شناسایی حملات DDoS با استفاده از شبکه‌های عصبی

نور حسن هادی[۱]، عباس مهدی زاده[۲و*]

۱- گروه کامپیوتر، دانشکده فنی مهندسی، دانشگاه امام رضا (ع)، مشهد، ایران

۲- استادیار، گروه کامپیوتر، مؤسسه آموزش عالی فردوس، مشهد، ایران

*** نویسنده مسئول: mehdiizadeh@gmail.com**

چکیده

امروزه با توسعه فناوری و رشد اینترنت اشیا، حملات سایبری در حال افزایش است. دستگاه‌های اینترنت اشیا می‌توانند در معرض تهدیدات و خطرات مختلف هکرها و بدافزارها از جمله سرقت اطلاعات، جعل و انکار سرویس قرار گیرند و خسارات مادی و معنوی زیادی به افراد و سازمان‌ها وارد کنند. بنابراین لازم است تدابیر امنیتی در این زمینه اتخاذ شود. تحقیقات اخیر در مورد مکانیسم‌های امنیتی دستگاه‌های اینترنت اشیا و سیستم‌های تشخیص نفوذ و ناهنجاری، نشان دهنده رواج استفاده از الگوریتم‌های یادگیری ماشین (ML) برای شناسایی ترافیک مخرب است که با استفاده از یک شبکه عصبی قادر به یادگیری یک مدل برای نمایش توالی از ارتباطات بین رایانه‌ها در یک شبکه است و با تحلیل و انتخاب ویژگی‌های درست، حملات متراکم با دقت بیشتری تشخیص داده می‌شوند. در این تحقیق، یک مدل تشخیص نفوذ مبتنی بر استخراج و انتخاب ویژگی و طبقه‌بندی شبکه‌های عصبی پیشنهاد شده است. ابتدا مجموعه داده جمع آوری شده و سپس داده‌های ورودی از قبل پردازش می‌شوند تا نویزها و داده‌های از دست رفته و افزونه حذف شوند و در ادامه اجرای فرآیند انتخاب ویژگی مبتنی بر الگوریتم بهینه سازی انتخابی، ابعاد مجموعه داده‌های کاهش می‌یابد. پس از آن، با استفاده از طبقه‌بندی شبکه‌های عصبی می‌توان یک مدل تشخیص نفوذ را برای یافتن حملات سیستم مبتنی بر IDS ساخت. نتایج ارزیابی ها نشان می‌دهد که روش پیشنهادی نسبت به روشهای موجود توانسته است نتایج به مراتب بهتری را ارائه کند.

کلمات کلیدی: اینترنت اشیا (IoT)، سیستم تشخیص نفوذ، طبقه‌بندی، شبکه عصبی، انتخاب ویژگی چندهدفه

۱. مقدمه

امروزه به دلیل تجاری‌سازی اینترنت و شبکه‌های محلی، حملات به سیستم‌های کامپیوتری در حال افزایش است. هدف بیشتر حملات سایبری تضعیف فرآیندهای امنیتی معمول در سیستم‌ها و واکنش بیش از حد به مهاجمان است. این اقدامات می‌تواند شامل خواندن اطلاعات امن یا محرمانه یا فقط هک کردن سیستم‌ها یا فایل‌های کاربر باشد[۱]. برای مقابله با این حملات، سیستم‌های تشخیص نفوذی معرفی شدند که وظیفه شناسایی و تعیین هرگونه استفاده غیرمجاز از سیستم، سوء استفاده یا آسیب توسط کاربران داخلی و خارجی را بر عهده دارند. تشخیص و جلوگیری از نفوذ یکی از مکانیسم‌های اصلی در تعیین امنیت شبکه‌ها و سیستم‌های کامپیوتری است و معمولا در کنار فایروال‌ها و به عنوان مکمل امنیتی برای آنها استفاده می‌شود.

یک سیستم تشخیص نفوذ قوی (IDS) دستگاه‌های شبکه را نظارت می‌کند و رفتار غیرعادی یا مخرب را در الگوهای فعالیتی شبکه تشخیص می‌دهد. یک سیستم تشخیص نفوذ نقش مهمی‌در امنیت شبکه‌های کامپیوتری ایفا می‌کند و از داده‌های کاربر و دستگاه‌های اینترنت اشیا در برابر فعالیت‌های مخرب محافظت می‌کند. علاوه بر این، اینگونه سیستم ها نه تنها هدفش شناسایی موفقیت آمیز مزاحمان در فعالیت‌های مخرب، بلکه نظارت بر تلاش‌های نقض امنیت با ارائه اطلاعات

به موقع در مورد سیستم امنیتی فعلی است [۲]. سیستم‌های تشخیص نفوذ به صورت نرم افزاری و سخت افزاری ایجاد می‌شوند و هر کدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزیت‌های سیستم‌های سخت افزاری است و عدم نقض امنیت توسط متجاوزان از دیگر ویژگی های این گونه سیستم‌هاست. اما سهولت استفاده از نرم افزار، سازگاری با پلتفرم‌ها و سیستم عامل‌های مختلف، چنین سیستم‌هایی را به انتخاب بهتری تبدیل می‌کند. به طور کلی سه کارکرد اصلی IDS عبارتند از: نظارت و ارزیابی، تشخیص و پاسخ. بر این اساس، هر IDS را می‌توان براساس روش‌های تشخیص نفوذ، معماری و انواع پاسخ های نفوذ دسته بندی کرد. اما بطور کلی یک سیستم تشخیص نفوذ مناسب باید بتواند رفتارهای استاندارد و غیرعادی جلسه را تشخیص دهد. این شامل هر رویداد، حالت، محتوا یا رفتاری است که با معیار از پیش تعریف شده، عادی یا غیرعادی تلقی می‌شود. بنابراین، زمینه تشخیص نفوذ به عنوان یک زمینه تحقیقاتی مهم توسعه یافته است، زیرا از نظر تئوری امکان راه اندازی یک سیستم بدون آسیب پذیری وجود ندارد. در یک شبکه جهانی، بسیاری از خدمات آنلاین و بسیاری از سرورهای بزرگ روی این سیستم در حال اجرا هستند. چنین شبکه‌هایی برای مهاجمان جذاب تر می‌شوند و بنابراین شبکه برای دفاع از این سیستم‌ها به مدل‌های تشخیص نفوذ هوشمند نیاز دارد [۱].

IDS های کارآمد معمولاً از طریق استفاده از تکنیک‌های داده کاوی ایجاد می‌شوند زیرا می‌توانند نفوذ را به روشی مناسب تشخیص دهند. با این حال، پیاده سازی و نصب چنین سیستم‌هایی می‌تواند به طور طبیعی پیچیده باشد. ویژگی‌های ذاتی سیستم را می‌توان به مجموعه‌ای از مشکلات متمایز براساس پارامترهای شایستگی، دقت و قابلیت استفاده تقسیم کرد، عمدتاً آن دسته از تکنیک‌هایی که مبتنی بر تشخیص ناهنجاری هستند، در مقایسه با تکنیک‌های تشخیصی قبلی که مبتنی بر امضاهای دست‌نویس هستند، درصد بیشتری از موارد نادرست را نسبت به موارد مثبت نشان می‌دهند. از این رو، پردازش داده‌ها و تشخیص نفوذ آنلاین برای این تکنیک‌ها دشوار است. سیستم تشخیص نفوذ در مراحل مختلفی مانند جمع آوری داده‌ها، پیش پردازش، انتخاب ویژگی (FS) و طبقه بندی عمل می‌کند [۲].

در روش پیشنهادی ابتدا سعی می‌کنیم داده‌های ورودی را با استفاده از روش بهینه سازی GOA کاهش دهیم. سپس با طبقه بندی درخت تصمیم یا شبکه عصبی MLP که یک الگوریتم یادگیری ماشینی است، سعی خواهیم کرد فضایی مشابه فضای آموخته شده را پیش بینی کنیم.

برای پرداختن به این الزامات، این مطالعه یک روش انتخاب ویژگی دو مرحله‌ای جدید را پیشنهاد می‌کند. در مرحله اول تمامی‌داده‌های آموزشی در فضایی با ابعاد کوچک نگاشت می‌شوند و در مرحله دوم انتخاب ویژگی براساس الگوریتم GOA انجام می‌شود. از آنجایی که این دو کار اکنون از هم جدا شده اند، اکنون می‌توان از ساختارهای پیچیده و ساده درخت تصمیم برای مراحل بعدی، پیش بینی و طبقه بندی استفاده کرد. سه مرحله روش پیشنهادی به شرح زیر است:

مرحله پیش پردازش: در مرحله اول و نرمال سازی داده‌ها، ابتدا ویژگی ها توسط سیستم استخراج و کدهای ابعادی نرمال سازی شده و شبکه عصبی براساس این کدهای ابعادی آموزش داده می‌شود.

