Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083948

ARTICLE

Analytical Modeling of Transcoding Artifacts for Detecting SIMBox-Routed Calls

Hyunghoon Kim[1], Wonsuk Choi[2], Kyungho Joo[3,*] and Hyojin Jo[1,*]

[1]Graduate School of Information, Yonsei University, Seoul, Republic of Korea
[2]Graduate School of Information Security, Korea University, Seoul, Republic of Korea
[3]School of AI Software, Soongsil University, Seoul, Republic of Korea

*Corresponding Authors: Kyungho Joo. Email: khjoo@ssu.ac.kr; Hyojin Jo. Email: hyojin.jo@yonsei.ac.kr

Received: 14 April 2026; Accepted: 15 June 2026; Published: 27 July 2026

ABSTRACT: As mobile networks evolve toward next-generation architectures in which cellular and IP-based voice services are increasingly converged, SIMBox-based call routing has emerged as an important issue in modern telecommunication networks. By converting Voice over IP (VoIP) traffic into local cellular calls, SIMBox appliances allow IP-originated calls to appear as domestic cellular calls. Although SIMBox usage is not inherently fraudulent, detecting SIMBox-routed calls is important for identifying abnormal call-routing behavior and supporting network-side and client-side security applications. In this paper, we propose a client-side framework for SIMBox-routed call detection. Calls routed through SIMBox infrastructure are identified by exploiting acoustic artifacts introduced by VoIP-to-VoLTE codec transcoding. Variable-length call recordings are segmented into fixed-duration windows and analyzed using supervised classifiers to capture spectral patterns associated with transcoding operations. Unlike network-side SIMBox detection methods that rely on carrier-controlled metadata such as call detail records, subscriber identifiers, or cell-location patterns, our approach focuses on acoustic evidence observable from the call audio itself. We evaluate the proposed framework using two datasets: 25,706 codec-processed voice samples generated from 12,853 source recordings with the ITU-T G.191 Software Tool, and 200 real-world call samples collected through a commercial SIMBox platform (DINSTAR UC2000-VE) under realistic call-routing conditions. Among five evaluated models, the CNN-based classifiers achieve the best performance, reaching an F1-score of 1.00. These results show that audio-level codec artifacts can serve as reliable indicators for detecting SIMBox-routed voice calls at the client side, and offer a promising direction for securing converged voice services in 5G and beyond.

KEYWORDS: SIMBox-routed call detection; VoIP-to-VoLTE transcoding; client-side call analysis

1 Introduction

SIMBox-based call routing has emerged as an important issue in modern telecommunication networks [1]. The Communications Fraud Control Association (CFCA) estimated that global telecommunications fraud losses reached $38.95 billion in 2023 [2]. A SIMBox bridges internet-based telephony and cellular networks by converting Voice over IP (VoIP) calls into local cellular calls, thereby allowing IP-originated or overseas-originated calls to appear as ordinary domestic cellular calls.

SIMBox appliances are also deployed for legitimate purposes, such as reducing international call termination costs [3]. Nevertheless, detecting SIMBox-routed calls remains important because SIMBox devices can obscure the original call-routing path and introduce abnormal routing behavior that is difficult to verify from the client side. Recent studies indicate that SIMBox-based bypass remains an active and evolving

threat to cellular networks. Kouam et al. describe international bypass, also known as SIMBox-based bypass, as a practice in which international cellular voice traffic is diverted from regulated routes and re-originated as local calls in the destination country [4]. Their findings further show that modern SIMBox operations can adopt diverse behavior-mimicking strategies, which makes reliable SIMBox detection a continuing challenge for mobile operators.

Existing approaches for detecting voice phishing or SIMBox activity can be broadly categorized into three groups according to the analyzed data source: (1) Call Detail Record (CDR)-based analysis at core network switches, (2) audio-based analysis exploiting physical characteristics of voice signals, and (3) signaling-based detection using cellular control-plane data. CDR-based methods analyze communication metadata collected at core network switches to identify abnormal calling patterns associated with SIMBox devices [5,6]. Although effective in early studies, modern SIMBox appliances increasingly employ Human Behavior Simulation (HBS) techniques that mimic legitimate subscriber behavior, thereby reducing the effectiveness of CDR-based detection [1]. Audio-based approaches analyze characteristics of received speech signals, such as codec fingerprints, packet loss patterns, or signal energy anomalies, to infer the presence of VoIP–GSM gateways [3,7,8]. Although these methods are promising, most prior work focuses on legacy GSM environments and does not account for artifacts introduced by VoIP-to-VoLTE codec transcoding, which commonly occurs in modern SIMBox operations. Finally, signaling-based approaches detect SIMBox activity using cellular control-plane data at base stations [9,10]. These methods analyze device behavior during network attachment and can identify SIMBox devices before calls are established. However, they require access to operator infrastructure and do not analyze voice signals observable from the client side.

