

Research Article

Improving the efficiency of an intrusion detection system using a decision tree adjusted with an improved random fractal search, combined with the whale algorithm

Shahrzad Rahimi ¹  | Masood Niazi Torshiz ^{2*}  | Seyyed Abed Hosseini ³ 

¹Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, Shahrzad.rahimi00@gmail.com

²Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, Masood.niazi@iau.ac.ir

³Department of Electrical Engineering, Islamic Azad University, Mashhad, Iran, Sa.hosseini@iau.ac.ir

Correspondence

*Masood Niazi Torshiz, Assistant Professor, Department of Computer Engineering, Islamic Azad University, Mashhad, Iran
Masood.niazi@iau.ac.ir

Main Subjects: Improving intrusion detection system

Received: 23 October 2025

Revised: 18 January 2025

Accepted: 12 February 2025

<https://doi.org/...>

Abstract

With the increasing use of the Internet, a large amount of information is exchanged between different communication devices. Data must be securely transmitted between communication devices, and therefore, network security is one of the dominant research areas for the current network scenario. Intrusion detection systems are widely used in conjunction with other security mechanisms such as firewalls and access control. On the other hand, the evolution in attack scenarios has been such that finding efficient and optimal intrusion detection systems with frequent updates has become a big challenge. Implementation of an intrusion detection system using machine learning techniques and updated intrusion data sets is one of the effective modeling solutions for an intrusion detection system. In this article, an improved random fractal search algorithm with chaos maps is introduced. To improve the mining ability of this algorithm, the relationships of the whale algorithm have been used in combination with it. This algorithm has been used to cluster infiltrated and normal data. Then, a decision tree classifier was used to classify the infiltrated data on the NSL-KDD

dataset. To evaluate the proposed method, its results have been compared with the case without clustering. In terms of the error and sensitivity criteria for the test data, the proposed method is equal to 0.2211, 0.5816, and the decision tree without clustering is equal to 0.2477, 0.5692, respectively. In terms of specificity and accuracy criteria, the proposed method has obtained better results compared to the decision tree without clustering. Therefore, the results showed that the proposed method has better efficiency and performance compared to the decision tree method without clustering.

Keywords: Intrusion Detection System, Random Fractal Search Algorithm, Chaotic Mapping, Whale Optimization Algorithm, Clustering.

پژوهشی

بهبود کارایی سیستم تشخیص نفوذ با استفاده از درخت تصمیم تنظیم شده با جستجوی فرکتال تصادفی بهبودیافته ترکیب شده با الگوریتم نهنگ

شهرزاد رحیمی^۱ | مسعود نیازی ترشیز^۲ | سیدعابد حسینی^۳

چکیده:

با افزایش استفاده از اینترنت، حجم زیادی از اطلاعات بین دستگاه های ارتباطی مختلف رد و بدل می شود. داده ها باید به طور ایمن بین دستگاه های ارتباطی منتقل شوند و بنابراین، امنیت شبکه یکی از حوزه های تحقیقاتی غالب برای سناریوی شبکه فعلی است. سیستم های تشخیص نفوذ به طور گسترده همراه با مکانیسم های امنیتی دیگر مانند فایروال و کنترل دسترسی استفاده می شوند. از سویی دیگر تکامل در سناریوهای حمله به گونه ای بوده است که یافتن سیستم های تشخیص نفوذ کارآمد و بهینه با به روزرسانی های مکرر به یک چالش بزرگ تبدیل شده است. پیاده سازی سیستم تشخیص نفوذ با استفاده از روش های یادگیری ماشین و مجموعه داده های نفوذ بروز شده یکی از راه حل های مدل سازی مؤثر سیستم تشخیص نفوذ است. در این مقاله یک الگوریتم جستجوی فرکتال تصادفی بهبودیافته با نگاشت های آشوب معرفی شده است. برای بهبود توانایی استخراج این الگوریتم، از روابط الگوریتم نهنگ به طور ترکیبی با آن استفاده شده است. از این الگوریتم به منظور خوشه بندی داده های نفوذ و عادی استفاده شده است. سپس از یک طبقه بندی کننده درخت تصمیم برای دسته بندی داده های نفوذ بر روی مجموعه داده NSL-KDD استفاده شده است. برای ارزیابی روش پیشنهادی نتایج آن با حالت بدون خوشه بندی مقایسه شده است. از نظر معیار خطا و حساسیت برای داده آزمایشی روش پیشنهادی به ترتیب برابر با ۰.۲۲۱۱، ۰.۵۸۱۶ و درخت تصمیم بدون خوشه بندی به ترتیب برابر با ۰.۲۴۷۷ و ۰.۵۶۹۲ به دست آمده است. از نظر معیارهای خصوصیت و صحت نیز روش پیشنهادی نتایج بهتری در مقایسه با درخت تصمیم بدون خوشه بندی به دست آورده است؛ بنابراین نتایج نشان داد که روش پیشنهادی کارایی و عملکرد بهتری در مقایسه با روش درخت تصمیم بدون خوشه بندی دارد.

^۱ گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

Shahrzad.rahimioo@gmail.com

^۲ گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

Masood.niazi@iau.ac.ir

^۳ گروه برق، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

Sa.hosseini@iau.ac.ir

نویسنده مسئول

* مسعود نیازی ترشیز، استادیار، گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

Masood.niazi@iau.ac.ir

موضوع اصلی: بهبود سیستم تشخیص نفوذ

تاریخ دریافت: ۱۰ خرداد ۱۴۰۴

تاریخ بازنگری: ۳۰ مرداد ۱۴۰۴

تاریخ پذیرش: ۲۷ شهریور ۱۴۰۴

<https://doi.org/...>

کلیدواژه: سیستم تشخیص نفوذ، الگوریتم جستجوی فرکتال تصادفی، نگاشت آشوبی، الگوریتم بهینه سازی نهنگ، خوشه بندی

۱- مقدمه

یکی از مهم‌ترین و قوی‌ترین ابزارهای حفظ امنیت شبکه، سیستم تشخیص نفوذ و حملات احتمالی از سوی مهاجمان به شبکه است. پس از تشخیص، راهکارهای لازم جهت جلوگیری از نفوذ انجام می‌گردد تا درصد موفقیت مهاجم کاهش یابد. هدف مهاجم، دستیابی غیرمجاز به اطلاعات، خرابکاری در سیستم‌ها و انتشار ویروس و حمله به کامپیوترهای شبکه است. برای ایجاد امنیت کامل در یک سیستم کامپیوتری، علاوه بر دیوارهای آتش و دیگر تجهیزات جلوگیری از نفوذ، سیستم‌های دیگری به نام سیستم‌های تشخیص نفوذ^۱ موردنیاز هستند تا بتوانند در صورتی که نفوذگر از دیواره‌ی آتش، آنتی‌ویروس و دیگر تجهیزات امنیتی عبور کرد و وارد سیستم شد، آن را تشخیص داده و چاره‌ای برای مقابله با آن بیندیشند. این فناوری یکی از مباحث روز فناوری در عصر جدید اطلاعات و ارتباطات است که با فراگیر شدن کاربرد آن در صنایع مختلف، مسئله‌ی امنیت و حریم خصوصی آن توجه زیادی را به سمت خود جلب نموده است و به موضوعی مهم در این حوزه تبدیل شده است [۱-۳].

سامانه‌های تشخیص نفوذ به صورت سامانه‌های نرم‌افزاری و سخت‌افزاری ایجاد شده و هرکدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزایای سیستم‌های سخت‌افزاری است و عدم شکست امنیتی آن‌ها توسط نفوذگران، قابلیت دیگر این گونه سیستم‌ها است. مهاجمان به طور مداوم، روش‌های حمله را برای دور زدن سیستم دفاعی طراحی می‌کنند. بسیاری از حملات برای به دست آوردن اعتبار کاربر که به آنها امکان دسترسی به شبکه و داده‌ها را می‌دهد، از بدافزارها یا مهندسی اجتماعی دیگر استفاده می‌کنند. یک سیستم تشخیص نفوذ شبکه برای امنیت شبکه بسیار مهم است، زیرا این امکان را فراهم می‌کند تا ترافیک مخرب را ردیابی کرده و پاسخ مناسب به آن داده شود. [۴].

هدف اصلی یک سیستم تشخیص نفوذ اطمینان از اطلاع کارکنان فناوری اطلاعات در هنگام حمله یا نفوذ شبکه است. یک سیستم شناسایی نفوذ شبکه هم ترافیک ورودی و خروجی در شبکه را کنترل می‌کند، هم داده‌های موجود بین سیستم‌های درون شبکه را مرور می‌کند. سیستم شناسایی نفوذ شبکه نظارت بر ترافیک شبکه را انجام می‌دهد و در صورت شناسایی فعالیت مشکوک یا تهدیدهای شناخته شده، هشدار ایجاد می‌کند، بنابراین کارکنان بخش فناوری اطلاعات می‌توانند با دقت بیشتری بررسی کرده و اقدامات مناسب را برای مسدود کردن یا متوقف کردن حمله انجام دهند [۵].

در این تحقیق مسئله‌ی تشخیص نفوذ شبکه‌های رایانه‌ای مورد توجه قرار گرفته شده است. بدین منظور ارائه‌ی یک روش مناسب برای رسیدن به این هدف موردنیاز است؛ بنابراین یک روش جدید مبتنی بر بهینه‌سازی جستجوی فرکتال تصادفی بهبود یافته با نداشت‌های آشوب معرفی شده است که برای بهبود توانایی استخراج آن از الگوریتم نهنگ استفاده شده است. از این الگوریتم به منظور خوشه‌بندی داده‌های شبکه استفاده شده است که در آن هدف خوشه‌بندی به دو دسته‌ی عادی و نفوذ یافته است. سپس به کمک یک روش طبقه‌بندی مناسب و ساده به نام درخت تصمیم داده‌های مربوط به نفوذ یافته طبقه‌بندی می‌شوند. ساختار مقاله به صورت زیر است. در بخش دوم مروری بر کارهای گذشته بیان شده است. در بخش سوم روش تحقیق ارائه شده است. در بخش چهارم شبیه‌سازی‌ها و نتایج ارزیابی‌ها آورده شده‌اند. در نهایت در بخش پنجم نتیجه‌گیری و پیشنهادها ارائه شده است.

