نشریه علمی «علوم و فناوری‌های پدافند نوین»
سال دوازدهم، شماره ۲، تابستان ۱۴۰۰؛ ص ۱۵۸- ۱۴۳

علمی- پژوهشی

توسعه یک سیستم تشخیص نفوذ مبتنی بر خوشه‌بندی فازی و الگوریتم بهینه‌سازی نهنگ

رضا نظری[۱]، سید مصطفی فخراحمد[۲*]

۱- دانشجوی دکتری دانشگاه تهران، ۲- دانشیار دانشگاه شیراز

(دریافت: ۱۳۹۸/۰۴/۳۱، پذیرش: ۱۳۹۹/۱۱/۰۲)

چکیده

امروزه شبکه‌های کامپیوتری در جهان کاربردهای فراوانی پیدا کرده‌اند. به‌دلیل استفاده گسترده از اینترنت، سیستم‌های رایانه‌ای، مستعد سرقت اطلاعات هستند که منجر به ظهور سیستم‌های تشخیص نفوذ (IDS) شده است. امنیت شبکه در پاسخ به افزایش اطلاعات حساس، به یک موضوع اساسی در علوم کامپیوتر تبدیل شده است. در پژوهش حاضر سیستم تشخیص نفوذ غیرنظارتی مبتنی بر خوشه‌بندی فازی (FCM) با بهره‌گیری از الگوریتم بهینه‌سازی نهنگ (WOA) پیشنهاد شده است و با مجموعه داده استاندارد تشخیص نفوذ KDD Cup ۹۹ مورد آزمایش قرار گرفت. در این روش به‌منظور جداسازی فعالیت‌های نفوذی از فعالیت‌های عادی، خوشه‌بندی C- میانگین فازی مورد استفاده قرارگرفته و از الگوریتم بهینه‌سازی نهنگ برای به‌دست آوردن تفکیک بهینه بین این فعالیت‌ها استفاده شده است. جهت کمک به FCM، از WOA استفاده شده است تا از مراکز خوشه‌های اولیه مناسب به‌جای مراکز تصادفی استفاده کند. نتایج تجربی بر روی مجموعه داده KDD Cup ۹۹ حاکی از بهبود نرخ همگرایی، صحت و همچنین نرخ هشدار اشتباه توسط الگوریتم WOA-FCM در قیاس با سایر روش‌های غیر نظارتی می‌باشد. از همین‌رو، یافته‌های پژوهش حاضر می‌تواند در زمینه حل مسائل پیچیده مرتبط با IDS مؤثر واقع شود.

کلیدواژه‌ها: سیستم تشخیص نفوذ (IDS)، خوشه‌بندی C- میانگین فازی (FCM)، الگوریتم بهینه‌سازی نهنگ (WOA)، منطق فازی، خوشه‌بندی فازی، WOA-FCM

Developing an Intrusion Detection System Based on Fuzzy Clustering and Whale Optimization Algorithm

R. Nazari, S. M. Fakhrahmad*
Shiraz University, Shiraz
(Received: 22/07/2019; Accepted: 21/01/2021)

Abstract

Nowadays, computer networks are being widely used in the world. Due to the widespread use of the internet, computer systems are prone to information theft and this has led to the emergence of intrusion detection systems (IDS). Thus, network security has become an essential subject in computer science responding to the increase of sensitive information. The current research has used fuzzy C-means (FCM) and Whale optimization algorithm (WOA) to propose an unsupervised machine learning intrusion identification system and has tested it with the KDD Cup 99 standard intrusion detection dataset. In this method, fuzzy C-means has been applied in order to distinguish intrusive activities from normal activities and Whale optimization algorithm has been used to achieve optimal separations among these activities. In order to help FCM, the WOA has been applied to start with suitable cluster centers rather than randomly initialized centers. Experimental results on KDD Cup 99 dataset showed that the proposed method offers higher detection accuracy and a lower false alarm rate compared to other similar algorithms. Therefore, the findings of the present study would be effective in solving complex problems related to IDS.

Keywords: Intrusion Detection System (IDS), Fuzzy C-Means (FCM), Whale Optimization Algorithm (WOA), Fuzzy Logic, Fuzzy Clustering, WOA-FCM

*Corresponding Author E-mail: fakhrahmad@shirazu.ac.ir

۱- مقدمه

در دنیای امروز اطلاعات به مهم‌ترین و حساس‌ترین دارایی‌های سازمان‌ها و افراد تبدیل شده و به‌همین دلیل اهمیت این دارایی‌ها افزایش قابل توجهی پیدا کرده است. گستردگی و نفوذ اطلاعات در یک سازمان مدیریت امنیت اطلاعات را پیچیده‌تر می‌کند و فقدان سیستمی کامل و مدون برای مدیریت امنیت اطلاعات، کاربران را با مشکلات زیادی مواجه می‌سازد. عموم سازمان‌ها تجهیزاتی مانند فایروال و آنتی‌ویروس دارند اما استفاده از این ابزارها کافی نیست. برای جلوگیری از حملات نسل جدید مثل باج‌افزارها، بدافزارهای پیچیده و حملات پیشرفته نیاز به جمع‌آوری، معمولی‌سازی و تحلیل اطلاعات از منابع مختلف وجود دارد که این کار بدون داشتن ابزار تخصصی امکان‌پذیر نیست. سیستم تشخیص نفوذ[1]، فایروال و هانی‌پات از فناوری‌هایی هایی هستند که می‌توانند از بروز چنین حملاتی به شبکه تا حد زیادی جلوگیری کنند. دستیابی به اطلاعات از مهم‌ترین راهبرد-های شناخت وضعیت نظامی دشمن است که در ابعاد نظامی و امنیتی از آن به‌عنوان یکی از شاخص‌های بسیار مهم در امور دفاعی بهره می‌برند. یکی از منابع مهم کسب اطلاعات، نفوذ به پایگاه‌های داده و شبکه‌های کامپیوتری استفاده شده در سیستم دفاعی است. از این‌رو این شبکه‌ها همواره در معرض حمله‌های سایبری دشمن می‌باشند. به‌همین دلیل حفظ امنیت آن‌ها یک امر حیاتی به‌شمار می‌آید. تشخیص حمله سایبری مسئله‌ای مهم و چالش برانگیز در حیطه فناوری اطلاعات است. در چنین سناریویی، مهاجمین سازوکارهای جدید را با سازوکارهای پلی مورفیک به منظور فرار از سیستم‌های تشخیص نفوذ مورد استفاده قرار می‌دهند. این مسئله منجر به از دست رفتن اطلاعات و افزایش آسیب‌پذیری امنیتی می‌شود. به‌منظور پیشگیری از نفوذ به شبکه‌های کامپیوتری و محافظت از امنیت اطلاعات و داده‌ها، از سامانه‌های تشخیص نفوذ که از مهم‌ترین ابزارهای امنیتی محسوب می‌شوند، استفاده می‌شود، که بر همین اساس تشخیص نفوذ خودکار، یکی از حیطه‌های مهم پژوهشی طی بیش از دو دهه اخیر به حساب می‌آید [۳-۱].

۱-۱- تشخیص نفوذ

به مجموعه عملیات غیرقانونی که تلاش می‌کنند یکپارچگی[2]، محرمانگی[3] و در دسترس بودن[4] یک منبع را به مخاطره بیاندازد نفوذ می‌گویند. نفوذها به دودسته داخلی و خارجی تقسیم می شوند. نفوذهای داخلی به مجموعه نفوذهایی گفته می‌شود که توسط افراد مجاز در شبکه داخلی، از درون خود شبکه انجام می‌شود و نفوذهای خارجی توسط افراد مجاز و یا غیرمجاز از خارج شبکه به درون شبکه داخلی صورت می‌گیرد. فرآیند نظارت بر ترافیک شبکه به‌منظور شناسایی فعالیت‌های مخرب در شبکه‌های کامپیوتری را تشخیص نفوذ می‌نامند [۳- ۱].

با توجه به استفاده وسیع سیستم‌ها و شبکه‌های کامپیوتری، استفاده از سیستم تشخیص نفوذ در جهان برای امور دولتی، نظامی، تجاری و موارد دیگر گسترش فراوانی داشته است. سیستم تشخیص نفوذ، دسترسی کاربر به سیستم کامپیوتری را با اجرای قوانینی خاص، بازبینی و محدود می‌کند. این قوانین بر اساس دانش افراد متخصص و باتجربه‌ای که سناریوهای مختلفی از حملات را ساخته‌اند استخراج می‌شود. سیستم تمامی تخلفات کاربران را شناسایی می‌کند و اقدامات لازم برای متوقف کردن حمله را انجام می‌دهد [۴]. در حقیقت سیستم‌های تشخیص نفوذ جزء اصلی یک زیرساخت امنیتی برای تشخیص تهدیدات، قبل از آسیب گسترده و جدی به سیستم‌ها هستند [۵].

وظیفه عمده ایجاد یک سیستم تشخیص نفوذ از منظر دسته‌بندی، تولید نوعی دسته‌بندی است که قادر است داده‌های موجود را به دسته‌های مشکوک و معمولی طبقه‌بندی کند. به‌عبارت‌دیگر، عملکرد اصلی یک سیستم تشخیص نفوذ عبارت از دسته‌بندی عملکرد درونی سیستم به دودسته عمده عملکرد معمولی و عملکرد مشکوک می‌باشد. سیستم‌های تشخیص نفوذ به‌طور معمول قادر به تعیین نوع حملات یا دسته‌بندی عملکردها به گروه‌های مشخص می‌باشد که به این منظور چندین روش محاسبات نرم، شبکه‌های عصبی- فازی، روش‌های استنتاج فازی و الگوریتم‌های ژنتیک را مورد استفاده قرار می‌دهد [۶].

۱-۲- انواع تشخیص نفوذ

نفوذگرها معمولاً از استراق سمع ترافیک شبکه، شکستن کلمات رمز، عیب‌های نرم‌افزاری و نقاط ضعف طراحی شبکه، کامپیوترهای شبکه و یا سرویس‌ها برای نفوذ به سیستم‌ها و شبکه‌ها استفاده می‌کنند. جهت مقابله با نفوذگران روش‌های متعددی تحت عنوان روش‌های تشخیص نفوذ ایجاد شده است. راه‌های شناسایی نفوذ در سامانه‌های تشخیص نفوذ به دودسته تشخیص سوءاستفاده[5] و تشخیص ناهنجاری[6] تقسیم می‌شوند [۷، ۸].

[1] Intrusion Detection Systems (IDS)
[2] Integrity
[3] Confidentiality
[4] Availability
[5] Misuse Detection
[6] Anomaly Detection

۱-۲-۱- تشخیص سوء استفاده

ایده اصلی در این روش تشخیص، الگوها[1] یا امضاها هستند که مهاجمان را می‌تواند به کمک این الگوها شناسایی کنند. در سیستم‌های تشخیص نفوذ مبتنی بر امضاء یا تشخیص سوء استفاده، الگوهای نفوذ (امضاء) که قبلاً رخ‌داده‌اند را در پایگاه داده خود ذخیره می‌کنند. به‌طوری‌که هر الگو انواع مختلفی از یک نفوذ خاص را در برگرفته و در صورت بروز چنین الگویی در سیستم، وقوع نفوذ اعلام می‌شود. موضوع اصلی در سیستم‌های تشخیص سوءاستفاده این است که چگونه الگوی رفتار مخرب نوشته شود که با فعالیت‌های عادی تشابه پیدا نکند. این دسته از روش‌ها در شناسایی نفوذهای مرسوم و مشهور با نرخ پایینی از خطا عمل می‌کنند اما در حملاتی که تاکنون از سوی سیستم شناخته‌نشده‌اند دچار ضعف می‌باشند و در صورت بروز حملات جدید در سطح شبکه، نمی‌توانند آن‌ها را شناسایی کنند. پس این روش در مقابله با نفوذهای ناشناخته به‌راحتی شکست می‌خورد. مشکل دیگر این روش به‌روزرسانی پایگاه دانش آن است که باید همواره انجام شود و این عملی زمان‌بر و دشوار است [۵، ۹، ۱۰].

