

علمی - پژوهشی

ارائه یک روش بهبود یافته نهان کاوی گفتار در داده‌های صوتی مبتنی بر VoIP با استفاده از رویکرد یادگیری عمیق

حجت‌اله مقدسی^۱، حمید دهقانی^{۲*}

۱- دانشجوی دکتری ۲- دانشیار، دانشگاه صنعتی مالک اشتر

(دریافت: ۱۴۰۲/۰۳/۱۰، بازنگری: ۱۴۰۲/۰۴/۱۲، پذیرش: ۱۴۰۲/۰۵/۲۳، انتشار: ۱۴۰۲/۰۶/۱۷)

DOR: <https://dorl.net/dor/20.1001.1.26762935.1402.14.2.2.4>

چکیده

امروزه پروتکل انتقال صدا از طریق اینترنت (VoIP) به صورت گسترده در ارتباطات بلادرنک و شبکه‌های اجتماعی مورد استفاده قرار گرفته و به حامل مناسبی برای روش‌های نهان‌نگاری تبدیل شده است. در راستای مقابله با این تهدیدات، روش‌های متعدد نهان‌کاوی ابداع شده‌اند که در میان راه‌حل‌های ارائه شده، ترکیب روش‌های پردازش سیگنال و یادگیری ماشین، امکان ایجاد نهان‌کاوهای با دقت بسیار بالا را فراهم نموده است. در این مقاله یک رویکرد ترکیبی از روش‌های پردازش سیگنال گفتار و الگوریتم‌های هوش مصنوعی استفاده شده است. در این تحقیق ابتدا پیش‌پردازش داده بر روی سیگنال صوتی فشرده شده با کدک G.729 صورت می‌گیرد که ویژگی‌های درون‌فریمی و همبستگی‌های بین‌فریمی را با دقت خوبی استخراج می‌کند. سپس نتایج به دست آمده به یک شبکه یادگیری عمیق داده شده تا آموزش داده‌های پاک از داده‌های نهان نگاشته انجام گیرد. ارزیابی نتایج حاصل از پیاده‌سازی، میزان بهبود را هم در بخش صحت تشخیص و در بحث زمان محاسبات شامل می‌شود. روش پیشنهادی برای دو خانواده مهم نهان‌نگاری یعنی QIM و PMS مورد ارزیابی قرار گرفته و برای نرخ‌های مختلف ادغام روش مذکور تست و پیاده‌سازی شده است. نکته مهم دیگر تست برخط بودن روش ارائه شده بوده که برای فایل‌های ۱۰۰۰ میلی‌ثانیه‌ای، زمان پاسخ‌گویی کمتر از ۵ میلی‌ثانیه بوده که نشان از سرعت بالای مدل پیشنهادی در مرحله اجرا می‌باشد.

کلیدواژه‌ها: نهان‌نگاری، نهان‌کاوی، یادگیری عمیق، نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون، نهان‌نگاری مدولاسیون فرکانس گام

An Improved Speech Steganalysis Based on VoIP Using Deep Learning Approach

H. Moghadasi, H. Dehghani*

Malek Ashtar University of technology

(Received: 2023/05/31, Revised: 2023/07/03, Accepted: 2023/08/14, Published: 2023/09/08)

Abstract

Today, Voice over Internet Protocol (VoIP) is widely used in real-time communication and social networks and has become a suitable carrier for steganography methods. To confronting these threats, many steganalysis methods have been invented, among the proposed solutions, the combination of signal processing and machine learning methods has made it possible to create steganalysis methods with high accuracy. In this paper, a combined approach of speech signal processing methods and artificial intelligence algorithms is used. In this research, first, data pre-processing is done on compressed audio signal with G.729 codec, which extracts intra-frame features and inter-frame correlations with good resolution. Then the obtained results are given to a deep learning network to train cover data from stego data. The results of the implementation include the improvement in both the detection accuracy and the computation time. This method has been analyzed for two important steganography families, QIM and PMS, and the proposed method has been tested and implemented for different embedding rates. Another important point is the real-time test of the presented method, which for 1000 millisecond files, the response time was less than 5 millisecond, which shows the high speed of the proposed model in the execution phase.

Keywords: Steganography, Staganalysis, Deep Learning, Quantization Index Modulation (Qim), Pitch Modulation Ssteganography (Pms).

*Corresponding Author E-mail: hamid_deh@yahoo.com

۱. مقدمه

امنیت اطلاعات یک حوزه مهم است که در سالهای اخیر پیشرفت‌های زیادی کرده است و با فن‌هایی از قبیل نهان‌نگاری^۱ آمیخته شده است. نهان‌نگاری، هنر و علم انتقال پیام‌های مخفی به طریقی است که به‌جز گیرنده موردنظر، هیچ شخص دیگری از وجود پیام مخفی آگاهی پیدا نکند [۱]. در سال‌های اخیر، موضوع نهان‌نگاری از مهم‌ترین روش‌های مورداستفاده توسط مهاجمین بوده و از این‌رو آشنایی و مقابله با روش‌های نهان‌نگاری امری ضروری به نظر می‌رسد. در نهان‌نگاری، اغتشاش و تغییری که از لحاظ ادراکی یا تحلیلی قابل کشف باشد، نباید در درون سیگنال میزبان رخ دهد. منظور از کلمه ادراک، قوه ادراک آدمی (مثل بینایی یا شنوایی) است که نایستی قادر به تشخیص حضور وجود پیام در سیگنال میزبان شود و منظور از کلمه تحلیل این است که با استفاده از تحلیل‌های آماری نایستی وجود پیام مخفی تشخیص داده شود. مسئله پنهان کردن اطلاعات و انتقال مخفیانه آن موضوع جدیدی نیست، اما اولین کتاب واقعی در این زمینه توسط اسکات نوشته شد که در آن به چگونگی مخفی کردن پیام در بین نت‌های موسیقی اشاره شد. پنهان‌سازی پیام‌های محرمانه در فواصل خالی در حدود یک نقطه، گوشواره زنان، پاشنه‌های کفش پیک‌های خبرسان و استفاده از جوهرهای نامرئی روش‌های دیگر پنهان کردن اطلاعات بوده است. از دهه ۸۰ میلادی روش‌های پنهان‌سازی اطلاعات به‌صورت دیجیتال مطرح شدند [۲ و ۳]. به سیگنالی که ارسال مخفی آن هدف اصلی سیستم است، سیگنال پیام یا داده جاسازی شده گفته می‌شود. به سیگنالی که به‌عنوان بستری برای مخفی کردن سیگنال پیام به کار می‌رود، سیگنال پوشش^۲ گفته می‌شود. سیگنال میزبانی را که در آن سیگنال پیام جاسازی شده است، سیگنال نهان‌نگاری شده^۳ می‌گویند. سه اصطلاح مهم که در ارزیابی سیستم‌های نهان‌نگاری به کار می‌روند، ظرفیت^۴، مقاومت^۵ و شفافیت^۶ می‌باشد. ظرفیت حداکثر مقدار داده‌ای است که می‌توان در سیگنال میزبان جاسازی کرد، بدون این‌که وجود این داده قابل کشف باشد. مقاومت به‌صورت توانایی برای استخراج پیام مخفی پس از دست‌کاری سیگنال نهان‌نگاری شده توسط دشمن تعریف می‌شود. شفافیت به‌صورت توانایی برای اجتناب از سوءظن درباره وجود یک پیام مخفی تعریف می‌شود. این اصطلاح برای سنجش این موضوع استفاده می‌شود که پس از جاسازی پیام در سیگنال میزبان، تغییر ایجاد شده تا چه حد قابل تشخیص است [۱]. در بحث نهان‌نگاری، سیگنال میزبان می‌تواند از نوع صوت، تصویر یا ویدئو باشد که در

