Journal of Systems Smartening and Data Processing
Vol. 2 - No. 2 - Summer 2024

Prevention and detection of denial of service attacks using random forest and decision tree algorithms

Seyed Ebrahim Dashti[1], Hadis farajpour[2]

[1]Department of Electrical and Computer Engineering, Jahrom Branch, Islamic Azad University, Jahrom, Iran
sayed.dashty@gmail.Com

[2] Department of Electrical and Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
h.farajpour1995@gmail.com

Received: 03 August 2024	**Revised:** 07 December 2024	**Accepted:** 08 December 2024

Abstract:

Cloud computing has ushered in a new era of innovation with incredible scalability and accessibility. Consumers who subscribe to a cloud-based service have unlimited access to applications and technology. In addition to reducing prices, this concept has increased the reliability and availability of supplies. Programs that are based on the cloud are available on demand from anywhere in the world at low cost, this unlimited access has many opportunities and challenges. Although cloud computing faces security concerns, it has thrived due to its incredibly instant services. Among the important challenges is the increase in the occurrence of Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, which has increased the demand for effective defense mechanisms. Detection of such anomalies in the computer network is usually done through network-based intrusion detection and prevention systems (NIDPS). Although NIDPSs allow for the detection of all known attacks, they are not immune to persistent changes in anomalies over time. Machine learning algorithm provides algorithms that, due to continuous evolution, can effectively reduce cyber threats by detecting new anomalies. Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks are one of the most important security threats in computer networks. These attacks can significantly disrupt access to services and lead to loss of revenue and reputation. Detection and prevention of These attacks are of particular importance. The proposed random forest and decision tree algorithms are among the efficient techniques in the field of machine learning that can be used to detect and prevent denial of service attacks, the results confirm the effectiveness of the proposed method.

Keywords: Distributed Denial of Service, Intrusion Prevention, Machine Learning, Cloud Computing, Random Forest, Tree Decision.

Corresponding Author: Dr. SeyedEbrahim Dashti
Corresponding Author Address: Department of Electrical and Computer Engineering- Mellat Blvd. -Islamic Azad University- Kazerun Branch, Iran

پیشگیری و تشخیص حملات انکار سرویس با استفاده از الگوریتم جنگل تصادفی و درخت تصمیم

سید ابراهیم دشتی[1]، استادیار، حدیث فرج پور[2]، کارشناسی ارشد

۱- دانشکده برق و کامپیوتر، واحد جهرم، دانشگاه آزاد اسلامی، جهرم، ایران
Sayed.Dashty@gmail.Com

۲- دانشکده برق و کامپیوتر ، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران
h.farajpour1995@gmail.com

تاریخ ارسال مقاله: ۱۴۰۳/۰۵/۱۳ **تاریخ بازنگری مقاله:** ۱۴۰۳/۰۹/۱۷ **تاریخ پذیرش مقاله:** ۱۴۰۳/۰۹/۱۸

چکیده: : رایانش ابری با مقیاس پذیری و دسترسی باورنکردنی عصر جدیدی از نوآوری را ایجاد کرده است. مصرف کنندگانی که مشترک سرویسی مبتنی بر ابر هستند، دسترسی نامحدودی به برنامه ها و فناوری دارند. این مفهوم علاوه بر کاهش قیمت ها، قابلیت اطمینان و دسترسی به عرضه ها را افزایش داده است. برنامه هایی که مبتنی بر ابر هستند، در صورت تقاضا از هر کجای دنیا با هزینه کم در دسترس هستند، این دسترسیهای نامحدود فرصتها و چالشهای زیادی را در بر داشته است. اگرچه محاسبات ابری با نگرانی های ایمنی مواجه است، اما به دلیل خدمات آنی فوق العاده اش پیشرفت کرده است. از جمله چالشهای مهم، افزایش وقوع حملات انکار سرویس (DoS)و انکار توزیع شده سرویس (DDoS)است که تقاضا برای مکانیسم های دفاعی مؤثر افزایش یافته است. تشخیص چنین ناهنجاری هایی در شبکه کامپیوتری معمولاً از طریق سیستم های تشخیص نفوذ و پیشگیری مبتنی بر شبکه(NIDPS)انجام می شود. اگرچهNIDPS ها امکان رهگیری همه حملات شناخته شده را می دهند، اما در برابر تغییرات مداوم ناهنجاریها در طول زمان مقاوم نیستند. الگوریتم یادگیری ماشین الگوریتم هایی را ارائه می کند که بدلیل تکامل پیوسته می توانند با تشخیص ناهنجاریهای جدید به طور موثری تهدیدات سایبری را کاهش دهند. حملات انکار سرویس (DoS) و انکار سرویس توزیع‌شده (DDoS) از مهم‌ترین تهدیدات امنیتی در شبکه‌های کامپیوتری هستند. این حملات می‌توانند به طور قابل توجهی دسترسی به سرویس‌ها را مختل کرده و منجر به از دست رفتن درآمد و اعتبار شوند. تشخیص و پیشگیری از این حملات از اهمیت ویژه‌ای برخوردار است. الگوریتم‌های پیشنهادی جنگل تصادفی و درخت تصمیم از جمله تکنیک‌های کارآمد در حوزه یادگیری ماشین هستند که می‌توانند برای تشخیص و پیشگیری از حملات انکار سرویس مورد استفاده قرار گیرند، نتایج کارآمدی روش پیشنهادی را تایید می کند.

کلمات کلیدی: انکارسرویس توزیع شده، سیستم پیشگیری از نفوذ، یادگیری ماشین، محاسبات ابری، جنگل تصادفی، درخت تصمیم.

نام نویسنده‌ی مسئول: دکتر سیدابراهیم دشتی
نشانی نویسنده‌ی مسئول: دانشکده برق و کامپیوتر، واحد جهرم، دانشگاه آزاد اسلامی

۱- مقدمه

حمله DDoS یک نوع توزیع شده از حالت حمله است که در آن یک مهاجم تعداد زیادی از ماشین های حمله را کنترل می کند و دستورالعمل های حمله DoS را برای دستگاه ارسال می کند. در آخرین گزارش امنیت اینترنت، حملات DDoS همچنان یکی از تهدیدهای اصلی امنیت سایبری است. حملات مبتنی بر(Denial-of-Service) DoS به طور مداوم در حال پیشرفت هستند وپیچیدگی و دامنه آن را افزایش می دهند و در نتیجه انجام تشخیص به موقع و دقیق را دشوارترمی کنند. حملات DoS سنتی، که هدف آنها ناتوان کردن یک منبع از ارائه خدمات به مشتریان واقعی است، به طور فشرده در ادبیات از طریق طرح های مختلف برای محافظت از زیرساخت های شبکه مورد مطالعه قرار گرفته است[۱].

امروزه تضمین امنیت سیستم های فناوری اطلاعات و ارتباطات (Information and Communications Technology) ICTبه عنوان یک دغدغه اصلی برای دفاع از شرکت ها و مرزهای ملی مطرح شده است. شناسایی، بررسی،تجزیه و تحلیل و پاسخ به تهدیدات و حملات سایبری برای اطمینان از امنیت سایبری موثر ضروری است[۲]. در فضای مجازی، عوامل مخرب از چندین روش برای حمله به اهداف خود استفاده می کنند. با توجه به بررسی جامع پیشنهاد شده است[۳]. برخی از حملات رایج عبارتند از: ۱-فیشینگ،۲-بدافزار،۳-انکار سرویس(/ (DoS ۴- انکار سرویس توزیع شده (DDoS).

استفاده از رایانش ابری به سرعت در حال افزایش است. این که آیا به طورکامل یا تا حد زیادی دولت ها و شرکت ها زیرساخت های فناوری اطلاعات خود را به ابر منتقل کرده اند. زیرساخت مبتنی بر ابر در مقایسه با زیرساخت های مرسوم سنتی ودر محل، مزایای مختلفی را ارائه می کند. حذف هزینه های مربوط به بهره برداری و نقص،و همچنین دسترسی به مواد در صورت درخواست، تنها تعدادی از مزایا هستند. بااین حال، نگرانی های زیادی وجود دارد که مصرف کنندگان ابری دارند، و تحقیق به این مسائل می پردازد. اکثر این پرس و جوها بر حفاظت از مفاهیم و اطلاعات عملیاتی متمرکزهستند. بسیاری از حملات مربوط به امنیت را می‌توان به صورت متعارف پیشگیری کرد سیستم های فناوری اطلاعات که از محاسبات ابری استفاده نمی کنند. جرایم متمرکز مبتنی بر ابر در حال حاضر از نوآوری های خود استفاده می کنند. بسیاری از آسیب پذیری های امنیتی در رایانش ابری در مقایسه با پیشینیان خود در محیط های محاسبات غیر ابری منحصربه فرد هستند، زیرا داده ها و منطق تجاری در یک سرور ابری خارجی ذخیره می شوند که نظارت قابل دسترسی ندارد. حمله انکار خدمات (DoS) یکی ازتکنیک هایی است که اخیراً مورد توجه قرار گرفته است. حوادث انکار سرویس به جای افرادی که از آن پشتیبانی می کند، متوجه سرور می شوند. مهاجمان DoS سعی می‌کنند بامخفی کردن کاربران واقعی به سرورهای زنده سرازیر شوند تا ظرفیت سرویس برای رسیدگی به درخواست های دریافتی را بیش از حد بارگذاری کنند. رایانش ابری یک سرویس مبتنی بر اینترنت است که کاربران را قادر می سازد تا بااشتراک گذاری منابع محاسباتی قابل تنظیم (شامل سرور، ذخیره سازی، نرم افزار کاربردی،سرویس ها، شبکه ها و غیره) دسترسی پیدا کنند تا به منابع محاسباتی در صورت تقاضا دسترسی آنلاین داشته باشند.رایانش ابری به عنوان ترکیبی از فناوری های نوظهور و مدل های کسب وکار، در سال های اخیر به دلیل مزایای آن در مقیاس فوق العاده، مجازی سازی، قابلیت اطمینان بالا، مقیاس پذیری خوب و خدمات براساس تقاضا، به سرعت توسعه یافته است. برای غلبه بر این مشکل، چندین پرسش به طور همزمان به سرور ارسال می‌شود.

