Intrusion Detection using MFCC, VQA and LBG Algorithm

Charu Chhabra¹ Archit Kumar²

¹²Maharshi Dayanand University, CBS Group of Institutions, Jhajjar, Haryana, India

Abstract—An Intrusion Detection system is a system whose main responsibility is to detect suspicious and malicious system activity. The main goal of any Intrusion Detection system is to alert the system administrator with any kind of an intrusion. In order to discriminate between normal and different types of attacks we use classification and clustering based algorithms. Many misuse detection system includes the rule based expert systems. A Speaker Recognition is one of the most useful biometric recognition techniques in this world where insecurity is a major threat. Many organizations like banks, institutions, industries etc are currently using this technology for providing greater security to their vast databases. The Vector Quantization algorithms (VQA) have been used to detect the intrusion and also by reporting with an accurate performance of 96% with minimum possible false alarms. The system extracts a small amount of data from the speaker’s voice signal that can later be used to represent that speaker is an intruder or not. Mel Frequency Cepstrum Coefficients (MFCC) have been used to extract a small amount of data from the speaker’s voice signal that can later be used to represent that speaker. LBG Algorithm is used for vector quantization.

Key words: Intrusion Detection system, ANN, Expert system, Audio/Video Processing, Neural Networks, MFCC, VQA algorithm, False Alarm, LBG algorithm.

I. INTRODUCTION

An intrusion detection system exhibition scheme can be classified into two categories:- Misuse detection and anomaly detection. Intrusion detection is the process of monitoring the events occurring in the network system and restoring and maintaining the data confidentiality, integrity, availability. An anomaly detection system is the one which identifies the activities which vary from the identified pattern for users. It involves the creation of knowledge bases which contains the profile of monitored activities. The other type of intrusion detection system is the misuse detection system. The main job of the misuse detection system is to monitor the activity of the user with the known behaviors of the attackers which are attempting to harm the system. A misuse detection system uses the rule based approach for detecting the intrusion. An IDS is the one which monitors the data congestion thereby identifying the behaviour of malicious attacks and intruders. We have got the motivation of developing an IDS using the expert system that is already established in all the biometric systems which works on pattern recognition. In this Era when people are very much fond of technology and system networks it is very important to protect the systems with an malicious attack thereby keeping the identity of the authenticated user in the database and immediately generating an alarm if the system finds an intruder accessing the system. An intrusion detection system thereby working as an audio detection system as one of the measures can be used in the speaking atm’s and voice transactions in banks or any other secure arena. Other biometric means like fingerprints, retinal scans etc have proven to be very reliable means of detecting an intrusion, but even voice detection based intrusion detection system have proven to be a non-evasive biometric which can be collected with or even without the user’s knowledge.

A. Intrusion Detection Systems

An Intrusion detection system maintains the set of historical profiles for users and then they match the record with the profiles already saved in the database. A rule based architecture is used for detecting the intrusion. The biometric is actually an application of pattern recognition system and of course of classification which have been recently used past few years and also proven to be very much efficient. Pattern classification is a series of steps, starting with input, moving to segmentation, data extraction and translation then finally classification. An automatic intrusion detection system such as discussed can be used in various security applications. The applications of the IDS developed can be used in wide areas. For example, users have to speak a PIN (Personal Identification Number) in order to gain access to the laboratory door, or users have to speak their credit card number over the telephone line to verify their identity. By checking the voice characteristics of the input utterance, using an automatic Intrusion Detection system similar to the one that we will describe, the system is able to add an extra level of security.

B. PREVIOUS WORK

Bolt, Beraneck, and Newman; the Dalle Molle Institute for Perceptual Artificial Intelligence Pattern Recognition is one of the most important areas which have been studied in computer science. In a standard formulation of the problem, we are given a pattern and a data and it is required to find all occurrences of the pattern in the data. Since the publication of the Bayer-Moore and Knuth-Morris-Pratt algorithm, several hundreds of papers have been published dealing with pattern recognition. The main limitation with the work was the generation of false alarm rate. The requirement is an algorithm that generated minimum possible false alarm rate. The main aim of this IDS was to report for even to stop any criminal activity by preventing the fraudulent transactions.

