سیستم تشخیص نفوذ همکارانه مبتنی بر دسته‌بند بیزی

لیلی ذوالفقاری پور[1]، احسان طیرانی راد[2]

[1] آزمایشگاه تخصصی آپا دانشگاه فردوسی مشهد
leila.zolfaghari@mail.um.ac.ir

[2] آزمایشگاه تخصصی آپا دانشگاه فردوسی مشهد
tayarani@um.ac.ir

چکیده

امروزه بحث امنیت شبکه بیش از پیش مورد توجه پژوهشگران قرار گرفته است و تشخیص نفوذ به‌عنوان یکی از اجزای اصلی برقراری امنیت در شبکه‌های کامپیوتری شناخته می‌شود. سیستم‌های تشخیص نفوذ از تعدادی اشکالات مانند میزان بالای هشدارهای نادرست، بازده تشخیص کم و کارایی پایین رنج می‌برند. در این مقاله یک سیستم تشخیص نفوذ همکارانه مبتنی بر دسته‌بند بیزی برای بهبود مشکلات موجود ارائه شده است؛ به طوری‌که ترافیک به‌صورت جریانی وارد واحد آموزش دسته‌بندها شده و هر کدام از موتورهای تشخیص با استفاده از دانش اولیه ایجاد می‌شوند. در ادامه، در صورت مشاهده داده جدید، از سایر سیستم‌های موجود همکار، برای تشخیص داده جدید کمک خواسته می‌شود. هر کدام از سیستم‌ها با توجه به واحد آموزش خود نسبت به داده جدید تشخیص می‌دهند و نتایج را می‌فرستند. سپس با توجه به نتایج، رأی‌گیری می‌شود و مجدداً نتیجه تشخیص به‌هنگام می‌شود و داده جدید به همراه نتیجه به واحد آموزش اضافه می‌شود. ارزیابی این سیستم با استفاده از مجموعه داده NSL-KDD انجام و نتایج حاصل با دسته‌بند بیز ساده و بهبود یافته آن مقایسه شده است. نتایج شبیه‌سازی حاکی از آن است که سیستم پیشنهادی از لحاظ کارایی، دقت و نیز نرخ هشدارهای نادرست عملکرد مناسب‌تری نسبت به سیستم‌های موجود دارد.

کلمات کلیدی

امنیت شبکه، تشخیص نفوذ، تشخیص نفوذ همکارانه، دسته‌بند بیز افزایشی.

۱- مقدمه

حملات و نفوذها به شبکه‌های کامپیوتری همزمان با رشد این شبکه‌ها به نحو چشم‌گیری افزایش یافته است. به‌منظور مقابله با نفوذگران به شبکه‌ها و سیستم‌های کامپیوتری، روش‌های متعددی توسعه داده شده است که روش‌های تشخیص نفوذ نامیده می‌شوند. هدف از فرایند تشخیص نفوذ، شناسایی استفاده‌های غیر مجاز، سوءاستفاده‌ها و نیز آسیب‌های محتمل به سیستم‌ها و شبکه‌های کامپیوتری است. به‌طورکلی روش‌های تشخیص نفوذ به دو دسته اصلی تشخیص سوءاستفاده و تشخیص رفتار غیرعادی تقسیم می‌شوند [۱].

روش تشخیص امضاء براساس الگوی نفوذهای شناخته شده عمل می‌کند. در این روش، مسأله تشخیص نفوذ به یک مسأله دسته‌بندی تبدیل می‌شود. به عبارت دیگر، سیستم تشخیص نفوذ قادر است حمله‌هایی را که پیشتر الگوی آنها را در یک مرحله آموزشی فرا گرفته، تشخیص دهد. مهمترین خصیصه این روش آن است که سیستم امنیتی قادر است حمله‌های شناخته شده را با دقتی بالا و نرخ هشدار نادرست خیلی کم تشخیص دهد. منظور از هشدار نادرست، هشداری است که هنگام عدم وقوع حمله توسط سیستم تشخیص نفوذ اعلام می‌گردد [۲].

وجود هشدار نادرست حتی به میزان کم چنانچه بار ترافیکی عادی شبکه بالا باشد، باعث وقوع هشدارهای متعدد و خسته‌کننده می‌گردد. به همین دلیل نرخ

تولید هشدار نادرست توسط یک سیستم تشخیص نفوذ بایستی تا حد امکان پایین باشد. البته ذکر این نکته ضروری است که پایین نگه داشتن نرخ مزبور سبب کاهش توانایی سیستم در تشخیص حملات محتمل می‌گردد. به عبارتی بایستی میان دقت تشخیص بالا و نرخ هشدار نادرست پایین نوعی تعادل برقرار نمود [۳].

در سال‌های اخیر، تشخیص نفوذهای اینترنتی، پیچیده‌تر و سخت‌تر شده است. نفوذها معمولاً با کمک نرم‌افزارهای مخرب (با نام نرم‌افزارهای مخرب) از جمله کرم‌ها، ویروس‌ها، تروجان‌ها، و بدافزارهای جاسوسی انجام می‌شوند. تکنیک‌های نفوذ اخیر به اتصال تعداد زیادی گره به شکل یک باتنت تمایل دارند [۴]، و از آن گره به‌خطر افتاده برای راه‌اندازی حملات توزیع‌شده مانند حمله ممانعت از سرویس توزیع‌شده (DDoS[۱]) [۵] استفاده می‌کنند. ابزار IDS[۲] از منظر محیط نظارت، به دو دسته مبتنی بر میزبان (HIDS[۳]) و مبتنی بر شبکه (NIDS[۴]) تقسیم می‌شوند [۶]. سیستم‌های تشخیص نفوذ سنتی در انزوا کار می‌کنند و به‌راحتی می‌توانند با حملات جدید و ناشناخته به‌خطر بیافتند. همکاری میان IDSها، هر IDS را به استفاده از اطلاعات اشتراکی و استفاده از تجربه دیگرIDSها برای رسیدن به تشخیص دقیق‌تر نفوذ تشویق می‌کند. شبکه‌ای که IDSها برای تبادل اطلاعات با یکدیگر به آن متصل هستند، شبکه تشخیص نفوذ همکارانه (CIDN[۵]) نامیده می‌شود [۷]. ابزار CIDS[۶] به عنوان یک راه حل توافقی در نظر گرفته می‌شود که از اطلاعات منابع متعدد برای به دست‌آوردن درک بهتر هدف و تأثیر حملات اینترنتی پیچیده استفاده می‌کند. ابزار CIDS از عهده مشکلات کلاسیک سیستم‌های تشخیص نفوذ مانند حملات روز صفر و نرخ‌های هشدار بالا و چالش‌های معماری مانند نقطه منفرد شکست در طراحی متمرکز نیز بر می‌آید [۸].

