مجله علمی- پژوهشی "پدافند الکترونیکی و سایبری"
سال پنجم، شماره ۳، پاییز ۱۳۹۶، ص ۹۳-۷۹

ارائه یک ره‌یافت جدید مبتنی بر روش ترکیبی به منظور آشکارسازی نفوذ در شبکه

سعید پارسا[1*]، سید حمیدرضا اعرابی[2]

۱- دانشیار، ۲- دانشجوی کارشناسی ارشد، دانشگاه علم و صنعت ایران

(دریافت: ۹۵/۰۹/۲۴، پذیرش: ۹۵/۱۱/۲۵)

چکیده

نقش یک سامانه تشخیص نفوذ برای آشکارسازی ناهنجاری‌ها در شبکه از اهمیت زیادی برخوردار است. حملات جدید و ناشناخته موجب ناکارآمدی راه‌کارهای شناسایی مبتنی بر امضاء و در نتیجه استفاده از راه‌کارهای شناسایی مبتنی بر ناهنجاری شده است. این راه‌کارها نیز علی‌رغم توانایی بالا در تشخیص ناهنجاری‌ها، از نرخ مثبت کاذب بالایی رنج می‌برند. برای غلبه بر این مشکل، ایده استفاده از آشکارسازهای ترکیبی مطرح شده است. در این مقاله، راه‌کاری نوین مبتنی بر روش آشکارسازی ترکیبی با یک معماری چهارلایه‌ای پیشنهاد شده است. لایه اول از واحد تحلیل‌گر جریان داده‌ها و واحد طبقه‌بندی تشکیل شده است که برای طبقه‌بندی نوع سرویس‌های شبکه از ترکیب روش آماری n-گرام و الگوریتم ژنتیک استفاده می‌کند. در لایه تشخیص نفوذ، یک واحد آشکارساز مبتنی بر امضاء و واحدهای آشکارساز مبتنی بر ناهنجاری به شکل ترکیبی پیاده‌سازی شده‌اند که متناسب با برچسب نوع سرویس‌ها فراخوانی می‌شوند. سپس، درنتیجه پردازش این واحدها، لایه تصمیم‌گیری فراخوانی می‌شود. این لایه ماهیت حمله و نوع پاسخ را تشخیص داده و لایه مدیریت وقایع را فرا می‌خواند. در این لایه ضمن اطلاع‌رسانی هشدارها به مدیر شبکه، در صورت نیاز، اعمال واکنشی و اقدامات امنیتی لازم نیز انجام خواهد شد. نتایج حاصل از ارزیابی اعتبارسنجی چندلایه‌ای، بهبود دقت تشخیص نفوذ را ۹۹/۸۱٪ نشان می‌دهد که در نتیجه کاهش میزان نرخ مثبت کاذب را در پی خواهد داشت.

واژه‌های کلیدی: تشخیص نفوذ، نرخ مثبت کاذب، مدیریت وقایع، طبقه‌بندی نوع سرویس، اعمال واکنشی، اعتبارسنجی چندلایه‌ای.

۱- مقدمه[1]

هدف از این پژوهش، کاهش نرخ مثبت کاذب در تشخیص نفوذ در شبکه است. اهمیت موضوع در تشخیص دقیق‌تر نفوذ در شبکه است. روش پیشنهادی مبتنی بر ادغام دو روش شناخته‌شده مبتنی بر امضاء[1] و مبتنی بر ناهنجاری[2] در ترافیک شبکه است [۲-۱]. برای کاهش نرخ مثبت کاذب با شناسایی و طبقه‌بندی سرویس‌های ترافیک شبکه بر اساس برچسب نوع سرویس‌ها می‌توان مدل‌های مناسب که با داده‌های مربوط به هر نوع سرویس به‌طور مجزا آموزش‌دیده‌اند را برای تشخیص نفوذ مورد استفاده قرار داد. در صورت عدم شناسایی نوع سرویس از جریان داده ترافیک شبکه می‌بایست مدل‌های موجود را با مجموعه داده‌های آموزشی[3] به‌روز بررسی کرد. روش‌های مبتنی بر امضاء نیازمند به‌روزرسانی مکرر پایگاه داده حملات در فواصل زمانی کوتاه هستند. به‌علاوه حملاتی که از روش‌های رمزنگاری داده‌ها استفاده می‌نمایند توسط این روش به‌سختی قابل تشخیص هستند. در این روش، الگوهای نفوذ از پیش‌ساخته (امضاء) به‌صورت الگوهای حمله نگهداری می‌شوند، به‌طوری‌که هر الگو انواع متفاوتی از یک نفوذ خاص را دربرگرفته و درصورت بروز چنین الگویی در سامانه، وقوع نفوذ اعلام می‌شود. در این روش‌ها، معمولاً تشخیص‌دهنده دارای پایگاه داده‌ای از امضاءها یا الگوهای حمله است و سعی می‌کند با بررسی ترافیک شبکه، الگوهای مشابه با آن‌چه را که در پایگاه داده خود نگهداری می‌کند، بیابد. این دسته از روش‌ها تنها قادر به تشخیص نفوذهای شناخته‌شده می‌باشند و درصورت بروز حملات جدید در سطح شبکه، نمی‌توانند آن‌ها را شناسایی کنند و نیازمند به‌روزرسانی پایگاه داده‌ای[4] از امضاءها یا الگوهای حمله در فواصل زمانی کوتاه هستند. در این روش نرخ مثبت کاذب پایین است و از مزایای این روش دقت در تشخیص نفوذهایی است که الگوی آن‌ها قبلاً شناسایی و در پایگاه داده‌ها ثبت شده است.

* رایانامه نویسنده مسئول: parsa@iust.ac.ir

1- Signature-Based
2- Anomaly-Based
3- Training Data Set
4- Data Base

روش آشکارسازی ناهنجاری به این صورت است که براساس داده‌های طبیعی مدل‌هایی از فعالیت‌های قانونی ساخته می‌شود و در صورت انحراف از مدل مذکور حمله یا ناهنجاری درنظر گرفته می‌شود. برای رسیدن به این هدف، دو فاز درنظر گرفته می‌شود: فاز آموزش و فاز آزمایش. در فاز آموزش با استفاده از روش‌های یادگیری ماشین براساس داده‌های آموزشی مدل طبیعی[1] ساخته می‌شود. در فاز آزمایش مدل یاد گرفته شده در برابر نمونه‌های جدید با استفاده از داده‌های آموزشی که از قبل ساخته شده‌اند، آزمایش می‌شود. بر همین اساس، این نوع آشکارسازها توانایی تشخیص حملات ناشناخته و جدید را دارند.

هدف دیگر از این پیشنهاد، ارائه روشی کارآمد با هزینه کم‌تر به‌منظور تشخیص فعالیت‌های بدخواه شبکه است. روش پیشنهادی سعی در افزایش دقت تشخیص آشکارسازها و کاهش نرخ مثبت کاذب نسبت به سایر روش‌های دیگر را دارد. این کار از طریق بررسی معایب روش‌های تشخیص نفوذ و بررسی راه‌کارهای هوشمندسازی فرآیند تشخیص نفوذ با استفاده از مدل‌های یادگیری ماشین و با هدف ارائه روشی جدید براساس روش‌های طبقه‌بندی و تشخیص محقق می‌گردد.

در آشکارسازهای مبتنی بر امضاء، نرخ مثبت کاذب پایین است و آشکارسازهای ناهنجاری در تشخیص حمله‌های شناخته‌شده و ناشناخته توانایی بالایی دارند. در روش آشکارسازی ناهنجاری دو چالش عمده وجود دارد. چالش اول، پیاده‌سازی یک واحد تحلیل‌گر جریان و طبقه‌بندی نوع سرویس ترافیک شبکه با دقت بالا است [۳]. چنان‌چه این بخش به‌درستی انجام گیرد خروجی‌های به‌دست‌آمده در مراحل بعد نیز تأثیرگذار خواهند بود؛ و در نهایت علی‌رغم توانایی بالا تشخیص آسیب‌پذیری‌های ناهنجاری، این روش از میزان بالای نرخ مثبت کاذب رنج می‌برد. بنابراین، چالش دیگر در مرحله تشخیص ماهیت و نوع حمله است. این روشی مناسب است که علاوه‌بر تشخیص تعداد بیش‌تر حملات نسبت به روش‌های دیگر، بتواند نرخ مثبت کاذب کم‌تری داشته باشد [۴-۶].

در این پژوهش، سامانه ی مبتنی بر روش ترکیبی معرفی‌شده استفاده شده است. به‌این‌ترتیب، به‌طور هم‌زمان از نرخ آشکارسازی بالای حملات شناخته‌شده در دسته اول و توانایی آشکارسازی حملات ناشناخته و جدید در دسته دوم، بهره‌مند می‌شویم. در بین روش‌های مختلف امنیتی در شبکه‌های کامپیوتری، نقش سیستم‌های آشکارسازی نفوذ[2] (IDS) به‌عنوان عاملی برای تشخیص ناهنجاری‌ها و حملات، از اهمیت بسیار بالایی برخوردار است. راه‌کار پیشنهادی از چهار لایه تشکیل شده است: ۱) لایه تحلیل و طبقه‌بندی ترافیک شبکه. ۲) لایه تشخیص نفوذ. ۳) لایه تصمیم‌گیری و ۴) لایه مدیریت ثبت وقایع و هشدارها. در ادامه به‌مرور بر کارهای مرتبط انجام‌شده در زمینه سامانه‌های تشخیص نفوذ و بررسی چالش آن‌ها پرداخته شده است و همین چالش‌ها، انگیزه‌ای برای ارائه روش پیشنهادی در بخش سوم ایجاد کرده است. در بخش چهارم، روش پیشنهادی در مقایسه با روش‌های قبلی به‌صورت عملی و نظری مورد ارزیابی قرارگرفته و در نهایت به جمع‌بندی و نتیجه‌گیری در بخش پنجم پرداخته شده است.

۲- کارهای مرتبط انجام‌شده

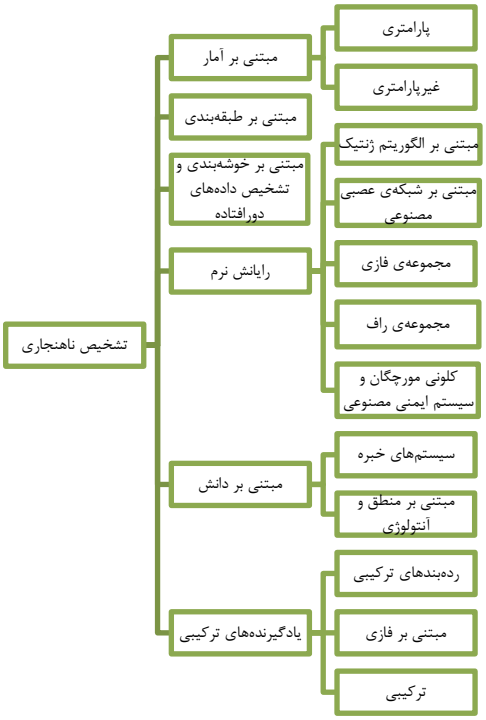
تحقیقات بسیار زیادی در زمینه تشخیص نفوذ در شبکه‌های رایانه ای انجام شده است. در حالت کلی، تحقیقات موجود در این زمینه در شش شاخه اصلی قابل دسته‌بندی هستند [۷ و ۳۷].

