مجله علمی - پژوهشی "پدافند الکترونیکی و سایبری"
سال سوم، شماره ۳، پاییز ۱۳۹۴، ص ۳۰- ۱۹

امنیت سامانه‌های فرماندهی نظامی تحت وب با استفاده از ترکیب دسته‌بندهای تک‌کلاسی

امینه جمالی‌فرد[1*]، حسین شیرازی[2]

۱- کارشناس‌ارشد مهندسی کامپیوتر، دانشکده فرماندهی و کنترل، دانشگاه صنعتی مالک اشتر، تهران

۲- دانشیار، دانشکده فرماندهی و کنترل، دانشگاه صنعتی مالک اشتر، تهران

(دریافت: ۹۳/۱/۱۸؛ پذیرش: ۹۴/۱۰/۲۲)

چکیده

فعالیت‌های نفوذگرانه تحت وب به شبکه‌های اطلاعاتی فرماندهی نیروهای نظامی به‌منظور تخریب، غیرعملیاتی کردن یا سرقت زیرساخت‌های اطلاعات راهبردی؛ از ترفندهای رایج دشمن در عصر جنگ‌های الکترونیک است. تامین امنیت سامانه‌های فرماندهی و کنترل تحت وب از ملاحظات دفاعی ضروری در سطح فرماندهی نوین نظامی است. در این مقاله سیستمی برای تشخیص نفوذ به سامانه‌های حساس و محرمانه نظامی تحت وب با استفاده از ترکیب دسته‌بندهای تک‌کلاسی پیشنهاد شده است. در مرحله آموزش بردارهای ویژگی استخراج شده مرتبط با هر درخواست HTTP، وارد سیستم شده و مدل درخواست عادی توسط هر دسته‌بند یادگیری می‌شود؛ سپس با استفاده از روش‌های مختلف ترکیب دسته‌بندهای تک‌کلاسی؛ بار دیگر مدل درخواست عادی HTTP یادگیری می‌شود. برای ترکیب دسته‌بندها از استراتژی ترکیبی نوینی، جهت تصمیم‌گیری گروهی استفاده شده است. استفاده از تصمیم‌گیری گروهی و استراتژی S-OWA برای ترکیب دسته‌بندهای تک‌کلاسی، قابلیت اطمینان و دقت سیستم تشخیص نفوذ به سامانه‌های نظامی تحت وب را در حد قابل توجهی بهبود بخشیده است.

واژه‌های کلیدی: فرماندهی نظامی، سامانه‌های نظامی تحت وب، تشخیص نفوذ، ترکیب دسته‌بندهای تک‌کلاسی، عملگر S-OWA

۱- مقدمه

آمریکا در سال ۱۹۶۹ شبکه اینترنت را جهت به‌کارگیری در عرصه ارتباطات نظامی وارد این عرصه کرد تا این شبکه کارآیی خود را پس از گذشت سال ها در فرآیند ارتباطات نشان داده و به اثبات برساند. پس از آن اینترنت وارد عرصه علم، دانشگاه‌ها و مراکز پژوهشی شد تا در دهه هشتاد قرن بیستم میلادی از چارچوب قبلی خود خارج و به بخشی از فضای جهانی مبدل شود؛ به این ترتیب اینترنت به جدیدترین سلاح معاصر و به‌صورت بالقوه تاثیرگذارترین و ویرانگرترین آن‌ها تبدیل شد. اکنون کشورهای جهان تلاش می‌کنند تا افسار این شبکه سرکش را جهت رام کردن و تحت نظر گرفتن آن به‌دست گیرند. در عصر حاضر بسیاری از زیرساخت‌های زندگی بشر بر پایه اینترنت است و تقریبا هیچ جنبه‌ای از زندگی بشر را نمی‌توان یافت که در آن از ارتباطات تحت وب بی‌نیاز باشد. در کاربردهای نظامی و امنیتی به‌دلیل حساسیت و محرمانگی موضوع از یک سو و نیاز به ارتباطات امن و کارآمد از سوی دیگر؛ فراهم آوردن سیستم‌هایی که قابلیت اطمینان بالا و کارآمدی قابل قبولی داشته باشند از ضرورت و اولویت ویژه‌ای برخوردار است. در شرایط تهدیدات نظامی و امنیتی دنیای کنونی، هیچ کشوری بدون داشتن زیرساخت‌های امن ارتباطی به‌ویژه در حوزه‌های نظامی و امنیتی؛ از گزند دشمنان مصون نیست. شاید یکی از بارزترین و مهم‌ترین تهدیدات اینترنتی صهیونیستی؛ تولید ویروس استاکس‌نت[1] به منظور اختلال در مراکز هسته‌ای ایران بود که بر اساس گزارش‌های پایگاه اطلاع‌رسانی سازمان پدافند غیرعامل کشور این ویروس‌ها پس از تولید توسط شرکت‌های مهندسی، ماه‌ها درنیروگاه‌های هسته‌ای دیمونا در صحرای نقب[2] مورد آزمایش قرار گرفتند. این سلاح سایبری نمونه‌ای قابل توجه از اقدامات و گام‌های دشمن در تغییر مفاهیم جنگ‌های سنتی قلمداد می‌شود، جنگ‌هایی که از جنگ با توپ و مواد منفجره به جنگ سایبری تبدیل شده‌اند [۱ و ۲].

امروزه در عرصه فرماندهی نیروهای نظامی، با توجه به اهمیت سرعت و قابلیت اطمینان ارسال داده ها از یک سو و حجم بالای اطلاعات مبادله شده از سوی دیگر، تبادل اطلاعات محرمانه و حساس نظامی با چالش های متعددی روبرو شده است. از این رو استفاده از زیرساخت های امن تحت وب که ارتباطات فرماندهان نظامی و نیروهای تحت کنترل را به‌صورت محافظت شده ای فراهم می کند، از دغدغه‌های مدیران ارشد نظامی است. این ضرورت و تهدید جدی ما را بر آن داشت تا درصدد طراحی سیستم تحت وب امنی برآییم. در این مقاله تلاش شده است پاسخی نوین و بومی برای این نیاز ضروری ارائه شود.

رایانامه نویسنده پاسخگو: am.jamalyfard@yahoo.com

1-Stuxnet
2-Negev Desert

در این مقاله روشی برای تشخیص درخواست‌های HTTP ناهنجار در برنامه های کاربردی تحت وب ارائه می‌شود که از ترکیبی از دسته بندهای تک کلاسی استفاده می‌نماید. نرخ تشخیص[1] و نرخ هشدار نادرست[2] این روش در مقایسه با سایر روش های ارائه‌شده، شناسایی مناسب و با خطای اندکی را نشان می‌دهند. در بخش دوم مقاله، پیشینه تحقیق بررسی می‌شود؛ در بخش سوم تعاریف اولیه را معرّفی می‌گردد. در بخش چهارم به ارائه مدل پیشنهادی مقاله در امنیت برنامه‌های کاربردی تحت وب پرداخته می‌شود. در روش تشخیص ناهنجاری پیشنهادی به منظور بهبود کارآیی از سیستم مبتنی بر ترکیب چند دسته بند تک کلاسی بهره می‌گیریم. بدین‌منظور از عملگر S-OWA برای ترکیب دست‌بندها استفاده می‌شود؛ سپس در بخش پنجم، روش پیشنهادی را مورد ارزیابی قرار داده و نتایج تجربی و مشاهدات حاصل را بیان می‌گردد. در بخش ششم، نتیجه‌گیری، بحث و بیان پژوهش‌های آینده آمده است.

۲- پیشینه تحقیق

نخستین سیستم تشخیص نفوذ مبتنی بر تشخیص ناهنجاری، برای حفاظت از برنامه های کاربردی تحت وب، این گونه پیشنهاد شد که فعالیت خرابکارانه خودش را در پارامترهایی که در درخواست HTTP یافت می‌شود؛ نشان می‌دهد. از تخمین پارامتر بیزین برای تشخیص ناهنجاری برنامه‌های کاربردی تحت وب استفاده شد. تکنیک‌هایی ارائه‌شده که روی تحلیل پارامترهای درخواست‌های HTTP تمرکز دارند و اساسا شامل ترکیبی از مدل‌های تشخیص مختلف می‌باشند. این مدل‌ها روی طول ویژگی‌ها، توزیع کاراکتری ویژگی‌ها، استنتاج ساختاری، حضور یا عدم حضور ویژگی‌ها و ترتیب ویژگی‌های درخواست HTTP تمرکز دارند [۳].

