فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Fall 2024, 5 (3), 47-54
DOI: 10.71856/impcs.2024.903654

Received: 22 Jun 2024
Accepted: 3 Aug 2024

مقاله پژوهشی

Improving the Accuracy of the Intrusion Detection System in the IoT by Machine Learning and Clustering Algorithms

Javad Pashaei Barbin[1*], Mahdi Jalali[2]

1. Assistant Professor, Department of Computer Engineering, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran. **Corresponding Author*, javad.pashaei@iau.ac.ir
2. Assistant Professor, Department of Electrical Engineering, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran.

Abstract

Introduction: The recent rise of the Internet of Things (IoT) has led to increasing attacks in IoT. Manufacturers of IoT devices are interested in reducing costs by ignoring security regulations that cause widespread damage and impede the growth of the IoT. The proliferation of IoT-based attacks will continue as long as IoT manufacturers incorporate accountability and security mechanisms into their devices. The proliferation of IoT-based attacks will continue as long as IoT manufacturers incorporate accountability and security mechanisms into their devices. Until then, the Internet of Things has the potential to become an environment for future cyber-attacks, which will pose great challenges.

Method: In this research, the solutions for establishing security in the Internet of Things have been investigated and have provided a solution based on the combination of support vector machine and K-means algorithm. First, preprocessing is applied to the data set and the data that has no effect on the result are deleted. Then, the support vector machine algorithm is applied to the data set and the intrusion or non-intrusion status is determined. This proposed method achieves better results by applying k-means to the data set, and the combination of support vector machine algorithms and k-means improves the accuracy of the proposed method.

Results: The results showed that the proposed method is more efficient than previous methods. this study sought to improve the security challenge in wireless sensor networks. The proposed method of this research is to use a combination of support vector machine and chi-mean, which showed very good performance compared to previous methods. According to the studies and the proposed method, it can be found that the best method in detecting and detecting intrusion is the use of K-Means algorithm, which can be achieved with 98.35% accuracy using the support vector machine method and K-Means algorithm.

Discussion: The most important criterion for determining the performance of an algorithm is the Accuracy criterion. This criterion calculates the total accuracy of a category. This criterion indicates what percentage of the total data set is properly categorized. This criterion is the evaluation based on the accuracy and the accuracy of the proposed method is better than the previously presented methods.

Keywords: Intrusion detection, Machine Learning, Data Mining, Support vector Machine, K-Means.

فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

بهبود دقت سیستم تشخیص نفوذ در اینترنت اشیا با استفاده از الگوریتم‌های یادگیری ماشین و خوشه‌بندی

تاریخ دریافت: ۰۲ /۰٤/ ۱٤۰۳
تاریخ پذیرش: ۱۳/ ۰۵/ ۱٤۰۳
دورهٔ پنجم، پاییز ۱٤۰۳
شمارهٔ دوم، صص: ٤۷-۵٤

جواد پاشائی باربین[1*]، مهدی جلالی[2]

۱- استادیار، گروه کامپیوتر، واحد نقده، دانشگاه آزاد اسلامی، نقده، ایران. (نویسندهٔ مسئول)
javad.pashaei@iau.ac.ir
۲- استادیار، گروه برق، واحد نقده، دانشگاه آزاد اسلامی، نقده، ایران.

چکیده: افزایش ا ستفاده از اینترنت ا شیا منجر به افزایش حملات در این شبکه‌ها شده‌ا ست. سازندگان د ستگاه‌های اینترنت ا شیا علاقه‌مند به کاهش هزینه‌ها با نادیده‌گرفتن مقررات امنیتی هستند که باعث آسیب گسترده و مانع از رشد اینترنت اشیا می‌شود. گسترش حملات مبتنی بر اینترنت ا شیا تا زمانی ادامه خواهدا شت که سازندگان اینترنت ا شیا مکانیزم‌های پا سخگویی و امنیتی را در دستگاه‌های خود بگنجانند. تا آن زمان، اینترنت اشیا این پتانسیل را دارد که به محیطی برای حملات سایبری آینده تبدیل‌شود که چالش‌های بزرگی را به همراه خواهدداشت. از این رو، در این تحقیق راهکارهای برقراری امنیت اینترنت اشیا بررسی‌شده و راه‌حلی مبتنی بر ترکیب ماشین بردار پشتیبان و الگوریتم k-means ارائه‌شده‌است. نتایج نشان‌می‌دهد که دقت روش پیشنهادی 98.35 درصد است که کارآمدی روش پیشنهادی را نشان‌می‌دهد و قابلیت پیاده‌سازی برای تشخیص خطا به‌صورت عملی را دارد.

واژه‌های کلیدی: تشخیص نفوذ، یادگیری ماشین، داده‌کاوی، ماشین بردار پشتیبان، k-means.

