International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 05, ISSN No. 2455-2143, Pages 60-66
Published Online September 2023 in IJEAST (http://www.ijeast.com)

METADATA AND CONTAINER STRUCTURE ANALYSIS FOR AUDIO AUTHENTICATION

P. S. Marathe
Regional Forensic Science Laboratory
Ganeshkhind Road, Pune 411007

G. C. Wayal
Regional Forensic Science Laboratory, Nashik

V. S. Pawade, Dr. S. V. Ghumatkar
Directorate of forensic Science Laboratories, Mumbai

Abstract: **Audio recordings of crime scenes are frequently utilized as digital evidence due to the universal usage of mobile phones, but they must first be authenticated before being used as evidence in court. The technique of identifying whether an audio recording is real or has been altered or manipulated is known as audio authenticity. In order to investigate and validate audio, this study provides a cutting-edge authentication technique that can distinguish the difference between real and fraudulent audio. Herein, we analyzed audio recording samples from different mobile handsets like iPhone, Samsung, One Plus etc. It is important to note that the proposed method authenticates audio files without regard to the audio content, i.e., without concern to the speaker or the speech. The audio recordings were subjected to manipulation that a human cannot recognize them auditory. The present results show that it is possible to verify the authentication of audio recordings generated through mobile phones or any recorder using metadata and the container structure of the recorded and edited audio file.**

I. INTRODUCTION

The audio and video recordings are created at any time using readily available digital multimedia tools including audio-video recorders, cell phones, and digital cameras. It is simple to copy the captured audio or video information in a variety of storage devices and transfer it from one location to another. With the use of the internet, the multimedia information may quickly go viral. Normally people also make their opinion based on the CCTV footages, videos and audio recordings without knowing the information about the authenticity of the data. Because of the numerous sources and editing software options available, with which the material may be quickly altered, the originality and authenticity of the audio video data have therefore become a crucial concern. Audio recordings are presented in front of the court of law as an evidence in the crime cases like bribery, sexual assaults, child exploitation, kidnapping, armed robberies, murders etc., their admissibility in the court depends on proven authenticity. Recent developments in technology offer easy access of manipulation in the digital media file and therefore acceptance of the digital media files become an important and unavoidable issue in the court. Forgery detection research has been prompted by questions about how to verify multimedia data, however, investigations in audio are often still limited in comparison to those in image and video.

A calculative devised approach exploits the audio recording device's residual ability to pick up magnetic fields from electrical power lines to help with authentication. ENF, nominally 60 Hz in the United States and 50 Hz in many other parts of the world [1]. Therefore, it may be possible to determine if an audio recording was produced at the stated time and location by comparing the measured ENF extracted from the audio recording with a database of known ENF readings from the electrical grid [2,3, 4, and 5].If well-designed audio equipment (such as condensers or piezoelectric microphones) or battery-operated devices are utilized to capture the recording, the ENF-based techniques might not always be suitable [6]. In the study by Zhau et al [7] they have proposed a technique to detect the forgery in audio considering acoustical environmental signature i.e. background noise in audio signals. They have calculated the difference in the noise signals and found out the audio recording location. After the case trial of United States versus McKeever case, the judge's ruling consisted of some requirements that also are used now with some variations [8].In paper [6] authors beautifully explained the classification of audio authentication methods into container-based authentication and content-based authentication. The source and integrity of the files needed for forensic investigation have been efficiently determined via file signature analysis in the study by D. Hamdi [9]. Hex data analysis of the audio recordings from different Android

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 05, ISSN No. 2455-2143, Pages 60-66
Published Online September 2023 in IJEAST (http://www.ijeast.com)

mobile phones was carried out in the MSc thesis by GINA [10].

