6TH INTERNATIONAL CONFERENCE ON APPLIED RESEARCH IN COMPUTER, ELECTRICAL AND INFORMATION TECHNOLOGY
March 6, 2022
Tbilisi - Georgia
CINVU
Kosar University
Avicenna International Community College LLC

ارائه راهکاری جدید جهت تشخیص نفوذ در اینترنت اشیاء با استفاده از تکنیک‌های داده‌کاوی

محمد جلالی[1] ، رضا روشنی[2]

۱-کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه فنی و حرفه‌ای، تهران، ایران

۲-دکتری تخصصی، گروه مهندسی کامپیوتر، دانشگاه فنی و حرفه‌ای، تهران، ایران

چکیده

در سال‌های اخیر خدمات مبتنی بر شبکه نوظهور رشد زیادی داشتند، یکی از این موارد اینترنت اشیا[1] است. اینترنت اشیاء با چالش‌هایی مواجه است که از مهم‌ترین آن‌ها می‌توان به امنیت داده‌های موجود، جلوگیری از نفوذ و حملات اشاره کرد. به همین دلیل سیستم تشخیص نفوذ به عنوان یکی از مهم‌ترین راهکارها جهت ارتقاء امنیت اینترنت اشیاء معرفی شد. در سیستم‌های تشخیص نفوذ، نفوذهای غیرمجاز به شبکه با الگوریتم‌هایی تشخیص داده می‌شود و نتایج آن به اطلاع کاربر اطلاع می‌رسد. در این کار، اشیاء با استفاده از الگوریتم جنگل تصادفی روشی برای تشخیص نفوذ در دستگاه‌های همراه اینترنت ارائه می‌شود. روش پیشنهادی در سه مرحله نرم افزار وکا شبیه‌سازی می‌شود. در مرحله اول، داده‌ها از وب سایت تحلیل و آنالیز ترافیک دستگاه‌های همراه و مقالات جمع آوری می‌گردد. در مرحله دوم، از تکنیک CFS که یک روش کاهش ویژگی است و براساس همبستگی‌ها پایه‌گذاری شده است استفاده می‌گردد. سپس الگوریتم BestFirst و الگوریتم جنگل تصادفی بر روی ویژگی‌های مهم انتخاب شده بر روی داده‌ها اجرا خواهد شد. در نهایت، برای اعتبارسنجی و ارزیابی روش پیشنهادی از فاکتورهای FT، TP، Recall و Precision استفاده می‌شود. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی در تمامی این فاکتورها کارایی بالاتری دارد.

کلمات کلیدی: اینترنت اشیاء، دستگاه‌های همراه، سیستم‌های تشخیص نفوذ، داده کاوی، الگوریتم جنگل تصادفی

[1] Internet Of Things

6TH INTERNATIONAL CONFERENCE ON APPLIED RESEARCH IN COMPUTER, ELECTRICAL AND INFORMATION TECHNOLOGY
March 6, 2022
Tbilisi - Georgia
CINVU
Kosar University
Avicenna International Community College LLC

۱. مقدمه

با گسترش روزافزون اینترنت، شاهد حضور گسترده آن در تمامی حوزه‌های زندگی افراد هستیم. تاکنون کاربرد اصلی از اینترنت، استفاده افراد به منظور گرفتن سرویس خاصی بوده است. اکنون شاهد شکل نوین استفاده از اینترنت هستیم که با عنوان اینترنت اشیاء شناخته می‌شود (Altman And Keren, 2016). در اینترنت اشیاء قابلیت اتصال اشیاء از طریق اینترنت مهیا می‌شود. اینترنت اشیاء یک پارادایم جدید است که به سرعت در حال رشد است. اصطلاح اینترنت اشیاء چند سالی هست که معروف شده است و در سال‌های اخیر، به دلیل پیشرفت تکنولوژی‌های بی‌سیم بیشتر به آن توجه شده است. ایده اصلی توجه به انواع شی مانند RFID، NFC، سنسور، دیسک، تلفن‌های همراه و غیره است که می‌تواند با اشیاء دیگر که آدرس مجزا دارند تعامل داشته باشد (Giusto et al., 2010). اشیاء در IoT قدرت قابل توجهی برای دیدن، شنیدن، فکر کردن و اجرای کارها از طریق گفتگو با همدیگر و به اشتراک گذاشتن اطلاعات و همگام سازی اطلاعات دارند. IoT این اشیاء را از طریق دستکاری فن آوری آن مانند محاسبات همه جایی و فراگیر، دستگاه‌های تعبیه شده، فن‌آوری‌های ارتباطی، شبکه‌های حسگر، پروتکل‌ها و برنامه‌های کاربردی از حالت سنتی به هوشمند تبدیل می‌کنند. همچنین زمانی‌که در اینترنت اشیا، فرکانس رادیویی معرفی شد، به نظر می‌رسد وجود این فن‌آوری برای آن ضروری است. فن‌آوری‌های مشابه RFID، مانند ارتباطات زمینه‌ای نزدیک، ماشین به ماشین و ارتباطات وابسته به وسایل نقلیه به وسایل نقلیه، که می‌توان از آن برای پیاده‌سازی ایده‌های نوین در اینترنت اشیا استفاده کرد را نام برد. امکان اینکه با استفاده از فن‌آوری‌های گوناگون اینترنت اشیاء زندگی کاربران آسان و راحت شود وجود دارد. در نتیجه، اینترنت اشیا تاثیر بسیار زیادی بر حوزه خانواده مانند کمک به زندگی، خانه‌های هوشمند، خودروهای هوشمند و غیره دارد. همچنین در بخش کسب و کار، اینترنت اشیاء پیشرفت قابل توجهی در تولید و خدمات از جمله ارائه خدمات بهتر، تولید بیشتر و کیفیت برتر دارد (Altman And Keren, 2016). روش‌های مهمی برای مدیریت اینترنت اشیاء با استفاده از تکنیک‌های جدید وجود دارد:

- استفاده از روش‌های خودکار برای سازماندهی و حفظ داده‌ها بر اساس محتوا.
- تحکیم امنیت داده‌های اینترنت اشیا بدون در نظر رفتن این که این داده‌ها از کجا می‌آیند و کجا نگهداری می‌شوند.