مرحله انتخاب ویژگی: در مرحله بعد از روش GOA برای انتخاب ویژگی استفاده می‌شود که سعی می‌شود بهترین نمایش داده‌ها را در ابعاد کوچک به دست آورد. در پایان، صفات براساس امتیاز ویژگی (تابع تناسب) رتبه بندی شده و سپس صفاتی که بالاترین امتیاز را دارند انتخاب می‌شوند.

مرحله طبقه بندی: در این مرحله سعی می‌شود داده‌های نرمال شده و ترسیم شده توسط GOA طبقه بندی و گروه بندی شوند. در این بخش می‌توان از روش شبکه عصبی MLP بهبود یافته با الگوریتم دیگری استفاده کرد.

به منظور ارزیابی راه حل پیشنهادی، روش پیشنهادی به همراه یکی از جدیدترین روش‌ها در این زمینه با استفاده از زبان برنامه نویسی متلب پیاده سازی شده و نتایج به صورت گرافیکی در نرم افزار اکسل مقایسه خواهد شد.

در این تحقیق، طرحی برای استخراج و انتخاب ویژگی‌ها و طبقه‌بندی براساس یادگیری ماشین ارائه شده است. ایده این است که فرآیند انتخاب ویژگی را تقسیم بندی کنیم. روش پیشنهادی می‌تواند هم برای برنامه‌های تحت نظارت و هم برای برنامه‌های بدون نظارت اعمال شود. آزمایش‌های گسترده انجام شده بر روی مجموعه داده‌های مختلف، برتری روش پیشنهادی را نسبت به سایر روش‌های موجود از نظر دقت طبقه بندی، عملکرد خوشه بندی (دقت) با استفاده از ویژگی های محدود انتخاب شده تایید می‌کند.

ساختار این مقاله، ابتدا مباحث پایه در مورد اینترنت اشیا و موارد متداول حملات در این شبکه از جمله حملات DDoS بررسی می‌شود و مفاهیم اولیه در مورد سیستم‌های تشخیص نفوذ مطرح می‌گردد و ابعاد نرم افزاری استفاده از الگوریتم‌های یادگیری ماشین در این شبکه‌ها مورد بررسی قرار گرفته است که شامل شبکه‌های عصبی می‌باشد. سپس سیستم تشخیص نفوذ پیشنهادی بر مبنای پیش پردازش داده‌ها و در ادامه انتخاب ویژگی براساس الگوریتم ملخ و در ادامه طبقه‌بندی با الگوریتم شبکه عصبی ارائه شده است که به تشریح امکانات ارائه شده توسط این شبکه‌ها و معیارهای ارزیابی آنها می پردازد. علاوه بر این، جزئیات روش پیشنهادی مورد بحث قرار گرفته است.

در ادامه روش پیشنهادی توسط شبیه ساز MATLAB مورد ارزیابی قرار گرفته است و طرح پیشنهادی از جنبه‌های مختلف مورد ارزیابی قرار گرفته است. در نهایت نتایج ارزیابی‌های انجام شده مورد بررسی قرار گرفته او پیشنهاداتی برای تحقیقات آتی ارائه شده است.

۲. پیشینه تحقیق

یکی از جدیدترین الگوریتم‌های بهینه‌سازی معرفی شده در سال ۲۰۱۷، الگوریتم بهینه‌سازی ملخ GOA است [۳]. الگوریتم ملخ یک الگوریتم فرابتکاری الهام گرفته از طبیعت است که رفتار ملخ‌ها در طبیعت و حرکت گروهی ملخ‌ها را به سمت منابع غذایی تقلید و شبیه‌سازی می‌کند. مدل ریاضی الگوریتم GOA تقلیدی از رفتار ملخ‌ها در طبیعت برای حل مسئله بهینه‌سازی است. نتایج شبیه‌سازی ها نشان می‌دهد که الگوریتم ملخ قادر به ارائه نتایج برتر در مقایسه با الگوریتم‌های شناخته شده و اخیر در ادبیات است. نتایج شبیه‌سازی مشکلات واقعی نیز ثابت کرد که الگوریتم ملخ قادر است مشکلات واقعی را با فضای ناشناخته حل کند.

الگوریتم‌های الهام گرفته از طبیعت فرآیند جستجو را به دو قسمت اکتشاف و بهره برداری تقسیم می‌کنند. در اکتشاف، عوامل جستجو تشویق می‌شوند که ناگهان حرکت کنند، در حالی‌که تمایل به حرکت محلی در حین عملیات دارند. به فرایند یافتن بهترین مقادیر برای متغیرهای یک مشکل خاص به منظور به حداقل رساندن یا به حداکثر رساندن یک تابع هدف، بهینه‌سازی گفته می‌شود. ابتدا، پارامترهای مشکل باید مشخص شوند. دوم، محدودیت‌های اعمال شده بر پارامترها باید مشخص شوند. این محدودیت‌ها مشکلات بهینه‌سازی را به دو دسته با و بدون محدودیت تقسیم می‌کنند.

تکنیک طبقه‌بندی یکی از متداول‌ترین روش‌های یادگیری مدل به منظور پیش‌بینی در داده‌کاوی می‌باشد. در روش‌های پیش‌بینی از مقادیر بعضی از ویژگی‌ها برای پیش‌بینی کردن مقدار یک ویژگی مشخص استفاده می‌شود. طبقه‌بندی فرآیندی برای پیدا کردن مدلی به منظور مشخص کردن کلاس اشیا با توجه به ویژگی‌های آن‌ها می‌باشد[۴]. در الگوریتم های طبقه‌بندی، مجموعه داده اولیه به دو مجموعه داده‌های آموزشی و آزمایشی تقسیم می‌شود. با استفاده از مجموعه داده‌های آموزشی مدل ساخته می‌شود و از مجموعه آزمایشی برای اعتبارسنجی و محاسبه دقت مدل استفاده می‌شود.

یکی از پرکاربردترین الگوریتم‌های داده کاوی، الگوریتم درخت تصمیم است. در داده کاوی، درخت تصمیم یک مدل پیش بینی است به طوری که می‌توان از آن برای هر دو مدل رگرسیون و کلاس استفاده کرد. هنگامی که درخت برای وظایف

طبقه بندی استفاده می‌شود، به عنوان درخت طبقه بندی شناخته می‌شود و زمانی که از آن برای فعالیت‌های رگرسیون استفاده می‌شود، درخت تصمیم گیری رگرسیون نامیده می‌شود [۵].

سیستم‌های رایانه‌ای به نام شبکه‌های عصبی مصنوعی (ANN) که به عنوان سیستم‌های پیوندی نیز شناخته می‌شوند، از شبکه‌های عصبی بیولوژیکی که زیربنای مغز حیوانات هستند، مدل‌سازی شده‌اند. این سیستم‌ها نمونه‌های داده‌های آموزشی را مطالعه می‌کنند تا در مورد فعالیت‌های مختلف بیاموزند. به عنوان مثال، در شناسایی تصویر، شبکه‌های عصبی می‌توانند نمونه عکس‌هایی را مطالعه کنند که قبلاً به صورت دستی برچسب «گربه‌پسند» یا «بدون گربه» داشتند و از این نتایج تحلیلی، تشخیص تصاویر از جمله گربه‌ها را یاد بگیرند. استفاده در تصاویر اضافی براساس شبکه ای از اجزای به هم پیوسته به نام نورون های مصنوعی، یک ANN (مشابه نورون های بیولوژیکی در یک مغز بیولوژیکی). یک سیگنال می‌تواند از طریق هر ارتباط (سیناپسی) بین نورون ها از یک نورون به نورون دیگر ارسال شود.

شبکه‌های عصبی از جمله الگوریتم‌های یادگیری ماشینی هستند که با دریافت داده‌های ورودی و خروجی سعی در پیش بینی فرآیند تبدیل ورودی به داده‌های خروجی با استفاده از رگرسیون دارند. در این حالت با داده‌های ورودی جدید می‌توان خروجی را پیش بینی کرد. کاربر حداکثر تعداد لایه‌ها را در شبکه‌های عصبی پرسپترون چند لایه تنظیم می‌کند. نورون های زیادی در هر لایه وجود دارند که اطلاعات دریافتی از لایه قبل و خروجی را به لایه بعد پردازش می‌کنند. توجه داشته باشید که اکثر برنامه‌ها از یک لایه ورودی استفاده می‌کنند که تعداد نورون های آن توسط کاربر انتخاب می‌شود. علاوه بر لایه ورودی، یک لایه خروجی نیز وجود دارد و تعداد نورون های آن به اندازه خروجی های آن است. توابع سیگموئید توسط نورون ها در لایه خروجی یک شبکه عصبی پرسپترون چند لایه در لایه اول استفاده می‌شود که برای تخمین رگرسیون استفاده می‌شود [۶].

