

Improved Sensor Sampling Method for the Joint Dictionary Learning and Compressive Data Gathering in WSNs with the Aid of Information Theory**Gholamreza Imanian¹, Ph.D. Candidate, Mohammad Ali Pourmina¹, Associate Professor, Ahmad Salahi², Associate Professor**¹Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran²Iran Telecommunication Research Center, Tehran, Iran
rez.iman@gmail.com, pourmina@srbiau.ac.ir, salahi@itrc.ac.ir**Abstract**

In the last decade, to reduce the costs of environmental monitoring, the data aggregation based on the joint dictionary learning and Compressive Sensing (CS) technique in Wireless Sensor Networks (WSNs) has been considered. It has been shown that when the dictionary obtained with the learning technique based on principal component analysis is used for CS-based data gathering of environmental signals, using a deterministic node selection method for data collection in WSNs can outperform random node selection ones. In this article, a deterministic and non-random sampling design for use in the CS-based data aggregation method is presented. This method is based on estimating the amount of mutual information of sensor data and is obtained by sampling all of them in a short part of the data collection round named the training phase. In the next and main stage of the data collection period, only the nodes that provide the most information about the non-sampled nodes are scheduled to sample. Simulation results for real signals in MATLAB software environment show that when the number of sampling sensors comprises still about 25% of the total network nodes, average energy savings of more than 12% can be achieved over a reference sampling method.

Keywords: compressive sensing, data aggregation, dictionary learning, environmental monitoring, wireless sensor network.**Received:** 16 December 2020**Revised:** 15 February 2021**Accepted:** 8 April 2021**Corresponding Author:** Dr. Mohammad Ali Pourmina

<https://dorl.net/dor/20.1001.1.23223871.1401.13.49.3.2>

مقاله پژوهشی

بهبود روش نمونه‌برداری تجمیع داده مبتنی بر سنجش فشرده و یادگیری لغت‌نامه در شبکه حسگر بی‌سیم به کمک نظریه اطلاعات

غلامرضا ایمانیان^۱، دانشجوی دکتری، محمدعلی پورمینا^۱، دانشیار، احمد صلاحی^۲، دانشیار

۱- دانشکده مهندسی برق و کامپیوتر- واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

۲- پژوهشگاه ارتباطات و فن‌آوری اطلاعات، تهران، ایران

rez.iman@gmail.com, pourmina@srbiau.ac.ir, salahy@itrc.ac.ir

چکیده: در دهه اخیر با هدف کاهش هزینه‌های نظارت محیطی، فرایند تجمیع داده مبتنی بر روش مشترک سنجش فشرده و یادگیری لغت‌نامه در شبکه‌های حسگر بی‌سیم مورد توجه قرار گرفته‌است. در این مقاله یک طرح نمونه‌برداری قطعی و غیر تصادفی برای استفاده در این روش تجمیع داده ارائه شده‌است. این طرح مبتنی بر برآورد کمیت اطلاعات متقابل داده حسگرها است که با نمونه‌برداری از تمام آنها در بخش کوتاهی از دوره جمع‌آوری داده به‌نام مرحله آموزش به‌دست می‌آید. در مرحله بعدی و اصلی دوره جمع‌آوری داده گره‌هایی نمونه‌برداری می‌شوند که بیشترین اطلاعات را درباره گره‌های نمونه‌برداری نشده در اختیار بگذارند. نتایج شبیه‌سازی‌ها با سیگنال‌های واقعی نشان می‌دهد که حتی زمانی که تعداد حسگرهای نمونه بردار تنها شامل ۲۵ درصد از کل گره‌های شبکه است می‌توان به‌طور متوسط به بیش از ۱۲ درصد صرفه‌جویی در مصرف انرژی نسبت به روش نمونه‌برداری مرجع دست یافت.

کلمات کلیدی: جمع‌آوری داده، سنجش فشرده، شبکه حسگر بی‌سیم، نظارت محیطی، یادگیری لغت‌نامه.

تاریخ ارسال مقاله: ۱۳۹۹/۹/۲۶

تاریخ بازنگری مقاله: ۱۳۹۹/۱۱/۲۷

تاریخ پذیرش مقاله: ۱۴۰۰/۱/۱۹

نام نویسنده‌ی مسئول: دکتر محمدعلی پورمینا

نشانی نویسنده‌ی مسئول: تهران- انتهای بزرگراه شهید ستاری- میدان دانشگاه- بلوار شهدای حصارک- دانشگاه آزاد اسلامی

واحد علوم و تحقیقات

۱- مقدمه

در دو دهه گذشته به علت تولید انبوه حسگرهای کوچک با قابلیت‌های ارتباطی و محاسباتی علاوه بر قابلیت سنجش، شبکه‌های حسگر بی‌سیم در زمینه‌هایی مانند نظارت بر محیط، ردیابی و جستجو کاربرد وسیعی پیدا کرده‌اند [۴-۱]. از جمله مهم‌ترین اهداف شیوه‌های تجمیع داده در این شبکه‌ها کاهش مصرف انرژی در گره‌های میانی (نه چاهک) است که در نتیجه آن طول عمر شبکه افزایش خواهد یافت. با توجه به این حقیقت که در شبکه‌های حسگر بی‌سیم، غالباً، بیشتر مصرف انرژی در ارتباطات بین گره‌ها اتفاق می‌افتد نه در محاسبات محلی آنها، کاهش تعداد بسته‌های ارسالی و یا کاهش طول آنها، اثر مهمی بر روی مصرف انرژی و کارآمدی کلی شبکه می‌گذارد [۵]. بر این اساس با توجه به همبستگی‌های مکانی و زمانی موجود در داده‌هایی که توسط شبکه‌های حسگر بی‌سیم جمع‌آوری می‌شوند خصوصاً در کاربردهایی که به نظارت سیگنال‌های محیطی^۱ مربوط می‌شود فشرده‌سازی داده‌ها از اهمیت زیادی برخوردار است.

همچنین در چند سال گذشته نظریه سنجش فشرده^۲ (CS) به‌عنوان روشی که فرآیندهای سنجش و فشرده‌سازی را یکجا باهم ترکیب می‌کند، در کاربردهایی مانند نظارت مبتنی بر شبکه‌های حسگر بی‌سیم [۶،۷]، سنجش طیفی^۳ [۸،۹] و تصویربرداری تشدید مغناطیسی [۱۰] که جمع‌آوری داده هزینه‌بر است مورد توجه قرار گرفته است. نظریه CS بازسازی موثر سیگنال نمونه‌برداری شده را با تعداد نمونه‌هایی بسیار کمتر از نرخ نایکوئیست تضمین می‌کند به شرط آنکه سیگنال در یک فضای پایه مناسب دارای بازنمایی تنک باشد. علاوه بر این اگر عملگر نمونه‌برداری به صورت یک ماتریس سنجش^۴ مشخص شود به جهت بازسازی موفق سیگنال در روش مرسوم CS اغلب ماتریس‌های تصادفی چگال مورد استفاده قرار می‌گیرند [۱۱].

استفاده از CS در شبکه‌های حسگر بی‌سیم با کار باژوا و دیگران روی یک شبکه تک پرش آغاز شد [۱۲]. بعدها این شیوه توسط کارهای دیگران [۱۳،۱۴] به شبکه‌های چند پرش توسعه داده شد. زیانگ و دیگران [۱۴] نشان دادند که استفاده از ماتریس‌های تصادفی چگال ممکن است نتواند تعداد ارسال‌ها و در نتیجه انرژی مصرفی کل را در مقایسه با روش‌های تجمیع داده بدون CS کاهش دهد. پژوهش‌های دیگر افزایش انرژی-کارآمدی با استفاده از کاربرد ماتریس‌های سنجش تنک در شبکه‌های حسگر بی‌سیم را هدف قرار دادند [۱۵-۱۸]. در بین این پژوهش‌ها، لی [۱۵] و کوور [۱۶] به صورت مستقل نشان دادند تا زمانی که فضای مناسبی برای بازنمایی به اندازه کافی تنک داده‌های شبکه‌ای وجود نداشته باشد در برخی اوقات عملکرد CS در بازسازی موفق داده‌های شبکه‌ای در حد روش‌های تجمیع داده‌ای است که از فرو نمونه‌برداری^۵ داده‌ها و سپس درون‌یابی برای بازسازی داده‌ها بهره می‌برند.

سیگنال‌های واقعی مرتبط با کاربردهای نظارت محیطی غالباً غیر ایستاد بوده و ساختار همبستگی آنها در طول دوره‌های طولانی جمع‌آوری داده ثابت نمی‌ماند. در پژوهشی دیگر کوور [۱۹] برای تنک‌سازی داده‌ها به جای لغت‌نامه‌های ثابتی مانند تبدیل کسینوسی گسسته و تبدیل موجک، استفاده از روش یادگیری لغت‌نامه مبتنی بر تبدیل تجزیه مؤلفه اصلی^۶ (PCA) را پیشنهاد کرده است. دست‌یابی به یک فضای پایه تنک ساز بر اساس فرآیندی کاملاً منطبق^۷ با مشخصات واقعی سیگنال‌ها مزیت این روش است. برخلاف لغت‌نامه‌های ثابت مزیت دیگر این روش آنست که برای وضعیتی که پراکنده‌گی حسگرها به‌صورت نامنظم و غیر مشبک^۸ است نیز قابل استفاده خواهد بود.

