پنهان سازی داده در سیگنال صحبت فشرده شده با آشکار سازی کور

رضا صالح،حسین خوشبین قماش

دانشگاه فردوسی مشهد - دانشکده مهندسی - گروه برق

Khoshbin@ferdowsi.um.ac.ir Reza_saleh59@yahoo.com

کلید کلیدی - پنهان سازی داده ، فشرده سازی ، تکنیک طیف گسترده، کیفیت شنیداری و آشکار سازی کور

چکیده- در این مقاله روشی برای پنهان سازی‌داده در سیگنال صحبت که عملیات فشرده سازی شکل موجی[1] برروی آن صورت گرفته و به صورت کور آشکار سازی[2] می شود ، ارائه شده است. ویژگی مهمی که این روش دارد آن است که گیرنده برای آشکار سازی به سیگنال اصلی نیازی ندارد ، بنابراین سیگنال صحبت به یک عبارت خاص مقید نبوده و می تواند تغییر کند. نتایج نشان می دهد که در این روش با حفظ کیفیت شنیداری[3]، نرخ خطا حتی در نسبت سیگنال به نویز های پایین حداقل می شود.

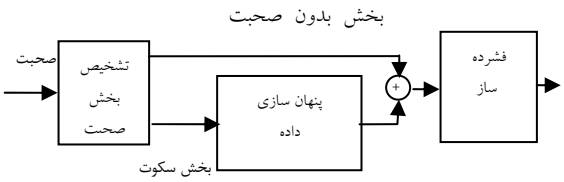
۱- مقدمه

یکی از شیوه های متداول برای جلوگیری از دسترسی افراد غیر مجاز به اطلاعات خاص، رمزنگاری می باشد. در این روش ابتدا اطلاعات به داده های دیگر نگاشت داده شده و سپس ارسال می گردد. گیرنده مجاز با داشتن الگوریتم رمزگشایی‌و کلید مربوطه ، داده ها را به شکل اصلی بر - می گرداند. به دلیل تصادفی بودن داده های رمز شده دشمن می تواند با پی بردن به اهمیت آن درصدد کشف رمز و تخریب آن بر آید که این یکی از مشکلات عمده این روش است. برای‌رفع این‌مشکل می توان داده مورد نظر را در سیگنالی دیگر(سیگنال میزبان) مانند تصویر ، صوت و یا صحبت پنهان نمود. این عمل باید به گونه ای صورت پذیرد که تغییر قابل توجهی در سیگنال میزبان ایجاد نکند. . با توجه به گستردگی استفاده از سیگنال صحبت در کاربردهایی از قبیل تلفن ، موبایل و بی سیم می توان از این بستر مناسب و در دسترس برای انتقال داده های اضافی استفاده کرد. از آنجایی که دراغلب کاربردها عملیات فشرده سازی بر روی سیگنال صحبت انجام می شود، الگوریتم پنهان سازی داده باید در مقابل فشرده سازی مقاومت خوبی داشته باشد.

اغلب کارهایی که‌تاکنون‌انجام شده است بر روی سیگنال های تصویر و صوت بوده است و کار زیادی بر روی صحبت ، بویژه با در نظر گرفتن اثر فشرده سازی‌صورت نگرفته است [۱]. در تحقیقی که در مقاله [۲] انجام شده است، اطلاعات

1 - Waveform compression
2 - Blind detection
3 - Perceptual quality

باینری در سیگنال صحبتی که فرستنده و گیرنده از آن مطلع هستند، پنهان گردیده و با استفاده از یک روش خاص



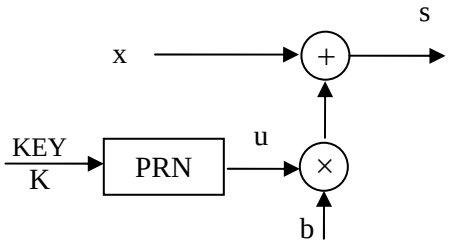
شکل ۱: بلوک دیاگرام بخش پنهان سازی داده در سیگنال صحبت.