- مبتنی بر آمار[3]
- مبتنی بر طبقه‌بندی[4]
- مبتنی بر خوشه‌بندی و تشخیص داده‌های دورافتاده[5]
- مبتنی بر رایانش نرم[6]
- مبتنی بر دانش[7]
- مبتنی بر یادگیرنده‌های ترکیبی[8]

در ادامه به شرح برخی از روش‌های تشخیص ناهنجاری پرداخته در روش‌های مبتنی بر آمار، فعالیت طبیعی ترافیک شبکه در قالب یک نمایه[9]، براساس معیارهایی مثل میزان ترافیک، تعداد بسته‌ها برای هر قوانین، میزان اتصالات، تعداد نشانی‌های IP[10] متفاوت ساخته می‌شود. وقتی رویدادهای شبکه رخ می‌دهند، براساس همان معیارها یک نمایه جاری تعریف‌شده و یک نمره ناهنجاری به‌وسیله مقایسه دو نمایه تخمین زده می‌شود.

1- Normal Model

2- Intrusion Detection System
3- Statistical
4- Classification Based
5- Clustering and Outlier-Based
6- Soft Computing
7- Knowledge-Based
8- Combination Learners
9- Profile
10- Internet Protocol



شکل (۱): دسته‌بندی روش‌های تشخیص ناهنجاری [۳۷]

روش‌های مبتنی بر دانش روش‌های تشخیص ناهنجاری، مبتنی بر مشخصه هستند که مدل مطلوب توسط یک فرد خبره به‌طور دستی برحسب مجموعه‌ای از قوانین (مشخصه‌ها) ساخته می‌شود. به‌کارگیری و توسعه این روش‌ها در مقیاس بالا پیچیده‌تر است.

رایج‌ترین روش‌های مورد استفاده در سال‌های اخیر، مبتنی بر روش‌های یادگیری ماشین هستند. یک شبکه بیزی[1] مدلی است که رابطه‌های احتمالی بین متغیرهای مورد نظر را کدگذاری می‌کند. این روش معمولاً برای تشخیص نفوذ در ترکیب با طرح‌های آماری به‌کار رفته است. این روش مزایای متعددی از جمله قابلیت کدکردن وابستگی‌های بین متغیرها و رویدادهای از پیش تعیین‌شده و هم‌چنین قابلیت ترکیب‌کردن دانش پیشین و داده‌ها را شامل می‌شود. ابتدا داده‌ها با استفاده از روش تجزیه و تحلیل مؤلفه‌های اصلی[2] پیش‌پردازش[3] و سپس با استفاده از یک شبکه بیزین، نمونه‌ها به دو دسته طبیعی و ناهنجار دسته‌بندی شده‌اند [۸]. هم‌چنین از ترکیبی از چند شبکه بیزین برای تشخیص نفوذ در سامانه بر روی مجموعه داده‌های Kdd'۹۹ استفاده شده است [۹]. الگوریتم‌های ژنتیک روش‌هایی برای جستجو هستند که در پیداکردن جواب دقیق یا تقریبی برای مسائل بهینه‌سازی یا جستجو به‌کار می‌روند. به‌دلیل انعطاف‌پذیری و مقاوم‌بودن در انجام جستجوی سرتاسری، در مسئله آشکارسازی ناهنجاری به شکل‌های مختلف مورد استفاده قرار گرفته است. از الگوریتم ژنتیک به‌طور مستقیم برای رده‌بندی استفاده شده است [۱۰]. در این مقاله، یک سامانه تشخیص نفوذ با استفاده از الگوریتم ژنتیک معرفی و پیاده‌سازی شده است. در تحقیقات اخیر در زمینه تشخیص نفوذ از الگوریتم ژنتیک و منطق فازی استفاده شده است [۱۱]. درخت تصمیم به‌عنوان یک «مدل پیش‌بینی‌کننده براساس یادگیری ماشین و آمارها به‌منظور ایجاد یک ساختار درختی برای مدل‌کردن الگوهای داده‌ای» معرفی می‌شود که در نمونه‌ای از کارهایی که با استفاده از درخت تصمیم‌گیری و خوشه‌بندی k-Means نسبت به تشخیص نفوذ در شبکه مورد استفاده قرار گرفته است [۱۳-۱۲].

مقالات متعددی از شبکه عصبی برای تشخیص نفوذ استفاده کرده‌اند. شبکه عصبی یک ساختار توزیع‌شده موازی به شکل یک گراف جهت‌دار است. هر گره را در این گراف یک نمونه پردازشگر یا یک واحد پردازشگر یا یک نرون[4] می‌نامند. این گراف از یک‌لایه ورودی و یک‌لایه خروجی تشکیل شده است. لایه‌های بین ورودی و خروجی را لایه‌های مخفی می‌نامند. هر اتصال گراف دارای وزن و جهت است که جهت تأثیر و میزان تأثیر گره مبدأ به گره مقصد را مشخص می‌کند. وزن‌های اتصالات در یک فاز آموزش یا یادگیری تعیین می‌شوند. تعیین وزن‌ها به کمک قواعد یادگیری شبکه و نمونه‌های ورودی و خروجی انجام می‌پذیرد. نمونه‌هایی از پژوهش‌های ارائه‌شده در زمینه تشخیص نفوذ با استفاده از شبکه عصبی وجود دارد [۱۵-۱۴]. روش‌های متعدد دیگری نیز در این دسته قرار می‌گیرند. ازجمله روش‌های مبتنی بر روش سامانه امن [۱۶]، مدل مارکف، ماشین بردار پشتیبان، خوشه‌بندی، منطق فازی.

منظور از رده‌بندهای ترکیبی این است که چند رده‌بند وجود دارد و الگوریتمی که نتیجه نهایی را اعلام می‌کند، از ترکیب پاسخ چند رده‌بند استفاده می‌کند.

هم‌چنین روش‌های دیگری نیز وجود دارند که از آشکارسازهای ترکیبی استفاده نموده‌اند. ابتدا از یک آشکارساز ناهنجاری

1- Bayesian Network
2- Principle Component Analysis
3- Preprocessing

4- Neurons

به‌منظور ایجاد فهرستی از موارد مشکوک استفاده می‌شود. سپس از یک تشخیص‌دهنده امضاء برای رده‌بندی موارد به سه دسته هشدارهای نادرست، حملات شناخته‌شده و حملات ناشناخته استفاده می‌شود [۱۷]. این روش براساس این فرض بنا شده است که آشکارساز ناهنجاری با نرخ آشکارسازی بالا کار می‌کند. چرا که نفوذهای شناسایی‌نشده در مرحله اول، در مرحله دوم هم توسط تشخیص‌دهنده امضاء شناسایی نمی‌شوند.

در مرحله اول از یک جنگل تصادفی به‌منظور تشخیص بدرفتاری استفاده شده است [۱۸]. به این ترتیب، در این مرحله نفوذهای شناخته‌شده آشکار می‌شوند. در مرحله بعد، از توانایی جنگل تصادفی در آشکارسازی داده‌های پرت، به‌منظور شناسایی نفوذهای ناشناخته استفاده می‌شود. در پژوهش دیگر آشکارساز نفوذ ترکیبی پیشنهادشده که از واحدهای آشکارساز ناهنجاری و آشکارسازی بدرفتاری و یک سامانه پشتیبان تصمیم‌گیری[1] (DSS) به‌منظور ترکیب خروجی دو واحد مذکور تشکیل شده است. در واحد آشکارساز ناهنجاری از ساختار SOM[2] برای مدل‌سازی رفتار طبیعی استفاده‌شده و هر انحرافی از رفتار طبیعی به‌عنوان حمله شناخته می‌شود. در واحد آشکارسازی بدرفتاری از یک درخت تصمیم به‌منظور رده‌بندی انواع حملات استفاده شده است [۱۹].

در آشکارسازهای مبتنی بر امضاء، نرخ مثبت نادرست پایین است و آشکارسازهای ناهنجاری[3] در تشخیص حملات شناخته‌شده و ناشناخته توانایی بالایی دارند. با ترکیب این دو ایده، سامانه با نام HIDS[4] معرفی شده که نتایج تجربی خوبی به همراه داشته است. در این مقاله توانایی روش در تشخیص حملات با کاوش در ترافیک ناهنجاری موجود در اتصالات اینترنتی ارتقا داده شده است [۲۰].

در بعضی سامانه‌ها، به‌جای ترکیب روش‌های آشکارسازی ناهنجاری و روش‌های مبتنی بر امضاء، چند آشکارساز ناهنجاری که هر یک با روش‌های مختلف کار می‌کنند باهم ترکیب شده‌اند. در این‌جا هدف اصلی کاهش نرخ بالای مثبت‌های کاذب[5] و در عین حال داشتن دقت قابل قبول در نرخ آشکارسازی است.

1- Decision Support System
2- Self-Organizing Map
3- Anomaly Detection
4- Host-based Intrusion Detection System
5- False Positive

یکی دیگر از مباحث مهم در شبکه موضوع طبقه‌بندی ترافیک است. روش مرسوم اولیه برای تعیین هویت کاربردهای شبکه، براساس قوانین و درگاه مربوطه است. این روش، در مورد کاربردهایی که به‌صورت پویا درگاه مورد استفاده را عوض می‌کنند یا در مواردی‌که خود را در درون کاربردهای دیگر پنهان می‌کنند، پاسخ‌گو نیست. به‌منظور حل این مشکل، در سال‌های اخیر، روش‌های مختلفی برای رده‌بندی ترافیک شبکه ارائه شده است. به‌طوری‌که این روش‌ها به دو دسته مبتنی بر امضاء و مبتنی بر آمار تقسیم می‌شوند. در ادامه، این روش‌ها را بررسی کرده و محدودیت‌های آن‌ها را مورد توجه قرار می‌دهیم.