۲- مروری بر کارهای گذشته

با وجود استفاده از سیستم‌های تشخیص نفوذ به منظور شناسایی حملات مختلف، تعداد و پیچیدگی حملات سایبری ناشناخته افزایش یافته است. این منجر شده است که توزیع و ناهمگنی برنامه‌ها، خدمات شبکه‌های اینترنت محلی را پیچیده و چالش برانگیز می‌کند [۱]. حملات سایبری مخرب مسائل جدی امنیتی را ایجاد می‌کند که نیاز به یک سیستم جدید شناسایی، نفوذپذیر و

¹ Intrusion Detection System

قابل اعتماد برای شناسایی نفوذ را می طلبد. روش سیستم های تشخیص نفوذ ابزاری فعال برای تشخیص نفوذ است که برای شناسایی و طبقه بندی به موقع حملات یا نقض سیاست های امنیتی به طور خودکار در زیرساخت های سطح شبکه و سطح میزبان استفاده می شود. بر اساس رفتارهای نفوذی، سیستم تشخیص نفوذ را می توان به دودسته سیستم تشخیص نفوذ مبتنی بر شبکه^۱ و سیستم تشخیص نفوذ مبتنی بر میزبان^۲ طبقه بندی نمود و معمولاً برای دستیابی به حداکثر کارایی و امنیت این سیستم های تشخیص نفوذ به صورت ترکیبی نیز مورد استفاده قرار می گیرند که با نام سیستم های تشخیص نفوذ توزیع شده^۳ شناخته می شوند یک سیستم های تشخیص نفوذ که از رفتار شبکه استفاده می کند، سیستم تشخیص نفوذ مبتنی بر شبکه نامیده می شود. رفتارهای شبکه با استفاده از تجهیزات شبکه از طریق بازتاب توسط دستگاه های شبکه مانند سوئیچ ها، روترها و تپ های شبکه جمع آوری می شوند و به منظور شناسایی حملات و تهدیدهای احتمالی پنهان در ترافیک شبکه، تجزیه و تحلیل می شوند [۱]. در تحقیقات، روش های مختلفی برای سیستم تشخیص نفوذ توسط محققان انجام شده است. در [۲] یک سیستم تشخیص نفوذ مبتنی بر داده مفید با استفاده از هوش مصنوعی، به ویژه روش های یادگیری ماشین توسعه یافته است. در این مطالعه، دو روش طبقه بندی مختلف برای تشخیص نفوذ، که هر کدام موارد استفاده منحصر به فرد خود را دارند، مقایسه شدند. مقاله [۳] یکی از کارهای اولیه در مورد استفاده از طبقه بندی بود. در این کار، نویسندگان از فاصله همینگ استفاده کردند تا نشان دهند تماس های سیستم مخرب چگونه می توانند با "توالی عادی موجود" متفاوت باشند. نویسندگان در [۴] یک ماشین بردار پشتیبان^۴ و درخت تصمیم را پیشنهاد می کنند و بر اساس ادعای خود موفق ترین روش از نظر دقت در نظر گرفته می شود که معمولاً از پشتیبان گیری پشتیبانی می کند. الگوریتم های ماشینی با این حال، این روش تعداد ایستا از خوشه ها را فرض می کند که هنوز باید تعیین شوند. در [۵]، نویسندگان یک الگوریتم مبتنی بر یک الگوریتم ژنتیک چندهدفه برای شکل دادن به سیستم های تشخیص نفوذ پیشنهاد کردند. این الگوریتم خوشه بندی را به عنوان یک تابع بهینه سازی تعریف می کند که هدف آن یافتن حداقل فاصله خوشه و حداکثر فاصله بین اهداف خوشه است. سپس یک الگوریتم ژنتیک چندهدفه را حل می کند. در مقاله [۶] توابع بهینه ساز برای تعیین بهترین خوشه ها تعریف می شود و سپس با کمک الگوریتم های بهینه سازی، تعداد خوشه های بهینه مورد نیاز پیدا می گردد. نویسندگان در [۷] یک سیستم تشخیص نفوذ مبتنی بر ناهنجاری را با استفاده از روش SOM فازی پیشنهاد کردند. در [۸] یک مدل تشخیص نفوذ پیشنهاد شد که روش نیویز^۵ و DL را با هم استفاده کرد و از الگوریتم ژنتیک نیز برای انتخاب ویژگی خوب بهره برد. در [۹] نیز یک مدل تشخیص برای ارتباطات نفوذی پیشنهاد شده است. این کار تحقیقاتی پنج روش طبقه بندی را آزمایش می کند. همچنین در [۱۰] یک رویکرد طبقه بندی کننده ترکیبی با استفاده از الگوریتم درختی برای تشخیص نفوذ شبکه ارائه گردید و مدل بر روی مجموعه داده kdd با صحت ۸۹/۲۴٪ ارزیابی شده است. مقاله [۱۱] مدلی بر اساس الگوریتم جنگل تصادفی برای انتخاب یک ویژگی مهم ارائه کرد که این مدل با استفاده از NSL-KDD مورد ارزیابی قرار گرفت. در [۱۲] یک مدل یادگیری عمیق برای تشخیص نفوذ شبکه با استفاده از الگوریتم فرا ابتکاری دوگانه بهینه سازی ازدهام ذرات توسعه داده شد.

در [۱۳] کاربرد فناوری داده کاوی در تشخیص نفوذ شبکه و نگهداری امنیت بررسی شده است. در این مرجع از روش خوشه بندی k-means بهبود یافته به منظور تشخیص نفوذ و نگهداری امنیت استفاده شد. یک الگوریتم بهینه سازی گرگ خاکستری لاپلاسن مبتنی بر یادگیری مقابل با ماشین یادگیری پشتیبان در سیستم تشخیص نفوذ در [۱۴] ارائه شد. کوریشف و همکاران یک دسته بندی کننده فازی مبتنی بر الگوریتم بهینه سازی نهنگ برای تشخیص نفوذ در [۱۵] ارائه نمودند. روش تشخیص خوشه-بندی ویژگی نفوذ شبکه بر اساس ماشین بردار پشتیبان و الگوریتم بلوک ICA در [۱۶] به کار برده شد. در این مرجع ابتدا

¹ Network-based Intrusion Detection System

² Host-based Intrusion detection system

³ Distributed-based Intrusion detection system

⁴ Support Vector Machine

⁵ Naive Bayes classifier

ویژگی‌های غیرمفید حذف شد تا دقت خوشه‌بندی افزایش یابد. سپس از ماشین بردار پشتیبان برای طبقه‌بندی دسته‌های مختلف استفاده شد. یک پروتکل خوشه‌بندی نامتقارن مبتنی بر کیفیت خدمات جدید با سیستم تشخیص نفوذ در شبکه‌های حسگر بیسیم در [۱۷] انجام شد. در این مقاله سیستم فازی عصبی تطبیقی برای انتخاب سر خوشه‌ها به کار برده شد. سپس از الگوریتم شکار گوزن برای یافتن سرخوشه‌های بهینه استفاده شد. در [۱۸] یک سیستم تشخیص نفوذ با روش خوشه‌بندی K-Means بهبودیافته برای شبکه‌های حسگر بیسیم مبتنی بر یادگیری متحد ارائه شد. یک سیستم تشخیص نفوذ برای اینترنت اشیا با استفاده از مدل روش‌های بهینه‌سازی گرگ خاکستری، توده ذرات و جنگل تصادفی ترکیبی در [۱۹] پیشنهاد شد. در [۲۰] از الگوریتم بهینه‌سازی ملخ برای تشخیص غیرعادی بودن شبکه برای ایجاد سیستم تشخیص نفوذ مؤثر استفاده شد. مقالات [۲۱-۲۲] هم از روش بهینه‌سازی نهنگ برای تشخیص نفوذ استفاده کرده‌اند.

با بررسی تحقیقات انجام‌شده مشخص می‌شود که آینده فناوری سیستم‌های تشخیص نفوذ به سمت یادگیری ماشین و هوش مصنوعی می‌رود و پیشرفت‌های فعلی حاصل تحقیق و توسعه در این حوزه هستند. همچنین برای بهبود دقت فناوری سیستم‌های تشخیص نفوذ، کاهش خطاها، تشخیص موارد مثبت/منفی کاذب و غیره و باتوجه به پیچیدگی این سیستم‌ها، آن‌ها باید به‌درستی درک، توسعه و تجزیه و تحلیل شوند.

۳- روش تحقیق

در این بخش روش پیشنهادی تشخیص نفوذ به کمک خوشه‌بندی داده‌ها با الگوریتم فرکتال تصادفی بهبودیافته با الگوریتم نهنگ به‌منظور افزایش صحت تشخیص نفوذ معرفی می‌شود. خوشه‌بندی به‌عنوان یک رویکرد کشف الگوی بدون نظارت، فرایند تقسیم داده‌ها به گروه‌های مجزا است، به‌طوری‌که اشیای یک دسته بیشترین شباهت را به هم داشته باشند و با اشیای سایر کلاس‌ها متفاوت باشند. این مسئله که کاربردهای گسترده‌ای در حوزه‌های تحلیل پایگاه‌داده، پردازش متون و تصاویر، کاوش در شبکه‌های اجتماعی و کشف تقلب دارد، با معرفی مفهوم خوشه‌بندی K-means تاکنون محققین بسیاری را به خود جذب کرده است. توسعه و بهبود این الگوریتم در دهه‌های اخیر با ترکیب الگوریتم‌های فراابتکاری و K-means وارد فاز جدیدی شده است و باتوجه به مزایا و معایب این الگوریتم‌ها، تعداد زیادی از آنها تاکنون ارائه شده‌اند. در این تحقیق برای اولین بار الگوریتم جستجوی فرکتال اتفاقی به‌منظور بهینه‌سازی خوشه‌بندی k-means در کاربرد تشخیص نفوذ ارائه شده است. مقایسه عملکرد این الگوریتم با الگوریتم مشهور ژنتیک و بهینه‌سازی ازدحام ذرات نشان می‌دهد که الگوریتم جستجوی فرکتال موفق شده جواب بهینه سراسری را با کمترین خطا به دست آورد [۲۳].

