جداسازی حملات در سامانه‌های تشخیص نفوذ با استفاده از فنون ماشین بردار پشتیبان و شبکه‌های عصبی خودسازمان‌ده

محمد نظری فرخی[۱]
ابراهیم نظری فرخی[۲]
سعید سعیدی[۳]
نرگس صالح پور[۴]

تاریخ دریافت: ۱۳۹۶/۰۶/۱۶ تاریخ پذیرش: ۱۳۹۶/۰۸/۱۱

چکیده

با توجه به رشد شبکه‌های رایانه‌ای، امنیت شبکه نیز به‌عنوان یک چالش بزرگ مطرح شده است. سامانه‌های تشخیص نفوذ برای اطمینان از پردازش و ذخیره امن داده‌ها بر روی شبکه توسعه داده شد؛ همچنین به‌عنوان یک مؤلفه اصلی برای امنیت شبکه محسوب می‌شوند. ازآنجاکه سامانه‌های تشخیص نفوذ سنتی پاسخ‌گوی حملات جدید نیستند؛ ازاین‌رو، سامانه‌های تشخیص نفوذ مبتنی بر داده‌کاوی در این پژوهش پیشنهاد می‌شوند. استفاده از فنون داده‌کاوی برای افزایش دقت و درستی سامانه‌های تشخیص نفوذ است و طبیعی است که باعث افزایش امنیت شبکه می‌شود. به‌دلیل گستردگی فنون داده‌کاوی، روش‌های شبکه‌های عصبی خودسازمان‌ده و ماشین بردار پشتیبان به‌منظور تشخیص و پیش‌بینی نفوذ در این تحقیق مورد استفاده قرار گرفته است؛ بنابراین، ترکیب دو فن یادشده، تشخیص ناهنجاری را تا حدودی بهبود می‌دهد. باوجوداین با استفاده از شبکه‌های عصبی خودسازمان‌ده و ماشین بردار پشتیبان می‌توان رفتارهای عادی و ترافیک شبکه را در یک دسته و حملات یا نرون‌های ناهمگن را نیز در دسته‌ی دیگر طبقه‌بندی نمود. برای آموزش و ارزیابی روش پیشنهادی از مجموعه داده‌ی KDD CUP 99 استفاده شده است. به عبارتی این روش به دلیل استفاده از یادگیری رقابتی برای آموزش در این پژوهش مورد استفاده قرار گرفته است؛ بنابراین، دقت روش پیشنهادی را با فن یادگیری مبتنی بر شبکه‌های عصبی خودسازمان‌ده و درخت تصمیم-گیری مقایسه می‌کند، نتایج نشان می‌دهد سامانه پیشنهادی دارای دقت بالا برای تشخیص نفوذ است هم‌چنین زمان کمتری نیز نسبت فن‌های مورد مقایسه دارا است.

واژگان کلیدی: *سامانه‌های تشخیص نفوذ، ماشین بردار پشتیبان، شبکه‌های عصبی خودسازمان‌ده، حملات، داده‌کاوی.*

---

۱. کارشناس ارشد مهندسی نرم افزار علوم و تحقیقات لرستان(mnf1365@chmail.com)
۲. کارشناس ارشد مدیریت فناوری اطلاعات دانشگاه علوم و تحقیقات تهران(e60_itmgtn@yahoo.com)
۳. کارشناس ارشد اطلاعات و حفاظت‌اطلاعات و عضوهئیت علمی دانشکده علوم و فنون فارابی، نویسنده مسئول (shamsshams202@gmail.com)
۴. مهندسی کامپوتر علوم و تحقیقات لرستان(Salehpour_narges@yahoo.com)

مقدمه

ایدهٔ اصلی سامانه‌های تشخیص نفوذ ابتدا در سال ۱۹۸۰ توسط جیمز اندرسون[1] ارائه شد. او برای اولین بار سوءاستفاده‌های رایانه‌ای را از مسیرهای ممیزی که با الگوهای رفتاری مطابقت داشت، ردیابی نمود. در دنیای امروزی، رایانه و شبکه‌های رایانه‌ای متصل به اینترنت نقش عمده‌ای در ارتباطات و انتقال اطلاعات ایفا می‌کنند. در این میان افراد سودجو با دسترسی به اطلاعات مهم قصد نفوذ و به هم ریختن نظم سامانه‌ها را در پیش‌گرفته‌اند. هک، کرک و نفوذ اصطلاحاتی هستند که امروزه در شبکه‌های رایانه‌ای، ازجمله خطرات نفوذ به شبکه‌های رایانه هستند که به محرومیت از خدمات وب می‌انجامد. امنیت شبکه‌های رایانه‌ای، نیز یکی از سازوکارهای تشخیص نفوذ است که برای کشف دسترسی‌های غیرمجاز در شبکه در تلاش است. دسترسی‌های غیرمجاز در شبکه‌های رایانه‌ای ازجمله فعالیت‌های مخرب، مانند حملات ویروس، دسترسی غیرمجاز و سرقت اطلاعات را مورد تجزیه‌وتحلیل قرار می‌دهند. البته در چند سال اخیر امنیت سایبری (رایانه‌ای) نیز موردتوجه قرارگرفته است و با فعالیت‌هایی که امروزه به‌صورت غیرقانونی در جهان شبکه در حال رشد هستند، امنیت شبکه را به‌عنوان یک چالش بزرگ در نظر می‌گیرند. امنیت سامانه‌های رایانه‌ای به دودستهٔ فعال و غیرفعال تقسیم می‌شوند. امنیت فعال شامل محافظت و واکنش در هنگام حمله است. دیوار آتش مثال خوبی برای امنیت فعال است. در امنیت فعال، برای جلوگیری از سوءاستفاده از یک سرویس یا اتصال خاص، دسترسی به آن‌ها پالایه می‌شود. اما در امنیت غیرفعال، اگر رخدادهای بدی در سامانه اتفاق بیفتد، فقط آگاه‌سازی صورت می‌گیرد. یعنی هیچ واکنش هم‌زمانی برای محافظت صورت نمی‌گیرد. سامانه تشخیص نفوذ مثالی از امنیت غیرفعال است. سامانه تشخیص نفوذ در مورد چگونگی حمله و یا این‌که چه کسی سعی در نفوذ به سامانه را دارد، اطلاعاتی در اختیار کاربر قرار می‌دهد. با توجه به بررسی‌های انجام‌شده در مرکز سی. إس. آی[2]، در دوازدهمین اجلاس سالانه جرائم رایانه‌ای و بررسی امنیت که با حضور بیش از ۴۰۰ شرکت پیشگام درزمینهٔ امنیت سامانه‌های رایانه‌ای در مؤسسات، نهادهای حکومتی، دانشگاه و... در ایالت متحده آمریکا انجام‌شده بود. می‌توان دریافت که با افزایش سرمایه‌گذاری بر روی صنعت سامانه‌های رایانه، میزان حملات به‌طور چشم‌گیری کاهش یافته است. با توجه به آمار، از سال ۲۰۰۰ تا ۲۰۰۷ استفاده‌کنندگان از اینترنت به بیش از یک میلیارد می‌رسید. باوجوداینکه، امروزه تعداد