۱-۲-۲- تشخیص آنومالی

روش تشخیص آنومالی یا رفتار غیرعادی توسط دنینگ پیشنهادشده است. در این روش درواقع الگوهای رفتار عادی کاربران ملاک عمل قرار داده می‌شود و درنتیجه هرگونه رفتار مغایر با آن به‌عنوان تلاشی جهت نفوذ به سیستم شناسایی می‌گردد. پس مدلی از رفتارهای عادی برخلاف روش اول ایجاد می‌شود و اگر در رفتارهای مشاهده‌شده، رفتاری مغایر با این مدل‌ها رخ دهد به‌عنوان رفتار غیرعادی تشخیص داده می‌شود. در روش تشخیص رفتار غیرعادی به جمع‌آوری ترکیبی از عملکرد کاربران، فرآیندهای موجود در سیستم و ساختمان شبکه در موقعیت معمول پرداخته می‌شود. سپس عملکردهای شبکه و رایانه با معیار استاندارد تعیین‌شده از قبل مورد مقایسه قرار می‌گیرد. در این مقایسه در صورت مشاهده رفتارهایی که منطبق با رفتار عادی نیست تحت عنوان عملکرد غیرعادی شناخته می‌شود. برای نمونه ترافیک HTTP روی یک پورت غیراستاندارد یا ترافیک خیلی زیاد UDP در مقایسه با TCP از رفتارهای غیرعادی می‌باشند. واضح است که تشخیص رفتار غیرعادی، قابلیت تشخیص حملات جدید را دارد و تنها نیاز است که داده‌های عادی در هنگام ساختن پروفایل[2] به آن معرفی شود. با توجه به این‌که نمونه‌های رفتار غیرعادی در داده‌های آموزشی کم هستند مشکل اساسی این روش تعیین مرز بین رفتارهای عادی و غیرعادی است [۵، ۱۰].

تشخیص رفتار غیرعادی به دودسته استاتیک و پویا تقسیم می‌شود. در روش استاتیک فرض می‌شود که رفتارها هیچ‌وقت تغییر نمی‌کنند. در نوع دوم، تشخیص به‌صورت پویا انجام می‌شود، به این صورت که الگوهایی از رفتارهای معمولی کاربران نهایی استخراج می‌شود که به این الگوها پروفایل می‌گویند. این روش، توانایی تشخیص انواع جدیدی از نفوذها را دارد و تنها به داده‌های معمولی برای درست کردن پروفایل نیازمند است [۱۱].

۱-۲-۳- روش‌های ترکیبی

روش‌های ترکیبی درواقع مشتمل بر روش‌های تشخیص سوءاستفاده و تشخیص رفتار غیرعادی است که امروزه با توجه به نقاط قوت و ضعف هر یک از روش‌های مذکور، استفاده از روش ترکیبی ترجیح داده می‌شود [۱۲].

۱-۳- انواع معماری سیستم تشخیص نفوذ

معماری‌های مختلف سامانه تشخیص نفوذ عبارت‌اند از [۱۳]:

۱. سامانه تشخیص نفوذ مبتنی بر میزبان (HIDS)[3]

۲. سامانه تشخیص نفوذ مبتنی بر شبکه (NIDS)[4]

۳. سامانه تشخیص نفوذ توزیع‌شده (DIDS)[5]

۱-۳-۱- سامانه تشخیص نفوذ مبتنی بر میزبان

سیستم مبتنی بر میزبان روی کامپیوترهای میزبان نصب می‌شوند و فعالیت‌ها و فایل‌های سیستم را مورد بررسی قرار می‌دهند. این سامانه می‌تواند حملات و تهدیداتی را روی سیستم‌ها تشخیص دهد که توسط سامانه‌های تشخیص نفوذ مبتنی بر شبکه قابل تشخیص نیستند. در سیستم‌های تشخیص نفوذ مبتنی برمیزبان، فقط از میزبان های حاوی سیستم محافظت به عمل می‌آید و به‌صورت پیش فرض کارت واسط شبکه[6] آن‌ها در حالت با قاعده[7] عمل می‌کند. با توجه به محل استقرار، این سیستم‌ها حاوی همه اطلاعات محلی اضافی و پیاده‌سازی‌های امنیتی می‌باشند که به‌عنوان نمونه می‌توان به اتصال‌های سیستم، فراخوانی‌ها و تغییرات فایل‌های سیستمی اشاره کرد. این موضوع باعث می‌شود در زمان ترکیب با ارتباطات شبکه‌ای، داده‌های مناسبی برای جستجوی رویدادهای احتمالی فراهم شود [۱۳، ۱۴].

[1] Signatures
[2] Profile
[3] Host - based Intrusion Detection System
[4] Network - based Intrusion Detection System
[5] Distributed - based Intrusion Detection System
[6] Network Interface Card
[7] Non promiscuous mode

۱-۳-۲- سامانه تشخیص نفوذ مبتنی بر شبکه

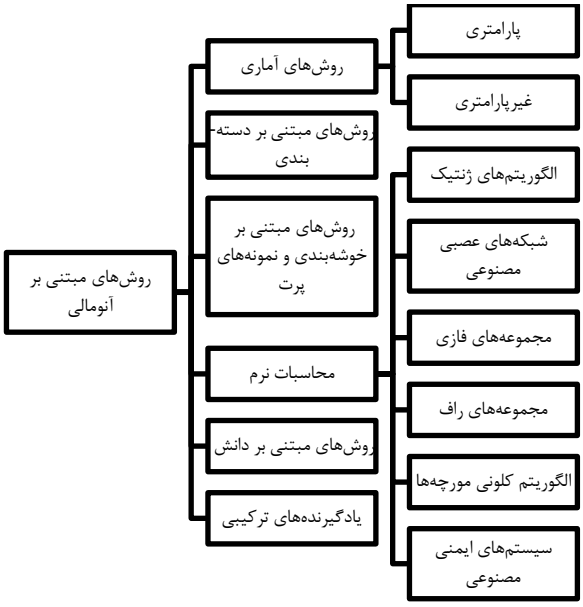
سیستم مبتنی بر شبکه از ترافیک شبکه به‌عنوان منبع اطلاعاتی استفاده می‌کند و داده شبکه را در سرور و درگاه قبل از اینکه به کاربر نهایی برسد، مورد بازرسی قرار می‌دهند. این سیستم‌ها به‌منظور جستجوی تلاش‌هایی که برای حمله صورت می‌گیرد، به بررسی بسته‌ها و پروتکل‌های ارتباطات فعال می‌پردازد. به‌عبارت‌دیگر معیار آن‌ها تنها بسته‌هایی است که در شبکه رد و بدل می‌شود. ازآنجایی‌که تشخیص را به یک سیستم منفرد محدود نمی‌کنند، گستردگی بیشتری دارند و فرایند تشخیص را به‌صورت توزیع‌شده انجام می‌دهند. بااین‌حال این سیستم‌ها در برابر بسته‌های رمز شده و یا شبکه‌هایی با سرعت و ترافیک بالا کارایی خود را از دست می‌دهند [۱۳، ۱۵].

۱-۳-۳- سامانه تشخیص نفوذ توزیع‌شده

سامانه تشخیص نفوذ توزیع‌شده از چند سیستم تشخیص نفوذ مبتنی بر شبکه یا سیستم تشخیص نفوذ مبتنی بر میزبان یا ترکیبی از این دو نوع به همراه یک ایستگاه مدیریت مرکزی تشکیل‌شده است. به این صورت که هر سیستم تشخیص نفوذ موجود در شبکه گزارش‌های خود را برای ایستگاه مدیریت مرکزی ارسال می‌کند. ایستگاه مرکزی مسئولیت امنیتی سیستم را بر عهده دارد، گزارش‌های رسیده را بررسی و پایگاه قوانین تشخیص هر یک از سیستم‌ها را به‌روزرسانی می‌کند. شبکه بین سامانه مدیریت مرکزی با سیستم‌های تشخیص نفوذ مبتنی بر شبکه می‌تواند خصوصی باشد یا از زیرساخت موجود جهت ارسال داده‌ها استفاده شود که در این صورت برای امنیت بیشتر از رمزنگاری یا شبکه‌های خصوصی مجازی (VPN)[1] استفاده خواهد شد [۱۳، ۱۶].

۱-۴- پیشینه تحقیق

قدمت سیستم‌های تشخیص نفوذ به دهه هشتاد میلادی بر می‌گردد و از آن زمان پژوهش‌های بسیاری در این زمینه انجام شده و مقالات متعددی چاپ گردیده است. از مطالعه‌ی منابع موجود چنین بر می‌آید که ترکیب سیستم‌های خبره و روش‌های آماری در دهه ۱۹۸۰ و اوایل دهه ۱۹۹۰ رواج فراوانی داشته است. افراد خبره امنیتی مدل‌های تشخیص را ارائه می‌کردند و بعد از آن در اواسط دهه ۱۹۹۰ تا پایان دهه، شناسایی رفتارهای عادی از غیرعادی از حالت دستی به صورت اتوماتیک تغییر پیدا کرد. تقسیم بندی روش‌های تشخیص نفوذ آنومالی توسط بویان و همکاران [۱۷] مشتمل بر ۶ دسته مجزا می باشد که در شکل (۱) نمایش داده شده است. این روش‌ها شامل روش‌های آماری[2]، روش‌های مبتنی بر دسته‌بندی[3]، روش‌های مبتنی بر خوشه‌بندی و نمونه‌های پرت[4]، محاسبات نرم[5]، روش‌های مبتنی بر دانش[6] و در نهایت یادگیرنده‌های ترکیبی[7] می‌باشد.



شکل ۱. روش‌های مورد استفاده مبتنی بر تشخیص آنومالی.

قبلاً تشخیص نفوذ به‌وسیله رویکردهای قانون محور انجام می‌شد، درحالی‌که کارشناسان یک مجموعه از قوانین را برای شرایط عادی و غیرعادی تعریف می‌کردند. این سیستم‌ها برای حملات شناخته‌شده بهتر عمل می‌کرد اما برای تشخیص حملات ناشناخته ناموفق بود. در اواخر دهه ۱۹۹۰ محققین روی توسعه سیستم‌های تشخیص نفوذ خودکار تمرکز کردند. بیشتر محققین از الگوریتم‌های داده‌کاوی و یادگیری ماشین جهت تشخیص حملات ناشناخته استفاده می‌کردند. در میان روش‌های تشخیص نفوذ مختلف، روش منطق فازی[8] نقش بسیار مهمی را بازی می‌کرد. از آنجا که روش پایه مورد استفاده در این پژوهش که در بخش آتی بحث و بررسی خواهد شد بر اساس خوشه‌بندی فازی است؛ لذا در اینجا لازم است تعدادی از سیستم‌های تشخیص نفوذ مبتنی بر خوشه‌بندی، اجمالا مورد بحث قرار گیرد. از مطالعه منابع موجود چنین برمی‌آید که جیان لیانگ و همکارانش یک سیستم تشخیص نفوذ با استفاده از الگوریتم خوشه‌بندی K-Means را توسعه دادند [۱۸]. آزمایش بر روی مجموعه داده‌های استاندارد KDD-۹۹ انجام گرفت. علی‌رغم این‌که خاتمه پذیری این الگوریتم تضمین‌شده است ولی جواب نهایی آن واحد نبوده و همواره جوابی بهینه نمی‌باشد. به‌طورکلی روش خوشه‌بندی K-Means دارای مشکلاتی است، که در ادامه به آن پرداخته می‌شود [۱۹].

