ISSN: 2745-3673
http://www.Rcsj.ir

مجله‌ی علمی پژوهش در علوم رایانه (سال چهارم)
شماره ۱۴ / تابستان ۱۳۹۸ / ص ۶۷-۵۸

تشخیص حملات ترافیک انتقالی شبکه با استفاده از منطق فازی مبتنی بر گروه مدل مخفی مارکوف

سیده معصومه موسوی[۱]، سیده طاهره موسوی [۲]

[۱] کارشناسی ارشد، گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد علوم تحقیقات، اهواز، ایران.

[۲] دانشجوی کاردانی، گروه فناوری اطلاعات، دانشکده فنی و مهندسی، علمی کاربردی جهاد دانشگاهی، اهواز، ایران.

نام نویسنده مسئول:
سیده معصومه موسوی

چکیده

با توجه به رشد روزافزون حجم تبادل اطلاعات از طریق شبکه‌های کامپیوتری، مساله امنیت در ارتباطات کامپیوتری، تبدیل به یک رکن اساسی در طراحی شبکه‌های ارتباطی شده است و با توجه به رشد تصاعدی فعالیت‌های غیرمجاز در شبکه، نیاز به راهکارهای مفید و موثر در مقابله با این نوع فعالیت‌ها وجود دارد. حمله‌های صفر روزه[۱] یکی از خطرناک‌ترین تهدیدهای هستند که کامپیوترهای شبکه را تهدید می‌کنند و در معنای لغوی به حمله‌هایی گفته می‌شود که تا به حال توسط سیستم شناخته نشده‌اند، بنابراین ابزارهای دفاعی که مبتنی بر یک سری قوانین می‌باشند درمقابل حملات صفرروزه ناتوان هستند. اخیرا ابزارهای دفاعی مبتنی ناهنجاری با الگوریتم‌های یادگیری ماشین برای شناسایی این حملات استفاده می‌شوند. با توجه به اینکه این روش‌ها حمله‌های صفر روزه را تا حد قابل قبولی خنثی کرده‌اند، از محبوبیت خوبی برخوردار شده‌اند. در این سیستم ها مدلی بر پایه داده های آماری از فعالیت شبکه ساخته می شود. چنانچه در هر لحظه بار ترافیکی شبکه از مرزی که بین فعالیت های عادی و غیر عادی توسط سیستم مشخص شده تخطی کند، سیستم هشداری مبنی بر وقوع حمله میدهد. در این مقاله یک سیستم تشخیص رفتار غیرعادی با استفاده از منطق فازی و طبقه بندی کننده چندگانه و زنجیره مارکوف ارائه شده است. نتایج آزمایشگاهی نشان میدهد که الگوریتم پیشنهادی نسب به سایر روش های قابل مقایسه عملکرد خوبی را از خود نشان داده است.

واژگان کلیدی: حملات شبکه، منطق فازی، تشخیص ناهنجاری، مدل مارکوف.

[1] Zero day attack

مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۱۴، تابستان ۱۳۹۸، ص ۶۷- ۵۸

ISSN: 2745-3673
http://www.Rcsj.ir

مقدمه

سیستم های تشخیص نفوذ وظیفه ی شناسایی و تشخیص هرگونه استفاده ی غیرمجاز به سیستم، سوء استفاده و یا آسیب رسانی توسط هر دو دسته ی کاربران داخلی و خارجی را بر عهده دارند. تشخیص و جلوگیری از نفوذ امروزه به عنوان یکی از مکانیزم های اصلی در برآوردن امنیت شبکه ها و سیستم های رایانه ای مطرح است و عموما در کنار دیواره های آتش و به صورت مکمل امنیتی برای آنها مورد استفاده قرار میگیرند. سامانه های تشخیص نفوذ به صورت سامانه های نرم افزاری و سخت افزاری ایجاد شده و هر کدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزایای سیستم های سخت افزاری است عدم شکست امنیتی آن ها تسط مهاجمان، قابلیت دیگر این گونه سیستم ها می باشد. اما استفاده ی آسان از نرم افزار، قابلیت سازگاری در شرایط نرم افزاری و تفاوت سیستم های عامل مختلف، عمومیت بیشتری را به سامانه های نرم افزاری می دهد و عموما این گونه سیستم ها انتخاب مناسب تری هستند. به طور کلی سه عملکرد اصلی IDS عبارت است از: نظارت و ارزیابی، کشف و واکنش.

۱- تاریخچه

با افزایش سرعت، کارایی، تعداد و ارتباط کامپیوترها در دهه ۱۹۷۰، نیاز به سیستم های امنیتی رشد بسیاری پیدا کرد. در سال های ۱۹۷۷ و ۱۹۷۸ سازمان بین المللی استاندارد، جلسه ای را مابین دولت ها و ارگان های بازرسی EDP تشکیل داد که نتیجه آن جلسه، تهیه گزارشی از وضعیت امنیت، بازرسی و کنترل سیستم ها در آن زمان بود. در همین زمان وزارت نیروی کشور آمریکا به علت نگرانی اوضاع امنیتی سیستمهای خود، تحقیق بسیار دقیقی را مورد بازرسی و امنیت سیستم های کامپیوتری شروع کرد. این کار توسط فردی به نام جیمز آندرسون انجام شد. آندرسون اولین فردی است که مقاله ای در رابطه با لزوم بازرسی خودکار امنیت سیستم ها ارائه داد. گزارش آندرسون که در سال ۱۹۸۰ تهیه شده را میتوان به عنوان هسته اولیه مفاهیم تشخیص نفوذ معرفی کرد. در این گزارش مکانیزم هایی برای بازرسی امنیت سیستم ها معرفی شد و همچنین مشخص شده است که در صورت بروز خرابی در سیستم چگونه با آن مقابله شود. در سال های ۱۹۸۴ تا ۱۹۸۶ دوروتی دنینگ و پیتر نیومان تحقیقاتی در زمینه امنیت سیستم های کامپیوتری انجام دادند که نتیجه آن تولید یک سیستم تشخیص نفوذ به صورت بلادرنگ بود که براساس سیستم های خبره عمل می کرد. این سیستم IDES نامگذاری شد. در این پروژه ترکیبی از تشخیص ناهنجاری و تشخیص سوء استفاده مورد بررسی قرار گرفت. ایده مطرح شده در این پروژه به عنوان پایه خیلی از سیستم های تشخیص نفوذ که از آن به بعد ایجاد شدند مورد استفاده قرار گرفت. گزارش اندرسون و تحقیقاتی که بر روی پروژه IDES صورت گرفت، شروع کننده ی زنجیره ای از تحقیقات در رابطه با سیستم های تشخیص نفوذ بودند.