1.James Anderson
2.CSI

کاربران اینترنت تا ۲۱۴ درصد رشد داشته است و با رشد کاربران اینترنت، نفوذ به سامانه‌های رایانه‌ای از طریق شبکه نیز افزایش‌یافته است؛ بنابراین، امنیت شبکه به یک مسئله جدی برای کاربران شخصی و شرکت‌های بزرگ تبدیل‌شده است. به‌طورکلی امنیت، شامل ضداسپم، ضدویروس، ورود و خروج حسابرسی‌ها و دیوارهای آتش است که ضداسپم بر روی کاربران نهایی تمرکز دارد و رایانامه هایی که در اسپم قرار می‌گیرند را مسدود می‌کند. فنون ضدویروس نیز معمولاً برای شناسایی نرم‌افزار، بلوک و از بین بردن ویروس‌های رایانه‌ای و برنامه‌های مخرب استفاده می‌شوند. یکی دیگر از راه‌های معمول برای نفوذ پیدا کردن به شبکه‌های رایانه‌ای استفاده از نام کاربری و رمز عبور است که در سال ۱۹۸۰ توسط دنینگ[1] شناخته شد. ازجمله سامانه‌های تشخیص تجاری تریپ وایر[2]، امنیت مانیتور کین[3]، پلیس سایبر[4]، امنیت واقعی[5]، رنجر شبکه[6]، سنتراکس[7]، امنیت سامانه‌های تشخیص نفوذ سیسکو[8]، اژدها[9]، اینترُورت[10] را می‌توان نام برد. برای اطلاعات بیشتر می‌توان به مراجع (G.Vigna, 2002)، (Richardson, 2007)، (S. Zhao, 2007) و (Beqiri, 2009)مراجعه کرد. در حال حاضر سامانه عامل‌ها و شبکه‌های رایانه‌ای از حس‌گرهای تشخیص نفوذ برای شناسایی حملات و نفوذگران استفاده می‌کنند. برخی، پاسخ به نفوذهای غیرقانونی سامانه را، پایان دادن اتصال به شبکه در نظر می‌گیرند. نمونه‌هایی از سطح نفوذ در شبکه شامل محرومیت از راه دور، ورود تروجان‌ها و جاسوسان به سامانه را می‌توان نام برد(Shanmugam, 2005)، (Jena, 2013). هدف ما در این پژوهش تشخیص حملات با استفاده از شبکه‌های عصبی خودسازمان‌ده و ماشین بردار پشتیبان است.

بیان مسئله

با گسترش روزافزون استفاده از ارتباطات و انتقال اطلاعات در بستر شبکه‌ها و اینترنت، حجم وسیعی از مبادلات تجاری و ... از طریق شبکه‌های رایانه‌ای صورت می‌پذیرد؛ بنابراین، با در اختیار بودن اینترنت بر عموم، امنیت اطلاعات به‌عنوان یک موضوع جهانی است که موجب

1.Denning
2.Tripwire
3.Kane Security monitor
4.Cybercop
5.RealSecure
6.NetRanger
7.Centrax
8.Cisco Secure IDS
9.Dragon
10.Intruvert

افزایش خطر برای سامانه‌ها می‌شود؛ چراکه با تهیه کدها و نرم‌افزارهـا مـی‌تـوان بـه مقابلـه بـا حملات پرداخت. در این میان وظیفهٔ ضدویروس و دیـواره‌هـای آتـش[1] ایـن اسـت کـه ابتـدا حملات را شناسایی و سپس به مقابله با حملات بپردازند.

در این تحقیق هدف، کشف حملات سایبری (رایانه‌ای) با استفاده از فنون داده‌کـاوی قبـل از بروز فاجعه است. در طی این روند بایـد ابتـدا داده‌هـا جمـع‌آوری، سـپس رفتـار ایـن داده‌هـا بررسی‌شده و در مقابل رفتارهایی که مشابه هستند را به‌عنوان یک حمله تشخیص داد. یکی از دلایل استفاده از فنون داده‌کاوی در حوزهٔ تشـخیص نفـوذ، توانـایی آن‌هـا در تشـخیص انـواع حملات است. در این میان نمونه‌های جدید تحت عنوان حمله یا بهنجـار شـناخته مـی‌شـوند. سامانه تشخیص نفوذ عبارت است از سامانه‌ای که توانایی تشخیص تغییرات و رفتارهای خاصی در یک میزبان و یا شبکه را دارا باشد. هدف از تشخیص نفوذ به‌ظاهر ساده است اما بـااین‌حـال کار سختی است و در همهٔ سامانه‌های رایانه‌ای فقط شواهدی از نفوذ قابل‌شناسایی است. ایـن شواهد گاهی به‌عنوان حمله جلوه می‌نمایند، درصورتی‌که هیچ تظاهری از حمله وجـود نـدارد. تشخیص نفوذ با استفاده از فنون داده‌کاوی پیچیده‌تر و دقیق‌تر از روش‌های مبتنی بـر امضـا و مبتنی بر ناهنجاری است. با استفاده از رویکردهای داده‌کاوی می‌تـوان رفتارهـای بهنجـار را از نابهنجار تشخیص داد.

فنون داده‌کاوی بسیار متنوع هستند؛ اما از مهم‌تـرین آن‌هـا مـی‌تـوان بـه شـبکه‌هـای عصـبی خودسازمان‌ده و ماشین بردار پشتیبان اشاره نمود. این تحقیق ازآن‌جهت دارای اهمیـت اسـت که با ترکیب ماشین بردار پشتیبان با شبکه‌های عصبی خودسازمان‌ده بـه‌عنـوان جداکننـده‌ای مطلوب برای جدا کردن ویژگی‌های مفید از غیرمفید مورد استفاده قرار می‌گیرد؛ بنابراین، ایـن جداسازی، تا حدودی تشخیص نفوذ در شبکه‌های رایانه‌ای را آسان‌تر می‌کند.

سؤالات پژوهش

۱. با توجه به گستردگی فنون داده‌کاوی، استفاده از کدام فن برای تشـخیص نفـوذ مناسـب خواهد بود؟
۲. آیا می‌توان عملکرد ماشین بردار پشتیبان را برای تشخیص ناهنجاری بهبود بخشید؟
۳. آیا ترکیب الگوریتم‌های یادگیری ماشین با شـبکه‌هـای عصـبی خودسـازمان‌ده مـا را بـه مزایای تشخیص نفوذ می‌رساند؟

1.Firewall

فرضیه‌های تحقیق

۱. با به‌کارگیری روش شبکه‌های عصبی خودسازمان‌ده و ماشین بردار پشتیبان، تشخیص نفوذ در شبکه‌های رایانه‌ای افزایش می‌یابد.
۲. دقت روش یادگیری ماشین و شبکه‌های عصبی خودسازمان‌ده در تشخیص حملات پذیرفتنی است.
۳. زمان تشخیص نفوذ با استفاده از روش یادگیری ماشین و شبکه‌های عصبی خودسازمان‌ده بهبود خواهد یافت.

روش تحقیق

مجموعه داده‌ی مورداستفاده در این تحقیق KDD CUP 99 است. با توجه به مقالات متعددی که مورد مطالعه قرار گرفته است، به‌دلیل بزرگ بودن حجم دیتاست، ما در این تحقیق برای افزایش سرعت و دقت از ۱۰ درصد دیتاست برای افزایش سرعت و دقت استفاده می‌کنیم. در غیر این صورت کارایی واقعی نشان داده نمی‌شود؛ بنابراین، در ابتدا تعدادی از رکوردهای دیتاست ۱۰ درصدی را به‌عنوان داده‌ی آموزش و تعدادی از رکوردهای آن را به‌عنوان مجموعه داده آزمون در نظر می‌گیریم(A. A.Olusola, 2010). در روش پیشنهادی درواقع از دو پالایه برای دسته‌بندی انجام می‌شود. ابتدا حملات از طریق شبکه‌های عصبی خودسازمان‌ده به حملات مشکوک و ترافیک بهنجار طبقه‌بندی می‌شوند. سپس بار دیگر برای جداسازی بهتر ترافیک‌های بهنجار از مهاجم با استفاده از ماشین بردار دسته‌بندی صورت می‌پذیرد.

تشخیص نفوذ: تشخیص نفوذ عبارت است از فرایند شناسایی و پاسخ به فعالیت‌های مخرب که به‌صورت هدفمند، منابع شبکه را مورد تهاجم قرار می‌دهند. اگرچه سطوح بسیاری وجود دارند که از دسترسی به شبکه‌های رایانه‌ای محافظت می‌کنند.

اما نفوذ گران راه‌هایی را برای ورود به شبکه می‌یابند تا خسارات عمده‌ای را در سطح شبکه به وجود آورند. هدف سامانه‌های تشخیص نفوذ، جلوگیری از حمله نیست بلکه کشف و شناسایی حملات و تشخیص اشکالات امنیتی در شبکه‌های رایانه‌ای و درنهایت، اعلام آن به مدیران سامانه است.