[1] Virtual Private Networks
[2] Statistical
[3] Classification Based
[4] Clustering and Outlier Based
[5] Soft Computing
[6] Knowledge Based
[7] Combination Learners
[8] Fuzzy Logic

- روالی مشخص برای محاسبه مراکز اولیه خوشه‌ها وجود ندارد.
- جواب نهایی به انتخاب خوشه‌های اولیه وابسته است.
- اگر در مرحله‌ای از الگوریتم تعداد داده‌های متعلق به خوشه‌ای صفر شد، راهی برای بهبود ادامه روش وجود ندارد.
- در این روش فرض بر آن است که تعداد خوشه‌ها از ابتدا مشخص است اما معمولاً در کاربردهای زیادی تعداد خوشه‌ها مشخص نمی‌باشد.

برای غلبه بر این نقاط ضعف، بارتی و همکاران دو نسخه مختلف از الگوریتم K-Means اولیه را طراحی کردند [۲۰].

رن و همکاران [۲۱] از الگوریتم FCM[1] برای تشخیص نفوذ استفاده کردند. این مدل تشخیص نفوذ بر اساس تقسیم‌بندی فازی و خوشه‌بندی داده‌ها ساخته شد. نتایج آزمایش‌ها نشان داد که الگوریتم به‌طور مؤثری قادر به جداسازی داده‌های عادی و غیرعادی است.

گورویی و همکاران [۲۲] یک الگوریتم خوشه‌بندی FCM نیمه نظارتی را توسعه دادند. این روش بر نقاط ضعف FCM غلبه می‌کند. برای مثال حساسیت مقادیر اولیه و همگرایی کمینه‌های محلی با استفاده از داده‌های برچسب‌دار برای بهبود توانایی یادگیری الگوریتم FCM استفاده می‌شود.

سمپات و سونوانی [۲۳] نیز یک سیستم تشخیص نفوذ را با استفاده از الگوریتم بهبودیافته خوشه‌بندی بر پایه FCM توسعه دادند. IDFCM[2] یک نسخه از FCM اولیه است که به‌طور مرتب مراکز خوشه را به‌روزرسانی می‌کند. نتایج آزمایش‌ها حاکی از آن است که IDFCM یک تشخیص نفوذ با دقت بهتری نسبت به FCM ابتدایی ارائه می‌کند.

حامید و همکاران [۲۴] یک الگوریتم خوشه‌بندی ترکیبی برای تشخیص نفوذ ارائه کردند. این الگوریتم ترکیبی از الگوریتم MFPCM[3] و خوشه‌بندی نمادین فازی است. این روش از ۳۰ ویژگی با حساسیت و قدرت تفکیک مطلوب استفاده می‌کند.

گاناپاتی و همکاران [۲۵] یک سامانه تشخیص نفوذ بر اساس FCM وزن‌دار و الگوریتم ژنتیک ایمنی پیشنهاد دادند. FCM وزن‌دار، یک شکل اصلاح‌شده FCM است که یک سیستم را برای تشخیص حملات با دقت بیشتر ساخته است. الگوریتم ژنتیک ایمنی باعث بهبود عملکرد و احتمال رسیدن به نقطه بهینه سراسری می‌شود و همچنین موجب حل مشکلات چندبعدی می‌گردد. خزائی و راد [۲۶] یک روش جدید را بر اساس الگوریتم FCM برای بهبود عملکرد تشخیص نفوذ ارائه کردند. آزمایش‌ها بر روی مجموعه داده‌های KDD Cup-۹۹ انجام شد. کومار و همکارانش [۲۷] یک روش جدید تشخیص نفوذ را توسعه دادند که شامل خوشه‌بندی ویژگی‌های فازی بود. در این روش خوشه‌بندی ویژگی فازی برای کاهش ابعاد، فراخوان سیستمی مورداستفاده قرار گرفت.

پاندیزواری و کومار [۲۸] یک سامانه تشخیص ترکیبی برای محیط‌های ابری ارائه دادند. فعالیت این مدل ترکیبی در سه مرحله انجام می‌شود. در مرحله نخست، با استفاده از خوشه‌بندی فازی، قابلیت یادگیری شبکه عصبی مصنوعی را بهبود می‌بخشد. در مرحله دوم، مؤلفه‌های مختلف شبکه عصبی مصنوعی بر اساس مقادیر خوشه‌هایشان آموزش می‌بینند و در آخرین مرحله مؤلفه تجمیع فازی برای ترکیب کردن نتایج شبکه عصبی مصنوعی به‌کاربرده می‌شوند.

کارتیک و ناگاپان [۲۹] یک سامانه تشخیص نفوذ با استفاده از شبکه عصبی بیزین و روش KFCM[4] را توسعه دادند. این روش ترکیبی مشتمل بر دو مرحله است. در مرحله‌ی اول برای به دست آوردن مراکز خوشه‌ها از مؤلفه فازی Bisector استفاده شد. در مرحله‌ی دوم مرکز ثقل‌های[5] به‌دست‌آمده از مرحله‌ی قبل را برای یادگیری شبکه بیزین مورداستفاده قراردادند. این آزمایش‌ها بر روی مجموعه داده‌های استاندارد KDD Cup-99 انجام گردید. روستام و تالیتا [۳۰] یک الگوریتم تشخیص نفوذ بر اساس KFCM پیشنهاد کردند.

خزائی و فائز [۳۱] یک روش طبقه‌بندی ترکیبی برای تشخیص نفوذ شبکه توسعه دادند. این روش ترکیبی خوشه‌بندی فازی را با شبکه عصبی پرسپترون چندلایه ترکیب می‌کند. نمونه های آموزشی در مرحله اول با استفاده از خوشه‌بندی فازی دسته بندی شدند و داده‌های نامناسب پس از تشخیص به مجموعه داده‌های دیگری انتقال یافتند. علاوه بر این پرسپترون چندلایه با استفاده از برچسب‌های جدید آموزش داده خواهد شد. برای طبقه‌بندی فازی، از شبکه عصبی ARTMAP استفاده شد.

سورانا [۳۲]، یک سیستم تشخیص نفوذ ترکیبی را با استفاده از FCM و شبکه عصبی توسعه داد. این رویکرد داده‌های آموزشی را با استفاده از FCM به گروه‌های کوچک‌تر تقسیم می‌کند. بعداً شبکه‌های عصبی با استفاده از این زیرمجموعه‌ها آموزش داده شد. در پایان نتایج شبکه عصبی تجمیع شدند.

[1] Fuzzy C-Means
[2] Improved Dynamic Fuzzy C-Means
[3] Modified Fuzzy Possiblistic C-Means
[4] Fuzzy Kernel C-Means
[5] Centroid

کومار و هاریش [۳۳] الگوریتم RSKFCM[1] را پیشنهاد کردند که اطلاعات فضایی و معیار فاصله‌ای هسته را مورداستفاده قرارداد. نتایج آزمایش‌ها نشان داد که این روش یک تشخیص نفوذ با دقت بهتری نسبت به خوشه‌بندی‌های مبتنی بر FCM پیشین ارائه کرد.

در سال جاری برای غلبه بر مشکل همگرایی FCM و نقاط مرکزی اولیه، میشرا و نیک یک الگوریتم ترکیبی برای سیستم تشخیص نفوذ پیشنهاد دادند. در این سیستم ترکیبی از FCM به همراه روش بهینه‌سازی ازدحام ذرات[2] استفاده شد. الگوریتم بهینه‌سازی ازدحام ذرات برای غلبه بر نقایص FCM و دستیابی به بهینه‌سازی سراسری و همگرایی سریع مورد استفاده قرار گرفت [۳۴].

شواهد ناشی از بحث فوق حاکی از آن است که محققین در تلاش بودند که یک روش تشخیص نفوذ قوی و مؤثرتری را ارائه دهند. بعلاوه اغلب روش‌های موجود بر اساس رویکردهای یادگیری نظارتی و غیر نظارتی هستند. روش‌های تشخیص نفوذ مبتنی بر یادگیری نظارتی از انعطاف‌پذیری خوبی در مراحل آموزش و آزمون برخوردارند. به عبارت دیگر قادرند راهبرد های اجرایی خود را با وصول اطلاعات جدید بروز کنند. چناچه مقادیر مناسب برای حدود آستانه انتخاب گردند، نرخ تشخیص حمله‌های شناخته شده در این روش‌ها بسیار بالا خواهد بود. نقص اصلی یادگیری نظارتی این است که به حجم بالایی از نمونه های آموزشی نیاز دارد و اگر اطلاعات مربوط به حمله‌های جدید به داده‌های آموزش اضافه نشده و به مدل اعمال نشود این روش‌ها امکان شناسایی حملات جدید را ندارند. از طرف دیگر یادگیری غیرنظارتی قادر به تشخیص حملات ناشناخته می‌باشد و در مقایسه با سایر روش‌های دسته‌بندی یا روش‌های آماری، این روش عملکرد پایدارتری را نشان می‌دهد. ولی محدودیت‌های اصلی آن به شرح زیر می‌باشد: نرخ هشدار اشتباه بالا، عدم تشخیص انواع اختصاصی حملات و تأثیر پذیری نامطلوب از ابعاد.

در حال حاضر مهم‌ترین چالش در استفاده از سیستم‌های تشخیص نفوذ، حجم بالای هشدارهای تولیدی توسط آن‌هاست که عملاً امکان رسیدگی به هشدارها را از بین می‌برد. از چالش‌های مهم دیگری که سیستم‌های تشخیص نفوذ با آن درگیر هستند، ارائه روشی است که قادر به تشخیص حملات امنیتی با دقت بیشتر و میزان کمتر هشدارهای اشتباه باشد. از جمله چالش‌های دیگر که بایستی مورد توجه واقع شود این است که مهاجمان سایبری به منظور مخفی کردن فعالیت‌های خرابکارانه خود در صدد تغییر رفتار و روش‌های خود هستند تا با بهره‌گیری از شیوه حملات جدید و دستیابی به حفره های امنیتی در سیستم‌های مورد نظر به اهداف شوم خود دست یابند. به‌علاوه اخیراً همگان بر این موضوع اتفاق نظر دارند که اغلب حملات نوین و پیشرفته از چند مرحله تشکیل شده‌اند که هر مرحله از یک سناریوی جداگانه برخوردار است که در نهایت با کشف مجموعه سناریوهای دخیل در یک حمله چند مرحله‌ایی می‌توان به راهبرد و هدف واقعی حمله کننده دست یافت. در حال حاضر اغلب سیستم‌های تشخیص نفوذ موجود، همه ویژگی‌های بسته‌های شبکه را به منظور بررسی و کشف ساختار حملات مورد استفاده قرار می‌دهند، حال آن‌که برخی از این ویژگی‌ها مرتبط نبوده و زائد به نظر می‌رسند. این امر منجر به طویل شدن فرآیند تشخیص و کاهش کارآیی می‌شود. ماهیت خاص سیستم‌های نفوذ، تولید هشدارهای زیاد را ایجاب می‌کند که با توجه به حجم زیاد هشدارهای تولید شده مدیریت آن‌ها توسط عامل انسانی امکان‌پذیر نیست. به‌عنوان یکی دیگر از چالش‌های اصلی در سیستم‌های تشخیص نفوذ، به حجم بالای داده‌ها می‌توان اشاره نمود که برخی از آن‌ها در تشخیص نفوذ تاثیری ندارند. از سویی دیگر با در نظر گرفتن ترافیک بالا، کاستن از میزان هشدار اشتباه در سیستم تشخیص نفوذ اهمیت ویژه‌ای دارد.