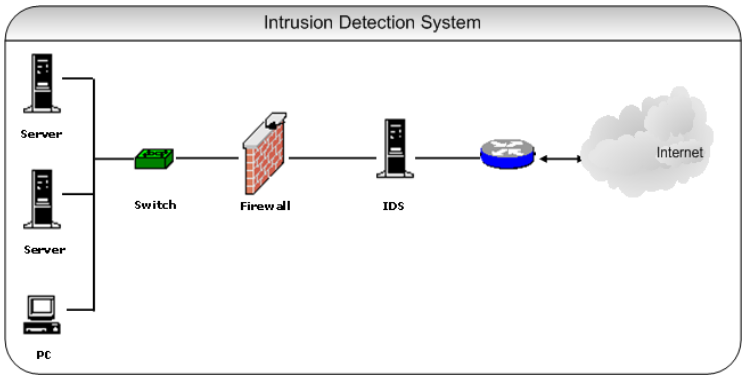
۲- رویکردهای تحلیل

در فرآیندهای تشخیص نفوذ بعد از معرفی منابع اطلاعات و مشخص شدن نوع دسته بندی آنها، نیاز بعدی تعیین آنالیزگر می باشد. در آنالیزگر، اطلاعات از منابع اطلاعات استخراج میشوند و با توجه به سیاست های امنیتی، انواع حملات و غیره مورد بررسی قرار میگیرند. در سیستم های تشخیص نفوذ، روش های آنالیز به دو دسته کلی تشخیص سو استفاده و تشخیص ناهنجاری و یا ترکیبی از آنها تقسیم می‌شوند:

- تشخیص سو استفاده: در این روش، آنالیزگر به دنبال نشانه ای میگردد که بیانگر یک عمل خلاف باشد. برای انجام این کار، ابتدا اطلاعات فیلتر می شوند تا الگوهایی که بیانگر نوع حمله و یا سایر سیاست های امنیتی باشد پیدا شوند.
- تشخیص ناهنجاری: در این روش آنالیزگر به دنبال موارد غیرمعمول میگردد. برای انجام این کار، اطلاعات جمع آوری شده بررسی میشوند تا الگوهایی که نشان دهنده اعمال غیرمعمول هستند پیدا شوند. در برخی موارد از این دو روش در کنار یکدیگر استفاده می کنند. در این سیستم ها، روش تشخیص ناهنجاری وظیفه تشخیص حملات جدید و ناشناخته را دارد و تشخیص سوءاستفاده وظیفه حفاظت تشخیص ناهنجاری را بر عهده میگیرد.

با این کار این تضمین به وجود می آید که اطلاعات و الگوهای جمع آوری شده برای تشخیص ناهنجاری امن باشد. در شکل(۱) دیاگرامی از سیستمی که از دو روش استفاده می کند نشان داده شده است.

ISSN: 2745-3673
http://www.Rcsj.ir



شکل۱: نمای کلی سیستم های تشخیص نفوذ

یکی از مواردی که در رابطه با تحلیل داده ها مطرح است، زمانبندی می باشد. آنالیز داده ها میتواند به دو صورت بلادرنگ و مد دسته‌ای باشد.

مد دسته‌ای: منظور از آنالیز مد دسته ای این است که اطلاعات مربوط به یک دوره زمانی جمع آوری می شوند و سپس به آنالیزگر داده میشوند. استفاده از این نوع زمانبندی در سیستم های قدیمی استفاده میشده است. به علت اینکه پهنای باند ارتباطی و قدرت پردازشی در سیستم های قدیمی به حدی نبوده است که سیستم ها بتوانند بصورت بلادرنگ عمل کنند.

بلادرنگ: با بالارفتن قدرت پردازشی و همچنین افزایش پهنای باند ارتباطی، اکثر سیستم های جدید از این روش استفاده میکنند. در این روش با هر رویدادی که رخ میدهد و یا در هر فاصله زمانی کوتاه، منبع اطلاعات به آنالیزگر داده می‌شود.

۳- کارهای گذشته

در این بخش از مقاله تاریخچه تعدادی از روش‌های مبتنی بر ناهنجاری که تا به حال بررسی شده‌اند را شرح خواهیم داد.در مدلی که توسط والدز و همکاران ارائه شده همبسته سازی در سه مرحله انجام میشود. مرحله اول تجمیع[۲] رویدادهای سطح پایین در قالب ریسمان حمله می باشد. هشدارهای براساس معیار شباهت بصورت خوشه های مشابه در می آیند. دیبر و وسپی، هر دو عمل تجمیع و همسبته سازی هشدارهای را در یک سیستم چند حسگر انجام می‌دهند. آن ها یک مدل معنایی از هشدارهای و یک ماژول تطبیق دهنده را پیاده سازی کرده اند.