مراحل طرح پیشنهادی به‌صورت زیر است. ابتدا پیش‌پردازش داده‌ها انجام می‌شود. داده‌هایی که به‌صورت نمادین هستند کمی‌سازی می‌شوند. سپس دو ویژگی ابتدایی در نظر گرفته‌شده و در خوشه‌بندی استفاده می‌شود. در مرحله خوشه‌بندی داده به کمک الگوریتم جستجوی فرکتال تصادفی بهبودیافته پیشنهادی خوشه‌بندی با نمونه‌های مختلف انجام می‌شود. در حالت خوشه‌بندی داده‌ها به عادی و نفوذ یافته تقسیم می‌شوند. سپس داده‌های نفوذ که خود شامل چهار دسته هستند به همراه داده‌های عادی که از قبل تشخیص داده شده‌اند به روش دسته‌بندی درخت تصمیم ارسال می‌شوند. روش درخت تصمیم بر این داده‌ها اعمال می‌شود و نتایج مربوط به ماتریس درهم‌ریختگی شامل دقت، خصوصیت، معیار F به دست می‌آیند.

در این بخش ابتدا درخت تصمیم و الگوریتم جستجوی فرکتال تصادفی معرفی می‌شود. در ادامه روش پیشنهادی استفاده از آشوب در الگوریتم فرکتال ترکیب‌شده با نهنگ به همراه فلوچارت آن پیشنهاد می‌شود. سپس استفاده از این الگوریتم در خوشه‌بندی داده‌های مربوط به تشخیص نفوذ بیان شده است.

درخت تصمیم^۱ یک نمایش سلسله مراتبی از تمام راه حل های ممکن برای یک مسئله ی تصمیم گیری یا طبقه بندی است [۲۴]. فضای مسئله داده شده را به کلاس های از پیش تعریف شده به دنبال دنباله ای از آزمون ها برای هر مقدار مشخصه طبقه بندی می کند. این آزمون ها توسط گره های داخلی درخت تصمیم انجام می شود. شاخه ها نشان دهنده نتایج احتمالی آزمون هستند و گره های خارجی نشان دهنده کلاس های فضاهای پارتیشن بندی شده هستند. هر چند ساخت، استفاده و تفسیر درخت تصمیم آسان است؛ اما با وجود این مزایا، درختان روش ایده آل برای یادگیری پیش بینی نیستند. درخت تصمیم دقت خوبی را با مجموعه داده شناخته شده ارائه می دهد. اما مدل ممکن است مجموعه داده ناشناخته را به طور دقیق طبقه بندی نکند. این به دلیل بیش برآزش^۲ است؛ بنابراین نرخ خطا در مقایسه با برخی از الگوریتم های دیگر بیشتر است.

۳-۱- الگوریتم جستجوی فرکتال تصادفی

الگوریتم جستجوی فرکتال از سه قانون زیر استفاده می کند:

- ۱- هر ذره یک انرژی پتانسیل الکتریکی دارد.
 - ۲- هر ذره منتشر و باعث تولید ذرات تصادفی دیگر می شود و انرژی هر ذره بین ذرات تولید شده، تقسیم می شود.
 - ۳- تنها کمی از بهترین ذرات در هر نسل بعدی باقی می ماند و باقی ذرات نادیده گرفته می شود.
- مهم ترین چالش جستجوی فرکتال، عدم تبادل داده بین افراد است؛ لذا در سال ۲۰۱۹ این الگوریتم توسط سلیمی [۲۳] بهبود یافت. الگوریتم بهبود یافته را الگوریتم جستجوی فرکتال اتفاقی^۳ می نامند.
- دو فرایند اصلی در فرکتال اتفاقی وجود دارد:

- فرایند انتشار: هر ذره در اطراف موقعیت جاری خود برای ارضای خاصیت استخراج^۴ انتشار می یابد. این فرایند شانس یافتن مینیمم سراسری را افزایش می دهد و همچنین از به دام افتادن در بهینه محلی جلوگیری می کند.
- فرایند به روزآوری: الگوریتم چگونگی به روز شدن یک نقطه در گروه را مبتنی بر موقعیت دیگر نقاط در گروه، شبیه سازی می کند.

۳-۲- بهبود الگوریتم فرکتال تصادفی به کمک آشوب

به منظور بهبود در الگوریتم جستجوی فرکتال تصادفی یک راه استفاده از نگاشت های^۵ آشوبی به جای جمعیت اولیه تصادفی و یا جمعیت اولیه از طریق روش یادگیری مبتنی بر مقابل است. نگاشت های آشوبی خاصیت شبه تصادفی دارند. به علت خاصیت غیر تکراری بودن تئوری آشوب، می توان الگوریتم های بهینه سازی را با آشوب طوری ترکیب نمود تا تنوع جمعیت افزایش یابد. این به جستجوی سراسری با سرعت بالاتر در مقابل با جستجوهای تصادفی که وابسته به احتمالات است، کمک می کند. از آنجایی که تنوع جمعیت افزایش می یابد، قابلیت همگرایی الگوریتم بیشتر می شود. همچنین از افتادن در بهینه محلی می تواند جلوگیری کند [۲۵].

الگوهای آشوبی تک بعدی سیستم های ساده ای هستند که قابلیت ایجاد حرکت های آشوبی را دارند و برای این کار مناسب اند. به عنوان نمونه الگوی آشوبی لجستیک به صورت معادله (۱) است [۲۵]:

$$x_{k+1} = ax_k(1 - x_k)$$

(۱)

¹ Decision Tree

² Overfitting

³ Stochastic Fractal Search - SFS

⁴ Exploitation

⁵ Maps

۷ بهبود کارایی سیستم تشخیص نفوذ با استفاده از درخت تصمیم... / شهرزاد رحیمی - مسعود نیازی ترشیز - سیدعابد حسینی

در این رابطه x_k ، k امین عدد آشوبناک و k نشانه عدد تکرار است. که $x \in (0,1)$ تحت این شرایط اولیه که $x_0 \in (0,1)$ و $\{0.0, 0.25, 0.75, 0.5, 1.0\}$ بر اساس تجربه $a=4$ در نظر گرفته می شود.

در قسمت تولید جمعیت اولیه از الگوی آشوبی استفاده می کنیم. همچنین در قسمت مربوط به انتشار نیز از نگاشت آشوبی استفاده می کنیم. در مرحله به روزآوری نیز از نگاشت آشوبی استفاده می کنیم. انتظار می رود که تنوع جمعیت افزایش یابد و جستجوی سراسری الگوریتم بهبود یابد.

۳-۳- بهبود الگوریتم جستجوی فرکتال تصادفی آشوبی با الگوریتم نهنگ

برای بهبود عملکرد الگوریتم از نظر توانایی استخراج آن، از فاز استخراج در الگوریتم نهنگ استفاده می کنیم. توضیحات آن در ادامه آورده شده است.

جهت مدل سازی ریاضی رفتار حباب - تور نهنگ های گوژپشت، دو روش به صورت زیر طراحی شده است:

۱. مکانیسم محاصره انقباضی: این رفتار به وسیله افزایش مقدار a در معادله ی ۳ حاصل می شود. به یاد داشته باشید که محدوده ی نوسان A به وسیله a کاهش می یابد. به عبارت دیگر، A مقداری تصادفی در فاصله ی a تا $-a$ است و a در طی تکرارها، از مقدار ۲ تا ۰ کاهش می یابد. با انتخاب مقادیر تصادفی A در فاصله ی ۱ تا -۱، می توان مکان جدید عامل جستجو را در هر کجای بین مکان اصلی عامل و مکان بهترین عامل کنونی، تعریف کرد.
۲. مکان در حال به روزرسانی مارپیچی: این روش در ابتدا فاصله بین نهنگ قرار گرفته در مختصات X, Y و طعمه موجود در X^*, Y^* را محاسبه می کند. معادله ای مارپیچی بین موقعیت نهنگ و طعمه ایجاد می شود تا حرکت حلزونی شکل نهنگ گوژپشت را تقلید کند:

$$X(t+1) = D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (2)$$

در این معادله

$$D = |X^*(t) - X(t)| \quad (3)$$

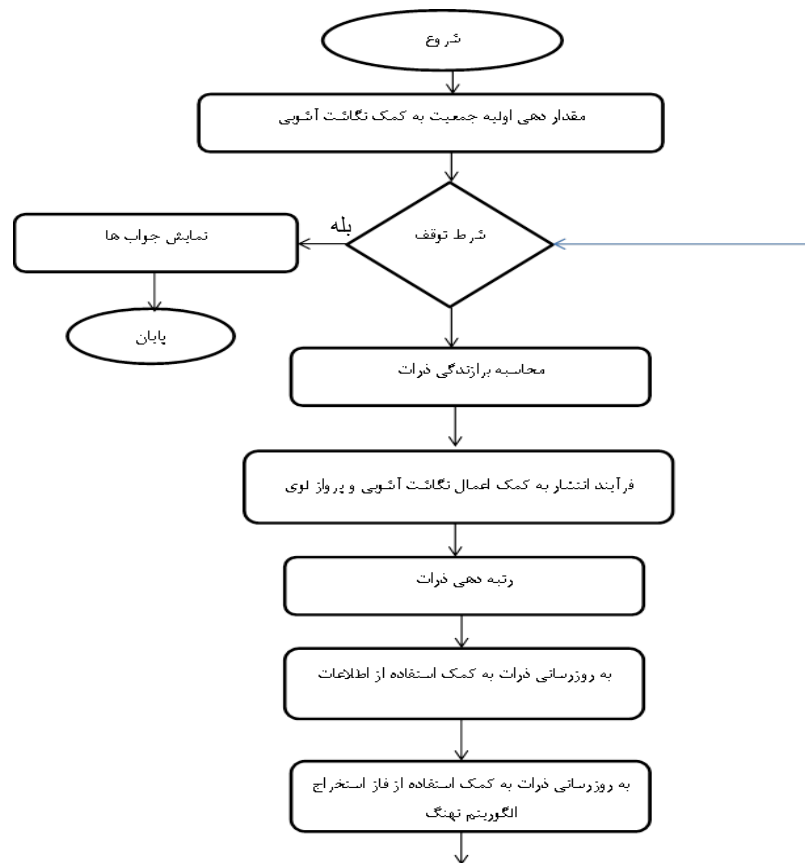
و به فاصله ی i امین نهنگ تا طعمه (بهترین راه حل به دست آمده تا اینجا) اشاره دارد، b ثابتی برای تعریف شکل مارپیچ لگاریتمی است، l عددی تصادفی بین ۱ تا -۱ و ۰ ضرب نقطه ای (المان در المان) است.

