نشریه علمی «علوم و فناوری‌های پدافند نوین»
سال چهاردهم، شماره ۲، تابستان ۱۴۰۲؛ ص ۱۱۱-۱۰۱

علمی - پژوهشی

ارائه یک روش بهبودیافته نهان‌کاوی 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: 20.1001.1.26762935.1402.14.2.2.4

چکیده

امروزه پروتکل انتقال صدا از طریق اینترنت (VoIP) به‌صورت گسترده در ارتباطات بلادرنگ و شبکه‌های اجتماعی مورداستفاده قرار گرفته و به حامل مناسبی برای روش‌های نهان‌نگاری تبدیل‌شده است. در راستای مقابله با این تهدیدات، روش‌های متعدد نهان‌کاوی ابداع شده‌اند که در میان راه‌حل‌های ارائه‌شده، ترکیب روش‌های پردازش سیگنال و یادگیری ماشین، امکان ایجاد نهان کاوهایی بادقت 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 فشرده‌شده با کدک G.729 صورت می‌گیرد که ویژگی‌های درون‌فریمی و همبستگی‌های بین‌فریمی را بادقت خوبی استخراج می‌کند. سپس نتایج به‌دست‌آمده 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 ارائه‌شده 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 حوزهٔ مهم است که در سالیان اخیر پیشرفت‌های زیادی کرده است و با فن‌هایی از قبیل نهان‌نگاری [1] آمیخته‌شده است. نهان‌نگاری، هنر و علم انتقال پیام‌های مخفی به طریقی است که به‌جز گیرندهٔ 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 ازاین‌رو 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 کلمهٔ ادراک، قوهٔ ادراک آدمی (مثل بینایی یا شنوایی) است که نبایستی قادر به تشخیص حضور وجود پیام در سیگنال میزبان شود و منظور از کلمهٔ تحلیل این است که با استفاده از تحلیل‌های آماری نبایستی وجود پیام مخفی تشخیص داده شود. مسئلهٔ پنهان‌کردن اطلاعات و انتقال مخفیانهٔ آن موضوع جدیدی نیست، اما اولین کتاب واقعی در این زمینه توسط اسکات نوشته شد که در آن به چگونگی مخفی‌کردن پیام در بین نت‌های موسیقی اشاره شد. پنهان‌سازی پیام‌های محرمانه در فواصل خالی در حدود یک نقطه، گوشوارهٔ زنان، پاشنه‌های کفش پیک‌های خبررسان و استفاده از جوهرهای نامرئی روش‌های دیگر پنهان‌کردن اطلاعات بوده است. از دههٔ ۸۰ میلادی روش‌های پنهان‌سازی اطلاعات به‌صورت دیجیتال مطرح شدند [۳ و۲].به سیگنالی که ارسال مخفی آن هدف اصلی سیستم است، سیگنال پیام یا دادهٔ جاسازی‌شده گفته می‌شود. به سیگنالی که به‌عنوان بستری برای مخفی‌کردن سیگنال پیام به کار می‌رود، سیگنال پوشش [2] گفته می‌شود. سیگنال میزبانی را که در آن سیگنال پیام جاسازی شده است، سیگنال نهان‌نگاری‌شده [3] می‌گویند. سه اصطلاح مهم که در ارزیابی سیستم‌های نهان‌نگاری به کار می‌روند، ظرفیت [4]، مقاومت [5] و شفافیت [6] 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 دربارهٔ وجود یک پیام مخفی تعریف می‌شود. این اصطلاح برای سنجش این موضوع استفاده می‌شود که پس از جاسازی پیام در سیگنال میزبان، تغییر ایجادشده تا چه حد قابل‌تشخیص 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 حوزهٔ زمان و حوزهٔ تبدیل تقسیم‌بندی می‌شوند. الگوریتم‌های حوزهٔ تبدیل نسبت به الگوریتم‌های حوزهٔ زمان در مقابل حملات مقاوم‌تر می‌باشند، اما مقاومت آن‌ها در سیگنال‌هایی که تعداد مؤلفه‌های حوزهٔ تبدیلشان بسیار کم می‌باشد، رضایت‌بخش نیست [۴]. در ضمن این الگوریتم‌ها نسبت به الگوریتم‌های حوزهٔ 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 [7] گفته می‌شود. تمرکز ما در بحث نهان‌کاوی در این مقاله بر روی سیگنال‌های گفتار مبتنی بر IP یا همان VoIP است. باتوجه‌به 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 درنهایت به‌صورت بسته‌های داده ارسال می‌شوند. در طی مراحل تبدیل آنالوگ به دیجیتال، کدک [8] ها نقش مهمی ایفا کرده و نقش فشرده‌سازی و عملیات عکس آن را انجام می‌دهند. کدک‌های مختلفی تعریف‌شده‌اند 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 [9]) 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 به‌دست‌آمده است. برای هر فریم ۱۰ میلی‌ثانیه‌ای، سیگنال صوتی برای استخراج پارامترهای مدل CELP آنالیز می‌شود که نتیجه آن پارامترهای ضرایب فیلتر پیش‌بینی خطی [10]، نمایه و گین کد بوک‌های ثابت و تطبیقی [11] است. این پارامترها در بخش کد کننده، کد شده و فرستاده می‌شوند. در بخش کدگشایی، از این پارامترها برای به‌دست‌آوردن 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) [12] برای محاسبه ضرایب فیلتر LP، روی هر فریم ۱۰ میلی‌ثانیه‌ای انجام می‌شود. این ضرایب به زوج طیف خطی (LSP) [13] تبدیل می‌شود و با روش کوانتیزاسیون برداری