سیستم تشخیص ناهنجاری شبکه‌ای پیشنهادشده که از فاصله Mahalanobis به‌عنوان راهی در تشخیص درخواست‌های ناهنجار در مجموعه داده‌هایی با ویژگی‌های چندگانه و مقیاس‌دهی هر متغیر بر مبنای انحراف معیار استاندارد و کوواریانس؛ استفاده می‌کند [۴].

گروهی از محققان حوزه امنیت وب، تشخیص‌دهنده ناهنجاری محتوا بر اساس تحلیل n-gram پیشنهاد دادند که از فیلترهای bloom استفاده می‌کرد و مقاومت در برابر حملات مشابه و چندریخت را فراهم می‌نمود [۵].

گروهی از محققان نشان دادند که چگونه می‌توان سیستمی ارائه داد که با استفاده از الگوریتم استنتاج DFA به‌همراه هیوریستیک‌های کاهش اتوماتا خطر مثبت نادرست را کمینه کند. روش آن‌ها الگوریتم آموزش دارای قابلیت کار با داده‌های غیرایستا با طول دلخواه را فراهم می‌کند [۶]. ابزاری آماری مبتنی بر یادگیری ماشین برای دفاع در برابر حملات تزریق ارائه‌شده که از رویکرد ترکیبی زنجیره‌های مارکوف برای مدل کردن درخواست‌های HTTP و استخراج الگوریتم آموزشی مرتبط، استفاده شده است [۷].

۳- تعاریف اولیّه

آسیب‌پذیری‌های تحت وب به‌عنوان بخش عمده‌ای در زمینه امنیت سیستم‌های رایانه‌ای مطرح می‌باشند. به‌منظور شناسایی حملات شناخته‌شده تحت وب، سیستم‌های تشخیص نفوذ به تعداد زیادی از امضای حملات مجهز می‌شوند. متاسفانه، همگام‌سازی سیستم با به‌روزرسانی تغییرات رخداده در زمینه حملات اینترنتی امر بسیار سختی است؛ همچنین ممکن است آسیب‌پذیری‌های سیستم با نصب برنامه‌های کاربردی تحت وب مشخصی رخ دهد. به همین دلیل سیستم های تشخیص نفوذ بهتر است به‌صورت سیستم‌های تشخیص ناهنجاری پیاده‌سازی شوند.

تشخیص ناهنجاری در حیطه مسائلی است که در میان داده ها برای یافتن الگوهایی که با رفتار از پیش مورد انتظار مغایرت دارند؛ تلاش می‌شود. این الگوهای ناهمگون اکثرا مربوط به داده های پرت، مشاهدات ناسازگار، موارد استثناء، موارد انحرافی، ویژگی‌های گمراه‌کننده و فعالیت‌های مختل‌کننده در زمینه‌های کاربردی گوناگون می‌باشد. واژه‌های داده پرت[3] و داده ناهنجار[4] به‌صورت متناوب در متون تخصصی این حیطه به‌جای یکدیگر به‌کار می‌روند که معادل هم می‌باشند [۸].

برنامه کاربردی تحت وب هر نوع برنامه کاربردی است که از جستجوگر وب به‌عنوان سرویس‌گیرنده استفاده می‌کند. تمامی‌وب سایت‌های موجود بر روی اینترنت از پروتکل HTTP استفاده می‌نمایند. با این که پروتکل HTTP با استفاده از پروتکل‌های دیگری نظیر IP و TCP ماموریت خود را انجام می‌دهد، ولی این پروتکل HTTP است که به‌عنوان زبان مشترک ارتباطی بین سرویس‌گیرنده و سرویس‌دهنده وب به رسمیت شناخته‌شده و از آن استفاده می‌گردد. در واقع مرورگر وب صدای خود را با استفاده از پروتکل HTTP به گوش سرویس‌دهنده وب می‌رساند و تقاضای سرویس می‌کند [۹].

درخواست HTTP مجموعه‌ای از خطوط متنی[5] است (با CRLF از یکدیگر جدا شده‌اند) که به سرویس‌دهنده وب ارسال می‌شود و شامل خط درخواست[6]، فیلدهای سرپیام درخواست[7] و بدنه درخواست[8] می‌باشد. خط درخواست از سه بخش

1-Detection Rate
2-False Alarm Rate

3-Outlier
4-Anomaly
5- Text line
6- Request line
7- Request header fields
8- Request body

تشکیل‌شده که با فاصله از یکدیگر جدا شده‌اند؛ این سه بخش نام متدی که می‌باستی اعمال شود، مسیر محلی منبع درخواست و نسخه پروتکلی که مورد استفاده قرار می‌گیرد را مشخص می‌کند. نخستین کلمه‌ای که در درخواست HTTP ظاهر می‌شود کلمه method است؛ بیشتر درخواست‌های HTTP از نوع متد GET هستند ولی انواع دیگری از متد همانند POST و HEAD نیز وجود دارند. بعد از method ، مسیر منبع (URI) ذکر می‌شود که عموما یک فایل، یک فهرست در سیستم فایل یا ترکیبی ازهر دو است. آخرین بخش، نسخه پروتکل استفاده‌شده توسط سرویس‌گیرنده را مشخص می‌کند. (عموما HTTP/1.0 یا HTTP/1. 1)

خط در خواست معمولا به شکل زیر است:

GET / path/to/file/index. Html HTTP/1.1

در ادامه خط درخواست اولیه در درخواست HTTP، فیلدهای سرپیام درخواست وجود دارند که اطلاعاتی درباره درخواست هستند. خطوط فیلد سرپیام به فرمت سرپیام عادی‌هستند: یک خط برای هر سرپیام به‌صورت "مقدار: نام سرپیام"[1] که با CRLF خاتمه می‌یابد. در نسخه ۱/۰ پروتکل HTTP معمولا ۱۶ سرپیام وجود دارد، با این‌وجود هیچ یک اجباری نیستند. در نسخه ۱/۱ پروتکل HTTP با ۴۶ سرپیام مشخص می‌شود، که تنها سرپیام Host در درخواست اجباری است. سرپیام‌های درخواست مجموعه‌ای از خطوط اختیاری هستند که اطلاعاتی اضافی درباره درخواست، سرویس‌گیرنده و یاهر دو ارائه می‌دهند (جستجوگر، سیستم عامل و غیره). هر یک از این خطوط از نامی تشکیل شده که نوع سرپیام را مشخص می‌کند و با (:) و مقدار سرپیام دنبال می‌شود [۱۰].

جهت تامین امنیت سامانه‌های تحت وب، یک سیستم تشخیص نفوذ سعی دارد حملات احتمالی به داده‌ها و منابع محاسباتی سیستم را تشخیص دهد. دو دیدگاه کلی موجود در مسئله تشخیص نفوذ وجود دارد: دیدگاه‌های مبتنی بر امضاء و دیدگاههای مبتنی بر تشخیص ناهنجاری؛ سیستم های مبتنی بر امضا، خصوصیات و مشخصه های شناخته‌شده حملات از پیش شناخته‌شده و مشخص را به‌کار می برند و از یک سیستم ساده تشخیص مبتنی بر قانون استفاده می کنند. این سیستم‌ها به سادگی پیاده‌سازی می‌شوند ولی نیازمند دانش اولیه ازهر نوع حمله هستند و نمی‌توانند حملات جدیدی که قبلا مشاهده نکرده اند را شناسایی کنند، می توان به سادگی آنها را با سناریوهای حمله جدید موردهجوم قرار داد و بیشتر در حوزه‌های عمومی و تجاری مورد اقبال و توجه هستند. در سیستم‌های مبتنی بر تشخیص ناهنجاری ازهوش مصنوعی، تکنیک‌های یادگیری ماشینی و داده کاوی برای پردازش اطلاعات تولیدشده توسط حسگرهای شبکه؛ جهت یافتن رخدادهای غیرعادی شبکه استفاده می‌شود. مدلی از رفتارهای عادی سیستم تحت شرایط کنترل‌شده خاصی ایجاد می‌شود. ساختن مدل رفتار سیستم در فاز آموزش و در بخش برون خط[2] تشخیص ناهنجاری است. هنگامی که این مدل ساخته شد، وضعیت شبکه پی‌درپی با این مدل مقایسه می‌شود تا انحراف‌ها از شرایط عادی کشف شوند. این بخش قسمت برخط[3] تشخیص ناهنجاری را تشکیل می‌دهد [۱۱].

