MOBILE TRUST

TELECOMMUNICATIONS

Contents

Preface....... 3

1. Terms and definitions....... 3

2. Strong cryptographic transformations....... 4

2.1. Strong voice transformations. The quantitative characteristics of encryption strength . . 4

2.1.1 Speech masking 4

2.1.2. Intelligibility of voice decrypted with the wrong key 5

2.1.3. The size of rearranged voice segments 5

2.1.4. The complexity of cryptographic transformations 5

2.1.5. Time and efforts required for decryption of encrypted speech 6

2.2. Analog analysis of strength 7

3. Assessing the strength of cryptographic transformations 7

3.1. Strong cryptographic voice transformations SCR4 7

3.1.1. General description 7

3.1.2. Description of transformations 7

3.1.3. Numerical characteristics of SCR4 security 7

3.1.3.1. Speech masking. 7

3.1.3.2. The number of encryption options 9

3.1.3.3. Speech signal fragment volume....... 9

3.3.1.4. Interception with a phony key....... 9

3.1.4. Analogues (Similar products) 10

3.2. Justification of the cryptographic strength of voice encryption transformations based on single-sideband time permutations with aliquant switching 12

3.2.1. General characteristics of transformations 12

3.2.2. Description of transformations....... 12

3.2.3. Characteristics of transformations' complexity 12

3.2.3.1. Number of ways to encrypt 1-second long voice segment 12

3.2.3.2. The qualitative characteristics of transformations' complexity 13

3.2.3.3. Speech masking....... 13

3.1.3.4. The average size of rearranged voice segments 14

3.2.4. Analogues 14

4. Summary....... 14

Preface

This manual describes details of cryptographic voice transformations designed to secure voice during conversations over mobile network and in Skype.

1. Terms and definitions

This section contains basic terms and definitions used in the further sections. Other terms are explained as you progress through the text.

Segments of voice signal – randomly selected sounds or their parts limited by time and frequency. Such segments can be stored entirely in the encrypted signal, but they must be short enough to be unintelligible.

Cryptographic strength of transformations – resistance to decryption of encrypted data. High encryption strength makes it difficult (or even impossible) to break the encryption and analyze the original message.

Decryption - the process of decoding data that has been encrypted. If the malefactor doesn’t have a correct key, he will try to figure out the general meaning of the message. The original voice signal might be distorted during decryption.

Residual intelligibility of voice – the number of correctly identified voice units – words, sentences or syllables. When the cryptographic transformations are strong, the residual intelligibility is close to zero.

Voice camouflage – the percentage of securely protected voice segments. The higher the percentage, the more secure your conversation is.

The size of voice segment – the duration of the voice segments multiplied by the width of the frequency band. This is a dimensionless value which describes the complexity of cryptographic transformations.

It will be harder to break the encryption and to obtain the original message if the size of fragments is small and there are multiple ways to rearrange them.

2. The strength of cryptographic transformations

This section contains general information on cryptographic strength.

2.1.Strong voice transformations. Quantitative characteristics of strength.

Mosaic cryptographic transformations rearrange small voice segments by frequency and time.

Cryptographic strength of mosaic cryptographic transformations is defined by several numeric parameters.

2.1.1 Voice camouflage

There are transformations that do not rearrange the voice segments randomly enough to make them unintelligible. Malefactors have to listen to them a couple of times to understand the meaning. Strong cryptographic transformations will make your voice almost completely unintelligible.

2.1.2. Intelligibility of voice decrypted with the wrong key

If the malefactor has similar voice encryption device he can intercept the signal and listen to decrypted voice, but to do so he'll need a correct key. If he uses a wrong key, he will hear rearranged segments instead of the original voice message.

If the transformations are not strong enough, criminals will still be able to make out some words. Strong transformations mix the voice segments twice to make the it almost unintelligible.

2.1.3. The size of rearranged voice segments

The size of the rearranged segments depends on their duration and the width of the frequency band. The small size makes encryption stronger, but also affects the quality of decrypted voice. That's why it's important to maintain balance between cryptographic strength and quality of voice. Usually the duration of the rearranged segments doesn't exceed the duration of the sound (100 ms), and the frequency band width can reach 2000 Hz. Voice segment can be defined by its volume, which is a duration multiplied by the width of the frequency band. If the rearranged segments differ in these parameters, we can calculate the average value.