به یاد داشته باشید که نهنگ گوژپشت، حول طعمه در امتداد یک دایره ی انقباضی و هم زمان در مسیر مارپیچی شکلی به شنا درمی آید. جهت مدل سازی این رفتار هم زمان، فرض کرده ایم که نهنگ با احتمال ۵۰ درصد از بین مکانیسم محاصره ی انقباضی و یا مدل مارپیچی یکی را انتخاب می کند تا موقعیت نهنگ ها در طول بهینه سازی به روزرسانی شود. مدل ریاضی بدین صورت است:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D, & p < 0.5 \\ D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t), & p \geq 0.5 \end{cases} \quad (4)$$

که در آن p عددی تصادفی بین ۰ تا ۱ است.

از این فاز استخراج پس از اتمام مراحل الگوریتم جستجوی فرکتال تصادفی استفاده می کنیم. با اعمال این فاز، توانایی استخراج الگوریتم اصلی جستجوی فرکتال تصادفی بهبود خواهد یافت.

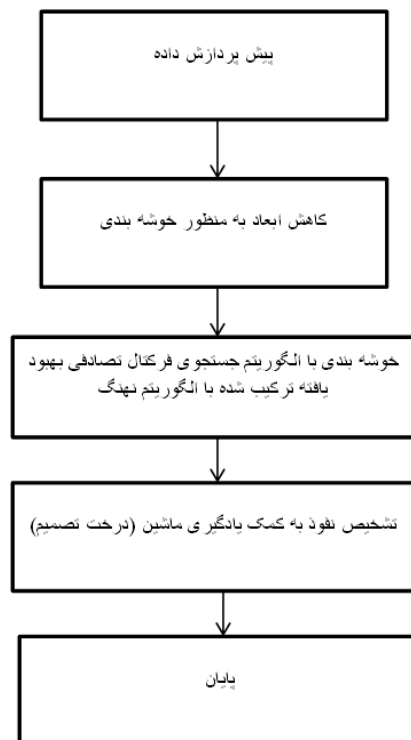


شکل ۱: الگوریتم فرکتال تصادفی بهبودیافته
Figure 1: Improved Random Fractal Algorithm

در این مقاله معیارهای مختلفی برای ارزیابی به کار برده شده است که مهم ترین آن ها در بخش بعدی توضیح داده شده اند.

۳-۴- روش جستجوی فرکتال تصادفی ترکیب شده با الگوریتم نهنگ و درخت تصمیم

طبقه بندی پدیده ها یا متغیرها از ارکان هر علمی است و تحلیل خوشه ای یکی از روش های تحلیل چند متغیره است که برای طبقه بندی عناصر یا متغیرها و تشخیص گروه های همگن به کار می رود. تحلیل خوشه ای طبقه بندی عناصر یا متغیرها به گروه های همگن است به گونه ای که عناصر (یا متغیرهای) هر گروه دارای بیشترین شباهت باهم و کمترین شباهت با عناصر (یا متغیرهای) گروه های دیگر باشند. خوشه بندی به عنوان یک رویکرد کشف الگوی بدون نظارت، فرآیند تقسیم داده ها به گروه های مجزا است، به طوری که اشیای یک کلاس بیشترین شباهت را به هم داشته باشند و با اشیای سایر کلاس ها متفاوت باشند. این مسئله که کاربردهای گسترده ای در حوزه های تحلیل پایگاه داده، پردازش متون و تصاویر، کاوش در شبکه های اجتماعی و کشف تقلب دارد. در این مقاله با توجه به ویژگی های روش خوشه بندی k-means از این روش به منظور خوشه بندی داده های عادی و نفوذ یافته استفاده می شود. فلوچارت مربوط به الگوریتم تشخیص نفوذ مبتنی بر خوشه بندی با الگوریتم جستجوی فرکتال تصادفی بهبود یافته جهت تشخیص حملات شبکه در شکل ۲ ارائه شده است.



شکل ۲: فلوچارت سیستم تشخیص نفوذ پیشنهادی

Figure 2: Flowchart of the Proposed Intrusion Detection System

باتوجه به شکل ۲ در مرحله اول پیش پردازش داده انجام می شود. در این مرحله داده ها باید طوری آماده شوند که در کاربرد خوشه بندی و طبقه بندی بتوان استفاده نمود. در مرحله پیش پردازش مراحل زیر انجام می شود:

- نوع پروتکل داده ها شامل TCP, UDP, ICMP است که به ترتیب برابر با ۱، ۲ و ۳ تعیین می کنیم.
- پایگاه داده شامل پنج دسته است که این دسته ها شامل Prob, R2L, U2R, DOS, Normal هستند.

در مرحله خوشه بندی، کل داده ها به دودسته حمله (نفوذ) و نرمال تقسیم بندی می شوند. این عمل به کمک الگوریتم پیشنهادی جستجوی فرکتال تصادفی بهبود یافته ترکیب شده با الگوریتم نهنگ انجام می شود.

پس از آنکه مرحله خوشه بندی به اتمام رسید، داده ها به عادی و حمله تقسیم بندی می شوند. این مرحله طوری انجام می شود که به طور کامل داده های عادی تشخیص داده می شوند. اما برای دسته بندی داده های حملات باید از طبقه بندی کننده استفاده کرد. در این تحقیق از طبقه بندی کننده درخت تصمیم به دلیل ساختار ساده و دقت مناسب آن استفاده می شود. در مرحله طبقه بندی داده ها به پنج دسته طبقه بندی می شوند. پیش از این مرحله داده های دسته پنجم (عادی) تشخیص داده شده بودند که در شبیه سازی و ارزیابی عملکرد طبقه بندی کننده لحاظ می شوند. داده کل شامل داده های آموزشی و آزمایشی است، دو مثال مختلف برای ارزیابی روش پیشنهادی با تعداد نمونه های متفاوت در نظر گرفته شده اند. برای نشان دادن عملکرد روش پیشنهادی نتایج آن با حالت بدون خوشه بندی مقایسه شده است.

۴- ارزیابی روش پیشنهادی

به منظور ارزیابی عملکرد الگوریتم پیشنهادی از مجموعه داده NSL-KDD استفاده می شود که در آن داده ها به پنج دسته نرمال و نفوذهای نوع Prob, R2L, U2R, DOS تقسیم شده اند. این مجموعه داده شامل ۴۱ ویژگی است. این مجموعه داده متشکل از ۱۲۵۹۷۳ نمونه است.

۴-۱- انواع معیارهای ارزیابی

در این بخش معیارهای ارزیابی برای دسته‌بندی بررسی می‌شود. برای ارزیابی باید برچسبی که دسته‌بند به حمله نسبت داده با برچسب مرجع مقایسه شود. برای این کار از ماتریس سردرگمی استفاده می‌شود [۲۶].

معیارهای عملکرد سیستم‌های تشخیص نفوذ ماتریس سردرگمی برای به‌تصویرکشیدن دسته‌های واقعی و پیش‌بینی‌شده در حملات امنیت سایبری استفاده می‌شود. ماتریس سردرگمی با عبارات زیر نشان داده می‌شود.

مثبت واقعی (TP): یک نمونه حمله به‌درستی به‌عنوان یک حمله پیش‌بینی می‌شود.

منفی واقعی (TN): یک نمونه عادی به‌درستی به‌عنوان یک نمونه غیر حمله یا عادی پیش‌بینی شده است.

مثبت کاذب (FP): یک نمونه عادی به‌اشتباه به‌عنوان حمله پیش‌بینی می‌شود.

منفی کاذب (FN): یک حمله واقعی به‌اشتباه به‌عنوان غیر حمله یا نمونه عادی پیش‌بینی می‌شود.

موارد مثبت کاذب در مواردی که یک فعالیت عادی شبکه به‌عنوان حمله طبقه‌بندی می‌شود، می‌تواند زمان ارزشمند مدیران امنیتی را تلف کند. منفی‌های کاذب بدترین تأثیر را بر سازمان‌ها دارند؛ زیرا حمله به‌هیچ‌وجه شناسایی نمی‌شود.

باتوجه به پارامترهای مطرح‌شده معیارهای ارزیابی مختلفی ارائه شده است که ازجمله مهم‌ترین آن‌ها می‌توان به معیارهای خصوصیت^۱، حساسیت^۲ و صحت^۳ اشاره کرد.

معیار صحت: مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی است و صحت کل یک دسته‌بندی را محاسبه می‌کند. این معیار نشان‌دهنده این موضوع است که چند درصد از کل مجموعه داده به‌درستی دسته‌بندی شده است.

$$\text{Accuracy} = \frac{T_p + T_N}{T_p + F_p + T_N + F_N} * 100 \quad (5)$$

معیار حساسیت: تعیین می‌کند که نمونه‌های حمله تا چه میزان در دسته مثبت قرار داده شده‌اند، یا به‌عبارت دیگر چه تعدادی از حملات به‌درستی تشخیص داده شده‌اند.

$$\text{Sensitivity} = \frac{T_p}{T_p + F_N} * 100 \quad (6)$$

معیار خصوصیت: تعیین می‌کند که چه تعدادی از نمونه‌های نرمال به‌درستی تشخیص داده شده‌اند.

$$\text{Specificity} = \frac{T_N}{F_p + T_N} * 100 \quad (7)$$

معیار دقت: یک متریک عمومی جهت اندازه‌گیری مفیدبودن الگوریتم پیشنهادی است که به‌صورت رابطه (۴) محاسبه می‌شود.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

معیار فراخوانی: یکی دیگر از متریک‌های ارزیابی در سیستم‌های تشخیص نفوذ است که به‌صورت رابطه (۲-۵) قابل محاسبه است.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

معیار F: از ترکیب دو پارامتر دقت و فراخوانی محاسبه می‌شود.

$$F - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

۴-۲- نتایج شبیه‌سازی

¹ Specificity

² Sensitivity

³ Accuracy

الف- حالت اول تعداد ۲۰ هزار نمونه

ابتدا ۲۰ هزار نمونه برای داده آموزش و ۲۰ هزار نمونه برای داده آزمایش در نظر می گیریم. الگوریتم جستجوی فرکتال تصادفی بهبود یافته ترکیب شده با الگوریتم نهنگ پیشنهادی را اجرا می کنیم. پس از اجرای الگوریتم نتیجه ی به دست آمده به این صورت است که تمام ۱۰۶۹۳ نمونه مربوط به داده های عادی به یک خوشه و مابقی به خوشه دیگر تعلق می یابند. در مرحله ی طبقه بندی از آنجایی که خوشه عادی درست تشخیص داده شده است، داده های مربوط به کلاس آن تخصیص می یابند، اما طبقه بندی برای دسته بندی داده های حملات انجام می شود. در ادامه نتایج مربوط به داده آموزشی در حالت ۲۰ هزار نمونه به کمک روش پیشنهادی آورده شده اند.

