فصلنامهٔ پدافند غیرعامل و امنیت (علمی ترویجی)
سال هفتم/ شمارهٔ بیست‌وهشتم/ پاییز ۱۳۹۸
صص: ٦٣ - ٨٤

Quarterly of Passive Defense & Security
Vol.7, No.28, Autumn 2019
P-P: 63 - 84

تهدیدشناسی امنیت وبگاه‌ها در جنگ‌های سایبری با استفاده از داده‌کاوی و روش سه‌شاخه‌ای ذهن

وحید یادگاری[1]
صمد سهراب[2]
احمدرضا متین فر[3]

DOR: 20.1001.1.26765756.1398.8.28.4.5

چکیده:

زمینه و هدف: بر اساس آمار نت‌کرافت[4] تا سال ۲۰۱۸ بیش از ۱/۸ میلیارد سایت اینترنتی در دنیا راه‌اندازی شده و حجم اطلاعات در اینترنت تا سال ۲۰۲۰ میلادی به بیش از ۴۴ زتابایت داده در سراسر جهان خواهد رسید. از این رو عملیات در قلمرو اطلاعات به اندازه عملیات در زمین، دریا، هوا و فضا واجد ارزش و اهمیت گردیده و عبارت جنگ سایبری با تمرکز بر جنگ‌افزارهای اطلاعاتی برجسته شده است. از انواع روش‌های جنگ‌های سایبری انجام حملات به وبگاه‌های بانکی، خدمات دولتی، نظامی و اطلاعاتی و... با هدف تخریب، تغییر، به سرقت بردن اطلاعات یا از دسترس خارج کردن خدمات برخط می‌باشد. از این رو هدف این تحقیق اول شناخت تهدیدات وبگاه‌های اینترنتی از طریق مدل سه‌شاخه‌ای ذهن و دوم راهکارهای شناسایی هوشمند آنها از طریق داده‌کاوی است. نتیجه‌کار می‌تواند در طرح‌ریزی و یا بهینه نمودن روش‌های سامانه‌های دفاعی و مقابله‌ای مؤثر واقع شود.

روش تحقیق: با مطالعه اسنادی و کتابخانه‌ای، روش‌های مورد استفاده در سایر کارهای ارائه شده مرور گردید و با جمع‌بندی صورت پذیرفته، یک الگوریتم بهینه و ترکیبی با بهره‌گیری از الگوریتم‌های واریانس، آنتروپی و جنگل تصادفی موازی که الگوریتم‌های پرکاربرد داده‌کاوی است ارائه گردید.

پیاده‌سازی و نتیجه تحقیق:

برای پیاده‌سازی مدل پیشنهادی، در محیط آزمایشگاهی، ترافیک مصنوعی برای مختل نمودن خدمات یک وبگاه ایجاد و پس از جمع‌آوری ترافیک شبکه، روش پیشنهادی مورد آزمایش قرار گرفت. برای اطمینان از

۱. کارشناس ارشد مهندسی فناوری اطلاعات دانشگاه تربیت مدرس *V.yadegari58@yahoo.com*
۲. کارشناس ارشد مهندسی امنیت فناوری اطلاعات دانشگاه تعالی قم
۳. استادیار دانشگاه جامع امام حسین

4. *Netcraft*

کیفیت خروجی مدل، مدل پیشنهادی بر روی مجموعه داده استاندارد (EPA-HTTP) نیز آزمایش و نتایج آن با سایر روش‌ها مقایسه گردید که عملکرد بهتری داشت.

کلیدواژه: تهدیدات سایبری[1]، حملات وب[2]، جنگ سایبری[3]، داده‌کاوی[4]، تحلیل ترافیک[5]، مدل سه‌شاخه‌ای ذهن

جستارگشایی

واضح است که ظهور اینترنت و فناوری وب، نحوه تعاملات، تجارت و انتشار و تبادل اطلاعات را دگرگون کرده است. از اینترنت می‌توان تحت عنوان «موتور خلاقیت» یاد کرد، اما همین موتور می‌تواند موجب شود تا تهدیدات سایبری با سرعت بیشتر از روش‌های دفاع سایبری پیشرفت کنند. از آنجا که تهدیدات سایبری ماهیتی پیچیده و دائماً در حال رشد دارند، شبکه‌های دفاعی برای آنکه بتوانند همپای آن‌ها رشد کنند، با دشواری مواجه هستند. یکی از این تهدیدات پیچیده و مهم، ساماندهی و انجام حملات سایبری بر روی وبگاه‌های دولتی، تجاری، امنیتی و... است. این‌گونه حملات با توجه به گسترش روزافزون وبگاه‌ها و نرم‌افزارهای تحت وب به‌صورت فزاینده‌ای در حال زیاد شدن هستند. حملاتی که قادر هستند روند فعالیت یک وبگاه را مختل نمایند، اطلاعات را سرقت نمایند، سایت را به‌عنوان سرپل یک اقدام مجرمانه در اختیار قرار گرفته و بهره‌برداری نمایند و... . بدون شک برای پیشگیری از وقوع تهدیدات تاکنون سامانه‌ها و روش‌های مختلفی پیشنهاد و اجرایی شده است به طوری که امروز شاهد توسعه سامانه‌های هوشمند رصد و آنالیز ترافیک وبگاه‌ها هستیم لیکن آن چه مسلم است پویایی و پیچیدگی اقدامات است که باعث می‌گردد الگوریتم‌های شناسایی هوشمند حملات نیز ماهیتی پویا داشته و متناسب با تهدیدات به‌روزآوری و بهینه گردند. با این رویکرد در این تحقیق سعی شده است با ارائه یک الگوریتم هوشمند مبتنی بر داده‌کاوی و تحلیل رخدادهای یک وبگاه، روند شناسایی یکی از پرکاربردترین حملات وبگاه‌ها موسوم به حملات منع خدمت ساده و توزیع‌شده وب مورد تجزیه و تحلیل قرار گیرد. رخدادهای وب در لایه هفتم از مدل OSI[6] (لایه‌های منطقی شبکه) تولید می‌شود. تحلیل ترافیک این لایه که تحت عنوان رخداد یا لاگ در وب سرور ثبت می‌شود، می‌تواند دربرگیرنده وضعیت عادی و یا حملات صورت گرفته به وب سرور باشد. به‌طور مثال، حمله انکار خدمت ساده و توزیع‌شده[7] آسیب‌های

1. *Cyber Threats*
2. *Web Attack*
3. *Cyberwarfare*
4. *Data Mining*
5. *Traffic analysis*
6. *Open System Interconnection*
7. *Distributed Denial of Service(DDOS), Denial-Of-Service (DOS)*

فراوانی به وب سرورها ایجاد کرده است به طوری که بر اساس گزارش سایت آمازون، حمله منع خدمت ساده و توزیع‌شده علت از دست دادن ۶۰۰۰۰۰ دلار در طول ۱۰ ساعت خرابی بوده است. به همین ترتیب، در طول حملات منع خدمت ساده و توزیع‌شده علیه *eBay*، در دسترس بودن (پایگاه اینترنتی) سایت *eBay.com* از ۱۰۰٪ به ۹/۴٪ تنزل یافته بود. در ژانویه ۲۰۰۱، مایکروسافت در طول یک دوره چندروزه حمله خدمت ساده و توزیع‌شده در سایت خود حدود ۵۰۰ میلیون دلار را از دست داد. در سال ۲۰۱۱، حملات خدمت ساده و توزیع‌شده پنج وبگاه با رده‌بالا، یعنی ویزا[1]، مسترکارت[2]، سونی[3]، وردپرس[4] و سازمان سیا را ویران کرد. امروزه، حملات خدمت ساده و توزیع‌شده قادر به تخریب قوی یک سایت در یک حمله ساده هستند. بر طبق برآوردها انتظار می‌رود که هزینه‌های قطع ناگهانی ۲۴ ساعته برای یک شرکت تجارت الکترونیک بزرگ ۳۰ میلیون دلار باشد (*Adi,2016*).