[1] Steganography
[2] Cover Signal
[3] Stego Signal
[4] Capacity
[5] Robustness
[6] Transparency

[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

(VQ)[1] پیش‌بینی‌کننده دوبخشی با ۱۸ بیت کوانتیزه می‌شود (نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون (QIM) [2] روی این ۱۸ بیت اعمال می‌شود که در بخش ۳-۱ توضیح داده می‌شود). سیگنال تحریک با روند جستجوی آنالیز - با - سنتز (AbS)[3] 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[4]، 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 مداربسته فرکانس گام با جستجو حول تناوب مدار باز فرکانس گام انجام می‌شود. با این آنالیز، تناوب مداربسته 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)[5] 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. ازاین‌رو نهان کاوی متناظر با آن نیز بر روی این دو خانواده متمرکز شده است. در ادامه به رویکردها و روش‌های نهان کاوی 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 [۱۰- ۸]. دستهٔ دوم از آنالیز ضرایب پیش‌بینی خطی و بر اساس مدولاسیون شاخص کوانتیزاسیون برای پنهان‌کردن اطلاعات استفاده می‌کنند [۱۳-۱۱]. دستهٔ سوم اطلاعات را با اصلاح مستقیم مقدار بعضی از کد المنت‌ها در جریان گفتار فشرده جاسازی می‌شود [۱۶-۱۴]. لازم به ذکر است بسیاری از تحلیل‌های نهان کاوی بر اساس همبستگی درون‌فریمی و بین‌فریمی پایه‌ریزی شده است [۲۰-۱۷]. نکته مهم دیگر در فرایند نهان‌کاوی مرحله پیش‌پردازش داده [۲۴-۲۱] به‌عنوان ورودی شبکه جهت فرایند آموزش داده است. یکی از مهم‌ترین و جدیدترین روش‌ها در این حوزه که مبنای مقایسه بسیاری از روش‌ها بوده، روش شبکه نهان کاو ادغام ویژگی (SFFN)[6] است [۲۵]. این روش یکی از مهم‌ترین مدل‌ها برای شناسایی نهان‌نگاری موازی ناهمگن (HPS) [7] 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، مورداستفاده 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) [8] و مکانیزم توجه برای سیگنال‌های صوتی مبتنی بر روش AbS-LPC[9] معرفی شد [۲۶]. در این مقاله روش آشکارسازی نهان‌نگاری با استفاده از کد المنت‌ها، Bi-LSTM[10] و CNN[11] همراه مکانیزم توجه [12] ارائه شده است. ابتدا کد المنت‌های هر فریم به بردارهای مالتی هات[13] تبدیل شده‌اند. سپس هرکدام از این بردارها به یک بردار با اندازهٔ 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)[14] 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 مورداستفاده 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

[1] Vector Quantization
[2] Quantization Index Modulation
[3] Analysis-By-Synthesis
[4] Open-Loop Pitch
[5] Pitch Modulation Steganography
[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

مدل، شبکه عصبی بازگشتی کوتاه‌مدت دو جهتی برای ضبط اطلاعات متنی طولانی‌مدت در حامل‌ها استفاده می‌شود و متعاقباً از CNN برای گرفتن ویژگی‌های محلی و سراسری و همچنین ویژگی‌های حامل زمانی استفاده می‌شود. روش دیگر نهان کاوی سیگنال‌های صوتی VoIP استخراج و ادغام ویژگی (KFEF)[1] بر اساس لایه ادغام و کدگذاری سیگنال و مکانیزم توجه می‌باشد. این روش برای نهان‌نگاری ناهمگن و بر اساس تشخیص نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون و مدولاسیون فرکانس گام کار می‌کند [۲۹]. همچنین روش نگاشت کد المنت برای نهان کاوی سریع و برخط و بر اساس همبستگی‌های موجود در سیگنال داده میزبان، تجزیه‌وتحلیل می‌شود [۳۰]. برای بهره‌برداری بهتر از همبستگی در کلمات کد، کد المنت‌ها به یک فضای معنایی[2] نگاشت می‌شوند. در این روش تنها یک‌لایه پنهان برای استخراج همبستگی بین این کلمات کد استفاده می‌شود. درنهایت، بر اساس ویژگی‌های همبستگی استخراج شده، از یک طبقه‌بند برای دسته‌بندی حامل‌های جریان ورودی استفاده می‌کند. برای افزایش کارایی این مدل پیشنهادی از یک چارچوب ساده تقطیر دانش[3] در آموزش نیز استفاده می‌شود.

