مقایسه سیستم تشخیص نفوذ مبتنی‌بر الگوریتم کرم شب تاب با SVM و الگوریتم تکامل تفاضلی

محمد بشارتلو[۱]، عطیه رحیمی زاده[۲]،محمد گرجی[۳]

گروه مهندسی کامپیوتر،موسسه آموزش عالی علوم و فناوری آریان،moohaamaad.beshaaratloo68@gmail.com

چکیده: از آنجاییکه داده مورد استفاده در سیستم تشخیص نفود حجم بالایی دارد، یکی از مسائل ضروری در این سیستم ها حفظ ویژگی‌های با بهترین کیفیت در مجموعه داده می‌باشد بطوریکه این ویژگی‌های منتخب قادر باشند، ساختار مجموعه داده‌ها را به درستی نشان دهند. بنابراین ضرورت خواهد داشت تا ویژگی‌های زائد و نامرتبط از مجموعه داده‌ها حذف و بهترین زیرمجموعه ویژگی از بین مجموع ویژگی‌ها تعیین گردد. در این مقاله، الگوریتم کرم شب تاب به عنوان استراتژی جستجو برای انتخاب بهترین زیرمجموعه ویژگی‌ها و طبقه‌بند درخت تصمیم و svm برای تعیین کیفیت ویژگی‌های انتخاب شده مورد استفاده قرار می‌گیرد. مجموعه داده KDD Cup 99 برای ارزیابی روش پیشنهادی بکار برده می‌شود. نتایج شبیه‌سازی‌های انجام شده نشان می دهد که زیرمجموعه ویژگی بدست آمده توسط الگوریتم پیشنهادی به نرخ تشخیص و دقت بالاتر و نرخ هشدار نادرست پایین‌تر در مقایسه با نتایج بدست آمده با استفاده از الگوریتم کرم شب تاب و استفاده از تمام ویژگی‌ها دست می‌یابد.

کلمات کلیدی: سیستم تشخیص نفود، انتخاب ویژگی، الگوریتم کرم شب تاب، درخت تصمیم،svm

۱. مقدمه

در دنیای امروز، کامپیوتر و شبکه‌های کامپیوتری متصل به اینترنت نقش عمده‌ای در ارتباطات و انتقال اطلاعات ایفا می‌کنند. در این بین افراد سودجو با دسترسی به اطلاعات مهم مراکز خاص یا اطلاعات افراد دیگر و با قصد اعمال نفوذ یا اعمال فشار و یا حتی به هم ریختن نظم سیستم‌ها، عمل تجاوز به سیستم‌های کامپیوتری را در پیش گرفته‌اند. **Hacker**، **Cracker** و **Intruder** کلماتی هستند که امروزه کم و بیش در محافل کامپیوتری مطرح می‌باشند و اقدام به نفوذ به سیستم‌های دیگر کرده و امنیت آن‌ها را به خطر می‌اندازد. بنابراین لزوم حفظ امنیت اطلاعاتی و حفظ کارآیی در شبکه‌های کامپیوتری که با دنیای خارج ارتباط دارند، کاملاً محسوس است. از آنجا که

* **Corresponding author**
Email: moohaamaad.beshaaratloo68@gmail.com

از نظر تکنیکی ایجاد سیستم‌های کامپیوتری (سخت‌افزار و نرم‌افزار) بدون نقاط ضعف و شکست امنیتی عملاً غیرممکن است، تشخیص نفوذ در تحقیقات سیستم‌های کامپیوتری با اهمیت خاصی دنبال می‌شود. سیستم‌های تشخیص نفوذ (**IDS**)برای کمک به مدیران امنیتی سیستم در جهت کشف نفوذ و حمله به کار گرفته شده‌اند. هدف یک سیستم تشخیص نفوذ جلوگیری از حمله نیست و تنها کشف و احتمالاً شناسایی حملات و تشخیص اشکالات امنیتی در سیستم یا شبکه‌ی کامپیوتری و اعلام آن به مدیر سیستم است. عموماً سیستم‌های تشخیص نفوذ در کنار دیواره‌های آتش و به صورت مکمل امنیتی برای آن‌ها مورد استفاده قرار می‌گیرند.

از آنجا که ساخت بخش‌های نرم‌افزاری و سخت‌افزاری در سیستم‌های کامپیوتریبدون ضعف و شکست امنیتی در عمل غیرممکن است، مسئله تشخیص نفوذ در این سیستم‌ها از اهمیت خاصی برخوردار است. سیستم‌های تشخیص نفوذ (IDS[1]) برای کمک به مدیران شبکه در جهت شناسایی نفوذ و بخورد با آن به کار گرفته شده‌اند. باید توجه داشت که هدف از یک سیستم تشخیص نفوذ، جلوگیری از حمله نمی‌باشد بلکه کشف و شناسایی حملات و تشخیص اشکالات امنیتی در آن سیستم است. از سیستم‌های تشخیص نفوذ به همراه دیواره آتش به صورت مکمل امنیتی استفاده می‌شود .