ازجمله وظایف سامانه‌های تشخیص نفوذ عبارتند از (Shanmugam, 2005):

- تأیید خطاهای سامانه،
- بررسی یکپارچگی سامانه و فایل‌های داده،
- تشخیص رفتارهای غیرعادی و نابهنجار سامانه،
- شناسایی حملات و هشدارها.

سامانه‌های تشخیص نفوذ: بر اساس منابع دریافتی اطلاعات به‌منظور شناسایی و تشخیص فعالیت‌های غیرمجاز رایانه بر اساس منابع دریافت اطلاعات سامانه‌های تشخیص نفوذ به دو دسته تقسیم می‌شوند:

۱- سامانه‌های مبتنی بر میزبان[1]

در این سامانه‌ها اطلاعات از چندین میزبان جمع‌آوری می‌شود و به میزبان مرکزی ارسال می‌گردند. در میزبان مرکزی عملیات پیچیده‌تر، انجام می‌شود. سامانه‌های تشخیص نفوذ مبتنی بر میزبان از رخدادهای سامانه‌ای و اطلاعات مربوط به برنامه‌های کاربردی به‌عنوان منابع اطلاعاتی استفاده می‌کنند. منابع اطلاعاتی که در سامانه‌های مبتنی بر میزبان نظارت می‌کنند عبارت‌اند از (M. K. Asif, 2013):

- سامانه فایل،
- رویدادهای شبکه،
- فراخوانی‌های شبکه.
- سامانه‌های مبتنی بر شبکه[2]

در این سامانه‌ها، بر فایل‌ها و داده‌های مربوط به ترافیک شبکه، سایر اجزای شبکه و سامانه‌های کنترل امنیت مانند دیوارهای آتش نظارت دارند. سامانه‌های مبتنی بر شبکه، بسته‌های عبوری در سطح شبکه را به‌عنوان منبع اطلاعات جمع‌آوری می‌کنند و احتمال آسیب رساندن به شبکه را کاهش می‌دهند. سامانه‌های تشخیص نفوذ مبتنی بر شبکه، اطلاعات ترافیک شبکه را مورد تجزیه‌وتحلیل قرار می‌دهند(M. K. Asif, 2013)، (G. Wang, 2010).

۲- دسته‌بندی حملات

- **حملات Prob:** در حملات prob، مهاجم اطلاعاتی را درباره شبکه جمع‌آوری می‌کند و با توجه به اطلاعات کسب‌شده به شبکه آسیب می‌رساند. حملات prob از فنون مهندسی اجتماعی برای نفوذ استفاده می‌کنند. این حملات اغلب به تخصص فنی زیادی نیاز ندارند.
- **حملات DOS:** مهاجمان، برخی محاسبات را که وظیفهٔ رسیدگی به درخواست‌های قانونی را دارند از کاربران سلب می‌کنند. این نوع حملات معمولاً از نقاط ضعف نرم‌افزارها جهت

1.Host Intrusion Detection System
2.Network intrusion detection system

آسیب رساندن استفاده می‌کنند و با ایجاد سربار در کانال‌های ارتباطی مانع ارتباط‌های قانونی سامانه می‌شوند. به عبارتی سرویس‌دهنده را از پاسخ‌گویی خارج می‌کنند

- **حملات U2R:** در این حمله مهاجم به حساب کاربری بر روی سامانه دسترسی پیدا می‌کند و به سامانه آسیب می‌رساند. در این نوع حمله، مهاجم سعی دارد با استفاده از یک ماشین خارجی به‌صورت غیرمجازی به ریشهٔ [1] سامانه دستیابی پیدا کند. از شایع‌ترین حملات، حملات سرریز بافر[2] را می‌توان نام برد.
- **حملات R2L:** در این نوع حمله مهاجم یکسری بسته را به ماشین روی شبکه ارسال می‌کند و به‌صورت غیرقانونی به‌عنوان کاربر به اطلاعات شبکه دسترسی می‌یابد(Shanmugam, 2005).

استفاده از هوش محاسباتی در سامانه‌های تشخیص نفوذ": امروزه سامانه‌های تشخیص نفوذ مبتنی بر هوش محاسباتی[3] مورد مطالعه قرارگرفته است. به دلیل این‌که در سامانه‌های تشخیص نفوذ مشکلاتی از قبیل بزرگی شبکه، حجم ترافیک و توزیع داده‌ها بسیار نامتعادل است بنابراین روش‌های پیشگیری از نفوذ مانند دیوارهای آتش و رمزگذاری نمی‌توانند از شبکه در مقابل بدافزارها و حملات سامانه‌ای محافظت کنند؛ بنابراین، فنون داده‌کاوی و هوش محاسباتی روش‌های مؤثری در تشخیص نفوذ هستند. ازجمله ویژگی‌های هوش محاسباتی می‌توان به سازگاری، تحمل خطا، سرعت محاسباتی بالا و در برابر خطا انعطاف‌پذیر بودن را اشاره نمود (Banzhaf, 2010).

شبکه‌های عصبی مصنوعی در هوش محاسباتی: شبکه‌های عصبی مصنوعی شامل مجموعه‌ای به نام واحد پردازش است با توجه به این‌که سلول‌های عصبی بر اساس مکان‌شناسی در ارتباط تنگاتنگ با یکدیگر هستند و دارای توانایی یادگیری هستند. شبکه‌های عصبی، ابزار داده‌کاوی و پردازش تصویر بسیار محبوبی است که منشأ خود را از تلاش برای الگو کردن روند تفکر بشر برگرفته است. شبکه‌های عصبی مصنوعی به‌عنوان الگوریتمی است که می‌تواند به‌طور مؤثر بر روی شبکه‌های رایانه‌ای اجرا شوند. همان‌طور که می‌دانیم مغز انسان شامل سلول‌های عصبی است که سیگنال‌های فعال‌سازی حساب کاربری را ارسال می‌کند و نتیجهٔ آن، افکار هوشمند است. (Banzhaf, 2010).

1.Root
2.Buffer Overflow
3.Computational Intelligence

انواع شبکه‌های عصبی مصنوعی در هوش محاسباتی عبارتند از:

- **شبکه‌های عصبی پیش‌خور:** اولین و ساده‌ترین شبکه‌های عصبی مصنوعی، شبکه‌های عصبی پیش‌خور است. سه نوع شبکه عصبی پیش‌خور در الگوهای نفوذ وجود دارد:
- **شبکه عصبی پیش‌خور چندلایه**[1]: در اوایل از شبکه عصبی پیش‌خور چندلایه برای تشخیص کاربر در سامانه‌های تشخیص نفوذ استفاده شد و قادر به تشخیص رفتار طبیعی، غیرطبیعی، استفاده از اطلاعات و نشانی میزبان است. هم‌چنین رویدادهای غیرعادی را نیز شناسایی می‌کند(Banzhaf, 2010).
- **شبکه‌های عصبی بر پایه تابع رادیال (RBF)**[2]: RBF فاصلهٔ بین ورودی و مراکز سلول‌های عصبی مخفی را اندازه‌گیری می‌کند. شبکهٔ RBF بسیار سریع‌تر از الگوریتم پس‌انتشار[3] است و برای نمونه‌هایی با حجم بزرگ مناسب است. شبکه‌های RBF برای یادگیری و شناسایی حملات طبیعی از خوشه‌های محلی استفاده می‌کنند(Banzhaf, 2010).
- **شبکه‌های عصبی تکراری**[4]: شبکه‌های عصبی تکراری، کلاسی از شبکه‌های عصبی است که بین واحدهای آن ارتباط وجود دارد. این شبکه‌ها می‌توانند از حافظهٔ داخلی خود به‌عنوان ورودی استفاده کنند. شبکه‌های عصبی تکراری دارای قابلیت پیش‌بینی رویداد بعدی در ورودی‌های شبکه است. درصورتی‌که خروجی پیش‌بینی‌شده اشتباه باشد، رویداد واقعی هشدار می‌دهد. البته پیش‌بینی رویدادهای بعدی با توجه به رویدادهای قبلی است(Banzhaf, 2010).

۳-۲- فنون یادگیری ماشین: فنون یادگیری ماشین به توانایی یک برنامه رایانه‌ای برای یادگیری و افزایش عملکرد آن در قالب مجموعه‌ای از وظایف در زمان را گویند. یادگیری ماشین به دودسته، یادگیری با نظارت و یادگیری بدون نظارت تقسیم می‌شوند.

