelligibility loss. These tests are applied to the permutation matrices to eliminate those permutations which do not significantly reduce intelligibility. The tests that we apply to screen permutations are the LPC distance measure (Gray & Markel (1976), Hamming distance measure and the Euclidean distance measure (Ecker 1985). A threshold is set for each distance measure and permutations which fall below threshold are discarded. Note that for the LPC measure a standard speech segment must be used and scrambling actually performed to evaluate the distance.

The generation and the selection of the permutation requires considerable processing time. However this processing can be carriedout before the call is set up and therefore does not cause any delay during the conversation. For each call 16 permutations are selected using the key. These permutation are stored in a RAM and used to permute 16 contiguous speech segments. The permutations are reused on a 16 frame multiframe basis.

SYNCHRONIZATION

Synchronization in the proposed system is achieved by sending a codeword at approximately the start of each multi-frame. The codeword is frequency shift keyed in order to lie in the voice band and is inserted in the first speech silence occurring after 16 frames. We search for this silent frame over the next 16 frames (ie approximately 0.5 sec). It is very unlikely that a silent frame will not be found over 0.5 sec of speech. In any case, if no silent frame is found after the search we clear the 33rd frame and insert the synchronization code. Thus it is ensured that a synchronization sequence is present at least every second.

QUALITY OF THE RECEIVED SIGNAL

The quality of the received signal depends on the following factors:(1) bandwidth restriction of channel.(2) finite wordlength effects of FFT and IFFT implementations.(3) amplitude, phase and group delay distortion of channel. Each of these factors will be considered in turn.

(1) In order that the scrambling process does not increase the bandwidth, permutation is restricted to DFT components lying between 300 Hz and 3000 Hz. The components lying below 300 Hz and above 3000 Hz must be set to zero since they are not permuted. If the number of frequency components N is large then no perceptible distortion in the received speech occur due to the bandwidth restriction. However, if the value of N is reduced to less than 128, deterioration in the quality of recovered signal is perceived. Reduction of N also reduces the cryptanalytic strength of the system. However, the delay of the scrambling process increases as N is increased. As a result of this trade-off between quality and processing delay, a frame length of N=256 has been chosen in the system. The corresponding value of M is 87.

(3) Finite wordlength of signal and twiddle factors could introduce significant noise in the system. It can be shown that for radix 2 FFT with power-of-two scaling between stages the noise increases as half bit per stage. Thus for 256 point FFT as much as 4 bits will be lost due to roundoff noise. This figure indicates that one has to use at least 16 bits representation for the signal to obtain a reasonable SNR.

(4) A telephone channel over which the scrambled speech is transmitted causes significant distortion due to group delay characteristics of the channel. However, this distortion may be overcome by using an equalizer. Adaptive update of the equalizer taps may be carried out using Widrow's algorithm (Widrow & Stearn, 1984).

CRYPTANALYSIS OF THE SYSTEM

We will assume that the attacker has complete knowledge of the system,and has the necessary hardware to synchronize and perform FFT. Thus the security of the system will be assumed to reside entirely within the selection of the key. A brute force attack on the system would then be to generate all M! possible inverse permutations and is obviously impractical. For example for M=87 there are approximately $2.1*10^{132}$ different permutations to be tried.

If the attacker knew the original speech x and the corresponding scrambled speech y for N frames then the attack proposed by Diffie & Hellman (1979) can be used. Since the attacker knows the plaintext-ciphertext pair $\{x_i, y_i\}$ for i $=\{1,2....N\}$ the matrix F^1PF relating x and y can be found by the inversion of a matrix of order N. In the proposed system we use at least 16 different permutation matrices to encrypt a multiframe of speech. These permutation matrices are selected before a call is set up based on the key. The cryptanalytic strength of this system is greater than for the constant permutation system. In fact the number of plaintext-ciphertext samples required to attack this system is $(16N^2)$. This corresponds to access of $(16.N^2).125 \cdot 10^{-6}$ sec (ie. around 3 mins) of original and scrambled speech.

A fact that is ignored in the above comments on cryptanalysis of the system is that speech has a considerable amount of redundancy. Thus the adjacent sample correlation may enable a practical cryptanalytic attack on the system. Further work on the cryptanalytic strength of the system taking into account characteristics of speech needs to be carried out.

SIMULATION RESULTS

Some simulation results are presented in Fig. 1 to 4. Fig 1 and 2 show the raw spectrum of the original and encrypted speech for a given speech segment. Fig 3 and 4 show the LPC based smoothed spectrum for the original and encrypted speech.

