توسعه پایدار و ساختمان هوشمند
https://sdsb.buqaen.ac.ir

جلد ۱، شماره ۴، زمستان ۱۴۰۳

رویکردی برخط و انطباقی برای شناسایی برنامه‌ها در ترافیک رمزنگاری‌شده شبکه با استفاده از مدل‌یادگیری شبکه عصبی عمیق

محمد امین رستگار*[۸۵]، مسعود نوفرستی[۸۶]

۱- دانشکده مهندسی-گروه مهندسی کامپیوتر - واحد علوم تحقیقات- دانشگاه آزاد اسلامی- تهران- ایران.

۲- دانشکده مهندسی-گروه مهندسی کامپیوتر - واحد علوم تحقیقات- دانشگاه آزاد اسلامی- تهران- ایران.

چکیده

با گسترش روزافزون استفاده از رمزنگاری در ترافیک شبکه، به‌ویژه از طریق پروتکل‌هایی مانند TLS و SSL، کارایی روش‌های سنتی شناسایی برنامه‌ها مبتنی بر تحلیل محتوای بسته‌ها به‌شدت کاهش یافته است. در شبکه‌های مدرن با پهنای باند بالا، شناسایی بلادرنگ و دقیق برنامه‌ها نه‌تنها با چالش‌هایی از نظر سرعت، حجم و تنوع داده‌ها مواجه است، بلکه هزینه‌بر نیز می‌باشد. در این پژوهش، یک رویکرد تطبیقی برای طبقه‌بندی جریان‌های ترافیکی ارائه می‌شود که با بهره‌گیری از تکنیک بازرسی عمیق بسته‌ها (DPI) به‌عنوان منبع مرجع، مدل طبقه‌بندی را به‌صورت پویا و مداوم به‌روزرسانی می‌کند. در هسته طبقه‌بندی پیشنهادی، از شبکه عصبی عمیق (DNN) به‌عنوان الگوریتم یادگیری بهره گرفته شده که قادر است الگوهای پیچیده و پنهان در ترافیک رمزنگاری‌شده را با دقت بالا شناسایی کند. ارزیابی تجربی این روش بر روی دو مجموعه‌داده عمومی ISCX VPN-nonVPN و ISCX Tor-nonTor انجام شده و نتایج به‌دست‌آمده نشان‌دهنده عملکرد بهتر DNNدر برخی آزمایش‌ها است که در دسته‌بندی جریان‌های پیچیده‌تر یا در مواقعی که داده‌های برچسب‌خورده به‌صورت لحظه‌ای کم هستند، عملکرد بهتری نسبت به طبقه‌بندهای سنتی دارد.

واژگان کلیدی: شناسایی برنامه‌های کاربردی، طبقه‌بندی ترافیک شبکه، طبقه‌بندی جریان داده‌ها، شبکه عصبی عمیق

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۱۰/۰۳

تاریخ پذیرش: ۱۴۰۳/۱۲/۰۹

* نویسنده مسئول، پست الکترونیکی: **mohammadamin.rastegar@iau.ir**

Journal of Sustainable Development and Smart Building, vol. ۱, no. ٤
Digital Object Identifier https://doi.org/۱۰,۲۲۰۳٤/sdsb.۱,٤,٤۰

مقدّمه

ظرفیت ترافیک شبکه‌ها به طور مستمر در حال افزایش است و امروزه شبکه‌های ۱۰۰ گیگابیتی در مراکز داده به کار گرفته می‌شوند [۱] این رشد فزاینده، نیاز به تحلیل برخط ترافیک را برای مدیران شبکه بیش از پیش ضروری ساخته است. طی دهه گذشته، طبقه‌بندی ترافیک شبکه به دلایلی همچون پیاده‌سازی مکانیزم‌های تضمین کیفیت خدمات (QoS) امنیت، نظارت و مهندسی شبکه، توجه ویژه‌ای را به خود جلب کرده است [۲].

امروزه طبقه‌بندی ترافیک شبکه‌های پهن‌باند به صورت بلادرنگ برای سیستم‌های تشخیص نفوذ خودکار جهت شناسایی الگوهای حملات و حملات انکار سرویس حیاتی است [۳]. همچنین این فناوری در خدمات شبکه‌ای، مانند مدیریت تعرفه‌های اینترنتی، برای شناسایی دقیق برنامه‌های مختلف در پهنای باند ضروری می‌باشد.

ترافیک شبکه‌های پهن‌باند را می‌توان به عنوان دنباله‌ای عظیم، نامحدود و پیوسته از جریان‌ها (Flow Stream) مدل کرد [۴]. هر جریان از بسته‌های متوالی با پنج‌تایی یکسان < آیپی مبدأ، پورت مبدأ، آیپی مقصد، پورت مقصد، پروتکلTCP/UDP > تشکیل می‌شود. از دیدگاه ساختار داده، ترافیک شبکه، یک دنباله نامتناهی از جریان‌هاست که به صورت $FlowSet=\{f_{1},...f_{n}\}$ نمایش داده می‌شود.