۲. روش تحقیق

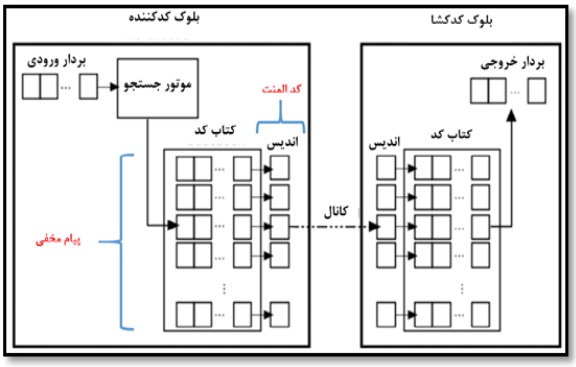
در ادامه این مقاله در بخش‌های بعدی باتوجه‌به اینکه پایه نهان‌کاوی ما بر پایه دو نهان‌نگاری QIM و PMS می‌باشد، به توضیح و تحلیل آن می‌پردازیم. در بخش ۳-۱ به تحلیل نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون، در بخش ۳-۲ به تحلیل نهان‌نگاری مدولاسیون فرکانس گام و در بخش ۳-۳ به معرفی مدل پیشنهادی خود مبتنی بر هوش مصنوعی برای طراحی یک سامانه نهان کاو می‌پردازیم.

۲-۱. تحلیل نهان‌نگاری مدولاسیون شاخص کوانتیزاسیون

نهان‌نگاری کوانتیزاسیون ضرایب LSP در کدگذاری QIM روی G.729 اعمال می‌شود. 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 می‌کنیم. در نهایت نحوه نهان‌نگاری روی این کوانتیزاسیون را در الگوریتم QIM بررسی می‌کنیم. کوانتیزاسیون برداری روشی است که برای فشرده‌سازی سیگنال‌های صوتی در کدکننده‌های 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 می‌شود. این مجموعه محدود به‌صورت $Y=\{y_i: i=1,2,\ldots,N\}$ است. هر بردار y_i به‌عنوان یک بردارکد یا کد واژه[4] و مجموعه این بردارها (Y) به‌عنوان کدکتاب[5] شناخته می‌شود. فضای RK را می‌توان به نواحی مجزا به نام V_i افراز کرد، به صورتی که هر ناحیه به یک کد واژه مرتبط است و به‌عنوان همسایگی آن در نظر گرفته می‌شود. این نواحی به‌صورت رابطه (۱) تعریف می‌شود:

$$V_i = \{x \in R^k : \|x - y_i\| \le \|x - y_j\|, \text{for all } j \neq i\}$$
$$\bigcup_{i=1}^{N} V_i = R^k \quad , \quad \bigcap_{i=1}^{N} V_i = \phi \quad , \quad \text{for all } j \neq i \qquad (1)$$

در VQ بردارهای موجود در هر ناحیه از فضا، روی کد واژه‌های مرتبط با آن نواحی تصویر می‌شود. به‌عبارت‌دیگر، کد واژه درون یک ناحیه به‌عنوان نماینده هر یک از بردارهای موجود در آن ناحیه در نظر گرفته می‌شود [۳۱]. کد واژه‌ای که نماینده یک بردار است، از باقی کد واژه‌ها به آن بردار نزدیک‌تر است. در عمل برای تصویرکردن یک بردار روی یک کد واژه، یک متریک مشخص (مثلاً فاصله اقلیدسی) آن بردار با هر یک از کد واژه‌ها محاسبه می‌شود و کد واژه‌ای که کمترین فاصله را با آن بردار داشته باشد به‌عنوان نماینده آن بردار انتخاب می‌شود. کدگذاری یک بردار به‌وسیله VQ به دو بخش کدگذار و کدگشا تقسیم می‌شود که در شکل (۱) نشان‌داده‌شده است. در کدگذار (بلوک کدکننده) بردار ورودی روی یکی از کد واژه‌ها (نزدیک‌ترین کد واژه) تصویر می‌شود و نمایه آن کد واژه در فرستنده ارسال می‌شود. در بخش کدگشا (بلوک کدگشا)، کد واژه مرتبط با این نمایه انتخاب می‌شود و به‌عنوان برداری خروجی نمایان می‌شود. بدین ترتیب در VQ به‌جای ارسال یک بردار ورودی k بعدی، تنها یک نمایه یک‌بعدی از شبیه‌ترین کد واژه فرستاده می‌شود. هر چه کد کتاب جامع‌تر باشد، دقت کدگذاری بالاتر و سرعت آن کمتر می‌شود، چراکه هر بردار ورودی با تمام کد واژه‌ها مقایسه می‌شود. از طرفی بر اساس اینکه پیام مخفی چه باشد، یک دسته‌بندی روی کتاب کد صورت‌گرفته و نمایه (کد المنت) مربوطه جهت ارسال در کانال انتقال انتخاب می‌گردد.



شکل ۱. کدگذاری یک بردار (فشرده‌سازی) در روش VQ.

[1] Key Feature Extraction and Fusion Network
[2] Semantic Space
[3] Knowledge Distillation
[4] Codeword
[5] Codebook

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. ضرایبLP به ضرایب LSP (q_1 تا q_{10}) تبدیل می‌شود و ضرایب LSP به‌صورت ضرایب LSF[1] (w_1 تا w_{10}) بازنویسی می‌شود ($q_i = \cos(w_{10})$). درنهایت 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 دوحالته MA[2] است و یکی از حالت‌های آن را که برای این فریم بهینه‌تر است، انتخاب می‌کند. L1، L2 و L3 مربوط به سه اندیس کد واژه از سه کدکتاب مورداستفاده 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) به‌صورت 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.

$$\hat{l}_i = \begin{cases} \mathbf{L1}_i(\mathrm{L1}) + \mathbf{L2}_i(\mathrm{L2}) & i=1,\ldots,5 \\ \mathbf{L1}_i(\mathrm{L1}) + \mathbf{L3}_{i-5}(\mathrm{L3}) & i=6,\ldots,10 \end{cases} \qquad (۲)$$