با توجه به اهمیت تهدیدشناسی امنیت فضای سایبری حوزه وبگاه‌ها و نفوذ، جاسوسی و یا اختلال از طریق آن، نیاز است مطابق با اسناد کشورهای پیشرو در حوزه فناوری و دانش، مروری بر محیط اطلاعاتی و فضای سایبری انجام شود تا متناسب با قابلیت‌های فضای سایبری، روش‌های نفوذ سایبری دشمنان در زمان‌های مختلف به خصوص زمان صلح معرفی شود. همچنین نیاز است تا متناسب با ویژگی‌های نفوذ از دیدگاه مقام معظم رهبری، روش‌های نفوذ در حوزه امنیت و سایبری بررسی گردد. برای بررسی و تحلیل مسئله نفوذ و تهدیدشناسی سایبری حوزه وبگاه‌ها، از مدل سه‌شاخه‌ای ذهن استفاده شده است.

بیان مسئله:

در عصر حاضر، شاهد گسترش روزافزون وبگاه‌ها و نرم‌افزارهای تحت وب برای توسعه و ارائه خدمات برخط در زمینه‌های اقتصادی، سیاسی، نظامی و...هستیم، به طوری که برابر گزارش سایت معتبر نت کرافت تا پایان سال ۲۰۱۸، قریب به ۱/۸ میلیارد وبگاه رسماً ثبت گردیده است. این مهم در کنار خلق مزایای فوق‌العاده، متاسفانه مورد سوءاستفاده افراد و دولت‌های متخاصم قرار گرفته و همواره سعی دارند با انجام حملات سازمان‌یافته و جنگ‌های سایبری، اقدام به تخریب، اخلال و سرقت اطلاعات از طریق وبگاه‌ها نمایند که نمونه‌های آن را می‌توان در حملات سایبری به سایت سازمان‌های دولتی کشور در سنوات اخیر مشاهده نمود. بدون شک برای پیشگیری از وقوع تهدیدات و حملات و مصون ماندن در جنگ‌های سایبری و به‌طور ویژه حملات وبگاه‌ها، تاکنون سامانه‌ها و روش‌های مختلفی پیشنهاد و اجرایی شده است به طوری که امروز شاهد توسعه سامانه‌های هوشمند رصد و آنالیز ترافیک وبگاه‌ها هستیم لیکن آن چه مسلم است پویایی و

1. *Visa*
2. *MasterCard*
3. *Sony*
4. *WordPress*

پیچیدگی اقدامات است که باعث می‌گردد الگوریتم‌های شناسایی هوشمند حملات نیز ماهیتی پویا داشته و متناسب با تهدیدات به‌روزآوری و بهینه گردند. با این رویکرد در این تحقیق سعی شده است با ارائه یک الگوریتم هوشمند مبتنی بر داده‌کاوی و تحلیل رخدادهای یک وبگاه، روند شناسایی یکی از پرکاربردترین حملات وبگاه‌ها موسوم به حملات منع خدمت ساده و توزیع‌شده وب مورد تجزیه و تحلیل قرار گیرد.

اهمیت و ضرورت:

واقعیت این است که جنگ سایبری را باید حوزه‌ای جدید و مستقل از جنگ و دفاع در نظر گرفت. جنگ سایبری حوزه‌ای جدید از جنگ است که طی آن بازیگران دولتی و غیردولتی، تلاش می‌کنند از اینترنت به عنوان یک سلاح جهت آسیب رساندن به دشمن استفاده نمایند. به هر حال تحت تأثیر رویدادهای مختلف مثل حملات سایبری روسیه به گرجستان، حملات متقابل ایران، رژیم صهیونیستی و ایالات متحده آمریکا و...، امروزه بیشتر کشورها اهمیت اینترنت به عنوان یک سلاح خطرناک و ویرانگر و در یک کلام جنگ سایبری به عنوان حوزه مستقل از جنگ و دفاع را به رسمیت شناخته‌اند و مباحثی همچون تدوین سند دفاع و تهاجم سایبری، تشکیل ارتش سایبری، تدوین استانداردها و سامانه‌های هوشمند دفاعی و...را مدنظر قرار داده‌اند (ترابی،۱۳۹۷:۲۶). از انواع روش‌های جنگ‌های سایبری، حمله به منابع اینترنتی و وبگاه‌های هر کشور با هدف سرقت اطلاعات و یا اخلال در امور و از کاراندازی سایت‌ها و خدمات است که متاسفانه در سال‌های گذشته سایت‌های مربوط به وزارت خانه‌ها، نشریات و... جمهوری اسلامی ایران نیز از این تهدیدات مصون نماندند. بنابراین ضرورت دارد تا با هوشمندسازی موضوع امنیت و دفاع در مقابله با این دسته از تهدیدات، هوشمندانه عمل نمود. این موضوع یک فرآیند پویا و مستمر است و می‌بایستی همواره در جستجوی روش‌های بهتر بود. در صورتی که رویکرد جستجوی روش‌های نوین پیگیری نشود، وقوع حملات با الگوریتم‌های پیشرفته، قدرت دفاع، پیشگیری و مقابله را به حداقل رسانده و امنیت ملی کشور را از طریق مختل نمودن سامانه‌های اقتصادی و بانکی، نظامی و امنیتی، خدمات دولت الکترونیک و... دچار آسیب خواهد کرد.

سؤال اصلی تحقیق:

تهدیدهای امنیتی وبگاه‌ها و روش‌های هوشمند مبتنی بر داده‌کاوی برای شناسایی و دسته‌بندی حملات کدامند؟

سؤال‌های فرعی تحقیق:

۱= تهدیدات امنیتی وبگاه‌ها مبتنی بر سه‌شاخگی ذهن کدام است؟

۲- الگوریتم‌های داده‌کاوی پرکاربرد برای شناسایی و دسته‌بندی حملات کدامند؟

۳- نتایج تطبیقی الگوریتم پیشنهادی با سایر روش‌ها چیست؟

مفاهیم و مبانی نظری پژوهش:

✓ فضای سایبر:

در عصر اطلاعات شاهد شکل‌گیری فضایی هستیم که در آن فعالیت‌های گوناگونی از قبیل اطلاع‌رسانی، داده‌ورزی، ارائه خدمات، مدیریت و کنترل و ارتباطات از طریق سازوکارهای الکترونیکی و مجازی انجام می‌پذیرد. از این فضا با نام «فضای تبادل اطلاعات» (فتا) و یا فضای سایبر یاد می‌شود. (سند راهبردی امنیت فضای تبادل اطلاعات (افتا))

✓ جنگ سایبری:

عبارت است از «اقدامات آفندی غیر متحرکی که به منظور کسب برتری اطلاعاتی از طریق تحت تأثیر قرار دادن سامانه اطلاعاتی و شبکه‌های رایانه‌ای دشمن اتخاذ می‌گردند». بر اساس این تعریف به نظر می‌رسد که جنگ سایبری زیرشاخه‌ای از جنگ اطلاعاتی بوده و شامل اقداماتی است که در فضای سایبری در تقابل با فضا و دنیای واقعی صورت می‌پذیرد. بستر جنگ سایبری عبارت است از هر سامانه واقعیت مجازی که دربرگیرنده مجموعه‌ای از رایانه‌ها و شبکه‌ها است. یکی از شاخص‌ترین محیط‌های جنگ سایبری اینترنت و شبکه‌های (مرتبط یا غیرنظامی) مرتبط است که به نحوی اطلاعات را به اشتراک می‌گذارند. با در نظر گرفتن جنبه نظامی جنگ سایبری می‌توان آن را در جنگی در حوزه اینترنت قلمداد نمود. (توکل،۱۳۸۵)

✓ حملات فضای سایبری وبگاه‌ها:

حمله در فضای سایبری تلاشی خطرناک است تا یک منبع قابل دسترسی از طریق شبکه به گونه‌ای مورد تغییر یا استفاده قرار گیرد که مورد نظر نبوده است. حملات شبکه را به ۳ دسته الف) دسترسی غیرمجاز به منابع اطلاعات ب) دست‌کاری غیرمجاز اطلاعات ج) اختلال در ارائه خدمات می‌توان تقسیم نمود (واحدی،۹۱:۱۳۹۱). حملات به وبگاه‌ها دسته‌ای از حملات مهم فضای سایبری می‌باشند. از مهم‌ترین حملات وبگاه‌ها به جعل درخواست[1]، حملات منع خدمت ساده و توزیع‌شده و... می‌توان اشاره کرد.