CONCLUSIONS

An encryption system is described in which the DFT coefficients of speech segments are permuted to destroy speech intelligibility. There are several advantages to be gained by using this frequency domain permutation approach compared to the time domain permutation approaches. It can be easily shown that the frequency domain permutation of a DFT frame is equivalent to a linear combination of the samples of the corresponding frame in the time domain. It follows that the encrypted speech signal cannot be descrambled by an inverse permutation in the time domain and this property gives extra cryptanalytic strength to the scheme. Furthermore our investigations reveal that the processing delay of this approach is significantly less than time domain schemes since frames short as 32msec will give adequate security and voice quality. A typical time domain permutation system requires a frame time of around 256 msec Mitchell & Piper (1985). Intelligibility tests carried out by simulation of the system indicates significantly less residual intelligibility compared to time domain techniques. Fig 1 to 4 illustrate the effectiveness of the proposed scheme in destroying formant and pitch information of speech. A simple cryptanalysis of the system indicates that an attack of the system requires at least 3 minutes of original and encrypted speech together with synchronization information. The proposed system is currently being implemented using Austek A4162 FFT processor interfaced to TMS320C25 signal processor.

REFERENCES

Asenstorfer,J.A, Gray,P., and Dawson,E.P. *(1987) Nonlinear Shift registers for use in cryptographic processes*, Digest of papers of IREECON 87, pp. 395-398.

Diffie,W.,and Hellman, M.E.(1979) Privacy and authentication: An introduction to cryptography, Proceedings of the IEEE, vol.67, No.3, March, pp.397-427.

Ecker, A. (1985) *Time division multiplexing scramblers: selecting permutations and testing the system*, Lecture notes in Computer Science, Advances in Cryptography.

Gray, A.H., and Markel, J.D., (1976), *Distance measures for speech processing*, IEEE Transactions on Acoustics Speech, and signal processing, ASSP-24, No.5, October.

Lee, L.S and Chou, G.C (1986) *A general theory of asynchronous speech encryption techniques*, IEEE Journal on selected areas in communication, vol SAC-4, No. 2, March pp. 280-287.

Mitchell, C.J and Piper, F.C (1985) *A classification of time element speech scramblers*, Journal of Institution of Electronic and Radio Engineers, vol. 55, No. 11/22, pp.391-396, Nov/Dec.

Sakurai, K., Koga, T. and Muratani, T. (1984) *A speech scrambler using Fast Fourier Transform Techniques*, IEEE Selected Areas in Communications, vol SAC-2, No. 3, May.

Sloane,N.J.A (1983) *Encryption by random rotations*, Lecture notes in Computer Science, Number 149, Spring-Verlang, pp.71-128.

Widrow,B., and Stearns, S.D., (1984), *Adaptive signal processing*, Prentice-Hall Englewood Cliffs.

ACKNOWLEDGEMENT

This work was supported by CSIRO/QIT Collaborative Research Grant 89/7.

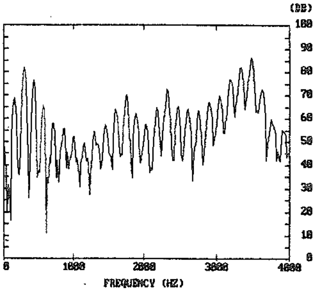

Fig. 1 Raw spectrum of original speech

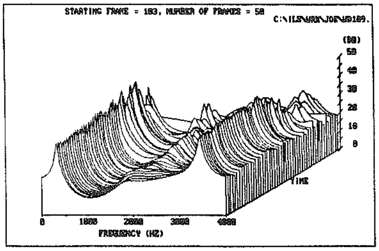

Fig. 3 LPC smoothed spectrum of original speech

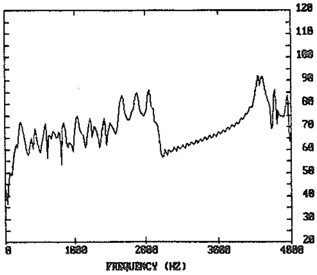

Fig. 2 Raw spectrum of encrypted speech

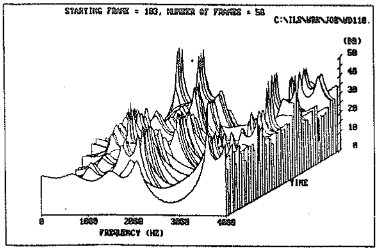

Fig. 4 LPC smoothed spectrum of encrypted speech