۱. مقدمه

اینترنت اشـیا برای اولین بار توسـط کوین اشـتون در سـال ۱۹۹۹ استفاده‌شد که در آن همه چیز، از جمله اشیاء بی‌جان، هویت دیجیتالی خود را دارد و به رایانه‌ها اجازه‌می‌دهد تا آن‌ها را سازماندهی و مدیریت کنند. در آینده‌ای نه چندان دور، بسیاری از کاربردهای اینترنت اشیا در خانه‌های هوشـمند، کارخانه‌های هوشـمند، مزارع هوشـمند، دفاتر هوشمند، سیستم‌های حمل و نقل هوشمند، بیمارستان‌های هوشمند، دانشـگاه‌های هوشـمند و غیره به هم متصـل‌شـده و توسـط فناوری اطلاعات مورداستفاده قرارخواهندگرفت[۱].

اینترنت اشـیا[۱] (IoT) در حال حاضـر یک الگوی محبوب اسـت که دنیایی را متصـور می‌باشـد که در آن انواع اشـیاء فیزیکی به اینترنت متصل‌شده و قادر به برقراری ارتباط و همکاری با یکدیگر برای رسیدن به اهداف مشـترک هسـتند. این ارتباط فراتر از ارتباطات ماشـین به ماشین[۲] (M2M) است، زیرا تمام داده‌های حسگرها یا فرستنده‌ها از طریق اینترنت یا اینترانت به سـرورها ارسـال‌می‌شـوند و گیرنده‌ها (دا شبورد یا برنامه‌ها) داده‌های مورد نیاز را از سرورها دریافت‌می‌کنند. در ارتباط ماشین به ماشین، دستگاه ها به یکدیگر متصل‌می شوند و عملاً داده‌های زیادی بین دسـتگاه‌ها رد و بدل نمی‌شـود و جایی برای ذخیره داده ها وجود ندارد [۲]. با رشـد روزافزون اینترنت اشـیا و کاربردهای آن در جامعه، اهمیت امنیت اطلاعات به طور فزاینده‌ای به عنوان بخش کلیدی این حوزه مورد توجه قرارگرفته‌اسـت. چالش‌ها در حوزه‌های امنیت و حریم خصـوصـی، از جمله مقاومت در برابر حمله، احراز هویت داده‌ها و کنترل دسترسی حریم خصوصی کاربران، باید در حین مطالعه مورد توجه قرارگیرند، در حالی که محدودیت‌های شـدید در منابع محاسـباتی، قدرت، حافظه مصـرف انرژی، هزینه و ناهمگونی پروتکل‌ها و دستگاه ها نیز غالب هستند [۳]. یکی از رایج‌ترین حملات، حملات انکار سـرویس[۳] (DOS) و حملات انکار سـرویس توزیع شـده[۴] (DDOS) اسـت. تعداد حملات انکار سـرویس در سـال ۲۰۱۵ در سـه ماهه اول سـال نسـبت به سـال قبل ۳۴ درصـد افزایش یافته و تعداد حملات به بیش از ۵ گیگابیت در ثانیه رسـیده‌اسـت. حملات به سـرورهای وب‌سـایت‌های مهم مانند توییتر، آمازون و نیویورک تایمز، باعث توجه بیشتر کارشناسان و متخصصان به این حملات شده‌است [۴].

ظهور اخیر اینترنت اشـیا (IoT) منجر به افزایش حملات DDoS مبتنی بر اینترنت اشـیا شـده و با ظهور اینترنت اشـیا (IoT) پتانسـیل حملات DDoS افزایش یافته‌است. سازندگان دستگاه‌های اینترنت اشیا علاقه‌مند به کاهش هزینه ها با نادیده گرفتن مقررات امنیتی هسـتند که باعث آسیب گسترده و مانع از رشد اینترنت اشیا می شود. افزایش حملات DDoS مبتنی بر اینترنت اشیا، که در سال‌های اخیر شاهد آن بوده‌ایم، احتمالاً تا زمانی ادامه خواهدداشت که سازندگان اینترنت اشیا، مسئولیت و مکانیسم‌های امنیتی را در دستگاه‌های خود بگنجانند. تا آن زمان، اینترنت اشیا پتانسیل تبدیل شدن به محیطی برای حملات سایبری آینده را دارد و چالش‌های بزرگی را ایجادمی‌کند [۵].

یکی دیگر از جنبه‌های مهم امنیت اینترنت اشیا، امنیت تجهیزات در دنیای واقعی است. افزایش عظیم حملات انکار سرویس توزیع شده در اینترنت اشـیا، نگرانی‌های جدی را برای کاربران این شـبکه‌ها به‌وجودآورده‌است. علاوه بر این، محدودیت‌های منابع محدودیت‌هایی را بر اقدامات امنیتی اعمال‌می‌کند که می‌توانند در دستگاه‌های هوشمند گنجانده‌شوند. این نیز یکی از نگرانی‌های جدی در مورد محیط اینترنت اشیا و انگیزه تحقیقات بیشتر در مورد سیستم‌های سبک وزن است [۶].