پوراس و همکارانش تاثیر هشدارها بر ماموریت شبکه را مورد توجه قرار داده اند. یک پایگاه دانش از مشخصات شبکه ی مورد حفاظت برای اولویت بندی هشدارهای مورد استفاده قرار میگیرد، همچنین شکل ساده ای از تصدیق هشدارهای برای حذف هشدارهایی که سرویس مورد نظر آن ها وجود ندارد نیز در نظر گرفته شده است. در مدل M2D2 از روش تشریح رسمی توانایی های هر IDS استفاده شده این توانایی ها با نگاه به حوزه مورد نظارت و محل استقرار IDS بیان می شوند و به کمک آنها مثبت های کاذب به کمک رای گیری بین چندین IDS شناسایی میشوند. چونگ و همکارانش یک زبان به نام CAML برای مدال سازی حملات معرفی نمودند که در آن پیش شرط و پس شرط هر حمله بیان میشود. پیش شرط مجموعه ای از شرایط است که برای انجام یک حمله بایستی مهیا باشد، و پس شرط، پیامدهای آن حمله است. در HIGSAW نیز برای تشخیص روابط سببی از مفاهیم مشابهی استفاده شده است. در شرایط حملات بوسیله توانایی ها و مفاهیم بیان میشوند. نینگ و همکارانش با تعریف مفاهیم پیش نیاز و پیامد روابط سببی بین هشدارهای را شناسایی می کنند. آن ها به کمک این مفاهیم گرافی از هشدارهای مرتبط را ساخته و به کمک الگوریتم های دستکاری گراف ها اندازه ی گراف را کم میکنند.
کین و همکارانش یک موتور همبسته سازی بیزین را برای استخراج روابط آماری پیشنهاد کرده اند. آنها فرض کرده اند که اگر یک همبستگی آماری قوی بین هشدارهای وجود داشته باشد آنگاه هشدارها با یکدیگر رابطه ی سببی دارند. میزان مربوط بودن هشدارهای از طریق محاسبه احتمال شرطی بین هر جفت فراهشدارها محاسبه میشود. آنها برای کم کردن بار محاسباتی سیستم آن را بصورت دو ماژول همبسته سازی آنلاین و آفلاین طراحی کرده اند.

ژو و همکارانش از شبکه عصبی چندلایه پرسپترون و همچنین ماشین بردار پشتیبان برای محاسبه احتمال همبستگی دو هشدار استفاده کرده اند. آنها با نگهداری ماتریسی از احتمالات وقوع هشدارهای مختلف به دنبال یکدیگر و بروز کردن ماتریس با هر همبستگی جدید، اطلاعات مور نیاز برای ورودی های شبکه عصبی و ماشین بردار پشتیبان خود را فراهم می‌کنند. Krugel سرویس‌های خاص

[2] aggregation

ISSN: 2745-3673
http://www.Rcsj.ir

تشخیص نفوذ را که ترکیبی از نوع، طول و چگونگی توزیع بایت‌های یک بسته ترافیک انتقالی بود را برای بدست آوردن سرویس های مورد نیاز یک مدل آماری در ترافیک نرمال توصیف می‌کرد [۲]. Netad ۴۸ بایت اول بسته ترافیک انتقالی را بررسی نموده است. او از تعدادی مدل های مجزای ساخته شده که مربوط به رایج‌ترین پروتکل‌های شبکه بوده‌اند در سیستم خود استفاده می‌کرد و همچنین به منظور کشف وقایع نادر، به هرکدام از بسته ها نمره ناهنجاری تخصیص می‌داد [۳]. در Payl، نفوذ ها از طریق آنالیز توزیعی بایت‌ها در HTTPPayl تشخیص داده می‌شده‌اند [۴]. نسخه بهبود یافته Payl در مقاله آمده است.به طور خاص این نسخه جدید تعدادی مدل برای هر طول بسته می‌سازد و ارتباط ترافیک ورودی و خروجی را از حیث تشخیص انتشار حملات(کرم ها) مطمئن می سازد [۴،۵].

یک راه حل به نام Anagram پیشنهاد داد که بر اساس مدل های n-gram استخراج شده از دو ترافیک نرمال و آلوده، ساخته شده-بود [۶]. Anagram همه n-gram های استخراج شده از ترافیک نرمال را ذخیره می‌کند، و فیلتر Bloom را به سیستم آموزش می دهد و همچنیـن این فیلترهای Bloom، n-gramهایی را که از بسته های مخرب استخراج شده اند را شناسایی و ذخیره می‌کنند. در زمان تشخیص، به بسته بر اساس تعداد n-gram های مخرب مشاهده نشده نمره تعلق می‌گیرد.تعدادی از n-gram های مخرب برای وزن‌دهی به نمره‌ها استفاده می‌شود. ولی مهمترین مشکل در این روش تشخیص تمایز بسته برای قرار گرفتن در کلاس خوب و یا بد بوده است.

