فصلنامه پژوهش‌های حفاظتی ـ امنیتی دانشگاه جامع امام حسین(علیه‌السلام)
سال پنجم، شماره‌ی ۱۸ (تابستان ۱۳۹۵): صص ۱۲۱ - ۱۴۱

روشی کارآمد برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم

شهره شهاب‌احمدی[۱]
قاسم میرجلیلی[۲]

تاریخ دریافت: ۱۳۹۴/۱۲/۱۵
تاریخ پذیرش: ۱۳۹۵/۰۵/۱۵

چکیده

در شبکه‌های حسگر بی‌سیم، مانند سایر شبکه‌ها، همواره امکان نفوذ عامل‌های بدخواه وجود دارد. همچنین ویژگی‌های خاص شبکه‌های حسگر بی‌سیم و محدودیت‌های ذاتی که در منابع دارا هستند، این شبکه‌ها را در برابر انواع مختلفی از حملات امنیتی بسیار آسیب‌پذیر می‌کند. یکی از روش‌های مؤثر برای برقراری امنیت و قابلیت اطمینان در شبکه‌های حسگر بی‌سیم، استفاده از روشی جهت تشخیص رفتارهای ناهنجار در شبکه می‌باشد. در این راستا، سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری به‌عنوان یکی از راه‌کارهای مهم برای برآوردن این هدف مطرح می‌گردند. در این مقاله، یک روش توزیعی با کارایی و دقت مناسب جهت تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم با استفاده از معیار آنتروپی و معیار جدید شباهت نقطه‌ای تجمعی پیشنهاد می‌شود. در این روش، ابتدا داده‌های هر گره حسگر با استفاده از یک روش بدون نظارت و غیر پارامتری تقسیم‌بندی شده و سپس با استفاده از یک سطح آستانه تطبیقی در هر سطح از شبکه، ناهنجاری‌ها شناسایی می‌شوند. دقت تشخیص این روش در شناسایی داده‌های ناهنجار بیشتر از ۹۹٪ می‌باشد.

کلید واژه‌ها: *شبکه‌های حسگر بی‌سیم، امنیت، تشخیص ناهنجاری، آنتروپی، شباهت نقطه‌ای تجمعی*

۱- کارشناس ارشد فناوری اطلاعات، گرایش مخابرات امن، گروه کامپیوتر، دانشگاه یزد. Sh.shahabahmadi@stu.yazd.ac.ir
۲- استاد، دانشکده مهندسی برق، دانشگاه یزد. mirjalily@yazd.ac.ir

۱- مقدمه

توانایی حس کردن و جمع‌آوری اطلاعات در شبکه‌های حسگر بی‌سیم، فرصت‌های منحصربه‌فردی را برای استفاده از آن، جهت کنترل توزیع‌شده در بسیاری از کاربردهای صنعتی و محیطی فراهم کرده است. بنابراین شبکه‌های حسگر بی‌سیم به یک سیستم کلیدی برای بسیاری از فرآیندهای کنترلی از طریق جمع‌آوری داده‌های حس شده و قابلیت‌های تصمیم‌گیری، تبدیل شده‌اند (Suthaharan,2010).

۱-۱ بیان مسئله

در اکثر شبکه‌های حسگر بی‌سیم، امنیت سناریوهای کاربردی، یک موضوع بسیار مهم است، به‌ویژه شبکه‌هایی که جهت به‌کارگیری در محیط‌های جنگی و تجاری طراحی شده‌اند. با توجه به اهمیت سطح امنیت در این شبکه‌ها، حصول اطمینان از سطح مورد انتظار سخت‌تر از سایر شبکه‌های بی‌سیم همتایش می‌باشد. در حقیقت بنا به دلایلی مانند: پخشی بودن ارتباطات بی‌سیم، محدودیت منابع گره‌های حسگر، مستعد ابتلا به حملات فیزیکی بودن و غیره، امنیت در شبکه‌های حسگر بی‌سیم دارای چالش‌های زیادی است که ممکن است در انواع دیگر از شبکه‌های بی‌سیم دیده نشود. راه‌حل‌های امنیتی مانند احراز هویت، رمزنگاری و یا مدیریت کلید به‌عنوان خط دفاعی اول می‌تواند امنیت شبکه را بالا ببرد. بااین‌وجود، این راه‌حل‌ها به‌تنهایی نمی‌توانند از تمام حملات ممکن جلوگیری کنند. به‌عنوان‌مثال طیف وسیعی از حملات در شبکه می‌تواند توسط گره‌هایی به وجود بیاید که ظاهراً مشروع به نظر می‌رسند اما درواقع این‌گونه نیستند و امنیت شبکه را به خطر می‌اندازند، به همین علت وجود یک خط دفاعی دوم مانند سیستم‌های تشخیص نفوذ ضروری است (Abduvaliyev,2013).

تهدیدات امنیتی در شبکه‌های حسگر بی‌سیم می‌توانند:

۱) بر یکپارچگی شبکه از طریق تخریب مسیرها و ساختار گره‌ها تأثیر بگذارند.

۲) فرآیندهای مسیریابی را تغییر دهند.

۳) عملیات نامشروع شبکه‌ای را اعمال کنند.

۴) تغییراتی غیرقانونی در شبکه انجام دهند یا داده‌های تحریف‌شده بسازند (Kumarage,2013:791).

تمام موارد نام برده می‌توانند به‌راحتی از طریق پیاده‌سازی حملات روی این نوع از شبکه‌ها به تجهیزات نظارتی صنعتی، القا شوند. استقرار شبکه‌های حسگر بی‌سیم، بدون هیچ مراقبتی در سراسر

ناحیه‌های بزرگ جغرافیایی در بیشتر کاربردها و عدم تضمین امنیت فیزیکی گره‌های حسگر، این شرایط را تشدید می‌کند.

این حملات امنیتی اغلب به‌عنوان ناهنجاری در جریان داده‌ها ظاهر می‌شوند که پیامدهایی جدی برای فرآیندهای تصمیم‌گیری در بر خواهند داشت و به‌راحتی می‌توانند اجزاء کلیدی برنامه‌های کاربردی را تهدید کنند (Pietro, 2009:1501). اگر این داده‌های ناهنجار بدون شناسایی رها شوند منجر به عملیات و تصمیم‌های کنترلی اشتباه روی زیرساخت‌های صنعتی خواهند شد که درنهایت می‌توانند به تلفات شدید اقتصادی، صدمات انسانی و خرابی‌های محیط زیستی منجر شوند. تحقیقات اخیر فقط روی امنیت ارتباطات شبکه‌های حسگر بی‌سیم، اغلب از طریق تعریف پروتکل‌های مسیریابی امن‌تر، مدیریت کیفیت رسانه، کلیدهای امنیتی و قابلیت اطمینان تمرکز کرده‌اند. یک نیاز ضروری برای ایجاد امنیت در شبکه‌های داده محور[1]، اطمینان از یکپارچگی داده‌های حس شده است. بنابراین، روش‌هایی برای تشخیص کارآمد و دقیق ناهنجاری‌های بالقوه در داده‌های حس شده، در پیاده‌سازی شبکه‌های حسگر بی‌سیم مهم است. در این راستا، سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری به علت ویژگی‌های خاصی که دارند برای شبکه‌های حسگر بی‌سیم مناسب‌ترین شیوه تشخیص نفوذ می‌باشند (Xie,2011).