شرح مجموعه داده NSL-KDD (آزمایشگاه امنیت ملی-کشف دانش و داده کاوی) شکل پیشرفته KDD۹۹ برای پیشی گرفتن از محدودیت‌های آن است. در ابتدا رکوردهای تکراری در مجموعه‌های آموزشی و تست حذف می‌شوند. دوم، رکوردهای مختلفی از KDD۹۹ اصلی برای دستیابی به نتایج قابل اعتماد از سیستم‌های طبقه بندی کننده انتخاب شده اند. سوم، موضوع توزیع احتمال نامتعادل حذف شد. مجموعه داده NSL-KDD دارای ۱۲۵۹۷۳ نمونه آموزشی و ۲۲۵۴۴ نمونه آزمایشی با ۴۱ ویژگی است که ۳۸ ویژگی سازگار و سه طبقه بندی شده (مقدار گسسته) هستند. ۶ متغیر پیوسته رد شدند زیرا به طور قابل ملاحظه ای ۰ بودند. مجموعه داده‌های آموزشی شامل ۲۳ برچسب بالقوه (عادی + ۲۲ برچسب مربوط به انواع مختلف نفوذ) است. به طور همزمان، مجموعه داده‌های آزمون دارای ۳۸ کلاس است، که نشان می‌دهد که اطلاعات آزمون هیچ نفوذی در دوره آموزشی وجود ندارد. ۲۳ کلاس آموزشی و ۳۸ کلاس تستی به طور کلی ۲۱ کلاس دارند. دو کلاس فقط در مجموعه آموزشی رخ می‌دهد و ۱۷۶ کلاس برای اطلاعات آزمون استثنایی بودند. حدود ۱۶,۶ درصد از نمونه‌های موجود در مجموعه داده‌های آزمایشی مربوط به کلاس‌های استثنایی است که در طول دوره آموزشی وجود ندارند. این واریانس در توزیع کلاس، دشواری بیشتری را برای طبقه بندی کننده‌ها ایجاد می‌کند [۷].

کلاس‌های آموزشی/آزمایی مربوط به یکی از پنج طبقه‌بندی بالقوه بود: NORMAL، PROBE، R۲L، U۲R و DoS . کلاس NORMAL اشاره دارد که هیچ ناهنجاری ارائه نشده است. این پنج کلاس به‌عنوان آخرین کلاس‌های منجر به نتایج در نظر گرفته می‌شوند. این کلاس‌ها هنوز برای تعریف نفوذ مفید هستند، و هنوز هم بسیار نامتعادل (یک ویژگی مهم از داده‌های نفوذ) هنوز هم تعدادی نمونه در هر کلاس به اندازه کافی بزرگ برای ارائه نتایج مهم‌تر هستند.

۳. کارهای مرتبط

از دهه‌های گذشته تا به امروز، تحقیقات و پیشرفت‌ها در زمینه سیستم های تشخیص نفوذ به دلیل دگرگونی آن همچنان ادامه دارد. در ادبیات، تحقیقات زیادی با استفاده از الگوریتم‌های یادگیری ماشین کلاسیک، مانند درخت تصمیم، بیزی ساده (NB)، رگرسیون لجستیک، خوشه‌بندی حداکثر انتظار، ماشین بردار پشتیبان، مدل پنهان مارکوف (HMM) و K نزدیکترین همسایه (KNN) ارائه شده است.

تسانگ و همکاران [۵] یک روش IDS ژنتیکی چند منظوره فازی (MOGFIDS) را برای تشخیص ناهنجاری ها پیشنهاد کرده اند. این روش می‌تواند با یافتن مجموعه ای از ویژگی های بهینه، به عنوان مجموعه ای از ویژگی های پنهان عمل کند. علاوه بر این، روش آنها یک سیستم مبتنی بر قانون ژنتیکی فازی (GFRBS) است که از یک چارچوب تکاملی مبتنی بر چندین عامل هوشمند تکامل یافته است. این چارچوب برای ساخت GFRBS بر روی قابلیت تفسیر و دقت برای سیستم‌های تشخیص نفوذ طراحی شده است. نویسندگان از مجموعه داده KDD برای آموزش و آزمایش استفاده کردند. علاوه بر این، این مدل به عنوان قوانین IF-THEN فازی، با نرخ تشخیص۹۲،۷۷ (DR) ٪ و دقت ۷۴،۷۴٪ در طبقه بندی ترافیک شبکه عادی استخراج شده است. این تکنیک چهار دسته اصلی حملات را طبقه بندی می‌کندDoS :، Probe، U۲R و .R۲L با این حال، عدم دقت برای U۲R و R۲L باعث می‌شود این روش در زمینه سیستم‌های تشخیص نفوذ دقیق نباشد.

یک سیستم تشخیص نفوذ هوشمند توسط گاناپاتی و همکاران توسعه داده شده است. [۸] برای شناسایی حملات در شبکه‌های بی سیم. نویسندگان یک الگوریتم تشخیص فاصله مبتنی بر وزن (WDBOD) را برای افزایش پیش‌بینی ثابت برای محاسبه عدم تطابق-K نزدیک‌ترین همسایه ایجاد کردند. در این مدل، دقت تشخیص حملات DoS و Probe در مجموعه داده KDDCUP۹۹ بیش از ۹۹ درصد است.

با استفاده از الگوریتم ژنتیک فازی، جانگ سویساگ و همکاران. [۹] یک سیستم تشخیص نفوذ بلادرنگ را برای شناسایی انواع شناخته شده و ناشناخته حملات معرفی کرد. نویسندگان از مجموعه داده RLD۰۹ برای آموزش و آزمایش استفاده کردند. مجموعه داده مورد استفاده دارای دو دسته اصلی حملات DoS و Probe است. علاوه بر این، مجموعه داده دارای ۱۷ دسته حمله جزئی است که به دو حمله اصلی همراه با ترافیک عادی دسته بندی می‌شوند. میانگین دقت نتیجه آزمایش تقریباً ۹۷٪ با مثبت کاذب ۱،۱۳ (FP) و منفی کاذب ٤،۱۰ (FN) بود.

یک مدل سیستم تشخیص نفوذ هوشمند دیگر توسط گاناپاتی و همکاران توسعه داده شد[۱۰]. برای طبقه بندی و انتخاب ویژگی ها. الگوریتم طبقه بندی SVM چند کلاسه مبتنی بر قانون پیشرفته (IREMSVM) نامیده می‌شود. این الگوریتم نسخه اصلاح شده الگوریتم پیشرفته مبتنی بر عامل هوشمند کلاسیک SVM در روش نمونه گیری کلاسی است. با استفاده از قوانین و به دست آوردن اطلاعات از مجموعه دادهKDDCUP۹۹ ، نویسندگان روش جدیدی را برای انتخاب ویژگی ها پیشنهاد کردند. یک روش مبتنی بر قانون برای انتخاب تاپل‌ها اعمال می‌شود. دقت طبقه بندی برای دسته‌های DoS و Probe با استفاده از ۱۹ ویژگی در مقایسه با سایر دسته‌های حمله بسیار بالا بود.

امبوسعیدی و همکاران [۱۱] مدلIDS ، یعنی LSSVM-IDS را پیشنهاد کرد. در این مدل، نویسندگان یک الگوریتم انتخاب ویژگی به نام انتخاب اطلاعات متقابل ویژگی انعطاف‌پذیر (FMIFS) را با حداقل مربعات پیشنهادی SVM ترکیب کردند. الگوریتم FMIFS تکاملی از الگوریتم Battiti است که هدف اصلی آن کاهش افزونگی ویژگی است. در این مدل، از سه مجموعه داده برای ارزیابی مدل استفاده می‌شود، یعنیKDDCUP۹۹ ، NSLKDDCUP۹۹ و Kyotoنتایج به‌دست‌آمده، با استفاده از مجموعه برچسب اصلاح‌شدهKDDCUP۹۹ ، نرخ تشخیص پایینی را برای حملات U۲R و R۲L با دقت فراهم می‌کند. به طور کلی ۷۸،۸۶٪ را نشان داد.

یک مدل جدید برای IDS براساس شبکه یادگیری و بهینه سازی ازدحام ذرات (PSO-FLN) توسط علی و همکاران توسعه داده شده است. [۶] نویسندگان از مجموعه داده KDDCUP۹۹ برای آزمایشات استفاده کردند. در این مدل، نویسندگان دریافتند که تعداد نورون‌های پنهان دقیقاً عملکرد کلی سیستم را کنترل کرده و بر آن تأثیر می‌گذارد. نتایج نشان داد که این مدل در آزمون دقت بهتر از سایر روش‌های یادگیری عمل می‌کند. علاوه بر این، نویسندگان دریافتند که حملات R۲L از دقت کمتری نسبت به سایر دسته‌های حملات برخوردار هستند.