"انکار سرویس" که از رایانه های متعددی برای حمله و آسیب رساندن به یک سرویس در یک زمان استفاده می کند. در میان خطرات مهم و احتمالاً فاجعه بار، در میان بسیاری دیگر،تعداد فزاینده حملات انکار سرویس توزیع شده مشاهده شده است. یک چهارم یا بیشتراز سازمان های جهان حمله انکار سرویس توزیع شده را تجربه کرده اند. حملات انکار سرویس توزیع شده به Rackspace،ارائه دهنده خدمات رایانش ابری، ضربه سختی زد. یکی دیگر از حملات انکار سرویس توزیع شده گسترده علیه سرورهای ابری آمازون EC2 راه اندازی شد که به عنوان نمونه ای از یک حمله عمل می‌کند. فعالیت های شرکت به شدت مختل شد، پول از دست رفت و اثرات فوری و بلند مدت بر مشاغل مورد حمله داشت. در سال‌های اخیر، حملات DDoS بیشتر شده است و بات نت مورد استفاده مهاجمان بزرگ تر شده است و میزان استفاده از ترافیک شبکه به ارتفاع G۱۰۰۰ رسیده است.

۲-کارهای مرتبط

راه حل های مبتنی بر هوش مصنوعی (AI) مانند یادگیری ماشینی (ML) و الگوریتم های یادگیری عمیق (DL) را می توان برای مقابله با مشکلات تشخیص تهدیدات سایبری به کار برد، زیرا از نظر نرخ تشخیص نتیجه امیدوارکننده هستند[۴]. با توجه به نظرسنجی های ارائه شده در [۵،۶] الگوریتم های ML می توانند نقش مهمی در این زمینه ایفا کنند و دانشمندان را قادر می سازند تا به مشکلات امنیتی سایبری خاصی مانند وظایف تشخیص DDoS/DoS رسیدگی کنند. به طور خاص، رویکردهای ML برای یادگیری الگوهای حمله DDoS/DoS برای شناسایی این حملات قبل از اختلال در خدمات شبکه هدف استفاده می شوند[۷].

حملات انکار سرویس توزیع شده (DDos) به ویژه برای اینترنت و تجارت (SaaSنرم افزار به عنوان سرویس) در چند فصل گذشته آسیب رسان بوده است. بیش از ۷۵ درصد از اقدامات متقابل شناخته شده در برابر حملات DDoS از خدمات ارائه شده توسط ابر استفاده می کنند[۸]. خسارات مالی به یکی از بدترین نتایج ممکن حمله انکار سرویس توزیع شده در فضای ابری اشاره دارد. بر اساس برخی برآوردها، میانگین قیمت یک حمله انکار سرویس توزیع شده ۴۸۲۰۰۰ دلار است. برخی از زیان های مالی متحمل شده در سه ماهه اول ۲۰۱۵ در افشای جدید Neustar به تفصیل شرح داده شده است. مطالعات نشان می دهد که به طور متوسط بیش از ۷۲ هزار دلار دریک ساعت به سرقت می رود. حملات انکار سرویس توزیع شده (DDos) اهمیت جدیدی در رایانش ابری پیدا می کند. این تغییر مستقیماً از مشکلات عملیاتی ناشی از حمله به شبکه قربانی ناشی می شود[۹].

در محیط محاسبات ابری، فناوری حمله DoS نیزدستخوش تغییرات جدیدی شده است و به اشکال مختلف ظاهر می شود. حمله ممکن است از خارج از خوشه سرور یا از داخل خوشه سرور انجام شود. در حال حاضر، روش حمله محبوب تر این است که مهاجم به یک پلتفرم محاسبات ابری یا خوشه سرور خاص حمله کند. این روش حمله صدمات فراوان و روش های مختلف را به همراه دارد و انجام سریع موقعیت یابی و عیب یابی عیب کار دشواری است. هدف از توضیح یک مدل ML ارائه قوانین تصمیم منطقی استنباط شده توسط خود مدل به شکلی است که برای انسان قابل درک باشد. این ویژگی به طور کلی به عنوان تفسیرپذیری یک مدل ML تعریف می شود[۱۰]. به طور کلی، هرچه قابلیت تفسیر مدل بالاتر باشد، قابلیت اطمینان سیستم پیش بینی بهتر است. بنابراین، قابلیت همکاری بالا به این معنی است که بتوانیم درک کنیم که یک مدل ML چگونه پیش بینی می کند، یعنی اینکه چگونه هر ویژگی داده به پیش بینی کمک می کند.

یک درخت تصمیم (DT) برای تفسیر قوانین تصمیم استخراج شده برای نوشتن قوانین طبقه بندی نفوذ شبکه استفاده می شود. با این حال، با وجود ایده آل بودن مدل های DT از نظر توانایی آنها در ارائه توضیح قابل درک از پیش بینی، همان طور که برای هر مدل ML ،تصمیمات استنباط شده توسط یادگیرنده با به روزرسانی فراپارامترهای آن تغییر می کند. بنابراین، با توجه به دو مدل متمایز DT که به نرخ تشخیص برابری دست می یابند، تنها مدلی که منجربه تعداد کمتر قوانین استخراج شده می شود، ترجیح داده می شود، زیرا منجر به تولید یک مجموعه قوانین بهینه به معنای حداقل تعداد قوانین تولید شده می شود. با این حال، با وجود ایده آل بودن مدل های DT از نظر توانایی آنها در ارائه توضیح قابل درک از پیش بینی، همان طور که برای هر مدل ML ، تصمیمات استنباط شده توسط یادگیرنده با به روزرسانی فراپارامترهای آن تغییر می کند. بنابراین، با توجه به دو مدل متمایز DT که به نرخ تشخیص برابری دست می یابند، تنها مدلی که منجربه تعداد کمتر قوانین استخراج شده می شود، ترجیح داده می شود، زیرا منجر به تولید یک مجموعه قوانین بهینه به معنای حداقل تعداد قوانین تولید شده می شود.

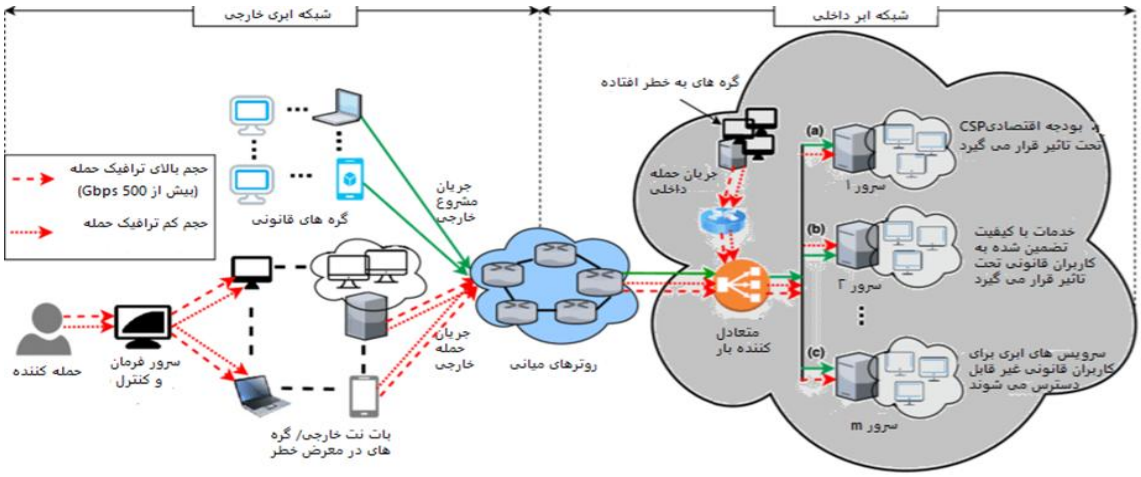
۳- حمله DDoS و ویژگی های ابری

حمله انکار سرویس (Deinal of Service) DoSیک حمله مخرب به سرور هدف از طریق روش های غیرعادی است که منجر به ناتوانی آن در ارائه خدمات به کاربران عادی شبکه می شود. در حال حاضر، حملات انکار سرویس توزیع شده (DDoS)به دستاوردهای زیادی در رایانش ابری دست یافته‌اند، جایی که هکرها از مدل «پرداخت درزمان خرید» استفاده می کنند. روش

ها و اشکال زیادی برای حمله DoS وجود دارد که در موارد زیر خلاصه می شود: اشغال و مصرف غیرقانونی منابع کامپیوتری مانند CPU ،پهنای باند شبکه و فضای ذخیره سازی. تغییر یا حتی از بین بردن اطلاعات پیکربندی سرور مورد نظرتغییر یا حتی تخریب تجهیزات گره کلیدی در شبکه فیزیکی؛ و دسترسی به خدمات با برنامه نویسی. ابر به رایانه های عظیمی نیازدارد که بتوانند به چندین کاربر در یک محیط استاندارد خدمات رسانی کنند. هدف یک مهاجم ممکن است همیشه محدود به «انکار خدمات» نباشد، اما می تواند شامل کاهش سودآوری مشترکین ابری باشد. نحوه جلوگیری از حملاتی مانند این از زمان شروع رایانش ابری موضوع داغ بوده است.

حمله انکار سرویس توزیع شده(DDoS)که به عنوان حمله انکار اقتصادی پایداری (EDoS) یا حمله مصرف منابع متقلبانه (FRC)نیز شناخته می شود.حملات عبارت FRC (Fraudulent Resource Consumption) در بسیاری از آثار دیگر برای توصیف این نوع حمله استفاده شده است. حملات پراکنده انکار سرویس که صفحات وب را هدف قرار می دهد و هکرها ربات ها و تروجانها را در سیستم های در معرض خطر در سراسر اینترنت نصب می کنند. اگر سرویس هدف در فضای ابری میزبانی شود، یک حمله DDoS به عنوان یک حمله EDoSاجرامی شود. "بوترها" مشاغلی هستند که مشتریان خود را با یک بات نت به یکدیگر متصل می کنند تا حملات انکار سرویس توزیع شده DDoS را در صفحات وب رقبای خود راه اندازی کنند. حملاتی از این دست می توانند توسط همه چیز از رقابت تجاری گرفته تا رقابت سیاسی و باج خواهی و جنگ سایبری کامل بین کشورها تحریک شوند[۱۱].

