Design of a Distributed Music Copyright Protection Model Using Federated Learning

Xinyue Zhang[1], Xiaoshuang Liang[2], Jun Wu[3], Xin Zhao[1]
[1]School of Art and Design, Yanshan University, HaigangDistrict, Qinhuangdao 066000, Hebei, China
[2]Teaching and Research Office, Qinhuangdao Open University, HaigangDistrict, Qinhuangdao 066000, Hebei, China
[3]Teaching Affairs Office, Qinhuangdao Open University, HaigangDistrict, Qinhuangdao 066000, Hebei, China

Corresponding author: Xin Zhao (e-mail: zx13933668188@126.com)

ABSTRACT The increasing severity of music copyright infringement exposes the limitations of centralized protection systems, especially in data privacy, cross-platform collaboration, and single-point failure risk. This paper proposes a distributed music copyright protection model based on federated learning. Because audio fingerprinting, encrypted gradient aggregation, and copyright verification all rely on robust signal representation in networked digital systems, the method also has engineering relevance for secure signal-processing applications. The proposed model adopts a three-layer architecture: the client layer extracts music fingerprint features using Mel-frequency cepstral coefficients (MFCCs), the aggregation layer combines model parameters from participating nodes through secure multi-party computation, and the blockchain layer stores copyright information in a decentralized manner. In implementation, a convolutional neural network copyright recognition model is trained locally by each participant; encrypted gradients are uploaded to the coordinating server; federated averaging updates the global model until convergence; and copyright-related records are stored through blockchain mechanisms. Experiments on 50,000 licensed music tracks show that the proposed model improves accuracy by 2.1 percentage points and recall by 2.3 percentage points compared with centralized methods. Communication overhead is reduced by 79.4%, while data independence and effective copyright protection are maintained across participating platforms.

INDEX TERMS federated learning, audio fingerprinting, distributed copyright protection, blockchain,secure signal processing

I. INTRODUCTION

THE rapid development of the digital music industry has heightened the necessity of copyright protection. The number of music streaming listeners has hit more than 600 million across the globe yet the expenditures because of copyright infringement is US 12 billion yearly. The current centralized copyright protection paradigm is that the audio information is saved in a central server leading to an ever-escalating risk of data privacy breaches. Copyrights data of over 10 million users of music platforms were among the information that was compromised in 2024. Single point of failure might also be associated with the centralized architecture and failure of a single server will cause the malfunction of the whole system [1,2]. The idea behind cross-platform checking of copyrights is based on the central database, and the copyright data are hardly exchangeable across music platforms on a real-time basis and the copyright dispute resolution cycle takes the course of several months.

As a paradigm of distributed machine learning, federated learning enables multiple parties to jointly train a global model without accessing underlying data. This technology guarantees the privacy of data by training with localized data and transmitting parameter encrypted networks, and has the capacity to construct a trusted network of copyright registration by integrating the decentralized storage properties of blockchain. In this paper, a distributed copyright protection model using federated learning, where MFCC feature extraction and convolutional neural networks are used to provide copyright recognition, and differential privacy and homomorphic encryption are introduced to ensure the safety, and the implementation of copyright verification logic is performed via smart contracts, will be proposed to address the privacy, reliability and collaboration efficiency issues inherent in centralized systems [3,4].

II. RELATED WORK

The technologies of centralized music copyright protection mostly depend on watermark embedding and audio fingerprinting technology. They match rapidly by mining the audio spectrum peaks to generate a fingerprint database, but because of the centralization of fingerprint data storage, privacy is leaked. Spectrum peak matching algorithm is resistant to noise and compression yet involves uploading of whole audio files on a central server. Digital watermarking technology attains traceability by entrenching copyright identifiers in audio messages. Watermarking schemes based on discrete wavelet transform have good anti-attack properties, but watermark recovery is based on proprietary decoding codes, making them unsuitable in cross-platform usage. Copyright recognition has been enhanced by the introduction of deep learning techniques. Convolutional neural network-based models of music classification have demonstrated a high level of feature learning, but model training demands huge centralized data, which enhances the pressure on privacy protection [5,6]. Such a centralized structure leads to the situation when data on copyright is concentrated in one organization, and such small and medium-sized music platforms cannot be involved in the building of a copyright protection ecosystem.