Perdisciسیستم تشخیص نفوذی به نام McPAD ارائه داده است. در این مدل یک الگوریتم استخراج ویژگی پیاده‌سازی شده است که می‌تواند یک تقریب برای مدل n-gram در نظر گرفته شود. در مدل McPAD بسته‌های ترافیک انتقالی در ترافـیک نرمـال به صـورت 2-v-gram اسـتفاده می‌شـود کـه فرکانس نسبی جفت بایت ها از موقعیت ۰ تا v تغییر می کند و هر کدام موقعیت خود را از طریق دیگری پیدا می کند. مجموعه ترافیک انتقالی در v+۱ فضای ویژگی مختلف نشان داده می‌شوند و یک کلاس برای طبقه‌بندی هر فضای ویژگی آموزش داده شده است و هر کدام از طبقه‌بندی‌کننده‌ها بر اساس ویژگی‌های فضای خود آموزش می‌بینند [۷]. اخیرا مدل های مارکوف و مدل های مخفی مارکوف برای مدل سازی مسائل امنیتی کامپیوتری استفاده می‌شوند در حالی که قبلا از آنها فقط در برنامه‌هایی نظیر تشخیص‌گفتار، تشخیص دست خطها وتجزیه تحلـیل دنباله‌هـای زیسـت شـناسی اسـتفاده می شده است [۸،۹].

در زمینه امنیت کامپیوتر استفاده از مدل مخفی مارکوف در طرف سرور مرسوم شده است. HMM و n-gram از لحاظ ریشه تئوری یکی هستند .HMM می‌تواند به خوبی n-gram داده‌ها را به صورت حالات محدود نمایش دهد [۱۳،۱۲،۱۱،۱۰]. عـلاوه بر اینـکه مـدل HMM از قـدرتـی یکـسان با مـدل n-gram برخوردار است، HMM در زمینه مدل کردن دنباله ها درمقایسه با n-gram ازمزایای زیادی برخوردار اسـت . n-gram از جهت میزان ویژگی‌هایی که می‌تواند مدل کند به 256^n محـدود می‌شـود، ایـن در حـالی اسـت که HMM می تواند هر طولی از دنباله ها را به خوبی و بدون محدودیت بدون تغییر در پیچیدگی محاسباتی پردازش کند.رعایت تعادل(Miao,2013) بین کنترل عملکرد سیستم و نرخ شناسایی حملات بازپخش نیاز به تعیین تدابیر امنیتی است به شکلی که هزینه کنترل شود. در این مقاله یک سیستم مبتنی بر بازی با هزینه بهینه و امن برای تشخیص حملات بازپخش ارائه شده است. در این مقاله فرض بر این شده است که مهاجم میتواند اندازه سنسور را ضبط کند، سای پنجره T را انتخاب کند و در هر زمان- مرحله تصمیم گرفته میشود که تحویل پیام اصلی به درستی ارسال شود.

سیستم های بهم پیوسته (Tan,2014) مانند سرورهای وب، سروهای پایگاه داده، سروهای محاسبات ابری و غیره تحت حملات مهاجمان قرار میگیرند. یکی از رایج ترین و بدترین این حملات DoS است که یک تهدید جدی برای سیستم های کامپیوتری است. در این مقاله یک سیستم تشخیص DoS با استفاده از تجزیه و تحلیل چند متغییره همبستگی با استخراج ویژگی های هندسی ترافیک شبکه ارائه شده است. روش پیشنهادی مبتنی بر MCA است و به تشخیص ناهنجاری های می پردازد و مبتنی بر ناهنجاری است. شبکه های موبایل(Sohail,2013) کاملا خود ساخت یافته نشان دهنده سیستم های توزیع شده پیچیده اند که ممکن است بخش اعظمی از یک سیستم پیچیده مانند سیستم مدیریت بحران باشند. با توجه به پیچیدگی ذاتی شبکه های همراه و محدودیت منابع، همواره توسعه و ارائه راه حل هایی امنیتی بسیار مورد نیاز می باشد. از آنجا که شبکه های موبایل نیاز به شناسایی مداوم، متمایز و منحصر به فرد هر گره در پروتکل های امنیتی هستند، حملات Sybil یک تهدید جدی برای آنها محسوب میشود. در این مقاله یک طرح سبک وزن برای شناسایی حملات Sybil بدون استفاده از اعتماد متمرکز یا هر سخت افزار اضافی دیگری مانند آنتن جهت دار یا سیستم موقعیت یاب جهانی ارائه شده است.

۴- سیستم پیشنهادی

روش پیشنهادی، بر اساس ساختار سیستم‌های تشخیص نفوذ مبتنی بر بی‌نظمی پیاده‌سازی شده است و از مدل مخفی مارکوف و استنتاج فازی برای حل مسئله تشخیص بی‌نظمی استفاده می‌کند. دو وظیفه‌ی مهم در تشخیص بی‌نظمی، مدلسازی یا دسته‌بندی الگوها و

مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۱۴، تابستان ۱۳۹۸، ص ۶۷- ۵۸

ISSN: 2745-3673
http://www.Rcsj.ir

تصمیم‌گیری می‌باشد که اگر به‌خوبی انجام شوند، آنگاه می‌توان عملکرد بهتری را در تشخیص بی‌نظمی مشاهده نمود. سیستم پیشنهادی دارای دو راهکار مهم می‌باشد که باعث می‌شوند، مدل‌سازی یا دسته‌بندی الگوها و تصمیم‌گیری به‌خوبی انجام شوند و درنتیجه نرخ خطا کاهش و نرخ تشخیص سیستم افزایش یابد.

۱،۴. راهکار اول

برای اینکه بتوان خطای ناشی از تعداد محدودی از دسته‌بندی کننده‌ها را در تصمیم‌گیری از بین برد، سیستم پیشنهادی از رأی اکثریت استفاده می‌کند. بدین ترتیب که با استفاده از دسته‌بندی کننده چندگانه برای هر ویژگی استخراج‌شده از یک ترافیک انتقالی، مدلی را تولید می‌کنیم که خروجی‌های به‌دست‌آمده از هر دسته‌بندی کننده درنهایت باهم ادغام می‌شوند و بدین ترتیب خطای برخی دسته‌بندی کننده نادیده گرفته می‌شود.