در اینترنت اشیاء، مسائل چالشی زیادی از نظر فنی و اجتماعی وجود دارند که هنوز جای بحث دارند و باید به بیشتر آن‌ها پرداخته شود. برای اینکه IoT بطور کامل مورد پذیرش قرار بگیرد، باید میزان هماهنگی در بین دستگاه‌های آن بیشتر شود، سطح هوشمندی آن‌ها از طریق افزایش توانایی در سازگاری و رفتار مستقل در کنار تضمین اعتماد، حفظ حریم خصوصی و امنیت افزایش یابد. مسأله استانداردسازی، مشکلات آدرس دهی، مقیاس پذیری و غیره از جمله مسئله‌هایی هستند که نیاز است درباره آن‌ها تحقیقات بیشتر صورت گیرد (Saraswati et al., 2016).

نفوذ به عنوان مجموعه‌ای از عملیات‌ها تعریف می‌شود که تلاش می‌کنند یکپارچگی، محرمانگی و در دسترس بودن یک منبع را به مخاطره بیاندازد. سیستم‌های تشخیص نفوذ اجازه می‌دهند تا در صورت دیدن یک رفتار غیر عادی در شبکه که در آن نفوذگر بعد از گذر از سیستم امنیتی شبکه قصد دسترسی غیرعادی به شبکه را دارد، شناسایی کند. جلوگیری از نفوذ

به صورت کامل میسر نمی‌باشد، اما لازم است اقداماتی انجام شود که به صورت خودکار و آنی رفتارهای کاربران مورد پایش قرار گرفته و در صورت مشاهده رفتار نفوذگرانه از آن جلوگیری به عمل آید. معمولاً آسیب پذیری سیستم به دلیل ضعف در نرم افزارهای امنیتی و یا اشکال در ساختار پیکربندی شبکه جهت کنترل دسترسی به اطلاعات شبکه است. به طور کلی روش‌های تشخیص نفوذ به دو دسته اصلی تشخیص سوءاستفاده و تشخیص رفتار غیرعادی تقسیم می‌شوند. در روش تشخیص سوء استفاده از الگوهای نفوذ شناخته شده برای شناسائی نفوذها استفاده می‌شود. در حالی که در روش‌های تشخیص رفتار غیر عادی، رفتار عادی کاربران ملاک عمل قرار داده می‌شود (Olejniczak et al., 2017).

سیستم تشخیص نفوذ، دسترسی کاربر به سیستم کامپیوتر را با اجرای قوانین خاص، بازبینی و محدود می‌کند. قوانین، مبتنی بر دانش متخصص می‌باشد که از مسئولان با تجربه‌های که سناریوهای حمله را ساخته‌اند، استخراج شده است. سیستم همه‌ی تخلفات توسط کاربران را شناسایی کرده و اقدامات لازم برای متوقف کردن حمله بر روی پایگاه داده را انجام می‌دهد. به مسئله تشخیص نفوذ در امنیت کامپیوتر بصورت گسترده پرداخته شده است. سیستم تشخیص نفوذ، نرم‌افزار یا سخت‌افزاری است که عمل تشخیص نفوذ را به طور خودکار انجام می‌دهد. با روند رو به رشد استفاده از شبکه‌های کامپیوتری به خصوص اینترنت و مهارت رو به رشد کاربران و مهاجمان این شبکه‌ها، و نیز نقاط آسیب پذیری مختلف در نرم افزارها، ایمن‌سازی سیستم‌ها و شبکه‌های رایانه‌ای از اهمیت بیشتری نسبت به گذشته برخوردار شده است. روش‌های مختلفی برای تشخیص نفوذ در سیستم پیشنهاد شده است (Gai et al., 2015; Namjouye Rad And Dadgarpour, 2021). در این کار ما بر روی مهم‌ترین متغیرهایی که کارایی مدل پیشنهادی را نشان می‌دهند متمرکز می‌شویم که عبارت اند از: نسبت دقت موفقیت کلی از نسبت تعداد دسته‌بندی‌های درست به کل دسته‌بندی‌ها، پارامتر فراخوانی که بیان کننده این موضوع است که ماشین یادگیری چه نسبتی از کلاس‌های مثبت را درست پیش‌بینی می‌کند، پارامتر دقت که بیان می‌کند چند درصد از رکوردهایی که ماشین یادگیری، مثبت معرفی کرده واقعاً مثبت هستند و همچنین معیار اندازه‌گیری F[1] که یک معیار مناسب برای ارزیابی دقت یک آزمایش است و از دو پارامتر دقت[2] و فراخوانی[3] استفاده می‌کند و مقدار آن در بهترین حالت یک و در بدترین حالت صفر است. از این معیار در مواردی استفاده می‌شود که نتوان اهمیت ویژه‌ای را برای هر یک از دو معیار فراخوانی و دقت نسبت به یکدیگر قائل شد. یکی دیگر از متغیرهای مهم در مدل پیشنهادی معیار خطای دسته‌بندی می‌باشد که دقیقاً برعکس معیار نسبت دقت کلی موفقیت می‌باشد.