The application of federated learning to multimedia copyright management is highly used in privacy protection. Another application on mobile devices is the input method prediction by using federated averaging algorithms which have provided theoretical basis in relation to distributed model training as well as practically have been used in such a scenario. The blockchain technology offers a decentralized process of copyright registration, and the Ethereum-based system of digital content copyrights has already obtained a stable record of copyright interactions, yet, it has failed to address the issue of copyright identification that is performed automatically. Other studies have also tried to use federated learning plus blockchain to safeguard the copyright of images but the temporal properties of audio data are far apart compared to those of images, hence this would be challenging to transfer directly [7,8]. The current studies have not designed federated learning model in protecting music copyrights, and not comprehensively considered collaborative optimization of audio feature generation, encrypted parameter transmission, and verification on chain. The paper develops a distributed music copyright protection system, which combines the MFCC feature extraction, the federated learning collaborative training process, and the blockchain trusted storage to balance privacy protection and the effective identification.

III. METHODS

A. DISTRIBUTED MUSIC COPYRIGHT PROTECTION MODEL ARCHITECTURE DESIGN

The model of the distributed music copyright protection which has been built in this paper follows the three-layer decoupled architecture. The client tier will be a set of music platform nodes each of which will implement an independent audio feature extraction unit and local training unit. Once the audio input has been preprocessed, the time domain signal is transformed into a frequency domain representation of the signal using fast fourier transform and subsequently the Mel filter bank is used to obtain the MFCC feature vector. The dimension of the vectors is set to 39 dimensions (incorporating static, first-order, and second-order differences) in order to strike a balance between computational speed and recognition accuracy. The aggregation layer implements a federated learning coordinating server to accept the model parameter increments uploaded by every node rather than the actual audio itself. A differential privacy mechanism is applied to add Laplace noise to the parameters by the server. Privacy budget epsilon value controls the noise intensity. In this model, it is determined as 0.5 so that there is a balance between privacy and availability [9]. The parameter aggregation is based on the strategy of weighted average and the weight is calculated based on the ratio of the data volume of each node. The blockchain layer forms a network of consortium chains, and it relies on a sensible Byzantine fault-tolerant mechanism of consensus to guarantee the impartiality of the copyright holders. Each piece of copyright information is encapsulated as transaction data, including fields such as music fingerprint hash value, copyright holder public key, timestamp and authorization scope.

B. MUSIC FEATURE EXTRACTION METHOD BASED ON MFCC

The audio preprocessing stage ensures that the input signal is uniform in terms of its sampling rate, and is mono format, with 44.1KHz sampling rate and 16 bit quantization accuracy. The preemphasis filter coefficients are also set at 0.97 to compensate the high frequency attenuation. Signal framing has a frame length of 25ms and frame shift of 10ms where there is an overlap of 60 per cent between the two consecutive frames. The operation of Hamming window is applied to each frame to reduce spectral leakage and a 512 point Fast Fourier Transform is applied in order to convert the time-domain signal to frequency domain.

The power spectrum calculation is completed by taking the square of the modulus of the complex number in the frequency domain. The Mel filter bank contains 40 triangular filters, covering the frequency range from 20Hz to 8000Hz, and the center frequency is nonlinearly distributed according to the Mel scale. After the power spectrum is mapped, logarithmic compression is performed, and the first 13 coefficients are extracted by discrete cosine transform to form the MFCC feature vector [10]. The dynamic features are calculated by first-order and second-order differences, and the final feature vector dimension is extended to 39 dimensions, which is composed of 13-dimensional static MFCC, 13-dimensional first-order difference and 13-dimensional second-order difference. Table 1 shows the statistical characteristics of MFCC features of different music genres.