با توجه به بررسی‌های انجام شده بر روی مدارک علمی موجود در این حیطه، مشکلاتی به شرح زیر شناسایی شدند که قادر هستند ما را در دستیابی به راه‌حل مناسب یاری کنند:

- لحاظ نمودن همه اطلاعات به‌دلیل زیاد بودن حجم اطلاعات و داده‌های امنیتی سخت بوده و تصمیم‌گیری در این زمینه را با مشکل مواجه نموده است.
- روش های امنیتی موجود در شناسایی حملات جدید پنهان و چند مرحله ای کارآیی کافی را ندارند.
- سازوکارهای پیچیده و نگهداری ناکافی تاریخچه حملات امکان شناسایی حملات پیچیده را فراهم نمی‌کند.
- بهره‌گیری از منابع غیر همگن منجر به تفسیر پیچیده داده‌ها شده است.
- به‌دلیل میزان بالای هشدارهای اشتباه امکان بازسازی یک نفوذ امنیتی با سطح اعتماد مناسب فراهم نیست.
- فقدان فرآیندی که قادر به ایجاد ارتباط موثر بین هشدارها از فعالیت‌های مخرب با سایر اطلاعات باشد.
- مدیران امنیتی قادر به دریافت جزئیات اطلاعات نیستند.

با این حال، برای حذف کردن مشکلات پیش گفت در این تحقیق، تلاش بر آن است تا یک روش خوشه‌بندی فازی به‌منظور تشخیص آنومالی شبکه مطرح شود. در این مقاله، به‌منظور فائق آمدن بر کاستی‌های سیستم‌های غیرنظارتی موجود، روشی

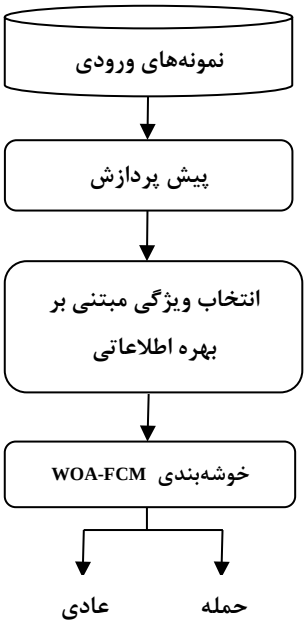
[1] Robust Spatial Kernel FCM
[2] Particle Swarm Optimization (PSO)

مشتمل بر سه مرحله پیش پردازش، انتخاب ویژگی و خوشه‌بندی پیشنهاد می‌شود که در آن از سیستم تشخیص نفوذ غیرنظارتی مبتنی بر FCM با بهره‌گیری از الگوریتم بهینه‌سازی نهنگ (WOA) جهت خوشه‌بندی داده‌ها در دنیای واقعی معرفی می‌گردد.

ساختار ارائه مطالب در این پژوهش بدین شکل می‌باشد: در بخش دوم ابتدا به بیان روش خوشه‌بندی FCM خواهیم پرداخت و سپس الگوریتم بهینه‌سازی نهنگ را مورد بررسی قرار می‌دهیم. در ادامه، الگوریتم ترکیبی WOA-FCM بر اساس ترکیب الگوریتم بهینه‌سازی نهنگ (WOA) و خوشه‌بندی FCM جهت خوشه‌بندی داده‌ها در دنیای واقعی معرفی می‌شود. در بخش سوم، مجموعه داده مورد استفاده، آماده‌سازی داده‌ها و فرایند معمولی سازی و همچنین نحوه ارزیابی روش پیشنهادی و بحث و نتیجه‌گیری را ارائه می‌دهیم. سپس در بخش چهارم جمع‌بندی و نتیجه‌گیری مطالب ارائه شده را بیان می‌کنیم.

۲- روش تحقیق

در این بخش، الگوریتم ترکیبی WOA-FCM بر اساس ترکیب الگوریتم‌های WOA و FCM جهت خوشه‌بندی داده‌ها در دنیای واقعی معرفی می‌شود. با عنایت به اهمیت بسیار زیاد و غیر قابل انکار معیارهای صحت، دقت، نرخ تشخیص و همچنین نرخ هشدار اشتباه در سیستم‌های تشخیص نفوذ، در این پژوهش سعی بر آن است که یک سیستم خودکار تشخیص حمله‌ها به شبکه‌های کامپیوتری با استفاده از سوابق گذشته اتصالات کاربران به شبکه ارائه نماییم؛ به‌طوری که صحت، دقت و نرخ تشخیص افزایش و نرخ هشدار اشتباه سیستم کاهش یابد. برای رسیدن به این هدف، نیاز به انتخاب روش خوشه‌بندی مناسب و سپس انتخاب معیارهای مناسب جهت تحلیل نتایج خوشه‌بندی داریم. نوع خوشه‌بندی انتخابی، در میزان دقت سامانه تشخیص نفوذ تاثیر بسزایی دارد، بنابراین الگوریتم خوشه‌بندی مورد نظر را در جهت اهداف خود بهبود می‌دهیم. امروزه منطق فازی به عنوان یک ابزار موفق در طراحی سیستم‌های تشخیص الگو و سامانه‌های تشخیص نفوذ محسوب می‌شود. به‌همین منظور در طراحی این سیستم با توجه به دو دلیل عمده از منطق فازی استفاده می‌کنیم. نخست این که در تشخیص نفوذ، ویژگی‌های کمّی فازی زیادی حائز اهمیت هستند و دلیل دوم استفاده از منطق فازی، جدا نمودن وضعیت عادی و غیرعادی و تبیین مرزی مشخص، میان این دو حالت است. با توجه به بررسی‌های انجام شده بر روی مدارک علمی موجود در این حیطه، الگوریتم FCM را می‌توان به‌عنوان نوعی جستجوی محلی در نظر گرفت. بنابراین، قرارگیری در کمینه محلی و حساسیت نسبت به مقادیر مراکز خوشه‌های اولیه، مشکل اصلی الگوریتم FCM است؛ بدین ترتیب مقادیر متفاوت برای مراکز خوشه‌های اولیه، نتایج متفاوتی را ایجاد می‌کند. با توجه به موارد پیش گفت، در می‌یابیم که به طور اجتناب ناپذیری باید از روش‌های فرا ابتکاری که قادر هستند پاسخ‌های تصادفی تولید شده را به سمت نتایج مناسب به‌روزرسانی کنند، استفاده نمود؛ چرا که رویکردهای مورد استفاده در حوزه تشخیص نفوذ، فضای پاسخی وسیع و بسیار پیچیده‌ای دارند. استفاده از قابلیت‌های خوشه‌بندی فازی به همراه بهره‌گیری از روش‌های فرا ابتکاری قادر به از بین بردن عدم قطعیت در تشخیص وضعیت عادی از غیر عادی بوده و نیز توانایی جستجوی مناسب فضای پاسخ موثر را دارد. این پژوهش بر آن است تا با استفاده از مفاهیم روش‌های فرا ابتکاری و استفاده از الگوریتم WOA برای بهینه‌سازی فرایند انتخاب مراکز خوشه‌های اولیه الگوریتم FCM، در مقایسه با روش‌های قبلی به نتایج قابل قبول‌تری دست‌یابد. در این مقاله، به منظور فائق آمدن بر کاستی‌های سیستم‌های غیرنظارتی موجود، روشی مشتمل بر سه مرحله‌ی پیش پردازش، انتخاب ویژگی و خوشه‌بندی پیشنهاد می‌شود که در آن از سیستم تشخیص نفوذ غیرنظارتی مبتنی بر FCM با بهره‌گیری از الگوریتم بهینه‌سازی نهنگ (WOA) جهت خوشه‌بندی داده‌ها در دنیای واقعی معرفی می‌گردد. شکل (۲) بلوک دیاگرام سیستم پیشنهادی را نمایش می‌دهد.



شکل ۲. بلوک دیاگرام سیستم پیشنهادی

۲-۱- خوشه‌بندی Fuzzy C-Means

منطق فازی نوعی از منطق‌های چند ارزشی یا منطق احتمالاتی است؛ این منطق با استدلالی سروکار دارد که به‌جای ثابت و دقیق بودن، تقریبی است. در منطق کلاسیک، یک عنصر فقط قادر به نشان دادن دو حالت درست یا نادرست است و متغیرهای منطق فازی در مقایسه با آن متعلق به یک مجموعه است. به بیان واضح‌تر درجه‌ی تعلق عناصر مجموعه با استفاده از تابع عضویت در مجموعه‌های فازی در بازه بسته صفر و یک قرار می‌گیرند [۳۵، ۳۶]. یکی از مهم‌ترین و پرکاربردترین روش‌های خوشه‌بندی روش C- میانگین فازی می‌باشد. FCM حالت توسعه‌یافته و بهبودیافته روش خوشه‌بندی K-Means هست و برخلاف K-Means هر داده می‌تواند متعلق به خوشه‌های مختلف با درجه‌های عضویت متفاوت باشد. الگوریتم FCM اولین بار توسط راسپینی معرفی شد و سپس دان و بزدک الگوریتم راسپینی را از الگوریتم خوشه‌بندی سخت به خوشه‌بندی فازی تعمیم دادند [۳۷]. الگوریتم FCM روشی برای خوشه‌بندی داده‌ها بر اساس بهینه‌سازی تابع هدف است. نتیجه خوشه‌بندی، درجه عضویت هر نقطه داده به مرکز خوشه‌بندی است که از طریق یک مقدار عددی نمایش داده می‌شود. بخش اصلی خوشه‌بندی فازی، تعیین معیار شباهت است که از طریق آن، فاصله میان الگوها مشخص می‌گردد. از آنجا که بیشتر روش‌های غیر نظارتی از معیار فاصله اقلیدسی استفاده می‌کنند؛ در این الگوریتم، از فاصله اقلیدسی به‌عنوان معیار شباهت استفاده‌شده که دلیل آن اخذ نتیجه مطلوب‌تر در مقایسه با سایر معیارهای فاصله‌ای بوده است. بنابراین تابع برازش به شکل زیر تعریف می‌شود (فرمول ۱):

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{i,j})^m \left\|c_i - x_j\right\|_{norm}^2 \qquad (۱)$$

که در این معادله، $u_{i,j}$ درجه عضویت x_j به c_i از i اُمین مرکز خوشه است. x_j درواقع j امین بُعد داده و c_i i امین مرکز خوشه است. $U = [u_{i,j}]_{c\times n}$ ماتریس تعلق است که درجه عضویت x_j به خوشه i-ام را نشان می‌دهد. $||_{norm}$ || نماد معمولی سازی ماتریس و $\|c_i - x_j\|$ فاصله اقلیدسی x_j از i امین مرکز خوشه است. به کمک روش ضریب لاگرانژ برای کمینه‌سازی تابع هدف J_m با استفاده از شرط معادله شماره (۲) خواهیم داشت:

$$\sum_{i=1}^{c}(u_{i,j}) = 1, \forall j = 1,2,\dots,n \qquad (۲)$$

$$u_{i,j} = \left[\sum_{k=1}^{c}\left[\frac{\left\|x_j - c_i\right\|}{\left\|x_j - c_k\right\|}\right]^{\frac{2}{m-1}}\right]^{-1} \qquad (۳)$$

$$c_i = \frac{\sum_{j=1}^{n}(u_{i,j})^m\, x_j}{\sum_{j=1}^{n}(u_{i,j})^m} \qquad (۴)$$

به‌وسیله دو فرمول محاسبه‌شده، الگوریتم FCM به شکل زیر توصیف می‌شود:

۱- تعیین تعداد خوشه‌ها c، مقداردهی اولیه مراکز خوشه‌ها $c_i^{(\cdot)}$ که $1 \le i \le c$، تعداد تکرار بیشینه و آستانه خطا ε جهت مشخص کردن زمان توقف الگوریتم.

