

The Presentation of a Hybrid Anomaly Detection Model Using Inverse Weight Clustering and Machine Learning in Cloud Environments

A. Jafar Gholi Beik¹, M. E. Shiri Ahmad Abadi^{2*}, A. Rezakhani³

*Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran

(Received: 28/10/2020, Accepted: 15/08/2021)

ABSTRACT

Today, due to highly advanced attacks and intrusions, it has become very difficult to detect IoT attacks in cloud environments. Other problems with cloud systems include low error detection accuracy, false positive rates, and long computation times. In the proposed method, we present a hybrid intrusion detection model including a clustering algorithm and a machine-based random forest classification for the fog and cloud environments. Also, to control the network traffic in the physical layer and also to detect the anomalies between IoT devices, calculations are performed on the fog and the edges of the cloud, so that after preprocessing, the incoming traffic to the fog and cloud are checked and if necessary, they are directed to an anomaly detection module. A random forest-based learning classification is used to identify the type of each attack. Both the general and cloud data have been used for this research. The overall accuracy, the mean false positive rate and the anomaly detection rate of the proposed intrusion detection system are 98.03, 17% and 96.30 respectively, which is notable in comparison to previous methods .

Keywords: IDS, Cloud Computing, Fog Computing, Anomaly Detection, IoT

*Corresponding Author Email: shiri@aut.ac.ir

مدل ترکیبی تشخیص ناهنجاری با استفاده از خوشه‌بندی وزنی معکوس و یادگیری ماشین در بستر محیط‌های ابری

عادلہ جعفر قلی بیک^۱، محمد ابراهیم شیری احمدآبادی^{۲*}، افشین رضاخانی^۳

۱- مربی، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد بروجرد، بروجرد، ایران، ۲- دانشجوی دکتری، گروه کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران،

۳- دانشجوی دکتری، گروه کامپیوتر، دانشگاه آیت‌الله بروجردی، بروجرد، ایران

(دریافت: ۱۳۹۹/۰۸/۰۷، پذیرش: ۱۴۰۰/۰۵/۲۴)

چکیده

امروزه به دلیل حملات و نفوذهای بسیار پیشرفته، شناسایی حملات در اینترنت اشیاء در بستر محیط‌های ابری بسیار دشوار شده است. از مشکلات دیگر سیستم‌های ابری می‌توان به پایین بودن دقت در تشخیص خطا، نرخ مثبت کاذب و زمان محاسبات طولانی اشاره کرد. در روش پیشنهادی یک مدل تشخیص نفوذ ترکیبی شامل یک الگوریتم خوشه‌بندی و یک طبقه‌بندی جنگل تصادفی مبتنی بر ماشین، برای محیط‌های ترکیبی مه و ابر ارائه می‌دهیم. همچنین برای کنترل ترافیک شبکه در لایه فیزیکی و همچنین تشخیص ناهنجاری در بین دستگاه‌های اینترنت اشیاء محاسبات در مه و لبه‌های ابر انجام خواهد شد به این صورت که پس از پیش پردازش، ترافیک ورودی به مه و ابر بررسی و در صورت نیاز به یک ماژول تشخیص ناهنجاری هدایت می‌شوند. برای شناسایی نوع هر حمله از یک طبقه‌بندی یادگیری مبتنی بر جنگل تصادفی استفاده شده است. از داده‌های عمومی و داده‌های ابری برای تحقیق استفاده شده است. دقت کلی سیستم تشخیص نفوذ پیشنهادی ۹۸/۰۳ و متوسط نرخ مثبت کاذب ۱۷٪ و نرخ تشخیص ناهنجاری ۹۶/۳۰ بوده است که نسبت به روش‌های گذشته قابل ملاحظه است.

کلیدواژه‌ها: سیستم تشخیص نفوذ، محاسبات ابری، محاسبات مه، تشخیص ناهنجاری، اینترنت اشیاء.

۱- مقدمه

حمله‌ها به مغلوب کردن سیستم‌های تشخیص نفوذ در محیط‌های ابری ادامه می‌دهند [۲]. حملات منع سرویس توزیع شده^۴ برای درهم شکستن توانایی تامین کننده ابر در جهت جذب حملات مبتنی بر سرور، و مهار و کنترل منابع شبکه‌سازی و محاسباتی مرکز داده به منظور طبقه‌بندی حملات منع سرویس توزیع شده با حجم‌های غیر منتظره به اندازه کافی بزرگ می‌شوند. به خاطر افزایش حجم حمله، آسیب پذیر بودن به یک دلیل اصلی نگرانی تبدیل شده است. از بین رفتن بسته، تاخیر بالا برای ترافیک اینترنتی آنهایی که ترافیک شبکه برای پیمایش شبکه‌های اشباع شده با حمله منع سرویس توزیع شده به وقوع می‌پیوندد. از حملات منع سرویس توزیع شده نه تنها برای مختل نمودن و وقفه در سرویس‌ها، بلکه همچنین برای منحرف نمودن منابع امنیتی استفاده می‌شود در حالی که انواع دیگر حملات مثل تراکنش‌های کلاهبردارانه و متقلبانه شروع می‌شوند. ابزارهای نفوذ و حمله پیچیده‌تر شده‌اند و سیستم‌های تشخیص نفوذ ابری را با حجم عظیمی از داده‌های ترافیک شبکه، رفتارهای پویا و پیچیده و انواع جدیدی از حمله‌ها به چالش می‌کشند [۳]. واضح است که یک سیستم تشخیص نفوذ برای ابر باید حجم

ادغام ابر با اینترنت اشیاء^۱، که به نام ابر اشیاء شناخته می‌شود، دارای مزیت‌هایی می‌باشد. به‌عنوان مثال، کمک می‌کند تا روند آماده‌سازی مدیریت منابع اینترنت اشیاء و خدمات اینترنت اشیاء با هزینه‌ای معقول‌تر و کارآمدی بیشتری انجام شود. علاوه بر این، جریان پردازش داده‌های اینترنت اشیاء را ساده کرده و نصب و ادغامی سریع و کم هزینه را فراهم می‌کند [۱]. مشخصات عملیات ابری از قبیل پرداخت به ازای مصرف و بر حسب تقاضا بودن آنها، در حال تغییر مدل محاسباتی شرکت‌ها بوده و در حال جایگزین نمودن زیرساخت‌های درون‌سازمانی^۲ به مرکز داده‌های برون‌سازمانی^۳ است، مراکزی که از طریق اینترنت قابل دسترسی هستند و توسط ارائه‌دهندگان میزبانی ابر مدیریت می‌شوند. با این حال، مسائل امنیتی زیادی از جمله تشخیص نفوذ در انتقال به این الگوی محاسباتی ایجاد می‌شوند. صرف نظر از پیشرفت مهم فناوری‌های امنیت اطلاعات در سال‌های اخیر، نفوذها و