۱-۲ معرفی سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری

ناهنجاری در یک شبکه می‌تواند با تجزیه‌وتحلیل داده‌های اندازه‌گیری شده و ویژگی‌های ترافیک مربوطه شناسایی شود. این تشخیص به دو جزء نیازمند است ۱) تعریفی از ناهنجاری‌های مشاهده‌شده در داده‌ها ۲) یک الگوریتم که بتواند با دقت و به‌موقع با مصرف کمترین انرژی این ناهنجاری‌ها را شناسایی کند. به‌طورکلی، ناهنجاری در یک مجموعه داده «یک مشاهده یا مجموعه‌ای از مشاهدات است که در مقایسه با سایر داده‌ها در مجموعه داده ناسازگار است» (Rajasegarar,2009:232).

تشخیص ناهنجاری یک شاخه از تشخیص نفوذ است که به علت داشتن ساختاری انعطاف‌پذیر و منابع پسند، برای شبکه‌های حسگر بی‌سیم مناسب‌ترین شیوه تشخیص می‌باشد (Xie,2011). در شبکه‌های حسگر بی‌سیم روش‌های تشخیص نفوذ بسیاری وجود دارد که از راه‌کارهای تشخیص ناهنجاری استفاده می‌کنند. این نوع سیستم‌ها معمولاً بر تجزیه‌وتحلیل این‌که آیا رفتار یک گره حسگر، با توجه به مفروضات خاص و معیارها، می‌تواند عادی در نظر گرفته شود یا غیرعادی استوار هستند. اکثر محققان این شیوه را

1. Data-centric

به‌عنوان روش اصلی برای تشخیص نفوذ در نظر گرفته‌اند، به‌این‌علت که معتقد هستند، پیاده‌سازی آن از روش‌های تشخیص نفوذ مبتنی بر بدرفتاری یا مشخصات، ساده‌تر می‌باشد (Abduvaliyev,2013).

تشخیص ناهنجاری زمانی مورداستفاده قرار می‌گیرد که تشخیص رفتار غیرعادی بدون داشتن اطلاعات قبلی در مورد آن حائز اهمیت باشد. چنین رفتارهای غیرعادی در سیستم می‌تواند در اثر حملات خرابکارانه یا نفوذ در شبکه، خرابی حسگرها در شبکه، یا بروز پدیده‌های غیرعادی در حوزه عملیاتی به وجود آید. در حال حاضر، علاقه به اینکه چگونه از شبکه‌های حسگر بی‌سیم در برابر حملات امنیتی دفاع شود رو به افزایش است. در این زمینه، تشخیص نفوذ مبتنی بر ناهنجاری نقش مهمی را در ایجاد امنیت، به‌عنوان جایگزین و یا تکمیل‌کننده خط دفاع اولیه بازی می‌کند (Rajasegarar,2009).

در شبکه‌های حسگر بی‌سیم، چالش‌های قابل‌توجهی برای تشخیص ناهنجاری‌ها وجود دارد که هنوز به‌صورت رضایت بخشی در مطالعات موجود رسیدگی نشده‌اند و موارد زیر را شامل می‌شوند:

۱- حجم زیاد داده‌های حس شده.

۲- وجود یک محیط ناهمگون با توزیع داده پویا، که نیاز به یک روش طبقه‌بندی بدون نظارت و غیر پارامتری دارد.

۳- توزیع چند تراکمی و ناهمگون در هر نمونه از مشاهدات جمع‌آوری‌شده.

۴- محدودیت منابع که به روش‌های کارآمد پردازش داده با هدف کاهش منابع محاسباتی و ارتباطی نیاز دارد.

به چالش‌هایی که در بالا ذکر شد در این مقاله با یک چارچوب پردازش داده که ناهنجاری‌ها را در سطوح متفاوتی از ساختار سلسله‌مراتبی شناسایی می‌کند، رسیدگی شده است. به‌طور خلاصه، در روش پیشنهادی، فضای داده در هر گره حسگر در یک پنجره زمانی مشخص ΔT به چندین ناحیه با یک روش کاملاً بدون نظارت و غیر پارامتری تقسیم می‌شود. این نواحی با استفاده از دو معیار آنتروپی نقطه‌ای تجمعی[1] و تشابه نقطه‌ای تجمعی[2] در هر سطح از ساختار سلسله‌مراتبی به دست می‌آیند. در هر سطح داده‌های ناهنجار با استفاده از یک مقدار آستانه تطبیقی شناسایی می‌شوند و به سطح بعدی منتقل می‌شوند. این عمل تا رسیدن به گره ریشه و مشخص شدن نتیجه نهایی ناهنجاری ادامه دارد. برای ارزیابی این روش، از یک مجموعه داده چند تراکمی مربوط به یک کاربرد محیطی رایج در شبکه حسگر بی‌سیم،

1. Cumulative point-wise entropy
2. Cumulative point-wise similarity

استفاده شده است (Suthaharan,2010)، که در این مجموعه داده، داده‌های حس شده مربوط به چهار گره حسگر که نوسانات دما و رطوبت را در یک دوره ۶ ساعتی اندازه‌گیری می‌کنند جمع‌آوری شده است.

در ادامه، مقاله به این صورت سازمان‌دهی شده است که در بخش ۲ کارهای تحقیقاتی انجام‌شده در زمینه تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم بیان می‌شود. در بخش ۳ روش پیشنهادی تشریح شده و در بخش ۴ نتایج ارزیابی و شبیه‌سازی بیان شده است.

۲- کارهای مرتبط

در گذشته، کارهای زیادی در رابطه با تشخیص ناهنجاری انجام شده است (Chandola,2009-Hodge,2004-Romaswamy,2000). این شیوه‌ها به‌این‌علت که یا به اطلاعات قبلی از توزیع داده‌های ورودی نیاز دارند یا طراحی مناسبی با توجه به محدودیت منابع ندارند، به‌صورت مستقیم قابل‌استفاده در شبکه‌های حسگر بی‌سیم مقیاس بزرگ نیستند.

یک عامل کلیدی نگران‌کننده دیگر، ماهیت طبیعی خود داده‌ها می‌باشد، به این صورت که ممکن است هر دامنه از داده‌های مشاهده‌شده دارای توزیع‌های پویا و ناهمگون با تراکم‌های متفاوت باشند. روش‌های موجود را می‌توان در دسته‌های متفاوتی تقسیم‌بندی کرد. به‌عنوان‌مثال، ۱) بر اساس تابع توزیع داده‌ها ۲) بر اساس تراکم و ۳) بر اساس خوشه‌بندی. در دسته‌بندی اول فرض اساسی این است که تابع توزیع احتمالات نقاط داده‌ای که برای ناهنجاری تجزیه‌وتحلیل خواهند شد از قبل شناخته‌شده هستند، به‌طور مثال، توزیع گوسی[1]. در ابتدا پارامترهای تابع توزیع مورد ارزیابی قرار می‌گیرند، سپس داده‌هایی که شباهت کمی با توزیع داده شده دارند، به‌عنوان ناهنجاری مشخص می‌شوند. در این روش، توزیع‌های تک متغیره یا چند متغیره می‌توانند استفاده شوند. در روش‌های تک متغیره، متغیرهای اندازه‌گیری به‌صورت مستقل با استفاده از توزیع‌های جداگانه مدل می‌شوند. ناهنجاری‌های پیداشده معمولاً در خارج از داده‌های توزیع‌شده اتفاق می‌افتند. در روش‌های چند متغیره، یک توزیع مشترک برای تمام متغیرهای اندازه‌گیری، معین می‌شود. این روش‌ها، مدل کردن همبستگی بین متغیرها را ممکن می‌سازند. بعلاوه، تجزیه‌وتحلیل‌های چند متغیره ممکن است ناهنجاری‌هایی را که در تجزیه‌وتحلیل‌های تک متغیره مشاهده نمی‌شوند، آشکار کنند. بااین‌حال، این امر منجر به یک مسئله یادگیری پیچیده‌تری (به علت تعداد پارامترهای بیشتری که برای ارزیابی موردنیاز است) می‌شود. روش‌های پارامتری زمانی مناسب هستند که اساساً نوع توزیع داده به‌خوبی

1. Gaussian distribution

شناخته‌شده باشد، که این امر شدیداً وابسته به کاربرد است. زمانی که گره‌های حسگر متحرک باشند و همچنین هنگامی‌که توزیع داده‌ها در طول عمر شبکه تغییر پیدا می‌کند، این روش می‌تواند منجر به طبقه‌بندی‌های اشتباه و مشکل در تشخیص شود.