2.1.4. The complexity of cryptographic transformations

The complexity of cryptographic transformations is defined by several qualitative and quantitative characteristics.

The main characteristic of transformations is the number of ways to encrypt 1-second long voice signal. To figure out the number of possible encryption options one should calculate the gamma consumption rate.

But this number does not directly affect the cryptographic strength.

Time-frequency transformation (TFT) is a transformation of a voice signal, e.g. its compression or spreading of the frequency band. This procedure is reversed when the signal is received. Strong transformation of voice signal reduces its intelligibility (by 2-4 times) and the ability to identify the speaker.

Frame structure and slide scale

There are two ways to rearrange voice signals by time. Frame structure rearrangement divides the signal into frames with fixed duration, and rearranges the segments within the frame. But the hackers can try all the possible rearrangement patterns to figure out the correct one. But the second way – “slide scale” – secures against frame selection, which makes the decryption process far more complicated.

Durations of rearranged voice segments

The rearranged voice segments may differ in duration.

If the segments are the same in length, their duration can be easily determined. If the segments differ in length, one must manually determine the duration, which makes it difficult to decrypt the message.

Camouflaging of switching makes decryption more complex and makes it difficult to figure out the original pattern of rearranged voice segments. Camouflage procedure evens out the transition points between bordering voice segments. Previously such procedures have never been carried out in any known encryption devices so it’s not clear whether they affect cryptographic strength or not.

2.1.5. Time and effort required to decrypt the voice signal

The structure of the rearranged segments remains the same, so it’s still possible to decrypt the signal. But the more complex the transformations, the more skillful the person has to be to break the encryption; he will also need more time.

To decrypt the most complex transformations a person’s skill must be on par with the ones of secret service specialists; he also must have adequate equipment and enough time (one needs a couple of hours to decrypt one second of voice)

The main characteristic of the cryptographic strength is the amount of time and effort required to break the encryption.

But figuring out the time needed for decryption might be a challenging task as it requires live testing. That’s why analog method is used more often.

2.2. Analog method

The easiest way to assess cryptographic strength is to compare them to the other transformations that have been already tested. To do so, one must choose the closest transformation analog, then discover the differences and figure out how they affect the transformations.

3. Assessing the strength of existing cryptographic transformations

3.1.3. Numerical characteristics of SCR4 security

3.1.3.1. Speech masking

Due to quadruple time-frequency transformation an encrypted signal resembles bird twitter: the frequency of the main pitch of the speech signal is increased fourfold. So it is absolutely impossible to understand the speech and identify the speaker, i.e. the encrypted speech does not contain any evidence of individuality and sex.

A spectrogram of the original speech signal (Fig. 1) and a spectrogram of an encrypted signal (Fig. 2) illustrate the transformation.