۲. کارهای مرتبط

در این بخش به مرور انواع روش‌های تشخیص نفوذ می‌پردازیم. در (Yan et al., 2015) روشی برای سیستم‌های توزیع شده تشخیص نفوذ[4] ارائه شد. در این سیستم در هر ناحیه از رایانش ابری IDSها مستقر شده‌اند و هر IDS سه ماژول دارد:

[1] F-Measure
[2] Precision
[3] Recall
[4] Distributed Intrusion Detection System (DIDS)

بلاک، ارتباطات و ماژول همکاری. ماژول بلاک برای بسته‌های مخربی که از گره مبدا فرستاده شده استفاده می‌شود و ماژول ارتباطات برای پیام هشداری که برای تشخیص حملات خاص که توسط خود IDS یا دیگر IDSها ارسال شده، می‌باشد و ماژول همکاری برای جمع آوری پیام‌های هشداری است که تصمیم می‌گیرد آیا هشدار درست است یا نه. بررسی نتایج این روش نشان می‌دهد که دقت آن نسبت به روش‌های مشابه خود بهتر است.

در (Moyo, 2016) روشی برای تشخیص نفوذ در شبکه با استفاده از نایوبیز ارائه کردند. نایوبیز یکی از روش‌های طبقه‌بندی بر اساس احتمالات می‌باشد که با استفاده از قضیه بیز، و همراه با فرضیه نایو درباره ویژگی‌های مستقل؛ کار می‌کند. در این روش فرض بر آن است که ارزش هر ویژگی مستقل از ارزش‌های دیگر ویژگی‌هاست. این فرض با عنوان مستقل شرطی شناخته می‌شود. بر خلاف فرضیه نایو و ساده‌سازی، طبقه‌بندی کننده نایوبیز باید برای استفاده در موقعیت‌های پیچیده واقعی ارتقا یابد. آن‌ها روش پیشنهادی خود را با مجموعه داده KDD99 و با استفاده از ابزار Weka آزمایش کردند. نتایج آزمایشات نشان داد روش پیشنهادی آن‌ها نسبت به سایر روش‌های مشابه دارای دقت بیشتری می‌باشد.

در (Borisenko et al., 2016; Ferrari And de Castro, 2015) روشی برای جلوگیری از حملات با الگوریتم خوشه‌بندی موسوم به LBG ارائه شد. با توجه به اینکه در الگوریتم خوشه‌بندی k-means به انتخاب اولیهٔ خوشه‌ها بستگی دارد و باعث می‌شود در تکرارهای مختلف الگوریتم، نتایج متفاوتی از خوشه‌بندی را داشته باشیم و در بسیاری از کاربردها نمی‌توان از آن استفاده کرد. برای رفع این مشکل الگوریتم خوشه‌بندی LBG پیشنهاد شد که تا حدودی توانایی غلبه به این مشکل را داشت. در این روش ابتدا الگوریتم، تمام داده‌ها را به صورت یک خوشه در نظر می‌گیرد و سپس برای این خوشه یک بردار مرکز محاسبه می‌کند، سپس این بردار را به دو بردار می‌شکند و داده‌ها را با توجه به این دو بردار خوشه‌بندی می‌کند. در مرحلهٔ بعد این دو نقطه به چهار نقطه شکسته می‌شوند و الگوریتم ادامه پیدا می‌کند تا تعداد خوشهٔ مورد نظر تولید شوند. الگوریتم LBG یک الگوریتم رقمی‌سازی برداری است که با استفاده از آن می‌توان یک Codebook مناسب بدست آورد. Codebook، مجموعه مراکز بازه‌های رقمی‌سازی است. بطور کلی الگوریتم LBG یک الگوریتم نوع پیمایشی است.

در (Iyenger And Ganapathy, 2015) روشی برای تشخیص حملات با تئوری آشوب ارائه شد. در این روش، تمام ترافیک شبکه پس از جمع آوری بسته‌های ترافیک شبکه و جریان اطلاعات نمونه گیری می‌شود. آن‌ها شناسایی ناهنجاری شبکه‌ی خود و الگوریتم شناسایی DDoS را در ۴ گام ارائه کردند. در مرحله اول، جمع آوری بسته‌های ترافیک شبکه و جریان اطلاعات به صورت بلادرنگ انجام می‌شود. سپس در گام دوم، پیش پردازش ترافیک شبکه با روابط قبلی محاسبه و سپس پیش‌بینی آن انجام می‌شود. در ادامه در مرحله سوم، بر اساس نظریه آشوب خطای پیش‌بینی تحلیل می‌شود و ترافیک غیر نرمال را تشخیص می‌دهند و در نهایت گام چهارم شناسایی DDOS انجام خواهد شد.

در (Guha et al., 2016) روشی برای تشخیص نفوذ با استفاده از شبکه‌های عصبی مصنوعی ارائه شد. معماری این روش در سه سطح تنظیم شده است. در سطح اول نرون‌های لایه ورودی قرار دارند. در سطح دوم نرون‌های لایه مخفی و نهایتاً در سطح آخر نرون‌های خروجی که نشان دهنده حمله یا عدم حمله‌اند قرار دارند. در این روش نیز برای عملیات

آموزش، اعتبارسنجی و آزمایش شبکه از مجموعه داده‌های KDD استفاده کردند. نتایج آزمایشات آن‌ها نشان می‌دهد روش پیشنهادی در دسته‌بندی انواع درخواست‌ها که مبتنی بر ایجاد ترافیک برای شبکه بوده‌اند بسیار موفق عمل کرده و نسبت به سایر روش‌ها دارای کیفیت و کارآمدی بیشتری می‌باشد.