طرح پیشنهادی در این تحقیق، طراحی یک سیستم تشخیص نفوذ همکارانه با استفاده از دسته‌بند بیز است. این سیستم پیشنهادی از چهار مؤلفه اصلی واحد آموزش، تشخیص، تجمیع و به‌روزرسانی تشکیل شده است. در ابتدا ترافیک ورودی به‌صورت جریانی وارد واحد آموزش دسته‌بند شده و به این صورت در این سیستم IDSها با استفاده از موتور تشخیص خود تشخیص محلی را انجام داده و با استفاده از تشخیص خود و تشخیص سایرین یک عمل تجمیع اعمال می‌شود و تشخیص نهایی انجام می‌شود و اصلاح به این صورت است که به‌صورت مستمر نتایج تشخیص به‌هنگام می‌شود. تابع تجمیع می‌تواند به شکل‌های مختلفی مانند میانگین، شمارش، مد و غیره باشد. از سوی دیگر همکاری بین موتورهای تشخیص‌ها با ارسال اطلاعات به یکدیگر برای اصلاح و به‌روزرسانی موتور تشخیص محلی انجام می‌شود. هدف از طراحی، بهبود کارایی، دقت تشخیص و نیز کاهش نرخ هشدارهای نادرست می‌باشد.

در ادامه این مقاله در بخش ۲ به مرور کارهای پیشین پرداخته می‌شود. در بخش ۳ سیستم پیشنهادی ارائه می‌شود. بخش ۴ نتایج آزمایش‌ها را مورد بحث قرار می‌دهد و در بخش آخر نتیجه‌گیری مقاله آورده شده است.

۲- کارهای پیشین

سیستم‌های تشخیص نفوذ بسیاری برای محافظت از شبکه در برابر انواع مختلف حملات ارائه شده است. در شبکه‌های ترافیک نرخ بالا و مقیاس بزرگ امروزی، IDSهای متمرکز سنتی نمی‌توانند با توجه به مقدار زیاد اطلاعات رد و بدل شده و پردازش شده به طور کارآمد عمل کنند. بنابراین IDSهای توزیع‌شده ارائه شده‌اند. برخی از آن‌ها مانند سیستم پیشنهادی توسط اربچر و همکاران [۹] شامل اشیاء حسگر هوشمند برای جذب ترافیک و تحلیلگر مرکزی برای تشخیص حمله است. حسگرها یک ساختار بهینه‌سازی شده برای محتوای اطلاعات به نام بردارهای ویژگی مبتنی بر جریان اطلاعات به منظور کاهش سطر اندازه داده ایجاد می‌کنند. پس از آن به تحلیلگر مرکزی برای تشخیص حمله منتقل می‌شود و بار اضافی بر روی شبکه را به طور خاص هنگامی که اندازه شبکه رشد می‌کند قرار خواهد داد. همچنین تحلیلگر مرکزی آسیب‌پذیر به نقطه منفرد شکست است.

در برخی از روش‌های دیگر مانند روش ارائه شده توسط گاناون و همکاران [۱۰]، پردازش داده‌ها در شیوه توزیع‌شده انجام شده است. در این سیستم مجموعه‌ای از واحدهای تجزیه و تحلیل برای افزایش سرعت پردازش تعریف شده است. خروجی این واحدهای تجزیه و تحلیل به میزبان‌های مختلف فرستاده می‌شوند و پس از پردازش، نتایج به دست آمده به سیستم ارسال می‌شوند. در نتیجه بار پردازش از یک سیستم مرکزی حذف شده است. با این حال، اشکال اصلی این سیستم، افزایش بار شبکه با توجه به ارتباطات اضافی در بین گره‌ها است.

SMLDIDS[۷] یک مدل جدید به نام سیستم تشخیص نفوذ مبتنی بر امضا با بکارگیری عامل‌های متحرک ارائه می‌دهد که می‌تواند تهدیدهای حتمی با نرخ موفقیت بالا توسط پایگاه داده‌های متعدد و کوچک را کشف کند [۱۱]. و مکانیزمی برای به روزرسانی پایگاه داده‌های امضا کوچک در فواصل منظم با استفاده از عامل‌های متحرک فراهم کند. براساس مزیت فناوری عامل متحرک، سیستم پیشنهادی ساسیکومار و منجولا [۱۲] یک سیستم تشخیص نفوذ توزیع‌شده مبتنی بر عامل است که از سه لایه تشکیل شده است. لایه‌ها شامل عامل میزبان و عامل شبکه، عامل‌های متحرک، عامل‌های پاسخ و تصمیم‌گیری است. MAD-IDS[۸] سیستم‌های تشخیص نفوذ مبتنی بر داده‌کاوی با استفاده از عامل متحرک را معرفی می‌کند سیستم MAD-IDS شیوه‌های داده‌کاوی و روش عامل متحرک به منظور شناسایی حملات شناخته شده و ناشناخته را ادغام می‌کند و ساختار توزیع شده این سیستم شامل عوامل مختلف که قادر به حرکت از یک ایستگاه به دیگری، به ترتیب: پویشگر، فیلتر، تشخیص استفاده نادرست، تشخیص ناهنجاری، قانون‌کاوی و عامل گزارشگر است [۱۳].

در [۱۴]، COLIDE[۹] (موتور تشخیص نفوذ همکارانه) یک چارچوب جدید برای تشخیص نفوذ کارآمد در شبکه‌های M2M[۱۰] ارائه می‌دهد بدون اینکه هزینه‌های بالای انرژی و ارتباطی را در گره‌های میزبان و لبه‌های شرکت‌کننده را تحمیل کند. چارچوب در نظر گرفته شده به چالش‌هایی مانند انعطاف‌پذیری، محدودیت منابع و ماهیت همکارانه شبکه M2M می‌پردازد. این پژوهش یک توصیف سیستم دقیق همراه با ارزیابی رسمی و تجربی آن با استفاده از سیستم Contiki ارائه می‌دهد. ارزیابی برای سناریوهای ارتباطی مختلف نشان می‌دهد که رویکرد پیشنهادی آنها سربار را از لحاظ مصرف انرژی و مصرف حافظه محدود کرده است.