ضرایب LSF با استفاده از رابطه (۳) کدگشایی می‌شود. در این رابطه $\hat{p}_{i,k}$ ضرایب پیش‌بینی‌کننده دوحالته MA است و به ازای $i = 1,2,\ldots,10$ و $k = 1,2,3,4$ ، چهار 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.

$$\hat{\omega}_i^{(m)} = \left(1 - \sum_{k=1}^{4} \hat{p}_{i,k}\right) \hat{l}_i^{(m)} + \sum_{k=1}^{4} \hat{p}_{i,k}\ \hat{l}_i^{(m-k)} \quad i = 1,2,\ldots 10 \qquad (۳)$$

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 نهان‌نگاری QIM برای پنهان کردن 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 نهان‌نگاری 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 نهان‌نگاری مدولاسیون فرکانس گام

در کدگذار G.729 هر ۱۰میلی‌ثانیه معادل ۸۰ نمونه است و یک فریم را تشکیل می‌دهد ($\{s(n)\}_{n=1,2,\ldots 80}$). هر فریم به ۲ زیرفریم ۴۰ نمونه‌ای تقسیم می‌شود. برای پیش‌بینی دوره تناوب pitch ابتدا سیگنال اولیه s(n) به سیگنال وزن‌دار sw(n) تبدیل می‌شود. سپس دوره تناوب pitch برای کل فریم به‌صورت مدار باز با استفاده از مقادیر sw(n) تخمین زده می‌شود. دوره تناوب pitch (T_{op})، در محدوده ۲۰ تا ۱۴۳ نمونه جستجو می‌شود. T_{op} با استفاده از رابطه همبستگی متقابل[3] طبق رابطه (۴) استخراج می‌شود [۳۲]:

$$\mathrm{R(k)} = \frac{\sum_{n=1}^{80} sw(n)sw(n-k)}{\sqrt{\sum_{n=1}^{80} sw^2(n-k)}}, 20 \le k \le 143 \qquad (۴)$$

پارامتر R(k) به ازای مقادیر مختلف k محاسبه می‌شود و آن مقدار k که R(k) بیشتری بدهد به‌عنوان دوره تناوب مدار باز pitch (T_{op}) در نظر گرفته می‌شود. پس از محاسبه T_{op}، پیش‌بینی دوره تناوب pitch به‌صورت مداربسته برای هر زیرفریم انجام می‌شود. برای دو زیر فریم، تناوب مداربسته 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 $[\mathrm{int}(T_1) - 5\frac{2}{3}, \mathrm{int}(T_1) + 4\frac{2}{3}]$ جستجو می‌شود. انتخاب تناوب‌های مداربسته T_i حول مقادیر مرجع به‌گونه‌ای است که سیگنال اولیه با سیگنال بازسازی‌شده 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:

[1] Line Spectral Frequency (LSF)
[2] Switched MA Predictor (MA : moving average)
[3] Cross-correlation

$$T_i = \mathrm{int}(T_i) + \frac{frac}{3},\ frac \in \{-1,0,1\} \quad (۵)$$

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:

$$P_i = \begin{cases} 3(\mathrm{int}(T1)-19)+frac\text{-}1 & ;\ \text{if } T_1=[19,\ldots,85],\ frac=[-1,0,1] \\ (\mathrm{int}(T1)-85)+197 & ;\ \text{if } T_1=[86,\ldots,143],\ frac=0 \end{cases} \quad (۶)$$

$$P_2 = 3(\mathrm{int}(T_2)-t_{\min})+frac+2$$
$$t_{\min} = \mathrm{int}(T_1)-5,\ if\ t_{\min}<20\ then\ t_{\min}=20 \quad (۷)$$

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

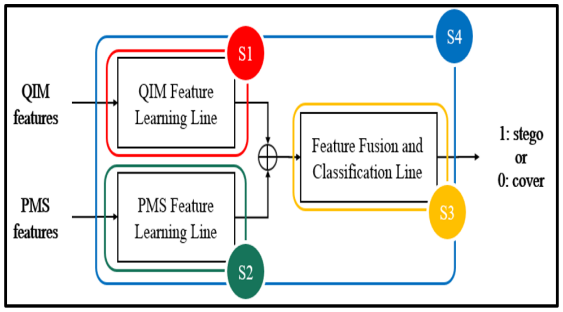
$$\text{if } P_1<197$$
$$\mathrm{int}(T_1)=[\frac{(P_1+2)}{3}]+19,\ frac=P_1-3\,\mathrm{int}(T_1)+58 \quad (۸)$$
$$\text{else}$$
$$\mathrm{int}(T_1)=P_1-112,\ frac=0$$

$$\mathrm{int}(T_2)=[\frac{(P_2+2)}{3}]-1+t_{\min},\ frac=P_2-2-3([\frac{(P_2+2)}{3}]-1) \quad (۹)$$
$$t_{\min}=\mathrm{int}(T_2)-5,\ if\ t_{\min}<20\ \ then\ \ t_{\min}=20$$