۲-۱- رده‌بندی ترافیک مبتنی بر امضاء

در این روش در محتوای بسته‌ها، رشته‌های مشخصی جستجو شده و بر این اساس، نوع کاربرد تعیین می‌شود. با تحلیل ۲۰ ترابایت از ترافیک نقطه به نقطه Kazaa مشخصه امضای ترافیک[6] KaZaA به‌دست می‌آید [۲۱]. اگرچه تحلیل معنایی ترافیک، دقت کار را بالا می‌برد، اما استفاده مستمر از آن در کاربردهای زمان واقعی عملی زمان‌بر بوده و غیرممکن است؛ بنابراین فقط به‌منظور تشکیل امضای لایه کاربرد، ترافیک سطح بسته بررسی می‌شود و در ادامه براساس امضای به‌دست‌آمده، ترافیک نقطه به نقطه[7] فیلتر می‌شود [۲۲]. آزمایش‌ها نشان می‌دهند نرخ مثبت نادرست و منفی نادرست، کمتر از ۵٪ است. این روش در مورد ترافیک رمزشده SSL به‌کار گرفته شده است [۲۳]. روش‌های مبتنی بر امضاء، اگرچه دقت خوبی دارند، در مواردی‌که دست‌دهی اولیه قوانین طول متغیری دارد، مانند Gnutella، قادر به تشخیص نوع کاربرد نیست. مشکل دیگر آن است که این روش می‌تواند تنها یک کاربرد مانند KaZaA یا گروهی از کاربردها مانند شناسایی ترافیک گپ و گفت[8] را شناسایی کند [۲۴].

۲-۲- رده‌بندی ترافیک مبتنی بر آمار

استفاده از خواص آماری ترافیک شبکه یا مدل‌سازی رفتار ترافیک مسئله جدیدی نیست. متغیرهای آماری مانند طول بسته، فاصله زمانی بین بسته‌ها و مدت‌زمان جریان بسته‌ها قادر هستند رفتار بعضی قوانین‌ها را توصیف کنند. با ظهور کاربردهای شبکه‌ای جدید، مسئله این است که چگونه متغیرهای آماری را به انواع

6- Traffic Signature
7- Peer to Peer
8- Chat

کاربردها ارتباط دهیم. راه‌حل طبیعی این مشکل به‌کارگیری روش‌های یادگیری ماشین[1] است. مسائل مربوط به طبقه‌بندی ترافیک[2] و آینده آن مطرح شده است [۲۷-۲۵]. عملکرد برخی از روش‌های موجود در زمینه طبقه‌بندی ترافیک بررسی شده است. فقط با بررسی پنج بسته ابتدایی در یک اتصال TCP کاربردها شناسایی می‌شوند. با توجه به این‌که چند بسته اولیه مربوط به فاز مذاکره است، پیغام‌های رد و بدل‌شده اطلاعات مفیدی راجع به نوع کاربرد ارائه می‌دهد. به این ترتیب، می‌توان از این روش در کاربردهای برخط استفاده نمود. یک فاز یادگیری آنلاین و یک فاز رده‌بندی آنلاین وجود دارد [۲۸]. همچنین سه روش خوشه‌بندی[3] K-means، خوشه‌بندی GMM و خوشه‌بندی طیفی به‌کار گرفته شده است و اتصالاتی که متعلق به هیچ خوشه‌ای نباشد به‌عنوان ناشناخته تلقی می‌شود. بنابراین، این روش قادر است کاربردهای جدید یا حالت‌های کاری جدید کاربردهای قبلی را هم شناسایی نماید.

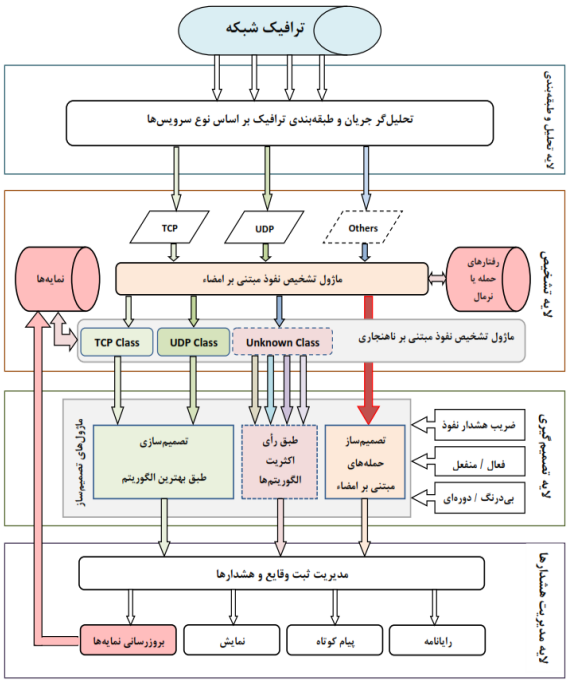
اگرچه این روش‌ها قادر به انجام رده‌بندی ترافیک هستند، تعداد سرویس‌های قابل شناسایی در آن‌ها بسیار محدود است. همچنین تعریف رده‌های مختلف سرویس‌ها، دقت لازم را ندارد. بنابراین، به‌طور دقیق نمی‌توان سرویس‌ها را از هم جدا کرد. یکی از کارهای شاخص در این زمینه که قادر است طیف وسیعی از سرویس‌ها را تشخیص دهد از روش نیمه نظارتی[4] استفاده نموده است [۲۹].

۳- معماری سامانه پیشنهادی

در این مقاله، یک سیستم آشکارساز نفوذ ترکیبی[5] پیشنهاد می‌شود که فعالیت‌های بدخواه موجود در شبکه را تشخیص داده و آن‌ها را به راهبر شبکه گزارش می‌دهد. شکل (۲) نمای کلی معماری سامانه پیشنهادی را نشان می‌دهد.

در این راه‌کار، چهار لایه پردازشی پیش‌بینی شده است:

- لایه تحلیل و طبقه‌بندی ترافیک شبکه[6]
- لایه آشکارسازی نفوذ[7]
- لایه تصمیم‌گیری[8]
- لایه مدیریت ثبت وقایع و هشدارها[9]



شکل (۲): نمای کلی معماری سامانه پیشنهادی

لایه نخست از دو واحد تحلیل جریان داده و طبقه‌بندی نوع سرویس تشکیل شده است. واحد طبقه‌بندی از ترکیب یک روش آماری برمبنای n-گرام و الگوریتم ژنتیک استفاده می‌نماید و وظیفه دسته‌بندی و برچسب‌گذاری داده‌ها را به عهده دارد. از الگوریتم n-گرام برای ایجاد بردار ویژگی مربوط به برنامه‌های مختلف استفاده‌شده و از الگوریتم ژنتیک[10] برای وزن‌دهی به هر یک از خانه‌های بردار ویژگی در جهت بهبود دقت[11] و کارایی[12] الگوریتم طبقه‌بندی بهره گرفته شده است. در لایه تشخیص نفوذ، یک واحد آشکارساز مبتنی بر امضاء و واحدهای آشکارساز مبتنی بر ناهنجاری به شکل ترکیبی پیاده‌سازی شده‌اند که متناسب با برچسب نوع سرویس‌ها فراخوانی می‌شوند. درنتیجه پردازش این واحدها لایه تصمیم‌گیری فراخوانی می‌شود. این لایه ماهیت حمله و نوع پاسخ را تشخیص داده و لایه مدیریت وقایع را فرامی‌خواند. در این لایه ضمن اطلاع‌رسانیِ هشدارها به مدیر شبکه، در صورت نیاز، اعمال واکنشی[13] و اقدامات امنیتی لازم نیز انجام خواهد شد.

1- Machine Learning
2- Traffic Classification
3- Clustering
4- Semi-Supervised
5- Hybrid Intrusion Detection System
6- Network Traffic Analysis and Classification Layer
7- Intrusion Detection Layer
8- Decision Making Layer
9- Notifications and Event Log Management Layer
10- Feature Vector
11- Accuracy
12- Performance
13- Responsive Actions

۳-۱- لایه تحلیل بسته و طبقه‌بندی ترافیک

امروزه طبقه‌بندی ترافیک شبکه[1]، به‌طور دقیق و حساس، به دلایلی توجه زیادی را به خود جلب کرده است، ازجمله: نقش پر اهمیت آن در زمینه‌های مختلف مثل برنامه‌ریزی شبکه، تأمین کیفیت خدمات[2]، نگاشت کلاس سرویس[3]. درگذشته، طبقه‌بندی ترافیک با تکیه بر یک درگاه و قوانین مشخص انجام می‌شد [۲۷]. استفاده از شماره درگاه بنا به دلایل زیر کارایی قابل قبولی ندارد:

- ظهور برنامه‌های جدید نظیر به نظیر[4] که برای پنهان‌ماندن از دید سازمان‌ها، از درگاهی‌های متفاوت و گاه متغیری استفاده می‌نمایند.
- اختصاص پویای درگاه[5] به برخی از برنامه‌ها.
- کپسوله‌سازی[6] سرویس‌های مختلف در یک برنامه.

به‌منظور غلبه بر این مشکل، تحقیقات قابل توجهی درزمینه طبقه‌بندی ترافیک شبکه انجام شده است.

رایج‌ترین دستاورد موفق در این زمینه، بازرسی[7] محتویات بار شبکه[8] و جستجوی حروف رشته‌ای برای مدل‌سازی برنامه کاربردی است. برای بیش‌تر برنامه‌های کاربردی، مراحل دست‌دهی[9] قوانین اولیه معمولاً به‌طور متفاوت انجام می‌شود؛ بنابراین از این ویژگی‌ها[10] می‌توان به‌منظور طبقه‌بندی استفاده نمود. علاوه‌بر این، امضاءهای قوانین می‌تواند یا از طریق مستندات عمومی مثل FCها یا تحلیل‌های تجربی برای دستیابی به رشته بیتی دقیق هر دو نوع ترافیک TCP و UDP مدل شود.

برای درک بهتر این موضوع، جدول (۱) امضاءهای یازده برنامه که در آن برای مشاهده امضاءها، حروف الفبا به فرم طبیعی و حروفهای غیرالفبایی در مبنای شانزده (با شروع "x۰") نشان داده شده است. همان‌طورکه در جدول (۱) مشاهده می‌گردد، برای تمایز راحت‌تر، به هرکدام از برنامه‌ها مجموعه‌ای از حروفهای یکتا اختصاص داده شده است.

1- Network Traffic Classification
2- Quality of service
3- Class of service mapping
4- Peer to Peer
5- Dynamic Port
6- Encapsulation
7- Auditing
8- Payload
9- Hand-Shaking
10- Features

جدول (۱): امضای برخی از برنامه‌های متداول

برنامه/ سرویس	مسیر	آفست	امضاء
Bit Torrent	مبدأ	1	BitTorrent
HTTP Image Transfer	مقصد	4	image/
HTTP Web	مبدأ	0	GET
Secure Web	مقصد	0	0x16 0x03
MSN Messenger	مبدأ	0	MSG
MS-SQL	مقصد	0	0x04 0x01 0x00 0x25 0x00 0x00 0x01 0x00 0x00 0x00 0x15 0x00 0x06 0x01 0x00 0x1B
POP	مقصد	0	+OK
SMTP	مبدأ	0	EHLO
Windows File Sharing	مقصد	4	|FF|SMB
Yahoo! Messenger	مقصد	0	YMSG

با توجه به جدول بالا می‌توان دریافت ترافیک شبکه امن، با ردگیری عبارت "۱۶۰۳" در ابتدای جریان بار شبکه تمایز داده شده است. به همین ترتیب، ترافیک برنامه Yahoo! Messenger با عبارت اسکی "YMSG" در بار شبکه قابل شناسایی است. هرچند این امضاءها لزوماً از ابتدای بار شبکه آغاز نمی‌شود. به‌عنوان مثال، برای شناسایی ترافیک انتقال‌دهنده تصویر HTTP، باید عبارت رشته‌ای "image/" را که از پنجمین مکان بار شبکه آغاز می‌شود جستجو کنیم. این نقطه شروع در جدول (۱) در ستون offset درج شده است. علاوه‌بر این، دانستن این موضوع که کدام سمت از ارتباط، سرویس‌دهنده[11] یا سرویس‌گیرنده[12]، امضاء را تولید می‌کنند نیز حائز اهمیت است. برای مثال آغازگر ارتباط در HTTP WEB سرویس‌گیرنده است. امضایی که برای شناسایی HTTP WEB به‌عنوان مثال رشته "GET"، درنظر گرفته شده است، توسط سرویس‌گیرنده ارسال می‌شود.