آشکار سازی ، در حضور عملیات فشرده سازی به کیفیت شنیداری و مقاومت خوب دست یافته اند. در این مقاله نیز یک روش برای پنهان سازی داده در سیگنال صحبت که عملیات فشرده سازی بر روی انجام می شود ، ارائه گردیده است با این تفاوت که پنهان سازی درحوزه فرکانس صورت گرفته و برای آشکارسازی احتیاجی به سیگنال اصلی صحبت (سیگنال صحبت قبل از پنهان سازی داده) در گیرنده وجود ندارد و به عبارت دیگر آشکار سازی به صورت کور انجام می شود. به کمک این ویژگی فرستنده می تواند هر عبارت دلخواهی را ارسال کند و به یک نمونه از پیش تعیین شده، محدود نگردد. کدکننده ای که در این روش برای فشرده سازی مورد استفاده قرار گرفته است ، کدکننده G.723 از نوع کدرهای شکل موجی [۳] می باشد.

۲- پنهان سازی داده در سیگنال صحبت

شکل(۱) بلوک دیاگرام سیستم پنهان سازی داده در سیگنال صحبت را نشان می دهد. ابتدا بخش . پس از پنهان سازی داده ، بخش های صحبت و سکوت به یکدیگر افزوده شده و عملیات فشرده سازی بر روی آنها انجام می گردد و از طریق کانال برای گیرنده ارسال می شود. در گیرنده بعد از انجام فرآیند عکس فشرده سازی ، سیگنال صحبت بازیابی شده و بعد از جدا کردن بخش های سکوت ، داده ها از روی باقی مانده سیگنال صحبت ، استخراج می گردد.

یک سیگنال گفتار هم شامل بخش صحبت و هم بخش سکوت است ، که بخش سکوت دارای اهمیت چندانی نمی باشد و حتی در برخی از پردازش ها از قبیل فشرده سازی برای کاهش نرخ بیت ممکن است حذف گردیده و در طرف مقابل فقط از یک سیگنال شبه سکوت استفاده



شکل۲: مدل کلی سیستم طیف گسترده.

شود [۳] . به همین دلیل بخش سکوت باید تشخیص داده شده و اطلاعات مورد نظر فقط در بخش صحبت دار سیگنال قرار گیرد. برای تشخیص یک فریم N نمونه‌ای سکوت، می‌توان - از معیار میانگین قدر مطلق اندازه نمونه‌های فریم استفاده نمود. اگر این معیار از یک حد آستانه (T) بیشترباشد فریم به عنوان فریم‌صحبت‌تشخیص‌داده می‌شود.

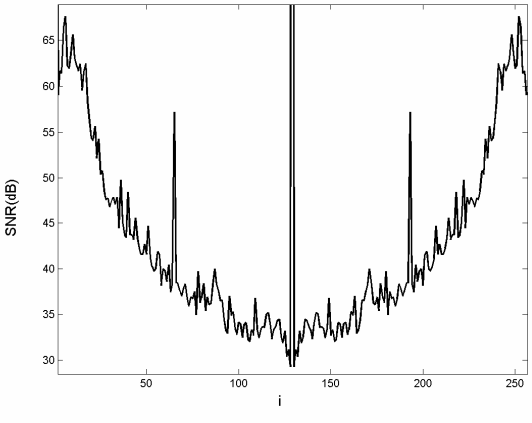
$$[\frac{1}{N}\sum_{i=1}^{N}|\, s_i\,|]>T \qquad (۱)$$

در رابطه (۱) ، s_i نمونه های فریم مورد نظر و Tحد آستانه‌ای است که به صورت تجربی تعیین می شود.

