A COMPARISON OF THE CRYPTOGRAPHIC REQUIREMENTS FOR DIGITAL SECURE SPEECH SYSTEMS OPERATING AT DIFFERENT BIT RATES

C J Mitchell

Racal-Comsec Limited, UK

1. INTRODUCTION

Speech is one of the most fundamental forms of communication available to mankind. Because of its continuing importance in modern communication systems, protecting confidential conversations remains a problem of great importance.

There are, essentially, two fundamental methods for securing speech communications, namely "analogue" and "digital". These two words describe the different forms of output for these two types of security device. In an analogue encryptor, the output is an analogue signal which is a transform of the original speech signal (where this transform normally takes place in the time and/or frequency domains), whereas the output from a digital encryptor is simply a sequence of binary digits.

The chief advantage of using an analogue encryptor is that (by a suitable choice for the encrypting transformation) the scrambled signal can be kept to the same bandwidth as the original speech signal. If this is the case, then the same transmission medium can be used for the encrypted signal as was used for the original "clear" speech signal. This has obvious advantages, especially if, as can be arranged, the encrypted speech signal has essentially the same robustness against channel noise as does the clear speech. However, because the transmitted signal is a transform of the original, highly redundant speech signal, traces inevitably remain of the original speech in the scrambled version, and the transmitted signal is definitely not "noise-like". Analogue systems are not discussed further here, and for further information the reader is referred to [2] chapter 9, [3], [5] and [6] .

Thus, to achieve the highest possible level of security it is necessary to use a digital encryption system. Such systems use an analogue to digital converter to change the speech from an analogue signal into a sequence of bits. This sequence of bits is then enciphered (using similar techniques to those employed in data encryption systems) and transmitted as an encrypted sequence of bits. For further reading on digital speech encryption systems see [1] , [2] chapter 9, [4] and [6].

The bit rate of a digital speech system (i.e the number of bits transmitted by a device in every second of operation) depends on the analogue to digital (A/D) method used; this in turn affects the available options for the encryption system. The different problems associated with operating at low bit rates as compared with medium bit rates form the subject matter for this paper. In particular we consider and compare the difficulties associated with designing a digital speech encryptor operating at 2400 bits/second (b/s) and a digital speech encryptor operating at 16000 b/s.

2. ANALOGUE TO DIGITAL CONVERSION TECHNIQUES

An A/D converter operating at a rate as low as 2400 b/s is by necessity a sophisticated device such as a vocoder; the low bit rate system considered here will therefore be referred to as System V. An A/D system operating at 16000 b/s could be a much simpler system such as a delta-modulator, and we will thus refer to this system as System D. Thus purely from the cost and complexity point of view, the higher bit rate system (System D) is at a distinct advantage. It is assumed for the purposes of this paper that System V uses a vocoder operating at 2400 b/s and that System D uses a delta-modulator operating at 16000 b/s.

The primary advantage of System V, i.e. the system with a 2400 b/s bit rate, is that data at this rate can be communicated using a narrow band channel, given that a suitably sophisticated modem is available. In many cases this means that a channel that is normally used for clear speech (e.g. a standard telephone line or a narrow band HF link) can also be used for digitally encrypted speech; the ability to send secure speech through such channels is something that is normally associated only with analogue speech security devices. This is a very important advantage, and perhaps the major reason for the existence of vocoders.

There is associated with System V an inevitable decrease in the redundancy contained in the transmitted data as compared with higher bit rate systems such as System D. This in turn affects the performance of the systems in differing channel noise conditions. For example, for a typical vocoder, performance starts to deteriorate at an error rate of around 1%. If the bit error probability is as high as 2% then noticeable phonic corruption occurs; at error rates of 3-4% the vocoder synchronisation becomes unreliable and even when the vocoder is in synchronism the recovered speech is virtually unintelligible. However, by adding a modest amount of error protection redundancy to the output from a vocoder (say 25%, which would increase 2400 b/s to 3000 b/s) such a system could be used over channels with a significantly higher error rate. The increase in bit rate is small enough for the system to remain usable over most normal audio circuits.

However using a typical 16000 b/s delta-modulator, at an error rate of 1% there is only a slight increase in background noise in the demodulated speech, and even at error rates as high as 10%, although the background noise is comparable with that of a noisy HF link, the recovered speech is still perfectly comprehensible.