در جدول ۱ نتایج مربوط به پیش بینی تعداد داده ها در هر دسته به کمک روش پیشنهادی نشان داده شده است.

جدول ۱: نتیجه داده آموزشی با روش پیشنهادی

Table 1: Result of the Training Data Using the Proposed Method

دسته	پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵
دسته واقعی ۱	۷۳۳۹	۰	۰	۵	۲
دسته واقعی ۲	۰	۱۰	۰	۱	۰
دسته واقعی ۳	۰	۵	۱۴۸	۰	۶
دسته واقعی ۴	۱	۰	۰	۱۷۸۷	۳
دسته واقعی ۵	۰	۰	۰	۰	۱۰۶۹۳

باتوجه به جدول ۱ مشاهده می شود که تعداد ۷۳۳۹ نمونه به دسته ۱ اختصاص یافته اند که درست است، اما به تعداد ۵ نمونه و ۲ نمونه به ترتیب به دسته های ۴ و ۵ اختصاص یافته اند که نادرست است. از سطر سوم این جدول مشاهده می شود که تعداد ۱۰ نمونه به دسته دوم اختصاص یافته اند که درست بوده ولی به تعداد یک نمونه به دسته ۴ تخصیص یافته که نادرست است. از سطر چهار این جدول مشاهده می شود که به تعداد ۱۴۸ نمونه به دسته سوم به طور درست تخصیص یافته اند. اما ۵ نمونه و ۶ نمونه به ترتیب به دسته های نادرست ۲ و ۵ اختصاص یافته اند. از سطر پنجم مشاهده می شود که به تعداد ۱۷۸۷ نمونه به دسته چهارم تخصیص یافته اند که درست بوده اما به تعداد ۱ و ۳ نمونه به دسته های ۱ و ۵ تخصیص یافته اند که نادرست است. از سطر آخر این جدول مشاهده می شود که به تعداد ۱۰۶۹۳ نمونه به دسته ۵ اختصاص یافته اند که درست است. هیچ نمونه ای به دسته های نادرست در این سطر اختصاص نیافته است. در جدول ۲ نتیجه ی مربوط به ماتریس درهم ریختگی چند دسته ای به کمک روش پیشنهادی نشان داده شده است.

جدول ۲: نتایج ماتریس درهم ریختگی مربوط به معیارهای TP, FP, FN, TN داده آموزش روش پیشنهادی

Table 2: Confusion Matrix Results for the TP, FP, FN, TN Metrics of the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۷۳۳۹	۱	۷	۱۲۶۵۳
دسته واقعی ۲	۱۰	۵	۱	۱۹۹۸۴
دسته واقعی ۳	۱۴۸	۰	۱۱	۱۹۸۴۱
دسته واقعی ۴	۱۷۸۷	۶	۴	۱۸۲۰۳
دسته واقعی ۵	۱۰۶۹۳	۱۱	۰	۹۲۹۶

در جدول ۲ مشاهده می شود که معیار مثبت درست برای دسته های ۱ الی ۵ به ترتیب برابر با ۷۳۳۹، ۱۰، ۱۴، ۱۷۸۷، ۱۰۶۹۳ شده اند. تعداد مثبت های نادرست برای دسته های ۱ الی ۵ به ترتیب برابر با ۱، ۵، ۰، ۶ و ۱۱ شده اند.

در جدول ۳ نتایج مقایسه ای مربوط به معیارهای دقت، خطا، حساسیت، خصوصیت، صحت، امتیاز F1، ضریب همبستگی میتوز آورده شده اند.

جدول ۳: نتایج مقایسه میان روش پیشنهادی و طبقه بندی کننده درخت تصمیم بدون گام خوشه بندی

Table 3: Comparison Results Between the Proposed Method and the Decision Tree Classifier Without the Clustering Step

معیارها	روش پیشنهادی		درخت تصمیم بدون خوشه بندی	
	نتایج	داده آموزش	داده آزمایش	داده آموزش
دقت	0.9989	0.7789	0.9982	0.7523
خطا	0.0011	0.2211	0.0018	0.2477
حساسیت	0.9673	0.5816	0.9671	0.5692
خصوصیت	0.9996	0.9248	0.9995	0.9186
صحت	0.9324	0.8616	0.9121	0.7676
نرخ مثبت نادرست	3.6812e-04	0.0752	5.1588e-04	0.0814
امتیاز F1	0.9459	0.6082	0.9326	0.5773
ضریب همبستگی متیوز	0.9476	0.5942	0.9335	0.5442

در جدول ۳ مشاهده می شود که روش پیشنهادی برای معیار دقت در داده آموزشی مقدار ۰/۹۹۸۹۸ بدست آورده است در حالیکه درخت تصمیم بدون خوشه بندی برابر با ۰/۹۹۸۲ بدست آورده است. معیار دقت بدست آمده برای داده آزمایشی به کمک روش پیشنهادی برابر با ۰/۷۷۸۹ و روش درخت تصمیم بدون خوشه بندی برابر با ۰/۷۵۲۳ شده است. معیار حساسیت برای داده آموزشی بدست آمده از روش پیشنهادی و درخت تصمیم بدون خوشه بندی به ترتیب برابر با ۰/۹۶۷۳ و ۰/۹۶۷۱ شده است. معیار حساسیت برای داده آزمایشی بدست آمده از روش پیشنهادی برابر با ۰/۵۸۱۶ و روش درخت تصمیم بدون خوشه بندی برابر با ۰/۵۶۹۲ شده است

نتیجه ی مربوط به معیار صحت برای داده آموزشی و آزمایشی از روش پیشنهادی به ترتیب برابر با ۰/۹۳۲۴ و ۰/۸۶۱۶ شده است. درحالی که این نتیجه برای روش درخت تصمیم بدون خوشه بندی به ترتیب برابر با ۰/۹۱۲۱ و ۰/۷۶۷۶ شده است. در مجموع از جدول ۳ نتیجه می شود که روش پیشنهادی برای تمام معیارها موفق تر عمل نموده است. در جدول ۴ نتایج مربوط به پیش بینی دسته های ۱ الی ۵ برای داده آزمایشی به کمک روش پیشنهادی نشان داده شده است.

جدول ۴: نتایج داده تست روش پیشنهادی

Table 4: Test Data Results of the Proposed Method

دسته	پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵
دسته واقعی ۱	۵۴۷۵	۰	۰	۶۲	۱۰۸۴
دسته واقعی ۲	۰	۲۴	۰	۰	۳۵
دسته واقعی ۳	۲۰۱	۵	۸۳	۱۶۴	۲۰۹۴
دسته واقعی ۴	۱۴۳	۰	۱	۱۳۹۲	۶۳۳
دسته واقعی ۵	۰	۰	۰	۰	۸۶۰۴

در جدول ۴ مشاهده می شود که در سطر دوم پیش بینی برای دسته اول برابر با ۵۴۷۵ نمونه شده است. اما به تعداد ۶۲ و ۱۰۸۴ نمونه به ترتیب برای دسته ۴ و ۵ پیش بینی شده اند که نادرست است. در سطر سوم جدول مشاهده می شود که به تعداد ۲۴ نمونه برای دسته دوم پیش بینی شده اند. اما به تعداد ۳۵ نمونه به اشتباه به دسته ۵ اختصاص یافته اند. از سطر چهارم مشاهده می شود که به تعداد ۲۰۱ نمونه به دسته اول، ۵ نمونه به دسته دوم، ۸۳ نمونه به دسته سوم، ۱۶۴ نمونه به دسته چهارم تخصیص یافته اند که نادرست است. در سطر آخر به تعداد ۸۶۰۴ نمونه به دسته پنجم اختصاص یافته است که کاملاً درست تشخیص داده شده اند.

در جدول ۵ نتایج ماتریس در هم ریختگی چند دسته ای داده آزمون بدست آمده از روش پیشنهادی نشان داده اند.

جدول ۵: نتایج ماتریس در هم ریختگی چند دسته‌ای داده آزمون روش پیشنهادی

Table 5: Multi-class Confusion Matrix Results for the Test Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۴۷۵	۳۴۴	۱۱۴۶	۱۳۰۳۵
دسته واقعی ۲	۲۴	۵	۳۵	۱۹۹۳۶
دسته واقعی ۳	۸۳	۱	۲۴۶۴	۱۷۴۵۲
دسته واقعی ۴	۱۳۹۲	۲۲۶	۷۷۷	۱۷۶۰۵
دسته واقعی ۵	۸۶۰۵	۳۸۴۶	۰	۷۵۵۰

از جدول ۵ مشاهده می‌شود که برای دسته واقعی ۱ به تعداد ۵۴۷۵ مثبت درست، ۳۴۴ مثبت نادرست، ۱۱۴۶ منفی نادرست، ۱۳۰۳۵ منفی درست تشخیص داده شده‌اند.

در ادامه نتایج مربوط به روش درخت تصمیم بدون خوشه‌بندی در جداول ۶ الی ۸ آورده شده است.