در یک مطالعه اخیر، یک مدل چند طبقه‌بندی پیشنهادی برای تشخیص ناهنجاری‌های شبکه مبتنی بر یادگیری ماشین توسط نویر و همکاران ارائه شد[۱۲]. این مدل از الگوریتم آنلاین برآوردگر وابستگی متوسط (AODE) نامیده می‌شود که نسخه پیشرفته الگوریتم NB است. براساس ویژگی تک والد، AODEویژگی های همه پیش بینی های چند طبقه بندی ۱-وابستگی را به طور متوسط انجام می‌دهد. در این مدل، نویسندگان از مجموعه داده UNSW-NB۱۵ استفاده کردند و دقت ۸۳,۴۷٪ را با فاصله ۶,۵۷٪ گزارش کردند. علاوه بر این، این مدل از دقت بالایی در تشخیص حملات کرم‌ها در مقایسه با سایر دسته‌های حملات برخوردار است.

نانسی و همکاران [۱۳] مدلی برای انتخاب و طبقه بندی ویژگی پیشنهاد کرد. در انتخاب ویژگی، نویسندگان مدل جدیدی به نام الگوریتم انتخاب ویژگی بازگشت پویا (DRFSA) ایجاد کردند. این مدل از هر دو روش بسته بندی و فیلترینگ بهره می برد. برای طبقه بندی، یک درخت تصمیم هوشمند با گسترش الگوریتم درخت تصمیم گیری سنتی با قوانین زمانی و فازی ساخته می‌شود. در این کار، از مجموعه داده KDDCUP۹۹ برای ارزیابی الگوریتم پیشنهادی استفاده شده است. دقت تشخیص DoS و Probe در مقایسه با U۲R و R۲L قابل قبول است که بسیار پایین است.

الهفناوی و ابوناسر [۱۴] در مطالعه خود یک چارچوب ترکیبی از الگوریتم فازی اصلاح شده ژنتیکی (HGFA) را برای تولید خروجی های بهینه برای متخصصان امنیتی در طبقه بندی حملات اصلی و جزئی معرفی کردند. مدل تطبیقی با استفاده از الگوریتم ژنتیک فازی (GFA) تکامل یافته است. هر GFA از الگوریتم‌های ژنتیک دوگانه (GA) تشکیل شده است. قسمت بیرونی برای تکامل مجموعه‌های فازی و قسمت داخلی برای تکامل قوانین فازی است GFA .خارجی به GFAداخلی در مرحله آموزش کمک می‌کند، جایی که بهترین فرد در GFA خارجی با GFA داخلی ضعیف تعامل می‌کند تا راه حل‌های جدیدی ایجاد کند که قابلیت پیش بینی حملات جهش یافته را افزایش می‌دهد. هر دو GFA از طریق فرآیند بهینه‌سازی با یکدیگر ارتباط برقرار می‌کنند تا بهترین قوانین را برای حملات دسته‌ای معمولی، عمده و جزئی ایجاد کنند.

در این تحقیق [۱۵]، یک سیستم تشخیص نفوذ مبتنی بر ANFIS برای شناسایی حملات در شبکه‌ها پیشنهاد شده است. ANFIS ترکیبی از مدل تداخل فازی و ANN است که نسبت به سایر تکنیک‌ها برتری دارد. علاوه بر این، در این مطالعه، نویسندگان از الگوریتم CSO بهینه‌سازی جستجوی Crow برای بهینه‌سازی مدل ANFIS برای افزایش عملکرد آن نسبت به تشخیص نفوذ استفاده کردند که یک مزیت برای سیستم IDS است. مدل آنها برای حل مشکلات تشخیص نفوذ استفاده شده است و مدل با استفاده از مجموعه داده آشنا NSL-KDD اعتبار سنجی شده است. مدل پیشنهادی با سایر تکنیک‌های موجود مانندBPNN ، FC-ANN، GA-ANFIS و PSO ANFIS مقایسه شده است. نتایج تشخیص نفوذ براساس مجموعه داده NSL-KDD در مقایسه با آن مدل‌ها بهتر و کارآمدتر بود زیرا نرخ تشخیص ۹۵,۸۰٪ و نتیجه FAR ۳.۴۵٪ بود.

در این مقاله [۱۵]، یک سیستم تشخیص نفوذ معرفی شده است که از مفاهیم داده کاوی و یادگیری ماشین برای تشخیص الگوهای نفوذ شبکه استفاده می‌کند. در این رویکرد، یک شبکه عصبی مصنوعی (ANN) به عنوان یک تکنیک یادگیری در تشخیص نفوذ استفاده می‌شود. الگوریتم فرابتکاری با رویکرد swarm-based برای کاهش خطاهای تشخیص نفوذ

استفاده می‌شود. در این روش از الگوریتم بهینه سازی ملخ (GOA) برای یادگیری بهتر و دقیق تر شبکه‌های عصبی مصنوعی برای کاهش میزان خطای تشخیص نفوذ استفاده می‌شود. نقش الگوریتم GOAMLP به حداقل رساندن خطای تشخیص نفوذ در شبکه عصبی با انتخاب پارامترهای مفید مانند وزن و بایاس است. پیاده‌سازی در نرم‌افزار MATLAB انجام می‌شود و با استفاده از مجموعه داده‌های KDD و UNSW نشان می‌دهد که روش پیشنهادی ترافیک و حملات غیرعادی و مخرب را با دقت بالا شناسایی می‌کند. روش GOAMLP نسبت به تکنیک‌های پیشرفته موجود مانندRF ، XGBoost و یادگیری تعبیه‌شده ANN با الگوریتم‌هایBOA ، HHO و BWO در تشخیص نفوذ شبکه، عملکرد بهتری دارد و دقیق‌تر است.

در تحقیقات [۱۶]، یک بهینه سازی محدود مبتنی بر ماشین یادگیری گسترده برای تشخیص نفوذ شبکه پیشنهاد شده است. در این مطالعه، یک استراتژی یادگیری تطبیقی افزایشی برای به دست آوردن تعداد مطلوب نورون های نهفته پیشنهاد شده است. معیارهای بهینه سازی و یک روش مقایسه ای برای افزایش نورون های نهفته با جستجوی باینری توسعه داده شده است. با توجه به تعداد زیاد آزمایش‌های انجام‌شده در این مطالعه، نتایج نشان داد که رویکرد پیشنهادی در ساخت مدل‌هایی با نرخ تشخیص حمله و سرعت یادگیری خوب مؤثر بوده است.

در مطالعه [۱۷]، یک طراحی سازگار از سیستم‌های تشخیص نفوذ مبتنی بر ماشین‌های آموزشی افراطی ارائه شده است. در سیستم پیشنهادی، قابلیت شناسایی حملات شناخته شده و جدید فراهم شده و با توجه به روند جدید الگوهای اطلاعاتی ارائه شده توسط کارشناسان امنیتی، به روش‌های مقرون به صرفه به روز رسانی شده است.

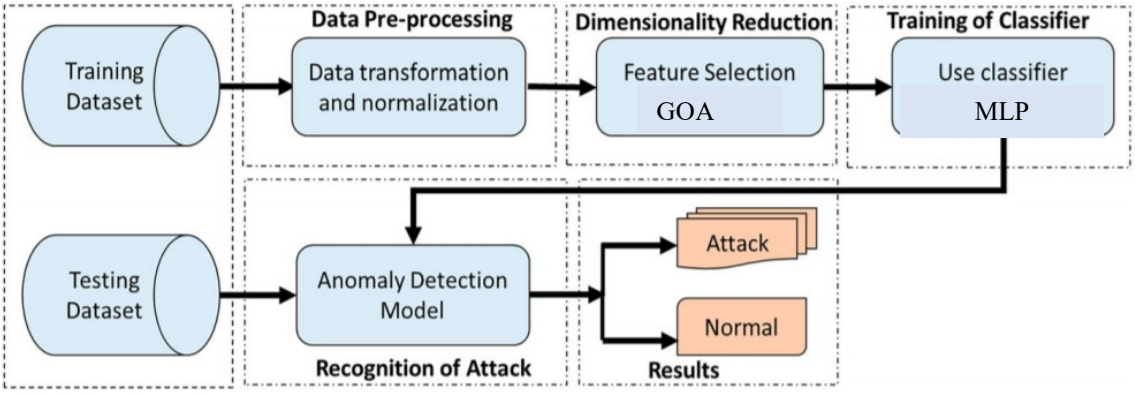
در مطالعه [۱۸]، عملکرد چهار الگوریتم طبقه بندی شناخته شده، ماشین بردار پشتیبان، بیز نامشخص، درخت تصمیم و جنگل تصادفی، با استفاده از اسپارک آپاچی، ابزار پردازش داده‌های بزرگ را برای تشخیص نفوذ به ترافیک شبکه ارزیابی کرد. در این تحقیق عملکرد کلی با توجه به دقت تشخیص، زمان ساخت و زمان پیش بینی شده مورد ارزیابی قرار گرفته است. نتایج تجربی بر روی NB۱۵-UNSW، در یک مجموعه داده کلی برای سیستم تشخیص نفوذ شبکه، یک مزیت مهم برای طبقه‌بندی جنگل تصادفی در میان سایر طبقه‌بندی‌های شناخته شده از نظر دقت تشخیص و زمان پیش‌بینی، با استفاده از مجموعه کاملی از ویژگی‌ها با تمام ویژگی‌ها نشان می‌دهد.

