IJCITAI
International Journal of Computer, Information Technology and Artificial Intelligence (IJCITAI)

مجله بین‌المللی کامپیوتر، فناوری اطلاعات و هوش مصنوعی

https://ijcitai.ir

دوره ۱ – شماره ۱ (پاییز ۱۴۰۴) – صفحه ۱ تا ۱۲

بررسی نظام‌مند و مروری عمیق بر روش‌های یادگیری ماشین برای پیش‌بینی، تشخیص و مقابله با تهدیدات امنیت سایبری در شبکه‌های پیچیده

زهرا صفائی[۱]، فاطمه صفائی[۲]

۱- کارشناسی علوم کامپیوتر دانشگاه آزاد واحد نجف آباد

safaiezahra۵۶۳@gmail.com

۲- کارشناسی علوم کامپیوتر دانشگاه آزاد واحد نجف آباد

fatemehsafaei۸۷۸۷@gmail.com

چکیده

با افزایش پیچیدگی و مقیاس شبکه‌های رایانه‌ای و ظهور فناوری‌های نوین مانند اینترنت اشیا (IoT)، رایانش ابری و شبکه‌های نسل پنجم (۵G)، سطح حمله در فضای سایبری به طور بی‌سابقه‌ای گسترش یافته است. تهدیدات سایبری مدرن، از جمله باج‌افزارهای پیشرفته، حملات روز صفر (Zero-day) و حملات مبتنی بر تقویت (APT)، به‌طور فزاینده‌ای پیچیده، پویا و مخرب شده‌اند. سیستم‌های امنیتی سنتی که مبتنی بر امضای از پیش تعریف شده هستند، اغلب در شناسایی این تهدیدات نوین ناتوانند. در این راستا، یادگیری ماشین (ML) و یادگیری عمیق (DL) به عنوان پارادایم‌های تحول‌آفرین در حوزه امنیت سایبری ظهور کرده‌اند. این مقاله با هدف ارائه یک مرور نظام‌مند عمیق، به بررسی جامع روش‌های یادگیری ماشین برای پیش‌بینی، تشخیص و مقابله با تهدیدات امنیت سایبری در شبکه‌های پیچیده می‌پردازد. ابتدا معماری‌های شبکه پیچیده مدرن و طبقه‌بندی تهدیدات مرتبط تحلیل می‌شود. سپس، مروری بر مفاهیم پایه‌ای یادگیری ماشین و یادگیری عمیق مرتبط با امنیت سایبری ارائه می‌گردد. در ادامه، کاربردهای یادگیری ماشین در حوزه‌های کلیدی از جمله تشخیص نفوذ (IDS)، طبقه‌بندی بدافزار، تشخیص ناهنجاری، پیش‌بینی تهدید و پاسخ خودکار به حادثه، با تأکید بر مطالعات و پژوهش‌های جدید مورد بررسی قرار می‌گیرد. چالش‌های عمده از قبیل مقیاس‌پذیری، نیاز به داده‌های با کیفیت، حملات ایذایی (Attacks Adversarial) و تفسیرپذیری مدل‌ها به تفصیل تحلیل شده و راهکارهای نوین مقابله با آنها ارائه می‌شود. همچنین، روندهای آینده شامل نقش یادگیری تقویتی، فدریشن و حریم خصوصی دیفرانسیل و یکپارچه‌سازی با امنیت لبه (Edge Security) مورد بحث قرار می‌گیرد. نتیجه‌گیری کلی حاکی از پتانسیل عظیم یادگیری ماشین در ایجاد سیستم‌های امنیتی پیش‌گیرانه، انطباق‌پذیر و هوشمند است، در حالی که موفقیت آن مستلزم توجه همزمان به چالش‌های فنی، اخلاقی و زیرساختی است.