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

۲-۳. 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 به‌دست‌آمده 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 به‌صورت مجزا تعلیم می‌بینند (مراحل S1 و S2 در شکل (۲)). برای این منظور در انتهای هر یک از این دو مسیر، یک‌لایه 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 می‌بینند (مرحله S3 در شکل (۲)). در نهایت تمام وزن‌های شبکه در فرآیندی به نام Fine tune مجدداً تعلیم می‌بینند 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 مورداستفاده در این تحقیق یک‌ثانیه‌ای با استاندارد G.729 است که یک فایل با فرمت g729. به‌صورت متنی قابل‌خواندن است و kB ۱حجم دارد. با باز کردن این فایل به صورت متنی، B ۱۰۰۰ در اختیار خواهیم داشت. ازآنجاکه در استاندارد G.729، به هر فریم ms ۱۰ ای bit ۸۰ اختصاص داده می‌شود، هر B ۱۰ 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 B ۱۰ خواهیم داشت. جدول (۱) محل قرارگیری پارامترهای کدگذاری (تخصیص بیت) را درbit ۸۰ موجود در هر سطر مشخص می‌کند.

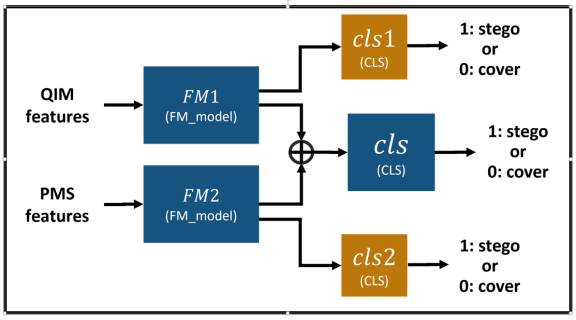
جدول ۱. محل قرارگیری هر پارامتر در رشته بیت ارسالی.

پارامتر	تخصیص بیت
Line spectrum pairs (LSP)	۱ و ۸-۲ و ۱۳-۹ و۱۸-۱۴
Adaptive –codebook delay (ACD)	۲۶-۱۹ و۵۲-۵۶
Pitch-delay parity	۲۷
Fixed-codebook index	۴۰-۲۸ و۶۹-۵۷
Fixed-codebook sign	۴۴-۴۱ و ۷۳-۷۰
Codebook gains	۵۱-۴۵ و۸۰-۷۴

از این bit ۸۰ سه پارامتر L1 (۸-۲ :۷bit)، L2 (۱۳-۹ bit) و L3 (۱۸-۱۴ :۵bit) به‌عنوان ویژگی‌های معرفی‌کننده QIM انتخاب می‌شود. همچنین از پارامترهای ACD تحت عنوان زوج (a_2, a_1) استفاده می‌شود. درحقیقت a_1 و a_2 حالت کدشده تناوب‌های مداربسته pitch در دو نیم فریم (P_1, P_2) است. a_1 و a_2 از تلفیق بخش صحیح و کسری P_1 وP_2 به‌دست 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 به‌صورت چهار پارامتر $P_{1,int}$، $P_{2,int}$، $P_{1,frac}$ و $P_{2,frac}$ به‌عنوان ویژگی‌های معرفی‌کننده PMS استفاده می‌شود که در آن

$P_{i,frac} \in \{-1,0,1\}$, $P_i = P_{i,int} + \frac{P_{i,frac}}{3}$ این ارتباط را نشان می‌دهد.

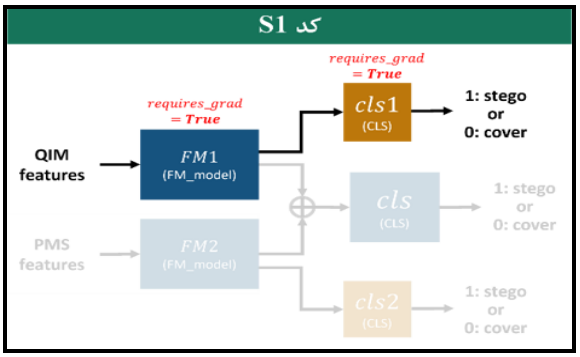
ما روش پیشنهادی خود را روش سریع دوگانه (Dual-FM)[1] می‌نامیم. برای پیاده‌سازی شبکه نهان‌کاوی Dual-FM پنج زیرشبکه تعریف می‌شود؛ دو زیرشبکه FM_model با نام‌های *FM1* و *FM2* برای تعلیم ویژگی‌های QIM و PMS و سه زیرشبکه طبقه‌بندی *CLS* با نام‌های *CLS*1 ,*CLS*2 و *CLS*3 برای طبقه‌بندی دو حالت Stego یا Cover بودن که در شکل (۳) نشان داده شده است. از دو زیرشبکه *CLS*1 و *CLS*2 صرفاً برای تعلیم مجزای دو زیرشبکه *FM*1 و *FM*2 در مراحل اولیه تعلیم استفاده می‌شود.



شکل ۳. زیرشبکه‌های تشکیل‌دهنده شبکه Dual-FM.