Despite these efforts, existing methods remain limited in two important respects. CDR- and signaling-based approaches rely on operator-side infrastructure, whereas audio-based approaches have primarily focused on legacy GSM environments. To the best of our knowledge, no prior work has investigated client-side detection of SIMBox-routed calls by leveraging acoustic artifacts introduced by VoIP-to-VoLTE codec transcoding. To address this gap, we propose a client-side SIMBox-routed call detection framework that exploits acoustic artifacts introduced during VoIP-to-VoLTE codec transcoding. Specifically, our framework detects SIMBox-routed calls by analyzing acoustic distortions caused by codec transitions such as G.711/G.729 to AMR-WB. Since call recordings vary in duration, the input audio is divided into fixed-size time windows, and each segment is analyzed independently. Segment-level predictions are then aggregated using majority voting to produce a call-level decision.

In summary, this paper makes the following contributions:

- We propose a client-side SIMBox-routed call detection framework that leverages acoustic artifacts introduced by VoIP-to-VoLTE codec transcoding. Unlike network-side approaches that rely on carrier-controlled metadata, the proposed method uses only audio evidence observable from the received call.
- We evaluate the proposed framework using two datasets: 25,706 codec-processed voice samples generated from 12,853 source recordings with the ITU-T G.191 Software Tool, and 200 real-world call samples collected through a commercial SIMBox platform (DINSTAR UC2000-VE) under realistic call-routing conditions. Among five evaluated models, the CNN-based models achieve the best performance with an F1-score of 1.0.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 provides background. Section 4 presents the proposed detection framework. Section 5 reports the experimental setup and results. Section 6 introduces discussion and future work. Finally, Section 7 concludes the paper.

2 Related Work

Prior research on SIMBox detection can be broadly categorized into three approaches depending on the analyzed data source: CDR-based analysis, audio signal analysis, and signaling-based detection.

2.1 CDR-Based SIMBox Detection

A major line of research detects SIMBox fraud using communication metadata collected in operator networks. Murynets et al. [5] analyzed SIMBox fraud in mobility networks using CDR-derived behavioral features and showed that fraudulent SIM cards can often be identified through abnormal usage patterns. Veloso et al. [6] further studied interconnect bypass fraud as a data-stream mining problem and demonstrated that repeated and distributed calling behaviors can be exploited for fraud detection.

These studies established the effectiveness of metadata-based detection in identifying anomalous SIMBox usage. Nevertheless, such approaches rely on the assumption that fraudulent devices exhibit distinguishable behavioral irregularities in operator-side records. As summarized by Kouam et al. [1] modern SIMBox systems increasingly employ evasion strategies such as Human Behavior Simulation (HBS), which are specifically designed to mimic legitimate subscriber behavior. This evolution reduces the reliability of purely CDR-based detection, particularly when attackers adapt their traffic patterns to resemble benign users.

2.2 Audio-Based Detection of Suspicious Call Paths

Another line of work analyzes the received speech signal itself to infer information about the call path. Balasubramaniyan et al. [7] proposed *PinDr0p*, which uses single-ended audio features to determine call provenance, showing that telephony paths leave measurable traces in the delivered audio. Reaves et al. [3] further demonstrated that cellular interconnect bypass fraud can be mitigated at the network edge by exploiting characteristics associated with SIMBox-mediated call delivery. In a related forensic context, Sampaio and Nascimento [8] showed that codec processing history can be inferred from speech signals by detecting AMR double compression using compressed-domain features.

Taken together, these studies suggest that audio signals can encode useful evidence about the communication path and codec transformations applied during transmission. However, prior work has primarily focused on legacy telephony environments, generic call provenance, or recompression detection, rather than explicitly modeling the acoustic artifacts introduced when VoIP codecs are transcoded into VoLTE codecs in modern SIMBox deployments. Consequently, the problem of detecting SIMBox-routed calls through *VoIP-to-VoLTE transcoding artifacts* remains insufficiently explored.

2.3 Signaling-Based SIMBox Detection

Recent work has also explored SIMBox detection using signaling information available at the cellular edge. Oh et al. [9] proposed a device model fingerprinting approach that identifies suspicious devices from signaling behavior. More recently, Kouam et al. [10] introduced *SigN*, which detects SIMBox activity through latency anomalies observed during cellular operations. These methods are attractive because they can identify suspicious devices before calls are fully established.

Despite this advantage, signaling-based approaches require access to operator infrastructure and control-plane data, which may limit their deployability in practice.

3 Background

3.1 VoIP and VoLTE Codecs

A codec (coder–decoder) encodes and decodes digital audio signals while compressing them for efficient transmission. In telecommunication systems, speech codecs are designed to represent human voice signals efficiently within the frequency range relevant to speech communication. Different communication systems adopt different codecs depending on bandwidth constraints and target quality of service. Table 1 summarizes representative codecs used in VoIP and VoLTE environments.

Table 1: Representative codecs used in VoIP and VoLTE environments.