Audio recordings often go through a few basic checks to look for obvious signal alterations [8, 11]. Basic Audio-Authentication techniques like critical listening, waveform analysis and spectrum analysis are the preliminary tests to be performed on the audio recording files [8, 12]. [16] Authentication framework proposed consisted of File structure analysis, Global analysis, Local analysis and Device verification. Anti-forensic techniques have evolved to undermine the reliability of digital evidence, including the modification of audio recordings using a sophisticated audio editor. To ascertain the origin of an audio recording in question and identify whether it has been tampered or not, forensic authentication analysis should be carried out. Audio latency, composition of file structure and the device based log history examination were reported in the paper by N I Park et al [13]. They also compared the file structures of the original audio files and the audio files edited using in-built application "voice memo" in iPhones. Again Forensic authentication for audio recordings generated by the voice recorder application in Samsung Galaxy Watch4 series was carried in the paper presented by NI Park et al [14]. The hex data of the audio files provide clear indications of the edition, which may be used to demonstrate the integrity [15, 16]. Overall understanding from the literature is that the multimedia file's container structure analysis gives all necessary information that is unique to the recording equipment and that the tampering with audio files may be skillfully managed using this analysis.

Over 1 billion people of Bharat use mobile phones, among them more than 800 million people avail the internet facility. Daily digital consumption report says that people spend averagely 1.4 hours on their mobile daily. Bharat has seen a steady increase in audio usage over the past two years, including podcasts, audio books, original audio programmes, talk shows, and more. There are already more than 150 million daily listeners in the nation, according to reports. Increased smartphone use in both rural and urban areas of India, low-cost mobile connectivity, and the availability of a large variety of on-demand audio material spanning genres, languages, moods, special events, and other factors have all contributed to the growth of audio platforms. When listening to audio content, people typically lose focus and get swept away without properly analyzing the authenticity and integrity of the content. It becomes forensically important to verify the integrity and authenticity of digital media, such as CCTV footage, mobile recorded videos or audio recordings of riots, murders, fights, dacoits, or political leader speeches, when they are shared on digital platforms and may be used as evidence in court.

Till now no authentication tool is in the market which will surely tell whether the audio file is original or not. Hence on the basis of null hypothesis, if we could prove that the file is not edited then automatically it infers that the file is authentic.

II. MATERIALS AND METHODS

Basically following the null hypothesis we in this study tried to find out the ways in which we could demonstrate that the audio file is not authentic by comparing its container structure and metadata. For this purpose ten mobile phones with different makes were selected and their built in audio recorder was used to record the random audio recording. These audio recordings were subjected to manipulations either by the software or by online audio editors available on internet. The details of the mobile brands selected and the manipulation opted are listed in Table 1. M4A is a file extension for an audio file encoded with Advanced Audio Coding (AAC) was observed in the seven selected mobile phones and rest were having MP3 format. The versions of OS which the mobiles phones were bearing are also listed in the table. The edited audio files were compared with their original counterparts on the basis of the parsed file structure and the results are discussed in the later section

S.R. No.	**Make**	**Model name**	**OS/iOS version**	**Format of recorded file**	**Manipulation by**
1	Apple iPhone 14 pro	A2890	iOS 16.6	M4A	Received on WhatsApp
2	Apple iPhone 13	A2633	iOS 16.6	M4A	Converted in Goldwave
3	Apple iPhone 11	A2221	iOS 16.6	M4A	Trimmed in ASPOSE online cutter
4	Apple iPhone 7	A1784	iOS 15.7.8	M4A	Converted to mp3 using format factory application
5	One Plus 8	IN2011	Oxygen OS 13.1	MP3	Trimmed in 123APPS mp3 online cutter
6	One Plus	Nord 2E IV 2201	Oxygen OS 13	MP3	Edited in Audacity
7	One Plus	Nord 2T	Oxygen OS 13	MP3	Sent as a document on

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 05, ISSN No. 2455-2143, Pages 60-66
Published Online September 2023 in IJEAST (http://www.ijeast.com)

					WhatsApp
8	Samsung	S20	Android OS 13	M4A	Playback rate increased in AVS Audio editor
9	Samsung	Galaxy M30s	Android OS 11	M4A	Sent as a drive attachment on mail
10	Samsung	Galaxy M21	Android OS 12	M4A	Edited by Ringtone maker app on mobile

III. OBSERVATIONS

Writing Application

In most of the cases trace of writing application was observed, which was an important evidence to identify file edition. Writing application in the form of Lavf which is basically an audio/video library containing demuxers and muxers widely used by most of the video editors [17]. Versions of Lavf differed for every editor. Version Lavf58.29.100 was observed for iPhone 11 audio file when it was trimmed in ASPOSE online cutter, Lavf58.45.100 for One plus nord 2E file when it was edited in Audacity software, Lavf59.27.100 for One plus 8 IN2011 audio file when it was trimmed in 123APPS mp3 online cutter and Lavf58.12.100 for iPhone 7 audio file when the file was converted to mp3 using format factory application. The same is depicted in Figure 1.