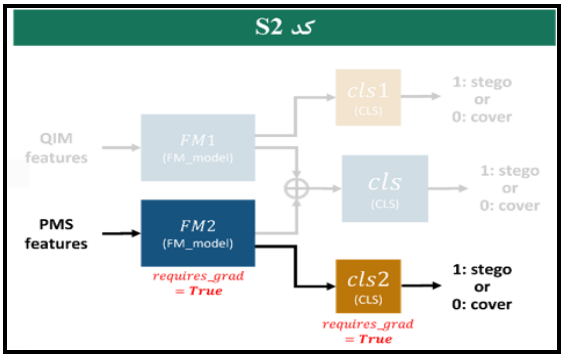
کلاس CLS از کلاس Torch.nn.Module ارث‌بری می‌کند و زیرشبکه طبقه‌بندی شبکه نهان کاوی Dual-FM را تعریف می‌کند. این زیرشبکه از یک‌لایه Linear و یک Softmax بعد از آن، تشکیل شده است. همچنین از عملیات Dropout در تعلیم این زیرشبکه استفاده شده است. کلاس FM_model از یک‌لایه Embedding و دولایه Linear تشکیل می‌شود. 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 قابل‌تنظیم 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 required_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 required_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، به‌روزرسانی بشود یا نشود. پس از تعلیم شبکه و در مرحله ارزیابی (validation)، وضعیت requies_grad برای تمام زیرشبکه‌ها در وضعیت False قرار می‌گیرد. در ادامه مراحل گفته شده را گام‌به‌گام توضیح می‌دهیم.

گام۱ (S1): در این بخش ورودی‌ها یعنی ویژگی‌های سه‌گانه L1، L2 و L3 (مطابق با توضیحات بخش ۲) برای آموزش نهان‌نگاری QIM استفاده می‌شود.



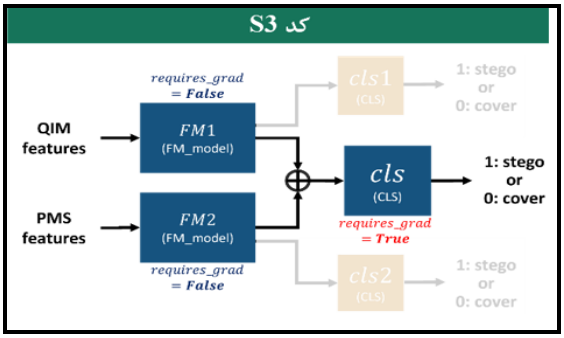
شکل ۴. مسیر اجرای کد S1 در فرآیند آموزش.

گام۲ (S2): در این بخش ورودی‌ها یعنی ویژگی‌های چهارگانه P1,int ، P2,int، P1,frac وP2, frac (مطابق با توضیحات بخش ۳) برای آموزش نهان‌نگاری PMS استفاده می‌شود.



شکل ۵. مسیر اجرای کد S2 در فرآیند آموزش.

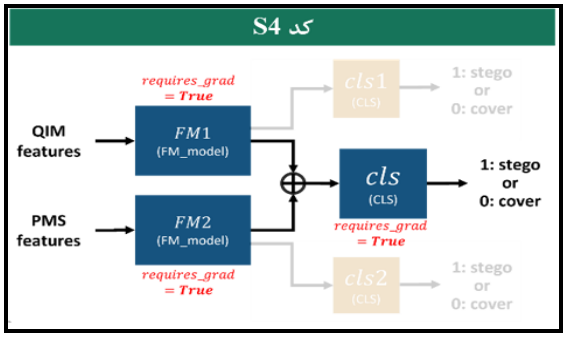
گام۳ (S3): در این بخش وزن‌های دو زیرشبکه S1 و S2 که در مراحل قبل اصلاح شده، ثابت می‌ماند و تنها وزن‌های زیرشبکه CLS در فرآیند تعلیم به‌روزرسانی می‌شود.



شکل ۶. مسیر اجرای کد S3 در فرآیند آموزش.

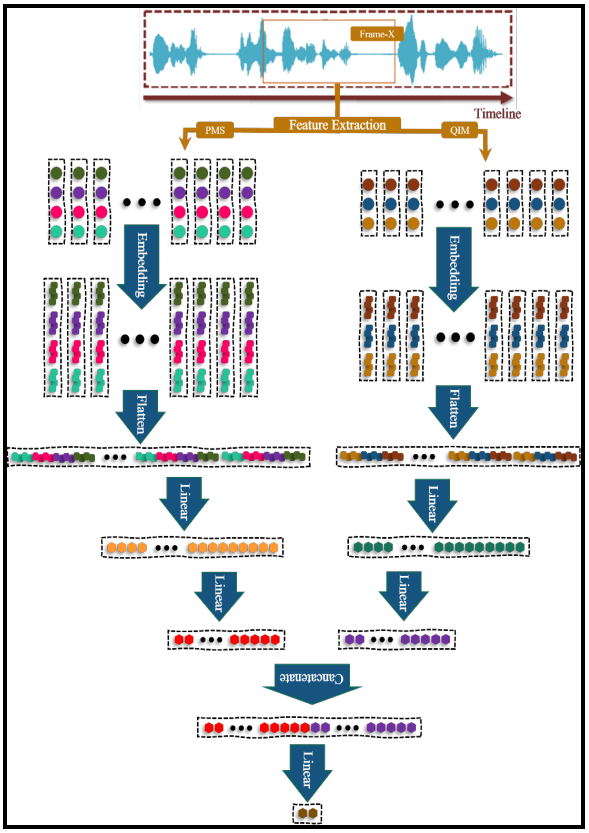
[1] Dual-Fast Method

گام۴ (S4): در این بخش از نظر مسیر forward وضعیتی مشابه کد S3 دارد. در این بخش وزن‌های همهٔ زیرشبکه‌ها در فرآیند تعلیم به‌روزرسانی می‌شود.



شکل ۷. مسیر اجرای کد S4 در فرآیند آموزش.