✓ مدل سه‌شاخگی ذهن:

در مدل سه‌شاخگی پدیده سازمان و مدیریت برحسب سه دسته عوامل ساختاری، رفتاری و زمینه‌ای بررسی و تجزیه‌وتحلیل می‌شود. علت نام‌گذاری این مدل به سه‌شاخگی آن است که ارتباط بین عوامل ساختاری رفتاری و زمینه‌ای به گونه‌ای است که هیچ پدیده یا رویداد سازمانی نمی‌تواند خارج از تعامل این سه‌شاخه صورت پذیرد. در نتیجه برای حل مشکلات و بحران‌های سازمان مدل سه‌شاخگی می‌تواند جزء بهترین گزینه‌ها باشد. تهدیدات سایبری یک جامعه یا سازمان را بر مبنای این مدل می‌توان ترسیم نمود.

1. Cross Site Request Forgery

✓ راهبرد سایبری ایالات متحده:

راهبرد سایبری جدید ایالات متحده در سال ۲۰۱۸ تدوین و ابلاغ شده است. هدف از تدوین این راهبرد جدید آن است که «آمریکایی‌ها و متحدانش هر روز در فضای سایبری تحت حمله قرار دارند.» بنابراین با اقدامات تهاجمی، آمریکا می‌بایستی «ساختارهای بازدارندگی» در فضای سایبری ایجاد کند. ایالات متحده آمریکا در راهبرد اعلامی در خصوص ایران نیز اعلام کرده است که حملات سایبری علیه آمریکا، اسرائیل و دیگر متحدان و شرکای آمریکا در منطقه مثل تولید موشک‌های بالستیک از اقدامات پرخطر و نامشروع ایران است. با تلفیق و ترکیب مفاد دو راهبرد اعلامی می‌توان به این نتیجه رسید که منابع سایبری کشور در رأس اقدامات تهاجمی سایبری ایالات متحده قرار دارد.

✓ هوش امنیتی و داده‌کاوی:

جمع‌آوری، بهنجارسازی، همبسته‌سازی و تحلیل بزرگ داده‌های امنیتی سایبری مثل رخدادهای یک وب به منظور خودکار نمودن فرآیندهای کاهش مخاطرات را هوش امنیتی می‌گویند. هوش امنیتی مبتنی بر داده‌کاوی است. از اهم سامانه‌های هوش امنیتی می‌توان به سامانه‌های تجمیع و داده‌کاوی رخدادها با عنوان مدیریت رخدادها و اطلاعات امنیتی[1] اشاره کرد. (نادری:۱۳۹۶)

✓ داده‌کاوی[2] و شناسایی حملات وب:

داده‌کاوی به مفهوم استخراج اطلاعات نهان یا الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ است. با توجه به حجم بالای رخدادهای یک وب سرور، از فن‌های داده‌کاوی برای شناسایی و کشف الگوها به‌صورت خودکار استفاده می‌گردد. از الگوریتم‌های داده‌کاوی به‌صورت کلی می‌توان به خوشه‌بندی، کلاسه‌بندی، کشف قانون وابستگی و... اشاره کرد (شیری،۱۳۹۳)

✓ شناسایی حملات با استفاده از الگوریتم‌های کلاسه‌بندی[3] در داده‌کاوی:

در سال ۲۰۱٤ گروهی از محققین دانشگاه USC در دانشگاه سانتیاگو، این سؤال را مطرح کردند که آیا برای حل مسائل دسته‌بندی داده‌ها در دنیای واقعی به صدها الگوریتم دسته بندی کننده نیاز داریم؟ در پاسخ به این سؤال، الگوریتم‌های مختلف دسته‌بندی روی بیش از صد مجموعه داده استاندارد پیاده‌سازی شد و درنهایت محققین به این نتیجه رسیدند که چند تا از الگوریتم‌های دسته‌بندی یعنی جنگل تصادفی[4] و ماشین‌های بردار پشتیبان[5] و شبکه‌های عصبی مصنوعی[6] بهترین عملکرد را دارند و به‌جز در موارد خاص، می‌توانیم در اکثر مسائل دنیای واقعی از همین چند الگوریتم اصلی و مناسب استفاده کنیم و به‌نوعی این الگوریتم‌ها (برابر شکل ۱) را می‌توان الگوریتم‌های برتر نامید. (*Manuel Fern´andez-Delgado,2014*)

1. *security information and event management (SIEM)*
2. *Data mining*
3. *Classification*
4. *Parallel Random Forest(parRf-t)*
5. *Support Vector Machines(SVM)*
6. *Artificial-Neural-Network(ANN)*

شکل ۱: الگوریتم‌های برتر کلاسه‌بندی

پیشینه پژوهش:

* هینگستون[1] و دیگران در سال ۲۰۱۷، تجزیه و تحلیل ترافیک حملات انجام شده با استفاده از دو فن یادگیری ماشین، یعنی درخت تصمیم[2] و ماشین‌های بردار[3] را پیاده‌سازی کردند.
* صابو لاکشمی[4] در سال ۲۰۱۷، یک مکانیزم تشخیصی و دفاعی یکپارچه برای ایجاد و شناسایی حملات منع سرویس با استفاده از الگوریتم‌های یادگیری ماشین‌ها مانند شبکه عصبی عقب[5]، نقشه خودسازمان‌ده و دستگاه بردار پشتیبانی و شناسایی آدرس آی پی واقعی از منبع حمله تفسیر شده با استفاده از مکانیزم دفاعی مبتنی بر آنتروپی ارائه داد.
* عباس وثوقی در سال ۱۳۹۷ در مقاله‌ای تحت عنوان مکانیزم های دفاعی در برابر حملات توزیع‌شده انکار سرویس، به انواع حملات انکار سرویس در تمامی لایه‌های شبکه پرداخته و ساختار و رفتار حملات را در هر لایه تشریح و مطالعه تطبیقی انجام داده است. (وثوقی، ۱۳۹۷)
* فرشته قهرمانی و بهمن آراسته در سال ۱۳۹۷، مقاله‌ای تحت عنوان تشخیص حملات منع سرویس با استفاده از الگوریتم‌های *K-means* و *Auto-encoder* ، عنوان نموده که پیشگیری قطعی از رخداد حملات انکار سرویس توزیع‌شده ممکن نیست ولی می‌توان با ترکیب روش‌های مختلف، قدرت شناسایی را بهبود بخشید. در این مقاله روشی برای تشخیص حملات انکار سرویس توزیع‌شده با ترکیب الگوریتم‌های *Auto-encoder* و *K-means* ارائه شده است و روش پیشنهادی بر روی نرم‌افزار وکا[6] طراحی و پیاده‌سازی شده است.
* هاتمر[7] و دیگران در سال ۲۰۱۸ به بررسی امنیت فناوری اطلاعات و ارتباطات به‌ویژه حملات منع سرویس در لایه کاربردی پرداخته است. در این مقاله توصیفی از الگوریتم اصلی طراحی شده برای تشخیص به موقع حملات کاربردی منع سرویس و سپس نتایج آزمایش تجربی فرآیند طراحی شده است. این الگوریتم بر روی تشخیص *HTTP GET* حمله سیل متمرکز شده است که موجب خرابی سرور مورد حمله می‌شود. شناسایی مناسب حمله از تجزیه و تحلیل ترافیک ورودی قادر به جلوگیری