برای پنهان سازی داده ، از تکنیک طیف گسترده[۴] در حوزه فرکانس‌استفاده شده‌است. درروش طیف گسترده داده هایی که می خواهیم پنهان شود ، برروی‌حجم زیادی‌از سیگنال گسترده می‌شود. این کار را می‌توان بر روی نمونه های زمانی [۴] ، و یا فرکانسی [۵] اجرا نمود. داده ها به گونه ای روی پارامتر مورد نظر گسترده می شوند که تغییر ایجاد شده توسط گوش قابل درک نباشد. مدل کلی مربوط به سیستم طیف گسترده در شکل (۲) مشاهده می شود. x سیگنال میزبان با مقادیر ۱- یا ۱+ ، bرشته داده ای با عناصر ۱- یا ۱+ است و s

4 - Spread spectrum

سیگنال خروجی می باشد.[5] PRNG به کمک کلید مخفی K ،رشته چیپ u با میانگین صفر و واریانس δ_uرا تولید می کند که گیرنده و فرستنده از این کلید مطلع هستند. سیگنال خروجی به

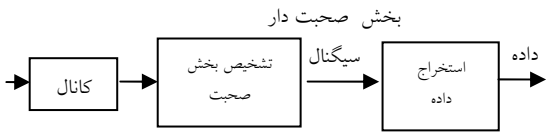


شکل۳: نمودار نسبت سیگنال به نویز ضرایب DFT در عبور از کانال فشرده ساز.

صورت رابطه (۲) تولید می شود.

$$s = x + bu \qquad (۲)$$

همانطور که می دانیم فشرده سازی باعث ایجاد اعوجاج در سیگنال می گردد. با توجه به یک سری از آزمایش ها مشخص شده است که میزان اعوجاج نمونه های حوزه فرکانسی نسبت به نمونه های حوزه زمانی در مقابل فشرده سازی کمتر است [۶] . بنابراین در این طرح داده های مورد نظر را بر روی ضرایب DFT[6] پخش کرده ایم. البته میزان این اعوجاج برای هر یک از ضرایب متفاوت است . برای تعیین اینکه کدامیک از این ضرایب دارای مقاومت بیشتری است از ۱۳ نفر خواسته شد که برای چند دقیقه صحبت کنند . در مرحله بعد هر یک از این سیگنال ها فریم بندی شده و ازآنها تبدیل DFT ، ۲۵۶ نقطه ای گرفته شده و نسبت سیگنال به نویز ضرایب DFT این فریم ها قبل و بعد از فشرده سازی محاسبه گردید و بعد از میانگین گیری نمودار شکل (۳) حاصل گردید. همانطور که مشاهده می شود ضرایب پایین نسبت به ضرایب بالا دارای اعوجاج کمتری در مقابل فشرده سازی هستند. به همین دلیل ضرایب ۵ تا ۵۴ (و با توجه به تقارن ، ضرایب ۲۰۴ تا ۲۵۳) برای پنهان سازی داده انتخاب گردیدند. برای پنهان سازی دو رشته چیپ تصادفی $u1$ و $u2$ ($u1$ و $u2$ قرینه



شکل۴: بلوک دیاگرام مربوط به بخش استخراج داده.

یکدیگر هستند) به طول ۵۰ نمونه که فرستنده و گیرنده از آن مطلع هستند انتخاب شدند. برحسب اینکه داده مورد نظر ۱- یا ۱+ (و یا ۰ یا ۱) باشد ، یکی از این دو رشته چیپ به عنوان رشته u انتخاب می شود. اگر F_i نمایش اندازه تبدیل DFT باشد ، این پارامتر به صورت رابطه (۳) تغییرمی کند.

$$F_i \leftarrow [F_i(1 + alpha \times u_{i-4})] \quad i = 5,...,54 \qquad (۳)$$

پارامتر $alpha$ در رابطه (۳) برروی وضوح شنیداری و خطا تاثیر می گذارد که باید به طور مناسبی انتخاب شود. پس از پنهان سازی داده در حوزه فرکانس با گرفتن عکس DFT سیگنال حوزه‌زمانی‌بازیابی‌می گردد. در مرحله بعد فریم های سکوت به فریم های صحبت افزوده شده و پس از فشرده سازی توسط کدکننده G.723 از طریق کانال برای گیرنده ارسال می شود.

۳- استخراج داده

بلوک دیاگرام مربوط به بخش استخراج داده در شکل (۴) مشاهده‌می‌شود. بعداز انجام عکس فشرده سازی ، فریم های صحبت از فریم‌های سکوت جدا می‌شوند و ازبخش صحبت داده ها استخراج می گردند. با توجه به رابطه (۲) ، استخراج بیت پنهان شده به کمک علامت پارامتر r که به صورت رابطه (۴) معرفی می شود، صورت می گیرد.