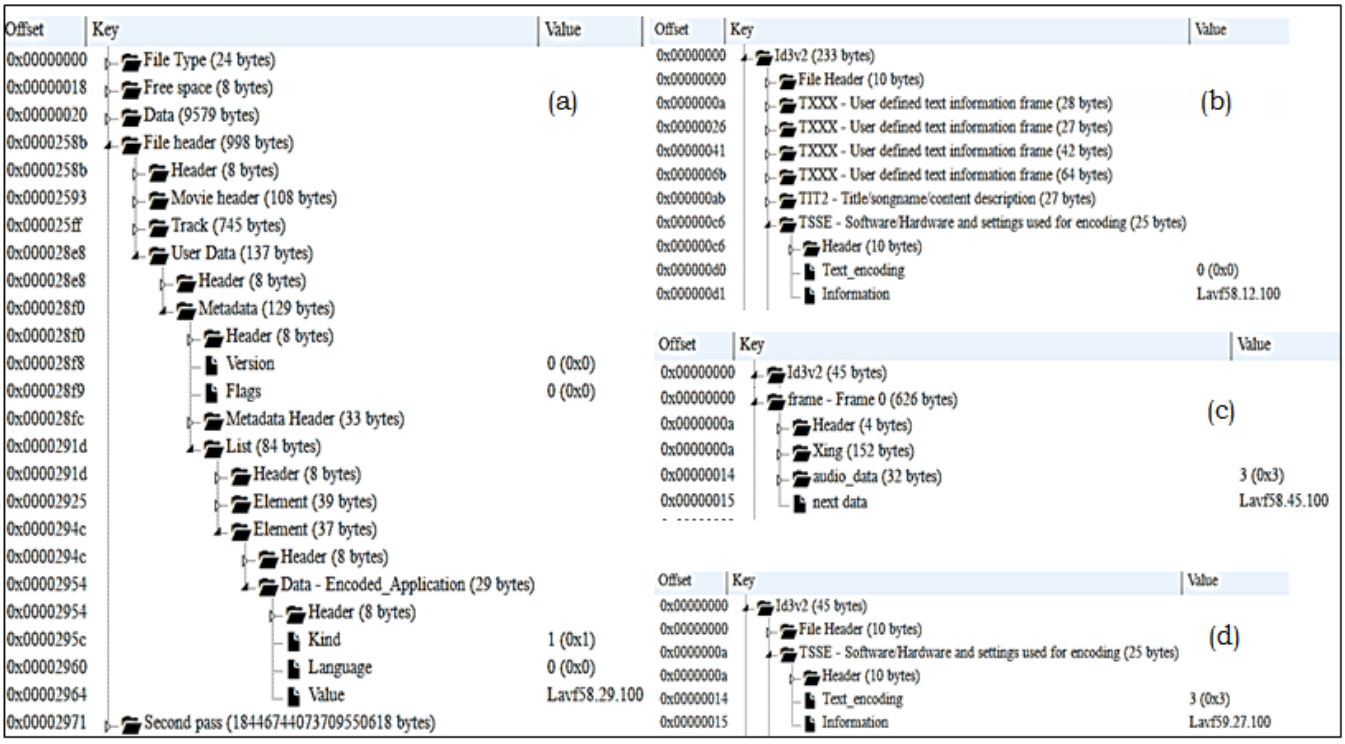

Figure 1 showing the trace of writing application in edited files which were recorded in (a)iPhone 11, (b)iPhone 7, (c)One plus nord 2E and (d) One plus 8 IN2011.

Writing application indicating to AVS library was observed (Figure 2) in the media file info tab of the file recorded in Samsung S20 and edited in online AVS editor (playback rate was increased).