وظیفه اصلی یک سیستم تشخیص نفوذ به سامانه‌های تحت وب، تشخیص رفتارهای غیرعادی با توجه به رخدادهای این برنامه‌ها و رفتارهای درخواست‌های HTTP می‌باشد. فرض کنید سامانه ما با تعدادی درخواست متخاصم روبروست که با به‌کارگیری سناریوی حمله‌های مختلف که برخی از آن‌ها برای ما ناشناخته است سعی در مختل کردن عملیات آن دارند. در این‌صورت مساله‌ای که با آن روبرو هستیم این است که چگونه می‌توان به دنباله درخواست‌های HTTP سامانه تحت وب در طول زمان برچسب عادی یا غیرعادی زد. حل این مساله وظیفه اصلی سیستم‌های تشخیص نفوذ مرسوم می‌باشد [۱۲].

در روش‌های مبتنی بر یادگیری ماشین، ابتدا رفتار عادی با استفاده از دسته بندهای تک‌کلاسی یادگیری می‌شود و سپس هر انحرافی از این رفتار عادی به عنوان ناهنجاری در نظر گرفته می‌شود. استفاده از بهترین دسته‌بندهای تک‌کلاسی یا ترکیبی از آن‌ها برای یادگیری رفتار عادی همواره به‌عنوان راه حلی پیش روی محققین می‌باشد [۱۳].

۴- مدل پیشنهادی

۴-۱- ساختار

در حوزه تشخیص ناهنجاری در برنامه‌های کاربردی تحت وب که مورد بحث ما در این پژوهش می‌باشد؛ تنها مجموعه دادگان درخواست‌های HTTP عادی در دسترس است و در مرحله آموزش ما می‌خواهیم رفتار عادی درخواست‌های HTTP را مدل کنیم تا در مرحله تشخیص، درخواست‌های ورودی به سیستم پیشنهادی ما با این مدل عادی مقایسه شوند. برای تهیه مدل عادی از دسته‌بندهای تک‌کلاسی استفاده می‌کنیم. در واقع هر کدام از دسته‌بندهای تک‌کلاسی را به‌صورت مستقل یک سیستم تشخیص ناهنجاری برای درخواست‌های HTTP در نظر می‌گیریم و مراحل آموزش و تشخیص را برای‌هر یک انجام می‌دهیم.

۴-۲- استخراج ویژگی

یکی از مهم‌ترین ملزومات برای پیشنهاد یک سیستم تشخیص ناهنجاری برای برنامه‌های کاربردی تحت وب بر مبنای

1- Header-Name: value

2- Offline

3-On line

پروتکل HTTP؛ شناخت رفتار عادی این پروتکل است تا بتوان ناهنجاری‌ها و حمله‌ها را که به‌عنوان انحراف از حالت عادی تعریف می‌شود، تشخیص داد. گام نخست در شناخت رفتار عادی پروتکل HTTP، توصیف دقیق و جامع ویژگی‌ها و رفتار آن می‌باشد. این توصیف اغلب با تعریف ویژگی صورت می‌گیرد و به تبع آن رفتار عادی به‌عنوان قیدی روی مقدار ویژگی‌های تعریف شده یا رابطه‌ای مابین ویژگی‌ها تعریف می‌شود. برای تعریف ویژگی‌هایی از درخواست‌های HTTP که در مساله تشخیص ناهنجاری تعیین‌کننده باشند، نیازمند شناخت حملات و نحوه تاثیر آن‌ها روی بخش‌های مختلف این درخواست‌ها هستیم. با استفاده از دانش خبره در حملات وب، ۲۸ ویژگی موثر در تشخیص ناهنجاری شناسایی شده است. (جدول ۱) [۶ و ۱۳] هر کدام از این ویژگی‌ها به نوعی در حملات متداول تحت وب، تحت تاثیر فعالیت‌های مخرب مهاجمان قرار گرفته‌اند. بعضی ویژگی‌ها به طول درخواست، طول فیلد Path یا Header بستگی دارد زیرا طول فیلدها برای تشخیص حملات سرریز بافر اهمیت دارند. همچنین مشاهده‌شده که کاراکترهای غیرالفبایی- غیرعددی در بسیاری از حملات سرریز مشاهده شده‌اند. با این‌حال، چهار گونه کاراکتر در این لیست لحاظ شده است: حروف، اعداد، کاراکترهای غیرالفبایی- غیرعددی و سایر کاراکترها. کاراکترهای غیرعددی- غیرالفبایی معنای خاصی در تعدادی از زبان‌های برنامه‌نویسی دارند. این کاراکترها در جدول (۱)، کاراکترهای خاص نامیده شده‌اند.

جدول (۱). ویژگی‌های درخواست HTTP

نام ویژگی	نام ویژگی
طول فیلد Path	طول سرپیام "Accept-Charset"
طول سرپیام "Accept"	طول Header
طول درخواست	طول سرپیام"Accept-Encoding"
طول سرپیام "Cookie"	طول سرپیام "Accept-Charset"
طول سرپیام "Content-Type"	طول سرپیام "Accept-Language"
طول سرپیام "Referer"	طول سرپیام "Content-Length"
شناسه متد	طول Host
تعداد کاراکترهای خاص در Header	طول سرپیام "User-Agent"
تعداد کاراکترهای دیگر در Header	تعداد آرگومان‌های درخواست
تعداد حروف در path	تعداد اعداد در Header
تعداد کاراکترهای خاص در Path	تعداد حروف در Header
Min ASCII char در Request	تعداد اعداد در فیلد Path
Max ASCII char در Request	تعداد کاراکترهای دیگر در فیلد Path
تعداد بایت‌های متمایز	تعداد Cookieها

۴-۳- دسته‌بندهای تک‌کلاسی

در فرآیند آموزش در روش‌های دسته‌بندی دو یا چند کلاسی، داده‌های مربوط به همه کلاس‌ها موجود می‌باشد. درصورتی‌که در مساله تشخیص ناهنجاری پیشنهادی، در هنگام توصیف رفتار عادی هیچ مجموعه‌داده حمله‌ای وجود ندارد و در فرآیند آموزش فقط داده‌های مربوط به یک کلاس (کلاس رفتار عادی که به‌طور عام‌تر کلاس هدف نامیده می‌شود) موجود می‌باشد [۱۴]. در این‌گونه مسائل مجبور به استفاده از دسته‌بندی‌های تک‌کلاسی‌هستیم تا بتوانیم مشخصات یک کلاس موجود را یادگیری نماییم.

در روش‌های دسته‌بندی تک‌کلاسی دو مولفه اصلی باید مشخص شود. مولفه اول عبارت است از اندازه‌گیری مقدار فاصله (d(x)) یا شباهت (p(x)) یک شیء x در فضای ویژگی به کلاس هدف (کلاس رفتار عادی) و مولفه دوم، حد آستانه روی مقدار فاصله یا شباهت می‌باشد. در فرآیند تشخیص، یک شیء جدید x برچسب عادی می‌خورد، اگر فاصله آن از کلاس عادی کوچک‌تر از حد آستانه باشد (d(x)<θ) و اگر شباهت آن به کلاس عادی بزرگ‌تر از حد آستانه باشد با برچسب (p(x)<θ) مشخص می‌شود. دسته‌بندی‌های تک‌کلاسی را با توجه به روشی که در حل مساله دسته‌بندی تک‌کلاسی به‌کار می‌گیرند و مدلی که از کلاس هدف ارائه می‌کنند در سه گروه قرار می‌دهند، روش‌های مبتنی بر مرز (مانند SVM و SVDD) ، روش‌های مبتنی بر چگالی (PDE و MOG) و روش‌های مبتنی بر دوباره‌سازی[1] (مانند SOM، K-means و PCA).

در این مقاله برای یادگیری رفتار عادی درخواست HTTP از دسته‌بندی‌های تک‌کلاسی زیر استفاده می‌کنیم:

۱- دسته‌بندی تک‌کلاسی SVDD: یک ابرکره[2] را بر داده‌های کلاس موجود (به‌عنوان کلاس‌هدف) محاط می‌کند. محدوده ابرکره توسط اشیائی از کلاس‌هدف تعیین می‌شود. این اشیاء بردارهای پشتیبان نامیده می‌شوند. در SVDD فاصله شیء x از کلاس هدف طبق رابطه زیر محاسبه می‌شود:

$$D_{SVDD}(x) = k(x,x) - 2\sum_i \alpha_i * k(x,x_i) + \sum_{i,j} \alpha_i\alpha_j * k(x_i,x_j) \quad (۱)$$

که k نشان دهنده تابع‌هسته، x_i و α_i ضریب لاگرانژ منتسب به بردار پشتیبان x_i است.