در (David And Thomas, 2015) روشی برای تشخیص نفوذ با آنتروپی ارائه شد. از آنجا که حملات FLASH CROWD و DOS حاکی از ایجاد یک سربار اضافی با هدف ایجاد اختلال و متوقف نمودن فعالیت‌های یک وب می‌باشد، به نوعی سیستم هدف دچار درهم ریختگی، بی نظمی و کهولت می گردد که میزان این بهم ریختگی و آشفتگی را می‌توان با محاسبه آنتروپی هر IP محاسبه نمود. مدل پیشنهادی آن‌ها در پوشش کامل همه مشخصات بهینه و مطلوب، موفق بوده است. ولی از همه حملات نرخ پایین منفی کاذب رنج می‌برد، چراکه توانایی تشخیص و پیشگیری همه حملات جمعیت فلش را ندارد. همچنین، برای اعتبارسنجی و ردیابی برخی از درخواست‌ها با شکست مواجه می‌شود. ضمن اینکه در مدل پیشنهادی حملات منع سرویس شناسایی می‌شوند و امکان شناسایی مبداء حملات منع سرویس توزیع شده بواسطه پیچیدگی بیشتر، بطور دقیق میسر نیست و نیاز است از طریق الگوریتم‌های داده‌کاوی و همبسته سازی مدل‌های پیشنهادی را توسعه داد.

در (Kebande And Venter, 2014) روشی برای تشخیص حملات مبتنی بر یک سیستم مصنوعی ارائه شد. روش پیشنهادی از چندین تابع که هر کدام کار مخصوص به خود را انجام می‌دهند تشکیل شده است. این توابع شامل: تابع کلاس بندی (از این تابع برای دسته‌بندی حملات در دو دسته حمله و غیر حمله استفاده می‌شود)، تابع ایجاد الگو (این تابع در یک بازه زمانی پیام‌های رسیده را ثبت می کند. اگر تعداد این پیام‌ها از حد آستانه بیشتر باشد از آن الگویی تشکیل می‌دهد و به عنوان یک حمله احتمالی در نظر گرفته می‌شود در غیر اینصورت الگوی حاصله کنار گذاشته می‌شود)، تابع ارسال و دریافت الگوی حمله احتمالی (وظیفه این تابع مطلع کردن ایستگاه پایه و سایر ماشین‌ها از حمله احتمالی است)، تابع تشخیص و تطابق الگوی حمله (این تابع عملیات تطابق ویژگی‌های حمله احتمالی با حمله‌های موجود در مجموعه داده‌ها را انجام می‌دهد) و تابع محدودیت نرخ (این تابع برای معلق کردن گره‌هایی است که دچار حمله شده‌اند. چنانچه حجم پیام‌های ارسالی بیش از حد آستانه اشغالی پهنای باند باشد و احتمال وقوع حمله اعلام شود این تابع گره‌های مشکوک به حمله را معلق می کند). همچنین در سیستم پیشنهادی آن‌ها تابعی برای تولید نسل جدید از شناساگرها در نظر گرفتند. این تابع مبتنی بر الگوریتم ژنتیک عمل می کند. آن‌ها برای ارزیابی روش پیشنهادی خود از مجموعه داده KDD استفاده کردند. نتایج آزمایشات آن‌ها نشان دهنده این است که روش پیشنهادی آن‌ها نسبت به سایر روش‌ها دارای دقت بالاتری می‌باشد.

۳. روش پیشنهادی

از مهم‌ترین روش‌های تحلیل هشدارهای نفوذی در محیط شبکه استفاده از روش‌های داده‌کاوی است (Safavian And Landgrebe, 1991). ما از الگوریتم شناخته شده درخت تصمیم که یک مدل پیش‌بینی کننده در داده کاوی

می‌باشد، استفاده می‌کنیم. در مسئله تشخیص نفوذ انواع حملات به شبکه‌های کامپیوتری به چهار دسته DoS، Probe، U2R، R2L تقسیم‌بندی شده است. همچنین اگر دسته حملات رایج را در نظر بگیریم، ورودی‌ها به یکی از ۵ دسته فوق تعلق خواهند داشت. در واقع هر کدام از این کلاس‌ها خود شامل چندین حمله شناخته شده می‌باشد. برای مثال DoS شامل حملاتی نظیر Neptune,Smurf, back و غیره است. در اکثر راه‌های مربوط به حل این مساله، دسته‌بندی برای ۵ کلاس مذکور صورت گرفته و به نظر می‌رسد تفکیک هر یک از حملات به عنوان زیرکلاس، دقت تشخیص حمله را بالا می‌برد و موجب می‌شود که عکس‌العمل مناسب‌تری در مقابل حمله مورد نظر پیش بینی گردد.

۳،۱. درخت تصمیم

درخت تصمیم یکی از مشهورترین و قدیمی ترین روش‌های داده‌کاوی جهت ساخت مدل دسته‌بندی است. در الگوریتم‌های دسته‌بندی مبتنی بر درخت تصمیم دانش خروجی به صورت یک درخت از حالات مختلف مقادیر ویژگی‌ها ارائه می‌شود. نمایش درخت به شکل درخت تصمیم سبب شده است که دسته‌بندهای مبتنی بر درخت تصمیم کاملاً قابل تفسیر باشند (Safavian And Landgrebe, 1991). در حالت کلی درخت تصمیم رسم شده برای یک مجموعه داده آموزشی، واحد و یکتا نیست و بر اساس یک مجموعه داده، درخت‌های تصمیم مختلفی می‌توان به دست آورد. ساختار آن یک فرآیند بالا به پایین است. اساساً این الگوریتم یک الگوریتم حریصانه است که از گره ریشه شروع می‌شود و برای هر گره غیر از برگ ابتدا یک صفت برای تست نمونه انتخاب می‌شود سپس مجموعه نمونه به چند زیرمجموعه نمونه با توجه به نتیجه آزمایش تقسیم می‌شود. هر زیر مجموعه نمونه یک گره برگ جدید را تشکیل می‌دهد و در نهایت این فرآیند تقسیم تا زمانی تکرار می‌شود که به شرایط خاص پایانی برسیم.