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 05, ISSN No. 2455-2143, Pages 60-66
Published Online September 2023 in IJEAST (http://www.ijeast.com)

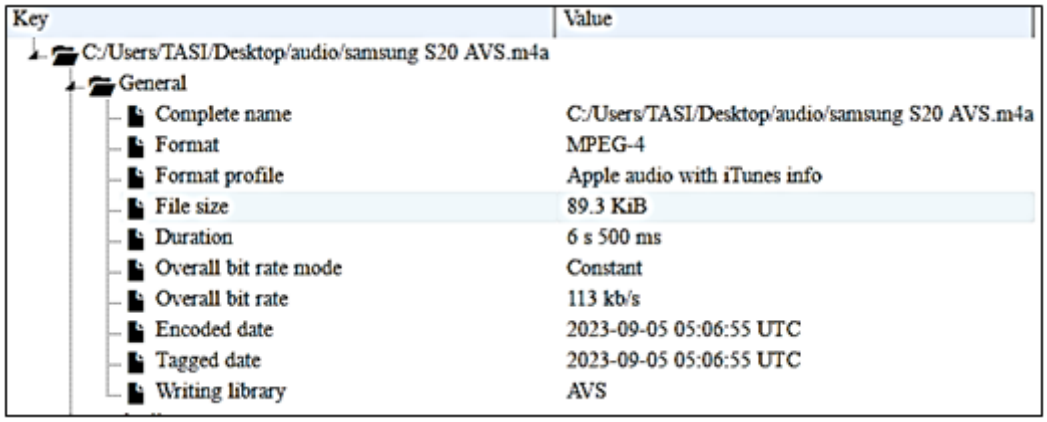

Figure 2 showing the trace of writing application AVS for Samsung S20 edited file

Container structure analysis

No changes in the container structure of the original file, which was captured on a Samsun M30s, were noticed when it was delivered as an email attachment. Similar to this, no change in the container structure was seen when the recorded file from the One Plus Nord 2T device was delivered as a document attachment via WhatsApp. The container structure of the M30s original file and attachment file are presented in Figure 3.