- **یادگیری با نظارت:** این الگوریتم به ازای هر ورودی دارای یک مقدار خروجی و یا تابع مشخص است. شبکه‌های عصبی مصنوعی، الگوریتم‌های خوشه‌بندی k-means، شبکه‌های بیزین[5] و ماشین بردار پشتیبان، ازجمله الگوریتم‌های یادگیری با نظارت هستند(Jena, 2013).

1.Multi-layered feed forward- back-propagation
2.Radial basis function
3.Back-Propagation
4.Recurrent neural network
5.Bayesian

- **یادگیری بدون نظارت**: در این الگوریتم فقط مقدار ورودی‌ها مشخص است. و اطلاعاتی در مورد خروجی در دسترس نیست. مقدار بعدی بر اساس موقعیت فعلی ورودی پیش‌بینی می‌شود(W. Lee, 2001).

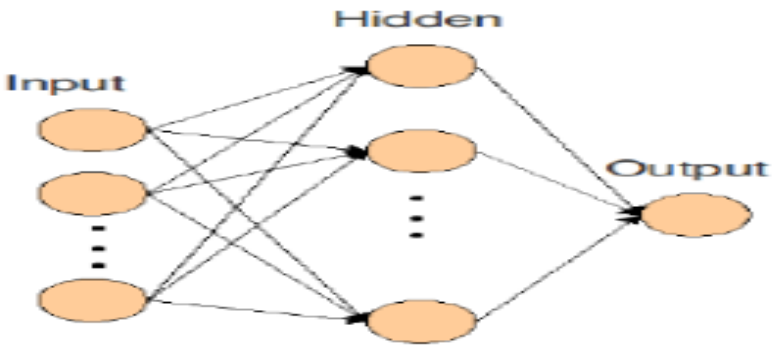
درخت‌های تصمیم‌گیری: درخت‌های تصمیم‌گیری نیز یکی از روش‌های طبقه‌بندی به شمار می‌روند که از داده‌ها به‌صورت دوقسمتی استفاده می‌کنند. یک قسمت به‌عنوان داده‌های آموزشی و قسمت دیگر به‌عنوان داده‌های آزمون هستند. با توجه به این‌که داده‌های آموزشی جهت یادگیری استفاده می‌شوند، مسئلهٔ اصلی اندازه درخت تصمیم‌گیری است و هر چه کوچک باشد بهتر است. داده‌های آموزشی باید به‌درستی طبقه‌بندی شوند. این امر باعث افزایش دقت می‌گردد. هر گره در درخت تصمیم‌گیری نشان‌دهندهٔ یک مشخصه است. هر گره دارای تعدادی یال است که بر اساس مقادیر ممکن، مشخصهٔ موجود در گره والد برچسب می‌خورد و هر یال یا دو تا گره را به هم وصل می‌کند یا یک گره را به یک برگ وصل می‌کند. الگوریتم‌های ID3 و **C4.5** برای ساخت درخت تصمیم استفاده می‌شوند(W. Lee, 2001).

شبکه‌های عصبی مصنوعی: شبکه‌های عصبی مصنوعی برای یادگیری توابعی با مقادیر حقیقی، گسسته و برداری کاربرد دارند. شبکه‌های عصبی از تعداد دلخواهی سلول یا درون تشکیل شده است و مجموعه‌ی ورودی را به خروجی ربط می‌دهند. با توجه به این‌که امروزه هدف شبکه‌های عصبی مصنوعی، دستیابی به کارایی شبه انسانی است. شبکه‌های عصبی مصنوعی بدون این‌که دانش قبلی نسبت به پدیده یا سامانه داشته باشند از اطلاعات نامشخص و داده‌های غیردقیق آگاه می‌شوند(W. Lee, 2001). در حال حاضر شبکه‌های عصبی مصنوعی برای حل مشکلات تشخیص الگو درداده کاوی استفاده می‌شوند(Beqiri, 2009).

الگوریتم‌های شبکه‌های عصبی در تشخیص نفوذ عبارتند از:

۱- الگوریتم پس انتشار خطا: در شبکه‌های عصبی، نرون‌ها برای حل مشکلات خاص با یکدیگر هماهنگ هستند. به عبارتی شبکه‌های عصبی، سامانه‌های انطباقی هستند که از طریق شبکه با توجه به فناوری اطلاعات و ارتباطات در فاز یادگیری قرار دارند. هم‌چنین از روابط بین ورودی الگوها و پیچیدگی خروجی برای پیدا کردن الگوهای موجود در داده‌ها استفاده می‌کنند. الگوریتم پس انتشار خطا، راهی برای تعیین خطا در خروجی از لایه قبل را با توجه به خروجی از لایه جاری ارائه می‌دهد(Mike, 2002)، (Guo, 2013).

شکل (۱)، شبکه کاملاً متصل با لایه ورودی، پنهان و واحد خروجی را نشان می‌دهد(S.Canto, 2010).



شکل ۱-شبکه کاملاً متصل با لایه ورودی، پنهان و واحد خروجی(S.Canto, 2010).

۲- شبکه‌های عصبی خودسازمان‌ده (SOM)[1]: شبکه‌های عصبی خودسازمان‌ده، یک ابزار مؤثر برای تجزیه‌وتحلیل داده‌های چندبعدی است. الگوریتم شبکه‌های عصبی خودسازمان‌ده می‌تواند داده‌های گسسته را به خوشه تبدیل کند و به‌عنوان یک الگوریتم بدون نظارت محسوب می‌شود. رودز[2] و همکاران پس از بررسی بسته‌های شبکه اظهار داشتند که هر پروتکل در لایه شبکه دارای یک ساختار و تابع منحصربه‌فرد است. شبکه‌های عصبی خودسازمان‌ده از سه لایه الگو ساخته شده است که در آن دولایه از شبکه‌های عصبی خودسازمان‌ده رفتار غیرعادی کاربر و ترافیک شبکه را تشخیص می‌دهد و لایه سوم برای حل‌وفصل سردرگمی ناشی از درون ناهمگن در نظر گرفته می‌شود. شبکه‌های عصبی خودسازمان‌ده، از روش یادگیری رقابتی برای آموزش استفاده می‌کنند و مبتنی بر مشخصه‌های خاصی از مغز انسان، توسعه یافته است(Banzhaf, 2010).

یک نوع الگو شبکه عصبی است که در شبکه‌های عصبی خودسازمان‌ده الگوریتم پیاده‌سازی و طرح‌ریزی مشخصه‌های غیرخطی از فضای چندبعدی به فضای یک‌بعدی مورد استفاده قرار می‌گیرد. الگوریتم آموزشی شبکه‌های عصبی خود سازنده به‌صورت زیر خواهد بود.

محاسبه فاصله بین الگو و تمام سلول‌های عصبی:

$$dij=||xk-wij||$$

انتخاب نزدیک‌ترین درون به‌عنوان درون برنده:

$$wij: dij = min(dmn)$$

به‌روزرسانی هر درون با توجه به قاعده:

$$(wwinner,wij)\ ||xk-wij||wij=wij+\alpha h$$

[1]. Self-organizing map
[2].Rhodes

این روند تا زمانی که یک معیار توقف خاص به دست آید، تکرار می‌شود. معمولاً معیار توقف، تعداد ثابتی از تکرار است. برای اثبات هم‌گرایی و ثبات نقشه، نرخ یادگیری و شعاع همسایگی در هر تکرار کاهش می‌یابد. بنابراین، هم‌گرایی به سمت صفر میل خواهد کرد. فاصله اندازه-گیری بین بردارها فاصله اقلیدسی است، ولی از سایر اندازه‌گیری‌های فواصل مانند فاصله Mahalanobis نیز می‌توان استفاده نمود(A.Mukherjee, 1997)، (T. Kohonen, 1997)، (A.Astela, 2007)، (T.Kohonen, 2001).

