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 Koeln, Germany

Heiko Knospe
Institute of Communication Engineering
Cologne University of Applied Sciences
50679 Koeln, 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 with 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 calls being distributed to a large number of receivers. Privacy protection is assured since only a fingerprint but not the complete audio data of a call is stored, which does not allow a reconstruction of the content or identification of the speaker.

1 Introduction

With modern computer and telecommunication systems voice calls can be automatically generated for many calls and prerecorded speech messages can then be played. As especially in IP-based networks the costs for voice call are quite low, Spam over IP Telephony (SPIT) can be a serious problem in the near future. Several activities are currently ongoing in order to identify SPIT and to protect IP telephones from being flooded with SPIT. A number of approaches are considered:

The rejection of voice calls can be performed 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 an additional expenditure of time for the user. Furthermore, there is the risk of false acceptances or false rejections.

2 Content-based music identification

In the following, a method adapted from the area of automatic music identification is proposed. Applying slight modifications this method can be used to identify SPIT and to automatically construct blacklists. 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). A typical application would be a mobile phone which captures an extract of a registered audio track. By comparing its "acoustic fingerprint" to those in the database a 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 detected when exactly the same track is played, thus detection by humming or singing would fail.

3 Spectral based Identification of SPIT

The method which is described in the following allows the identification of SPIT calls. Spectral analysis of the audio signal similar to the content based music identification method is applied. Replayed calls are identified, marked and the caller identity is stored in a blacklist. 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 network are analyzed. The spectral parameters SFM and the SCF are computed. Replayed calls have very similar characteristics regarding SCF and SFM. These features have the property that they are not significantly influenced by speech coding systems or by other intended modifications of the audio signal. Thus it

would be difficult for the sender to modify the audio data 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. If a high similarity is identified between a call and at least one previous call, both calls are marked as replayed (probably SPIT) and the caller IDs are added to a black-list. Future calls from this caller ID are blocked to prevent the users from being disturbed. The identification even succeeds when there are slight differences between the SFM/SCF feature vectors (e.g. caused by noise, the used speech coder, different order of speech blocks). Figure 1 shows the general set-up of the system.

In addition, a white-list can be created which stores those caller IDs that are permitted to send identical calls (e.g. alarm calls). However, calls by a user who wants to access prerecorded messages (e.g. weather forecast) are not affected by the SPAM filter as it only analyzes audio signals from the calling party.

It should be noted that the identification requires at least two fully established calls for a successful replay detection. Furthermore, SPIT calls with varying caller IDs could in fact be detected but not be blocked beforehand.

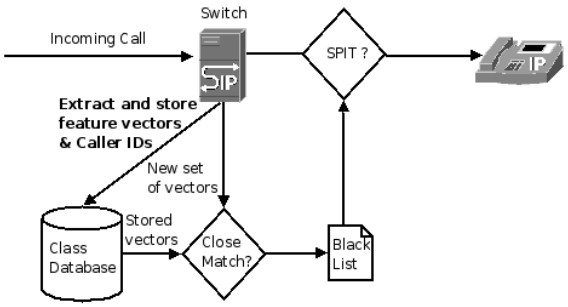

Figure 1: Identification of SPIT based on audio features: The caller ID and feature vectors of each call are stored and compared to the feature vectors in the database. In case of a high degree of similarity a SPIT call is identified.

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 to determine 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 for each 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, slightly modified audio sequences were generated:

- Change of the pitch of the signal (max. 10%)
- Extraction of small sequences (ca. 10 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 modifications of the audio signals in the described form do not hinder the identification of SPIT. However, a significant decrease 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 the 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 characterize SPIT. The method helps to generate blacklists of spammer caller-IDs.

An advantage of the method is that an identification of replayed calls is possible after a few such 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.) for several applications. Finally privacy is preserved as no content-based data (e.g. speaker, 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, Aug 2006.