در [۱۵]، یک روش تشخیص نفوذ همکارانه و سازگار بر اساس SVM[۱۱]ها و درخت‌های تصمیم دو کلاسه پیشنهاد شده است. مدل تشخیصی به نام CAIDM[۱۲] ایجاد و پیاده‌سازی شده است. مدل E-CARGO[۱۳] به عنوان یک ابزار برای توصیف تشخیص نفوذ و مدل‌سازی آن استفاده می‌شود. در این مقاله، نقش‌ها، گروه‌ها و عوامل همه مورد مطالعه و کاربرد قرار می‌گیرد، چهار نوع توابع شناسایی SVM وجود دارد که طراحی شده و پیاده‌سازی می‌شوند و

عوامل مرتبط ایجاد می‌شوند. این عوامل با استفاده از خواص مختلف ساخته شده‌اند تا به ترتیب برای حملات برای TCP[۱۴]، UDP[۱۵]، ICMP[۱۶] و پروتکل‌های لایه کاربرد استفاده شوند.. نتایج آزمایش نشان می‌دهد که مدل تشخیص نفوذ همکارانه و سازگار بهینه (CAIDM) بر اساس SVM ها و درخت‌های تصمیم دو کلاس دقیق‌تر و کارآمد تر از سیستم آشکارساز با مجموعه‌ای از SVM های تک نوع است.

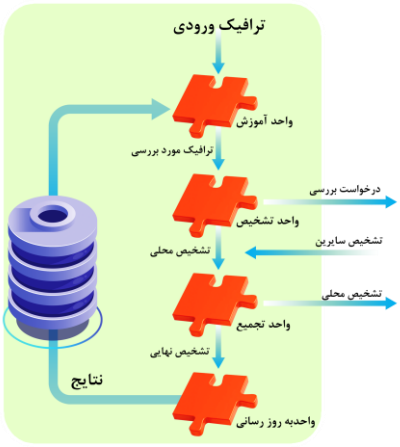
۳- معرفی سیستم پیشنهادی

در این بخش یک سیستم تشخیص نفوذ همکارانه مبتنی بر بیز ارائه شده است. در این سیستم از دسته‌بند بیز برای انجام عمل تشخیص نفوذ استفاده می‌شود. در مدل پیشنهادی یک سیستم چند مؤلفه‌ای در نظر گرفته شده است که هر کدام از این مؤلفه‌ها برای تشخیص یک نفوذ به صورت بهینه با یکدیگر همکاری می‌نمایند.

از آنجا که روش پیشنهادی از آموزش افزایشی استفاده می‌کند، در نتیجه هر بار از ترافیک ورودی به عنوان داده آموزشی استفاده شده و حجم آموزش هر عامل افزایش می‌یابد که این کار باعث افزایش دقت تشخیص و به تبع آن کاهش تعداد هشدارهای نادرست می‌گردد. از طرفی عامل‌ها در تشخیص هر ترافیک جدید با یکدیگر همکاری نموده و از نظر اکثریت برای تعیین نوع برچسب ترافیک ورودی استفاده می‌شود. در مدل پیشنهادی، با توجه به این که هر یک از عامل‌ها با استفاده از میزان قابل توجهی ترافیک ورودی آموزش داده شده‌اند، همکاری آن‌ها باعث می‌شود پیش‌بینی درست‌تری صورت گیرد که این عمل نیز به نوبه خود باعث افزایش دقت و کاهش تعداد هشدارهای نادرست می‌گردد.

در مدل پیشنهادی در این مقاله یک سیستم همکارانه مطابق شکل (۱) در نظر گرفته شده است که شامل چهار بخش می‌باشد:

- واحد آموزش
- واحد تشخیص
- واحد تجمیع
- واحد به‌روزرسانی



شکل (۱): معماری یک همکار در سیستم تشخیص نفوذ پیشنهادی

در ادامه کار هر بخش تشریح شده و به طرز کار سیستم و طریقه تعاملات این بخش‌ها پرداخته می‌شود.

بخش آموزش: نقش بخش آموزش در معماری پیشنهادی جمع‌آوری کننده ترافیک ورودی می‌باشد و در مؤلفه‌های نصب شده بر روی تمام میزبان‌های تحت نظارت قرار دارد و در این بخش پیش‌پردازش بر روی ترافیک ورودی انجام می‌شود و از این ترافیک برای آموزش مؤلفه‌ها استفاده می‌شود و نتایجی که از تشخیص ترافیک‌های ورودی به‌دست می‌آید به همراه آن ترافیک نیز به مجموعه آموزش اضافه می‌شود.

بخش تشخیص: این مؤلفه در سیستم پیشنهادی مهم‌ترین مؤلفه می‌باشد و در واقع نقش نهایی تشخیص نفوذ را بر عهده دارد. در این سیستم پیشنهادی موتور تشخیص، دسته‌بند بیز می‌باشد.

بخش تجمیع: نتایج حاصل از تشخیص محلی و تشخیص سایرین در این بخش جمع‌آوری شده و تصمیم‌گیری نهایی برای تشخیص در این بخش صورت می‌گیرد.

بخش به‌روزرسانی: نتایج نهایی از بخش قبلی را دریافت می‌کند و با استفاده از آنها تشخیص محلی را به‌روز می‌کند و ترافیک ورودی جدید به همراه نوع آن به بخش آموزش ارسال می‌شود تا به‌روزرسانی برای این ترافیک جدید توسط بخش آموزش انجام شود .

۳-۱- تعاملات بین بخش‌های سیستم پیشنهادی

همان‌طور که گفته شد در معماری پیشنهادی بخش آموزش نقش جمع‌آوری ترافیک ورودی را بر عهده داشته و در تمام مؤلفه‌های نصب شده بر روی میزبان‌های تحت نظارت وجود دارد. مؤلفه‌های نصب شده بر روی میزبان‌های تحت نظارت هر کدام بخشی از ترافیک ورودی را دریافت می‌کند. سپس هر کدام از مؤلفه‌ها، پیش پردازشی بر روی مجموعه ترافیک دریافتی انجام می‌دهند.