۳- تشخیص نفوذ با استفاده از شبکه‌های عصبی: در سامانه‌های تشخیص نفوذ، شبکه‌های عصبی برای تشخیص فعالیت‌های مشکوک به کار گرفته می‌شوند. شبکه‌های عصبی مصنوعی از یک گروه غیرخطی عناصر پردازش‌شده و به‌هم‌پیوسته تشکیل‌شده است. همه شبکه‌های عصبی مصنوعی، یک مجموعه از سلول‌های عصبی به نام نرون‌ها را به‌عنوان ورودی دریافت می‌کنند. شبکه‌های عصبی مصنوعی به‌صورت موازی، اطلاعات را پردازش می‌کنند و به الگوی یادگیری خاصی محدود نمی‌شوند. البته توانایی شناسایی غیرمجاز از سامانه و تصمیم‌گیری درباره فعالیت‌های مزاحمان را در بردارند. کندی[1] از شبکه عصبی چندلایه پیش‌خور و استفاده از ۱۰۰۰۰ بسته جمع‌آوری‌شده در یک محیط شبیه‌ساز شبکه برای تشخیص نفوذ استفاده کرد. اما زمان آموزش شبکه‌های عصبی نزدیک به ۲۶ ساعت به طول می‌انجامید(Cannady, 1998).

۴- استفاده از شبکه‌های عصبی برای تشخیص هرزنامه و کرم: به‌طور مستقیم نفوذها و حملات، از طریق نرم‌افزارهای مخرب ازجمله تروجان به سامانه وارد می‌شوند. سامانه توزیع‌شده‌ای که به تروجان آلوده است، اطلاعات خود را از دست می‌دهد. کرم رایانه نیز یک کد اجرایی در قالب یک تصویر باینری است که این کد، در شبکه تکثیر می‌شود و به‌راحتی بهسی. پی.یو[1] و شبکه آسیب می‌رساند. استاپل[2] و همکاران به شناسایی نرم‌افزارهای مخرب بر اساس رفتار طبیعی و غیرطبیعی پرداختند. هم‌چنین فعالیت‌های جعلی شرکت‌های بزرگ بدون هیچ هزینه‌ای به اسپم در رایانامه مخاطبان ارسال شد که این عمل، بدون رضایت گیرندگان صورت گرفت. این عمل نیز به‌عنوان نفوذ شناخته می‌شود. موضوع جالب، پنهان بودن منشأ فرستنده و مسیر پیام است که یک موضوع تحریک‌کننده به‌حساب می‌آید؛ بنابراین، جدا کردن اسپم از غیر اسپم یک موضوع مهم است. چی هانگ وو[3] روش تشخیص اسپم را با استفاده از شبکه عصبی مصنوعی ارائه داد. شبکه‌های عصبی مصنوعی از یک رویکرد مبتنی بر

1.Cannady

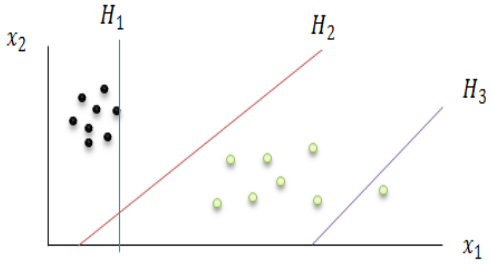
قواعد، هرزنامه‌ها را شناسایی می‌کنند و از طریق پروتکل ساده رایانامه [۴]، نامه‌ها را دریافت می‌نماید(C. Bitter, 2012).

از معایب اصلی سامانه‌های تشخیص نفوذ مبتنی بر شبکه‌های عصبی مصنوعی، به‌دقت تشخیص پایین، به‌خصوص برای حملات تکراری و تشخیص پایداری ضعیف می‌توان اشاره نمود. بدین منظور برای حل دو مشکل یادشده از یک رویکرد جدید به نام روش خوشه‌بندی مرحله‌ای مبتنی بر شبکه‌های عصبی [۵] به کار گرفته می‌شود. توسط فن خوشه‌بندی مرحله‌ای، مجموعه آموزش ناهمگن به چندین زیرمجموعه همگن تقسیم می‌شود. بدین ترتیب پیچیدگی هر یک از زیرمجموعه‌های آموزش کاهش می‌یابد و درنتیجه عملکرد تشخیص نیز افزایش می‌یابد. رویکرد خوشه‌بندی مرحله‌ای مبتنی بر شبکه‌های عصبی برای حملات با تکرار کم، اثربخش است. مانند حملات کاربر به ریشه (U2R) [۶] و حملات از راه دور به نزدیک، ازنظر تشخیص دقت و پایداری مفید است(G. Wang, 2010).

الگوریتم ماشین بردار پشتیبان (SVM): ماشین بردار پشتیبان از دیگر فنون داده‌کاوی است که در سال ۱۹۶۲ این الگوریتم توسط وپنیک[۱] و کورتر[۲] ابداع شد. ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است. که برای طبقه‌بندی دودویی و رگرسیون استفاده می‌شود. مبنای کار ماشین بردار پشتیبان بر اساس دسته‌بندی خطی داده‌ها است در تقسیم خطی داده‌ها سعی بر انتخاب خطی است که حاشیهٔ اطمینان بیش‌تری داشته باشد. الگوریتم ماشین بردار پشتیبان در الگوریتم‌های تشخیص الگو دسته‌بندی می‌شوند و جایی که نیاز به تشخیص الگو یا دسته‌بندی اشیا باشد از کلاس‌های خاص استفاده می‌کند. هدف الگوریتم ماشین بردار پشتیبان انتخاب بهترین خط است. بدین سبب که کم‌ترین میزان خطا را برای طبقه‌بندی داده‌ها دارند. هم‌چنین تعمیم ماشین بردار پشتیبان به مشخصات هندسی داده‌های آموزشی بستگی دارد نه به ابعاد ورودی. پایه معادلات ماشین بردار پشتیبان به‌صورت دو کلاس تعریف می‌شوند. مجموعه داده‌ها به‌وسیله خط از یکدیگر جدا می‌شوند. مجموعه داده‌های آموزشی دارای برچسب هستند و به‌وسیلهٔ ابر صفحه، نقاط متعلق به دو کلاس، از یکدیگر جدا

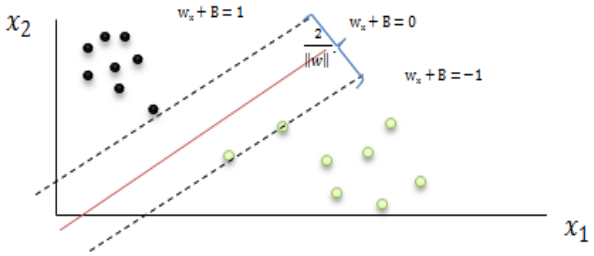
1.CPU
2.Stopel
3.Chih-Hung Wu
4.Simple mail transfer*protocol*
5.FuzzyClusteringBasedonNeural Networks
6. User 2 Root

می‌شوند. به این جداکننده مرز تصمیم‌گیری گویند. شکل (۲) به تعدادی از مرزهای تصمیم‌گیری اشاره دارد.



شکل ۲- مرز تصمیم‌گیری(Jena, 2013).

حاشیه مرز نیز به کوتاه‌ترین فاصله میان نزدیک‌ترین نقطه در هر دو طرف نیم‌صفحه اشاره می‌کند. در شکل (۳) به‌طور بدیهی قابل شهود است و ازنظر ریاضی نیز قابل اثبات است که ابر صفحه حداکثر خطا را پیدا می‌کند(Jena, 2013). شکل (۳) حداکثر حاشیهٔ ابر صفحه برای آموزش ماشین بردار پشتیبان با دو نمونه از کلاس را نشان می‌دهد.



شکل ۳- حداکثر حاشیه ابر صفحه(Jena, 2013).

ابر صفحه برای به حداکثر رساندن حاشیه بین نزدیک‌ترین نقاط خوشه، از نقاط روی مرز که بردار پشتیبان نامیده می‌شوند، استفاده می‌کند. بردار پشتیبان نیز از زیرمجموعه‌ای از داده‌های آموزشی استفاده می‌کند. داده‌ها با استفاده از ابر صفحه، با تعدادی بردار پشتیبان از یکدیگر جدا می‌شوند.. ابر صفحه مرز بین دو کلاس را مشخص می‌نماید. درصورتی‌که مرز بین دو کلاس توسط ماشین بردار پشتیبان قابل‌تشخیص نباشد برای حل این مشکل از توابع هسته ازجمله توابع خطی، چندجمله‌ای و گاوسی استفاده می‌شود. ماشین بردار پشتیبان در کران درجه دوم دارای محدودیت است که این مسئله به مشکل در بهینه‌سازی می‌انجامد(W.H.