۲. مطالعات پیشین

سـال ۲۰۱۹، سـهیب حنیف و همکاران روشـی برای تشـخیص نفوذ اینترنت اشیا با استفاده از شبکه‌های عصبی مصنوعی در مجموعه داده UNSW-15 ارائه‌کرده‌اند. دسـتگاه‌های اینترنت اشـیا به‌دلیل قدرت کم، نیازهای محاسـباتی پایین و محیط کنترل‌شـده، با حملات سـایبری زیادی مواجه‌می‌شوند. پیاده‌سازی سیستم تشخیص حمله بسیار دشوار است. این مقاله یک شبکه عصبی مصنوعی را برای شناسایی حملات IoT برای حل مشکلات احراز هویت پیشنهاد می‌کند. شبکه‌های عصبی مصـنوعی شـامل لایه‌های ورودی، خروجی و پنهان هسـتند. تکنیک پیشنهادی قادر به تشخیص موثر حملات است و دقت متوسط ۸۴٪ با میانگین ضریب خطای مثبت ۸٪ را ارائه‌می‌دهد [۷].

سال ۲۰۲۰، رحمان و همکاران یک سیستم تشخیص نفوذ مبتنی بر یادگیری ماشـین برای شـهرهای هوشـمند با اینترنت اشـیا معرفی‌کردند. با رشـد دسـتگاه‌های متصـل، اسـتفاده از سـیسـتم‌های تشخیص نفوذ افزایش یافته‌است. این مقاله محدودیت سیستم تشخیص نفوذ را برای دسـتگاه‌های با منابع محدود با پیشـنهاد دو روش، به نام نیمه توزیع‌شـده و توزیع‌شـده، که اسـتخراج ویژگی و انتخاب را ترکیب‌می‌کنند، توصـیف می‌کند. به منظور توزیع وظایف محاسـباتی، مدل‌های یادگیری ماشـین موازی مبتنی بر مجموعه داده‌های حمله تقسیم‌بندی شده، توسعه داده شده‌اند. بر اساس مقایسه آثار موجود، نتایج عددی نشـان‌می‌دهد که روش پیشـنهادی دارای دقت تشـخیص قابل‌مقایسه با سیستم‌های تشخیص نفوذ بوده و کارایی آن از نظر زمان و دقت بهتر است [۸].

نادیا چابونی و همکاران روشی را با استفاده از یادگیری ماشین لبه برای سـیسـتم‌های تشـخیص و پیشـگیری از نفوذ ارائه دادند. در این مقاله، یک سیستم تشخیص و پیشگیری از نفوذ برای لایه سرویس توسط استاندارد oneM2M معرفی‌شده‌است. این یک سیستم تشخیص نفوذ برای خدمات oneM2M مبتنی بر یادگیری ماشـین لبه اسـت. تمرکز آزمایش‌ها بر کاهش ابعاد ویژگی‌ها در یادگیری ماشـین و اندازه مدل‌های آموزشی است [۹].

سـال ۲۰۱۹، عبدالعزیز الدیاز و همکاران روشـی را با اسـتفاده از الگوریتم‌های پیشگیری از نفوذ برای بهبود امنیت سایبری در اینترنت

اشیا مدرن ارائه‌کردند. اینترنت اشیا یک الگوی به‌سرعت در حال تکامل است که پتانسیل تغییر تعاملات سازمان‌ها و افراد را دارد. این فناوری در زمینه‌های مختلف از جمله سلامت، یادگیری و آموزش، مدیریت منابع و پردازش اطلاعات کاربردهای فراوانی دارد. یک تکنیک پیشگیری برای بهبود امنیت سایبری در دستگاه‌های IoT در برابر حملات DDos پیشنهاد شده‌است. رویکرد پیشنهادی مبتنی بر تجزیه و تحلیل حملات پهن باند است که در درجه اول بر روی DDos تمرکز دارد و عملکرد شبکه را کاهش‌می‌دهد [۱۰].

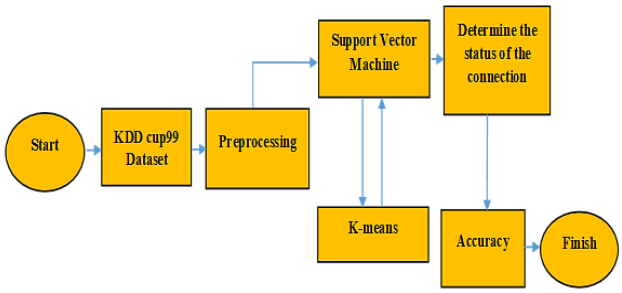
در تحقیقی که اخیراً در زمینه عملکرد سیستم تشخیص نفوذ، به منظور بررسی تأثیر رفتار سیستم تشخیص سوءاستفاده و همچنین تشخیص ناهنجاری با استفاده از منطق فازی مبتنی بر دستگاه آلفا ارائه‌شده‌است نتایج به‌دست‌آمده میزان دقت را تا ۹۱/۲۶ درصد و تشخیص هشدارهای کاذب را تا میزان ۹۰/۹۶ در صد نشان‌داده‌است [۱۱].