این اطلاعات به روش‌های مبتنی بر امضاء برای بهبود کارایی آن‌ها با جستجوی امضاءهای کوتاه در منبع یا مقصد بار شبکه، کمک چشم‌گیری می‌کند. علی‌رغم دقت نسبتاً مناسبی که این روش در شناسایی برنامه‌های شبکه دارد، اما بین ۲۰٪ تا ۴۰٪ شناسایی جریان شبکه شکست می‌خورد. به همین دلیل، به روش دقیق‌تری برای دسته‌بندی سرویس شبکه نیاز است. در زیر به برخی از دلایل شکست این روش اشاره شده است.

11- Server
12- Client

هنگامی‌که نسخه جدیدی از یک برنامه منتشر می‌گردد، تمامی امضاءهای مربوط به آن برنامه باید به‌روزرسانی شوند که این امر معمولاً به‌صورت سریع امکان‌پذیر نیست.

تعداد زیادی از برنامه‌ها توسط توسعه‌دهندگان کوچک در سراسر جهان منتشر می‌شوند که شرکت‌های شبکه از آن‌ها اطلاع ندارند. هم‌چنین، برای آن‌که شرکت‌ها ترافیک‌شان را فیلتر نکنند، از تعداد زیادی از قوانین‌ها با امضاءهای مختلف استفاده می‌نمایند. به‌طور معمول، این امضاءها معادل‌هایی از برنامه‌های مشهور مانند Bit Torrent هستند که تغییر کوچکی در آن‌ها داده شده است.

ترافیک رمزشده یکی از چالش‌های بزرگ در طبقه‌بندی ترافیک[1] است. رمزگذاری داده‌ها[2] با رشد سریع سرویس‌های تجاری بسیار مشهور شده‌اند، اما در کنار این مسئله بسیاری از فعالیت‌های خراب‌کارانه از این روش برای پنهان‌کردن خود استفاده می‌نمایند.

با توجه به مشکل ذکرشده، در ادامه این بخش، روشی جدید ارائه خواهد شد که بتواند با دقت مناسب، نوع سرویس ترافیک را طبقه‌بندی نموده و بر مشکلات بالا نیز فائق آید. بدین ترتیب، برنامه‌های جدید و هم‌چنین سرویس‌های نظیر به نظیر نیز به‌درستی در دسته‌های متناسب با ویژگی‌های مشابه قرار می‌گیرند. با این‌حال هنوز مسئله دسته‌بندی ترافیک‌های رمزشده به قوت خود باقی است.

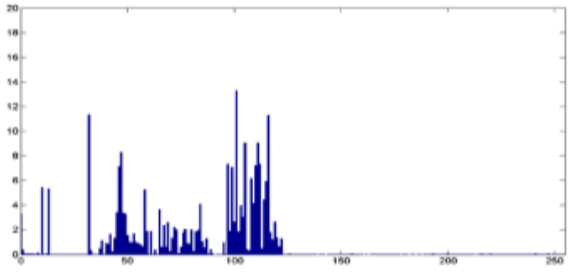
۳-۱-۱- مدل ۱- گرام وزنی

n-گرام‌ها ابزارهای مستقل از زبانی هستند که برای اندازه‌گیری میزان شباهت متون مورد استفاده قرار می‌گیرند. در پردازش متون و اسناد معمولاً یک پنجره متحرک بر روی قسمت‌های مختلف متن حرکت کرده و در هر بخش تعداد تکرارهای n- گرام‌ها را محاسبه می‌نماید. این روش اولین بار توسط Damshek معرفی شد و پس از آن در بسیاری از سامانه‌های تحلیل متن مورد استفاده قرار گرفت [۳۰]. درصورتی‌که این روش را بر روی بسته‌های شبکه اعمال نماییم و ۱- گرام آن‌ها را محاسبه کنیم، می‌توان به نمایش معناداری از توزیع حروفهای یک بسته پی ببریم. بدین ترتیب، بسته‌هایی که باهم شباهت دارند را می‌توان با استفاده از دسته‌بندی هیستوگرام از تعداد دفعات تکرار هر حروف مشخص نمود. Wang و همکارانش برای نخستین بار از این روش برای تشخیص ناهنجاری در شبکه استفاده نمودند. آن‌ها در مقاله خود، هر بسته را با استفاده از این روش به یک بردار با ۲۵۶ خانه تبدیل کردند که هر خانه نماینده تعداد دفعات تکرار هر یک از حروفهای اسکی بود. سپس یک بسته طبیعی با استفاده از محاسبات آماری محاسبه‌شده و هر بسته جدید با محاسبه فاصله ماهالانوبیس[3] در یکی از دسته‌های طبیعی و یا ناهنجار قرار می‌گرفت. هرچند در نظریه می‌توان از n- گرام نیز استفاده نمود اما به‌دلیل پیچیدگی‌های محاسباتی به‌ندرت در موضوع شبکه استفاده می‌گردد. به‌عنوان مثال، درصورتی‌که از ۲- گرام استفاده شود، بردار خروجی شامل ۶۵۵۳۶ خانه خواهد بود. ایده استفاده از n- گرام‌ها علاوه‌بر دسته‌بندی ترافیک در تشخیص نفوذ نیز کاربرد دارد [۳۱].

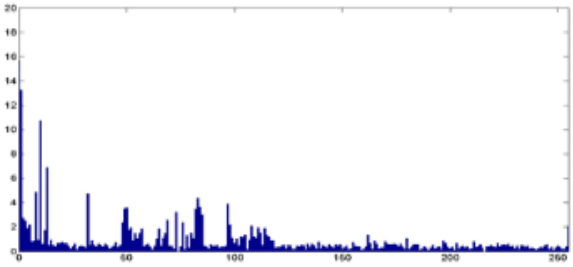
در این مقاله از این روش برای طبقه‌بندی نوع سرویس ترافیک استفاده شده است. بدین منظور، m بایت ابتدایی هر بسته را که ۵۱۲ بایت درنظر گرفته شده است استخراج نموده و تعداد تکرار هر حروف اسکی را محاسبه می‌نماییم. چون‌که برخی از سرویس‌ها امضای خود را در بسته مبدأ و مقصد، به‌عنوان داده‌های مجزا در نظر می‌گیرند. به‌عبارت دیگر، هر حروف اسکی دو مقدار دارد، یکی برای داده‌ای که سرویس‌گیرنده به سمت سرویس‌دهنده فرستاده و آغازکننده ارتباط است و یکی برای داده‌ای که از سمت سرویس‌دهنده به سرویس‌گیرنده ارسال می‌گردد. شکل (۳) نشان‌دهنده نمونه‌هایی از جریان http و تعداد تکرارهای حروفها برای ۶۴ بایت ابتدایی بسته‌ها است.

در اشکال (۳ و ۴) مقایسه ۱- گرام برخی سرویس‌ها نشان داده شده است. به‌طوری‌که محور x معرف کد اسکی حروفها و محور y معرف فراوانی تکرار آن‌ها است. همان‌گونه‌که مشخص است، نوع سرویس‌های متنی به‌راحتی از سرویس‌هایی مانند تورنت و یا سرویس‌های رمزشده قابل تمییز است. دلیل این است که در سرویس‌های مبتنی بر متن مانند وب قسمت بیش‌تری از داده‌ها مربوط به حروف الفبا هستند، اما به‌عنوان‌مثال، در داده‌های رمزگذاری توزیع یکنواخت‌تری وجود دارد، بنابراین این داده‌ها اطلاعات بسیار مفیدی برای تشخیص نوع سرویس است. برای دسته‌بندی نوع سرویس ابتدا یک بردار ۵۱۲ خانه‌ای تعریف می‌شود و در ۲۵۶ خانه‌ی اول آن تعداد تکرار حروفها در بار شبکه مبدأ و در ۲۵۶ خانه بعدی تعداد تکرار حروفهای اسکی در بار شبکه مقصد قرار داده می‌شود.

1- Traffic Classification
2- Data Encryption
3- Mahalanobis



شکل (۳): ۱-گرام سرویس Http



شکل (۴): ۱-گرام سرویس SMTP

در بردارهای بالا، تمامی حروفهای موجود دارای اهمیت یکسانی هستند. با این حال ممکن است در برخی از قوانین‌ها تعدادی از حروفها دارای اهمیت زیادتری بوده و برخی از آن‌ها اصلاً اهمیت نداشته باشند. بدین منظور، در بردار تعداد تکرار هر حروف در بسته به هریک از خانه‌ها وزنی را نسبت داده و از نرم‌افزار خاصی برای مدل‌کردن و محاسبه مجموعه وزن متناظر با الگوریتم یادگیری استفاده شده است. سپس با به‌کارگیری الگوریتم فرآیند تکاملی سعی می‌شود تا بهترین مجموعه وزن که منجر به کم‌ترین خطا در طبقه‌بندی می‌گردد را پیدا نماییم. بدین منظور، از یک روش مبتنی بر الگوریتم ژنتیک استفاده می‌شود که در ادامه توضیح داده خواهد شد.

۳-۱-۲- آموزش وزن‌ها با استفاده از الگوریتم ژنتیک

در این بخش، راه‌کاری مبتنی بر الگوریتم ژنتیک برای یافتن بهترین مجموعه وزن‌ها معرفی می‌گردد. الگوریتم ژنتیک یک الگوریتم غیرقطعی و تصادفی[1] است که براساس فرآیند تکاملی در دنیای واقعی بنا شده و قادر است مسائل پیچیده‌ای که در زمان کوتاهی به پاسخ نمی‌رسند را حل نماید. در این روش یک جمعیت از زیرمجموعه‌های کاندید تولید می‌کنیم. در هر بار تکرار الگوریتم، با استفاده از عملگرهای جهش و بازترکیبی بر روی عناصر جمعیت قبلی، عناصر جدیدی تولید شده و با استفاده از یک تابع ارزیابی[2]، میزان شایستگی عناصر جمعیت فعلی را مشخص کرده و عناصر بهتر را به‌عنوان جمعیت نسل بعد انتخاب می‌کنیم.