۲- محاسبه مقادیر جدید u و c با استفاده از فرمول‌های شماره (۳ و ۴).

۳- مقدار تفاوت میان مراکز خوشه‌های جدید و درجه عضویت جدید فاز دوم از مقادیر قبلی آن‌ها محاسبه می‌شود، چنانچه مقدار به‌دست‌آمده کمتر از آستانه خطا ε یا تعداد تکرار برابر با مقدار بیشینه باشد، الگوریتم به پایان می‌رسد؛ در غیر این صورت، مرحله دوم انجام می‌شود.

الگوریتم FCM را می‌توان به‌عنوان نوعی جستجوی محلی در نظر گرفت؛ بنابراین مشکل اصلی الگوریتم FCM، قرارگیری در کمینه محلی و حساسیت نسبت به مقادیر مراکز خوشه‌های اولیه است؛ بدین ترتیب مقادیر متفاوت برای مراکز خوشه‌های اولیه، نتایج متفاوتی را در پی خواهد داشت. در ادامه یکی از روش‌های بهینه‌سازی جهت بهبود الگوریتم Fuzzy C-Means موردبررسی قرار می‌گیرد.

۲-۲- الگوریتم بهینه‌سازی نهنگ

الگوریتم بهینه‌سازی نهنگ[1] توسط جلیلی و لوئیس برای بهینه‌سازی مسائل عددی معرفی‌شده است. این الگوریتم رفتار هوشمندانه نهنگ‌های گوژپشت هنگام شکار را شبیه‌سازی می‌کند. این رفتار خاص برای یافتن غذا، «تغذیه حباب تور» نام دارد و تنها میان نهنگ‌های گوژپشت مشاهده می‌شود. در این روش شکار، نهنگ‌ها حین محاصره طعمه، حباب‌هایی را در مسیری دایره‌ای شکل ایجاد می‌کنند. به بیان ساده‌تر، شکار حباب تور بدین شکل است که نهنگ‌ها حدوداً ۱۲ m به زیرآب رفته و سپس حباب‌ها را در الگویی حلزونی شکل حول طعمه ایجاد کرده و درنهایت به دنبال حباب‌ها به سمت بالا شنا می‌کنند [۳۸].

۲-۲-۱- شکار محاصره‌ای

نهنگ‌های گوژپشت می‌توانند مکان طعمه‌ها را شناسایی کرده و آن‌ها را محاصره کنند. در الگوریتم WOA فرض بر این است که شکار هدف، بهترین راه‌حل موجود یا نزدیک به حالت مطلوب

[1] Whale Optimization Algorithm (WOA)

است. بعد از شناسایی بهترین عامل جستجو، عوامل دیگر جستجو تلاش می‌کنند که موقعیت خود را نسبت به بهترین عامل جستجو، به‌روزرسانی کنند. این رفتار را می‌توان در معادلات ذیل فرموله کرد:

$$\vec{D} = \left|\vec{C}.\vec{X^*}(t) - \vec{X}(t)\right| \quad (۵)$$

$$\vec{X}(t+1) = \vec{X^*}(t) - \vec{A}.\vec{D} \quad (۶)$$

که در اینجا، t نشان‌دهنده تکرار جاری، X^* بردار مکان بهترین راه‌حل به‌دست‌آمده، X بردار مکان هر عامل، || قدر مطلق و . نماد ضرب درایه‌ای[1] است. بردار A و C بدین ترتیب محاسبه می‌شوند:

$$\vec{A} = 2\vec{a}.r - \vec{a} \quad (۷)$$

$$\vec{C} = 2r \quad (۸)$$

که $\vec{a}$ به‌صورت خطی در دوره تکرار از ۲ تا ۰ کاهش می‌یابد و r یک عدد تصادفی بین ۰ و ۱ است.

۲-۲-۲- شکار محاصره‌ای

راهبرد حباب تور ترکیبی از دو روش است که می‌تواند بدین صورت مدل‌سازی شود:

الف) سازوکار محاصره‌ی انقباضی

این رفتار نهنگ‌ها با کاهش مقدار $\vec{a}$ در معادله‌ی (7) شبیه‌سازی می‌شود. شایان توجه است که بازه نوسان $\vec{A}$ نیز به‌وسیله $\vec{a}$ کاهش می‌یابد. به‌بیان‌دیگر، $\vec{A}$ مقداری تصادفی در بازه $[-a, a]$ بوده که a در طی تکرارها از ۲ به ۰ کاهش می‌یابد. با انتخاب مقادیر تصادفی$\vec{A}$ در بازه $[-۱, ۱]$، موقعیت جدید عامل جستجو می‌تواند در هرجایی میان موقعیت اصلی عامل و موقعیت بهترین عامل فعلی تعریف شود.

ب) به‌روزرسانی موقعیت حلزونی

در این روش، یک معادله حلزونی میان موقعیت نهنگ و طعمه جهت شبیه‌سازی حرکت مارپیچی نهنگ گوژپشت به شکل زیر ایجاد می‌شود:

$$\overrightarrow{D'} = \left|\vec{X^*}(t) - \vec{X}(t)\right| \quad (۹)$$

$$\vec{X}(t+1) = \overrightarrow{D'}.e^{bl}.cos(2\pi l) + \vec{X^*}(t) \quad (۱۰)$$

که در اینجا $\overrightarrow{D'}$ فاصله میان نهنگ و شکار، b ثابت توصیف‌کننده شکل لگاریتمی، l مقداری تصادفی بین $[-۱, ۱]$ و . نماد ضرب درایه‌ای است.

در حقیقت، نهنگ‌های گوژپشت هم‌زمان در مسیری حلزونی شکل و دایره‌ی انقباضی شنا می‌کنند. با فرض احتمال ۵۰ درصد، انتخاب یکی از این دو حرکت در جریان تکرار الگوریتم شبیه‌سازی می‌شود. مدل ریاضی بدین‌صورت می‌باشد:

$$\vec{X}(t+1) = \begin{cases} \vec{X^*}(t) - \vec{A}.\vec{D} & if \quad p < 0.5 \\ \overrightarrow{D'}.e^{bl}.cos(2\pi l) + \vec{X^*}(t) & if \quad p \geq 0.5 \end{cases} \quad (۱۱)$$

که در اینجا p عددی تصادفی بین $[۰, ۱]$ است.

۲-۲-۳- جستجو برای شکار

تقریباً تمامی الگوریتم‌های فرا ابتکاری، جواب بهینه را با استفاده از انتخاب تصادفی جستجو می‌کنند. در روش حباب تور، موقعیت جواب بهینه نامعلوم است، بنابراین نهنگ‌های گوژپشت به‌صورت تصادفی به دنبال شکار می‌گردند. برخلاف فاز استخراج که $\vec{A}$ در بازه $[-۱, ۱]$ بود، در این فاز $\vec{A}$، برداری از مقادیر تصادفی بزرگ‌تر از ۱ یا کمتر از ۱- فرض شده تا عامل جستجو را قادر به دور شدن از نهنگ مرجع کند. این دو عمل به ترتیب زیر فرموله می‌شوند:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}} - \vec{X}\right| \quad (۱۲)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \quad (۱۳)$$

در این معادلات X_{rand} به‌عنوان بردار موقعیت تصادفی انتخاب شده از جمعیت فعلی در نظر گرفته‌شده است. الگوریتم بهینه‌سازی نهنگ با مجموعه‌ای از راه‌حل‌های تصادفی شروع به کار می‌کند. در هر تکرار، عوامل جستجو، موقعیت خود را بر اساس توضیحات بالا به‌روزرسانی می‌کنند. شاخص $\vec{a}$ به‌منظور فراهم آوردن اکتشاف و استخراج، از مقدار ۲ تا ۰ متغیر است. بهترین راه‌حل جهت به‌روزرسانی مکان عوامل جستجو زمانی انتخاب می‌شود که $\vec{A}$ در بازه $[-۱, ۱]$ باشد درحالی‌که به‌طورمعمول یک عامل جستجوی تصادفی در حالت $\vec{A}$ بزرگ‌تر از ۱ یا کمتر از ۱- انتخاب می‌شود. با در نظر گرفتن مقدار p، الگوریتم WOA قادر است یکی از حرکات دایره‌ای یا مارپیچی را انتخاب کند. درنهایت، این الگوریتم با تحقق شرایط خاتمه، پایان می‌پذیرد.

[1] Element-by-element Multiplication

۲-۳- الگوریتم پیشنهادی

در این بخش، الگوریتم ترکیبی WOA-FCM بر اساس ترکیب الگوریتم‌های WOA و FCM جهت خوشه‌بندی داده‌ها در دنیای واقعی معرفی می‌شود. الگوریتم FCM را می‌توان به‌عنوان نوعی جستجوی محلی در نظر گرفت. بنابراین، قرارگیری در کمینه محلی و حساسیت نسبت به مقادیر مراکز خوشه‌های اولیه، مشکل اصلی الگوریتم FCM است؛ بدین ترتیب مقادیر متفاوت برای مراکز خوشه‌های اولیه، نتایج متفاوتی را ایجاد می‌کند. ازآنجاکه الگوریتم FCM نسبت به مقادیر اولیه خوشه‌ها حساس است، این مقاله از الگوریتم WOA برای بهینه‌سازی فرایند انتخاب مراکز خوشه‌های اولیه الگوریتم FCM استفاده می‌کند.

در این بخش سعی داریم مسئله خوشه‌بندی را با بهره‌گیری از WOA حل کنیم. با الهام از مقوله خوشه‌بندی، فرض می‌کنیم که هر عامل جستجو نشان‌دهنده k مرکز خوشه است (k از پیش تعریف‌شده و نشان‌دهنده تعداد خوشه‌هاست). بدین ترتیب هر عامل جستجوی X_i به‌صورت زیر ایجاد می‌شود:

$$X_i = (Z_{i1}, Z_{i2}, \dots, Z_{ik}) \quad (۱۴)$$

که در اینجا z_{ij} به j امین بردار مرکز خوشه i-امین عامل جستجو در خوشه c_{ij} اشاره می‌کند. بنابراین، یک گروه بیانگر تعداد خوشه‌های کاندید برای بردارهای مجموعه داده‌ها است.