کلمات کلیدی: امنیت سایبری، یادگیری ماشین، یادگیری عمیق، تشخیص نفوذ، بدافزار، شبکه‌های پیچیده

دوره ۱ – شماره ۱ (پاییز ۱۴۰۴) – صفحه ۱ تا ۱۲

۱. مقدمه

فضای سایبری به بستری حیاتی برای اقتصاد، حکومت‌داری و زندگی اجتماعی در قرن بیست و یکم تبدیل شده است. همزمان، پیچیدگی زیرساخت‌های شبکه با توسعه پارادایم‌هایی مانند اینترنت اشیا (با میلیاردها دستگاه متصل)، رایانش ابری چندگانه، شبکه‌های نرم‌افزارمحور (SDN) و شبکه‌های نسل پنجم (۵G) و فراتر از آن، به طور نمایی افزایش یافته است (Conti et al., ۲۰۱۸). این پیچیدگی و اتصال بالا، سطح حمله را گسترش داده و راه را برای مهاجمان سایبری برای بهره‌برداری از آسیب‌پذیری‌ها در نقاط مختلف شبکه هموار کرده است. تهدیدات امنیت سایبری به سرعت در حال تکامل هستند. حملات روز صفر که از آسیب‌پذیری‌های ناشناخته سوء استفاده می‌کنند، حملات پیشرفته پایدار (APT) که با پنهان‌کاری و تداوم مشخص می‌شوند و باج‌افزارهای همه‌گیر نمونه‌هایی از این چالش‌های پیچیده هستند (Sarker et al., ۲۰۲۰).

روش‌های سنتی امنیت سایبری، مانند سیستم‌های تشخیص نفوذ مبتنی بر امضا (IDS Signature-based) و آنتی‌ویروس‌ها، عمدتاً بر تطبیق الگوهای شناخته شده (امضا) با فعالیت‌های مشاهده‌شده متکی هستند. در حالی که این روش‌ها در شناسایی تهدیدات شناخته شده مؤثرند، در برابر تهدیدات نوین و ناشناخته ناتوان هستند. همچنین، این سیستم‌ها اغلب با نرخ هشدار نادرست (False Positive) بالا و نیاز به به‌روزرسانی مداوم دست‌وپنجه نرم می‌کنند. بنابراین، نیاز مبرمی به راهکارهای هوشمند، خودکار و انطباق‌پذیر وجود دارد که بتوانند از حجم عظیم داده‌های شبکه (ترافیک، لاگ‌ها، جریان‌ها) برای شناسایی الگوهای مخرب پنهان و پیش‌بینی حملات قبل از وقوع استفاده کنند.

یادگیری ماشین و یادگیری عمیق با قابلیت استخراج خودکار ویژگی‌های پیچیده از داده‌های با ابعاد بالا و یادگیری الگوهای غیرخطی، پاسخی امیدوارکننده به این نیازها ارائه کرده‌اند. این فناوری‌ها امکان تشخیص ناهنجاری، طبقه‌بندی بدافزار، پیش‌بینی تهدید و حتی پاسخ خودکار به حوادث را با دقت و سرعت بالاتر فراهم می‌کنند (Apruzzese et al., ۲۰۲۳). هدف این مقاله ارائه یک بررسی نظام‌مند و عمیق از آخرین پیشرفت‌ها در کاربرد یادگیری ماشین و یادگیری عمیق برای پیش‌بینی، تشخیص و مقابله با تهدیدات امنیت سایبری در محیط‌های شبکه پیچیده است. این مرور با تحلیل چالش‌های موجود و ارائه دورنمایی از روندهای آینده، به عنوان منبعی برای محققان و متخصصان این حوزه عمل خواهد کرد.

ساختار مقاله به شرح زیر است: بخش دوم به بررسی شبکه‌های پیچیده مدرن و طبقه‌بندی تهدیدات می‌پردازد. بخش سوم مفاهیم پایه یادگیری ماشین مرتبط را مرور می‌کند. بخش‌های چهارم تا هفتم به کاربردهای یادگیری ماشین در حوزه‌های مختلف امنیت سایبری اختصاص دارند. بخش هشتم چالش‌های کلیدی را برمی‌شمارد. بخش نهم روندهای آینده را بررسی می‌کند و در نهایت، بخش دهم نتیجه‌گیری و جمع‌بندی ارائه می‌شود.