با توجه به این اصل کار که DDoS منابع سیستم را مصرف می کند و باعث می شود سیستم نتواند خدمات عادی را ارائه دهد، مدیران شبکه می توانند سیستم را برای بهبود تحمل سیستم در برابر حملات DDoS بهینه سازی و تقویت کنند و حتی برخی از بسته های حمله DDoS را مسدود کنند. رویکرد رایانش ابری چندین فرصت و مزیت را برای مشتریان فراهم می کند.با این وجود، مهاجمان DDoS نیز به این ویژگی ها دسترسی دارند و ممکن است آنها را مفید بدانند. به منظور انجام "انکار سرویس"، مهاجمی که یک حمله DDoS را راه اندازی می کند، سیل درخواست های جعلی را ارسال می کند. اگرچه فناوری حمله DDoS متنوع است، اما شباهتهای زیادی به پدیده های ناشی از سیستم دارد. بنابراین، از طریق پیاده سازی یک سیستم تشخیص توزیع شده، تلاش خواهیم کرد تا رفتار حملات DDoS را در اولین بارپیدا کنیم و منبع و ویژگی های حملات را دقیقاً تعیین کنیم. از طریق سیستم تجزیه و تحلیل ترافیک غیرعادی شبکه و ابزار تشخیص DDoS،رفتار DDoSدرشبکه را به موقع پیدا کنید، مشکلات را به موقع پیدا کنید و توانایی تشخیص و تجزیه وتحلیل کلی سیستم را بهبود بخشید. به همین دلیل،یک رویکرد مبتنی بر یادگیری ماشینی برای شناسایی حملات انکار سرویس توزیع شده ((DDoSدر ابرارائه کردیم.Random Forest و Decision Tree دو تکنیک مختلف دسته بندی یادگیری ماشین هستند که به سیستم کمک می کنند حمله انکارسرویس توزیع شده را شناسایی کند.



شکل(۱): معماری ابری تحت تاثیر حملات DDoS

بیشتر حملات انکار سرویس (DDoS)توزیع شده در فضای ابری،حملات Flood SYN یا Crowd Flash هستند. تجزیه و تحلیل نشان داد که حملات بانرخ پایین انکار سرویس TCP و کاهش عملکرد دو مورد از رایج ترین دسته های حملات هستند [۱۲]. حمله Flood SYN یکی از انواع حملات انکار سرویس(DoS)است که در آن مهاجم تلاش می کند با ارسال تعداد زیادی بسته های SYN به سرور هدف،منابع آن را اشغال کرده و آن را از کار بیندازد.این حمله از فرآیند دست دادن سه مرحله ای (three-way handshake)در پروتکل TCP استفاده می کند.این فرآیند شامل مراحل زیر است:

کلاینت به سرور بسته SYN می فرستد:این بسته نشان دهنده در خواست برقراری اتصال است.

سرور به کلاینت بسته SYN-ACK می فرستد:سرور درخواست را می پذیرد و منتظر تایید نهایی از کلاینت می ماند.

کلاینت به سرور بسته ACK می فرستد:این مرحله نهایی است که اتصال TCP را برقرار می کند.

در حمله Flood SYN ،مهاجم تعداد زیادی بسته های SYN به سرور ارسال می کند اما مرحله سوم ارسال (ACK)را انجام نمی دهد. این باعث می شود که سرور منتظر پاسخ های نهایی بماند و منابع زیادی را به این اتصالات نیمه باز اختصاص دهد، در نتیجه سرور قادر به پاسخگویی به در خواست های واقعی کاربران نخواهد بود و از کار می افتد.

از سوی دیگر، حملات DDoS با هدف قطع دسترسی و قدرت پردازش خدمات محاسبات ابری انجام می شود. این تأثیر منفی بر عملکرد و دسترسی به منابع رایانش ابری دارد. در حال حاضر هیچ روش قابل اعتمادی برای شناسایی یا فیلتر کردن حملات DDoS وجود ندارد، بنابراین آنها ابزاری قابل اعتماد برای افرادی هستند که به دنبال انجام حملات سایبری هستند. اخیراً، دانشمندان آزمایش تکنیک های یادگیری ماشین(ML)را برای توسعه تاکتیک های مؤثر مبتنی بر ML برای شناسایی حملات انکار سرویس توزیع شده (DDoS)آغاز کرده اند.

۴- سیستم تشخیص و پیشگیری از نفوذ شبکه

راه حل های سیستم تشخیص نفوذ و پیشگیری از نفوذ شبکه (NIDPS) معمولاً به عنوان یک خط دفاعی استفاده می شود که می تواند به طور مؤثر بین ترافیک قانونی و غیرعادی تمایز قائل شود. معماری کلی NIDPS ارائه شده در[۱۳] متشکل از دنباله ای از ماژول ها است. به طور خاص، اولین ماژول به جمع آوری بسته های شبکه واگذارمی شود. این نوع عملکرد بسته به حالت کار NIDPS متفاوت است. در واقع، صرف نظر از اینکه حالت پیشگیری فعال باشد، دریافت بسته در حالت درون خطی انجام می شود زیرا پردازش بلادرنگ باید تضمین شود. از طرف دیگر، اگر سیستم به عنوان یک آشکارساز عمل کند، حالت انعکاس پورت (غیرفعال) فعال می شود. این اجازه می دهد تا به طور موقت بسته های در انتظار پردازش بافر شوند. هنگامی که یک بسته ضبط می شود، یک ماژول رمزگشا ترافیک را تجزیه و تحلیل می کند تا از مطابقت با استانداردهای ارتباطی شبکه اطمینان حاصل کند. در غیر این صورت، بسته دور انداخته می شود و یک هشدار ایجاد می شود. پس از آن، یک ماژول پیش پردازنده یکپارچه سازی، مونتاژ مجدد، و انطباق جلسه را در پروتکل های سطح پایین انجام می دهد، در حالی که پلاگین های سطح بالا بسته ها را بر اساس چندین پروتکل کاربردی تأیید می کنند. علاوه بر ماژول های قبلی، ماژول تشخیص، بسته های شبکه تحلیل شده را با الگوهای تعریف شده در پایگاه دانش، یعنی مجموعه قوانین درگیر، مقایسه می کند. این جزء نقش کلیدی ایفا می کند زیرا بخش اصلی فرآیند تصمیم گیری است. بنابراین، هر چه نرخ تشخیص تضمین شده توسط مجموعه قوانین درگیر بالاتر باشد، عملکرد کلی فرآیند بهترخواهد بود[۱۴].

۴-۱- سیستم های تشخیص نفوذ (IDS)

منابع، داده ها و برنامه های ذخیره شده دراینترنت عمومی به دلیل اتصال سیستم در معرض حمله قرار دارند. سیستم های تشخیص نفوذ (IDS)در فضای ابری برای نظارت بر رفتار مخرب در شبکه و سیستم میزبان استفاده می شوند. از آنجایی که اطلاعات غیرقانونی زیادی به صورت آنلاین ایجاد میکند، شناسایی یک حمله انکار سرویس توزیع شده (DDos)برای سیستم های تشخیص نفوذ IDSچالش برانگیز است. تجزیه و تحلیل امنیت سایبری می تواند از طریق استفاده از روش های داده کاوی به شناسایی نفوذها کمک کند. بسیاری از رویکردهای متمایز با روش های یادگیری ماشینی به عنوان پایه و اساس آنها توسعه

داده شده است[۱۵]. سیستم های سنتی تشخیص نفوذ (IDS)از یک رویکرد مبتنی بر امضا برای شناسایی DoS پیروی کنید، ترافیک شبکه جاری را در زمان واقعی نظارت کنید و به دنبال دنباله ها یاالگوهایی باشید که ممکن است با امضای حمله که قبلاً معرفی شده مطابقت داشته باشد. امضاها را می توان بر اساس هدر بسته ها و آدرس های شبکه، که حاوی توالی هایی از داده ها هستند که به عنوان یک حمله آشنا شناخته می شوند، شناسایی کرد.

IDS مبتنی بر امضا یک راه حل پرکاربرد است ،زیرا مبتنی بر نرم افزار منبع بازاست و به طور موثر در شناسایی حملات رایج کار می کند. اگرچه این راه حل با حملاتی که قبلاً شناخته شده است نتایج عالی دارد (الگوهای موجود در پایگاه داده امضا)،نمی تواند هر حمله سایبری را با موفقیت شناسایی کند، علاوه بر این، تشخیص الگوهایHTPP از طریق ترافیک رمزگذاری شده بسیار پیچیده است. علاوه بر این، با استفاده از IDS سنتی، حملات ناشناخته بدون توجه به زیرساخت شبکه می گذرد و تنها زمانی که آنها کشف شده و امضای آن به پایگاه داده منتقل می شود، قابل شناسایی است. بنابراین، در طول این فرآیند، دستگاه های شبکه در معرض قرار می گیرند و عواقب آن ممکن است طاقت فرسا باشد.

سیستم های تشخیص نفوذ شبکه مبتنی بر ناهنجاری(NIDS)ترافیک شبکه ورودی را به منظورشناسایی هر گونه رفتار تهاجمی یا نامشروع تجزیه و تحلیل می کنند. این رویکرد نه تنها برای شناسایی حملات شناخته شده، بلکه همچنین برای شناسایی مناسب عمل میکند. سیستم در زمان واقعی ترافیک شبکه، پردازش و جمع آوری بسته ها را در جریان های مکالمه نظارت می کند و ویژگی های شبکه معنی دار اضافی را به عنوان آمار استخراج می کند که به صورت پویا برای تشخیص ناهنجاری تجزیه و تحلیل می کند.

الگوریتم های یادگیری ماشین در پیشگیری و تشخیص حملات انکار سرویس:

استفاده از الگوریتم‌های یادگیری ماشین در پیشگیری و تشخیص حملات انکار سرویس (Denial of Service) یا (DoS)و حملات توزیع‌شده انکار سرویس (Distributed Denial of Service) یا (DDoS)به دلیل توانایی این الگوریتم‌ها در تحلیل حجم زیادی از داده‌ها و شناسایی الگوهای پیچیده و ناهنجاری‌ها، به سرعت در حال رشد و توسعه است. در ادامه به برخی از الگوریتم‌های یادگیری ماشین که برای تشخیص و پیشگیری از این نوع حملات مورد استفاده قرار می‌گیرند، اشاره می‌کنیم:

۱-درخت تصمیم (Decision Tree)

به سادگی قابل فهم و تفسیر است و به خوبی با داده‌های بزرگ کار می‌کند.