سال ۲۰۱۴، نایاکی و کومار یک چارچوب تشخیص DDoS مبتنی بر میزبان به نام BRAIN، با در نظرگرفتن کوتاه‌ترین زمان شناسایی حملات DDOS، پیشنهادکردند. این روش تشخیص DDoS بر اساس مدل تصمیم‌گیری تکامل مصنوعی فازی پیشنهاد شده‌است. آن‌ها همچنین چندین آزمایش را برای نشان‌دادن برتری مدل پیشنهادی خود نسبت به سایر الگوریتم‌های تشخیص DDoS انجام‌داده‌اند. در این مقاله حملات DDoS به سه سطح تقسیم‌می‌شود و که از این سطوح اطلاعات حمله برای شناسایی پارامترهای روش تشخیص DDoS استفاده‌می‌کنند. نتایج نشان‌می‌دهد که افزودن سخت‌افزار برای تشخیص به‌طور قابل‌توجهی دقت را افزایش‌می‌دهد. آن‌ها معتقدند که چارچوب پیشنهادی برای تشخیص DDoS کم هزینه، سازگار و بسیار دقیق با دقت ۹۹,۸ درصد است [۱۲].

سال ۲۰۱۷، چانگ‌لانگ یین و همکاران نحوه مدل سازی سیستم تشخیص نفوذ مبتنی بر یادگیری عمیق را بررسی‌کردند و یک روش یادگیری عمیق برای تشخیص نفوذ با استفاده از شبکه‌های عصبی تکراری پیشنهادکردند. علاوه بر این، آن‌ها عملکرد مدل را در کلاس‌های باینری و چندکلاسه و تعداد نورون‌ها و تأثیرات یادگیری مختلف بر عملکرد مدل پیشنهادی مطالعه‌کرده‌اند. روش کار با شبکه‌های عصبی مصنوعی، جنگل‌های تصادفی، ماشین‌های بردار پشتیبان و سایر روش‌های یادگیری ماشینی پیشنهادشده توسط محققان قبلی مقایسه‌شده‌است. نتایج تجربی نشان‌می‌دهد که مدل پیشنهادی برای مدل سازی یک مدل طبقه‌بندی با دقت بالا منا سب است و عملکرد آن بهتر از روش‌های طبقه‌بندی یادگیری ماشین سنتی می‌باشد [۱۳].

۳. روش پیشنهادی

پس از بررسی روش‌های مورد استفاده در سیستم‌های تشخیص نفوذ که عمدتاً روش‌های یادگیری ماشین هستند، در این بخش روش پیشنهادی که مبتنی‌بر ترکیبی از روش‌های ماشین بردار پشتیبان و k-means در تشخیص نفوذ در اینترنت اشیا توضیح داده‌می شود. (شکل ۱)

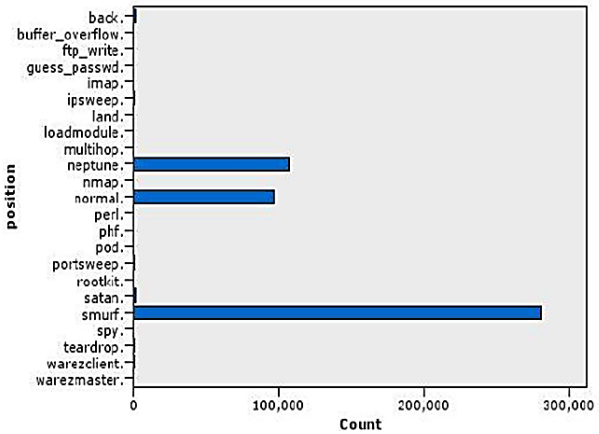


شکل ۱: فلوچارت روش پیشنهادی

ابتدا پیش پردازش روی مجموعه داده اعمال‌شده و داده‌هایی که تاثیری در نتیجه ندارند، حذف‌می‌شوند. سپس الگوریتم ماشین بردار پشتیبان بر روی مجموعه داده اعمال شده و وضعیت نفوذ یا عدم نفوذ مشخص‌می‌شود. این روش پیشنهادی با اعمال k-means به مجموعه داده ها به نتایج بهتری دست‌می‌یابد و ترکیب الگوریتم های ماشین بردار پشتیبان و k-means دقت روش پیشنهادی را بهبود می‌بخشد.