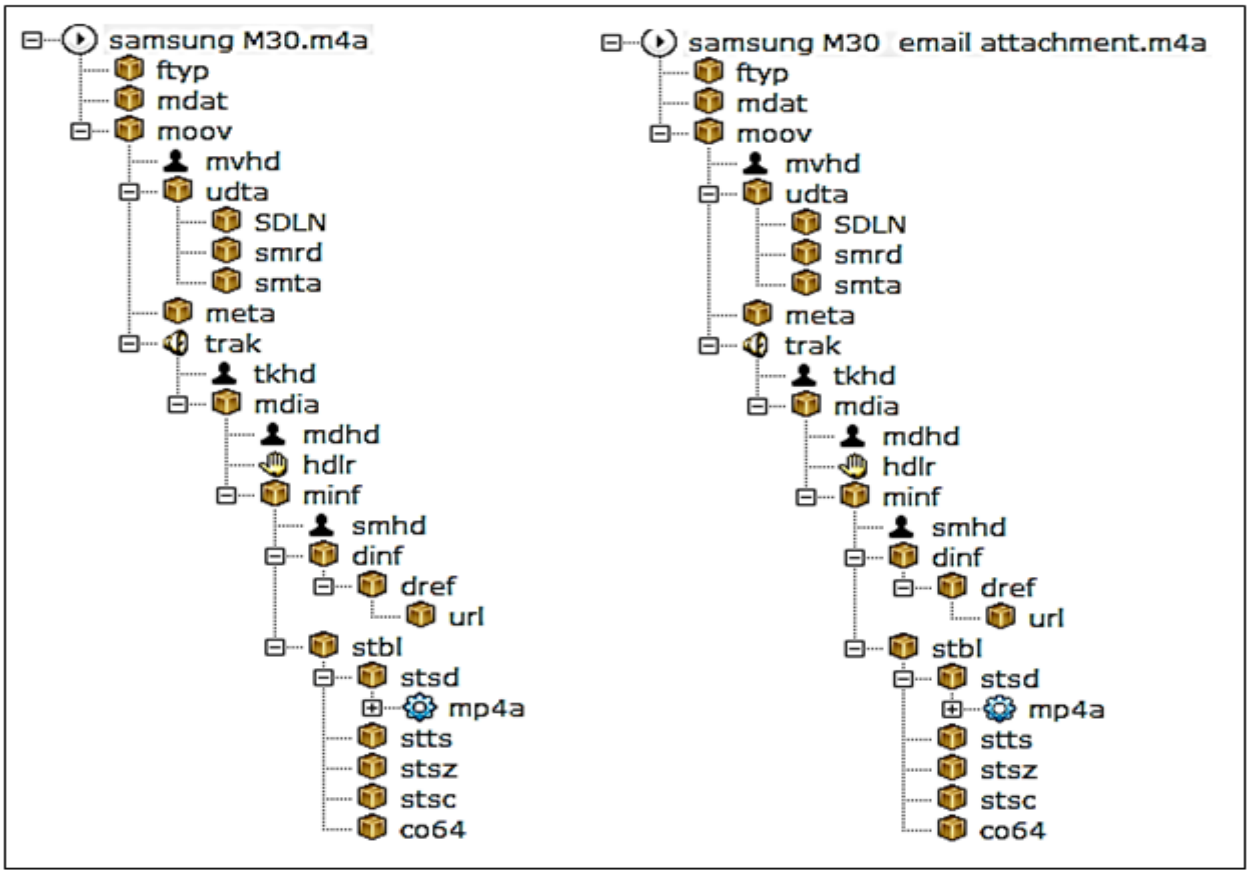

Figure 3 showing the container structure of the original and attachment file

Extra parent box ‘beam’ (Figure 4) observed in the file which was sent through WhatsApp messenger gained extra attention. Although the 'beam' parent box's source could not be entirely confirmed, some study suggested that it could

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 05, ISSN No. 2455-2143, Pages 60-66
Published Online September 2023 in IJEAST (http://www.ijeast.com)

have come from the Beamr [18,19] encoding tools, which some businesses use to optimize multimedia content and reduce file sizes.

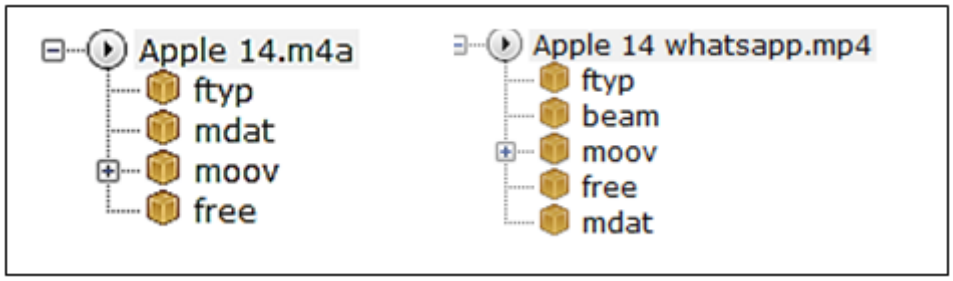

Figure 4Showing Extra 'beam' parent box was added when the file was shared on WhatsApp

The recorded file in the Samsung M21 was edited by Ringtone maker app on mobile. The original file showed four child boxes of "moov" parent box namely "mvhd", "udta", "meta" and "trak" whereas in edited file only two child boxes named "mvhd" and "trak" were observed (Figure 5).

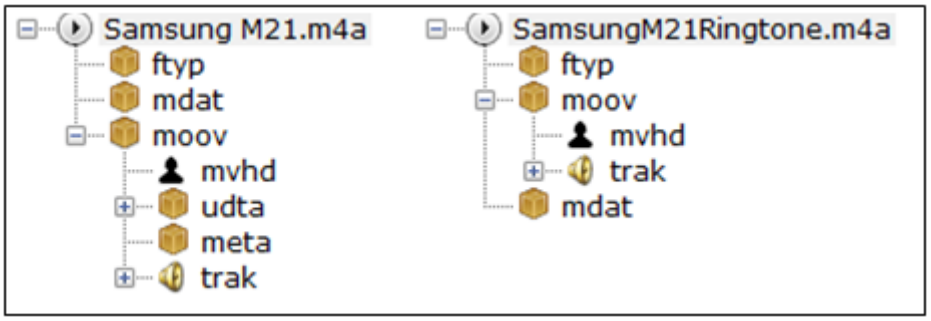

Figure 5Showing the comparison of container structure of original and edited files