برای شناسایی حملات اینترنت اشیا، کریشنا و همکاران [۱۹] یک روش بهینه سازی ترکیبی با استفاده از الگوریتم بهینه سازی متاهیوریستیک شیر و الگوریتم بهینه سازی کرم شب تاب (ML-F) پیشنهاد کردند. روش پیشنهادی آنها از مجموعه داده‌های NSL-KDD و NBaIoT استفاده می‌کند. مجموعه داده شامل چهار نوع مختلف حمله است Denial of Service (DoS)، User to Root (U۲R)، Probing وRemote to Local (R۲L) در روش پیشنهادی آنها، داده‌های ورودی برای حذف نویزهای ناخواسته و داده‌های از دست رفته و ناقص پیش پردازش می‌شوند. برای این کار از ۴ نوع روش پیش پردازش استفاده می‌شود که عبارتند از: پاکسازی داده‌ها، نرمال سازی، تبدیل داده‌ها و یکپارچه سازی داده‌ها. روش ترکیبی پیشنهادی ML-F نسبت به روش طبقه‌بندی گرادیان موجود، عملکرد بالاتری در طبقه‌بندی حملات دارد.

۴. روش پیشنهادی

حملات شبکه یکی از مواردی است که ممکن است سرورها را از دسترس خارج کند. اکثر حملات شامل ارسال بسته‌های زیادی به یک یا چند میزبان در شبکه است. سوئیچ نمی‌تواند آدرس مقصد بسته‌های دریافتی را پیدا کند، اگر آدرس مقصد جعلی باشد، همانطور که اغلب اتفاق می افتد، و باید بسته را به منبع ارسال کند. جمع آوری و جداسازی بسته‌های واقعی و جعلی می‌تواند در منبع بسیار مشغول باشد. این باعث می‌شود مبدا برای بسته‌های واقعی جدید در دسترس نباشد و ممکن است مبدا را بیش از حد بارگذاری کند تا شبکه به درستی کار نکند. حتی اگر یک سیستم پشتیبان وجود داشته باشد،

ممکن است با همین چالش روبرو شود. در همین حال، برای دفاع موثر در برابر حملات سایبری، سیستم‌های تشخیص نفوذ بسیار کارآمد هستند.



شکل ۱- چارچوب روش پیشنهادی

در این مقاله رویکردی برای شناسایی حملات DDoS ارائه شده است که سعی می‌کند با استفاده از رویکردی مبتنی بر انتخاب ویژگی GOA، جلساتی را که بخشی از حملات DDoS هستند شناسایی کند. نوآوری روش پیشنهادی در استفاده از MO-GOA برای انتخاب و کاهش ویژگی و استخراج ویژگی براساس RFE است.

رویکرد پیشنهادی برای شناسایی حملات DDoS شامل سه مرحله اساسی است. در مرحله اول مجموعه داده‌های دریافتی (UNSW_NB۱۵ و KDD) که شامل ویژگی های مختلف (۴۱ ویژگی) و ویژگی های آنها می باشد، سپس این داده‌ها پیش پردازش شده و در مرحله دوم با استفاده از الگوریتم MOGOA ارزیابی شده و ویژگی های اضافی آن حذف می‌شود. در مرحله آخر، مجموعه داده بهینه شده با استفاده از شبکه عصبی پرسپترون چند لایه استفاده می‌شود و جلسات را به دو مجموعه سالم و حمله طبقه بندی می‌کند. در ادامه هر یک از این مراحل بررسی خواهد شد.

برای اجرای ایده پیشنهادی، اطلاعات جلسه دستگاه اینترنت اشیا باید بررسی شود تا از حمله یا ایمن بودن آن اطمینان حاصل شود. هر گره اینترنت اشیا اطلاعات جلسه خود را برای مدت زمان مشخصی در این فایل ذخیره می‌کند. هر جلسه در شبکه اینترنت اشیا دارای مجموعه ای از اطلاعات است، از جمله اینکه چه پروتکلی در جلسه استفاده شده است، چه پرچم‌هایی فعال هستند، چند بایت اطلاعات ارسال و دریافت شده، یا مدت زمان هر جلسه و غیره. و همه این مواردی که در یک جلسه وجود دارد ویژگی آن جلسه نامیده می‌شود. اکنون باید سیستمی طراحی شود تا این اطلاعات را دریافت کند و تشخیص دهد که آیا حمله عادی است یا خیر. داده‌های ناقص در مجموعه داده‌ها همیشه اجتناب ناپذیر است. این می‌تواند به دلیل خطای انسانی، خرابی سیستم، مشکلات انتقال داده یا مسائل دیگر باشد. اما داده‌های ناقص می‌تواند عواقب بدی برای سیستم داشته باشد. این مشکل زمانی حادتر می‌شود که سیستم مورد استفاده مبتنی بر یادگیری ماشین باشد. زیرا در چنین سیستم‌هایی صحت اطلاعات فرض شده و هرگونه نقص در داده‌های دریافتی می‌تواند منجر به خروجی نادرست سیستم شود.

در مرحله اول (پیش پردازش داده ها)، داده‌های ورودی از یک مجموعه داده معین برای تجزیه و تحلیل تجربی گرفته می‌شود. سپس داده‌های ورودی از قبل گرفته می‌شوند تا داده‌های ناقص و داده‌های از دست رفته حذف شوند. در روش پیشنهادی، برای پیش پردازش داده‌های ورودی، روش‌هایی مانند تمیز کردن داده‌ها، نرمال سازی، تبدیل، یکپارچه سازی داده‌ها انجام می شود.

پاکسازی داده‌ها فرآیند تهیه داده‌ها و تصحیح داده‌هایی است که ناقص، نادرست، نامربوط، کپی شده یا نادرست تنظیم شده اند. داده‌های مربوط به داده‌ها ضروری نیست، تجزیه و تحلیل می‌شود زیرا از نتایج نادرست جلوگیری می‌کند. پاک کردن داده‌ها به سادگی دقت را پاک نمی‌کند بلکه اطلاعات را نیز حذف می‌کند. پاکسازی داده‌ها شامل حذف داده‌ها، تغییر داده‌های نادرست، حذف اطلاعات ناخواسته بدون حذف داده‌های مهم است. هدف اصلی پاک کردن داده‌ها در مجموعه داده‌ای بود که تجزیه و تحلیل داده‌ها را استاندارد می‌کرد و برای یافتن داده‌های مناسب برای پرس و جوها به راحتی در دسترس بود.

از آنجایی که داده‌های ناقص یا نامشخص وجود داشت، برای بهبود کیفیت، داده‌های از دست رفته باید با حذف داده‌های ناخواسته تصحیح شوند. فرآیند عادی سازی Min-Max نقش مهمی‌در یکپارچه سازی و همچنین عادی سازی داده‌ها ایفا می‌کند. هر مقدار مشخصه ای که دارای حداقل مقدار باشد به ۰ و حداکثر مقدار به ۱ تبدیل می‌شود. تمام مقادیر اعشار بین ۰ و ۱ به معادلات تبدیل می‌شوند. هنگامی که حداکثر نرمال سازی برای داده‌های بدون ساختار انجام می‌شود، به دلیل داده‌های ترافیکی آلوده، عدم اطمینان در داده‌ها وجود خواهد داشت. بنابراین، استخراج چنین ویژگی هایی از ساختارهای پیچیده مختلف به تعیین زمینه داده‌ها کمک می‌کند.

در مرحله پاکسازی داده ها، ویژگی هایی که حاوی مقادیر null هستند و هیچ تاثیری ندارند حذف می‌شوند. پس از پیش پردازش داده‌های ورودی، مقادیر مشخصه به طور خودکار به فرآیند انتخاب ویژگی کمک می‌کند، که به بهبود دقت کمک می‌کند. مقادیر مشخصه انتخاب نشده که غیرضروری، اضافی یا نامربوط هستند برای طبقه بندی حملات مفید نیستند. بنابراین، از تکنیک‌های انتخاب ویژگی برای انتخاب ویژگی های برجسته برای تعیین دقت فضای جستجو استفاده می‌شود.

گام بعدی طبقه بندی توسط MLP می باشد. هنگامی‌که ویژگی‌های اضافی داده‌ها حذف شدند، در دسترس شبکه عصبی قرار خواهند گرفت. شبکه‌های عصبی از جمله الگوریتم‌های یادگیری ماشینی هستند که با دریافت داده‌های ورودی و خروجی سعی در پیش‌بینی فرآیند تبدیل ورودی به داده‌های خروجی با استفاده از رگرسیون دارند. در این حالت با داده‌های ورودی جدید می‌توان خروجی را پیش‌بینی کرد. یک شبکه براساس پیش‌بینی حجم ترافیک در زمان واقعی مدیریت می‌شود. وابستگی‌های بلند مدت نقش حیاتی در پیش‌بینی دقیق و کارآمد دارند. MLP اطلاعات زمانی را از جریان ترافیک برای تجزیه و تحلیل داده‌های رفتار شبکه استخراج می‌کند و پیش‌بینی می‌کند که یک برنامه مخرب است یا خیر. این مشکل در پیش‌بینی ترافیک بلادرنگ را افزایش می‌دهد که منجر به مشکل دقت پایین می‌شود.

