Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5520009, 10 pages
https://doi.org/10.1155/2021/5520009

Research Article

Research on Music Multiterminal Audio Authentication Based on Wireless Network

Dongmei Li

Henan Polytechnic Institute, Nanyang, Henan 473000, China

Correspondence should be addressed to Dongmei Li; 2007003@hnpi.edu.cn

Received 27 January 2021; Revised 20 February 2021; Accepted 11 March 2021; Published 27 March 2021

Academic Editor: Chi-Hua Chen

The current music multiterminal audio authentication algorithm does not consider the mutation of music signal, which leads to poor tamper detection ability and long time of audio authentication. By analyzing the characteristics and key technologies of wireless network, a wireless multiterminal audio system is established. The short-term energy calculation method is used to consider the sudden change of music signal. The music signal is divided into note segments, and chroma features of half order notes are extracted. The robust hash value is calculated by nonuniform quantization method. The dynamic time warping algorithm is used to align the notes, and the Hamming distance between the hash values of each two corresponding notes is calculated to obtain the measurement values of error series, statistical characteristics, and time distribution characteristics. According to the measurement value, the fuzzy classification method is applied to calculate the membership degree of the signals belonging to two different types of operation, and the authentication confidence degree is obtained. The tampered area of the music signal that has not passed the authentication is detected, and the music multiterminal audio authentication is realized. Experimental results show that the proposed algorithm has good tamper detection ability and can effectively shorten the audio authentication time.

1. Introduction

With the maturity of multimedia compression technology and the rapid popularization of Internet, the creation, storage, and transmission of multimedia digital works such as image, video, and audio have become extremely convenient [1]. The massive music information represented by MP3 has been widely spread on the Internet. With the wide use of modern audio processing tools, the processing of digital audio signal becomes very simple. However, it also means that the tampering of audio semantic information can be carried out at a lower cost [2]. For example, the semantics of audio signal may change fundamentally after a simple rearrangement or removal of a few small segments, which cannot be detected only by human auditory perception. Audio authentication technology is an effective technical means to protect the integrity and authenticity of music, voice, and other audio data [3]. It can ensure that the received audio data in the transmission process, without malicious editing and tampering by a third party, that is, in the sense of human perception system and the original audio, is exactly the same. The technology is widely used in many fields: government departments, national security, court defense, trade secrets, news, recorded speech, music recording and distribution, military, and so on.

In order to get secure multimedia applications, it is more important to protect and authenticate the authenticity and integrity of audio content. The ideal audio content authentication system should be able to accurately distinguish the content keeping operation and malicious tampering and can accurately locate the tampered area. At present, a large number of scholars have done research on a music multiterminal audio authentication algorithm and achieved some theoretical results. In reference [4], a dynamic authentication watermarking algorithm for IOT signals is proposed to detect network attacks. The watermark algorithm is based on the long-term and short-term memory structure of deep learning, which enables the Internet of things devices to extract a group of random features from the generated signals and embed these features into the signals dynamically. This

algorithm enables IOT gateway which collects signals from iotd to effectively verify the reliability of signals. In reference [5], a physical layer authentication scheme based on angle of arrival (AOA) estimation is proposed to cross verify the reported location information. Considering the multiantenna roadside unit under location deception attack, according to the Cramer Rao bound of AOA estimation and the existence of effective estimator, the basic limit of AOA estimation is given. The problem of determining whether the received signal is from the claimed GPS position is described as a two-sided hypothesis testing problem, and its solution is given by Wald test statics [6]. The closed form of correct decision probability (PD) and false alarm probability (PF) are given. The key to obtain reliable AOA measurement is the high accuracy of array response vector. But the above algorithm does not consider the signal mutation, resulting in poor tamper detection ability and long audio authentication time [7].

In order to solve the above problems, a music multiterminal audio authentication algorithm based on a wireless network is proposed. Using the characteristics and key technologies of wireless network, a wireless multiterminal audio system is established. Using the method of calculating short-term energy, considering the mutation of music signal, the music signal is segmented, chroma feature is extracted, and the robust hash value is calculated. The dynamic time warping algorithm is used to align the notes, the Hamming distance of the hash value is calculated, the error sequence is obtained, and the measurement index value is counted. The fuzzy classification method is applied to calculate the membership degree of the signal, and the authentication confidence is obtained. The tampered area is detected, and the music multiterminal audio authentication is realized. Experimental results show that the audio authentication time of the proposed algorithm is short, and it has good tamper detection ability [8].

The research contributions of the thesis include the following:

(1) By analyzing the characteristics and key technologies of wireless networks, a wireless multiterminal audio system is established

(2) The dynamic time warping algorithm is used to align the notes, and the Hamming distance between the hash values of each two corresponding notes is calculated to obtain the error sequence, the measurement value of the statistical feature, and the time distribution feature

(3) According to the measured value, the fuzzy classification method is used to calculate the membership degrees of the signals belonging to two different types of operations to obtain the authentication confidence. Detect the tampering area of the music signal that has not passed the authentication, and realize the music multiterminal audio authentication

The organization structure of the thesis is as follows. The second part discusses the related technologies of wireless network, the third part discusses the music multiterminal audio authentication algorithm, the fourth part conducts an experimental simulation, and the fifth part summarizes the paper.