۲- اختلاط مدل‌های گوسی(MOG)[3]: اختلاط مدل‌های

1-Reconstruction
2-Hyper Sphere
3-Mixture of Gaussian model

گوسی یک ترکیب خطی از توزیع نرمال است که تابع چگالی آن طبق رابطه زیر بدست می آید:

$$P_{MOG}(x) = \frac{1}{N_{MOG}} \sum_i \alpha_i P_N(x; \mu_i, \sum_i) \quad (۲)$$

α_i ضریب اختلاط است. MOG بایاس کمتری نسبت به یک تابع توزیع نرمال دارد و در عوض به داده‌های بیشتری برای آموزش نیاز دارد. در صورتی که MOG با داده‌های کمتری آموزش داده می‌شود واریانس بیشتری از خود نشان می‌دهد. وقتی تعداد مدل‌های گوسی، N_{MOG}، مشخص باشد، میانگین و کوواریانس هر کدام از مدل‌های گوسی با روش بیشینه‌سازی امید ریاضی[1] تعیین می‌شود.

۳- تصمیم چگالی پارزن[2] (PDE): تخمین چگالی پارزن روشی برای تخمین چگالی احتمال یک متغیر تصادفی می باشد. برای‌هر شیء x، چگالی تخمینی از رابطه زیر به‌دست می‌آید:

$$P_{PDE}(x) = \frac{1}{N} \sum_i K_h(x - x_i) \quad (۳)$$

که تابع‌هسته مورد استفاده (اغلب گوسی)، N تعداد اشیاء موجود در مجموعه‌داده آموزش، x_i، iامین شیء موجود در مجموعه‌داده آموزش و h عرض‌هسته است که با آموزش و با استفاده از روش بیشینه مقدار احتمال[3] تعیین می‌شود.

۴- ماشین بردار پشتیبان (SVM): این روش ابرصفحه‌هائی با حداکثر حاشیه را به‌دست می‌آورد که دسته‌های داده‌ها را ازهم جدا کنند. هدف، پیدا کردن بهترین خط (ابر صفحه) که دو دسته را از هم جدا کند. در حالت دو بعدی معادله این خط به‌صورت زیر است [۹]:

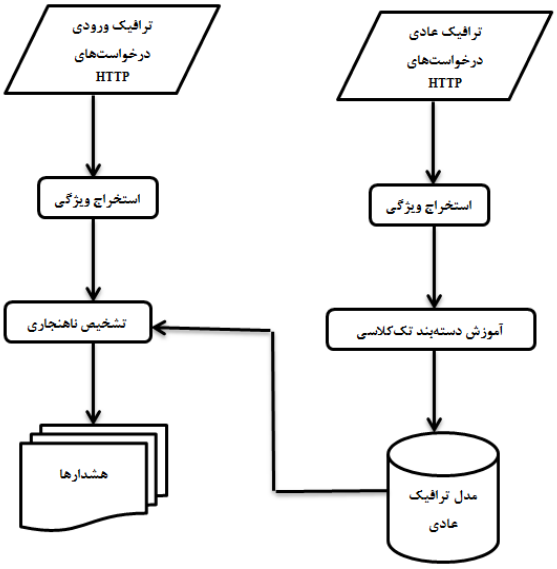
$$w_1 X_1 + w_2 X_2 + b = 0 \quad (۴)$$

در حالت n بعدی خواهیم داشت:

$$\sum_{i=0}^{n} w_i x_i + b = 0 \quad (۵)$$

مدل سیستم تشخیص ناهنجاری پیشنهادی برای هر درخواست HTTP در شکل (۱) نشان داده شده است.

برای تشخیص حمله، با استفاده از یادگیری ماشین و با رویکرد تشخیص ناهنجاری ابتدا مدل رفتار عادی برنامه کاربردی تحت وب بر مبنای پروتکل HTTP یادگیری شده و سپس با اعمال درخواست‌های HTTP، انحراف از حالت عادی اندازه‌گیری می‌شود. در روش پیشنهادی، مرحله آموزش دسته‌بندها به‌صورت برون‌خطی انجام می‌شود. به‌عبارت دیگر مدل رفتار عادی درخواست‌ها مبتنی بر پروتکل HTTP، توسط دسته‌بندهای تک‌کلاسی قبل از شروع به استفاده از برنامه کاربردی تحت وب یادگیری می‌شود و سپس از مدل‌های یادگیری‌شده درهنگام کارکرد برنامه کاربردی برای تشخیص ناهنجاری استفاده می‌شود. روش پیشنهادی، در مرحله یادگیری رفتار عادی برنامه کاربردی تحت وب، به مجموعه داده حمله ندارد و صرفا از مجموعه‌دادگان رفتار عادی برنامه کاربردی برای ساختن مرزهای تصمیم بهره می‌گیرد.



شکل (۱). آموزش و تشخیص سیستم تشخیص ناهنجاری.

ما در روش پیشنهادی خود، مساله تشخیص ناهنجاری را به صورت مساله تصمیم‌گیری گروهی دنبال می‌کنیم و روش‌های متداول ترکیب و نوعی روش میانگین مرتب شده وزن‌دار[4] (OWA) موسوم به S-OWA، را در آن‌ها به‌کار می‌گیریم. در حالت کلی، فرآیند تصمیم‌گیری گروهی عبارت است از حالتی که دو یا چند متخصص، هر کدام با عقاید و ویژگی‌های منحصر به‌فرد خود سعی می‌کنند تا یک تصمیم مشترک بگیرند. مهمترین مساله‌ای که در تصمیم‌گیری گروهی مطرح می‌شود این است که چگونه نظرات متخصصین با هم ترکیب شود طوری‌که تصمیم گرفته‌شده در جهت ارضای معیار مشخصی باشد[۵].

دسته‌بندهای تک‌کلاسی به سختی می‌توانند تمامی مشخصات[5] داده را در نظر بگیرند. ترکیب دسته‌بندها به‌همین منظور مطرح می‌شود. محققان به‌طور مستمر به دنبال بهبود کارآیی روش‌های پیشنهادی در مسایل دسته‌بندی می‌باشند و ترکیب دسته‌بندها یکی از راه‌های دست‌یابی به این‌هدف است. ترکیب دسته‌بندها منجر به بهبود کارآیی و استحکام دسته‌بندی در ازای افزایش پیچیدگی می‌شود. فرض کنید برای فرآیند

1-Expectation-maximization
2-Parzen density estimator
3-Maximum likelihood

4- Order weighted averaging
5-Characteristics

یادگیری، یکی از دسته‌بندها با توجه به قدرت آن دسته‌بند در تشخیص حمله‌های موجود انتخاب شود و در آینده حمله‌ای جدید در شبکه اعمال شود که در نقطه کور دسته‌بند مورد استفاده قرار داشته باشد، در نتیجه حمله تشخیص داده نخواهد شد. درصورتی‌که استفاده از چند دسته‌بند، به شرط این‌که دسته‌بندهای انتخاب‌شده، رویکردهای یادگیری متفاوتی داشته باشند و مکمل یکدیگر باشند، احتمال مواجهه با حالت مذکور را کاهش می‌دهد و نقطه کور یک دسته‌بند با دسته‌بندهای دیگر پوشش داده می‌شود.

روش‌های مختلفی مانند میانگین‌گیری، رای اکثریت، انتخاب دسته‌بند کمینه، انتخاب دسته‌بند بیشینه، انتخاب دسته‌بند میانه و قالب‌های تصمیم نظیر S-OWA برای ترکیب خروجی دسته‌بندها پیشنهاد شده است.

