بهبود امنیت انتقال اطلاعات در مواجه با رخدادهای اتفاقی توسط الگوی ماشین بردار پشتیبان و شناسایی الگوریتم‌های دفاعی کاربردی

امیرحسین رحیمی دشتی
موسسه آموزش عالی غیر انتفاعی لامعی گرگانی

رضا روشنی
۱. موسسه آموزش عالی غیر انتفاعی لامعی گرگانی
۲. گروه مهندسی کامپیوتر، دانشگاه ملی مهارت، تهران، ایران

چکیده

در سیستم تشخیص نفوذ داده‌های یادگیری می‌تواند اطلاعات ترافیک شبکه و یا اطلاعات کارت‌های اعتباری مشتریان و کاربران محیط ابری باشد و خصوصیت مورد نظر عادی یا غیر عادی بودن یک ارتباط است. درخت تصمیم یک ابزار برای پشتیبانی از تصمیم است که از درختان برای مدل کردن استفاده می‌کند و در جاهایی بکار می‌رود که لازم است استراتژی با بیشترین احتمال به هدف برسد. یکی از ویژگی‌های مناسب درخت تصمیم، قابلیت ترکیب با روش های دیگر است به گونه‌ای که در این تحقیق، نتیجه درخت تصمیم با روش ماشین بردار پشتیبان ترکیب شده تا نتایج بهتری بدست آید. الگوریتم ماشین بردار پشتیبان جزء الگوریتم‌های دسته‌بندی می‌باشد و با استفاده از داده‌های یادگیری این توان را پیدا می‌کند که خصوصیت مورد نظر را در داده جدید پیش‌بینی کند. بنابراین ما در این تحقیق، روشی ترکیبی ماشین بردار پشتیبان و درخت تصمیم را پیشنهاد می‌کنیم که موجب افزایش دقت سیستم تشخیص نفوذ می‌شود.

واژگان کلیدی: داده کاوی، سیستم تشخیص نفوذ، الگوریتم تدافعی، ماشین بردار پشتیبان، درخت تصمیم