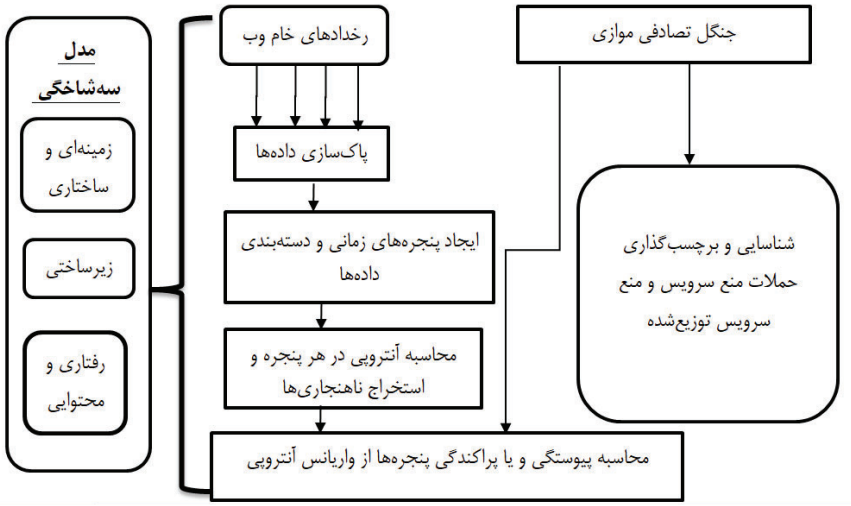
1. *Hingston*
2. *Decision Tree*
3. *Vector Machines*
4. *Subbulakshmi*
5. *Back-Propagation Neural Networks*
6. *Weka*
7. *Hottmar*

از تصادف، از سرور اتفاق می‌افتد. برای شناسایی چنین حمله‌ای، یک الگوریتم اصلی طراحی شده توسط گروه ما استفاده شد.

- انیمش[1] و دیگران در سال ۲۰۱۸ مقاله‌ای با عنوان تشخیص حمله منع سرویس توزیع‌شده با استفاده از رویکرد یادگیری ماشین ارائه کرده‌اند که در این مقاله پس از بررسی آماری از حملات منع سرویس در جهان و برآورد تقریبی از هزینه‌های این حملات، یک روش آموزش ماشین برای پیش‌بینی از این حملات را بیان می‌کند. برای این کار از یک الگوریتم جدید انتخاب رتبه‌بندی وزن استفاده می‌کند که بهتر از سایر روش‌های پایه از نظر دقت تشخیص و سربار برای مدل ساخت مدل است.

روش‌شناسی تحقیق:

پس از دسته‌بندی تهدیدات بر مبنای سه‌شاخگی ذهن با استفاده از نظریات خبرگان و گزارش‌های شرکت‌های امنیتی، یکی از تهدیدات با استفاده از روش‌های داده‌کاوی مورد بررسی قرار خواهد گرفت. برای این کار ترافیک یک وبگاه از طریق وب‌سرور با استفاده از روش ترکیبی آنتروپی و جنگل تصادفی موازی که یکی از الگوریتم‌های پرکاربرد داده‌کاوی است تحلیل و حملات سایبری وبگاه‌ها شناسایی و مقابله می‌شود.



شکل ۲: چارچوب پیشنهادی

✓ دسته‌بندی تهدیدات بر مبنای مدل سه‌شاخگی ذهن:

در این بخش تهدیدهای فضای سایبری در حوزه حملات وب بر اساس مدل سه‌شاخه بررسی می‌گردد.

1. *Animesh*

از آن جایی که وبگاه‌ها بخشی از فضای سایبری نیروهای مسلح و کشور است، دشمنان برای نفوذ به بخش‌ها و لایه‌های مختلف آن از قابلیت‌ها، ابزارها و روش‌های پیچیده و متفاوت نفوذ استفاده حداکثری را می‌برند. بنابراین در ادامه بر مبنای مدل سه‌شاخه‌ای ذهن، اهم تأثیرات و پیامدها و راهکارهای مقابله‌ای بر اساس منابع آشکار علمی و خبری و نظر خبرگان حوزه‌ها ارائه می‌گردد. از آن جا که در این تحقیق تهدیدات گروه زیرساختی و به‌طور ویژه حملات اخلال وب و روش‌های شناسایی آن از طریق یک الگوریتم بهینه در داده‌کاوی مد نظر است در ادامه این مهم مورد توجه قرار می‌گیرد

جدول ۱– دسته‌بندی تهدیدات بر مبنای مدل سه‌شاخگی ذهن

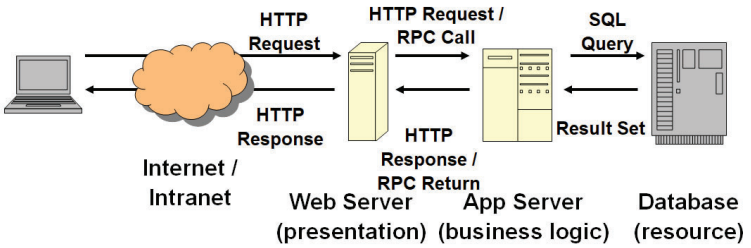
نفوذ زمینه‌ای و ساختاری	۱ - راه‌اندازی سایت‌های فارسی زبان توسط معاندین ۲ - تولید خبر و شایعه توسط معاندین علیه نظام در سایت‌های وابسته ۳ - فعال‌سازی شبکهٔ حامیان غرب در ایران از طریق ظرفیت‌های سایبری ۵ - جمع‌آوری اطلاعات از طریق نفوذ به زیرساخت‌ها ۶ - فعال شدن شرکت‌ها و صنایع سخت‌افزاری و نرم‌افزاری خارجی مرتبط با حوزه میزبانی و طراحی وبگاه‌ها و...
نفوذ زیرساختی	۱- انجام حملات اختصاصی وب مثل SQL Injection، DDOS،LI،Fi و... ۲- ارسال بدافزارها با هدف تخریب و یا جمع‌آوری اطلاعات از وبگاه‌ها و...
نفوذ رفتاری	۱-فریب متخصصان سست‌عنصر و ناآگاه به همکاری با دشمنان که طراحی و مدیریت وبگاه‌ها را عهده‌دار هستند. ۲-پیشنهاد و ترغیب متخصصان به بهره‌برداری از نرم‌افزارهای غیربومی بدافزار ۳- مهندسی اجتماعی بر روی کاربران و طراحان سایت‌ها و...

جدول ۱–دسته‌بندی تهدیدات بر مبنای مدل سه‌شاخگی ذهن

✓ رخدادهای خام وب و پاک‌سازی داده‌ها:

برای شناسایی حملات می‌بایستی ترافیک یک وبگاه که در یک وب‌سرور تحت عنوان لاگ یا رخدادهای وب قرار دارد پردازش و آنالیز شود (برابر شکل ۳). رخدادهای وب سرور در وب سرور آپاچی در فایل ورود به سیستم دسترسی ذخیره می‌شود و برای انجام پیش‌پردازش اطلاعات فایل موردنظر را در قالب

فایل‌های CSV خوانده می‌شوند و در مرحله بعد با تفکیک فیلدهای اطلاعاتی، پیش‌پردازش اطلاعات انجام می‌شود. حجم رخدادهای ثبت‌شده با توجه به کارکرد سایت متغیر است.