این مقاله ما حوزه صوت را مورداستفاده قرار داده‌ایم. نهان‌نگاری صوتی عموماً به دو گروه حوزه زمان و حوزه تبدیل تقسیم‌بندی می‌شوند. الگوریتم‌های حوزه تبدیل نسبت به الگوریتم‌های حوزه زمان در مقابل حملات مقاوم‌تر می‌باشند، اما مقاومت آن‌ها در سیگنال‌هایی که تعداد مؤلفه‌های حوزه تبدیلشان بسیار کم می‌باشد، رضایت‌بخش نیست [۴]. در ضمن این الگوریتم‌ها نسبت به الگوریتم‌های حوزه زمان پیچیده‌تر و از زمان بیش‌تری استفاده می‌کنند [۵]. در مقابل به کلیه عملیاتی که به‌منظور آشکارسازی یا تشخیص حضور یا عدم حضور اطلاعات پنهان‌شده می‌پردازد، نهان‌کاوی^۷ گفته می‌شود. تمرکز ما در بحث نهان‌کاوی در این مقاله بر روی سیگنال‌های گفتار مبتنی بر IP یا همان VoIP است. با توجه به اینکه صوت می‌بایست جهت انتقال از طریق شبکه IP به بسته‌هایی از جنس داده تبدیل شوند، دیجیتال شدن سیگنال صوت باید انجام پذیرد. در این سیستم‌ها در گام اول ابتدا صدا به حالت دیجیتال درآمده، سپس فشرده می‌شود و در نهایت به‌صورت بسته‌های داده ارسال می‌شوند. در طی مراحل تبدیل آنالوگ به دیجیتال، کدک^۸‌ها نقش مهمی ایفا کرده و نقش فشرده‌سازی و عملیات عکس آن را انجام می‌دهند. کدک‌های مختلفی تعریف شده‌اند که هر کدام استاندارد خاص خود را به‌منظور نمونه‌برداری و ارسال صدا تعریف کرده‌اند. یکی از کدک‌های معروف در بحث VoIP رشته G است که ما به دلیل استفاده و کاربرد گسترده، کدک G.729 را مورد ارزیابی قرار می‌دهیم. کدگذاری G.729 بر پایه مدل پیش‌بینی خطی با کد تحریک (CELP)^۹ بنا شده است [۶]. عملیات کد کردن در این کدک روی فریم‌های صوتی ۱۰ میلی‌ثانیه‌ای انجام می‌شود. هر فریم این کدک مرتبط با ۸۰ نمونه است که با نرخ نمونه‌برداری ۸۰۰۰ نمونه بر ثانیه به‌دست آمده است. برای هر فریم ۱۰ میلی‌ثانیه‌ای، سیگنال صوتی برای استخراج پارامترهای مدل CELP آنالیز می‌شود که نتیجه آن پارامترهای ضرایب فیلتر پیش‌بینی خطی^{۱۰}، نمایه و گین کد بوک‌های ثابت و تطبیقی^{۱۱} است. این پارامترها در بخش کدکننده، کد شده و فرستاده می‌شوند. در بخش کدگشایی، از این پارامترها برای به‌دست آوردن پارامترهای سیگنال استفاده می‌شوند. در بخش کدگذار این کدک، سیگنال ورودی از یک فیلتر بالاگذر عبور می‌کند و در بلوک پیش‌پردازش مقیاس‌دهی سیگنال صورت می‌گیرد. آنالیز پیش‌بینی خطی (LP)^{۱۲} برای محاسبه ضرایب فیلتر LP، روی هر فریم ۱۰ میلی‌ثانیه‌ای انجام می‌شود. این ضرایب به زوج طیف خطی (LSP)^{۱۳} تبدیل می‌شود و با روش کوانتیزاسیون برداری

⁷ Steganalysis

⁸ Codec

⁹ Code-Excited Linear-Prediction (Celp)

¹⁰ Linear-Prediction Filter Coefficients

¹¹ Adaptive And Fixed-Codebook Indices And Gains

¹² Linear Prediction

¹³ Line Spectrum Pairs

¹ Steganography

² Cover Signal

³ Stego Signal

⁴ Capacity

⁵ Robustness

⁶ Transparency

حوزه که مبنای مقایسه بسیاری از روش‌ها بوده، روش شبکه نهان کاو ادغام ویژگی (SFFN)^۶ است [۲۵]. این روش یکی از مهم‌ترین مدل‌ها برای شناسایی نهان‌نگاری موازی ناهمگن (HPS)^۷ در VoIP است که به طور گسترده در فریم‌های داده برای تشخیص وجود پیام‌های پنهان که در آن فریم‌های رسانه‌های جریانی با چندین روش نهان‌نگاری پنهان‌شده، مورد استفاده قرار می‌گیرد. شبکه نهان کاو ادغام ویژگی از سه زیرشبکه تشکیل شده است که شامل یک شبکه یادگیری ویژگی، یک شبکه ادغام ویژگی و یک شبکه طبقه‌بندی می‌باشد. این مدل می‌تواند با این سه زیرشبکه، ویژگی‌های نهان کاو را برای روش‌های نهان‌کاوی موازی ناهمگن به طور مؤثر استخراج کند. روش دیگر نهان‌کاوی کد المنت (CE)^۸ و مکانیزم توجه برای سیگنال‌های صوتی مبتنی بر روش Abs^۹ LPC معرفی شد [۲۶]. در این مقاله روش آشکارسازی نهان‌نگاری با استفاده از کد المنت‌ها، Bi-LSTM^{۱۰} و CNN^{۱۱} همراه مکانیزم توجه^{۱۲} ارائه شده است. ابتدا کد المنت‌های هر فریم به بردارهای ماتری هات^{۱۳} تبدیل شده‌اند. سپس هر کدام از این بردارها به یک بردار با اندازه ثابت نگاشت شده‌اند تا نمایش بهتری از داده‌ها ارائه شود. سپس شبکه‌های بازگشتی و کانولوشنی قرار داده شده‌اند که به ترتیب برای استخراج اطلاعات متنی و خصوصیات محلی بردارهای تعبیه‌شده استفاده شده‌اند. سپس از شبکه مکانیزم توجه استفاده شده است تا ضرایب متفاوتی بسته به اهمیت ویژگی‌ها به آن‌ها اختصاص داده شود. روش دیگر استخراج همبستگی سریع (FCEM)^{۱۴} برای تشخیص سریع سیگنال‌های نهان‌نگاشته صوتی است [۲۷]. در این مقاله، از سازوکارهای توجه به دلیل محاسبات بسیار موازی پذیر و انعطاف‌پذیر در مدل‌سازی همبستگی‌ها استفاده شده است. این روش برای تشخیص نهان‌نگاری مبتنی بر مدولاسیون شاخص کوانتیزاسیون در جریان VoIP مورد استفاده قرار می‌گیرد. در این روش یک شبکه عصبی سبک‌وزن به نام مدل استخراج همبستگی سریع بر اساس یک مکانیزم توجه چندگانه برای استخراج ویژگی‌های همبستگی از سیگنال فشرده‌شده صوتی استفاده شده است. این طرح در صحت تشخیص قابل‌مقایسه با مدل‌های شبکه‌های عصبی بازگشتی و شبکه‌های عصبی کانولوشنی است. روش دیگر در این حوزه، شبکه کانولوشنی و بازگشتی خاص (CNN-LSTM) برای تشخیص سیگنال‌های صوتی VoIP می‌باشد [۲۸]. معماری اصلی شبکه شامل شبکه‌های عصبی کانولوشنی و شبکه‌های عصبی بازگشتی است که روش‌های کاملاً متفاوتی را برای درک سیگنال‌های مختلف اتخاذ می‌کنند. در این

(VQ)^۱ پیش‌بینی‌کننده دوبخشی با ۱۸ بیت کوانتیزه می‌شود (نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون (QIM)^۲ روی این ۱۸ بیت اعمال می‌شود که در بخش ۳-۱ توضیح داده می‌شود). سیگنال تحریک با روند جستجوی آنالیز - با - سنتز (Abs)^۳ انتخاب می‌شود که در آن خطای بین سیگنال صوتی اولیه و بازسازی‌شده، کمینه می‌شود. پارامترهای سیگنال تحریک (پارامترهای کدبوک ثابت و تطبیقی) در هر نیم فریم ۵ میلی‌ثانیه‌ای (۴۰ نمونه) تعیین شده و ضرایب LP برای نیم فریم دوم استفاده می‌شود. برای نیم فریم اول از ضرایب درون‌یابی شده LP استفاده می‌شود [۶]. در ادامه با محاسبات ریاضی روی کل فریم، تناوب مدار باز فرکانس گام^۴، بر پایه سیگنال صوتی وزن‌دار شده، تخمین زده می‌شود. سپس عملیات زیر برای هر نیم فریم اجرا می‌شود. با استفاده از سیگنال هدف و پاسخ ضربه (پاسخ ضربه فیلتر سنتز وزن‌دار)، برای یافتن تناوب و گین کدبوک تطبیقی، آنالیز مدار بسته فرکانس گام با جستجو حول تناوب مدار باز فرکانس گام انجام می‌شود. با این آنالیز، تناوب مدار بسته فرکانس گام در هر نیم فریم با رزولوشن ۱ به ۳ به دست می‌آید. این تناوب در نیم فریم اول با ۸ بیت و در نیم فریم دوم با ۵ بیت کد می‌شود (نهان‌نگاری مدولاسیون فرکانس گام (PMS)^۵ روی این ۱۳ بیت اعمال می‌شود که این روش در بخش ۳-۲ توضیح داده می‌شود). لازم به ذکر است در بحث نهان‌نگاری شبکه‌های صوتی مبتنی بر VoIP به طور شاخص دو خانواده اصلی نهان‌نگاری QIM و PMS مورد ارزیابی و تحلیل قرار می‌گیرد. از این‌رو نهان‌کاوی متناظر با آن نیز بر روی این دو خانواده متمرکز شده است. در ادامه به رویکردها و روش‌های نهان‌کاوی در حوزه VoIP می‌پردازیم.