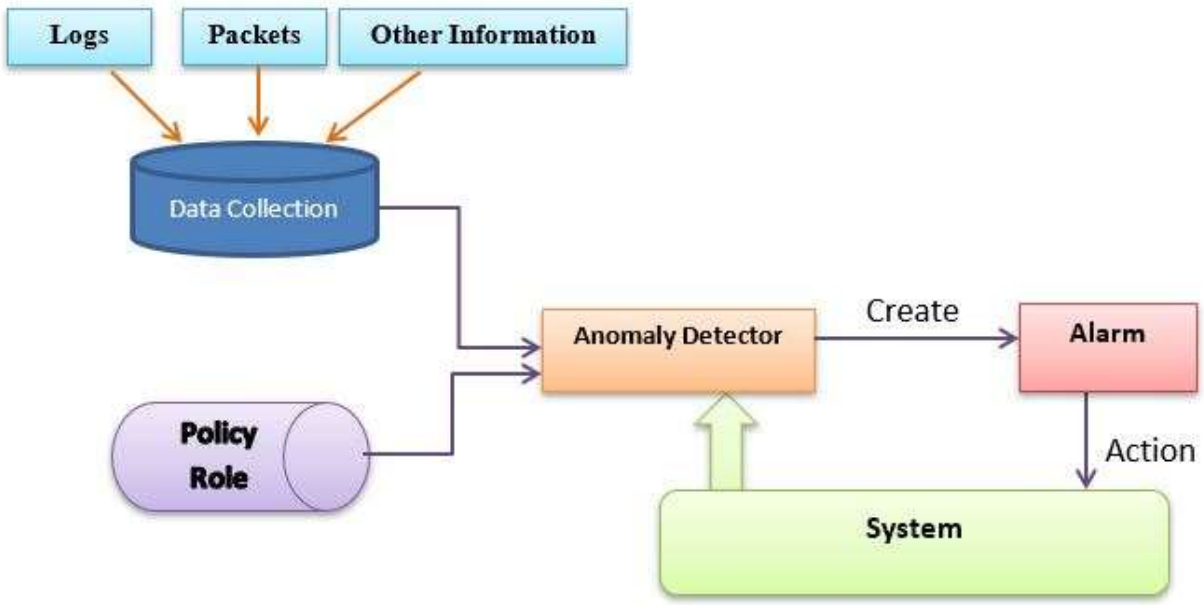
۱. مقدمه

انواع سیستم‌های تشخیص نفوذ و نوع نگرش آنها به حمالت و روش‌های برخورد با حمالت در شبکه کامپیوتری از جمله مباحثی است که در این بخش به آن پرداخته شده است. در ادامه ی به طبقه‌بندی سیستم‌های تشخیص نفوذ از دیدگاه‌های مختلف پرداخته، جزئیات هر نمونه و نمونه‌هایی از کارهای انجام گرفته در هر زمینه می‌شود. سپس الگوریتم‌های مورد استفاده در این پژوهش شامل الگوریتم‌های متاهیورستیک SVM و الگوریتم تدافعی تشریح می‌گردند.تاکنون روش‌های مختلفی برای تشخیص نفوذ استفاده شده‌است که به نتایج مختلفی با توجه به مشخصه های در نظر گرفته شده با استفاده از روشهای گوناگون رسیده اند. با افزایش روزانه جمعیت جهان و افزایش تعداد کاربران اینترنت و خدمات اینترنتی و شبکه سیل تعداد کاربرانی که قصد نفوذ به شبکه را دارند رو به افزایش است و نیاز به روشهای نوین در تشخیص این قبیل حمالت نیز باید به طبع افزایش یابد. استفاده از علم داده کاوی برای بدست آوردن الگوی درون اطلاعات توانسته کمک بسیاری در زمینه تشخیص نفوذ بکند چرا که با تشخیص و کشف قوانین در مجموعه های داده توانسته اطلاعات و روش های جدیدی را بوجود آورد. یکی از تکنیکهای بکارگیری داده کاوی که در سال های اخیر بیشتر مورد توجه کارشناسان دنیای امنیت قرار گرفته است روشهای ترکیبی است که در آن از چندین روش به صورت ترکیبی استفاده میشود. که در این پژوهش از ترکیب دو الگوریتم تدافعی و الگوریتم بهینه سازی SVM برای سیستم پیشنهادی استفاده شده است که این دو الگوریتم با انتقال داده ها و پاسخ ها با یکدیگر مدل تشخیص نفوذ را ایجاد نموده اند.

برحسب نظر بسیاری از محققان پس از نیروی انسانی با ارزش ترین و حیاتی ترین سرمایه یک سازمان اطلاعات آن سازمان خواهد بود. اطلاعات دارایی هایی هستند که قدرت سازمان ها را شکل می دهند (Power Is Information). دنیای کنونی دنیای اطلاعات و ارتباطات است ولی استفاده از داده ها و اطلاعات به صورت امن و بدور از دسترسی های غیر مجاز خود دغدغه ای جدی می باشد چرا که افرادی سودجو وجود دارند که به واسطه آن با رخنه کردن به داخل شبکه سعی میکنند اطلاعات مورد نیاز و حساس و محرمانه را برداشته و برای مقاصد دیگر و منفی استفاده کنند، از جمله حمالتی به حسابهای مالی کاربران، برداشتن اطلاعات شخصی دیگران و غیره. با توجه به اینکه امنیت از اهداف اولیه طراحی اینترنت نبود است، در دهه های اخیر ایمن سازی این شبکه ها در برابر حملات از اهمیت بسیاری برخوردار شده است. امروزه جهت تأمین امنیت، سیستم ها و ابزارهای امنیتی متفاوتی از جمله سیستمهای تشخیص نفوذ در شبکه ها استفاده میشوند.

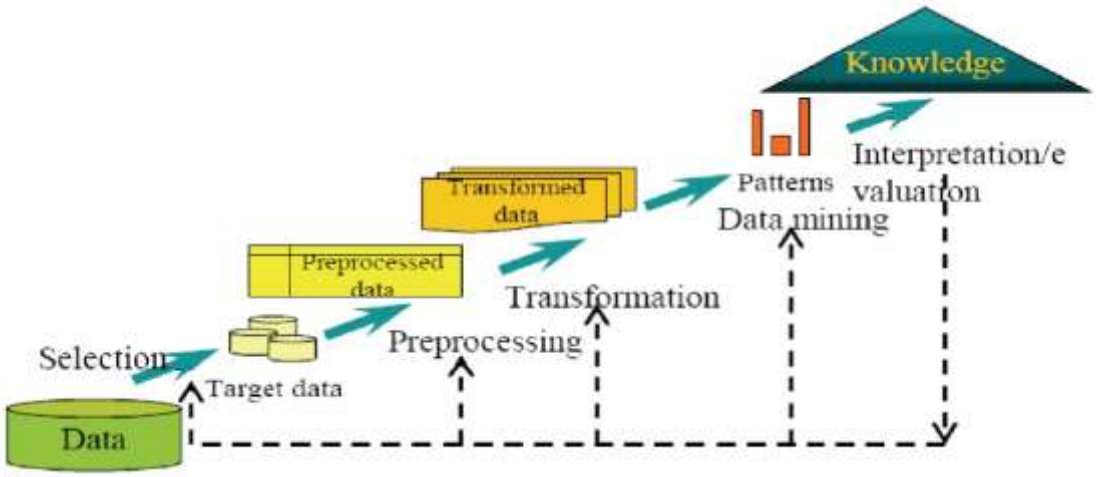
سیستم های رایانه ای به طرز قابل توجهی در دنیای کنونی وارد گردیده اند به نحویکه دنیای کنونی را بدون این سیستم ها نمی توان تصور نمود. اگر چه این سیستم ها سبب بالا بردن کیفیت های زندگی در جهات مختلف گردیده اند اما نوع جدیدی از تهدیداتی را نیز سبب گردیده اند که بی توجهی به این تهدیدات می توان سبب بروز چالشها و زیان های فراوانی گردد. یکی ازاین تهدیدات و شاید مهمترین این چالشها به تهدیدات امنیتی است لذا تشخیص نفوذ در تجهیزات و سیستم های رایانه ای از اهمیت ویژه ای برخوردار است. به زبان ساده نفوذ عبارت است از هر نوع اقداماتی که مثلث امنیت شامل صحت ، محرمانگی و یا دسترسی را به خطر بیندازد . نفوذ ممکن است از درون شبکه داخلی و توسط افراد مجاز / غیر مجاز صورت پذیرد و یا از خارج سازمان نشات گرفته باشد . مهاجمان از روش هایی نظیر دستیابی به کلمات عبور ، استفاده از عیوب نرم افزار ها و یا سرویس ها و مشکلات طراحی شبکه ای اقدام به نفوذ می نمایند که بر اساس نوع فعالیت نفوذگر ، اقدامات مجرمانه اعم از دسترسی غیر مجاز ، استراق سمع و... صورت می پردازد روش ها و نرم افزار های متنوعی در جهت مقابله با این مهاجمان به وجود آمده اند که سیستم های تشخیص نفوذ (**IDS**) یکی از انواع سیستم ها می باشد. اینگونه سیستم ها با نظارت بر ترافیک شبکه و رفتارهای کاربران به تشخیص رفتارهای مخرب کاربران سیستم های رایانه ای می پردازد هر چند این نفوذ از درون یا بیرون سازمان سرچشمه گرفته باشد. در حقیقت سیستم تشخیص نفوذ با رصد نمودن ترافیک موجود در شبکه و بسته های مبادله شده داده ها و اطاعات غیر ضروری و غیر مفید را حذف می نمایند و در مرحله بعد با تحلیل و