1.Vapnik
2.Cortes

Chen, 2005)، (D.Meyer, 2012)، (S. Mukkamala, 2005) و(R. x. ZHANG, 2010). هدف ماشین بردار پشتیبان تولید یک الگو بر اساس داده‌های آموزشی است به بیانی دیگر مشکل سامانه‌های تشخیص نفوذ در طبقه‌بندی الگو را بیان می‌کند(R. x. ZHANG, 2010). با بررسی الگوی فعالیت کاربران با استفاده از شبکه‌های عصبی و ماشین‌های بردار پشتیبان بسیاری از نفوذها نیز قابل‌تشخیص است. الگوهای ذخیره‌شده باید به‌طور مداوم به‌روز شوند که این امر مستلزم تخصص و ابزار هوشمند است. ماشین بردار پشتیبان یک ابزار عملی برای مجموعه داده‌های بزرگ است که داده‌ها را با تعیین مجموعه داده‌های بردار پشتیبان دسته‌بندی می‌کنند. روش‌های تشخیص نفوذ بر اساس الگوریتم‌های تشخیص الگو آموزش‌دیده و فعالیت‌های مخرب را شناسایی می‌کند(R. x. ZHANG, 2010). با توجه به تحقیقات صورت گرفته، عملکرد ماشین بردار پشتیبان در حل مشکلات دسته‌بندی و رگرسیون بهتر از شبکه‌های عصبی مصنوعی است(Y. X. Xia, 2009)، (S. Mukkamala G. J., 2002).

اگرچه ماشین بردار پشتیبان روش امیدوارکننده‌ای برای دسته‌بندی داده‌ها است اما مشکل اصلی آن پیچیدگی آموزش‌های بسیار، وابسته به مجموعه داده‌ها است. بااین‌حال محاسبهٔ ماشین بردار پشتیبان ازنظر شرایط زمانی و مصرف حافظه بسیار پرهزینه است. ماشین بردار پشتیبان داده‌ها را از دسته‌های قبلی برای یادگیری در بردار پشتیبان به‌صورت تدریجی فشرده‌سازی می‌کند. نمونهٔ دیگری از سامانه تشخیص نفوذ، مبتنی بر ماشین بردار پشتیبان و خوشه‌بندی سلسله‌مراتبی، BIRCHاست. ازجمله توانایی‌های استفاده از این الگوریتم، کاهش زمان آموزش است که باعث بهبود عملکرد ماشین بردار پشتیبان می‌شود. این سامانه دارای عملکرد بهتری در تشخیص حملات DOS و Prob است(S. J. Horng, 2011)، (Y. X. Xia, 2009).

با توجه به این‌که از معایب ماشین بردار پشتیبان ، طولانی بودن زمان آموزش است که باعث استفاده محدود از آن شده است. ازجمله برنامه‌های کاربردی در داده‌کاوی و بیوانفورماتیک به پردازش مجموعه داده‌های بزرگ نیاز دارند. و برای تجزیه‌وتحلیل مجموعه داده‌های بزرگ از خوشه‌بندی سلسله‌مراتبی به‌صورت پویا به نام درخت خودسازمان‌ده(DGSOT) استفاده می‌شود؛ بنابراین، با این روش دقت در طبقه‌بندی بهبود، و اندازه مجموعه آموزش کاهش می‌یابد(L. Khan, 2006). وپنیک نشان داد که برای حل مشکل بهینه‌سازی درجه دوم چگونه از روش تشخیص الگو استفاده کند. در رابطه زیر این بهینه‌سازی به‌وضوح دیده می‌شود.

(۴)

$$\min imize : w(\alpha) = -\sum_{i=1}^{l} \alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} yiyj\alpha_i\alpha_j k(x_i, x_j)$$

بنابراین

(۵)

$$\sum_{i=1}^{l} y_i\alpha_i \quad \forall_i : 0 \le \alpha_i \le C$$

راه‌حل معادله:

L: تعداد نمونه‌های α از بردار L است و هر جز هوش مصنوعی مربوط به نمونه‌های (XI,YI) است که در معادله رابطه (۴) به حداقل رسیده است و در رابطه (۵) بهینه‌شده است. ازجمله کاربردهای ماشین بردار پشتیبان، شبیه‌سازی مسیر، پیش‌بینی کیفیت، بهینه‌سازی زمان پیوند اعضا و کاهش هزینه، تشخیص چهره، طبقه‌بندی ژن‌های سرطان و تجزیه‌وتحلیل بحران اقتصادی ارز است(S. Mukkamala, 2005)، (S. Zhao, 2007). ژیانگ[1] و وانگ برای بهینه‌سازی ساختار هسته‌ای ماشین بردار پشتیبان ویژگی‌های اضافی و اثرات منفی را کشف و شناسایی کردند و برای بهینه‌سازی اطلاعات هسته ماشین بردار پشتیبان ، راندمان پارامتر تابع یادگیری ماشین بردار پشتیبان را افزایش دادند که این امر به حل مشکلات در محاسبات می‌انجامد(Chen, 2012). استفاده از الگوریتم ژنتیک و ماشین بردار پشتیبان نیز برای تشخیص خودکار مجموعه‌ای از ویژگی‌ها مناسب است. انتخاب ویژگی‌های مهم برای افزایش قابلیت تعمیم و سرعت بخشیدن به فرایند یادگیری و بهبود الگو است. محاسبهٔ فرایند فیتنس[2]، انتخاب کراس آور[3]وجهش که برای اجرای حداکثر تعداد نسل میسر هستند و با نخبه‌گرایی در هر نسل بهترین رشته برای آن نسل ایجاد می‌شود؛ بنابراین، محل به‌دست‌آمده بهترین مجموعه دسته‌بندی است(S. Saha, 2012). الگوی پیشرفته ماشین بردار پشتیبان با هسته تابع وزن‌دار، دارای عملکرد بهتری نسبت به ماشین بردار پشتیبان معمولی است. الگوریتم جدید ماشین بردار پشتیبان با توجه به سطوح وزنی که برای رتبه‌بندی ویژگی‌های تشخیص نفوذ داده مورداستفاده قرار می‌گیرد . برای تشخیص نفوذ، دقت، زمان محاسبه و میزان خطا بسیار کم‌تر است. فرمت کلی هسته تابع جدید به‌صورت:

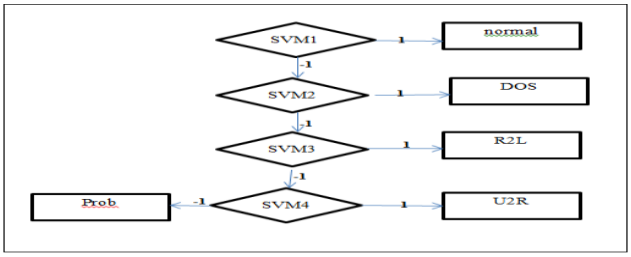
$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha iyik\left(wxi, wx\right) + b\right) \quad (۵)$$