نهان‌نگاری اطلاعات مبتنی بر سیگنال گفتار بر اساس مکان تعبیه به سه دسته تقسیم‌بندی می‌شود. دسته اول تعبیه آنالیز فرکانس گام [۷] و بر اساس اصلاح راهبرد جستجوی کدبوک‌های ثابت و یا کدبوک‌های تطبیقی انجام می‌پذیرد [۸-۱۰]. دسته دوم از آنالیز ضرایب پیش‌بینی خطی و بر اساس مدولاسیون شاخص کوانتیزاسیون برای پنهان کردن اطلاعات استفاده می‌کنند [۱۱-۱۳]. دسته سوم اطلاعات را با اصلاح مستقیم مقدار بعضی از کد المنت‌ها در جریان گفتار فشرده‌سازی می‌شود [۱۴-۱۶]. لازم به ذکر است بسیاری از تحلیل‌های نهان‌کاوی بر اساس همبستگی درون‌فریمی و بین‌فریمی پایه‌ریزی شده است [۱۷-۲۰]. نکته مهم دیگر در فرایند نهان‌کاوی مرحله پیش‌پردازش داده [۲۱-۲۴] به‌عنوان ورودی شبکه جهت فرایند آموزش داده است. یکی از مهم‌ترین و جدیدترین روش‌ها در این

⁶ Steganalysis Feature Fusion Network

⁷ Heterogeneous Parallel Steganography

⁸ Code Element

⁹ Analysis-By-Synthesis Linear Predictive Coding

¹⁰ Bidirectional- Recurrent Neural Network

¹¹ Convolutional Neural Network

¹² Attention Mechanisms

¹³ Multi-Hot

¹⁴ Fast Correlation Extract Model

¹ Vector Quantization

² Quantization Index Modulation

³ Analysis-By-Synthesis

⁴ Open-Loop Pitch

⁵ Pitch Modulation Steganography

مدل، شبکه عصبی بازگشتی کوتاه مدت دو جهتی برای ضبط اطلاعات متنی طولانی مدت در حامل ها استفاده می شود و متعاقباً از CNN برای گرفتن ویژگی های محلی و سراسری و همچنین ویژگی های حامل زمانی استفاده می شود. روش دیگر نهان کاوی سیگنال های صوتی VoIP استخراج و ادغام ویژگی (KFEF)^۱ بر اساس لایه ادغام و کدگذاری سیگنال و مکانیزم توجه می باشد. این روش برای نهان نگاری ناهمگن و بر اساس تشخیص نهان نگاری مدولاسیون شاخص کوانتیزاسیون و مدولاسیون فرکانس گام کار می کند [۲۹]. همچنین روش نگاشت کد المنت برای نهان کاوی سریع و برخط و بر اساس همبستگی های موجود در سیگنال داده میزبان، تجزیه و تحلیل می شود [۳۰]. برای بهره برداری بهتر از همبستگی در کلمات کد، کد المنت ها به یک فضای معنایی^۲ نگاشت می شوند. در این روش تنها یک لایه پنهان برای استخراج همبستگی بین این کلمات کد استفاده می شود. در نهایت، بر اساس ویژگی های همبستگی استخراج شده، از یک طبقه بند برای دسته بندی حامل های جریان ورودی استفاده می کند. برای افزایش کارایی این مدل پیشنهادی از یک چارچوب ساده تقطیر دانش^۳ در آموزش نیز استفاده می شود.

۲. روش تحقیق

در ادامه این مقاله در بخش های بعدی با توجه به اینکه پایه نهان کاوی ما بر پایه دو نهان نگاری QIM و PMS می باشد، به توضیح و تحلیل آن می پردازیم. در بخش ۳-۱ به تحلیل نهان نگاری مدولاسیون شاخص کوانتیزاسیون، در بخش ۳-۲ به تحلیل نهان نگاری مدولاسیون فرکانس گام و در بخش ۳-۳ به معرفی مدل پیشنهادی خود مبتنی بر هوش مصنوعی برای طراحی یک سامانه نهان کاوی می پردازیم.

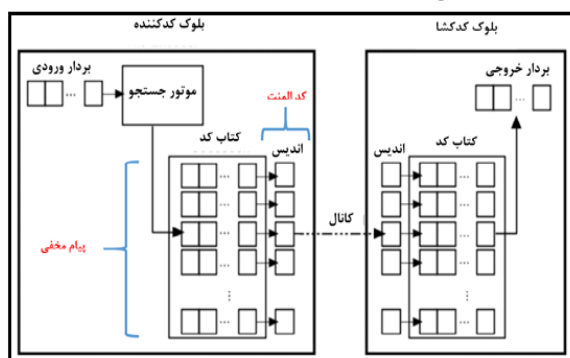
۲-۱. تحلیل نهان نگاری مدولاسیون شاخص کوانتیزاسیون

نهان نگاری کوانتیزاسیون ضرایب LSP در کدگذاری QIM روی G.729 اعمال می شود. در این بخش ابتدا روش کوانتیزاسیون برداری (VQ) که پایه و اساس این نوع نهان کاوی را تحلیل کرده و سپس عملیات کوانتیزاسیون برداری ضرایب LSP در کدگذاری G.729 را بررسی می کنیم. در نهایت نحوه نهان نگاری روی این کوانتیزاسیون را در الگوریتم QIM بررسی می کنیم. کوانتیزاسیون برداری روشی است که برای فشرده سازی سیگنال های صوتی در کدکننده های مختلف کاربرد دارد. کدکننده G.729 نیز از این روش در الگوریتم خود بهره برده است. در VQ، تعدادی بردار k بعدی در فضای برداری R^k ، روی مجموعه محدودی از بردارها در این فضا تصویر می شود. این مجموعه محدود به صورت $Y = \{y_i\}$

$$V_i = \{x \in R^k : \|x - y_i\| \leq \|x - y_j\|, \text{ for all } j \neq i\} \quad (1)$$

$$\bigcup_{i=1}^N V_i = R^k, \quad \bigcap_{i=1}^N V_i = \emptyset, \quad \text{for all } j \neq i$$

در VQ بردارهای موجود در هر ناحیه از فضا، روی کد واژه های مرتبط با آن نواحی تصویر می شود. به عبارت دیگر، کد واژه درون یک ناحیه به عنوان نماینده هر یک از بردارهای موجود در آن ناحیه در نظر گرفته می شود [۳۱]. کد واژه ای که نماینده یک بردار است، از باقی کد واژه ها به آن بردار نزدیک تر است. در عمل برای تصویر کردن یک بردار روی یک کد واژه، یک متریک مشخص (مثلاً فاصله اقلیدسی) آن بردار با هر یک از کد واژه ها محاسبه می شود و کد واژه ای که کمترین فاصله را با آن بردار داشته باشد به عنوان نماینده آن بردار انتخاب می شود. کدگذاری یک بردار به وسیله VQ به دو بخش کدگذار و کدگشا تقسیم می شود که در شکل (۱) نشان داده شده است. در کدگذار (بلوک کدکننده) بردار ورودی روی یکی از کد واژه ها (نزدیک ترین کد واژه) تصویر می شود و نمایه آن کد واژه در فرستنده ارسال می شود. در بخش کدگشا (بلوک کدگشا)، کد واژه مرتبط با این نمایه انتخاب می شود و به عنوان برداری خروجی نمایان می شود. بدین ترتیب در VQ به جای ارسال یک بردار ورودی k بعدی، تنها یک نمایه یک بعدی از شبیه ترین کد واژه فرستاده می شود. هر چه کد کتاب جامع تر باشد، دقت کدگذاری بالاتر و سرعت آن کمتر می شود، چراکه هر بردار ورودی با تمام کد واژه ها مقایسه می شود. از طرفی بر اساس اینکه پیام مخفی چه باشد، یک دسته بندی روی کتاب کد صورت گرفته و نمایه (کد المنت) مربوطه جهت ارسال در کانال انتقال انتخاب می گردد.



شکل ۱. کدگذاری یک بردار (فشرده سازی) در روش VQ.

^۴ Codeword

^۵ Codebook

^۱ Key Feature Extraction and Fusion Network

^۲ Semantic Space

^۳ Knowledge Distillation

بیت‌های مخفی در سیگنال صوتی، در کدگذاری سیگنال وارد می‌شود و در آن یک تغییر جزئی ایجاد می‌کند. این تغییر در مرحله کوانتیزاسیون ضرایب LSP است و در کدکتاب‌های L1، L2 و L3 اعمال می‌شود. در پنهان‌نگاری QIM، کد واژه‌های موجود در هر کدکتاب به دو دسته تقسیم می‌شود؛ دسته اول مرتبط با مقدار «یک» و دسته دوم مرتبط با مقدار «صفر» است. برای پنهان‌سازی بیت یک، کد واژه بهینه از دسته اول انتخاب می‌شود و برای پنهان‌سازی بیت صفر، انتخاب کد واژه بهینه از دسته دوم است. در بخش کدگشا برای استخراج بیت‌های مخفی، بررسی می‌شود که کد واژه دریافتی از کدام دسته است که بدین ترتیب بیت پنهان شده شناسایی می‌شود.