روش‌های مبتنی بر تراکم، ناهنجاری‌ها را بر اساس تراکم محلی هر مشاهده شناسایی می‌کنند (Breunig,2000- Xie,2013). همسایه‌های محلی هر مشاهده بر اساس یک فاصله r که شامل k نزدیک‌ترین نقاط داده باشد، تعریف می‌شود. درحالی‌که این روش‌ها با در نظر گرفتن کمترین مقدار فرضیات و پارامترها قادر به کنترل توزیع‌های چند تراکمی هستند، اما با ریز خوشه‌هایی از داده‌های ناهنجار که تراکمی مشابه با داده‌های عادی دارند دچار مشکل هستند. برای رسیدگی به این مشکل، به‌عنوان‌مثال پارامتر k می‌تواند بزرگ‌تر تعیین شود که این حساسیت به یک پارامتر باعث عدم اطمینان و کاهش نرخ تشخیص در زمانی که مشاهدات دارای نقاط تکی یا ریز خوشه‌هایی از ناهنجاری هستند، خواهد شد. یکی از محققان با استفاده از انتگرال همبستگی محلی بر اساس اندازه‌های آماری مشتق شده از تراکم‌های محلی، به این مسئله رسیدگی کرده است (Papadimitrion,2003). این روش ناهنجاری‌ها را به‌صورت غیر پارامتری و مقاوم با در نظر گرفتن توزیع‌های پویا و تراکم‌های متفاوت شناسایی می‌کند. بااین‌حال قادر به تشخیص رفتارهای غیرعادی که تراکم مشابهی با تراکم داده‌های عادی دارند، نیست.

شیوه‌های مبتنی بر تقسیم‌بندی/ خوشه‌بندی، زمانی که با روش‌های غیر پارامتری پیاده‌سازی می‌شوند مناسب‌ترین روش برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم مقیاس بزرگ هستند. این روش‌ها تقسیم‌بندی‌هایی کاملاً بدون نظارت، بدون هیچ دانش یا فرضیات قبلی را که می‌توانند بر اساس مدل‌های یادگیری تطبیق داده شوند فراهم می‌کنند. استفاده از این روش‌ها برای شبکه‌هایی با توپولوژی سلسله‌مراتبی که در آن‌ها ناهنجاری‌ها به‌صورت تدریجی در سطوح متفاوت شبکه ارزیابی می‌شوند، بهترین راه‌کار می‌باشد. بااین‌حال، اغلب روش‌های خوشه‌بندی موجود برای تشخیص ناهنجاری، فرآیندهایی مرکز محور و غیر بهینه با پیچیدگی بالا هستند که منابع سیستمی زیادی را مصرف می‌کنند (Rokach,2009- Xu,2005).

در اغلب الگوریتم‌های تشخیص ناهنجاری مبتنی بر خوشه‌بندی، از دو ویژگی فاصله یا تراکم برای طبقه‌بندی داده‌ها استفاده شده است. چنین روش‌هایی ازآنجایی‌که فقط از آمارگان مرتبه دوم داده استفاده می‌کنند دارای محدودیت هستند و بیشتر برای داده‌های با خوشه‌های فوق کروی[1] یا فوق بیضوی[1]

1. Hyper-spherical

مناسب‌اند. نگرانی دیگر چگونگی تعیین تعداد خوشه‌های مورد انتظار به‌صورت پویا بدون داشتن دانش قبلی از داده‌ها می‌باشد. این نگرانی توسط Yao با به‌کارگیری معیار آنتروپی نقطه‌ای تجمعی برای خوشه‌بندی مستقیم داده‌ها با استفاده از مدل استنتاج فازی برطرف شده است (Yao, 2000). اساس استفاده از این معیار را می‌توان به این صورت توضیح داد که به‌طورکلی اگر داده‌ها دارای خوشه‌های مجزا از هم باشند دارای یک ساختار منظم هستند، در غیر این صورت دارای یک ساختار نامنظم یا تصادفی می‌باشند. بر اساس نظریه آنتروپی، می‌دانیم که مقدار آنتروپی در مجموعه‌های منظم کم است و در مجموعه‌های نامنظم زیاد است. اگر یک مجموعه داده را بر اساس نقاط داده مجزا تجزیه کنیم، یک ساختار منظم داریم به این معنی که برای هر نقطه داده، تعدادی نقطه داده وجود دارد که به آن نزدیک هستند (به یک خوشه تعلق دارند) و سایر نقاط از آن دور هستند. پس اگر مقدار آنتروپی را برای هر نقطه داده به دست آوریم، نقطه داده‌ای با کمترین مقدار آنتروپی یک کاندید خوب برای مرکز خوشه شدن می‌باشد. اما چنانچه مجموعه داده دارای نقاط ناهنجار باشد این تعریف برای انتخاب نقطه داده برای مرکز خوشه شدن معتبر نیست و در ابتدا باید نقاط ناهنجار از مجموعه داده حذف شوند. Yao برای محاسبه آنتروپی از معیار فاصله اقلیدسی بین نقاط داده استفاده کرده است. بااین‌حال، این روش از مقادیر آستانه از قبل تعریف‌شده β و γ برای برطرف کردن تأثیر نویز و تعریف مرزهای تشابه هر خوشه استفاده می‌کند که باعث عدم تطبیق‌پذیری این کار برای تشخیص ناهنجاری در زمینه توزیع‌های چند تراکمی پویا شده است. از طرف دیگر روش‌های خوشه‌بندی مبتنی بر تراکم، برای تشخیص رفتارهای ناهنجاری که ناحیه‌ای با تراکم مشابه با نقاط عادی تشکیل می‌دهند، نمی‌توانند استفاده شوند. به‌این‌ترتیب، روش‌های خوشه‌بندی توزیع‌شده داده‌ها که به‌طور مستقیم از ویژگی‌های اطلاعاتی بیشتر از آمارگان مرتبه دوم (به‌عنوان‌مثال استفاده از معیار آنتروپی مناسب) استفاده می‌کنند دارای برتری هستند و برای داده‌های مشاهده‌شده پویای چند تراکمی مقاوم‌اند.