Network	Codec	Bitrate (kbps)
VoIP	G.711	64
VoIP	G.722	48/56/64
VoIP	G.723.1	5.3/6.3
VoIP	G.726	16/24/32/40
VoIP	G.729	8
VoIP	GSM	13
VoLTE	AMR-NB	7.72–12.2
VoLTE	AMR-WB	6.6–23.85
VoLTE	AMR-WB+	6–48
VoLTE	EVS	5.9–128

Note: Abbreviations: VoIP, Voice over IP; VoLTE, Voice over LTE; EVS, Enhanced Voice Services.

3.2 Mel-Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used features in speech and audio processing because they approximate the nonlinear frequency perception of the human auditory system [11]. The MFCC extraction pipeline consists of the following steps.

1. Pre-emphasis and framing: The raw audio signal is first passed through a pre-emphasis filter to amplify high-frequency components. The signal is then divided into overlapping short-time frames, typically 20–40 ms in length with approximately 50% overlap.
2. Windowing and FFT: Each frame is multiplied by a Hamming window to reduce spectral leakage and is then transformed into the frequency domain using the Fast Fourier Transform (FFT), yielding the power spectrum.
3. Mel filterbank: The power spectrum is processed using a bank of triangular filters spaced according to the mel scale, which is defined as

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right), \tag{1}$$

where f denotes frequency in hertz. The inverse relationship is given by

$$f = 700\left(10^{m/2595} - 1\right). \tag{2}$$

The mel scale is approximately linear below 1 kHz and logarithmic above 1 kHz, reflecting the nonlinear characteristics of human auditory perception.

4. Log compression and DCT: The logarithm of the mel filterbank energies is computed, and the Discrete Cosine Transform (DCT) is applied to decorrelate the coefficients, producing the final MFCC feature vector. In this study, 40 MFCC coefficients are extracted from each audio sample.

MFCC features are particularly suitable for SIMBox detection because codec transcoding systematically alters the spectral envelope of speech signals. The mel-scale filterbank emphasizes perceptually relevant frequency bands in which such transcoding artifacts can become more pronounced.

4 Proposed Method

4.1 Key Observation

SIMBox-based call routing is a representative form of bypass routing in telecommunication networks, in which international calls transmitted over VoIP are converted into local calls using SIMBox equipment. The message flow of a SIMBox-routed international call proceeds as follows: a call originating from a VoIP endpoint is routed over VoIP through a relay or gateway to a SIMBox located in the destination country. The SIMBox then forwards the call to the recipient as a local call [1]. During this process, the voice signal undergoes two critical transformations: (a) codec transcoding from VoIP codecs (e.g., G.711 at 8 kHz and G.729 at 8 kHz) to VoLTE codecs (e.g., AMR-WB at 16 kHz), resulting in irreversible spectral information loss; and (b) packet loss during VoIP transmission, with packet loss concealment (PLC) algorithms introducing additional detectable artifacts.

Fig. 1 illustrates the speech conversion process using the ITU-T G.191 Software Tools [12]. A 44 kHz speech waveform file is provided as input to the simulator, which consists of a codec module (encoder/decoder) and a packet loss generator. By applying the configured VoIP and VoLTE processing conditions, the simulator produces converted speech waveform files. In this study, the converted signals were generated using (1) the VoIP codec G.711 (8 kHz), (2) the VoLTE codec AMR-WB (16 kHz), and (3) G.711-to-AMR-WB transcoding.

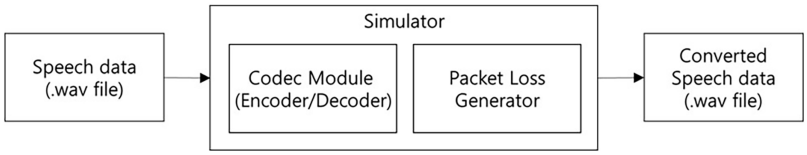

Figure 1: Speech conversion process using the ITU-T G.191 Software tools.

Fig. 2 compares the spectrograms of the original and processed speech signals. The original 44 kHz signal in Fig. 2a contains frequency components extending up to approximately 15 kHz. As shown in Fig. 2b, G.711 encoding removes most components above 4 kHz. In contrast, the AMR-WB-encoded signal in Fig. 2c preserves frequency components up to approximately 8 kHz while attenuating higher-frequency regions. Fig. 2d shows the signal obtained after sequential processing with G.711 and AMR-WB. The absence of spectral components above 4 kHz indicates that the information removed by G.711 encoding is not recovered by subsequent AMR-WB encoding. This characteristic pattern forms a distinctive double-compression spectral signature of SIMBox-routed call transcoding.