۲-۲. تحلیل پنهان‌نگاری مدولاسیون فرکانس گام

در کدگذار G.729 هر ۱۰ میلی‌ثانیه معادل ۸۰ نمونه است و یک فریم را تشکیل می‌دهد ($\{s(n)\}_{n=1,2,\dots,80}$). هر فریم به ۲ زیرفریم ۴۰ نمونه‌ای تقسیم می‌شود. برای پیش‌بینی دوره تناوب pitch ابتدا سیگنال اولیه $s(n)$ به سیگنال وزن دار $sw(n)$ تبدیل می‌شود. سپس دوره تناوب pitch برای کل فریم به صورت مدار باز با استفاده از مقادیر $sw(n)$ تخمین زده می‌شود. دوره تناوب pitch (T_{op})، در محدوده ۲۰ تا ۱۴۳ نمونه جستجو می‌شود. با استفاده از رابطه همبستگی متقابل^۳ طبق رابطه (۴) استخراج می‌شود [۳۲]:

$$R(k) = \frac{\sum_{n=1}^{80} sw(n)sw(n-k)}{\sqrt{\sum_{n=1}^{80} sw^2(n-k)}}, \quad 20 \leq k \leq 143 \quad (4)$$

پارامتر $R(k)$ به ازای مقادیر مختلف k محاسبه می‌شود و آن مقدار k که $R(k)$ بیشتری بدهد به عنوان دوره تناوب مدار باز pitch (T_{op}) در نظر گرفته می‌شود. پس از محاسبه T_{op} ، پیش‌بینی دوره تناوب pitch به صورت مدار بسته برای هر زیرفریم انجام می‌شود. برای دو زیر فریم، تناوب مدار بسته T_1 و T_2 در بازه $[19\frac{1}{3}, 84\frac{2}{3}]$ با رزولوشن $\frac{1}{3}$ و در بازه $[85, 143]$ با رزولوشن 1 انتخاب می‌شود. تناوب بهینه T_1 حول مقدار T_{op} در بازه $[T_{op}-3\frac{2}{3}, T_{op}+3\frac{2}{3}]$ و تناوب بهینه T_2 حول جزء صحیح T_1 در بازه $[\text{int}(T_1)-5\frac{2}{3}, \text{int}(T_1)+4\frac{2}{3}]$ جستجو می‌شود. انتخاب تناوب‌های مدار بسته T_1 حول مقادیر مرجع به گونه‌ای است که سیگنال اولیه با سیگنال بازسازی شده کمترین اختلاف را داشته باشد. هر T_i از دو جزء صحیح و کسری تشکیل می‌شود، به صورتی که در رابطه (۵) نشان داده شده است:

در کدگذاری G.729 در بخش کدگذار، آنالیز پیش‌بینی خطی (LP) روی هر فریم ۱۰ ms انجام می‌شود و ضرایب LP (a_0 تا a_{10}) استخراج می‌شود. ضرایب LP به ضرایب LSP (q_1 تا q_{10}) تبدیل می‌شود و ضرایب LSP به صورت ضرایب LSF (w_1 تا w_{10}) بازنویسی می‌شود ($q_i = \cos(w_{10})$). در نهایت ضرایب LSF با روش کوانتیزاسیون بردار (VQ) در قالب ۱۸ بیت کوانتیزه شده و در مجموعه بیت‌های ارسالی قرار می‌گیرد. این ۱۸ بیت از چهار بخش L0 (۱ بیت)، L1 (۷ بیت)، L2 (۵ بیت) و L3 (۵ بیت) تشکیل شده است. L0 مربوط به پیش‌بینی کننده دو حالته MA^۱ است و یکی از حالت‌های آن را که برای این فریم بهینه تر است، انتخاب می‌کند. L1، L2 و L3 مربوط به سه اندیس کد واژه از سه کدکتاب مورد استفاده در این کوانتیزاسیون است. کدگذاری G.729 برای کوانتیزاسیون برداری ضرایب LSP از سه کدکتاب L1، L2 و L3 استفاده می‌کند. کدکتاب L1 شامل ۱۲۸ کد واژه ۱۰ بعدی است که با اندیس‌های ۷ بیتی در کدکتاب آدرس‌دهی می‌شود. هر یک از کدکتاب‌های L2 و L3 شامل ۳۲ کد واژه ۵ بعدی است که با اندیس‌های ۵ بیتی در کدکتاب آدرس‌دهی می‌شود. برای بازسازی ضرایب LSF در بخش کدگشا ابتدا هر یک از ابعاد کد واژه‌های L1 ام، L2 ام و L3 ام (به ترتیب از سه کدکتاب L1، L2 و L3) به صورت زیر با هم جمع می‌شود و ۱۰ ضریب \hat{l}_1 تا \hat{l}_{10} محاسبه می‌شود که در رابطه (۲) نشان داده شده است.

$$\hat{l}_i = \begin{cases} L1_i(L1) + L2_i(L2) & i=1, \dots, 5 \\ L1_i(L1) + L3_{i-5}(L3) & i=6, \dots, 10 \end{cases} \quad (2)$$

ضرایب LSF با استفاده از رابطه (۳) کدگشایی می‌شود. در این رابطه $\hat{p}_{i,k}$ ضرایب پیش‌بینی کننده دو حالته MA است و به ازای $i=1, 2, \dots, 10$ و $k=1, 2, 3, 4$ ، چهل مقدار دارد. این ضرایب بسته به صفر یا یک بودن L0 مقادیر متمایزی دارد. اندیس‌های بالای پارامترها که در پرانتز نوشته شده، مربوط به شماره فریم است. فریم فعلی شماره m بوده و از ضرایب \hat{l}_i چهار فریم قبلی استفاده شده است. $\hat{\omega}_i$ ها همان ضرایب LSF اند.

$$\hat{\omega}_i^{(m)} = \left(1 - \sum_{k=1}^4 \hat{p}_{i,k} \right) \hat{l}_i^{(m)} + \sum_{k=1}^4 \hat{p}_{i,k} \hat{l}_i^{(m-k)} \quad i=1, 2, \dots, 10 \quad (3)$$

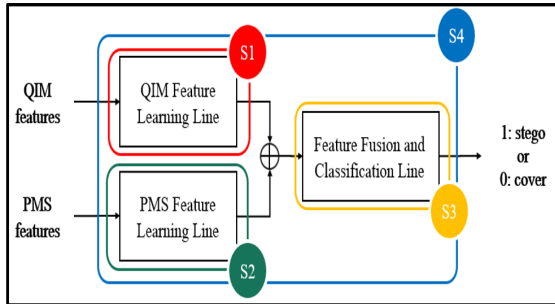
در کوانتیزاسیون برداری ضرایب LPC، بیت L0 (که ضرایب $\hat{p}_{i,k}$ را تعیین می‌کند) و اندیس‌های L1، L2 و L3 از کدکتاب‌های مربوطه به نحوی انتخاب می‌شود که اختلاف ضرایب LSF اولیه و کدگشایی شده کمینه شود. ترتیب انتخاب کد واژه‌های بهینه از سه کدکتاب، به ترتیب از L1 تا L3 است. در هر مرحله برای یافتن کد واژه بهینه، تمام کد واژه‌های موجود در کدکتاب جستجو و بررسی می‌شود. الگوریتم پنهان‌نگاری QIM برای پنهان کردن

^۳ Cross-correlation

^۱ Line Spectral Frequency (LSF)

^۲ Switched MA Predictor (MA : moving average)

تعلیم می‌بیند و اصلاح می‌شود (مرحله S4 در شکل (۲)). برای اجرای این چهار مرحله، از چهار کد پایتون با نام‌گذاری S1 تا S4 استفاده می‌شود.



شکل ۲. مراحل تعلیم شبکه پیشنهادی در سامانه نهان کاو.

فایل‌های صحبت مورد استفاده در این تحقیق یک‌ثابته‌ای با استاندارد G.729 است که یک فایل با فرمت g.729. به صورت متنی قابل خواندن است و ۱ kB حجم دارد. با باز کردن این فایل به صورت متنی، ۱۰۰۰ B در اختیار خواهیم داشت. از آنجا که در استاندارد G.729، به هر فریم ۱۰ ms ای ۸۰ bit اختصاص داده می‌شود، هر ۱۰ B مخصوص یک فریم خواهد بود. در هر فایل ۱۰۰ فریم خواهیم داشت که هر یک از این ۱۰۰ فریم را در یک سطر از فایل txt قرار داده می‌شود. بدین ترتیب در هر سطر ۱۰ خواهیم داشت. جدول (۱) محل قرارگیری پارامترهای کدگذاری (تخصیص بیت) را در ۸۰ bit موجود در هر سطر مشخص می‌کند.

جدول ۱. محل قرارگیری هر پارامتر در رشته بیت ارسالی.