بازبینی این بسته ها یک سری ویژگی و خصیصه جمع آوری می نمایند آنگاه بر اساس الگوریتم ها و روش های مختلف به بررسی این خصیصه ها می نمایندآنگاه پس از ارزیابی و بررسی این فعالیت ها ،میزان احتمال وجود حمله بررسی می گردد.
مهمترین و حساس ترین بخش یک سیستم تشخیص نفوذ نیز دقیقاً همین بخش می باشد که به اصلاح به این بخش تشخیص دهنده گفته می شود. شکل ۱ نمای کلی از یک سیستم تشخیص نفوذ را نشان می دهد.



شکل ۱ : نمای سیستم تشخیص نفوذ به صورت ساده

از نظر نوع معماری سیستم های تشخیص نفوذ به سه دسته تقسیم می شوند : سلسله مراتبی، شبکه ای و ترکیبی. وارسی داده ، پیش پردازش و استفاده از الگوریتم های داده کاوی می باشند که در نهایت الگوها و یا روابط تازه و قابل درکی را کشف خواهند نمود. داده کاوی در حقیقت فرایند جستجوی خودکار در میان حجم بالایی از داده ها (شامل پایگاه های داده ، **DataWareHouse** و...) برای یافتن و تشکیل الگوها قابل فهم با کمک قوانین همبستگی جهت کشف دانش می باشد. در شکل۲ فرایند کشف دانش نشان داده شده است.



در شکل ۲ فرایند کشف دانش

۲. کارهای مرتبط

در ((Guo, C., Ping, Y., Liu, N., & Luo, S. S. (2016)) و همکاران یک روش ترکیبی دو سطحی برای تشخیص نفوذ با استفاده از الگوریتم ها و عملکردهای RF KNN و K-means ارائه داده اند که این روش ترکیبی جهت دستیابی به نرخ تشخیص بالا با میزان مثبت کاذب کم بوده است. این روش ترکیبی شامل دو مؤلفه تشخیص ناهنجاری و یک مؤلفه تشخیص سوء استفاده می باشد (مؤلفه ۱ تشخیصی ناهنجاری جهت طبیعی یا غیر طبیعی بودن رفتار استفاده شده است و در صورت غیر طبیعی بودن به مؤلفه ۲ تشخیص ناهنجاری جهت ارزیابی بیشتر ارسال می شود در غیر این صورت به مؤلفه تشخیصی سوء استفاده ارسال می شود در صورتی که مؤلفه ۲ تشخیصی ناهنجاری تشخیص دهد که رفتار طبیعی است و مؤلفه تشخیصی سوء استفاده تشخیص دهد که حمله نیست در نتیجه ارتباط نرمال است در غیر این صورت اگر مؤلفه ۲ تشخیص ناهنجاری تشخیص دهد که رفتار غیر طبیعی است و مؤلفه تشخیصی سوء استفاده تشخیص دهد که حمله است پس در نتیجه حمله رخ داده است و ارتباط غیر نرمال است)، توسط این مؤلفه‌ها تشخیص می دهند که ارتباط داخل شبکه نرمال یا حمله است. آزمایشات روی ؛ حمله dos،U2R Probe و R2L انجام شده است و به این نتیجه رسیدند که با در نظر داشتن اینکه تشخیص ناهنجاری اغلب موجب مثبت کاذب می‌شود این روش بطور مؤثر ناهنجاری شبکه با میزان مثبت کاذب پایین را تشخیص می دهد.

در ((Lin, W. C., Ke, S. W., & Tsai, C. F. (2015)) و همکاران یک رویکرد بازنمایی از ویژگی های جدید، یعنی مرکز خوشه و رویکرد نزدیکترین همسایه CANN پیشنهاد کردند. آن‌ها در ابتدا مروری بر روش های یادگیری ماشین با نظارت و بدون نظارت داشته اند (با نظارت از K - نزدیکترین همسایه K-NN که الگوریتم های طبقه بندی غیر پارامتری معمولی در یادگیری ماشین تاثیر دارد استفاده کرده اند هدف از این الگوریتم اختصاص یک برچسب به نمونه داده بدون برچسب به کلاس از K نزدیکترین همسایه آن و در یادگیری بدون نظارت از الگوریتم خوشه بندی K-means استفاده کرده اند که یک راه ساده و آسان برای طبقه بندی یک مجموعه داده از طریق یک تعداد معینی از خوشه است . هدف از الگوریتم K-means در اینجا این است که نقاط K از یک مجموعه داده، به بهترین وجه را پیدا کند. نقاط k مرکز خوشه یا مرکز ثقل از هر خوشه است). سپس آن ها روش هایی برای تشخیص نفوذ از جمله CANN و پردازش آن ارائه داده اند (روش پیشنهاد شده دو فاصله برای تعیین ویژگی های جدید مد نظر قرار می‌دهد، فاصله اول بین یک نقطه خاص داده Di و تمام داده های دیگر در یک خوشه سپس، کوتاه ترین فاصله بین دو نمونه داده به نمایندگی Di و نزدیکترین