شکل ۳: ترافیک و رخدادهای وبگاه

آنتروپی:

آنتروپی اطلاعات که به نام آنتروپی شانون هم شناخته می‌شود (متأثر از نام *Claude E. Shannon* ریاضی‌دان آمریکایی) حاکی از تمایل سامانه‌ها به کهولت و بی‌نظمی است. سامانه‌های بسته به مرور زمان از هم گسیخته می‌شوند، زیرا انرژی یا داده جدید از محیط دریافت نمی‌کنند ولی سامانه‌های باز، آنتروپی منفی دارند یعنی می‌توانند خود را ترمیم کرده با حفظ ساختار خود زنده بمانند و حتی با وارد کردن انرژی اضافی یعنی ورود انرژی بیش از صدور آن رشد می‌کند. آنجا که حملات منع سرویس حاکی از ایجاد یک سربار اضافی با هدف ایجاد اختلال و متوقف نمودن فعالیت‌های یک وب است، به نوعی سیستم هدف؛ دچار درهم‌ریختگی، بی‌نظمی و کهولت می‌گردد که میزان این به‌هم‌ریختگی و آشفتگی را می‌توان با محاسبه آنتروپی هر *IP* محاسبه نمود. برای محاسبه آنتروپی *IP* از فرمول زیر استفاده می‌کنیم. نتیجه آنتروپی بیانگر رفتار طبیعی (نرمال)، غیرمعمول و بحرانی یک *IP* در یک وب سرور است. (*Subbulakshmi,2017*)

$$p(x_i) = s_i / \sum_{i=1}^{n} s_i$$

$$\text{Entropy} = -\sum_{x} p(x_i)\log p(x_i) = E[\log\{\frac{1}{p(x_i)}\}]$$

فرمول (۱)–محاسبه آنتروپی

واریانس:

یکی از شاخص‌های پراکندگی است که برای به دست آوردن آن باید ابتدا میانگین داده‌ها را به دست آوریم، سپس هر کدام از داده‌ها را از میانگین کم نموده سپس حاصل به دست آمده را به توان ۲ می‌رسانیم

و در نهایت همه را با هم جمع نموده تقسیم بر تعداد داده‌ها می‌کنیم (Johnson Singh,2016).

$$\sigma 2 = \frac{\sum(Xi - \overline{X}}{N}$$

فرمول (۲) –محاسبه واریانس

ایجاد پنجره‌های زمانی:

پس از پیش‌پردازش اولیه، تجمیع داده‌ها صورت می‌پذیرد. در این مرحله می‌بایستی با توجه به حجم رخدادها و نوع فیلدهای موجود، خلاصه‌سازی و کاهش ابعاد و ایجاد فیلدهای جدید در یک محدوده زمانی خاص (در اکثر مقالات پنجره‌های زمانی به‌صورت ۲۰ ثانیه‌ای تشکیل‌شده‌اند) صورت پذیرد. (Johnson Singh,2016)

$$\Delta t = (< X1 \text{ و } S1 >< X2 \text{ و } S2 > \cdots)$$

فرمول (۳) –ایجاد پنجره زمانی

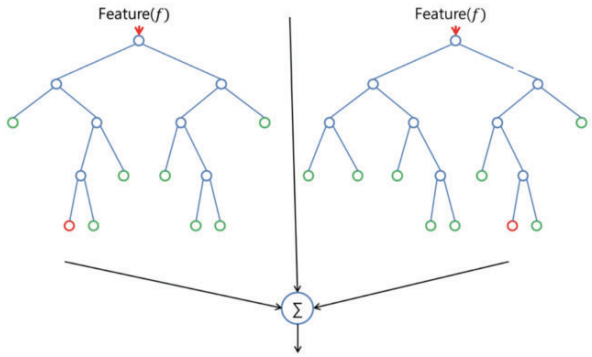
$X1$: اشاره به یک آی پی در پنجره زمانی دارد

$S1$: اشاره به تعداد رخدادهای وب از آی پی X در محدوده زمانی Δt دارد.

جنگل تصادفی موازی:

جنگل تصادفی یک الگوریتم یادگیری نظارت شده محسوب می‌شود. همان‌طور که از نام آن مشهود است، این الگوریتم جنگلی را به طور تصادفی می‌سازد. «جنگل» ساخته شده، در واقع گروهی از «درخت‌های تصمیم[1] است. یکی از مزایای جنگل تصادفی قابل استفاده بودن آن، هم برای مسائل دسته‌بندی و هم رگرسیون است که غالب سامانه‌های یادگیری ماشین کنونی را تشکیل می‌دهند. در تصویر ۴، می‌توان دو جنگل تصادفی ساخته شده از دو درخت را مشاهده کرد.

1. Decision Trees



شکل ۴: جنگل تصادفی موازی

پیاده‌سازی و نتیجه تحقیق:

مدل پیشنهادی با داده‌های مشخص و استاندارد در محیط مطلب پیاده شده است که در ادامه خروجی هر مرحله مشاهده و تشریح خواهد شد.

داده‌های تحقیق:

برای پیاده‌سازی مدل پیشنهادی، در محیط آزمایشگاهی، ترافیک مصنوعی برای مختل نمودن خدمات یک وبگاه ایجاد و پس از جمع‌آوری ترافیک شبکه، روش پیشنهادی مورد آزمایش قرار گرفت. برای اطمینان از کیفیت خروجی مدل، مدل پیشنهادی بر روی مجموعه داده استاندارد ($EPA-HTTP$) نیز آزمایش و نتایج آن با سایر روش‌ها مقایسه گردید که عملکرد بهتری داشت.

در جداول ۲ الی ۶ نتیجه این مرحله ایجاد پنجره‌های زمانی نمایش داده شده است

جدول ۲– پنجره زمانی ۱

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۱	۲۵۲,۱۷۲,۱,۲۰۲	۱۲۶,۷۱ ۶٤,۲۲۲	۱۲۱۳
۲	۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱۲۶,۷۱ ۶٤,۲۲۲	۱۲٤۲
۳	۲۰۱,۱۶۶,۵۸,۵۱	۱۲۶,۷۱ ۶٤,۲۲۲	۲۲۱
٤	۱۹۰,۲۷,۹۵,۱۹۲	۱۲۶,۷۱ ۶٤,۲۲۲	۱۸۵۶
۵	۲۵۵,۲۹۹,۱۷۳,۵۱	۱۲۶,۷۱ ۶٤,۲۲۲	۷۳۹
۶	۱۷۲,۸۹,۷۵,٤۰	۱۲۶,۷۱ ۶٤,۲۲۲	۱۱۸۲

جدول ۳- پنجره زمانی ۲

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۱	۲۵۲,۱۷۲,۱,۲۰۲	۱۲۶,۷۱ .۶٤,۲۲۲	۶۱۹
۲	۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۶۳۷
۳	۲۰۱,۱۶۶,۵۸,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۷۸
٤	۱۹۰,۲۷,۹۵,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۰۲۵
۵	۲۵۵,۲۹۹,۱۷۳,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۳۲۲
۶	۱۷۲,۸۹,۷۵,٤۰	۱۲۶,۷۱ .۶٤,۲۲۲	۶۰۰

جدول ۴- پنجره زمانی ۳

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۱	۲۵۲,۱۷۲,۱,۲۰۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۲۹
۲	۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۷۸
۳	۲۰۱,۱۶۶,۵۸,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۳۰۱
٤	۱۹۰,۲۷,۹۵,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۸۰۳
۵	۲۵۵,۲۹۹,۱۷۳,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۲۱۷
۶	۱۷۲,۸۹,۷۵,٤۰	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۰۱

جدول ۵- پنجره زمانی ۴

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۱	۲۵۲,۱۷۲,۱,۲۰۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۵۳
۲	۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۷۳
۳	۲۰۱,۱۶۶,۵۸,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۳۳٤
٤	۱۹۰,۲۷,۹۵,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۷۳٤
۵	۲۵۵,۲۹۹,۱۷۳,۵۱	۱۲۶,۷۱ .۶٤,۲۲۲	۷۷
۶	۱۷۲,۸۹,۷۵,٤۰	۱۲۶,۷۱ .۶٤,۲۲۲	۱۰۹۹

جدول ۶- پنجره زمانی ۵

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۱	۲۵۲,۱۷۲,۱,۲۰۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۳۶
۲	۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱۲۶,۷۱ .۶٤,۲۲۲	۱۲۷۷