با ظهور یادگیری عمیق (Deep Learning) و موفقیت آن در حوزه‌های مختلف از جمله بینایی ماشین و پردازش زبان طبیعی، توجه به این تکنیک در حوزه شبکه نیز افزایش یافته است[۵]. به‌ویژه در طبقه‌بندی ترافیک شبکه، استفاده از شبکه‌های عصبی عمیق (DNN)[۱] امکان استخراج خودکار الگوهای پیچیده را از ویژگی‌های خام فراهم کرده و عملکرد مدل‌های طبقه‌بندی را در سناریوهای پویا و داده‌های رمزنگاری‌شده بهبود داده است [۶].

به طور کلی سه روش اصلی در ادبیات طبقه‌بندی ترافیک شبکه وجود دارد:

1. روش مبتنی بر پورت: که با توجّه به شماره پورت‌های استاندارد عمل می‌کند. مطالعاتی مانند [۷]. نشان داده‌اند که این روش امروزه کارایی لازم را ندارد.
2. روش مبتنی بر محتوای بسته‌ها :(DPI) که با بررسی محتوای بسته‌ها و تطبیق الگوها عمل می‌کند. اگرچه این روش، دقت بالایی دارد، اما در شبکه‌های واقعی با محدودیت‌هایی مواجه است.
3. روش‌های مبتنی بر یادگیری ماشین و یادگیری عمیق: که از ویژگی‌های آماری مستقل از پورت و محتوای بسته‌ها استفاده می‌کند. این روش‌ها نه‌تنها هزینه محاسباتی کمتری نسبت به DPI دارند، بلکه توانایی طبقه‌بندی ترافیک رمزنگاری‌شده را نیز دارا می‌باشند. به‌ویژه، شبکه‌های عصبی عمیق (DNN) با قابلیت استخراج ویژگی‌های سطح بالا از داده‌های خام، دقت طبقه‌بندی را در محیط‌های پویا افزایش داده‌اند.

ویژگی‌های پویا و متنوع شبکه‌های امروزی، باعث تغییرات مداوم در الگوهای ترافیک می‌شود، بنابراین، الگوریتم‌های طبقه‌بندی باید بتوانند به صورت پویا با این تغییرات سازگار شده و داده‌های قدیمی غیرمرتبط را کنار بگذارند [۷].

در این پژوهش، یک روش انطباقی و برخط برای طبقه‌بندی ترافیک شبکه ارائه می‌شود که از تکنیک‌های یادگیری عمیق بهره می‌برد. روش پیشنهادی از چهار ماژول اصلی تشکیل شده است:

1. استخراج ویژگی‌ها: تشخیص جریان‌ها و نگهداری ساختار داده FlowSet
2. بازرسی عمیق بسته‌ها :(DPI) تولید برچسب‌ها برای جریان‌های ترافیکی
3. پردازش جریان: مدیریت جریان‌ها بین ماژول‌ها.
4. طبقه‌بندی جریان: ساخت و به‌روزرسانی تابع طبقه‌بندی.

در ماژول طبقه‌بندی، از شبکه‌های عصبی عمیق (DNN) به عنوان مدل پایه استفاده شده است. این شبکه‌ها با استفاده از چندین لایه پنهان، توانایی شناسایی الگوهای پیچیده و غیرواضح در داده‌های جریان را دارند و نسبت به مدل‌های کلاسیک، عملکرد بهتری در مواجهه با داده‌های متغیر و رمزنگاری‌شده نشان می‌دهند.

در این پژوهش، یک چارچوب انطباق‌پذیر و برخط برای طبقه‌بندی ترافیک شبکه ارائه شده است که با هدف پاسخ‌گویی به چالش‌های ناشی از تغییرات پویا در الگوهای ترافیکی طراحی گردیده است. این چارچوب با ترکیب تکنیک های یادگیری عمیق و پردازش جریانی، قادر ا ست به صورت لحظه‌ای، جریان‌های شبکه را تحلیل کرده و با دقت بالا به طبقه‌بندی آن‌ها بپردازد. یکی از نوآوری‌های کلیدی این روش، به‌کارگیری شبکه‌های ع صبی عمیق (DNN) در بخش طبقه‌بندی جریان‌هاست که موجب بهبود چشم‌گیر دقت و کارایی مدل به‌ویژه در مواجهه با ترافیک‌های رمزنگاری‌شده شده است. همچنین، چارچوب پیشنهادی بر روی مجموعه داده‌های ا ستاندارد و رایج در حوزه ترافیک شبکه، مورد ارزیابی قرار گرفته و نتایج نشان می‌دهد که رویکرد پیشنهادی، نه‌تنها عملکرد قابل

[۱] Deep Neural Network

قبولی در طبقه‌بندی بلادرنگ دارد، بلکه در شرایط واقعی و پیچیده نیز قابل اطمینان است.

مروری بر کارهای مرتبط

در سال‌های اخیر، استفاده از روش‌های یادگیری ماشین برای طبقه‌بندی ترافیک شبکه مورد توجه گسترده‌ای قرار گرفته است. با این حال، مسئله‌ی شناسایی کاربرد و توصیف دقیق ترافیک رمزنگاری‌شده هنوز به‌صورت کامل در ادبیات پژوهشی پوشش داده نشده است.