The metadata for a presentation is stored in the single Movie Box which is abbreviated as 'moov'. This is mandatory box and its quantity is exactly one. This box contains at least two 'trak' child boxes, one to store the video track information, and one for the audio track. Both boxes have a similar structure. 'Trak' box which is required to contain the media data and there shall be at least one media track within an ISO file [20]. The movie header child box 'mvhd' is contained in 'moov' box. This box is mandatory and defines overall information which is media-independent, and relevant to the entire presentation considered as a whole. 'mvhd' presented the date of creation and date of modification of the audio. It also displayed the time scale which is an integer that specifies the time-scale for the entire presentation; this is the number of time units that pass in one second [20].

Next child box in 'moov' container which is not mandatory is 'udta' i.e user data box. It will be placed in either of the following boxes, Movie Box 'moov', Track Box 'trak', Movie Fragment Box 'moof' or Track Fragment Box 'traf' [20]. This box contains objects that declare user information about the containing box and its data. This 'udta' child box is missing in the edited file shown in Figure 5. A 'meta' box contains descriptive or annotative metadata. The 'meta' box is required to contain a 'hdlr' box indicating the structure or format of the 'meta' box contents. That metadata is located either within a box within this box (e.g. an XML box), or is located by the item identified by a primaryitem box [20]. It is not a mandatory box hence it is present in original and not present in the edited file shown in Figure 5.

M4A media files have similar container structure as that of MP4 files and which is totally different from the MP3 format. M4A uses container structure and MP3 format contains MP3 frames where each frame consists of header and data blocks. When the audio file recorded in the M4A format in iPhone 13 and was saved using Gold wave software it got converted into MP3. The containers structure altogether changed to frames and data blocks. ID3V2 tag was added in the structures of the MP3 edited files. ID3 is a metadata container most often used in conjunction with the MP3 audio file format. It allows information such as the title, artist, album, track number, and other information about the file to be stored in the file itself. In ID3v2, an extensible set of "frames" located at the start of the file are used.

IV. CONCLUSION

This study showed how to easily determine the edition of an audio file without hearing the audio. We concentrated our attention in this investigation on a M4A and MP3 audio clip that had been purposefully edited using both installed editing software and some online editor resources. The original recording for the file was made with a mobile device. The metadata information and the file container structure are interrelated. The files that were shared by email and as a document over WhatsApp retained their original metadata and structure. This implies that the two techniques mentioned above are the safest ways to transport digital data since there is no chance of a disruption. The findings of our analysis of both original and changed digital audio files shown that irrespective of the type of editing applied, the metadata properties, file container boxes, and information raised alongside container boxes all reflect the editing process. The method's greatest advantage is unquestionably its capacity to save time.

In this case, a quick workaround exists for the container structure evaluation and metadata analysis. We can quickly tell if a file has been altered or not based on the hash value if both the original and edited files are provided for analysis. However, file structure and metadata can surely be helpful if we are just given one audio file and asked to remark on the legitimacy of the file. We anticipate that file structure- and metadata-based audio forensics techniques will be utilized in forensic situations.

V. ACKNOWLEDGEMENT

We would like to thank Mr. Sanjay Kumar Verma, IPS, Director General, Legal and Technical, Directorate of Forensic Science Laboratories, Mumbai, India for providing research environment and guidance throughout the work

VI. REFERENCE

[1]. Maher, R.C. (2010). Overview of Audio Forensics. In: Sencar, H.T., Velastin, S., Nikolaidis, N., Lian, S. (eds) Intelligent Multimedia Analysis for Security Applications. Studies in Computational Intelligence, vol 282. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-11756-5_6.

