iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI: http://doi.org/10.22214/ijraset.2017.11002

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI, November 2017- Available at www.ijraset.com

Development of Streaming Audio and Video System for Multimodal Biometric Security System

Binsu C. Kovoor[1]
[1] *Information Technology Department, Cochin University of Science and Technology*

***Abstract*: Use of multiple biometric features increases the reliability of a security system. When audio and video are simultaneously selected and remote access is to be provided, efficiency of streaming plays an important role. In this study, a prototype system is developed for streaming audio and video to use for multimodal biometric security system. The visual and speech signals are transmitted over networks for the authorization of a remote user. The Real-time Transport Protocol (RTP) and Session Description Protocol (SDP) protocols are implemented in this work for communicating in the network. The results of this study indicate that the success rate is reliable and the proposed system can be used for real time applications.**
***Keywords*: Streaming Video, Multimodal Biometric, Real-time Transport Protocol, Session Description Protocol, VideoLAN.**

I. INTRODUCTION

The establishment and verification of the identity of a person in an automatic, reliable and time-efficient manner is of paramount importance as the society is becoming increasingly dependent on the use of information technology even for everyday tasks. Of the many automatic identification technologies, the methods based on biometrics have gained considerable attention due to its robustness as well as reliability. Biometrics is described as the science of recognizing an individual based on human physiological or behavioral traits.

Using human face as a key to security, the biometrics face recognition technology has received significant attention in the past several years due to its potential for a wide variety of applications in both law enforcement and non-law enforcement. As compared with other biometrics systems using fingerprint/palm print and iris, face recognition has distinct advantages because of its non-contact process. Face images can be captured from a distance without touching the person being identified; the identification does not require interacting with the person, nor his consent or knowledge. In addition, face recognition serves the crime deterrent purpose because face images that have been recorded and archived can later help identify a person.

Even though, the biometric identification systems out-perform peer technologies, the uni-modal biometric systems have to contend with a variety of problems - including noisy data, intra-class variations, restricted degrees of freedom, non-universality and spoof attacks. Many of these limitations can be addressed by deploying multimodal biometric systems that integrate the evidences presented by multiple sources of information. Voice of a person holds certain unique characteristics which can be utilized for personal authentication.

Voice is a very intuitive behavioral and ubiquitous biometric which can be captured by modern personal computer. Further, no expensive special hardware is required, only a microphone is necessary. Thus using a video image of a person speaking a pass phrase can be used for authentication, which is going to make the identification system more reliable.

The emerging field of biometric authentication over a network calls for the use of robust person authentication techniques as well as secure computer network protocols. The identification or authentication system over networks permits a user, the access of the services from a remote location. The video consisting of the speech and image of the face of a remote user is a robust biometric feature, which can be utilised for authentication, especially in a networked environment. Appropriate client-server architecture is required for reducing unacceptable long delays in transmission of video biometrics over network.

In the present study, a streaming video system is developed, that can perform many functions, such as admission control, request handling, data retrieval, guaranteed stream transmission and stream encryption. Streaming is the technology where the client plays the content as and when it comes over the network, instead of downloading the entire content as in traditional approach and playing it afterwards. The system consists of video servers, the network, and the client receivers. The video server is the network equipment receiving the recorded information and transmitting the data to the clients.

II. SYSTEM IMPLEMENTATION

The different modules developed in the implementation of Multimodal biometric security system include streaming video server and video LAN client.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI, November 2017- Available at www.ijraset.com