It is clear that both systems can tolerate a certain error rate. In addition note that relatively small changes in the error rate can make the difference between perfectly adequate communications and no communication at all. This is especially true for System V, where, as noted above, doubling the error rate from 1-2% to 2-4% can change the quality of the recovered speech from virtually unaffected to almost completely unintelligible. Thus it is worth stressing that, for both System V and System D, any cipher system which propagates errors will have the effect of making certain channels, which are usable for clear communications, unusable for secure communications.

3. DIGITAL ENCRYPTION TECHNIQUES

3.1 Basic Concepts

In this section some of the terms that are to be used are defined. The heart of every cryptographic system is its encryption algorithm. In order to decide how the clear information is to be encrypted, such algorithms require one or more secret keys. The choice

of algorithm is clearly of fundamental importance; the requirements for such algorithms are not discussed here, the interested reader is referred to [2] sections 4.4 and 4.5. However, the choice of algorithm must take into account the amount of structure and redundancy in the information being enciphered; as noted above, vocoder output contains substantially less redundancy than delta-modulator output, and so the requirements for the algorithm may vary between the two systems.

As in [2] the term "base key" is used here for a key which is selected by the user and changed by the user at regular intervals; it will most probably be used to encipher a number of different messages. In conjunction with this base key it is also essential to use some form of "message key", i.e. a key which is different for each message and is normally generated by the encryption device (and is transmitted to the receiver at the start of an encrypted transmission). For a discussion of why such a key is essential see, for example, [2] section 8.2.

In addition it may often be convenient to use a key which changes very infrequently and actually changes the function performed by the algorithm. Such keys are usually included in order to allow the user to "customise" commercially available equipment, and are commonly referred to as "customer option" keys.

3.2 Cipher Feedback Encryptors

One type of encryption system that could be used for a digital speech encryption system is cipher feedback. A cipher feedback system is shown in block diagram form in Figure 1.

The major advantage of cipher feedback systems is that they are "self-synchronising", i.e. no separate synchronisation signal is required. This is because the bit used to encipher (or decipher) any bit of plaintext (or ciphertext) is a function only of the base key and the previous n bits of ciphertext, and not of its position within the ciphertext.

However this major advantage is also intimately associated with the main disadvantage of cipher feedback systems, namely error propagation. If any bit of the enciphered speech is received in error, not only will that bit be incorrect after deciphering, but in addition between 0 and n of the next n bits will also be in error after decryption. This is because the bit used to decipher any received bit is a function of the base key and the previous n ciphertext bits. Hence, on average, each bit in error in transmission causes (n/2)+1 errors in the deciphered data (e.g. if n=20, then one error in transmission will cause around 11 errors in the deciphered bit sequence!).

This error propagation can easily degrade the communications channel to such an extent that in clear operation speech seems virtually unaffected by errors, whereas it is not possible to communicate in secure. As noted above, when using a vocoder over a channel with a 1-2% error rate, a simple doubling of the error rate can mean the difference between completely satisfactory operation and complete loss of vocoder synchronisation. As far as the user is concerned, for radio links this can mean quite a severe range reduction.

Because of this major problem, if the systems are ever to be used over non-perfect channels (which is normally the case with telephone and radio links) then cipher feedback must be rejected as a possible encryption scheme for both System V and System D. This is especially true for System V where the A/D method is particularly sensitive to relatively small increases in the error rate.

3.3 Stream Cipher Systems

The other widely used technique for enciphering digitised speech is stream ciphering. Note that block ciphers are not particularly suitable for speech encryption because they not only propagate errors but also introduce a time delay. A stream cipher system is shown in block diagram form in Figure 2.

Stream ciphers require a separate synchronisation signal. This is because the bit used to encipher (or decipher) any bit of plaintext (or ciphertext) is a function of the base key, message key and the position of the bit within the ciphertext.

The major advantage of stream ciphers is that they do not propagate errors. If any bit of encrypted speech is received in error, then that is the only bit that will be incorrect after decryption. Thus, in principle at least, the use of a stream cipher technique makes feasible a digital speech cipher which will enable secure communication over any channel which allows clear communication at the required bit rate.