ردیف	آدرس منبع	آدرس مقصد	تعداد تراکنش
۳	۲۰۱,۱۶۶,۵۸,۵۱	۱۲۶,۷۱ ۶۴,۲۲۲	۳۳۷
٤	۱۹۰,۲۷,۹۵,۱۹۲	۱۲۶,۷۱ ۶۴,۲۲۲	۱۷۱۳
۵	۲۵۵,۲۹۹,۱۷۳,۵۱	۱۲۶,۷۱ ۶۴,۲۲۲	۱۱۶
۶	۱۷۲,۸۹,۷۵,٤۰	۱۲۶,۷۱ ۶۴,۲۲۲	۱۱٤۸

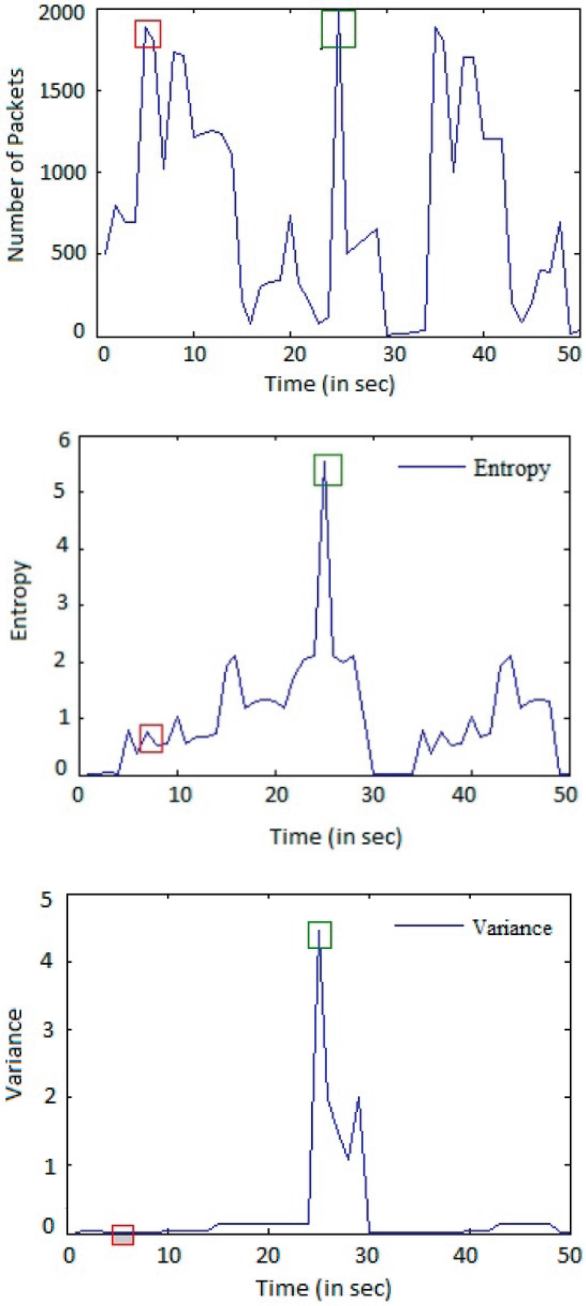
در جدول زمانی بالا، پس از این که تعداد تراکنش متقابل هر آی پی با هم در یک زمان خاص مقایسه شده است. از آن جا که حملات مختل‌سازی و منع سرویس و خدمات، ایجاد سربار اضافی و غیرمتعارف در سایت است، محاسبه تعداد تراکنش‌ها برای به دست آوردن ناهنجاری‌ها و حملات الزامی است.

✓ محاسبه آنتروپی و واریانس آنتروپی در هر پنجره و استخراج ناهنجاری‌ها:

نتایج آنتروپی هر آی پی در پنجره زمانی و محاسبه میانگین و واریانس آنتروپی‌ها در هر پنجره با استفاده از رابطه ۱ برابر جدول شماره ۷ و نمودارهای مربوطه برابر شکل ۳ است.

آدرس منبع	آنتروپی								
	اول بار	دومین بار	سومین بار	چهارمین بار	پنجمین بار	متوسط	واریانس	تقریبی
۲۵۲,۱۷۲,۱,۲۰۲	۱,۰۱۸	۱,۰۲۲	۰,۵٤٤	۰,۶۶۹	۰,۶۷۸	۰,۷۸۶	۰,۰۳۸۶۸۷	۰,۰۳۹
۲۲۷,۱٤۸,۱۲۰,۱۹۲	۱,۰۰۶	۱,۰۱٤	۰,۵۲۱	۰,۶۵۷	۰,۶۶۳	۰,۷۷۲	۰,۰٤۰۲۸۵	۰,۰٤
۲۰۱,۱۶۶,۵۸,۵۱	۱,۹۱۸	۲,۲۱۱	۱,۱۹۲	۱,۲۹	۱,۳۱۹	۱,۵۸۶	۰,۱۹۲۹۹۸	۰,۱۶۳
۱۹۰,۲۷,۹۵,۱۹۲	۰,۷۹۹	۰,۷۵۱	۰,۳۸۷	۰,۵۲۷	۰,۵٤۸	۰,۶۰۲	۰,۰۱۳۹۱	۰,۰۱٤
۲۵۵,۲۹۹,۱۷۳,۵۱	۱,۳۰۲	۱,۱۷۹	۱,۷۳۹	۲,۰۵۲	۲,۱۰۶	۱,۶۷۶	۰,۱٤۳٤۲۶	۰,۱٤۳
۱۷۲,۸۹,۷۵,٤۰	۱,۰۳۲	۱,۰۳۸	۰,۵۸۵	۰,۷۳۸	۰,۷٤۵	۰,۸۲۸	۰,۰۳۱۹۵۱	۰,۰۳۲

جدول ۷– ایجاد پنجره‌های زمانی، محاسبه آنتروپی و واریانس آنتروپی

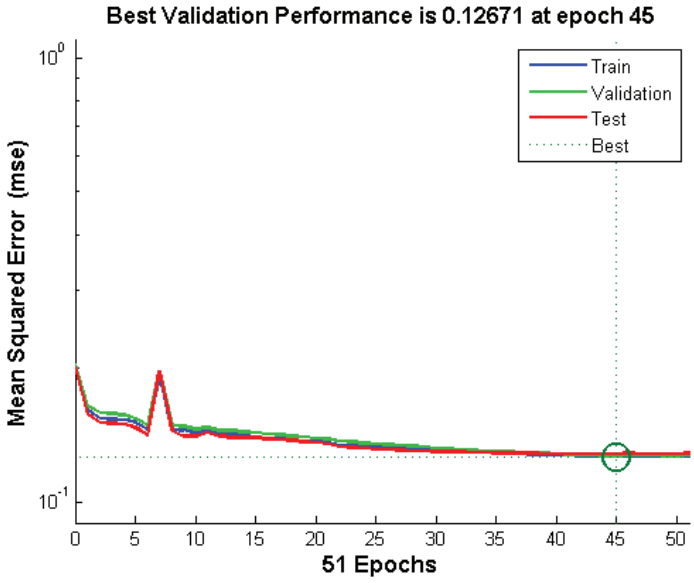


شکل ۶- خروجی آنتروپی، واریانس آنتروپی و حملات

با محاسبه آنتروپی، کهولت و بی‌نظمی هر آی پی به دست می‌آید. از آنجاکه حملات منع سرویس و یا خدمت از مصادیق بی‌نظمی می‌باشند، با استفاده از این روش، حملات شناسایی گردیده‌اند. البته این روش بر روی همه داده‌ها اجرا نشده و فقط در فضای نمونه اجرا گردیده است و در ادامه می‌بایستی با استفاده از روش‌های داده‌کاوی، مدل حملات به دست آمده برای شناسایی و طبقه‌بندی وضعیت هنجار و یا ناهنجار سایر آی‌پی‌ها که مشکوک به اخلال در خدمات وبگاه می‌باشند گسترش یابد.