۱,۳. مجموعه داده

با توجه به حجم بالای داده‌ها در پایگاه داده kdd.data_10_percent ، از ۱۰ درصد رکوردها به‌عنوان نمونه آموزشی استفاده‌می شود. این پایگاه شامل ۴۹۴۰۲۱ رکورد ارتباطی است که شامل هر چهار نوع حمله ذکر شده می‌باشد. این چهار دسته از حملات شامل ۲۲ نوع حمله و یک حالت عادی است. علاوه بر این پایگاه داده، یک پایگاه داده آزمایشی نیز وجوددارد که شامل ۳۱۱۰۲۹ رکورد است که نشان‌دهنده ۳۷ نوع حمله و یک حالت عادی است. در نتیجه ۱۵ حمله بیشتر از پایگاه داده آموزشی دارد. شکل‌های ۲ و ۳ نحوه توزیع حملات در پایگاه داده را نشان‌می‌دهد. با توجه به نرخ پایین برخی از حملات نسبت به سایر حملات در پایگاه داده، در صد پراکندگی آن‌ها صفر است. شرح کامل ویژگی‌های مورد استفاده در پایگاه داده در جدول زیر قابل‌مشاهده است.



شکل ۲ : پراکندگی حملات در پایگاه داده

Value	Proportion	%	Count
smurf.		56.84	280790
neptune.		21.7	107201
normal.		19.69	97278
back.		0.45	2203
satan.		0.32	1589
ipsweep.		0.25	1247
portsweep.		0.21	1040
warezclient.		0.21	1020
teardrop.		0.2	979
pod.		0.05	264
nmap.		0.05	231
guess_passwd.		0.01	53
buffer_overflow.		0.01	30
land.		0.0	21
warezmaster.		0.0	20
imap.		0.0	12
rootkit.		0.0	10
loadmodule.		0.0	9
ftp_write.		0.0	8
multihop.		0.0	7
phf.		0.0	4
perl.		0.0	3
spy.		0.0	2

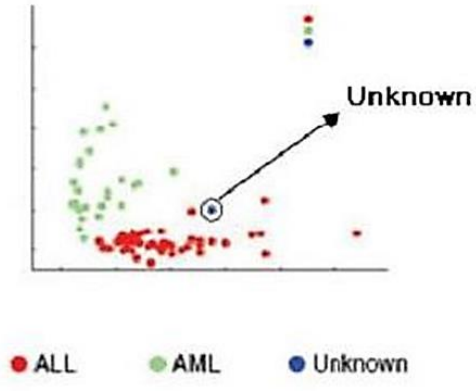
شکل ۳ : نرخ و درصد حملات در پایگاه داده

۲،۳. ماشین بردار پشتیبانی

ماشین بردار پشتیبانی، که به اختصار SVM[۵] نامیده‌می‌شود، یک الگوریتم کامپیوتری است که به‌عنوان مثال نحوه اختصاص برچسب‌های مرتبط به اشیاء مختلف را یادمی‌گیرد. هدف این الگوریتم شناسایی و تمایز الگوهای پیچیده در داده‌ها است که بسته به کاربرد در خوشه‌بندی، طبقه‌بندی، رتبه‌بندی، پاکسازی و غیره استفاده می‌شود. ماشین بردار پشتیبان به‌طور کلی از چهار مفهوم اساسی تشکیل شده‌است: ابر جداکننده، ابر حاشیه حداکثر، حاشیه نرم و عملکرد هسته.

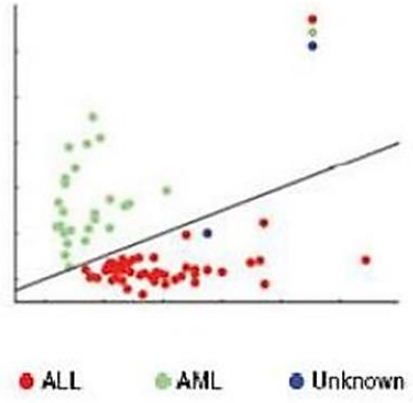
۱،۲،۳. هایپرپلان جداکننده

هایپرپلان اساساً اصطلاحی است که فضای بالای سه بعدی را پوشش می‌دهد. در فضای یک بعدی هایپرپلان فقط یک نقطه، در فضای دوبعدی یک خط و در فضای سه‌بعدی یک صفحه و در فضای بیش از سه‌بعدی یک ابر صفحه است. اما برای راحتی، همه اینها هایپرپلان نامیده‌می شود. شکل ۴ را در نظر بگیرید، در این شکل در کلاس‌های ALL و AML فضای دوبعدی با ناحیه مربوط به ALL در پایین شکل و AML در بالای شکل داریم. همچنین یک نقطه رنگی در نزدیکی کلاس ALL وجوددارد که کلاس آن مشخص نیست و باید طبقه‌بندی‌شود.



شکل ۴ : داده‌ها در فضای دوبعدی

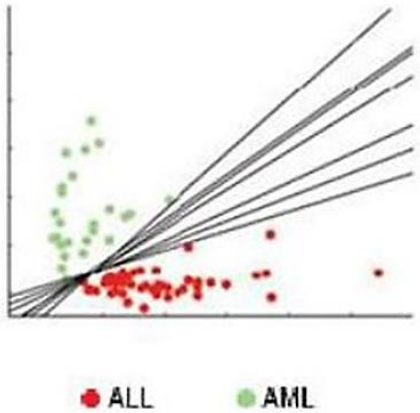
این دو کلاس را می‌توان با یک خط از هم جدا کرد که در شکل ۵ دیده‌می شود و بر این اساس تکلیف نقطه مجهول نیز مشخص‌می شود. این نقطه بخشی از کلاس ALL است.