* رایانامه نویسنده مسئول: shiri@aut.ac.ir

¹ Intern of Thing

² On-premises

³ Off-premises

⁴ Distributed Denial-of-Service(DDoS)



k-means است که بر اساس شباهت بین رکوردها عملیات خوشه‌بندی را انجام می‌دهد. پس از پایان یافتن الگوریتم k-means در روش پیشنهادی خوشه‌بندی وزنی معکوس [۷] با ایجاد سنتروییدهای جدید مجدداً خوشه‌بندی انجام می‌شود. این روش مشکلات احتمالی روش قبلی برای خوشه‌بندی و انتخاب داده اولیه اشتباه را ندارد و دقت بسیار بالایی دارد. در واقع داده‌ها پس از خوشه‌بندی توسط یک الگوریتم یادگیری ماشین کلاس‌بندی می‌شوند که پس از آزمایش و آزمایش روش‌های متعدد، روش جنگل تصادفی انتخاب شده است و داده‌ها به دو دسته عادی و غیر عادی دسته‌بندی می‌شوند. شناسایی حملات به ویژه حملات منع سرویس^۱ و منع سرویس توزیع شده و تشخیص نفوذ در لایه مه انجام می‌پذیرد. برای غلبه به مشکلات محیط‌های ابری در تاخیر زمان پاسخگویی [۸] و صرفه‌جویی در هزینه پهنای باند و همچنین ایمنی و حریم خصوصی و افزایش امنیت محاسبات در لبه ابر پیشنهاد شده است [۹]. یک معماری به نام پردازش مه در سطح سیستم جهت توزیع، محاسبه و ذخیره‌سازی و کنترل شبکه قبل از ورود به ابر فراهم شده است [۱۰]. در ادامه، بخش ۲ مروری بر کارهای مرتبط را ارائه می‌دهد. بخش ۳ روش پیشنهادی و مجموعه داده‌های استفاده شده در این روش توضیح داده خواهد شد. در بخش ۴ نتایج آزمایشگاهی و مقایسه روش‌ها انجام خواهد شد و همچنین جزئیات آزمایش‌های انجام شده و معیارهای عملکرد استفاده شده برای ارزیابی سیستم تشخیص نفوذ پیشنهادی را ارائه می‌دهد. در بخش ۵ نتیجه‌گیری و موضوعات مهم برای آینده تحقیق ارائه خواهد شد.

۲- کارهای مرتبط

این بخش به معرفی کارهای پیشینی می‌پردازد که برای بهبود عملکرد تشخیص نفوذ در محیط‌های ابری و سایر حملات و ناهنجاری اختصاص داده شده‌اند. میرزایی و همکارانش [۱۱] یک الگوریتم مبتنی بر تشخیص و ترکیب انجمن‌های مشابه را برای شناسایی گره‌های ناهنجار را انجام داده‌اند. شوشیان و همکارانش [۱۲] یک مدل طبقه‌بندی جدیدی در روش‌های مهم‌سازی برای مدل‌سازی به حملات سایبری مهم، روشی مبتنی بر فن جایگزین حمله‌ارایه داده‌اند یادگیری و همکارانش [۱۳] حملات منع سرویس وب با استفاده از آنروپی و الگوریتم ماشین بردار پشتیبان. حملات منع سرویس لایه وب و یا کاربردی از طریق ایجاد مصنوعی حجم زیاد ترافیک بر روی وب سرور تولید و باعث اختلال در سرویس‌دهی وب می‌گردد. در این تحقیق

عظیمی از داده‌های ترافیک شبکه را تحلیل کند، رفتار حمله‌های جدید را به خوبی تشخیص داده و به دقت بالایی با کمترین اشتباه دست‌یابد. با این حال پیش‌پردازش، تحلیل و تشخیص نفوذها در محیط‌های ابری با استفاده از روش‌های رایج و سنتی از لحاظ محاسباتی، صرف وقت و بودجه بسیار پرهزینه است. بنابراین، تشخیص نفوذهای کارآمد در محیط‌های ابری نیاز به استفاده از روش‌های جدید توزیع شده و هوشمند از قبیل روش‌های یادگیری ماشین دارند [۴]. ما در این مقاله، یک سیستم تشخیص نفوذ ترکیبی مبتنی بر خوشه‌بندی و دسته‌بندی بر اساس الگوریتم‌های یادگیری ماشین برای محیط‌های ابری ارائه می‌دهیم. هدف مقاله تشخیص ناهنجاری [۵] در بین دستگاه‌های اینترنت اشیا است. از طرف دیگر با ادغام مه و ابر و قرار گرفتن نودهای مه در لبه‌های ابر می‌توان علاوه بر کنترل ترافیک ورودی به مه، امنیت و کاهش زمان پاسخگویی را به همراه داشت. ابتدا داده‌های جمع‌آوری شده توسط دستگاه‌های اینترنت اشیا شامل خانه‌های هوشمند، شهرهای هوشمند و دوربین‌های هوشمند به نودهای لایه مه فرستاده می‌شوند. عملیات پیش‌پردازش و کاهش ابعاد در نودهای مه انجام می‌گیرد. در مرحله بعدی داده‌های پیش‌پردازش شده جهت خوشه‌بندی و تشخیص ناهنجاری مورد بررسی قرار می‌گیرند. توسط روش خوشه‌بندی وزنی معکوس (Inverted Weighted Cluster) که توسعه یافته روش خوشه‌بندی k-means [۶] است داده‌ها به ۵ خوشه تقسیم می‌شوند و سپس تابع توزیع عادی بررسی و تشخیص ناهنجاری در نودهای مه انجام می‌گیرد. خروجی تابع توزیع عادی دو حالت است یا داده‌ها عادی هستند یا غیر عادی و حمله می‌باشند. تشخیص نوع حملات توسط روش جنگل تصادفی انجام می‌گیرد. در مجموعه داده مخصوص اینترنت اشیا [۶] ابتدا کاهش ابعاد انجام می‌گردد و سپس خوشه‌بندی و بر چسب‌گذاری می‌شوند و سپس توسط یک دسته‌بند تک کلاسه دقت روش جهت تشخیص ناهنجاری بررسی می‌شود. در معماری توزیع شده پیشنهادی، ابتدا ترافیک ورودی به نودهای مه با استفاده از الگوریتم پنجره لغزان مبتنی بر زمان بر روی هر مسیریاب در لبه مه بررسی و سپس به یک ماژول تشخیص ناهنجاری هدایت می‌شود. در هر پنجره زمانی، داده‌های ناهنجار ترافیک شبکه بر روی هر مسیریاب با یک سرور ذخیره مرکزی هماهنگ می‌شوند. سپس، یک دسته‌بند یادگیری مبتنی بر جنگل تصادفی به منظور تشخیص نوع هر حمله مورد استفاده قرار می‌گیرد. در این روش علاوه بر شناسایی حالت سیستم که یا عادی است و یا عادی نیست قابلیت تشخیص و نوع هر حمله نیز قابل تشخیص است. سیستم پیشنهادی در بستر مه و ابر و محاسبات در مه و لبه‌های ابر انجام می‌شود. روش خوشه‌بندی وزنی معکوس که پیشرفته روش خوشه‌بندی

^۱ DoS (Denial of Service)

و طبقه‌بندی شده است و بر روی زمان بندی آن تحقیق شده است.