جدول ۶: نتیجه داده آموزشی با روش درخت تصمیم بدون خوشه بندی

Table 6: Training Data Results Using the Decision Tree Method Without Clustering

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۷۳۳۹	۰	۰	۵	۲
دسته واقعی ۲	۰	۱۰	۰	۱	۰
دسته واقعی ۳	۰	۵	۱۴۸	۰	۶
دسته واقعی ۴	۱	۰	۰	۱۷۸۷	۳
دسته واقعی ۵	۲	۲	۳	۶	۱۰۶۸۰

در جدول ۶ مشاهده می‌شود که به تعداد ۷۳۳۹ نمونه مربوط به دسته ۱ پیش‌بینی شده است که درست است. اما برای دسته‌های دوم و سوم به تعداد صفر نمونه پیش‌بینی شده است. اما به تعداد ۵ و ۲ نمونه برای دسته‌های ۴ و ۵ پیش‌بینی شده‌اند. در جدول ۷ نتایج معیارهای مربوط به ماتریس درهم‌ریختگی نشان داده شده است.

جدول ۷: نتایج معیارهای TP, FP, FN, TN داده آموزشی روش درخت تصمیم بدون خوشه بندی

Table 7: TP, FP, FN, TN Metrics Results for the Training Data Using the Decision Tree Method Without Clustering

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۷۳۳۹	۳	۷	۱۲۶۵۱
دسته واقعی ۲	۱۰	۷	۱	۱۹۹۸۲
دسته واقعی ۳	۱۴۸	۳	۱۱	۱۹۸۳۸
دسته واقعی ۴	۱۷۸۷	۱۲	۴	۱۸۱۹۷
دسته واقعی ۵	۱۰۶۸۰	۱۱	۱۳	۹۲۹۶

در جدول ۷ مشاهده می‌شود که معیار مثبت درست بدست آمده از دسته اول برابر با ۷۳۳۹، مثبت نادرست برابر با ۳، منفی نادرست برابر با ۷ و منفی درست برابر با ۱۲۶۵۱ شده است. معیار مثبت درست، مثبت نادرست، منفی نادرست و منفی درست برای دسته دوم به ترتیب برابر با ۱۰، ۷، ۱ و ۱۹۹۸۲ شده است. این معیارها برای دسته سوم به ترتیب برابر با ۱۴۸، ۳، ۱۱، ۱۹۸۳۸ شده است. این معیارها برای دسته چهارم برابر با ۱۷۸۷، ۱۲، ۴ و ۱۸۱۹۷ شده است. این معیارها برای دسته پنجم برابر با ۱۰۶۸۰، ۱۱، ۱۳ و ۹۲۹۶ شده است. در جدول ۸ نتایج داده آزمایشی روش درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۸: نتایج داده آزمایشی روش درخت تصمیم بدون خوشه بندی

Table 8: Test Data Results of the Decision Tree Method

دسته	پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵
دسته واقعی ۱	5475	0	0	62	1084
دسته واقعی ۲	0	24	0	0	35
دسته واقعی ۳	201	5	83	164	2094
دسته واقعی ۴	143	0	1	1392	8072
دسته واقعی ۵	55	7	9	461	8072

در جدول ۸ مشاهده می شود که از سطر دوم به تعداد ۵۴۷۵ نمونه به درستی برای دسته اول اختصاص یافته اند. نمونه ای برای دسته ۲ و ۳ پیش بینی نشده است. برای دسته چهارم ۶۲ نمونه و دسته پنجم ۱۰۸۴ نمونه پیش بینی شده است. از سطر سوم مشاهده می شود که برای دسته دوم، به تعداد صفر نمونه به دسته اول، ۲۴ نمونه به درستی به دسته دوم، صفر نمونه به دسته های سوم و چهارم و ۳۵ نمونه به دسته پنجم تخصیص یافته است. از سطر چهارم مشاهده می شود که به تعداد ۲۰۱ نمونه به دسته اول، ۵ نمونه به دسته دوم، ۸۳ نمونه به درستی به سوم، ۱۶۴ نمونه به دسته چهارم، ۲۰۹۴ نمونه به دسته پنجم اختصاص یافته است. از سطر پنجم مشاهده می شود که به تعداد ۱۴۳ نمونه به دسته اول، صفر نمونه به دسته دوم، ۱ نمونه به دسته سوم و ۱۳۹۲ نمونه به دسته چهارم و ۸۰۷۲ نمونه به دسته پنجم پیش بینی شده است. از سطر ششم مشاهده می شود که ۵۵ نمونه به دسته اول، ۷ نمونه به دسته دوم، ۹ نمونه به دسته سوم، ۴۶۱ نمونه به دسته چهارم و ۸۰۷۲ نمونه برای دسته پنجم پیش بینی شده است.

در جدول ۹ نتایج معیارهای TP,FP,FN,TN به دست آمده از روش درخت تصمیم بدون خوشه بندی آورده شده است.

جدول ۹: نتایج معیارهای TP,FP,FN,TN داده آزمایشی روش درخت تصمیم بدون خوشه بندی

Table 9: TP, FP, FN, TN Metrics Results for the Test Data Using the Decision Tree Method Without Clustering

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۴۷۵	۳۹۹	۱۱۴۶	۱۲۹۸۰
دسته واقعی ۲	۲۴	۱۲	۳۵	۱۹۹۲۹
دسته واقعی ۳	۸۳	۱۰	۲۴۶۴	۱۷۴۴۳
دسته واقعی ۴	۱۳۹۲	۶۸۷	۷۷۷	۱۷۱۴۴
دسته واقعی ۵	۸۰۷۲	۳۸۴۶	۵۳۲	۷۵۵۰

در جدول ۹ مشاهده می شود که از نظر معیار مثبت درست برای تمام دسته ها اعداد بالایی به دست آمده است. اما مثبت نادرست و منفی نادرست برای دسته های ۱، ۳، ۴ و به خصوص پنجم زیاد است. در حالی که روش پیشنهادی مقادیر کمتری بدست آورده است.

ب- حالت دوم نتایج مربوط به داده ۱۰۰ هزار نمونه ای

در این حالت تعداد ۱۰۰ هزار نمونه برای داده آموزشی در نظر گرفته شده است. برای داده تست تعداد ۲۰ هزار نمونه در نظر گرفته شده است. ابتدا الگوریتم جستجوی فرکتال تصادفی بهبود یافته ترکیب شده با الگوریتم نهنگ را اجرا می کنیم. هدف آن خوشه بندی داده ها به دو خوشه عادی و حمله است. پس از اجرا نتایج به دست آمده مربوط به اندیس ماتریس خوشه های به دست آمده را ذخیره می کنیم. از نتایج به دست آمده مشاهده شد که به تعداد ۵۳۴۴۵ نمونه به دسته واقعی ۵ اختصاص یافته که مربوط به داده های عادی است و تماماً به درستی شناخته شده اند. اما مابقی نمونه ها به منظور طبقه بندی درست نیاز به یک دسته بندی کننده است که در اینجا از درخت تصمیم استفاده شده است. نتایج بدست آمده از این خوشه بندی به مرحله ی بعدی طبقه بندی داده می شوند.

۱۵ بهبود کارایی سیستم تشخیص نفوذ با استفاده از درخت تصمیم... / شهرزاد رحیمی - مسعود نیازی ترشیز - سیدعابد حسینی

در جدول ۱۰ نتایج مربوط به داده آموزشی با روش پیشنهادی نشان داده شده است. در جدول ۱۱ نتایج مربوط به معیارهای TP,FP,FN,TN آورده شده است.

جدول ۱۰: نتیجه مربوط به پیش بینی دسته های داده آموزشی با روش پیشنهادی

Table 10: Results of Training Data Category Prediction Using the Proposed Method

پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵	دسته
۳۶۴۴۶	۰	۰	۵	۹	دسته واقعی ۱
۰	۳۳	۲	۱	۷	دسته واقعی ۲
۰	۱	۷۷۰	۲	۱۵	دسته واقعی ۳
۰	۰	۱	۹۲۴۱	۲۲	دسته واقعی ۴
۰	۰	۰	۰	۵۳۴۴۵	دسته واقعی ۵

در جدول ۱۰ مشاهده می شود که برای دسته واقعی ۱ به تعداد ۳۶۴۴۶ پیش بینی شده است. به تعداد صفر نمونه به دسته های دوم و سوم اختصاص یافته است. به تعداد ۵ و ۹ نمونه برای دسته های چهارم و پنجم پیش بینی شده است. برای دسته واقعی دوم، به درستی ۳۳ نمونه پیش بینی شده است. از سطر چهارم جدول ۱۰ مشاهده می شود که به تعداد صفر نمونه مربوط به دسته اول پیش بینی شده است. برای دسته چهارم مشاهده می شود که به تعداد ۹۲۴۱ نمونه بدرستی تشخیص داده شده اند. اما به تعداد ۱ نمونه به دسته سوم و ۲۲ نمونه به دسته پنجم به اشتباه تخصیص یافته است. از سطر آخر مشاهده می شود که به درستی ۵۳۴۴۵ نمونه به درستی به دسته واقعی ۵ اختصاص یافته اند.

در جدول ۱۱ معیارهای TP,FP,FN,TN مربوط به داده آموزشی بدست آمده از روش پیشنهادی نشان داده شده است.

جدول ۱۱: نتایج معیارهای TP,FP,FN,TN داده آموزش روش پیشنهادی

Table 11: TP, FP, FN, TN Metrics Results for the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۳۶۴۴۶	۰	۱۴	۶۳۵۴۰
دسته واقعی ۲	۳۳	۱	۱۰	۹۹۹۵۶
دسته واقعی ۳	۷۷۰	۳	۱۸	۹۹۲۰۹
دسته واقعی ۴	۹۲۴۱	۸	۲۳	۹۰۷۲۸
دسته واقعی ۵	۵۳۴۴۵	۵۳	۰	۴۶۵۰۲

در جدول ۱۲ نتایج پیش بینی دسته بندی مربوط به داده آزمایشی با روش پیشنهادی که تعداد آن ۲۰ هزار نمونه است، نشان داده شده است.

جدول ۱۲: نتیجه داده آزمایشی با روش پیشنهادی

Table 12: Test Data Results Using the Proposed Method

پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵	دسته
۵۳۶۶	۰	۰	۶۳	۱۱۹۲	دسته واقعی ۱
۰	۲۱	۷	۰	۳۱	دسته واقعی ۲
۱	۵	۱۳۵	۲۲۲	۲۱۸۴	دسته واقعی ۳
۱۶۹	۰	۰	۱۳۷۰	۶۳۰	دسته واقعی ۴
۰	۰	۰	۰	۸۶۰۴	دسته واقعی ۵

در جدول ۱۳ نتایج مقایسه ای میان روش پیشنهادی و درخت تصمیم بدون خوشه بندی نشان داده شده است.