۳،۲. انتخاب ویژگی

کارایی یک سیستم تشخیص الگو بسیار وابسته به روش انتخاب ویژگی است. از آنجایی که با افزایش تعداد ویژگی‌ها هزینه محاسباتی یک سیستم نیز افزایش می‌یابد، طراحی و پیاده‌سازی سیستم‌ها با کمترین تعداد ویژگی‌های ممکن ضروری به نظر می‌رسد. در اکثر مسائل با ابعاد بالا، انتخاب ویژگی‌های موثر بر مسئله و حذف ویژگی‌های دیگر، می‌تواند تا حد زیادی دقت دسته‌بندی را بالا برده و از پیچیدگی پردازش داده‌ها در مراحل مختلف بکاهد. با توجه به اینکه در نرم افزار وکا سه روش برای نمونه گیری وجود دارد لذا این روش‌ها در ادامه توضیح داده می شوند:

CFS[1] نوعی روش کاهش ویژگی‌است که براساس همبستگی‌ها پایه گذاری شده است. این الگوریتم نمره بالایی به ویژگی‌هایی می‌دهد که دارای وابستگی بیشتری به کلاس و وابستگی ضعیفی با یکدیگر دارند. به عنوان مثال، k نمایانگر

[1] Correlation Feature Selection

تعداد صفات، rcf[1] نمایانگر میانگین ارتباط بین ویژگی‌ها و ویژگی کلاس و rff[2] بیانگر میانگین وابستگی بین ویژگی‌ها با یکدیگر می‌باشد.

آنتروپی میزان خلوص (بی‌نظمی یا عدم خالص بودن) مجموعه‌ای از مثال‌ها را مشخص می‌کند. اگر مجموعه S شامل مثال‌های مثبت و منفی از یک مفهوم هدف باشد آنتروپی S نسبت به این دسته‌بندی بولی تعریف می‌شود.

۴. یافته‌ها

برای بررسی کارایی روش پیشنهادی از روش تحلیلی و شبیه سازی نرم افزار داده‌کاوی Weka استفاده می‌شود. از مجموعه دادگان استاندارد موجود در مقالات و سایت‌های معتبر با کاربردهای مختلف جهت ارزیابی استفاده می‌گردد و خروجی‌های حاصل شده را مورد تجزیه و تحلیل قرار می‌دهیم.

۴،۱. جمع آوری داده

داده‌های استفاده شده در روش پیشنهادی مربوط به اطلاعات شبکه‌ها در بسترهای مختلف مانند ویندوز، لینوکس و سایر سیستم عامل‌ها می‌باشد که مورد حمله قرار گرفته‌اند. داده‌های مورد نظر از وب‌سایت‌های مختلف که در جدول ۱ نشان داده شده است جمع آوری شدند.

جدول ۱: سایت‌های مربوط به جمع آوری داده‌ها

ردیف	آدرس وب‌سایت‌ها
۱	https://www.av-test.org/en/statistics/malware/, Last visit 2021
۲	http://www.netresec.com/?page=AboutNetresec, Last visit 2021
۳	https://www.evilfingers.com/index.php, Last visit 2021
۴	http://www.malware-traffic-analysis.net/about.html, Last visit 2021
۵	https://www.payload-security.com/products/overview/full, Last visit 2021
۶	http://www.unb.ca/research/iscx/, Last visit 2021

۴،۲. نتایج اجرا

تحلیل داده‌ها در چهار مدل متفاوت انجام می‌شود. در روش اول از الگوریتم BestFirst برای انتخاب بهترین ویژگی‌ها استفاده می‌شود. در روش دوم از الگوریتم InfoGainAttributeEval برای انتخاب بهترین ویژگی‌ها استفاده خواهد شد. در ادامه اجرای الگوریتم j.48 بر روی BestFirst انجام خواهد شد و در نهایت اجرای الگوریتم جنگل تصادفی بر روی BestFirst انجام می‌شود.

[1] relationship between class and feature
[2] relationship between feature and feature

در ابتدا الگوریتم BestFirst برای انتخاب بهترین ویژگی‌ها انجام می شود برای اینکار الگوریتم درخت تصمیم گیری اجرا می‌شود. سپس در روش دوم از الگوریتم InfoGainAttributeEval برای انتخاب بهترین ویژگی‌ها استفاده می‌شود و همانند روش اول اجرای الگوریتم درخت تصمیم‌گیری را داریم.

پس از اجرای هر دو روش بالا تمام داده‌های خروجی را ثبت می‌کنیم. سپس یک بار الگوریتم درخت تصمیم گیری را بروی کل داده‌ها اجرا می‌کنیم. علت این کار آزمون و خطا جهت رسیدن به نتیجه بهتر از الگوریتم‌ها می‌باشد.

پس از اجرای الگوریتم BestFirst و اجرای الگوریتم J.48 بر روی BestFirst نتایج به شرح زیر می باشد.