پارامتر	تخصیص بیت
Line spectrum pairs (LSP)	۱ و ۲-۸ و ۹-۱۳ و ۱۴-۱۸
Adaptive-codebook delay (ACD)	۵۶-۵۲ و ۱۹-۲۶
Pitch-delay parity	۲۷
Fixed-codebook index	۵۷-۶۹ و ۲۸-۴۰
Fixed-codebook sign	۷۰-۷۳ و ۴۱-۴۴
Codebook gains	۷۴-۸۰ و ۴۵-۵۱

از این ۸۰ bit سه پارامتر L1 (۲-۸ vbit)، L2 (۹-۱۳ bit) و L3 (۱۴-۱۸ bit) به عنوان ویژگی‌های معرفی‌کننده QIM انتخاب می‌شود. همچنین از پارامترهای ACD تحت عنوان زوج (a_2, a_1) استفاده می‌شود. در حقیقت a_1 و a_2 حالت گذشته تناوب‌های مدار بسته pitch در دو نیم فریم (P_1, P_2) است. a_1 و a_2 از تلفیق بخش صحیح و کسری P_1 و P_2 به دست آمده است. P_1 نشان‌دهنده تناوب نیم فریم اول است و P_2 فاصله تناوب نیم فریم دوم را از تناوب نیم فریم اول مشخص می‌کند. اجزاء صحیح و کسری P_1 و P_2 از پارامترهای a_1 و a_2 استخراج می‌شود و به صورت چهار پارامتر $P_{1,int}, P_{1,frac}, P_{2,int}, P_{2,frac}$ به عنوان ویژگی‌های معرفی‌کننده PMS استفاده می‌شود که در آن

$$T_i = \text{int}(T_i) + \frac{\text{frac}}{3}, \text{frac} \in \{-1, 0, 1\} \quad (5)$$

در کدگذار، اجزاء صحیح و کسری هر یک از دو تناوب T_1 و T_2 با هم تلفیق می‌شود و پارامترهای P_1 و P_2 به دست می‌آید. P_1 در ۸ بیت (۲۵۶ حالت) و P_2 در ۵ بیت (۳۲ حالت) به این صورت کد می‌شود که در روابط (۶) و (۷) بیان شده است:

$$P_1 = \begin{cases} 3(\text{int}(T_1)-19)+\text{frac}-1 & ; \text{if } T_1=[19,\dots,85], \text{frac}=[-1,0,1] \\ (\text{int}(T_1)-85)+197 & ; \text{if } T_1=[86,\dots,143], \text{frac}=0 \end{cases} \quad (6)$$

$$P_2 = 3(\text{int}(T_2) - t_{\min}) + \text{frac} + 2 \\ t_{\min} = \text{int}(T_2) - 5, \text{if } t_{\min} < 20 \text{ then } t_{\min} = 20 \quad (7)$$

در سمت کدگشا، اجزاء صحیح و کسری هر یک از تناوب‌های مدار بسته T_1 و T_2 از دو پارامتر P_1 و P_2 به صورت روابط (۸) و (۹) استخراج می‌شود:

$$\text{if } P_1 < 197 \\ \text{int}(T_1) = \lfloor \frac{(P_1+2)}{3} \rfloor + 19, \text{frac} = P_1 - 3 \text{int}(T_1) + 58 \quad (8)$$

$$\text{else} \\ \text{int}(T_1) = P_1 - 112, \text{frac} = 0$$

$$\text{int}(T_2) = \lfloor \frac{(P_2+2)}{3} \rfloor - 1 + t_{\min}, \text{frac} = P_2 - 2 - 3(\lfloor \frac{(P_2+2)}{3} \rfloor - 1) \quad (9) \\ t_{\min} = \text{int}(T_2) - 5, \text{if } t_{\min} < 20 \text{ then } t_{\min} = 20$$

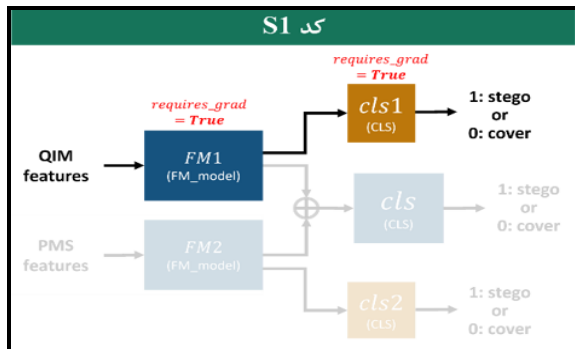
در ادامه مقاله از پارامترهای L1 و L2 و L3 به عنوان نماینده‌های (ویژگی‌ها) نهان‌نگاری QIM و از P_1 و P_2 به عنوان نماینده‌های (ویژگی‌ها) نهان‌نگاری PMS به عنوان ورودی یک شبکه یادگیری عمیق در مرحله آموزش شبکه نهان کاو استفاده می‌شود.

۳-۲. مدل پیشنهادی

مدل پیشنهادی از سه زیرشبکه تعلیم ویژگی، تلفیق ویژگی و طبقه‌بندی تشکیل می‌شود. زیرشبکه اول برای تعلیم ویژگی‌های نهان‌کاوی مرتبط با دو روش QIM و PMS از دادگان ورودی استفاده می‌شود. زیرشبکه دوم، ویژگی‌های به دست آمده از این دو روش را با هم تلفیق می‌کند و به یک ویژگی توأم تبدیل می‌کند. در نهایت در زیرشبکه سوم با استفاده از این ویژگی توأم، احتمال وجود پیام مخفی پیش‌بینی می‌شود. روند تعلیم این شبکه‌ها به این صورت است که ابتدا هر یک از مسیرهای بالا و پایین به صورت مجزا تعلیم می‌بیند (مراحل S1 و S2 در شکل (۲)). برای این منظور در انتهای هر یک از این دو مسیر، یک لایه Linear موقت برای طبقه‌بندی به برچسب 0 و 1 قرار می‌گیرد. در مرحله بعد، وزن‌های تعلیم‌دیده این دو مسیر ثابت می‌ماند و تنها مسیر تلفیق ویژگی و طبقه‌بندی تعلیم می‌بیند (مرحله S3 در شکل (۲)). در نهایت تمام وزن‌های شبکه در فرآیندی به نام Fine tune مجدداً

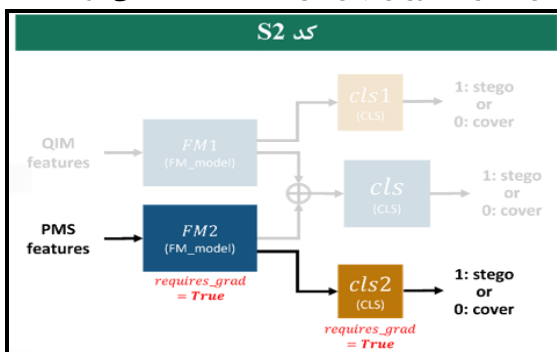
به‌روزرسانی یا شود یا نشود. پس از تعلیم شبکه و در مرحله ارزیابی (validation)، وضعیت `requires_grad` برای تمام زیرشبکه‌ها در وضعیت `False` قرار می‌گیرد. در ادامه مراحل گفته شده را گام به گام توضیح می‌دهیم.

گام ۱ (S1): در این بخش ورودی‌ها یعنی ویژگی‌های سه‌گانه L1، L2 و L3 (مطابق با توضیحات بخش ۲) برای آموزش نهان‌نگاری QIM استفاده می‌شود.



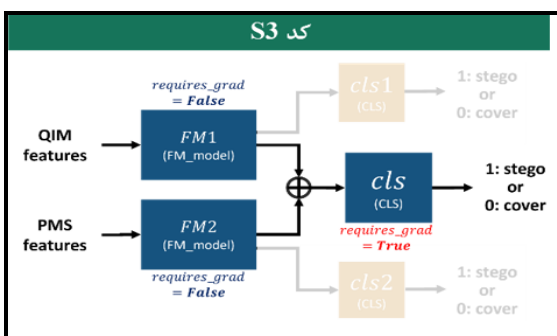
شکل ۴. مسیر اجرای کد S1 در فرآیند آموزش.

گام ۲ (S2): در این بخش ورودی‌ها یعنی ویژگی‌های چهارگانه `P1,int`، `P1,frac`، `P2,int`، `P2,frac` (مطابق با توضیحات بخش ۳) برای آموزش نهان‌نگاری PMS استفاده می‌شود.



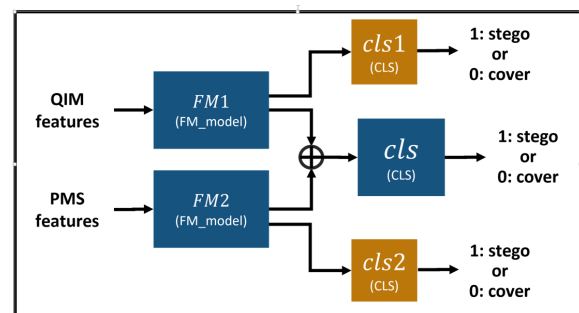
شکل ۵. مسیر اجرای کد S2 در فرآیند آموزش.

گام ۳ (S3): در این بخش وزن‌های دو زیرشبکه S1 و S2 که در مراحل قبل اصلاح شده، ثابت می‌ماند و تنها وزن‌های زیرشبکه CLS در فرآیند تعلیم به‌روزرسانی می‌شود.



شکل ۶. مسیر اجرای کد S3 در فرآیند آموزش.