در پژوهش [۲]، کارایی ویژگی‌های زمانی مبتنی بر جریان را برای شناسایی ترافیک VPN و Tor بررسی کرده‌اند. آن‌ها با استفاده از سه الگوریتمJ۴۸ ، جنگل تصادفی (Random Forest) وKNN ، ویژگی‌های زمانی را بر روی مجموعه‌داده‌های عمومی ISCX VPN-nonVPN و Tor-nonTor ارزیابی کردند. نتایج آن‌ها نشان داد که این ویژگی‌ها برای توصیف ترافیک رمزنگاری‌شده بسیار مؤثرند.

در پژوهش دیگری [۸]، از الگوریتم‌های یادگیری ماشین برای شناسایی برنامه‌هایی نظیر فیس‌بوک، توییتر، اسکایپ و غیره استفاده کردند. آن‌ها چهار الگوریتم طبقه‌بندی مختلفJ۴۸ ، Random Forest، KNNو Bayes Net را روی مجموعه‌ای متشکل از ۱۱۱ ویژگی، آزمایش کردند. نتایج نشان داد که الگوریتم KNN با k=۱ دقت ۹۳/۹۴٪ را برای مجموعه‌داده ISCX و Random Forestدقت ۹۰/۸۷٪ را برای مجموعه داده داخلی آن‌ها ارائه داده است. آن‌ها همچنین موفق به کاهش تعداد ویژگی‌ها به ۱۲ ویژگی مؤثر برای هر مجموعه داده بدون کاهش دقت شدند.

با ورود یادگیری عمیق به این حوزه، [۵] برای نخستین بار استفاده از روش‌های انتها به انتها (End-to-End) با استفاده از شبکه عصبی پیچشی یک‌بعدی (۱D-CNN) را برای طبقه‌بندی ترافیک رمزنگاری‌شده پیشنهاد دادند. این روش نیز با استفاده از مجموعه‌داده ISCX ارزیابی شده و نشان داد که دقت الگوریتم در شناسایی ترافیک VPN نسبت به ترافیک غیر VPN-بهتر است.

جریان بی‌پایان ترافیک ورودی، محیطی پویا با تغییرات غیرمنتظره ایجاد می‌کند که مستلزم به کارگیری روش‌های تحلیلی برای پاسخ‌گویی به چالش‌های پردازش در شبکه‌های پهن‌باند است، از جمله یادگیری افزایشی، پردازش برخط و مدیریت داده‌های دورافتاده (Outlier).[۴]. تعریفی رسمی از یک سیستم شناسایی رفتار برای شبکه‌های پهن‌باند ارائه دادند. با در نظر گرفتن محدودیت‌های موجود در سیستم‌های شناسایی رفتار، آن‌ها روشی را پیشنهاد کردند که در آن ساختاری داده‌ای به نام «فرا-رویداد» (meta-event) تعریف می‌شود و روابط بین این فرا-رویدادها برای استخراج رفتار کاربران نهایی مورد استفاده قرار می‌گیرد.

در ادامهٔ این مسیر، [۹] روشی مبتنی بر یادگیری عمیق برای شناسایی کاربرد و توصیف ترافیک شبکه ارائه کردند که در آن فرایندهای استخراج ویژگی و طبقه‌بندی در قالب یک چارچوب یکپارچه پیاده‌سازی شده است. آن‌ها دو مدل مختلف شامل خودرمزگذار انباشته (SAE) و شبکه عصبی پیچشی (CNN) را در این چارچوب به کار گرفته و روش خود را بر روی مجموعه‌داده‌های ISCX و Tor ارزیابی کردند. در این پژوهش، مدل CNN موفق به کسب امتیاز F۱ برابر با ۰/۹۱ برای شناسایی کاربرد و ۰/۸۷ برای توصیف نوع ترافیک شده است.

تکنیک‌های بررسی عمیق بسته‌ها (DPI) ترافیک شبکه را بر اساس امضاهای برنامه‌ها پردازش کرده و اطلاعاتی درباره رفتارهای اخیر ارائه می‌دهند. در این راستا، [۷]. رویکردی مبتنی بر جریان و مبتنی بر DPI برای طبقه‌بندی جریان‌های شبکه ارائه دادند که هدف آن شناسایی برنامه‌های جدید و به‌روزرسانی تطبیقی مدل طبقه‌بندی است. این روش، جریان‌ها را به صورت آنلاین با استفاده از تکنیک DPI برچسب‌گذاری کرده و مدل طبقه‌بندی را به‌طور متناسب به‌روزرسانی می‌کند. رویکرد پیشنهادی با استفاده از ابزار MOAپیاده‌سازی شده و بر روی چندین مجموعه‌داده ارزیابی شده است. در این مطالعه، استفاده از الگوریتم‌های طبقه‌بندی جریان Adaptive Random Forestو Knn با PAW باعث افزایش دقت در طبقه‌بندی مجموعه‌داده‌ها شده است.