۳- روش پیشنهادی

هدف از مدل، تشخیص و شناسایی ناهنجاری در داده‌هایی است که در لایه کاربرد معماری اینترنت اشیا با دستگاه‌های اینترنت اشیا مبادله می‌شود و کنترل ترافیک در لایه فیزیکی است. به منظور انجام این کار، مدل نیازمند طراحی فرایندی است که داده‌ها را از دستگاه‌های اینترنت اشیا به‌عنوان ورودی بگیرد، داده‌های ورودی را به‌طور سیستماتیک قبل از ورود به لایه کاربرد به واسطه رایانش مه و ابر پردازش کند و دو دسته داده "عادی" یا "غیر طبیعی" پیش‌بینی کند. همچنین طی فرایندی نوع حملات قابل شناسایی باشند. نوآوری روش پیشنهادی، خوشه‌بندی دقیق داده‌های ورودی با استفاده از روش خوشه‌بندی وزنی معکوس و استفاده از جنگل تصادفی جهت تشخیص ناهنجاری و کنترل ترافیک شبکه و احراز هویت در مه و لایه‌های ابر با استفاده از روش‌های یادگیری ماشین است. شناسایی و تشخیص نفوذ و حمله در مه انجام می‌شود. جهت بهبود امنیت و افزایش دقت و سرعت پاسخگویی، نودهای مه در لایه‌های ابر مستقر شده‌اند. شناسایی و تشخیص نفوذ در لایه مه و قبل از ورود به ابر انجام و نتیجه به ابر ارسال می‌گردد. تشخیص حملات به ویژه حملات منع سرویس و منع سرویس توزیع شده با دقت بالا و کاهش نرخ مثبت کاذب و پاسخگویی سریع از نوآوری روش پیشنهادی است. فرآیند تشخیص ناهنجاری در مه فرایندی است که مسئول دسته‌بندی (خوشه‌بندی)، طبقه‌بندی و پیش‌بینی است که شامل سه مرحله می‌باشد:

- **پیش پردازش داده‌های ورودی:** این مرحله مربوط به قالب بندی داده‌های ورودی به شیوه‌ای که پردازش آن آسان‌تر باشد کاهش تعداد ویژگی‌ها و انتخاب بهترین ویژگی‌ها باعث افزایش دقت و کاهش زمان محاسبه و مدل‌سازی خواهد شد.

- **پردازش:** این مرحله مربوط به خوشه‌بندی داده‌های ورودی به ۵ خوشه است. یک خوشه عادی و ۴ خوشه حمله است با استفاده از روش خوشه‌بندی وزنی معکوس و استفاده از تابع توزیع عادی و روش خوشه‌بندی تک‌کلاسی بردار ماشین پشتیبان^۴.

- **پیش‌بینی‌ها:** این مرحله مربوط به استفاده از جنگل تصادفی ساخته شده برای پیش‌بینی نوع حملات است.

برای شناسایی این دسته از حملات، لاگ‌های وب سرور با ایجاد پنجره‌های زمانی ۲۰ ثانیه‌ای و محاسبه میزان فعالیت هر آی‌پی دسته‌بندی گردیده و سپس آن‌تروپی مربوط به هر آی‌پی در پنجره زمانی محاسبه و از طریق واریانس آن‌تروپی پنجره‌های زمانی دارای پیوستگی تعیین و در مرحله بعد از طریق الگوریتم ماشین بردار پشتیبان، شبکه آموزش داده می‌شود تا پنجره‌های زمانی ناهنجار و درنهایت آی‌پی‌های که منجر به حملات منع سرویس و یا منع سرویس توزیع شده‌اند. به‌طور مشابه، مدی و همکارانش [۱۴] از سیستم تشخیص نفوذ اسنورت و دسته‌بندی یادگیری ماشین برای تشخیص ناهنجاری‌های موجود در ترافیک شبکه بین ماشین‌های مجازی استفاده کرده‌اند. چارچوب پیشنهادی آنها، یک سیستم تشخیص نفوذ شبکه^۱ را با زیرساخت ابر ادغام می‌کند. این چارچوب بر اساس اسنورت و دسته‌بند درخت تصمیم^۲ است. هدف اصلی این چارچوب پیشنهادی، تشخیص حمله‌های شبکه در ابر با نظارت بر ترافیک شبکه است. نویسندگان از مجموعه‌داده‌گان NSL-KDD و KDD برای ارزیابی سیستم خود استفاده نموده‌اند [۱۵]. ویسینگ و همکاران [۱۶] روشی را پیشنهاد کرد که دو آشکارساز را با هم ترکیب می‌کند. یک ردیاب ویژگی و یک ردیاب آماری. ردیاب ویژگی با استفاده از اسنورت رویدادها را بر اساس پروتکل‌های شبکه جدا می‌کند آشکارساز آماری با استفاده از بسته‌های داده از آنجا با ردیاب ویژگی همکاری می‌کند تعیین کنید که آیا یک رویداد یک حمله است یا خیر. اگر میزان بسته‌های بدست آمده از آستانه قبلی بیشتر باشد، این مورد به‌عنوان حمله در نظر گرفته می‌شود. رضایت‌بخشی در تشخیص حمله‌ها به‌دست آمده است. ادهمد و دوستان [۱۷] یک روش تشخیص با نظارت برای حمله منع سرویس بر اساس شبکه عصبی پیشخور پیشنهاد کرده‌اند. این روش شامل سه مرحله اصلی است: (۱) جمع‌آوری ترافیک ورودی شبکه، (۲) انتخاب ویژگی‌های مربوط به تشخیص حمله منع سرویس با استفاده از یک روش انتخاب ویژگی مبتنی بر همبستگی بدون نظارت^۳، (۳) دسته‌بندی ترافیک ورودی شبکه به دو دسته ترافیک عادی یا ترافیک منع سرویس. رویکرد پیشنهادی آنها به عملکرد خوبی بر روی مجموعه‌داده‌گان UNSW-NB15 و NSL-KDD دست یافته است. در تحقیق لویی [۱۸] و همکارانش در پردازش داده در مقیاس بزرگ به‌طور فزاینده‌ای از چارچوب‌های نگاشت کاهش در بستر رایانش ابری استفاده می‌کند. در این نگاشت کاهش بهبود یافته‌اند و همچنین روش‌های مربوط به نگاشت کاهش و هادوپ به‌طور خلاصه شرح داده شده است. سپس انواع زمان‌بندی‌های نگاشت کاهش بررسی