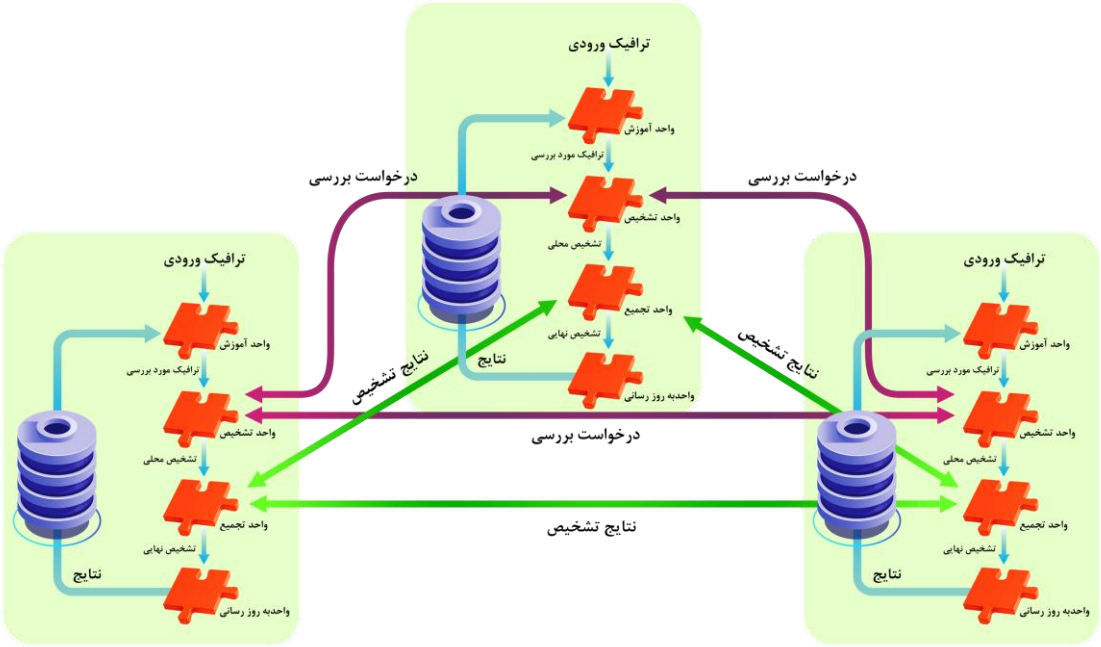
مدل پیشنهادی دارای تعدادی عامل تشخیص نفوذ است و برای تشخیص یک نفوذ به صورت بهینه با یکدیگر همکاری می‌نمایند. در مرحله ابتدایی هرکدام از همکاران موتور تشخیص محلی را با توجه به دانش خود ایجاد می‌نمایند. سپس در مراحل بعد برای تشخیص همکاری کرده و براساس نتایج حاصل موتور تشخیص خود را اصلاح می‌نمایند.

۳-۲- دسته‌بند بیز افزایشی همکارانه

سیستم پیشنهادی یک سیستم تشخیص نفوذ همکارانه مبتنی بر بیز می‌باشد. در پیاده‌سازی تعداد موتورهای تشخیص را به صورت پیش‌فرض ۵ عدد در نظر گرفتیم که هر کدام از این موتورهای تشخیص خود به صورت جداگانه بخشی از ترافیک ورودی را جمع‌آوری می‌کنند. در حقیقت ترافیکی که دریافت می‌کنند با بقیه متفاوت است. سپس در مرحله ابتدایی هرکدام از همکاران موتور تشخیص محلی را با توجه به دانش خود ایجاد می‌نمایند. سپس در مراحل بعد برای تشخیص همکاری کرده و براساس نتایج حاصل موتور تشخیص خود را اصلاح می‌نمایند.

بعد از مرحله ایجاد موتور تشخیص نوبت به بررسی ترافیک ورودی جدید می‌رسد. هر کدام از موتورهای تشخیص به طور مستقل ترافیک ورودی جدید را مورد بررسی قرار می‌دهند. به عبارتی هر کدام از همکاران با استفاده از موتور محلی خود نوع ترافیک را تشخیص می‌دهند و درخواست بررسی این ترافیک را برای سایرین ارسال می‌کنند به این دلیل که از سایر سیستم‌های

موجود برای تشخیص نهایی مشاهدات جدید کمک خواسته می‌شود و هر کدام از سیستم‌ها با توجه به واحد آموزش خود نوع مشاهده جدید را تشخیص می‌دهند و نتایج تشخیص محلی هر موتور تشخیص برای سایرین فرستاده می‌شود. این نتایج وارد واحد تجمیع هر کدام از همکاران می‌شود. سپس با توجه به نتایج رأی‌گیری می‌شود و نوع ترافیک جدید مبتنی بر نفوذ یا نرمال‌بودن آن براساس نتیجه رأی‌گیری مشخص می‌شود. سپس هر کدام از موتورهای تشخیص با در نظر گرفتن نتایج نهایی به دست آمده به‌هنگام می‌شوند و هر موتور تشخیص با استفاده از همین مشاهدات جدید که نوع آن‌ها مشخص شده است مجدداً به تعداد داده‌های آموزش خود اضافه می‌کند و هر بار که ترافیک جدیدی مشاهده شد و نوع آن با همکاری سایر موتورهای تشخیص مشخص شد این ترافیک به همراه نوع آن مجموعه داده آموزشی موتورها اضافه می‌شود. معماری کلی سیستم پیشنهادی در شکل (۲) نشان داده شده است.



شکل (۲): معماری سیستم تشخیص نفوذ همکارانه مبتنی بر دسته‌بند بیزی

۳-۳- تشریح الگوریتم بیز افزایشی

این الگوریتم مجموعه داده‌های تست را تحت عنوان $(t_1,t_2,\ldots,t_n)$ گرفته به‌گونه‌ای که هر یک از t_iها نمایانگر یکی از داده‌های مجموعه تست NSL-KDD بوده سپس با استفاده از روش دسته‌بند بیز [۱۶] هر نمونه محاسبه می‌شود. در نهایت داده‌های تست را به مجموعه آموزشی دسته‌بند بیز اضافه می‌کنیم (شکل (۳)).