1.Xiong
2.Fitness
3.crossover

تابع ذکرشده برای پیدا کردن وزن با حداکثر حاشیهٔ ابر صفحه برای جدا کردن دو کلاس بر مبنای جداسازی‌کننده غیرخطی است. که L تعداد رکوردهای آموزش،$y_i y_i y_i y_i * \epsilon\epsilon\{-1,1\}$ برچسب اطلاعات آموزش داده‌شده است.$c0 \gg \alpha_i \ll c \ll \alpha_i \gg cW_{xi}W_{xi} \ll \alpha_i \ll c$ و $W_{xi}W_{xi} \gg \alpha_i \ll c$ بردار پشتیبان بر اساس وزن است(J. T. Yao, 2006). ازآن‌جایی‌که IPSنیز به‌عنوان ابزاری است که با تجزیه‌وتحلیل الگوهای کاربران، حملات را تشخیص می‌دهند. IPS مبتنی بر یادگیری با استفاده از ماشین بردار پشتیبان برای کشف و شناسایی حملات سوءاستفاده، کاربرد دارد. استفاده از ماشین بردار پشتیبان بر اساس IPS هشدارهای غلط و حملات جدید را به حداقل می‌رسانند(Lee, 2006). پیاده‌سازی سامانه‌های تشخیص نفوذ با استفاده از ماشین بردار پشتیبان احتمال خطر، خطای تعمیم و خطا در آموزش را کاهش می‌دهد. روش PCNN[1] مبتنی بر الگوریتم تشخیص نفوذ هشدارهای اشتباه را کاهش می‌دهد. هم‌چنین استخراج ویژگی از طریق PCNN عملکردهای تشخیص نفوذ مبتنی بر ماشین بردار پشتیبان را نیز بهبود می‌بخشد. به عبارتی استخراج ویژگی کارآمدترین روش برای تقویت عملکرد سامانه تشخیص نفوذ است(H. H. Gao, 2005). برای شناسایی ناهنجاری در شبکه از دو نوع مختلف از الگوریتم ماشین بردار پشتیبان به نام، الگوریتم با ناظر C-SVM و الگوریتم بدون ناظر One-classsvm استفاده می‌شوند(Q. Zheng, 2004). LIBSVM به‌عنوان کتابخانهٔ ماشین بردار پشتیبان است که توسط چانک و همکاران توسعه یافت و به‌راحتی با کدهای جاوا ادغام‌شده است و فایل‌های باینری آن قادر به اجرای سامانه هستند(Jena, 2013). نتایج تجربی نشان می‌دهند که عملکرد ماشین بردار پشتیبان در مشکلات طبقه‌بندی و رگرسیون بهتر از شبکه‌های عصبی است (Y. X. Xia, 2009). ژانگ و همکاران برای شناسایی چهار نوع از حملات از ماشین بردار پشتیبان استفاده نمودند. آن‌ها نسبت تشخیص درست را ۰.۹۴۵۴ادعا کردند. بر اساس الگوی ارائه‌شده، همان‌طور که در شکل (۴) نشان داده‌شده است که ماشین بردار پشتیبان برای شناسایی ترافیک بهنجار، ترافیک DOS،ترافیکR2L، ترافیکU2R و ترافیک Probاستفاده می‌شود. از ماشین بردار پشتیبان ابتدا برای شناسایی ترافیک بهنجار استفاده می‌شود اگر خروجی یک باشد. یعنی سامانه در حالت بهنجار است در غیر این صورت یعنی سامانه با نفوذ روبه‌رو است. ماشین بردار پشتیبان دوم برای شناسایی حملات DOS، ماشین بردار پشتیبان سوم برای شناسایی حملاتR2Lو ماشین بردار پشتیبان چهارم برای شناسایی حملاتU2R و

1.Principal Component Neural Networks

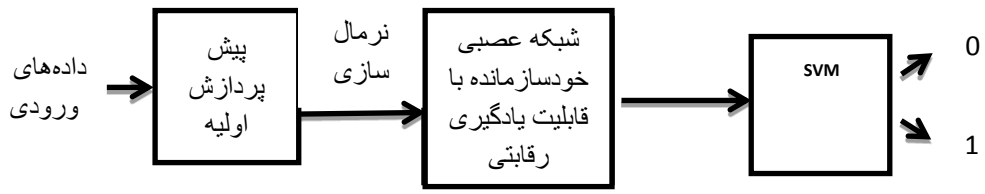
Probاستفاده می‌شود. با توجه به شکل (۴) روند شناسایی نفوذ لمس‌شدنی است (G. Xiaoqing, 2010).



شکل ۴ -شناسایی حملات بر اساس ماشین بردار پشتیبان(G. Xiaoqing, 2010).

روش پیشنهادی

در روش پیشنهادی سامانه‌های مبتنی بر شبکه عصبی، ابتدا شبکه عصبی را با تعداد زیادی حملات آموزش می‌دهیم، سپس در مرحله بعدی داده‌های آزمون را به این شبکه عصبی وارد می‌نماییم. در این مرحله، شبکه عصبی بر اساس پایگاه دانش و آگاهی خود سعی دارد که رفتارهای بهنجار به کلاس صفر و حملات با رفتارهای غیرعادی و آنومالی را به کلاس یک انتساب دهد. این مرحله با تحمل درصدی از خطا ترافیک بهنجار یا حمله را تشخیص می‌دهد. این پژوهش یک سامانه تشخیص نفوذ مبتنی بر یادگیری ماشین با استفاده از شبکه عصبی خودسازمان‌ده را با استفاده از میزان یادگیری رقابتی بیان می‌کند. از آن‌جا که مجموعه داده مورد استفاده در پژوهش KDD CUP99 است. مجموعه رکوردها در کل ۴۱ ویژگی دارند که معرف ترافیک وارده به شبکه است. در مرحله اول برای شناسایی حملات، از یادگیری بر اساس شبکه‌های عصبی خودسازمان‌ده استفاده می‌شود و طبقه‌بندی فقط روی ویژگی‌هایی انجام می‌گیرد که سودمند باشند. بر این اساس هر چه ابعاد آموزش کاهش یابد، سرعت آموزش در شبکه عصبی افزایش می‌یابد. مرحله دوم برای دسته‌بندی و جداسازی حملات از یکدیگر صورت می‌پذیرد. در این حالت ابعاد مسئله نسبت به حالت اول بیشتر است و آموزش، یادگیری نسبت به مرحله اول دارای زمان بیش‌تری است. شکل (۵) روش پیشنهادی را نشان می‌دهد.



شکل ۵- سامانه پیشنهادی تشخیص نفوذ با دو کلاس‌بندی حملات.

ارزیابی

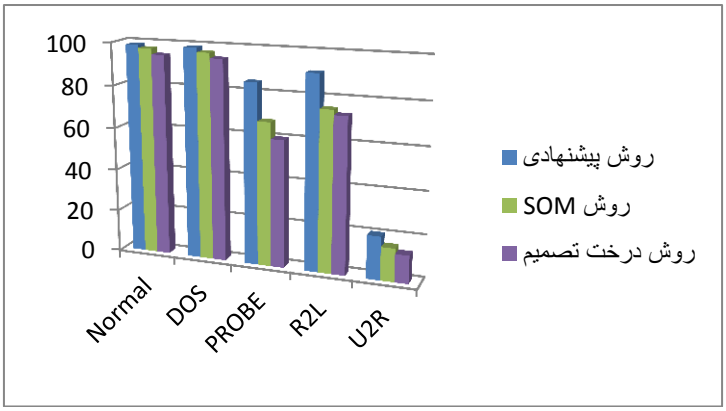
جدول (۱) میزان دقت تشخیص الگوریتم پیشنهادی با دو روش تشخیص نفوذ برای پنج حمله اصلی را مقایسه می‌کند. در این مقایسه‌ها تعداد رکوردهای آموزشی دو هـزار رکـورد اسـت. بـا توجه به اطلاعات جدول (۱) مشخص است که تشخیص نفوذ در الگـوریتم پیشـنهادی از روش درخت تصمیم و شبکه عصبی خودسازمان‌ده بهتر است.

جدول ۱- محاسبه دقت تشخیص در سه روش مورد بررسی در تحقیق(Jena, 2013).

نام حمله	درخت تصمیم‌گیری	شبکه عصبی SOM	روش پیشنهادی
Normal	٪۹۴.۵۰	٪۹۸.۴۰	٪۹۸.۷۳
Dos	٪۹۷.۱۰	٪۹۶.۹۰	٪۹۸.۸۳
Probe	٪۸۳.۳۰	٪۶۷.۶۰	٪۸۵
R2l	٪۸.۴۰	٪۷.۳۰	٪۸.۶۰
U2R	٪۱۳.۲۰	٪۱۵.۷۰	٪۲۰.۴۲

نمودار شکل (۶) نمودار مقایسه دقت الگوریتم پیشنهادی، شبکه عصبی خودسازمان‌ده و درخت تصمیم-گیری نشان داده شده است.