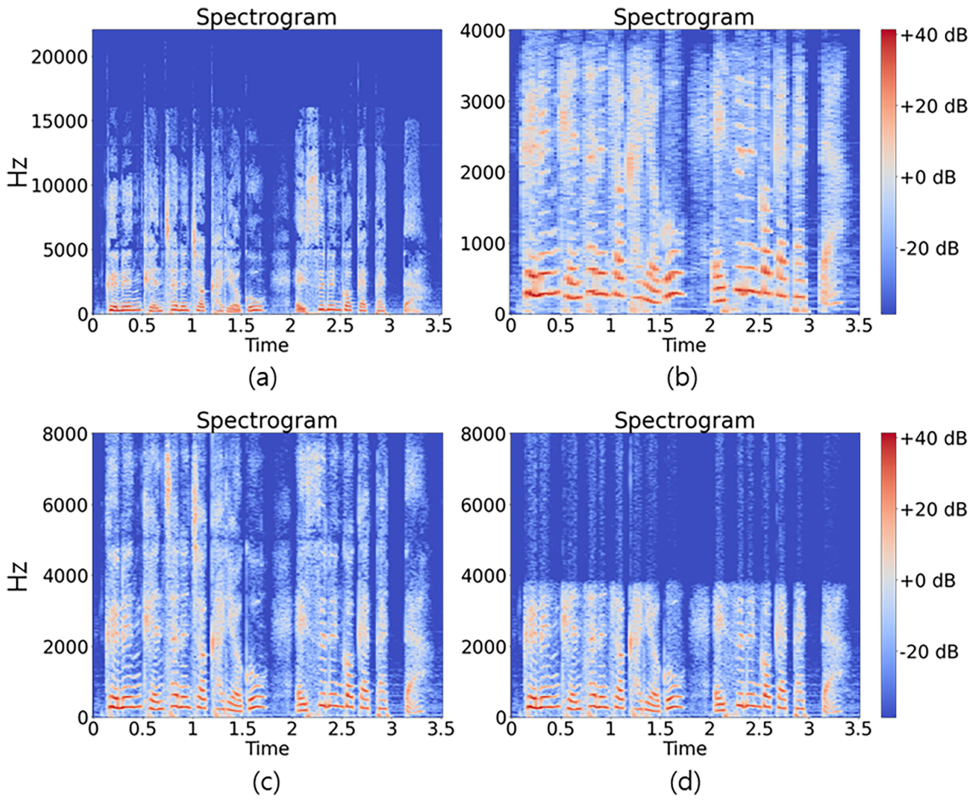

Figure 2: Spectrograms. (**a**) 44 kHz original, (**b**) G.711, (**c**) AMR-WB, (**d**) G.711→AMR-WB.

4.2 SIMBox-Routed Call Detection

The proposed method adopts a client-side SIMBox-routed call detection pipeline as shown in Fig. 3.

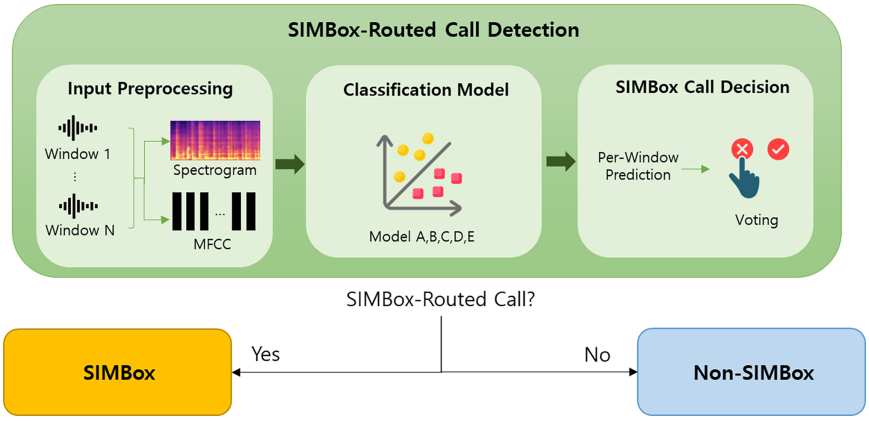

Figure 3: Overview of the proposed method.

The framework determines whether an incoming call has been routed through a SIMBox by analyzing codec transcoding artifacts. Since call recordings are variable in length, each audio stream is segmented into fixed-duration time windows, and each segment is independently classified. The segment-level predictions are then aggregated using a majority-voting scheme to obtain the final call-level decision.

The proposed SIMBox-routed call detection pipeline consists of three components: input preprocessing, detection modeling, and SIMbox call decision as shown in Fig. 3. First, variable-length voice calls are segmented into fixed-duration windows and converted into model-specific input representations. Next, each segment is classified using one of the supervised detection models. Finally, the segment-level predictions are aggregated through majority voting to determine the SIMBox call decision.

Input Preprocessing. The input audio is first divided into fixed-duration. Each time-windowed segment is then processed independently and converted into the input representation required by the classifier. For the deep learning models, each audio segment is transformed into a spectrogram using the Short-Time Fourier Transform (STFT). In contrast, for the traditional machine learning models, a 40-dimensional MFCC feature vector is extracted from each time-windowed audio segment.

Classification Model. Five models are designed as supervised binary classifiers. Among them, Models A and B are deep learning-based classifiers, whereas Models C, D, and E are traditional machine learning classifiers. The configurations of these models are summarized in Table 2.

Table 2: Summary of model configurations.