در این مقاله، از این روش برای تشخیص مجموعه وزن‌های بهینه برای هر خانه بردار تعداد تکرارها استفاده می‌گردد. بدین منظور مراحل زیر انجام می‌پذیرد:

تولید جمعیت اولیه: بدین منظور ابتدا تعدادی بردار به اندازه بردار تعداد تکرار حروفها ایجاد می‌نماییم. به‌عبارت‌دیگر، تعدادی بردار با ابعاد ۵۱۲ خانه ایجاد می‌شود. این بردارها همان ژن‌های مربوط به جمعیت اولیه هستند.

وزن‌دهی اولیه به هر موقعیت بردار تعداد تکرارها: پس از ایجاد جمعیت اولیه برای هر یک از ژن‌های ایجادشده (بردارهای تعداد تکرار حروفها) به هر یک از خانه‌های آن یک وزن تصادفی اختصاص می‌دهیم، به‌طوری‌که مجموع تمام وزن‌ها در بردار برابر یک باشد.

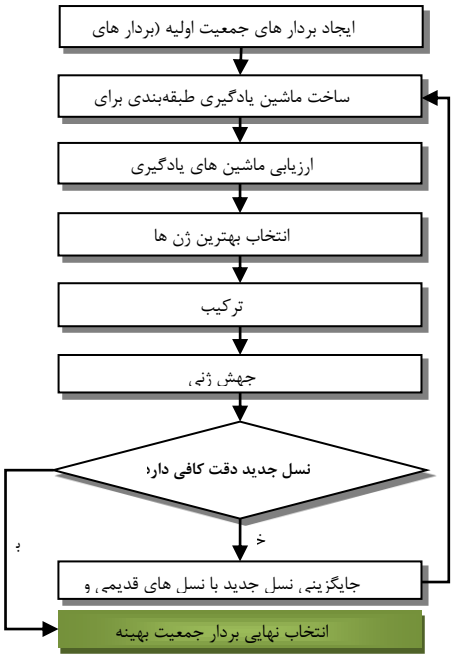
ارزیابی میزان کارایی هر ژن یا محاسبه تابع ارزیابی: در این مرحله با توجه به وزن‌های داده‌شده یک‌بار عملیات طبقه‌بندی را بر روی‌داده‌ها انجام می‌دهیم تا میزان کارایی هر ژن (هر بردار) مشخص شود. بدیهی است هرچه این میزان بیش‌تر باشد به معنی کارایی بالای مجموعه انتخاب شده است.

انتخاب بهترین اعضای مجموعه: در این مرحله دو ژن که دارای بیش‌ترین مقدار تابع ارزیابی هستند، انتخاب می‌شوند. به‌عبارت‌دیگر، بهترین اعضای گروه که داده‌ها را با بیش‌ترین دقت دسته‌بندی کرده‌اند، انتخاب می‌شوند. در مرحله بعد، دو ژن والد با هم ترکیب شده و دو ژن جدید تولید می‌نمایند. در مرحله جهش ژنی ممکن است برخی از وزن‌ها به‌صورت تصادفی تغییر نمایند. این مرحله براساس جهش ژنی در دنیای واقعی بنا شده است که به ژن‌ها اجازه می‌دهد در برخی موارد وزن‌هایی متفاوت با وزن‌های والدین خود داشته باشند. در مرحله بعد، پس از جهش ژنی، مجموعه ژن‌های ایجادشده به جمعیت اولیه اضافه می‌شوند و مراحل ۳ تا ۷ تا هنگامی‌که تعداد خاصی از نسل‌ها تولید شده و یا میزان خطا کم‌تر از یک مقدار خاص باشد ادامه می‌یابد.

شکل (۵) نمایی از سامانه وزن‌دهی به بردارهای تعداد تکرار حروف با استفاده از الگوریتم ژنتیک را نمایش می‌دهد.

1- Stochastic

2- Fitness Function



شکل (۵): ساختار وزن‌دهی با الگوریتم ژنتیک

هم‌چنین از دو روش برای آموزش و آزمون داده‌ها استفاده شده است. در روش اول، داده‌های مجموعه داده به‌صورت مستقیم و بدون پیش‌پردازش به سامانه تحویل داده شده است. در دومین روش، داده‌ها ابتدا پیش‌پردازش شده و پس از کاهش ابعاد دوباره مراحل آموزش و آزمون صورت پذیرفته است. برای کاهش ابعاد ویژگی‌ها و به‌منظور کاهش هزینه‌های محاسباتی و افزایش کارایی از روش‌های تحلیل مؤلفه‌های اصلی[1] (PCA) و تحلیل الگوی متمایز خطی[2] (LDA) استفاده شده است [۳۳-۳۲].

۳-۲- روش‌های پیاده‌سازی آشکارسازی مبتنی بر امضاء و مبتنی بر ناهنجاری

سامانه خبره برای پردازش حقایق و استنتاج نتایج منطقی از وقایع رخداده در سامانه با توجه به زنجیره‌ای از الگوها یا سناریوهای نفوذ استفاده می‌کند. از مزایای این روش می‌توان به ارائه حملات در قالب قواعد توسط کاربر بدون نیاز به دانستن نحوه عملکرد سامانه خبره و امکان اضافه‌کردن قواعد جدید بدون تغییر قواعد قبلی اشاره کرد ولی معایب آن کارایی پایین و از طرفی برای حجم زیاد داده‌ها و بیان ترتیب در قواعد، نامناسب است. به‌همین دلیل، از روش‌های مبتنی بر گذار حالت برای پیاده‌سازی واحد تشخیص مبتنی بر امضاء در این پژوهش استفاده شده است. این روش‌ها از مدل‌های گرافیکی نظیر شبکه‌های مارکوف و بیزی استفاده می‌کنند، علت استفاده از این مدل‌ها، استفاده از روش‌های انطباق الگو، سرعت و قابلیت آن‌ها است. واحدهای تشخیص مبتنی بر ناهنجاری با استفاده از مدل‌های یادگیری ماشین پیاده‌سازی می‌شوند. این مدل‌ها قادر هستند رفتارها را برحسب طبیعی و غیرطبیعی دسته‌بندی کنند. در روش تحلیل کمی، نمایه‌ها با معیارهای عددی بیان می‌شوند و در تحلیل آماری نمایه‌ها با معیارهای آماری با استفاده از انحراف معیار از یک توزیع طبیعی با میانگین مشخص بیان می‌شوند. در سامانه پیشنهادی از روش‌های داده‌کاوی به‌منظور دسته‌بندی رفتارهای طبیعی و غیرطبیعی استفاده شده است. روش‌های یادگیری ماشین در داده‌کاوی به‌طورکلی به دو دسته بدون نظارت و با نظارت تقسیم‌بندی می‌شوند.

استفاده از تکنیک‌های یادگیری ماشین در شرایطی مناسب است که هیچ‌گونه دانش اولیه در مورد الگوهای داده‌ها وجود ندارد، مدل ایجادشده دارای دانشی است که آن را از مجموعه داده‌های آموزش یاد گرفته است. این دانش ساختار داده را در بردارد و الگوهای موجود در آن‌را می‌شناسد. دانش به‌دست‌آمده برای مجموعه داده‌هایی که هیچ اطلاعاتی از آن‌ها در اختیار نیست آزمایش می‌شود. در این روش‌ها معمولاً به انسان‌های خبره برای تعیین ملزومات مورد نظر به‌منظور تشخیص نفوذ نیازی نیست. به همین دلیل، بسیار سریع عمل کرده و مقرون به‌صرفه هستند. البته در سال‌های اخیر ایده ترکیب سامانه‌های تشخیص ناهنجاری با سامانه تله عسل[3] به‌عنوان مکمل سامانه‌های تشخیص نفوذ برای هدایت ترافیک مشکوک به تله‌ها استفاده شده است. سازوکار این سامانه‌ها بدین گونه است که از اغفال و فریب مهاجم جهت جمع‌آوری اطلاعات بیش‌تر از نحوه عملکرد آن و به‌منظور جمع‌آوری بدافزارها استفاده می‌شود.

۳-۳- لایه تشخیص نفوذ

همان‌طورکه در شکل (۲) نشان داده شد، در لایه تشخیص نفوذ یک واحد آشکارساز مبتنی بر امضاء و واحدهای آشکارساز مبتنی بر ناهنجاری به شکل ترکیبی پیاده‌سازی شده‌اند.

1- Principle Component Analysis
2- Linear Discriminant Analysis
3- Honeypot

ابتدا واحد آشکارساز مبتنی بر امضاء فراخوانی می‌گردد و چنان‌چه جریان داده‌ای با الگوهای موجود در پایگاه داده حملات تطابق نداشته باشد با توجه به برچسب نوع سرویس، واحد آشکارساز ناهنجاری که از قبل با مجموعه داده آموزشی خاص همان سرویس یادگیری شده است، فراخوانی می‌گردد. در این‌جا منظور از واحد آشکارساز مدل یادگیری ماشینی است که از قبل با مجموعه داده آموزشی[1] مرتبط دانش مورد نیاز از رفتارهای طبیعی و غیرطبیعی را کسب نموده و قادر است رفتار داده‌های آزمایشی[2] را بررسی کند. چنان‌چه نوع سرویس شبکه غیرقابل شناسایی بوده باشد واحدهای آشکارساز مختلف که به‌صورت ترکیبی پیاده‌سازی شده‌اند بررسی می‌شوند. به‌واسطه معماری و سازوکار تشخیص در سامانه پیشنهادی، این امکان فراهم شده است که هر واحد آشکارساز ناهنجاری مختص یک نوع سرویس خاص شبکه مانند TCP، UDP، ICMP و سایر سرویس‌ها پیاده‌سازی شوند. از طرفی واحدهای آشکارساز با مجموعه داده آموزشی متناسب با هر نوع سرویس از قبل یادگیری شده‌اند که همین امر باعث بهبود کارایی سامانه هم از نظر دقت و هم از لحاظ سرعت پردازش می‌گردد و سپس نتیجه پردازش‌ها به‌منظور تصمیم‌سازی نوع پاسخ[3] و واکنش[4] به رویداد حمله به لایه تصمیم‌گیری ارسال می‌شود.