2. Wireless Network Technology

2.1. Type of Wireless Network. The so-called wireless network refers to the network that can realize the interconnection of various communication devices without wiring. Wireless network technology covers a wide range, including global voice and data networks that allow users to establish long-distance wireless connections, as well as infrared and RF technologies that optimize short-range wireless connections [9]. According to the different network coverage, a wireless network can be divided into wireless wide area network (WWAN), wireless local area network (WLAN), wireless metropolitan area network (WMAN), wireless personal area network (WPAN), and wireless mesh network [10].

(1) WWAN: it mainly refers to the data communication through mobile communication, satellite, etc., with the largest coverage. Representative technologies include 3G, 4G, and 5G, and the general data transmission rate is above 2 Mb/s. As the standards of 3GPP and 3GPP2 are becoming more and more mature, some international standardization organizations are aiming at the next generation mobile communication system, which can provide higher wireless transmission rate and flexible and unified all IP network platform, generally known as 3G, enhanced IMT-2000, System Beyond IMT-2000, or 4G.

(2) WLAN: it is generally used for wireless communication between regions, and its coverage is small. The representative technology is IEEE802.11 series, which is also called WiFi network. The data transmission rate is 11~56 Mb/s, even higher.

(3) WMAN: it is a type of wireless network connecting several wireless LANs. Mobile data communication mainly through mobile phones or vehicle devices can cover most areas of the city [11]. The representative technology is IEEE802.20 series, which mainly studies the mobile broadband wireless access (MBWA) technology and the formulation of relevant standards. The standard emphasizes more on mobility, which is developed from the broadband wireless access (BBWA) of IEEE802.16.

(4) WPAN: at present, there are two wireless personal area network standards: wireless personal area network (WPAN, IEEE802.15.1) (Bluetooth) and low-speed wireless personal area network (LR-WPAN, IEEE802.15.4) (ZigBee). Bluetooth devices are generally battery devices with a coverage radius of 10 meters. ZigBee is more committed to ultralow power consumption networks. For example, for devices that can last about 10 years without changing the battery, ZigBee has a coverage radius of 50 meters [12].

(5) Wireless mesh network: wireless mesh network is a multihop ad hoc network, which is composed of mesh routers and mesh clients. Mesh routers constitute the backbone network and connect with a wired internet network, which is responsible for providing multihop wireless Internet connection for mesh clients. Wireless mesh network provides a wider network topology by storing and forwarding messages between AP. It can extend the existing wired or wireless network. Its biggest characteristic is that AP can not only act as an access point but also store and forward messages, playing the role of wireless router [13].

2.2. Characteristics of Wireless Network. Compared with a wired network, the main feature of wireless network is to completely eliminate the limitations of wired network and realize the wireless transmission of information. Specific wireless network features are as follows:

(1) Strong mobility: wireless network transmits network signals by transmitting radio waves. As long as it is within the range of transmission, it can use the corresponding receiving equipment to realize the connection to the corresponding network. This greatly gets rid of the limitation of space and time, which is beyond the traditional network.

(2) Strong expansibility: wireless network can be carried out by wireless signal at any time, and its network expansion performance is relatively strong, which can effectively realize the network expansion and configuration settings. Users will also become more efficient and convenient in accessing information. Wireless network not only expands the space range of people to use the network but also improves the use efficiency of the network [14].

(3) Low cost: generally speaking, the process of installing the wired network is more complicated. In addition to a large number of network cables and network cable connectors, the later maintenance cost of the wired network is very high [15]. The wireless network does not need to lay a large number of network cables and install a wireless network transmitting device. At the same time, it also creates a very convenient condition for the later network maintenance, which greatly reduces the cost of network installation and later maintenance.

2.3. Key Technologies of Wireless Network. Wireless network can effectively sense the changes of the external environment and then carry out deeper understanding and learning and effectively adjust and configure the relevant resources within the communication network, so as to meet the changes of the external environment. By fully learning from wireless cognitive network technology, it can not only solve the conflict between the growing demand of spectrum and limited spectrum resources but also effectively solve the problem of spectrum resource shortage and promote the reasonable improvement of spectrum application efficiency [16].

(1) Spectrum sharing: spectrum sharing can help users maximize the application probability of spectrum by managing interference items. Spectrum can be classified from different levels, according to different network architectures divided into component distribution and centralized. Centralization refers to the centralized processing of users' information by the central server and the distributed computing of cognitive terminals to determine the idle spectrum. Through different ways of spectrum allocation, it can be divided into cooperative and noncooperative. In the process of spectrum sharing, the filling sharing method is adopted, which can reduce the interference of primary users to the maximum while spectrum is idle.

(2) Spectrum sensing: in wireless network technology, spectrum sensing is one of the core technologies. This technology can provide valuable spectrum for the majority of users through spectrum hole, time domain, and frequency domain discovery. In essence, there are three kinds of signal detection methods that can detect the primary user autonomously, namely, cyclostationary feature detection, matched filter detection, and energy detection [17]. Among them, the detection of energy has good performance and easy operation, but it is easily restricted by objective factors, which makes the main signal difficult to identify. Detection-matched filter can effectively and quickly detect user information on the basis of clear user information, but in this process, many conditions need to be ensured, such as special receiver, frequency, and synchronous timing. The detection of cyclostationary feature can identify the noise energy and detect the main signal, but the calculation process is complex [18].