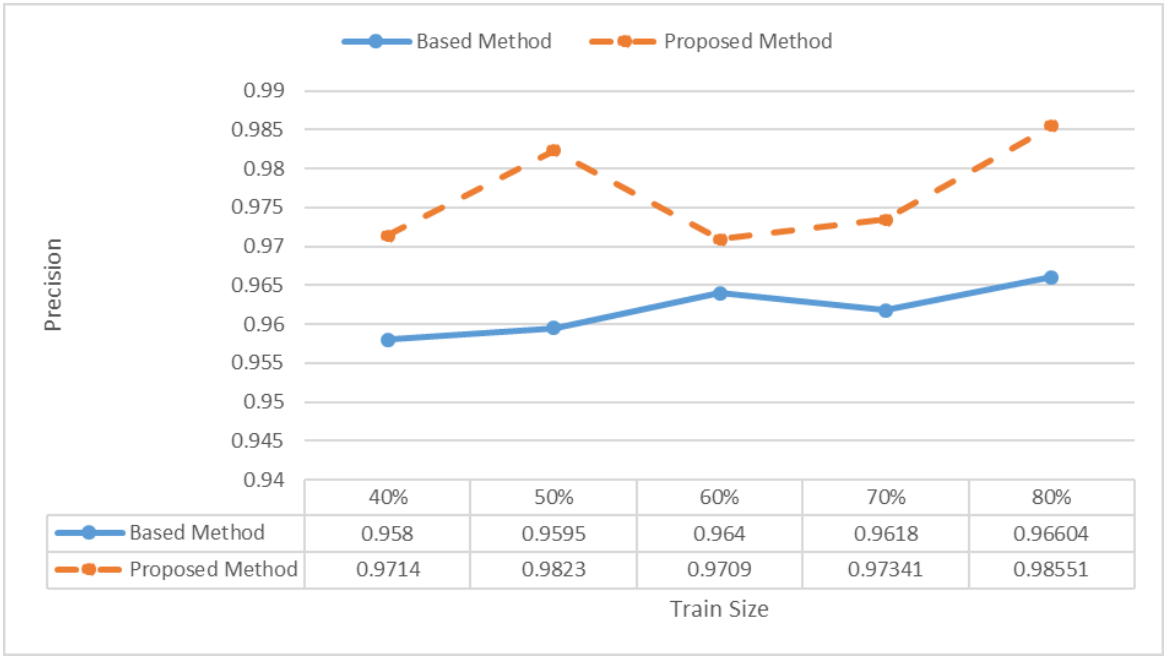
در مرحله آخر کیفیت شبکه عصبی MLP آموزش داده شده را ارزیابی می‌کنیم. برای این کار از مجموعه داده‌های آزمایش استفاده می‌شود. در این فرآیند برای هر نمونه داده از بخش تست، ویژگی های ورودی جدا شده و در اختیار شبکه عصبی MLP قرار می گیرد. خروجی شبکه عصبی یک پیش بینی مقدار برای ویژگی هدف است. از سوی دیگر، مقدار ویژگی هدف برای نمونه داده وارد شده به شبکه عصبی MLP در مجموعه داده موجود است. مقایسه این دو مقدار می‌تواند کیفیت شبکه آموزش دیده را مشخص کند. به طوری که اگر مقدار پیش بینی شده با مقدار واقعی یکسان باشد، ۱ نقطه مثبت برای شبکه عصبی در نظر گرفته می‌شود. اما اگر مقدار پیش بینی شده با مقدار واقعی متفاوت باشد، نمره منفی به شبکه داده می‌شود. این فرآیند برای تمامی نمونه داده‌های بخش تست انجام می‌شود و خروجی این بخش در قالب پارامترهای ارزیابی سیستم پیشنهادی ظاهر می‌شود.

۵. ارزیابی نتایج

سیستم‌هایی که برای شناسایی حملات در شبکه‌های کامپیوتری طراحی شده اند، از حساس ترین بخش‌های یک سیستم امنیتی شبکه هستند. این سیستم‌ها باید قبل از استفاده در شبکه به خوبی آزمایش شوند. بنابراین سیستم‌هایی که در

کارهای تحقیقاتی ارائه می‌شوند باید دارای تست‌های لازم جهت اخذ تاییدیه استفاده در شبکه کامپیوتری باشند. ما یک سیستم برای شناسایی حملات DDOS ارائه کردیم. در این بخش، ابتدا به بررسی شرایط انجام آزمون ها، مجموعه داده‌های مورد استفاده، معیارهای ارزیابی و در نهایت نتایج ارزیابی می پردازیم. برای ارزیابی روش‌هایی که ماهیت طبقه بندی دارند، اغلب از چهار معیار استفاده می‌شود: مثبت واقعیTP، مثبت کاذب FP، منفی واقعی TN و منفی کاذب FN. مجموعه داده NSL-KDD رایج ترین مجموعه داده مورد استفاده در محیط اینترنت اشیا است. مجموعه داده NSL-KDD از بخش‌های مختلف مجموعه داده اصلی KDD Cup ۹۹، بدون اضافه و تکرار تشکیل شده است و شامل ۴۱ ویژگی است که به عنوان اتصالات معمولی یا انواع حمله برچسب گذاری شده است. NSL-KDD یک مجموعه داده است که برای حل برخی از مشکلات ذاتی مجموعه داده KDD'۹۹ که ذکر شده است، پیشنهاد شده است. تعداد رکوردها در NSL-KDD و مجموعه‌های تست معقول است. این مزیت، اجرای آزمایش‌ها را در مجموعه کامل بدون نیاز به انتخاب تصادفی بخش کوچکی مقرون به صرفه می‌کند. در نتیجه، نتایج ارزیابی کارهای تحقیقاتی مختلف سازگار و قابل مقایسه خواهد بود. NSL-KDD شامل چهار حمله مانند DoS، U۲R، R۲L و حمله Probe است.

اولین معیار مورد ارزیابی شامل Precision می باشد. نتایج بررسی این معیار به ازای مقادیر مختلف اندازه مجموعه داده آموزش را میتوان در شکل ۲ مشاهده کرد.

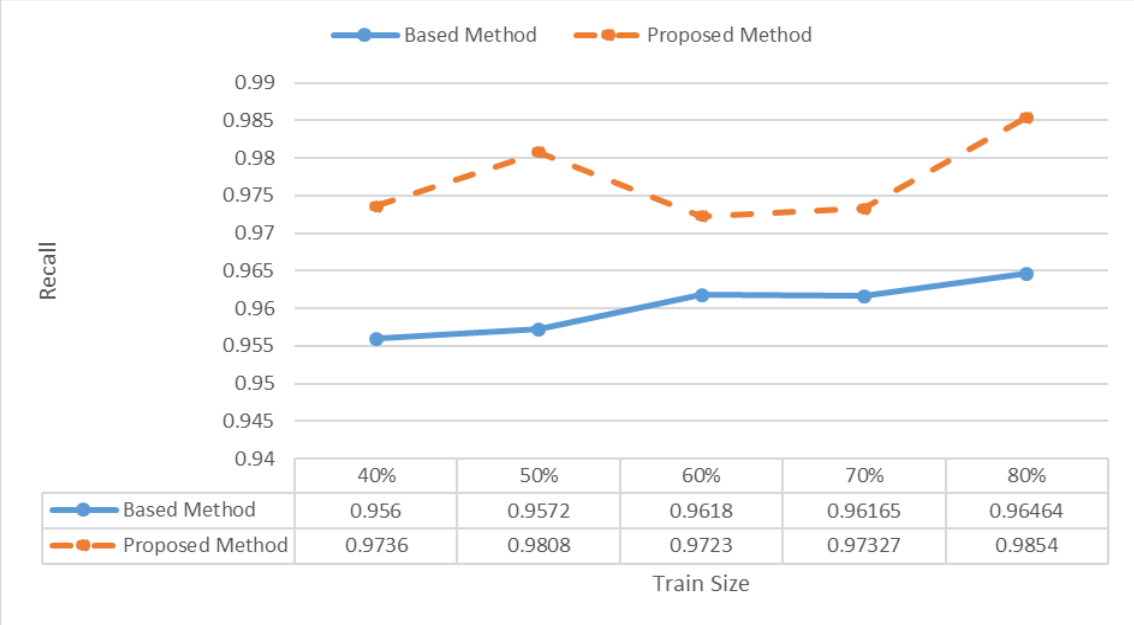


	40%	50%	60%	70%	80%
Based Method	0.958	0.9595	0.964	0.9618	0.96604
Proposed Method	0.9714	0.9823	0.9709	0.97341	0.98551

شکل۲ - مقادیر Precision به ازای روش‌های مورد مقایسه

همانطور که در نتایج این شکل مشخص است روش پیشنهادی توانسته است به ازای تمامی بازه مورد سنجش نسبت به روش مورد مقایسه نتایج بهتری را ارائه کند. به نحوی که تغییر در اندازه مجموعه داده آموزش نیز نتوانسته است بر برتری روش پیشنهادی نسبت به روش مورد مقایسه تاثیر گذار باشد.