جدول ۱۳: مقایسه روش پیشنهادی و درخت تصمیم بدون خوشه بندی

Table 13: Comparison Between the Proposed Method and the Decision Tree Without Clustering

نتایج	روش پیشنهادی				درخت تصمیم بدون خوشه بندی
	داده آموزش	داده آزمایش	داده آموزش	داده آزمایش	
دقت	0.9993	0.7748	0.9991	0.7615	
خطا	6.5000e-04	0.2252	9.0000e-04	0.2384	
حساسیت	0.9483	0.5702	0.9483	0.5640	
خصوصیت	0.9997	0.9233	0.9997	0.9201	
صحت	0.9930	0.8472	0.9757	0.8118	
نرخ مثبت نادرست	2.5337e-04	0.0767	3.1541e-04	0.0799	
امتیاز F1	0.9683	0.6008	0.9610	0.5875	
ضریب همبستگی متیوز	0.9693	0.5882	0.9612	0.5658	

در جدول ۱۳ مشاهده می شود که مقدار معیار دقت به دست آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۹۹۳ و داده آموزشی برابر با ۰/۹۹۹۱ است. معیار دقت برای داده ای آزمایشی به دست آمده از روش پیشنهادی برابر با ۰/۷۷۴۸ و ۰/۷۶۱۵ است. معیار خطای به دست آمده از روش پیشنهادی برای داده آموزش برابر با 6.5000e-04 ، داده آزمایش برابر با ۰/۲۲۵۲ ، این معیار برای روش درخت تصمیم بدون خوشه بندی برابر با 9.0000e-04 داده آزمایش برابر با ۰/۲۳۸۴ است. معیار حساسیت به دست آمده از روش پیشنهادی برای داده آموزش برابر با ۰/۹۴۸۳ ، داده آزمایش برابر با ۰/۵۷۰۲ ، این معیار حساسیت به دست آمده به کمک درخت تصمیم بدون خوشه بندی برای داده آموزش برابر با ۰/۹۴۸۳ و ۰/۵۶۴ شده است. معیار خصوصیت به دست آمده به کمک روش پیشنهادی برای داده آموزش برابر با ۰/۹۹۹۷ ، داده آزمایش برابر با ۰/۹۲۳۳ شده است. در این معیار به کمک روش دیگر برابر برای داده آموزشی برابر با ۰/۹۹۹۷ شده ولی برای داده آزمایشی برابر با ۰/۹۲۰۱ شده است. معیار صحت به دست آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۹۳۰ و برای داده آزمایشی برابر با ۰/۸۴۷۲ شده است. این معیار به دست آمده از روش دیگر برای داده آموزشی برابر با ۰/۹۷۵۷ ، ۰/۸۱۱۸ است. نرخ مثبت نادرست به دست آمده از روش پیشنهادی برای داده آموزشی برابر با 2.5337e-04 و برای داده آزمایشی برابر با ۰/۰۷۶۷ است. این معیار برای داده آموزشی برابر با 3.1541e-04 و برای داده آزمایشی برابر با ۰/۰۷۹۹ شده است. معیار امتیاز F1 به دست آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۶۸۳ و برای داده آزمایشی برابر با ۰/۶۰۰۸ شده است. این معیار برای داده آموزشی برابر با ۰/۹۶۱ و داده آزمایشی برابر با ۰/۵۸۷۵ شده است. در ادامه نتایج مربوط به درخت تصمیم بدون خوشه بندی ارائه شده است. در جدول ۱۴ نتایج مربوط به پیش بینی دسته ها از درخت تصمیم بدون خوشه بندی نشان داده شده است.

جدول ۱۴: نتایج مربوط به داده آموزش روش درخت تصمیم بدون خوشه بندی

Table 14: Training Data Results Using the Decision Tree Method Without Clustering

دسته	پیش بینی کلاس ۱	پیش بینی کلاس ۲	پیش بینی کلاس ۳	پیش بینی کلاس ۴	پیش بینی کلاس ۵
دسته واقعی ۱	۳۶۴۴۶	۰	۰	۵	۹
دسته واقعی ۲	۰	۳۳	۲	۱	۷
دسته واقعی ۳	۰	۱	۷۷۰	۲	۱۵
دسته واقعی ۴	۰	۰	۱	۹۲۴۱	۲۲
دسته واقعی ۵	۹	۳	۵	۸	۵۳۴۲۰

در جدول ۱۴ مشاهده می‌شود که برای دسته شماره ۱، به تعداد ۳۶۴۴۶ نمونه درست پیش‌بینی شده است. در حالی که به تعداد ۵ نمونه به دسته چهار و نمونه به دسته پنج به‌طور نادرست اختصاص یافته است. در جدول ۱۵ نتایج معیارهای TP,FP,FN,TN مربوط به داده آموزش با روش پیشنهادی نشان داده شده است.

جدول ۱۵: نتایج معیارهای TP,FP,FN,TN مربوط به داده آموزش با روش پیشنهادی

Table 15: TP, FP, FN, TN Metrics Results for the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۳۶۴۴۶	۹	۱۴	۶۳۵۳۱
دسته واقعی ۲	۳۳	۴	۱۰	۹۹۹۵۳
دسته واقعی ۳	۷۷۰	۸	۱۸	۹۹۲۰۴
دسته واقعی ۴	۹۲۴۱	۱۶	۲۳	۹۰۷۲۰
دسته واقعی ۵	۵۳۴۲۰	۵۳	۲۵	۴۶۵۰۲

در جدول ۱۵ مشاهده می‌شود که معیار TP به‌دست آمده از دسته‌های ۱ الی ۵ به ترتیب برابر با ۳۶۴۴۶، ۳۳، ۷۷۰، ۹۲۴۱، ۵۳۴۲۰ می‌باشند. از این جدول مشاهده می‌شود که مقادیر مثبت نادرست و منفی نادرست در مقایسه با روش پیشنهادی بیشتر شده‌اند. معیارهای مثبت درست، منفی درست به‌دست آمده از درخت تصمیم بدون خوشه‌بندی در مجموع کمتر از روش پیشنهادی شده است. این نشان می‌دهد که روش پیشنهادی از نظر معیارهای TP,FP,FN,TN به‌طور کلی موفق‌تر عمل نموده است. در جدول ۱۶ نتایج مربوط به داده آزمون به کمک روش درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۱۶: نتایج دسته‌بندی مربوط به داده آزمایش روش درخت تصمیم بدون خوشه‌بندی

Table 16: Classification Results for the Test Data Using the Decision Tree Method Without Clustering

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۵۳۶۶	۰	۰	۶۳	۱۱۹۲
دسته واقعی ۲	۰	۲۱	۷	۰	۳۱
دسته واقعی ۳	۱	۵	۱۳۵	۲۲۲	۲۱۸۴
دسته واقعی ۴	۱۶۹	۰	۰	۱۳۷۰	۶۳۰
دسته واقعی ۵	۶۳	۲	۲	۱۹۸	۸۳۳۹

در جدول ۱۶ مشاهده می‌شود که پیش‌بینی‌ها برای داده آزمایشی در حد مناسبی انجام شده است اما روش پیشنهادی در مقایسه با درخت تصمیم بدون خوشه‌بندی موفق‌تر عمل نموده است. در جدول ۱۷ نتایج مربوط به TP,FP,FN,TN داده آزمایشی به کمک روش درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۱۷: نتایج معیارهای TP,FP,FN,TN داده آزمایش روش پیشنهادی

Table 17: TP, FP, FN, TN Metrics Results for the Test Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۳۶۶	۱۷۰	۱۲۵۵	۱۳۲۰۹
دسته واقعی ۲	۲۱	۵	۳۸	۱۹۹۳۶
دسته واقعی ۳	۱۳۵	۷	۲۴۱۲	۱۷۴۴۶
دسته واقعی ۴	۱۳۷۰	۲۸۵	۷۹۹	۱۷۵۴۶
دسته واقعی ۵	۸۶۰۴	۴۰۳۷	۰	۷۳۵۹

در جدول ۱۷ مشاهده می‌شود که برای دسته واقعی ۱، معیار مثبت درست ۵۳۶۶، مثبت نادرست ۱۷۰، منفی نادرست ۱۲۵۵، منفی درست برابر با ۱۳۲۰۹ است. از این جدول مشاهده می‌شود که معیار مثبت درست در مقایسه با مثبت نادرست خیلی بیشتر است.

باتوجه به جداول ۱۰ الی ۱۷ مشاهده می شود که برای مجموعه داده ۱۰۰ هزارتایی نیز در مجموع نتایج به دست آمده از روش پیشنهادی خوشه بندی با الگوریتم جستجوی فرکتال تصادفی بهبود یافته ترکیب شده با الگوریتم نهنگ بهتر از حالت بدون خوشه بندی است.