(3) Dynamic access: in wireless network technology, dynamic spectrum access technology can be divided into open sharing mode, multilayer access mode, and dynamic special application mode. Among them, in the dedicated dynamic mode, the primary user can completely control the spectrum and at the same time can choose the technology and service mode at will. Open sharing mode can share a variety of systems, and there is no interference between them. Compared with the above two modes, it is found that the multilayer access mode can completely get rid of the impact of this user's transmit power, which can not only effectively expand the scope of application but also further improve the information capacity and throughput [19].

2.4. Wireless Multiterminal Audio System. Wireless multiterminal audio system mainly includes three modules: control point (CP), digital media render (DMR), and digital media server (DMS). Among them, the control points are generally

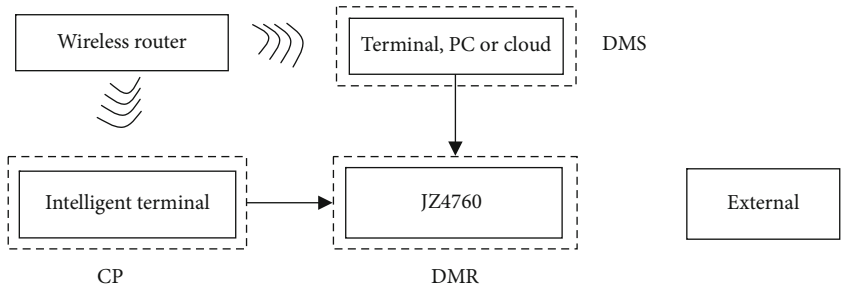

Figure 1: Connection diagram of each module of wireless multiterminal audio system.

mobile phones, tablet computers, and other intelligent terminals. DMR refers to the development board with wireless WiFi function. This paper uses Junzheng development board of MIPS processor. DMR can be either a home computer or an intelligent terminal. The connection of each module of wireless multiterminal audio system is shown in Figure 1.

It can be seen from Figure 1 that there is a logical relationship between the modules of the wireless multiterminal audio system and the wireless network plays a role in data output. The control point (mobile terminal) is first added to the same LAN with the wireless audio chip, and then the self-developed software is opened. The mobile terminal will find all the available devices on the LAN. Several devices can be selected to join the same group. After joining the group, the audio resources on the mobile terminal or other servers can be selected to play [20].

3. Music Multiterminal Audio Authentication Algorithm

3.1. Basic Framework of Music Multiterminal Audio Authentication Algorithm. A music multiterminal audio authentication algorithm is mainly divided into two stages: protection stage and authentication stage. The basic framework of the music multiterminal audio authentication algorithm is shown in Figure 2.

In the protection phase, firstly, an effective note segmentation algorithm is used to segment the music signal into a series of unequal length note segments. Then, based on each note segment, half order chroma features containing rich music semantic information are extracted. Finally, these feature vectors are transformed into binary hash authentication codes by a nonuniform quantization method, which are stored in a trusted second-order third-party certification center for future music certification. In the authentication stage, the music to be authenticated first needs to go through the same process of note segmentation, feature extraction and hash value calculation as in the protection stage to get the hash sequence of the music. Then, note alignment is carried out to reduce the different effects of note segmentation caused by various distortions in the transmission process [21]. Then, the h-value between the hash sequence of the current music segment and the authentication code is calculated in terms of notes Hamming distance to get an error sequence. Based on the error sequence, three new metrics reflecting hash difference, statistical characteristics, and time distribution characteristics are calculated. According to the three metrics, membership degrees of signals belonging to two different operations are calculated by using a fuzzy classification method, and the final "verification confidence" is obtained. Finally, for the music signals that have not passed the authentication, the system will also detect the tampered area testing [22].

3.2. Music Multiterminal Audio Authentication Algorithm Protection Stage

3.2.1. Music Segmentation. This paper adopts the method of calculating short-term energy [10] and, at the same time, considers the sudden change of the music signal in the high and low frequency parts for onset detection. The specific steps are as follows:

(1) The mute segments at the beginning and end of the music are removed, and the method of calculating short-term energy is used to detect the mute frame

(2) The music is decomposed in each frequency band as Table 1, and five subband signals are obtained

(3) The signal on each subband is divided into frames with a length of 60 ms, and there is a 50% overlap between two adjacent frames. For the medium and high frequency part of the signal (subband 2~5), use the energy difference to define the onset detection function:

$$E_i(n) = \sum_{m=(n-1)h+1}^{nh} |x_i(m)|^2, \quad i = 2, 3, 4, 5, \tag{1}$$

$$\mathrm{ons}_i(n) = E_i(n) - \sum_{a=1}^{10} \frac{E_i(n-a)}{a}, \quad i = 2, 3, 4, 5, \tag{2}$$

in which $x_i(m)$ represents the amplitude value of the i subband signal at time m, $E_i(n)$ represents the n frame energy of the i subband signal, h represents the frame length, and $\mathrm{ons}_i(n)$ represents the onset detection function of the i subband signal. Formula (2) uses the index in the method of differential weighting; the closer the frame to the current time is, the higher the weight [23].

