Analyse spektraler Parameter des Audiosignals zur Identifikation und Abwehr von Telefon-SPAM

Christoph Pörschmann, Heiko Knospe

Institut für Nachrichtentechnik
Fachhochschule Köln
Betzdorfer Str. 2
50679 Köln
Christoph.Poerschmann@fh-koeln.de
Heiko.Knospe@fh-koeln.de

Abstract: Es wird ein Verfahren vorgestellt, welches die Sprachdaten von Telefongesprächen analysiert und einen „akustischen Fingerabdruck“ berechnet. Das Verfahren ermöglicht die Erkennung von wieder eingespielten Anrufen und kann zur Abwehr von SPIT (Spam over IP Telephonie) eingesetzt werden. Es werden spektrale Parameter mit Referenzdaten von vorhergehenden Anrufen vergleichen und Ähnlichkeiten zwischen den Anrufen identifiziert. Das Verfahren ist robust bezüglich unterschiedlichen Kodierungen und Störungen des Audiosignals. Aus datenschutzrechtlicher Hinsicht ist das Verfahren unbedenklich, da sich aus dem „akustischen Fingerabdruck“ weder Inhalt noch Person des Sprechers bestimmen lassen.

1 Einleitung

Mit modernen Telekommunikationssystemen lassen sich aufgezeichnete Sprachnachrichten einer großen Zahl von Telefonteilnehmern automatisiert zustellen. Speziell in IP-basierten Netzen sind Aufwand und Kosten für solche (in der Regel unerwünschten) SPAM-Anrufe gering. IP-basierter Telefonie-SPAM, der in der Regel als SPIT bezeichnet wird, kann in absehbarer Zukunft ein ernst zu nehmendes Problem darstellen, vergleichbar mit SPAM-Mails. Es gibt derzeit zahlreiche Aktivitäten, SPAM-Anrufe zu identifizieren und den Empfänger vor diesen Anrufen zu schützen.

2 Ansätze zur Identifikation von Telefon-SPAM

In den heutigen Telefonnetzen (leitungsorientierte Festnetze und mobile Netze, IP-basierte Sprachnetze) werden in der Regel alle eingehenden Gespräche an den Empfänger signalisiert. Der Empfänger muss den unerwünschten Anruf zunächst erkennen und die Verbindung dann aktiv trennen.

Das Problem von SPITs wird insbesondere im Zusammenhang mit IP-Telefonie (Voice over IP) erörtert [RJ07]. Zur Erkennung und Bekämpfung sind insbesondere folgende Ansätze bekannt:

Die Ablehnung von Anrufen kann im Netz auf Grundlage einer „Black List“ von Anrufer-Kennungen erfolgen bzw. für die Annahme von Anrufen kann eine „White List“ herangezogen werden. Darüber hinaus können SPITs durch Filterung auf der Grundlage von authentifizierten Anrufer-Identitäten [PJ06] oder auf der Grundlage von Trust- oder Security-Attributen bekämpft werden. Alle genannten Verfahren basieren auf einer Einordnung der Anrufer bzw. einer Verifikation ihrer Identität.

Außerdem könnten unbekannte Anrufer einer sprachbasierten „Challenge-Response“ Prozedur zu unterzogen werden, die maschinelle Anrufer erkennen soll [Br07]. Diese führt zu einer Beeinträchtigung des Anrufers und zu einem erhöhten zeitlichen Aufwand durch die Legitimation. Zudem besteht die Möglichkeit einer falschen negativen Erkennung durch einen intelligenten Anrufautomaten oder einer falschen positiven Erkennung durch eine fehlerhafte Erkennung einer menschlichen Antwort.

3 Verfahren zur Identifikation von Audiodaten

Im Folgenden wird ein Verfahren vorgestellt, das aus dem Bereich der automatischen Musikidentifikation stammt. Durch eine leichte Veränderung und Anpassung kann dieses Verfahren auch zur Erkennung identischer Sprachabschnitte und somit zu Identifikation von SPITs eingesetzt werden.

Zunächst wird ein sog. „akustischer Fingerabdruck“ von Musikstücken erstellt; aus den Audiodaten werden dazu verschiedene Parameter extrahiert, und in einem hierfür gängigen Verfahren werden die Spectral Flatness Measure (SFM) sowie der Spectral Crest Factor (SCF) verwendet. Diese Parameter werden zusammen mit dem Titel und Interpreten in einer Datenbank abgelegt [ACF01].

Wird dann mit einem Mobiltelefon oder einem anderen Gerät ein Ausschnitt eines registrierten Musikstücks aufgezeichnet, so kann es durch einen Vergleich des Fingerabdrucks mit den Daten aus der Datenbank identifiziert werden.