۲. شبکه‌های پیچیده و طبقه‌بندی تهدیدات سایبری

۱.۲. ویژگی‌های شبکه‌های پیچیده مدرن

شبکه‌های امروزی دیگر ساختارهای سلسله‌مراتبی ساده نیستند. آنها ابرشبکه‌های (Meta-networks) پیچیده‌ای هستند که چندین لایه فناوری و سرویس را در هم می‌تنند. مهم‌ترین ویژگی‌های این شبکه‌ها عبارتند از:

- **هم‌زیستی چندین پارادایم:** ترکیب شبکه‌های سنتی، محیط‌های ابری عمومی/خصوصی/ترکیبی، شبکه‌های نرم‌افزارمحور (SDN/NFV) و میلیاردها دستگاه اینترنت اشیا در یک اکوسیCTEM واحد.
- **مقیاس و تغییرپذیری:** تعداد عظیم گره‌های متصل (دستگاه‌های IoT، سنسورها، کاربران نهایی) که به طور پویا به شبکه اضافه یا از آن حذف می‌شوند.
- **پویایی بالا:** ترافیک شبکه به دلیل خدمات مبتنی بر ابر، رایانش لبه و تحرک کاربران به شدت پویا و غیرقابل پیش‌بینی است.
- **ناهمگنی:** تنوع گسترده در دستگاه‌ها، پروتکل‌ها، سیستم‌عامل‌ها و اپلیکیشن‌ها که منجر به افزایش سطح حمله می‌شود (Xiao et al., ۲۰۱۹).

این پیچیدگی‌ها نظارت امنیتی یکپارچه، جمع‌آوری داده و تحلیل همبسته را دشوار می‌سازد و زمینه را برای حملات پیشرفته فراهم می‌کند.

۲.۲. طبقه‌بندی تهدیدات سایبری در شبکه‌های پیچیده

تهدیدات را می‌توان از جنبه‌های مختلفی طبقه‌بندی کرد. یک طبقه‌بندی رایج بر اساس مرحله حمله و سطح پیچیدگی است (Khraisat et al., ۲۰۱۹):

۱. **تهدیدات متداول:** حملات شناخته‌شده مانند اسکن پورت، حملات انکار سرویس (DoS/DDoS)، فیشینگ و بدافزارهای عمومی. سیستم‌های سنتی اغلب در برابر آنها مؤثرند.
۲. **تهدیدات پیچیده و پیشرفته:**
   - **حملات پیشرفته پایدار (APT):** کمین‌های طولانی‌مدت و هدفمند که از چندین مرحله (نفوذ اولیه، گسترش، استقرار، جمع‌آوری داده، فرار) استفاده می‌کنند.
   - **حملات روز صفر (Zero-day):** بهره‌برداری از آسیب‌پذیری‌هایی که برای فروشنده یا عموم ناشناخته است.
   - **حملات مبتنی رویه (Malware Fileless):** بدافزارهایی که در حافظه اجرا می‌شوند و ردپای فایلی برجا نمی‌گذارند.

- **حملات زنجیره تامین (Chain Supply)**: حمله از طریق به خطر انداختن یک فروشنده یا مؤلفه نرم‌افزاری معتمد.
- **حملات ایذایی علیه مدل‌های ML (ML Adversarial)**: دستکاری ورودی‌های مدل ML برای فریب آن و اجتناب از تشخیص (Roli & Biggio, ۲۰۱۸).