در جدول ۲، صحت جزئیات خاصیت‌ها نسبت به خاصیت class برای اجرای الگوریتم درخت تصمیم گیری بر روی الگوریتم BestFirst و در جدول ۳، صحت جزئیات خاصیت‌ها نسبت به خاصیت class برای اجرای الگوریتم جنگل تصادفی بر روی الگوریتم BestFirst و در جدول ۴، صحت جزئیات خاصیت‌ها نسبت به Class برای اجرای الگوریتم درخت تصمیم‌گیری بر روی کلیه داده‌های اولیه نشان داده می‌شود.

همچنین جهت تحلیل بهتر و مقایسه عملکرد الگوریتم‌های Attribute Selection الگوریتم جنگل تصادفی را طبق روالی که توضیح داده شد، بر روی کلیه داده‌های جمع‌آوری شده اجرا می‌نماییم، نتایج به دست آمده را جهت تحلیل بیشتر و مقایسه با سایر روش‌ها ثبت می‌کنیم. جدول ۵، نشان دهنده صحت جزئیات خاصیت‌ها نسبت به Class برای اجرای الگوریتم جنگل تصادفی RandomForest بر روی کلیه داده‌های اولیه می‌باشد.

جدول۲: نتایج اجرای الگوریتم درخت تصمیم گیری بر روی الگوریتم BestFirst

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
۰.۹۸۳	۰.۱۲۲	۰.۹۳۵	۰.۹۸۳	۰.۹۵۸	۰.۸۸۱	۰.۹۹	۰.۹۹۳	BENIGN
۰.۸۷۸	۰.۰۱۷	۰.۹۶۶	۰.۸۷۸	۰.۹۲	۰.۸۸۱	۰.۹۹	۰.۹۸۲	Mal
۰.۹۴۵	۰.۰۸۵	۰.۹۴۶	۰.۹۴۶	۰.۹۴۵	۰.۸۸۱	۰.۹۹	۰.۹۸۹	Weighted Avg

جدول۳: نتایج اجرای الگوریتم جنگل تصادفی بر روی الگوریتم BestFirst

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
۰.۹۸۱	۰.۱۲	۰.۹۳۶	۰.۹۸۱	۰.۹۵۸	۰.۸۸	۰.۹۹	۰.۹۹۴	BENIGN
۰.۸۸	۰.۰۱۹	۰.۹۳۶	۰.۸۸	۰.۹۲	۰.۸۸	۰.۹۹	۰.۹۸۲	Mal
۰.۹۴۵	۰.۰۸۴	۰.۹۴۶	۰.۹۴۵	۰.۹۴۴	۰.۸۸	۰.۹۹	۰.۹۹	Weighted Avg

جدول٤: نتایج اجرای الگوریتم درخت تصمیم‌گیری بر روی کلیه داده‌های اولیه

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
۰.۹۸۸	۰.۱۱۴	۰.۹۴	۰.۹۸۸	۰.۹۶۳	۰.۸۹۴	۰.۹۹۱	۰.۹۹۲	BENIGN
۰.۸۸۶	۰.۰۱۲	۰.۹۷۶	۰.۸۸۶	۰.۹۲۹	۰.۸۹۴	۰.۹۹۱	۰.۹۸۱	Mal
۰.۹۵۱	۰.۰۷۸	۰.۹۵۳	۰.۹۵۱	۰.۹۵۱	۰.۸۹۴	۰.۹۹۱	۰.۹۸۸	Weighted Avg

جدول۵: نتایج اجرای الگوریتم جنگل تصادفی RandomForest بر روی کلیه داده‌های اولیه

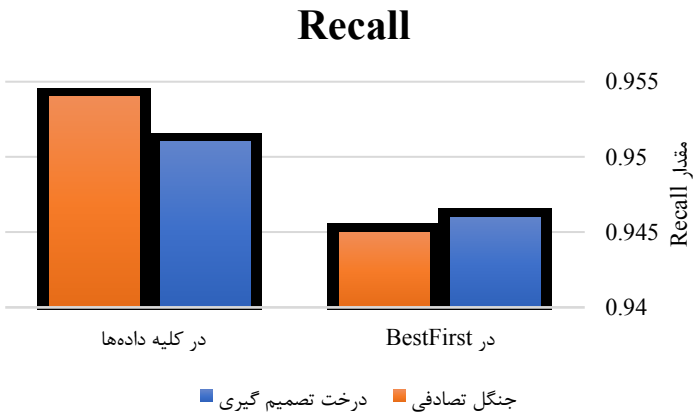
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
۰.۹۸۶	۰.۱۰۳	۰.۹۴۵	۰.۹۸۶	۰.۹۶۵	۰.۹۰۱	۰.۹۹	۰.۹۹۳	BENIGN
۰.۸۹۷	۰.۰۱۴	۰.۹۷۳	۰.۸۹۷	۰.۹۳۴	۰.۹۰۱	۰.۹۹	۰.۹۸	Mal
۰.۹۵۴	۰.۰۷۱	۰.۹۵۵	۰.۹۵۴	۰.۹۵۴	۰.۹۰۱	۰.۹۹	۰.۹۸۸	Weighted Avg

جدول٦: نتایج کلی حاصل شده از آزمایشات انجام شده

روش / فاکتور	فاکتور			
	Recall	Precision	FP	TP
درخت تصمیم گیری برروی BestFirst	۰.۹۴۶	۰.۹۴۶	۰.۰۸۵	۰.۹۴۵
جنگل تصادفی بر روی BestFirst	۰.۹۴۵	۰.۹۴۶	۰.۰۸۴	۰.۹۴۵
درخت تصمیم گیری بر روی کلیه داده‌ها	۰.۹۵۱	۰.۹۵۳	۰.۰۷۸	۰.۹۵۱
جنگل تصادفی بر روی کلیه داده‌ها	۰.۹۵۴	۰.۹۵۵	۰.۰۷۱	۰.۹۵۴