Zwei Eigenschaften dieses bereits in kommerziellen Produkten eingesetzten Verfahrens sind dabei wichtig: Zum einen sind die identifizierten Parameter robust gegenüber den Sprachcodern (z.B. GSM, AMR), Hintergrundrauschen und anderen Störungen. Zum anderen erfolgt eine Erkennung nur dann, wenn exakt das gleiche Stück wie in der Datenbank abgespielt wird, eine Identifikation durch Nachsingen, Nachsummen oder Nachspielen ist nicht möglich.

4 Identifikation von SPITs anhand spektraler Parameter

Mit Hilfe des im Folgenden erläuterten Verfahrens können durch eine geeignete Analyse des Audiosignals SPIT-Anrufe detektiert werden. Wiedereingespielte Anrufe werden auf der Grundlage der eingehenden Audio-Daten identifiziert, geeignet kennzeichnet und dann in einer Black-List abgelegt. Danach ist es möglich, weitere Anrufe mit dieser Anruferkennung zu blockieren.

4.1 Beschreibung des Verfahrens

Zur Identifikation werden alle oder ein Teil der in einem Telekommunikationsnetz vermittelten Gesprächsdaten einer Klassifikation unterzogen. Aus den dekodierten Audiodaten werden verschiedene Parameter extrahiert, hierfür wurden die Spectral Flatness Measure (SFM) oder der Spectral Crest Facor (SCF) gewählt. Es werden ähnliche oder die gleichen Verfahren wie bei der Identifikation von Musikstücken eingesetzt [ACF01]. Dabei wird genutzt, dass SPITs zueinander sehr ähnliche Eigenschaften bezüglich einzelner Merkmale des Audiosignal aufweisen (z.B. SCF und SFM). Die Merkmale haben die Eigenschaft, dass Sie durch Einflüsse von Sprachkodierern oder durch den Versender der SPITs gezielt vorgenommene Veränderungen kaum zu beeinflussen sind und somit auch in einem solchen Falle eine eindeutige Klassifizierung der SPIT ermöglichen.

Die akustischen Merkmalsvektoren und die Anrufer-Kennungen werden in einer Class Database abgelegt. Die eigentlichen Gesprächsdaten und –inhalte werden nicht gespeichert. Während der Übermittlung der Gesprächsdaten werden die Merkmalsvektoren der eingehenden Audiodaten ermittelt und mit den gespeicherten Vektoren in der Datenbank verglichen. Wird eine hohe Übereinstimmung mit mindestens einem früheren Anruf identifiziert, so wird dieser als SPIT-Anrufe identifiziert. Hierbei werden wieder eingespielte Anrufe auch dann identifiziert, wenn sie sich geringfügig unterscheiden (z.B. durch Rauschen, Einflüsse der Sprachcoder, Verwendung unterschiedlicher aneinander gereihter Sprachblöcke).

Zur Abwehr weiterer Anrufe werden die Anrufer-Kennungen dieser eingespielten Anrufe in einer Black-List gespeichert und eingehende Anrufe bereits bei der Signalisierung abgewiesen. In einer White-List werden Anrufer-Kennungen von regelmäßig erwünschten automatischen Anrufen (z.B. Weckrufe) hinterlegt. Anrufe des Benutzers bei speziellen Bandansagen, etc. sind von den SPAM-Abwehrverfahren nicht betroffen, da nur Anrufe erfasst werden, die bei dem Teilnehmer eingehen.

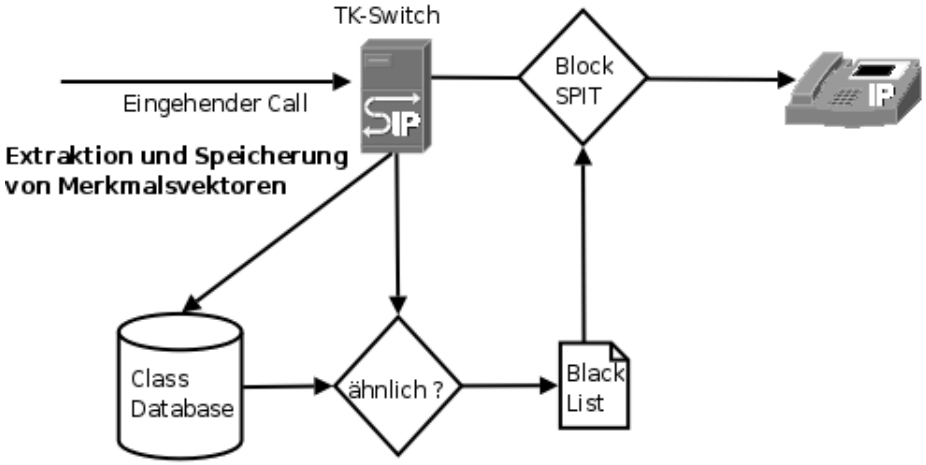

Abbildung 1: Erkennung von SPIT-Anrufen basierend auf Merkmalen der Audiodaten: Die Anruferkennung und die Merkmalsvektoren jedes Anrufs werden gespeichert und ein Vergleich mit Merkmalsvektoren aus der Datenbank durchgeführt. Bei hoher Ähnlichkeit wird ein wieder eingespielter Anruf erkannt.