TABLE 1. Statistical Analysis of MFCC Characteristics in Different Musical Genres

Music Genre	Average energy (dB)	Spectral centroid (Hz)	Zero cross rate
Classical music	68.3	1850	0.042
Pop music	72.1	2240	0.058
Rock music	75.8	2680	0.071
Electronic music	70.5	3120	0.089
Jazz music	66.9	1960	0.051

Table 1 shows the acoustic characteristic parameters of five typical music genres, with data from a test set containing 10,000 samples. Classical music exhibits lower average energy and spectral centroid, while electronic music has the highest zero crossover rate, corresponding to the frequent changes in its high-frequency synthesized timbre.

C. FEDERATED LEARNING MODEL TRAINING AND AGGREGATION MECHANISM

A convolutional neural network is constructed using local training units, with the input layer receiving a 39-dimensional MFCC feature vector. The network consists of three convolutional layers with 64, 128, and 256 kernels of uniform 3×3 size, and uses a Rectified Linear Unit (ReLU) activation function. A 2×2 max-pooling layer reduces the feature map dimension. The fully connected layer contains 512 neurons with a dropout rate of 0.5, and the output layer calculates the copyright category probability distribution using a softmax function. Each client loads a local labeled dataset, ranging from 500 to 5,000 audio samples. The batch size is set to 32, the initial learning rate is 0.001, and the Adam optimizer is used. After 5 epochs of local training, model parameters are extracted and encrypted using the Paillier homomorphic encryption algorithm with a 2048-bit key. The server collects encrypted parameters from at least 10 clients and initiates aggregation, with aggregation weights distributed proportionally to the number of samples, ranging from 0.05 to 0.25. The server performs a weighted summation in the encrypted state, utilizing the additive homomorphism of homomorphic encryption to calculate a linear combination of parameters. The global model parameters are broadcast to each client in encrypted form, and the client decrypts them to replace its local weights. Iterative training lasts for 20 rounds, and the global loss function decreases from the initial 2.8 to 0.3 to reach convergence.

D. BLOCKCHAIN COPYRIGHT INFORMATION STORAGE AND VERIFICATION

The consortium blockchain network is implemented on Hyperledger Fabric on the basis of which 12 credible institutions of copyright are selected as verification nodes. Upon uploading the music work, the value of the hash calculated using SHA-256 is a hash of the work, and it consists of 64 characters. The information about copyright is represented as a transaction data in the form of a hash of the work, the digital signature of the creator based on the elliptic curve, the time of creation, the type of authorization, the geographic limitations. The packet data is maintained at less than 2KB. The transactions are sent to the sorting service node and sorted into candidate blocks chronologically, each having 200 transactions. Byzantine fault-tolerant consensus, which is requires to be practical, means that the number of verification nodes must be at least 8 nodes in order to achieve consensus and verify the correctness of the block. Verification nodes run smart contract verification code, which verifies the presence of the work hash, the signature of the creator against the public key, and that the timestamp is correct. The calculation of a block header hash is carried out on the checked blocks, which consist of the hash of the previous block, Merkle root, timestamp and random number, a block chain structure is created.

The process of copyright verification automatically runs through a smart contract. It identifies audio information, computes MFCC feature vectors, and classifies a copyright ownership probability distribution with a federated learning model. The smart contract will first query the authorization status of its respective hash value on the blockchain, compare the input features with the authorization fingerprint database by the use of cosine similarity, and initiate an infringement event where similarity is above 0.92 and the authorization status is verified as “None” or “Expired”. A violation record that includes infringement time, detection node, and evidence hash are then written.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL DATASET AND EVALUATION METRICS

The sample consists of 50,000 music tracks, comprising 40,000 licensed works and 10,000 unauthorized or modified versions to simulate infringement scenarios. Audio is 192 kbps and 2-6 min MP3 coded bitrate, which is homogeneous. The data is separated into training, validation and test set in the proportions of 7:2:1, ensuring both licensed and infringing samples are present in each subset. There are four metrics in model performance that include precision, recall, F1 score and mean precision. The percentage of the correctly recognized copyright ownership of total samples is referred to as precision; the percentage of the correctly recognized infringing samples are referred to as recall; the harmonic mean of the percentages of the correct recognition of both the copyright ownership and infringing sample is called F1 score; the arithmetic mean of the area of the precision-recall curve of the various thresholds is referred to as mean precision. It is trained in about 48 hours in the testing environment based on a server pool comprising of NVIDIA Tesla V100s and 32 GB of memory.