در سال ۲۰۱۵، Kumarage و همکارانش روشی با نام «تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم با استفاده از تقسیم‌بندی پویای داده به‌وسیله معیار آنتروپی» بر اساس یک روش تقسیم‌بندی غیر پارامتری بدون نظارت و تطبیق‌پذیر با استفاده از آنتروپی نقطه‌ای تجمعی و میانگین تراکم نسبی[2] ارائه دادند (Kumarage,2015). آن‌ها از یک مدل سلسله‌مراتبی برای نمایش شبکه استفاده کرده‌اند که در آن، عمل تشخیص ناهنجاری در چند گام در هر سطح انجام می‌شود. در این روش، یک پنجره زمانی ΔT در نظر گرفته شده است که در هر پنجره زمانی، گره‌های حسگر مشاهدات خود را تجمیع می‌کنند و عملیات

1. Hyper-elliptical
2. Average relative density

تعریف شده را برای تشخیص ناهنجاری بر روی داده‌هایی که جمع‌آوری کرده‌اند، انجام می‌دهند. به این صورت که در ابتدا میانگین تراکم‌های نسبی نقاط داده محاسبه می‌شوند. سپس داده‌های مشاهده‌شده در نواحی منسجم متفاوتی که E-Regions نامیده شده است با استفاده از معیار آنتروپی نقطه‌ای تجمعی هر نقطه داده تقسیم‌بندی می‌شوند. برای این کار در ابتدا مقدار آنتروپی نقطه‌ای تجمعی هر نقطه داده در هر گره حسگر محاسبه می‌گردد. سپس نقطه‌ای با کمترین مقدار آنتروپی به‌عنوان نقطه‌ای که یک ناحیه اطراف آن می‌تواند شکل بگیرد انتخاب می‌شود. ممکن است نقطه انتخاب‌شده یک نقطه ناهنجار باشد که در ناحیه‌ای پراکنده از مجموعه داده حضور دارد، در این صورت آن نقطه داده با استفاده از یک مقدار آستانه تطبیقی، حذف شده و نقطه بعدی انتخاب می‌گردد. سپس یک ناحیه از نقاطی که مقدار آنتروپی نقطه‌ای تجمعی آن‌ها به‌اندازه یک انحراف معیار استاندارد از مقدار آنتروپی نقطه‌ای تجمعی نقطه انتخاب‌شده فاصله دارد، ایجاد می‌شود. این عمل به‌صورت بازگشتی ادامه پیدا می‌کند تا تمام مجموعه داده به نواحی تعریف‌شده تقسیم‌بندی شود. در هر مرحله، نقاط داده‌ای که به نواحی اختصاص داده شده‌اند، حذف و ناحیه بعدی بر اساس نقاط باقی‌مانده در مجموعه داده محاسبه می‌شود. در آخر بر اساس نواحی تعریف‌شده و تراکم‌های نسبی مربوط به هر مشاهده، ناهنجاری‌های تکی و گروهی شناسایی می‌شوند. این روش توانسته نرخ تشخیص خوبی را برای شناسایی نقاط ناهنجار به دست آورد.

۳- روش پیشنهادی

در این بخش یک روش تقسیم‌بندی بدون نظارت و غیر پارامتری جدید برای داده‌های پویا و ناهمگون با استفاده از نسبت بین معیارهای آنتروپی نقطه‌ای تجمعی و تعریف معیار جدید شباهت نقطه‌ای تجمعی ارائه شده است. در این روش، شناسایی ناهنجاری با استفاده از یک مقدار آستانه تطبیق‌پذیر که از ویژگی‌های آماری داده‌ها مشتق شده است، در گام‌های متفاوت، در هر سطح از شبکه انجام می‌شود. هر گام تنها روی دامنه اطلاعاتی در دسترس در همان سطح شبکه تکیه می‌کند و هیچ‌گونه پارامتر یا فرضیه‌ای را در نظر نمی‌گیرد، که این امر منجر به مقاومت و دقت تشخیص برای داده‌های پویا و با تراکم‌های متفاوت می‌شود. در طرح پیشنهادی، یک شبکه حسگر با ساختار سلسله‌مراتبی در نظر گرفته شده است، به‌طوری‌که برای هر گره برگ، یک گره والد وجود خواهد داشت که اطلاعات را از فرزندانش دریافت کرده و از وضعیت آن‌ها آگاه است.

۳-۱ معیار شباهت و آنتروپی نقطه‌ای تجمعی

برای محاسبه معیار شباهت، یک مجموعه از مشاهدات $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$ را در نظر بگیرید که هر نقطه داده X_i یک مشاهده m بعدی است به‌صورت $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$. شباهت بین دو نقطه داده X_i و X_j که با S_{ij} نشان داده می‌شود به‌صورت زیر محاسبه می‌گردد (Yao, 2000: 383):

$$S_{ij} = \exp^{-\alpha D_{ij}} \quad (۱)$$

D_{ij} که فاصله اقلیدسی بین دو نقطه داده X_i و X_j است و از رابطه (۲) به دست می‌آید، می‌تواند هر مقدار بین صفر تا بی‌نهایت را دارا باشد. به همین علت در رابطه (۱)، مقدار S_{ij} با استفاده از تابع exp، نرمالیزه شده است، به‌گونه‌ای که مقدار نگاشت شده در محدوده [۰،۱] می‌باشد. با توجه به این رابطه، می‌توان استنباط کرد که مقدار شباهت یک نقطه داده با داده‌هایی که فاصله کمی با آن‌ها دارد، بسیار زیاد است (نزدیک به یک) و برای داده‌هایی که با آن‌ها بیشترین فاصله را دارد بسیار کم است (نزدیک به صفر).

$$D_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots (x_{im} - x_{jm})^2} \quad (۲)$$

برای محاسبه معیار شباهت، مقدار α به‌صورت خودکار با جایگزین کردن S_{ij} با مقدار ۰/۵ و D_{ij} با میانگین فاصله بین تمام جفت نقاط داده در رابطه (۱) به دست می‌آید. بنابراین، α به‌صورت وفقی با استفاده از رابطه (۳) بر اساس داده‌های موجود در مجموعه داده با پرهیز از فرض‌های دلخواهانه محاسبه می‌گردد، که D در این رابطه، همان میانگین فاصله تمام جفت نقاط موجود در مجموعه داده می‌باشد (Yao, 2000: 383).

$$\alpha = -\ln 0.5/D \quad (۳)$$

با توجه به آنچه تاکنون مطرح گردید، معیار شباهت نقطه‌ای تجمعی یک نقطه داده X_i به‌صورت زیر محاسبه می‌شود:

$$S_i = \sum_{j \in X}^{j \neq i} S_{ij} \quad (۴)$$

برای محاسبه مقدار آنتروپی بین دو نقطه داده X_i و X_j به‌صورت زیر عمل می‌کنیم (Kumarage, 2015:6):

$$e_{ij} = -S_{ij} \log S_{ij} - (1 - S_{ij}) \log(1 - S_{ij}) \quad (۵)$$

و مقدار آنتروپی نقطه‌ای تجمعی برای هر مشاهده $i \in X$ به‌صورت زیر به دست می‌آید:

$$e_i = -\sum_{j \in X}^{j \neq i} (S_{ij} \log S_{ij} + (1 - S_{ij}) \log(1 - S_{ij})) \quad (۶)$$

همان‌طور که از روابط فوق استنباط می‌شود برای نقاط داده‌ای که X_i از آن‌ها بیشترین و کمترین فاصله را دارد، مقدار e_{ij} بسیار کم است (نزدیک به صفر) و برای نقاطی که از آن‌ها فاصله‌ای برابر با فاصله میانگین فاصله‌اش با سایر نقاط در مجموعه داده را دارد، بیشترین مقدار را داراست (نزدیک به یک).