ما روش پیشنهادی خود را روش سریع دوگانه (Dual-FM)^۱ می‌نامیم. برای پیاده‌سازی شبکه نهان‌کاوی Dual-FM پنج زیرشبکه تعریف می‌شود؛ دو زیرشبکه FM_model با نام‌های FM1 و FM2 برای تعلیم ویژگی‌های QIM و PMS و سه زیرشبکه طبقه‌بندی CLS با نام‌های CLS1، CLS2 و CLS3 برای طبقه‌بندی دو حالت Stego یا Cover بودن که در شکل (۳) نشان داده شده است. از دو زیرشبکه CLS1 و CLS2 صرفاً برای تعلیم مجزای دو زیرشبکه FM1 و FM2 در مراحل اولیه تعلیم استفاده می‌شود.



شکل ۳. زیرشبکه‌های تشکیل‌دهنده شبکه Dual-FM.

کلاس CLS از کلاس `Torch.nn.Module` ارث‌بری می‌کند و زیرشبکه طبقه‌بندی شبکه نهان‌کاوی Dual-FM را تعریف می‌کند. این زیرشبکه از یک لایه `Linear` و یک `Softmax` بعد از آن، تشکیل شده است. همچنین از عملیات `Dropout` در تعلیم این زیرشبکه استفاده شده است. کلاس `FM_model` از یک لایه `Embedding` و دو لایه `Linear` تشکیل می‌شود. لایه `Embedding` هر یک از اعداد طبیعی ورودی را به برداری از اعداد اعشاری تبدیل می‌کند. پس از تعریف این پنج زیرشبکه، مسیر `forward` شبکه در تابع `forward_pred` تعریف می‌شود. در شبکه Dual-FM سه مسیر `forward` قابل تنظیم است که در مراحل مختلف تعلیم و در کدهای پایتون S1 تا S4 از آن استفاده می‌شود. روند تعریف تابع `forward_pred` و تنظیم وضعیت `requires_grad` در این چهار کد، همانند روند کلی شکل (۲) و بر مبنای زیرشبکه‌های تشکیل‌دهنده شبکه پیشنهادی شکل (۳) می‌باشد. در شکل‌های (۴) تا (۷) مسیر `forward_pred` و وضعیت `requires_grad` در هر یک از چهار کد تعلیم مشخص شده است. در چهار کد تعلیم S1 تا S4، وضعیت `requires_grad` برای پارامترهای هر یک از زیرشبکه‌ها تنظیم می‌شود. `requires_grad` تعیین می‌کند که وزن‌های موجود در یک زیرشبکه در حین فرآیند تعلیم،

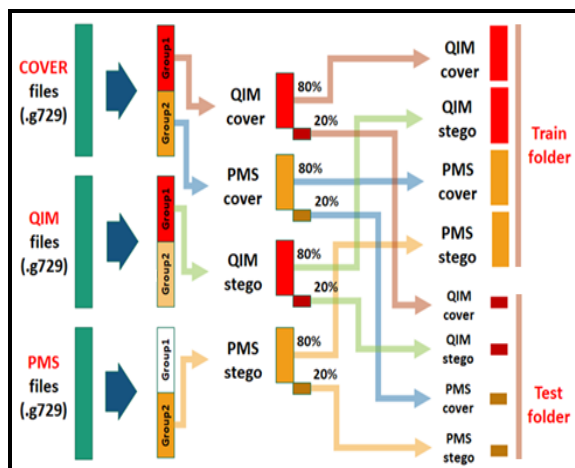
^۱ Dual-Fast Method

۴. نتایج و بحث

دادگان مورد استفاده در این تحقیق از یک دیتاست شامل ۷۲ ساعت صوت انگلیسی برگرفته شده است. این صوت ها با فرمت wav. به صورت فایل های ۳۰ دقیقه ای و به عنوان دادگان خام در دسترس است. این صوت ها برای ورود به شبکه های نهان کاوی شش مرحله زیر را طی می کنند:

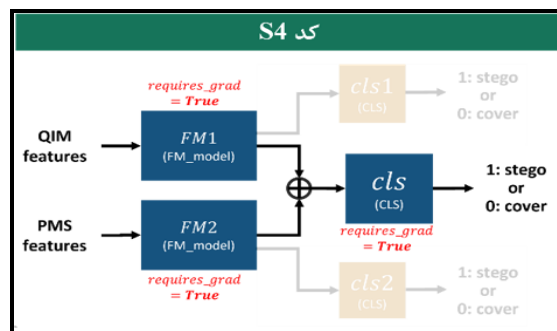
- ۱- شکسته شدن به قطعات صوتی کوچک در حد ثانیه و ذخیره سازی به صورت صوت خالص با فرمت pcm.
- ۲- فشرده سازی با استفاده از نرم افزار Stegocder و تولید فایل های cover و stego (QIM و PMS) با فرمت g729.
- ۳- تقسیم بندی دادگان g729. به دو شاخه تعلیم و تست و تقسیم بندی مجدد این دو شاخه به دو زیرشاخه QIM و PMS
- ۴- پیش پردازش فایل های با فرمت g729. و استخراج ویژگی های مرتبط با دو نهان نگاری QIM و PMS در قالب فایل های متنی با فرمت txt.

همان طور که گفته شد، تولید فایل های stego و cover با استفاده از نرم افزار stegocder [۳۳] انجام می شود. این نرم افزار، فایل های صوتی را با قالب g729. فشرده سازی می کند و حجم فایل ها را با نسبت ۱۶ به ۱ کاهش می دهد. نهان نگاری های PMS و QIM در حین عملیات فشرده سازی انجام می شود. فایل های با قالب g729. به دست آمده از خروجی نرم افزار، هم نام با فایل های pcm. است و اینکه در کدام یک از سه پوشه Cover، PMS یا QIM قرار گرفته است، مدل آن فایل را مشخص می کند. نیمی از فایل های موجود در پوشه Cover را برای نهان نگاری QIM (Group1) و نیم دیگر را برای نهان نگاری PMS (Group2) در نظر می گیریم و فایل های stego معادل را از دو پوشه QIM و PMS انتخاب می کنیم. همچنین ۸۰ درصد کل فایل ها را برای تعلیم شبکه و ۲۰ درصد باقیمانده را برای تست شبکه در نظر می گیریم که در شکل (۹) این تقسیم بندی نشان داده شده است.



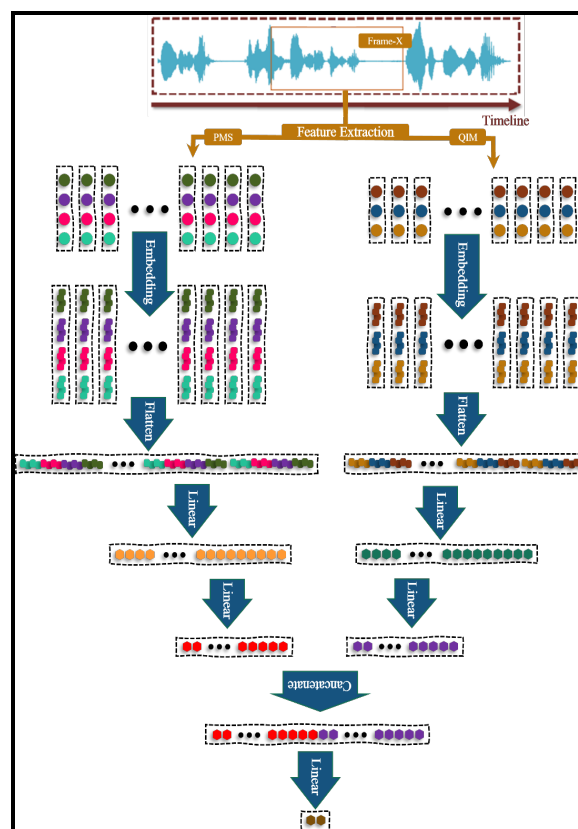
شکل ۹. دسته بندی فایل های g729.

گام ۴ (S4): در این بخش از نظر مسیر forward وضعیتی مشابه کد S3 دارد. در این بخش وزن های همه زیر شبکه ها در فرآیند تعلیم به روز رسانی می شود.



شکل ۷. مسیر اجرای کد S4 در فرآیند آموزش.

باتوجه به توضیحات گفته شده در بخش های قبلی ساختار کلی نهان کاو پیشنهادی در شکل (۸) نشان داده شده است.



شکل ۸. شبکه نهان کاو پیشنهادی.

مطابق با شکل (۸) بردار های به دست آمده در انتهای این دو مسیر حاصل از QIM و PMS در کنار هم قرار می گیرد و یک بردار یکپارچه را تشکیل می دهد. در نهایت طبقه بندی سیگنال صوتی اولیه به وضعیت Cover یا Stego به وسیله یک لایه Linear و تابع Softmax انجام می شود. این شبکه سرعت و دقت بالاتری نسبت به مدل های مرسوم هوش مصنوعی دارد که در ادامه نتایج پیاده سازی آن نشان داده می شود.