باتوجه‌به توضیحات گفته‌شده در بخش‌های قبلی ساختار کلی نهان کاو پیشنهادی در شکل (۸) نشان‌داده‌شده است.



شکل ۸. شبکه نهان‌کاو پیشنهادی.

مطابق با شکل (۸) بردارهای به‌دست‌آمده در انتهای این دو مسیر حاصل از QIM و PMS در کنار هم قرار می‌گیرد و یک بردار یکپارچه را تشکیل می‌دهد. در نهایت طبقه‌بندی سیگنال صوتی اولیه به وضعیت Cover یا Stego به‌وسیله یک‌لایه Linear و تابع Softmax انجام می‌شود. این شبکه سرعت و دقت بالاتری نسبت به مدل‌های مرسوم هوش مصنوعی دارد که در ادامه نتایج پیاده‌سازی آن نشان داده می‌شود.

۴. نتایج و بحث

دادگان مورداستفاده در این تحقیق از یک دیتاست شامل ۷۲ ساعت صوت انگلیسی برگرفته شده است. این صوت‌ها با فرمت wav. به‌صورت فایل‌های ۳۰ دقیقه‌ای و به‌عنوان دادگان خام در دسترس است. این صوت‌ها برای ورود به شبکه‌های نهان کاوی شش مرحله زیر را طی می‌کنند:

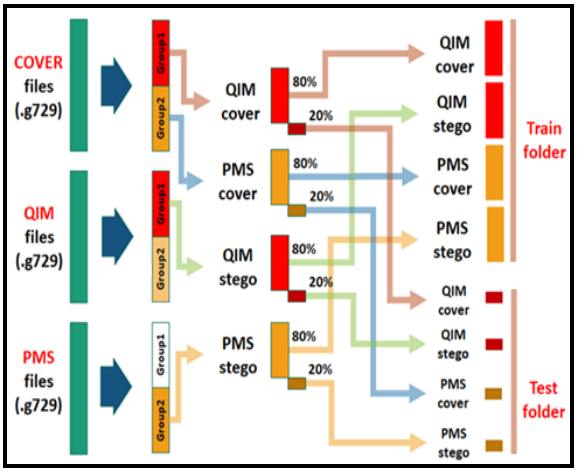
۱- شکسته شدن به قطعات صوتی کوچک در حد ثانیه و ذخیره‌سازی به‌صورت صوت خالص با فرمت pcm.

۲- فشرده‌سازی با استفاده از نرم‌افزار Stegcoder و تولید فایل‌های cover و stego (QIM و PMS) با فرمت g729.

۳- تقسیم‌بندی دادگان g729. به دو شاخه تعلیم و تست و تقسیم‌بندی مجدد این دو شاخه به دو زیرشاخه QIM و PMS

۴- پیش‌پردازش فایل‌های با فرمت g729. و استخراج ویژگی‌های مرتبط با دو نهان‌نگاری QIM و PMS در قالب فایل‌های متنی با فرمت txt.

همان‌طور که گفته شد، تولید فایل‌های stego و cover با استفاده از نرم‌افزار stegcoder [۳۳] انجام می‌شود. این نرم‌افزار، فایل‌های صوتی را با قالب g729. فشرده‌سازی می‌کند و حجم فایل‌ها را با نسبت ۱۶ به ۱ کاهش می‌دهد. نهان‌نگاری‌های PMS و QIM در حین عملیات فشرده‌سازی انجام می‌شود. فایل‌های با قالب g729. به‌دست‌آمده از خروجی نرم‌افزار، هم‌نام با فایل‌های pcm. است و اینکه در کدام‌یک از سه پوشه Cover، PMS یا QIM قرار گرفته است، مدل آن فایل را مشخص می‌کند. نیمی از فایل‌های موجود در پوشه Cover را برای نهان‌نگاری QIM (Group1) و نیم دیگر را برای نهان‌نگاری PMS (Group2) در نظر می‌گیریم و فایل‌های stego معادل را از دوپوشه QIM و PMS انتخاب می‌کنیم. همچنین ۸۰ درصد کل فایل‌ها را برای تعلیم شبکه و ۲۰ درصد باقیمانده را برای تست شبکه در نظر می‌گیریم که در شکل (۹) این تقسیم‌بندی نشان‌داده‌شده است.



شکل ۹. دسته‌بندی فایل‌های g729 .

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 (۲) و جدول (۳)، دقت نهان‌کاوی دو شبکه SFFN و Dual-FM در دو حالت نشان‌داده‌شده 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 ۲. 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
۱۰	۱۰۰۰	۶۰/۳۱	۶۱/۵۲	۵۲/۳۲	۵۴/۵۷
۴۰	۱۰۰۰	۸۱/۶۹	۸۳/۱۸	۷۶/۸۰	۷۶/۰۵
۷۰	۱۰۰۰	۹۳/۴۱	۹۲/۲۱	۹۰/۶۷	۹۰/۲۵
۷۰	۱۰۰۰	۹۷/۹۲	۹۷/۶۰	۹۶/۶۵	۹۶/۴۴

جدول ۳. 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
۱۰	۱۰۰۰	۶۲/۶۴	۶۷/۵۷	۵۳/۱۲	۵۳/۳۲
۴۰	۱۰۰۰	۸۹/۱۰	۸۹/۳۵	۷۶/۶۳	۷۵/۴۱
۷۰	۱۰۰۰	۹۷/۰۷	۹۶/۸۵	۹۰/۹۹	۹۰/۶۱
۱۰۰	۱۰۰۰	۹۸/۹۳	۹۸/۸۵	۹۷/۱۴	۹۶/۹۴

در هر دو جدول (۲) و جدول (۳) نتایج صحت تشخیص سامانه نهان نکاو برای طول قطعه‌بندی صوتی برای داده‌های ms ۱۰۰۰ و برای نرخ ادغام مختلف به‌دست 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 (۴) نشان داده شده است.