C. PROPOSED WORK

We have chosen different Audios, took 5 samples of same text speech from each Audio and extracted Mel-frequency Cepstral coefficients from their speeches, vector quantized those MFCCs using Linde, Buzo and Gray (LBG) algorithm for VQ and formed code books for each Audio. Kept 1 code book of each Audio as a reference and then calculated the Euclidean distances between these code books and the MFCCs of different speeches of each Audio and made use of these distances between codebooks to identify the corresponding Audio. I recorded the speech of a person who is not in the above 9 Audios and calculated the MFCCs and formed a codebook using LBG VQ, calculated the distance between this codebook and the MFCCs which I kept as reference and proved him as an imposter as he doesn’t match with any one in my database. Thus both Audio
identification and verification is done which is nothing but Intrusion Detection. All this work is carried out in MATLAB, version 7.

II. INTRUSION DETECTION AS BIOMETRICS

A. INTRODUCTION

The Biometrics are truly based on intrinsic characteristics such as voice, fingerprint and retinal scans. The main advantage of the same is that they cannot be lost or forgotten as they are completely based on the behavioral patterns of the user. They do not contain the concept of PIN or passwords which can be stealing or forged. The main advantage of biometric system is that it can act as an identification system and also as the verification system. There can be different methods for biometrics like fingerprints and retinal scans and also audio processing.

B. IDS USING AUDIO AND VIDEO DETECTION SYSTEM

The IDS is the system which traps the video and audio of all users. It maintains the database and then checks for the intruder. This system has two sessions. First being the enrolment session wherein each registered user has to provide the speech signal of his/her audio recorded.

In the second session named as testing phase, the audio signal is matched with the one stored in the reference model.

C. PERFORMANCE METRICS

The most commonly discussed performance measure of a biometric is its Identifying Power. The terminologies which defines the power of ID are known as False Rejection Rate (FRR), or Type I Error, and False Acceptance Rate (FAR) [1], or Type II Error. If this tolerance setting is tightened to make it harder for impostors to gain access, it also will become harder for authorized people to gain access (i.e., as FAR goes down, FRR rises). One can easily learn from the decision matrix and threshold graphs as shown below:

![Decision Matrix](image1)

Fig. 1: Decision Matrix for a system

Another figure as shown below represents the selection of required threshold for minimizing errors:

![Selection of threshold](image2)

Fig. 2: Selection of threshold function to minimize errors

III. CLASSIFICATION OF IDS

The intruder detection system having the capability of the audio and video recognition can be classified into two categories as shown above. Speech recognition, verification or identification systems works by matching the patterns which are generated by the signal processing which is another concept of digital signal processing. Any voice based security system can be categorized as intrusion detection and audio verification system.

IV. MEL-FREQUENCY CEPSUTRUM COEFFICIENTS TECHNIQUES

A. INTRODUCTION

The speech signal is converted into a waveform using Digital signal processing. The speech signal is a slowly timed varying signal (it is called quasi-stationary). An example of speech signal is shown in Figure 2. When examined over a sufficiently short period of time (between 5 and 100 msec), its characteristics are fairly stationary. However, over long periods of time (on the order of 1/5 seconds or more) the signal characteristic change to reflect the different speech sounds being spoken. Therefore, short-time spectral analysis is the most common way to characterize the speech signal.

![Example of speech signal](image3)

Fig. 3: Classification of IDS

A wide range of possibility is there for representing the speech signals for ID tasks such as Linear Prediction Coding (LPC), MFCC is perhaps the best known and most reliable techniques.

B. MFCC PROCESSOR

A block diagram of the MFCC processor is as shown in the figure below. In this the speech is recorded at the sampling rate of 10000 HZ. Then the aliasing is done in order to exhibit the analog-digital conversion.

![MFCC Processor](image4)
C. FRAME BLOCKING

In this step the continuous speech signal is blocked into frames of \( N \) samples, with adjacent frames being separated by \( M \) \((M < N)\). The first frame consists of the first \( N \) samples. The second frame begins \( M \) samples after the first frame, and overlaps it by \( N - M \) samples and so on. This process continues until all the speech is accounted for within one or more frames. Typical values for \( N \) and \( M \) are \( N = 256 \) (which is equivalent to \( \sim 30 \) msec windowing and facilitate the fast radix-2 FFT) and \( M = 100 \).