سیستم تشخیص نفوذ برای بسیاری از سازمان‌ها و مراکز دولتی و خصوصی ضرورت است. برخی از فواید این سیستم‌ها عبارتند از [۳]: کارایی بالا در مقایسه با سیستم‌های دستی در زمینه شناسایی نفوذ ، منبع دانش جامع از انواع حمله، توانایی بررسی حجم بالای داده، توانایی هشدار تقریبا بلادرنگ با هدف کاهش خسارت ، پاسخ خودکار مانند قطع ارتباط کاربر، غیر فعال سازی حساب کاربر، اعمال مجموعه دستورهای خودکار ، افزایش میزان بازدارندگی، توانایی گزارش‌دهی.

سیستم های تشخیص نفوذ در سال‌های اخیر، به طور قابل ملاحظه‌ای به منظور افزایش امنیت شبکه‌های کامپیوتری بکار گرفته شده است. از طرفی، به علت حجم بالا و کیفیت پایین رویدادنامه‌ها[2]، نیاز به پردازش بیشتر آن‌ها وجود خواهد داشت. بنابراین مدیران شبکه ترجیح می‌دهند تا ارتباط این رویدادنامه‌ها را به صورت دستی فراهم نمایند. اما به علت انعطاف پذیر نبودن این روش، زمانبر بودن و نداشتن دید عمومی نسبت به رویدادنامه‌ها، یافتن ارتباط بین اطلاعات اجزاء شبکه و رویدادهای غیر اصلی که می‌توانند یک تهدید به شمار آیند، غیر ممکن خواهد بود. در این راستا، روش‌های همبسته‌سازی رویدادنامه‌ها به منظور افزایش کیفیت هشدار و بررسی دقیق وضعیت امنیتی ارائه شدند [۲].

بالا رفتن مهارت و کامپیوتری شبکه‌های از شبکه‌ها، با افزایش استفاده و سیستم‌ها امروزه، ایمن‌سازی است. این خوردار بر گذشته به نسبت بیشتری اهمیت از نرم افزارها، در آسیب‌پذیر نقاط وجود با کاربران

[1] Intrusion detection system
[2] Log

شبکه در و دسترس‌پذیری[۲۳] جامعیت محرمانگی[۱]، اساسی رکن سه فراهم آوردن معنای به امنیت تأمین امنیتی سیاست‌های از تخطی یا تشخیص و امنیتی پیشگیری یکی از دو صورت، به می‌تواند می‌باشد که روی بر تمرکز و با شبکه‌های کامپیوتری امنیت زمینه در انجام شده تاکنون اکثر تحقیقات پذیرد. انجام اعتبار و[۴] هویت‌شناسی مکانیزم‌های بکارگیری مانند پیشگیرانه روش‌های استفاده از با امنیت تامین است . اما از آنجاییکه صورت پذیرفته آتش[۸]) (دیواره حفاظ و رمزنگاری[۷] سنجی[۵]، کنترل دسترسی[۶]، غیر سیستم‌های روی بر بررسی و تحقیق روش‌های پیشگیرانه پس از اعمال نفوذ کاربردی ندارند، لذا و شبکه‌های سیستم‌ها در حمله‌ها وظیفه تشخیص که تهاجم تشخیص سیستم‌های پیشگیرانه مانند است برخوردار خاصی اهمیت از کامپیوتری را بر عهده دارندیز

، از معیار ارزیابی آنتروپی[۱] جهت تشخیص منظم بودن ترافیک شبکه استفاده]۴[در مدل آماری ارائه شده در معیار آنتروپی که در سال ۱۹۸۴ ارائه شد، نمایانگر میزان عدم قطعیت و تصادفی بودن یک پردازش .شده است اتفاقی است. به علاوه معیار آنتروپی جهت فشرده‌سازی داده با کمترین میزان اتلاف داده می باشد. از این رو ارائه دهندگان این مدل با استفاده از مفهوم فشرده‌سازی آنتروپی یک روال تحلیل موثر بر مبنای فشرده‌سازی دنباله معیارهای شبکه طراحی کردند. آن‌ها مشاهده نمودند که وقوع حمله پویش[۹] در یک بازه زمانی خاص، بر روی آنتروپی کلی میزبان تاثیر می‌گذارد. به طور کلی اگر از یک منبع تعداد زیادی جریان شبکه ارسال شود(منبع پویشگر)، آنتروپی مرتبط با توزیع آدرس‌های مبدا دچار کاهش ناگهانی می‌شود. همچنین در همان زمان آنتروپی مرتبط با توزیع آدرس‌های مقصد دچار افزایش ناگهانی می‌شود چرا که یک منبع مشغول برقراری ارتباط با تعداد زیادی مقصد است. استفاده از این تغییرات آنتروپی، می‌تواند به آشکارسازی حملات در شبکه کمک نماید.