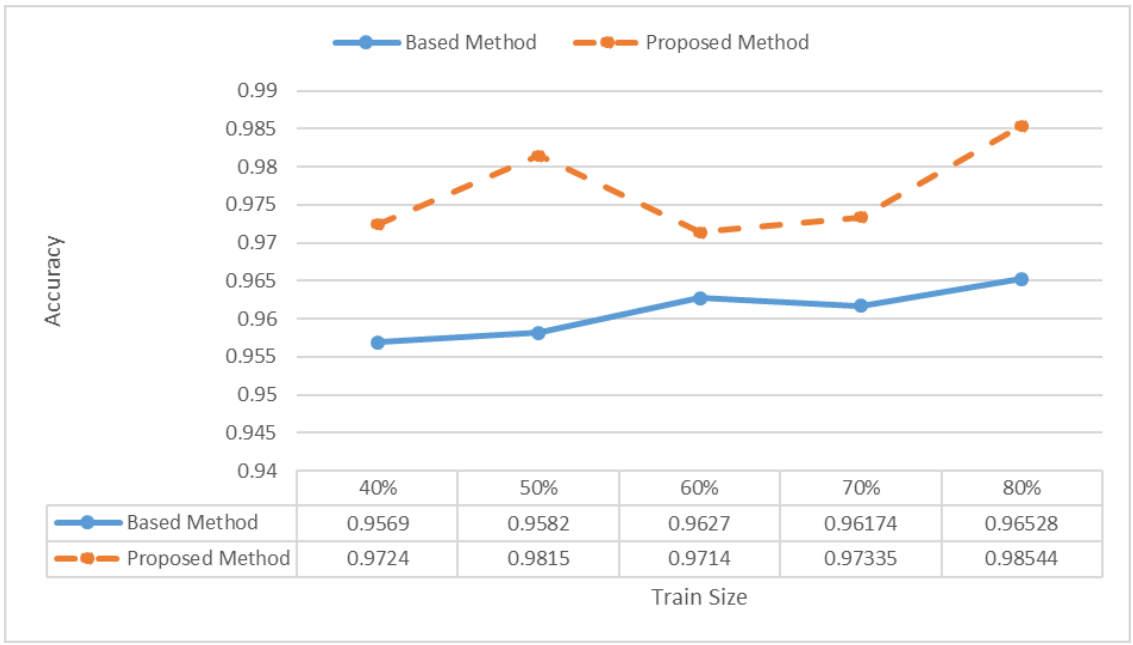
دومین معیار مورد ارزیابی شامل Recall است. نتایج بررسی این معیار در بازه مورد سنجش به ازای تغییر در مجموعه داده آموزش در شکل ۳ قابل مشاهده است.



شکل ۳ مقادیر Recall به ازای روش‌های مورد مقایسه

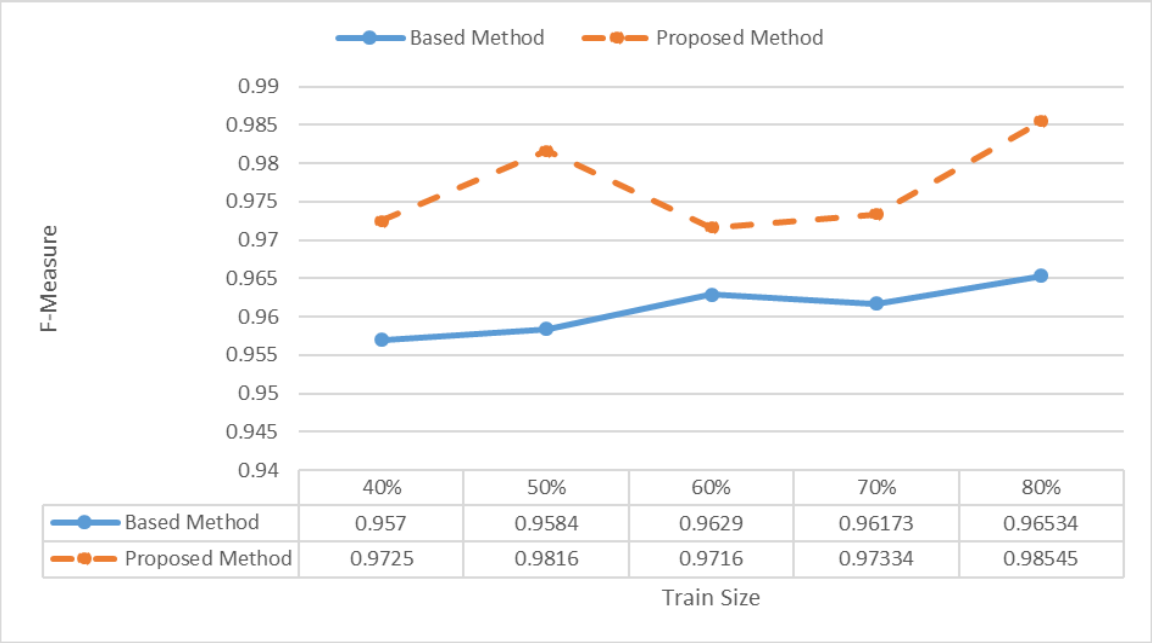
همانطور که در نتایج این شکل نیز مشخص است روش پیشنهادی همواره توانسته است نتایج بهتری را نسبت به روش مورد مقایسه از خود نشان دهد .این موضوع در تمامی بازه مورد سنجش صادق است.

سومین معیار مورد ارزیابی Accuracy است. نتایج بررسی این معیار که نشان دهنده میزان تصمیمات درست گرفته شده توسط سیستمهای مورد مقایسه نسبت به تمامی‌تصمیمات اتخاذ شده می باشد در شکل ۴ نشان داده شده است.



شکل ۴- مقادیر Accuracy به ازای روش‌های مورد مقایسه

بررسی این معیار نیز نشان می‌دهد که روش پیشنهادی توانسته است همانند دو معیار قبلی نتایج بهتری را به دست آورد. آخرین معیار مورد ارزیابی F-Measure می باشد. نتایج بررسی این معیار در شکل ۵ قابل مشاهده است.



	40%	50%	60%	70%	80%
Based Method	0.957	0.9584	0.9629	0.96173	0.96534
Proposed Method	0.9725	0.9816	0.9716	0.97334	0.98545

شکل ۵- مقادیر F-Measure به ازای روش‌های مورد مقایسه

در نمودار های شکل‌های ۲ و ۳ برتری با روش پیشنهادی می باشد و میتوان متوجه شد که برتری روش پیشنهادی نسبت به روش پایه براساس نتایج معیارهای Precision و Recall بوده است.

بررسی نتایج بدست آمده نشان می‌دهد که روش پیشنهادی به ازای تغییر در اندازه مجموعه داده آموزش و به ازای تمامی معیارهای مورد بررسی توانسته است نتایج بهتری را نسبت به روشهای موجود به دست آورده است و می‌توان به این نکته پی برد که رویکرد پیشنهادی با تکیه بر الگوریتم تکاملی ملخ که یک الگوریتم ارائه شده جدید و دارای قدرت اکتشاف بالا است توانسته است در انتخاب ویژگی های مفید برای طبقه بندی داده‌ها بهتر عمل کند. انتخاب ویژگی های مفیدتری از داده‌ها باعث شده است که کیفیت مجموعه داده بهبود پیدا کرده و این مسئله منجر به ارائه یک مجموعه داده با پیچیدگی کمتر و در عین حال ویژگیهای قوی تری باشد. در طرف مقابل نیز با مقایسه طبقه بندی مورد استفاده در روش پیشنهادی ثابت می‌شود که استفاده از شبکه‌های عصبی برای طبقه بندی مجموعه داده‌های پیچیده کارایی بالاتری نسبت به طبقه بندهای غیرخطی دارد. در واقع شبکه عصبی ارائه شده در رویکرد پیشنهادی توانسته است با طبقه بندی مناسب که بر روی داده‌ها انجام می‌دهد نشست‌های سالم و حمله را را بهتر طبقه بندی کند. شبکه‌های عصبی برای به کار گرفته شدن در مجموعه داده‌های پیچیده توانایی بالاتری دارد. از این رو انتخاب ما برای طبقه بندی شبکه عصبی بوده است.