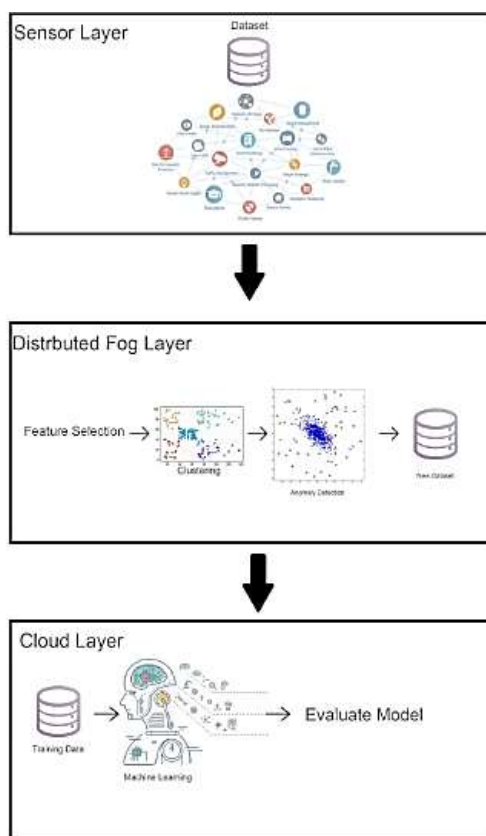
¹ Network Intrusion Detection System (NIDS)

² Decision Tree

³ Unsupervised Correlation-Based Feature Selection (CFS)

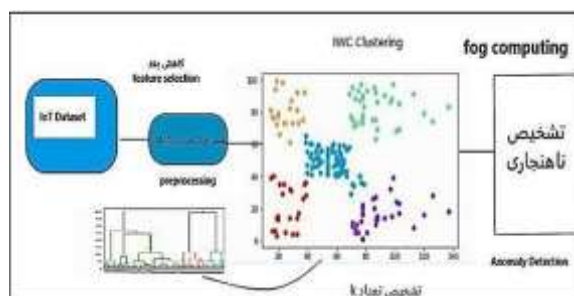
⁴ Support Vector Machine

تصمیمات کلان در لایه ابر انجام خواهد گرفت. این روش برای لایه کاربرد اینترنت اشیا که نیاز به پاسخگویی سریع و بدون وقفه دارد بسیار کارآمد است. از طرفی ناهنجاری‌های بین دستگاه‌های اینترنت اشیا به خوبی قابل تشخیص است.



شکل (۱): معماری روش پیشنهادی

شکل شماره (۲) عملکرد روش پیشنهادی را در محیط ترکیبی مه و ابر تشریح می‌کند. ذخیره سازی و پردازش و ارتباطات ایمن در لایه مه و در گره‌های مستقل از ابر انجام می‌شود. با پردازش در مه میزان تاخیر و ناهنجاری و تشخیص حملات بهبود یافته و محیط ایمن تری برای اینترنت اشیا فراهم می‌شود. دستگاه‌های که در لبه ابر مورد استفاده قرار می‌گیرند شامل موبایل، سرورها، بردهای رزبری پای و دستگاه‌های مختلف اینترنت اشیا برای جمع‌آوری داده‌ها هستند.



شکل (۲): عملکرد روش پیشنهادی

• **تشخیص حملات:** تشخیص حملات منع سرویس و منع سرویس توزیع شده در محیط مه و به صورت بلادرنگ و ارسال گزارش به محیط ابر.

در اینترنت اشیا تعداد زیادی از تجهیزات همواره در حال تولید داده‌ها ساخت یافته و یا غیر ساخت یافته با حجم و تنوع مختلف می‌باشند. از طرفی به دلیل تامین سطح بالایی از امنیت و تامین حریم خصوصی در رایانش ابری، یکی دیگر از مهمترین مشکلات مربوط به اطلاعات در اینترنت اشیا به واسطه بهره‌گیری از آن برطرف می‌شود. به طور کلی داده‌های ورودی به مرحله پیش پردازش می‌رود و سپس توسط روش خوشه‌بندی وزنی معکوس برای ساخت پنج خوشه مورد استفاده قرار می‌گیرند. سپس از الگوریتم تشخیص ناهنجاری جهت تشخیص شرایط عادی و غیر عادی استفاده می‌شود. ترافیک ورودی به نودهای مه بررسی و تشخیص حمله و ناهنجاری در آنها بررسی می‌شود. خوشه‌بندی براساس شباهت بین ویژگی‌های داده‌ها انجام می‌شود. پس از خوشه‌بندی، داده‌های ورودی پیش پردازش شده به دو گروه تقسیم می‌شوند: یک گروه برای آموزش و ساخت مدل کلاسیفایر و گروه دیگر، برای آزمایش صحت کلاسیفایر جهت تفکیک داده‌های عادی و غیرطبیعی، استفاده می‌شوند. تعدادی از داده‌ها برای بررسی دقت کلاسیفایر و ارزیابی مدل استفاده خواهد شد. سیستم به صورت عادی آموزش داده خواهد شد و سپس هر رفتاری که خارج از حالت عادی سیستم باشد به‌عنوان ناهنجاری در نظر گرفته خواهد شد. در ادامه سیستم تشخیص نفوذ پیشنهادی برای تشخیص حملات استفاده می‌شود. داده‌های ترافیک شبکه ورودی به مه در لبه‌های ابر، در هر پنجره زمانی در مسیرهای مه ثبت می‌شوند. سپس داده‌های ثبت شده مورد پیش‌پردازش قرار گرفته و به اولین مرحله تشخیص ناهنجاری هدایت می‌شوند که از یک الگوریتم یادگیری ماشین استفاده می‌کند. این امر اجازه تشخیص و مسدود کردن ترافیک مشکوک را می‌دهد، در عین اینکه اجازه دسترسی به ترافیک عادی را نیز می‌دهد. ترافیک مشکوک بر روی هر مسیرهای شبکه با یک سرور مرکزی هماهنگ می‌شوند. یک دسته‌بندی یادگیری گروهی مبتنی بر جنگل تصادفی نیز برای دسته‌بندی داده‌های ترافیک شبکه موجود بر روی سرور ذخیره مرکزی مورد استفاده قرار می‌گیرد تا نوع هر حمله را تشخیص دهد. شکل (۱) معماری روش پیشنهادی را نشان می‌دهد. در روش پیشنهادی داده‌ها توسط سنسورها که در اینجا دستگاه‌های اینترنت اشیا هستند جمع‌آوری می‌شوند. سپس برای پیش پردازش و تشخیص نفوذ و ناهنجاری به لایه مه فرستاده می‌شوند. پردازش‌های نهایی و

انتخاب ویژگی فیلتر و روش ضریب همبستگی و اطلاعات متقابل^۱ و همچنین با استفاده از بدست آوردن برای تک تک ویژگی‌ها تعداد ویژگی‌ها را از ۷۸ ویژگی به تعداد ۲۱ ویژگی کاهش می‌دهیم. در پایگاه داده فوق داری هشت کلاس می‌باشد. یک کلاس عادی و هفت کلاس حمله می‌باشد.