موفق شدند در ابتدا با استفاده از یک ساختار با نام طرح اولیه اجتماع داده‌های جریان]۵[محققان در مرجع یک جدول یک بعدی درهم است که جهت دسترسی سریع به داده‌های شبکه را ذخیره نمایند. طرح اولیه، توسط این جدول می‌توان تعداد رخداد مرتبط با یک رویداد را شمارش کرد. ذخیره شده بسیار مناسب می‌باشد. طرح اولیه این امکان را فراهم می‌کند که یک سری ویژگی‌های آماری در مورد تغییرات ترافیک در طول زمان به دست آید. سپس یک سیستم شناسایی مبتنی بر ناهنجاری بر مبنای تغییرات شدید و ناگهانی این ویژگی‌های آماری ذخیره شده در طرح اولیه، نفوذ را شناسایی و واکنش نشان می‌دهد].۵[

نام این نیز یک زیربنای عمومی برای شناسایی حملات محدود کننده سرویس معرفی گردید.]۶[در مرجع نامگذاری شد که وظیفه آن جمع آوری جریان‌های شبکه از منابع و مکان‌های مختلف بود. TOPASسیستم[۱۰]

[1] Entropy
[2] Explore attack

این زیربنا دارای بخش‌های شناسایی زمان حقیقی متفاوتی است که توسط مدیر شبکه پیکربندی می‌شود. بخش‌های مختلف این زیربنا قادر به شناسایی انواع حملات است اما تمرکز اصلی این پژوهش بر شناسایی حملات محدود کننده سرویس و حملات محدود کننده سرویس توزیع شده، می‌باشد.

با یاری جستن از معیارهای تحلیل شبکه اجتماعی،نظریه‌ای با مفهوم "نفوذگر کسی است که فارغ از مرزهای اجتماعی موجود، با اجتماعات مختلف ارتباط برقرار کند" اثبات گردید. در این مرجع با استفاده از معیارهای تحلیل شبکه‌های اجتماعی مانند ضریب همبستگی دسته و مرکزیت میانی،روشی موثر و ساده برای شناسایی نفوذگران ارائه گردید. مدل ارائه شده در این بررسی، به عنوان مکمل سیستم‌های مبتنی بر امضا در تحقیقات مورد استفاده قرار می‌گیرد [۷].

بدون توجه به شماره پورت مدل‌سازی IRC، ترافیک مرتبط با کانال IRC[1] در مدل ارائه شده در مرجع [۸] در پژوهش مذکور دو هدف کلی دنبال شده است. اول اینکه طی یک روند چند مرحله‌ای، فرماندهان حملات شناسایی می‌شوند. در ابتدا با بررسی وقایع ثبت شده مرتبط با عملیات پویش، هرزنامه و ویروس‌های مرتبط با ارتباطات مشکوک شناسایی می‌شوند. این ارتباطات مشکوک که بطور معمول از طریق کانال IRC برقرار می‌شود، می‌تواند میان فرمانده احتمالی و شبکه تحت کنترل او باشد. سپس این الگوی ارتباطی مشکوک با مدل ساخته شده از جریان شبکه مقایسه می‌شود تا از صحت شناسایی مطمئن گردند. دومین هدفی که در پژوهش مورد نظر دنبال شده، این است که پس از شناسایی فرمانده حمله، سیستم بتواند سیستم‌های آلوده تحت کنترل را از لحاظ رفتاری دسته‌بندی کند. به همین منظور یک الگوریتم دسته‌بندی سلسله مراتبی ارائه داده شد که کامپیوترهای آلوده را بر مبنای فعالیت‌های غیرمرتبط دسته‌بندی می‌کند.

در به مقایسه و بررسی الگوریتم‌های داده کاوی شبکه بیزین و بردار پشتیبان برای تشخیص نفوذ به مطالعه داده‌های شد ذکر بالا در که مدل‌هایی از استفاده پرداخته شده است. در این تحقیق محققان سعی کردند که با آزمایش کنند. نتیجه این پژوهش نشان می‌دهدکه بهترین الگوریتم با توجه به مدل روی بر را تست و آموزشی می‌باشد که دارای دقت ۸۳/۲۹ درصد است HNBنوع داده آموزشی، الگوریتم

، پژوهشی تحت عنوان مروری بر سیستم های تشخیص نفوذ با استفاده از ماشین بردار [۱۰] در مرجع پشتیبان ارائه گردید. محققین عنوان کردند که در روش ماشین پشتیبان با روش دسته‌بندی، می‌توان حملات را از یکدیگر متمایز نمود و خطای تعمیم را به حداقل رساند. اما روش ماشین بردار پشتیبان نیز به مدت زمان طولانی برای آموزش نیاز دارد. بنابراین با استفاده از روش‌های ترکیبی حاصل از ترکیب الگوریتم خوشه‌بندی با ماشین بردار پشتیبان می‌توان زمان آموزش را به نسبت قابل توجه کاهش و دقت در شبکه را نیز بهبود بخشید [۱۰].