۲،۴. راهکار دوم

در اکثر کارهای گذشته برای اینکه بتوانند ترافیک انتقالی نرمال و غیر نرمال را تفکیک کنند، از یک مقدار آستانه استفاده می‌کردند که در حالتی که شباهت بسته‌های نرمال و غیر نرمال فراوان است سیستم دچار ضعف بوده است. برای اینکه بتوان نرمال و غیر نرمال بودن بسته‌های ترافیک انتقالی را در حالتی که بسته‌ها شباهت فراوانی باهم دارند تفکیک کرد، از استنتاج فازی استفاده کرده‌ایم. با توجه به دو راهکار ذکرشده، سیستم تشخیص نفوذ پیشنهادی به نرخ تشخیص بالایی می‌رسد درحالی‌که پایین‌ترین نرخ مثبت کاذب را ارائه می‌دهد.

۳،۴. سیستم تشخیص نفوذ پیشنهادی

به‌طورکلی، سیستم‌های تشخیص نفوذ مبتنی بر الگوریتم‌های یادگیری ماشین در دو فاز آموزش و آزمایش توسعه داده می‌شوند. در فاز آموزش، مدل‌ها و ماژول‌های سیستم با توجه به داده‌های آموزشی پیکربندی می‌شوند.. در مرحله آزمایش، سیستم ساخته‌شده با توجه به داده‌های عملیاتی نرمال و غیر نرمال ارزیابی می‌شود. برای هر ترافیک انتقالی در این سیستم چندین گروه HMM ساخته می‌شود. هر گروه HMM برای مدل‌سازی هر ویژگی استخراج‌شده از ترافیک انتقالی، آموزش داده می‌شود. مقدار احتمالی حاصل از HMM های درون یک گروه HMM ترکیب‌شده و خروجی نهایی موردنظر شکل می‌گیرد. سپس با فازی سازی خروجی هر گروه HMM، فرایند استنتاج با به‌کارگیری قوانین و مجموعه‌های فازی، نرمال و غیر نرمال بودن ترافیک انتقالی را مشخص می‌کند.

۴،۴. پیش‌پردازش داده‌ها و استخراج ویژگی‌ها

۱،۴،۴. مدل داده‌ای

ورودی سیستم پیشنهادی، مجموعه‌ای از بردارهای اتصال دریافت شده توسط سروری می‌باشد که ترافیک انتقالی را میزبانی می‌کند. جدول (۱) یک بردار اتصال ترافیک انتقالی را نشان می‌دهد که در پایگاه داده KDDCup وجود دارد. همان‌طور که دراین جدول مشاهده می‌کنید، این بردار درواقع شامل پارامترهای ارسال‌شده به سرور موردنظر می‌باشد. درنتیجه بردار را می‌توان به‌صورت $b=\{(a_1, v_1), (a_2, v_2), (a_n, v_n)\}$ نشان داد، که در آن $a_i \in A$ می‌باشد و A مجموعه همه‌ی ویژگی‌های مربوط به بردار اتصال می‌باشد و v_i مقدار آن‌ها می‌باشد. این ویژگی‌ها با توجه به ارزشی که دارند، نقش خود را در فرآیند تصمیم‌گیری ایفا می‌کنند. این پژوهش، بر روی ۵ ویژگی مهم ترافیک انتقالی KDD Cup 1999تمرکز دارد که در جدول (۱) نشان داده‌شده‌اند.

جدول۱ : ویژگی‌های استخراج‌شده از ترافیک انتقالی KDD Cup

خصیصه انتخاب‌شده	نوع	نام‌گذاری شده
src_bytes	پیوسته	Src
dst_bytes	پیوسته	Dst
logged_in	پیوسته	Log
dst_host_count	پیوسته	Hst
dst_host_srv_count	پیوسته	srv

ISSN: 2745-3673
http://www.Rcsj.ir

۲،۴،۴. آموزش

در این مرحله به یادگیری و ارزیابی مدل پیشنهادی می‌پردازیم. در مرحله بعدی تصمیم‌گیری انجام می‌شود. سیستم پیشنهادی از الگوریتم Baum-Welch که در (L.R. Rabiner,1989) آمده است برای آموزش HMM استفاده می‌کند. از آنجائی که طراحی یک HMM به آزمون‌وخطا بستگی دارد و یک امر تجربی است، با تغییر در پارامترهای تعداد حالات، حالت ابتدایی، ماتریس توزیع سمبل و ماتریس گذر حالت، می‌توان بهترین وضعیت آموزش را به دست آورد. زمانی که از تعداد حالات کمی استفاده می‌شود، دقت آموزش هر HMM کاهش می‌یابد ولی در عوض زمان کم‌تری صرف آموزش می‌شود، اما زمانی که از تعداد حالات بیشتری استفاده می‌شود، دقت آموزش هر HMM افزایش و زمان بیشتری صرف یادگیری می‌شود و این روند تا جایی ادامه پیدا می‌کند که سیستم در حالت تقریباً پایداری قرار می‌گیرید؛ بنابراین با توجه به اینکه دقت یادگیری بازمان یادگیری رابطه مستقیم دارند و برای اینکه بتوانیم یک تعادل در دقت و زمان برقرار کنیم، از تعداد ۱۵ حالت استفاده کرده‌ایم.