۳-۲- سیستم تشخیص نفوذ پیشنهادی

در این بخش روال در نظر گرفته شده برای تشخیص نفوذ در مه و ابر و پیش از ورود به لایه کاربرد در بین دستگاه‌های اینترنت اشیا می‌باشد و همچنین اجزای سیستم تشخیص نفوذ پیشنهادی ارائه شده است.

۳-۲-۱- اجزای سیستم تشخیص نفوذ

سیستم تشخیص نفوذ پیشنهادی از اجزای زیر تشکیل شده است:

ماژول جمع‌آوری‌کننده داده‌های ترافیک شبکه در مه و

لبه‌های ابر: این ماژول در کنترل‌کننده ترافیک شبکه مه نزدیک به هر مسیریاب لبه شبکه فراهم‌کننده ابری پیاده‌سازی می‌شود. این امر، ثبت داده‌های ترافیک ورودی شبکه به سرویس‌های مه میزبانی شده بر روی همان زیرساخت را ممکن می‌سازد. این ماژول، ثبت داده‌ها را در یک پنجره زمانی مشخص انجام می‌دهد. در هر پنجره زمانی، داده‌های ثبت شده از ترافیک شبکه در یک سرور ذخیره‌سازی ذخیره می‌شوند.

ماژول پیش-پردازش داده‌های ترافیک شبکه: این

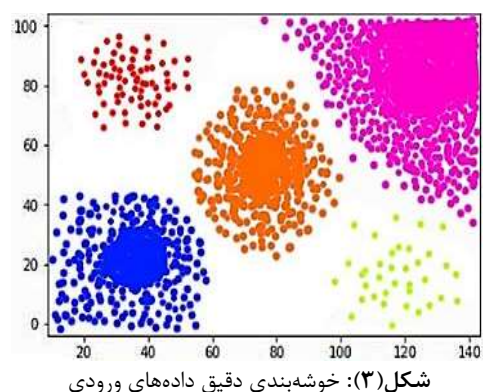
ماژول وظیفه پیش-پردازش داده‌های ترافیک شبکه را در طول هر پنجره زمانی بر عهده دارد. وظایف قالب‌بندی، پاکسازی و عادی‌سازی داده‌های ترافیک شبکه در نودهای مه انجام می‌شود. این ماژول برای قرار گرفتن در نزدیکی ماژول کنترل‌کننده داده‌های ترافیک شبکه مه و ابر طراحی شده است. در این روش از چند گره مه جهت احراز هویت، پردازش، ذخیره‌سازی، تشخیص ناهنجاری استفاده شده است. هر گره دارای پایگاه داده برای ذخیره‌سازی و ارتباط است.

ماژول تشخیص ناهنجاری: این ماژول به‌دسته‌بندی

داده‌های ترافیک شبکه در هر پنجره زمانی به دو دسته ترافیک عادی یا ناهنجاری می‌پردازد. با استفاده از خوشه‌بندی و سپس دسته‌بندی و انجام آزمایش و آمایش با استفاده از روش جنگل تصادفی بر روی مجموعه داده می‌توان سرعت تشخیص ناهنجاری و دقت تشخیص را افزایش داد.

ناهنجاری در بین دستگاه‌های اینترنت اشیا به دلایل مختلف از جمله سیستم عامل، حافظه محدود، تنوع دستگاه‌های اینترنت اشیا و عوامل مختلف دیگر رخ می‌دهد. لذا به روز رسانی و همچنین فاز احراز هویت و اعتبار سنجی از عوامل مهم در شناخت ناهنجاری بین دستگاه‌های اینترنت اشیا است. فاز خوشه‌بندی دقیق داده‌ها و تشخیص نفوذ و ناهنجاری در لایه مه انجام می‌شود. جمع‌آوری ترافیک ورودی به شبکه و همچنین انتخاب ویژگی‌ها جهت تشخیص حملات با استفاده از الگوریتم‌های انتخاب ویژگی و دسته‌بندی ترافیک ورودی به شبکه به دو گروه عادی و غیرعادی در لایه مه انجام می‌گردد. مدیریت گره‌ها و تصمیمات نهایی و کلان در لایه ابر انجام می‌شود. در روش پیشنهادی داده‌ها را به درستی و بدون خطا به دو دسته عادی و عادی خوشه‌بندی و برچسب گذاری شده است. در این پژوهش از مجموعه داده [۱۹] CIC-IDS2017 استفاده شده است. برای تشخیص تعداد خوشه‌های بهینه و افزایش دقت خوشه‌بندی ابتدا از روش خوشه‌بندی سلسله مراتبی دندوگرام آن را ترسیم می‌نماییم و سپس خوشه‌بندی بر روی داده‌های ورودی انجام خواهد شد.

شکل شماره (۳) خوشه‌بندی دقیق داده‌های ورودی توسط الگوریتم پیشنهادی را نشان می‌دهد و سپس جهت تشخیص و پیشگویی حملات از الگوریتم‌های یادگیری ماشین جنگل تصادفی استفاده شده است. یک خوشه مربوط به داده‌های عادی و بقیه مربوط به مجموعه حملاتی است که در سال‌های اخیر در شهرها، خانه‌های هوشمند، دوربین‌ها و شهرهای هوشمند اتفاق افتاده است.

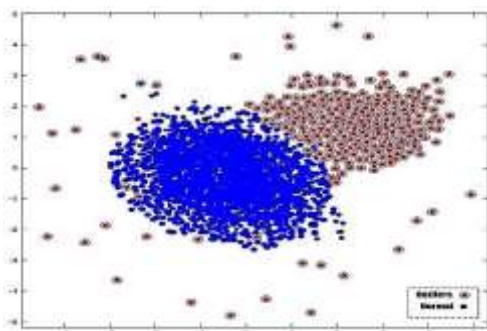


۳-۱- مجموعه داده CIC-IDS2017

در پایگاه داده CIC-IDS2017 که پایگاه داده مخصوص داده‌های اینترنت اشیا است که توسط دوربین‌های هوشمند، خانه‌های هوشمند و شهرهای هوشمند جمع‌آوری شده است. در مجموع ۷۴۳/۸۳۰/۲ رکورد دارد و شامل ۷۸ ویژگی است. براساس روش

¹ Mutual Information

ناهنجاری انجام می‌شود. سپس یادگیری جمعی جنگل تصادفی برای تشخیص نوع هر نفوذ مورد استفاده قرار می‌گیرد. نتایج به دست آمده از کل سیستم تشخیص نفوذ پیشنهادی با نتایج دسته‌بندی استاندارد جنگل تصادفی جمعی مقایسه شده است. شکل (۴) تشخیص ناهنجاری را در روش پیشنهادی نشان می‌دهد. اگر داده‌های ورودی از یک حد مجاز بزرگتر یا مساوی باشند حالت عادی است و اگر کوچکتر یا مساوی باشند حالت غیر عادی یا حمله رخ داده است.