در [۱۱]،کاربرد الگوریتم ژنتیک در تشخیص نفوذ شبکه‌های کامپیوتری بررسی گردید. در این مقاله به مروری جامع بر روی سیستم تشخیص نفوذ مبتنی بر الگوریتم ژنتیک پرداخته شد. همچنین تکنیک‌های تشخیص نفوذ شامل الگوریتم ژنتیک در دهه اخیر به طور مختصر بررسی گردید

[1] Internet Relay Chat

در این مقاله، یک روش انتخاب ویژگی برای سیستم تشخیص نفوذ پیشنهاد می‌شود تا زیرمجموعه بهینه از ویژگی‌ها را تولید نماید. روش پیشنهادی براساس الگوریتم تکامل کرم شب تاب برای جستجوی زیرمجموعه بهینه ویژگی‌ها است. همچنین، svm به عنوان طبقه‌بند بکار برده می‌شود تا کیفیت زیرمجموعه‌های ویژگی تولید شده را بهبود دهد.

ادامه متن مقاله به این صورت سازماندهی شده است: قسمت ۲ معرفی مختصری از svm و الگوریتم کرم شب تاب را ارائه می‌نماید. روش انتخاب ویژگی پیشنهادی براساس الگوریتم کرم شب تاب و svm در قسمت ۳ توضیح داده می‌شود. نتایج آزمایشها در قسمت ۴ گزارش می‌شود. در نهایت، در قسمت ۵ نتیجه‌گیری و پیشنهادات برای کارهای آتی بیان می شود.

۲. معرفیSVM و الگوریتم کرم شب تاب و الگوریتم تکامل تفاضلی

svm .۱.۲

یکی از قوی ترین طبقه بندی ها در حوزه یادگیری ماشین، ماشین بردار پشتیبان است که در سال ۱۹۹۵ توسط vapnik بر پایه تئوری یادگیری آماری ارائه شد. این روش یکی از روش های یادگیری با نظارت است که برای طبقه بندی و رگرسیون به کار برده می شود. ایده اصلی ماشین بردار پشتیبان ساخت ابر صفحه ای به عنوان سطح تصمیم گیری می باشد به طوری که حاشیه جداسازی بین دو کلاس داده های مثبت و منفی را حداکثر سازد. به طور دقیق تر، ماشین بردار پشتیبان جدا کننده ای است که ریسک جداسازی را با به حداکثر رساندن حاشیه ی بین دو کلاس داده به حداقل می رساند.

۲.۲. الگوریتم کرم شب تاب

الگوریتم کرم شب تاب(FA) در اواخر سال ۲۰۰۷ توسط xin-she yang معرفی شده است که ایده ی اصلی آن از ارتباط نوری میان کرم های شب تاب الهام گرفته شده است. این الگوریتم برای بهینه سازی توابع غیر خطی ناهموار از رفتار گونه های طبیعی الهام می گیرد. این الگوریتم را می توان از مظاهر هوش ازدحامی دانست که در آن از همکاری (و احتمالا رقابت) اعضای ساده و کم هوش، مرتبه ی بالاتری از هوشمندی ایجاد می شود که قطعا توسط هیچ یک از اجزا قابل حصول نیست. یکی از الگوریتم های موفق و در عین حال کم هزینه به شمار می رود.

۳,۲.الگوریتم تکامل تفاضلی

الگوریتم تکاملی تفاضلی ((DEنخستین بار در سال ۱۹۹۵ توسط استورن و پرایس معرفی شد.این دو نشان دادند که این الگوریتم توانایی خوبی در بهینه سازی توابع غیرخطی مشتق ناپذیر دارد که به عنوان روشی قدرتمند و سریع برای مسائل بهینه سازی در فضاهای پیوسته معرفی شده است.

الگوریتم (DEجهت غلبه بر عیب اصلی الگوریتم ژنتیک، یعنی فقدان جستجوی محلی دراین الگوریتم ارائه شده است، تفاوت اصلی بین الگوریتم های ژنتیکی و الگوریتم (DEدر عملگر انتخاب selection operatorsمی باشد.

در اپراتور انتخاب GA،شانس انتخاب یک جواب به عنوان یکی از والدین وابسته به مقدار شایستگی آن می باشد،اما در الگوریتم DEهمه جواب ها دارای شانس مساوی جهت انتخاب شدن می باشند. یعنی شانس انتخاب شدن آنها وابسته به مقدار شایستگی آنها نمی باشد، پس از این که یک جواب جدید با استفاده از یک اپراتور جهش خود-تنظیم و اپراتور crossoverتولید شد،جواب جدید با مقدار قبلی مقایسه می شود و در صورت بهتر بودن جایگزین می گردد.

در این الگوریتم بر خلاف دیگر الگورتیم ها که اول عملگر crossover و سپس عملگر mutation انجام می شود به گونه ای که ابتدا عملگرجهش اعمال شده و سپس عملگر تقاطع اعمال می شود تا بدین وسیله نسل جدید ایجاد گردد.

۳. انتخاب ویژگی براساس الگوریتم کرم شب تاب

گامهای نسخه باینری الگوریتم کرم شبتاب بصورت زیر است:

۱-تولید جمعیت اولیه کرم شبتاب‌ها: جمعیت اولیه کرم شبتاب‌ها را بطور تصادفی با کمک رابطه زیر تولید می نماییم.

$$V_i(j) = round(rand),\ \ j = 1, \dots, d\ and\ i = 1, \dots, N$$

که $rand$ تابعیست که یک عدد تصادفی در بازه $[0,1]$ تولید می نماید و $round$ تابعی است که عدد تصادفی تولید شده توسط تابع $rand$ را به نزدیکترین عدد صحیح (صفر یا یک) گرد می نماید. d تعداد کل ابعاد راه حل است و N تعداد کل کرم شبتابهای موجود در جمعیت است.