طرح نمونه‌برداری که در روش مشترک CS و PCA توسط کوور استفاده می‌شود به‌طوری است که حسگرها به‌صورت تصادفی در شکاف‌های زمانی دوره جمع‌آوری داده انتخاب شده و داده‌های خود را از طریق مسیریابی چند پرش ارسال می‌کنند. اخیراً هوشمند و دیگران [۲۰] روش جدیدی به‌نام انتخاب قطعی گره‌ها مبتنی بر همبستگی بهبودیافته^۹ (ECB-DNS) ارائه کرده‌اند که در آن حسگرها داده‌های خود را براساس یک زمان‌بندی از پیش تعیین شده و نه تصادفی ارسال می‌کنند. چاهک، این زمان‌بندی را در بخش آغازین دوره جمع‌آوری داده که مرحله آموزش^{۱۰} نامیده می‌شود بر اساس همبستگی بین داده‌گرها محاسبه کرده و در کل شبکه توزیع می‌کند. به کمک این روش که داده حسگرها به‌صورت منظم در مکان و زمان نمونه‌برداری می‌شود بهبود قابل ملاحظه‌ای در کیفیت بازسازی داده‌ها در چاهک نسبت به نمونه‌برداری تصادفی به‌دست آمده است. عملکرد بهتر در بازیابی داده‌ها نتیجه تاثیر مثبت نمونه‌برداری قطعی بر روی فرآیند یادگیری لغت‌نامه مبتنی بر PCA است.

معیاری که توسط هوشمند و دیگران برای انتخاب گره‌ها در شکاف‌های زمانی استفاده می‌شود دو اشکال دارد. اول اینکه این معیار علاوه بر همبستگی بین داده گره‌ها به مقیاس داده‌ها نیز وابسته است به‌طوری‌که گره‌های با داده‌های بزرگ‌تر به خطا سهم بزرگ‌تری را در این معیار ایجاد کرده و در نتیجه انتخاب گره‌های مناسب را با مشکل مواجه می‌کنند. دوم اینکه در فرآیند استخراج جدول زمانی برای ارسال داده گره‌های شبکه، چاهک اطلاعات گره‌های انتخابی پیشین در یک شکاف زمانی را برای انتخاب گره‌های دیگر در آن شکاف در نظر نمی‌گیرد. نوآوری ما در این کار ارائه یک روش نمونه‌برداری قطعی جدید مبتنی بر اطلاعات متقابل بین گره‌های شبکه است که در آن اشکال‌های فوق برطرف شده‌است. نتایج به‌دست آمده از شبیه‌سازی برای داده‌های واقعی مورد آزمایش، دستیابی به بهبود عملکرد را برای محدوده وسیعی از مصرف انرژی در گره‌های شبکه تایید می‌کند.

ادامه مقاله به‌صورت زیر سازماندهی شده‌است. در بخش ۲ فرمول بندی مساله با توصیفی مقدماتی از فرآیند تجمیع داده مبتنی بر CS و فرآیند یادگیری لغت‌نامه مبتنی بر PCA ارائه خواهد شد، همچنین روش مرجع برای نمونه‌برداری حسگرها از سیگنال‌های شبکه حسگر بی‌سیم توضیح داده می‌شود. در بخش ۳ روش نمونه‌برداری قطعی پیشنهادی و روش نمونه‌برداری مرجع توصیف می‌شود. در بخش ۴ توصیف سیگنال‌ها و شبکه‌های مورد آزمایش و مدل انرژی استفاده شده معرفی می‌گردد. در بخش ۵ نتایج عددی حاصل از شبیه‌سازی‌ها برای ارزیابی و مقایسه روش پیشنهادی با روش مرجع مورد توجه قرار گرفته‌است. در نهایت، در بخش ۶ جمع بندی مطالب ارائه می‌گردد.

۲- بیان مساله تجمیع داده در شبکه حسگر بی‌سیم

در این بخش، جمع‌آوری داده از یک شبکه حسگر بی‌سیم را در قالب روش مشترک CS و یادگیری PCA مرور می‌شود. در مرجع [۱۹]، نویسندگان ثابت کرده‌اند که وقتی با رویکرد قاعده بیز به مساله تجمیع داده پرداخته شود شیوه بازسازی داده‌ها در چاهک با شیوه فوق، برای جمع‌آوری سیگنال‌های محیطی بهینه است. به‌منظور تطبیق تجمیع داده با جمع‌آوری مشاهداتی از سیگنال‌های شبکه که در حالت کلی دارای مشخصات آماری متغیر هستند از روشی متمرکز و دو مرحله‌ای استفاده می‌شود. طول دوره جمع‌آوری داده به شکاف‌هایی زمانی تقسیم بندی شده و داده‌ها به ترتیب در دو مرحله آموزش و نظارت نمونه‌برداری می‌شوند ولی ارسال سنجش‌ها به چاهک تا انتهای دوره جمع‌آوری داده به تاخیر می‌افتد. بر این اساس در کاربردهایی که تاخیر در جمع‌آوری داده مجاز است به کمک این روش بخش قابل توجه مصرف انرژی در ارسال داده به ظرفیت ترابری^{۱۱} آنها اختصاص خواهد یافت نه سرآیند بسته‌ها.

۲-۱- تجمیع داده مبتنی بر CS

داده‌های جمع‌آوری شده از N گره شبکه حسگر بی‌سیم را می‌توان به‌صورت یک بردار N بعدی به‌صورت زیر نشان داد:

$$\mathbf{r} = [r_1 \ r_2 \ \dots \ r_N]^T \in \mathbb{R}^N \quad (1)$$

که در آن r_i نشان‌دهنده مقداری است که توسط گره i ام جمع‌آوری شده‌است. برای اینکه توالی زمانی داده‌ها در نظر گرفته شود سیگنال جمع‌آوری شده در لحظه t را با r_t نمایش می‌دهیم. به کمک CS [۱۹]، در هر لحظه t تنها با جمع‌آوری بخش ناقصی از داده‌های N گره شبکه به‌صورت بردار کوچکتر y_t به طول L ($L < N$) می‌توان بردار داده r_t را به‌صورت تقریبی به‌دست آورد به شرط آنکه این بردار در یک لغت‌نامه مانند رابطه (۲) به‌صورت تنگ رابطه (۳) قابل نمایش باشد:

$$(N \leq \tilde{N}) \mathbf{D}_t \in \mathbb{R}^{N \times \tilde{N}} \quad (2)$$

$$\mathbf{r}_t = \mathbf{D}_t \mathbf{s}_t \quad (3)$$

در رابطه بالا s_t یک بردار K -تنگ به طول \tilde{N} ($K \ll \tilde{N}$) است که همان نمایش تنگ r_t است. منظور ما از K -تنگ آن است که تنها K مؤلفه از \tilde{N} مؤلفه بردار با اهمیت است و بقیه در مقایسه با میانگین انرژی محاسبه شده روی تمام مؤلفه‌ها قابل صرف نظر کردن هستند.

داده‌های خوانده شده از تعداد L گره h سگر بر اساس روش نمونه‌برداری که در بخش ۳ ذکر خواهد شد از بین داده‌های کل شبکه انتخاب شده و برای چاهک ارسال می‌شود تا آنها را در بردار y_t ذخیره کند. داده‌های خوانده شده جزئی (y_t) و داده‌های خوانده شده کل شبکه (r_t) مطابق رابطه زیر به هم مرتبط هستند:

$$y_t = M_t r_t \quad (4)$$

در این رابطه ماتریس M_t با ابعاد $L \times N$ یک ماتریس نمونه‌برداری است که بجز دقیقا یک ۱ در هر سطر و حداکثر یک ۱ در هر ستون بقیه عناصر آن صفر است. ماتریس M_t همان ماتریسی است که در فرآیند CS از آن با عنوان ماتریس سنجش نام برده می‌شود. با قرار دادن رابطه (۳) در (۴) خواهیم داشت:

$$y_t = M_t D_t s_t \quad (5)$$

با فرض مشخص بودن دو ماتریس D_t و M_t در گره چاهک (به ترتیب به بخش‌های ۲-۲ و ۳ مراجعه کنید)، این گره می‌تواند جواب‌های دستگاه معادلات نامعین^{۱۲} رابطه (۵) را بر اساس نظریه CS از حل یک مساله بهینه سازی با تقریب خوبی به دست آورد [۲۱, ۲۲]. چاهک با به کار بردن این جوابها در رابطه (۳) تقریبی از بردار r_t که آنرا با \hat{r}_t نمایش می‌دهیم به دست خواهد آورد.

۲-۲- یادگیری لغت‌نامه مبتنی بر PCA

PCA روشی استاندارد است که با استفاده از آن، مجموعه‌ای از مشاهدات داده‌ای N بعدی مثل X در زیر فضایی با ابعادی کوچک‌تر از N به صورت بهینه تقریب زده می‌شود [۲۳]. در حقیقت روش PCA، یکی پس از دیگری، به دنبال جهت‌های متعامدی برای نمایش داده‌ها می‌گردد که واریانس تصویر داده‌ها در آن جهت‌ها ماکزیمم شود. بنابراین داده‌ها در فضای تصویر کننده PCA که از بردارهای ویژه ماتریس کوواریانس داده‌ها به دست می‌آید به صورت تنک قابل نمایش خواهند بود. تبدیلات متعارف مانند تبدیل فوریه گسسته، تبدیل موجک و تبدیل کسینوسی گسسته غالباً به عنوان لغت‌نامه سیگنال‌های محیطی در تجمیع داده مبتنی بر CS موفق عمل نمی‌کنند [۱۶]، به همین علت در کار [۱۹] لغت‌نامه تنک‌ساز مبتنی بر PCA پیشنهاد شده و عملکرد بهتر آن در تجمیع داده مبتنی بر CS نسبت به تبدیلات متعارف فوق نشان داده شده است. چاهک به کمک الگوریتم PCA می‌آموزد که چگونه لغت‌نامه تنک‌ساز را برای هر یک از شکاف‌های زمانی مرحله نظارت، با استفاده از داده‌های به دست آمده در شکاف‌های زمانی قبلی استخراج کند. الگوریتم یادگیری لغت‌نامه مبتنی بر PCA بگونه‌ای است که در زیر توضیح داده می‌شود [۲۰]:

۱. مطابق روش PCA، بردارهای ویژه به دست آمده از تجزیه مقادیر ویژه^{۱۳} ماتریس کوواریانس داده‌های جمع‌آوری شده در مرحله آموزش را به عنوان بردارهای پایه لغت‌نامه تنک‌ساز D_t در اولین شکاف زمانی مرحله نظارت در نظر می‌گیریم.

۲. با فرض معلوم بودن D_t و M_t در شکاف زمانی فعلی بردار داده‌های بازسازی شده \hat{r}_t را بر اساس شیوه CS به دست می‌آوریم.

۳. بردار \hat{r}_t را به نمونه‌های ذخیره شده قبلی این بردار^{۱۴} در بافری به طول H شکاف زمانی^{۱۵} با شرط $B_t = \{\hat{r}_t, \hat{r}_{t-1}, \dots, \hat{r}_{t-H+1}\}$ ملحق می‌کنیم.

۴. ماتریس کوواریانس داده‌های بازسازی شده $\hat{\Sigma}_t$ را به عنوان تقریبی از ماتریس کوواریانس داده‌های واقعی با استفاده از نمونه‌های جمع شده در ماتریس B_t به ابعاد $N \times H$ محاسبه می‌کنیم.

۵. با استفاده از تبدیل PCA از روی $\hat{\Sigma}_t$ لغت‌نامه تنک‌ساز را برای استفاده در شکاف زمانی بعدی (D_{t+1}) به دست می‌آوریم.

۶. چنانچه دوره جمع‌آوری داده به پایان نرسیده است به مرحله ۲ رفته و دوباره گام‌های الگوریتم را تکرار می‌کنیم.

۳- روش نمونه‌برداری حسگرها در شبکه حسگر بی‌سیم

در تمام شکاف‌های زمانی مرحله آموزش که مدت نسبتاً کوتاهی از کل زمان تجمیع را به خود اختصاص می‌دهد همه گره‌های حسگر نمونه‌برداری را انجام می‌دهند ولی در مرحله نظارت ضرورتاً تمام گره‌ها در هر شکاف زمانی نمونه‌برداری را انجام

نمی‌دهند، بلکه بر اساس یک زمان‌بندی از پیش تعیین شده این کار صورت می‌گیرد. گره چاهک با توجه به سنجش‌هایی که در مرحله آموزش از کل گره‌های شبکه جمع‌آوری کرده است تصمیم می‌گیرد که کدام گره باید در کدام شکاف زمانی برای نمونه‌برداری انتخاب شود. زمان‌بندی استخراج شده که حاوی اطلاعات باینری نمونه‌برداری یا عدم نمونه‌برداری در شکاف‌های زمانی است به تمام گره‌ها ارسال می‌شود و به این ترتیب هر گره تنها با یک عمل مقایسه می‌تواند تصمیم بگیرد که در یک شکاف زمانی معین عمل سنجش را انجام دهد یا نه. براین اساس اگر طبق زمان‌بندی استخراج شده تعدادی گره مجاز به نمونه‌برداری در شکاف زمانی t باشند ماتریس نمونه‌برداری نظیر آن شکاف زمانی M_t ماتریسی تمام صفر است به جز دقیقاً یک ۱ در هر سطر که موقعیت ستونی این عناصر غیر صفر مطابق گره‌های مجاز به نمونه‌برداری در این شکاف انتخاب می‌شود (به بخش ۲-۱ مراجعه کنید). در ادامه روشی به نام انتخاب گره مبتنی بر همبستگی متقابل پیشرفته^{۱۶} (ECB-DNS) [۲۰] برای استخراج زمان‌بندی فوق توصیف می‌شود که به عنوان روش مرجع جهت مقایسه با روش پیشنهادی در نظر گرفته شده است. قبل از آشنایی با این دو روش انتخاب گره لازم است که یک سری تعاریف مرتبط با این روش‌ها تشریح شود. پارامترها به صورت زیر تعریف می‌شوند:

مقدار گره i : مقدار کمیت سنجش شده از سیگنال محیطی پراکنده در شبکه که توسط حسگر واقع در موقعیت گره i ام خوانده می‌شود. مطابق روابط استفاده شده در بخش ۲-۱ می‌توان این کمیت را در هر شکاف زمانی t با r_{it} نشان داد.

میانگین گره i : میانگین آماری به دست آمده از روی مقادیر گره i در θ شکاف زمانی که بصورت رابطه زیر محاسبه می‌شود:

$$\eta_i = \frac{\sum_{t=1}^{\theta} r_{it}}{\theta} \quad (6)$$

واریانس گره i : واریانس آماری به دست آمده از روی مقادیر گره i در θ شکاف زمانی که بصورت رابطه زیر محاسبه می‌شود:

$$\sigma_i^2 = \frac{\sum_{t=1}^{\theta} (r_{it} - \eta_i)^2}{\theta - 1} \quad (7)$$

همبستگی گره i و j : همبستگی متقابل آماری بین مقادیر گره‌های i و j در θ شکاف زمانی که بصورت رابطه زیر محاسبه می‌شود:

$$\rho_{ij} = \frac{\sum_{t=1}^{\theta} (r_{it} - \eta_i)(r_{jt} - \eta_j)}{(\theta - 1)\sigma_i\sigma_j} \quad (8)$$

۳-۱- روش انتخاب گره ECB-DNS [۲۰]

در روش ECB-DNS دقیقاً تعداد L گره حسگر در هر شکاف زمانی $t \in \{1, 2, \dots, T\}$ از دوره جمع‌آوری داده T براساس همبستگی مکانی داده حسگرها انتخاب می‌شود. برای اینکه از لحاظ محاسباتی با مدلی ساده و قابل کنترل^{۱۷} روبرو شویم فرض می‌شود که داده‌ها دارای توزیع توأم نرمال هستند. این فرض به طور متداول برای مدل سازی سیگنال‌های محیطی جمع‌آوری شده از تجهیزات شبکه‌های حسگر بی سیم و نیز تولید مصنوعی آنها مورد استفاده قرار می‌گیرد [۲۷-۲۴]. براساس این فرض واریانس شرطی مقدار سنجش شده یک گره (X_i) به شرط مقدار گره دیگر (X_j) به صورت رابطه زیر قابل محاسبه است:

$$\text{Var}(X_i | X_j) = \sigma_i^2 (1 - \rho_{ij}^2) \quad (9)$$

در روش ECB-DNS از این واریانس شرطی به عنوان کمیتی استفاده می‌شود که بیانگر اطلاعات یک گره درباره گره دیگر است. با فرض اینکه تعداد گره‌های شبکه (N) طوری باشد که در کل دوره جمع‌آوری داده T دقیقاً بتوان در هر شکاف زمانی L گره انتخاب کرد یعنی $T = N/L$ ، هر بار که در یک شکاف زمانی گره‌ای مانند Z از بین گره‌هایی که تاکنون انتخاب نشده گزینش می‌شود، باید مجموع واریانس مقادیر گره‌های دیگر که تاکنون انتخاب نشده‌اند مشروط به مقدار این گره که به صورت زیر محاسبه می‌شود کمترین مقدار باشد:

$$m_j = \sum_i \sigma_i^2 (1 - \rho_{ij}^2) = \sum_i \sigma_i^2 - \sum_i \sigma_i^2 \rho_{ij}^2 \quad (10)$$

چون در رابطه بالا مجموع $\sum \sigma_i^2$ مستقل از گره انتخابی مقداری ثابت است، بنابراین انتخاب گره‌ای که کمترین مقدار واریانس مشروط کل^{۱۸} (m_j) را دارا باشد معادل انتخاب گره‌ای است که بیشترین مقدار کمیت جدید زیر را داشته باشد:

$$m'_j = \sum_i \sigma_i^2 \rho_{ij}^2 \quad (11)$$

پارامتر اخیر در حقیقت میزان همبستگی مقدار سنجش شده در گره حسگر z با مقادیر سنجش شده در گره‌های حسگر دیگر را نشان می‌دهد. در روش ECB-DNS ما ابتدا تعداد T گره را که به ترتیب بیشترین همبستگی را با بقیه گره‌ها دارند انتخاب کرده و به شکاف‌های زمانی $\{1, 2, \dots, T\}$ اختصاص می‌دهیم. سپس همین تعداد گره را از بین بهترین گره‌های باقیمانده بر اساس معیار کمترین m_j یا بیشترین m'_j انتخاب کرده و مجدداً به شکاف‌های زمانی $\{1, 2, \dots, T\}$ نسبت می‌دهیم و این فرایند را تکرار می‌کنیم تا کل گره‌ها به شکاف‌های زمانی منتسب شوند. به این ترتیب در هر گام، به‌طور عادلانه، بهترین گره‌ها در بین تمام T شکاف زمانی توزیع شده و الگوریتم با انتساب آخرین گره، خاتمه می‌یابد. با تکرار زمان‌بندی به‌دست آمده با روش فوق می‌توان این روش را به حالتی که دوره جمع‌آوری داده بزرگتر و مضرب صحیحی از عدد N/L است تعمیم داد. همچنین وقتی که این عدد صحیح نیست با نمونه‌برداری مجدد بعضی گره‌ها در شکاف‌های زمانی می‌توان مطمئن شد که دقیقاً تعداد L گره در هر شکاف برای نمونه‌برداری انتخاب می‌شود.

۳-۲- روش پیشنهادی

ما در این بخش برای تعیین زمان‌بندی نمونه‌برداری حسگرها روشی ارائه می‌کنیم که عملکرد بهتر خود را نسبت به روش ECB-DNS طبق نتایج عددی به‌دست آمده در بخش ۵ نشان داده است. روش جدید هم در انتخاب معیار و هم در نحوه انتخاب گره‌ها در شکاف‌های زمانی با روش ECB-DNS متفاوت است. معیار واریانس مشروط کل (m_j) و معیار معادل آن (m'_j) همچنان که در روابط (۱۰) و (۱۱) مشهود است علاوه بر همبستگی متقابل بین داده‌ها به مقدار واریانس آنها نیز وابسته است. پیروی این واقعیت چون واریانس داده‌ها نیز از مقیاس این داده‌ها تاثیر می‌پذیرد، گره‌هایی که داده‌های بزرگتری دارند اگرچه همبستگی مشابهی داشته باشند به‌خطا سهم بزرگتری را در معیارهای فوق بوجود می‌آورند. به همین دلیل ما به‌جای استفاده از معیار واریانس مشروط برای گره‌ها از کمیت اطلاعات متقابل^{۱۹} متغیرهای تصادفی X_i و X_j که بر مبنای مفهوم آنتروپی تعریف می‌شود استفاده خواهیم کرد:

$$I(X_i; X_j) \triangleq H(X_i) + H(X_j) - H(X_i, X_j) \quad (12)$$

این کمیت بر خلاف معیار استفاده شده در الگوریتم ECB-DNS نسبت به متغیرهای تصادفی X_i و X_j تغییر ناپذیر با مقیاس^{۲۰} است. بازم در اینجا فرض می‌کنیم که داده‌ها دارای توزیع توأم نرمال هستند ولی در نهایت الگوریتم به‌دست آمده را به‌منظور سنجش کارآمدی بر روی یک حالت عمومی به‌کار خواهیم برد. با این فرض که n متغیر تصادفی X_1, X_2, \dots, X_n دارای توزیع توأم گوسی باشند می‌توان نشان داد که آنتروپی توأم این متغیرهای تصادفی از رابطه زیر به‌دست می‌آید [۲۸]:

$$H(X_1, X_2, \dots, X_n) = \ln(\sqrt{(2\pi e)^n \Delta}) \quad (13)$$

در رابطه اخیر Δ دترمینان ماتریس کوواریانس متغیرهای تصادفی است. به این ترتیب برای دو متغیر تصادفی X_i و X_j کمیت اطلاعات متقابل به کمک روابط (۱۲) و (۱۳) به‌صورت زیر قابل محاسبه است:

$$I(X_i; X_j) = \ln(\sqrt{2\pi e \sigma_i^2}) + \ln(\sqrt{2\pi e \sigma_j^2}) - \ln(\sqrt{(2\pi e)^2 \sigma_i^2 \sigma_j^2 (1 - \rho_{ij}^2)}) = -0.5 \ln(1 - \rho_{ij}^2) \quad (14)$$

رابطه بالا که البته در حالت خاص توزیع گوسی برای کمیت اطلاعات متقابل به‌دست آمده است نشان می‌دهد که این کمیت از مقدار واریانس هر دو متغیر تصادفی X_i و X_j مستقل بوده و فقط به همبستگی متقابل آن دو وابسته است. در روش جدید که انتخاب گره مبتنی بر اطلاعات متقابل^{۲۱} (IB-DNS) نام‌گذاری می‌شود مانند آنچه در مورد روش ECB-DNS توضیح داده شد گره‌ها را به‌صورت گام به گام به شکاف‌های زمانی نسبت می‌دهیم و این فرایند را آنقدر تکرار می‌کنیم که دقیقاً تعداد L گره به هر شکاف زمانی اختصاص یابد. به این ترتیب که در اولین گام به‌جای معیار m_j برای انتخاب گره‌ها از کمیت c_j که به‌صورت زیر محاسبه می‌شود استفاده می‌کنیم:

$$c_j \triangleq \sum_{\substack{i=1 \\ i \neq j}}^N I(X_i; X_j) = \sum_{\substack{i=1 \\ i \neq j}}^N (-0.5 \ln(1 - \rho_{ij}^2)) = -0.5 \ln \left(\prod_{\substack{i=1 \\ i \neq j}}^N (1 - \rho_{ij}^2) \right) \quad (15)$$

در حقیقت c_j مرتبط با کل اطلاعاتی است که گره j در باره تک تک گره‌ها ارائه می‌دهد. با توجه به اینکه تابع لگاریتم تابعی اکیدا یکنوا است می‌توان به جای جستجوی بیشترین مقدار کمیت c_j در بین گره‌های حسگر کمترین مقدار کمیت معادل زیر را جستجو کرد:

$$c'_j = \prod_{\substack{i=1 \\ i \neq j}}^N (1 - \rho_{ij}^2) \quad (16)$$

در الگوریتم ECB-DNS از گام دوم به بعد وقتی در یک شکاف زمانی به دنبال بهترین گره می‌گردیم، اطلاعاتی را که گره‌های انتخاب شده قبلی در آن شکاف نسبت به گره‌های انتخاب نشده فعلی در اختیار می‌گذارند در نظر نمی‌گیریم. در الگوریتم جدید این نقص به این ترتیب برطرف شده است که فرض می‌کنیم در گام n ($n > 1$) و در شکاف زمانی $t \in \{1, 2, \dots, T\}$ گره‌های انتخابی قبلی به ترتیب $s^t(1)$ ، $s^t(2)$ ، ... و $s^t(n-1)$ باشند، برای بیان میزان اطلاعاتی که گره‌های انتخابی قبلی به همراه گره j در باره گره i دلخواه می‌گذارند از کمیتی که در زیر تعریف شده استفاده می‌کنیم:

$$I(X_i; X_j, X_{s^t(1)}, X_{s^t(2)}, \dots, X_{s^t(n-1)}) \triangleq H(X_i) + H(X_j, X_{s^t(1)}, X_{s^t(2)}, \dots, X_{s^t(n-1)}) - H(X_i, X_j, X_{s^t(1)}, X_{s^t(2)}, \dots, X_{s^t(n-1)}) \\ = \ln(\sigma_i \sqrt{2\pi e}) + \ln(\sqrt{(2\pi e)^n \delta_{i,j}^{n,t}}) - \ln(\sqrt{(2\pi e)^{n+1} \Delta_{i,j}^{n,t}}) = -0.5 \ln \left(\frac{\Delta_{i,j}^{n,t}}{\sigma_i^2 \delta_{i,j}^{n,t}} \right) = -0.5 \ln \left(\frac{\Delta'_{i,j}^{n,t}}{\delta'_{i,j}^{n,t}} \right) \quad (17)$$

$\Delta_{i,j}^{n,t}$ و $\Delta'_{i,j}^{n,t}$ به ترتیب دترمینان ماتریس کوواریانس و ماتریس همبستگی داده‌های $X_{s^t(1)}$ ، $X_{s^t(2)}$ ، ... و $X_{s^t(n-1)}$ به انضمام دو مقدار سنسش شده دیگر X_i و X_j است. $\delta_{i,j}^{n,t}$ و $\delta'_{i,j}^{n,t}$ نیز به ترتیب دترمینان ماتریس‌هایی است که پس از حذف سطر و ستون مربوط به داده X_i از دو ماتریس کوواریانس و همبستگی فوق‌الذکر حاصل می‌شوند. نتیجه به دست آمده در رابطه (۱۷) بازهم استقلال کمیت مورد استفاده ما را از واریانس متغیرها و در نتیجه مقیاس داده گره‌ها نشان می‌دهد. به این ترتیب در الگوریتم جدید با کمک کمیت اطلاعات متقابل که در رابطه (۱۷) محاسبه شد کمیت c_j را اصلاح و به صورت زیر باز تعریف می‌کنیم:

$$c_j \triangleq \sum_{\substack{i \in S_1(n,t) \\ i \neq j}} I(X_i; X_j, X_{s^t(1)}, X_{s^t(2)}, \dots, X_{s^t(n-1)}) = \sum_{\substack{i \in S_1(n,t) \\ i \neq j}} \left(-0.5 \ln \left(\frac{\Delta'_{i,j}^{n,t}}{\delta'_{i,j}^{n,t}} \right) \right) = -0.5 \ln \left(\prod_{\substack{i \in S_1(n,t) \\ i \neq j}} \left(\frac{\Delta'_{i,j}^{n,t}}{\delta'_{i,j}^{n,t}} \right) \right) \quad (18)$$

$S_1(n,t)$ مجموعه تمام گره‌هایی است که تا شکاف زمانی $t \in \{1, 2, \dots, T\}$ و گام $n \in \{1, 2, \dots, L\}$ هنوز انتخاب نشده‌اند. می‌توان به جای پیدا کردن ماکزیمم c_j از کمیت معادل آن c'_j که به صورت زیر تعریف می‌شود استفاده کرد:

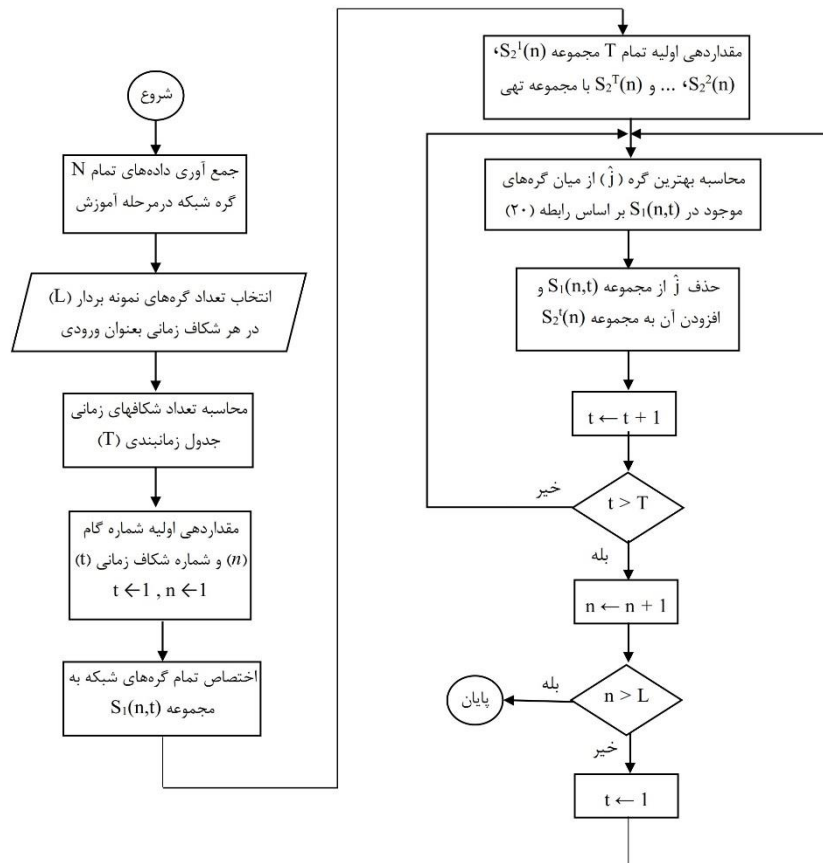
$$c'_j \triangleq \frac{\prod_{\substack{i \in S_1(n,t) \\ i \neq j}} \Delta'_{i,j}^{n,t}}{\prod_{\substack{i \in S_1(n,t) \\ i \neq j}} \delta'_{i,j}^{n,t}} \quad (19)$$

بنابراین گره بهینه \hat{j} که بیشترین اطلاعات را در مورد گره‌های j حسگر انتخاب نشده در اختیار می‌گذارد به صورت زیر به دست می‌آید:

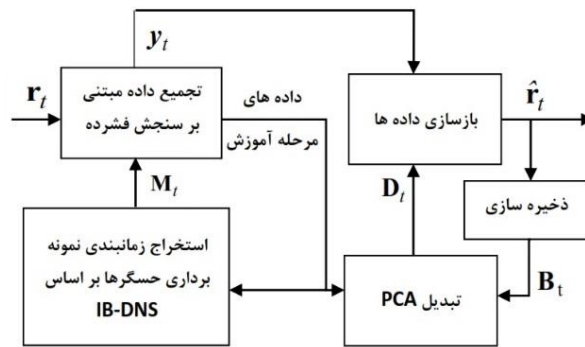
$$\hat{j} = \operatorname{argmin}_{j \in S_1(n,t)} \left(\frac{\prod_{\substack{i \in S_1(n,t) \\ i \neq j}} \Delta'_{i,j}^{n,t}}{\prod_{\substack{i \in S_1(n,t) \\ i \neq j}} \delta'_{i,j}^{n,t}} \right) \quad (20)$$

همچنین فرض می‌کنیم $S_2^1(n)$, $S_2^2(n)$, ... و $S_2^T(n)$ تعداد T مجموعه هستند که هر یک شامل گره‌های انتخاب شده تا گام n در تمام T شکاف زمانی هستند، به طوریکه $S_2^t(n) = \{s^t(1), s^t(2), \dots, s^t(n-1)\}$. در این صورت الگوریتم IB-DNS را می‌توان به صورت زیر ارائه کرد:

۱. در اولین گام و اولین شکاف زمانی ($t=1$ و $n=1$) مجموعه $S_1(n,t)$ با مجموعه تمام N گره شبکه حسگر و نیز همه مجموعه‌های $S_2^t(n)$ که $t \in \{1, 2, \dots, T\}$ است با مجموعه تهی مقدار دهی اولیه می‌شوند.
 ۲. در هر گام جدید بهترین گره‌ها را بر اساس رابطه (۲۰) برای تمام شکاف‌های زمانی از بین اعضاء مجموعه $S_1(n,t)$ انتخاب کرده و به ترتیب به مجموعه‌های $S_2^1(n)$, $S_2^2(n)$, ... و $S_2^T(n)$ اختصاص می‌دهیم. هر بار که گره‌ای را به یکی از این مجموعه‌ها اضافه می‌کنیم آن گره را از مجموعه $S_1(n,t)$ حذف می‌کنیم.
 ۳. اگر به آخرین گام رسیدیم ($n=L$)، انتخاب گره‌ها را خاتمه می‌دهیم در غیر این صورت به مرحله ۲ می‌رویم.
- الگوریتم IB-DNS در شرایطی که دوره جمع‌آوری داده در رابطه $T=N/L$ صدق کند با اختصاص دقیقاً L گره به هر شکاف زمانی خاتمه می‌یابد. مراحل این الگوریتم به صورت فلوجارت در شکل (۱) نشان داده شده است. در وضعیت‌هایی که رابطه بالا برقرار نباشد برای انتخاب گره‌ها مانند آنچه که در روش ECB-DNS پیشنهاد شده است عمل می‌کنیم. با هدف مشخص کردن ارتباط الگوریتم IB-DNS با اجزای دیگر شرکت کننده در فرایند تجمیع داده مبتنی بر CS و یادگیری لغت نام PCA، جایگاه این بخش به صورت بلوک دیاگرام در شکل (۲) به تصویر کشیده شده است.



شکل (۱): فلوجارت استخراج زمانبندی نمونه برداری حسگرها با روش پیشنهادی IB-DNS
Figure (1): Flowchart of the proposed IB-DNS method for extraction of sensor sampling schedule



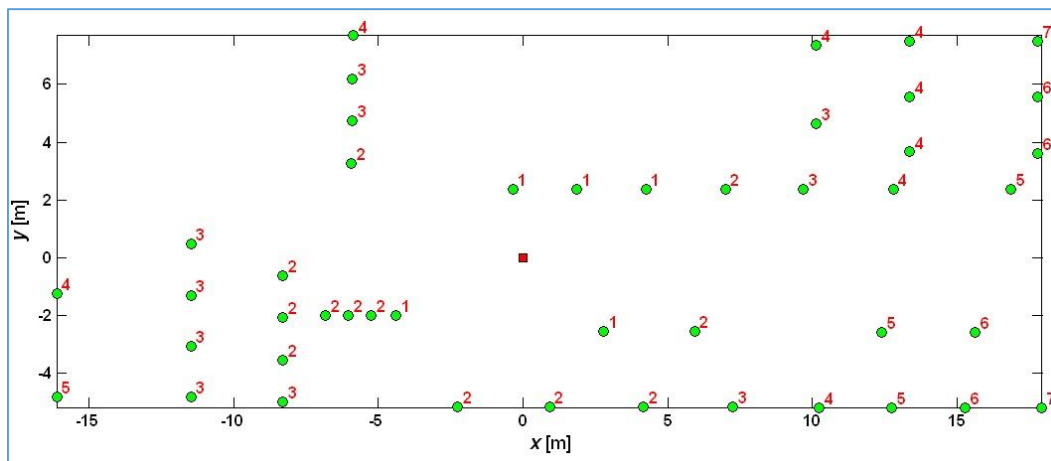
شکل (۲): بلوک دیاگرام جایگاه و ارتباطات بخش زمانبندی IB-DNS با دیگر اجزای سیستم تجمیع داده
Figure (2): Location and connections of the IB-DNS scheduling module with other components of the aggregated data system

۴- فرضیه‌های تحقیق

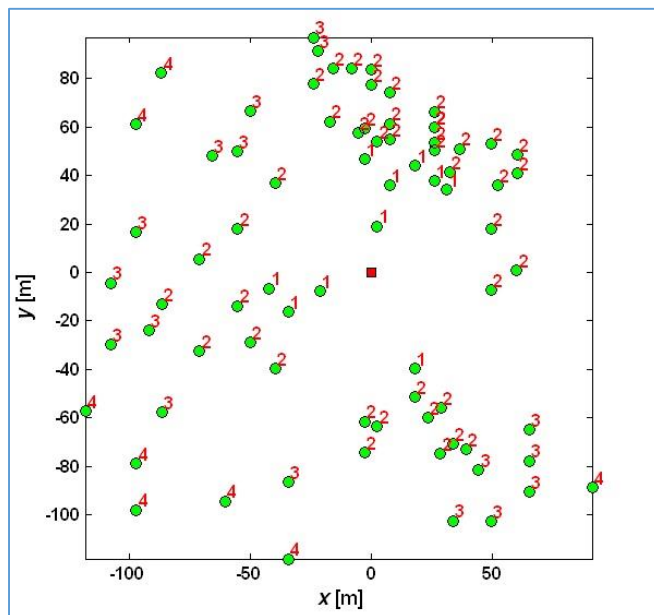
در این بخش سیگنال‌ها و شبکه‌های حسگر مورد آزمایش و مدل مصرف انرژی توصیف خواهد شد.

۴-۱- توصیف سیگنال‌ها و شبکه‌های مورد آزمایش

داده‌های مورد آزمایش، سیگنال‌های واقعی جمع‌آوری شده از تجهیزات شبکه‌های حسگر بی‌سیم مستقر در دو دانشگاه پادوا و لوزان هستند. این داده‌ها علاوه بر اینکه در شبکه‌هایی با دو مقیاس متفاوت جمع‌آوری شده‌اند دارای مشخصات آماری متفاوت و همبستگی‌های متنوع مکانی و زمانی هستند. اولی تاریخچه‌ای از سیگنال‌های محیطی ثبت شده است که توسط تعداد ۴۵ گره حسگر بی‌سیم مستقر در طبقه همکف ساختمان مهندسی فن‌آوری اطلاعات دانشگاه پادوا [۱۹] جمع‌آوری شده و دومی توسط تعداد ۸۱ گره حسگر بی‌سیم در فضای باز پردیس دانشگاه لوزان [۲۹] جمع‌آوری شده است. هر دو مجموعه توسط گره‌های حسگر بی‌سیم مجهز به فرستنده-گیرنده رادیویی سازگار با استاندارد IEEE802.15.4 ثبت شده‌اند. نحوه قرارگیری گره‌های حسگر در شبکه حسگر بی‌سیم اول و دوم به ترتیب در شکل‌های (۳) و (۴) نشان داده شده است. در این شکل‌ها موقعیت گره چاهک و تعداد پرش لازم برای ارسال بسته‌های داده از هر گره به گره چاهک بر اساس مسیریابی پیشبرد جغرافیایی^{۲۲} که در هر دو شبکه مورد استفاده قرار گرفته، نشان داده شده است. زیرمجموعه‌ای از داده‌های جمع‌آوری شده در این دو شبکه که برای شبیه‌سازی استفاده شده‌اند عبارتند از: دمای محیط، در صد رطوبت، شدت روشنایی و جهت وزش باد. کوور و دیگران [۱۹] با قالب‌بندی داده‌های خام حسگرها در چارچوب‌های زمانی ۵ دقیقه‌ای این داده‌ها را در فایل‌هایی با ساختار قابل استفاده در نرم‌افزار متلب مرتب کرده‌اند.



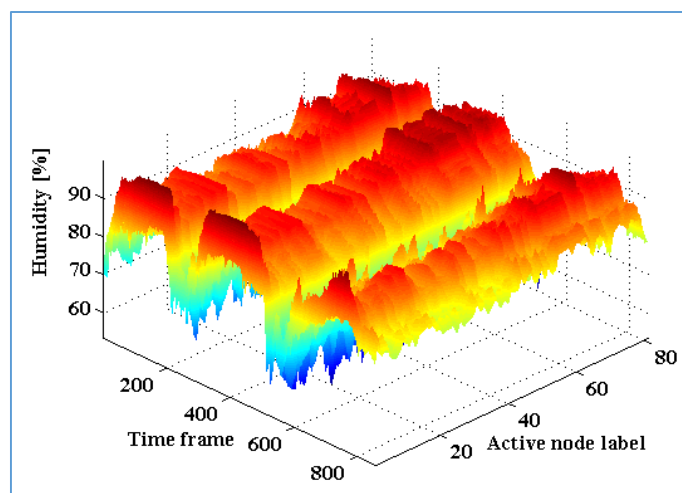
شکل (۳): نحوه قرارگیری گره‌های حسگر در شبکه حسگر بی‌سیم دانشگاه پادوا
Figure (3): Wireless Sensor Network deployment spotting of PADOVA university



شکل (۴): نحوه قرارگیری گره‌های حسگر در شبکه حسگر بی‌سیم دانشگاه لوزان

Figure (4): Wireless Sensor Network deployment spotting of EPFL

علاوه بر سیگنال‌های جمع‌آوری شده هر حسگر، اطلاعات موقعیت هر حسگر و تعداد پرش مربوط به مسیریابی هر حسگر تا گره چاهک نیز در این فایل‌ها ارائه شده است. در شکل (۵)، به‌عنوان نمونه، سیگنالی با تغییرات آرام مشخصات آماری که دارای همبستگی شدید مکانی است از داده‌های شبکه دوم نشان داده شده است.



شکل (۵): نمایش تصویری داده‌های واقعی سنسور شده مربوط به رطوبت محیط

Figure (5): Representation of real measured data related to humidity

۴-۲- مدل انرژی مصرفی

برای تعیین کل انرژی مصرفی شده، هزینه اعمال انجام شده در هر حسگر را به دو بخش هزینه محاسبات^{۲۳} و هزینه انتقال داده^{۲۴} تفکیک می‌کنیم. هزینه محاسبات شامل هزینه اعمالی است که توسط بخش پردازنده حسگر اجرا می‌شود. این هزینه با توجه به روش نمونه‌برداری که در بخش ۳ ذکر شد شامل یک عمل مقایسه در پردازنده حسگر است. علاوه بر عمل مقایسه چنانچه در حسگر عمل سنسور نیز انجام شود باید مصرف انرژی مربوط به فرآیند اکتساب داده^{۲۵} در هزینه محاسبات گنجانده شود. هزینه انتقال داده مربوط است به ارسال با پرش چند گانه^{۲۶} داده‌های سنسور شده از گره‌ها به چاهک و ارسال داده‌های

جدول زمان‌بندی از چاهک به گره‌ها در جهت عکس. برای تبدیل هزینه‌های مذکور به اعدادی برحسب واحد انرژی ژول، سخت‌افزار به کار رفته در گره‌های حسگر TmoteSky [۳۰] را مبنا قرار داده و از روش توصیف شده در مرجع [۳۱] استفاده می‌کنیم.

۴-۲-۱- هزینه محاسبات

ابتدا تعداد اعمال مقایسه انجام شده در تمام گره‌های حسگر شبکه شمارش می‌شود. سپس عدد به دست آمده در مدت زمان لازم برای اجرای یک عمل مقایسه (TCMP) ضرب شده تا مجموع زمان محاسبات در گره‌ها به دست آید. در صورتی که عمل سنجش نیز توسط یک یا چند گره حسگر انجام شود باید مدت زمان لازم برای تکمیل فرآیند اکتساب داده (T_{ACQ}) را به زمان محاسبات در آن گره‌ها افزود. در نهایت ما کل زمان محاسبات را در توان مصرفی یک حسگر که واحد میکروکنترلر آن فعال و بخش رادیویی آن غیرفعال است (P_{UC}) ضرب می‌کنیم تا هزینه محاسبات برحسب ژول به دست آید.

۴-۲-۲- هزینه انتقال داده

در انتقال یک بیت به صورت تک پرش، بخش رادیویی دو گره شرکت‌کننده در ارتباط یکی در وضعیت فرستنده و دیگری در وضعیت گیرنده انرژی مصرف می‌کند. انرژی مصرفی در گره فرستنده از حاصل ضرب زمان ارسال یک بیت (T_{bit}) در توان مصرفی مربوط به وضعیت فرستنده (P_{TX}) و انرژی مصرفی در گره گیرنده از حاصل ضرب همین زمان در توان مصرفی مربوط به وضعیت گیرنده (P_{RX}) به دست می‌آید. در نتیجه ما با جمع این دو، انرژی مصرفی مربوط به انتقال یک بیت در یک پرش (E_{bit}) را برآورد می‌کنیم. با توجه به اینکه سخت‌افزار به کار رفته در TmoteSky شامل واحد میکروکنترلری از نوع MSP430 [۳۲] و بخش فرستنده-گیرنده رادیویی از نوع CC2420 [۳۳] است، پارامترهای مدل انرژی مصرفی که در بالا توصیف شد با استفاده از برگه‌های اطلاعات فنی این سخت‌افزارها استخراج شده و در جدول (۱) نمایش داده شده است.

۵- نتایج آزمایش‌ها

در این بخش ما به مقایسه روش‌های نمونه‌برداری قطعی IB-DNS و ECB-DNS در فرآیند تجمیع داده مبتنی بر روش مشترک CS و یادگیری لغت‌نامه مبتنی بر PCA می‌پردازیم. روش تولید مصنوعی سیگنال‌های محیطی دارای همبستگی‌های مکانی-زمانی که در کار هوشمند و دیگران [۲۰] مورد استفاده قرار گرفته به دلیل اینکه سیگنالی ایستاد تولید می‌کند نمی‌تواند بیانگر مشخصات آماری متغیر با زمان سیگنال‌های محیطی واقعی باشد. به همین دلیل ما برای بررسی نتایج از شبکه‌های حسگر بی‌سیم واقعی و سیگنال‌های محیطی جمع‌آوری شده در آنها استفاده می‌کنیم.

۵-۱- مشخصات شبیه‌سازی

ما آزمایش‌هایی را با انتخاب دو سیگنال از هر یک از دو شبکه ذکر شده در بخش ۴-۱ انجام داده‌ایم، یکی با تغییرات سریع و دیگری با تغییرات آرام از لحاظ آماری. ما میانگین خطای بازسازی را در یک دوره جمع‌آوری داده که به صورت رابطه زیر تعریف می‌شود محاسبه می‌کنیم:

$$\varepsilon = \frac{\sum_{t=1}^T \|r_t - \hat{r}_t\|_2}{T} \quad (21)$$

به طوری که r_t سیگنال اصلی N گره حسگر در لحظه t است و \hat{r}_t نسخه بازسازی شده آن در چاهک است که از روش مهمی و دیگران [۲۱] به دست می‌آید (نگاه کنید به بخش ۲-۱). علاوه بر این در شبیه‌سازی‌ها، میانگین انرژی مصرف شده در یک شکاف زمانی در هر گره (E_{ave}) با در نظر گرفتن دو هزینه انتقال داده و محاسبات برآورد می‌شود. در ارتباط با هزینه انتقال داده و محاسبات در شبیه‌سازی‌ها فرض بر این است که داده‌های سنجش شده در حسگرها توسط ۱۶ بیت کوانتیزه می‌شود و

همچنین طول دوره جمع‌آوری داده برابر $T=100$ شکاف زمانی است. نتایج با اجرای ۱۰۰ بار شبیه‌سازی روی داده‌های واقعی و با میانگین‌گیری خطای بازسازی دوره جمع‌آوری داده ارائه شده‌است.

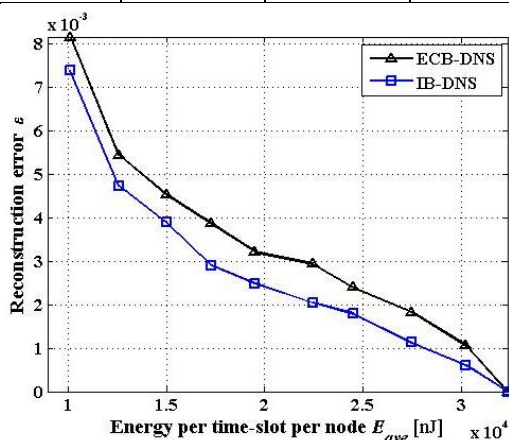
۵-۲- نمودارهای عملکرد

به‌منظور مقایسه کمی عملکرد دو روش انتخاب‌گره با یکدیگر ما منحنی‌های مصالحه‌ای^{۲۷} را رسم کرده‌ایم که تغییرات میانگین خطای بازسازی (ϵ) را بر حسب میانگین انرژی مصرفی (E_{ave}) نشان می‌دهد. در تمام شکل‌ها، منحنی‌های مصالحه با تغییر مقدار L به‌عنوان پارامتر آزاد در محدوده‌ای بین ۵ درصد تعداد گره‌های شبکه تا تعداد کل این گره‌ها به‌دست آمده‌است. تغییرات مقدار L در گام‌هایی با فواصل یکسان بوده و بگونه‌ای است که با جهت افزایش میانگین انرژی مصرفی (E_{ave}) هم‌خوانی دارد. منحنی‌هایی که به صورت ECB-DNS مشخص شده با استفاده از استخراج زمان‌بندی نمونه‌برداری حسگرها بر مبنای روش هوشمند و دیگران (بخش ۳-۱) به‌دست آمده درحالی‌که برای دیگری که با IB-DNS مشخص شده از روش نمونه‌برداری مبتنی بر اطلاعات متقابل (روش پیشنهادی) استفاده شده‌است. عملکرد روش‌های IB-DNS و ECB-DNS برای سیگنال‌هایی با تغییرات آرام در شبکه با مقیاس کوچک‌تر دانشگاه پادوا (با تعداد ۴۵ گره حسگر) و شبکه با مقیاس بزرگ‌تر دانشگاه لوزان (با تعداد ۸۱ گره حسگر) به ترتیب در شکل‌های (۶) و (۷) نشان داده شده‌است. همچنین مقایسه روش‌های نمونه‌برداری حسگرها در همین دو شبکه به ترتیبی که در بالا ذکر شد، برای سیگنال‌هایی با تغییرات سریع در شکل‌های (۸) و (۹) به تصویر کشیده شده‌است. مقایسه بین عملکردهای IB-DNS و ECB-DNS نشان می‌دهد که در مورد هر دو مجموعه داده‌های واقعی مورد آزمایش وقتی تمام پارامترها بجز روش انتخاب‌گره‌های حسگر یکسان است، تجمیع داده مبتنی بر CS با استفاده از روش اول بر روش دوم برتری دارد. بهبود عملکرد نتیجه اثر بهتر استفاده از روش IB-DNS نسبت به روش ECB-DNS بر فرآیند یادگیری لغت‌نامه مبتنی بر PCA است. در مورد هر چهار شکل اخیر این نکته قابل ذکر است که زمانی که تعداد گره‌های انتخاب شده با تعداد کل گره‌های شبکه برابر می‌شود، انتخاب روش نمونه‌برداری بر عملکرد فرایند تجمیع بی‌اثر خواهد شد و منحنی‌های عملکرد یکدیگر را قطع می‌کنند. از آنجایی که در عمل انواع مختلفی از سیگنال‌های محیطی توسط یک شبکه حسگر بی‌سیم نمونه‌برداری می‌شود به‌منظور مطالعه عملکرد مقرون به صرفه^{۲۸} روش پیشنهادی، به‌طور میانگین برای هر چهار نوع سیگنال مورد آزمایش، ما میزان کاهش مصرف انرژی را که استفاده از روش IB-DNS نسبت به روش ECB-DNS نتیجه می‌دهد به‌صورت نمودار میله‌ای در شکل (۱۰) رسم کرده‌ایم.

Table (1): parameters of energy consumption model

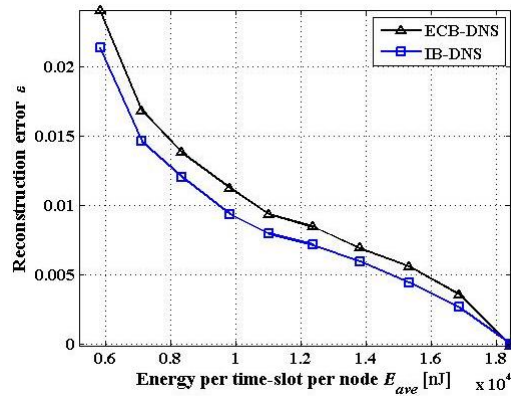
جدول (۱): پارامترهای مدل انرژی مصرفی

$P_{\mu C}$	P_{TX}	P_{RX}	f_{CPU}	τ_{bit}	ϵ_{bit}	τ_{CMP}	τ_{ACQ}
$\Delta/4mW$	$\Delta\Delta/5mW$	$\epsilon\Delta/4mW$	$1/0.48MHz$	$4\mu s$	$495/6nJ$	$\Delta/7\mu s$	$30\mu s$



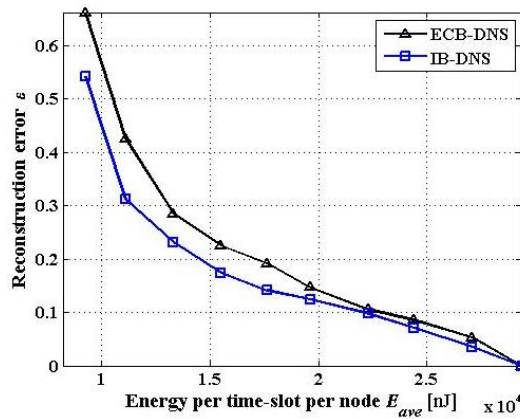
شکل (۶): مقایسه تاثیر دو روش نمونه‌برداری حسگرها بر عملکرد فرآیند تجمیع داده در شرایط استفاده از سیگنال با تغییرات آرام دمایی محیط

Figure (6): Performance comparison of two different data aggregation approach for slowly changing temperature signal



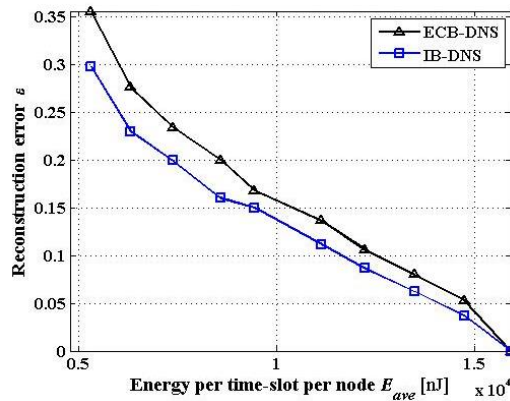
شکل (۷): مقایسه تاثیر دو روش نمونه‌برداری حسگرها بر عملکرد فرآیند تجمیع داده در شرایط استفاده از سیگنال با تغییرات آرام رطوبت

Figure (7): Performance comparison of two different data aggregation approach for slowly changing humidity signal



شکل (۸): مقایسه تاثیر دو روش نمونه‌برداری حسگرها بر عملکرد فرآیند تجمیع داده در شرایط استفاده از سیگنال با تغییرات سریع شدت روشنایی

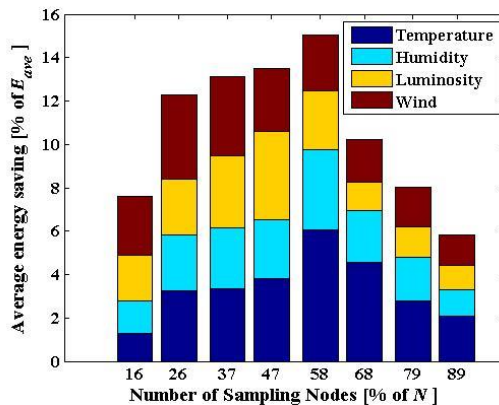
Figure (8): Performance comparison of two different data aggregation approach for fast-changing luminosity signal



شکل (۹): مقایسه تاثیر دو روش نمونه‌برداری حسگرها بر عملکرد فرآیند تجمیع داده در شرایط استفاده از سیگنال با تغییرات سریع جهت وزش باد

Figure (9): Performance comparison of two different data aggregation approach for fast-changing wind direction signal

در این شکل تغییرات میانگین انرژی صرفه‌جویی شده^{۲۹} در هر حسگر بر حسب تعداد حسگرهای نمونه‌بردار (پارامتر L) به نمایش گذاشته شده است. با توجه به این شکل می‌توان نتیجه گرفت که در محدوده‌ای از منحنی‌های مصالحه [شکل‌های (۶) تا (۹)] که مقدار L تقریباً بین ۲۶ تا ۶۸ درصد تعداد کل گره‌های حسگر شبکه تغییر می‌کند.



شکل (۱۰): تغییرات میانگین انرژی صرفه جویی شده در هر حسگر با استفاده از روش نمونه برداری IB-DNS نسبت به ECB-DNS
Figure (10): Average energy saving results from applying the IB-DNS approach compared to the ECB-DNS one

بهبود عملکرد با روش نمونه برداری جدید نسبت به بخش‌های دیگر منحنی‌ها قابل توجه است به طوری که حداقل ۱۰ درصد صرفه جویی در مصرف انرژی حاصل خواهد شد. خارج از محدوده فوق در چهار شکل مذکور و در جایی که مقدار متوسط انرژی مصرفی زیادتر از بقیه نواحی است به دلیل نزدیک شدن مقدار L به تعداد کل گره‌های موجود در شبکه، مطابق انتظار، نوع نمونه برداری کم کم اثر خود را در بهبود عملکرد از دست می‌دهد و دو روش نمونه برداری از لحاظ مصرف انرژی با یکدیگر قابل مقایسه می‌شوند.

