Analysis of Spectral Parameters of Audio Signals for the Identification of Spam Over IP Telephony

Christoph Pörschmann
Institute of Communication Engineering
Cologne University of Applied Sciences
50679 Köln, Germany

Heiko Knospe
Institute of Communication Engineering
Cologne University of Applied Sciences
50679 Köln, Germany

Abstract

A method is presented which analyses the audio speech data of voice calls and calculates an "acoustic fingerprint". The audio data of the voice calls are compared to each other based on spectral parameters and voice calls are identified which have a high degree of similarity. The method, which is resistant to various modifications of the audio signal, can be used to detect SPIT which is typically characterized by similar or identical voice data in a large number of calls. Privacy protection is assured since only a small fingerprint but not the complete audio data of a call is stored, which does not permit the reconstruction of the content or the identification of the speaker.

1 Introduction

With modern computer and telecommunication systems voice calls can be automatically set up and prerecorded speech messages can then be played. As especially in IP-based networks the costs for voice calls are quite low, Spam over IP Telephony (SPIT) can become a serious problem in the near future. Several activities are currently ongoing in order to identify SPIT and to protect telephone networks from being flooded with SPIT. A number of approaches are considered: The rejection of voice calls can be based on a black-list of caller IDs and/or a white-list of allowed callers. Furthermore, calls can be filtered on the basis of authenticated caller identities and by analyzing trust or security attributes. These approaches are based on the classification of the callers or a verification of their identity. Another approach is to use a challenge-response procedure to identify machine-based calling systems. However, this leads to disturbances and to extended call set-up times. Furthermore, there is the risk of false acceptances or false rejections.

In this article, a method adapted from the area of music identification is proposed which by an appropriate analysis of the audio speech signals can be used to identify SPIT and to automatically establish black-lists.

2 Content-based music identification

In a first step a so-called "acoustic fingerprint" of the audio track is created. Several parameters are extracted from the audio data: the Spectral Flatness Measure (SFM) and the Spectral Crest Factor (SCF) of a music track are computed (for sequenced time windows and for different frequency bands). Together with title and artist, these spectral parameters are stored in a database (Allamanche et al., 2001). Due to their properties with respect to music identification the SCF and the SFM have been standardized as low level signal features within the MPEG-7 framework. A typical application is a mobile phone capturing an extract of a registered audio track. By comparing its "acoustic fingerprint" to those in the database the captured sequence can be identified. Two characteristics of this method (which is already commercially available) are of great importance: The identified parameters are resistant to influences caused by voice coding (e.g. GSM, AMR), background noise and other modifications. Furthermore, a match is only recognized when exactly the same track is played, thus detection by humming or singing fails.

3 Spectral based identification of SPIT

The method which is described in the following allows the identification of SPIT calls. A spectral analysis of the audio signal similar to the content-based music identification method is applied. Replayed calls are identified and the caller identifier is marked for the black-list. It is then possible to block further calls originating from this caller ID.

3.1 Description of the method

To identify SPIT some or all incoming voice calls in a telephone network are analyzed. The spectral parameters SFM and the SCF are computed. Replayed calls have very similar characteristics regarding SCF and SFM parameters. As already shown by Allamanche et al. (2001) these features have the property that they are not significantly influenced by speech coding systems or by other modifications of the audio signal. It would thus be difficult for the caller to modify the audio data automatically in such a way that the identification fails.

The SFM/SCF feature vectors of incoming calls and the corresponding caller IDs are stored in a class database. The comparison between the fingerprints of the actual call and the ones in the class database is performed applying a standard distance metric. If the distance is below a certain limit a similarity between the two sequences is determined and both calls are marked as duplicate (probably SPIT). After a certain number of duplicates, the caller ID is added to a black-list. Further calls from this caller ID are blocked during the signalling phase. The identification still succeeds when there are slight differences between the SFM/SCF feature vectors (e.g. caused by noise, speech coding, different order of speech blocks). Figure 1 shows the general set-up of the system.