A. Streaming Video Server

A streaming video server is an application server, which can stream videos to the network. In the streaming process, a client node gets linked to a particular network group, which is served by a server and from that instance it can use the resource provided by the server in the group. The video files available in the server are streamed across the network. When the client makes a request and if the client is in the permitted group the server adds it to the current session of streaming. Thus the client will be able to utilize the channel. The different methods of streaming implemented in the proposed work are Unicast, Multicast as well as Broadcast technology [1]. Depending on the range of the network and the clients, an appropriate streaming method can be chosen. Due to the delay between deliveries of packets in real-time data, a jitter may occur [2] which may be prevented by using timestamp and playback buffers. A playback buffer is used to store the data until they are played back which does not start until the time units of data are equal to a threshold value. While the data are stored in the buffer at a possibly variable rate, they are extracted and played back at a fixed rate. The amount of data in the buffer shrinks or expands, but as long as the delay is less than the time taken to play back the threshold amount of data, there will be no jitter. The Protocol used in the streaming video server is Real-time Transport Protocol (RTP) which is the Internet-standard protocol for the transport of real-time data including audio and video [3]. RTP consists of data as well as a control part. The data part of the RTP is a thin protocol providing support for applications with real-time properties such as continuous media, including timing reconstruction, loss detection, security and content identification. The control part referred to as Real Time Transport Control Protocol controls the flow and quality of data and allows the recipient to send feedback to the source. Fig. 1 shows the RTP header format where V is the version bits. Since the version used in the system is 2, V is assigned the value 2. P is the padding bit, which when set, the packet contains one or more additional padding bytes at the end which are not part of the payload. When X the Extension bit is set, the fixed header is followed by exactly one header extension. CC, which is the CSRC Count of bit length 4, indicates the number of CSRC identifiers that follow the fixed header. M, Marker which is a single bit, has its interpretation defined by profile. It is intended to allow significant events such as frame boundaries to be marked in the packet stream. PT, Payload type is of bit length 7 and identifies the format of the RTP payload and determines its interpretation by the application. In the system, the value of PT is 33 which correspond to MPEG2 Video.

RTP Header

0 31

V=2	P	X	CC	M	PT	Sequence Number
Time stamp						
Synchronization source (SSRC) identifier						
Contributing source (CSRC) identifiers						
Extension header						
RTP payload						

Fig 1: RTP Header format

Sequence number of bit length 16 bits, increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence. The initial value of the sequence number is random. The timestamp reflects the sampling instant of the first octet in the RTP data packet. For each succeeding packet, the value is the sum of the preceding timestamp plus the time the first byte is produced. SSRC is the 32-bit numeric Synchronization source identifier which identifies the source of stream of RTP packets and is made independent of network address. All packets from a synchronization source form part of the same timing and sequence number space, so that a receiver groups packets by synchronization source for playback [4]. In the proposed system, synchronization sources include the sender of a stream of packets derived from a signal source, a microphone and a camera. A synchronization source may change its data format and the SSRC identifier is a randomly chosen value meant to be globally unique within a particular RTP session. A participant need not use the same SSRC identifier for all the RTP sessions in a multimedia session; the binding of the SSRC identifier is provided through RTCP. CSRC, contributing source 32 bits is an array of 0 to 15 CSRC elements identifying the contributing sources for the payload contained in this packet. Although RTP is itself a transport layer protocol, the RTP packet is not encapsulated directly in an IP datagram. Instead, RTP is treated like an application

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI, November 2017- Available at www.ijraset.com

program and is encapsulated in a UDP user datagram [5]. RTP uses temporary even-numbered UDP port. The next odd numbered port is used by RTCP. RTP Payload is the data transported by RTP in a packet, for example, audio samples or compressed video data. The system defines Transport Stream (TS), a packet-based protocol for transmission applications. TS packets have a fixed length of 188 bytes, including a minimum 4-byte header. In the proposed system, the continuous MPEG-2 bit stream (also called Elementary Stream) from source encoder is broken into 18,800-byte segments. These video segments are usually called Packetized Elementary Stream (PES) packets. As shown in Fig. 2, an MPEG-2 TS packet is encapsulated in an RTP packet. The RTP packet is then carried by a UDP packet, which is carried by an IP packet and sent to the network.