شکل ۵ : طبقه‌بندی خطی در فضای دوبعدی

۲،۲،۳. هایپرپلان با حداکثر حاشیه

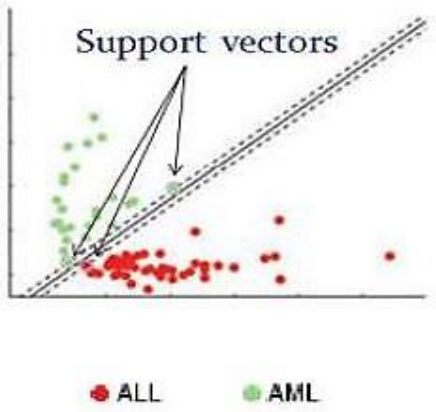
هدف ماشین بردار پشتیبان جداسازی دو کلاس ALL و AML در فضای دوبعدی با یک خط است. خطوط زیادی وجوددارد که این کار را انجام‌می‌دهند. شکل ۶ نمای این خطوط را نشان‌می‌دهد.



شکل ۶ : خطوط متمایزکننده دو کلاس

سوال این است که کدام یک از این خطوط بهتر از خطوط جداکننده دیگر است؟ الگوریتم SVM خط وسط را به‌عنوان ابر صفحه جداکننده

انتخاب‌می‌کند. به عبارت دیگر، خطی را انتخاب‌می‌کند که حداکثر فاصله را از هر کلاس داشته‌باشد. این خط در شکل ۷ مشاهده‌می‌شود.



شکل ۷ : خطوط متمایزکننده دو کلاس با استفاده از بردار پشتیبان

۴. الگوریتم k-means

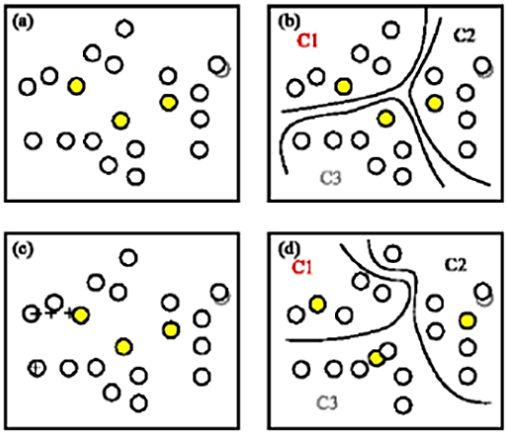
مراحل الگوریتم k-means به صورت زیر است:

۱- ابتدا K نقطه به عنوان نقاط مراکز خوشه انتخاب‌می‌شوند.

۲. هر نمونه داده به خوشـــه‌ای اختصـــاص‌داده‌می‌شـــود که مرکز آن کمترین فاصله را با آن داده دارد.

۳- سپس میانگین را محا سبه‌کرده و یک مرکز جدید برای هر خو شه قرار داده‌می‌شود.

۴- مراحل ۲ و ۳ تکرار می‌شـــود تا زمانی که دیگر تغییراتی در مراکز خوشه‌ها ایجادنشود.



شکل ۸ : فرآیند خوشه‌بندی با استفاده از k-means الف) تعیین مرکزیت اولیه ب) اختصاص نمونه‌ها ج) محاسبه میانگین خوشه‌ها د) خوشه‌های جدید

باتو جه به م عایبی که این الگوریتم دارد می‌توا ند در برخی از کاربردها پا سخ قابل‌قبولی ارائه‌دهد. این معایب عبارتند از: وابـ ستگی به مراکز اولیه، وابستگی به تعداد خوشه‌ها و علف‌های هرز.

۵. نتایج و بحث

ابتدا مشخصات سیستم کامپیوتری که شبیه‌سازی با آن انجام‌شده‌است معرفی و مجموعه داده‌ها شرح داده‌می‌شود، سپس پیاده‌سازی توصیف-شده و در نهایت ارزیابی انجام کار ارائه‌می‌گردد.

۱،۵. مشخصات سیستم

این پروژه بر روی سیستمی با مشخصات زیر انجام‌شده‌است:

CPU: سری Corei5 اینتل

ظرفیت حافظه رم: ۴ گیگابایت

ظرفیت هارد: ۵۰۰ گیگابایت

۲،۵. معرفی مجموعه داده

در سال‌های ۱۹۹۸ و ۱۹۹۹، مجموعه استانداردی از داده‌ها برای مقایسه و ارزیابی سیستم‌های تشخیص نفوذ، از جمله انواع نفوذها در یک سیستم شبیه‌سازی‌شده نظامی توسعه‌یافت. داده‌های این پایگاه داده از داده‌های خام TCP در شبکه LAN نیروی هوایی ایالات متحده در طول ۹ هفته استخراج‌شده‌است. پایگاه داده به‌دست‌آمده بسیار بزرگ بوده و شامل یک بخش داده‌های آموزشی و یک بخش داده‌های آزمایشی است. بخش داده‌های آموزشی شامل تقریباً ۵ میلیون رکورد داده و بخش داده‌های آزمون شامل حدود ۲ میلیون رکورد داده‌است. در عمل ۱۰٪ از داده‌ها برای سیستم‌های تشخیص نفوذ استفاده می‌شود و این پایگاه داده به صورت (kdd (data_10_percent نمایش داده‌می‌شود. این پایگاه دارای ۵ نوع داده مختلف شامل ۴ نوع حمله و یک سری داده معمولی می‌باشد. ۴ نوع حمله نیز خود شامل زیرمجموعه‌ای از انواع مختلف حملات هستند. ده درصد از پایگاه داده خود حدود پانصد هزار داده است و تعداد زیادی است که معمولاً در سیستم‌های تشخیص همه این داده ها استفاده‌نمی‌شود. زیرا هر یک از رکوردهای داده دارای ۴۱ ویژگی مختلف است. اگر بخواهیم از کل این ده درصد پای‌گاه داده استفاده‌کنیم، آموزش سیستم تشخیص نفوذ زمان زیادی می‌برد. یکی از کارهای انجام-شده در این پروژه، انتخاب مجموعه‌ای مناسب از این پایگاه داده است که حجم آن زیاد نباشد و شامل انواع مختلف حملات و به تعداد مناسب باشد.

این تحقیق با استفاده از زبان برنامه‌نویسی متلب و ارزیابی‌ها بر روی پایگاه داده KDD 99 CUP انجام‌شده‌است. پایگاه داده مورد استفاده در این تحقیق پایگاه داده‌ای است که سال ۱۹۹۹ توسط دارپا ارائه‌شده‌است و یکی از معتبرترین پایگاه‌های داده برای آزمایش سیستم‌های تشخیص نفوذ شناخته‌می‌شود. این پایگاه داده که KDDCup99 نام‌دارد توسط آزمایشگاه MIT لینکلن در دسترس محققان قرارگرفته‌است.

۳،۵. پیاده‌سازی

در قسمت اول، کدهای مربوط به خواندن و بارگذاری پایگاه داده در محیط متلب انجام‌می‌شود و سپس اطلاعات از پایگاه داده خوانده‌شده و به فرمت قابل‌فهم محیط متلب تبدیل‌می‌شود و در نهایت در یک فایل ذخیره‌می‌شود. در این فایل کد مربوط به حملات به‌صورت زیر تعریف شده‌است:

کد حمله DOS شماره ۱ و حمله U2R شماره ۲، حمله R2L شماره ۳، حمله PROB شماره ۴ و در نهایت وضعیت غیرحمله یا NORMAL با شماره ۵ مشخص‌می شود. ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای آزمایش انتخاب‌می‌شـــوند. پس از اجرا تعداد داده‌های آموزشـــی ۳۴۵۸۱۴ و تعداد داده‌های آزمایش ۱۴۸۲۰۷ می‌باشـد. حملات نیز از ۱ تا ۵ دسته بندی‌می شوند. انتخاب ویژگی توسط الگوریتم K-means انجام و نتایج وارد SVM می‌شود. از تابع Confusion یا تابع خالی بودن و سردرگمی برای تشخیص صحیح نتایج استفاده شده‌است. سپس با استفاده از داده‌های آموزشی مدل مورد نظر را آموزش‌می‌دهیم و پس از آموزش کامل مدل را با اســتفاده از داده‌های آزمون ارزیابی‌می‌کنیم. در این بخش، الگوریتم پیشنهادی از نظر پارامترهای نرخ تشخیص، دقت و نرخ هشدار نادرست با الگوریتم‌های موجود مقایسه‌می شود. معیارهای مختلف مورد اسـتفاده برای این منظور، بررسـی عملکرد و نتایج تجربی به شرح زیر است:

۱- نرخ تشخیص (DR)

۲- نرخ هشدار نادرست (FAR)

۳- مثبت کاذب (FP)

۴- منفی کاذب (FN)

۵. مثبت واقعی (TP)

۶- منفی واقعی (TN)

معیارهای ارزیابی به‌صورت زیر محاسبه‌می‌گردد:

$$Detection\ Rate = \frac{TP}{TP + FP} \quad (۱)$$

$$False\ Alarm\ Rate = \frac{FP}{TN + FP} \quad (۲)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۳)$$

۴.۵. ارزیابی روش پیشنهادی

مهمترین معیار برای تعیین عملکرد یک الگوریتم، معیار دقت است. این معیار، دقت کلی یک دسته را محاسبه‌می‌کند و نشان‌می‌دهد که چند درصد از کل مجموعه داده‌ها به‌درستی طبقه‌بندی‌شده‌است. معیار ارزیابی بر اساس دقت نشان‌داده‌شده در رابطه (۳) است و دقت روش پیشنهادی بهتر از روش‌های ارائه‌شده قبلی است.