Model	Input	Architecture	Hyperparameters
A	Spectrogram (129 bins, 60-frame, 50% overlap)	$[\text{Conv2D}(c \in \{16, 32, 64, 64\}, k \in \{5, 5, 5, 3\}) \rightarrow \text{BN} \rightarrow \text{ReLU} \rightarrow \text{MaxPool}]^{\times 4} \rightarrow \text{FC} \rightarrow \text{ReLU} \rightarrow \text{Dropout}(0.3) \rightarrow \text{FC} \rightarrow \text{ReLU} \rightarrow \text{FC}(2)$	Adam, lr = 10^{-3}, epochs = 50, batch = 256, Dropout = 0.3
B	Mel-spectrogram (128 bins, 30-frame, 50% overlap)	$[\text{Conv2D}(c \in \{16, 8, 4\},\ k = 3) \rightarrow \text{Sigmoid} \rightarrow \text{MaxPool}]^{\times 3} \rightarrow [\text{LSTM}(h = 64)]^{\times 2} \rightarrow \text{FC}(64) \rightarrow \text{FC}(32) \rightarrow \text{FC}(16) \rightarrow \text{FC}(2)$	Adam, lr = 10^{-3}, epochs = 50, batch = 256
C	MFCC (40-dim,80 ms)	SVM	RBF kernel, $C = 10$
D	MFCC (40-dim, 80 ms)	Random Forest	$n = 100$, max_depth = 30
E	MFCC (40-dim, 80 ms)	k-NN	$k = 20$

SIMBox Call Decision. Each time window is independently classified, producing a binary output for each segment. Let $y_i \in \{0, 1\}$ denote the prediction for the i-th segment, where $y_i = 0$ indicates legitimate VoLTE traffic, i.e., non-SIMBox-routed traffic, and $y_i = 1$ indicates SIMBox-routed traffic (double compression). For an input call divided into n segments, the segment-level predictions are represented as $y_1, y_2, \ldots, y_n \in \{0, 1\}$.

The final call-level decision, denoted by S, is determined through majority voting over the segment-level predictions. Specifically,

$$S = \begin{cases} 1, & \text{if } \sum_{i=1}^{n} y_i \geq \left\lceil \frac{n}{2} \right\rceil, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where $S = 1$ indicates a SIMBox-routed call and $S = 0$ indicates a legitimate call. This aggregation strategy improves robustness against transient misclassifications and increases decision confidence for longer calls.

5 Experimental Setup and Results

5.1 Experimental Setup and Dataset

We constructed two complementary datasets to evaluate the proposed SIMBox-routed call detection method: a simulation dataset generated using the ITU-T G.191 Software Tool and a real-world dataset collected through a commercial SIMBox call environment. The simulation dataset was used to evaluate the proposed method under controlled codec-transcoding conditions, whereas the real-world dataset was used to verify whether the codec-transcoding artifacts remain observable in practical SIMBox call-routing conditions.

For the simulation dataset, we used 12,853 Korean-language source recordings in `.wav` format. For each source recording, we generated two processed samples: (1) a VoIP sample and (2) a VoIP-to-VoLTE transcoded sample. Specifically, the VoIP sample was generated by applying a VoIP codec condition, and the VoIP-to-VoLTE sample was generated by sequentially applying VoIP and VoLTE codec processing using the ITU-T G.191 Software Tool. As a result, the simulation dataset contains 25,706 samples in total, consisting of 12,853 VoIP samples and 12,853 VoIP-to-VoLTE samples. In addition, we constructed a SIMBox call environment using commercial off-the-shelf equipment to validate the proposed method under realistic conditions. Table 3 summarizes the experimental platform.

Table 3: Equipment used in the SIMBox call environment.

Component	Equipment	Purpose
SIMBox	DINSTAR UC2000-VE	VoIP-to-VoLTE gateway that converts VoIP calls into local mobile calls
SIP Server	3CX SIP Server Software	Deployed on Google Cloud Platform; provides VoIP networking and SIP trunk connectivity
LTE Router	CNR-L580W	Provides internet connectivity for the SIMBox

Note: Abbreviations: SIP, Session Initiation Protocol.

The SIMBox device (DINSTAR UC2000-VE) functions as a VoIP-to-VoLTE gateway that converts incoming VoIP calls into local cellular calls using installed SIM cards. The 3CX SIP server, deployed on Google Cloud Platform, provides the VoIP backbone and SIP trunk connectivity, while the LTE router supplies internet access for the SIMBox device. This configuration, as shown in Fig. 4, emulates a realistic SIMBox deployment scenario in which VoIP calls are routed through a domestically located SIMBox before reaching recipients via the local VoLTE network. Using this environment, we collected the real-world dataset consisting of 200 call recordings: 100 VoIP samples and 100 VoIP-to-VoLTE samples.

Table 4 summarizes the composition of the two datasets used in our evaluation. In both datasets, VoIP-to-VoLTE samples represent the SIMBox-routed class, whereas VoIP samples are used as the non-SIMBox baseline class.