۳-۲ روش تقسیم‌بندی:

در روش پیشنهادی، برای تقسیم‌بندی فضای داده در هر گره در هر پنجره زمانی مشخص ΔT، برای هر یک از نقاط داده با استفاده از روابطی که در بخش قبل بیان شد مقدار آنتروپی نقطه‌ای تجمعی و مقدار شباهت نقطه‌ای تجمعی را محاسبه می‌کنیم. طبق آنچه تاکنون بیان شد می‌توان نتیجه گرفت نقاطی که در قلب نواحی متراکم و نواحی پراکنده قرار دارند دارای مقدار e_i کمی هستند و نقاطی که در نواحی متراکم و فشرده فضا قرار دارند دارای S_i بزرگ و آن‌هایی که در نواحی پراکنده و پرت هستند دارای S_i کوچک هستند. درواقع معیار S_i می‌تواند نشان‌دهنده میزان تراکم نقاط در اطراف نقطه داده و موقعیت آن داده نسبت به سایر نقاط مجموعه داده باشد. با این فرض که بیشتر نقاط موجود در مجموعه داده نقاط عادی هستند می‌توان بیان کرد نقاط داده ناهنجار دارای شباهت نقطه‌ای تجمعی کم و آنتروپی نقطه‌ای تجمعی کم و نقاط داده عادی دارای شباهت نقطه‌ای تجمعی زیاد و آنتروپی نقطه‌ای تجمعی کم هستند. با تقسیم مقدار شباهت نقطه‌ای تجمعی بر آنتروپی نقطه‌ای تجمعی معیاری به دست می‌آید که با دقت بسیار بالایی نقاط داده مشابه را در یک ناحیه قرار می‌دهد. نسبت بین شباهت نقطه‌ای تجمعی و آنتروپی نقطه‌ای تجمعی را برای هر نقطه داده X_i با R_i نشان خواهیم داد، که مجموعه $R = \{R_1, R_2, \dots, R_i, \dots, R_n\}$ را برای هر نقطه داده $X_i \in X$ به شکل زیر تعریف می‌کنیم:

$$R_i = S_i / e_i \qquad (۷)$$

مقدار R_i برای نقاط در نواحی پراکنده (نقاط داده ناهنجار) کوچک و برای نقاط در نواحی فشرده (نقاط عادی) مقدار بزرگی خواهد شد. الگوریتم تقسیم‌بندی به این صورت عمل می‌کند که در ابتدا کمترین مقدار R_i انتخاب می‌شود و نقاطی که به‌اندازه یک انحراف معیار از آن فاصله دارند همگی در یک دسته قرار می‌گیرند و آن‌ها را از مجموعه داده حذف می‌کند. این رویه به‌صورت بازگشتی برای نقاط باقی‌مانده در مجموعه داده انجام می‌شود تا زمانی که دیگر نقطه‌ای در مجموعه داده باقی نماند. به‌این‌ترتیب فضای داده به نواحی موردنظر برای تشخیص ناهنجار یا عادی بودنشان، تقسیم می‌شود. لازم به ذکر است برای کم کردن اثرات نویز و نقاط پرت و بهبود بخشیدن به تقسیم‌بندی داده‌ها در نواحی موردنظر، ابتدا نقاط پرت را

حذف می‌کنیم. یک نقطه داده $X_i \in X$ یک نقطه پرت در مجموعه داده در نظر گرفته می‌شود درصورتی‌که مقدار S_{i} محاسبه‌شده برای آن نقطه کمتر از ۵ درصد میانگین S_{i} های کل گره باشد.

۳-۳ تشخیص ناهنجاری

در انتهای هر پنجره زمانی $\Delta \mathrm{T}$، هر گره حسگر بعد از تقسیم‌بندی فضای داده به نواحی موردنظر، تشخیص ناهنجاری را برای مشخص شدن وضعیت هر ناحیه انجام می‌دهد. برای بالا بردن دقت تشخیص ناهنجاری، آن را در دو قسمت گره‌های برگ و گره‌های والد به‌صورت جداگانه بیان خواهیم کرد:

الف- تشخیص ناهنجاری در گره‌های برگ:

در این حالت، ابتدا میانگین S_i ها و e_i های نواحی به‌دست‌آمده در هر گره برگ را محاسبه کرده و نسبت بین این دو مقدار را برای هر ناحیه به‌صورت جداگانه به دست می‌آوریم. فرض کنید $\mathrm{E} = \{\mathrm{E}_1, \mathrm{E}_2, \dots, \mathrm{E}_{\mathrm{i}}, \dots, \mathrm{E}_{\mathrm{k}}\}$ مجموعه نواحی به‌دست‌آمده در یک گره و k تعداد نواحی موجود برای آن گره باشد. برای هر ناحیه، مقدار میانگین S_i ها و e_i ها را که به ترتیب با $\overline{\mathrm{E}}_{\mathrm{S_i}}$ و $\overline{\mathrm{E}}_{\mathrm{e_i}}$ نشان داده می‌شوند، محاسبه کرده و مجموعه $\bar{R}$ را در هر گره به‌صورت زیر به دست می‌آوریم:

$$\bar{R} = \{\bar{R}_1, \bar{R}_2, \dots, \bar{R}_{\mathrm{K}}\} \qquad (۸)$$

$$\bar{R}_{\mathrm{i}} = \frac{\overline{\mathrm{E}}_{\mathrm{S_i}}}{\overline{\mathrm{E}}_{\mathrm{e_i}}}$$

برای تشخیص ناهنجار بودن یک ناحیه از قواعد زیر پیروی می‌شود:

۱) اگر در مجموعه $\bar{R}$، تمام مقادیر به‌دست‌آمده از یک کوچکتر بودند ناهنجاری در مجموعه داده وجود ندارد، اما اگر مشاهده شد بعضی از نواحی دارای مقدار $\bar{R}_{\mathrm{i}}$ بزرگتر از یک و بعضی کمتر از یک هستند، ناحیه‌هایی که دارای $\bar{R}_{\mathrm{i}}$ بزرگتر از یک هستند مستعد ناهنجاری‌اند و شرط بعدی برای آن نواحی باید چک شود:

۲) بین $\overline{\mathrm{E}}_{\mathrm{S_i}}$ نواحی میانگین گرفته و چنانچه ناحیه‌ای، $\overline{\mathrm{E}}_{\mathrm{S_i}}$ اش از میانگین به‌اندازه یک انحراف معیار استاندارد کمتر بود، ناحیه به‌عنوان ناهنجار معرفی می‌شود.

برای تشخیص ناهنجاری فرض شده اغلب داده‌های موجود در مجموعه داده، داده‌های عادی هستند و چنانچه جمعیت ناحیه‌ای که دارای $\bar{R}_{\mathrm{i}}$ بزرگ‌تر از یک است، برابر با داده‌های عادی باشد، تصمیم‌گیری در هر سطح به سطح بعدی واگذار می‌گردد. اگر در گره ریشه این رویه همچنان برقرار بود ناهنجاری اعلام نخواهد شد.

ب- تشخیص ناهنجاری در گره‌های والد:

بعد از تقسیم‌بندی داده‌ها، و مشخص شدن نواحی ناهنجار در گره‌های برگ، اطلاعاتی که در ادامه آمده است به گره‌های والد در سلسله‌مراتب بالاتر شبکه مخابره می‌شوند:

۱- میانگین داده‌های محلی هر ناحیه

۲- داده‌های محلی که به‌صورت ناهنجار شناسایی‌شده‌اند

در گره والد، چنانچه گره برگ داده ناهنجاری را گزارش نکرده باشد، رویه انجام شده برای تشخیص ناهنجاری در گره برگ انجام می‌شود. اما چنانچه داده ناهنجار توسط گره برگ گزارش‌شده باشد، برای پیدا کردن ناهنجاری در این سطح بین $\bar{R}_{\mathrm{i}}$ ها میانگین گرفته و هر ناحیه که $\bar{R}_{\mathrm{i}}$ اش کمتر از یک انحراف معیار استاندارد از میانگین $\bar{R}_{\mathrm{i}}$ ها بود، مستعد ناهنجاری اعلام می‌شود و شرط ۲ بیان‌شده در قسمت تشخیص ناهنجاری در گره‌های برگ برای آن اجرا خواهد شد. اگر ناحیه‌ای در سطوح قبلی، یک ناحیه عادی معرفی‌شده بود و در سطح جدید به‌عنوان یک ناحیه ناهنجار شناسایی شود، گره والد مربوطه از گره فرزندش درخواست می‌کند تمام نقاط داده مربوط به آن ناحیه را ارسال کند و در مقابل اگر ناحیه‌ای که در سطوح قبلی ناهنجار شناسایی‌شده بود در سطح جدید عادی شناسایی شود، بعد از اتمام مرحله تشخیص ناهنجاری، تنها میانگین داده‌های آن ناحیه به سطح بعدی ارسال خواهد شد. بر این اساس، تشخیص ناهنجاری در هر سطح از شبکه بر اساس سطوح متفاوتی از دقت ارزیابی می‌شود. این روند به‌صورت بازگشتی تا رسیدن به گره ریشه که در آن نتیجه نهایی ناهنجاری اعلام خواهد شد در سلسله‌مراتب شبکه ادامه پیدا می‌کند.