یکی از چالش‌های اصلی در مدیریت داده‌های دورافتاده (Outlier) در شبکه‌های پهن‌باند، نبود هیچ‌گونه اطلاعات قبلی در مورد جریان‌های ورودی آینده است. [۱۰]. با معرفی ساختار داده‌ای «فرا-رویدادهای گسترش‌یافته» (extended-meta-events) بر روی پنجره لغزان، الگوریتم خوشه‌بندی برخطی را برای جریان‌های شبکه‌ای در محیط‌های پهن‌باند پیشنهاد دادند. این الگوریتم برای استخراج تدریجی اطلاعات از حجم بالای ترافیک پیوسته طراحی شده است، به‌طوری‌که برخی از داده‌های دورافتاده پیش از تحلیل رفتار شبکه، فیلتر می‌شوند. در این روش، مفهوم «بافر» در ساختار فرا-رویدادهای گسترش‌یافته، برای مدیریت Outlierها در نظر گرفته شده و داده‌های دورافتاده‌ای که با هیچ‌یک از فرا-رویدادهای موجود سازگاری ندارند، در بافری کلی به نام «پنجره» نگهداری می‌شوند.

در شبکه‌های پهن‌باند، تحلیل تطبیقی رفتار متغیر شبکه اهمیت بالایی دارد. در این راستا، [۱۰]. روش ACoPE به‌عنوان یک رویکرد

نیمه‌نظارتی تطبیقی برای اجرای سیاست‌های پیچیده معرفی شده است. این روش با شناسایی روابط میان جریان‌های ترافیکی و به‌کارگیری کنترل آماری فرایند، تغییرات رفتاری را شناسایی و از اطلاعات ماژول DPI برای سازگاری با تغییرات استفاده می‌کند. ارزیابی ACoPE در سناریوهای مختلف، نظیر برنامه‌های پرترافیک، برنامه‌های تغییریافته، و حملات توزیع‌شده منع سرویس، اثربخشی آن را در اجرای سیاست‌های پیچیده و تطبیق با تغییرات رفتاری نشان می‌دهد.

روش پیشنهادی مبتنی بر طبقه‌بندی جریانی با DNN و DPI

در این پژوهش، رویکردی نوین ارائه شده است که با بهره‌گیری هم‌زمان از تکنیک‌های بازرسی عمیق بسته‌ها (DPI) و الگوریتم‌های یادگیری عمیق مبتنی بر شبکه عصبی (DNN)، به طبقه‌بندی جریانی ترافیک شبکه می‌پردازد. در این چارچوب، برچسب‌های استخراج‌شده توسط ماژول DPI به‌عنوان داده‌های حقیقت مبنا (ground truth) مورد استفاده قرار می‌گیرند تا مدل طبقه‌بندی به‌صورت پیوسته آموزش دیده و به‌روزرسانی شود. این ترکیب، امکان شناسایی دقیق و تطبیقی برنامه‌ها را در مواجهه با ترافیک رمزنگاری‌شده و متغیر فراهم می‌سازد.

فرض کنیم در زمان t، تابع طبقه‌بندی CFt برای نگاشت جریان‌ها به برچسب‌ها در اختیار باشد. رویکرد پیشنهادی برای هر جریان $flow \in FlowSet$ ، یک برچسب $label \in LabelSet$ تخصیص می‌دهد که رابطه آن به‌صورت $(flow, label) \in CF_t$ تعریف می‌شود. به‌منظور به‌روزرسانی انطباقی تابع طبقه‌بندی، خروجی‌های ماژول DPI برای تولید CFt+۱ استفاده می‌شوند:

$$CF_{t+1} = Update(CF_t, application) \quad (1$$

چارچوب پشنهادی از چهار ماژول اصلی تشکیل شده است:

۱ .استخراج ویژگی‌ها (Feature Extractor)

در این مؤلفه، ترافیک شبکه پردازش شده و جریان‌ها (flows) ساخته می‌شوند. برای هر جریان، ویژگی‌هایی آماری مانند تعداد بسته‌ها، حجم داده، مدت زمان جریان و نرخ ارسال محاسبه می‌شود. ساختار هر جریان به صورت زیر در نظر گرفته می‌شود:

$$f = <source\text{-}IP, source\text{-}port, destination\text{-}IP, destination\text{-}port, TCP/UDP\ protocol, fe_i, \ldots fe_m> \quad (2$$

برای استخراج جریان ها از ترافیک، ابزار CICFlowMeter مورد استفاده قرار گرفته است که توانایی شناسایی جریان‌های TCP/UDPرا دارد و ۸۳ ویژگی آماری را برای هر جریان تولید می‌کند.

۲ . بازرسی عمیق درون‌خطی بسته‌ها (Inline Deep Packet Inspector)

این ماژول با تطبیق بسته‌های ورودی هر جریان، با الگوهای از پیش تعریف‌شده، اقدام به شناسایی برنامه یا سرویس مرتبط می‌نماید. خروجی حاصل از این فرایند، به‌صورت برچسب‌هایی از برنامه‌های شناسایی‌شده است که در مرحله آموزش مدل طبقه‌بندی به کار گرفته می‌شوند. در واقع، ماژول DPI نقش یک منبع قابل‌اعتماد برای تولید داده‌های برچسب‌خورده را ایفا می‌کند و با فراهم‌سازی داده‌های دقیق و معتبر، زمینه‌ساز بهبود عملکرد و دقت طبقه‌بندی یادگیری عمیق می‌شود.