۴-۴- استفاده از روش تولید وزن‌های S-OWA برای ترکیب دسته‌بندها

با استفاده از روش تولید وزن‌های S-OWA می‌توان نظرات دسته‌بندها با یکدیگر ترکیب نمود. برای ترکیب دسته‌بندهای تکی ذکرشده از عملگر S-OWA استفاده شد. این روش برای جمع‌بندی نظر دسته‌بندها درباره درخواست‌های HTTP به‌کار بسته‌شد. درجه orness را به‌صورت زیر تعریف می‌کنیم [۱۴]:

$$\frac{1}{2}\sum_{i=1}^{n} w_i = \frac{1}{2} \qquad (۵)$$

برای به‌دست آوردن تعریف جدید، این فرمول را تغییر می‌دهیم و از تساوی زیر استفاده می‌کنیم:

$$\frac{1}{2}\sum_{i=1}^{n} w_i = \frac{1}{2} \qquad (۶)$$

فرمول درجه orness را به صورت زیر بازنویسی می‌کنیم

$$orness(W) = \frac{1}{2} + \sum_{i=1}^{n}(\frac{(n-i)}{(n-1)} - \frac{1}{2})w_i = \frac{1}{2} + \sum_{i=1}^{n}\frac{(n-2i+1)}{2(n-1)}w_i$$

$$orness\,(W) = \frac{1}{2} + \sum_{i=1}^{n} q_i w_i \qquad (۷)$$

حال در نظر می‌گیریم وضعیتی را که n زوج باشد، n = 2m؛ همچنین i = k ، k <= m و i=n+1-k خواهیم داشت:

$$q_k = \frac{(n-2k+1)}{2(n-1)}$$

$$q_{n+1-k} = \frac{n-2n-2+2k+1}{n-1} = \frac{-n+2k-1}{2(n-1)} = -q_k \qquad (۸)$$

همچنین

$$orness(W) = \frac{1}{2}\sum_{i=1}^{n} q_k\,(w_k - w_{n+1-k}) \qquad (۹)$$

اگر n فرد باشد، بنابراین n = 2m+ 1 و خواهیم داشت:

$$orness(W) = \frac{1}{2}\sum_{i=1}^{n} q_k(w_k - w_{n+1-k}) + q_{m+1}\,w_{m+1} \qquad (۱۰)$$

$$q_{m+1} = \frac{2m+1-2(m-1)+1}{2(n-1)} \qquad (۱۱)$$

پس برای orness خواهیم داشت

$$orness(W) = \frac{1}{2}\sum_{i=1}^{n} q_k\,(w_k - w_{n+1-k}) \qquad (۱۲)$$

با استفاده از این عبارت مستقیما روشی را برای ساختن عملگرهای S-OWA با وزن‌هایی با درجه orness از پیش تعیین شده بیان می‌کنیم.

فرض می‌کنیم درجه orness با نام Ω از پیش داده شده باشد. می‌توانیم فرض کنیم که تمامی وزن‌های به‌جز w_1 و w_n مساوی باشند. با این فرض تابع orness به‌سادگی به‌صورت زیر درمی‌آید:

$$orness(W) = \frac{1}{2} + q_1(w_1 - w_n) = \frac{1}{2} + \frac{1}{2}(w_1 - w_n) \qquad (۱۳)$$

با درجه orness از پیش تعیین‌شده Ω می‌توان به تعریف واضحی برای تفاوت میان اولین و آخرین وزن رسید:

$$w_1 - w_n = 2(\Omega - 0.05) \qquad (۱۴)$$

می‌توان w_1 و w_n را هر عددی در بازه بین صفر و یک انتخاب کرد به‌طوری که شرط فوق را برآورده نمایند. سپس مجموع وزن‌های باقی‌مانده می‌بایستی بین صفر و یک باقی بماند. بنابراین داریم

$$\mathrm{wi} = \frac{1}{n}[\,1 - (\mathrm{w1} - \mathrm{wn})\,],\ \mathrm{i} = 2,3,\ldots,\mathrm{n}-1 \qquad (۱۵)$$

سپس فرآیند فوق را با الگوریتم زیر تغییر اندکی می‌دهیم:

1. $\Delta = 2(\Omega - 0.05)$
2. Let
$$L = \frac{1}{n}\,(1 - |\Delta|$$
3. $for\ i = 2,\ldots,n-1$
$$w_i = L$$
4. $if\ \Delta > 0\ then$
$w_1 = L + \Delta\,,\qquad w_n = L$
$if\ \Delta \le 0\ then$
$w_1 = L\,,\qquad w_n = L + L$

با چنین فرآیند وزن‌های S-OWA تولید می‌شود؛ در واقع ما پس از تولید وزن‌ها به خروجی‌هر دسته‌بند؛ یکی از وزن‌های تولیدشده توسط عملگر S-OWA را اختصاص می‌دهیم. برای ما در این پژوهش چنانچه $\Delta > 0$ باشد، $w_1 = L + \Delta$ و $w_i = L$ برای $i = 1,..n$ خواهد بود. در این حالت برای مقدار ترکیب

یافته که در واقع خروجی حاصل از ترکیب خروجی دسته‌بندهاست خواهیم داشت

$$F(a_1, \dots a_n) = \Delta\, Max_i[\alpha_i] + L \sum_{i=1}^{n} \alpha_i = \Delta\, Max_i[\alpha_i] + \frac{(1-\Delta)}{n} \sum_{i=1}^{n} \alpha_i \quad (۱۶)$$

بنابراین خواهیم داشت

$$F(a_1, \dots a_n) = \Delta\, Max_i[\alpha_i] + (1-\Delta)\, Ave(a_1, \dots, a_n) \quad (۱۷)$$

فرمول فوق عملگر S-OWA نامیده شده است [۹]. اگر نتایج حاصله به‌گونه‌ای باشد که $\Delta <= 0$ باشد خواهیم داشت:

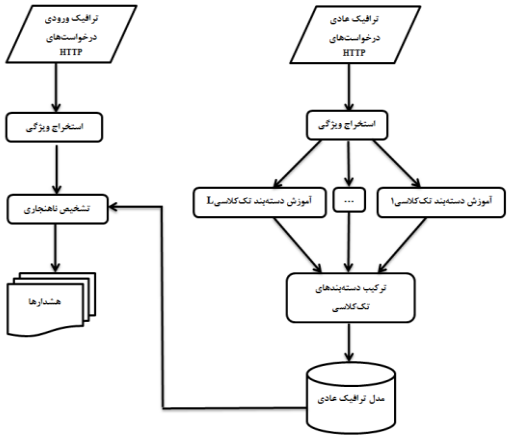
$$F(a_1, \dots a_n) = \Delta\, Min_i[\alpha_i] + (1-\Delta)\, Ave(a_1, \dots, a_n) \quad (۱۸)$$

۴-۵- روش پیشنهادی

همان طور که بیان شد، در تشخیص ناهنجاری برنامه‌های کاربردی تحت وب؛ تنها مجموعه‌دادگان درخواست‌های HTTP عادی در دسترس است و در مرحله آموزش ما می‌خواهیم رفتار عادی درخواست‌های HTTP را مدل کنیم تا در مرحله تشخیص، درخواست‌های ورودی به سیستم پیشنهادی ما با این مدل عادی مقایسه شوند. دیدیم برای تهیه مدل عادی از رفتار برنامه‌های کاربردی تحت وب از دسته‌بندهای تک‌کلاسی استفاده شد. در این بخش به‌جای استفاده از هر کدام از دسته‌بندهای تک‌کلاسی به‌صورت مستقل، به‌عنوان یک سیستم تشخیص ناهنجاری برای درخواست‌های HTTP، از ترکیب نظرات این دسته‌بندها استفاده می‌کنیم و با این سیستم ترکیبی مراحل آموزش و تشخیص را برای سیستم پیشنهادی انجام می‌دهیم.

در معماری سیستم ترکیبی پیشنهادی نیز برای تشخیص ناهنجاری، ابتدا مدل رفتار عادی برنامه کاربردی تحت وب بر مبنای پروتکل HTTP با استفاده از دسته‌بند ترکیب یافته از دسته‌بندهای تک‌کلاسی؛ یادگیری‌شده و سپس با اعمال درخواست‌های HTTP، انحراف از حالت عادی اندازه‌گیری می‌شود. در اینجا نیز، مرحله آموزش دسته‌بند به‌صورت برون‌خطی[1] انجام می‌شود. به عبارت دیگر مدل رفتار عادی درخواست‌ها مبتنی بر پروتکل HTTP، توسط دسته‌بند تک‌کلاسی ترکیبی قبل از شروع به استفاده از برنامه کاربردی تحت وب یادگیری می‌شود و سپس از مدل یادگیری‌شده درهنگام کارکرد برنامه کاربردی برای تشخیص ناهنجاری استفاده می‌شود. همچنین، در مرحله یادگیری رفتار عادی برنامه کاربردی تحت وب، به مجموعه داده حمله دسترسی وجود ندارد و صرفا از مجموعه داده رفتار عادی برنامه کاربردی برای ساختن مرزهای تصمیم بهره‌گیری می‌شود.