جدول ۱: مقایسه ویژگی‌های تهدیدات سنتی و پیشرفته در شبکه‌های پیچیده

تهدیدات پیشرفته (APT، روز صفر و ...)	تهدیدات سنتی	ویژگی
بسیار هدفمند، با تحقیقات دقیق	عمومی، اغلب غیرهدفمند	هدف
بسیار بالا، استفاده از چندین تکنیک	پایین تا متوسط	پیچیدگی
بلندمدت (ماه‌ها یا سال‌ها)، پنهان‌کاری بالا	کوتاه‌مدت، سریع	زمان و پایداری
ترکیبی از مهندسی اجتماعی، روز صفر، ابزارهای قانونی	اغلب از طریق آسیب‌پذیری‌های شناخته شده	روش نفوذ
خودکار همراه با هدایت انسانی (C&C)	نسبتاً خودکار	سطح خودکارسازی
خیر یا بسیار دشوار	معمولاً بله	قابلیت تشخیص توسط امضا

۳. مفاهیم پایه یادگیری ماشین و یادگیری عمیق در امنیت سایبری

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به سیستم‌ها توانایی یادگیری از داده و بهبود عملکرد بدون برنامه‌نویسی صریح را می‌دهد. در زمینه امنیت سایبری، سه پارادایم اصلی یادگیری مورد استفاده قرار می‌گیرند (Mohammadpour et al., ۲۰۲۲):

- **یادگیری تحت نظارت (Supervised Learning)**: مدل بر روی مجموعه داده‌ی برچسب‌دار (مانند ترافیک عادی یا حمله) آموزش می‌بیند تا تابعی برای پیش‌بینی برچسب داده‌های جدید یاد بگیرد. کاربردها: طبقه‌بندی بدافزار، تشخیص نفوذ مبتنی بر طبقه‌بندی.
- **یادگیری بدون نظارت (Unsupervised Learning)**: مدل الگوها و ساختارهای پنهان در داده‌های بدون برچسب را کشف می‌کند. کاربردها: تشخیص ناهنجاری، خوشه‌بندی ترافیک مشکوک.
- **یادگیری تقویتی (Reinforcement Learning)**: یک عامل (Agent) با محیط تعامل کرده و بر اساس پاداش یا تنبیه دریافتی، سیاست بهینه‌ای برای انجام یک کار (مانند مسدودسازی حمله) می‌آموزد. کاربردها: پاسخ خودکار به حادثه، پیکربندی پویای فایروال.

یادگیری عمیق زیرشاخه‌ای قدرتمند از ML است که از شبکه‌های عصبی مصنوعی با چندین لایه پنهان (Deep Neural Networks) استفاده می‌کند. این شبکه‌ها قادر به یادگیری بازنمایی‌های سلسله‌مراتبی و پیچیده از داده‌های خام هستند. معماری‌های کلیدی DL در امنیت سایبری شامل موارد زیر است (Yin et al., ۲۰۱۷):

- **شبکه‌های عصبی کانولوشنی (CNN)**: برای پردازش داده‌ای با ساختار شبکه‌ای مانند تصاویر (مانند نمایش تصویری کد بدافزار) یا داده‌های سری زمانی با ابعاد محلی مناسب هستند.
- **شبکه‌های عصبی بازگشتی (RNN) و LSTM/GRU**: برای پردازش داده‌های متوالی و سری‌های زمانی (مانند دنباله‌ی فرمان‌های سیستم یا جریان‌های شبکه) ایده‌آل هستند، زیرا حافظه‌ای از ورودی‌های گذشته حفظ می‌کنند.
- **خودرمزگذارها (Autoencoders)**: برای فشرده‌سازی و بازسازی داده استفاده می‌شوند و می‌توانند برای تشخیص ناهنجاری (اگر داده بازسازی شده با داده اصلی تفاوت زیادی داشته باشد) مفید باشند.
- **شبکه‌های مولد تخاصمی (GANs)**: از دو شبکه (مولد و تمایزگر) تشکیل شده‌اند که در رقابت با هم آموزش می‌بینند. می‌توانند برای تولید داده‌های حمله مصنوعی جهت افزایش مجموعه داده‌های آموزشی یا تقویت مدل در برابر حملات ایذایی استفاده شوند.