۵،۴. ترکیب خروجی‌های HMM

برای اینکه بتوان خروجی‌های تولیدشده از هر HMM را باهم ترکیب کنیم، به ازای یک دنباله ورودی Sequence، خروجی i امین HMM، Hmm_i نامیده می‌شود، که به‌صورت زیر بیان می‌شود:

$$P(Sequence|Hmm_i)\ P(Hmm_i|Sequence)\ P(Sequence)\ /\ p(Hmm_i)$$

بنابراین در مرحله آموزش داریم:

$$Train = \max\{P(Sequence|Hmm_i)\},\ i \in [1, N$$

و در مرحله تشخیص داریم:

$$Detection = Average\{P(Sequence|Hmm_i)\},\ i \in [1, N]$$

در اینجا N تعداد کل HMM های درون یک ensemble است؛ بنابراین با استفاده از قانون بزرگ‌ترین در مرحله آموزش و میانگین در مرحله تشخیص، می‌توان سیستم را به بالاترین نرخ تشخیص رساند و خطاهای برخی از HMMها را در مرحله تشخیص یا میانگین‌گیری از بین برد. این امر درواقع دلیل اصلی استفاده از گروه HMM می‌باشد.

۶،۴.ساخت قوانین فازی

قوانین فازی ساخته‌شده نسبت به حساسیت و تأثیر هر یک از ۵ ویژگی انتخاب‌شده متفاوت می‌باشند و باید انتظار داشت که هرکدام نسبت به ورودی دریافت شده از خروجی هر گروه HMM، تصمیم متفاوتی را اتخاذ کنند؛ بنابراین برای هر ویژگی استخراج‌شده سه مجموعه low، medium و high وجود دارد که احتمال حاصل از گروه HMM مربوطه را نشان می‌دهد. بنابراین با توجه به این مجموعه‌های فازی می‌توان قوانین فازی موردنظر را به وجود آورد؛ بنابراین اگر قوانین فازی را به شکل زیر تعریف کنیم:

$$R = \sum_{i=1}^{n} R_i$$

و ۵ ویژگی انتخاب‌شده را به شکل زیر داشته باشیم:

$$F = \sum_{i=1}^{5} f_i$$

از اینرو، درجه عضویت هر ویژگی به مجموعه فازی شکل‌گرفته را به شکل زیر نشان می‌دهیم:

$$\mu_R(Fi),\ \mu_R(Fi) \in \{low, medium, high\}$$

به عنوان مثال برای یک ترافیک انتقالی، اگر احتمال مناسبی برایDst ، Hst،Log وsrv به دست آید، اما احتمال به‌دست‌آمده برای Src نامناسب باشد، آنگاه ترافیک انتقالی نرمال می‌باشد. همچنین اگر احتمال به‌دست‌آمده برای Log,Hst و srv نامناسب باشد و احتمال به‌دست‌آمده برای Dst,Hst مناسب باشد، ترافیک انتقالی غیر نرمال برچسب می‌خورد؛ بنابراین با توجه به این فرضیات قوانین موردنظر شکل می‌گیرند.

ISSN: 2745-3673
http://www.Rcsj.ir

۷،۴. ارزیابی

برای ارزیابی سیستم پیشنهادی از ۲۰۰۰۰۰ داده غیر مخرب و ۲۷۰۰۰۰ داده مخرب که به‌صورت تصادفی از مجموعه داده‌ها استخراج‌شده‌اند، استفاده کرده‌ایم. نتایج به‌دست‌آمده از ماتریس درهم‌ریختگی در جدول (۲) نمایان شده است نشان می‌دهد که سیستم پیشنهادی در تشخیص حملات موفق بوده است و تمامی حملات را شناسایی کرده است اما دارای خطای نسبتاً پایینی در شناسایی داده‌های غیر مخرب به‌عنوان مخرب بوده است.

جدول ۲: جدول درهم‌ریختگی حاصل ارزیابی سیستم پیشنهادی.

سیستم تشخیص نفوذ	با تصمیم‌گیری فازی			
تعداد حمله‌های کشف‌شده	Ipsweep	Portsweep	Neptune	smurf
	۱۰۰۰۰	۱۰۰۰۰	۵۰۰۰۰	۲۰۰۰۰۰
تعداد بسته‌های غیر مخربی که مخرب شناخته‌شده	۴۸۰			
تعداد بسته‌های مخرب که غیر مخرب شناخته‌شده	۷۹۰			
نرخ کشف	٪۹۹,۷۱			
نرخ مثبت کاذب	٪۰,۲۴			
نرخ منفی کاذب	٪۰,۲۹			