شکل (۲) نحوه عملکردهر درخواست HTTP در روش‌های ترکیبی را نشان می‌دهد. درخواست‌های تولیدشده مربوط به‌هر برنامه کاربردی، به بردار ویژگی تبدیل می‌شوند و بردار ویژگی تولید شده متناسب را به‌عنوان ورودی به دسته‌بندهای تک‌کلاسی می‌دهند. سپس خروجی دسته‌بندها (به‌عنوان معیار شباهت بردار ویژگی به کلاس رفتار عادی) باهم ترکیب می‌شوند تا وضعیت عادی یا غیرعادی بودن‌هر درخواست بر مبنای آن شکل بگیرد.



شکل (۲). نمودار سیستم ترکیبی تشخیص ناهنجاری.

در مورد مساله مورد پژوهش ما نیز از روش‌های معمول ترکیب به‌همان صورت استفاده می‌شود و فرآیند تصمیم‌گیری گروهی انجام می‌پذیرد. در واقع هر دسته‌بند تک‌کلاسی به‌عنوان عنصری شرکت‌کننده در تصمیم‌گیری گروهی نظر خود را پیرامون عادی یا ناهنجار بودن درخواست HTTP ورودی اعلام می‌کند و در نهایت با یکی از روش‌های ترکیب تجمیع نظرات دسته‌بندها پیرامون آن درخواست صورت می‌پذیرد و تصمیم نهایی اعلام می‌شود. ما در پژوهش خود از روش ترکیب با استفاده از عملگر S-OWA به عنوان استراتژی ترکیب، استفاده می‌نماییم و تصمیم‌گیری گروهی را انجام می‌دهیم.

برای تشخیص حمله، با استفاده از یادگیری ماشین و با رویکرد تشخیص ناهنجاری ابتدا مدل رفتار عادی برنامه کاربردی تحت وب بر مبنای پروتکل HTTP یادگیری شده و سپس با اعمال درخواست‌های HTTP، انحراف از حالت عادی اندازه‌گیری می‌شود. در روش پیشنهادی، مرحله آموزش دسته‌بندها و ترکیب آنها به صورت برون‌خطی انجام می‌شود[۱۴-۱۶]. به عبارت دیگر مدل رفتار عادی درخواست‌ها مبتنی بر پروتکل HTTP، توسط دسته‌بندهای تک‌کلاسی یا ترکیب آنها قبل از شروع به کار برنامه کاربردی تحت وب یادگیری می‌شود و سپس از مدل‌های یادگیری شده درهنگام کارکرد برنامه کاربردی برای تشخیص ناهنجاری استفاده می‌شود. روش پیشنهادی، در مرحله یادگیری رفتار عادی برنامه کاربردی تحت وب، نیازی به مجموعه‌داده حمله ندارد و صرفا از مجموعه‌داده رفتار عادی برنامه کاربردی برای ساختن مرزهای تصمیم بهره می‌گیرد.

1- Offline

برای یادگیری رفتار عادی پروتکل ما چهار دسته‌بند تک‌کلاسی SVDD، MOG، PDE و SVM را روی بردار خصیصه‌های استخراج شده از مجموعه‌داده CSIC2012 به‌کار می‌بریم. سپس از روش‌های ترکیب با استفاده از عملگر S-OWA برای ترکیب نتایج حاصله از آن دسته‌بندها می‌پردازیم. معماری کلی روش پیشنهادی در شکل (۲) نمایش داده شده است.

۵- نتایج تجربی و آزمایشات

۵-۱- مجموعه‌داده

مجموعه‌داده مورد استفاده در آزمایشات انجام شده در این مقاله، مجموعه‌داده CSIC2012 می‌باشد. این مجموعه‌داده شامل درخواست‌های نرمال یا ناهنجار متعلق به تمامی صفحات وب مرتبط با یک برنامه کاربردی تحت وب تجاری است و پارامترهای مرتبط با درخواست‌های HTTP آن شامل مقادیر مختلفی می‌باشد. مجموعه‌داده CSIC2012 شامل حملات وب نوینی نظیر تزریق injection، سرریز بافر، تزریق CRLF و XSS می‌باشد.

۵-۲- معیارهای ارزیابی

۵-۲-۱- نرخ تشخیص و نرخ‌هشدار نادرست

از دو معیار نرخ تشخیص (DR) و نرخ‌هشدار نادرست (FAR) برای ارزیابی کارآیی سیستم تشخیص ناهنجاری پیشنهادی برای برنامه‌های کاربردی تحت وب می‌توان استفاده کرد. برای این دو معیار داریم:

$$DR = \frac{TP}{TP + FN} \quad (۱۹)$$

$$FAR = \frac{FP}{FP + TN} \quad (۲۰)$$

که در آن TP تعداد درخواست‌های HTTP ناهنجاری‌هستند که به درستی تشخیص داده شده اند و FN تعداد درخواست‌های ناهنجاری‌هستند که به عنوان عادی تشخیص داده شده‌اند. FP تعداد درخواست‌های عادی هستند که که به نادرستی ناهنجار تشخیص داده شده‌اند و TN تعداد درخواست‌هایی است که به درستی عادی تشخیص داده شده‌اند.

به صورت ایده آل سیستم تشخیص ناهنجاری می‌بایستی نرخ تشخیص ۱۰۰٪ و نرخ‌هشدار نادرست ۰٪ داشته باشد. با این حال در عمل این امر به سختی محقق می‌شود.

۵-۲-۲- منحنی ROC[1]

ایده اساسی سیستم‌های تشخیص ناهنجاری محاسبه احتمال ناهنجار بودن درخواست‌های HTTP، بر اساس نتایج آزمون تشخیص ناهنجاری می‌باشد. تحلیل‌های ROC برای مشخص کردن دقت واقعی نتایج تشخیص است. برای بررسی عملکرد سیستم‌های تشخیصی منحنی‌های ROC از اهمیت ویژه‌ای برخوردارند [۱۷]. تحلیل‌های ROC رویکردی استاندارد است که برای مشخص‌کردن حساسیت و ویژگی تشخیص‌ها به‌کارمی‌روند. برای این منظور، منحنی ROC برای تعریف کردن رابطه حساسیت و ویژگی سیستم تشخیصی به‌کار می‌رود.

منحنی‌ها بین صفر و یک قرار می‌گیرند. منحنی‌های که درهمسایگی نیمساز ۴۵ درجه هستند معرف سیستم‌های تشخیصی نامناسب هستند وهمچنین نمودارهای که مساحت زیر منحنی ROC و مساوی یا کمتر از مساحت بالای منحنی باشد نشان دهنده آزمایشی غیرموفقیت آمیز هستند.

۵-۲-۳- مقادیر AUC

سطح زیر نمودار ROC (AUC[2]) به‌عنوان یک معیار معمول و شناخته‌شده برای مقایسه روش‌های دسته‌بندی و داده‌کاوی به‌کار می‌رود. شش الگوریتم مختلف دسته‌بندی روی شش مجموعه‌دادگان پزشکی واقعی مورد آزمایش قرار گرفته و اعلان شده است که AUC خواص دقت بهتری نسبت به ROC از خود نشان می‌دهد و معیار بهتری برای مقایسه الگوریتم‌های دسته‌بندی می‌باشد [۱۱].

معیارهایی که برای ارزیابی دسته‌بندهای تک‌کلاسی به‌عنوان تشخیص‌دهنده ناهنجاری در این پژوهش به‌کار برده‌ایم، علاوه بر نرخ تشخیص و نرخ هشدار نادرست، سطح زیر نمودار AUC نیز بوده است.

۵-۳- ارزیابی نتایج

به‌منظور مقایسه کارآیی روش ترکیب دسته‌بندها در مقایسه با استفاده از دسته‌بندها به‌صورت مستقل؛ نتایج مقایسه‌ای در جدول (۲) ؛ روی بردارهای ویژگی استخراج شده از مجموعه‌داده CSIC2012 حاصل شده است. همه نتایج با استفاده از پردازنده Intel Core i5 ، 2.53 GHz متعلق به سیستم HP Pavilion dv3 Notebook PC با سیستم عامل Microsoft Windows 7؛ به‌دست آمده است.