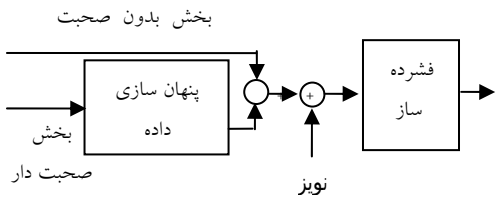
5 - Pseudo random noise generator
6 - Discrete Fourier transform

$$r = \frac{< s,u >}{< u,u >} = \frac{< x + bu,u >}{\delta_u} = b + c_x \qquad (۴)$$

$$c_x = \frac{< x,u >}{\delta_u} \qquad (۵)$$

در این‌صورت $\hat{b}$ ، بیت استخراج شده به صورت‌رابطه (۶) تخمین زده می شود.

$$\hat{b} = sign(r) \qquad (۶)$$



شکل۵: افزایش نویز سفید گوسی به سیگنال صحبت.

بعد از فریم بندی بخش صحبت ، از هر یک از فریم ها تبدیل DFT گرفته و اندازه ضرایب ۵ تا ۵۴ را در اولین رشته چیپ ، $u1$، ضرب کرده و حاصل جمع می گیریم ، اگر حاصل (lim) عددی مثبت باشد ، بیت به عنوان ۱+ و در غیر اینصورت به عنوان ۱- آشکار سازی می شود.

$$\lim = \sum_{i=5}^{54} [F_i \times u_{i-4}] \qquad (۷)$$

$$\hat{b} = \begin{cases} +1 & \lim > 0 \\ -1 & \lim < 0 \end{cases} \qquad (۸)$$

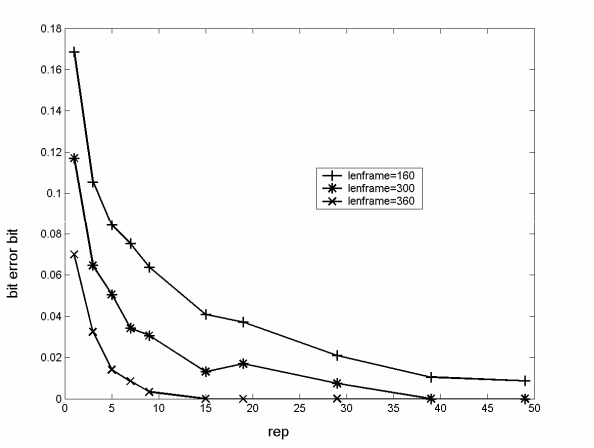
که $\hat{b}$ در رابطه (۸) بیت استخراج شده می باشد.

۴- افزایش نویز سفید گوسی

برای بررسی اثر نویز بر عملکرد سیستم مطابق شکل (۵) ، نویز سفید گوسی به سیگنال صحبت افزوده می شود. همانطور که مشاهده می شود ، افزایش نویز بعد از پنهان سازی داده و قبل از فشرده سازی صورت می گیرد. برای این منظور نویز با نسبت های مختلف سیگنال به نویز به سیگنال صحبت افزوده می شود و اثر آن بر دقت آشکار سازی در گیرنده مورد بررسی قرار می گیرد.

۵- شبیه سازی و نتایج

سیگنال صحبت با فرکانس ۸ kHz نمونه برداری شده و هر نمونه با ۱۶ بیت نمایش داده می‌شود. برای‌فریم‌بندی - سیگنال‌صحبت‌از فریم هایی به طول ۲۰ ms (معادل ۱۶۰ = N نمونه)استفاده‌می گردد. برای تشخیص فریم های سکوت ، میانگین قدر مطلق نمونه های فریم با سطح آستانه T= ۲۰۰ مقایسه می شود(مقدار این پارامتر برای فریم کاملا" سکوت حدود ۸۰ می باشد که برای اطمینان ، سطح آستانه ۲۰۰ در نظر گرفته شده است. البته این مقادیر به



شکل ۶: نمودار خطا بر حسب پارامتر rep.