۵- نتیجه گیری

در این مقاله یک الگوریتم بهینه سازی جستجوی فرکتال تصادفی بهبود یافته با نگاشت های آشوبی ترکیب شده با الگوریتم بهینه سازی نهنگ معرفی شد. از این الگوریتم برای خوشه بندی داده ی NSL-KDD به دو خوشه حمله و عادی استفاده شده است. در این مرحله الگوریتم پیشنهادی به داده اعمال شد و نتایج آن نشان داد که به طور کامل داده های عادی از حملات جدا شده اند و هر کدام در خوشه خود جای گرفته اند. دو مثال برای نشان دادن این مورد انجام شد که مثال اول با ۲۰ هزار نمونه برای داده آموزش، ۲۰ هزار نمونه برای داده آزمایش و مثال دوم شامل ۱۰۰ هزار نمونه برای داده آموزش و ۲۰ هزار نمونه برای داده آزمایش انجام شد. نتایج نشان داد که در هر دو مثال مذکور خوشه های مربوط به عادی و نفوذ یافته به درستی و به طور کامل از هم تفکیک شده اند. این خوشه ها ذخیره شده و به مرحله ی بعدی طبقه بندی اعمال شده اند. سپس از روش درخت تصمیم برای طبقه بندی داده های حملات استفاده شده است. در این مرحله دسته مربوط به داده های نرمال از پیش از آن درست تشخیص داده شده و تنها فقط دسته بندی برای حملات باید انجام شود. برای ارزیابی روش پیشنهادی نتایج آن با حالت درخت تصمیم بدون خوشه بندی مقایسه شده است. هر ۵ دسته باید ارزیابی شوند تا مقایسه عادلانه با روش مشابه درخت تصمیم بدون خوشه بندی انجام شود. نتایج نشان داده اند که از نظر معیارهای TP, FP, FN, TN روش پیشنهادی برای تعداد نمونه های ۲۰ هزارتایی داده آموزش، ۲۰ هزارتایی داده آزمایش، ۱۰۰ هزارتایی داده آموزش، ۲۰ هزارتایی داده آموزش موفق تر از روش درخت تصمیم بدون خوشه بندی عمل نموده است. نتایج از نظر معیارهای دقت، خاصیت، حساسیت در هر دو مثال برای داده های آموزشی و آزمایشی انجام شد. نتایج برای مثال اول نشان داد که معیار دقت به دست آمده از روش پیشنهادی برابر با ۰/۹۹۸۹ در حالی که روش دیگر برابر با ۰/۹۹۸۲ به دست آورده است. معیار حساسیت به دست آمده از روش پیشنهادی برابر با ۰/۹۶۷۳ شده است در حالی که این معیار از روش دیگر برابر با ۰/۹۶۷۱ شده است. از نظر خصوصیت روش پیشنهادی برابر با ۰/۹۹۹۶ و روش درخت تصمیم بدون خوشه بندی برابر با ۰/۹۹۹۵ شده است. نتایج به دست آمده برای داده آزمایشی برای معیار دقت از روش پیشنهادی برابر با ۰/۷۷۸۹، در حالی که برای روش دیگر برابر با ۰/۷۵۲۳ شده است. از نظر معیار حساسیت با روش پیشنهادی برابر با ۰/۵۸۱۶ و روش دیگر برابر با ۰/۵۶۹۲ شده است. از نظر معیار خصوصیت روش پیشنهادی مقدار ۰/۹۲۴۸ و روش دیگر برابر با ۰/۹۱۸۶ به دست آورده است.

نتایج به دست آمده برای مثال دوم حالت ۱۰۰ هزار نمونه داده آموزشی و ۲۰ هزار نمونه داده آزمایشی به این صورت شده اند. برای داده آموزشی از نظر معیار دقت روش پیشنهادی توانسته است که مقدار ۰/۹۹۹۳ و روش درخت تصمیم بدون خوشه بندی برابر با ۰/۹۹۹۱ به دست آورد. از نظر معیار حساسیت روش پیشنهادی توانسته است که مقدار ۰/۹۴۸۳ و روش دیگر برابر با ۰/۹۴۸۳ به دست آورد. از نظر معیار خاصیت روش پیشنهادی مقدار ۰/۹۹۹۷ و روش دیگر ۰/۹۹۹۷ به دست آورده است. اما برای داده های آزمایشی روش پیشنهادی برای معیار دقت برابر با ۰/۷۷۴۸، در حالی که روش درخت تصمیم بدون خوشه بندی برابر با ۰/۷۶۱۵ به دست آورده است. از نظر معیار حساسیت ۰/۵۷۰۲، ۰/۵۶۴۰ شده است. برای معیار خاصیت روش پیشنهادی توانسته که مقدار ۰/۹۲۳۳ و روش درخت تصمیم بدون خوشه بندی مقدار ۰/۹۲۰۱ به دست آورد. در مجموع از نتایج گرفته شده نتیجه می شود که روش پیشنهادی خوشه بندی با الگوریتم جستجوی فرکتال تصادفی بهبود یافته با نگاشت های آشوب ترکیب شده با الگوریتم نهنگ موفق تر از حالت بدون خوشه بندی عمل نموده است.

به منظور پیشنهاد برای تحقیقات آینده، می توان از این الگوریتم از نوع باینری آن برای انتخاب ویژگی برای افزایش دقت طبقه بندی استفاده نمود. سپس نتایج آن را برای داده NSL-KDD و داده های مشابه در تحقیقات استفاده نمود. یک پیشنهاد دیگر استفاده از روش های طبقه کننده پیچیده تر مانند یادگیری عمیق بهبود با الگوریتم پیشنهادی جستجوی فرکتال تصادفی بهبود یافته ترکیب شده با الگوریتم نهنگ است که می تواند با روش پیشنهادی برای افزایش دقت طبقه بندی استفاده شود.

مراجع

- [1] Diro, A. A., & Chilamkurti, N., Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, 82, 761-768, 2018
- [2] Roseline Oluwaseun Ogundokun, Joseph Bamidele Awotunde, Peter Sadiku, Emmanuel Abidemi Adeniyi, Moses Abiodun, Oladipo Idowu Dauda, An Enhanced Intrusion Detection System using Particle Swarm Optimization Feature Extraction Technique, *Procedia Computer Science*, Volume 193, 2021, Pages 504-512, ISSN 1877-0509.
- [3] S. Hofmeyr, S. Forrest, and A. Sornayaji, "Lightweight intrusion detection for networked operating systems," *Journal of Computer Security*, vol. 5, no. 2, 1997.
- [4] A. Puri and N. Sharma, "A novel technique for intrusion detection system for network security using hybrid svm-cart," *IJEDR*, vol. 5, no. 2, pp. 155–161, 2017.
- [5] P. Kabiri and A. A. Ghorbani, "Research on intrusion detection and response: A survey." *IJ Network Security*, vol. 1, no. 2, pp. 84–102, 2005.
- [6] M. Shojafar, R. Taheri, Z. Pooranian, R. Javidan, A. Miri and Y. Jararweh, "Automatic Clustering of Attacks in Intrusion Detection Systems," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), 2019, pp. 1-8, doi: 10.1109/AICCSA47632.2019.9035238.
- [7] A. Karami, "An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities," *Expert Systems with Applications*, vol. 108, pp. 36–60, 2018.
- [8] M. Tabash, M. Abd Allah, and B. Tawfik, "Intrusion detection model using naive bayes and deep learning technique," *3e International Arab Journal of Information Technology*, vol. 17, no. 2, 2020.
- [9] A. Ghazali, W. Nuaimy, A. Al-Atabi, and I. Jamaludin, "Comparison of classification models for Nsl-Kdd dataset for network anomaly detection," *Academic Journal of Science*, vol. 4, no. 1, pp. 199–206, 2015.
- [10] J. Kevric, S. Jukic, and A. Subasi, "An effective combining classifier approach using tree algorithms for network intrusion detection," *Neural Computing & Applications*, vol. 28, no. S1, pp. 1051–1058, 2017.
- [11] A. Hadi, "Performance analysis of big data intrusion detection system over random forest algorithm," *International Journal of Applied Engineering Research*, vol. 13, no. 2, pp. 1520–1527, 2018.
- [12] W. Elmasry, A. Akbulut, and A. H. Zaim, "Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic," *Computer Networks*, vol. 168, Article ID 107042, 2020
- [13] Zhu, Y., Gaba, G. S., Almansour, F. M., Alroobaea, R., & Masud, M. (2021). Application of data mining technology in detecting network intrusion and security maintenance. *Journal of Intelligent Systems*, 30(1), 664-676.
- [14] Anitha, P., & Kaarthick, B. (2021). Oppositional based Laplacian grey wolf optimization algorithm with SVM for data mining in intrusion detection system. *Journal of Ambient Intelligence and Humanized Computing*, 12(3), 3589-3600.
- [15] Koryshev, N., Hodashinsky, I., & Shelupanov, A. (2021). Building a fuzzy classifier based on whale

- optimization algorithm to detect network intrusions. *Symmetry*, 13(7), 1211.
- [16] Zhang, J., Sun, J., & He, H. (2021). Clustering Detection Method of Network Intrusion Feature Based on Support Vector Machine and LCA Block Algorithm. *Wireless Personal Communications*, 1-15.
 - [17] Maheswari, M., & Karthika, R. A. (2021). A novel QoS based secure unequal clustering protocol with intrusion detection system in wireless sensor networks. *Wireless Personal Communications*, 118(2), 1535-1557.
 - [18] Xie, B., Dong, X., & Wang, C. (2021). An Improved-Means Clustering Intrusion Detection Algorithm for Wireless Networks Based on Federated Learning. *Wireless Communications and Mobile Computing*, 2021.
 - [19] Keserwani, P. K., Govil, M. C., Pilli, E. S., & Govil, P. (2021). A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO-PSO-RF model. *Journal of Reliable Intelligent Environments*, 7(1), 3-21.
 - [20] Dwivedi, S., Vardhan, M., & Tripathi, S. (2021). Building an efficient intrusion detection system using grasshopper optimization algorithm for anomaly detection. *Cluster Computing*, 24(3), 1881-1900.
 - [21] Haider AL-Husseini, et al. Whale Optimization Algorithm-Enhanced Long Short-Term Memory Classifier with Novel Wrapped Feature Selection for Intrusion Detection, *Journal of Sensor and Actuator Networks*, Nov 2024, 13, 73, <https://doi.org/10.3390/jsan13060073>
 - [22] Rajashekar Kandakatla, et al. Whale-Optimized Probabilistic Selection For Enhanced Intrusion Detection In Cloud Environments, *Journal of Theoretical and Applied Information Technology*, June 2024. Vol. 102. No. 12
 - [23] Salimi, H., Stochastic fractal search: a powerful metaheuristic algorithm. *Knowledge-Based Systems*, 2015. 75: p. 1-18
 - [24] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. 1991;21(3):660–674.
 - [25] Misaghi, M., & Yaghoobi, M. (2019). Improved invasive weed optimization algorithm (IWO) based on chaos theory for optimal design of PID controller. *Journal of Computational Design and Engineering*, 6(3), 284-295.
 - [26] Kumar Ahuja, Dr. Gulshan. (2015). Evaluation Metrics for Intrusion Detection Systems-A Study. *International Journal of Computer Science and Mobile Applications*. 11.