۳ .پردازش جریان‌ها (Stream Processor)

ماژول پردازش جریان، مسئول نگهداری صفی از جریان‌ها است که برخی از آن‌ها دارای برچسب DPI هستند. این ماژول شامل یک صف FIFO بدون حافظه است. هنگامی که صف پر می‌شود، سیگنال توقف برای ماژول استخراج ویژگی ارسال می‌شود و پس از خالی شدن صف، مجدداً فعال می‌گردد. در صورتی که نمونه‌های کافی برای آموزش و آزمون در دسترس نباشد، ماژول طبقه‌بندی نیز متوقف می‌شود و سپس با تأمین نمونه‌ها دوباره فعال می‌گردد.

۴. طبقه‌بندی مبتنی بر شبکه عصبی عمیق (DNN)

در این مؤلفه، یک شبکه عصبی عمیق (DNN) به‌عنوان هسته، طبقه‌بندی استفاده می‌شود. این شبکه با استفاده از زوج‌های (flow,label) که تو سط DPI برچسب‌گذاری شده‌اند، آموزش داده می شود و تابع طبقه‌بندی CFt را به صورت تدریجی به‌روز می‌کند. معماری شبکه، شامل چندین لایه‌ی کاملاً متصل (fully connected)و تابع فعال‌سازی ReLU است که به کمک تکنیک‌هایی مانند دراپ‌آوت و نرمال‌سازی دسته‌ای، دقت و تعمیم‌پذیری مدل افزایش یافته است.

شبکه عصبی طراحی‌شده، توانایی کار در دو حالت جریانی (Stream Mode) و دسته‌ای (Batch Mode) را دارد. در حالت جریانی، مدل به‌صورت برخط با داده‌های جدید به‌روزرسانی می‌شود و در حالت دسته‌ای، آموزش به‌صورت نوبتی با استفاده از مجموعه‌های آموزشی انجام می‌شود. به‌کارگیری DNN باعث شده تا مدل، به‌صورت خودکار روابط پیچیده بین ویژگی‌ها و نوع ترافیک

را یاد بگیرد و در طبقه‌بندی ترافیک‌های رمزنگاری‌شده عملکرد بالایی داشته باشد.

در این پژوهش، شبکه عصبی عمیق (DNN) به‌عنوان طبقه‌بند اصلی در ماژول طبقه‌بندی جریان‌ها مورد استفاده قرار گرفته است. دلیل اصلی انتخاب این مدل، توانایی آن در یادگیری روابط غیرخطی و پیچیده میان ویژگی‌های آماری جریان‌های شبکه و برچسب‌های مربوط به برنامه‌های کاربردی است؛ قابلیتی که به‌ویژه در مواجهه با ترافیک‌های رمزنگاری‌شده یا داده‌هایی با ابعاد بالا اهمیت پیدا می‌کند.

ورودی به مدل DNN برداری از ویژگی‌های آماری استخراج شده از جریان‌های شبکه است. این ویژگی‌ها توسط ابزار CICFlowMeter از ترافیک شبکه به دست می‌آیند و شامل اطلاعاتی نظیر مدت زمان جریان، تعداد بسته‌ها، حجم کل داده‌های ارسال و دریافت‌شده، میانگین طول بسته‌ها و نرخ ارسال داده هستند. برای افزایش کارایی مدل، تنها زیرمجموعه‌ای از این ویژگی‌ها که بیشترین تأثیر را در طبقه‌بندی دارند، انتخاب می‌شوند.

معماری شبکه عصبی مورد استفاده شامل چندین لایه کاملاً متصل (fully connected) است. لایه ورودی مطابق با تعداد ویژگی‌ها تعریف شده و در لایه‌های میانی، از توابع فعال‌سازی ReLU و لایه‌های Dropout برای جلوگیری از بیش‌برازش استفاده شده است. همچنین، در انتهای شبکه لایه‌ای با تابع Softmax برای خروجی چندکلاسه وجود دارد که برچسب احتمالی هر جریان را تعیین می‌کند. نرمال‌سازی ویژگی‌ها و استفاده از Batch Normalization در لایه‌های داخلی، موجب تسریع هم‌گرایی و افزایش پایداری آموزش می‌شود.

فرایند آموزش مدل با داده‌های دارای برچسب که از ماژول بازرسی عمیق بسته‌ها (DPI) به دست می‌آیند، به‌صورت دوره‌ای و پیوسته انجام می‌شود. این آموزش بر اساس تابع هزینه Cross-Entropy و با استفاده از الگوریتم Adam صورت می‌گیرد. در معماری پیشنهادی، مدل DNN قابلیت به‌روزرسانی تدریجی نیز دارد، به‌گونه‌ای که می‌تواند در طول زمان، خود را با الگوهای جدید ترافیکی وفق دهد.