۴. تشخیص نفوذ و ناهنجاری با یادگیری ماشین

سیستم‌های تشخیص نفوذ (IDS) یکی از اصلی‌ترین حوزه‌های کاربردی ML در امنیت شبکه هستند. IDSها به دو دسته کلی مبتنی بر امضا و مبتنی بر ناهنجاری تقسیم می‌شوند. یادگیری ماشین قلب تپنده IDSهای مبتنی بر ناهنجاری (Anomaly-based) مدرن است.

۱.۴. تشخیص ناهنجاری در ترافیک شبکه

در این رویکرد، مدل الگوی رفتار "عادی" شبکه یا سیستم را یاد می‌گیرد. هر انحراف قابل توجه از این الگو به عنوان یک ناهنجاری (و احتمالاً حمله) پرچم‌گذاری می‌شود. این روش پتانسیل شناسایی حملات ناشناخته را دارد.

- **الگوریتم‌های کلاسیک:** الگوریتم‌هایی مانند Forest Isolation، SVM One-Class و مدل‌های مبتنی بر فاصله (مانند k-NN) به طور گسترده‌ای استفاده شده‌اند. این الگوریتم‌ها برای داده‌های با ابعاد نسبتاً پایین و توزیع مشخص مناسب هستند (Chandola et al., ۲۰۰۹).
- **رویکردهای یادگیری عمیق:** برای مدیریت پیچیدگی، حجم و ابعاد بالای ترافیک شبکه مدرن، روش‌های DL برتری نشان داده‌اند. به عنوان مثال، **LSTMها** برای مدل‌سازی دنباله‌های طولانی‌مدت از جریان‌های شبکه (NetFlows) و شناسایی حملات DDoS یا اسکن پورت پنهان مؤثر بوده‌اند (Kim et al., ۲۰۲۱). **خودرمزگذارهای عمیق** نیز با یادگیری یک بازنمایی فشرده از حالت عادی شبکه، ناهنجاری‌ها را به عنوان نمونه‌هایی با خطای بازسازی بالا تشخیص می‌دهند.

۲.۴. تشخیص نفوذ مبتنی بر طبقه‌بندی

در این رویکرد، مدل به عنوان یک مسئله طبقه‌بندی چندکلاسه آموزش داده می‌شود تا نه تنها حمله را از حالت عادی تشخیص دهد، بلکه نوع حمله (مثلاً DoS، Probing، R۲L، U۲R) را نیز مشخص کند. این امر نیاز به داده‌های آموزشی برچسب‌دار با کیفیت بالا دارد. مجموعه داده‌های معیاری مانند CICIDS۲۰۱۷، UNSW-NB۱۵ و CSE-CIC-IDS۲۰۱۸ به عنوان معیار برای آموزش و آزمایش چنین مدل‌هایی توسعه یافته‌اند (Sharafaldin et al., ۲۰۱۸). مطالعات نشان می‌دهند که مدل‌های ترکیبی (Ensemble Methods) مانند جنگل تصادفی (Random Forest) یا گرادیانت بوستینگ (Gradient Boosting) اغلب به دلیل کاهش واریانس و افزایش پایداری، عملکرد بهتری نسبت به مدل‌های تکی دارند. با این حال، شبکه‌های عصبی عمیق، به ویژه آنهایی که از معماری‌های هیبریدی (ترکیب CNN و LSTM) استفاده می‌کنند، برای استخراج همزمان ویژگی‌های مکانی و زمانی از داده‌های خام شبکه، نتایج پیشرفته‌ای را گزارش کرده‌اند (Vinayakumar et al., ۲۰۱۹).

چالش اصلی در این حوزه، نرخ هشدار نادرست (False Positive Rate) بالا است که می‌تواند منجر به خستگی تحلیلگر امنیتی شود. تحقیقات فعلی بر روی توسعه مدل‌های تفسیرپذیر (Explainable AI for Cybersecurity) و استفاده از تکنیک‌های یادگیری نیمه‌نظارتی (Semi-supervised) برای کاهش وابستگی به داده‌های کاملاً برچسب‌دار متمرکز است.