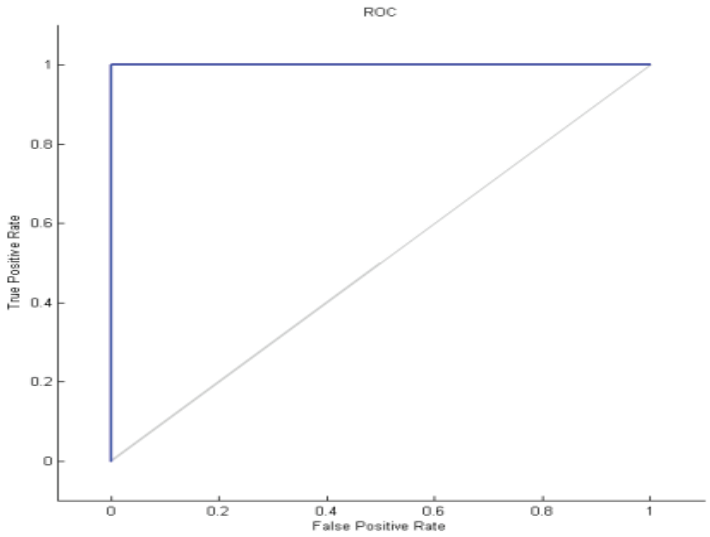
✓ استفاده از جنگل تصادفی موازی برای کلاسه‌بندی و برچسب‌گذاری حملات:

پس از محاسبه آنتروپی و واریانس آنتروپی و تعیین رفتارهای نرمال و ناهنجار که منجر به بروز منع خدمت شده است، از یک الگوریتم دسته‌بندی برای یادگیری و بسط مدل استفاده می‌کنیم. در این مرحله از الگوریتم جنگل تصادفی موازی (در قسمت قبل به‌طور کامل تشریح گردید) استفاده گردید که نتایج آن به شرح زیر است.



شکل ۷– آموزش داده‌های ورودی

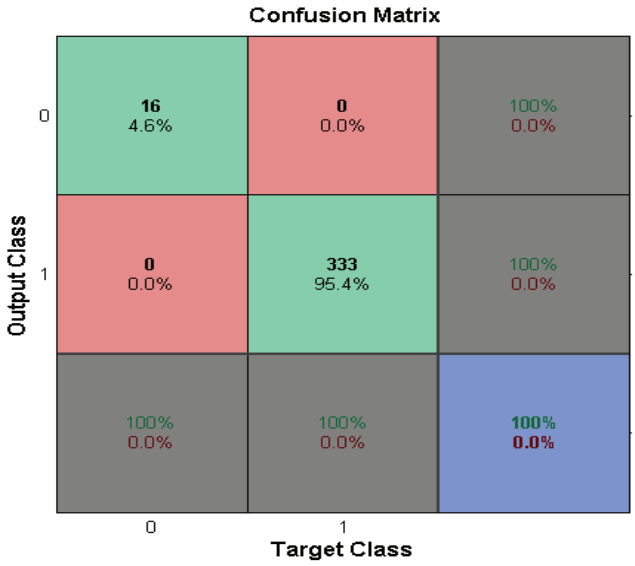
در قدم اول نیاز است نمودار تأیید اعتبار برای آموزش داده‌های ورودی را بررسی کنیم. این نمودار در شکل ۷ ارائه شده است. تحلیل این نمودار بیانگر این موضوع است که بهترین عملکرد اعتبارسنجی، در زمان مقرر، ۰/۰۱۲۶۷۱ است. در شکل ۷، روند اعتبارسنجی، روند آموزش و خطای تست (به ترتیب با خطوط سبز، آبی و قرمز) با افزایش دوره زمان یا حجم آموزش کاهش می‌یابد. در مرحله بعدی، نمودار مشخصه عملکرد سامانه بررسی و سپس ماتریس درهم‌ریختگی مثل بخش قبل ترسیم و محاسبات لازم انجام می‌شود.



شکل ۸-منحنی مشخصه عملکرد سامانه

نمودار مشخصه عملکرد سامانه روشی برای بررسی کارایی دسته‌بندهاست. درواقع منحنی‌های مشخصه عملکرد سامانه منحنی‌های دوبعدی هستند که در آن‌ها DR یا همان نرخ تشخیص صحیح دسته مثبت (TPR[1]) روی محور Y و به‌طور مشابه FAR یا همان نرخ تشخیص غلط دسته منفی (FPR[2]) روی محور X رسم می‌شوند. به‌بیان‌دیگر یک منحنی مشخصه عملکرد سامانه مصالحه نسبی میان سودها و هزینه‌ها را نشان می‌دهد. منحنی مشخصه عملکرد سامانه اجازه مقایسه تصویری مجموعه‌ای از دسته‌بندی‌کننده‌ها را می‌دهد، همچنین نقاط متعددی در فضای مشخصه عملکرد سامانه قابل توجه است. نقطه پایین سمت چپ (۰ ,۰) استراتژی را نشان می‌دهد که در یک دسته‌بند مثبت تولید نمی‌شود. راهبرد (استراتژی) مخالف که بدون شرط، دسته‌بندهای مثبت تولید می‌کند، با نقطه بالا سمت راست (۱ ,۱) مشخص می‌شود. نقطه (۰ ,۱) دسته‌بندی کامل و بی‌عیب را نمایش می‌دهد. به‌طورکلی یک نقطه در فضای مشخصه عملکرد سامانه بهتر از دیگری است اگر در شمال غربی‌تر این فضا قرار گرفته باشد. همچنین در نظر داشته باشید منحنی‌های مشخصه عملکرد سامانه رفتار یک دسته‌بندی‌کننده را بدون توجه به توزیع دسته‌ها یا هزینه خطا نشان می‌دهند، بنابراین کارایی دسته‌بندی را از این عوامل جدا می‌کنند. فقط زمانی که یک دسته‌بند در کل فضای کارایی به‌وضوح بر دسته دیگری تسلط یابد، می‌توان گفت که بهتر از دیگری است. در ادامه ماتریس نتایج ماتریس درهم‌ریختگی[3] با بررسی و معیارهای بررسی محاسبه می‌گردد.

1. *True Positive Rate*
2. *FalsePositive Rate*
3. *Confusion Matrix*



شکل ۹– ماتریس به‌هم‌ریختگی

ماتریس درهم‌ریختگی، به ماتریسی گفته می‌شود که در آن عملکرد الگوریتم‌های مربوطه را نشان می‌دهند. معمولاً چنین نمایشی برای الگوریتم‌های یادگیری با ناظر استفاده می‌شود، اگرچه در یادگیری بدون ناظر نیز کاربرد دارد. هر ستون از ماتریس، نمونه‌ای از مقدار پیش‌بینی‌شده را نشان می‌دهد. درصورتی‌که هر سطر نمونه‌ای واقعی (درست) را در بردارد. در این مرحله ما از ماتریس درهم‌ریختگی برای صحت عملکرد الگوریتم پیاده‌سازی شده تا دقت و صحت دسته‌بندها و کلاسه‌های ایجادشده را مشخص کنیم (*M. V.Nidhi,2016*).