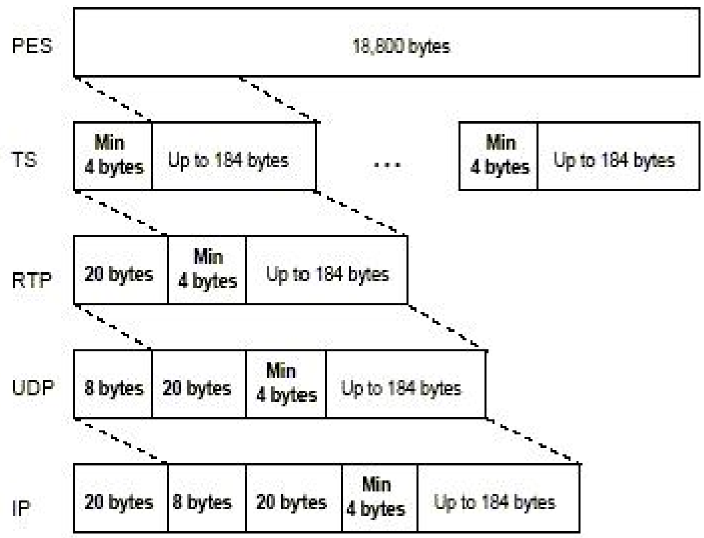


Fig 2: Encapsulation of MPEG-2 TS packets in IP packets

B. The VideoLAN Client

The VideoLAN Client (VLC) works on many platforms such as Linux and Windows. It can read MPEG 1, MPEG 2 and DVDs streamed on a network. VLC uses MPlayer for viewing movies that are streamed from the server. The protocol used in the VLC is Session Description Protocol (SDP). It is used to define a particular session within the client machine [6]. SDP serves two primary purposes, a means to communicate the existence of a session, and to convey sufficient information to enable joining and participating in the session. SDP includes session name and purpose, time the session is active, the media comprising the session and information to receive those media (addresses, ports, formats etc). SDP media information includes the type of media (video, audio etc), transport protocol (RTP/UDP/IP etc), format of the media (MPEG video, H.261 video etc).

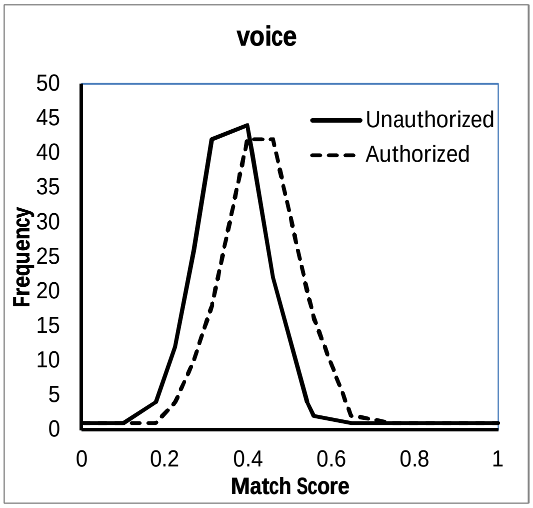
III. PROTOTYPE SYSTEM

In this work, combined face and speaker identification system on a network is developed to perform the various requirements of an on-line multi-modal biometric security system. At the time of ‘logging on’ in a remote server in the network, a snap of the face of the spoken user is captured using a 8.0 megapixel web camera and sound signals of a passphrase by the user is received using a microphone. The graphic user interface of the system was developed in Qt. The server streams video and audio signals to a client. The identification of face and speaker is performed and the access is permitted based on combined scores of two biometric features [7]. The data has been collected from six different users and each person performed ten enrolment sessions. In the enrolment phase, five image data and recitation of passphrase combination are streamed over the network and another five data combinations are used without streaming. The video and audio data with the same background lighting and noise level conditions as expected during testing phase is collected for enrolment. In the testing phase, the ten data, five data streamed over network and another five data without streaming have been used. The success score rate of data used with and without streaming is compared. It was found that in both the cases, the magnitude of success score rate is almost equal.

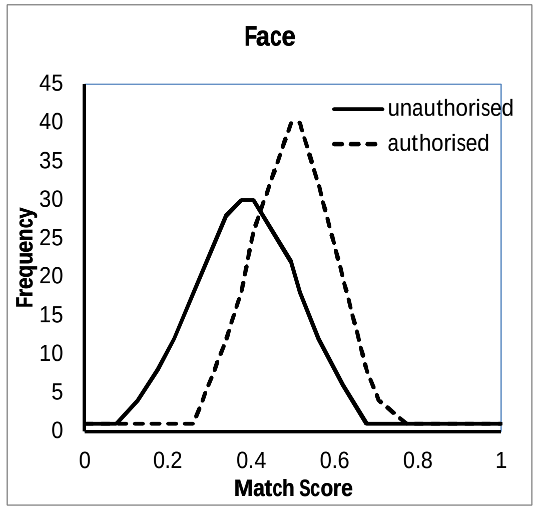
IV. RESULTS AND DISCUSSION

The system is tested to find out the rate of False Acceptance (FAR) and False Rejection (FRR). The False Acceptance means acceptance of impostors. The False Rejection means rejection of a true claimant. The performance is measured in using the parameters, FAR and FRR.If I_a is the number of impostors classified as true claimant and I_t is the total number of impostors in classification test, then FAR can be calculated as the ratio of I_a to I_t . If C_r is the total number of true claimant classified as impostors, C_t is the total number of true claimant classification test then FRR is the ratio of C_a to C_t.