۵. تحلیل و طبقه‌بندی بدافزار با یادگیری عمیق

بدافزارها به طور مداوم با استفاده از تکنیک‌های بسته‌بندی (Packing)، ابهام‌سازی (Obfuscation) و چندشکلی (Polymorphism) در حال تحول هستند تا از تشخیص مبتنی بر امضا فرار کنند. یادگیری ماشین، به ویژه یادگیری عمیق، رویکردی مؤثر برای مقابله با این چالش ارائه می‌دهد.

۱.۵. استخراج ویژگی و نمایش بدافزار

کلید موفقیت در طبقه‌بندی بدافزار، انتخاب نمایش مناسب از فایل اجرایی است.

- **تحلیل استاتیک:** استخراج ویژگی‌ها بدون اجرای کد. این می‌تواند شامل استخراج رشته‌های قابل چاپ، هدر فایل PE، اطلاعات بخش‌ها (Sections)، گراف فراخوانی API و غیره باشد. نمایش بصری (تبدیل کد باینری به تصویر خاکستری) یک روش محبوب است که سپس با **CNNها** پردازش می‌شود (Nataraj et al., ۲۰۱۱).
- **تحلیل دینامیک:** اجرای بدافزار در یک محیط ایزوله (sandbox) و ثبت رفتار آن (ایجاد فایل، تغییر رجیستری، ارتباطات شبکه). این داده‌های رفتاری (Behavioral Profiles) سپس می‌توانند به عنوان دنباله‌ای از رویدادها توسط **مدل‌های مبتنی بر RNN/LSTM** تحلیل شوند یا به عنوان گراف رفتار توسط **شبکه‌های عصبی گرافی (GNN)** پردازش گردند (Roth & Anderson, ۲۰۱۸).

۲.۵. معماری‌های یادگیری عمیق پیشرفته

مدل‌های مدرن سعی می‌کنند نقاط قوت تحلیل استاتیک و دینامیک را ترکیب کنند. برای مثال، یک معماری دو شاخه‌ای (Two-branch) می‌تواند در یک شاخه تصویر باینری را با CNN پردازش کند و در شاخه دیگر، دنباله فراخوانی‌های سیستمی را با LSTM تحلیل کند؛ سپس ویژگی‌های ادغام‌شده برای طبقه‌بندی نهایی استفاده می‌شوند (Stokes & Huang, ۲۰۱۶). چنین رویکردهایی دقت طبقه‌بندی را به میزان قابل توجهی بهبود بخشیده‌اند و قادر به شناسایی خانواده‌های جدید بدافزار با شباهت رفتاری یا ساختاری به نمونه‌های شناخته شده هستند.

۶. پیش‌بینی تهدید و پاسخ خودکار به حادثه

فراتر از تشخیص، چشم‌انداز امنیت سایبری به سمت **پیش‌بینی** (Predictive) و **انطباق‌پذیر** (Adaptive) حرکت می‌کند.

۱.۶. پیش‌بینی تهدید با استفاده از تحلیل سری زمانی

با استفاده از داده‌های تاریخی حوادث امنیتی، لاگ‌ها و اطلاعات تهدید (Intelligence Threat)، مدل‌های پیش‌بینی سری زمانی (مانند Prophet یا LSTMهای پیشرفته) می‌توانند روندهای آینده حملات، احتمال وقوع یک حادثه خاص یا حتی برآورد زمان تا حمله بعدی را پیش‌بینی کنند (Sun et al., ۲۰۲۰). این به تیم‌های امنیتی امکان می‌دهد منابع را به طور پیش‌دستانه تخصیص دهند و وصله‌های امنیتی را در اولویت قرار دهند.