D. WINDOWING

The next step is to window each individual frame in order to minimize the discontinuities in the signal. If we define the window as \( w(n) \), \( 0 \leq n \leq N-1 \), where \( N \) is the number of samples in each frame, then the result of windowing the signal is:

\[
y_l(n) = x_l(n)w(n), \quad 0 \leq n \leq N-1
\]

The figure shown above shows the hamming window.

Also we obtained the periodogram as shown below

E. MEL FREQUENCY WRAPPING

Every single audio having an actual frequency \( f \), measured in Hz, a pitch is measured on scale called the mel scale. The mel-frequency scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. For conversion purposes, we use the following formula:

\[
mel(f) = \frac{2595 \times \log_{10}(1 + f/700)}
\]

V. CEPSTRUM

By applying the procedure described above, for each speech frame of around 30msec with overlap, a set of mel-frequency cepstrum coefficients is computed. These are result of a cosine transform of the logarithm of the short-term power spectrum expressed on a mel-frequency scale. This set of coefficients is called an acoustic vector. Therefore each input utterance is transformed into a...
sequence of acoustic vectors. In the next section we will see how those acoustic vectors can be used to represent and recognize the voice characteristic of the Audio.

VI. FEATURE MATCHING

A. PATTERN RECOGNITION

The problem of Intrusion Detection belongs to a much broader topic in scientific and engineering so called pattern recognition. The goal of pattern recognition is to classify objects of interest into one of a number of categories or classes. The objects of interest are generically called patterns and in our case are sequences of acoustic vectors that are extracted from an input speech using the techniques described in the previous section. The classes here refer to individual Audios. Since the classification procedure in our case is applied on extracted features, it can be also referred to as feature matching.

B. LBG ALGORITHM STEPS

Step 1: Design a 1-vector codebook; this is the centroid of the entire set of training vectors (hence, no iteration is required here).

Step 2: Double the size of the codebook by splitting each current codebook \( y_n \) according to the rule

\[
Y^{+} = y_n (1 + \varepsilon) \quad Y^{-} = y_n (1 - \varepsilon)
\]

where \( n \) varies from 1 to the current size of the codebook, and \( \varepsilon \) is a splitting parameter (we choose \( \varepsilon = 0.01 \)).

Step 3: Nearest-Neighbor Search: for each training vector, find the codeword in the current codebook that is closest (in terms of similarity measurement), and assign that vector to the corresponding cell (associated with the closest codeword).

Step 4: Centroid Update: update the codeword in each cell using the centroid of the training vectors assigned to that cell.

Step 5: Iteration 1: repeat steps 3 and 4 until the average distance falls below a present threshold

Step 6: Iteration 2: repeat steps 2, 3 and 4 until a codebook size of \( M \) is designed.

VII. EUCLIDEAN DISTANCE CALCULATION

The distance from a vector to the closest codeword of a codebook is called a VQ-distortion. VQ distortion is nothing but the Euclidian distance between the two vectors and is given by the formula:

\[
d_{E}(x, y) = \sum_{t=1}^{N} (x_t - y_t)^2
\]

VIII. MATLAB AUDIO SPECTRUM

Fig. 10: Audio Spectrum on MATLAB

IX. CONCLUSION

The results obtained in this project using MFCC and VQ are applaudable. I have computed MFCCs corresponding to each speaker and these are vector quantized. The VQ distortion between the resultant codebook and MFCCs of an unknown speaker is taken as the basis for determining the speaker’s authenticity. Here I used MFCCs because they follow the human ear’s response to the sound signals. The performance of this model is limited by a single coefficient having a very large VQ distortion with the corresponding codebook. The performance factor can be optimized by using high quality audio devices in a noise free environment. There is a possibility that the speech can be recorded and can be used in place of the original speaker.

X. APPLICATIONS

Intrusion detection Systems with audio and video tracking mechanisms are being used in many applications such as:-

- Access Control Systems
- Telephone-Banking/Booking
- Biometric Login to telephone aided shopping systems Information and Reservation systems
- Security control for confidential information Forensic purposes
- Voice command and control
- Voice dialing in hands-free devices
- Credit card validation or personal identification number (PIN) entry in security systems.
- Banking Locker systems.
- Restoring and managing the secure and confidential information in defense organizations.

REFERENCES