۲- محاسبه مقدار شفافیت هر کرم شبتاب: مقدار شفافیت هر کرم شبتاب(راه حل) با تابع هدف آن مرتبط میباشد. در صورتیکه در مسئله ای که قصد داریم راه حل آن را بیابیم به دنبال یافتن راه حلی باشیم که تابع هدف را ماکزیمم نماید بنابراین مقدار شفافیت با مقدار تابع هدف رابطه مستقیم خواهد داشت یعنی مقدار شفافیت کرم شبتاب را برابر با مقدار تابع هدف قرار می دهیم.

۳- برای هر p از ۱ تا تعداد کرم شبتابها (N) به گام ۴ برو

۴- برای هر q از ۱ تا تعداد کرم شبتابها (N) به گام ۵ برو

۵- اگر مقدار شفافیت کرم شبتاب p کمتر از مقدار شفافیت کرم شبتاب q باشد گامهای ۶ و ۷ را اجرا می نماییم

۶- کرم شبتاب p به سمت کرم شبتاب q طبق رابطه زیر حرکت می نماید

$$V'_p = V_p + \beta(r) \times (V_p - V_q) + \alpha \times \left(rand - \frac{1}{2}\right)$$

که مقدار جذب بین دو کرم شبتاب طبق رابطه زیر محاسبه می شود

$$\beta(r) = \beta_0 e^{-\gamma r_{pq}^n}\ ,\quad n \geq 1$$

که جذب توسط تنظیم پارامترهای β_0 و γ می تواند بدست آید. $r_{p,q}$ فاصله بین دو کرم شبتاب p و q است که طبق رابطه زیر محاسبه می شود.

$$r_{pq} = V_p - V_q$$

سپس برای اینکه بتوانیم هر مولفه از کرم شبتاب را با مقادیر صفر یا یک مقداردهی نماییم ابتدا مقدار تابع سیگموید یا *tanh* را برای هر مولفه V'_p به ترتیب طبق یکی از روابط زیر محاسبه می نماییم

$$f(V'_p) = \frac{1}{1+\exp(-V'_p)}$$

$$f(V'_p) = \tanh(|V'_p|) = \frac{\exp(2*|V'_p|)-1}{\exp(2*|V'_p|)+1}$$

سپس موقعیت نهایی کرم شبتاب p را طبق رابطه زیر محاسبه می نماییم

$$V'_p(j) = \begin{cases} 1 & if\ f(V'_p(j)) > rand \\ 0 & otherwise \end{cases} \quad ,j = 1,..,d$$

۷- کرم شبتاب جدید را مورد ارزیابی قرار می دهیم و مقدار شفافیت آن را بدست می آوریم

۸- کرم شبتاب های فعلی و جدید را براساس مقدار روشنایشان مرتب می نماییم و بهترینهایشان را برای تکرار بعد انتخاب می نماییم

۹- در صورتیکه به ماکزیمم تکرار دست نیافتیم (شرط توقف برقرار نیست) به گام ۳ بازمی گردیم.

۴.انتخاب ویژگی براساس الگوریتم کرم تکامل تفاضلی

الگوریتم DE از گامهای زیر تشکیل شده است:

گام۱: جمعیت اولیه را در فضای جستجو با توزیع یکنواخت بطور تصافی ایجاد می نماییم.

$$x_i(j) = lb_j + (ub_j - lb_j) \times rand,\ \ j = 1,\dots,d \text{ and } i = 1,\dots,N$$

به ترتیب کران پایین و کران lb_j و ub_j تولید می نماید. $[0,1]$ تابعیست که یک عدد تصادفی در بازه $rand$که تعداد کل افراد موجود در جمعیت N تعداد کل ابعاد راه حل است و d-ام از فضای جستجو هستند. jبالای بعد است.

گام۲: عملگر جهش را بر روی هر بردار x_i (فرد از جمعیت) اجرا می نماییم تا بردار v_iرا بدست آوریم

گام۳: عملگر crossover را بر روی هر جفت بردار x_i و v_i اجرا می نماییم تا بردار u_i را بدست آوریم

گام ۴: عملگر انتخاب دو بردار x_i و u_i را برحسب مقدار تابع هدف مقایسه می نماید تا برداری را که برای نسل بعد زنده می ماند تعیین نماید که عملگر انتخاب بصورت زیر است (در صورتیکه تابع هدف بصورت ماکزیمم سازی تعریف شود آنگاه برداری که دارای مقدار تابع هدف بزرگتری است را در جمعیت نسل بعد قرار می دهیم)

$$x_i = \begin{cases} u_i & ,if\ f(u_i) \geq f(x_i) \\ x_i & ,\ otherwise \end{cases}$$

که $f(.)$ مقدار تابع هدف را به ازای ورودیش بازمی‌گرداند.