۳-۴- لایه تصمیم‌گیری

در این لایه متناظر با واحدهای آشکارساز نفوذ، واحدهای تصمیم‌ساز نفوذ طراحی شده است؛ یعنی متناسب با واحد آشکارساز نفوذ مبتنی بر امضاء یک واحد تصمیم‌ساز نفوذ وجود دارد و می‌توان در این واحد ماهیت نوع حمله را از پایگاه داده حملات استخراج نمود و نوع پاسخ مناسب را با توجه به تنظیماتی که از سوی مدیر شبکه تعیین شده است تعیین نمود. به‌عنوان نمونه، در حالت منفعل[5] صرفاً پیام هشدار از طریق سرویس اطلاع‌رسانی لایه مدیریت ثبت وقایع و هشدارها به مدیر شبکه ارسال و واکنش به رویداد حمله به آن‌ها واگذار می‌گردد و در حالت فعال[6] می‌تواند کد مخصوص به حذف نشست[7] حمله ایجادشده از سوی مهاجم[8] را فراخوانی و دسترسی آن را مسدود کند.

متناسب با ماژوهای آشکارساز مبتنی بر ناهنجاری نیز واحد تصمیم‌ساز جداگانه‌ای طراحی شده است. در این‌جا نیز چنان‌چه نوع سرویس شبکه از قبل شناسایی شده باشد، آشکارساز مناسب براساس برچسب نوع سرویس فراخوانی شده و نتیجه پردازش در صورت تشخیص ناهنجاری به پایگاه داده حملات اضافه می‌گردد. در این‌جا نیز متناسب با خط‌مشی تعیین‌شده از سوی مدیر شبکه نوع پاسخ و واکنش به رویداد حمله تصمیم‌سازی می‌شود اما درحالی‌که نوع سرویس شبکه قابل شناسایی نباشد جریان داده‌ای به گروهی از واحدهای آشکارساز با دقت بالا که به‌صورت ترکیبی پیاده‌سازی شده‌اند ارسال می‌گردد، سپس نتیجه پردازش واحدها جهت بررسی به واحد تصمیم‌ساز ارسال تا نسبت به نوع پاسخ و واکنش به رویداد حمله در صورت تشخیص نفوذ، تصمیم‌سازی و متناسب با نوع پاسخ تعیین‌شده لایه مدیریت ثبت وقایع و هشدارها فراخوانی می‌گردد.

۳-۴-۱- واحد تصمیم‌ساز نفوذ

در این بخش به بررسی واحد تصمیم‌ساز نفوذ می‌پردازیم. نوع پاسخ و یا نوع واکنش به رویداد حمله از سوی مدیر شبکه تعیین می‌گردد و توسط واحد تصمیم‌سازی این تنظیمات اعمال می‌گردد.[9] برای این منظور مدیر شبکه قادر است تنظیمات مورد نظر خود را در قالب خط‌مشی واکنش به رویداد حمله‌ها تعریف و از طریق واحدهای تصمیم‌ساز نفوذ آن‌ها را اعمال کند به‌عنوان یک نمونه کاربردی ویژگی ضریب هشدار نفوذ متناسب با حساسیت اطلاعات ترافیک شبکه محلی توسط مدیر شبکه تنظیم می‌گردد. در این پژوهش از مدل‌های یادگیری ماشین که با مجموعه داده‌های NSL-KDD از قبل آموزش دیده‌اند استفاده شده است؛ بنابراین با فرض این‌که این مدل‌ها براساس معیار دقت مرتب شده باشند نتیجه پردازش واحدهای آشکارساز (مدل‌های یادگیری ماشین) به‌ترتیب اولویت به لایه تصمیم‌گیری ارسال می‌گردد و برحسب میزان ضریب هشدار نفوذ تعیین‌شده واحد تصمیم‌ساز درصورت تشخیص ناهنجاری تا با وزن‌دهی به واحدهای آشکارساز و دخیل‌نمودن ضریب هشدار نفوذ سپس در صورت رأی اکثریت نسبت به پاسخ هشدار مناسب تصمیم‌سازی می‌گردد.

1- Training Data Set
2- Test Data Set
3- Response Type
4- Reaction
5- Passive
6- Active
7- Session
8- Attacker
9- Policy Enforcement

لازم به توضیح است که هرچقدر درصد ضریب هشدار نفوذ کم‌تر درنظر گرفته شود هشدارها با درصد نفوذ خیلی پایین اعلام می‌شوند و چنان‌چه این ضریب مقدار متوسط به بالا تعیین گردد هشدارهای نفوذ با درصد دقت خیلی بالاتری به لایه مدیریت ثبت وقایع و هشدار ارسال می‌گردد.

بدیهی است که هرچقدر ضریب هشدار نفوذ متناسب با ترافیک شبکه، دقیق‌تر تعیین گردد نرخ مثبت کاذب بسیار کم‌تری را در پی خواهد داشت. البته تصمیم‌سازی می‌تواند براساس رأی اکثریت بدون دخالت ضریب هشدار نفوذ و یا با دخالت ضریب هشدار نفوذ باشد.

۳-۴-۲- واکنش به رویداد نفوذ

به فاصله زمانی بین رخداد وقایع در منبع اطلاعات تا تحلیل آن‌ها توسط واحدهای آشکارساز، زمان‌بندی[1] می‌گویند. در سامانه پیشنهادی درصورتی‌که نوع آشکارسازی نفوذ، مبتنی بر امضاء باشد تشخیص نفوذ به محض وقوع و یا حتی قبل از آن، امکان پاسخ‌گویی فعال و پیش‌گیری از نفوذ وجود دارد و از زمان‌بندی بی‌درنگ[2] استفاده می‌شود ولی اگر نوع واحد آشکارساز، مبتنی بر ناهنجاری باشد درصورت کشف نفوذ پس از وقوع، امکان پاسخ‌گویی فعال وجود نداشته و از زمان‌بندی دوره‌ای[3] استفاده می‌گردد و در این شرایط صرفاً پیغام هشدار نفوذ به مدیر شبکه ارسال می‌گردد. در سامانه پیشنهادی، درصورتی‌که ماهیت نوع حمله مشخص شده باشد و نوع پاسخ[4] از سوی مدیر شبکه فعال درنظر گرفته شود برخی اعمال واکنشی نیز به‌صورت خودکار انجام می‌گردد به‌عنوان نمونه دسترسی مهاجم مسدود می‌شود و درصورتی‌که ماهیت حمله مشخص نباشد و نوع پاسخ منفعل تنظیم شده باشد پیغام هشدار در قالب نمایش بر روی صفحه[5]، ارسال پست الکترونیکی[6] و پیام کوتاه[7] به مدیران شبکه ارسال تا نسبت به واکنش مناسب به رویداد حمله، خود تصمیم‌گیری نمایند.

۳-۵- لایه مدیریت ثبت وقایع و هشدارها

گرچه تلاش‌های زیادی برای کاهش نرخ هشدارهای کاذب توسط سامانه‌های تشخیص نفوذ صورت گرفته است، اما با توجه به طبیعت پویای شبکه‌های کامپیوتری، داشتن سامانه تشخیص نفوذ بدون هیچ هشدار کاذبی دور از انتظار نیست. چراکه می‌توان نرخ هشدارهای کاذب را با جمع‌آوری تاریخچه[8] و پردازش انواع مختلفی از اطلاعات از منابع گوناگون مثل سامانه‌های تشخیص نفوذ، آنتی‌ویروس‌ها، رویدادنامه‌های[9] سامانه‌های عامل، رویدادنامه‌های سرویس‌های تحت وب کاهش دهیم. پیاده‌سازی یک واحد تحلیل رویداد بلادرنگ[10] از هشدارهای امنیتی، برای تحلیل هم‌بستگی[11] بین رویدادهای ثبت‌شده با هدف کاهش حجم هشدارها و اعلان‌ها، وارسی صحت هشدارها و استخراج حملات چندمرحله‌ای براساس جمع‌آوری اطلاعات و هم‌بسته‌سازی هشدارها[12] برای افزایش نرخ دقت تشخیص نفوذ و کاهش نرخ هشدارهای مثبت کاذب به‌صورت توأمان قابل قبول است. در این پژوهش به‌منظور پاسخ و واکنش به رویداد حمله از روش فعال[13] برای روش مبتنی بر امضاء و غیرفعال[14] برای روش مبتنی بر ناهنجاری متناسب با نوع پاسخ تعیین‌شده در لایه تصمیم‌گیری استفاده شده است و ضمن اطلاع‌رسانی هشدارها به مدیر شبکه، در صورت نیاز، اعمال واکنشی و اقدامات امنیتی لازم نیز انجام خواهد شد.

۴- انتخاب ویژگی با الگوریتم ژنتیک

الگوریتم ژنتیک یکی از روش‌هایی است که کاربرد زیادی در تشخیص نفوذ دارد. یکی از قسمت‌هایی که می‌توان از الگوریتم ژنتیک استفاده کرد، انتخاب ویژگی[15] است. از الگوریتم ژنتیک برای انتخاب ویژگی‌های مفید استفاده شده و ویژگی‌های انتخاب‌شده با استفاده از یک درخت تصمیم‌گیری به دسته‌های طبیعی و یا ناهنجار تقسیم شده‌اند [۳۴].

در این روش یک جمعیت از زیرمجموعه‌های کاندید تولید می‌کنیم، یعنی یک مجموعه ویژگی داده‌شده با یک‌رشته باینری به طول n نمایش داده می‌شود که با یک صفر یا یک در محل iام، بودن یا عدم حضور ویژگی i را در مجموعه نشان داده می‌شود. n تعداد کل ویژگی‌های در دسترس است، یعنی جمعیتی از کروموزوم‌ها نگهداری می‌شود. در هر بار تکرار الگوریتم با استفاده از یک تابع ارزیابی، میزان شایستگی عناصر جمعیت فعلی را

1- Timing
2- Real-Time
3- Batch
4- Response Type
5- Monitoring
6- Email
7- SMS

8- History
9- Logs
10- Real-time
11- Correlation
12- Alert Correlation
13- Active
14- Passive
15- Feature Selection

مشخص کرده و عناصر بهتر را به‌عنوان جمعیت نسل بعد انتخاب می‌کنیم. کروموزوم‌های جدید طبق فرآیند زیر از کروموزوم‌های قدیمی به وجود می‌آیند:

۱) باز ترکیبی[1] که بخش‌هایی از دو کروموزوم والد با هم ترکیب می‌شوند تا فرزندهای جدیدی ایجاد کنند.

۲) جهش[2] که بیت‌های یک والد به‌صورت تصادفی تخریب می‌شوند تا فرزند جدیدی ایجاد کند.

پیداشدن بهترین جواب در این روش تضمین نمی‌شود، ولی همیشه یک جواب خوب به نسبت مدت زمانی که به الگوریتم اجازه اجرا داده باشیم، پیدا می‌کند. روند این مرحله نیز تا حدودی همانند روند الگوریتم ژنتیک توضیح داده‌شده در بخش طبقه‌بندی ترافیک هست که در شکل (۵) نشان داده شده است.

۵- ارزیابی نتایج

مطابق شکل (۶) ماتریس درهم‌ریختگی چگونگی عملکرد الگوریتم دسته‌بندی را با توجه به مجموعه داده ورودی به تفکیک انواع دسته‌های مسئله دسته‌بندی، نمایش می‌دهد.

مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی دقت یا نرخ دسته‌بندی[3] است که این معیار دقت کل یک دسته‌بند را محاسبه می‌کند. در واقع این معیار مشهورترین و عمومی‌ترین معیار محاسبه کارایی الگوریتم‌های دسته‌بندی است که نشان می‌دهد دسته‌بند طراحی‌شده چند درصد از کل مجموعه رکوردهای آزمایشی را به‌درستی دسته‌بندی کرده است.

تعریع معیارهای مورد نیاز در ارزیابی:

مثبت واقعی[4] (TP): رویداد تحلیل به‌طور صحیح به‌عنوان نفوذ تشخیص داده شود.

مثبت کاذب[5] (FP): رویداد تحلیل از منظر امنیتی بی‌خطر است اما به‌عنوان مخرب تشخیص داده شود.

منفی واقعی[6] (TN): رویداد تحلیل به‌طور صحیح به‌عنوان بی‌خطر و طبیعی تشخیص داده شود.

منفی کاذب[7] (FN): رویداد تحلیل مخرب است، اما به‌عنوان بی‌خطر تشخیص داده شود.

وضعیت واقعی		نتیجه‌ی ارزیابی: مثبت	نتیجه‌ی ارزیابی: منفی
	مثبت	مثبت درست	منفی نادرست
	منفی	مثبت نادرست	منفی درست

شکل (۶): ماتریس درهم‌ریختگی

معیار دقت یا Accuracy از طریق فرمول زیر محاسبه می‌گردد:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

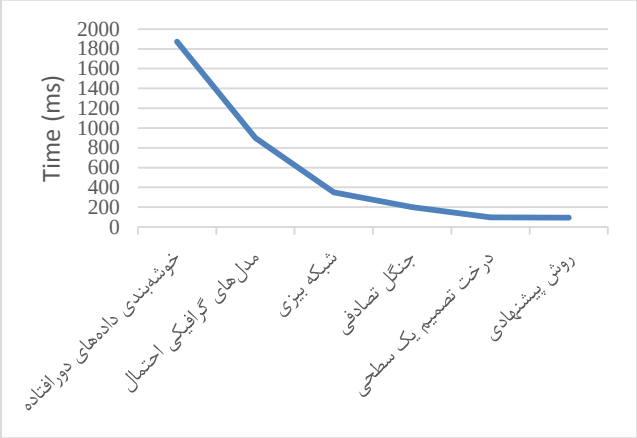
به‌منظور مقایسه روش پیشنهادی با سایر روش‌ها، از معیار دقت استفاده می‌نماییم. همان‌طورکه در جدول (۲) مشاهده می‌شود، روش پیشنهادی به‌دلیل استفاده از طبقه‌بندی ترافیک شبکه و آموزش چند ماشین یادگیری ترکیبی براساس طبقه‌بندی نوع سرویس‌ها، دارای نرخ مثبت کاذب بسیار کمی است. هم‌چنین دقت روش پیشنهادی در مقابل سایر روش‌ها، بالاتر است.

جدول (۲): مقایسه میزان دقت روش پیشنهادی با سایر روش‌ها

نام روش	TP	FP	ACC	Time(MS)
خوشه‌بندی داده‌های دورافتاده	۵۲/۳۴	۳۴/۷	۵۵/۷۸	۱۸۷۳
مدل‌های گرافیکی احتمال	۷۷/۵۶	۲/۰۷	۷۸/۸۳	۹۰۱
شبکه بیزی	۷۸/۱۲	۱/۷۷	۷۷/۳۶	۳۴۹
جنگل تصادفی	۷۶/۷۸	۱/۲۱	۷۹	۲۰۲
درخت تصمیم یک سطحی	۷۶/۸۷	۱/۲۱	۷۹	۹۸
روش پیشنهادی	۹۹/۸۳	۰/۰۹	۹۹/۸۱	۹۵

همان‌طورکه در نمودار (۱) نشان داده شده است زمان محاسبه روش پیشنهادی در مقایسه با سایر روش‌ها بسیار کم‌تر است و این به‌دلیل جداسازی و طبقه‌بندی ترافیک شبکه و کاهش ابعاد ویژگی قبل از تشخیص نفوذ است.

1- Crossover
2- Mutation
3- Classification Accuracy - Rate
4- True Positive Rate
5- False Positive Rate
6- True Negative Rate
7- False Negative Rate



نمودار (۱): نمودار مقایسه زمان محاسبه روش‌های مختلف با روش پیشنهادی

با استفاده از روش کاهش ابعاد ویژگی، مجموعه ویژگی‌های انتخاب‌شده با توجه به نوع حمله‌ها مطابق جدول (۳) دسته‌بندی می‌گردد.

جدول (۳): دسته‌بندی ویژگی‌ها، متناسب با هر حمله [۳۶]

نوع حمله‌ها	ویژگی‌های انتخاب‌شده
DoS	۳-۵-۲۴-۳۰
Probe	۱-۳-۵-۶-۲۳-۳۰
R2L	۱-۳-۵-۶-۱۰-۲۳-۲۴-۳۲
U2R	۳-۵-۶-۱۰-۱۱-۱۴-۱۷-۱۸-۳۲-۳۵

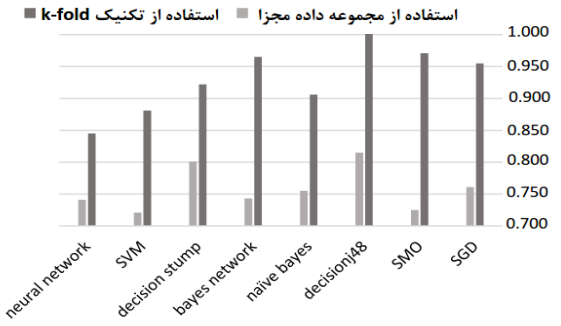
جدول (۴) تأثیر کاهش ابعاد ویژگی را در زمان محاسبه واحدهای آشکارسازی نفوذ نشان می‌دهد، همان‌طورکه مشاهده می‌گردد بعد از کاهش ابعاد ویژگی زمان محاسبه تشخیص نفوذ ۶۱/۸۴٪ کاهش یافته است.

جدول (۴): مقایسه زمان محاسبه قبل و بعد از کاهش ابعاد ویژگی

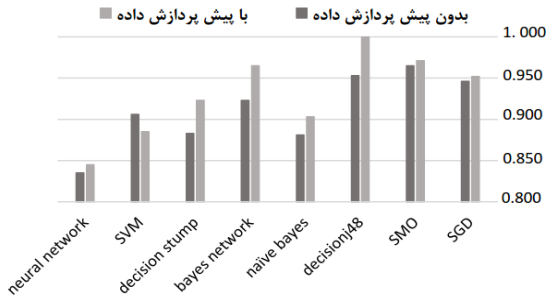
	قبل از کاهش ویژگی	بعد از کاهش ویژگی	زمان محاسبه
زمان (ثانیه)	۲۴۹	۹۵	84%.61

برای بررسی تأثیر مجموعه داده‌ها، یک‌بار از مجموعه داده مجزا[1] استفاده شده و در روش دیگر از روش k-fold با k=10 استفاده شده است. همان‌گونه‌که در نمودار (۱)نشان داده شده است، در صورت استفاده از روش k-fold با استفاده از روش پیشنهادی، درخت تصمیم‌گیری j48 به‌دقت بسیار بالای ۹۹،۸۱ رسیدیم.

1- Holdout

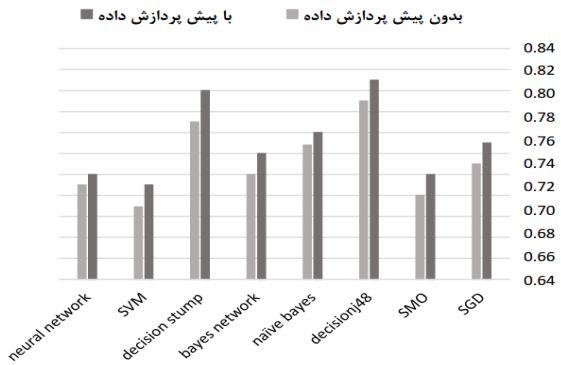


نمودار (۲): تأثیر انتخاب مجموعه داده‌ها بر نتایج نهایی روش پیشنهادی



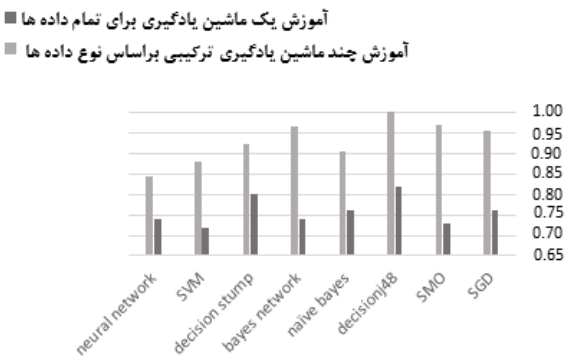
نمودار (۳): تأثیر پیش‌پردازش داده‌ها درصورت استفاده از 10-fold

آن‌چه از نمودار (۲) پیدا است، این‌که سطح دشواری پایگاه داده NSL-KDD به‌دلیل استفاده از مجموعه داده‌های آموزش و آزمون متفاوت بیش‌تر از سایر مجموعه‌هاست و می‌تواند مرجع جامع و مناسب‌تری برای ارزیابی الگوریتم‌های مختلف باشد. هم‌چنین در نمودار (۳) نتیجه پیش‌پردازش داده‌ها به‌طور مؤثری در نتایج داده‌ها تأثیر داشته است. هم‌چنین مطابق نمودار (۴) درصورت استفاده از مجموعه داده‌های مجزا در روش‌های مبتنی بر درخت‌های تصمیم‌گیری و مدل‌های گرافیکی، تأثیر بیش‌تری داشته است، اما در برخی روش‌ها ازجمله شبکه‌های عصبی و ماشین بردار پشتیبان تأثیر زیادی نداشته است.



نمودار (۴): تأثیر پیش‌پردازش داده‌ها درصورت استفاده از مجموعه داده‌های مجزا

متأسفانه، به‌دلیل عدم دسترسی به مجوزهای لازم برای جمع‌آوری داده‌های واقعی از شبکه‌های مختلف، امکان جمع‌آوری داده‌های واقعی و آموزش جامع سامانه پیشنهادشده به‌صورت واقعی وجود ندارد. از سوی دیگر، مجموعه داده‌ای که بتوان هر دو قسمت طبقه‌بندی ترافیک و تشخیص نفوذ را باهم پیاده نمود موجود نیست. با این حال، به‌منظور بررسی کارایی سامانه پیشنهادشده، ابتدا داده‌های مجموعه داده معرفی‌شده را براساس نوع سرویس جدا و برای هر نوع داده یک ماشین یادگیری تعریف کرده‌ایم و در نهایت، نتایج به‌دست‌آمده را ترکیب نموده‌ایم. نمودار (۴) نتایج حاصل را نمایش می‌دهد. همان‌گونه‌که مشاهده می‌گردد تمامی روش‌های بررسی‌شده با استفاده از روش پیشنهادی که تشخیص نوع سرویس، یادگیری یک ماشین مجزا برای هر نوع سرویس، بهبود عملکرد داشته‌اند.