برای هر دسته‌بند پارامترها طوری تنظیم شده‌اند که عملکرد آن دسته‌بند بهینه شود. در دسته‌بند MOG پارامتر regularization برای ماتریس کوواریانس ۰/۰۱ و تعداد تکرار[3] الگوریتم ۲۵ بوده است. برای دسته‌بند PDE، مقدار تخمین شباهت بیشینه[4] برای هموارسازی[5] تخمین چگالی Parzen، ۰/۰۵بوده است. پارامتر کرنل گاوسی برای دسته‌بند SVDD،

1- Receiver Operating Characteristics Curve

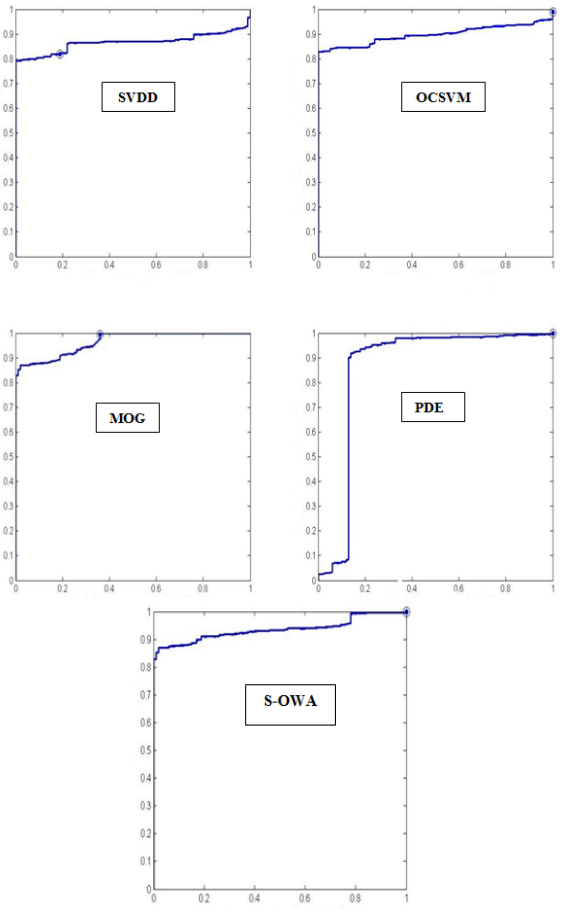
2- Area Under the Curve
3- Iteration
4- Maximum likelihood estimation
5- Smoothing

۰/۰۰۵ و پارامتر کرنل RBF برای دسته‌بند OCSVM، ۰/۰۲ در نظر گرفته شده است.

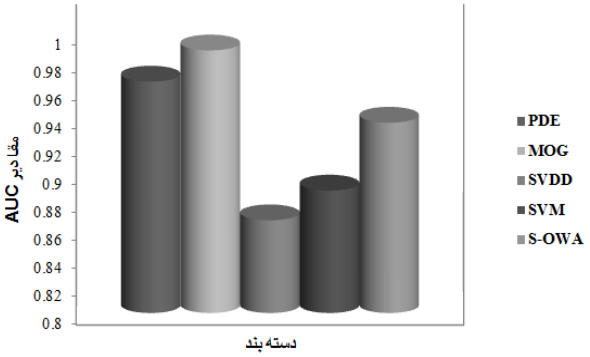
شکل (۳) نمودار ROC مربوط به داده‌های حمله‌های مختلف موجود در مجموعه‌دادگان را با استفاده از چهار دسته‌بند مذکور را نشان می‌دهد. در نمودارهای ROC، محور عمودی نشان‌دهنده نرخ تشخیص حمله و محور افقی نشان‌دهنده نرخ‌هشدار نادرست می‌باشد. همچنین مقدار AUC نیز با استفاده از این دسته‌بندها در شکل (۴) نشان داده شده است.

نرخ تشخیص و نرخ‌هشدار نادرست هر دسته‌بند یا ترکیب دسته‌بندها پس از مرحله یادگیری و در مرحله آزمایش؛ به‌عنوان مقیاس کارآیی ذکر شده است.

همان طوری که در نتایج مشاهده می‌شود؛ در ترکیب دسته‌بندهای تک‌کلاسی با استفاده از عملگر S-OWA نرخ تشخیص ترکیب دسته‌بندها نرخ تشخیص بهبود یافته و نرخ‌هشدار نادرست نیز در مقایسه با به‌کارگیری مستقل دسته‌بندها کاهش می‌یابد و کارآیی روش پیشنهادی و ایده ترکیب دسته‌بندها در تشخیص ناهنجاری برنامه‌های کاربردی تحت وب بخوبی اثبات می‌شود.



شکل(۳). منحنی ROC دسته‌بندهای تک‌کلاسی



شکل(۴). مقادیر AUC دسته‌بندهای تک‌کلاسی

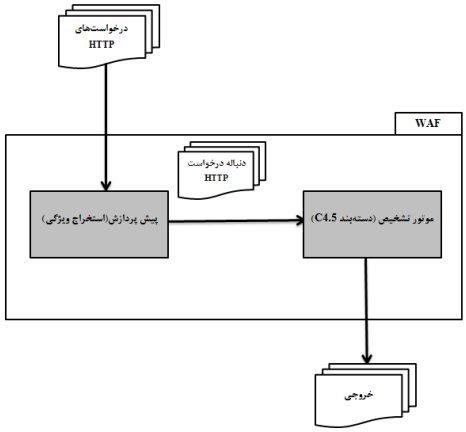
همان طوری که در نتایج مشاهده می‌شود؛ در ترکیب دسته‌بندهای تک‌کلاسی با استفاده از عملگر S-OWA نرخ تشخیص ترکیب دسته‌بندها نرخ تشخیص بهبود یافته و نرخ‌هشدار نادرست نیز در مقایسه با به‌کارگیری مستقل دسته‌بندها کاهش می‌یابد و کارآیی روش پیشنهادی و ایده ترکیب دسته‌بندها در تشخیص ناهنجاری برنامه‌های کاربردی تحت وب بخوبی اثبات می‌شود.

برای ارزیابی کارآیی روش پیشنهادی خود آن را با سیستم امنیت تحت وب نوینی مقایسه می‌نماییم. سیستم پیشنهاد شده که دیواره آتش برنامه کاربردی تحت وب[1] نام گرفته؛ پس از پیش پردازش داده‌ها، در موتور تشخیص خود از الگوریتم درخت تصمیم C4.5 استفاده نموده است. علت استفاده از این الگوریتم، کاربرد گسترده آن در حیطه تشخیص نفوذ و موفقیت الگوریتم مبتنی بر درخت تصمیم در مسابق تشخیص نفوذ DARPA بیان گردیده است. WAF پیشنهادی برای دسته‌بندی درخواست‌های HTTP با نمونه‌های عادی و حمله آموزش داده می‌شود. در مرحله آموزش، مجموعه‌دادگان اولیه برای آموزش دسته‌بندی درخت تصمیم به سیستم ارائه می‌شود. ساختار روش پیشنهادی در شکل (۵) آمده است. نخست، پیش‌پردازش برای استخراج ویژگی‌های و برچسب هر بسته HTTP انجام می‌شود. سپس، یک سوم مجموعه‌دادگان برای آموزش تشخیص‌دهنده با استفاده از الگوریتم C4.5 در نرم‌افزار WEKA؛ به‌کار گرفته می‌شود. نرم افزار درخت تصمیمی که موتور تشخیص WAF را بازنمایی می‌کند؛ به‌عنوان خروجی، بازمی‌گرداند [۱].

این تحقیق برای ارزیابی کارآیی روش خود از معیارهای نرخ تشخیص و نرخ‌هشدار نادرست استفاده کرده است.

با مقایسه نتایج ارزیابی WAF با روش‌هایی که ما در این پژوهش از آن‌ها بهره گرفتیم، مشاهده می‌شود نرخ تشخیص آن از بسیاری از روش‌های پیشنهادی به مراتب کمتر است و صرفا در حد روش‌هایی نظیر MOG و روش ترکیبی کمینه است.