The only remaining problem concerns obtaining cryptographic synchronisation; this problem is not to be underestimated. The purpose of synchronisation is to ensure that the encryptor's and decryptor's sequence generators (see Fig. 2) are running "in step". The problem of obtaining synchronisation can be solved however, and we consider possible methods for achieving synchronism in Section 4 below.

It is clear from the above discussion that for most, if not all, practical applications, encryption techniques which propagate errors have clear user disadvantages. Thus for both System V and System D, given that the synchronisation problems can be overcome, a stream cipher system (or some other encryption system which does not propagate errors) is much to be preferred.

4. METHODS OF CRYPTOGRAPHIC SYNCHRONISATION

As detailed above, if a stream cipher technique is to be used to secure the digital speech, a separate cryptographic synchronisation scheme is required. There are two types of synchronisation scheme in common use, namely "Initial Synchronisation" (or "Single Shot") systems and "Continuous Synchronisation" systems. These systems are discussed and their respective pros and cons are outlined in 4.1 and 4.2 below.

4.1 Initial Synchronisation

In these systems the transmitting device sends a single block of synchronisation information at the start of every transmission. This synchronisation block also needs either to contain or be immediately followed by the message key which provides one of the inputs to the sequence generator. Thereafter the receiving (deciphering) unit uses a crystal controlled clock, together with some form of timing recovery system, to keep it in step with the transmitting (enciphering) unit.

The chief advantage of this system is that, once the synchronisation block has been transmitted, no extra synchronisation information is sent, and the channel can be dedicated to sending encrypted speech information. This fact, combined with the use of a stream cipher (or similar non-error-propagating system), means that the channel will have exactly the same characteristics for secure transmissions as for clear transmissions; i.e. if the channel is good enough to be usable in clear then it will be good enough to be used in secure (given that the initial synchronisation block is received).

Because of the importance of the initial synchronisation (and the need to transmit the message key correctly) this block of data must be protected

against transmission errors. The block must contain sufficient error protection redundancy to ensure that not only is it detected but also that the message key is received correctly even in very bad error conditions. Certainly the synchronisation system should work with a very high probability in all those error conditions in which there is even the remotest chance of receiving intelligible speech.

The primary disadvantage of initial synchronisation systems is that they do not allow "late entry". That is, if a receiving device for some reason "misses" the initial synchronisation (for example a third party joining a net) then it is necessary for the user of the receiving device to communicate to the transmitting device that a new synchronisation block is required.

4.2 Continuous Synchronisation

In a continuous synchronisation system, the encrypted output sequence is periodically interrupted for the transmission of synchronisation updates; the period between updates would normally be of the order of 1-2 seconds and the length of the update synchronisation block would typically be of the order of 100-300 bits.

Every synchronisation update contains within it (or directly associated with it) a new message key, so that a receiver's sequence generator can be reloaded with the same (new) set of keys as are being used by the transmitting unit. Thus on receipt of any synchronisation block a receiving unit is immediately able to start decrypting the digitised speech data following the synchronisation block.

A desirable feature for continuous synchronisation systems is for each new message key to be a function of the previous one (the initial message key being randomly generated). This means that even if a receiving unit fails to detect a synchronisation update (or uses some inbuilt redundancy to detect that the update has been received in error) then the "expected" message key can be used to update the receiver's sequence generator and decryption can continue. This process is commonly known as "fly-wheeling".

It can be seen that, for a continuous synchronisation scheme, redundancy should be added to each synchronisation update to ensure that if it is received in error then the receiver will not use the corrupted message key but will use the "expected" message key; this prevents temporary loss of communication.

One of the main advantages of continuous synchronisation is that it offers a late entry facility. It offers an important additional advantage over single shot systems when the transmission channel is subject to varying noise conditions such as fades. If the channel is especially bad during the initial synchronisation, then the entire transmission will be lost, whereas with a continuous synchronisation system, even if the first few synchronisation signals are missed, as soon as a receiving unit detects a synchronisation update it can commence decrypting the enciphered speech.

The main disadvantage of continuous synchronisation is that "holes" have to be created in the encrypted binary sequence in order to make room for the blocks of synchronisation information. This can mean a degradation in the recovered speech quality, the actual effect depends to a very large extent on the A/D technique being employed.