برای اولین داده تست، احتمالات به روش معمولی محاسبه می‌شوند. اما برای داده‌های تست بعدی، چون داده تست قبلاً به مجموعه داده آموزشی اضافه شده است، احتمالات با استفاده از فرمول ارائه شده محاسبه می‌شود. در این فرمول چون یکی به تعداد کل داده‌ها اضافه شده $t+1$ در مخرج کسر گذاشته شده است. و $n^{t+1}(CK)$ بدان معنا است که ممکن است داده تست اضافه شده متعلق به کلاس CK باشد یا نباشد که در نتیجه آن به ترتیب $n^{t+1}(CK)= n^t (CK)+1$ و یا $n^{t+1}(CK)= n^t (CK)$ خواهد شد. همین‌طور در محاسبه $P^{t+1}(CK)$ هم باید این مسأله در نظر گرفته شود که آیا داده تست اضافه شده در کلاس CK قرار داشته یا خیر که این مسأله در مقدار $P^{t+1}(CK)$ مؤثر خواهد بود.

ورودی: مجموعه داده‌های آموزشی خروجی: دقت

شروع الگوریتم

۱. آموزش دسته‌بند بیز با استفاده مجموعه داده آموزشی NSL-KDD

۲. به ازای هر گروه از داده‌های تست مراحل زیر تکرار می‌شود:

الف) با استفاده از دسته‌بند بیز احتمال تعلق داده تست به هر کدام از کلاس‌ها تعیین شده و کلاسی که بالاترین احتمال را دارد به عنوان کلاس آن داده تست در نظر گرفته می‌شود.

$$P(C_K \mid a) = \arg_{c\in C} \max P(C_K)\prod_{i=1}^{n} P(a_i \mid C_K)$$

ب) با استفاده از نتایج مرحله الف، مجموعه داده تست به عنوان داده آموزشی به بیز اعمال می‌شود:

$$P^t(C_K) = \frac{n^t(C_K)}{t}$$

$$P^{t+1}(C_K) = \frac{n^{t+1}(C_K)}{t+1}$$

$$P^t(C_K \mid a) = \arg_{c\in C} \max P^t(C_K)\prod_{i=1}^{n} P(a_i \mid C_K)$$

$$P^t(C_K \mid a) = \arg_{c\in C} \max \frac{n^t(C_K)}{t} \prod_{i=1}^{n} \frac{P^t(a_i \cap C_k)}{P^t(C_K)}$$

$$n^{t+1}(C_K) = \begin{cases} n^t(C_K) & t+1 \notin C_k \\ n^t(C_K)+1 & t+1 \in C_k \end{cases}$$

$$P^{t+1}(C_K) = \begin{cases} \dfrac{P^t(C_K).t}{t+1} & t+1 \notin C_k \\ \dfrac{P^t(C_K).t+1}{t+1} & t+1 \in C_k \end{cases}$$

$$P^{t+1}(C_K \mid a) = \arg_{c\in C} \max \frac{n^{t+1}(C_K)}{t+1} \prod_{i=1}^{n} \frac{P^{t+1}(a_i \cap C_k)}{P^{t+1}(C_K)}$$

ج) محاسبه دقت

پایان الگوریتم

شکل (۳): الگوریتم دسته‌بند بیز افزایشی

٤- ارزیابی نتایج

٤-١- مجموعه داده NSL-KDD

برای ارزیابی روش پیشنهادی در حوزه تشخیص نفوذ به شبکه‌های کامپیوتری از مجموعه داده NSL-KDD، استفاده نموده‌ایم. این مجموعه داده برای حل برخی از مشکلات ذاتی مجموعه داده KDD'99 مانند حل مشکل افزونگی، پیشنهاد شده است [۱۰]. در جدول (۱)، ویژگی‌های این مجموعه داده ارائه شده است.

جدول (۱) : تعداد نمونه‌های مجموعه داده‌ی تشخیص نفوذ

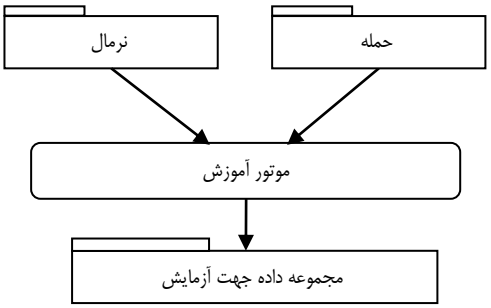
کلاس‌ها	NSL-KDD Train+	NSL-KDD Test+
Normal	۶۷۳۴۳	۹۷۱۰
DoS	۴۵۹۲۷	۷۴۵۸
U2R	۵۲	۶۷
R2L	۹۹۵	۲۸۸۷
Probe	۱۱۶۵۶	۲۴۲۲
مجموع	۱۲۵۹۷۳	۲۲۵۴۴

٤-١-١- داده آموزشی (Training data)

ما از مجموعه داده NSL-KDD که شامل ۱۲۵۹۷۳ رکورد است استفاده کردیم که می‌تواند برای آموزش موتور آموزش مورد استفاده قرار گیرد. این رکوردهای آموزشی شامل رکوردهای نرمال (غیر حمله) و رکوردهای حمله شناخته شده در میان چهار نوع حملات توزیع شده است: Dos[۱۷]، Probe، R2L[۱۸] و U2R[۱۹].

یک جستجوی در زمینه روش‌هایی انتخاب ویژگی که فقط بر روی دسته‌بند بیز ساده انجام می‌شود صورت گرفته است. نتیجه این جستجو این شد که بهترین روش انتخاب ویژگی بر روی دسته‌بند بیز مشخص شود و طبق نتایج کار نواکویک [۱۷] یکی از روش‌های انتخاب ویژگی One_R می‌باشد.