نمودار (۵): مقایسه روش پیشنهادی

همچنین به‌منظور بررسی الگوریتم ارائه‌شده با سایر روش‌ها، برخی از کارهای تحقیقاتی که از مجموعه داده‌های NSL-KDD استفاده نموده‌اند، انتخاب شده است. به‌منظور معتبربودن نتایج و برقراری شرایط اولیه یکسان در مقایسه‌ها، داده‌های آموزشی مجموعه داده NSL-KDD به‌منظور آموزش و آزمون انتخاب گردیده است. جدول (۵) بهترین نتیجه به‌دست‌آمده در هر یک از مقالات را نشان می‌دهد. همان‌گونه که مشاهده می‌گردد روش ذکرشده در این مقاله در صورت استفاده از مجموعه داده آموزش NSL-KDD، دارای بیش‌ترین دقت است.

جدول (۵): مقایسه روش پیشنهادشده با برخی از مقالات دیگر

نام روش	درصد دقت
Multinomial Naïve Bayes[8]	۹۶/۶
PCA+SVM[35]	۹۱/۵۴
Decision tree j48 [13]	۹۶/۳
Neural Network[14]	۸۷/۶
روش پیشنهادی	۹۹/۸۱

۶- نتیجه‌گیری

در این مقاله یک سامانه ترکیبی چهارلایه برای تشخیص نفوذ در شبکه‌های کامپیوتری ارائه شده است. این سامانه در لایه اول ترافیک شبکه را تحلیل و سپس آن را براساس نوع سرویس طبقه‌بندی می‌نماید. در لایه تشخیص نفوذ، یک واحد آشکارساز مبتنی بر امضاء و واحدهای آشکارساز مبتنی بر ناهنجاری به شکل ترکیبی پیاده‌سازی شده‌اند که متناسب با برچسب نوع سرویس‌ها فراخوانی می‌شوند. سپس، درنتیجه پردازش این واحدها، لایه تصمیم‌گیری فراخوانی می‌شود. این لایه ماهیت حمله و نوع پاسخ را تشخیص داده و لایه مدیریت وقایع را فرا می‌خواند. در این لایه ضمن اطلاع‌رسانی هشدارها به مدیر شبکه، در صورت نیاز، اعمال واکنشی و اقدامات امنیتی لازم نیز انجام خواهد شد. برای شبیه‌سازی سامانه تشخیص نفوذ پیشنهادی از مجموعه داده NSL-KDD استفاده شده است. نتایج حاصل از ارزیابی اعتبارسنجی چندلایه‌ای سامانه پیشنهادی، دقت بسیار بالای ۹۹/۸۱ و کاهش نرخ مثبت کاذب ۰/۰۹ را نشان داده است.

۷- مراجع

[1] M. Joshi, R. Agarwal, and A. V. Kumar, "Predicting rare classes:can boosting make any weak learner strong," Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, vol. 306, p. 297, 2002.

[2] K. Gisung, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," Expert Systems with Applications, vol. 41, no. 4, pp. 1690-1700, 2014.

[3] O. Depren, M. Topallar, E. Anarim, and M. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks," Expert systems with Applications, vol. 29, no. 4, pp. 713-722, 2005.

[4] C. Xiang and S. Lim, "Design of multiple-level hybrid classifier for intrusion detection system," IEEE Workshop on Machine Learning for Signal Processing, pp. 117-122, 2005.

[5] M. Sabhnani and G. Serpen, "Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context," In International Conference on Machine Learning, Models, Technologies and Applications, pp. 209-215, 2003.

[6] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," Information Sciences, vol. 177, no. 18, pp. 3799-3821, 2007.

[7] L. Hung-Jen et al., "Intrusion detection system: A comprehensive review," Journal of Network and Computer Applications, vol. 36, no. 1, pp. 16-24, 2013.

[8] K. Levent, T. A. Mazzuchi, and S. Sarkani, "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier," Expert Systems with Applications, vol. 18, no. 39, pp. 13492-13500, 2012.

[9] X. Liyuan and Y. Chen, "Bayesian model averaging of bayesian network classifiers for intrusion detection,"

Computer Software and Applications Conference Workshops, IEEE 38th International, 2014.

[10] Hoque, M. Sazzadul et al., "An implementation of intrusion detection system using genetic algorithm," arXiv preprint arXiv, pp. 1204-1336, 2012.

[11] H. Mostaque, "Current studies on intrusion detection system, genetic algorithm and fuzzy logic,"| arXiv preprint arXiv, pp. 1304-3535, 2013.

[12] Muniyandi, A. Prabakar, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading k-Means clustering and C4. 5 decision tree algorithm," Procedia Engineering, vol. 30, pp. 174-182, 2012.

[13] S. Shailendra and B. M. Mehtre, "Network intrusion detection system using j48 decision tree," Advances in Computing, Communications and Informatics, International Conference on IEEE, 2015.

[14] S. Devaraju and S. Ramakrishnan, "Performance comparison for intrusion detection system using neural network with KDD dataset," ICTACT Journal on Soft Computing, vol. 4, no. 3, pp. 743-752, 2014.

[15] A. Yousef and et al., "Flow-based anomaly intrusion detection system using two neural network stages," Compute. Sci. Inf. Syst., vol. 11, no. 2, pp. 601-622, 2014.

[16] O. Chung-Ming, "Host-based intrusion detection systems adapted from agent-based artificial immune systems," Neurocomputing, vol. 88, pp. 78-86, 2012.

[17] E. Tombini, H. Debar, L. Me, M. Ducasse, F. Telecom, and F. Caen, "A serial combination of anomaly and misuse IDSes applied to HTTP traffic," In Proceedings of the 20th Annual Computer Security Applications Conference, pp. 428-437, 2004.

[18] J. Zhang and M. Zulkernine, "A hybrid network intrusion detection technique using random forests," In Proceedings of the First International Conference on Availability, Reliability and Security (ARES), p. 8, 2006.

[19] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," Information Sciences, vol. 177, no. 18, pp. 3799-3821, 2007.

[20] S. Peddabachigari, A. Abraham, C. Grosan, and J. Thomas, "Modeling intrusion detection system using hybrid intelligent systems," Journal of Network and Computer Applications, vol. 30, no. 1, pp. 114-132, 2007.

[21] K. Gummadi, R. Dunn, S. Saroiu, S. Gribble, H. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," ACM SIGOPS Operating Systems Review, vol. 37, no. 5, pp. 314-329, 2003.

[22] S. Sen, O. Spatscheck and D. Wang, "Accurate, scalable in-network identi_cation of p2p tra_c using application signatures," In Proceedings of the 13th international conference on World Wide Web, pp. 512-521, 2004.

[23] L. Bernaille and R. Teixeira, "Early recognition of encrypted applications," In Proceedings of the 8th International Conference on Passive and Active Network Measurement, pp. 165-175, 2007.

[24] C. W. Dewes and A. Feldmann, "An analysis of internet chat systems," In Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement, pp. 51-64, 2003.

[25] V. Silvio and et al., "Reviewing traffic classification, Data Traffic Monitoring and Analysis," Springer Berlin Heidelberg, pp. 123-147, 2013.

[26] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," IEEE/ACM Transactions on Networking (TON), vol. 3, no. 3, pp. 226-244, 1995.

[27] D. Alberto, A. Pescape, and C. Kimberly, "Issues and future directions in traffic classification," IEEE network, vol. 26, no. 1, pp. 35-40, 2012.

[28] L. Bernaille, R. Teixeira, and K. Salamatian, "Early application identification," ACM Conference on Emerging Network Experiment and Technology, 2006.

[29] Shrivastav and A. Tiwari, "Network traffic classification using semi-supervised approach," Machine Learning and Computing (ICMLC), Second International Conference on. IEEE, 2010.

[30] M. Damashek, "Gauging similarity with n-grams: Language-independent categorization of text," Science, vol. 267, p. 843, 1995.

[31] K. Wang and S. Stolfo, "Anomalous payload-based network intrusion detection," Lecture Notes in Computer Science, pp. 203-222, 2004.

[32] D. l. Hoz, Eduardo and et al., "PCA filtering and probabilistic SOM for network intrusion detection," Neurocomputing, vol. 164, pp. 71-81, 2015.

[33] K. Fangjun, X. Weihong, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," Applied Soft Computing, vol. 18, pp. 178-184, 2014.

[34] B. Senthilnayaki, K. Venkatalakshmi, and A. Kannan, "An intelligent intrusion detection system using genetic based feature selection and Modified J48 decision tree classifier," Fifth International Conference on Advanced Computing (ICoAC).IEEE, 2013.

[35] A. Heba and F. Eid, "Principle components analysis and support vector machine based intrusion detection system," International Conference on Intelligent Systems Design and Applications.IEEE, 2010.

[36] N. Sharma, "A Novel Multi-Classifier Layered Approach to Improve Minority Attack Detection in IDS," in 2nd International Conference on Communication, Computing & Security, 2012.

[37] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network Anomaly Detection: Methods, Systems and Tools," IEEE Communications Surveys & Tutorials, vol. 16, no. 1, 2014.

Journal Of Electronical & Cyber Defence
Vol. 5, No. 3, 2017, Serial No. 19

A New Approach to Network Intrusion Detection Based on Hybrid Methods

S. Parsa*, S. H. R. Aarabi

*Iran University of Science and Technology

(Received: 14/12/2016, Accepted: 13/02/2017)

ABSTRACT

The role of intrusion detection systems has been considered significant in network anomaly detection. New and unknown attacks have proved that signature-based detection methods are inefficient, and raised the attention to anomaly-based detection methods. Despite their great ability in anomaly detection, these methods suffer from high rate of false-alarms. Therefore, the idea of using hybrid intrusion detection systems is developed in order to reduce the false-alarm rate. In this paper, we propose a four-layered model based on hybrid methods. The first layer consists of data flow analysis and service type classification modules. The service type classifier uses both an n-gram-based statistical technique, and an evolutionary algorithm. In the intrusion detection layer, a signature-based and several anomaly-based detection modules have been implemented with hybrid methods. These specific detection modules are called according to the type of service which has been identified through the first layer. The decision-making layer is then called based on the results of intrusion detection process. This layer identifies the attack nature and the type of response, and then calls the event management layer. In this layer, network administrator is notified appropriately; and, responsive actions are managed if needed. Applying the cross-validation method shows that intrusion detection has been improved and, in result, the false alarm rate has been reduced.

Keywords: Intrusion Detection, False-Positive Rate, Event Log, Cross Validation, Traffic Classification, Notification Service

* Corresponding Author Email: parsa@iust.ac.ir