این مسئله تحت عنوان یک مسئله کمینه‌سازی در نظر گرفته می‌شود و ما از تابع برازش زیر جهت استفاده در الگوریتم بهینه‌سازی نهنگ بهره بردیم:

$$d(s,z) = (\sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\|x_i - z_k\|^2)^{\frac{1}{2}} \quad (۱۵)$$

که در معادله بالا $s = (x_1, x_2, \dots, x_n)$ مجموعه‌ای از n داده که هرکدام از آن‌ها از m بعد یا ویژگی به‌صورت $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ تشکیل‌شده است. هدف نهایی در اینجا نسبت دادن هر نمونه x_i به یکی از k خوشه $z = (c_1, c_2, \dots, c_k)$ است به‌طوری‌که فاصله‌ی بین x_i و مرکز خوشه مربوطه به حداقل برسد. w_{ik} نمایانگر وزن مربوط به نمونه x_i در خوشه k می‌باشد که به‌صورت زیر تعریف می‌شود:

$$w_{ik} = \begin{cases} 1 & if\ \|x_i - z_k\|^2 = min_{1\le j\le n}\|x_i - z_j\|^2 \\ 0 & otherwise \end{cases} \quad (۱۶)$$

طبق توضیحات گفته شده، شبه کد الگوریتم پیشنهادی به شرح زیر است:

الگوریتم ۱. سیستم فازی غیر نظارتی جهت تشخیص نفوذ در شبکه بر مبنای الگوریتم بهینه‌سازی نهنگ

Initialize each search agent to contain k randomly cluster centers
while t < Iteration **do**
  for each search agent *i* **do**
    for each data vector x_p **do**
      Calculate the Euclidean distance of x_p to all cluster centers
      Assign x_p to the cluster c_{ij} such that

$$|x_p - z_{ij}| = min_{c=1,2,\dots,k}|x_p - z_{ic}|$$

      Calculate the fitness using Eq.17

$$fitness = \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij}|x_{ij} - z_{ij}| \quad (۱۷)$$

Where

$$w_{ij} = \begin{cases} 1 & if\ |x_i - z_{ij}| = min_{1\le m\le k}|x_i - z_{im}| \\ 0 & else \end{cases}$$

    end for
  end for
  X^* is the best search agent
  for each search agent **do**
    update a, A, C, I and p
    if $p < 0.5$ **then**
      if $|A| < 1$ **then**
        update search agent by Eq.6
      else if $|A| \ge 1$ **then**
        select random search agent
        update current search agent by Eq.13
      end if
    else if $p \ge 0.5$ **then**
      update the position of current search agent by Eq.10
    end if
  end for
  t = t + 1
end while
Initialize the cluster centers of FCM with position of the best cluster center vector. Then using this cluster centers, iterate the FCM algorithm.
Do Update the membership matrix by Eq.3
Refine the cluster centers by Eq.4
While (until it meets the convergence criteria)
Exit

۳- نتایج و بحث

یکی از مجموعه داده‌هایی که در این حوزه بسیار مورد استفاده قرارگرفته است؛ مجموعه داده KDD Cup ۹۹ می‌باشد که محققین متعددی به منظور ارزیابی آن را مورد استفاده قرار داده‌اند و همچنین این مجموعه‌داده در روش‌های مشابه

الگوریتم‌های پژوهش حاضر استفاده شده است. برای تهیه این مجموعه داده، حجم عظیم داده‌های گردآوری‌شده در پروژه DIDE[1] انجام‌شده با همکاری سازمان پروژه‌های تحقیقاتی پیشرفته دفاعی، وزارت دفاع ایالات‌متحده آمریکا و آزمایشگاه لینکلن دانشگاه MIT مورداستفاده قرار گرفت که هدف از گردآوری آن، ایجاد یک مجموعه داده استاندارد برای ارزیابی سیستم‌های تشخیص نفوذ بود. افراد خبره در حوزه امنیت اطلاعات اقدام به برچسب‌گذاری کلیه این رکوردهای این مجموعه داده نمودند به شیوه‌ای که به‌آسانی می‌توان تعلق هر رکورد به کلاس خاصی از حمله و عادی بودن رکورد را تشخیص داد. زیر بخش‌های پیش‌رو، آماده‌سازی مجموعه داده KDD Cup ۹۹، ارزیابی عملکرد و نتایج WOA-FCM را نشان می‌دهند.

۳-۱- آماده‌سازی KDD Cup 99

ویژگی‌های موجود در مجموعه داده KDD Cup ۹۹ [۳۹] متشکل از اشکال مختلفی همچون پیوسته و گسسته می‌باشد. ویژگی‌های عددی KDD99 مقیاس‌های مختلفی دارد و این نکته موجب انحراف به سمت برخی از ویژگی‌ها می‌شود؛ ازاین‌رو فرایند معمولی سازی ضرورت می‌یابد. فرایند معمولی سازی اعمال‌شده بر ویژگی‌های عددی به‌این‌ترتیب است:

۱- ویژگی‌های با مقیاس کوچک به‌صورت خطی با استفاده از فرمول ۱۸ به بازه [۰,۱] مقیاس‌بندی شده است.

$$\overline{X} = \frac{X - MinX}{MaxX - MinX} \qquad (۱۸)$$

که در اینجا X ویژگی با ارزش عددی بوده و $MinX$ و $MaxX$ به ترتیب مقدار کمینه و بیشینه‌ای است که X می‌تواند داشته باشد.

۲- ویژگی‌های با مقیاس بالا شامل ویژگی‌های src_bytes [میلیارد ۰,۱.۳] و dest_bytes [میلیارد ۰,۱.۳] از طریق اعمال مقیاس دهی لگاریتمی (با پایه ۱۰) معمولی سازی شده‌اند [۴۰].

۳- ویژگی‌های بولین که مقدارشان ۰ یا ۱ و یا در بازه [۰,۱] بوده معمولی سازی نشده‌اند.

تعداد ۴۱ ویژگی در مجموعه داده KDD Cup 99 وجود دارد. الگوریتم پیشنهادی با حذف ویژگی‌هایی با اهمیت کمتر در تشخیص نفوذ از ۳۰ ویژگی استفاده می‌کند؛ برای انجام این کار، روشی مبتنی بر بهره اطلاعاتی[2] به‌کار رفته است [۴۱].

بهره اطلاعاتی بر اساس آنتروپی ویژگی، میزان موضوعیت داشتن یک ویژگی داده‌شده را مورد ارزیابی قرار می‌دهد، به‌طور مثال نقش آن را در تعیین برچسب نوع کلاس بررسی می‌کند. اگر آن ویژگی موضوعیت داشته باشد، آنگاه آنتروپی‌های محاسبه‌شده نزدیک به ۰ بوده و بهره اطلاعاتی نیز نزدیک به ۱ خواهد بود. بهره اطلاعاتی برای ویژگی‌های گسسته محاسبه می‌شود بنابراین ویژگی‌هایی با مقادیر پیوسته بایستی گسسته سازی شوند. بدین منظور، ویژگی‌های پیوسته بر اساس تناوب نمونه‌ها، تقسیم‌بندی می‌شوند [۴۲، ۴۳].

S را مجموعه‌ای از نمونه‌های آموزشی با برچسب‌های متناظرشان در نظر بگیرید. فرض کنید m کلاس وجود دارد، مجموعه آموزشی شامل s_i نمونه از کلاس I و s بوده که بیانگر تعداد کل نمونه‌ها در مجموعه آموزش است. اطلاعات مورد انتظاری که برای دسته‌بندی یک نمونه داده شده مورد نیاز هستند به کمک فرمول زیر محاسبه شده‌اند [۴۲]:

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1} \frac{s_i}{s} log_2(\frac{s_i}{s}) \qquad (۱۹)$$

ویژگی F با مقادیر $\{f_1, f_2, \dots, f_v\}$ می‌تواند مجموعه آموزشی را به v زیرمجموعه $\{S_1, S_2, \dots, S_v\}$ تقسیم کند که در آن، S_j زیرمجموعه‌ای است که مقدار f_j برای ویژگی F را داراست. به‌علاوه فرض کنید S_j شامل s_{ij} نمونه از کلاس i است. آنتروپی ویژگی F به شرح زیر محاسبه می‌شود:

$$E(F) = \sum_{i=1} \frac{s_{1j} + \dots + s_{mj}}{s} * I(s_{1j}, \dots, s_{mj}) \qquad (۲۰)$$

بهره اطلاعاتی برای F می‌تواند بدین طریق محاسبه شود:

$$Gain(F) = I(s_1, \dots, s_m) - E(F) \qquad (۲۱)$$

در آزمایش‌های ما، بهره اطلاعاتی برای هر ویژگی محاسبه‌شده است. ویژگی‌های دارای مقادیر منفی شامل ویژگی‌های land، wrong_fragment، urgent، num-compromised، num_root، su_attempted، num_shell، is_host_login، is_guest_login و num-outbound_cmds است که در جدول (۱) نشان داده شده‌اند (متن‌های سایه‌دار، ویژگی‌های حذف‌شده هستند). حذف این ویژگی‌ها سرعت الگوریتم پیشنهادی را بدون تحت تأثیر قرار دادن صحت، بالا می‌برد.

[1] Darpa Intrusion Detection Evalution
[2] Information Gain

جدول ۱. ویژگی‌های مجموعه داده KDD Cup 99

نوع ویژگی	ویژگی‌ها
پیوسته	duration, src_bytes, dst_bytes, wrong_fragment, urgent, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_ creations, num_shells, num_access_files, num_outbound_cmds, count, srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate,dst_ host_srv_rerror_rate.
گسسته	protocol_type, service, flag, land, logged_in, is_host_login, is_guest_login.

۳-۲- ارزیابی عملکرد

مجموعه داده‌های KDD Cup ۹۹ به‌منظور ارزیابی سیستم تشخیص نفوذ پیشنهادی استفاده شده است، به این صورت که افراد خبره به هر رکورد موجود در این مجموعه داده برچسب حمله نسبت داده‌اند که با برچسب سیستم پیشنهادی به رکوردهای موجود در KDD Cup ۹۹ مقایسه گردیده است که با توجه به تشخیص درست یا نادرست رکوردها توسط سیستم پیشنهادی، چهار نتیجه محتمل به‌دست می‌آید که در جدول (۲) نشان داده شده‌اند و از آن به‌عنوان ماتریس درهم‌ریختگی[۱] یاد می‌شود.

جدول ۲. ماتریس درهم‌ریختگی

نوع رکورد	رکوردهای تشخیصی توسط IDS پیشنهادی		
	نوع برچسب	عادی (منفی)	**حمله (مثبت)**
رکوردهای واقعی	عادی (منفی)	TN (منفی واقعی)	**FP (مثبت کاذب)**
	حمله (مثبت)	FN (منفی کاذب)	**TP (مثبت واقعی)**

جهت ارزیابی کارایی سیستم‌های تشخیص نفوذ از توانایی سیستم در پیش‌بینی صحیح استفاده می‌شود. بر این اساس چهار تعریف داریم [۵]:

هشدار منفی واقعی (TN) : چنانچه سیستم داده‌های معمولی را درست تشخیص دهد.

هشدار مثبت واقعی (TP) : چنانچه سیستم داده‌های حمله را درست تشخیص دهد.

هشدار منفی کاذب (FN): چنانچه سیستم داده‌های حمله را معمولی تشخیص دهد.

هشدار مثبت کاذب (FP) : چنانچه سیستم داده‌های معمولی را حمله تشخیص دهد.

بر اساس تعاریف فوق معیارهای ارزیابی عددی به‌صورت زیر محاسبه می‌شوند:

- **نرخ هشدار منفی درست**[۲]: $\frac{TN}{TN+FP}$ که در منابع از آن به خصوصیت[۳] نیز یاد می‌شود.
- **نرخ هشدار مثبت درست**[۴]: $\frac{TP}{TP+FN}$ که آن را **نرخ تشخیص**[۵]، حساسیت[۶] و فراخوانی[۷] نیز می‌خوانند.
- **نرخ هشدار منفی نادرست**[۸]: $\frac{FN}{FN+TP} = 1 - sensitivity$
- **نرخ هشدار مثبت نادرست**[۹]: $\frac{FP}{FP+TN} = 1 - specificity$ که آن را **نرخ هشدار اشتباه** نیز می‌خوانند.
- **صحت**[۱۰]: $\frac{TN+TP}{TN+TP+FN+FP}$
- **دقت**[۱۱]: $\frac{TP}{TP+FP}$
- **معیار F**[۱۲]: $\frac{(\beta^2+1)\times Precision \times TPR}{\beta^2 \times Precision + TPR}$ که از آن با عنوان میانگین همساز[۱۳] دقت و نرخ تشخیص نیز یاد می‌کنند. β عددی مثبت است و برای وزن‌دهی استفاده می‌شود.
- **منحنی مشخصه عملکرد سیستم (ROC)**[۱۴]: این منحنی جهت نمایش تعادل بین کمیت‌های نرخ تشخیص و نرخ هشدار اشتباه در یک روش طبقه‌بندی مورداستفاده قرار می‌گیرد و عملکرد یک طبقه‌بند را به آستانه‌های ممکن کلی خلاصه می‌کند.