4.2 Implementierung und Messergebnisse

Das Verfahren wurde in Matlab realisiert. Dabei wurden im Rahmen der Standardisierung von MPEG-7 verwendeten Beispielimplementierungen zur Bestimmung des Spectral Flatness Measure (SFM) und der Spectral Crest Factor (SCF) verwendet.

Für jeden Anruf wurden 256 Merkmalsvektoren mit 28 Komponenten gespeichert, es ergibt sich somit ein Datenvolumen von 7168 Bytes. Die zu speichernden Merkmalsvektoren wurden mit Hilfe eines Vektorquantisierers nach dem Linde-Buzo-Gray-Algorithmus [LBG80] aus sämtlichen auftretenden Merkmalsvektoren ermittelt.

Insgesamt wurden 27 unterschiedliche Gespräche mit einer Dauer zwischen 20 s und 35 s aufgezeichnet. Die Abtastrate betrug 8 kHz. Die sich ergebenden Merkmalsvektoren wurden gespeichert. Weiterhin wurden Sequenzen erzeugt, bei denen gezielt Modifikationen der aufgenommenen Signale vorgenommen wurden:

- Änderung der Abspielgeschwindigkeit (max. 10%)
- Auswahl einer gekürzten Sequenz aus dem aufgenommenen Sprachsignal (ca. 10 s)
- Änderung des Pegels (max. 12 dB)
- Hinzufügen von Hintergrundrauschen (Weißes Rauschen)
- Lineare Verzerrungen (Tiefpass- , Hochpassfilterung)
- Nichtlineare Verzerrungen (Clipping)

Die Ergebnisse zeigen, dass für die vorgenommenen Veränderungen die ähnlichen Sprachsignale eindeutig zugeordnet werden könnten und somit auch bei Veränderungen der aufgenommenen Signale eine sichere Klassifikation möglich war. Einzig das Hinzufügen von Hintergrundrauschen erwies sich als kritisch. Hier wurde bereits bei einem Anheben des Hintergrundgeräusches auf 6 dB unter den Effektivwert des Sprachsignals Fehlerkennungen identifiziert. Um das Verfahren bezüglich des Hintergrundrauschens zu verbessern wird erwogen, die Identifikation zusätzlich auf die Ermittlung von Peaks im Frequenzzeitverlauf auszudehnen [Ma05]. Das Verfahren wurde am Institut für Nachrichtentechnik der FH Köln realisiert und im Rahmen von Diplomarbeiten getestet und weiterentwickelt. Ein Einbringen des Verfahrens in eine Industriekooperation ist geplant.

5 Zusammenfassung

Mit Hilfe des Verfahrens besteht die Möglichkeit, weitgehend identische Anrufe zuverlässig und ohne Beeinträchtigung der regulären Telefonkommunikation zu erkennen und „Black Lists“ aufzubauen.

Ein Vorteil des Verfahrens ist, dass eine eindeutige Identifikation schon nach wenigen SPAM-Anrufen, die von dem Telekommunikationsnetz erfasst werden, möglich ist. Vergleichbare Verfahren nach dem Stand der Technik erfordern eine wesentlich höhere Anzahl von SPAM-Anrufen für eine zuverlässige Erkennung.

Ein weiterer Vorteil des Verfahrens besteht in der hohen Zuverlässigkeit bei der Erkennung der SPAM-Anrufe, da Merkmale betrachtet werden, die in anderen Bereichen der Audiosignalerkennung bereits vielfach eingesetzt werden und die bereits auf Robustheit gegenüber Einflüssen von Codecs, Hintergrundgeräuschen, etc. untersucht wurden.

Schließlich ist das Verfahren aus datenschutzrechtlicher Sicht unbedenklich, da nur ein „akustischer Fingerabdruck“ des Gesprächs erstellt wird, aus dem sich weder Inhalt noch Person des Sprechers bestimmen lässt.

Literaturverzeichnis

[ACF01] Allamanche, E.; Cremer, M.; Fröba, B.; Hellmuth, O.; Herre; J. Kastner, T.: Content based Identification of Audio Material Using MPEG-7 Low Level Description, 2nd Annual International Symposium on Music Information Retrieval, 2001.

[Br07] Brouwer, S.: Spam protection system for voice calls, Patent EP 1742452, 2007.

[LBG80] Linde, Y.; Buzo, A.; Gray, R. M.: An Algorithm for Vector Quantizer Design, IEEE Transactions on Communications, pp. 702-710, 1980.

[Ma05] Mandel, M.: Audio Fingerprinting for Recognition, Internet: mr-pc.org/work/shazam.ps.

[PJ06] Peterson, J.; Jennings, C.: Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP). IETF, RFC 4474, 2006.

[RJ07] Rosenberg, J.; Jennings, C.: The Session Initiation Protocol (SIP) and Spam, IETF, Internet Draft, 2007.