گام ۵: متغیر شمارنده نسل را افزایش می دهیم و در صورتیکه به ماکزیمم نسلها دست نیابیم به گام ۲ بازمی گردیم و در غیر اینصورت اجرای الگوریتم خاتمه می یابد.

۱.۳. تابع برازندگی

سه معیار برای ارزیابی کارایی سیستم تشخیص نفوذ بکار برده می شود که عبارتند از: ۱- نرخ تشخیص حمله[۱] (ADR) ۲-نرخ مثبت اشتباه[۲] یا هشدار نادرست (FPR) ۳-دقت سیستم[۳] (SA) یا نرخ دقت[۴] (AR) [۲۴]. هر یک از این معیارها به ترتیب با روابط (۱۰)، (۱۱) و (۱۲) تعریف می شوند.

$$DR = \frac{\text{تعداد حملاتی که به درستی حمله تشخیص داده شده‌اند}}{\text{تعداد کل حملات در داده تست}} \times 100\% \qquad (۱۰)$$

$$FPR = \frac{\text{تعداد نفوذهای نرمال که به اشتباه حمله تشخیص داده شده‌اند}}{\text{تعداد کل نفوذهای نرمال در داده تست}} \times 100\% \qquad (۱۱)$$

$$AR = \frac{\text{تعداد نمونه‌های درست تشخیص داده شده}}{\text{کل نمونه‌های داده تست}} \times 100\% \qquad (۱۲)$$

هر سیستم تشخیص نفود می‌بایست نرخ تشخیص حمله را بهبود دهد و نرخ هشدار نادرست را کاهش دهد. بنابراین مقادیر بالاتر DR و AR و مقادیر کمتر برای FPR عملکرد دسته‌بندی بهتری را برای سیستم‌های تشخیص نفود نشان می‌دهد.

برازندگی تعریف شده برای تابع بهینه‌سازی مورد استفاده، میزان شانس انتخاب هر راه حل را برای حضور در نسل بعد تعیین می نماید. در این مقاله، تابع برازندگی براساس دو معیار DR و FPR تعریف می‌شود. بنابراین راه حلی که

[۱] **Attack detection rate**
[۲] **False positive rate**
[۳] **System accuracy**
[۴] **Accuracy rate**

بیشترین مقدار DR و کمترین مقدار FPR را بدست دهد، بالاترین مقدار برازندگی را خواهد داشت. ترکیب این دو معیار تابع برازندگی را بصورت (۱۳) تعریف می نماید:

$$Fitness = 035 * DR + 0.3 * (1 - FPR) + 0.35 * AR; \quad (۱۳)$$

این معادله بیان می نماید که DR و FPR اهمیت متفاوتی براساس دو ضریب α و β دارند که $\alpha \in [0,1]$ و $\beta = 1 - \alpha$ است. این پارامترها در طول روال شبیه‌سازی به صورت $\alpha = 0.7$ و $\beta = 0.3$ درنظر گرفته شده‌اند. برای محاسبه برازندگی هر فرد از جمعیت ابتدا براساس زیرمجموعه ویژگیهای انتخاب شده توسط آن فرد SVM با الگوریتم ID3 ساخته می شود و سپس هر داده تست براساس درخت ساخته شده طبقه بندی می شود و براساس برچسب‌های کلاسی بدست آمده برای داده‌ها، مقدار برازندگی طبق رابطه (۱۳) محاسبه می شود.

۴. نتایج شبیه‌سازی

در شبیه‌سازی‌های انجام شده از مجموعه داده KDD Cup 99 برای تعیین میزان کارایی الگوریتم پیشنهادی استفاده می‌گردد[۲۵]. رکوردهای موجود در این مجموعه داده دارای ۴۱ ویژگی به اضافه یک برچسب کلاس می باشد. رکوردها شامل ۲۱ نوع حمله می باشند بنابراین رکوردهای بصورت نرمال یا یکی از انواع حمله برچسب گذاری می شوند. چهار دسته حمله به شرح زیر در مجموعه داده KDD Cup 99 مورد بررسی قرار می‌گیرد:

حمله DoS[1]: منابع سیستم در این نوع حمله بیش از حد مورد استفاده قرار می‌گیرد و موجب رد شدن درخواست‌های نرمال برای در اختیار گرفتن منابع می‌شود.

حمله R2L: حمله کننده در این نوع حمله با نفوذ به ماشین قربانی به صورت غیرمجاز و از راه دور از طریق حدس زدن رمز عبور، از حساب قانونی کاربر سوء استفاده کرده و اقدام به ارسال بسته بر روی شبکه می‌نماید.

حمله U2R: این نوع حمله ها به طور موفقیت آمیزی در ماشین قربانی اجرا شده و حمله کننده دسترسی های کاربر ارشد محلی را در اختیار می‌گیرد.

حمله Probing: سیستم‌ها در این نوع از حمله‌ها، به منظور جمع‌آوری اطلاعات و یا یافتن قابلیت‌های آسیب‌پذیری نظارت و کاوش می شوند.