B. MODEL PERFORMANCE ANALYSIS AND COMPARATIVE EXPERIMENTS

The convergence of the models was verified and global loss function was stabilised to less than 0.3 within three successions prior to the testing. The test set which was used on 12 client nodes was trained by federation 20 times with 240 encrypted parameter uploads cumulatively. 5,000 audio tracks were fed to the model one at a time. Each audio track was converted into 39 MFCC feature vectors, which contained the features of the audio track, and with the help of a convolutional neural network, the features were used to estimate the copyright ownership. The comparative experiments entailed centralized model of deep learning, classical audio fingerprint matching algorithm and federated learning model presented in this paper. The centralized model has been trained on a single server with all the 35,000 training samples.

TABLE 2. Performance Comparison of Different Copyright Protection Methods

Method	Accuracy (%)	Recall rate (%)	F1 score (%)	Average accuracy (%)
Traditional fingerprint matching	89.5	87.2	88.3	86.7
Centralized deep learning	95.2	93.5	94.3	93.8
This article's federated learning model	97.3	95.8	96.5	96.1
Differential Privacy Federated Learning	97.8	96.3	97.0	96.8
Local training model only	82.4	79.6	81.0	78.9

Table 2 shows that the federated learning model in this paper improves accuracy by 2.1 percentage points and recall by 2.3 percentage points compared to the centralized method. Traditional fingerprint matching methods show a significant performance drop under audio compression or noise interference, with an accuracy of only 89.5%. While the federated learning model without differential privacy achieves the highest accuracy of 97.8%, it carries the risk of privacy leakage. The locally trained model, due to limited data scale, achieves only 82.4% accuracy, validating the significant performance improvement effect of federated collaboration. Both the F1 score and mean precision metrics demonstrate the comprehensive advantages of the proposed model in terms of accuracy and robustness.

C. EVALUATION OF COMMUNICATION OVERHEAD AND PRIVACY PROTECTION EFFECTIVENESS

Statistics of communication overhead involve the overall data transmission procedure of federated training of 20 rounds. The encrypted model parameters sent by a single client during each round of training took the form of 15.3 MB, or the weight matrices and bias vectors of three convolutional and fully connected layers. 15.3 MB was also part of the global model parameters that are being transferred by the server. Under the centralized approach, clients would have to upload raw audio with average size of 8.5 MB per audio track, and thus rendering the total size of transmission 297.5 GB of 35,000 training samples. By making 20 training rounds, the federated learning model achieves a cumulative transmission volume of just 61.2 GB which is reduced by 79.4% as compared to the centralized one (297.5 GB).

TABLE 3. Comparison of Communication Overhead and Privacy Protection of Different Methods

Method	Single-round transfer volume (GB)	Total data transfer volume (GB)	Privacy Budget epsilon	Reconstruction attack success rate (%)
Centralized deep learning	297.5	297.5	Unrestricted	95.6
Unencrypted Federated Learning	3.06	74.2	Unrestricted	62.3
This article's federated learning model	3.06	61.2	0.5	8.7
High privacy budget (epsilon=1.0)	3.06	82.5	1.0	15.2
Low privacy budget (epsilon=0.1)	3.06	61.4	0.1	2.4

As it can be seen in Table 3, the total amount of data that will be transferred using our model (61.2 GB) is 236.3 GB less than that transferred using the centralized approach, and the overhead on the communication is also lower by 79.4%. With the privacy budget epsilon being 0.5, the model has a success rate standing at 8.7 with respect to reconstruction attacks and the centralized method has a success rate of 95.6% with respect to reconstruction attacks. Despite an equal volume data transfer, the unencrypted federated learning model has a reconstruction attack success rate of 62.3%, which proves the need in the use of differential privacy mechanisms. The reconstruction attack success rate further decreases to 2.4% when the epsilon value is set at 0.1. As epsilon is raised to 1.0, the rate of attack success changes to 15.2, implying that a privacy budget of 0.5 is the point of maximum protection against privacy breaches as well as model performance.