4.3 Synchronisation Methods for System V and System D

First consider System V. As was described above, because of the advantages that it offers (in particular late entry), continuous synchronisation is normally to be preferred to initial synchronisation. However, to use continuous synchronisation it is necessary to "lose" blocks of data every so often, commonly every 1-2 seconds. For a vocoder system, the loss of even a short block could cause the receiving vocoder's synchronisation to be temporarily lost; this would result in a short but noticeable interruption in the recovered speech. So System V is obliged to use initial synchronisation because it does not "lose" any of the data.

With System D the situation is somewhat different. The insertion of a block of say 200 bits once every 1-2 seconds can be treated at the receiver so as to not have a significant effect on the quality of the recovered speech. What can be done for example is to repeat the previously received block of 200 digitised speech bits in order to fill in the gap left at the receiver by the transmission of the synchronisation block. With practical trials carried out using a 16000 b/s delta-modulator with 224-bit updates occurring at intervals between 1.0 and 1.5 seconds, such processing had a negligible effect on the recovered speech quality.

Thus, because of the advantages of continous synchronisation, this technique would seem to be the more logical one to use for System D. To minimise the speech degradation effects, the interval between synchronisation updates needs to be maximised; contrariwise the maximum time required for a receiver to obtain synchronisation and start deciphering speech data also needs to be minimised, and this time period is equal to the maximum delay between synchronisation updates. As a good compromise between these two conflicting requirements, a delay between synchronisation blocks of the order of 1-2 seconds would seem reasonable.

5. SUMMARY AND EXAMPLES

In the discussions above various conclusions have been drawn as to the methods of encryption and synchronisation which seem optimal for the two systems. These conclusions are summarised in Table 1.

As can be seen from the table, implementations of System D require relatively unsophisticated A/D circuitry. A 16000 b/s digital secure speech system can, however, only be used over wide band channels such as VHF radio links, and such links are normally FM. For an FM link no modem is required, and thus, apart from the radio itself and the encryption and synchronisation circuitry, the whole system only requires a relatively unsophisticated A/D converter.

Such a converter (e.g. a delta-modulator) can be built into a single small "encryption module" containing everything required to "secure" a VHF radio. An example of such a device is provided by the Racal Comsec MA4263 encryption module, a picture of which is given in Figure 3, illustrating the small size of the device. This small size is achieved by using customised LSI circuits to miniaturise the sophisticated encryption and synchronisation hardware. Because of its small size and relatively low cost implementation, the MA4263 can be built into hand portable VHF radios.

The MA4263 uses a delta-modulator circuit for its A/D conversion, and encrypts the digitised speech data using a stream cipher technique. Synchronisation is achieved by the transmission of short synchronisation blocks at pseudo-random intervals (the period between sychronisation blocks varies between approximately 1.0 and 1.5 seconds). The synchronisation blocks themselves are masked in order that they also appear pseudo-random. The combination of pseudo-random timing and pseudo-random appearance of synchronisation updates makes selective synchronisation jamming extremely difficult.

System V implementations have rather different applications. Because the A/D circuitry is far more

TABLE 1 - A comparison of the two systems

	System V	System D
A/D technique	A/D used is a vocoder operating at 2400 b/s, (a relatively sophisticated technique)	A/D used is a delta-modulator operating at 16000 b/s, (a relatively simple technique)
Type of channel	Suitable for narrow band channels (given a suitably sophisticated modem).	Suitable only for wide band channels (does not require a sophisticated modem).
Error tolerance	Maximum error rate A/D can tolerate is around 2%	Maximum error rate A/D can tolerate is around 10-15%
Encryption technique	Stream cipher or other encryption system which does not propagate errors.	Stream cipher or other encryption system which does not propagate errors.
Synchronisation system	Single shot synchronisation technique (no interference with channel).	Continuous synchronisation normally preferable (minimal interference with channel).

complex, and thus more expensive, than for System D, a device implementing System V would only normally be used on narrow band links. In addition it would be fair to say that with current techniques the speech quality available from a fairly simple 16000 b/s A/D is probably superior to that available from even the most sophisticated 2400 b/s vocoder system.