(مرحله S4) برای نهان‌نگاری QIM بی‌تأثیر است؛ ولی برای نهان‌نگاری PMS مرحله S3 مفید و در پیاده‌سازی اعمال می‌شود. ارزیابی نتایج برای شبکه نهان‌کاوی پیشنهادی (Dual-FM) نیز به نتایج مشابهی برای مراحل S3 و S4 می‌انجامد. هدف اصلی نهان‌کاوی این است که تعیین کند که آیا داده‌ای که در فضای سایبری مورد استفاده قرار می‌گیرد، دارای اطلاعات پنهانی است یا نه؟ به عبارت دیگر، نهان‌کاوی رسانه مورد بررسی را به دودسته پوششی یا نهان‌نگاشته تقسیم می‌کند. اگر روشی برای نهان‌کاوی انتخاب گردد، چهار حالت ممکن است درباره نتیجه استفاده از این روش به وجود آید:

1- True Positive (TP): به این معنا که یک رسانه نهان‌نگاشته شده به درستی به عنوان یک رسانه نهان‌نگاشته انتخاب شده باشد.

2- False Negative (FN): به این معنا که یک رسانه نهان‌نگاشته شده به صورت نادرست به عنوان یک رسانه پوششی انتخاب شده باشد.

3- True Negative (TN): به این معنا که یک رسانه پوششی به درستی به عنوان یک رسانه پوششی انتخاب شده باشد.

4- False Positive (FP): به این معنا که یک رسانه پوششی به صورت نادرست به عنوان یک رسانه نهان‌نگاشته انتخاب شده باشد.

حال بر اساس توضیحات گفته‌شده، صحت تشخیص طبق رابطه (9) محاسبه می‌گردد:

$$\text{Detection Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (9)$$

مطابق با رابطه (9) مقایسه مقادیر صحت تشخیص برای دو روش فوق برای نهان‌نگاری QIM و PMS در جدول (5) و جدول (6) نشان داده شده است.

جدول 5. مقادیر صحت تشخیص برای نهان‌نگاری QIM برای دو روش SFFN و Dual-FM (روش پیشنهادی).

Rate (%)	Length (ms)	QIM	
		SFFN	Dual-FM
10	1000	60/31	62/64
40	1000	81/69	89/10
70	1000	92/41	97/07
100	1000	97/92	98/93

جدول 6. مقادیر صحت تشخیص برای نهان‌نگاری PMS برای دو روش SFFN و Dual-FM (روش پیشنهادی).

Rate (%)	Length (ms)	PMS	
		SFFN	Dual-FM
10	1000	52/32	53/12
40	1000	76/80	76/63
70	1000	90/67	90/99
100	1000	96/65	97/14

در ادامه شبکه نهان‌کاوی مرسوم و مرجع SFFN و روش پیشنهادی (Dual-FM) در فرآیندی مشابه، در چهار مرحله S1 تا S4 تعلیم داده می‌شوند. مرحله چهارم تعلیم (S4)، برای تعلیم و اصلاح مجدد تمام وزن‌های شبکه است که به این عملیات در اصطلاح fine tune کردن شبکه می‌گویند. این فرآیند می‌تواند موجب بهبود یا افت دقت شبکه نهان‌کاوی شود؛ بنابراین نیاز است تا در مورد انجام این مرحله از تعلیم، تصمیم‌گیری شود. در جدول (2) و جدول (3)، دقت نهان‌کاوی دو شبکه SFFN و Dual-FM در دو حالت نشان داده شده است. در حالت اول از وزن‌های شبکه بعد از مرحله S3 و در حالت دوم از وزن‌های شبکه بعد از مرحله S4، برای فرآیند تست استفاده می‌شود.

جدول 2. نتایج حاصل از پیاده‌سازی مقایسه صحت تشخیص روش SFFN از مرحله تعلیم بعد از مرحله S4 و بعد از مرحله S3.

Rate (%)	Length (ms)	QIM		PMS	
		S3	S4	S3	S4
10	1000	60/31	61/52	52/32	54/57
40	1000	81/69	83/18	76/80	76/05
70	1000	92/41	92/21	90/67	90/25
70	1000	97/92	97/60	96/65	96/44

جدول 3. نتایج حاصل از پیاده‌سازی مقایسه صحت تشخیص Dual-FM (روش پیشنهادی) از مرحله تعلیم بعد از مرحله S4 و بعد از مرحله S3.

Rate (%)	Length (ms)	QIM		PMS	
		S3	S4	S3	S4
10	1000	62/64	67/57	53/12	53/32
40	1000	89/10	89/35	76/63	75/41
70	1000	97/07	96/85	90/99	90/61
100	1000	98/93	98/85	97/14	96/94

در هر دو جدول (2) و جدول (3) نتایج صحت تشخیص سامانه نهان‌نکو برای طول قطعه‌بندی صوتی برای داده‌های ms 1000 و برای نرخ ادغام مختلف به دست آمده است. ارزیابی نتایج دو جدول فوق برای نرخ‌های مختلف نهان‌نگاری نشان‌دهنده برتری نتایج برای نهان‌نگاری روش PMS قبل از اعمال مرحله S4 است. بررسی نتایج برای نهان‌نگاری روش QIM نشان از اثر خنثی در بعد از مرحله S4 است. نمایش عددی این برتری در جدول (4) نشان داده شده است.

جدول 4. تعداد برتری دو حالت بعد از مرحله S4 و بعد از مرحله S3.

Method	S3	S4	S3	S4
SFFN	2	2	3	1
Dual-FM	2	2	3	1

جدول (4) نشان می‌دهد نتایج شبکه نهان‌کاوی SFFN با اعمال مرحله سوم تعلیم (S4)، برای نهان‌نگاری PMS، 3 بار تضعیف در مقابل 1 بار بهبود (تقویت) ولی برای نهان‌نگاری QIM، 2 بار بهبودی (تقویت) در مقابل 2 بار تضعیف به دست آمده است. نتیجه آنکه برای روش SFFN نتایج پس از fine tune کردن

کدک مذکور، ویژگی‌های درون‌فریمی و همبستگی‌های بین‌فریمی را با دقت خوبی استخراج می‌کند. باتوجه‌به مدل ارائه‌شده و فرایند پردازش موازی، سرعت و صحت تشخیص روش پیشنهادی از روش SFFN به‌عنوان روش پایه و مرجع در این حوزه بالاتر می‌باشد. سادگی در مرحله پیاده‌سازی و صحت تشخیص قابل‌قبول حتی برای نرخ ادغام پایین از ویژگی‌های روش پیشنهادی می‌باشد. لازم به ذکر است روش مذکور برای فریم‌های تست ۱ ثانیه سرعت پاسخ‌گویی بسیار بالایی داشته که می‌تواند برای کاربردهای برخط استفاده شود.

۶. مراجع‌ها

- [1] Shamalizadeh Baei, M. A.; Norozi, Z.; Sabzinezhad, M.; Karami, M. R. "Designing an Image Steganography Algorithm Based on Entropy and ELBS2"; Adv. Defence Sci. & Technol. 2018, 02, 39-50 (In Persian).
- [2] Petitcolas, F. A.; Anderson, R. J.; Kuhn, M. G. "Information Hiding-A Survey"; IEEE Int. Joint Conf. Neural Networks 1999, 87(7), 1062-1078.
- [3] Tacticus, A. "How to Survive Under Siege"; Clarendon Press. 1990. DOI:10.1109/5.771065.
- [4] Lemma, A. N.; Aprea, J.; Oomen, W.; Van de Kerkhof, L. "A Temporal Domain Audio Watermarking Technique"; IEEE Trans. Signal Process. 2003, 51, 1088-1097. DOI: 10.1109/TSP.2003.809372.
- [5] Jafari, S. M. Sadnejad, S. R.; Saryazdi, S.; Jamshidi, V. "A New Audio Steganography Algorithm Based on Sample Clustering"; 7th ISCISC'10, 2010 (In Persian). DOI:10.11591/ijce.v12i1.pp320-330.
- [6] ITU, G. "Coding of Speech at 8 kbit/s Using Conjugatestructure Algebraic-Code-Excited Linear-Prediction (CSACELP)"; 1996.
- [7] Chen, B.; Wornell, G. W. "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding"; IEEE Trans. Inform. Theory 2001, 47, 1423-1443. DOI: 10.1109/18.923725.
- [8] Yan, S.; Tang, G.; Chen, Y. "Incorporating Data Hiding into G.729 Speech Codec"; Multimed. Tools Appl. 2016, 75, 11493-11512.
- [9] Ren, Y.; Wu, H.; Wang, L. "An AMR Adaptive Steganography Algorithm Based on Minimizing Distortion"; Multimed. Tools Appl. 2018, 77, 12095-12110. DOI:10.1007/s11042-015-2865-1.
- [10] Huang, Y.; Liu, C.; Tang, S.; Bai, S. "Steganography Integration into a Low-Bit Rate Speech Codec"; IEEE Trans. Inf. Foren. Sec. 2012, 7, 1865-1875. DOI:10.1109/TIFS.2012.2218599.
- [11] Huang, Y.; Tao, H.; Xiao, B.; Chang, C. "Steganography in Low Bit-Rate Speech Streams Based on Quantization Index Modulation Controlled by Keys"; Sci. China Tech. Sci. 2017, 60, 1585-1596. DOI:10.1007/s11431-016-0707-3.
- [12] Liu, P.; Li, S.; Wang, H. "Steganography Integrated into Linear Predictive Coding for Low Bit-Rate Speech Codec"; Multimed. Tools Appl. 2017, 76, 2837-2859. DOI: 10.1007/s11042-016-3257-x.
- [13] Xiao, B.; Huang, Y.; Tang, S. "An Approach to Information Hiding in Low Bit-Rate Speech Stream"; IEEE Glob. Telecomm. Conf. 2008, 1-5. DOI: 10.1109/GLOCOM.2008.ECP.375.