همسایه آن) این امر منجر به مقدار ویژگی بر اساس فاصله جدید برای ارائه هر داده جدید است و بعد از آن استخراج مراکز خوشه و نزدیکترین همسایه توسط k-meanS محاسبه می‌شود و داده های جدید ایجاد میگردد. در این تحقیق آن ها نرخ دقت، تشخیص و آلارم کاذب، را در نظر گرفتند. آن‌ها عملکرد CANN،Knn و حتی TAnn را در مجموعه داده های ۶ و ۱۹ بعدی مورد بررسی قرار دادند نتایج آن‌ها نشان داد که با توجه به نتایج اگر به طور متوسط دقت، سرعت کشف و شناسایی و نرخ هشدار غلط را در نظر بگیریم، برای مجموعه داده ۶ بعدی، CANN از نظر نرخ تشخیص و نرخ هشدار غلط به بهترین نحو انجام می‌شود، در حالی که برای مجموعه داده ۱۹ بعدی K-NN بهتر از نظر دقت به بهترین نحو انجام می شود.

در ((Singh, R., Kumar, H., & Singla, R. K. (2015) و همکاران یک روش بر اساس پروفایل ترافیک شبکه و یادگیری آنلاین زیاد پی در پی ماشین (OS-ELM) برای تشخیصی نفوذ ارائه داده اند. هکرها از ویژگی‌های پیشرفته مانند پورت پویا، IP آدرسی، Spoofing، محموله های رمزگذاری شده و غیره، برای جلوگیری از تشخیص استفاده می کردند. محققان اعتقاد داشتند که این نوع از نفوذ را می توان با کشف الگوها در دیتاست ترافیک شبکه شناسایی کرد. با توجه به دیتاست های بزرگ و نامتوازن سیستم تشخیصی نفوذ IDS بر اساس یادگیری ماشین با مشکل پردازش کل داده ها مواجه می شوند. بنابراین، برای شناسایی نفوذ به رفتار ترافیک شبکه لازم بود. در روش پیشنهادی از دو پروفایل استفاده کرده اند که پروفایل آلفا برای کاهش پیچیدگی زمانی استفاده می شود در حالی که ویژگی های بی ربط با استفاده از مجموعه ای از تکنیک های انتخاب ویژگی FSS بر اساسی فیلتر، همبستگی و سازگاری دور انداخته می شوند و به جای نمونه گیری، پروفایل بتا به منظور کاهش حجم دیتاست آموزشی استفاده کرده اند. در این تحقیق از سه روش ترکیبی از این پروفایل ها Alpha-FullFeatures , Alpha-FST , Alpha-FST-Beta ها را مقایسه کردند و نتایج نشان داد که روش Alpha-FST-Beta بهینه و کامل تر از دو روش دیگه می باشد. آزمایشات انجام شده توسط تکنیک هایی مانند رگرسیون، SVM و RF بوده است. نتایج تجربی نشان می دهد که روش پیشنهادی بهتر از روش های دیگر انجام می شود. به روشنی نشان می دهد که هیچ روش دیگر قادر به کاهش ابعاد و حجم نمونه با استفاده از پروفایل آلفا و بتا نخواهد بود. تجزیه و تحلیل مقایسه نشان می دهد که پیشنهاد IDS برای تشخیص نفوذ به شبکه های کامپیوتری کارآمد می باشد.

۳.روش پیشنهادی

روش های مورد استفاده در این پژوهش به صورت دقیق تر در این بخش توضیح داده شده اند. در حقیقت وظیفه و کاربرد اصلی سیستم پیشنهادی، تشخیص نفوذ و رفتار های غیر طبیعی در بستر شبکه می‌باشد. ضمنا توانایی تشخیص روش های نفوذ جدید نیز توسط این سیستم مقدور خواهد بود به نحویکه رفتار های غیر طبیعی جدید نیز توسط سیستم قابل تشخیص خواهد بود. روش پیشنهادی به تفکیک مراحل در چند بخش تشریح می گردد

جهت آموزش سیستم تشخیص نفوذ از دیتاست KDD-NSL استفاده گردیده که به صورت مختصر در ذیل توضیح داده شده است. در پژوهش انجام شده به تشخیص این مورد پرداخته شد که آیا کاربر وارد شده به شبکه قصد نفوذ به شبکه را دارد یا خیر.

۳-۱. مدل های دسته جمعی

روش دسته بندی جمعی چندین دسته بند را با هم ترکیب می کند تا به کارایی بالاتری دست یافته و آنها را بهبود دهد. این روش ،دقت دسته بندی را توسط نتایج پیش بینی چند دسته بند افزایش می دهد.گوناگونی دسته بندها در نمونه هایی که به اشتباه دسته بندی شده اند بر موفقیت رویکرد جمعی موثرند.این گوناگونی به چهار طریق قابل دسترسی است.

۱- از روش های مختلف آموزشی برای آموزش تمام دسته بندها استفاده گردد.