[1] **Denial of service**

مجموعه داده‌های آموزشی و تست از KDD Cup 99 به ترتیب شامل ۴۹۴۰۲۰ و ۳۱۱۰۲۸ نمونه می باشد. همانطور که مشاهده می‌شود ابعاد این مجموعه داده‌ها برای استفاده بسیار بزرگ است. به همین دلیل دو زیرمجموعه داده (آموزشی و تست) بطور تصادفی از آن‌ها استخراج می شود و برای حفظ نسبت هر نوع حمله در هر دو مجموعه آموزشی و تست، تعداد نمونه‌های هر حمله بر ۱۰۰ تقسیم می‌گردد. برای مثال، تعداد حمله‌های ipsweep در مجموعه داده‌های آموزشی و تست اصلی ۱۲۴۷ و ۳۰۶ می باشد در حالیکه تعداد آنها در مجموعه داده‌های استخراج شده ۱۲ و ۳ می باشد. جدول ۱ انواع مختلف حمله و تعداد رخداد متناظرشان را در داده‌های آموزشی و تست به ترتیب نشان می دهد. در آزمایش‌های انجام شده تعداد داده‌های آموزشی ۴۹۴۷ و تعداد داده‌های تست ۳۱۱۷ می باشند که بطور تصادفی انتخاب شده‌اند. از جدول ۱، probing(41;42) به معنی آن است که تعداد رکوردها از حمله Prob در مجموعه آموزشی ۴۱ است در حالیکه تعداد رکوردها از این نوع حمله در مجموعه تست برابر ۴۲ می باشد.

جدول۱. انواع مختلف حمله و تعداد رخداد متناظرشان به ترتیب در مجموعه داده آموزشی و تست.

Normal(973;606)			
Probing (41; 42)	DoS(3915; 2299)	U2R(5; 10)	R2L(13; 160)
ipsweep(12;3), Mscan(0;11), Nmap(2;1) Portsweep(11;4) Saint(0;7), Satan(16;16).	apache2(0;8), back(22;11), land(0; 0), mailbomb(0;50), Neptune(1072;580), processtable(0;8), Pod(3;1), udpstorm(0;0), Smurf(2808;1641), Teardrop(10;0),	buffer_overflow(3;1), httptunnel(0;3), loadmodule(0;0), perl(0;0), rootkit(2;2), xterm(0;2), Ps(0;2), Sqlattack(0;0),	ftp_write(0;0), imap(0;0), guesspasswd(2;44), named(0;0), multihop(0;0), phf(0;0), sendmail(0;0), snmpgetattack(0;77), snmpguess(0;24), spy(0;0,) warezclient(10;0), worm(0;0), warezmaster(1;15), xsnoop(0;0), xlock(0;0),

روش پیشنهادی در آزمایش‌های انجام شده که بر مبنای الگوریتم بهینه‌سازی کرم شب‌تاب با درخت تصمیم و الگوریتم کرم شب‌تاب با **svm** و حالت استفاده از تمام ویژگی‌ها مقایسه می‌شود. اندازه جمعیت اولیه در هریک از الگوریتم ها ۳۰ و ماکزیمم تعداد تکرار الگوریتم‌ها ۵۰ می باشد. آزمایش ۲۰ مرتبه بصورت مستقل اجرا می شود و مقادیر میانگین معیارهای ارزیابی سیستم تشخیص برای این ۲۰ اجرای مستقل گزارش می شود. نتایج بدست آمده از آزمایش در جدول ۲ گزارش شده است.

جدول۲- نتایج سه معیار ارزیابی سیستم تشخیص نفوذ برای روشهای مختلف.

روش	DR(%)	AR(%)	FPR(%)

الگوریتم تکامل تفاضلی	۹۲,۶۷۴	۹۲,۲۴۴	0.66۲
الگوریتم کرم شب‌تاب با SVM	**95.19**	**95.45**	**3.45**
درخت تصمیم با تمام ویژگی‌ها	71.087	73.267	17.685

همانطور که در جدول ۲ مشاهده می‌شود روش پیشنهادی مبتنی بر الگوریتم کرم شب‌تاب با **svm** از نظر دو معیار DR، AR نسبت به سایر روش ها به بهترین مقدار دست یافت و از نظر معیار FPR بعد از روش کرم شب تاب با درخت تصمیم بهتر عمل نمود. به طور کلی می توان نتیجه گرفت روش پیشنهادی می تواند زیر مجموعه ویژگی بهینه را به دست آورد. موفقیت روش پیشنهادی مبتنی‌بر کرم شب‌تاب با svm در انتخاب زیرمجموعه ویژگی بهینه نسبت به روش‌های دیگر خواهد بود.