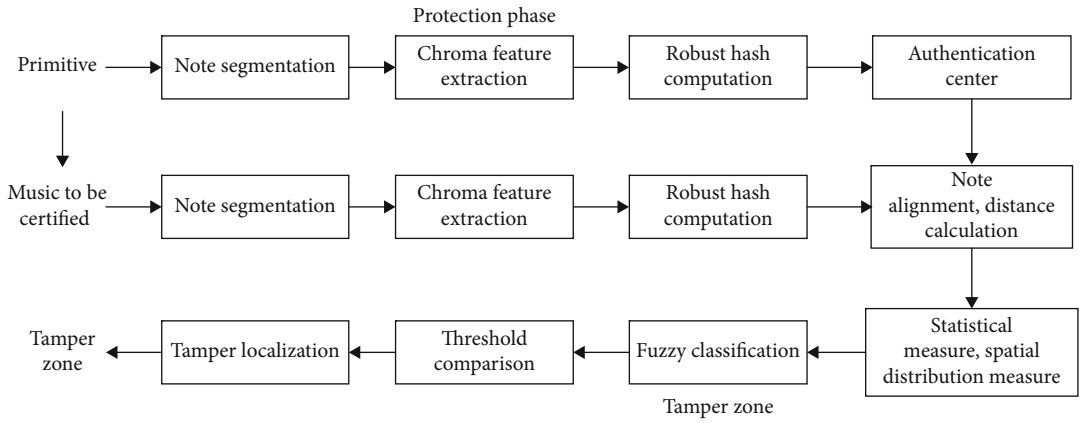

Figure 2: Basic framework of music multiterminal audio authentication algorithm.

Table 1: Frequency range of each subband during subband decomposition.

Subband No.	Freq range (Hz)
1	0~1024
2	1024~2048
3	2048~4096
4	4096~8192
5	8192~22050

For the low frequency part of the signal (subband 1), because the energy change is not obvious, the detection function is defined by using the change of the spectral coefficient [11]:

$$dX_n(k) = X(k, nh) - X(k, (n-1)h),$$
$$\text{ons}_1(n) = \frac{\sum_{k:DX_n(k)>0} dX_n(k)^2}{\sum_{k=1}^{N/2} |X(k, (n-1)h)|^2}, \quad (3)$$

in which $X(k, nh)$ represents the k FFT coefficient of the n frame of the first subband signal and N represents the Fourier transform length

(4) Perform a weighted summation of the detection functions on each subband to obtain the total onset detection function:

$$\text{ons}(n) = \sum_{i=1}^{5} \omega_{0i} \text{ons}_i(n), \quad (4)$$

in which $\omega_0 = [\omega_{01}, \omega_{02}, \omega_{03}, \omega_{04}, \omega_{05}]$ is the weighting coefficient

(5) Find the peak point of the detection function $\text{ons}(n)$ to determine the position of the note onset in the music

According to the above detection results, a music segment roughly corresponding to a note length is formed between two onsets [24].

3.2.2. Chroma Feature Extraction. For a music content authentication system, it is very important to select appropriate features that can fully reflect the semantic information of music. In this paper, we use chroma features, which are widely used in music retrieval, music structure analysis, and other fields [12]. The feature projects the whole spectrum distribution of music signal to 12 half order notes in a complete octave range, which has rich information of scale distribution and melody trend. Therefore, chroma is a 12-dimensional eigenvector, which is calculated frame by frame according to the following formula:

$$\text{Chroma}\left(K', n\right) = \sum_{K:P(K)=K'} X(K, n),$$
$$P(K) = \left[12 \cdot \log 2\left(\frac{K}{\text{NFFT}} \times \frac{f_s}{f_1}\right)\right] \bmod 12, \quad (5)$$

in which $\text{Chroma}(K', n)$ represents the K' dimension data of the chroma feature vector of the signal $x(n)$, $X(K, n)$ represents the K FFT coefficient of the signal $x(n)$, NFFT is the Fourier transform length, f_s represents the sampling frequency, and f_1 is the reference frequency [25].

In the specific implementation, first of all, each segment based on note segmentation is divided into fixed length frames; the frame length is 256. Then, chroma feature is extracted from each frame, and the mean value of chroma feature of each frame in each note segment is taken as the feature of the whole note. So far, each note segment of music signal is represented by a 12-dimensional feature vector.

3.2.3. Robust Hash Computation. Finally, the extracted Chroma features need to be quantized to generate a 36-bit authentication code for each note segment. The specific steps are as follows:

(1) The chroma vector was normalized:

$$p_o(i, j) = \frac{p(i, j)}{\sqrt{\sum_{k=1}^{12} p^2(i, k)}}, \quad (6)$$

in which $p(i,j)$ represents the j component of the feature corresponding to the i note and $p_o(i,j)$ represents the j component of the corresponding normalized feature vector. After the normalization operation, the value of each dimension component is between 0 and 1

(2) The normalized chroma features were quantized nonuniformly:

$$\widehat{p}(i,j) = \begin{cases} \text{floor}(p_o(i,j) \times 10, 0 \le p_o(i,j) < 0.7), \\ 7, 0.7 \le p_o(i,j) \le 1, \quad j = 1, 2, \cdots, 12. \end{cases} \tag{7}$$

In formula (7), $\widehat{p}(i,j)$ represents the j component of the quantized vector, and $\text{floor}(x)$ returns the largest integer not greater than x

(3) Express each $\widehat{p}(i,j)$ with the corresponding 3-bit binary, namely, $\widehat{p}(i,j) = [b_2 b_1 b_0]_2$, and connect all the binary bits to form a 36-bitHash code $h(i)$

Quantization of music features can not only reduce the storage space occupied by authentication information but also improve the robustness of various signal processing. The original music hash sequence is stored in a trusted third-party certification center for future music certification.