این الگوریتم یک درخت تصمیم یک سطحی ایجاد می‌کند و شامل مجموعه‌ای از قوانین است که همگی یک صفت خاص را می‌آزمایند One_R این صفت را مبنای تصمیماتش قرار می‌دهد و مجموعه قوانینی را انتخاب می‌کند که بهتر عمل کرده و کمترین نرخ خطا را دارند. الگوریتم One_R را بر روی دسته‌بند بیز به تعداد ویژگی‌ها اجرا کردیم سپس نتایج دقت آن را مورد بررسی قرار دادیم که در نهایت دسته‌بند بیز با ۲۹ ویژگی بیشترین دقت برای ۴۰٪ از داده‌های آموزشی را در بین تمامی حالات داشت همچنین جهت آزمایش مجموعه داده ما مقدار Fold را برابر ۱۰ قرار می دهیم تا موتور آموزش بتواند به نحوه بهتری آموزش ببیند. در شکل (۳) بلوک دیاگرام برای سیستم IDS ساخته شده را نشان می‌دهد. برای هر رکورد ورودی یک برچسب که شرح دهنده نوع اتصال است وجود دارد که هر اتصال شامل وضعیت نرمال یا حمله می‌باشد.



شکل (۳) : ساختار موتور آموزش

٤-١-٢- داده آزمایشی (Test data)

در این روش بعد از اینکه مدل آموزش داده می‌شود نیاز به داده‌هایی جهت آموزش می‌باشد که معمولاً از مجموعه داده اصلی داده‌هایی جهت آزمایش انتخاب می‌شوند. در این مدل ۱۰٪ از داده‌ها را جهت آزمایش انتخاب کردیم.

٤-٢- معیارهای ارزیابی

در این مقاله جهت ارزیابی روش پیشنهادی از پنج معیار، دقت، F-Measure، درصد قابلیت اعتماد به خروجی)، درصد موفقیت الگوریتم در تشخیص، نرخ تشخیص درست حملات و نرخ هشدار غلط بهره برده شده است که فرمول این پنج معیار به شرح زیر می‌باشد [۱۸]:

TN[۲۰]: درصد رکوردهای نرمال که به درستی طبقه‌بندی شده است.

TP[۲۱]: درصد رکورد حمله که به درستی طبقه‌بندی شده است.

FP[۲۲]: درصد رکوردهای که به اشتباه به عنوان حمله قرار گرفتند در حالی که در واقع آن‌ها فعالیت معتبر هستند.

FN[۲۳]: درصد رکوردهای که به اشتباه به عنوان فعالیت نرمال قرار گرفتند در حالی که در واقع آن‌ها حمله هستند.

$$ACC = \frac{TN + TP}{FP + FN + TP + TN} \qquad (۴)$$

$$F-Measure = 2 * \left(\left(\frac{1}{Precision}\right) + \left(\frac{1}{Recall}\right)\right) \quad (۵)$$

$$precision = \frac{TP}{FP+TP} \times 100 \quad (۶)$$

$$Recall = \frac{TP}{FP+TP} \times 100 \quad (۷)$$

$$DR = \frac{Total\_detected\_attacks}{Total\_attacks} \times 100 \quad (۸)$$

$$FPR = \frac{Total\_misclassifed\_process}{Total\_normal\_process} \times 100 \quad (۹)$$

٤-٣- نتایج حاصل از آزمایش صورت گرفته برای مقایسه بیز ساده با بیز افزایشی

در قسمت پیش پردازش داده با استفاده از انتخاب ویژگی تنها ۲۹ ویژگی را با روش One-R [۱۷] انتخاب کردیم و ارزیابی را انجام دادیم. نتایج آزمایش‌ها نشان می‌دهد دقت تشخیص با افزایش داده‌های ورودی افزایش می‌یابد. در پایان آموزش، روش پیشنهاد شده در این مقاله (INB[۲۴]) به لحاظ دقت نسبت به سیستم‌های تشخیص نفوذ مبتنی بر بیز ساده (NB[۲۵]) از ۹۰.۸۹ به ۹۵.۴۵ افزایش یافته است.

جدول (۲) : ارزیابی معیارهای معرفی شده

روش‌ها	Precision	F- measure	ACC	FPR	DR
NB	۹۲.۳۳	۰.۹۴	۹۰.۸۹	۰.۰۶	۰.۹۱۳
INB	۹۶.۷۲	۰.۹۷۴۶	۹۵.۴۵	۰.۰۴	۰.۹۵۱۶

٤-٤- مقایسه سیستم پیشنهادی با سیستم‌های موجود

در ادامه برای ارزیابی هر چه بیشتر رویکرد فوق، سیستم پیشنهادی با روش مقاله [۱۹] که نتایج خود را بر روی داده‌های NSL-KDD ارائه کرده‌اند مقایسه شده است. از نتایج به‌دست آمده می‌توان استنباط نمود که سیستم مورد پیشنهادی کارایی خوبی در تشخیص نفوذ در شبکه داشته و نرخ تشخیص و نرخ هشدارهای نادرست قابل قبولی نیز ارائه می‌نماید.

سه مجموعه داده بنام مجموعه داده ۱، مجموعه داده ۲، و مجموعه داده ۳ در آزمایش‌ها استفاده می‌شود. سه مجموعه داده از مجموعه داده NSL-KDD مشتق و پیش پردازش شده است. مجموعه داده ۱ و مجموعه داده ۲ توسط SNB[۲۶] استفاده شده در حالی که مجموعه داده ۳ توسط SSNB[۲۷] استفاده می‌شود. توزیع کلاس مجموعه داده ۱ شامل ۸۸۰۰ نمونه به صورت تصادفی انتخاب شده است که در جدول (۳) در ستون دوم نشان داده شده است و مجموعه داده ۲ شامل ۱۷۶۰۰ نمونه به طور تصادفی انتخاب شده است که در جدول (۳) در ستون سوم نشان داده شده است. توزیع مجموعه داده ۳ شامل ۸۸۰۰ نمونه به صورت تصادفی انتخاب شده از مجموعه داده NSL-KDD خوشه‌بندی شده است که در جدول (۳) در ستون چهارم نشان داده شده است. روش اعتبار سنجی متقابل برای آموزش و تست هر دو SNB و SSNB استفاده می‌شود. SNB برای اولین بار با استفاده از مجموعه داده ۱ شامل ۸۸۰۰ نمونه آموزش داده شده است و نتایج در جدول (۴) مقداردهی شده است. سپس SNB با دو برابر مقدار نمونه از SSNB یعنی ۱۷۶۰۰ آموزش داده است. نتایج در جدول (۵) مقداردهی شده است. کلاس‌های Normal، Probe، DOS، U2R و R2L برچسب به ترتیب ۰، ۱، ۲، ۳ و ۴ داده می‌شود. در جدول (۳)، نمونه‌ها به صورت تصادفی انتخاب شده‌اند در ستون چهارم جدول، مجموعه داده NSL-KDD با استفاده از الگوریتم خوشه‌بندی K-means به ۵ خوشه، خوشه‌بندی شده است. خوشه‌ها برای تشکیل مجموعه داده ۳ نمونه‌بندی شده است. SSNB با استفاده از مجموعه داده ۳ آموزش و تست داده شده است. نتیجه به دست آمده در جدول (۶) ذکر شده است. هر کدام از مجموعه داده‌ها با بیز افزایشی همکارانه آزمایش شده است و نتایج ارزیابی در جداول نمایش داده شده است.