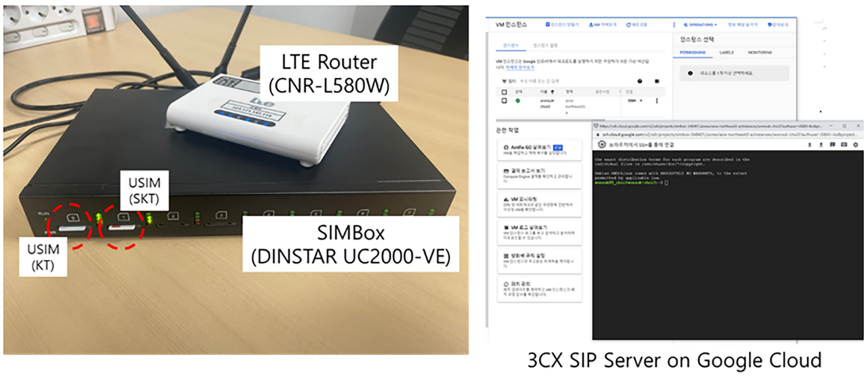

Figure 4: Experimental SIMBox call environment: SIMBox and LTE router (**left**), and the 3CX SIP server deployed on Google Cloud Platform (**right**).

Table 4: Summary of the evaluation datasets.

Dataset	Generation/Collection Method	Class	Samples per Class	Total Samples
Simulation	Generated from 12,853 source recordings using the ITU-T G.191 Software Tool	VoIP/VoIP-to-VoLTE	12,853/12,853	25,706
Real-world	Collected through DINSTAR UC2000-VE under practical call-routing conditions	VoIP/VoIP-to-VoLTE	100/100	200

5.2 Evaluation Metrics

The performance of the SIMBox-routed call detection models is evaluated using four standard classification metrics: accuracy, precision, recall, and F1-score. These metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{6}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. A true positive corresponds to a VoIP-to-VoLTE sample correctly classified as SIMBox-routed traffic, whereas a false positive denotes a VoIP sample incorrectly classified as SIMBox-routed traffic.

5.3 Evaluation on the Simulation Dataset

Table 5 reports the detection performance of the five evaluated models on the simulation dataset generated using the ITU-T G.191 Software Tool. For evaluation, 10-fold cross-validation was employed to assess the generalization performance of each model.

Table 5: Detection performance on the simulation dataset.

Model	Accuracy	Precision	Recall	F1-Score
A	1.000	1.000	1.000	1.000
B	1.000	1.000	1.000	1.000
C	1.000	1.000	1.000	1.000
D	1.000	1.000	1.000	1.000
E	1.000	1.000	1.000	1.000

5.4 Evaluation on the Real-World Dataset

Table 6 reports the detection performance of the five evaluated models on the real-world dataset collected through the commercial SIMBox platform. For evaluation, 10-fold cross-validation was employed to assess the generalization performance of each model. This evaluation verifies whether the transcoding-related patterns observed under controlled codec-processing conditions remain detectable under practical SIMBox call-routing conditions.

Table 6: Detection performance on the real-world dataset.

Model	Accuracy	Precision	Recall	F1-Score
A	1.000	1.000	1.000	1.000
B	1.000	1.000	1.000	1.000
C	0.995	1.000	0.990	0.995
D	0.990	1.000	0.980	0.990
E	0.995	1.000	0.990	0.995

5.5 Analysis of Model Performance

Among the evaluated models, Models A and B show the best overall performance. This result can be explained by the frequency-domain characteristics of VoIP-to-VoLTE transcoding. As shown in Fig. 2, G.711 encoding removes most spectral components above approximately 4 kHz, and the subsequent AMR-WB encoding process does not recover the lost high-frequency information. As a result, the VoIP-to-VoLTE transcoding path leaves a distinctive spectral cutoff and double-compression pattern.

Models A and B are both based on a CNN architecture and directly learn local time-frequency patterns from spectrogram representations. This makes them well suited for capturing frequency-domain artifacts such as high-frequency attenuation, spectral discontinuities, and repeated local patterns caused by codec transcoding. In contrast, traditional machine learning models based on MFCC features rely on compact hand-crafted representations, which may discard fine-grained frequency-bin information that is useful for distinguishing VoIP-to-VoLTE transcoding artifacts. This explains why the CNN-based model is particularly effective for SIMBox-routed call detection in our setting.

The majority voting mechanism across time windows further improves classification robustness by reducing the influence of transient segment-level misclassifications caused by silence intervals or low-energy frames. Nevertheless, additional real-world data collected from heterogeneous environments are required to further verify the robustness of the proposed approach in practical settings.

6 Discussion and Future Work

This study investigates client-side detection of SIMBox-routed calls based on acoustic artifacts introduced by VoIP-to-VoLTE codec transcoding. From this perspective, we identify two directions for future research. The first direction is to evaluate whether such audio-based detection remains effective in next-generation mobile networks, where voice services may rely on different codec chains, bandwidth configurations, and interworking mechanisms. The second direction is to explore how SIMBox-routed call detection can be incorporated into broader voice fraud risk assessment systems.