✓ مقایسه روش پیاده‌سازی شده با سایر کارهای مشابه:

مدل دسته‌بندی	ماتریس به‌هم‌ریختگی	دقت	صحت	اطمینان
Random Forest	۲۶۷(*a*) ۴(*b*) ۴(*c*) ۸۲(*d*)	۷۶,۹۷	۹۸۵۲,۰	٤٦٥,۰

مدل دسته‌بندی	ماتریس به‌هم‌ریختگی	دقت	صحت	اطمینان
RBF Network	۲۶۳(a) ۷(b) ۵(c) ۸۲(d)	۶۳,۹۶	۹۸۱۳,۰	۰۷۸۷,۰
MLP	۲۶۰(a) ۱۱(b) ۶(c) ۸۰(d)	۲۳,۹۵	۹۷۷۳,۰	۱۲۰۸,۰
SVM	۳۲۳(a) ۰(b) ۴(c) ۲۲(d)	۹۸.۹	۰۱,۱	۰
ANN+RPROP	۲۶(a) ۰(b) ۱(c) ۳۲۲(d)	۹۹.۸	۰۳,۱	۰
روش پیشنهادی (PRForest)	۱۶(a) ۰(b) ۴(c) ۳۳۳(d)	۱	۱	۰

جدول ۸–مقایسه روش پیاده‌سازی شده با سایر روش‌ها

نتیجه‌گیری

همان‌طور که در مقدمه و بیان مسئله بیان شد، انقلاب و جنگ‌های سایبری تأثیرات گسترده‌ای بر ابعاد و سطوح امنیت گذاشته است، به شکلی که می‌توان از انقلاب سایبری در عرصه امنیت صحبت کرد. در این راستا، انقلاب و جنگ‌های سایبری معنا و محتوای امنیت در حوزه‌های نظامی، سیاسی، اقتصادی، فرهنگی و اجتماعی در سطوح فردی، اجتماعی و دولتی را چنان دچار دگرگونی و دگردیسی عمیق نموده

که نادیده گرفتن و کم‌اهمیت جلوه دادن آن توسط کشورها، پیامدهای فاجعه‌بار گسترده‌ای به دنبال خواهد داشت. بدون تردید توسل به چنین رویکردهایی می‌تواند خود منشأ درک متفاوت و عمیقی از شرایط امنیتی دنیای کنونی و در نتیجه اتخاذ راهبردها، تاکتیک‌ها و اقدامات جدیدی گردد. مطالعه‌های صورت پذیرفته بر روی راهبردها و اقدامات صورت پذیرفته از سوی کشورهای پیشرو، بیانگر این موضوع است که راهی جز استفاده از سامانه‌های هوشمند دفاعی و پیشگیرانه که مبتنی بر داده‌کاوی هستند وجود ندارد و البته این که سامانه‌های هوشمند همواره بایستی با بهینه‌سازی الگوریتم‌های مورد استفاده و ارائه الگوریتم‌های جدید، کارایی و اثربخشی لازم را داشته باشند. با این رویکرد در این تحقیق بر روی شناسایی تهدیدات سایت‌های اینترنتی از طریق مدل سه‌شاخه‌ای ذهن پرداخته شد و با توجه به ابعاد سه‌گانه نفوذ زمینه‌ای و ساختاری، زیرساختی و رفتاری؛ به‌صورت ویژه تحقیق بر روی بعد زیرساختی که شامل حملات به وبگاه‌ها است تمرکز شده و حملات منع خدمت و یا سرویس از طریق ارائه یک الگوریتم بهینه داده‌کاوی با جنگل تصادفی موازی یا رندوم فارست مورد ارزیابی قرار گرفت که نتیجه بهتری نسبت به الگوریتم‌های قبلی ارائه شده در مطالعات قبلی داشت.

برای تحقیقات آتی پیشنهادهای زیر ارائه می‌گردد

۱- موضوع هوشمندسازی مقولهٔ امنیت با تمرکز بر الگوریتم‌های پیچیده و ترکیبی داده‌کاوی و فراابتکاری مورد توجه جدی باشد.

۲- در خصوص سامانه‌های امنیتی هوشمند مثل SOC,*WAF* و... همواره مطالعات تطبیقی صورت پذیرفته تا سامانه مناسب با سازمان تهیه گردد.

۳- راهکارهای راه‌اندازی و استفاده از زیرساخت‌های داخلی برای مدیریت و میزبانی وبگاه‌های مهم مورد توجه و مطالعه باشد.

۴- روش‌های نوین تست نفوذ به سایت‌های اینترنتی به‌صورت مستمر مطالعه و رفتارهای ناهنجار و مشکوک با آنالیز ترافیک شبکه خنثی گردد

۵- سایر حملات وب که در متن تحقیق عنوان شده مورد بررسی قرار گیرد.

۶- در خصوص تهدیدات زمینه‌ای و محتوایی که در مدل سه‌شاخه‌ای ذهن مطرح گردید نیز پیشنهادهای زیر ارائه می‌گردد

منابع

- **جزوات یادگیری ماشین** دکتر شیری (۱۳۹۱)-دانشگاه تهران
- توکل، اکبر، ۱۳۸۵، **مفهوم جنگ سایبری و کاربرد آن در جنگ آینده**، نشریه مطالعات دفاعی
- ترابی، قاسم،۱۳۹۷، **فضای سایبر از چشم‌انداز امنیتی**، پژوهشکده مطالعات کاربردی قلم
- قهرمانی، فرشته و بهمن آراسته، (۱۳۹۷)، **تشخیص حملات DDOS با استفاده از الگوریتم‌های K-means و Auto-encoder**، چهارمین کنفرانس ملی محاسبات توزیعی و پردازش داده تبریز: دانشگاه شهید مدنی آذربایجان، https://www.civilica.com/Paper-DCBDP04-DCBDP04_091.html
- کریمخانی زندی، حمید، (۱۳۹۷)، **امنیت و حریم خصوصی در شبکه‌های اجتماعی**، کنگره ملی سالانه‌ای ، برق و کامپیوتر، ساری: موسسه آموزش عالیه ساری، https://www.civilica.com/Paper-SETCO01-SETCO01_030.html
- قادریان، (۱۳۸۷)، پایان‌نامه کارشناسی ارشد با موضوع: **بهبود مدل کاربر در وبگاه به‌صورت خودکار با استفاده از معناشناسی با مفاهیم خاص دامنه**، تهران: دانشگاه امیرکبیر.
- وثوقی، عباس، (۱۳۹۷)، **مکانیزم‌های دفاعی در برابر حملات توزیع‌شده انکار** سرویس، دومین همایش بین‌المللی مهندسی برق، علوم کامپیوتر و فناوری اطلاعات، همدان: دبیرخانه دائمی کنفرانس، https://www.civilica.com/Paper-ECICONFE02-ECICONFE02_035.html

- *Adi, E, Baig, Z. &Hingston, P. (2017).* ***Stealthy Denial of Service (DoS) attack modelling and detection for HTTP/2 services****. Journal of Network and Computer Applications, 91, 1-13.*
- *Ďurčeková.V, Schwartz.L, Hottmar.V, Adamec.B. (2018).* ***Detection of Attacks Causing Network Service Denial.****Advances in Military Technology.2018, Vol. 13 Issue 1, p87-94. 8p.*
- *Gupta, A. (2018).* ***Distributed Denial of Service Attack Detection Using a Machine Learning Approah (Unpublished master's thesis).*** *University of Calgary, Calgary, AB.*
- *http://hdl.handle.net/1880/107615*
- *Johnson Singh, K. Thongam, K. & De, T. (2016).* ***Entropy-Based Application Layer DDoS Attack Detection Using Artificial Neural Networks****. Entropy, 18(10), 350.*
- *N.Chambers, B.Fry, J.McMasters, (2018).* ***Detecting Denial-of-Service Attacks from Social Media Text:Applying NLP to Computer Security,*** *Proceedings of*

NAACL-HLT 2018, pages 1626–1635,New Orleans, Louisiana, June 1 - 6, 2018. c 2018 Association for Computational Linguistics

- OWASP Top 10 2017-***The Ten Most Critical Web Application Security Risks***
- Subbulakshmi, T. (2017). ***A learning-based hybrid framework for detection and defence of DDoS attacks.*** International Journal of Internet Protocol Technology, 10(1), 51-60.
- Zargar, S.T. Joshi, J. and Tipper, D. 2013. ***A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks***. IEEE communications surveys & tutorials, 15(4), pp.2046-2069.
- Subbulakshmi, T. (2017). ***A learning-based hybrid framework for detection and defence of DDoS attacks.*** International Journal of Internet Protocol Technology, 10(1), 51-60.
- [8] Subbulakshmi, T. (2017). ***A learning-based hybrid framework for detection and defence of DDoS attacks.*** International Journal of Internet Protocol Technology, 10(1), 51-60