جدول (۳): توزیع کلاس در مجموعه داده ها

کلاس	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳
۰	۲۶۰۰	۶۲۴۰	۳۵۲۰
۱	۲۶۰۰	۴۱۰۷	۱۴۲۴
۲	۲۶۰۰	۶۲۴۰	۳۷۹۲
۳	۵۲	۵۲	۵
۴	۹۴۸	۹۶۱	۵۹

جدول (٤): نتایج دسته‌بندی برای SNB با استفاده از ۸۸۰۰ نمونه

معیارها	روش‌ها	Normal	Dos	Probe	U2R	R2L
FPR	[۱۹]	۰.۰۱	۰.۰۵۷	۰.۰۰۶	۰.۰۹۶	۰.۰۰۲
DR	[۱۹]	۰.۸۸۹	۰.۹۱۵	۰.۹۳۵	۰.۹۲۳	۰.۳۶۵
F-measure	[۱۹]	۰.۹۳۲	۰.۸۹۳	۰.۹۶	۰.۱۰۳	۰.۵۲۹
FPR	INB	۰.۰۰۹	۰.۰۴۱	۰.۰۰۴	۰.۰۷	۰.۰۰۱
DR	INB	۰.۹۱	۰.۹۴۳	۰.۹۶۸	۰.۹۴۹	۰.۸۱۲
F-measure	INB	۰.۹۵۶	.۰.۹۶۷	۰.۹۸	۰.۹۵۱	۰.۸۴

جدول (٥): نتایج دسته‌بندی برای SNB با استفاده از ۱۷۶۰۰ نمونه

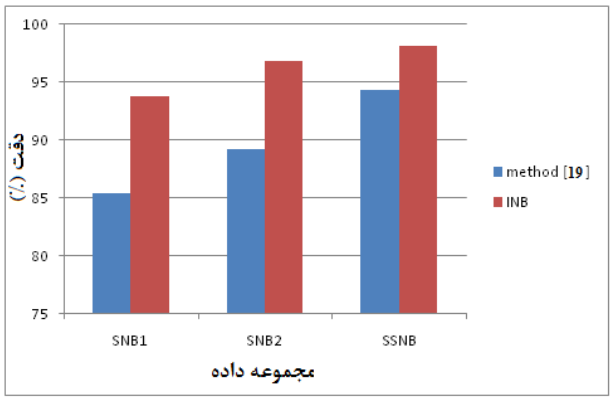
معیارها	روش‌ها	Normal	Dos	Probe	U2R	R2L
FPR	[۱۹]	۰.۰۱	۰.۰۴۳	۰.۰۰۷	۰.۰۶۲	۰.۰۰۲
DR	[۱۹]	۰.۹۱۳	۰.۹۱۱	۰.۹۴	۰.۸۸۵	۰.۳۷۷
F-measure	[۱۹]	۰.۹۴۵	۰.۸۸۸	۰.۹۶۳	۰.۰۷۷	۰.۵۳۵
FPR	INB	۰.۰۰۸	۰.۰۲۳	۰.۰۰۳	۰.۰۳	۰.۰۰۱
DR	INB	۰.۹۳۳	۰.۹۳۵	۰.۹۶۹	۰.۹۲	۰.۸۵
F-measure	INB	۰.۹۶۹	۰.۹۵	۰.۹۸۵	۰.۹۷۱	۰.۸۹

جدول (٦): نتایج دسته‌بندی برای SSNB

معیارها	روش‌ها	Normal	Dos	Probe	U2R	R2L
FPR	[۱۹]	۰.۰۱۲	۰.۰۴۳	۰.۰۱۲	۰.۰۰۵	۰.۰۰۱
DR	[۱۹]	۰.۹۳۸	۰.۹۷۸	۰.۹۴	۰.۶	۰.۷۶۳
F-measure	[۱۹]	۰.۹۵۹	۰.۸۸۹	۰.۹۶۱	۰.۱۲	۰.۷۸۳
FPR	INB	۰.۰۱	۰.۰۳	۰.۰۱	۰.۰۰۱	۰.۰۰۱
DR	INB	۰.۹۵	۰.۹۶	۰.۹۷۱	۰.۹۲۵	۰.۹۱
F-measure	INB	۰.۹۷	۰.۹۵۲	۰.۹۸۹	۰.۹۸	۰.۹۲۲

دقت INB پیشنهادی موقعی که از مجموعه داده ۱ و مجموعه داده ۲ استفاده می‌کند، به ترتیب ۹۳.۷۸٪ و ۹۶.۸۱٪ می‌باشد. دقت درSSNB با

توجه به اینکه تعداد داده‌ها کمتر است ولی دقت افزایش پیدا کرده و معادل ۹۸.۱۸ می‌باشد. که در شکل (۴) نمودار دقت روش پیشنهادی ما با روش پیشنهادی در مرجع [۱۹] مقایسه شده است و نشان می‌دهد که دقت روش ما با استفاده از سه مجموعه داده بیشتر از روش مذکور می‌باشد.