می‌تواند برای طبقه‌بندی ترافیک شبکه به ترافیک عادی و ترافیک حمله استفاده شود.

۲-ماشین بردار پشتیبانی (SVM)

کارایی بالا در شناسایی حملات و ناهنجاری‌ها با داده‌های خطی و غیرخطی.

می‌تواند ترافیک شبکه را به دو دسته ترافیک نرمال و ترافیک حمله طبقه‌بندی کند.

۳-جنگل تصادفی (Random Forest)

ترکیبی از چندین درخت تصمیم برای افزایش دقت و کاهش احتمال .overfitting

برای شناسایی ویژگی‌های مهم ترافیک شبکه که ممکن است نشان‌دهنده حملات باشند.

۴-الگوریتم K-Means

این الگوریتم‌ها می‌توانند ترافیک شبکه را به خوشه‌های مختلف تقسیم کنند و ناهنجاری‌ها را شناسایی کنند.

شناسایی ترافیک‌های غیرعادی که ممکن است نشانه‌ای از حملات DDoS باشند.

۵-رگرسیون لجستیک(Logistic Regression)

رگرسیون لجستیک برای پیشگیری و تشخیص حملات انکار سرویس (DoS)با تحلیل الگوهای ترافیک شبکه و شناسایی ویژگی‌های مشکوک کاربرد دارد. این مدل می‌تواند داده‌های شبکه را برای پیش‌بینی احتمال وقوع حملات بررسی کرده و در نتیجه به ایجاد سیستم‌های امنیتی پیشگیرانه کمک کند.

۶-تحلیل مؤلفه‌های اصلی (PCA)

کاهش ابعاد داده‌ها و شناسایی ویژگی‌های مهم.

شناسایی ناهنجاری‌ها با تجزیه و تحلیل داده‌های ترافیک شبکه.

۷- شبکه‌های عصبی کانولوشن (CNN)

تحلیل داده‌های ترافیک شبکه به صورت سلسله‌مراتبی و شناسایی الگوهای پیچیده.

تشخیص حملات DDoS از طریق تحلیل داده‌های ترافیک شبکه به صورت تصویری و...

۸- درخت تصمیم (Decision Tree)

درخت تصمیم یکی از روش‌های پرکاربرد در یادگیری ماشین است که از ساختار درختی برای تصمیم‌گیری استفاده می‌کند. این الگوریتم با تقسیم داده‌ها به مجموعه‌های کوچکتر و ایجاد ساختار درختی، به پیش‌بینی و تشخیص حملات کمک می‌کند.

درخت تصمیم شامل گره‌های تصمیم‌گیری و گره‌های برگ است. هر گره تصمیم‌گیری یک شرط خاص را بررسی می‌کند و گره‌های برگ نتیجه نهایی (یعنی حمله یا عدم حمله) را نشان می‌دهند.

در هر گره، بهترین ویژگی برای تقسیم داده‌ها انتخاب می‌شود. این کار معمولاً با استفاده از معیارهایی مانند اطلاعات متقابل (information gain)یا جینی (Gini)انجام می‌شود.

داده‌ها بر اساس ویژگی انتخاب شده به دو یا چند زیرمجموعه تقسیم می‌شوند.فرآیند انتخاب ویژگی و تقسیم داده‌ها تا زمانی که به یک شرط توقف برسیم، تکرار می‌شود. این شرط می‌تواند عمق درخت یا تعداد داده‌های هر گره باشد.

۹- الگوریتم جنگل تصادفی (Random Forest)

جنگل تصادفی یک مدل ترکیبی است که از ترکیب چندین درخت تصمیم مستقل تشکیل شده است. این مدل با ایجاد تنوع در درخت‌های تصمیم از طریق نمونه‌گیری بوت استرپ (Bootstrap Sampling)و انتخاب تصادفی ویژگی‌ها، باعث افزایش دقت و پایداری مدل می‌شود. جنگل تصادفی به دلیل قدرت بالای خود در تشخیص الگوهای پیچیده، یکی از روش‌های مؤثر در پیشگیری و تشخیص حملات است.

با توجه به ویژگیهای روشهای مختلف که در بالا بررسی شده، در این مقاله تصمیم گرفته شد از روش یادگیری ماشین استفاده شود و از درخت تصمیم و جنگل تصادفی بدلیل ویژگیهای آنها برای تشخیص حملات انکار سرویس که هدف این مقاله است استفاده شده است.

۵-روش پیشنهادی

استفاده از الگوریتم‌های یادگیری ماشین برای پیشگیری و تشخیص حملات انکار سرویس توزیع‌شده می‌تواند به طور قابل توجهی به افزایش امنیت شبکه‌ها کمک کند. این الگوریتم‌ها با تحلیل داده‌های ترافیک شبکه، شناسایی الگوهای غیرعادی و ناهنجاری‌ها، و طبقه‌بندی ترافیک به ترافیک نرمال و ترافیک حمله، می‌توانند به طور موثری از بروز این نوع حملات جلوگیری کنند و پاسخ مناسبی به تهدیدات امنیتی ارائه دهند. انتخاب الگوریتم مناسب بستگی به نوع داده‌ها، پیچیدگی حملات و نیازهای خاص شبکه دارد.

۵-۱- درخت تصمیم برای تشخیص DDoS/DoS

DT یک مدل ML است که به طور گسترده در ادبیات برای مقابله با آن استفاده می شود وظیفه تشخیص[۱۶]. DDoS/DoS تجزیه و تحلیل چندین الگوریتم ML که وظیفه طبقه بندی DDoS/DoS را انجام می دهند، انجام شد و نشان داد که در بین روش های معیار، DT می تواند بالاترین امتیاز دقت طبقه بندی را به دست آورد. تحلیل مشابهی در [۱۷]، که در آن مدل DT با مدل K-NN نزدیکترین همسایه (K-NN)مقایسه شد. درنتیجه این تجزیه و تحلیل، DT از K-NN هم در دقت طبقه بندی و هم در زمان اجرا، یعنی زمان آموزش و پیش بینی، بهتر عمل می کند. که در [۱۸] نویسندگان از یک مدل DT برای ارائه یک سیستم سربار محاسباتی کم با قابلیت شناسایی حملات DDoSاستفاده کرده اند. این با انتخاب حداقل تعداد ویژگی با استفاده از فیلتر کم واریانس به دست آمد. نتایج به دست آمده نشان دهنده اثربخشی در به حداقل رساندن بار CPU موردنیاز و در عین حال حفظ یک مقدار دقت امیدوارکننده است. که در [۱۹]، چندین طبقه بندی کننده مبتنی بر درخت در وظیفه

تشخیص حمله DDoS با استفاده از مجموعه داده CICDS2017 مورد ارزیابی قرار گرفتند. با استفاده از یک استراتژی انتخاب ویژگی کارآمد، بهترین راه حل در میان مقایسه شده، الگوریتم جزئی DT بوده است، هم از نظرزمان اجرا و هم از نظر دقت طبقه بندی. که در[۲۰]، یک مدل DT که توسط به دست آوردن اطلاعات هدایت می شود برای شناسایی حملات DDoS در زمان واقعی استفاده می شود. مرحله آزمایشی اثربخشی مدل را در ساخت یک DT با قابلیت شناسایی منبع حمله در میان حجم زیادی از ترافیک توزیع شده نشان می دهد. این به دلیل استراتژی هرس درخت متوسط است که استحکام مدل را حتی در مورد نویز تضمین می کند. که در[۲۱]، یک مدل DT برای رسیدگی به وظایف تشخیص DDoSاستفاده می شود.

کاربرد DT را برای شناسایی حملات DDoS با گرفتن ترافیک از یک شبکه تعریف شده نرم افزاری (SDN) و با استفاده از دو مجموعه داده پیشرفته بررسی کردند.

آزمایش ها نشان داد که مدل DT بهتر از ماشین بردار پشتیبان ((SVM و طبقه بندی کننده های ساده بیز عمل می کند. که در [۲۲]،یک ماژول حفاظتی DDoS بر اساس نسخه اصلاح شده DT ارائه شده است. این سیستم با انتخاب ویژگی های ترافیکی قابل توجه برای انتقال به مدل DT آموزش دیده با استفاده از ناخالصی جینی، اطلاعات شبکه را جمع آوری می کند. علاوه بر این، سیستم یک استراتژی هرس خطای بدبینانه با هدف اصلی کاهش زمان اجرای مدل را در بر می گیرد. آزمایش ها در یک محیطSDN انجام شد و اثربخشی سیستم را در تشخیص حملات DDoS نشان داد. مدل ارائه شده در [۱۷] از یک الگوریتم DT برای شناسایی حملات DDoS در یک سیستم ابری مبتنی بر SDN استفاده می کند.