In addition, a white-list can be created of those caller IDs that are permitted to send out identical calls (e.g. alarm calls). Furthermore, outgoing calls to prerecorded messages services (e.g. weather forecast) or messages from answering machines are not affected by this SPIT filter since it only analyzes the incoming audio data.

It should be noted that the identification requires at least two fully established calls and some seconds of incoming audio data for successful replay detection. Furthermore, SPIT calls with varying and spoofed caller IDs could in fact be detected and further analyzed but can not be blocked beforehand during the signalling phase. But most VoIP operators require anyway authentication of callers and trust user identifiers only from selected foreign networks.

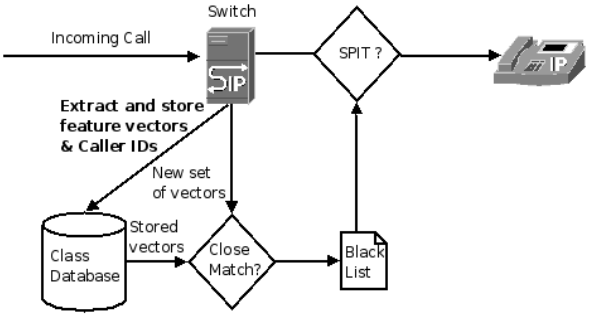

Figure 1: Identification of SPIT based on audio features: The caller ID and feature vectors of incoming calls are stored and compared to those in the database. In case of a high degree of similarity a probable SPIT call is identified.

3.2 Implementation and measurement results

The proposed method has been implemented in Matlab at Cologne University of Applied Sciences. Demo implementations from the MPEG-7 standardization were used for the determination of the SFM and the SCF parameters. For each voice call, 256 feature vectors with 28 components are stored. Thus 7168 bytes are required to store the acoustic fingerprint of one call. The set of 256 feature vectors is determined from the complete set of vectors by applying a vector quantization method (Linde et al., 1980).

28 different voice calls (8 kHz sampling rate) with a duration ranging from 20 to 35 seconds have been analyzed and the resulting sets of feature vectors have been stored. Furthermore, since a spammer might slightly modify the audio data, the following modifications were investigated:

- Change of the pitch of the signal (max. 10%)
- Extraction of small sequences (ca. 5 s)
- Amplitude modification (max. 12 dB)
- Add noise with different spectral characteristics
- Linear distortions (high- or low-pass filtering)
- Non-linear distortions (clipping)

The results show that even for the modified audio signals a robust identification can be achieved. Thus most of the described modifications do not hinder the identification of SPIT. However, a significant degradation in the identification can be observed when adding white noise with energy of more than 6 dB below the energy of the speech signal. In order to increase robustness regarding background noise it is currently considered to adapt the weighting of the identified parameters or to additionally determine the peaks in the spectrogram. A comparable approach in music identification is described in Wang (2006).

4 Conclusion

The described method allows the identification of calls with identical or very similar audio data which typically characterizes SPIT. The method helps to detect SPIT calls and to generate black-lists of spamming caller IDs. An advantage of the method is that an identification of replayed calls is possible after very few of these calls have been captured by the system. Comparable approaches require a higher number of SPIT calls in order to allow a clear identification. A second advantage is the high reliability of the feature comparison. Spectral Features are determined which have shown their resistance to different modifications (codec, background noise, etc.). Finally privacy is protected as no content-related data (e.g. audio content) is stored.

5 References

E. Allamanche, M. Cremer, B. Fröba, O. Hellmuth, J. Herre, T. Kastner, T. (2001). Content based Identification of Audio Material Using MPEG-7 Low Level Description, *2nd Annual International Symposium on Music Information Retrieval.*

Y. Linde, A. Buzo, R.M. Gray (1980). An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications*, 702-710.

A. Wang (2006). The Shazam music recognition service. *Communications of the ACM*, 49(8), 44–48.