1- Web application firewall (WAF)



شکل(۵). دیواره آتش WAF

نرخ‌هشدار نادرست WAF در حد دسته‌بندهای تک‌کلاسی است؛ هرچند از آن‌ها کمتر است. نکته قابل توجه نرخ‌هشدار نادرست بسیار پایین‌تر روش‌های ترکیبی در مقایسه با این روش است که تفاوت بسیار مشهود است و کارآیی بسیار بالاتر این روش‌ها نسبت به WAF را نشان می‌دهد. (جدول ۲)

جدول (۲). نتایج حاصل از روش پیشنهادی و مقایسه با روش دیواره آتش WAF

عنوان روش	نرخ تشخیص	نرخ‌هشدار نادرست
SVDD	۹۸/۹	۲/۲۸
MOG	۹۵/۸	۳/۸
PDE	۹۸/۶	۴
SVM	۹۷/۷۷	۳/۰۱
استراتژی S-OWA	۹۹	۰/۲
روش دیواره آتشWAF	۹۵/۷	۴/۶۸

۶- نتیجه‌گیری

با توجه به فراگیر شدن استفاده از فناوری اطلاعات و ارتباطات در عصر حاضر و استفاده از سامانه‌های تحت وب در کاربردهای حساس و محرمانه فرماندهی نظامی، تامین امنیت این سامانه‌ها از دغدغه‌های مدیران ارشد نظامی است. ارائه روشی که از دقت و صحت بالایی در برقراری امنیت سامانه‌های نظامی تحت وب برخوردار باشند؛ بسیار ضروری می‌نماید.

در این مقاله از ترکیب دسته‌بندهای تک‌کلاسی رایج به‌منظور تشخیص درخواست‌های HTTP ناهنجار در برنامه‌های کاربردی تحت وب استفاده شد و پردازش روش پیشنهادی روی درخواست‌های مجموعه‌داده CSIC2012 انجام گرفت. برای ترکیب دسته‌بندهای تک‌کلاسی از استراتژی عملگر S-OWA، استفاده شده است. استفاده از تصمیم‌گیری گروهی خصوصا با روش S-OWA، معیارهای کارآیی سیستم تشخیص ناهنجاری را بخوبی بهبود بخشیده است. فرآیند تشخیص ناهنجاری با تصمیم‌گیری گروهی بر مبنای عملگر S-OWA سبب بهبود نرخ‌هشدار نادرست به‌طور چشمگیری می‌شود و نرخ تشخیص مناسبی نیز دارد؛ به‌طوری که نرخ تشخیص به ۹۹ درصد و نرخ‌هشدار نادرست نیز به ۰/۲ درصد رسیده است.

در روش ذکرشده با ترکیب دسته‌بندهای تک‌کلاسی متداول با استفاده از عملگر S-OWA؛ نرخ تشخیص افزایش یافته و نرخ‌هشدار نادرست نیز بخوبی کاهش یافت.

پژوهش‌های آینده می‌توانند روی میزان تاثیر استفاده از روش‌های دیگر دسته‌بندی تک‌کلاسی به‌صورت مستقل و روش‌های دیگر ترکیب این دسته‌بندها؛ در بهبود کارآیی سیستم‌های تشخیص نفوذ مبتنی بر تشخیص ناهنجاری متمرکز شوند. می‌توان با مطالعات بیشتر ویژگی‌های دیگری برای توصیف درخواست‌های HTTP تعریف کرد تا به‌صورت جامع‌تری عملکرد درخواست‌ها را توصیف نماید. دسته‌بندهای مختلف دیگری را می‌توان با روش‌های متنوع ترکیب نمود و نتایج را مورد ارزیابی قرار داد. چنانچه مجموعه داده‌های جدیدتری هم در اختیار باشند، نتایج روش پیشنهادی را می توان بهتر و دقیق‌تر بررسی کرد.

۷- مراجع

[1] Iranian Passive Defense Organization, "Internet, the Newest and Most Effective Weapon," [Online]: Available: http://paydarymelli.ir/fa/news /2499, Accessed: 2014.

[2] Iranian Passive Defense Organization "Cyber Wars in 21th Century,"[Online]:Available: http://paydarymelli.ir/fa/news/2472, Accessed: 2014.

[3] C. Kruegel and G. Vigna, "Anomaly Detection of Web-based Attacks," In Proc of the 10th ACM Conference on Computer and Communications Security, ACM New York, pp. 251-261, 2003.

[4] H. T. Nguyen, " Reliable Machine Learning Algorithms for Intrusion Detection Systems, Ph.D. dissertation, Dept. of Computer Science, Gjøvik University College, Gjøvik, Norway, 2012.

[5] C. Torrano-Gimenez, H. T. Nguyen, G. Alvarez, and K. Franke, "Combining Expert Knowledge with Automatic Feature Extraction for Reliable Web Attack Detection," Security and Communication Networks, vol. 8, pp. 2750-2767, August 2012.

[6] K. L. Ingham, "Anomaly Detection for HTTP Iintrusion Detection: Algorithm Comparisons and the Effect of Generalization on Accuracy," Ph.D. dissertation, Dept. of Computer Science, The University of New Mexico, Albuquerque, USA, 2007.

[7] G. M. Nascimento, "Anomaly Detection of Web-based Attacks," M. S. Thesis, Dept. of Computer Science, University of Lisbon, Lisbone, Portugal, 2010.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A survey," ACM Computing Surveys (CSUR), vol. 41, p. 15, July 2009.

[9] D. M. J. Tax, "One-Class Classification," Ph.d dissertation, Dept. of Computer Science, Delft University, Delft, Netherland, 2001.

[10] S. Khandelwal, P. Shah, M. K. Bhavsar, and S. Gandhi, "Frontline Techniques to Prevent Web Application Vulnerability," International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 2, p. 208, 2013.
[11] X. Ling, J. Huang, and H. Zhang, "Advances in Artificial Intelligence: AUC: a Better Measure than Accuracy in Comparing Learning Algorithms," Advances in Artificial Intelligence, vol. 267, no. 1, pp. 329-341, May 2003.
[12] M. Rahmanimanesh, "Anomaly Detection of Adhoc Networks Using Nodes Validation," Ph.D dissertation, Dept. of Computer Engineering, Tarbiat Modares University, Tehran, Iran, 2013.
[13] M. Reformat and R. R. Yager, "Building ensemble classifiers using belief functions and OWA operators," Soft Computing, vol. 12, pp. 543-558, April 2008.
[14] D. Filev and R. R. Yager, "On the Issue of Obtaining OWA Operator Weights Fuzzy Sets and Systems," vol. 94, pp. 157-169, March 1998.
[15] C. Kruegel, G. Vigna, and W. Robertson, "A Multi-model Approach to the Detection of Web-based Attacks," Computer Networks, vol. 48, pp. 717-738, August 2005.
[16] T. Berners-Lee, R. Fielding, and H. Frystyk, "Hypertext Transfer Protocol, HTTP/1.0", 1996.
[17] A. P. Bradley, "The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms," Pattern recognition, vol. 30, pp. 1145-1159, July 1997.
[18] The HTTP Dataset CSIC2012, Department of Information Processing and Codification (T.I.C.), of the Institute of Applied Physics (I.F.A.), Spanish Scientific Research Council (C.S.I.C.), <http://iec.csic.es/dataset/>, 2012.
[19] D. M. J. Tax, "Dd tools 2012, the Data Description Toolbox for Matlab," version 1.9.1.< http://prlab.tudelft.nl/david-tax/dd_tools.html >, 2013.

Journal Of Electronical & Cyber Defence
Vol. 3, No. 3, 2016, Serial No.11

Web-based Military Management Systems Security Using Combination of One-class Classifiers

A. Jamalyfard*, H. Shirazi

*Malek-Ashtar University Of Technology, Tehran, Iran

(Received: 07/09/2013, Accepted: 12/01/2016)

ABSTRACT

Cyber attacks against the web-based military command systems is very common in the age of electronic warfare. Web application is one of the most widely used tools in the world wide web. Because of its dynamic nature, it is vulnerable to serious security risks. Web-based command and control systems security considerations are very important for the modern military managers. Anomaly based intrusion detection is an approach that focuses on new and unknown attacks.

A method for anomaly detection in web applications using a combination of one-class classifiers, is proposed. First, in preprocessing phase, normal HTTP traffic is logged and Features vector is extracted from each HTTP request. The proposed method consists of two steps; In the training phase, the extracted features vectors associated with each request, enter the system and the model of normal requests , using combination of one-class classifiers, is learned. In the detection phase, anomaly detection operation is performed on the features vector of each each HTTP request using learned model of the training phase. S-OWA operator is used to combine the one-class classifiers. The data used for training and test are from CSIC2012 dataset. Detection rate and false alarm rate obtained from experiments, shows better results than other methods.

Keywords: Military Management, Web-Applications, Intrusion Detection, Combination of One-class Classifiers, S-OWA Operator

* Corresponding Author Email: am.jamalyfard@gmail.com