3.3. Music Multiterminal Audio Authentication Algorithm Authentication Stage

3.3.1. Note Alignment. In this paper, a dynamic time warping algorithm is used to obtain the most reasonable correspondence between two note sequences. The distance between every two notes is defined as the Hamming distance of their corresponding Hash value:

$$D(i,j) = \|h_o(i) - h_1(j)\| = \frac{1}{36}\sum_{k=1}^{36}[h_o(i,k) \oplus h_1(j,k)], \tag{8}$$

in which $D(i,j)$ represents the Hamming distance between the i note of the original music and the j note of the music to be authenticated, $h_o(i)$ is the Hash code of the i note of the original music, and $h_1(j)$ is the j note of the music to be authenticated Hash code. On the basis of note alignment, the Hamming distance of each two corresponding note Hash values is calculated, and the content integrity determination is further performed according to the sequence of these distance values.

3.3.2. Measurement Index. On the basis of note alignment, the Hamming distance between the hash values of every two corresponding notes is calculated, and all the distance values form a sequence diff. Define the possible tampering points in diff as those points whose distance value exceeds the set threshold T, denoted as PMP, whose index value is stored in the vector POS, as shown in

$$\text{PMP} = \left\{ \text{diff}(i) \middle| \text{diff}(i) \ge T, 1 \le i \le N' \right\},$$

$$\text{POS} = \{\text{pos}(j) | \text{diff}(\text{pos}(j)) = \text{PMP}(j), \quad j = 1, 2, \cdots, \|\text{PMP}\|\},$$

$$T = \begin{cases} \max(PMP) \times 0.5, \max(PMP) \ge T_0, \\ \text{median}(PMP), \max(PMP) \ge T_0, \end{cases} \tag{9}$$

in which $\text{diff}(i)$ represents the distance value corresponding to the i note, T is an adaptive threshold, and its setting takes into account the value of diff under acceptable operations and malicious tampering. The threshold T_0 is used to roughly judge whether the signal has serious distortion. Based on the above concepts, three measurement indicators reflecting the characteristics of diff sequence statistics and time distribution are defined. They have strong distinguishing ability for maintaining content operations and malicious tampering.

(1) Average distortion (AD): the average distortion AD of a signal to be authenticated is defined as the average bit error rate of all PMP points:

$$\text{AD} = \frac{1}{L_{\text{pmp}}}\sum_{j=1}^{L_{\text{pmp}}} \text{PMP}(j), \tag{10}$$

in which $L_{\text{pmp}} = \|\text{PMP}\|$. The average amount of distortion reverses the degree of change in the music content. Obviously, malicious tampering usually has a larger AD value compared to allowable operations.

(2) Uniform measure (UD): the uniformity metric is aimed at reflecting the uniformity of the error distribution. Let $\text{DIS} = \{\text{dis}(j) | \text{dis}(j) = \text{pos}(j+1) - \text{pos}(j), j = 1, 2, \cdots, \|\text{PMP}\| - 1\}$ denote the number of notes between adjacent PMP points, and define the uniformity metric as the standard deviation of DIS:

$$\text{UD} = \left[\frac{1}{N_{\text{dis}}}\sum_{j=1}^{N_{\text{dis}}} \left(\text{dis}(j) - \frac{1}{N_{\text{dis}}}\sum_{j=1}^{N_{\text{dis}}} \text{dis}(j) \right)^2 \right]^{1/2}, \tag{11}$$

in which $N_{\text{dis}} = \|\text{DIS}\|$. Obviously, the distribution of PMP points is more uneven by malicious tampering, the corresponding UD value is also larger, and the UD value corresponding to the allowable operation is smaller.

(3) Maximum connected area size (MC): a group of continuous dense points in the diff sequence form a connected area, and the size of the connected area is defined as the number of all points in the area. MC refers to the number of points contained in the largest connected area. Generally speaking, the MC value of malicious tampering is much larger than the allowable operation. This is because under malicious

tampering, the error tends to be very tightly concentrated in some local areas, while for the allowable operation, the error is often scattered on the entire timeline.

3.3.3. Music Content Certification. In the fuzzy classification method, firstly, it is necessary to define the membership functions for these three indexes, respectively, and describe the degree of their values to the two fuzzy sets of "large" and "small." Define the membership function of the average distortion AD as follows:

$$X_{As}(\mathrm{AD}) = \begin{cases} 1, & 0 \le \mathrm{AD} \le th_1, \\ \dfrac{1}{th_1 - th_2}(\mathrm{AD} - th_2), & th_1 < \mathrm{AD} \le th_2, \\ 0, & \mathrm{AD} > th_2, \end{cases}$$
$$X_{Al}(\mathrm{AD}) = \begin{cases} 0, & 0 \le \mathrm{AD} \le th_1, \\ \dfrac{1}{th_2 - th_1}(\mathrm{AD} - th_1), & th_1 < \mathrm{AD} \le th_2, \\ 1, & \mathrm{AD} > th_2, \end{cases} \tag{12}$$