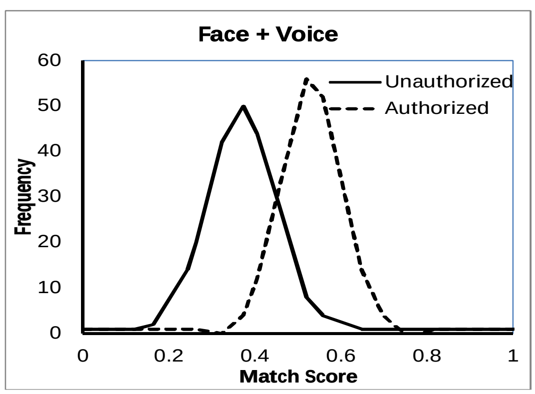
The authorized-unauthorised curve for two traits and their fusions are shown in the Figure 3. The minimum authorised and maximum unauthorized score for each trait and their fusion results is shown in authorized-unauthorised distribution curve as minA and maxU, respectively. The performance of the multimodal biometric identification system is evaluated using width of the confusion zone. The width of the confusion zone is the width between the control limit of the minimum authorized (minA) and maximum unauthorized point (maxU) in the authorized-unauthorized distribution curve. The width of the confusion zone is an indication of effectiveness of the system. If the width is less, system will be able to classify the test sample more correctly. The authorised and unauthorised zones are distinct and the confusion in identifying the test sample is low in such cases. The system is more reliable and robust, if the width of the confusion zone is lower. In general, the fusion of biometric traits in the identification system reduces the width of confusion zone. The width of the confusion zone in score-frequency response of bimodal system of speech and face is found to be low compared to their unimodal system.



a) minA = 0.2273, maxU= 0.5923



b) minA =0.292912, maxU= 0.653596



c) minA = 0.3278, maxU = 0.5906

Fig 3 : Authorised-Unauthorised Distribution Curve.

V. CONCLUSIONS

A multimodal biometric security system with streaming video and audio data has been developed to use in a networked environment.. The proposed system supports multicasting and is implemented using RTP. The RTP provides quality of service (QoS) feedbacks such as number of lost packets, inter arrival jitter and delay. The performance of the system was tested using the data received from six users. It was found that the system, when used remotely, performs as good as a system which is directly using the data without streaming. The excessive amount of time required to download audio and video data as files has made video streaming a popular alternative to reduce video startup latency. This indicates that the proposed streaming video and audio system is acceptable for a multimodal biometric security system

REFERENCES

[1] Rao K.R., Zoran.S.Bojkovic and Dragorad A. Milovanovic, 'Multimedia communicaton systems: techniques, standards, and networks' Prentice-Hall Of India Pvt Ltd, Delhi, P544, 2002
[2] Anna Watson and Martina Angela Sasse, 'Evaluating audio and video quality in low-cost multimedia conferencing systems', Interacting With Computers,8(3),255-275, 1996
[3] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, 'Rtp: a transport protocol for real-time applications.' Rfc 1889 ,1996.
[4] Chung-Ming Huang and Ruey-Yang Lee, 'Achieving multimedia synchronisation between live video and live audio streams using qos controls', Computer Communications, 19, 456-467,1996.
[5] Hoffman, D., and G. Fernando,1998, ,RTP payload for mpeg1/mpeg2 video' Rfc 2250, 1998.
[6] Behrouz A.Forouzan, 'Tcp/Ip Protocol Suite', Tata McGrawhill, Delhi, p861,2000.
[7] Plumpe, M.D. Quatieri, T.F.Reynolds, 'Modelling of the glottal flow derivative waveform with application to speaker identification' IEEE Transactions On Speech And Audio Processing, Volume: 7(5),pp 569-586,1999.

10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY