



تشخیص ناهنجاری توزیع شده در شبکه های حسگر بی سیم مبتنی بر تحلیل مولفه های اصلی

محمد احمدی لیوانی، مهدی آبادی

تهران، دانشگاه تربیت مدرس، دانشکده مهندسی برق و کامپیوتر، گروه مهندسی کامپیوتر

ahmadi.l@modares.ac.ir, abadi@modares.ac.ir

چکیده

در این مقاله، یک رویکرد توزیع شده انرژی کارا برای تشخیص ناهنجاری در داده های جمع آوری شده توسط گره های حسگر در یک شبکه حسگر بی سیم پیشنهاد می شود که با انجام پردازش درون شبکه ای هزینه ارتباطاتی را کاهش می دهد. ناهنجاری در داده های جمع آوری شده ممکن است به علت خرابی گره ها یا وجود گره های سازش کار در شبکه باشد. در رویکرد پیشنهادی، شبکه حسگر بی سیم به تعدادی گروه تقسیم می شود. هر گروه دارای گره سرگروه می باشد که بردارهای داده را از سایر گره های عضو گروه دریافت می کند. فرآیند ایجاد نمای عادی بین تمام گره های عضو هر گروه توزیع می شود. هر گره به جای ارسال تمام بردارهای داده تنها توصیفی از این بردارها را برای گره سرگروه ارسال می کند. گره سرگروه، از تحلیل مولفه های اصلی توزیع شده (DPCA) و الگوریتم خوشه بندی پهنا ثابت (FWC) برای ایجاد نمای عادی سراسری و تشخیص ناهنجاری استفاده می کند. همچنین، با استفاده از ضرایب وزن دار حاصل از منحنی فراموشی نمای عادی سراسری را به روزرسانی می کند. نتایج آزمایش های انجام شده با استفاده از داده های جمع آوری شده توسط گره های حسگر در آزمایشگاه اینتل-برکلی نشان می دهد که رویکرد پیشنهادی از نرخ تشخیص قابل مقایسه ای نسبت به رویکرد تشخیص ناهنجاری متمرکز برخوردار می باشد، در حالی که هزینه ارتباطی را به طور قابل ملاحظه ای کاهش می دهد.

واژه های کلیدی

شبکه حسگر بی سیم، تشخیص ناهنجاری توزیع شده، تحلیل مولفه های اصلی، گره سازش کار.

از شبکه های حسگر بی سیم در کاربردهای مختلفی از قبیل نظارت بر محیط و میداين جنگی استفاده می شود. به دلیل طبیعت بحرانی این کاربردها، تشخیص ناهنجاری^۴ در داده های جمع آوری شده توسط گره های حسگر از اهمیت زیادی برخوردار است [۳]. داده های ناهنجار یا پیرامونی^۵ به داده هایی گفته می شوند که از سایر داده های جمع آوری شده انحراف قابل توجهی داشته باشند [۴]. ناهنجاری در داده های جمع آوری شده ممکن است به علت خرابی گره ها یا وجود گره های سازش کار^۶ در شبکه باشد. در شبکه های حسگر بی سیم نسبت مصرف انرژی برای

۱- مقدمه !

هر شبکه حسگر بی سیم (WSN)^۱ از تعداد زیادی گره حسگر^۲ تشکیل می شود که در یک محیط پراکنده شده و داده های مختلفی را جمع آوری می کنند. گره های حسگر در توان پردازشی، انرژی و پهناي باند^۳ دارای محدودیت هایی هستند [۱]. این محدودیت ها، شبکه های حسگر را در برابر حملات مختلفی در معرض خطر قرار می دهند [۲].

⁴ Anomaly detection

⁵ Outlier

⁶ Compromised node

¹ Wireless Sensor Network

² Sensor node

³ Bandwidth

را برای گره اصلی ارسال می‌کنند. گره اصلی مولفه‌های اصلی داده‌های دریافتی از تمام گره‌های فرعی را به دست آورده و داده‌های ناهنجار را با استفاده از روش زیرفضا^۵ تشخیص می‌دهد. در روش فوق، گره‌های فرعی تمام داده‌ها را به گره اصلی ارسال می‌کنند که این فرآیند دارای هزینه ارتباطی زیادی می‌باشد و طول عمر شبکه را به طور قابل ملاحظه‌ای کاهش می‌دهد.

Rajasegarar و همکارانش از خوشه‌بندی [۹] و SVM^۶ [۱۰] برای تشخیص ناهنجاری توزیع شده استفاده کرده‌اند. در روش مبتنی بر خوشه‌بندی، شبکه حسگر دارای ساختار سلسله مراتبی است. گره‌ها در پایان هر بازه زمانی داده‌های جمع‌آوری شده را خوشه‌بندی کرده و توصیفی از خوشه‌های ایجاد شده را برای گره والد خود ارسال می‌کنند. گره والد با دریافت اطلاعات فوق، خوشه‌هایی که با هم همپوشانی دارند را ادغام می‌کند. این رویه تا زمانی ادامه پیدا می‌کند که توصیف خوشه‌ها به گره سرگروه برسد. گره سرگروه با دریافت اطلاعات فوق، فاصله بین خوشه‌ها را محاسبه کرده و خوشه‌های ناهنجار را تشخیص می‌دهد. روش فوق تنها در صورتی قادر به تشخیص ناهنجاری است که تعداد خوشه‌های ناهنجار از تعداد خوشه‌های عادی خیلی کمتر باشد، در غیر این صورت نرخ تشخیص به شدت کاهش پیدا می‌کند. در روش مبتنی بر SVM، گره سرگروه یک ربع ابرکره سراسری را پیدا کرده و برای تمام گره‌های عضو گروه ارسال می‌کند. گره‌های عضو داده‌هایی که خارج از ابرکره قرار می‌گیرند را به عنوان ناهنجاری تشخیص می‌دهند. شعاع ابرکره سراسری به صورت توزیع شده و با همکاری تمام گره‌های عضو گروه محاسبه می‌شود. روش فوق دارای سربار محاسباتی زیادی است.

۳- بیان مساله

شبکه حسگر بی‌سیم از تعدادی گره حسگر $N = \{s_i; i=1 \dots s\}$ تشکیل شده و در یک محیط همگن توسعه داده شده است. گره‌های حسگر همگام‌سازی شده‌اند و اندازه‌گیری‌های آن‌ها دارای توزیع یکسان ناشناخته می‌باشند. در هر بازه زمانی Δ_k ، هر گره حسگر s_i بردار داده x_k^i را اندازه‌گیری می‌کند. بردار داده x_k^i از ویژگی‌ها یا خصیصه‌های x_{kj}^i تشکیل می‌شود:

$$x_k^i = (x_{k1}^i, x_{k2}^i, \dots, x_{kd}^i) \quad (1)$$

ارتباطات رادیویی خیلی بیشتر از مصرف انرژی برای محاسبات در گره‌های حسگر است [۵-۷]. برای مثال در ذرات برکلی^۱ و حسگرهای سنسوریا^۲ نسبت بین مصرف انرژی برای ارتباطات رادیویی و مصرف انرژی برای محاسبات بین 10^3 تا 10^4 می‌باشد [۸]. بنابراین می‌توان از افزایش سربار محاسباتی جهت کاهش هزینه ارتباطی بهره برد و مصرف انرژی در شبکه را کاهش داد.

در این مقاله، یک رویکرد توزیع شده برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم پیشنهاد می‌شود که با انجام پردازش درون شبکه‌ای و توزیع محاسبات بین گره‌های حسگر هزینه ارتباطاتی را کاهش می‌دهد. در رویکرد پیشنهادی، از تحلیل مولفه‌های اصلی توزیع شده (DPCA^۳) و الگوریتم خوشه‌بندی پهنا ثابت (FWC^۴) برای ایجاد نمای عادی سراسری و تشخیص ناهنجاری در داده‌های جمع‌آوری شده توسط گره‌های حسگر استفاده می‌شود. همچنین، نمای عادی سراسری با استفاده از ضرایب وزن‌دار حاصل از منحنی فراموشی به‌روزرسانی می‌شود.

در ادامه، در بخش ۲ کارهای انجام شده در زمینه تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم شرح داده می‌شوند. در بخش ۳ مساله تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم بیان شده و در بخش ۴ الگوریتم تحلیل مولفه‌های اصلی توزیع شده معرفی می‌شود. در بخش ۵ رویکرد پیشنهادی برای تشخیص ناهنجاری توزیع شده در شبکه‌های حسگر بی‌سیم شرح داده می‌شود. در بخش ۶ نتایج آزمایش‌های انجام شده برای ارزیابی کارایی رویکرد پیشنهادی ارائه می‌شود و در نهایت در بخش ۷ نتیجه‌گیری به عمل می‌آید.

۲- کارهای مرتبط

برای تشخیص ناهنجاری در داده‌های جمع‌آوری شده توسط گره‌های حسگر روش‌های مختلفی پیشنهاد شده است که در ادامه به برخی از آن‌ها اشاره می‌شود.

Chatzigiannakis و همکارانش [۳] روشی مبتنی بر تحلیل مولفه‌های اصلی (PCA) برای تشخیص ناهنجاری پیشنهاد کرده‌اند. در این روش، شبکه حسگر براساس شباهت بین داده‌های جمع‌آوری شده توسط گره‌های حسگر به تعدادی گروه تقسیم می‌شود. هر گروه از یک گره اصلی و تعدادی گره فرعی تشکیل می‌شود. گره‌های فرعی داده‌های جمع‌آوری شده

¹ Berkeley nodes

² Sensoria sensors

³ Distributed Principal Component Analysis

⁴ Fixed-Width Clustering

⁵ Subspace method

⁶ Support Vector Machine

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_s \end{bmatrix} \quad (3)$$

که X_i شامل n_i بردار داده دریافتی از گره $s_i, i=1 \dots s$ می‌باشد. بنابراین، X یک ماتریس $n \times d$ می‌باشد که n نشان‌دهنده تعداد کل بردارهای داده دریافتی و d نشان‌دهنده تعداد خصیصه‌ها است.

$$n = \sum_{i=1}^s n_i \quad (4)$$

ماتریس کوواریانس S برای X با استفاده از رابطه (۵) محاسبه می‌شود:

$$S = \frac{1}{n} X^T (I - \frac{1}{n} e_n e_n^T) X \quad (5)$$

که $e_n \equiv (1, 1, \dots, 1)^T$ یک بردار به طول n را نمایش می‌دهد.

مولفه‌های اصلی X با تجزیه مقدار تکین (SVD) [۱۳] ماتریس S به دست می‌آیند:

$$S = V \Sigma^2 V^T \quad (6)$$

که Σ یک ماتریس قطری و V ماتریس مولفه‌های اصلی X را تشکیل می‌دهند. از اولین مولفه اصلی سراسری ψ برای تشخیص بردارهای داده ناهنجار استفاده می‌شود.

محاسبه مولفه‌های اصلی به روش فوق مستلزم ارسال بردارهای داده تمام گره‌های عضو به گره سرگروه می‌باشد که هزینه ارتباطی نسبتاً زیادی دارد و طول عمر شبکه را به طور قابل ملاحظه‌ای کاهش می‌دهد. بنابراین، در این مقاله از روش تحلیل مولفه‌های اصلی توزیع‌شده (DPCA) [۱۴] استفاده می‌شود که در آن نیازی به ارسال بردارهای داده تمام گره‌های عضو به گره سرگروه نمی‌باشد. در ادامه، چگونگی محاسبه مولفه‌های اصلی با استفاده از این روش شرح داده می‌شود:

هر گره حسگر s_i ، ابتدا ماتریس میانگین صفر شده X_i را محاسبه می‌کند:

$$\hat{X}_i = (I - \frac{1}{n_i} e_{n_i} e_{n_i}^T) X_i = X_i - e_{n_i} \bar{x}_i^T \quad (7)$$

که \bar{x}_i بردار میانگین-ستونی X_i است. سپس، تجزیه QR^۲ [۱۳] را برای \hat{X}_i محاسبه می‌کند:

که $X_k^i \in \mathbb{R}^d$ است. در هر پنجره زمانی، هر گره حسگر s_i مجموعه‌ای شامل n_i بردار داده $\{X_k^i : k=1 \dots n_i\}$ را جمع‌آوری می‌کند. یک بردار داده ناهنجار یا پیرامونی نامیده می‌شود، اگر از سایر بردارهای داده انحراف قابل توجهی داشته باشد.

در رویکرد متمرکز برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم، در پایان هر پنجره زمانی، هر گره حسگر s_i مجموعه بردارهای داده X_i خود را برای گره سرگروه s_G ارسال می‌کند. گره سرگروه s_G بردارهای داده دریافتی از گره‌های عضو را با هم ادغام کرده و مجموعه بردارهای داده X را تشکیل می‌دهد:

$$X = \bigcup_{i=1 \dots s} X_i \quad (2)$$

سپس الگوریتم تشخیص ناهنجاری را اجرا کرده و بردارهای داده ناهنجار $O \subseteq X$ را در X تشخیص می‌دهد.

رویکرد متمرکز برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم دارای چندین عیب می‌باشد. در این رویکرد حجم زیادی از داده‌ها در شبکه ارسال می‌شوند. این سربار ارتباطی موجب می‌شود تا طول عمر شبکه به طور قابل ملاحظه‌ای کاهش یابد [۱۱]. همچنین، با توجه به این که گره سرگروه داده‌های سایر گره‌ها را دریافت می‌کند، بنابراین بار ارتباطی زیادی به این گره تحمیل می‌شود. با توجه به نکات فوق مشخص می‌شود که رویکرد متمرکز طول عمر شبکه را به سرعت کاهش می‌دهد.

در این مقاله، رویکردی توزیع‌شده برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم ارائه می‌شود که در آن از ارسال تمام بردارهای داده به گره سرگروه جلوگیری می‌شود. همچنین، با به حداقل رساندن ارتباطات مصرف انرژی کاهش یافته و در نتیجه طول عمر شبکه افزایش می‌یابد.

۴- تحلیل مولفه‌های اصلی توزیع شده

تحلیل مولفه‌های اصلی (PCA^۱) تکنیکی قدرتمند برای شناسایی الگوها در داده‌ها است که مهم‌ترین مولفه‌های اصلی را پیدا می‌کند. این مولفه‌ها در واقع پراکندگی داده‌ها را نشان می‌دهند [۱۲]. با استفاده از PCA اولین مولفه اصلی ψ محاسبه می‌شود که توزیع تقریبی داده‌ها را نمایش می‌دهد. از PCA می‌توان برای تشخیص ناهنجاری در داده‌ها استفاده کرد. در ادامه، چگونگی محاسبه مولفه‌های اصلی به روش متمرکز شرح داده می‌شود:

فرض کنید X ماتریس بردارهای داده دریافتی توسط گره سرگروه s_G باشد:

^۲ Singular Value Decomposition

^۳ QR decomposition

^۱ Principal Component Analysis

دریافت می‌کند. این گروه‌بندی می‌تواند ثابت یا پویا باشد. در گروه‌بندی پویا، شبکه حسگر به صورت دوره‌ای و با تغییر شرایط محیطی مجدداً آرایش می‌شود. فرض می‌شود شبکه حسگر بی‌سیم در یک محیط همگن توسعه داده شده است. بنابراین، اندازه‌گیرهای حسگرها دارای توزیع یکسان ناشناخته می‌باشند. تشکیل گروه‌ها براساس همبستگی داده‌های جمع‌آوری شده توسط گره‌ها می‌باشد [۳، ۱۵، ۱۶]. برای گروه‌بندی شبکه‌های حسگر می‌توان از الگوریتم‌های ارائه شده در [۳، ۱۷] استفاده کرد.

رویکرد پیشنهادی از دو مرحله آموزش و تشخیص تشکیل می‌شود. در مرحله آموزش، از بردارهای داده عادی استفاده می‌شود. فرض کنید گروه $G \subseteq N$ شامل تعدادی گره حسگر باشد. هر گره حسگر $s_i \in G$ ابتدا بردارهای داده عادی X_i خود را با استفاده از رویه نرمال‌سازی داده‌ها مقیاس می‌دهد. سپس رویه توصیف داده‌ها را روی X_i اعمال کرده و داده‌های به دست آمده را برای گره سرگروه $s_G \in G$ ارسال می‌کند. گره s_G با استفاده از داده‌های دریافتی از گره‌های عضو نمای عادی^۱ سراسری را ایجاد می‌کند. از نمای ایجاد شده در مرحله تشخیص استفاده می‌شود.

در مرحله تشخیص، در پایان هر پنجره زمانی هر گره حسگر $s_i \in G$ ابتدا بردارهای داده جمع‌آوری شده X_i خود را با استفاده از رویه نرمال‌سازی داده‌ها مقیاس می‌دهد. سپس رویه توصیف داده‌ها را روی X_i اعمال کرده و داده‌های به دست آمده را برای گره سرگروه $s_G \in G$ ارسال می‌کند. گره s_G پس از دریافت داده‌های هر یک از گره‌های s_i رویه تشخیص ناهنجاری را اجرا می‌کند. از این رویه برای تشخیص ناهنجاری در داده‌های جمع‌آوری شده توسط گره‌های عضو گروه استفاده می‌شود. در نهایت، با استفاده از داده‌های عادی دریافتی از تمام گره‌های عضو گروه، نمای عادی سراسری را به‌روزرسانی می‌کند.

در رویکرد پیشنهادی، فرآیند ایجاد نمای عادی بین تمام گره‌ها توزیع می‌شود. همچنین هر گره، به جای ارسال تمام بردارهای داده تنها توصیفی از این بردارها را برای گره سرگروه ارسال می‌کند. بنابراین، فرآیند تشخیص ناهنجاری در مقایسه با رویکرد متمرکز با هزینه ارتباطی کمتری انجام می‌شود. در بخش‌های بعدی رویکرد پیشنهادی با جزئیات بیشتری مورد بحث قرار می‌گیرد.

۵-۱- نرمال‌سازی داده‌ها

خصیصه‌های بردارهای داده جمع‌آوری شده توسط گره‌های حسگر ممکن است دارای دامنه‌های متفاوتی باشند. بنابراین،

$$\hat{X}_i = Q_i R_i \quad (8)$$

که R_i یک ماتریس بالا مثلثی $d \times d$ است. در نهایت، \bar{x}_i و R_i را به گره سرگروه s_G ارسال می‌کند.

گره s_G دو به دو ماتریس‌های R_i و R_j دریافتی از گره‌های عضو را روی هم قرار داده و تجزیه QR آن‌ها را محاسبه می‌کند:

$$\begin{bmatrix} R_i \\ R_j \end{bmatrix} = Q_{(i,j)} R_{(i,j)} \quad (9)$$

این عمل در گره s_G تا $\ell = \lceil \log_2^s \rceil$ مرحله ادامه می‌یابد تا در نهایت $R_{(1,2,\dots,s)}$ به دست آید. همچنین، با استفاده از رابطه (۱۰) بردار میانگین-ستونی سراسری \bar{x} را برای X محاسبه می‌کند:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i \bar{x}_i \quad (10)$$

سپس، ماتریس H را تشکیل می‌دهد:

$$H = \begin{bmatrix} \sqrt{n_1} (\bar{x}_1 - \bar{x}) \\ \sqrt{n_2} (\bar{x}_2 - \bar{x}) \\ \vdots \\ \sqrt{n_s} (\bar{x}_s - \bar{x}) \\ R_{(1,2,\dots,s)} \end{bmatrix} \quad (11)$$

و تجزیه QR آن را محاسبه می‌کند:

$$H = QR \quad (12)$$

در ادامه، با استفاده از ماتریس R ماتریس کوواریانس سراسری S را برای X محاسبه می‌کند:

$$S = \frac{1}{n} R^T R \quad (13)$$

در نهایت، تجزیه مقدار تکین را روی S اعمال کرده و اولین مولفه اصلی سراسری φ را به دست می‌آورد. لازم به ذکر است که مولفه‌های اصلی سراسری محاسبه شده در روش توزیع شده با مولفه‌های اصلی سراسری محاسبه شده در روش متمرکز یکسان می‌باشند.

۵- رویکرد پیشنهادی

در این بخش، رویکردی توزیع شده برای تشخیص ناهنجاری در داده‌های جمع‌آوری شده توسط گره‌های حسگر در یک شبکه حسگر بی‌سیم پیشنهاد می‌شود. در رویکرد پیشنهادی، شبکه‌ی حسگر بی‌سیم به تعدادی گروه تقسیم می‌شود. هر گروه دارای گره سرگروه می‌باشد که داده‌ها را از سایر گره‌های عضو گروه

^۱ Normal profile

procedure FWC

input:

A set of data vectors $X_i = \{x_k^i : k=1 \dots n_i\}$

Cluster width w_c

output:

A set of clusters $C_i = \{C_j^i : j=1 \dots l_i\}$

begin

$C_i := \emptyset$

for each data vector $x_k^i \in X_i$ **do**

if $C_i = \emptyset$ **then**

Make a new cluster C_1^i with centroid x_k^i and radius w_c , $C_i := \{C_1^i\}$

else

Find the nearest cluster $C_j^i \in C_i$ to x_k^i

if $d(x_k^i, C_j^i) < w_c$ **then**

Add x_k^i to cluster C_j^i and update its centroid

else

Make a new cluster C_j^i with centroid x_k^i and radius w_c , $C_i := C_i \cup \{C_j^i\}$

end if

end if

end for

for each cluster $C_j^i \in C_i$

Find the outermost data vector x_k^i in cluster C_j^i

Set the radius of cluster C_j^i to $d(x_k^i, C_j^i)$

end for

end procedure

شکل ۱. شبه کد الگوریتم خوشه‌بندی پهنا ثابت (FWC)

با اجرای الگوریتم خوشه‌بندی پهنا ثابت، هر گره حسگر s_i مجموعه خوشه‌های $C_i = \{C_j^i : j=1 \dots l_i\}$ را ایجاد می‌کند که l_i برابر تعداد خوشه‌ها است. هر خوشه C_j^i با مرکز C_j^i و شعاع r_j^i نمایش داده می‌شود.

۵-۳- ایجاد نمای عادی سراسری

فرض کنید $C = \bigcup_{i=1}^S C_i$ توصیف خوشه‌های دریافتی در مرحله آموزش از گره‌های عضو توسط گره سرگروه s_G باشد. برای ایجاد نمای عادی سراسری، گره s_G ابتدا با استفاده از الگوریتم تحلیل مولفه‌های اصلی توزیع شده (بخش ۴) اولین مولفه اصلی سراسری φ را به دست می‌آورد. سپس با استفاده

برای کاهش تاثیر مقادیر خصیصه‌ها در محاسبه فاصله از فرآیند نرمال‌سازی داده‌ها استفاده می‌شود.

در رویکرد متمرکز، بردارهای داده جمع‌آوری شده توسط گره‌های حسگر به گره سرگروه ارسال می‌شوند. بنابراین، فرآیند نرمال‌سازی را می‌توان در گره سرگروه انجام داد. اما در رویکرد توزیع‌شده، فرآیند نرمال‌سازی در هر گره حسگر و بر روی بردارهای داده محلی انجام می‌شود. برای یکسان شدن نتایج فرآیند نرمال‌سازی در رویکرد توزیع‌شده با رویکرد متمرکز، هر گره حسگر باید کوچکترین و بزرگترین مقادیر سراسری خصیصه‌ها را در اختیار داشته باشد. برای پیدا کردن این مقادیر از رویه زیر استفاده می‌شود:

هر گره حسگر $s_i \in G$ بردارهای کوچکترین و بزرگترین مقادیر محلی (x_{\min}^i, x_{\max}^i) را برای X_i محاسبه کرده و آن‌ها را برای گره سرگروه s_G ارسال می‌کند.

گره s_G بردارهای (x_{\min}^i, x_{\max}^i) را از گره‌های عضو گروه دریافت می‌کند. سپس بردارهای کوچکترین و بزرگترین مقادیر سراسری (x_{\min}^g, x_{\max}^g) را محاسبه کرده و برای تمام گره‌های عضو گروه ارسال می‌کند.

هر گره حسگر s_i با استفاده از بردارهای سراسری (x_{\min}^g, x_{\max}^g) مقادیر خصیصه‌های بردارهای داده خود را به بازه $[0, 1]$ مقیاس می‌دهد. برای نرمال‌سازی داده‌ها در هر پنجره زمانی از بردارهای (x_{\min}^g, x_{\max}^g) پنجره (های) زمانی قبلی استفاده می‌شود.

۵-۲- توصیف داده‌ها

در رویکرد پیشنهادی برای کاهش سربار ارتباطی تنها توصیفی از بردارهای داده ارسال می‌شود. بدین منظور هر گره حسگر $s_i \in G$ در پایان هر پنجره زمانی سه‌تایی (\bar{x}_i, R_i, C_i) را محاسبه می‌کند که \bar{x}_i بردار میانگین-ستونی، R_i ماتریس حاصل از تجزیه QR و C_i توصیف خوشه‌های حاصل از اعمال الگوریتم خوشه‌بندی پهنا ثابت (FWC) [۱۸، ۴] روی بردارهای داده X_i است. از این داده‌ها در گره سرگروه s_G برای ایجاد نمای عادی سراسری و تشخیص ناهنجاری در بردارهای داده جمع‌آوری شده توسط گره حسگر s_i استفاده می‌شود. در ادامه، الگوریتم خوشه‌بندی پهنا ثابت شرح داده می‌شود.

الگوریتم خوشه‌بندی پهنا ثابت (FWC)

در الگوریتم خوشه‌بندی پهنا ثابت مجموعه‌ای از خوشه‌ها با پهنا ثابت w_c ایجاد می‌شود که پارامتر w_c توسط کاربر تعیین می‌شود. در شکل ۱ شبه کد الگوریتم خوشه‌بندی فوق نمایش داده شده است.

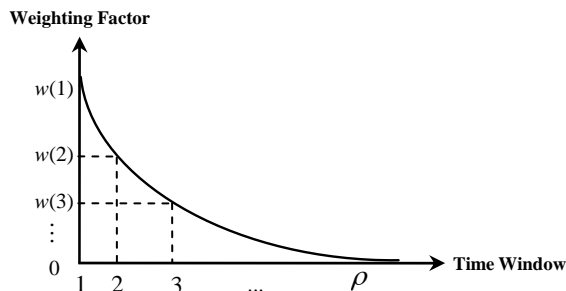
فرض کنید که $\bar{x}(t)$ بردار میانگین-ستونی سراسری و $S(t)$ ماتریس کوواریانس سراسری برای بردارهای داده عادی در پنجره زمانی t باشند. لازم به ذکر است که آمارهای فوق برای هر پنجره زمانی با استفاده از روابط (۱۰) و (۱۳) محاسبه می شوند. در رویکرد پیشنهادی، برای به روزرسانی نمای عادی سراسری، ابتدا بردار میانگین و ماتریس کوواریانس سراسری برای مجموعه بردارهای داده عادی در ρ پنجره زمانی قبلی با استفاده از روابط (۱۸) و (۱۹) محاسبه می شوند:

$$\bar{x}_\rho = \sum_{t=1}^{\rho} w(t) \bar{x}(t) \quad (18)$$

$$S_\rho = \sum_{t=1}^{\rho} w(t) S(t) \quad (19)$$

که $w(t)$ وزن نسبت داده شده به بردارهای داده در پنجره زمانی t است و با استفاده از منحنی فراموشی^۱ Ebbinghaus [۱۹] محاسبه می شود (شکل ۳). هدف از استفاده از این منحنی فراموشی، کاهش اهمیت بردارهای داده در پنجره های زمانی قدیمی هنگام به روزرسانی نمای عادی سراسری است.

در نهایت، تجزیه مقدار تکین روی S_ρ اعمال شده و اولین مولفه اصلی سراسری φ به دست می آید.



شکل ۳. منحنی فراموشی Ebbinghaus

۶- نتایج آزمایشها

برای ارزیابی کارایی رویکرد پیشنهادی از داده های واقعی استفاده شده است. این داده ها از ۵۴ گره حسگر Mica2Dot در آزمایشگاه تحقیقاتی اینتل-برکلی از ۸ فوریه تا ۵ آوریل ۲۰۰۴ میلادی جمع آوری شده است. داده های جمع آوری شده شامل مقادیر رطوبت، دما، نور و ولتاژ در هر ۳۱ ثانیه می باشد. در آزمایش های انجام شده، ابتدا شبکه حسگر با استفاده از الگوریتم گروه بندی ارائه شده در [۱۷] به تعدادی گروه تقسیم شد. سپس گروه شامل گره های ۲۷ تا ۴۲ انتخاب شد. برای تولید بردارهای داده ناهنجر، یکی از گره های فوق به صورت تصادفی انتخاب و به بردارهای داده جمع آوری شده توسط

از رابطه (۱۴) فاصله هر یک از خوشه های $C_j^i \in C$ را از φ محاسبه می کند (شکل ۲):

$$d(C_j^i, \varphi) = d_p(c_j^i, \varphi) + r_j^i, \quad (14)$$

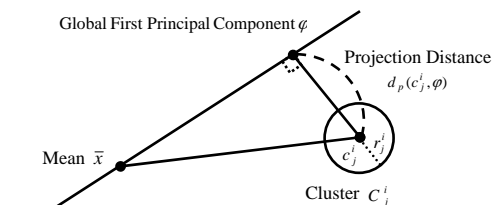
که c_j^i و r_j^i به ترتیب مرکز و شعاع خوشه C_j^i و $d_p(c_j^i, \varphi)$ فاصله تصویر c_j^i از φ است و با استفاده از رابطه (۱۵) محاسبه می شود:

$$d_p(c_j^i, \varphi) = \text{sqrt}(\|c_j^i - \bar{x}\|^2 - (\varphi^T \cdot (c_j^i - \bar{x}))^2) \quad (15)$$

که \bar{x} بردار میانگین-ستونی سراسری است. در نهایت، بیشترین فاصله خوشه ها از φ را به دست می آورد:

$$d_{\max} = \max\{d(C_j^i, \varphi)\}, i=1 \dots s, j=1 \dots l_i \quad (16)$$

از d_{\max} در مرحله تشخیص و برای تشخیص بردارهای داده ناهنجر استفاده می شود.



شکل ۲: فاصله خوشه C_j^i از اولین مولفه اصلی سراسری φ

۴-۵. تشخیص ناهنجاری

فرض کنید C_i توصیف خوشه های دریافتی در مرحله تشخیص از گره $s_i, i=1 \dots s$ باشد. گره سرگروه s_G برای تشخیص ناهنجاری در داده های جمع آوری شده توسط گره s_i ، ابتدا با استفاده از رابطه (۱۴) فاصله هر یک از خوشه های $C_j^i \in C_i$ را از φ محاسبه می کند. سپس در صورتی که فاصله فوق از d_{\max} بزرگتر باشد، خوشه C_j^i را به عنوان ناهنجر تشخیص می دهد:

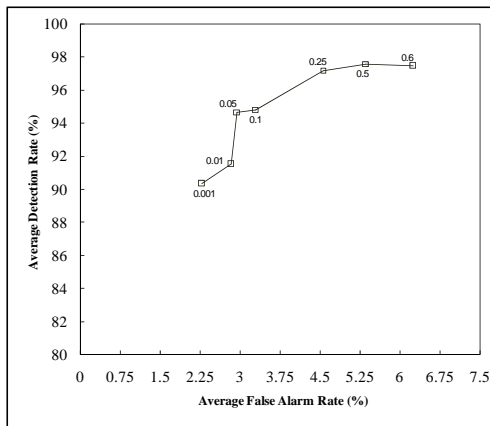
$$\begin{cases} d(C_j^i, \varphi) > d_{\max} & : \text{Anomaly} \\ d(C_j^i, \varphi) \leq d_{\max} & : \text{Normal} \end{cases} \quad (17)$$

که d_{\max} بیشترین فاصله خوشه ها از φ در مرحله آموزش است. در صورتی که تعداد خوشه های ناهنجر دریافتی از یک گره از یک حد آستانه بیشتر باشد، توصیف داده های دریافتی از آن گره نادیده گرفته می شود.

۵-۵. به روزرسانی نمای عادی سراسری

از آن جا که شرایط محیطی که شبکه حسگر بی سیم در آن توسعه می یابد ممکن است در طی زمان تغییر کند، بنابراین لازم است نمای عادی سراسری به روزرسانی شود.

¹ Forgetting curve

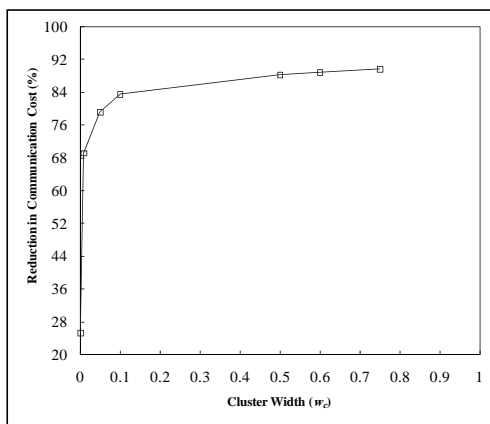


شکل ۵. متوسط نرخ تشخیص در مقابل نرخ هشدار نادرست به ازای مقادیر مختلف پهنای ثابت خوشه‌ها (w_c)

در شکل‌های ۶ و ۷ میزان متوسط کاهش هزینه ارتباطی (RCC^5) در رویکرد پیشنهادی در مقایسه با رویکرد متمرکز [۳] برای ۱۰۰ پنجره زمانی نمایش داده شده است. برای محاسبه کاهش هزینه ارتباطی در هر پنجره زمانی از رابطه (۲۰) استفاده می‌شود:

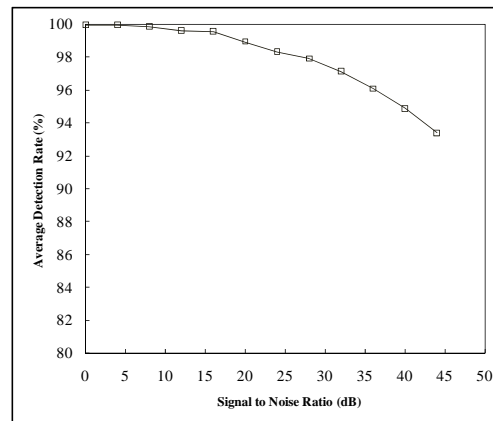
$$RCC = \frac{n - \tau}{n} \quad (20)$$

که n میزان داده‌های ارسالی در رویکرد متمرکز و τ میزان داده‌های ارسالی در رویکرد پیشنهادی است.



شکل ۶. میزان کاهش هزینه ارتباطی به ازای مقادیر مختلف پهنای ثابت خوشه‌ها (w_c)

این گره نویز با توزیع گوسی^۱ اضافه شد. میزان نویز اضافه شده با پارامتر نرخ سیگنال به نویز (SNR^2) مشخص می‌شود. از دو معیار نرخ تشخیص^۳ و نرخ هشدار نادرست^۴ برای ارزیابی رویکرد پیشنهادی استفاده شده است. نرخ تشخیص برابر با درصدی از بردارهای داده ناهنجار در یک پنجره زمانی است که با موفقیت تشخیص داده می‌شوند. نرخ هشدار نادرست برابر با درصدی از بردارهای داده عادی در یک پنجره زمانی است که به صورت نادرست به عنوان بردارهای داده ناهنجار تشخیص داده می‌شوند.



شکل ۴. متوسط نرخ تشخیص به ازای مقادیر مختلف SNR

در شکل ۴ متوسط نرخ تشخیص به ازای مقادیر مختلف SNR برای ۱۰۰ پنجره زمانی نمایش داده شده است. با توجه به این شکل مشخص می‌شود که با وجود کاهش میزان نویز اضافه شده به بردارهای داده، رویکرد پیشنهادی قادر است بردارهای داده ناهنجار را با نرخ بالایی تشخیص دهد. بنابراین، رویکرد پیشنهادی از حساسیت کمی در مقابل کاهش نویز برخوردار است. در آزمایش فوق، پارامترهای پهنای ثابت خوشه‌ها $w_c = 0.5$ و اندازه پنجره زمانی ۵۲ دقیقه در نظر گرفته شد.

در شکل ۵ متوسط نرخ تشخیص در مقابل نرخ هشدار نادرست به ازای مقادیر مختلف پهنای ثابت خوشه‌ها (w_c) برای ۱۰۰ پنجره زمانی نمایش داده شده است. در این آزمایش، پارامترهای نرخ سیگنال به نویز $SNR = 32 \text{ dB}$ و اندازه پنجره زمانی ۵۲ دقیقه در نظر گرفته شد. با توجه به شکل فوق مشخص می‌شود که با انتخاب $w_c = 0.5$ توازن مناسبی بین نرخ تشخیص و نرخ هشدار نادرست برقرار می‌شود.

¹ Gaussian noise

² Signal to Noise Ratio

³ Detection rate

⁴ False alarm rate

⁵ Reduction in Communication Cost

جدول ۱. مقایسه رویکرد پیشنهادی با رویکرد متمرکز [۳]

سربرار ارتباطی در شبکه حسگر	سربرار محاسباتی در هر گره عضو گروه	سربرار محاسباتی در گره سرگروه	رویکرد متمرکز [۳]
$O(nd)$	-	$O(nd^2)$	رویکرد متمرکز [۳]
$O(md^2)$	$O(n_i^2 d)$	$O(d^3 \log_2^m)$	رویکرد پیشنهادی

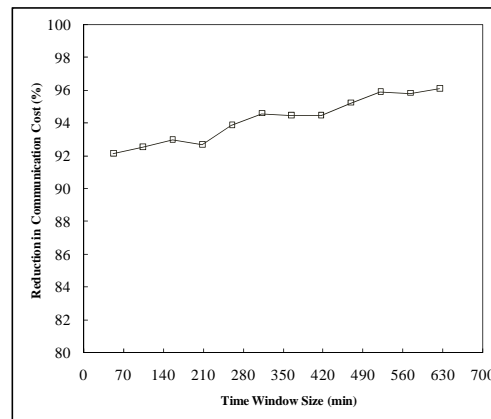
$$(m \ll n, d \ll n, n_i \ll n)$$

سپاسگزاری

این تحقیق با حمایت مالی مرکز تحقیقات مخابرات ایران و تحت قرارداد با کد شناسایی ۱۲۸-۱۲-۸۸ انجام شده است.

مراجع

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 104–112, August 2002.
- [2] A. Perrig, J. Stankovic, and D. Wagner, "Security in wireless sensor networks," *Communications of the ACM*, vol. 47, no. 6, pp. 53–57, June 2004.
- [3] V. Chatzigiannakis and S. Papavassiliou, "Diagnosing anomalies and identifying faulty nodes in sensor networks," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 637–645, May 2007.
- [4] C. E. Loo, M. Y. Ng, C. Leckie, and M. Palaniswami, "Intrusion detection for routing attacks in sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2, no. 4, pp. 313–332, November 2006.
- [5] M. Ilyas, I. Mahgoub, and L. Kelly, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, CRC Press, London, 2004.
- [6] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, pp. 364–369, Los Angeles, California, USA, 2005.
- [7] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energy aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, March 2002.
- [8] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: an information-directed approach," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1199–1209, 2003.
- [9] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Distributed anomaly detection in wireless sensor networks," in *Proceedings of the 10th IEEE International Conference on Communication systems*, pp. 1–5, Singapore, October 2006.
- [10] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, "Quarter sphere based distributed anomaly detection in wireless sensor networks," in *Proceedings of the IEEE International Conference of Communication*, pp. 3864–3869, Glasgow, UK, June 2007.



شکل ۷. میزان کاهش هزینه ارتباطی به ازای مقادیر مختلف اندازه پنجره زمانی

با توجه به شکل های ۶ و ۷ مشخص می شود که با افزایش پهنای خوشه ها هزینه ارتباطی به طور قابل ملاحظه ای کاهش می یابد، اما اندازه پنجره زمانی تاثیر قابل ملاحظه ای در کاهش هزینه ارتباطی ندارد.

۷- نتیجه گیری

در این مقاله، یک رویکرد توزیع شده مبتنی بر تحلیل مولفه های اصلی برای تشخیص ناهنجاری در شبکه های حسگر بی سیم پیشنهاد شده است. در رویکرد پیشنهادی، شبکه حسگر بی سیم به تعدادی گروه تقسیم می شود. هر گروه دارای گره سرگروه می باشد که داده ها را از سایر گره های عضو گروه دریافت می کند. فرآیند ایجاد نمای عادی بین تمام گره های عضو هر گروه توزیع می شود. هر گره به جای ارسال تمام بردارهای داده تنها توصیفی از این بردارها را برای گره سرگروه ارسال می کند. بنابراین، فرآیند تشخیص ناهنجاری در مقایسه با رویکرد متمرکز [۳] با هزینه ارتباطی کمتری انجام می شود.

رویکرد پیشنهادی با استفاده از داده های واقعی گره های حسگر در آزمایشگاه تحقیقاتی اینتل-برکلی مورد ارزیابی قرار گرفت. نتایج آزمایش های انجام شده نشان می دهد که حتی در صورت وجود نویز کم رویکرد پیشنهادی قادر است بردارهای داده ناهنجار را با نرخ بالایی تشخیص دهد. همچنین، با انتخاب مقادیر مناسب برای پارامترها، رویکرد پیشنهادی قادر است هزینه ارتباطی را در مقایسه با رویکرد تشخیص ناهنجاری متمرکز [۳] تا ۸۸ درصد کاهش دهد. در جدول ۱ سربرار محاسباتی و هزینه ارتباطی رویکرد پیشنهادی با رویکرد متمرکز [۳] مقایسه شده است. در این جدول، n تعداد کل بردارهای داده ارسالی به گره سرگروه در هر پنجره زمانی، n_i تعداد بردارهای داده جمع آوری شده توسط هر گره عضو گروه در هر پنجره زمانی، d تعداد ویژگی های قابل اندازه گیری و m تعداد اعضای گروه است. نتایج آزمایش های انجام شده نشان می دهد که رویکرد پیشنهادی از نرخ تشخیص قابل مقایسه ای نسبت به رویکرد تشخیص ناهنجاری متمرکز برخوردار می باشد، در حالی که هزینه ارتباطی را به طور قابل ملاحظه ای کاهش می دهد.

- [11] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, March 2002.
- [12] H. Nakayama, S. Kurosawa, A. Jamalipour, Y. Nemoto, and N. Kato, "Dynamic anomaly detection scheme for AODV-based mobile ad hoc networks," *IEEE Transaction On Vehicular Technology*, vol. 58, no 5, pp. 2471–2481, June 2009.
- [13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Third Edition, 1996.
- [14] Z. Bai, R.Chan, and F. Luk, "Principal component analysis for distributed data sets with updating," in *Proceedings of International workshop on Advanced Parallel Processing Technologies*, Singapore, 2005.
- [15] M. C. Vuran, B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 45, no. 3, pp. 245–259, June 2004.
- [16] H.-L. Huang and P. Antonelli, "Application of principal component analysis to high-resolution infrared measurement compression and retrieval," *Journal of Applied Meteorology*, vol. 40, no. 10, pp. 365–388, March 2001.
- [17] G. Li, J. He, and Y. Fu, "Group-based intrusion detection system in wireless sensor networks," *Computer Communications*, vol. 31, no. 18, pp. 4324–4332, December 2008.
- [18] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data," *Applications of Data Mining in Computer Security*, Kluwer, 2002.
- [19] F. Y. Luo, H. S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," *IEEE Journal Selected Areas Communication*, vol. 23, no. 4, pp. 839–850, April 2005.