۶. نتیجه گیری و کارهای آینده

تشخیص حملاتDDOS به عنوان یکی از سیستم‌های حیاتی برای شبکه‌های کامپیوتری همواره مورد توجه محققین زیادی قرار گرفته است. علیرغم آنکه برای پیشگیری از این نوع حملات تاکنون راه حل‌های متعددی ارائه شده است اما راه

حل قطعی برای آن ارائه نشده است. شبکه‌های عصبی به عنوان بخشی از الگوریتم‌های طبقه بندی تاکنون راه حل‌های مناسبی را برای این موضوع تحقیقاتی ارائه کرده اند. اما وجود داده‌های با کیفیت پایین یا پیچیده همواره یکی از چالشهای اساسی در این موضوع تحقیقاتی بوده است. ما در این مقاله به ارائه رویکردی به منظور تشخیص حملات DDOS با استفاده از الگوریتم تکاملی ملخ پرداخته ایم. در رویکرد پیشنهادی ابتدا مجموعه ای از داده‌ها به سیستم پیشنهادی وارد می‌شوند سپس سیستم پیشنهادی با استفاده از الگوریتم تکاملی ملخ به پالایش داده‌ها و حذف ویژگی های غیر ضروری می پردازد. با حذف ویژگی‌های غیرضروری علاوه بر کاهش پیچیدگی داده‌ها می‌توان به افزایش کیفیت داده‌ها نیز کمک کرد. در گام دوم از روش پیشنهادی مجموعه داده‌ها به دو قسمت آموزش و آزمایش تقسیم می‌شوند. قسمت آموزش برای یادگیری یک طبقه بند مورد استفاده قرار می گیرد. اما از قسمت آزمایش برای صحت سنجی عملکرد طبقه بند آموزش دیده استفاده خواهد شد. سپس داده‌های قسمت آموزش وارد یک شبکه عصبی پرسپترون چند لایه می‌شود. این شبکه عصبی با استفاده از داده‌های دریافتی اقدام به یادگیری داده‌های حمله و داده‌های سالم می‌کند. به نحوی که بتواند با استفاده از داده‌های نشست‌های مختلف سالم یا حمله بودن نشست را تشخیص دهد. سپس مجموعه داده آزمایش برای صحت سنجی عملکرد روش پیشنهادی به سیستم پیشنهادی وارد می‌شود. در این مرحله از سیستم پیشنهادی خواسته می‌شود تا با توجه به نشست‌های داخل مجموعه داده آزمایش نوع نشست‌ها را پیش بینی کند. در آخرین مرحله با استفاده از معیارهای ارزیابی به تعیین کیفیت روش پیشنهادی پرداختیم

به منظور ارزیابی رویکرد پیشنهادی از مجموعه داده بنچ کار NSL-KDD استفاده شده است. در این آزمایشات تلاش شده است تا پارامترهای مجموعه داده نیز مدنظر قرار بگیرد. برای این منظور ارزیابی ها براساس اندازه‌های مختلف مجموعه داده آموزش و آزمایش انجام شده است. نتایج ارزیابی هایی که در قالب معیارهای Precision, Recall, Accuracy و F-Measure ارائه شده است نشان می‌دهد که روش پیشنهادی نسبت به روش مورد مقایسه توانسته است نتایج به مراتب بهتری را ارائه کند.

با توجه به اینکه روز به روز بر تعداد حملات به شبکه‌های کامپیوتری و همچنین پیچیدگی این حملات افزوده می‌شود موضوع تشخیص حملات همچنان به عنوان یک موضوع تحقیقاتی باز مدنظر است. از این رو ما به محققین علاقه‌مند در این حوزه پیشنهاد می‌کنیم که با در نظر گرفتن چارچوب این تحقیق به ارائه رویکردهایی برای تشخیص این حملات بپردازند. چهارچوب پیشنهاد شده در این تحقیق شامل یک مرحله پایش داده‌ها و مرحله دوم طبقه بندی داده‌ها میباشد. از اینرو محققین علاقه‌مند می‌توانند با تغییراتی در سیستم انتخاب ویژگی این کار تحقیقاتی و همچنین ترکیب آن با طبقه بندهای مختلف به ارائه رویکردهای جدیدی بپردازند.

۷. فهرست مراجع

[۱] Gaikwad, D. P., & Thool, R. C. (۲۰۱۵). Intrusion detection system using bagging with partial decision treebase classifier. Procedia Computer Science, ٤٩, ٩٢-٩٨

[۲] Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (۲۰۱۸). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, ٢٥, ١٥٢-١٦٠.

[۳] Sh.Saremi a,b , S.ali.Mirjalili a,b , A.Lewis, "Grasshopper Optimisation Algorithm: Theory and application", ۲۰۱۷ Elsevier.

[٤] ggarwal, C. C. (۲۰۱۵). Data classification. In Data Mining (pp. ٢٨٥-٣٤٤). Springer,

[٥] M. Guclu, C. Bakir, V. Hakkoymaz, and B. Diri, "Comparisons on intrusion detection and prevention systems in distributed databases," Balkan Journal of Electrical and Computer Engineering, vol. ٧, pp. ٤٤٦–٤٥٥, ٢٠١٩

[٦] M. H. Ali, B. A. D. Al Mohammed, A. Ismail, and M. F. Zolkipli, ''A new intrusion detection system based on fast learning network and particle swarm optimization,'' IEEE Access, vol. ٦, pp. ٢٠٢٥٥–٢٠٢٦١, ٢٠١٨.

[٧] D.P. Gaikward, Ravindra c Thool, Intrusion detection system using bagging with partial decision tree base classifier, in: Proceeding of International Conference on Advanced in Computing, Communication and Control, ICAC٣(١٥, in: Procedia Computer Science, vol. ٤٩, Elsevier, ٢٠١٥, pp. ٩٢–٩٨

[٨] S. Ganapathy, N. Jaisankar, P. Yogesh, and A. Kannan, ''An intelligent system for intrusion detection using outlier detection,'' in Proc. Int. Conf. Recent Trends Inf. Technol. (ICRTIT), Jun. ٢٠١١, pp. ١١٩–١٢٣.

[٩] P. Jongsuebsuk, N. Wattanapongsakorn, and C. Charnsripinyo, ''Network intrusion detection with fuzzy genetic algorithm for unknown attacks,'' in Proc. Int. Conf. Inf. Netw. (ICOIN), Jan. ٢٠١٣, pp. ١–٥.

[١٠] S. Ganapathy, K. Kulothungan, S. Muthurajkumar, M. Vijayalakshmi, P. Yogesh, and A. Kannan, ''Intelligent feature selection and classification techniques for intrusion detection in networks: A survey,'' EURASIP J. Wireless Commun. Netw., vol. ٢٠١٣, no. ١, p. ٢٧١, Dec. ٢٠١٣.

[١١] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, ''Building an intrusion detection system using a filter-based feature selection algorithm,'' IEEE Trans. Comput., vol. ٦٥, no. ١٠, pp. ٢٩٨٦–٢٩٩٨, Oct. ٢٠١٦.

[١٢] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, ''Multi-classification of UNSW-NB١٥ dataset for network anomaly detection system,'' J. Theor. Appl. Inf. Technol., vol. ٩٦, no. ١٥, pp. ٥٠٩٤–٥١٠٤, Aug. ٢٠١٨.

[١٣] P. Nancy, S. Muthurajkumar, S. Ganapathy, S. V. N. Santhosh Kumar, M. Selvi, and K. Arputharaj, ''Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks,'' IET Commun., vol. ١٤, □□. ٥, □□. ٨٨٨–٨٩٥, □□□.

[١٤] Elhefnawy, R., Abounaser, H., & Badr, A. (۲۰۲۰). A Hybrid Nested Genetic-Fuzzy Algorithm Framework for Intrusion Detection and Attacks. IEEE Access

[۱۵] Moghanian, Shadi, et al. "GOAMLP: Network intrusion detection with multilayer perceptron and grasshopper optimization algorithm." IEEE Access ۸ (۲۰۲۰): ۲۱۵۲۰۲-۲۱۵۲۱۳

[۱۶] Wang, C. R., Xu, R. F., Lee, S. J., & Lee, C. H. (۲۰۱۸). Network intrusion detection using equality constrained-optimization-based extreme learning machines. Knowledge-Based Systems, ۱٤۷, ٦۸-۸۰.

[۱۷] Roshan, S., Miche, Y., Akusok, A., & Lendasse, A. (۲۰۱۸). Adaptive and online network intrusion detection system using clustering and extreme learning machines. Journal of the Franklin Institute, ۳٥٥(٤), ۱۷٥۲-۱۷۷۹.

[۱۸] Belouch, M., El Hadaj, S., & Idhammad, M. (۲۰۱۸). Performance evaluation of intrusion detection based on machine learning using Apache Spark. Procedia Computer Science, ۱۲۷, ۱-٦.

[۱۹] Krishna, E. P., & Thangavelu, A. (۲۰۲۱). Attack detection in IoT devices using hybrid metaheuristic lion optimization algorithm and firefly optimization algorithm. International Journal of System Assurance Engineering and Management, ۱-۱۴.