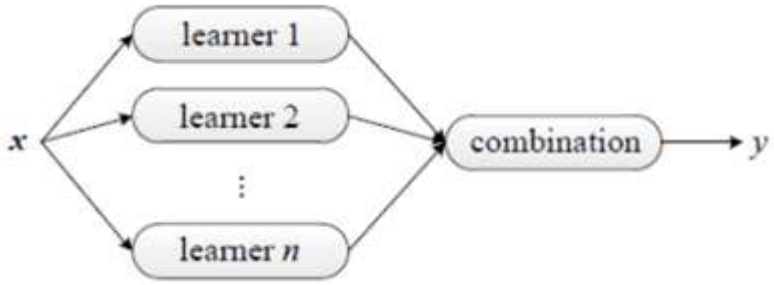
۲- بکارگیری پارامترهای مختلف آموزش

۳- بکارگیری ویژگی های مختلف جهت آموزش دسته بند

۴- بکارگیری الگوریتم های مختلف دسته بندی

۲-۳. روش های دسته جمعی

برای حل یک مسئله تکنیک های دسته جمعی یادگیرنده های متعدد را آموزش می دهند. در صورتی که روش های یادگیری عادی که برای ایجاد تک یادگیرنده از مجموعه داده آموزشی استفاده می کنند، روش دسته جمعی جهت ایجاد دسته ای از یادگیرندگان و سپس ترکیب نتایج آنها سعی می کند. ضمنا باید یادآوری نمود که یادگیری به روش دسته جمعی مبتنی بر کمیته ، یادگیری سیستم دسته بند چندگانه نیز اطلاق می شود. معماری دسته جمعی در شکل ۳ را نشان داده شده است. تکنیک یادگیری دسته جمعی عبارت است از مجموعه ای از یادگیرندگان که به این دسته ، یادگیرندگان پایه نامیده می شوند. یادگیرندگان پایه عموما از محموعه داده های آموزشی بوسیله الگوریتم های یادگیری پایه تولید می گردند که این الگوریتم ها می تواند از نوع الگوریتم های شبکه عصبی ،درخت تصمیم ویا انواع الگوریتم های یادگیری ماشین باشند.



شکل ۳ : یادگیریدسته جمعی

۳-۳. بوستینگ

سیستم های تشخیص نفوذ با دقت زیادی به تهدیدات و حمله های شبکه ای پی می برند.یادگیری ماشینی ای که بر مدل های تشخیص نفوذ مبتنی است جهت افزایش نرخ تشخیص تمام تلاش خود را می کند. آنچه که می تواند سبب کاهش دقت در سیستم های تشخیص نفوذ شود ،مشکل داده های نامتوازن است ،که معمولا در داده های مرتبط به حملات شبکه دیده می شود.

در حوزه یادگیری ماشین الگوریتم بوستینگ از توانایی بالایی برخوردار می باشد. بوستینگ[UO] فرا الگوریتم ترکیبی است که برای کاستن عدم توازن و همچنین کاهش میزان واریانس در مجموعه داده نامتوازن کاربرد دارد. این روش در حوزه یادگیری با نظارت در خانواده الگوریتم های یادگیری ماشین بحساب می آید. بوستینگ به زبان ساده به دسته ای از الگوریتم ها اشاره دارد که قادرند تا یادگیرنده های ضعیف را به یادگیرنده های قوی تبدیل نمایند. یادگیرنده های ضعیف در اجرا نتیجه ای فقط اندکی بهتر از حدس تصادفی(۵۰ به ۵۰) را خواهند داشت، در صورتی که یادگیرنده های قوی دارای نتیجه ای بسیار نزدیک به عملکرد واقعی و گاها کاملا درست بی عیب می باشند.