۸،۴. استنتاج فازی

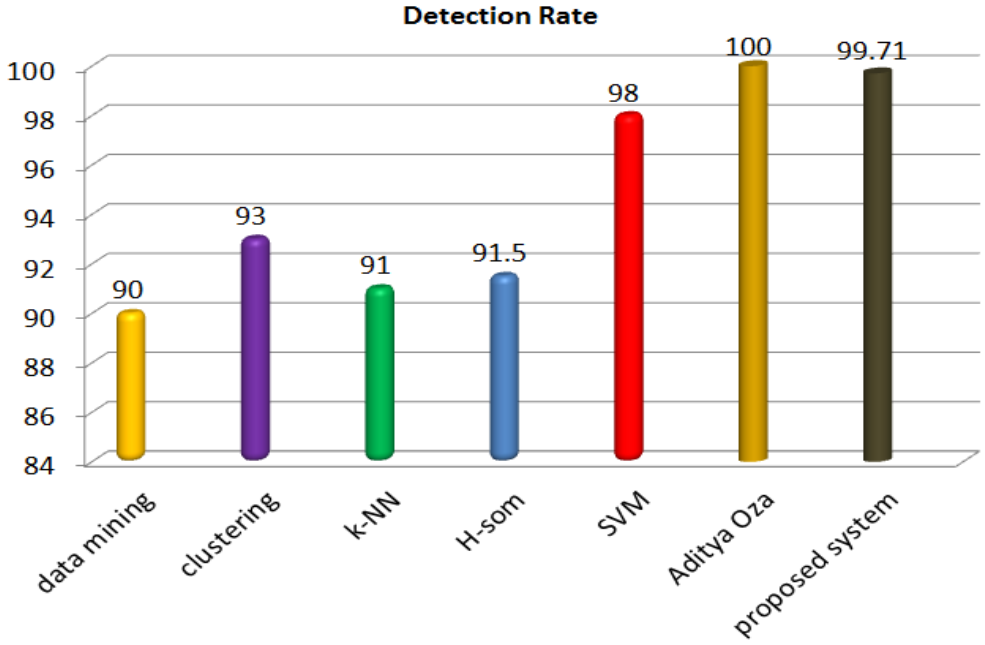
در مرحله‌ی غیرفازی‌سازی، مقادیر فازی به مقادیر حقیقی موردنیاز تبدیل می‌شود که می‌تواند یکی از دو مقدار غیر مخرب و مخرب باشد. در این کار از غیرفازی‌ساز ماکزیمم استفاده‌شده است. همان‌طور که در جدول (۳) مشاهده می‌کنید، می‌توان به‌سادگی فهمید که با افزایش قوانین فازی، نرخ کشف و نرخ مثبت کاذب در وضعیت خوبی قرار می‌گیرد که به ترتیب ٪۹۹,۷۱ و ٪۰,۲۹ است که نسبت به سیستم‌های گذشته بهبودیافته است.

جدول ۳: نتایج ارزیابی سیستم پیشنهادی با قوانین فازی

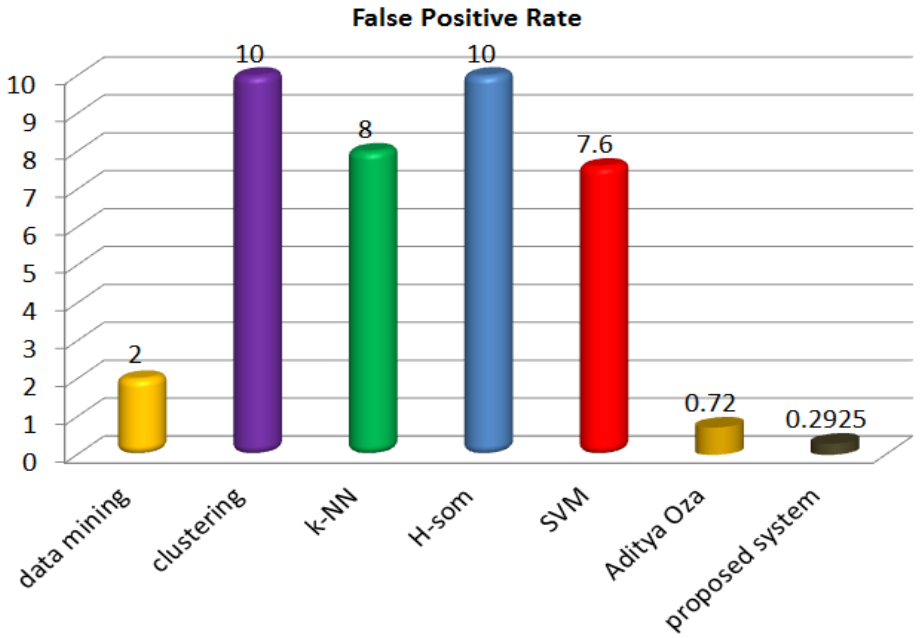
قوانین فازی	نرخ کشف	نرخ مثبت کاذب
۵	٪۹۵,۶۰	٪۰,۱۰
۱۰	٪۹۸,۲۰	٪۰,۲۰
۱۲	٪۹۹,۲۶	٪۰,۳۲
۲۵	٪۹۹,۷۱	٪۰,۲۹
۳۳	٪۱۰۰	٪۰,۵۸

هرچند با افزایش قوانین فازی پیچیدگی سیستم بیشتر می‌شود و سیستم کندتر عمل می‌کند، ولی باید بین پیچیدگی و نرخ مثبت کاذب تعادل برقرار کرد تا بتوان بهترین حالت را به وجود آورد.

ISSN: 2745-3673
http://www.Rcsj.ir



شکل۲: مقایسه نرخ تسشخیص سیستم پیشنهادی با آخرین مدل ارائه شده



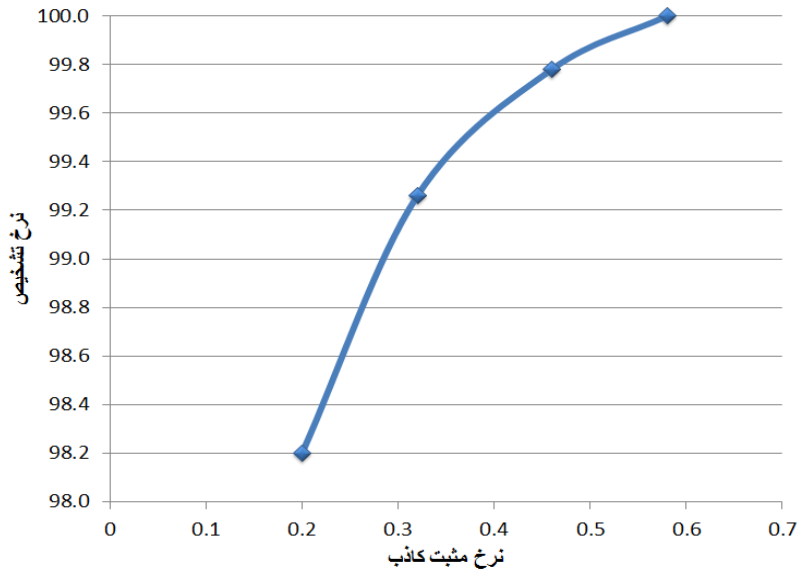
شکل۳: مقایسه نرخ مثبت کاذب سیستم پیشنهادی با آخرین مدل ارائه شده