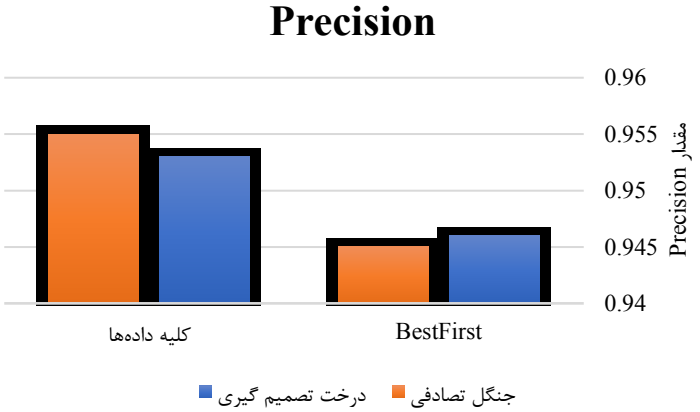
جدول ۶، نشان دهنده نتایج کلی بدست آمده از آزمایشات انجام شده می‌باشد. در این جدول نتایج هر چهار روش برای همه فاکتورها نشان داده شده است. پس از بررسی داده‌های به دست آمده به نتایج قابل قبولی دست یافتیم. هر چه FP عدد کمتری و TP عدد بیشتری را نشان دهد می‌توان نتیجه گرفت که برنامه دارای کیفیت و کارایی مناسب‌تری می‌باشد. در اینجا جنگل تصادفی بر روی کلیه داده‌ها دارای کمترین FP و بیشترین TP می‌باشد که بهترین روش پیشنهادی در اینجا مطرح می‌باشد.

در شکل ۱ نمودار فاکتور recall برای هر چهار روش وجود دارد. همانطور که در این شکل نشان داده شده است روش درخت تصمیم گیری برروی BestFirst برابر با ۰.۹۴۶، روش جنگل تصادفی بر روی BestFirst برابر با ۰.۹۴۵، روش درخت تصمیم گیری بر روی کلیه دادها برابر با ۰.۹۵۱ و روش جنگل تصادفی بر روی کلیه داده‌ها برابر با ۰.۹۵۴ می‌باشد.



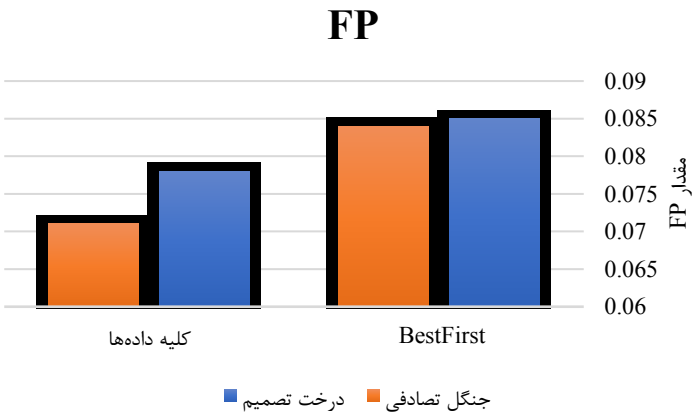
شکل ۱. نمودار فاکتور recall در هر چهار بخش

در شکل ۲ نمودار فاکتور precision برای هر چهار بخش وجود دارد. همانطور که در این شکل نشان داده شده است روش درخت تصمیم گیری برروی BestFirst برابر با ۰.۹۴۶، روش جنگل تصادفی بر روی BestFirst برابر با ۰.۹۴۵، روش درخت تصمیم گیری بر روی کلیه داده‌ها برابر با ۰.۹۵۳ و روش جنگل تصادفی بر روی کلیه داده‌ها برابر با ۰.۹۵۵ می‌باشد.



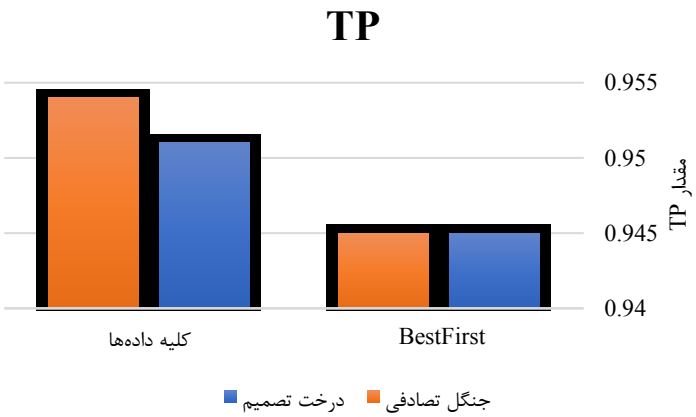
شکل ۲. نمودار فاکتور precision برای هر چهار روش

در شکل ۳ نمودار فاکتور FP برای هر چهار روش وجود دارد. همانطور که در این شکل نشان داده شده است روش درخت تصمیم گیری برروی BestFirst برابر با ۰.۰۸۵، روش جنگل تصادفی بر روی BestFirst برابر با ۰.۰۸۴، روش درخت تصمیم‌گیری بر روی کلیه داده‌ها برابر با ۰.۰۷۸ و روش جنگل تصادفی بر روی کلیه داده‌ها برابر با ۰.۰۷۱ می‌باشد.