۵. نتیجه‌گیری

در این مقاله، روش جدیدی مبتنی بر ترکیب الگوریتم کرم شب‌تاب با svm به منظور انتخاب ویژگی‌های بهینه مورد نیاز برای اجرای روال تشخیص نفوذ در سیستم‌های کامپیوتری ارئه گردید. داده مورد نیاز برای فرآیند آموزش و تست از مجموعه KDD Cup 99 انتخاب گردید. در ابتدا، روال اصلاح الگوریتم کرم شب‌تاب به منظور تطبیق با شرایط موجود در فرآیند انتخاب ویژگی ارائه گردید. سپس طبقه‌بند مبتنی‌بر SVM برای ارزیابی ویژگی‌های انتخاب شده در الگوریتم کرم شب‌تاب بکار برده شد. نتایج شبیه‌سازی‌های انجام شده نشان داد که زیرمجموعه ویژگی‌های بهینه انتخابی توسط الگوریتم کرم شب‌تاب **svm** پیشنهادی بیشترین مقدار معیارهای DR و AR و کمترین مقدار معیار FPR را بدست می‌دهد. همچنین در کارهای آینده، بررسی استفاده از تکنیک‌های دیگر برای طبقه‌بندی مانند ماشین بردار پشتیبان، شبکه‌های عصبی و روش‌های خوشه بندی پیشنهاد می شود.

مراجع

[1][۱] D. Zamboni, "Using internal sensors for computer intrusion detection," *Center for Education and Research in Information Assurance and Security, Purdue University*, 2001.

[2][۲] Y. Y. Chung and N. Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)," *Applied Soft Computing*, vol. 12, pp. 3014-3022, 2012.

[3][۳] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36 , pp. 16-24, 2013.

[4][۴] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks," *Expert systems with Applications*, vol. 29, pp. 713-722, 2005.

[5][۵] G. Wang ,J. Hao, J. Ma, and L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," *Expert systems with applications*, vol. 37, pp. 6225-6232, 2010.

[6][۶] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge-based systems*, vol. 78, pp. 13-21, 2015.

[7][۷] J. Zhang, H. Li, Q. Gao, H. Wang, and Y. Luo, "Detecting anomalies from big network traffic data using an adaptive detection approach," *Information Sciences*, vol. 318, pp. 91-110, 2015.

[8][۸] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, pp. 1690-1700, 2014.

[9][۹] X‐.s. Gan, J.-s. Duanmu, J.-f. Wang, and W. Cong, "Anomaly intrusion detection based on PLS feature extraction and core vector machine," *Knowledge-Based Systems*, vol. 40, pp. 1-6, 2013.

[10][۱۰]A. Karami and M. Guerrero-Zapata, "A fuzzy anomaly detection system based on hybrid pso-kmeans algorithm in content-centric networks," *Neurocomputing*, vol. 149, pp. 1253-1269, 2015.

[11][۱۱]S. Rastegari, P. Hingston, and C.-P. Lam, "Evolving statistical rulesets for network intrusion detection," *Applied Soft Computing*, vol .33 ,pp. 348-359, 2015.

[12][۱۲] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on software engineering*, pp. 222-232, 1987.

[13][۱۳] Y. Li and L. Guo, "An active learning based TCM-KNN algorithm for supervised network intrusion detection," *Computers & security*, vol. 26, pp. 459-467, 2007.

[14][۱۴] W.-H. Chen, S.-H. Hsu, and H.-P. Shen, "Application of SVM and ANN for intrusion detection," *Computers & Operations Research*, vol. 32, pp. 2617-2634, 2005.

[15][۱۵] A. Zainal, M. A. Maarof, and S. M. Shamsuddin, "Ensemble classifiers for network intrusion detection system," *Journal of Information Assurance and Security*, vol. 4, pp. 217-225, 2009.

[16][۱۶] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using an ensemble of intelligent paradigms," *Journal of network and computer applications*, vol. 28, pp. 167-182, 2005.

[17][۱۷] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," *Applied Soft Computing*, vol. 10, pp. 1-35, 2010.

[18][۱۸] K. Shafi and H .A. Abbass, "Biologically-inspired complex adaptive systems approaches to network intrusion detection," *Information Security Technical Report*, vol. 12, pp. 209-217, 2007.

[19][۱۹] C. Khammassi and S. Krichen, "A GA-LR Wrapper Approach for Feature Selection in Network Intrusion Detection," *Computers & Security*, 2017.

[20][۲۰] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Systems with Applications*, vol. 42, pp. 2670-2679, 2015.

[21][۲۱] S. Salzberg, "Book Review: C4. 5: Programs for machine learning by J," *Ross Quinlan*, pp. 1-6, 1994.

[22][۲۲] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, pp. 341-359, 1997.

[23][۲۳] X. He, Q. Zhang, N. Sun, and Y. Dong, "Feature selection with discrete binary differential evolution," in *Artificial Intelligence and Computational Intelligence, 2009. AICI ۰۹'International Conference on*, 2009, pp. 327-330.

[24][۲۴] R.-C. Chen, K.-F. Cheng, Y.-H. Chen, and C.-F. Hsieh, "Using rough set and support vector machine for network intrusion detection system," in *Intelligent Information and Database Systems, 2009. ACIIDS 2009. First Asian Conference on*, 2009, pp. 465-470.

[25] http://kdd.ics.uci.edu