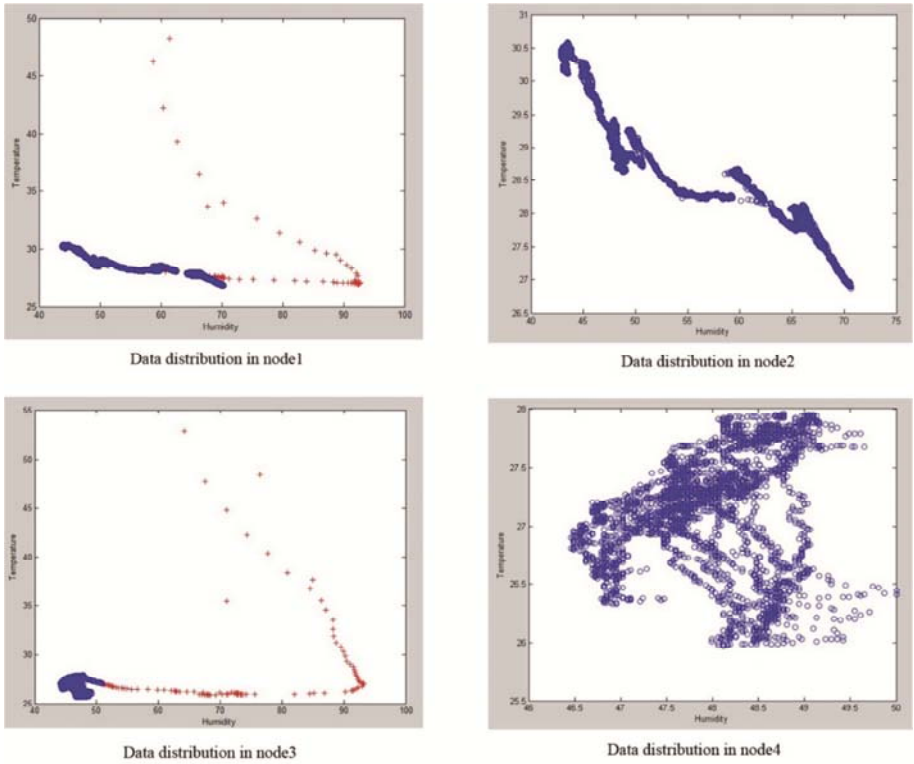
۴- ارزیابی و شبیه‌سازی

در این بخش، روش پیشنهادی با استفاده از یک مجموعه داده با توزیع پویا و چند تراکمی برای مشخص کردن دقت الگوریتم، مورد آزمایش قرار گرفته است. همچنین نتایج به‌دست‌آمده از شبیه‌سازی با روش پیشنهاد شده توسط Kumarage مقایسه شده است و میزان سربار ارتباطی ایجادشده از آن در مقایسه با روش‌های متمرکز و روش ارائه شده توسط Kumarage بیان‌شده است. آزمایش‌ها با استفاده از دو معیار حساسیت[1] و تشخیص[2] به‌عنوان دو معیار اصلی برای ارزیابی الگوریتم، انجام شده‌اند. روش ارائه شده برای یک مدل شبکه سلسله‌مراتبی سه سطحی از گره‌های حسگر، در محیط متلب پیاده‌سازی شده است، که در ادامه به شرح آن خواهیم پرداخت.

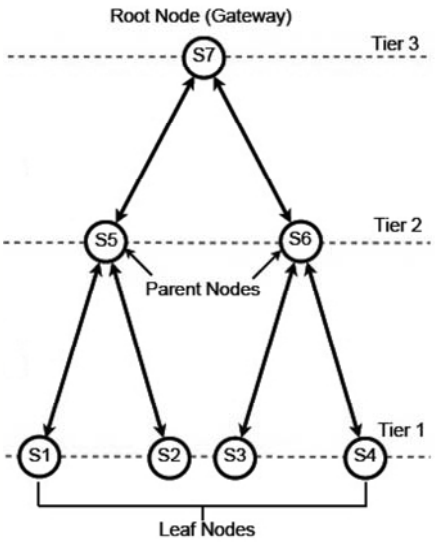
1. Sensitivity
2. Specificity

۴-۱ مجموعه داده

برای آزمایش صحت و مقاومت روش ارائه شده، از یک مجموعه داده چند تراکمی مربوط به یک کاربرد محیطی رایج در شبکه حسگر بی‌سیم، استفاده شده است. در این مجموعه داده، داده‌های حس شده مربوط به چهار گره حسگر (گره‌های ۱ تا ۴) با استفاده از ذرّه حسگر TelosB که نوسانات دما و رطوبت را در یک دوره ۶ ساعتی با فاصله‌های زمانی ۵ ثانیه‌ای اندازه‌گیری می‌کند، جمع‌آوری شده‌اند. ناهنجاری در ذره‌های حسگر ۱ و ۳ به‌صورت دستی با قرار دادن یک کتری آب جوش برای افزایش دما و رطوبت ایجاد شده است (suthaharan,2010). شکل ۱ کل مجموعه داده‌ها را با اجتماع تمام پنجره‌های زمانی در هر گره حسگر نمایش می‌دهد. در هر گره، ۴۰۰۰ داده در نظر گرفته شده است که در شکل، داده‌های ناهنجار با علامت بعلاوه و داده‌های عادی با دایره در هر گره حسگر مشخص شده‌اند. حسگرهای ۱ و ۳ دارای داده ناهنجار هستند درحالی‌که تمام داده‌ها در حسگرهای ۲ و ۴ عادی می‌باشند.



شکل ۱ ـ مجموعه داده‌ها در گره‌های ۱ تا ۴



شکل ۲ ـ ساختار سلسله‌مراتبی شبکه شبیه‌سازی‌شده (Kumarage,2015:4)

برای شبیه‌سازی روش پیشنهادی از یک مدل شبکه سلسله‌مراتبی سه سطحی نشان داده شده در شکل ۲ استفاده شده است، که داده‌های حسگرهای ۱ تا ۴، به ترتیب به گره‌های برگ S1 تـا S4 داده شـده‌انـد. بنابراین گره‌های S1 و S3 دارای ناهنجاری بوده درحالی‌که گره‌های S2 و S4 شامل فقط داده‌های عادی هستند. پنجره زمانی ΔT در آزمایش‌ها شامل ۴۰۰ نقطه داده برای هر گره برگ در هر دور می‌شود.