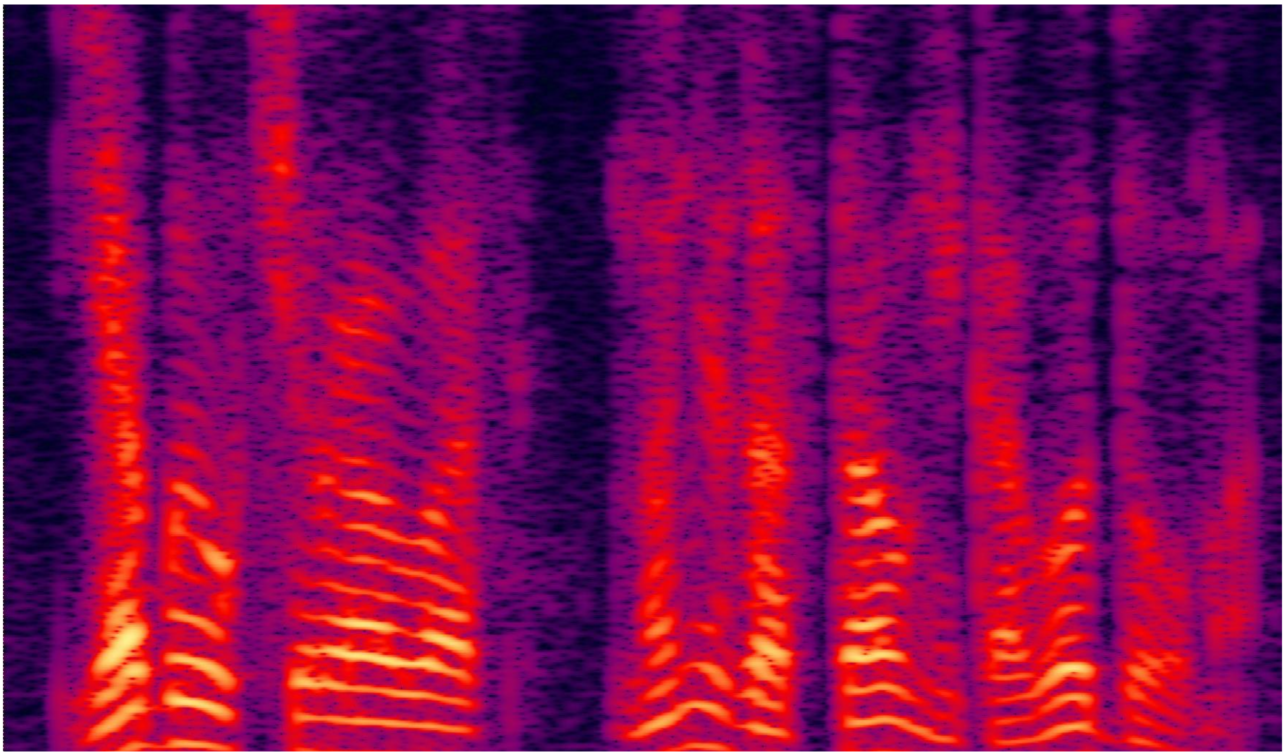

Fig. 1. A spectrogram of the outgoing speech signal

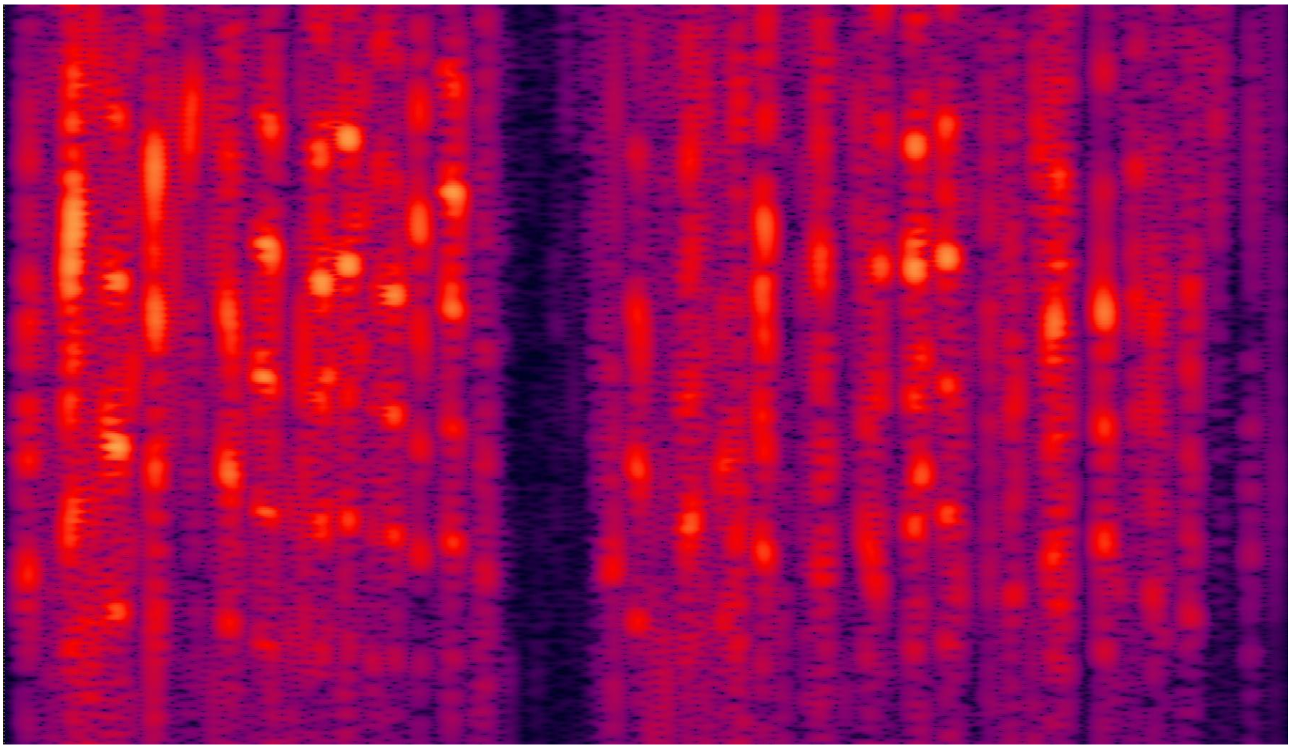

Fig. 2. A spectrogram of the signal encrypted by the SCR4 encryption device

3.1.3.2. The number of encryption options

One minimal fragment of a speech signal with the length of 60 milliseconds and a frequency bandwidth of 750 Hz can obtain 5 time delay options and occupy one of four possible frequency positions. In this case 12 characters of the binary gamma are consumed during the encryption of 4 such fragments. Thus, encoding of one second of a speech signal consumes 12 / 0.06 = 200 binary gamma characters. Therefore, the total number of options to encode 1 second of speech is $2^{200}=10^{65}$.

3.1.3.3. Speech signal fragment volume

The interval length is Δt=60 milliseconds, and the bandwidth is Δf= 750 Hz. Thus the volume of the speech signal fragment for the considered encrypting transformations is

$$V=\Delta t*\Delta f=0,06*750=45$$

3.1.3.4. Interception with a phony key

According to expert estimates, intelligibility is very low during interception. The spectrogram of the signal, encrypted using the SCR4 encoder and decrypted with a phony key, is shown in Figure 3. It shows that the speech signal fragments are mixed up in frequency and time.