روش پیشنهادی مشتمل بر سه مرحله پیش‌پردازش، انتخاب ویژگی و خوشه‌بندی است. برای حل مشکل همگرایی FCM و یافتن مراکز اولیه مناسب از الگوریتم بهینه‌سازی نهنگ استفاده کردیم. به‌صورت تجربی و متعاقب اجراهای متعدد مشخص شد که روش پیشنهادی قادر به وصول به جواب‌های نسبتاً مناسب در زمان حل معقول بر روی پارامترهای انتخابی گردید. جهت تنظیم پارامتر در الگوریتم بهینه‌سازی نهنگ، بیشینه تعداد تکرار WOA و اندازه جمعیت اولیه را به‌ترتیب برابر ۲۰۰ و ۵۰ قرار دادیم.

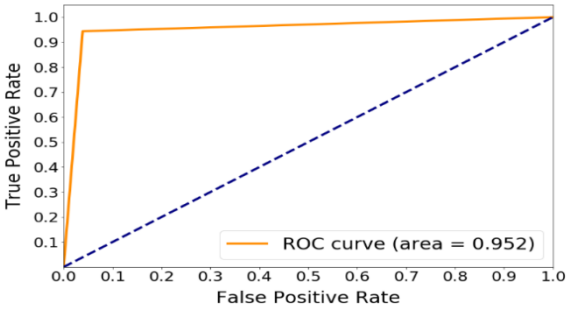
[1] Confusion matrix

[2] True Negative Rate (TNR)
[3] Specificity
[4] True Positive Rate (TPR)
[5] Detection Rate (DR)
[6] Sensitivity
[7] Recall
[8] False Negative Rate (FNR)
[9] False Positive Rate (FPR)
[10] Accuracy
[11] Precision
[12] F-Measure
[13] Harmonic Mean
[14] Receiver Operating Characteristics (ROC)

به منظور ارزیابی سیستم تشخیص نفوذ پیشنهادی، نتیجه اجرا بر روی مجموعه داده استاندارد تشخیص نفوذ KDD Cup ۹۹ با تنظیم پارامترهای ذکر شده با استفاده از معیارهای ارزیابی صحت، دقت، نرخ تشخیص، معیار F1 و نرخ هشدار اشتباه محاسبه و در جدول (۳) نشان داده شده است. همچنین شکل (۳) منحنی مشخصه عملکرد سیستم (ROC) روش WOA-FCM پیشنهادی را نمایش می‌دهند.

جدول ۳. نتیجه ارزیابی IDS پیشنهادی

معیار ارزیابی	روش WOA-FCM پیشنهادی
صحت (Accuracy)	٪ ۹۴/۶۹۷
دقت (Precision)	٪ ۹۹/۰۱۳
نرخ تشخیص (Recall)	٪ ۹۴/۳۵۴
معیار F1	٪ ۹۶/۶۲۷
نرخ هشدار اشتباه (FPR)	٪ ۳/۸۸۶



شکل ۳. منحنی مشخصه عملکرد سیستم (ROC)

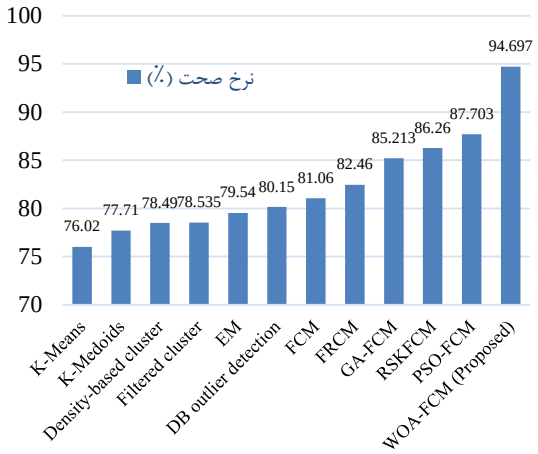
برای تشخیص نفوذ در سیستم‌ها معیارهای مختلفی مورد استفاده قرار می‌گیرد که تاکید عمده محققین بر روی دو معیار صحت و نرخ هشدار اشتباه است. از این‌رو در پژوهش حاضر نیز این دو معیار مد نظر قرار گرفته است؛ به بیانی دیگر سیستم تشخیص نفوذی را مطلوب قلمداد می‌کنیم که از صحت بالا و نرخ هشدار اشتباه پایین برخوردار باشد. به‌منظور قیاس روش پیشنهادی، آزمایش‌هایی بر روی مجموعه داده‌های استاندارد تشخیص نفوذ KDD Cup ۹۹ انجام شد و نتایج حاصل را با یازده روش تشخیص ناهنجاری بدون ناظر مقایسه کردیم. روش‌های به‌کار رفته در این مقایسه بدین ترتیب هستند: K-Means، K-Medoids، Density-based cluster، Filtered cluster، Expectation Maximization، Distance-based outlier detection، Fuzzy C-Means، Fuzzy Rough Clustering، GA-FCM، RSKFCM، FCM و PSO-FCM [۳۳، ۳۴، ۴۴، ۴۵].

جدول (۴) نشانگر مقایسه روش پیشنهادی با دیگر روش‌ها است. نتایج حاکی از آن است که روش WOA-FCM پیشنهادی در مقایسه با سایرین عملکرد بهتری دارد.

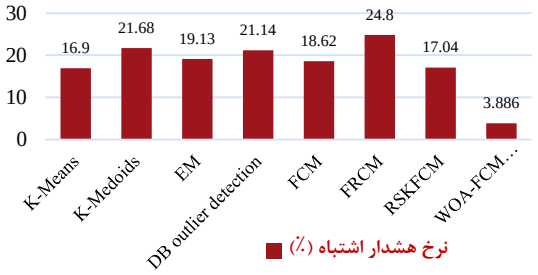
جدول ۴. نتیجه ارزیابی IDS پیشنهادی در مقایسه با دیگر روش‌ها

روش‌ها	نرخ صحت (٪)	نرخ هشدار اشتباه (٪)
K-Means	۷۶/۰۲	۱۶/۹۰
K-Medoids	۷۷/۷۱	۲۱/۶۸
Density-based cluster	۷۸/۴۹	–
Filtered cluster	۷۸/۵۳۵	–
Expectation Maximization	۷۹/۵۴	۱۹/۱۳
Distance-based outlier detection	۸۰/۱۵	۲۱/۱۴
Fuzzy C-Means	۸۱/۰۶	۱۸/۶۲
Fuzzy Rough Clustering	۸۲/۴۶	۲۴/۸۰
GA-FCM	۸۵/۲۱۳	–
RSKFCM	۸۶/۲۶	۱۷/۰۴
PSO-FCM	۸۷/۷۰۳	–
WOA-FCM (Proposed)	**۹۴/۶۹۷**	**۳/۸۸۶**

شکل (۴) نرخ صحت و شکل (۵) نرخ هشدار اشتباه مطالعه فعلی را در مقایسه با پژوهش‌های پیشین توسط سایر محققین با روش‌های متفاوت نمایش می‌دهد.



شکل ۴. مقایسه نرخ صحت IDS پیشنهادی با دیگر روش‌ها



شکل ۵. مقایسه نرخ هشدار اشتباه IDS پیشنهادی با دیگر روش‌ها

در جدول (۴)، شکل (۴) و شکل (۵) مقایسه روش پیشنهادی با دیگر روش‌ها را می‌توان ملاحظه نمود که نشان‌دهنده سرعت همگرایی بالاتر و در نتیجه صحت و نرخ تشخیص بهتر در روش WOA-FCM است. در بعضی از موارد در موقعیت‌های ویژه صدور هشدارهای نادرست باعث نقصان کارایی سیستم تشخیص نفوذ می‌شود؛ مثلا در مواردی که شبکه تحت نفوذ حمله‌های زیادی نیست، به عنوان وضعیت عادی شبکه شناخته می‌شود. مسئولین تامین امنیت شبکه معمولا ترجیح می‌دهند که مواجهه کمتری با هشدارهای اشتباه داشته باشند؛ زیرا در شرایطی که هشدارهای اشتباه کمتری در شرایط عادی شبکه وجود داشته باشد اختلال کم‌تری برای متخصصین شبکه ایجاد می‌کند و موجب ترغیب بیشتر آن‌ها به استفاده از سیستم تشخیص نفوذ پیشنهادی می‌گردد.

در حالت کلی با توجه به جدول (۴) و شکل (۴ و ۵) مشاهده می‌شود که روش پیشنهادی ما نسبت به دیگر روش‌ها عملکرد بهتری دارد. همان‌گونه که قبلاً اشاره شد، در روش پیشنهادی از الگوریتم WOA به‌منظور بهینه‌سازی فرایند انتخاب مراکز خوشه‌های اولیه الگوریتم FCM استفاده شد. یافته‌ها بر بهبود نرخ همگرایی، صحت، نرخ تشخیص و همچنین نرخ هشدار اشتباه توسط الگوریتم WOA-FCM در مقایسه با الگوریتم FCM ابتدایی و دیگر روش‌های مبتنی بر FCM دلالت می‌کند.

۴- نتیجه‌گیری

با توجه به کاربرد وسیع سیستم‌ها و شبکه‌های کامپیوتری، استفاده از سیستم تشخیص نفوذ در جهان برای امور دولتی، نظامی، تجاری و موارد دیگر گسترش فراوانی داشته است. در حقیقت سیستم‌های تشخیص نفوذ جزء اصلی یک زیرساخت امنیتی برای تشخیص تهدیدات قبل از آسیب گسترده و جدی به سیستم‌ها هستند. برای ایجاد سیستم‌هایی که قادر به ارتقا سطح امنیت باشند، چالش‌های گوناگونی وجود دارد که با پیشرفت فناوری و تغییر ویژگی‌های زیرساخت‌های ارتباطی و اطلاعاتی، حوزه جدیدی از پژوهش را برای محققان این عرصه ایجاد نموده است. امروزه منطق فازی به‌عنوان یک ابزار موفق در طراحی سیستم‌های تشخیص الگو و سامانه‌های تشخیص نفوذ محسوب می‌شود. در طراحی سیستم پیشنهادی با توجه به دو دلیل عمده از منطق فازی استفاده کردیم. نخست این که سامانه‌های تشخیص نفوذ با ویژگی‌های کمّی فازی بسیاری درگیر هستند و دلیل دوم، فقدان مرزی مشخص بین وضعیت عادی و غیر عادی است. در سال‌های اخیر، پژوهشگران به دنبال روش‌هایی بوده‌اند تا بتوانند یک سیستم تشخیص حملات به شبکه‌های کامپیوتری به‌صورت خودکار با نرخ صحت و دقت بالا و همچنین نرخ هشدار اشتباه پایین تولید کنند. روش پیشنهادی مشتمل بر سه مرحله پیش پردازش، انتخاب ویژگی و خوشه‌بندی است که در این پژوهش، سیستم تشخیص نفوذ غیر نظارتی مبتنی بر FCM با بهره‌گیری از الگوریتم بهینه‌سازی نهنگ پیشنهاد شد و با مجموعه داده‌های استاندارد تشخیص نفوذ KDD Cup ۹۹ مورد آزمایش قرار گرفت.