Because of its narrow band application, a System V implementation would also require a sophisticated modem in order to be able to transmit the required bit rate through the channel (e.g. a telephone or HF link). Both the vocoder and the modem are likely to be not only relatively expensive but also quite bulky, and so miniaturisation of the System V implementation is of lesser importance.

An example of a digital speech encryption device incorporating the features of System V is the Racal Comsec MA4433 full duplex vocoder encryptor. This unit incorporates a stream cipher encryption algorithm together with an initial "single shot" synchronisation system. Given that synchronisation is not lost this means that the encryption system can be used over all links for which clear communication is viable.

The MA4433 encryption unit is housed in a half 19" rack width extruded aluminium case; this unit includes both the encryption and synchronisation circuitry. Using a Racal Milgo Phoneplex 24 Vocoder, and a Racal Milgo 2400 b/s full duplex, 2-wire, telephone modem, a complete full duplex telephone encryption system can be packaged into a 19" rack width configuration 5U high.

REFERENCES

1. Beker, H.J., 1980, "Cryptographic Requirements for Digital Secure Speech Systems". Electronic Engineering, 52 no. 634, 37-46.

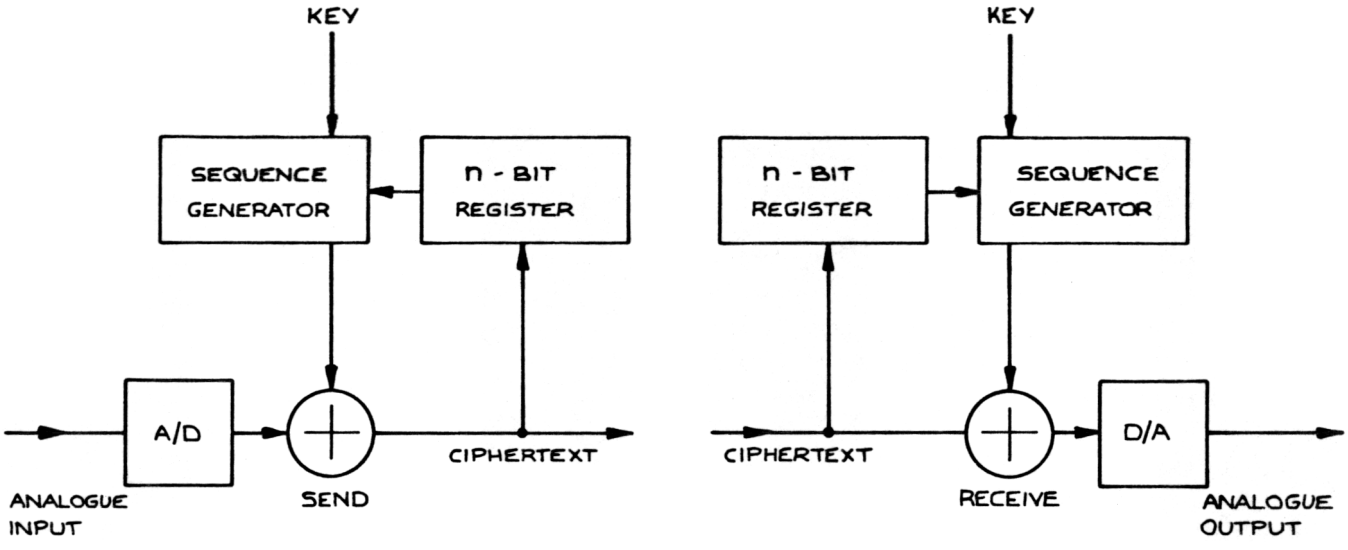
2. Beker, H.J., and Piper, F.C., 1982, "Cipher Systems".Northwood Books, London, England.