شرایط تست، میکروفن‌وکارت صدای کامپیوتر وابسته می- باشد) . بعد از تعیین فریم های‌صحبت ، از آنها تبدیل DFT ، ۲۵۶ نقطه ای گرفته و داده ها ی مورد نظر را با توجه با الگوریتم تشریح شده در بخش (۲) در ضرایب مربوطه جای‌می‌دهیم. برای انتخاب رشته های u_1 و u_2 از دو رشته تصادفی به طول ۵۰ استفاده شده است که فرستنده و گیرنده از آن مطلع هستند. مقدار پارامتر $alpha$ با توجه به معیار شنیداری ۰,۱۵ در نظرگرفته شده است. برای رسیدن به خطای کمتر در آشکار سازی هر بیت داده در تعداد rep فریم پنهان می گردد (rep عددی فرد انتخاب می شود) . در مرحله بعد با عکس DFT گرفتن ، به سیگنال حوزه زمان باز می گردیم و با افزودن فریم های سکوت سیگنال صحبت کامل

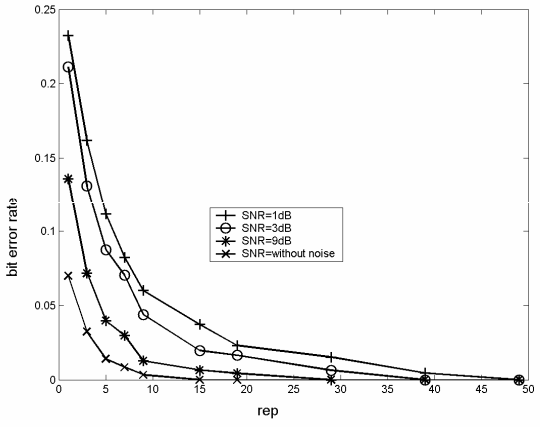
بازیابی می شود. سپس عملیات فشرده سازی بـر روی سـیگنال صورت گرفتـه و حاصـل بـرای گیرنـده ارسـال مـی شـود. در گیرنده بعد از انجام عکس فشرده سازی و تشخیص فریم های صحبت ، بیت های پنهان شده استخراج مـی شـوند. چـون هـر بیت rep بار تکرار شده است با توجه به معیار حداکثر تکـرار، بیت پنهان شده تخمین زده می شود. آزمایش برای ۵ گوینـده و با مقادیر مختلف rep تکرار شده است کـه نمـودار میـانگین خطا بر حسب پارامتر rep آن در شکل (۶) مشاهده می‌شـود.- برای بررسی اثرطول فریم بـر روی‌خطـا ، آزمـایش‌بـرای‌فـریم هایی به طول ۳۰۰ و ۳۶۰ نمونه هـم تکـرار گردیـد. بـرای ایـن کار از ضرایب ۱۰ تا ۱۰۹ (۱۰۰ نمونه) DFT ،۵۱۲ نقطه ای استفاده شد. نتایج این آزمایش در شکل (۶) ترسیم شده است. همانطور که مشاهده می شود با افزایش طول فریم خطا کاهش می یابد و حتی می توان به خطای صفر هم

جدول ۱: نرخ بیت داده پنهان شده در بخش صحبت بر حسب bps.

طـــول فریم / rep	۱۶۰	۳۰۰	۳۶۰
۱	۵۰	۲۶,۶۶	۲۲,۲۲
۳	۱۶۰۶۶	۸,۸۸	۷,۴۰
۵	۱۰	۵,۳۳	۴,۴۴
۷	۷,۱۴	۳,۸۰	۳,۱۷
۹	۵,۵۵	۲,۹۶	۲,۴۶
۱۵	۳,۳۳	۱,۷۷	۱,۴۸
۱۹	۲,۶۳	۱,۴۰	۱,۱۶
۲۹	۱,۷۲	.۹۱	.۷۶
۳۹	۱,۲۸	.۶۸	.۵۶
۴۹	۱,۰۲	.۵۴	.۴۵

رسید. البته باید توجه داشت که در قبال نرخ پـایین خطـا بـه نرخ داده های کمتر بسنده شده است. از آنجایی که نرخ بیـت در این نمودارها بر حسب پارامتر rep بیان شده اسـت ، مقـدار نرخ بیت داده بر حسب bps[7] در جدول (۱) مشاهده می شود.