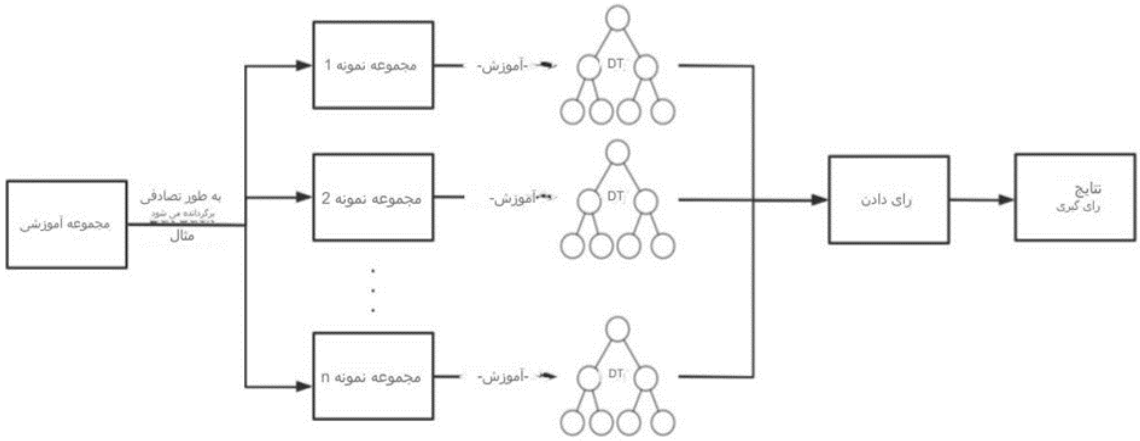
مدلهای یادگیری ماشینی (ML) را می توان به الگوریتم های پارامتری و ناپارامتریک تقسیم کرد. یک الگوریتم ML زمانی پارامتری گفته می شود که تابع آموخته شده به شکل شناخته شده ای متشکل از تعداد ثابتی از پارامترها ساده شود. از سوی دیگر، یک مدل ناپارامتریک هیچ فرضی در مورد نحوه استخراج تابع نگاشت نمی کند. یک مدل درخت تصمیم (DT) متعلق به کلاس مدل های ناپارامتریک مبتنی بر پارادایم یادگیری نظارت شده است . این نوع مدل از یک ساختار درخت مانند برای نشان دادن احتمال وقوع یک رویداد بر اساس داده های آموزشی داده شده به عنوان ورودی مدل استفاده می کند. ازگره ریشه، تصمیمات و برگ ها در یک ساختار درختی از بالا به پایین مرتب می شوند تا مدل تصمیم را نشان دهند. هر سطح از گره ها یک ویژگی در مجموعه داده را نشان می دهد،در حالی که یک شاخه یک مقدار ممکن یا مجموعه ای از مقادیر را نشان می دهد. علاوه بر گره های ریشه که پیش بینی کننده های قابل توجهی ارائه می کنند، گره های برگ نیزطبقه بندی نهایی را تعیین می کنند. . فرآیند فوق الذکر که امکان ساخت DT ها از داده های آموزشی را فراهم می کند، القاء DT نامیده می شود [۲۳].

زمانیکه حمله ای برنامه ریزی می شود، مهاجمان به دنبال حفره های امنیتی در پروتکل ها می گردند تا از یک سرویس به طور نامشروع استفاده کنند. در سرورهای HTTP،یک تهدید داخلی وجود دارد: اتصال تنها زمانی آزاد می شود که یک هدر کامل دریافت شود.این بدان معناست که مهاجمان می توانند از این آسیب پذیری برای انجام یک حمله استفاده کنند. در این سناریو یک مهاجم می تواند یک درخواست ساده ارسال کند و اتصال سوکت تا زمانی که درخواست تکمیل نشود بسته نخواهد شد. ارسال درخواست های ناقص با نسبت آهسته با تعداد زیادی مشتری، تمام اتصالات موجود یک وب سرور را اشغال می کند. اگر یک مهاجم تعداد زیادی از اتصالات همزمان را باز نگه دارد، وب سرور منتظر پایان اتصالات خواهد ماند و به یک وضعیت DoS می رسد.مهاجم درخواست های HTTP ناقصی را ارسال می کند تا سوکت را اشغال کند. اگر حملات از طریق ترافیک رمزگذاری شده انجام شود، تطبیق امضا در هدرها و پیام های HTTP دشوار خواهد بود.

۵-۲- جنگل تصادفی

پس از مقایسه های مختلف یادگیری ماشین، ما RF را به عنوان الگوریتم تشخیص حمله DDoS انتخاب کرده ایم.RF، به عنوان یک الگوریتم یادگیری یکپارچه، با ترکیب چندین مدل درخت تصمیم تولید می شود. در RF، هر درخت تصمیم به طور مستقل تولید می شود و فرآیند تولید آن بر اساس نمونه های تصادفی مختلف و زیر مجموعه های ویژگی ها است. اصل اساسی RF کاهش خطر بیش از حد برازش یک درخت تصمیم واحد با ساخت درخت های تصمیم گیری متعدد است، در نتیجه

توانایی تعمیم مدل را بهبود می بخشد.هر درخت تصمیم در RF بر روی یک زیرمجموعه نمونه برداری تصادفی از داده های اصلی آموزش داده می شود.RF طبقه بندی یا رگرسیون را با رای دادن یا میانگین گیری خروجی هر درخت تصمیم انجام می دهد.RF چندین پارامتر مهم دارد. مهمترین این پارامترها نرخ نمونه برداری، تعداد درخت تصمیم و تعداد ویژگی های هر درخت تصمیم است. با تنظیم این پارامترها، دقت و کارایی محاسباتی RFقابل کنترل است. RF طیف گسترده ای از کاربردها در بسیاری از زمینه ها، از جمله مالی،مراقبتهای بهداشتی، پردازش زبان طبیعی، بینایی کامپیوتری و ... از مزایای آن می توان به دقت بالا،توانایی رسیدگی به تعداد زیادی از ویژگی ها و نمونه ها، توانایی شناسایی ویژگی های مهم، توانایی رسیدگی به داده های از دست رفته و غیره اشاره کرد.



شکل(۲): :نمودار جریان الگوریتم جنگل تصادفی

۱.نمونه گیری تصادفی: نمونه های زیادی از مجموعه آموزشی اصلی با استفاده از روش نمونه گیری بوت استرپ(با نمونه برداری برگشتی) به عنوان مجموعه آموزشی جدید گرفته می شود.

۲.انتخاب ویژگی تصادفی: بخشی از مجموعه ویژگی های اصلی به طور تصادفی به عنوان مجموعه ویژگی های جدید انتخاب می شود و تنها این بخش از ویژگی ها در هنگام ساخت درخت تصمیم در نظر گرفته میشود.

۳.ساخت درخت تصمیم: درخت تصمیم بر اساس مجموعه آموزشی و مجموعه ویژگی های جدید ساخته می شودکه هر تقسیم گره بر اساس یک زیر مجموعه تصادفی از مجموعه ویژگی ها است.

۴.یکپارچه سازی : RFدرخت های تصمیم گیری متعددی تولید می شوند و در نهایت هر نمونه با رای دادن یا گرفتن میانگین طبقه بندی یا پسرفت می شود.

الگوریتمRF دارای مزایای زیر است:

۱.الگوریتم های RF توانایی بیشتری برای مدیریت داده های با ابعاد بالا و مجموعه داده های بزرگ مقیاس دارند.

۲.الگوریتم RF می تواند به طور موثر از مشکل بیش از حد برازش جلوگیری کند.

۳.الگوریتم RF می تواند داده های گسسته و پیوسته را مدیریت کند.

۴.الگوریتم RF می تواند رتبه بندی اهمیت متغیرها را ارائه دهد که برای انتخاب ویژگی مناسب است.

از روش جنگل تصادفی برای تشخیص حملات DDoS استفاده شده است. جنگل تصادفی یک الگوریتم یادگیری ماشینی است که با ترکیب چندین درخت تصمیم‌گیری، یک مدل قوی و دقیق برای پیش‌بینی ایجاد می‌کند. جنگل تصادفی به عنوان یک روش طبقه‌بندی برای تشخیص الگوهای حملات DDoS در ترافیک شبکه مورد استفاده قرار گرفته است.این روش به دلیل توانایی آن در مقابله با مسائل مانند بیش‌برازش (overfitting) و توانایی کار با مجموعه داده‌های بزرگ و پیچیده، انتخاب شده است. مدل جنگل تصادفی با تجزیه و تحلیل ویژگی‌های مختلف ترافیک شبکه، توانایی تشخیص و طبقه‌بندی ترافیک مخرب از ترافیک عادی را دارد.جنگل تصادفی یکی از الگوریتم‌های یادگیری ماشین محبوب است که در تشخیص و شناسایی حملات انکار سرویس (DoS) بسیار مؤثر است. در ادامه چندین مورد از کاربردهای جنگل تصادفی در این زمینه آورده شده است:

۱. تشخیص حملات انکار سرویس توزیع‌شده (DDoS) تشخیص الگوهای ترافیک مشکوک: جنگل تصادفی می‌تواند الگوهای ترافیک شبکه را تحلیل کرده و ترافیک غیرعادی که نشانه‌های حملات DDoS هستند را شناسایی کند.
تمایز بین ترافیک عادی و ترافیک حمله: این الگوریتم قادر است به دقت ترافیک عادی شبکه را از ترافیک ایجاد شده توسط حملات DDoS تفکیک کند.
۲. تشخیص حملات انکار سرویس مبتنی بر پروتکل
تحلیل بسته‌های شبکه: با استفاده از جنگل تصادفی می‌توان بسته‌های شبکه را مورد تحلیل قرار داد و بسته‌هایی که احتمالاً حامل حملات انکار سرویس مبتنی بر پروتکل هستند را شناسایی کرد.
تمایز بین پروتکل‌های مشروع و مخرب: جنگل تصادفی می‌تواند با تحلیل داده‌های بسته‌ها و پروتکل‌ها، ترافیک مشروع را از ترافیک مخرب جدا کند.
۳. تشخیص حملات انکار سرویس مبتنی بر حجم
تشخیص افزایش ناگهانی حجم ترافیک: جنگل تصادفی قادر است افزایش ناگهانی و غیرمعمول حجم ترافیک که نشانه‌ای از حملات انکار سرویس مبتنی بر حجم است را شناسایی کند.
ارزیابی همزمانی ترافیک غیرعادی: این الگوریتم می‌تواند همزمانی و همبستگی بین ترافیک‌های غیرعادی را تحلیل کرده و حملات احتمالی را شناسایی کند.
۴. تشخیص حملات انکار سرویس سطح پایین
شناسایی تغییرات کوچک در الگوی ترافیک: جنگل تصادفی می‌تواند تغییرات کوچک و جزئی در الگوی ترافیک را شناسایی کرده و حملات انکار سرویس سطح پایین را تشخیص دهد.
تحلیل زمان‌بندی بسته‌ها: این الگوریتم می‌تواند زمان‌بندی بسته‌های شبکه را تحلیل کرده و تفاوت‌های زمان‌بندی که نشانه‌ای از حملات انکار سرویس است را شناسایی کند.
۵. تشخیص حملات انکار سرویس مبتنی بر استفاده از منابع
شناسایی مصرف غیرعادی منابع: جنگل تصادفی می‌تواند مصرف غیرعادی منابع سیستمی مثل پردازنده و حافظه که نشانه‌ای از حملات انکار سرویس است را شناسایی کند.
تحلیل درخواست‌های سیستمی: این الگوریتم قادر است درخواست‌های سیستمی را تحلیل کرده و درخواست‌های غیرعادی که احتمالاً منجر به حملات انکار سرویس می‌شوند را شناسایی کند.
استفاده از جنگل تصادفی به دلیل قدرت آن در تحلیل داده‌های پیچیده و چندبعدی، دقت بالا در شناسایی الگوهای غیرعادی، و توانایی آن در کار با داده‌های بزرگ، یک ابزار مؤثر و قدرتمند در تشخیص و شناسایی حملات انکار سرویس است