جدول ۱: ارزیابی روش پیشنهادی بر اساس دقت

Method	Accuracy
GWOSVM (7 Wolves) [14]	96%
PSO [14]	89%
Proposed method	98.35%

جدول ۲: ارزیابی روش پیشنهادی بر اساس هشدار کاذب

Method	False Alarm
GWOSVM (7 Wolves) [14]	3%
PSO [14]	26%
Proposed method	1.47 %

جدول ۳: ارزیابی روش پیشنهادی بر اساس میزان تشخیص

Method	Detection Rate
GWOSVM (7 Wolves) [14]	96%
PSO [14]	93%
Proposed method	99.12%

۶. نتیجه

استفاده از اینترنت اشیا و شبکه‌های حسگر بی‌سیم به دلیل عملکرد خوب آن در مکان‌های مختلف مانند خانه، شهر و ... روزبه‌روز در حال افزایش است، اما این شبکه‌ها از نظر امنیتی با چالش مواجه هستند و هر لحظه ممکن است حملاتی رخ‌دهد. برای این منظور، این مطالعه به‌دنبال بهبود چالش امنیتی در شبکه‌های حسگر بی‌سیم است. روش پیشنهادی این تحقیق استفاده از ترکیب ماشین بردار پشتیبان و k-mean می‌باشد که نسبت به روش‌های قبلی عملکرد خوبی دارد. نتایج نشان‌می‌دهد که دقت روش پیشنهادی 98.35 درصد است. می‌توان سیستم پیشنهادی را با توجه به اینکه از مجموعه داده‌های واقعی در محیط‌های عملیاتی موجود برای تشخیص خطاها استفاده‌می‌کند، عملاً پیاده‌سازی‌کرد.

References

[1] M. Bauer, and J.W. Walewski, "The IoT Architectural Reference Model as Enabler, in Enabling Things to Talk," 2013, *Springer*. p. 17-25.

[2] E. De Coninck, , et al. , "Distributed neural networks for Internet of Things," the Big-Little approach. in International Internet of Things Summit. 2015. Springer.

[3] J.Cañedo, and A. Skjellum. , "Using machine learning to secure IoT systems. in Privacy, Security and Trust (PST)," 2016 14th Annual Conference on. 2016. IEEE.

[4] S. Mansfield-Devine, "The growth and evolution of DDoS" Network Security, 2015. 2015(10): p. 13-20.

[5] M. Nazarpour, N. Nezafati, , S. Shokouhyar, 'Using the Modified Colonial Competition Algorithm to Increase the Speed and Accuracy of the Intelligent Intrusion Detection System', Intelligent Multimedia Processing and Communication Systems (IMPCS),2023, 4(1), pp. 1-10. [Persian].

[6] V. Adat, and B. Gupta, "Security in Internet of Things: issues, challenges, taxonomy, and architecture" Telecommunication Systems, 2018. 67(3): p. 423-441.

[7] S. Hanif, T. Ilyas and M. Zeeshan, "Intrusion Detection In IoT Using Artificial Neural Networks On UNSW-15 Dataset," 2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT), 2019, pp. 152-156, doi: 10.1109/HONET.2019.8908122.

[8] M. A. Rahman, et al, "Scalable machine learning-based intrusion detection system for IoT-enabled smart cities," Sustainable Cities and Society 61: 102324.

[9] N. Chaabouni, M. Mosbah, A. Zemmari and C. Sauvignac, "A OneM2M Intrusion Detection and Prevention System based on Edge Machine Learning," NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, 2020, pp. 1-7, doi: 10.1109/NOMS47738.2020.9110473.

[10] A. Aldaej, "Enhancing Cyber Security in Modern Internet of things (IoT) Using Intrusion Prevention Algorithm for IoT (IPAI)," in IEEE Access, doi: 10.1109/ACCESS.2019.2893445.

[11] M. Akhlaghpour, 'Providing a Solution Based on Fuzzy Logic to Reduce False Positive Alarms in The Intrusion Detection System', Intelligent Multimedia Processing and Communication Systems (IMPCS), 2(4),2021, pp. 45-50. [Persian].

[12] R.S. Nayaki, and A.S. Kumar, “An Analysis of DDoS Attack Detection and Mitigation Using Machine Learning System,” International Journal on Recent and Innovation Trends in Computing and Communication, 2017. 5(10): p. 80-82.
[13] C. Yin, et al., “A deep learning approach for intrusion detection using recurrent neural networks;” IEEE Access, 2017. 5: p. 21954-21961.
[14] M. Safaldin, M. Otair, and L. Abualigah, “Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks,” J Ambient Intell Human Comput 12, 1559–1576 (2021). https://doi.org/10.1007/s12652-020-02228-z

پی‌نوشت

[1] Internet Of Things
[2] Machine to Machine
[3] Denial Of Service
[4] Distributed Denial Of Service
[5] Support Vector Machine