به منظور ارزیابی سیستم تشخیص نفوذ، مجموعه داده KDD Cup ۹۹، از پرکاربردترین مجموعه داده استاندارد محسوب می‌شود. علی‌رغم محدودیت‌هایی که برای این مجموعه‌داده قائل هستند؛ محققین متعددی به منظور ارزیابی آن را مورد استفاده قرار داده‌اند و همچنین این مجموعه داده در روش‌های مشابه الگوریتم پژوهش حاضر استفاده شده است. از این‌رو در این تحقیق به منظور ارزیابی سیستم تشخیص نفوذ پیشنهادی، این مجموعه داده مورد استفاده قرار گرفته است. با عنایت به اهمیت بسیار زیاد و غیر قابل انکار معیارهای صحت، دقت، نرخ تشخیص و همچنین نرخ هشدار اشتباه در سیستم‌های تشخیص نفوذ، این معیارها به‌عنوان معیارهای ارزیابی در نظر گرفته شد. یافته‌ها بر بهبود نرخ همگرایی، صحت و همچنین نرخ هشدار اشتباه توسط الگوریتم WOA-FCM در مقایسه با دیگر روش‌های غیر نظارتی دلالت می‌کند. علی‌رغم مزایای گفته‌شده این مدل پیشنهادی واجد مقدار محاسبات اضافی برای پیدا کردن مراکز خوشه‌های اولیه بهینه FCM می‌باشد که حتی با در نظر گرفتن این بار محاسباتی، روش پیشنهادی در مقایسه با دیگر مدل‌های موجود عملکرد بهتری دارد. بر همین اساس یافته‌های پژوهش حاضر می‌تواند در زمینه حل مسائل پیچیده مرتبط با IDS مؤثر واقع شود. جهت فعالیت‌های آینده می‌توان استفاده از سایر رویکردهای انتخاب ویژگی را مد نظر قرار داد. روش روش پیشنهادی می‌تواند با انتخاب خصیصه بر مبنای سایر روش‌ها و نیز تولید ابعاد گوناگون ویژگی از جمله روش‌های نگاشت فضای ویژگی به فضای جدید و استفاده از محاسبات تکاملی یا سایر رویکردهای جدید نتایج بهتری را رقم بزنند. از آنجایی که استفاده از هر الگوریتم فرا ابتکاری منجر به کاهش سرعت اجرا می‌شود؛ لذا به منظور برطرف کردن این نقیصه بهتر است فرآیند حل مسئله به وظایف کوچک‌تری تبدیل شده که جهت رسیدن به سرعت بالاتر به‌صورت هم‌زمان این زیر وظایف را اجرا کند و با محاسبات موازی به اجرای همزمان یک برنامه بر روی چند پردازنده برسد. علاوه بر موارد فوق، به منظور پژوهش می‌توان به تشخیص نفوذ به‌صورت چند سطحی توجه نمود به شکلی که در مرحله برخط رفتار غیرعادی در یک سطح شناسایی شده و بعد از آن نوع حمله‌ها در سطوح بعدی مشخص گردد. به‌طور کلی استفاده از خوشه‌بندی فازی مبتنی بر الگوریتم فرا ابتکاری ارائه شده عملکرد مطلوبی از خود نشان داده که امکان استفاده از آن را در حوزه‌های دیگر فراهم می‌کند که پیشنهاد می‌شود پژوهش‌های بیشتر با استفاده از الگوریتم‌های فرا ابتکاری تازه ابداع شده انجام شود.

۵- مرجع‌ها

[1] Al-Yaseen, W. L.; Othman, Z. A.; Nazri, M. Z. A. "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-Means for Intrusion Detection System"; Expert Systems with Applications 2017, 67, 296-303.

[2] Jun-lan, Y. A. O. "Intrusion Detection Technology and Its Future Trend"; Journal of Information Technology 2006, 4, 172–176.

[3] Ahmed, M.; Naser Mahmood, A.; Hu, J. "A Survey of Network Anomaly Detection Techniques"; Journal of Network and Computer Applications 2016, 60, 19-31.

[4] Abe, S.; Thawonmas, R. "A Fuzzy Classifier with Ellipsoidal Regions"; IEEE Transactions on Fuzzy Systems 1997, 5, 358-368.

[5] Wu, S. X.; Banzhaf, W. "The Use of Computational Intelligence in Intrusion Detection Systems: A Review"; Appaon"; IEEE Netw. 1994, 8, 26–41.

[10] Denning, D. E. "An Intrusion-Detection Model"; IEEE Transactions on Software Engineering 1987, SE-13, NO-2, 222–232.

[11] Chebrolu, S.; Abraham, A.; Thomas, J. P. "Feature Deduction and Ensemble Design of Intrusion Detection Systems"; Computers & Security 2005, 24, 295–307.

[12] Aljawarneh, S.; Aldwairi, M.; Yassein, M. B. "Anomaly-Based Intrusion Detection System through Feature Selection Analysis and Building Hybrid Efficient Model"; Journal of Computational Science 2018, 25, 152–160.

[13] Butun, I.; Morgera, S. D.; Sankar, R. "A Survey of Intrusion Detection Systems in Wireless Sensor Networks"; IEEE Communications Surveys & Tutorials 2014, 16, 266–282.

[14] Chawla, A.; Lee, B.; Fallon, S.; Jacob, P. "Host Based Intrusion Detection System with Combined CNN/RNN Model"; European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2018, 149-158.

[15] Ficke, E.; Schweitzer, K.; Bateman, R.; Xu, S. "Characterizing the Effectiveness of Network-Based Intrusion Detection Systems"; IEEE Military Communications Conference 2018, 76-81.

[16] Indhumathi, M.; Kavitha, S. "Distributed Intrusion Detection System for Cognitive Radio Networks Based on Weighted Fair Queuing Algorithm"; International Journal of Research in Computer Science, Engineering and Information Technology 2018, 426-436.

[17] Bhuyan, M. H.; Bhattacharyya, D. K.; Kalita, J. K. "Network Anomaly Detection: Methods, Systems and Tools"; IEEE Communications Surveys & Tutorials 2013, 16, 303–336.

[18] Jianliang, M.; Haikun, S.; Ling, B. "The Application on Intrusion Detection Based on K-Means Cluster Algorithm"; International Forum on Information Technology and Applications 2009, 1, 150–152.

[19] Ding, C.; He, X. "K-Means Clustering via Principal Component Analysis"; International Conference on Machine Learning, 29-37.

[20] Bharti, K.; Shukla, S.; Jain, S. "Intrusion Detection Using Unsupervised Learning"; International Journal of Computational Science and Engineering 1865, 2, 2010.

[21] Ren, W.; Cao, J.; Wu, X. "Application of Network Intrusion Detection Based on Fuzzy C-Means Clustering Algorithm"; Intelligent Information Technology Application 2009, 3, 19–22.

[22] Guorui, F.; Xinguo, Z.; Jian, W. "Intrusion Detection Based on the Semi-Supervised Fuzzy C-Means Clustering Algorithm"; Conference on Consumer Electronics, Communications and Networks 2012, 2667–2670.

[23] Sampat, R.; Sonawani, S. "Network Intrusion Detection Using Dynamic Fuzzy c Means Clustering"; Network 2015, 2, 135-141.

[24] Hameed, S. M.; Saad, S.; Alani, M. F. "An Extended Modified Fuzzy Possibilistic C-Means Clustering Algorithm for Intrusion Detection"; Lecture Notes on Software Engineering 2013, 1, 273-278.

[25] Ganapathy, S.; Kulothungan, K.; Yogesh, P.; Kannan, A. "A Novel Weighted Fuzzy C–Means Clustering Based on Immune Genetic Algorithm for Intrusion Detection"; Procedia Engineering 2012, 38, 1750–1757.

[26] Khazaee, S.; Rad, M. S. "Using Fuzzy C-Means Algorithm for Improving Intrusion Detection Performance"; International Financial Services Commission 2013, 27–29.

[27] Kumar, G. R.; Mangathayaru, N.; Narsimha, G. "An Approach for Intrusion Detection Using Fuzzy Feature Clustering"; The International Conference on Engineering & MIS 2016, 1–8.

[28] Pandeeswari, N.; Kumar, G. "Anomaly Detection System in Cloud Environment Using Fuzzy Clustering Based ANN"; Mob. Networks Appl. 2016, 21, 494–505.

[29] Principal, V. "Intrusion Detection System Using Kernel FCM Clustering and Bayesian Neural Network"; Data Bases 2013, 3, 391-399.

[30] Rustam, Z.; Talita, A. S. "Fuzzy Kernel C-Means Algorithm for Intrusion Detection Systems"; J. Theor. Appl. Inf. Technol. 2015, 81, 161-165.

[31] Khazaee, S.; Faez, K. "A Novel Classification Method Using Hybridization of Fuzzy Clustering and Neural Networks for Intrusion Detection"; Int. J. Mod. Educ. Comput. Sci. 2014, 6, 11-24.

[32] Surana, S. "Intrusion Detection Using Fuzzy Clustering and Artificial Neural Network"; Adv. Neural Networks, Fuzzy Syst. Artif. Intell. 2013, 209–217.

[33] Harish, B. S.; Kumar, S. V. A. "Anomaly Based Intrusion Detection Using Modified Fuzzy Clustering"; International J. of Interactive Multimedia and Artificial Intelligence 2017, 4, 54–59.

[34] Mishra, D.; Naik, B. "Detecting Intrusive Behaviors Using Swarm-Based Fuzzy Clustering Approach"; South Carolina Dental Association 2019, 837–846.

[35] Gaffarpour, R.; Pourmusa, A. A.; Ranjbar, A. M. "Presenting an Index for Evaluation of Power Network Security Using Fuzzy Set Theory"; Adv. Defence Sci. & Technol. 2019, 7, 289–304 (In Persian).

[36] Mendel, J. M. "Uncertain Rule-Based Fuzzy Systems"; Introduction and New Directions; Springer International Publishing, 2017.

[37] Bezdek, J. C.; Ehrlich, R.; Full, W. "FCM: The Fuzzy c-Means Clustering Algorithm"; Comput. Geosci. 1984, 10, 191–203.

[38] Mirjalili, S.; Lewis, A. "The Whale Optimization Algorithm"; Adv. Eng. Softw. 2016, 95, 51–67.

[39] "KDD-CUP 1999 Dataset"; http://kdd.ics.uci.edu/databases/kddcup99/, 2019.

[40] Revathi, M.; Ramesh, T. “Network Intrusion Detection System Using Reduced Dimensionality”; Indian J. Comput. Sci. Eng. 2011, 2, 61–67.

[41] Sabhnani, M.; Serpen, G. “Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context”; MLMTA 2003, 209–215.

[42] Kayacik, H. G.; Zincir-Heywood, A. N.; Heywood, M. I. “Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets”; Proceedings of the Annual Conference on Privacy, Security and Trust 2005, 94, 1723-1728.

[43] Kazemitabar, J.; Taheri, R.; Kheradmandian, Gh. “A Novel Technique for Improvement of Intrusion Detection via Combining Random Forrest and Genetic Algorithm”; Adv. Defence Sci. Technol. 2019, 10, 287–296 (In Persian).

[44] Syarif, I.; Prugel-Bennett, A.; Wills, G. “Unsupervised Clustering Approach for Network Anomaly Detection”; Networked Digital Technologies 2012, 135–145.

[45] Chimphlee, W.; Abdullah, A. H.; Sap, M. N. M.; Srinoy, S.; Chimphlee, S. “Anomaly-Based Intrusion Detection Using Fuzzy Rough Clustering”; International Conference on Hybrid Information Technology 2006, 1, 329–334.