3. Beker, H.J., and Piper, F.C., 1982, "Analogue Speech Scrambling". New Electronics, 15 no. 17, 28-32.

4. Beker, H.J., and Piper, F.C., 1982, "Digital Speech Scrambling". New Electronics, 15 no. 18, 94-100.

5. Jayant, N.S., 1982, "Analog Scramblers for Speech Privacy". Computers and Security, 1, 275-289.

6. MacKinnon, N.R.F., 1980. "The Development of Speech Encipherment". The Radio and Electronic Engineer, 50, 147-155.



igure 1 A Cipher feedback system

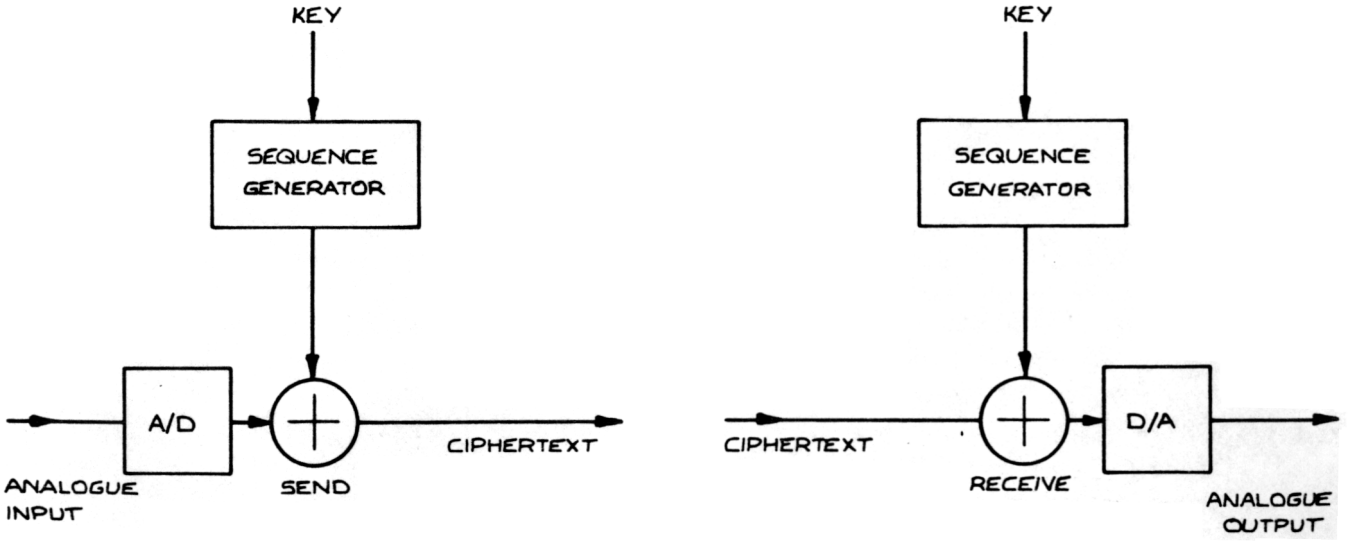


Figure 2 A Stream Cipher

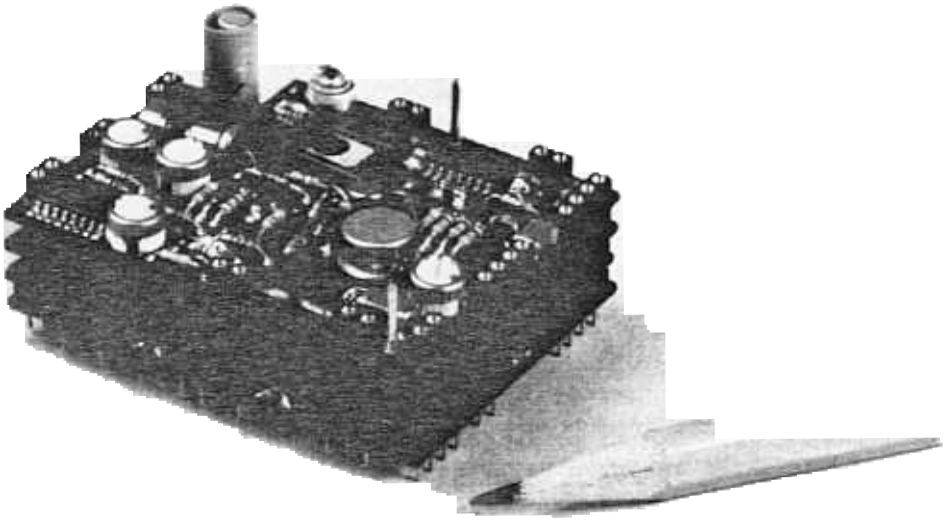


Figure 3 Racal Comsec MA4263 encryption module