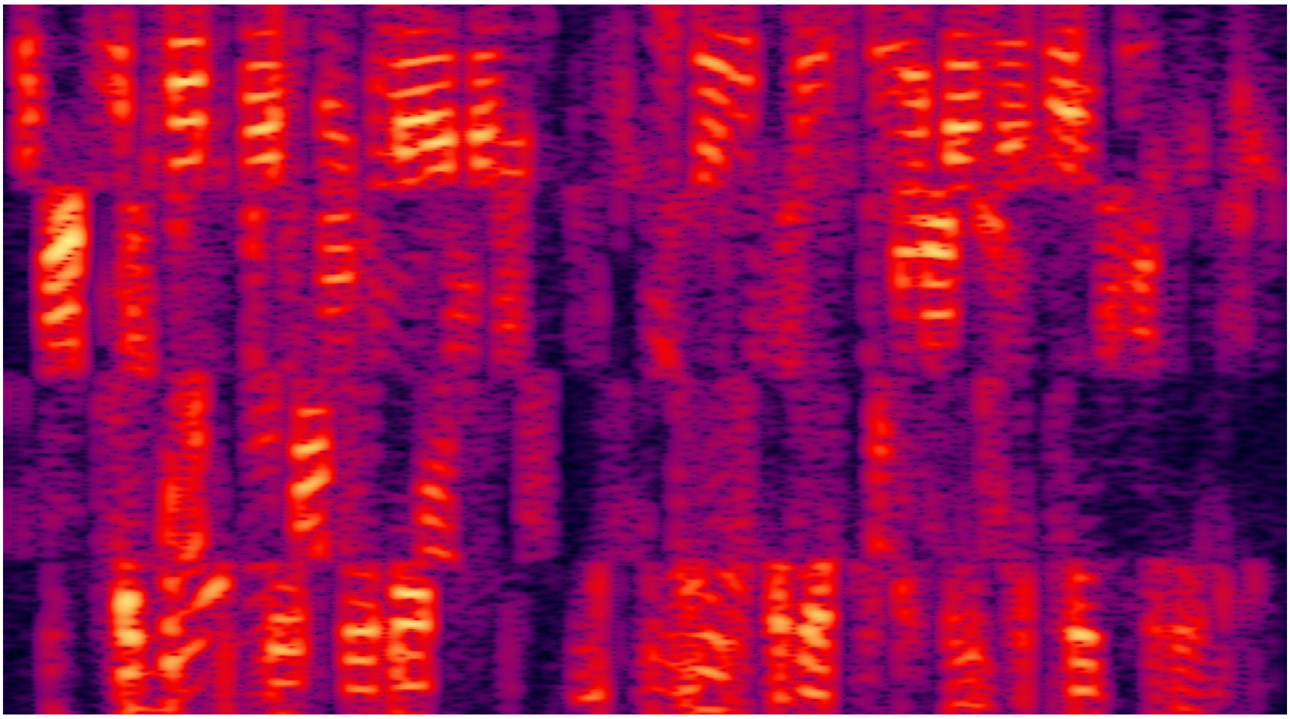

Fig. 3. The spectrogram of the signal, encrypted with the help of the SCR4 encryption device and decrypted with a phony key.

3.1.4. Analogues

The principles and parameters of SCR4 transformations (the duration and the frequency band of non-switched speech segments, the consumption of the encrypting gamma, the number of options to encrypt 1 second of speech, the sliding scale, multiple switching) are close to the known encryption transformations, which have been proven to be high-grade.

The difference is due to SCR4 frequency-time transformation. It increases the level of masking in the encrypted speech signal, as compared to the known transformations.

3.2. Justification of the cryptographic strength of voice encryption transformations based on single-sideband time permutations with aliquant switching

3.2.1. General characteristics of transformations

Transformations implement time permutations with “a sliding scale", with various length of permuted signal fragments and masking the borders of the permuted fragments.

3.2.2. Description of transformations

The speech signal, received at the input, is divided into fragments of various length, defined by the encryption gamma. Fragment lengths range from 20 to 70 ms, the average length is approximately 45 ms. 3 or 4 binary gamma symbols are consumed to develop the length value. Fragment permutation is performed using a delay buffer in the full frequency band, occupied by the signal, which is a transformation parameter. The bandwidth ranges from 2.7 KHz to 3.5. The procedure of selecting fragments from the buffer is given by the gamma. 3 or 4 gamma characters are used to select each fragment. In this case the fragment can be inverted with respect to time, or it can be read at the initial position, depending on the sign of the corresponding gamma. Thus, it takes from 7 to 9 binary gamma characters to encrypt a speech signal fragment with an average length of 45 milliseconds.

3.2.3. Transformation complexity features

3.2.3.1. The number of options to encrypt 1 second of speech.

Basing on the description of transformations, from 7 to 9 symbols are consumed to encrypt a speech signal fragment. Given that the average fragment length is 45 ms, it takes

N=(7 …9)/0,045=156…200 bits of encryption unit gamma to encrypt 1 second of speech.

Thus, the number of options to encrypt 1 second of a speech signal is from $2^{156} = 6{,}4*10^{46}$ to $2^{200}=10^{65}$.

3.2.3.2. Qualitative characteristics of transformation complexity

Different lengths of permuted speech signal fragments along with a sliding scale and masking the borders of permuted fragments characterize this transformation as a very complicated one.

To achieve higher cryptographic strength, aliquot switching is changed to non-aliquot, i.e. variable duration of non-switching segments.

To improve cryptographic security, the average duration of rearranged segments is decreased, which helps to improve residual voice intelligibility and to increase the number of encryption options.

3.2.3.3. Speech masking

According to preliminary estimates, a speaker can be identified, and sometimes it seems that familiar words can be heard. However, such an impression may be due to the fact that the encrypted signal is speech-like. Therefore, in order to perform the reliable assessment of speech masking, articulation measurements should be made, which include expert evaluation of residual speech intelligibility according to the results of interception of the encrypted voice signal, performed by experts. The spectrogram of the encrypted signal is shown in Fig. 4. As can be seen, the spectrogram has similarities with the spectrogram of the speech signal (see. Fig. 1), which also indicates that the signal, encrypted with these transformations, is speech-like.