در مرحلـه بعـد، اثـر افـزایش نـویز بـه سـیگنال صـحبت مـورد بررسی قرار می‌گیرد. برای‌ایـن‌منظـور مطـابق شـکل(۵)، نـویز سفید گوسی‌با نسبت های سیگنال به نویز ۱ dB ، ۳ dB و ۹ dB به سیگنال افزوده‌می شود . طول فریم هایی که فرآیند پنهان سازی بر روی آنها صورت گرفته است ، ۳۶۰ نمونه می باشـد. در نمـودار شـکل(۷) اثـر افـزایش نـویز بـر نـرخ خطـای آشکارسازی مشاهده می شود.همانطورکه‌ملاحظه مـی گـردد در نسبت سیگنال به نویز های پـایین نیـز مـی تـوان بـا افـزایش پارامتر rep به دقت خوبی در آشکار سازی دست پیدا کرد. بـا توجه به آزمایش های انجام شده کیفیـت شـنیداری سـیگنال صحبت بعد از پنهان سازی داده ، از کیفیت خـوبی برخـوردار است و تفاوت چندانی با سیگنال اصلی ندارد.



شکل۷: اثر نویز بر نرخ خطای آشکار سازی

۶- نتیجه گیری

در این مقاله روشی جدید برای پنهان سازی داده در سـیگنال-صحبتی‌که‌فشرده سازی بر روی‌آن صورت می گیرد ، معرفـی شده است. از ویژگی های مهم این روش ، مقاومـت در مقابـل فشرده سازی ، کیفیت شنیداری خوب ، آشکار سازی کـور و امکان آشکار سازی در نسبت های سیگنال بـه نـویز پـایین‌مـی باشد. آشکار سازی کور این امکان را مـی دهـد کـه گیرنـده

7 - bit per second

برای استخراج داده احتیاجی به سیگنال اصلی صحبت ندارد و این سیگنال می‌تواند هر عبارت دلخواهی باشد . از این ویژگی می توان در کاربردهای امنیتی استفاده کرد و یک داده را به صورت امن در خلال یک مکالمه برای طرف مورد نظر ارسال نمود. از آنجایی که در اغلب کاربردها، داده ای که- به این صورت پنهان می شود، یک عبارت کوتاه و یا یک کلید رمز است ، نرخ بیت یک پارامتر اساسی نمی باشد و می توان ازاین الگوریتم برای ارسال داده مورد نظر استفاده کرد. در ضمن می‌توان با بکار بردن دلخواه طول فریم صحبت ، طول تبدیل DFT ، ضرایب DFT و رشته چیپ های گسترش- دهنده، به گونه ای که فرستنده و گیرنده از آن مطلع باشند ، امنیت الگوریتم را بالا برد.

مراجع

[1] Nedeljko Cvejic, "Algorithms For Audio Watermarking And Steganography" ,*Thesis report*, Department of Electrical and Information Engineering ,Information Processing Laboratory, University of Oulu 2004.

[۲] رقیه دوست ، حسن آقایی و حسین شمسی ، *ایجاد کیفیت ادراکی و مقاومت بسیار مطلوب نهان نگاری ، در برابر فشرده سازی صحبت* ، سیزدهمین کنفرانس مهندسی برق ایران ، بهار ۱۳۸۴.

[3] A.Kondoz, *Digital Speech*, John Wiley & Sons, Inc., New York,2004.

[4] Paraskevi Bassia, Ioannis Pitas, Nikos Nikolaidis, "Robust audio watermarking in the time domain". *IEEE Transactions on Multimedia*, vol.3,p 232–241, June 2001.

[5] Darko Kirovski , Henrique S.Malvar , "Spread-Spectrum watermarking of audio signals",*IEEE Transaction on Signal Processing* ,vol.51,no.4,April 2003.

[6] Chung-Ping Wu , C.-C. Jay Kuo, "Fragile speech watermarking for content integrity verification", *IEEE International Conference on Acoustics, Speech and Signal Processing* ,May 2002.