شکل (٤): مقایسه دقت روش پیشنهادی با روش پیشنهادی در مرجع [۱۹]

۵- نتیجه گیری

تشخیص نفوذ یک بخش حساس جهت برقراری امنیت در سیستم‌های کامپیوتری است. در سیستم‌های تشخیص نفوذ، IDS ها سعی دارند نفوذهای غیر مجاز به شبکه را با توجه الگوریتم‌های خاص تشخیص دهند و می‌توان آن‌ها را به دو دسته تشخیص سواستفاده و غیرعادی تقسیم کرد. در این مقاله یک سیستم تشخیص نفوذ همکارانه مبتنی بر دسته‌بند بیزی برای بهبود مشکلات موجود در این سیستم‌ها توسعه داده شده است. نوآوری طرح پیشنهادی آموزش افزایشی دسته‌بند بیزی و یک سیستم تشخیص نفوذ همکارانه مبتنی بر دسته‌بند بیزی می‌باشد. در قسمت پیش پردازش داده با استفاده از انتخاب ویژگی تنها ۲۹ ویژگی را انتخاب کردیم و ارزیابی را انجام دادیم. در مقایسه با بیز ساده، نتایج دقت نشان می‌دهد که هر بار که داده‌ها را افزایش می‌دهیم دقت بیز ساده افزایش پیدا می‌کند. نتایج ارزیابی با سیستم‌های موجود نشان می‌دهد که سیستم پیشنهادی دارای دقت بیشتر و تعداد هشدارهای نادرست کمتری می‌باشد.

مراجع

[1] Aljawarneh, S., Aldwairi, M. and Yassein, M.B., "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model", Journal of Computational Science, Vol. 25, pp.152-160, 2018.

[2] Vert, G., Claesson-Vert, A. L., Roberts, J., and Bott, E. "A Technology for Detection of Advanced Persistent Threat in Networks and Systems Using a Finite Angular State Velocity Machine and Vector Mathematics", In Computer and Network Security Essentials, pp. 41-64, 2018.

[3] Grill, M., Pevný, T. and Rehak, M., "Reducing false positives of network anomaly detection by local adaptive multivariate smoothing", Journal of Computer and System Sciences, Vol. 83, pp.43-57, 2017.

[4] Hassan, N.A. and Hijazi, R., "Introduction to Online Threats and Countermeasures." In Open Source Intelligence Methods and Tools, pp. 21-94, 2018.

[5] Haque, M.R., Tan, S.C., Yusoff, Z., Lee, C.K. and Kaspin, R., "DDoS Attack Monitoring using Smart Controller Placement in Software Defined Networking Architecture." In Computational Science and Technology, pp. 195-203, 2019.

[6] Meng, W., Tischhauser, E.W., Wang, Q., Wang, Y. and Han, J., "When intrusion detection meets blockchain technology: a review", Ieee Access, Vol. 6, pp.10179-10188, 2018.

[7] Fung, C.J., Zhu, Q., Boutaba, R. and Başar, T., "Bayesian decision aggregation in collaborative intrusion detection networks", In Network Operations and Management Symposium (NOMS), pp. 349-356, 2010.

[8] Liu, M., Xue, Z., Xu, X., Zhong, C. and Chen, J., "Host-Based Intrusion Detection System with System Calls: Review and Future Trends", ACM Computing Surveys (CSUR), Vol. 51, p.98, 2018.

[9] Erbacher, R.F. and Hutchinson, S., "Distributed sensor objects for intrusion detection systems", In Information Technology: New Generations (ITNG), Ninth International Conference on, pp. 417-424, 2012.

[10] Gunawan, L.A., Vogel, M., Kraemer, F.A., Schmerl, S., Slåtten, V., Herrmann, P. and König, H., "Modeling a distributed intrusion detection system using collaborative building blocks", ACM SIGSOFT Software Engineering Notes, Vol. 36, pp.1-8, 2011.

[11] Uddin, M., Rahman, A.A., Uddin, N., Memon, J., Alsaqour, R.A. and Kazi, S., "Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents", IJ Network Security, Vol. 15, pp.97-105, 2013.

[12] Sasikumar, R. and Manjula, D., "Dynamic Distributed Intrusion Detection System Based on Mobile Agents with Fault Tolerance", Journal of Computer Science, Vol. 8, p.1092, 2012.

[13] Brahmi, I., Yahia, S.B. and Poncelet, P., "MAD-IDS: novel intrusion detection system using mobile agents and data mining approaches", In Pacific-Asia Workshop on Intelligence and Security Informatics, pp. 73-76, 2010.

[14] Arshad, J., Abdellatif, M.M., Khan, M.M. and Azad, M.A., "A novel framework for collaborative intrusion detection for M2M networks", In 9th International Conference on Information and Communication Systems (ICICS), pp. 12-17, 2018.

[15] Teng, S., Wu, N., Zhu, H., Teng, L. and Zhang, W., "SVM-DT-based adaptive and collaborative intrusion detection", IEEE/CAA Journal of Automatica Sinica, Vol. 5, pp.108-118, 2018.

[16] Al-Aidaroos, K.M., Bakar, A.A. and Othman, Z., "Naive Bayes variants in classification learning." In Information Retrieval & Knowledge Management, (CAMP), International Conference on, pp. 276-281, 2010.

[17] Novakovic, J., "The impact of feature selection on the accuracy of naïve bayes classifier", In 18th Telecommunications forum TELFOR, Vol. 2, pp. 1113-1116, 2010.

[18] Muda, Z., Yassin, W., Sulaiman, M. N., Udzir, N. I., "K-Means Clustering and Naive Bayes Classification for Intrusion Detection", Journal of IT in Asia, 2014.

[19] Subramanian, U. and Ong, H.S., "Analysis of the Effect of Clustering the Training Data in Naive Bayes Classifier for Anomaly Network Intrusion Detection," Journal of Advances in Computer Networks, Vol. 2, 2014.

پانویس ها

1 Distributed Denial Of Service
2 Intrusion Detection System
3 A Host-based Intrusion Detection System
4 Network Intrusion Detection System
5 Collaborative Intrusion Detection Network
6 Collaborative Intrusion Detection System
7 Signature-based Multi-Layer Distributed Intrusion Detection System
8 Mobile Agent using Data mining based Intrusion Detection System
9 COLlaborative Intrusion Detection Engine
10 Machine to Machine
11 Support Vector Machines
12 Collaborative and Adaptive Intrusion Detection Model
13 Environments Classes, Agents, Roles, Groups, and Objects
14 Transmission Control Protocol
15 User Datagram Protocol
16 Internet Control Message Protocol
17 Denial Of Service
18 Remote to User
19 User-To-Root
20 True Negative
21 True Positive
22 False Positive
23 False Negative
24 Incremental Naïve Bayes
25 Naïve Bayes
26 Supervised Naive Bayes
27 semi supervised Naive Bayes