۹،۴. عملکرد نهایی سیستم

برای نمایش عملکرد نهایی سیستم، از نمودار AUC استفاده می‌شود. منحنیROC۳ برای نمایش تعادل بین کمیت‌های نرخ تشخیص و نرخ مثبت کاذب در یک روش طبقه‌بندی استفاده می‌شود. مشخصه مهم این نمودار مساحت زیر شکل حاصل از آن است که به آنAUC۴ گویند. کمیت AUC برای یک طبقه‌بندی خوب برابر ۱ است. کمیت AUC برای سیستم پیشنهادی طبق شکل (۴) برابر ۰،۹۹ است.

[3] Receiver Operating Characteristic
[4] Area Under the Curve

ISSN: 2745-3673
http://www.Rcsj.ir



شکل ۴: نمودار AUC در حالتی که سیستم علاوه بر نرخ تشخیص بالا، به نرخ مثبت کاذب پایینی دست‌یافته است.

۵. نتیجه‌گیری

در این پژوهش یک سیستم تشخیص نفوذ برای کشف حمله‌های مخرب برای ترافیک انتقالی شبکه ارائه‌شده است که به‌جای استفاده از یک مدل مخفی مارکوف، از چندین مدل مخفی مارکوف به‌عنوان یک گروه HMM جهت به‌کارگیری تکنیک طبقه بندی چندگانه برای مدل‌سازی ویژگی‌های استخراج‌شده استفاده می‌کند. در این مقاله دو راهکار برای حل مشکل تشخیص بی نظمی ارائه شده است. همچنین این سیستم به‌جای استفاده از یک مقدار آستانه جهت تصمیم‌گیری غیر مخرب یا مخرب بودن ترافیک انتقالی، با تولید مجموعه‌ها و قوانین فازی و به‌کارگیری استنتاج فازی، به پایین‌ترین نرخ مثبت کاذب دست‌یافته است.

ISSN: 2745-3673
http://www.Rcsj.ir

منابع و مراجع

[1] Kru¨ gel C, Toth T, Kirda E. Service specific anomaly detection for network intrusion detection. In: SAC'02: proceedings of the2002 ACM symposium on APPLIED computing. New York, NY, USA: ACM; 2002. 201e208.
[2] MartinGarcia L. Programming with libpcap - sniffing the network from our own application. Hakin9 Magazine, http://recursos.aldabaknocking.com/libpcapHakin9LuisMartinGarcia.pdf;February 2008.
[3] Wang K, Stolfo SJ. Anomalous payload-based network intrusion detection. In: Jonsson E, Valdes A, Almgren M, editors. RAID. Lecture Notes in Computer Science, vol. 3224. Springer; 2004.p. 203e22.
[4] Damashek M. Gauging similarity with n-grams: languageindependent categorization of text. Science 1995; 267(5199):843e8.
[5] Wang K, Cretu GF, Stolfo SJ. Anomalous payload-based wormdetection and signature generation. In: Valdes A, Zamboni D, editors. RAID. Lecture notes in computer science, vol. 3858.Springer; 2005. p. 227e46.
[6] Wang K, Parekh JJ, Stolfo SJ. Anagram: a content anomaly detector resistant to mimicry attack. In: Zamboni D, Kru¨ gel C, editors. RAID. Lecture notes in computer science, vol. 4219.Springer; 2006. p. 226e48.
[7] Perdisci R, Ariu D, Fogla P, Giacinto G, Lee W. Mcpad: a multiple classifier system for accurate payload-based anomaly detection. Computer Networks 2009; 53(6):864e81 [Special Issue on Traffic Classification and Its Applications to Modern Networks].
[8] Rabiner L. A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 1989; 77(2):257e86.
[9] Ghmm: General Hidden Markov Model library, http://ghmm.org/.Gu¨ nter S, Bunke H. Optimizing the number of states, training iterations and gaussians in an hmm-based handwritten word recognizer. In: Proceedings of the Seventh Internationa lConference on Document analysis and recognition. IEEE Computer Society; 2003. 472.
[10] Suen CY. N-gram statistics for natural language understanding and text processing. IEEE Transactions on Pattern Analysisand Machine Intelligence PAMI 1979; 1(2):164e72.
[11] Wang K, Cretu GF, Stolfo SJ. Anomalous payload-based worm Detection and signature generation. In: Valdes A, Zamboni D, editors. RAID. Lecture notes in computer science, vol. 3858.Springer; 2005. p. 227e46.
[12] Durbin R, Eddy S, Krogh A, Mitchison G. Biological sequenceanalysis. Cambridge University Press; 2006.
[13] Fogla P, LeeW.Evading network anomaly detectionsystems: formal reasoning and practical techniques. In: CCS'06: Proceedings of the 13th ACM conference on computer and communications security. New York, NY, USA: ACM; 2006. 59e68.
[14] Davide, Ariu, Roberto, tronic, Giorgio, Giacinto," HMMPayl: An intrusion detection system based on Hidden Markov Models", Department of Electric and Electronic Engineering, University of Cagliari Piazza d'Armi, 09123 Cagliari, Italy.