استفاده از DNN در طبقه‌بندی جریان‌های شبکه، دقت طبقه‌بندی را به طور قابل‌توجهی افزایش می‌دهد و توانایی شناسایی برنامه‌ها در سناریوهای پیچیده و در حضور رمزنگاری را فراهم می‌سازد. همچنین، این مدل به دلیل ساختار تعمیم‌پذیر خود، به راحتی قابل استفاده در محیط‌های عملیاتی مختلف است و با تغییرات الگوی ترافیک، عملکرد خود را حفظ می‌کند.

نتایج ارزیابی

در این مطالعه، نتایج الگوریتم DNN با پنج الگوریتم مختلف طبقه‌بندی مورد ارزیابی قرار گرفته‌اند که عبارتند از: Naive Bayes (NB)، Random Hoeffding Tree (RHT)، Hoeffding Adaptive Tree (HAT)، Hoeffding Tree (HT)، و Adaptive Random Forest (ARF). برای ارزیابی این الگوریتم‌ها، از روش «آزمون دوره‌ای نگه‌داشته شده» استفاده شده است که طی آن داده‌ها به‌صورت پیوسته وارد سیستم شده و در فواصل زمانی معین با مدل ارزیابی می‌شوند.

داده‌های استفاده‌شده برای آموزش و ارزیابی، از دو دیتاست عمومی و شناخته شده در حوزه تحلیل ترافیک شبکه استخراج شده‌اند: دیتاست UNB ISCX VPN-nonVPN و دیتاست UNB ISCX Tor-nonTor. این دیتاست‌ها شامل ترافیک شبکه مربوط به کاربردهای مختلف از جمله مرور، انتقال فایل، چت، ایمیل، P۲P، استریمینگ و تماس صوتی هستند. برای مثال در دیتاست VPN-nonVPN، برنامه‌هایی مانند Skype، Facebook، Hangouts، AIM و BitTorrent در دسته‌های مختلف حضور دارند. دسته‌بندی برنامه‌ها در این دو مجموعه داده در جدول ۱ نشان داده شده است.

برای سنجش کیفیت طبقه‌بندی، از سه معیار رایج ارزیابی استفاده شده است: دقت (Accuracy)، صحت (Precision) و فراخوانی (Recall). معیار دقت میزان کلی صحت عملکرد الگوریتم را نشان می‌دهد و از نسبت مجموع نمونه‌های درست‌شناخته‌شده به کل نمونه‌ها محاسبه می‌شود. دقت‌عملیاتی نشان‌دهنده نسبت نمونه‌های صحیح شناسایی‌شده به کل نمونه‌های شناسایی‌شده در یک کلاس خاص است، در حالی که فراخوانی، توانایی الگوریتم در شناسایی تمام نمونه‌های واقعی یک کلاس را نشان می‌دهد.

جدول ۱: دسته‌بندی داده‌های در مجموعه داده ISCX VPN-nonVPN

کاربرد	دسته‌بندی
HTTPS	Browsing
Skype, SFTP, FTPS, SCP	File Transfer
Facebook, Hangouts, Skype, AIM, IC	Chat
SMTP, POP۳, IMAP	Email
Bittorrent	P۲P
YouTube, Vimeo	Streaming
Hangout, Skype voice call, Voip bust	VoIP

جدول زیر نتایج الگوریتم های مختلف را نشان می‌دهد.

ته بندی	دقت		صحت		یادآوری	
	عه داده VP	عه داده To	عه داده VP	عه داده To	عه داده VP	عه داده To
NB	۸۶	۶۲	۳۷	۴۲	۴۱	۴۴
RHT	۹۴	۸۳	۵۱	۷۰	۴۳	۵۶
HAT	۹۱	۶۷	۴۴	۶۱	۳۶	۴۱
HT	۸۷	۸۴	۳۸	۷۶	۳۹	۶۲
ARF	۹۷	۸۲	۷۳	۷۳	۵۶	۶۴
DNN	۹۶	۹۴	۸۱	۶۸	۷۷	۸۵

در ارز یابی عملکرد الگوریتم های طبقه‌بندی جر یان های ترافیکی رمزنگاری‌شـده، نتایج نشـان می‌دهد که شـبکه عصـبی عمیق (DNN) عملکرد بهتری نسبت به سایر الگوریتم‌ها دا شته ا ست. بر روی مجموعه‌داده VPN-nonVPN ، این مدل توانسـته با ثبت دقت ۹۶ درصد، صحت ۸۱ درصد و یادآوری ۷۷ درصد، بهترین نتایج را به‌دسـت آورد. این نشـان می‌دهد که DNN توانایی بالایی در شـناسـایی دقیق و درسـت برنامه‌ها حتی در شـرایطی که داده‌ها پیچیده یا به‌صورت بلادرنگ هستند، دارد.