جدول (۷) زمان موردنیاز برای اجرای هر یک از این مراحل و مجموع زمان صرف شده برای تعلیم دو شبکه را نشان می‌دهد. این زمان برای دادگان با طول قطعه‌بندی ۱۰۰۰ ms و نرخ نهان‌نگاری ۱۰۰٪ محاسبه شده است. در محاسبه زمان، چهار مرحله S1 تا S4 لحاظ شده و پارامتر epoch=50 تنظیم شده است.

جدول ۷. زمان صرف شده در مراحل تعلیم (برحسب دقیقه).

Step	SFFN	Dual-FM
S1	۵/۹۲	۴/۱۴
S2	۶/۲۷	۴/۲۹
S3	۳۰/۷۹	۷/۹۸
S3	۳۵/۷۹	۱۱/۴۲
SUM	۷۸/۷۳	۲۷/۸۳

جدول (۸) زمان لازم برای تست یک فایل ۱۰۰۰ ms را برای تشخیص Stego یا Cover بودن آن، در هر یک از دو شبکه نهان‌کاوی SFFN و Dual-FM نشان می‌دهد.

جدول ۸. زمان موردنیاز برای تست یک تک فایل (برحسب میکروثانیه).

SFFN	Dual-FM
۱۸۹۵۰	۴۹۸۵

نتایج دو جدول فوق نشان می‌دهد شبکه SFFN نسبت به شبکه Dual-FM سرعت تعلیم و تست کمتری دارد. دلیل این سرعت کم، زمان‌بر بودن لایه‌های بازگشتی LSTM و کانولوشن CNN در ساختار شبکه SFFN است. درحالی‌که در شبکه Dual-FM محاسبات لازم برای تست یک فایل، به‌صورت موازی و در قالب عملیات ماتریسی انجام می‌گیرد؛ برخلاف شبکه SFFN که در آن عملیات متوالی در زیرشبکه‌های LSTM برقرار است.

۵. نتیجه‌گیری

در این مقاله، یک روش نهان‌کاوی مبتنی بر هوش مصنوعی برای تشخیص دو خانواده معروف نهان‌نگاری QIM و PMS معرفی و تشریح داده شد. پیاده‌سازی انجام شده بر روی داده‌های صوتی مبتنی بر کدک G.729 به‌عنوان یکی از کدک‌های معروف در حوزه VoIP اعمال شده است. باتوجه‌به اینکه در روش‌های نهان‌نگاری مذکور با نمونه‌های فشرده شده سیگنال صحبت مواجه هستیم، تشخیص داده‌های پاک از داده‌های نهان‌نگاشته بخصوص در نرخ ادغام پایین کاری بسیار دشوار می‌باشد؛ بنابراین روش‌های نهان‌کاوی متداول که از روش‌های آماری و پردازش سیگنال استفاده می‌کند، کارایی چندانی نداشته و صحت تشخیص قابل‌قبولی را ارائه نمی‌دهد. از این‌رو روش‌های مبتنی بر مدل‌های هوش مصنوعی و به‌خصوص روش‌های مبتنی بر یادگیری عمیق در سالیان اخیر پیشنهاد شده است. در این مقاله از رویکرد یادگیری عمیق و ترکیب آن با روش‌های پیش‌پردازش داده استفاده شده است. در این تحقیق پیش‌پردازش سیگنال صوتی فشرده‌شده با

- [24] Wu, Z.; Guo, J. "MFPPD-LSTM: A Steganalysis Method Based on Multiple Features of Pitch delay Using RNN-LSTM"; *J. Inf. Sec. App.* 2023, 74, 103469. DOI: 10.1016/j.jisa.2023.103469.
- [25] Hu, Y.; Huang, Y.; Yang, Z.; Huang, Y. "Detection of Heterogeneous Parallel Steganography For Low Bit-Rate Voip Speech Streams"; *Neurocomputing* 2021, 419, 70-79. DOI:10.1016/j.neucom.2020.08.002.
- [26] Li, S.; Wang, J.; Liu, P.; Wei, M.; Yan, Q. "Detection of Multiple Steganography Methods in Compressed Speech Based on Code Element Embedding, Bi-LSTM and CNN with Attention Mechanisms"; *IEEE/ACM TASLAP* 2021, 29, 1556-1569. DOI:10.1109/TASLP.2021.3074752.
- [27] Yang, H.; Yang, Z.; Bao, Y.; Liu, S.; Huang, Y. "FCEM: A Novel Fast Correlation Extract Model for Real Time Steganalysis of VOIP Stream via Multi-head Attention"; In *IEEE ICASSP*. 2020, 2822-2826. DOI:10.1109/ICASSP40776.2020.9054361.
- [28] Yang, H.; Yang, Z.; Huang, Y. "Steganalysis of VoIP Streams with CNN-LSTM Network"; *ACM Workshop on Information Hiding and Multimedia Security* 2019, 204-209. <https://doi.org/10.1145/3335203.3335735>.
- [29] Wang, H.; Yang, Z.; Hu, Y.; Yang, Z.; Huang, Y. "Fast Detection of Heterogeneous Parallel Steganography for Streaming Voice"; *ACM Workshop on Information Hiding and Multimedia Security* 2021, 137-142. DOI: 10.1145/3437880.3460404.
- [30] Yang, H.; Yang, Z.; Bao, Y.; Liu, S.; Huang, Y. "Fast Steganalysis Method for VoIP Streams"; *IEEE Signal Proc. Let.* 2019, 27, 286-290. DOI:10.1109/LSP.2019.2961610.
- [31] "Vector Quantization - Mohamed Qasem" <http://mqasem.net/vector-quantization/> (accessed Oct. 25, 2021).
- [32] ITU, T.S.S.O. "Dual Rate Speech Coder for Multimedia Communication Transmitting at 5.3 and 6.3 kbit/s"; Recommendation g, 723. 1996.
- [33] Lin, Z.; Huang, Y.; Wang, J.; "RNN-SM: Fast Steganalysis of VoIP Streams Using Recurrent Neural Network"; *IEEE Trans. Inf. Foren. Sec.* 2018, 13, 1854-1868. DOI:10.1109/TIFS.2018.2806741.
- [14] Huang, Y. F.; Tang, S.; Yuan, J. "Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec"; *IEEE Trans. Inf. Foren. Sec.* 2011, 6, 296-306. DOI: 10.1109/TIFS.2011.2108649.
- [15] Liu, J.; Zhou, K.; Tian, H. "Least-Significant-Digit Steganography in Low Bitrate Speech"; *IEEE ICC* 2012, 1133-1137.
- [16] Lin, R. S. "An Imperceptible Information Hiding in Encoded Bits of Speech Signal"; *IEEE IHH-MSP* 2015, 37-40.
- [17] Xu, S.; Tian, H.; Quan, H.; Lu, J. "A Novel Global-Local Representations Network for Speech Steganalysis"; *5th Int. Conf. on AI and Pattern Recognition* 2022, 945-949.
- [18] Wang, J.; Yang, J.; Gao, F.; Xu, P. "Steganalysis of Compressed Speech Based on Global and Local Correlation Mining"; *IEEE Access*. 2022, 10, 78472-78483. <https://doi.org/10.1109/ACCESS.2022.3194051>.
- [19] Qiu, Y.; Tian, H.; Tang, L.; Mazurczyk, W.; Chang, C.C. "Steganalysis of Adaptive Multi-Rate Speech Streams with Distributed Representations of Codewords"; *J. Inf. Sec. App.* 2022, 68, 103250. DOI:10.1016/j.jisa.2022.103250.
- [20] Li, S.; Wang, J.; Liu, P. "General Frame-Wise Steganalysis of Compressed Speech Based on Dual-Domain Representation and Intra-Frame Correlation Leaching"; *IEEE/ACM Trans. ASLAB*. 2022, 30, 2025-2035.
- [21] Qiu, Y.; Tian, H.; Li, H.; Chang, C. C.; Vasilakos, A. V. "Separable Convolution Network With Dual-Stream Pyramid Enhanced Strategy for Speech Steganalysis"; *IEEE Trans. Inf. Foren. Sec.* 2023.
- [22] Yang, Z.; Yang, H.; Chang, C. C.; Huang, Y.; Chang, C. C. "Real-time Steganalysis for Streaming Media Based on Multi-channel Convolutional Sliding Windows"; *Knowl-Based Syst.* 2022.
- [23] Ren, Y.; Liu, D.; Liu, C.; Xiong, Q.; Fu, J.; Wang, L. "A Universal Audio Steganalysis Scheme based on Multiscale Spectrograms and DeepResNet"; *IEEE Trans. Depend. Sec.* 2022, 20, 665-679. DOI:10.1109/TDSC.2022.3141121.