6.1 Application to Next-Generation Mobile Networks

The first direction is to examine whether audio-based SIMBox detection can be applied to next-generation mobile networks. This study focuses on acoustic artifacts introduced by VoIP-to-VoLTE codec transcoding, such as G.711/G.729 to AMR-WB. These artifacts arise because narrowband VoIP codecs remove or distort parts of the speech spectrum, and the lost information cannot be recovered by subsequent VoLTE encoding. However, next-generation voice services may involve different codec chains, bandwidth configurations, and packet loss concealment mechanisms. In particular, 5G voice services, such as Voice over New Radio (VoNR), may use newer codecs such as EVS and may interwork with VoLTE or legacy networks. These changes can alter the acoustic signatures observed in this study.

Therefore, future work should validate the proposed approach under next-generation voice settings. This includes constructing datasets that contain real-world SIMBox-routed calls over 5G voice services and controlled codec-chain samples involving EVS and VoNR-related configurations. Using these datasets, future work should evaluate whether codec-transcoding artifacts can still be learned as discriminative signals for SIMBox-routed call detection.

6.2 Application to Voice Fraud Risk Assessment

The second direction is to investigate how SIMBox-routed call detection can support broader voice fraud risk assessment systems. Although SIMBox routing alone does not imply fraudulent intent, it can provide useful channel-level evidence that a received call may have followed an abnormal routing path. Therefore, the output of the proposed detector should be used as a risk indicator rather than as a standalone fraud decision. One possible use case is a client-side warning system. Such a system could first apply the proposed detector to identify calls exhibiting SIMBox-related acoustic artifacts. If a call is classified as SIMBox-routed, the result could then be combined with other risk indicators, such as caller reputation, user reports, historical calling patterns, or content-level cues obtained from speech transcripts. In this setting, SIMBox routing would indicate channel-level suspicion, while other signals would provide additional context for determining whether a stronger warning is needed.

For example, a lightweight extension could analyze the transcript of a SIMBox-routed call using keyword-based filtering. According to prior work [13], suspicious linguistic cues may include law-enforcement impersonation, financial-institution impersonation, urgency or fund-transfer requests, and solicitation of sensitive personal information. Terms related to warrants, investigators, account freezes, urgent transfers, OTP numbers, or verification codes may provide additional evidence when they appear in a SIMBox-routed call. Under this application scenario, a call detected as SIMBox-routed but without

additional content-level indicators could be treated as a lower-priority alert, whereas a call that is both SIMBox-routed and associated with fraud-indicative linguistic cues could receive a stronger warning. In addition, recent transcript-based and LLM-based scam detection studies suggest that future systems could go beyond fixed keyword lists by using semantic classifiers to capture contextual fraud cues and paraphrased scam scripts [14–16]. However, such extensions require careful consideration of privacy, bias, hallucination, and adversarial paraphrasing risks [17]. Future work should therefore explore privacy-preserving and robust content-analysis methods suitable for on-device or edge-based deployment.

7 Conclusion

This paper presented a client-side SIMBox-routed call detection framework for VoLTE networks. The proposed method identifies SIMBox-routed calls by analyzing acoustic artifacts introduced during VoIP-to-VoLTE codec transcoding. In particular, the framework segments variable-length call recordings into fixed-duration windows and classifies each segment using supervised models to capture spectral patterns associated with transcoding operations. We evaluated the proposed framework using two datasets: a simulation dataset consisting of 25,706 codec-processed samples generated from 12,853 source recordings using the ITU-T G.191 Software Tool, and a real-world dataset consisting of 200 call recordings collected through a commercial SIMBox platform (DINSTAR UC2000-VE). The experimental results show that the CNN-based models achieve the best overall performance, indicating that spectrogram-based learning is effective for capturing the spectral cutoff and double-compression artifacts caused by VoIP-to-VoLTE transcoding.

These results demonstrate that audio-level codec-transcoding artifacts can serve as reliable evidence for detecting SIMBox-routed voice calls at the client side, without relying on carrier-controlled metadata such as CDRs, subscriber identifiers, or cell-location information. Future work will proceed in two directions. First, we plan to evaluate whether transcoding artifacts remain discriminative in next-generation mobile networks, including 5G voice services and EVS-based codec chains. Second, we will investigate how SIMBox-routed call detection can be used as a risk indicator in downstream voice fraud assessment systems, where it may be combined with other signals such as caller reputation, user reports, or transcript-based content cues.

Acknowledgement: We thank Sanghoon Shim and Gyeongyeon Lee for their contributions to data acquisition and model implementation.