شکل (۴): تشخیص ناهنجاری

۴-۱- معیارهای ارزیابی

سیستم تشخیص نفوذ پیشنهادی به دسته‌بندی داده‌های ترافیک شبکه به صورت مثبت یا منفی می‌پردازد، که به ترتیب متناظر با ترافیک عادی یا حمله هستند. نتایج به دست آمده با استفاده از معیارهای ارزیابی زیر مورد ارزیابی قرار گرفته است:

دقت: درصدی از داده‌های ترافیک که به درستی دسته‌بندی شده‌اند [۲۵-۲۰].

$$\text{دقت} = \frac{\text{داده‌هایی که به درستی شده‌اند}}{\text{کل داده‌های ثبت شده ترافیک}} \quad (1)$$

نرخ مثبت کاذب^{۱۰}: درصدی از داده‌های ترافیک عادی که به عنوان داده‌های ترافیک حمله دسته‌بندی شده‌اند.

$$\text{نرخ مثبت کاذب} = \frac{FP}{TN + FP} \quad (2)$$

که در این رابطه، FP یا منفی کاذب تعداد داده‌های ترافیک عادی است که به اشتباه به عنوان داده‌های ترافیک حمله دسته‌بندی شده‌اند. TN نیز تعداد داده‌های ترافیک عادی است که به درستی دسته‌بندی شده‌اند.

زمان اجرا: کل مدت زمان مورد نیاز برای تشخیص نفوذها که شامل مدت زمان پیش-پردازش، مدت زمان تشخیص ناهنجاری و مدت زمان دسته‌بندی جنگل تصادفی است.

ماژول هماهنگ‌سازی ترافیک شبکه: این ماژول، مدل ترافیک مخرب شبکه را بر روی هر مسیریاب با یک سرور مرکزی هماهنگ می‌کند. این عمل هماهنگ‌سازی برای هر پنجره زمانی به طور جداگانه انجام می‌شود.

ماژول دسته‌بندی ترافیک حمله‌ها: این ماژول به دسته‌بندی داده‌های ترافیک مخرب شبکه می‌پردازد که با سرور ذخیره‌سازی مرکزی هماهنگ شده‌اند. یک دسته‌بندی یادگیری گروهی بر اساس جنگل تصادفی برای اجرای دسته‌بندی چند-کلاسی بر روی داده‌های ترافیک مخرب شبکه مورد استفاده قرار می‌گیرد. این امر، تشخیص نوع هر حمله و اجرای زودهنگام اقدامات مقابله‌ای در برابر آن حمله را ممکن می‌سازد.

۴- نتایج و بحث

در روش خوشه‌بندی وزنی معکوس، داده‌ها بر اساس شباهت بین رکورد داده‌ها به پنج گروه خوشه‌بندی می‌نماید. پس از تکمیل خوشه‌بندی، با استفاده از الگوریتم تشخیص ناهنجاری گروه‌ها تحت عناوین عادی^۱ و غیرعادی^۲ برچسب‌گذاری می‌شوند. داده‌های ورودی به دو بخش تقسیم می‌شوند: داده‌های آموزش^۳ (۶۰٪) و داده‌های آزمایش^۴ (۴۰٪). داده‌های آموزش که برای آموزش حالت عادی سیستم و آموزش رفتار عادی سیستم استفاده می‌شود و از روش خوشه‌بندی تک کلاسی استفاده می‌نماید. ایجاد درخت طبقه‌بندی جنگل تصادفی و داده‌های آزمایش برای آزمایش پیش‌بینی سیستم استفاده می‌شود. از نرم افزار آی فاگ سیم^۵ برای پیاده‌سازی محیط مه و ابر و همچنین از نرم افزار پایتون برای پیاده‌سازی الگوریتم‌ها و آزمایش سیستم استفاده شده است. از ماتریس درهم ریختگی^۶ به منظور ارزیابی سیستم از نظر مثبت صحیح^۷، منفی صحیح^۸، مثبت کاذب و منفی کاذب^۹ شده است. سپس جهت تشخیص صحت مدل تشخیص نفوذ پیشنهادی، نتایج به دست آمده با نتایج سایر های انجام شده مقایسه خواهد شد. به منظور شبیه‌سازی نفوذها در بستر ابری، یک روش نمونه‌برداری پنجره‌ای مبتنی بر زمان بر روی مجموعه آزمایشی نیز اعمال شده است. رکوردهای ثبت شده از ترافیک در هر پنجره زمانی متعاقباً به پنج بخش تقسیم می‌شوند که هر رکورد تنها به یک مسیریاب ارسال می‌شود. بر روی هر یک از مسیریاب‌ها وظایف پیش-پردازش و تشخیص

¹ Normal

² Abnormal

³ Train

⁴ Test

⁵ Ifogsim

⁶ Confusion matrix

⁷ True Positive Rate

⁸ True negative

⁹ False negative

¹⁰ False Positive Rate

است. مدل پیشنهادی را در شناسایی و تشخیص ناهنجاری نشان می‌دهد.

نتایج موجود در جدول (۱) مقادیر متوسط دقت و متوسط نرخ مثبت کاذب را در سیستم تشخیص نفوذ پیشنهادی برای تشخیص چهار نوع از دسته‌های حمله موجود در مجموعه داده استفاده شده در این مقاله را نشان داده است که حمله‌های منع سرویس، اسکن پورت^۲، PingScan و BruteForce هستند. دقت کلی سیستم تشخیص نفوذ پیشنهادی برای تمام دسته‌های حمله به مقدار ۹۸/۰۳ رسیده است، و متوسط FPR آن نیز ۰/۱۷ است. که نقطه قوت این تحقیق است. در روش پیشنهادی حملات منع سرویس با متوسط دقت ۹۹/۸۳ می‌باشد.