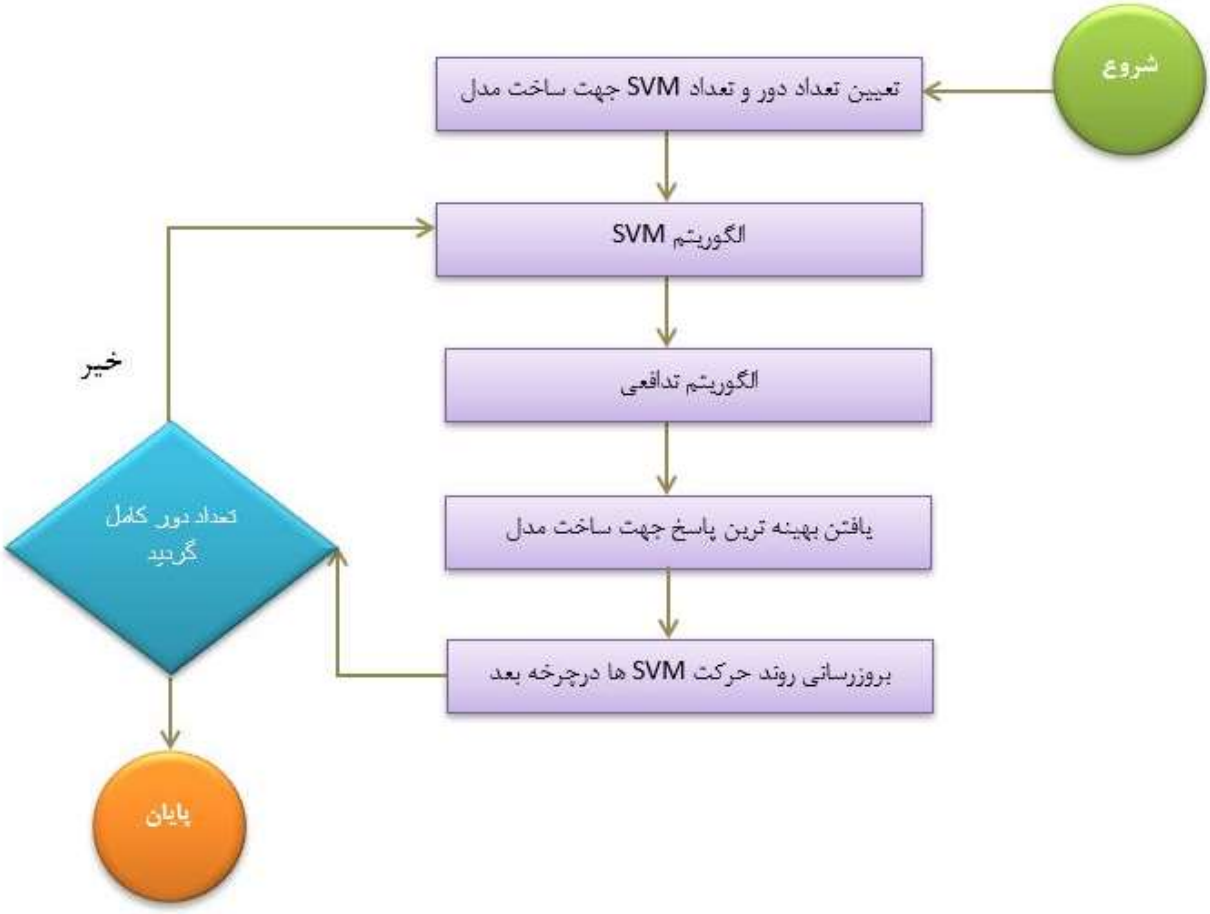
اگرچه بوستینگ در ساختار اصلی الگوریتم جای ندارد ولی بیشتر الگوریتم هایی که بر اساس بوستینگ ایجادو طراحی شده اند، تعدادی از یادگیرنده های ضعیف را به صورت چرخه های تکرار شونده آموزش می دهند و نتایج حاصل را به مجموعه های قبلی اضافه می نماید تا با اتمام چرخه و پایان اجرای الگوریتم یک طبقه بند قوی ایجاد گردد. تمام یادگیرنده های ضعیف در زمان اضافه شدن به مجموعه ، وزن دهی خواهند شد که عموما براساس مقدار دقت در طبقه بندی نمونه های آموزشی میزان وزن تعیین می شود. پس از افزوده شدن هریک از طبقه بندها، داده ها (نمونه های موجود) نیز وزن دهی خواهند شد (وزن نمونه ها بسته به نیاز اصلاح می گردد). وزن دهی به اینگونه می باشد که در هر یک از مراحل، در نمونه هایی که طبقه بندی به صورت صحیح اجرا شده است ، میزان وزن کاهش یافته و داده هایی که به درستی طبقه بندی نشده اند، مقدار وزن بیشتری برای نمونه تعیین می شود. پس از افزوده شدن هریک از طبقه بندها، داده ها (نمونه های موجود) نیز وزن دهی خواهند شد (وزن نمونه ها بسته به نیاز اصلاح می گردد). وزن دهی به اینگونه می باشد که در هر یک از مراحل، در نمونه هایی که طبقه بندی به صورت صحیح اجرا شده است ، میزان وزن کاهش یافته و داده هایی که به درستی طبقه بندی نشده اند، مقدار وزن بیشتری برای نمونه تعیین می شود. فروند و شاپیر در سال ۱۹۹۶روش جدیدی را با نام تدافعی برای ایجاد سیستم ترکیبی ارائه نمودند. این الگوریتم با استقبال فراوانی روبرو شد و به عنوان یکی از ۱۰ الگوریتم برتر داده کاوی شناخته شده است. کلمه تدافعی مخفف دو کلمه Adaptive Boosting می باشد که یک الگوریتم از نوع بوستینگ می باشد و یکی از اساسی ترین و پرکاربردترین روش های یادگیری چند گانه[KG] است، چرا که از سادگی در پیاده سازی ، محاسبات بسیار دقیق و مبانی نظری قوی برخوردار می باشد. اساس تدافعی به یادگیری بر مبنای PAC شکل گرفته است. روش PAC در حقیقت روشی برای ارزیابی کارایی یادگیرنده ها می باشد.میزان خطای آموزش در این یادگیری باید در حد قابل قبولی باشد. در PAC ثابت گردید که ترکیب یادگیرنده های ساده و ضعیف که نتایجی تنها بهتر از انتخاب تصادفی(نتایج بهتر از پنجاه درصد) دارند می توانند در تهایت امر یک دسته بند نهایی خوب را شکل دهند که این نکته اصل و مبنای ایده ایجاد بوستینگ می باشد.

۴. یافته ها

برای ارزیابی روش پیشنهادی از مجموعه داده Cup KDD ۱۹۹۹ استفاده می‌گردد. این مجموعه داده تنها و کاملترین مجموعه داده است که شامل بیش از ۴ میلیون رکورد که هر رکورد نمونه یک داده است، میباشد .هر نمونه دارای ۴۲ متغییر است که ۴۱ رکورد اول

را ویژگیها تشکیل میدهند و متغیر آخر بیان کننده برچسب هر نمونه است. نمونه ما زیر مجموعهای از داده فوق است که تحت نام Cup KDD NSLگردآوری شده است. این مجموعه به خوبی بیانگر مجموعه داده اصلی است. هر ثبت، ویژگی های اتصال متنوعی را در بر میگیرد، مانند: نوع خدمات، نوع پروتکل و تعداد دفعاتی که کاربر برای ورود به شبکه تالش میکند اما موفق نمیشود

در سیستم پیشنهادی از الگوریتم SVM به همراه الگوریتم تدافعی استفاده گردید که نحوه ارتباط میان دو الگوریتم در جهت ساختن یک مدل واحد در فلوچارت شکل ۳ ترسیم شده است.