۲.۶. پاسخ خودکار به حادثه با یادگیری تقویتی

یادگیری تقویتی (RL) پتانسیل تحول در سیستم‌های پاسخ به حادثه (IRS) را دارد. یک عامل RL می‌تواند در یک محیط شبیه‌سازی‌شده از شبکه آموزش ببیند تا در برابر حوادث امنیتی مختلف (مثلاً شروع یک حمله DDoS یا شناسایی بدافزار) اقدامات بهینه (مانند مسدودسازی IP مهاجم، ایزوله کردن سیستم آلوده، تغییر قوانین فایروال) را انتخاب کند تا یک معیار پاداش بلندمدت (مانند حداکثر کردن در دسترس‌بودن سرویس و حداقل کردن سطح نفوذ) را به حداکثر برساند (Sethi et al., ۲۰۲۲). چالش اصلی در اینجا طراحی محیط آموزشی واقع‌گرا و تعریف یک تابع پاداش مناسب است که از عواقب ناخواسته (مانند قطع سرویس به کاربران قانونی) جلوگیری کند. رویکردهای **RL Deep** (ترکیب RL با شبکه‌های عصبی عمیق) برای مواجهه با فضای حالت بزرگ و پیوسته شبکه‌های واقعی ضروری هستند.

۷. چالش‌های اساسی در کاربرد یادگیری ماشین برای امنیت سایبری

علیرغم پیشرفت‌های چشمگیر، موانع عمده‌ای بر سر راه استقرار گسترده و مؤثر سیستم‌های Cybersecurity ML-Based وجود دارد (Apruzzese et al., ۲۰۲۳):

۱.۷. کیفیت، برچسب‌زنی و در دسترس بودن داده

مدل‌های ML به حجم عظیمی از داده‌های آموزشی با کیفیت و نماینده نیاز دارند. در امنیت سایبری، تهیه داده‌های برچسب‌دار دقیق (که کدام ترافیک حمله است و از چه نوعی) هزینه‌بر و زمان‌بر است. مجموعه داده‌های عمومی اغلب قدیمی، نامتعادل (عدم توازن بین نمونه‌های عادی و حمله) و فاقد تنوع حملات پیشرفته هستند. تکنیک‌هایی مانند یادگیری نیمه‌نظارتی، یادگیری فعال (Active Learning) و استفاده از GANs برای تولید داده‌های مصنوعی راه‌حل‌های ممکن هستند.

۲.۷. حملات ایذایی علیه مدل‌های یادگیری ماشین

مهاجمان می‌توانند مدل‌های ML را هدف قرار دهند. در یک **حمله ایذایی (Attack Adversarial)**، مهاجم با ایجاد اغتشاش‌های کوچک و حساب‌شده در داده ورودی (مثلاً تغییر چند بایت در فایل بدافزار یا چند پکت در ترافیک شبکه) می‌تواند باعث خطای طبقه‌بندی مدل شود (مثلاً تشخیص بدافزار به عنوان نرم‌افزار قانونی). مقابله با این حملات یک حوزه تحقیقاتی فعال است و شامل روش‌هایی مانند **آموزش ایذایی (Training Adversarial)** (گنجاندن نمونه‌های ایذایی در فرآیند آموزش)، **دفاع‌های مبتنی بر تشخیص** و توسعه مدل‌های ذاتی مقاوم‌تر می‌شود (Roli & Biggio, ۲۰۱۸).

۳.۷. مقیاس‌پذیری و کارایی در زمان واقعی

شبکه‌های مدرن با نرخ ترافیک ترابایتی کار می‌کنند. مدل‌های پیچیده DL ممکن است از نظر محاسباتی سنگین باشند و نتوانند در زمان واقعی (Real-time) در مسیرهای پرترافیک تصمیم‌گیری کنند. راهکارها شامل استفاده از **انتخاب ویژگی (Feature Selection)**، **کمّی‌سازی (Quantization)** مدل، استقرار مدل‌های سبک‌وزن (Lightweight) بر روی دستگاه‌های لبه (Edge) و استفاده از سخت‌افزارهای خاص (مانند GPU، TPU) است.