۴-۲ ارزیابی و نتایج روش پیشنهادی:

صحت و دقت روش پیشنهادی در هر گره و در هر رده از سلسله‌مراتب بـر اسـاس نـرخ مثبـت کـاذب (FPR)[1] و نرخ منفی کاذب (FNR)[2] نشان داده شده است که با توجه به تعداد مثبـت‌هـای کـاذب (FP) و منفی‌های کاذب مشاهده‌شده (FN) می‌باشد. مثبت کاذب، یک نمونه اندازه‌گیـری شـده عـادی اسـت کـه به‌عنوان ناهنجاری شناسایی می‌شود درحالی‌که منفی کاذب، یک نمونه اندازه‌گیری شده ناهنجار اسـت کـه به‌عنوان داده عادی شناسایی می‌شود. نـرخ مثبـت کـاذب نسـبت بـین تعـداد مثبـت‌هـای کـاذب و تعـداد

1. False Positive Rate
2. False Negative Rate

اندازه‌گیری‌های عادی و نرخ منفی کاذب نسبت بین تعداد کاذب‌های منفی و تعداد ناهنجاری‌های شناسایی‌شده می‌باشد:

$$FPR = \frac{FP}{(FP+TP)} \quad (۹)$$

$$FNR = \frac{FN}{(FN+TN)} \quad (۱۰)$$

در روابط فوق از مقادیر مثبت واقعی (TP) و منفی واقعی (TN) استفاده شده است. با استفاده از مقادیر اندازه‌گیری شده، می‌توان دو معیار حساسیت و تشخیص را محاسبه نمود. حساسیت با تقسیم تعداد داده‌های عادی که عادی شناسایی‌شده‌اند بر تعداد کل داده‌های عادی، محاسبه می‌شود و نشان می‌دهد یک آزمون با چه احتمالی داده‌های عادی را در یک مجموعه داده درست شناسایی می‌کند، درحالی‌که تشخیص، با تقسیم تعداد داده‌های ناهنجاری که درست شناسایی‌شده‌اند بر تعداد کل ناهنجاری‌ها محاسبه می‌شود و نشان می‌دهد یک آزمون با چه احتمالی داده‌های ناهنجار را در یک مجموعه داده به‌درستی شناسایی می‌کند:

$$sensitivity = \frac{TP}{(TP+FP)} \quad (۱۱)$$

$$specificity = \frac{TN}{(TN+FN)} \quad (۱۲)$$

جدول ۱ نتایج حاصل از شبیه‌سازی روش پیشنهادی را در گره‌های برگ نشان می‌دهد. در گره S1 از ۵۸ نقطه داده ناهنجار موجود در مجموعه داده، روش پیشنهادی در مرحله تقسیم‌بندی، تعداد ۵ نقطه ناهنجار را در دسته داده‌های عادی دسته‌بندی کرد. این نقاط ناهنجار با داده‌های عادی کاملاً تلاقی داشته و امکان تفکیک آن‌ها بر اساس روش پیشنهاد شده وجود نداشت؛ بااین‌حال این روش در تقسیم‌بندی داده‌های ناهنجار و عادی در گره S3 با وجود ۱۰۰ نقطه ناهنجار در مجموعه داده توانست به عدد صفر برای FNR رسیده و تمام داده‌های ناهنجار را از داده‌های عادی تفکیک کند. در گره S1 هیچ داده‌ای از داده‌های عادی در دسته نقاط ناهنجار قرار نگرفتند، اما در گره S3 به علت تلاقی شدید نقاط در مرز داده‌های عادی و ناهنجار، تعداد ۳ نقطه عادی در دسته داده‌های ناهنجار قرار گرفتند که مقدار FPR بسیار کمی را برای این گره به وجود آورد. در گره‌های S2 و S4 تمام داده‌های درست در این سطح به‌درستی شناسایی شدند که برای هر دو گره، مقدار FPR صفر مشاهده شد. در فاز تشخیص ناهنجاری، یک ناحیه از داده‌های عادی در گره S1 در دسته داده‌های ناهنجار قرار گرفت که مقدار FPR برابر با ۲ درصد را برای این گره نشان داد.

جدول ۲ نتایج به‌دست‌آمده برای داده‌های مشابه را با استفاده از روش بیان شده در مقاله Kumarage نشان می‌دهد. همان‌طور که از دو جدول مذکور می‌توان مشاهده کرد، در فاز اجرایی هر دو الگوریتم در گره‌های برگ، برتری روش پیشنهادی به‌وضوح قابل‌مشاهده است.

جدول ۱ ـ دقت تقسیم‌بندی روش پیشنهادی برای تشخیص ناهنجاری در گره‌های برگ

Sensor Node	FNR%	FPR%	Sensitivity%	Specificity%
S1	0.86	2.47	99.14	97.53
S2	NAN	0.00	NAN	100
S3	0.00	0.10	100	99.90
S4	NAN	0.00	NAN	100

جدول ۲ ـ دقت تقسیم‌بندی روش ارائه‌شده توسط Kumarage برای تشخیص ناهنجاری در گره‌های برگ

Sensor Node	FNR%	FPR%	Sensitivity%	Specificity%
S1	2.93	2.05	97.07	97.95
S2	NAN	5.37	NAN	94.63
S3	3.90	2.03	96.10	97.97
S4	NAN	2.45	NAN	97.55

جدول ۳ نتایج حاصل از روش پیشنهادی در ردیف دوم از سلسله‌مراتب شبکه را در گره‌های والد نشان می‌دهد. در این فاز، میانگین داده‌های نواحی ایجاد شده از گره‌های S1 و S2 در گره S5 و میانگین نواحی ایجاد شده از گره‌های S3 و S4 در گره S6 به‌عنوان نماینده داده‌های هر ناحیه، برای شناسایی ناهنجاری در سطح بالاتر موردبررسی قرار می‌گیرند. در گره S1 ناحیه‌ای از داده‌های عادی که در مرحله قبل به‌عنوان یک ناحیه ناهنجار شناسایی‌شده بود، به‌عنوان یک ناحیه عادی شناسایی شد. در گره S2 تعداد ۴ داده از داده‌های عادی به‌صورت ناهنجار شناسایی شدند که مقدار ناچیزی به FPR اضافه کرد. در این جدول همچنین نتایج به‌دست‌آمده از شبیه‌سازی روش پیشنهادی Kumarage آمده است که برتری روش پیشنهادی نسبت به آن قابل‌مشاهده می‌باشد.

جدول ۳ ـ دقت طبقه‌بندی روش پیشنهادی و روش ارائه‌شده توسط Kumarage در گره‌های والد برای تشخیص ناهنجاری

Sensor Node	FNR%	FPR%	Sensitivity%	Specificity%
S1,S2 (Proposed Method)	0.86	0.05	99.14	99.95
S3,S4 (Proposed Method)	0.00	0.03	100	99.97
S1,S2 (Kumarage Method)	2.93	3.72	97.07	96.28
S3,S4 (Kumarage Method)	3.90	1.97	96.10	98.03

درنهایت نتایج حاصل از اجرای روش پیشنهادی در گره ریشه در جدول ۴ آمده است. در این گره، میانگین داده‌های نواحی ایجاد شده در سلسله‌مراتب‌های قبلی تمام گره‌ها به‌عنوان نماینده هر ناحیه برای شناسایی ناهنجاری در بالاترین سطح مورداستفاده قرار می‌گیرد. همان‌طور که در جدول مشاهده می‌کنید، در این سطح، داده‌ها به‌خوبی تقسیم‌بندی شده‌اند. نهایتاً برتری روش پیشنهادی بر روش Kumarage در این جدول مشاهده می‌شود.