در مجموعـه‌داده Tor-nonTor نیز DNN همچنـان عملکرد قابل‌قبولی دارد و با دقت ۹۴ درصـد، صـحت ۶۸ درصـد و یادآوری ۸۵ درصـد، از سـایر روش‌ها پیشـی گرفته اسـت. تفاوت در مقدار صحت بین دو مجموعه‌داده (VPN) و (Tor) را می‌توان به پیچیدگی بی شتر ترافیک در شبکه‌ی Tor نسبت داد که باعث کاهش نسبی صـحت شـده اسـت. با این وجود، مقدار بالای یادآوری در این مجموعه‌داده بیانگر توانایی بالای DNN در شـناسـایی نمونه‌های مثبت واقعی اسـت، حتی اگر درصـدی از نمونه‌ها به اشـتباه نیز طبقه‌بندی شده باشند.

در مقایسـه با روش‌های کلاسـیک مانند Naive Bayes (NB) و Hoeffding Tree (HT)، عملکرد DNN به‌مراتب پایدارتر و دقیق‌تر اسـت. برای مثال، در مجموعه VPN، NB تنها به دقت ۸۶ درصـد ر سیده و صحت آن ۳۷ در صد گزارش شده ا ست، که نشان‌دهندهٔ ضـعف آن در دسـته‌بندی صـحیح جریان‌ها در حضـور داده‌های پیچیده اسـت. همچنین الگوریتم‌هایی مانند Adaptive Random Forest (ARF)نیز عملکرد مناسـبی داشـته‌اند و در برخی معیارها مانند دقت در مجموعه VPN (۹۷ درصـد) به نتایج خوبی دسـت یافته‌اند، ولی همچنان در ترکیب کلی معیارها نسـبت به DNN پایین‌تر هستند.

به‌طور کلی، تحلیل نتایج نشان می‌دهد که استفاده از شبکه عصبی عمیق به عنوان طبقه‌بند یادگیرنده، در کنار به‌روزرسـانی انطباقی مدل با بهره‌گیری از داده‌های مرجع حاصـل از بازرسـی عمیق بسـته‌ها (DPI) ، رویکردی مؤثر برای طبقه‌بندی دقیق جریان‌های رمزنگاری‌شده در شبکه‌های مدرن به شمار می‌رود.

نتیجه‌گیری و پیشنهادات

در سامانه‌های تحلیل ترافیک شبکه به صورت بلادرنگ، شنا سایی دقیق و درون‌خطی (inline) برنامه‌های کاربردی شـبکه از اهمیت بالایی برخوردار است. با توجّه به این‌که شبکه‌های با پهنای باند بالا محیط‌هایی پویا و با جریان بی‌پایانی از ترافیک ایجاد می‌کنند، در این مقاله یک رویکرد نوین برای طبقه‌بندی جریان ترافیک شـبکه ارائه شده که از تکنیک بازرسی عمیق بسته‌ها (DPI) برای تشخیص اولیه بر نا مه های کاربردی بهره می‌برد و به‌طور تطبیقی مدل طبقه‌بندی را به‌روز سانی می‌کند. این سامانه از چهار ماژول ا صلی شـا مل اسـتخراج‌کننده ویژگی(Feature Extractor) ، بازرس درون‌خط بسـته‌ها(Inline Deep Packet Inspector) ، پردازش‌گر جریـان (Stream Processor) و طبقـه‌بنـد جریـان (Stream Classifier)تشکیل شده است.

در کنار روش‌های طبقه‌بندی کلاسیک، تمرکز ویژه‌ای بر به‌کارگیری شبکه‌های ع صبی عمیق (DNN) به‌عنوان طبقه‌بند ا صلی جریان‌ها انجام شده ا ست. شبکه‌های DNN به‌دلیل توانایی بالا در ا ستخراج ویژگی‌های غیرخطی و پیچیده از داده‌ها، قادرند الگوهای پنهان در جریان‌های رمزنگاری‌شـده یا غیر اسـتاندارد را نیز با دقت بالا شـناسـایی کنند. در این معماری، ویژگی‌های اسـتخراج‌شـده از جریان‌های شبکه به‌صورت برداری به مدل DNN منتقل می‌شوند و طبقه‌بندی به صورت بلادرنگ صورت می‌گیرد. مدل DNN می‌تواند با سـاختارهایی مانند Fully Connected Networks یا ترکیب آن باAutoEncoder ها یاCNN های یک‌بعدی تو سعه داده شود تا هم دقت طبقه‌بندی و هم انعطاف‌پذیری مدل در مواجهه با تغییرات ترافیکی حفظ شود.

پیاده‌سـازی این سـامانه با اسـتفاده از ابزار MOA صـورت گرفته و ارزیـابی آن بر روی دو مجموعـه‌داده عمومی-UNB ISCX VPN nonVPNو UNB ISCX Tor-nonTor انجام شده ا ست. در ارزیابی صورت‌گرفته، استفاده از DNN در برخی آزمایش‌ها نشان داد که در دسـته‌بندی جر یان های پیچیده‌تر یا در مواقعی که داده های برچسب‌خورده به‌صورت لحظه‌ای کم هستند، عملکرد بهتری نسبت به طبقه‌بندهای سنتی دارد.