شکل ۶-مقایسه دقت انواع حملات با الگوریتم یادگیری بر پایه روش پیشنهادی، شبکه عصبی خودسازمان‌ده و درخت تصمیم‌گیری.



از تحلیل نمودار شکل (۶) نشان داده می‌شود که دقت الگوریتم پیشنهادی از الگـوریتم شـبکه عصبی خودسـازمان‌ده و درخـت تصـمیم در حمـلات Probe، Dos،U2R، Normal،R2l بیشتر است. تشخیص حمله در روش شبکه خودسازمان‌ده کم‌تر از روش پیشـنهادی اسـت. بـا توجه به این موضوع که هر چه الگوریتم یادگیرنده بیش‌تر آموزش دیده باشد، میزان دقت نیـز افزایش می‌یابد؛ بنابراین، یکی از ویژگی‌های یک سامانه تشخیص نفوذ مناسـب، داشـتن زمـان آموزش کمتر است. یعنی زمانی که به زمان واقعی شباهت بیش‌تری داشـته باشـد. بـرای ایـن

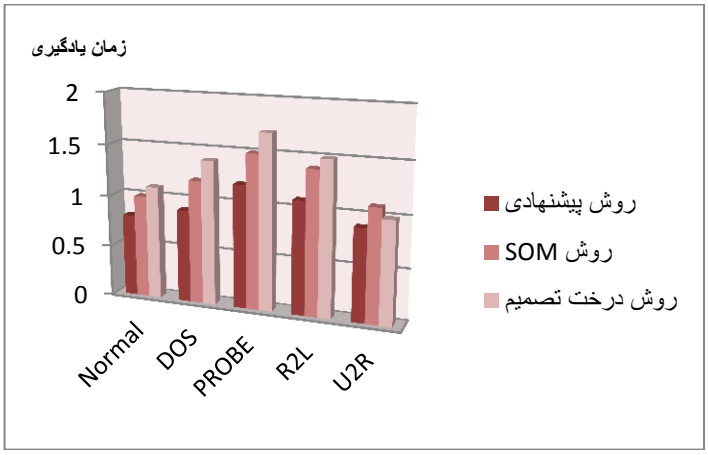
منظور جدول (۲) زمان یادگیری بین سه الگوریتم پیشنهادی، شبکه خودسـازمان‌ده و درخـت تصمیم را نشان می‌دهد.

جدول ۲- تشخیص در چهار روش مورد بررسی در تحقیق.

نام حمله	درخت تصمیم‌گیری	شبکه عصبی SOM	روش پیشنهادی
Normal	۱.۱	۱	۰.۸
Dos	۱.۴	۱.۲	۰.۹
Probe	۱.۷	۱.۵	۱.۲
R2l	۱.۵	۱.۴	۱.۱
U2R	۱	۱.۱	۰.۹

روش پیشنهادی دارای زمان کم‌تری برای یادگیری نسبت به سایر روش‌ها دارا است، شکل (۷) نشان‌دهنده زمان یادگیری در الگوریتم‌های پیشـنهادی، شـبکه‌هـای عصـبی خودسـازمان‌ده و درخت تصمیم‌گیری است.

شکل ۷- زمان یادگیری انواع حملات با الگوریتم یـادگیری بـر پایـه روش پیشـنهادی، شـبکه عصبی خودسازمان‌ده و درخت تصمیم‌گیری.



نتیجه‌گیری

باوجوداینکه امروزه شبکه‌های رایانـه‌ای نقـش عمـده‌ای در ارتباطـات و انتقـال اطلاعـات ایفـا می‌کنند. اما افراد سودجو با دسترسی به اطلاعات مهم قصد نفوذ بـه سـامانه‌هـا و شـبکه‌هـای رایانه‌ای را دارند. ازآن‌جایی‌که شبکه‌های عصبی اطلاعات را در خود نگه می‌دارند و زمـانی کـه نفوذگر وارد سامانه می‌شود با توجه به رفتارهای خرابکارانه و آنومالی‌هایی کـه از خـود برجـای می‌گذارد. با استفاده از شبکه‌های عصبی خودسازمان‌ده که قادر بـه ذخیـرۀ ناهنجـاری اسـت.

آنومالی تشخیص و حملات را بر اساس ویژگی‌های سودمندی که دارند طبقه‌بندی نموده، برای جداسازی حملات در شبکه‌های رایانه‌ای از ماشین بردار پشتیبان که یکی دیگر از فنون داده‌کاوی به شمار می‌رود، استفاده نمودیم؛ بنابراین، در روش ماشین بردار پشتیبان با روش دسته‌بندی می‌توان حملات را از یکدیگر متمایز نمود و خطای تعمیم را به حداقل رساند.

پژوهش‌های آینده

امروزه با وجود حجم انبوهی از حملات در شبکه‌های سایبری (رایانه‌ای) می‌توان با استفاده از روش‌ها و رویکردهای جدید، مدیریت، کنترل و پردازش اطلاعات به بهبود ساختار شبکه‌ها و کاهش تلفات سایبری (رایانه‌ای) کمک نمود. هدف ما در پژوهش آتی، این است که با تجزیه‌وتحلیل داده‌های بزرگ[1] به الگوهای پنهان حملات سایبری (رایانه‌ای) دست یافت.

1.Big Data

منابع

1. G.Vigna, R. a. (2002). Intrusion Detection:A Brief History and Overview. 27-30.
2. Richardson, R. (2007). ***Computer Crime and Security Survey. 1-30.***
3. S. Zhao, R. a. (2007***). Intrusion detection using the support vector machine enhanced with a feature-weight kernel***. 1-83.
4. Beqiri, E. (2009). ***Neural Networks for Intrusion Detection Systems***. 156–165.
5. Shanmugam, N. B. (2005). ***Artificial Intelligence Techniques Applied to IntrusionDetection***. 52-55.
6. Jena, P. a. (2013). ***ntrusion Detection Using Self-Training Support Vector Machines***. 1-35.
7. M. K. Asif, T. A. (2013). ***Network Intrusion Detection and its StrategicImportance***.140-144.
8. G. Wang, J. H. (2010). ***A new approach to intrusion detection using ArtificialNeural Networks andfuzzy clustering***. 6225–6232.
9. G. Xiaoqing, G. H. (2010). ***Network Intrusion Detection Method Based on Agent and SVM***. 1-4.
10. Banzhaf, S. X. (2010). ***The use of computational intelligence in intrusion etection systems***: A review. 1-35.
11. W. Lee, S. J. (2001). ***Real Time Data Mining-based Intrusion Detection***. 89 –100.
12. Mike, S. (2002). ***Network Security Principles and Practices***. Chapter 14.
13. Guo, J. L. (2013). ***implementation of neural network backpropagation in CUDA***. 1021-1027.
14. S.Canto, X. F. (2010). ***Parallel Training of a Back-Propagation Neural Network using CUDA***. 1-6.
15. Cannady, J. (1998). ***Artificial Neural Networks for Misuse Detection***. 368–381.
16. C. Bitter, J. N. (2012). ***An Introduction to the Use of Neural Networks for Network Intrusion Detection***. 5-24.
17. G. Wang, J. H. (2010). ***A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering***. 6225–6232.
18. Jena, P. a. (2013). ***Intrusion Detection Using Self-Training Support Vector Machines***. 1-35.

19. W.H. Chen, S. H. (2005). ***Application of SVM and ANN for intrusion detection***. 2617–2634.
20. D.Meyer. (2012). ***Support Vector Machines***. 1-8.
21. S. Mukkamala, A. H. (2005). ***Intrusion detection using an ensemble of intelligent paradigms***. 167-182.
22. R. x. ZHANG, Z. a. (2010). ***Intrusion Detection based on SVM and decision fusion***. 87-90
23. S. Mukkamala, G. J. (2002). ***Intrusion Detection Using Neural Networks and Support Vector Machines***. Department of Computer Science New Mexico Institute of Mining and Technology, 1702-1707.
24. Y. X. Xia, Z. H. (2009). ***A Framework for Distributed Incremental Intrusion Detection Based on SVM***. 369-372.