۵-۳- مراحل پیاده‌سازی روش پیشنهادی

۵-۳-۱- جمع‌آوری و آماده‌سازی داده‌ها

در این مرحله، داده‌های شبکه‌ای شامل ترافیک نرمال و ترافیک مربوط به حملات DoS و DDoS جمع‌آوری می‌شود. این داده‌ها باید به شکلی آماده‌سازی شوند که برای الگوریتم‌های یادگیری ماشین قابل استفاده باشند. این شامل فرآیندهای پیش‌پردازش مانند پاکسازی داده‌ها، نرمال‌سازی و استخراج ویژگی‌ها می‌شود.
پیاده‌سازی الگوریتم‌های درخت تصمیم و جنگل تصادفی با استفاده از داده‌های آماده‌شده، الگوریتم‌های درخت تصمیم و جنگل تصادفی پیاده‌سازی می‌شوند. این شامل انتخاب پارامترهای بهینه، آموزش مدل‌ها و ارزیابی عملکرد آنها با استفاده از معیارهای مختلف مانند دقت وفراخوانی می‌شود.
این مقاله به استفاده از دو الگوریتم یادگیری ماشین یعنی جنگل تصادفی (Random Forest) و درخت تصمیم (Decision Tree) برای تشخیص حملات DDoS اشاره شده است. این الگوریتم‌ها به دلیل قابلیت‌های پیش‌بینی قوی و دقت بالای تشخیص حملات مورد استفاده قرار گرفته‌اند.ترکیب الگوریتم جنگل تصادفی و درخت تصمیم برای شناسایی و تشخیص

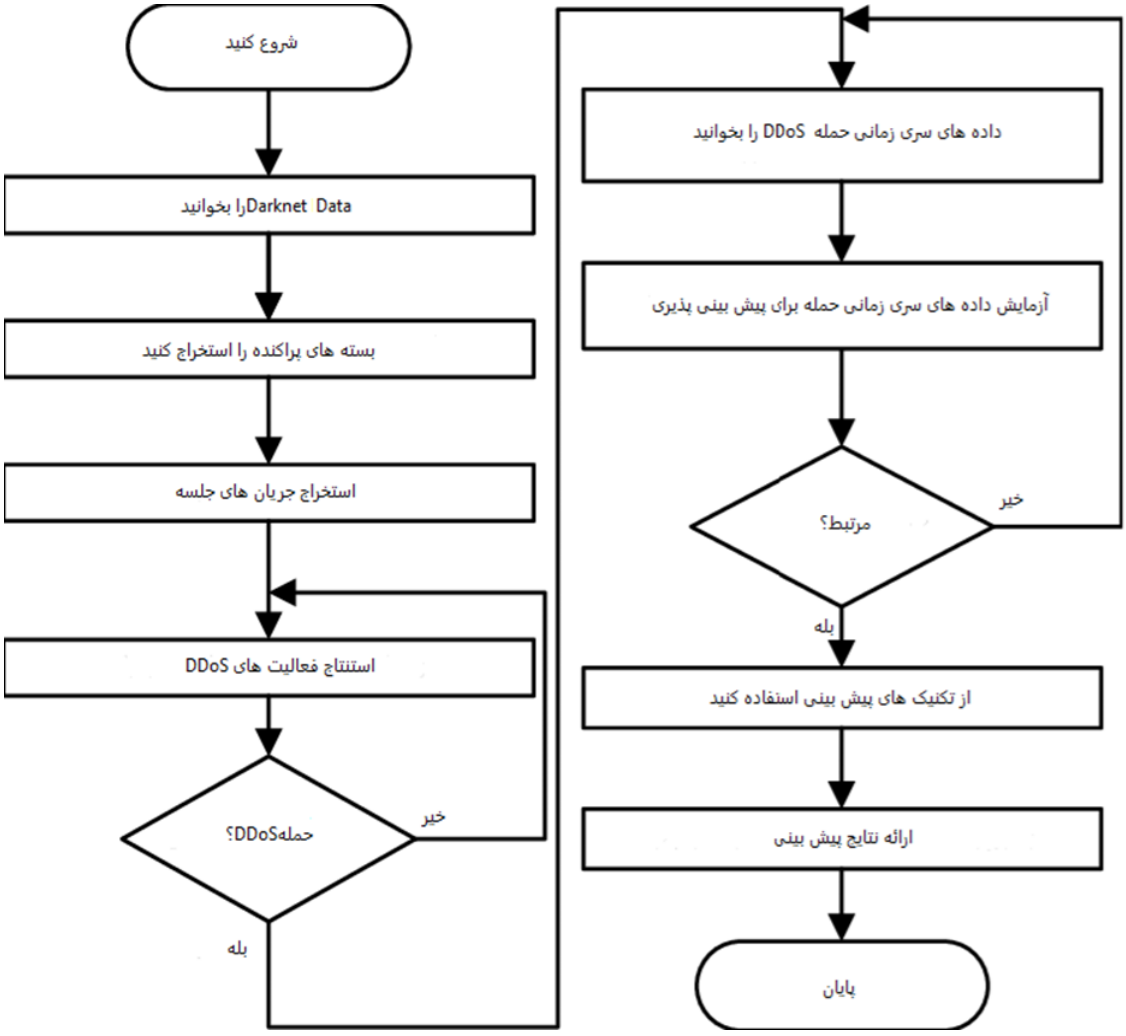
حملات انکار سرویس می تواند با هدف بهره برداری از مزایای هر دو روش و کاهش معایب آنها انجام شود.این ترکیب معمولا به صورت استفاده از جنگل تصادفی با درخت های تصمیم بهره می برد.

جمع آوری داده ها: جمع آوری داده های شبکه شامل ویژگی هایی مانند آدرس IP، تعداد بسته ها، ترافیک ورودی و خروجی

پیش پردازش داده ها: پاکسازی داده ها از نویز و داده های ناقص.نرمال سازی داده ها به منظور داشتن توزیع یکنواخت

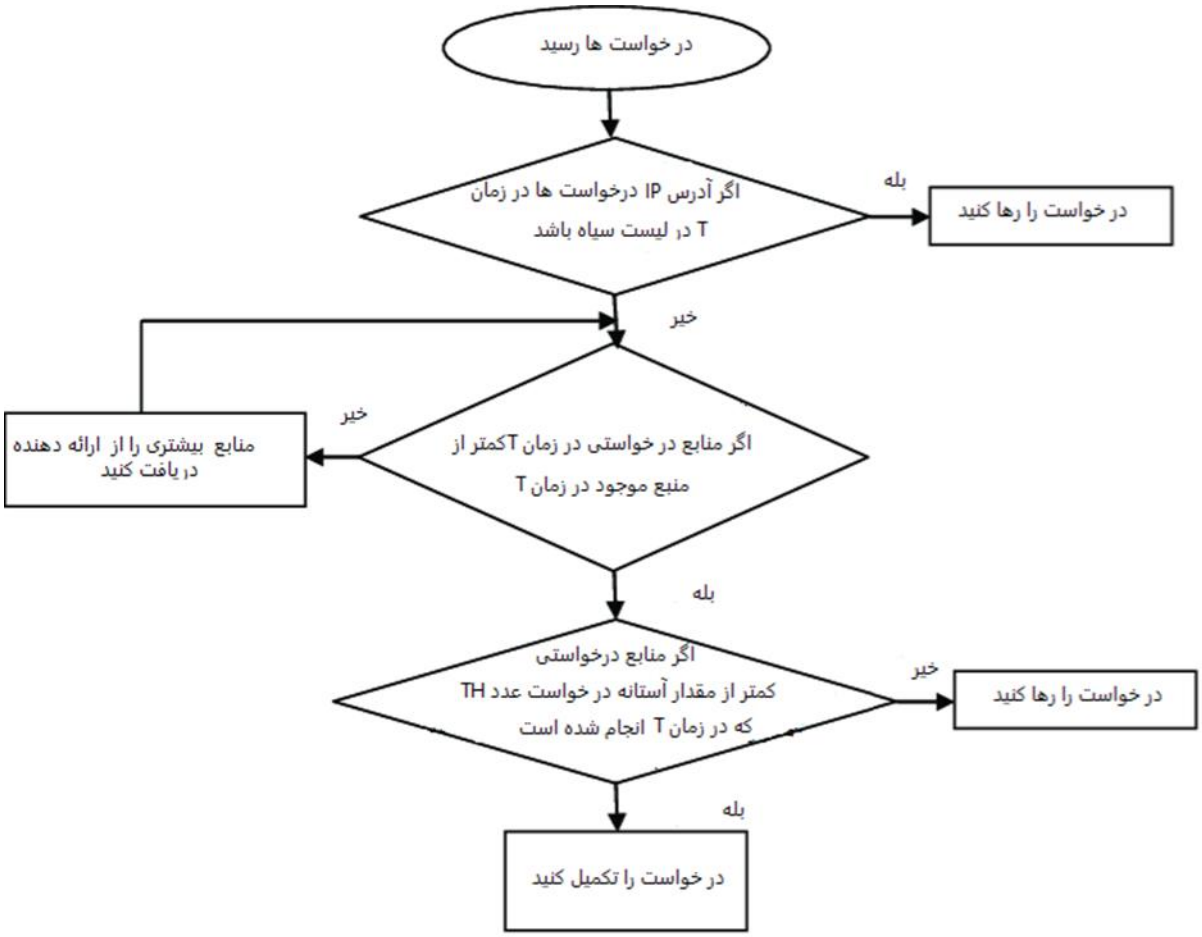
۵-۳-۲- الگوریتم های یادگیری ماشین

ساخت چندین درخت تصمیم با استفاده از نمونه های تصادفی از ویژگی ها و نمونه های آموزش ساخته می شود. استفاده از مدل جنگل تصادفی برای شناسایی حملات انکار سرویس در داده های جدید شبکه.پایش مدام شبکه و به روزرسانی مدل با داده های جدید به منظور بهبود دقت شناسایی.