in which $X_{As}(\mathrm{AD})$ and $X_{Al}(\mathrm{AD})$, respectively, represent the membership degrees of the smaller AD value and th_1 and th_2 are the two thresholds discussed above. Define the membership function of the uniform metric UD as

$$X_{Us}(\mathrm{UD}) = 1 - \frac{1}{1 + e - \alpha(\mathrm{UD} - \beta)},$$
$$X_{Ul}(\mathrm{UD}) = \frac{1}{1 + e - \alpha(\mathrm{UD} - \beta)}, \tag{13}$$

in which $X_{Us}(\mathrm{UD})$ and $X_{Ul}(\mathrm{UD})$, respectively, represent the degree of membership for which the value of UD is small and large. The β value of the parameter is the mean value of the music signal UD calculated under a series of content preservation and malicious tampering, and the parameter α control the entire S curve, especially the change speed at the sudden change point $\mathrm{UD} = \beta$. In this article, α and β are set to 25 and 0.08, respectively. The membership function that defines the maximum connected region size MC is as follows:

$$X_{Ms}(\mathrm{MC}) = \begin{cases} 1, & \mathrm{MC} \le \mu_1, \\ e^{-\left((\mathrm{MC}-\mu_1)^2\right)/2\sigma^2} & \mathrm{MC} > \mu_1, \end{cases}$$
$$X_{Ml}(\mathrm{MC}) = \begin{cases} 1, & \mathrm{MC} \ge \mu_2, \\ e^{-\left((\mathrm{MC}-\mu_2)^2\right)/2\sigma^2} & \mathrm{MC} < \mu_2, \end{cases} \tag{14}$$

in which $X_{Ms}(\mathrm{MC})$ and $X_{Ml}(\mathrm{MC})$, respectively, represent the membership degrees of the smaller MC value and the parameters μ_1 and μ_2 are set to $\mu_1 = 5$, $\mu_2 = 105$, respectively. Given a set of metric values $m = [\mathrm{AD}, \mathrm{UD}, \mathrm{MC}]$, its membership degree of fuzzy class C_i is expressed as follows:

$$X_{C_i}(m) = \sum_{j=1}^{3} w_j X_{C_{ij}}(m_j), \quad i = 1, 2, \cdots, 8, \tag{15}$$

in which $X_{C_i}(m)$ represents the degree of membership of m belonging to class iC_i, $X_{C_{ij}}(m_j)$ represents the degree of membership of m_j belonging to class C_{ij}, and the weight vector w_j reflects the importance of each index for content integrity authentication, which is determined by experiments:

$$D_y = \sum_{i=1}^{8} w_{yi} X_{C_i}(m),$$
$$D_n = \sum_{i=1}^{8} w_{ni} X_{C_i}(m), \tag{16}$$

in which D_y and D_n represent the final authenticity measure and nonauthenticity measure of the music signal, respectively, and w_{yi} and w_{ni} represent the weight contributions of various types of authentication passing and failing, respectively. Finally, the authentication credibility measure $authRatio = D_y/D_n$ is defined. If $authRatio > 1$, the music content is more likely to be authentic, and its authenticity increases with the increase of $authRatio$ value. If $authRatio < 1$, the music content is more likely to have been maliciously tampered with, and the possibility of tampering increases as the value of $authRatio$ decreases. If $authRatio = 1$, the system cannot make a decision.

3.3.4. Tamper Detection. For nonauthentic signals that fail to pass the music content authentication, all connected areas whose amplitude values are not less than a given threshold T are marked in the diff sequence. These areas correspond to the tampered part of the test signal. In this paper, the connected area is defined by the concept of 8 neighborhoods; therefore, the error of tamper detection is about 4 notes in the worst case (generally 1~2 s). Through the above steps, the music multiterminal audio authentication is completed.

4. Experimental Analysis

4.1. Experimental Environment and Data. In order to verify the effectiveness of the multiterminal audio authentication algorithm for music based on wireless networks, this article uses the voice set of the TIMIT voice library, 1280 segments of 4 s speech, of which there are 600 segments of English and 680 segments of Chinese, and the used speech parameters are as follows: sampling rate of 16000 Hz, bit rate of 256 kbps, single channel number audio channel, sampling precision of 16 bits, WAV format, frame length of 20 ms, and frame shift of 10 ms. The experimental hardware platform is Inter Core i3 processor, 2450 M, 8.00G memory, and 800G hard disk, and the experimental environment is Matlab R2012b under Windows7 operating system.