شکل۳. نمودار فاکتور FP برای هر چهار روش

در شکل ۴ نمودار فاکتور TP برای هر چهار روش وجود دارد. همانطور که در این شکل نشان داده شده است روش درخت تصمیم گیری برروی BestFirst برابر با ۰.۹۴۵، روش جنگل تصادفی بر روی BestFirst برابر با ۰.۹۴۵، روش درخت تصمیم گیری بر روی کلیه داده‌ها برابر با ۰.۹۵۱ و روش جنگل تصادفی بر روی کلیه داده‌ها برابر با ۰.۹۵۴ می‌باشد.



شکل ۰٤. نمودار فاکتور TP برای هر چهار روش

۵. بحث و نتیجه گیری

با توجه به توسعه انواع خدمات در بستر شبکه، تشخیص نفوذ به عنوان یک روش مهم جهت ارتقاء امنیت شبکه معرفی شده است. در سیستم‌های تشخیص نفوذ، نفوذهای غیرمجاز به شبکه با توجه به الگوریتم‌های خاص تشخیص داده شده و آنها را به کاربر اطلاع می‌دهد. در این کار، یک روشی جدید برای تشخیص نفوذ در شبکه‌های کامپیوتری ارائه شده است. روش پیشنهادی در سه مرحله اجرا شد. در مرحله اول داده‌ها از وب سایت‌های مختلف و مقالات جمع آوری شد و سپس در نرم افزار وکا بارگذاری شدند. در ادامه با استفاده از الگوریتم‌های مختلف آزمایش شده و نتایج آزمایشات نشان می‌دهد روش درخت تصمیم‌گیری برروی BestFirst برابر با ۰.۹۴۵، روش جنگل تصادفی بر روی BestFirst برابر با ۰.۹۴۵، روش درخت تصمیم گیری بر روی کلیه داده‌ها برابر با ۰.۹۵۱ و روش جنگل تصادفی بر روی کلیه داده‌ها برابر با ۰.۹۵۴ می باشد. در همه آزمایشات روش جنگل تصادفی نتایج بهتری از خود نشان داد. همچنین زمان اجرای این روش کمتر از سایر روش‌ها می‌باشد.

منابع

Altman, Y., And Keren, A. Y. (2016). System and method for automated configuration of intrusion detection systems. In: Google Patents.

Borisenko, K., Smirnov, A., Novikova, E., And Shorov, A. (2016). DDoS Attacks Detection in Cloud Computing Using Data Mining Techniques. Advances in Data Mining. Applications and Theoretical Aspects, Cham.

David, J., And Thomas, C. (2015). DDoS Attack Detection Using Fast Entropy Approach on Flow-Based Network Traffic. *Procedia Computer Science*, *50*, 30-36. https://doi.org/10.1016/j.procs.2015.04.007

Ferrari, D. G., And de Castro, L. N. (2015). Clustering algorithm selection by meta-learning systems: A new distance-based problem characterization and ranking combination methods. *Information Sciences*, *301*, 181-194. https://doi.org/10.1016/j.ins.2014.12.044

Gai, K., Qiu, M., Tao, L., And Zhu, Y. (2015). Intrusion detection techniques for mobile cloud computing in heterogeneous 5G. *Security and Communication Networks*, *PP*. https://doi.org/10.1002/sec.1224

Giusto, D., Iera, A., Morabito, G., And Atzori, L. (2010). *The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications*. https://doi.org/10.1007/978-1-4419-1674-7

Guha, S., Yau, S. S., And Buduru, A. B. (2016, 8-12 Aug. 2016). Attack Detection in Cloud Infrastructures Using Artificial Neural Network with Genetic Feature Selection. 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Auckland, New Zealand.

Iyenger, N. C. S. N., And Ganapathy, G. (2015). Chaotic Theory based Defensive Mechanism against Distributed Denial of Service Attack in Cloud Computing Environment. *International Journal of Security and Its Applications*, *9*(9), 197-212. https://doi.org/10.14257/ijsia.2015.9.9.18

Kebande, V. R., And Venter, H. S. (2014). A cognitive approach for botnet detection using Artificial Immune System in the cloud. 2014 Third International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), Beirut, Lebanon.

Moyo, L. (2016). *The classification performance of Bayesian Networks Classifiers: A case study of detecting Denial of Service (DoS) attacks in cloud computing environments* [Master's thesis, University of Fort Hare]. http://libdspace.ufh.ac.za/

Namjouye Rad, A. a., And Dadgarpour, M. (2021). Detection of network penetration by data mining and using machine learning via SVM algorithm. *Karafan Quarterly Scientific Journal*, *17*(4), 13-33. https://doi.org/10.48301/kssa.2021.128393

Olejniczak, M. N., Kozanecki, M., Saramak, J., Matusiak, M., Kadlubowski, S., And Matyjaszewski, K. (2017). Raman spectroscopy study on influence of network architecture on hydration of poly (2-(2-methoxyethoxy) ethyl methacrylate) hydrogels. *Journal of Raman Spectroscopy*, *48*(3), 465-473. https://doi.org/10.1002/jrs.5048

Safavian, S. R., And Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660-674. https://doi.org/10.1109/21.97458

Saraswati, A., Hagenbuchner, M., And Zhou, Z. Q. (2016). High resolution SOM approach to improving anomaly detection in intrusion detection systems. Australasian Joint Conference on Artificial Intelligence,

Yan, Q., Yu, F. R., Gong, Q., And Li, J. (2015). Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges. *IEEE communications surveys & tutorials*, *18*(1), 602-622. https://doi.org/10.1109/COMST.2015.2487361