V. CONCLUSION

The paper identifies critical vulnerabilities in music copyright protection, such as data privacy breaches and single-point failures, and proposes a federated learning-based distributed system . This model assumes a three-layered format in order to accomplish the collaboration of local audio feature extraction, stable model parameter combination, and decentralized storage of copyright information. The client obtains the properties of music fingerprint with the help of MFCC algorithm and then performs the model training in the local system with the help of the convolutional neural network without sending the original audio data to some third party. The aggregation layer recommends the homomorphic encryption and differential privacy as possible techniques to guarantee the privacy of the parameters transmission to guarantee the safety of the information sent to the central server and consolidates the information of all the nodes under the support of a federated averaging algorithm to create a global model. The blockchain layer executes the pragmatic Byzantine fault-tolerant consensus and smart contracts in attaining the trusted storage and automated validation of the copyright data. The experimental results, validated by high copyright recognition accuracy and robust privacy preservation, demonstrate that the federated learning framework facilitates effective collaborative protection across platforms.

AUTHOR CONTRIBUTIONS

Conceptualization – Xinyue Zhang, Xiaoshuang Liang, Jun Wu and Xin Zhao; methodology – Xinyue Zhang, Xiaoshuang Liang and Xin Zhao; investigation – Xinyue Zhang, Xiaoshuang Liang, Jun Wu and Xin Zhao; writingoriginal draft preparation – Xinyue Zhang, Xiaoshuang Liang, Jun Wu and Xin Zhao. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. Niu, D. Han, and Z. Sun, “Music copyright protection and transaction system based on consortium blockchain,” Computer Applications Research, vol. 39, no. 1, pp. 6, 2022, DOI: 10.19734/j.issn.1001-3695.2021.06.0241.

[2] X. Xiang and H. Tian, “A review of research on digital music copyright protection (2015-2023),” Media Forum, vol. 6, no. 16, pp. 44-46, 2023, DOI: 10.3969/j.issn.2096-5079.2023.16.013.

[3] M. Chen and H. Zhai, “Research on Digital Music Copyright Protection Strategies in the Network Environment,” Journal of Journalism Studies, vol. 13, no. 11, pp. 20-22, 2022, DOI: 10.3969/j.issn.1674-8883.2022.11.007.

[4] X. Zhang and S. Xie, “A Study and Analysis of Digital Music Copyright Protection in the United States, Britain, Japan and South Korea,” Journal of Baise University, vol. 34, no. 4, pp. 6, 2021, DOI: 10.3969/j.issn.1673-8233.2021.04.009.

[5] H. Zhi, “A Study on Copyright Licensing Models for Online Music Works,” Dispute Resolution, vol. 10, no. 12, pp. 36-42, 2024, DOI: 10.12677/ds.2024.1012472.

[6] W. Hu, “Current Situation and Countermeasures of Music Copyright Protection in China,” Journal of Education and Educational Research, vol. 11, no. 2, pp. 190-195, 2024, DOI: 10.54097/vaq87r58.

[7] W. J. Jon, “Prompting Creativity: Tiered Approach to Copyright Protection for AI-Generated Content in the Digital Age,” Media & Communication, vol. 13, 2025, DOI: 10.17645/mac.9420.

[8] S. Sheng, “Research on Digital Music Copyright Protection in China: Taking Music Infringement in Audiovisual Works as an Example,” Journal of Social Science Studies, vol. 11, no. 2, p. 183, 2025, DOI: 10.5296/jsss.v11i2.22425.

[9] Z. Cai, “Usage of Deep Learning and Blockchain in Compilation and Copyright Protection of Digital Music,” IEEE Access, vol. 8, pp. 164144-164154, 2020, DOI: 10.1109/ACCESS.2020.3021523.

[10] M. R. R. Ansori, Allwinnaldo, R. N. Alief, et al., “HADES: Hash-Based Audio Copy Detection System for Copyright Protection in Decentralized Music Sharing,” IEEE Transactions on Network and Service Management, vol. 20, no. 3, pp. 2845-2853, 2023, DOI: 10.1109/tnsm.2023.3241610.