4.2. Tamper Detection and Analysis. When the music is judged as authentication failure, it is often necessary to detect the tampered area. Taking a classical music "the Blue

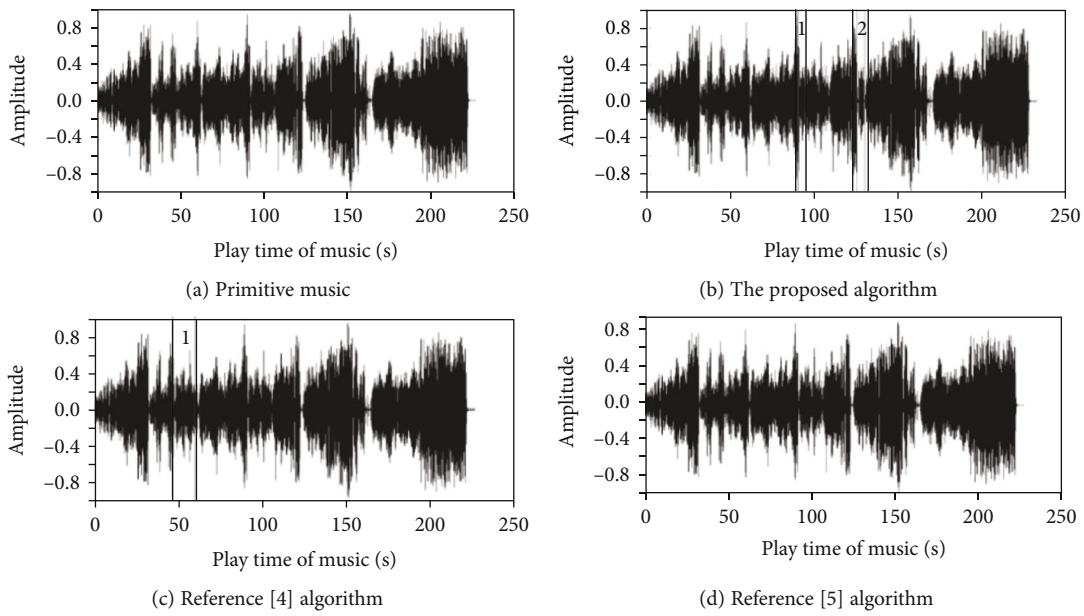

Figure 3: Comparison of detection results of different algorithms.

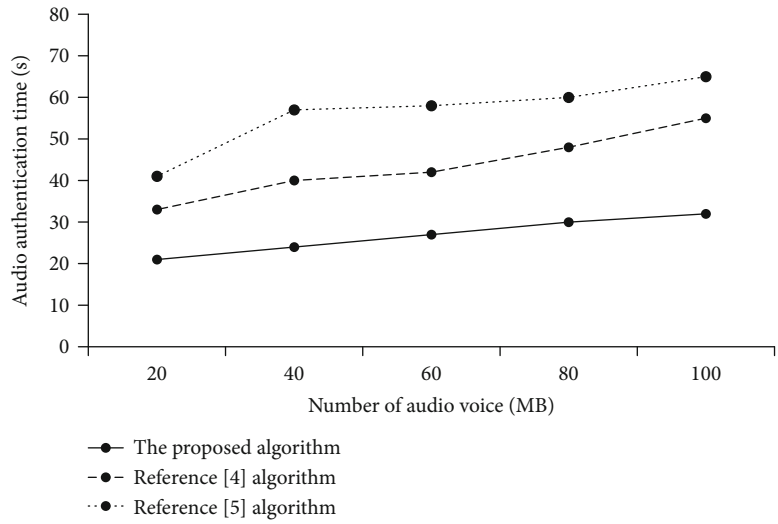

Figure 4: Comparison results of audio authentication time of different algorithms.

Danube" as an example, reference [4]'s algorithm, reference [5]'s algorithm, and the proposed algorithm are used to detect the tampered area, respectively, and the tampered area detection results of different algorithms are shown in Figure 3.

According to Figure 3, the tamper detection ability of reference [5]'s algorithm is poor, and the tampered area is not detected, followed by the tamper detection ability of reference [4]'s algorithm, which can locate one tampered area. The proposed algorithm has good tamper detection ability, which can effectively distinguish the admissible operation and malicious tampering. At the same time, most of the tampered areas can be accurately located by taking notes as the smallest unit, which has high accuracy in tamper detection.

4.3. Audio Authentication Time Analysis. Randomly extract 100 segments of speech from the speech database for operation, respectively, and use reference [4]'s algorithm, reference [5]'s algorithm, and the proposed algorithm for audio authentication and count the audio authentication time of different algorithms as shown in Figure 4.

According to the simulation results in Figure 4, as the number of audio voices increases, the audio authentication time of different algorithms will increase. The algorithm proposed in this paper is more sensitive to audio changes, and compared with the algorithm proposed in Reference [4] and Reference [5], it is more relaxed and the authentication time is shorter. When the audio voice volume reaches 100 MB, the audio authentication time of the algorithm in reference [4] is 67 s, the audio authentication time of the

algorithm in reference [5] is 59 s, and the audio authentication time of the proposed algorithm is only 32 s. Therefore, the audio authentication time of this algorithm is short.

5. Conclusion

This paper presents the research of multiterminal audio authentication for music based on a wireless network. According to the characteristics and key technologies of wireless networks, a wireless multiterminal audio system was established. The short-term energy calculation method is used to consider sudden changes in music signals. Segment the music signal, extract chroma features, and calculate a robust hash value. The dynamic time warping algorithm is used to align the notes, count the measured index values, use fuzzy classification to calculate the membership of the signal, obtain the authentication confidence, and detect the tampered area to realize the music multiterminal audio authentication. It can be seen from the experimental simulation that the algorithm proposed in the paper can effectively improve the tamper detection ability and shorten the audio authentication time. However, the research content of the paper still has some limitations. For example, the algorithm proposed in the paper does not consider different types of audio signals, and whether the algorithm has different performance results under different signals, these are the directions of future research. Therefore, in future research, the proposed algorithm will be extended to other types of audio signal content authentication, and it is hoped that an algorithm suitable for multiple signals can be obtained.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] H. Fang, X. Wang, and S. Tomasin, "Machine learning for intelligent authentication in 5G and beyond wireless networks," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 55–61, 2019.