جدول ۴. تعداد برتری دو حالت بعد از مرحله S4 و بعد از مرحله S3.

Method	S3	S4	S3	S4
SFFN	۲	۲	۳	۱
Dual-FM	۲	۲	۳	۱

جدول (۴) نشان می‌دهد نتایج شبکه نهان‌کاوی SFFN با اعمال مرحله سوم تعلیم (S4)، برای نهان‌نگاری PMS، ۳ بار تضعیف در مقابل ۱ بار بهبود (تقویت) ولی برای نهان‌نگاری QIM، ۲ بار بهبودی (تقویت) در مقابل ۲ بار تضعیف به‌دست‌آمده است. نتیجه آنکه برای روش SFFN نتایج پس از fine tune کردن (مرحله S4) برای نهان‌نگاری QIM بی‌تأثیر است؛ ولی برای نهان‌نگاری PMS مرحله S3 مفید و در پیاده‌سازی اعمال می‌شود. ارزیابی نتایج برای شبکه نهان کاو 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 مورداستفاده قرار می‌گیرد، دارای اطلاعات پنهانی است یا نه؟ به‌عبارت‌دیگر، 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 به‌وجود آید:

۱- True Positive (TP): به این معنا که یک رسانه نهان‌نگاشته شده به‌درستی به‌عنوان یک رسانه نهان‌نگاشته انتخاب شده باشد.

۲- False Negative (FN): به این معنا که یک رسانه نهان‌نگاشته شده به‌صورت نادرست به‌عنوان یک رسانه پوششی انتخاب شده باشد.

۳- True Negative (TN): به این معنا که یک رسانه پوششی به‌درستی به‌عنوان یک رسانه پوششی انتخاب شده باشد.

۴- False Positive (FP): به این معنا که یک رسانه پوششی به‌صورت نادرست به‌عنوان یک رسانه نهان‌نگاشته انتخاب شده باشد.

حال بر اساس توضیحات گفته‌شده، صحت تشخیص طبق رابطه (۹) محاسبه می‌گردد:

$$\text{Detection Accuracy}=\frac{TP+TN}{TP+TN+FN+FP} \quad (۹)$$

مطابق با رابطه (۹) مقایسه مقادیر صحت تشخیص برای دو روش فوق برای نهان‌نگاری QIM و PMS در جدول (۵) و جدول (۶) نشان‌داده‌شده است.

جدول ۵. مقادیر صحت تشخیص برای نهان‌نگاری QIM برای دو روش SFFN و Dual-FM (روش پیشنهادی).

Rate (%)	Length (ms)	QIM	
		SFFN	Dual-FM
۱۰	۱۰۰۰	۶۰/۳۱	۶۲/۶۴
۴۰	۱۰۰۰	۸۱/۶۹	۸۹/۱۰
۷۰	۱۰۰۰	۹۳/۴۱	۹۷/۰۷
۱۰۰	۱۰۰۰	۹۷/۹۲	۹۸/۹۳

جدول ۶. مقادیر صحت تشخیص برای نهان‌نگاری PMS برای دو روش SFFN و Dual-FM (روش پیشنهادی).

Rate (%)	Length (ms)	PMS	
		SFFN	Dual-FM
۱۰	۱۰۰۰	۵۲/۳۲	۵۳/۱۲
۴۰	۱۰۰۰	۷۶/۸۰	۷۶/۶۳
۷۰	۱۰۰۰	۹۰/۶۷	۹۰/۹۹
۱۰۰	۱۰۰۰	۹۶/۶۵	۹۷/۱۴

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 ms ۱۰۰۰ و نرخ نهان‌نگاری ٪۱۰۰محاسبه 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 ms ۱۰۰۰ 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 نهان کاوی متداول که از روش‌های آماری و پردازش سیگنال استفاده می‌کند، کارایی چندانی نداشته و صحت تشخیص قابل‌قبولی را ارائه نمی‌دهد. ازاین‌رو 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 کدک مذکور، ویژگی‌های درون‌فریمی و همبستگی‌های بین‌فریمی را بادقت 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. مرجع‌ها

[1] Shamalizadeh Baei, M. A.; Norozi, Z.; Sabzinezhad, M.; Karami, M. R. "Designing an Image Steganography Algorithm Based on Entropy and ELSB2"; 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. https://doi.org/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. https://doi.org/ 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). https://doi.org/10.11591/ijece.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. https://doi.org/ 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/ 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. https://doi.org/ 10.1109/ GLOCOM.2008.ECP.375.

[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. https://doi.org/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 IIH-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. https://doi.org/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. https://doi.org/10.1109/TDSC.2022. 3141121.

[24] Wu, Z.; Guo, J. “MFPD-LSTM: A Steganalysis Method Based on Multiple Features of Pitch delay Using RNN-LSTM”; J. Inf. Sec. App. 2023, 74, 103469. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1109/TIFS.2018.2806741.