شکل ۳: عملکرد روش پیشنهادی

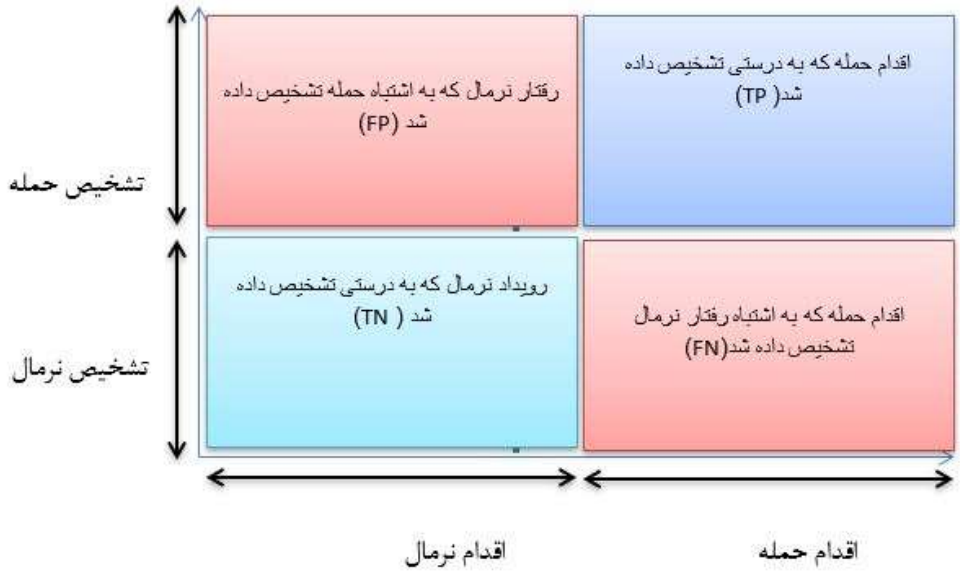
در اولین گام و جهت کاهش ابعاد فضای مجموعه داده نیاز به انتخاب خصایص می باشد تا میزان محاسبات و استفاده از منابع کاهش یافته و درک وغهم مسئله آسانتر گردد تا با صرف هزینه کمتر و به کمک داده کاوی استخراج دانش به صورت صحیح صورت پذیرد.اما باید این انتخاب خواص در حالتی صورت پذیرد که خواص و ویژگی داده ها تغییر نیابد و تنها خصیصه هایی حذف گردند که اثر بخش و تاثیگذار نباشند.

۴-۱. مقدار دهی اولیه به پارامترها

جهت اجرای شبیه سازی و ایجاد مدل نهایی و درنتیجه ارزیابی این مدل باید مقادیر اولیه در زبان برنام نویسی متلب مقداردهی گردند.این مقداردهی نیازمند دقت فراوان می باشد. به عنوان نمونه در میزان تعداد SVM ها باید به یک نقطه متعادل دست یافت چرا که تعیین تعداد SVM به مقدار کم سبب می گردد تا تمام فضای مسئله به درستی بررسی نشده و آموزش در بهینه ترین حالت ممکن صورت نپذیرد و میزان تعداد SVM زیاد نیز سبب اجرای محاسبات بسیار طولانی و زمانبر می گردد.

۴-۲. ارزیابی واکنش سیستم به رویدادها

به طورکلی واکنش های یک IDS در برابر رویدادهای یک سیستم را می توان به چهار گروه تقسیم بندی نمود که در شکل ۴ نشان داده شده اند.



شکل ۴:نتیجه ارزیابی حملات توسط سیستم

True Positive (TP) :تعداد رکوردهای نفوذ که بصورت یک فعالیت نفوذ طبقه بندی شده اند.
True Negative (TN) : تعداد فعالیت های نرمال که بصورت یک فعالیت نرمال طبقه بندی شده اند.
False positive (FP) :تعداد فعالیت های نرمال که بصورت یک فعالیت نفوذ طبقه بندی شده اند.
False Negative (FN) : تعداد رکوردهای نفوذ که بصورت یک رویداد نرمال طبقه بندی بندی گردیده اند.

در سیستم پیشنهادی مقادیر بالا در جدول ۱ نشان داده شده اند:

جدول ۱: نتیجه سیستم پیشنهادی در تشخیص نفوذ

	کلاس ۱	کلاس ۲	کلاس ۳	کلاس ۴	کلاس ۵
TP	۰.۹۹۹۱۱	۰.۹۹۹۴۹	۰.۹۹۶۲۸	۰.۸۶۶۶۷	۰.۹۹۸۵۷
FP	۰.۰۰۰۸۸۷۹۲	۰.۰۰۰۵۰۹۶۵	۰.۰۰۳۷۱۷۵	۰.۱۳۳۳۳	۰.۰۰۱۴۲۸۶
FN	۰.۰۰۰۶۸۴۹۷	۰.۰۰۰۰۸۳۱۳۹	۰.۰۰۰۲۳۹۸۶	۰.۰۰۰۲۶۴۷۲	۰.۰۰۰۲۶۲۴۶
TN	۰.۹۹۹۳۲	۰.۹۹۹۹۲	۰.۹۹۹۷۶	۰.۹۹۹۹۷	۰.۹۹۹۷۴