[2] D. Chen, N. Zhang, R. Lu, N. Cheng, K. Zhang, and Z. Qin, "Channel precoding based message authentication in wireless networks: challenges and solutions," *IEEE Network*, vol. 33, no. 1, pp. 99–105, 2019.

[3] W. I. Khedr, K. M. Hosny, M. M. Khashaba, and F. A. Amer, "Prediction-based secured handover authentication for mobile cloud computing," *Wireless Networks*, vol. 26, no. 6, pp. 4657–4675, 2020.

[4] A. Ferdowsi and W. Saad, "Deep learning for signal authentication and security in massive internet-of-things systems," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1371–1387, 2019.

[5] A. Abdelaziz, R. Burton, F. Barickman, J. Martin, J. Weston, and C. E. Koksal, "Enhanced authentication based on angle of signal arrivals," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4602–4614, 2019.

[6] R. Diamant, P. Casari, and S. Tomasin, "Cooperative authentication in underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 954–968, 2019.

[7] T. Wang, K. Hu, X. Yang, G. Zhang, and Y. Wang, "A trust enhancement scheme for cluster-based wireless sensor networks," *Journal of Supercomputing*, vol. 75, no. 5, pp. 2761–2788, 2019.

[8] B. Groza, L. Popa, and P. S. Murvay, "Highly efficient authentication for CAN by identifier reallocation with ordered CMACs," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6129–6140, 2020.

[9] S. Li, M. Cheng, Y. Chen et al., "Enhancing the physical layer security of OFDM-PONs with hardware fingerprint authentication: a machine learning approach," *Journal of Lightwave Technology*, vol. 38, no. 12, pp. 3238–3245, 2020.

[10] K. Machat, "Gadgets/gizmos, inventions, and technology in Toronto," *Journal of the Audio Engineering Society*, vol. 67, no. 4, pp. 236–236, 2019.

[11] J. Zhang, H. Zhong, J. Cui, M. Tian, Y. Xu, and L. Liu, "Edge computing-based privacy-preserving authentication framework and protocol for 5G-enabled vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7940–7954, 2020.

[12] H. Liu, P. Zhang, G. Pu, T. Yang, S. Maharjan, and Y. Zhang, "Blockchain empowered cooperative authentication with data traceability in vehicular edge computing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4221–4232, 2020.

[13] H. Vasudev, V. Deshpande, D. Das, and S. K. Das, "A lightweight mutual authentication protocol for V2V communication in internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6709–6717, 2020.

[14] X. Sun, "Simulation of internet of things terminal dynamic authentication based on cluster architecture," *Computer Simulation*, vol. 37, no. 4, pp. 312–316, 2020.

[15] G. Bansal, N. Naren, V. Chamola, B. Sikdar, N. Kumar, and M. Guizani, "Lightweight mutual authentication protocol for V2G using physical unclonable function," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7234–7246, 2020.

[16] Z. Ye, C. Hu, L. He, G. Ouyang, and F. Wen, "The dynamic time-frequency relationship between international oil prices and investor sentiment in China: a wavelet coherence analysis," *Energy Journal*, vol. 41, no. 1, 2020.

[17] Z. He, F. Zhou, X. Xia, F. Wen, and Y. Huang, "Interaction between oil price and investor sentiment: nonlinear causality, time- varying influence, and asymmetric effect," *Emerging Markets Finance and Trade*, vol. 55, no. 12, pp. 2756–2773, 2019.

[18] C.-H. Chen, "An arrival time prediction method for bus system," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4231-4232, 2018.

[19] H. M. Liu, X. H. Bi, Z. F. Ye, and W. L. Wang, "Arc promoting inpainting using exemplar searching and priority filling," *Journal of Image and Grapgics*, vol. 21, no. 8, pp. 993–1003, 2016.

[20] J. Y. Lin, D. X. Deng, J. Yan, and X. Lin, "Self-adaptive group based sparse representation for image inpainting," *Journal of Computer Applications*, vol. 37, no. 4, pp. 1169–1173, 2017.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural

similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[22] T. Y. Fu, L. Q. Jin, Z. Lei, and Z. Q. Li, "Face super-resolution method based on key points layer by layer," *Journal of Signal Processing*, vol. 32, no. 7, pp. 834–841, 2016.

[23] L. I. Gang, L. I. Haifang, S. H. A. N. G. Fangxin, and G. U. O. Hao, "Noise image segmentation model with local intensity difference," *Journal of Computer Applications*, vol. 38, no. 3, pp. 842–847, 2018.

[24] L. WANG and C. PAN, "Robust level set image segmentation via a local correntropy-based K-means clustering," *Pattern Recognition*, vol. 47, no. 5, pp. 1917–1925, 2014.

[25] Z. Chen, H. Cai, Y. Zhang et al., "A novel sparse representation model for pedestrian abnormal trajectory understanding," *Expert Systems with Applications*, vol. 138, p. 112753, 2019.