جدول (۱): نرخ مثبت کاذب و متوسط دقت در حملات

مطالعات	حملات	متوسط نرخ	متوسط دقت
روش پیشنهادی	Portscan	۰/۲۱	۹۷/۱
	PingScan	۰/۱۹	۹۷/۲
	DoS	۰/۱۱	۹۹/۸۳
	BruteForce	۰/۱۸	۹۸
	Overall	۰/۱۷	۹۸/۰۳
مطالعات قبلی [۲۰]	Portscan	۰/۲۴	۹۶/۷
	PingScan	۰/۲۲	۹۷/۸
	DoS	۰/۱۲	۹۹/۲
	BruteForce	۰/۲۶	۹۴/۵
	overall	۰/۲۱	۹۷/۰۵

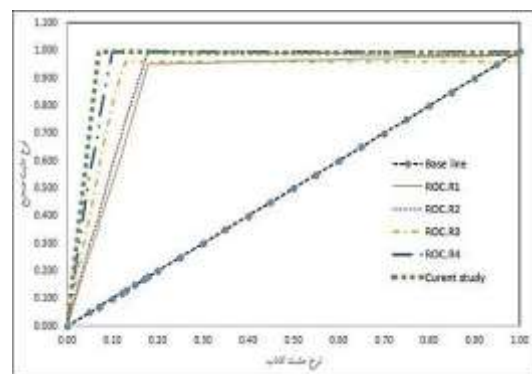
ماژول ترافیک شبکه به ثبت و ضبط ترافیک شبکه ورودی به مه و لبه‌های ابر بر روی هر مسیر یاب لبه‌ای شبکه در یک پنجره زمانی مشخص می‌پردازد. سپس داده‌های ثبت شده مورد عملیات پیش پردازش قرار گرفته است. تشخیص ناهنجاری با استفاده از یک مدل تشخیص ناهنجاری هدایت می‌شوند. سپس، ترافیک مشکوک بر روی هر مسیر یاب با سرور مرکزی هماهنگ می‌شود. توجه به شکل (۶) نرخ تشخیص، دقت کلاسیفایر، پرسیژن و ریکال را در مجموعه داده اینترنت اشیا نشان می‌دهد. ستون اول حالت عادی و ستون دوم حالت ناهنجاری را نشان می‌دهد. در روش پیشنهادی نرخ تشخیص و نرخ دقت و نرخ تشخیص خطا افزایش یافته است و همچنین نرخ مثبت کاذب منفی کاهش یافته که دقت مدل را نشان می‌دهد.

نمودارهای ROC و AUC: نمودار مشخصه عملکرد

دریافت‌کننده ROC و مقدار ناحیه زیر منحنی^۱ AUC و معمولاً برای نمایش نتایج در مسائل تشخیص دودویی در یادگیری ماشین مورد استفاده قرار می‌گیرند. نمودار ROC نشان می‌دهد که چگونه تعداد نمونه‌های مثبتی که به درستی دسته‌بندی شده‌اند، با تعداد نمونه‌های منفی که به اشتباه دسته‌بندی شده‌اند، تغییر می‌کند. مقدار AUC نیز دقت مدل را مشخص می‌کند [۲۰].

۴-۲- عملکرد سیستم تشخیص نفوذ پیشنهادی

ما در این بخش، نتایج به دست آمده از ارزیابی سیستم تشخیص نفوذ را ارائه می‌دهیم که با استفاده از مجموعه داده CIC-IDS2017 ارزیابی شده است. نتایج به دست آمده از ماژول تشخیص ناهنجاری در شکل (۵) نشان داده شده است. نتایج نمودارهای ROC و مقدار امتیاز AUC ماژول تشخیص ناهنجاری بر روی هر مسیر یاب در بستر ترکیبی مه و ابر نشان داده شده است. ماژول تشخیص ناهنجاری بر روی مسیر یاب سوم ۹۰/۳ است. همچنین، این ماژول به دقتی برابر با ۹۲/۲، ۹۳/۱، ۸۹ به ترتیب بر روی مسیر یاب‌های ۱، ۲ و ۴ دست یافته است. در روش پیشنهادی ماژول تشخیص ناهنجاری به بالاترین دقت یعنی ۹۳/۳ رسیده است. زمان اجرای کم و نرخ بالای مثبت کاذب نیز بر روی هر مسیر یاب مشاهده می‌شود. تشخیص حملات و تشخیص ناهنجاری در بین دستگاه‌های اینترنت اشیا و کاهش نرخ مثبت کاذب و کاهش زمان محاسباتی و پاسخ بلادرنگ از جمله چالش‌ها در محیط اینترنت اشیا و در محیط ابری است.



شکل (۵): بهبود مدل در روش پیشنهادی در تشخیص ناهنجاری

در روش پیشنهادی حملات منع سرویس نسبت به روش‌های گذشته بهبود یافته است. یکی از حملات رایج در بین دستگاه‌های اینترنت اشیا حملات شنود و حملات منع سرویس

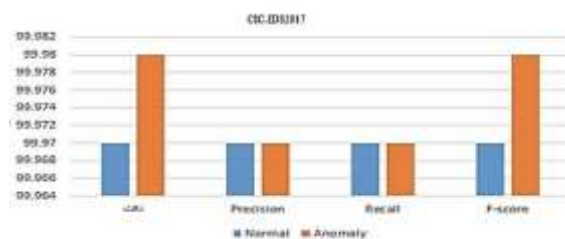
^۲ portscana

^۱ Area Under Roc

در اینترنت اشیا موثر است. سیستم تشخیص نفوذ پیشنهادی بر روی بستر ابری گوگل پیاده‌سازی شده و با استفاده از مجموعه داده عمومی CIC-IDS2017 مورد آزمایش قرار گرفته است. در آینده ارزیابی سیستم برای شناسایی و تشخیص کامل حملات منع خواب در نودهای مه و کاهش بار محاسباتی مورد بررسی قرار خواهد گرفت.