۶- نتیجه گیری

در این مقاله یک روش جدید برای استخراج جدول زمان بندی نمونه برداری حسگرهای شبکه بی سیم در شکاف‌های زمانی دوره جمع آوری داده ارائه شده است که IB-DNS نام گرفته است. چاهک با محاسبه کمیت اطلاعات متقابل داده‌های خوانده شده از کل حسگرهای شبکه در بخش ابتدایی هر دوره جمع آوری داده این جدول زمانی را به دست آورده و در کل شبکه پخش می‌کند. مقایسه روش جدید نمونه برداری با روش مرجع به کمک نتایج شبیه سازی‌ها بهبود قابل توجهی را در خطای سیگنال بازسازی شده در چاهک، خصوصاً برای سیگنال‌های با تغییرات آرام، نشان می‌دهد. در حالت کلی نیز که انواع مختلف سیگنال‌های محیطی در شبکه نمونه برداری می‌شوند، زمانی که تعداد گره‌های نمونه برداری شده تقریباً بین ۲۵ تا ۶۰ درصد کل گره‌ها تغییر می‌کند، میانگین صرفه جویی انرژی مصرفی در حسگرها کمتر از ۱۰ درصد نخواهد بود.

References

مراجع

- [1] D. Lin, W. Min, J. Xu, "An energy-saving routing integrated economic theory with compressive sensing to extend the lifespan of WSNs", IEEE Internet of Things Journal, vol. 7, pp. 7636-7647, Aug. 2020 (doi: 10.1109/JIOT.2020.2987354).
- [2] B. Sun, Y. Guo, N. Li, D. Fang, "Multiple target counting and localization using variational bayesian em algorithm in wireless sensor networks", IEEE Trans. on Communications, vol. 65, no. 7, pp. 2985-2998, July 2017 (doi: 10.1109/TCOMM.2017.2695198).
- [3] S. P. Tirani, A. Avokh, S. Azar, "WDAT-OMS: A two-level scheme for efficient data gathering in mobile-sink wireless sensor networks using compressive sensing theory", IET Communications, vol. 14, pp. 1826-1837, July 2020 (doi: 10.1049/iet-com.2019.0433).
- [4] M. Hasan Hoseini, F. Mesrinejad, H. Mahdavi-Nasab, "A routing method for tracking a moving target with reduced energy consumption", Journal of Intelligent Procedures in Electrical Technology, vol. 11, no. 43, pp. 29-47, Autumn 2020 (in Persian).
- [5] A. Bion, C. Pielli, M. Rossi, A. Zanella, D. Zordan, M. Kelly, M. Zorzi, "EC-CENTRIC: An energy-and context-centric perspective on IoT systems and protocol design", IEEE Access, vol. 5, pp. 6894-6908, Apr. 2017 (doi: 10.1109/ACCESS.2017.2692522).

- [6] S. Pakdaman Tirani, A. Avokh, "Impact of sink node placement onto wireless sensor networks performance regarding clustering routing and compressive sensing theory", *Journal of Intelligent Procedures in Electrical Technology*, vol. 7, no. 25, pp. 41-54, Spring 2016 (in Persian).
- [7] J. Huang and B. Soong, "Cost-aware stochastic compressive data gathering for wireless sensor networks", *IEEE Trans. on Vehicular Technology*, vol. 68, no. 2, pp. 1525-1533, Feb. 2019 (doi: 10.1109/TVT.2018.2-887091).
- [8] Y. Ma, X. Zhang, Y. Gao, "Joint sub-Nyquist spectrum sensing scheme with geo-location database over TV white space", *IEEE Trans. on Vehicular Technology*, vol. 67, no. 5, pp. 3998-4007, May 2018 (doi: 10.1109/TVT.2017.2776560).
- [9] X. Zhang, Y. Ma, Y. Gao, W. Zhang, "Autonomous compressive-sensing-augmented spectrum sensing", *IEEE Trans. on Vehicular Technology*, vol. 67, no. 8, pp. 6970-6980, Aug. 2018 (doi: 10.1109/TVT.2018.28-22776).
- [10] J. C. Ye, "Compressed sensing MRI: a review from signal processing perspective", *BMC Biomedical Engineering*, vol. 1, no. 1, pp. 1-17, Mar. 2019 (doi:10.1186/s42490-019-0006-z).
- [11] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, "A simple proof of the restricted isometry property for random matrices", *Constructive Approximation*, vol. 28, no. 3, pp. 253-263, Dec. 2008 (doi:10.1007/s00365-007-9003-x).
- [12] W. U. Bajwa, J. D. Haupt, A. M. Sayeed, R. D. Nowak, "Joint source-channel communication for distributed estimation in sensor networks", *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3629-3653, Oct. 2007 (doi: 10.1109/TIT.2007.904835).
- [13] C. Luo, F. Wu, J. Sun, C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks", *Proceedings of the ACM/MobiCom*, pp. 145-156, Beijing, China, Sept. 2009 (doi: 10.1145/1614320.16143-37).
- [14] L. Xiang, J. Luo, A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks", *Proceedings of the IEEE/SECON*, pp. 46-54, Salt Lake City, UT, USA, June 2011 (doi: 10.1109/SAHCN.2011.5984932).
- [15] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, A. Ortega, "Compressed sensing and routing in multi-hop networks", *Technical Report*, University of Southern California, 2009.
- [16] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, M. Zorzi, "On the interplay between routing and signal representation for compressive sensing in wireless sensor networks", *Proceeding of the IEEE/ITA Workshop*, pp. 206-215, La Jolla, CA, USA, Feb. 2009 (doi: 10.1109/ITA.2009.5044947).
- [17] W. Wang, M. Garofalakis, K. Ramchandran, "Distributed sparse random projections for refinable approximation", *Proceedings of the IEEE/IPSAN*, pp. 331-339, Cambridge, MA, USA, Apr. 2007 (doi: 10.1109/IPSAN.2-007.4379693).
- [18] R. Rana, W. Hu, C.T. Chou, "Energy-Aware Sparse approximation Technique (EAST) for rechargeable wireless sensor networks", *Proceedings of the EWSN*, pp. 306-321, University of Coimbra, Coimbra, Portugal, Feb. 2010 (doi:10.1007/978-3-642-11917-0_20).
- [19] G. Quer, R. Masiero, G. Pilonetto, M. Rossi, M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework", *IEEE Trans. on Wireless Communications*, vol. 11, no. 10, pp. 3447-3461, Oct. 2012 (doi: 10.1109/TWC.2012.081612.110612).
- [20] M. Hooshmand, M. Rossi, D. Zordan, M. Zorzi, "covariogram-based compressive sensing for environmental wireless sensor networks", *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1716-1729, Mar. 2016 (doi: 10.1109/JS-EN.2015.2503437).
- [21] E.J. Candès, J. Romberg, T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006 (doi: 10.1109/TIT.2005.862083).
- [22] H. Mohimani, M. Babaie-Zadeh, C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm", *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 289-301, Jan. 2009 (doi: 10.1109-9/TSP.2008.2007606).
- [23] I. Jolliffe, *Principal component analysis*: Springer, 2011.
- [24] D. Zordan, G. Quer, M. Zorzi, M. Rossi, "Modeling and generation of space-time correlated signals for sensor network fields", *Proceedings of the IEEE/GLOBECOM*, pp. 1-6, Houston, TX, USA, Dec. 2011 (doi: 10.1109/GLOCOM.2011.6133891).
- [25] M.C. Vuran, I.F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks", *IEEE/ACM Trans. on Networking*, vol. 14, no. 2, pp. 316-329, April 2006 (doi: 10.1109/TN-ET.2006.872544).
- [26] R. Cristescu, M. Vetterli, "On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks", *Proceedings of the IEEE/IPSAN*, pp. 159-164, Boise, ID, USA, Apr. 2005 (doi: 10.1109/IP-SN.2005.1440918).

- [27] A. Deshpande, C. Guestrin, S.R. Madden, J.M. Hellerstein, W. Hong, "Model-driven data acquisition in sensor networks", Proceedings of the VLDB, vol. 30, pp. 588-599, Toronto, Canada, Aug. 2004.
- [28] T.M. Cover, J.A. Thomas, Elements of information theory: John Wiley & Sons, 2012.
- [29] Sensorscope Data [Online]. Available: <https://zenodo.org/record/2654726/files/Sensorscope.zip>, Apr. 2019 (doi:10.5281/zenodo.2654726)
- [30] Moteiv Corporation, "TmoteSky datasheet", 2006.
- [31] Y. Liang, W. Peng, "Minimizing energy consumptions in wireless sensor networks via two-modal transmission", ACM/SIGCOMM Comput. Commun. Review, vol. 40, no. 1, pp. 12-18, Jan. 2010 (doi: 10.1145/1672308.1672311).
- [32] Texas Instruments, "MSP430x15x, MSP430x16x, SP430x161x mixed signal microcontroller", 2006.
- [33] Texas Instruments, "CC2420, single-chip 2.4GHz IEEE802.15.4 compliant and Zigbee (TM) ready RF transceiver", 2007.

زیرنویس‌ها

1. Environmental signals
2. Compressive sensing
3. Spectrum sensing
4. Sensing matrix
5. Downsampling
6. Principal component analysis
7. Adaptive
8. Non-lattice
9. Enhanced correlation based deterministic node selection
10. Training phase
11. Payload
12. Underdetermined system of equations
13. Eigenvalue decomposition

۱۴. چنانچه در شکاف‌های زمانی قبلی تمام یا برخی مقادیر واقعی داده گره‌ها در دسترس باشد از این مقادیر به‌جای مقادیر بازسازی شده استفاده می‌کنیم.

۱۵. در مورد سیگنال‌های غیر ایستادن مقدار H نباید بزرگ انتخاب شود، در مرجع [۱۹] نشان داده شده است که برای سیگنال‌های واقعی با انتخاب عدد ۲ بهترین نتایج حاصل می‌شود.

16. Enhanced correlation-based deterministic node selection
17. Computationally tractable
18. Overall conditional variance
19. Mutual information
20. Scale invariant
21. Information based deterministic node selection
22. Geographical forwarding
23. Transmission cost
24. Computation cost
25. Data acquisition
26. Multi-hop
27. Tradeoff curves
28. Achievable performance
29. Energy saving