Funding Statement: This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (2021-0-00511, Robust AI and Distributed Attack Detection for Edge AI Security, 50%) and (RS-2024-00439762, Developing Techniques for Analyzing and Assessing Vulnerabilities, and Tools for Confidentiality Evaluation in Generative AI Models, 50%).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Hyunghoon Kim, Kyungho Joo and Hyojin Jo; methodology, Hyunghoon Kim and Hyojin Jo; software, Hyunghoon Kim and Wonsuk Choi; validation, Hyunghoon Kim, Kyungho Joo and Hyojin Jo; data curation, Hyunghoon Kim and Wonsuk Choi; writing—original draft preparation, Hyunghoon Kim; writing—review and editing, Kyungho Joo and Hyojin Jo; supervision, Hyojin Jo; All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Hyojin Jo, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript

AMR-WB	Adaptive Multi-Rate Wideband
CDR	Call Detail Record
CFCA	Communications Fraud Control Association
EVS	Enhanced Voice Services
GSM	Global System for Mobile Communications
HBS	Human Behavior Simulation
LTE	Long-Term Evolution
MFCC	Mel-Frequency Cepstral Coefficient
PLC	Packet Loss Concealment
SIP	Session Initiation Protocol
SIMBox	Subscriber Identity Module Box
VoIP	Voice over IP
VoLTE	Voice over LTE

References

1. Kouam AJ, Viana AC, Tchana A. SIMBox bypass frauds in cellular networks: strategies, evolution, detection, and future directions. IEEE Commun Surv Tutor. 2021;23(4):2295–323.
2. Communications Fraud Control Association (CFCA). Global fraud loss survey 2023. Roseland, NJ, USA: CFCA; 2023.
3. Reaves B, Shernan E, Bates A, Carter H, Traynor P. Boxed out: blocking cellular interconnect bypass fraud at the network edge. In: Proceedings of the 24th USENIX Security Symposium; 2015 Aug 12–14; Washington, DC, USA.
4. Kouam AJ, Viana AC, Tchana A. Battle of wits: to what extent can fraudsters disguise their tracks in international bypass fraud? In: Proceedings of the ASIA CCS 2024; 2024 Jul 1–5; Singapore, Singapore. doi:10.1145/3634737.3657023.
5. Murynets I, Zabarankin M, Piqueras Jover R, Panagia A. Analysis and detection of SIMbox fraud in mobility networks. In: Proceedings of the IEEE INFOCOM 2014; 2014 Apr 27–May 2; Toronto, ON, Canada.
6. Veloso B, Tabassum S, Martins C, Espanha R, Azevedo R, Gama J. Interconnect bypass fraud detection: a case study. Ann Telecommun. 2020;75:583–96. doi:10.1007/s12243-020-00808-w.
7. Balasubramaniyan VA, Poonawalla A, Ahamad M, Hunter MT, Traynor P. PinDr0p: using single-ended audio features to determine call provenance. In: Proceedings of the ACM CCS 2010; 2010 Oct 4–8; Chicago, IL, USA. doi:10.1145/1866307.1866320.
8. Sampaio J, Nascimento F. Detection of AMR double compression using compressed-domain speech features. Forensic Sci Int Digit Investig. 2020;33:200907. doi:10.1016/j.fsidi.2020.200907.
9. Oh B, Ahn J, Bae S, Son M, Lee Y, Kang MS, et al. Preventing SIM box fraud using device model fingerprinting. In: Proceedings of the NDSS 2023; 2023 Feb 27–Mar 3; San Diego, CA, USA.
10. Kouam AJ, Viana AC, Martins P, Adjih C, Tchana A. SigN: SIMBox activity detection through latency anomalies at the cellular edge. In: Proceedings of the ASIA CCS 2025; 2025 Aug 25–29; Hanoi, Vietnam. doi:10.1145/3708821.3733902.
11. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Readings in speech recognition. IEEE Trans Acoust Speech Signal Process. 1980;28(4):357–66. doi:10.1016/b978-0-08-051584-7.50010-3.
12. Recommendation G.191. G.191: software tools for speech and audio coding standardization. Geneva, Switzerland: ITU-T; 2019.
13. Voice phishing on the rise again: an all-out response from blocking criminal tools to arresting offenders. korean national police agency. 2024 [cited 2026 Jan 1]. Available from: https://www.police.go.kr/component/file/ND_fileDownload.do?q_fileId=2dc2c440-8f31-4e46-961c-42a5acbe9071&q_fileSn=157605.

14. Shen Z, Wang K, Zhang Y, Ngai G, Fu EY. Combating phone scams with LLM-based detection: where do we stand? Proc AAAI Conf Artif Intell. 2025;39(28):29487– 9.
15. Ma Z, Wang P, Huang M, Wang J, Wu K, Lv X, et al. TeleAntiFraud-28k: an audio-text slow-thinking dataset for telecom fraud detection. In: Proceedings of the 33rd ACM International Conference on Multimedia; 2025 Oct 27–31; Dublin, Ireland. doi:10.1145/3746027.3755835.
16. Ma H, Su Q, Huang M, Kai W. Detecting continuously evolving scam calls under limited annotation: a LLM-augmented expert rule framework. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP; 2025 Sep 4–9; Suzhou, China.
17. Li W, Manickam S, Chong YW, Karuppayah S. Talking like a phisher: LLM-based attacks on voice phishing classifiers. arXiv:2507.16291. 2025.