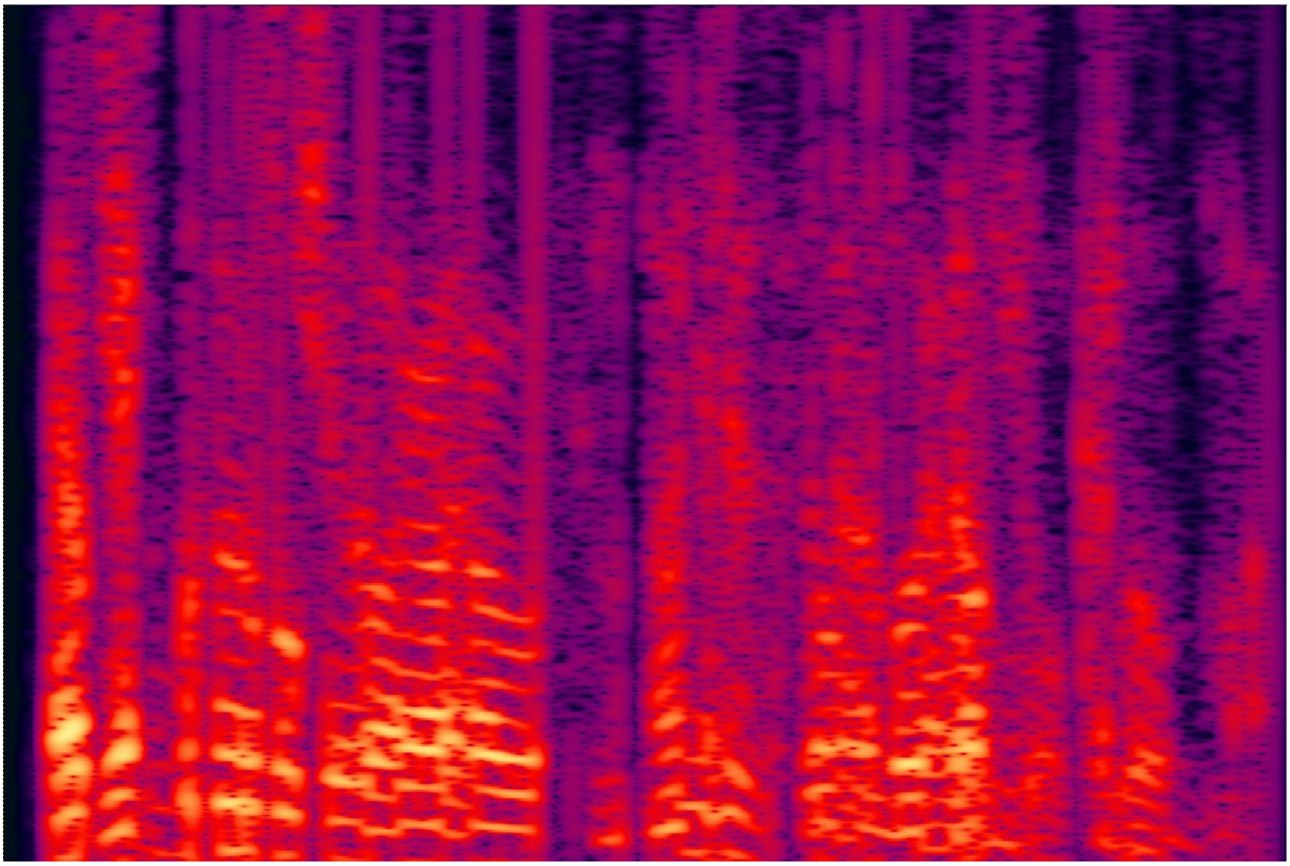

Fig. 4. The spectrogram of a signal encrypted with a single-sideband transformations.

3.2.3.4. The average volume of permuted speech signal fragments

The average fragment length is Δt = 45 ms, and the average bandwidth is approximately Δf = 3000 Hz. Thus, the average volume of permuted speech signal fragments for the considered encryption transformations is

$$V=\Delta t*\Delta f=0,045*3000=135,$$

it is several times higher than the volume of SCR4 transformations. It indirectly characterizes these transformations as less secure.

3.2.4. Analogues

The closest analogue is the known equipment, the cryptoscheme of which implements dual-channel frequency-time permutations with a sliding scale and various lengths of permuted fragments.

It is known that these transformations have rather high security and cannot be automatically decrypted.

4. Summary

The encryption device Stealthphone Hard in Crypto Voice over GSM mode performs a dynamic frequency-time transformation of each speech signal segment, followed by the permutation of signals in time.

Interception of the encrypted signal makes it impossible to restore the meaning of speech or identify the speaker.

Encryption is performed by permutation of speech fragments in time with additional spectral transformations. It is impossible to perform segmentation of the speech signal to recover keys. It is a specific feature of the encryption algorithm.