نمودار(۱): جریان رویکرد پیشنهادی

اجزای اصلی رویکرد استنتاج و پیش بینی DDoS در نمودار نشان داده شده است. به طور خلاصه،این رویکرد با استخراج داده های پراکنده و جریان های جلسه از ترافیک ارائه می شود. متعاقبا،فعالیت های DDoS استنباط می شوند و در نتیجه برای پیش بینی پذیری آزمایش می شوند.در نهایت،در صورت امکان،تکنیک های پیش بینی در ترافیک DDoS اعمال می شوند.



نمودار (۲): نمودار جریان برای الگوریتم پیشنهادی کاهش حمله DDoS

برای غلبه بر حمله DDoS ، اساساً سه راه حل پیشگیری، تشخیص و کاهش وجود دارد. از بین این سه، کاهش حمله ساده ترین راه برای راه حل حمله DDoS است. در راه حل پیشنهادی یک لیست آدرسهای IP معیوب بر اساس عملکرد آنها در طول تست تورینگ تهیه شده و به عنوان لیست سیاه نامگذاری شده است.یک آدرس IP در صورتی که نتواند آزمون تورینگ را پشت سر بگذارد یا در سابقه قبلی آدرس IP زمان زیادی برای پاسخ به سؤال پرسیده شده در طول آزمون تورینگ نیاز داشته باشد،آدرس IP معیوب نامیده می شود. بنابراین اگر درخواست ارسال کننده با آدرس IP در لیست سیاه باشد، سیستم اجازه دسترسی به منابع را نخواهد داد و درخواست خود را رها می کند. اگر درخواست از مرحله اول عبور داده شود، درمرحله دوم بررسی می شود که تعداد منابع درخواستی کمتر از منابع موجود است یا خیر. اگر منابع موجود بیشتر از تعداد منابع درخواستی نسبت به درخواست باشد،به مرحله سوم منتقل می شود، در غیر این صورت منابع بیشتری را برای ارائه درخواست ،درخواست می کند. در مرحله سوم درخواست مقداری از منابع با مقدار آستانه " Th"مقایسه می شود. " Th"مقدار آستانه درخواستی است که می تواند حداکثردر زمان T انجام شود. " Th"را می توان با مشاهده رفتار حداکثرتعداد درخواست های انجام شده در زمانی که هیچ حمله ای وجود ندارد محاسبه کرد. اگر درخواست برای منبع کمتر از Thباشد،منابع به درخواست تخصیص داده می شود در غیر این صورت این درخواست حذف می شود و به عنوان درخواست مخرب مشکوک می شود.

- هر زمان که تعداد درخواست ها در زمان T برسد، آدرس IPاین درخواست ها در لیست سیاه قرار می گیرد یا خیر. اگر آدرسIP درخواست در لیست سیاه قرار گیرد، به عنوان درخواست مخرب مشکوک می شود و حذف می شود.
- برای بقیه درخواست ها،منابع موجود بیشتر از تعداد منابع مورد نیاز درخواست ها در زمانT است. اگر بله، مرحله ۳ بروید.

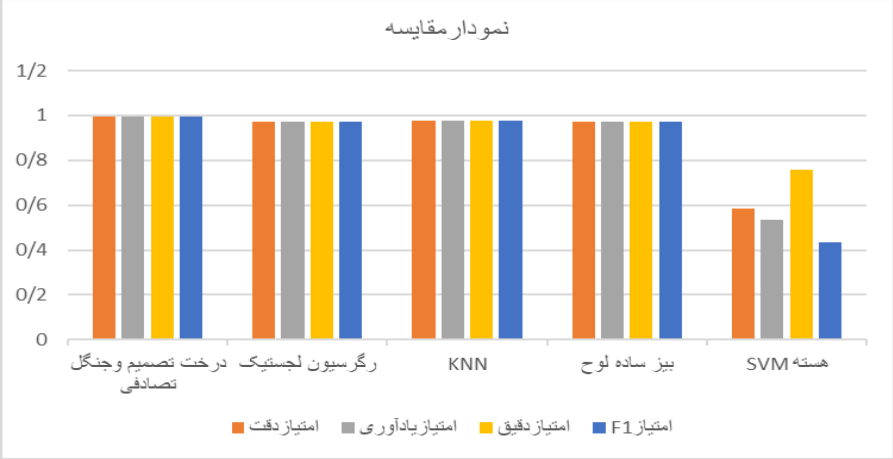
- سپس دوباره، این درخواست ها بررسی می شوند،منابع درخواستی کمتر از مقدار آستانه برای تعداددرخواست هایی هستند که می توانند برای یک منبع درزمان T وارد شوند.
- اگر تعداد درخواست ها از مقدار آستانه بیشتر است، تمام درخواستها را کنار بگذارید.
- برای افزودن منابع بیشتر از ارائه دهندگان درخواست کنید.

۶- ارزیابی

دودسته اصلی از الگوریتم های یادگیری ماشینی تحت نظارت که برای پیش بینی استفاده می شوند، الگوریتم های رگرسیون و طبقه بندی هستند. در حالی که روش های طبقه بندی مقادیر مقوله ای خروجی را پیش بینی می کنند، رویکردهای رگرسیونی مقادیر پیوسته خروجی را پیش بینی می کنند. هدف اصلی این مطالعه پیش بینی مقادیر طبقه بندیDDoS و حملات خوش خیم بر روی برچسب های هدف در مجموعه داده است.از محبوب ترین الگوریتم های طبقه بندی می توان به رگرسیون لجستیک، جنگل تصادفی، درخت تصمیم، Naive Bayes، K-Nearest Neighbour اشاره کرد. این تکنیک ها نه تنها سریع تر، بلکه به طور قابل ملاحظه ای دقیق تر از تکنیک های سنتی برای تشخیص حمله DDoS هستند.

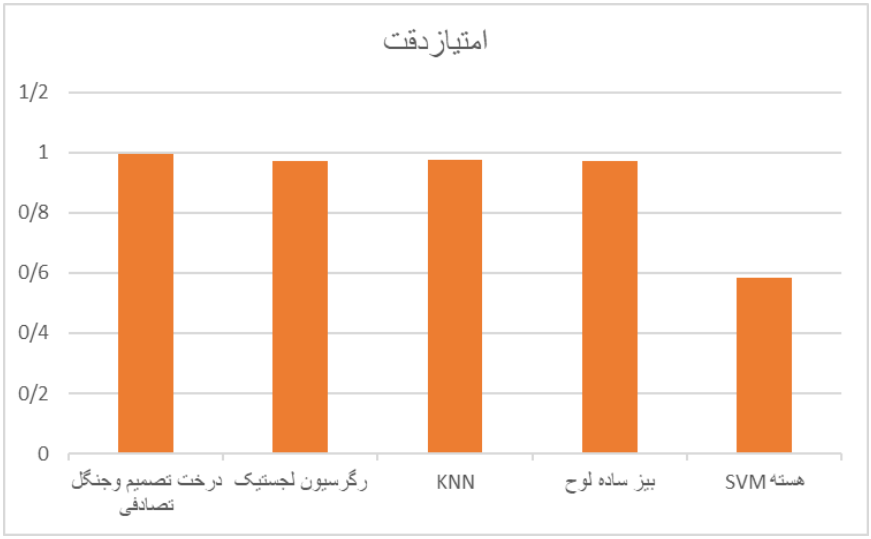
جدول (۱): معیارهای عملکرد

مدل طبقه بندی	امتیاز دقت	امتیاز یادآوری	امتیاز دقیق	امتیاز F1
رگرسیون لجستیک	۰.۹۷۱۳	۰.۹۷۰۱	۰.۹۷۱۸	۰.۹۷۰۹
درخت تصمیم و جدول تصادفی	۰.۰۰۵۹	۰.۹۹۵۴	۰.۹۹۶۳	۰.۹۹۵۸
نزدیک ترین همسایه	۰.۹۷۷۴	۰.۹۷۶۵	۰.۹۹۷۸	۰.۹۷۷۱
بیز ساده لوح	۰.۹۷۱۳	۰.۹۷۱۴	۰.۹۷۰۶	۰.۹۷۱۰
هستهSVM	۰.۵۸۴۰	۰.۵۳۶۹	۰.۷۵۷۳	۰.۴۳۴۳



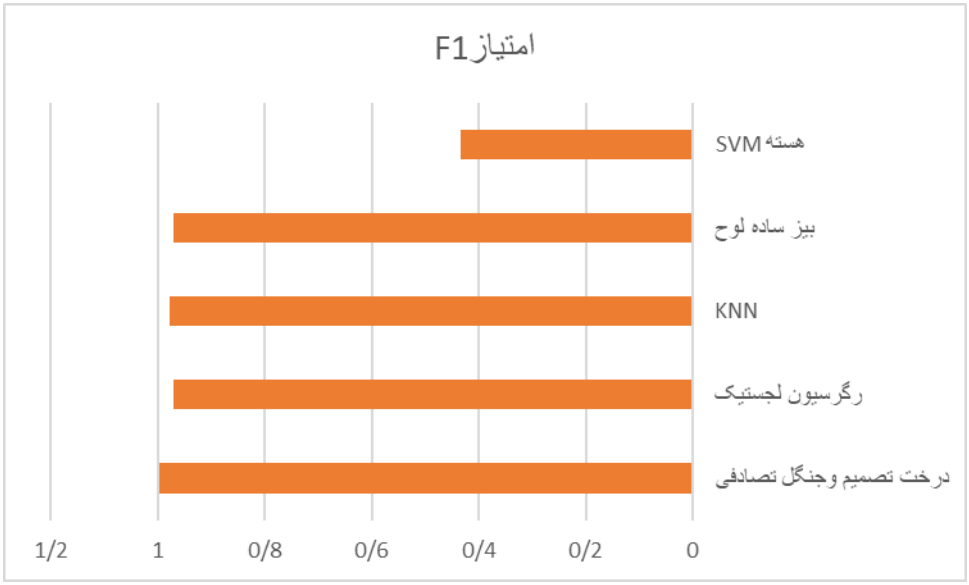
شکل(۳): :نمودار مقایسه جدول معیارهای عملکرد

(جدول ۱) دقت، فراخوانی و امتیاز F1 به دست آمده برای مدل های مختلف را نشان می دهد. مدل های درخت تصمیم و جنگل تصادفی به بالاترین دقت و امتیاز دست می یابند، در حالی که مدل Kernel SVM کمترین امتیاز را نشان می دهد. مدل های رگرسیون لجستیک، Naive Bayes و K-NN سطوح عملکردی را بین مدل های درخت تصمیم/جنگل تصادفی و مدل Kernel SVM نشان می دهند. هر دومدل درخت تصمیم و جنگل تصادفی بالاترین دقت را ارائه می دهند و آنها را برای پیش بینی حملات مبتنی بر شبیه سازی مناسب می کند.