یکی از مهمترین ملاک ها در ارزیابی کارایی یک سیستم تشخیص نفوذ نرخ تشخیص ، هشدار اشتباه و مقایسه بین این دو نوع نرخ می باشد.ضمنا ملاک های دیگری نیز وجود دارند که از اهمیت ویژه ای برخوردارند که از آن جمله می توان به تحمل خطای سیستم در مقابل حملات و کارایی سیستم در سه حوزه سرعت پردازش،سرعت انتشار وعکس العمل اشاره نمود.

۵. بحث و نتیجه‌گیری

هدف از این تحقیق حداکثر استفاده از اطلاعات موجود برای تشخیص نفوذ و تشخیص نوع حمله می‌باشد. استفاده از تدافعی و الگوریتم SVM در سیستم‌های تشخیص تهاجم باعث ایجاد انعطاف‌پذیری در این سیستم ها می‌شود و نیاز به بهنگام سازی مداوم این سیستم ها از بین می رود. از مزایای مهم سیستم‌های تشخیص تهاجم مبتنی بر روش پیشنهادی این است که علاوه بر آنکه قادر به شناسایی حملات موجود در مجموعه ی آموزشی هستند بلکه حتی حملات جدید را نیز شناسایی می کنند. در این پروژه، تدافعی در شناسایی و دسته بندی حملات شناخته شده و ناشناخته استفاده شد و میزان خطای حاصل از اعمال الگوریتم با مقالات مشابه که با سایر الگوریتم‌های تکاملی پیاده‌سازی شدند مقایسه گردید. ترکیب تدافعی و الگوریتم بهینه سازی SVM قادر به شناسایی ۹۹.۹ درصد حملات موجود در مجموعه آزمون بودند. در نتیجه می‌توان گفت که ترکیب الگوریتم های فرا ابتکاری SVM و تدافعی ، انتخاب مناسبی در سیستم های تشخیص نفوذ مبتنی بر روش سوء استفاده می باشد و به کمک آنها می توان به میزان تشخیص تهاجم خوبی دست یافت.

منابع

Guo, C., Ping, Y., Liu, N., & Luo, S. S-A two-level hybrid approach for intrusion detection. Neurocomputing, 214, 391-400.

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, 67, 296-303.

Lin, W. C., Ke, S. W., & Tsai, C. F. CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-based systems, 78, 13-21.

Singh, R., Kumar, H., & Singla, R. K. An intrusion detection system using network traffic profiling and online sequential extreme learning machine. Expert Systems with Applications, 42(22), 8609-8624 .

Aburomman, A. A., & Reaz, M. B. I. A novel SVM-kNN-PSO ensemble method for intrusion detection system. Applied Soft Computing, 38, 360-372.

Gupta, M., & Shrivastava, S. K. (2015). Intrusion Detection System based on SVM and Bee Colony. International Journal of Computer Applications, 111.(١٠)

Liao, H., Ding, S., Wang, M., & Ma, G. An overview on rough neural networks. Neural Computing and Applications, 27(7), 1805-1816..

Abadeh, M. S., Habibi, J., & Lucas, C. Intrusion detection using a fuzzy genetics-based learning algorithm. Journal of Network and Computer Applications, 30(1), 414-428.

Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on software engineering, (2), 222-232

Vapnik, V. The nature of statistical learning theory. Springer science & business media.

Mukkamala, S., & Sung, A. H. (2003). Identifying significant features for network forensic analysis using artificial intelligent techniques. International Journal of digital evidence, 1(4), 1-17.

Nia, F. Y., & Khalili, M. (2015, November). An efficient modeling algorithm for intrusion detection systems using C5. 0 and Bayesian Network structures. In Knowledge-Based Engineering and Innovation (KBEI), 2nd International Conference on (pp. 1117-1123). IEEE.

Natesan, P., Balasubramanie, P., & Gowrison, G. AdaBoost Algorithm with Single Weak Classifier in Network Intrusion Detection. In Network Security Attacks and Countermeasures (pp. 259-269). IGI Global.

Improving the security of information transmission in the face of accidental events by the Support Vector Machine model and identifying practical defense algorithms

AmirHossein Rahimi Dashti

Lamei Gorgani Institute of Higher Education, Gorgan, Iran

Reza Roshani

1. Lamei Gorgani Institute of Higher Education, Gorgan, Iran

2. Department of Mechanical Engineering, National University of Skills (NUS), Tehran, Iran

Abstract

In the intrusion detection system, learning data can be network traffic information or credit card information of customers and users of the cloud environment, and the desired characteristic is the normality or abnormality of a connection. A decision tree is a decision support tool that uses trees for modeling and is used in places where a strategy needs to achieve the goal with the highest probability. One of the suitable features of the decision tree is the ability to combine with other methods in such a way that in this research, the result of the decision tree is combined with the support vector machine method to obtain better results. . The support vector machine algorithm is part of the classification algorithms and by using the learning data, it finds the ability to predict the desired feature in the new data. Therefore, in this research, we propose a combined method of support vector machine and decision tree, which increases the accuracy of the intrusion detection system.

Keywords: Data mining, intrusion detection system, defensive algorithm, support vector machine, decision tree.