جدول ۴ ـ دقت طبقه‌بندی روش پیشنهادی و روش ارائه‌شده توسط Kumarage در گره ریشه برای تشخیص ناهنجاری

Sensor Node	FNR%	FPR%	Sensitivity%	Specificity%
S1,S2,S3,S4 (Proposed Method)	0.31	0.04	99.69	99.96
S1,S2,S3,S4 (Kumarage Method)	3.54	2.22	96.46	97.78

۴-۳ پیچیدگی ارتباطی:

با توجه به محدودیت ذاتی انرژی در شبکه‌های حسگر بی‌سیم، این‌که هر الگوریتم ارائه شده، سربار مصرف انرژی را کاهش دهد یک امر حیاتی است. ازآنجاکه در شبکه‌های حسگر بی‌سیم، بخش عمده انرژی مصرفی مربوط به ارتباطات داده‌ها می‌باشد، بنابراین تأکید ویژه‌ای روی کاهش مقدار داده‌های مخابره شده توسط هر حسگر وجود دارد. جدول ۵، روش پیشنهادی را با توجه به تعداد پیام‌های کنترلی مخابره شده با روش متمرکز و جدول ۶ روش پیشنهادی را با روش ارائه شده توسط Kumarage مقایسه

می‌کند. در هر دو جدول، کم شدن قابل‌توجهی از سربار ارتباطی در روش پیشنهادی کاملاً مشخص است. در روش متمرکز، کلیه داده‌های حس شده توسط حسگرها به ایستگاه پایه ارسال شده و تشخیص ناهنجاری به‌صورت متمرکز در آنجا انجام می‌شود.

جدول ۵ ـ مقایسه پیچیدگی ارتباطی روش پیشنهادی و روش‌های متمرکز

Sensor Node	Number of Messages (proposed Method)	Number of Messages (Centralized)	Overhead Reduction%
S1	195	4000	95.12
S2	48	4000	98.80
S3	167	4000	98.33
S4	61	4000	95.82
S5	144	8000	98.18
S6	228	8000	97.15

جدول ۶ ـ مقایسه پیچیدگی ارتباطی روش پیشنهادی و روش ارائه‌شده توسط Kumarage

Sensor Node	Number of Messages (Proposed Method)	Number of Messages (Kumarage Method)	Overhead Reduction%
S1	195	225	13.33
S2	48	347	86.16
S3	167	253	33.99
S4	61	251	75.69
S5	144	558	74.19
S6	228	466	51.07

۵- نتیجه‌گیری

در این مقاله، یک روش جدید برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم ارائه شده است. در این روش، داده‌ها در ناحیه‌های مختلفی به‌صورت پویا بر اساس نسبت بین شباهت نقطه‌ای تجمعی و آنتروپی نقطه‌ای تجمعی تقسیم‌بندی می‌شوند. این روش تقسیم‌بندی بر چالش چند تراکمی و پویا بودن مشاهدات در نظر گرفته شده، غلبه کرده است و درنتیجه برای هر ناحیه، رفتارهای ناهنجار با دقت بالا

شناسایی‌شده‌اند. مقدار آستانه برای تعریف تفاوت بین عادی و ناهنجار بودن به‌صورت وفقی از آمارگان دانش در دسترس در هر مرحله مشتق می‌شود. روش ارائه شده از پردازش درون‌شبکه‌ای استفاده کرده و قادر به شناسایی ناهنجاری در سطوح مختلف شبکه می‌باشد. نتایج آزمایش‌ها، دقت تشخیص بالایی را برای هر دو نوع داده‌های ناهنجار و عادی با نرخ مثبت کاذب و منفی کاذب کم نشان می‌دهند. کاهش سربار ارتباطی سبب گسترش کارایی این روش برای استفاده در شبکه‌های حسگری مقیاس بزرگ می‌شود، که این یک امر برجسته است.

۶- پیشنهادات

با توجه به بستر آماده‌ای که روش پیشنهادی برای تشخیص ناهنجاری در شبکه‌های حسگر بی‌سیم فراهم کرده است، پیشنهاد می‌شود برای مشخص شدن صحت و دقت عملکرد الگوریتم، این روش بر روی مجموعه داده‌های بیشتری بخصوص مجموعه داده‌هایی که ناهنجاری در آن‌ها به‌صورت واقعی رخ داده است، مورد آزمایش قرار گیرد. در روش پیشنهادی به جای معیار شباهت نقطه‌ای تجمعی می‌توان از معیارهای دیگری استفاده کرده و کارایی روش پیشنهادی را با آن معیارها مورد ارزیابی قرار داد. همچنین می‌توان روش پیشنهادی را با روش‌های منطق فازی ادغام کرده و کارایی آن را مورد آزمایش قرار داد.

منابع و مآخذ:

Suthaharan Shan, Leckie Christopher, Moshtaghi Masud, Karunasekera Shanika, and Rajasegarar Sutharshan (2010), Sensor data boundary estimation for anomaly detection in wireless sensor networks, in: Mobile Adhoc and Sensor Systems (MASS), IEEE 7th International Conference on, pp. 546-551.

Abduvaliyev Abror, Pathan Al-Sakib Khan, Zhou Jianying, Roman Rodrigo, and Wong Wai-Choong (2013), On the Vital Areas of Intrusion Detection Systems in Wireless Sensor Networks, in: IEEE communications surveyas & tutirials, vol. 15, no. 3, pp. 1223-1237.

Kumarage Heshan, Khalil Ibrahim, Tari Zahir, Zomaya Albert (2013), Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling, in:journal parallel and distributed computing,vol. 73,no. 6, pp. 790-806.

Pietro Di, Roberto, V. Mancini Luigi, Soriente Claudio, Spognardi Angelo, and Tsudik Gene (2009), Data security in unattended wireless sensor networks, in: IEEE Transaction, vol. 58, no. 11,pp. 1500-1511.

Rajasegarar, Leckie and Palaniswami, (2009), Detecting data anomalies in wireless sensor networks, in *Secur. Ad Hoc. Sens. Netw*, no.*3*, pp. 231-259.

Chandola, Varun, Banerjee Arindam, and Kumar Vipin (2009), Anomaly detection: A survey, in: ACM computing surveys (CSUR), vol. 41, no. 3, pp. 15.

Hodge, Victoria J., and Austin Jim (2004), A survey of outlier detection methodologies, in: Artificial Intelligence Review, vol.22, no. 2, pp. 85-126.

Ramaswamy, Sridhar, Rastogi Rajeev, and Shim Kyuseok (2000), Efficient algorithms for mining outliers from large data sets, in: ACM SIGMOD Record, Vol. 29, No. 2, pp. 427-438.

Breunig, M. Markus, Kriegel Hans-Peter, T. Ng Raymond, and Sander Jörg (2000), LOF: identifying density-based local outliers, in: ACM sigmod record, vol. 29, no. 2, pp. 93-104

Xie, Miao, Hu Jiankun, Han Song, and Chen Hsiao-Hwa (2013), Scalable hypergrid k-NN-based online anomaly detection in wireless sensor

networks, in: *Parallel and Distributed Systems, vol.* 24, no. 8, pp. 1661-1670.

Papadimitriou, Spiros, Kitagawa Hiroyuki, B. Gibbons Philip, and Faloutsos Christos (2003), Loci: Fast outlier detection using the local correlation integral, in: Data Engineering. Proceedings. 19th International Conference on IEEE, pp. 315-326.

Rokach Lior (2009), A survey of clustering algorithms, in: Data mining and knowledge discovery handbook. Springer US, pp. 269-298.

Xu Rui, and Wunsch Donald (2005), Survey of clustering algorithms, in: Neural Networks, vol. 16,no. 3, pp. 645-678.

Yao, J., M. Dash, S. T. Tan, and H. Liu (2000), Entropy-based fuzzy clustering and fuzzy modeling, Fuzzy Sets and Systems, vol. 113, no. 3, pp. 381-388.

Kumarage Heshan, Khalil Ibrahim, Tari Zahir (2015), Granular Evaluation of Anomalies in Wireless Sensor Networks using Dynamic Data Partitioning with an Entropy Criteria, Computers, IEEE Transactions on vol. 64, no. 9, pp. 2573-2585.