شکل(۴): امتیاز دقت در مدل های طبقه بندی شده یادگیری ماشین

یک پارامتر برای ارزیابی مدل های طبقه بندی دقت است.قابل ذکر است که در بین مدل های یادگیری ماشینی ارزیابی شده، جنگل تصادفی ودرخت تصمیم هر دو دارای بهترین امتیاز دقت ۰.۹۹۵۹ از همه مدل های ذکر شده در بالاهستند. امتیاز دقت KNN در حال حاضر ۰.۹۷۷۴ است، در حالی که نمرات دقت رگرسیون لجستیک و طبقه بندی ساده بیز هر دو ۰.۹۷۱۳ است. امتیاز دقت برای Kernel SVM کمترین است و برابر ۰.۵۸۴۰است.



شکل(۵): امتیاز فراخوان در مقابل مدل های طبقه بندی

(امتیازفراخوان معیاری برای تعیین میزان دقت یک مدل مقادیر مثبت واقعی در بین تمام مقادیر مثبت واقعی است.)
برای ارزیابی مدل های طبقه بندی دقت است.قابل ذکر است که در بین مدل های یادگیری ماشینی ارزیابی شده، جنگل تصادفی ودرخت تصمیم هر دو دارای بهترین امتیاز دقت۰/۹۹۵۹ از همه مدل های ذکر شده در بالاهستند.[25][24]. امتیاز دقت KNN در حال حاضر ۰/۹۷۷۴است، در حالی که نمرات دقت رگرسیون لجستیک و طبقه بندی ساده بیز هر دو ۰/۹۷۱۳ است. امتیاز دقت برای Kernel SVM کمترین است. امتیازF1 دقت یک مدل را در یک مجموعه داده به عنوان میانگین هارمونیک یادآوری و دقت، به این نام خوانده می شود. به طور مشابه، درخت تصمیم و جنگل تصادفی هر دو دارای بیشترین

امتیاز F1 با ۰.۹۹۵۸ هستند، در حالی که سایر مدل ها دارای امتیاز کلی F1 قابل احترامی هستند که SVM Kernel کمترین امتیاز را دارد (۰.۴۳۴۳).

۷- نتیجه‌گیری

در این مقاله، به بررسی روش‌های پیشگیری و تشخیص حملات انکار سرویس با استفاده از الگوریتم‌های جنگل تصادفی و درخت تصمیم پرداخته شد. نتایج نشان داد که این الگوریتم‌ها به دلیل دقت و کارایی بالای خود، می‌توانند به عنوان ابزارهای مؤثر در مقابله با حملات انکار سرویس مورد استفاده قرار گیرند. در آینده، انجام تحقیقات بیشتر برای بهبود این الگوریتم‌ها و ترکیب آنها با سایر روش‌های امنیتی می‌تواند به افزایش امنیت سایبری کمک کند.

References

مراجع

[1] J. Yuan, K. Mills, "Monitoring the macroscopic effect of DDoS flooding attacks", *IEEE Trans. Dependable Secur. Comput.*, vol. 2, no. 4, pp. 324–335, 2005, doi: 10.1109/TDSC.2005.50.

[2] H. Arora, T. Manglani, G. Bakshi, S. Choudhary, "Cyber Security Challenges and Trends on Recent Technologies", *in Proc. 6th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, 2022, pp. 115–118, doi: 10.1109/ICCMC53470.2022.9753967.

[3] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments", *Energy Reports*, vol. 7, pp. 8176–8186, 2021, doi: 10.1016/j.egyr.2021.08.126.

[4] I. H. Sarker, M. H. Furhad, R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions", *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–18, 2021, doi: 10.1007/s42979-021-00557-0.

[5] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade", *IEEE Access*, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[6] A. B. Nassif, M. A. Talib, Q. Nasir, F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review", *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[7] A. Aljuhani, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments", *IEEE Access*, vol. 9, pp. 42236–42264, 2021, doi: 10.1109/ACCESS.2021.3062909.

[8] K. J. Singh, K. Thongam, T. De, "Detection and differentiation of application layer DDoS attack from flash events using fuzzy-GA computation", *IET Inf. Secur.*, vol. 12, no. 6, pp. 502–512, 2018, doi: 10.1049/iet-ifs.2017.0500.

[9] T. Subbulakshmi, S. Mercy Shalinie, A. Ramamoorthi, "Detection and classification of DDOS attacks using machine learning algorithms", *Eur. J. Sci. Res.*, vol. 47, no. 3, pp. 334–346, 2010, doi: 10.1007/978-3-642-14478-3.

[10] Y. Zhang, P. Tino, A. Leonardis, K. Tang, "A Survey on Neural Network Interpretability", *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 5, no. 5, pp. 726–742, 2021, doi: 10.1109/TETCI.2021.3100641.

[11] R. E. Spiridonov, V. D. Cvetkov, O. M. Yurchik, "Data mining for social networks open data analysis", *in Proc. 2017 IEEE 2nd Int. Conf. Control Tech. Syst. (CTS)*, 2017, pp. 395–396, doi: 10.1109/CTSYS.2017.8109578.

[12] R. Biswas and J. Wu, "Optimal Filter Assignment Policy Against Distributed Denial-of-Service Attack", *IEEE Trans. Dependable Secur. Comput.*, vol. 19, no. 1, pp. 339–352, 2022, doi: 10.1109/TDSC.2020.2987301.

[13] A. Waleed, A. F. Jamali, A. Masood, "Which open-source IDS? Snort, Suricata or Zeek", *Comput. Networks*, vol. 213, 2022, doi: 10.1016/j.comnet.2022.109116.

[14] A. Coscia, V. Dentamaro, S. Galantucci, A. Maci, G. Pirlo, "Automatic decision tree-based NIDPS ruleset generation for DoS/DDoS attacks", *J. Inf. Secur. Appl.*, vol. 82, p. 103736, 2024, doi: 10.1016/j.jisa.2024.103736.

[15] N. S. Naz, S. Abbas, M. A. Khan, B. Abid, N. Tariq, M. F. Khan, "Efficient load balancing in cloud computing using multi-layered Mamdani fuzzy inference expert system", *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 3, pp. 569–577, 2019, doi: 10.14569/IJACSA.2019.0100373.

[16] M. Gohil and S. Kumar, “Evaluation of Classification algorithms for Distributed Denial of Service Attack Detection”, *in Proc. 2020 IEEE 3rd Int. Conf. Artif. Intell. Knowl. Eng. (AIKE),* 2020, pp. 138–141, doi: 10.1109/AIKE48582.2020.00028.

[17] J. J. Praba and R. Sridaran, “An SDN-based Decision Tree Detection (DTD) Model for Detecting DDoS Attacks in Cloud Environment”, *Int. J. Adv. Comput. Sci. Appl*., vol. 13, no. 7, pp. 54–64, 2022, doi: 10.14569/IJACSA.2022.0130708.

[18] G. Lucky, F. Jjunju, A. Marshall, “A Lightweight Decision-Tree Algorithm for detecting DDoS flooding attacks”, *in Proc. Companion 2020 IEEE 20th Int. Conf. Softw. Qual. Reliab. Secur. (QRS-C)*, 2020, pp. 382–389, doi: 10.1109/QRS-C51114.2020.00072.

[19] M. I. Kareem and M. N. Jasim, “DDOS Attack Detection Using Lightweight Partial Decision Tree algorithm”, *in Proc. 2nd 2022 Int. Conf. Comput. Sci. Softw. Eng. (CSASE),* 2022, pp. 362–367, doi: 10.1109/CSASE51777.2022.9759824.

[20] B. M. Kalema and V. V. Busobozi, Big Data Analytics for Data Quality Improvement to Enhance Evidence-Based Health Care in Developing Countries, 2020, doi: 10.1007/978-981-13-8253-6_4.

[21] C. O. Tinubu, A. S. Sodiya, O. A. Ojesanmi, E. O. Adeleke, A. O. Adebowale, “DT-Model: a classification model for distributed denial of service attacks and flash events”, *Int. J. Inf. Technol*., vol. 14, no. 6, pp. 3077–3087, 2022, doi: 10.1007/s41870-022-00946-5.

[22] Y. Chen, J. Pei, D. Li, “DETPro: A High-Efficiency and Low-Latency System Against DDoS Attacks in SDN Based on Decision Tree”, *in IEEE Int. Conf. Commun*., 2019, pp. 1–6, doi: 10.1109/ICC.2019.8761580.

[23] V. G. Costa and C. E. Pedreira, “Recent advances in decision trees: an updated survey”, vol. 56, no. 5, *Springer Netherlands*, 2023, doi: 10.1007/s10462-022-10275-5.

[24] I. Singh, S. K. Singh, R. Singh, S. Kumar, “Efficient Loop Unrolling Factor Prediction Algorithm using Machine Learning Models”, *in 2022 3rd Int. Conf. Emerg. Technol. (INCET)*, 2022, pp. 1–8, doi: 10.1109/INCET54531.2022.9825092.

[25] N. Nedjah, G. Martínez Pérez, B. B. Gupta, Eds., Lecture Notes in Networks and Systems 599: International Conference on Cyber Security, Privacy and Networking (ICSPN 2022), 2022.