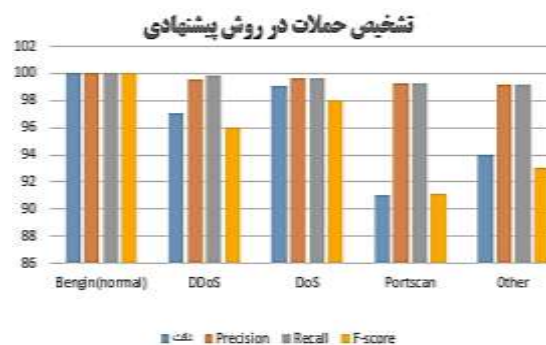
۶- مراجع

- [1] H. F. Atlam, A. Alenezi, R. J. Walters, G. B. Wills, and J. Daniel, "Developing an adaptive Risk-based access control model for the Internet of Things," in 2017 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData), 2017: IEEE, pp. 655-661.
- [2] S. Iqbal et al., "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service," *Journal of Network and Computer Applications*, vol. 74, pp. 98-120, 2016.
- [3] G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "DDoS attacks in cloud computing: Issues, taxonomy, and future directions," *Computer Communications*, vol. 107, pp. 30-48, 2017.
- [4] G. Somani, M. S. Gaur, D. Sanghi, M. Conti, M. Rajarajan, and R. Buyya, "Combating DDoS attacks in the cloud: requirements, trends", and future directions *IEEE Cloud Computing*, vol. 4, no. 1, pp. 22-32, 2017.
- [5] M. A. Lawal, R. A. Shaikh, and S. R. Hassan, "An anomaly mitigation framework for iot using fog computing," *Electronics*, vol. 9, no. 10, p. 1565, 2020.
- [6] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved k-means clustering algorithm," in 2010 Third International Symposium on intelligent information technology and security informatics, 2010: Ieee, pp. 63-67.
- [7] Barbakh, W., & Fyfe, C. (2007). Inverse weighted clustering algorithm. *Computing and Information Systems*, 11(2).
- [8] H. Neuschmied, M. Winter, K. Hofer-Schmitz, B. Stojanovic, and U. Kleb, "Two Stage Anomaly Detection for Network Intrusion Detection," in *ICISSP*, 2021, pp. 450-457.
- [9] S. Weisong, Z. Xingzhou, W. Yifan, and Z. Qingyang, "Edge computing: State-of-the-art and future directions," *Journal of Computer Research and Development*, vol. 56, no. 1, p. 69, 2019.
- [10] F. M. Ramos, D. Kreutz, and P. Verissimo, "Software-defined networks: On the road to the softwarization of networking," *Cutter IT journal*, 2015.



شکل(۶): تشخیص حالت عادی و غیر عادی در روش پیشنهادی

شکل (۷) نرخ تشخیص حملات و افزایش دقت مدل در شناسایی و تشخیص حملات در نودهای مه قبل از ورود به ابر را نشان می‌دهد در نمودار زیر نرخ تشخیص حملات منع سرویس و منع سرویس توزیع شده نسبت به آخرین تحقیق انجام شده بهبود یافته است. در هر قسمت مثلاً در قسمت تشخیص حالت عادی ستون اول دقت، ستون دوم پریسیژن، ستون سوم ریکال و ستون آخر دقت ارزیابی را نشان می‌دهد.



شکل(۷): تشخیص حملات در روش پیشنهادی

۵- نتیجه‌گیری

در این مقاله یک سیستم تشخیص نفوذ توزیع شده و همچنین یک سیستم تشخیص ناهنجاری با استفاده از روش ترکیبی برای محیط‌های مه و ابر برای شناسایی ناهنجاری بین دستگاه‌های اینترنت اشیا ارائه شده است. سیستم تشخیص نفوذ، مازول ترافیک شبکه به ثبت و ضبط ترافیک شبکه ورودی به مه و لبه‌های ابر بر روی هر مسیر یاب لبه‌ای شبکه در یک پنجره زمانی مشخص می‌پردازد. سپس داده‌های ثبت شده در گره‌های مه مورد عملیات پیش پردازش قرار گرفته است. ترافیک مشکوک بر روی هر مسیر یاب با سرور مرکزی هماهنگ می‌شود. سپس، یک دسته‌بندی یادگیری بر اساس جنگل تصادفی مورد استفاده قرار می‌گیرد تا داده‌های ترافیک شبکه موجود بر روی سرور ذخیره‌سازی مرکزی دسته‌بندی شده و نوع هر حمله تشخیص داده شود. احراز هویت، تشخیص ناهنجاری بین دستگاه‌های اینترنت اشیا، تشخیص ناهنجاری در گره‌های مه انجام و سپس نتایج به لایه ابر فرستاده می‌شود. این روش برای افزایش امنیت، کاهش زمان پاسخگویی برنامه‌های کاربردی که بلادرنگ هستند

- [18] L. Liu and Y. Zhai, "A Survey on MapReduce Scheduling in Cloud Computing," Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control, pp. 1710-1715, 2015.
- [19] M. M. Rashid, J. Kamruzzaman, M. M. Hassan, T. Imam, and S. Gordon, "Cyberattacks Detection in IoT-Based Smart City Applications Using Machine Learning Techniques," International Journal of Environmental Research and Public Health, vol. 17, no. 24, p. 9347, 2020.
- [20] M. Idhammad, K. Afdel, M. Belouch, Distributed Intrusion Detection System for Cloud Environments based on Data Mining techniques, *Procedia Computer Science*, Vol 127, pp35-41, 2018.
- [21] R. Beghdad, "Efficient deterministic method for detecting new U2R attacks," *Computer Communications*, Vol 32, pp1104-1110, 2009.
- [22] E .Kim, and S. Kim, "A Novel Anomaly Detection System Based on HFR-MLR Method", in *Mobile, Ubiquitous, and Intelligent Computing*, p p. 279-286, 2014.
- [23] C.A. Charu, and K.R. Chandan, "Data clustering: algorithms and applications," In: Chapman and Hall/CRC Boston, MA. 2013.
- [24] H. Pajouh, G. Dastghaibiyfard, and S. Hashemi," Two-tier network anomaly detection model," a machine learning approach. *Journal of Intelligent Information Systems*, Vol 48, pp. 61-74, 2017.
- [25] S. Mishra and A. Tripathi, "IoT platform business model for innovative management systems," *International Journal of Financial Engineering*, vol. 7, no. 03, p. 2050030, 2020.
- [11] M. Mirzaei, A. Mehabadi," Hybrid Anomaly Detection Method Using Community Detection in Graph and Feature Selection," *Journal of Electronical & Cyber Defence* Vol. 8, No. 1, 2020. (in persion)
- [12] K. Shoushian , A. J. Rashidi, M. Dehghani," Modeling of Cyber-Attacks Obfuscation, Based on Alteration Technique of Attack," *Journal of Electronical & Cyber Defence* Vol. 8, No. 1, 2020. (in persion)
- [13] V. Yadegari, A. Matinfar, "Detect Web Denial of Service Attacks Using Entropy and Support Vector Machine Algorithm," *Journal of Electronical & Cyber Defence* Vol. 6, No. 4, 2019. (in persion)
- [14] C. Modi, D. Patel, B. Borisanya, A. Patel, M. Rajarajan, A novel framework for intrusion detection in cloud, in: *Proceedings of the Fifth International Conference on Security of Information and Networks*, ACM, pp. 67-74, 2012.
- [15] S. Teng, C. Zheng, H. Zhu, D. Liu, and W. Zhang, "A cooperative intrusion detection model for cloud computing networks," *International Journal of Security and its applications*, vol. 8, no. 3, pp. 107-118, 2014.
- [16] S. Weisong, Z. Xingzhou, W. Yifan, and Z. Qingyang, "Edge computing: State-of-the-art and future directions," *Journal of Computer Research and Development*, vol. 56, no. 1, p. 69, 2019.
- [17] M. Idhammad, K. Afdel, M. Belouch, "Dos detection method based on artificial neural networks," *International Journal of Advanced Computer Science and Applications (ijacsa)*, vol 10, pp 14569, 2017.