[2]. Brixen, E.B., (2008), ENF—quantification of the magnetic field. In: Proc. Audio Eng. Soc. 33rd Conf. Audio Forensics—Theory and Practice, Denver, CO.

[3]. Cooper, A.J., (2008),The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings – an automated approach. In: Proc. Audio Eng. Soc. 33rd Conf. Audio Forensics—Theory and Practice, Denver, CO.

[4]. Grigoras, C., (2005), Digital audio recording analysis: the electric network frequency (ENF) criterion. Int. J. Speech Language and the Law 12, 63–76.

[5]. Grigoras, C., (2007), Application of ENF analysis method in authentication of digital audio and video recordings. In: Proc. Audio Eng. Soc. 123rd Conv. Paper 1273.

[6]. Zakariah, M., Khan, M.K. & Malik, H., (2018), Digital multimedia audio forensics: past, present and future. Multimed Tools Appl 77, 1009–1040. https://doi.org/10.1007/s11042-016-4277-2.

[7]. Zhao H, Malik H (2013) Audio recording location identification using acoustic environment signature. Information Forensics and Security, IEEE Transactions on 8:1746–1759

[8]. Maher R (2009) Audio forensic examination. Signal processing magazine, IEEE 26:84–94

[9]. D. Hamdi, F. Iqbal, T. Baker and B. Shah,2016, "Multimedia File Signature Analysis for Smartphone Forensics," 2016 9th International Conference on Developments in eSystemsEngineering (DeSE), Liverpool, UK, pp. 130-137, doi: 10.1109/DeSE.2016.22.

[10]. DeAngelis Gina Antoinette,2020,University of Colorado, MSc Thesis, “ANALYSIS OF AUDIO RECORDINGS MADE USING THE VOICE RECORDER APPLICATION ON ANDROID PHONES”,

[11]. B.E. Koenig and D.S. Lacey, 2009,‘‘Forensic Authentication of Digital Audio Recordings,’’ J. Audio Eng. Soc., vol. 57, no. 9, pp. 662-695.

[12]. S. Gupta, S. Cho and C. J. Kuo,2012, "Current Developments and Future Trends in Audio Authentication," in IEEE MultiMedia, vol. 19, no. 1, pp. 50-59, Jan. doi: 10.1109/MMUL.2011.74.

[13]. Park NI, Lee JW, Shim KS, Byun JS, JeonOY., 2021, A method of forensic authentication of audio recordings generated using the Voice Memos application in the iPhone. Forensic Sci Int. 2021 Mar;320:110702. doi: 10.1016/j.forsciint.2021.110702. Epub 2021 Jan 23. PMID: 33561789.

[14]. Park, NI, Lim, SH, Byun, JS, Kim, J-H, Lee, JW, Chun, C, et al. 2023;Forensic authentication method for audio recordings generated by Voice Recorder application on Samsung Galaxy Watch4 series. J Forensic Sci. 68: 139–153. https://doi.org/10.1111/1556-4029.15158.

[15]. Gangwar, D P &Pathania, Anju. (2020). AUTHENTICATION OF DIGITAL AUDIO RECORDING USING FILE'S SIGNATURE AND METADATA PROPERTIES.

[16]. MasterThesis by Daniel Lawn Rappaport, University of Colorado, ESTABLISHING A STANDARD FOR DIGITAL AUDIO AUTHENTICITY: A CRITICAL ANALYSIS OF

TOOLS, METHODOLOGIES, AND CHALLENGES, 2012.

[17]. Gangwar, D P & Scientist, Senior &Anand, (2021). AUTHENTICATION OF DIGITAL MP4 VIDEO RECORDINGS USING FILE CONTAINERS AND METADATA PROPERTIES. 10.21817/ijcsenet/2021/v10i2/211002004.

[18]. B. Video, "Beamr - Media Optimization for Better Monetization," Beamr. http://beamr.com.

[19]. S. Victor, A Master's Thesis, 'COMPARATIVE FILE STRUCTURE ANALYSIS OF VIDEO FILES SENT AND RECEIVED VIA WHATSAPP', B.F.A. University of the West Indies,2020.

[20]. ISO_IEC_14496-12_2015.