REAL-TIME END-TO-END SECURE VOICE COMMUNICATIONS OVER GSM VOICE CHANNEL

N.N. Katugampala, K.T. Al-Naimi, S. Villette, and A.M. Kondoz

Centre for Communication Systems Research, University of Surrey
Guildford, GU2 7XH, United Kingdom
phone: + 44 1483 689843, fax: +44 1483 686011, email: a.kondoz@surrey.ac.uk
web: www.ee.surrey.ac.uk/CCSR/

ABSTRACT

GSM is the most wide spread mobile communications system in the world. However the security of the GSM voice traffic is not guaranteed especially over the core network. It is highly desirable to have end-to-end secure communications over the GSM voice channel. In order to achieve end-to-end security, speech must be encrypted before it enters the GSM network. A modulation scheme that enables the transmission of encrypted voice and data over the GSM voice channel was designed[1]. A real-time prototype is implemented demonstrating the end-to-end secure voice communications over the GSM voice channel.

The modem technology presented facilitates the transmission of encrypted data and an encryption algorithm is not specified. The users may choose an algorithm and a hardware platform as necessary.

1. INTRODUCTION

The GSM system ensures subscriber identity confidentiality, subscriber authentication as well as confidentiality of user traffic and signalling. The ciphering algorithms used in GSM [1] have proved to be effective in ensuring traffic confidentiality. However the traffic confidentiality is only ensured across the radio access channel. Voice traffic is transmitted across the core circuit switched networks ‘in clear’ in the form of PCM or ADPCM speech which opens up the possibility of unauthorised access to GSM-to-GSM or GSM-to-PSTN conversations. Moreover, the encryption on the GSM speech channel is optional and controlled by the network operator, not the end user. Control by the end user may be preferable in some applications. For guaranteed end-to-end security the speech signal must be encrypted before entering the communications system.

Although the GSM data channel can be used for encrypted speech transmission, this approach suffers from a number of disadvantages. The GSM data channel has interoperability problems especially across the international networks [2]. The GSM data channel typically requires 28-31 seconds to establish a connection, of which approximately 18 seconds are taken up by the GSM modem handshaking time. In addition, the GSM data channel uses Automatic Repeat Request (ARQ) for error correction and has zero errors at the expense of increased delay. The average round-trip time of the GSM data channel is 0.5 seconds [3]. This value depends upon the size of the packets transmitted. In practice this translates into a delay, which exceeds the ITU-T specifications for one-way transmission times of 150ms for telephony services [4]. Although the proposed 3GPP standards specify the provision of low-latency data bearer channels, which could be used for end-to-end secure communications or telemetry operations, the deployment dates of such systems are as yet uncertain, and it will be quite some time before 3G mobile systems will be ubiquitously available.

On the other hand, the use of encryption on the speech channel is not straightforward. The GSM terminal has a speech compression/decompression process for efficient use of the bandwidth and this is heavily based on the assumption that the input signal will be speech. It uses the usual speech production model parameters such as pitch, vocal tract model parameters etc. to efficiently compress the input speech. If the speech signal is encrypted before it comes to the encoding block, as it will be randomised by the encryption process, it will not satisfy the expected speech characteristics. Hence it will fail to go through the GSM speech transcoding process with sufficient accuracy. A method was presented where after the encryption process the resultant bits are modulated back onto speech-like waveforms [5], which possess the required speech characteristics. This paper presents the progress made since the publication [6], in terms of developing a real-time prototype with reduced complexity and the results from testing on public GSM network voice calls.

2. GSM MODEM

The standard modems used in PSTN are not suitable for the compressed low bit rate speech channels. The main objective of speech compression is to reduce the number of bits re-

[1] This work was partially supported by Engineering and Physical Sciences Research Council (EPSRC) UK Portfolio Partnership in Integrated Electronics grant (reference number: GR/S72320/01).

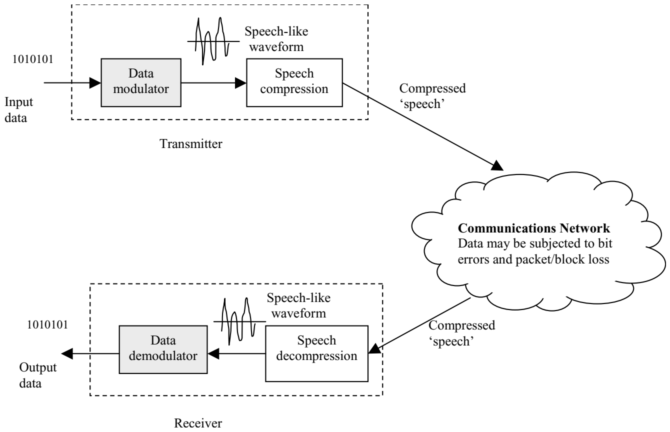

Figure 1: Modulation over the speech channel of a communications network

quired to represent speech, whilst still retaining an acceptable speech quality level [7]. A side-effect of this approach is that the resulting synthesised speech, whilst perceptually being similar to the input speech, i.e. it sounds very similar to the original, may have a fairly different waveform on a sample-by-sample basis. This objective difference prevents most data modems from operating over channels, which employ speech compression systems. This problem is compounded by the fact that in many networks, and in particular, mobile communication systems, the speech signal may undergo more than one set of compression/decompression stages, a phenomenon known as tandeming.

Therefore it was necessary to design a different modem for low bit rate speech channels [5], [6]. This modem can be used to transmit any form of general digital data, e.g. encrypted speech. Figure 1 depicts the relationship between the modulator, the demodulator, and the transmission path in a low bit rate voice communication system.

Figure 2 depicts a more detailed example for the GSM system. The example shown is a typical mobile terminal to mobile terminal communications path. The input speech signal is first compressed using a very low bit rate speech encoder [8], e.g. 1.2 kbps, in order to accommodate in the available bandwidth. The output bit stream of the speech encoder can be encrypted. The encrypted speech data is fed into the modulator, which converts it into a speech-like waveform to feed into the GSM handset. The GSM speech encoder in the handset compresses the modulated waveform. The resulting digital bit stream is transmitted over the communications channel, which includes a radio link, a speech decoder at the base station, a core transmission network, a speech encoder at the second base station and a downlink radio channel. The bit stream is received by the decoder of the receive terminal which converts it back to a speech-like waveform. The transcoding that takes place within the network, cause the waveform generated by the decoder to differ from that produced by the modulator at the transmit end. The demodulator is still able to extract the original transmitted data. For simplicity only simplex communication is illustrated, however full duplex secure voice communication is possible using the same techniques.

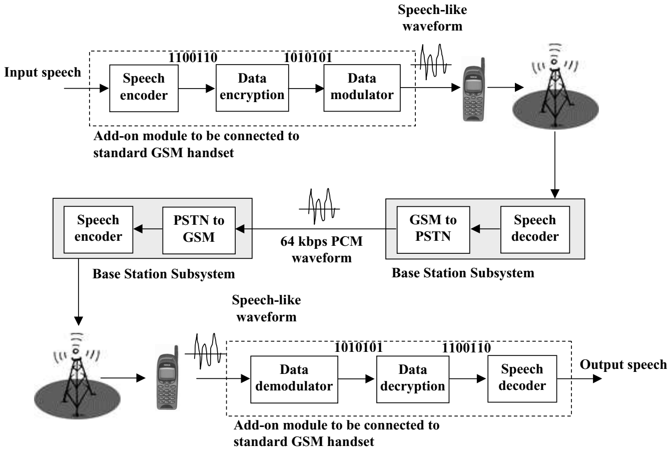

Figure 2: Overview of the complete system

3. REAL-TIME PROTOTYPE

There are several methods to interface the service access point of the communications network, e.g. GSM handset to the modem.

1. The modem, encryption/decryption, and the speech codec may be implemented on a personal digital assistant (PDA) with a GSM connection. Then the modulated waveform could be directly copied onto the GSM voice buffer.
2. The secure voice system is implemented as a separate add on module and the interface provided with a Bluetooth audio link.
3. The secure voice system is implemented as a separate add on module and the interface provided with cables using the hands free sockets of the GSM handsets.

It should be noted that Bluetooth provides a digital connection, while the hands free cables provide an analogue connection. Analogue connections add extra distortion and perform worse than the digital connections. An integrated PDA implementation, which directly accesses the GSM voice buffers, will not add any distortion due to the interface.

The demodulator needs to be frame synchronised before any demodulation of data begins. Synchronisation of the frame boundaries is achieved by using a different modulated signal with a much lower data rate (400 bps). This synchronisation signal is derived from a known set of data stored at both the modulator and the demodulator and transmitted at the beginning. This signal passes through a GSM voice call with virtually no errors.

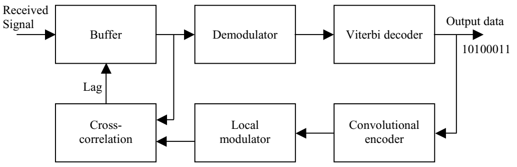

Figure 3: Drift compensation

An additional problem associated with the analogue connections is drifting of the digital samples of the received signal. The clock frequencies of the Analogue to Digital Converters (ADC) and the Digital to Analogue Converters (DAC) used may be slightly different, which results in stretching or shrinking of the received signal. The demodulator will loose the frame synchronisation due to this effect. Therefore the drift is continuously measured and corrected at the demodulator, as depicted in Figure 3. The output data bits are channel encoded and modulated using exactly the same processes as at the transmitter. Then the cross-correlation between the input to the demodulator and the output of the local modulator is estimated. If the maximum normalised cross-correlation value corresponds to a non-zero lag, the frame boundaries are adjusted using that lag value. However the output bit stream may contain bit errors and in order to avoid unnecessary jitter due to those errors, the correction is applied only when the maximum normalised cross-correlation value is greater than a suitable threshold. The cross-correlation values less than the set threshold indicate frames with bit errors, which may also be used by the user application, e.g. speech decoder, as a bad frame indicator.

A real-time prototype of the system is implemented, which demonstrates full duplex secure voice communication on GSM-to-GSM voice calls. Two laptop Personal Computers (PC) are used, each with a 2.8 GHz Intel Celeron processor, Microsoft Windows XP operating system, and 1.18 GB of RAM. The interface to the GSM handsets is provided using hands-free cables. Creative Sound Blaster Audigy 2 soundcards and various standard Nokia handsets are used. Microsoft Visual C++ library functions are used to read and write to the sound cards. Figure 4 depicts the real-time prototype implementation.

Table 1 shows the complexity and the memory requirements of the components of the system. Three configurations$^{1, 2, \text{and } 3}$ of the modem are available with different complexity and memory requirements. The second and the third configurations use sub optimum search techniques in the demodulator and have reduced complexity and memory requirements at the expense of increased bit error rates. Complexity is given in fixed point MIPS. The modem includes channel coding, Viterbi decoding, drift compensation, and synchronisation. Once the complexity reduction techniques currently being investigated are implemented the complete full duplex secure voice system, including encryption and decryption is expected to run on a modern PDA e.g. 200 to 400 MHz and 64 MB of RAM.

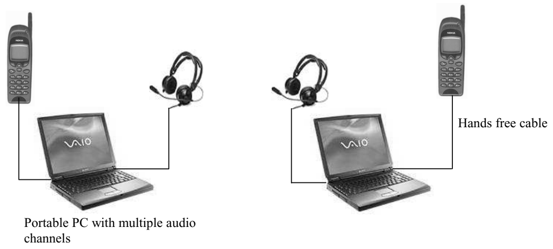

Figure 4: Real-time prototype implementation

A state of the art proprietary 1.2 kbps speech codec [8], [9] is used to provide communication quality speech on GSM-to-GSM voice calls. There is no noticeable degradation in the speech quality when tested with the worst-case modem configuration that has a 1% Frame Error Rate (FER) at 1.2 kbps, see Tables 1 and 2. The speech codec does not use Long Term Prediction (LTP) [10], [11] and there is no propagation of errors. The error resilience techniques employed in the speech decoder ensure that the frame erasures are concealed on typical GSM-to-GSM voice calls.

The extra end-to-end delay introduced by the system stays reasonable, 95ms for the algorithmic delay of the 1.2 kbps speech coder plus 40 ms for the modem give an overall extra delay of 135 ms in addition to the normal GSM speech channel delay. This is significantly less than the delay of the GSM data channel. As a result the proposed system provides a better quality of service than the existing systems, in addition to the improved security.

4. RESULTS

Table 2 shows the results obtained on GSM-to-GSM cross network voice calls on UK public networks, namely Vodafone and O2. This is the most challenging scenario for the proposed system due to the double tandem speech transcoding involved in GSM-to-GSM voice calls. The system works better on GSM-to-PSTN, PSTN-to-GSM, or PSTN-to-PSTN connections, due to one or no speech transcoding stages involved. Tables 1 and 2 clearly demonstrate the trade-off between the complexity and memory requirements, and the Bit Error Rate (BER).

In order to avoid the potential problems associated with analogue interfacing at the transmitter side a digital interface was emulated by copying a modulated waveform file onto a GSM handset, and playing the file while on a voice call to a second GSM handset. The second handset was connected to a PC via a hands-free cable. The PC runs the demodulator and the speech decoder and plays the output speech in real-time. This process transmits the modulated signal on a GSM-to-GSM voice call, however the modulated signal was transferred to the handset as a data file. It is quite clear from Table 2 that digital interfacing or an integrated PDA implementation would significantly improve the demodulation accuracy.

TABLE 1: Complexity and memory requirements

Component	Complexity MIPS		RAM	ROM	Executable
	Average	Worst	Kbytes	Kbytes	Kbytes
Speech codec	42	50	8	12	50
Modem[1]	250	260	15	700	35
Modem[2]	200	210	15	700	35
Modem[3]	40	50	15	35	35

TABLE 2: Results on GSM-to-GSM Voice Calls

Interface	Before channel decoding		After channel decoding		
	Rate kbps	BER %	Rate kbps	BER %	FER %
Digital/Analogue[1]	3.0	2.9	1.7	0.40	1.8
Digital/Analogue[1]	3.0	2.9	1.2	0.03	0.2
Digital/Analogue[2]	3.0	3.4	1.7	0.50	2.0
Digital/Analogue[2]	3.0	3.4	1.2	0.12	0.4
Digital/Analogue[3]	3.0	3.9	1.7	0.78	2.9
Digital/Analogue[3]	3.0	3.9	1.2	0.17	0.6
Analogue/Analogue[3]	3.0	6.0	1.2	0.35	1.0

5. CONCLUSION

Secure voice and data communications over the GSM voice channel has been enabled by modulating the encrypted data onto speech-like waveforms. A throughput of 3 kbps has been achieved with 2.9% Bit Error Rate (BER) on GSM-to-GSM connections. With the addition of error correcting codes a throughput of 1.2 kbps with 0.03% BER and 0.2% Frame Error Rate (FER) has been achieved. A real-time prototype has been implemented which produces communication quality speech using the secure channel on GSM-to-GSM connections. A state of the art proprietary 1.2 kbps speech codec, which does not noticeably degrade the speech quality on GSM-to-GSM connections, is used in the prototype. The latency of the secure channel is significantly lower than the latency of the GSM data channel, making it a more attractive alternative for real-time secure voice communications. It has been shown in this paper that end-to-end secure communication over the GSM voice channel is achievable.

REFERENCES

[1] C. Lo and Y. Chen, "Secure communication mechanisms for GSM networks," *IEEE Transactions on Consumer Electronics*, Vol. 45, No. 4, pp. 1074-1079, November 1999.

[2] M. Street, "Interoperability and international operation: An introduction to end to end mobile security," *IEE Secure GSM and Beyond: End to End Security for Mobile Communications,* London, February 2003.

[3] P. Challans, R. Gover, and J.P. Thorlby, "End to end data bearer performance characterisation for communications over wide area mobile networks," *IEE Secure GSM and Beyond: End to End Security for Mobile Communications,* London, February 2003.

[4] ITU-T Recommendation G.114, "One-way transmission time," Revision 3, May 2000.

[5] N. Katugampala, S. Villette, and A. Kondoz, "Secure voice over GSM and other low bit rate systems," *IEE Secure GSM and Beyond: End to End Security for Mobile Communications,* London, February 2003.

[6] N.N. Katugampala, K.T. Al-Naimi, S. Villette, and A.M. Kondoz, "Real-time data transmission over GSM voice channel for secure voice & data applications," *The 2nd IEE Secure Mobile Communications Forum,* London, September 2004.

[7] A. Kondoz, "Digital speech: coding for low bit rate communication systems," J. Wiley, New York, 1994.

[8] M. Stefanovic, Y. D. Cho, S. Villette, and A. M. Kondoz, "A 2.4/1.2 kb/s speech coder with noise pre-processor," *proceedings EUSIPCO 2000*, Tampere, Finland, pp. 4-8, September 2000.

[9] S. Villette, K. Al-Naimi, C. Sturt, A. M. Kondoz, and H. Palaz, "A 2.4/1.2 SB-LPC based speech coder: the Turkish NATO STANAG candidate," *Proceedings of the IEEE Speech Coding Workshop 2002*, Tsukuba, Japan, October 2002.

[10] S. Singhal and B. Atal, "Amplitude optimisation and pitch prediction in multipulse coders," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Vol. 37, pp. 317-327, March 1989.

[11] R. Ramachandran and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Vol. 37, No. 4, pp. 467-478, April 1989.