در ادامه این پژوهش، هدف ما کاهش سربار پردازش DPI با کاهش نسـبت ترافیکی اسـت که باید توسـط این ماژول بررسـی شـود. بدین‌منظور، سامانه می‌تواند نرخ خطای طبقه‌بندی را برر سی کرده

و تنها در صـــورت نیاز، ترافیک را به ماژول DPI ارســـال نماید. همچنین، لازم ا ست این سامانه در یک محیط واقعی با پهنای باند بالا به‌صورت آزمایشی پیاده‌سازی شده و عملکرد آن در شناسایی درون‌خط و تطبیقی برنامه‌های کاربردی به‌صـــورت عملی ارزیابی شود. بهره‌گیری بیشتر از معماری‌های DNN در سطح عملیاتی نیز می‌تواند باعث بهبود چشم‌گیر عملکرد سیستم شود.

منابع

[۱]. R. Manivannan, S. Senthilkumar, "Intrusion detection system for network security using novel adaptive recurrent neural network-based fox optimizer concept," International Journal of Computational Intelligence Systems, vol. ۱۸, No. ۱, p. ۳۷, ۲۰۲۵.

[۲]. G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," Proceedings of the ۲nd International Conference on Information Systems Security and Privacy (ICISSP), pp. ٤٠٧–٤١٤, ٢٠١٦.

[۳]. M. Noferesti, R. Jalili, "Inline high-bandwidth network analysis using a robust stream clustering algorithm," IET Information Security, vol. ۱۳, No. ٥, pp. ٤٨٦-٤٩٥, ٢٠١٩.

[٤]. M. Noferesti, R. Jalili, "HB۲DS: A behavior-driven high-bandwidth network mining system," Journal of Systems and Software, vol.۱۲۷, pp. ۲٦٦–۲۷۷, ۲۰۱۷. https://doi.org/۱۰,۱۰۱٦/j.jss.۲۰۱٦,۰۹,۰۰٤

[٥]. W. Wang, M. Zhu, J. Wang, X. Zeng, Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," ۲۰۱۷ IEEE International Conference on Intelligence and Security Informatics (ISI), pp. ٤٣–٤٨, ۲۰۱۷. https://doi.org/۱۰,۱۱۰۹/ISI.۲۰۱۷,۸۰۰٤۸٦۲

[٦]. V. Tong, C. Dao, H. A. Tran, D. Tran, H. T. T. Binh, N. T. Hoang, T. X. Tran, "Encrypted Traffic Classification Through Deep Domain Adaptation Network With Smooth Characteristic Function,' IEEE Transactions on Network and Service Management, ۲۰۲٥.

[۷]. Z. Nazari, M. Noferesti, R. Jalili, " DSCA: An inline and adaptive application identification approach in encrypted network traffic," In Proceedings of the ۳rd international conference on cryptography, security and privacy, pp. ۳۹-٤۳, ۲۰۱۹.

[۸]. B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz, M. E. Karsligil, " Application identification via network traffic classification. ۲۰۱۷ International Conference on Computing," Networking and Communications (ICNC), pp. ۸٤۳–۸٤۸, ۲۰۱۷. https://doi.org/۱۰,۱۱۰۹/ICCNC.۲۰۱۷,۷۸۷٦۲٤۱

[۹]. S. Mali, M. Gujral, A. K. Cherukuri, M. G. Ms, " Encrypted Network Traffic Classification Using Intelligent Techniques," Cureus, vol. ۲, No. ۱, ۲۰۲٥.

[۱۰]. M. Noferesti, R. Jalili, " ACoPE: An adaptive semi-supervised learning approach for complex-policy enforcement in high-bandwidth networks," Computer Networks, vol.۱٦٦, ۱۰٦۹٤۳, ۲۰۲۰.

.

An Online and Adaptive Approach for Identifying Applications in Encrypted Network Traffic Using a Deep Neural Network Learning Model

Mohammad Amin Rastegar*, Masoud Noferesti
۱- Faculty of Engineering-Department of Computer Engineering-Sciences and Research Branch- Islamic Azad University-Tehran- Iran,.
۲- Faculty of Engineering-Department of Computer Engineering-Sciences and Research Branch- Islamic Azad University-Tehran- Iran
.

Abstract

With the growing use of encryption in network traffic, especially through protocols like TLS and SSL, the effectiveness of traditional application identification methods based on packet content analysis has significantly declined. In modern high-bandwidth networks, real-time and accurate application identification faces challenges not only in terms of speed, volume, and traffic diversity but also due to high resource costs. This study presents an adaptive approach for traffic stream classification that dynamically and continuously updates the classification model by leveraging Deep Packet Inspection (DPI) as a reference source. At the core of the proposed classifier, a Deep Neural Network (DNN) learning algorithm is employed, capable of accurately identifying complex and hidden patterns in encrypted traffic. The proposed method is empirically evaluated on two public datasets, ISCX VPN-nonVPN and ISCX Tor-nonTor, and the results indicate that DNN performs better in certain scenarios—particularly in classifying more complex flows or when real-time labeled data is scarce—compared to traditional classifiers.

Keywords: Application Identification, Network Traffic Classification, Stream Classification, Deep Neural Network