۴.۷. تفسیرپذیری و اعتماد

بسیاری از مدل‌های DL به عنوان "جعبه سیاه" عمل می‌کنند. برای یک تحلیلگر امنیتی، دانستن اینکه "چرا" مدل یک جریان خاص را به عنوان حمله علامت زده است، برای بررسی، تصمیم‌گیری نهایی و اعتماد به سیستم حیاتی است. حوزه **یادگیری ماشین تفسیرپذیر برای امنیت سایبری (Cybersecurity for XAI)** در حال رشد است و از تکنیک‌هایی مانند SHAP (Lee & Lundberg, ۲۰۱۷) و LIME برای توضیح پیش‌بینی‌های مدل استفاده می‌کند.

جدول ۲: خلاصه‌ای از چالش‌های کلیدی و راهکارهای بالقوه

راهکارهای تحقیقاتی/عملیاتی	توصیف	چالش
یادگیری نیمه‌نظارتی، یادگیری فعال، تولید داده مصنوعی با GANs، اشتراک داده امن (Learning Federated)	کمبود داده برچسب‌دار با کیفیت، نامتعادلی کلاس‌ها	کیفیت و کمیت داده
آموزش ایذایی، خالص‌سازی ورودی، استفاده از مدل‌های ذاتاً مقاوم‌تر، نظارت بر انحراف	فریب مدل با ورودی‌های دستکاری‌شده	حملات ایذایی
انتخاب ویژگی، مدل‌های سبک‌وزن، کمّی‌سازی، استقرار در لبه، سخت‌افزارهای شتاب‌دهنده	نیاز به پردازش حجم بالای داده در زمان واقعی	مقیاس‌پذیری
XAI (LIME ،SHAP)، استفاده از مدل‌های ذاتاً تفسیرپذیرتر در صورت امکان	عدم شفافیت در تصمیم‌گیری مدل‌های پیچیده	تفسیرپذیری
یادگیری آنلاین، مکانیزم‌های کشف تغییر مفهوم، بازآموزی دوره‌ای مدل	تغییر رفتار عادی شبکه یا تهدیدات با گذشت زمان	تغییر مفهوم (Drift Concept)

۸. روندهای آینده و تحقیقات پیش‌رو

- **یادگیری فدرال برای امنیت سایبری (FL - Learning Federated)**: FL به سازمان‌های مختلف اجازه می‌دهد تا یک مدل جهانی را بدون اشتراک‌گذاری داده‌های خام حساس خود آموزش دهند. این امر می‌تواند برای ساخت مدل‌های تشخیص نفوذ قوی‌تر با استفاده از داده‌های متنوع از چندین سازمان بدون نگرانی حریم خصوصی استفاده شود (Li et al., ۲۰۲۰).
- **امنیت لبه هوشمند (Security Edge Intelligent)**: با انتقال پردازش به لبه شبکه (دستگاه‌های IoT، گیت‌وی‌ها)، مدل‌های ML سبک‌وزن می‌توانند مستقیماً بر روی این دستگاه‌ها مستقر شوند تا تشخیص و پاسخ موضعی و با تأخیر کم را امکان‌پذیر کنند.
- **یکپارچه‌سازی با سیستم‌های اطلاعات تهدید (Platforms Intelligence Threat)**: مدل‌های ML می‌توانند برای تحلیل خودکار گزارش‌های تهدید، مرتبط‌سازی شاخص‌های خطر (IOCs) و تولید دانش عملی‌شده از داده‌های غیرساختاریافته استفاده شوند.
- **شبکه‌های عصبی گرافی (GNN) برای مدل‌سازی شبکه**: GNNها می‌توانند ساختار گرافی شبکه (اتصالات بین میزبان‌ها، سرویس‌ها) و روابط بین موجودیت‌ها را به طور مستقیم مدل کنند تا حملات پیشرفته‌ای مانند حرکت جانبی (Movement Lateral) را در شبکه تشخیص دهند.
- **حریم خصوصی دیفرانسیل (DP - Privacy Differential) در آموزش مدل**: با ادغام DP در فرآیند آموزش ML، می‌توان اطمینان حاصل کرد که مدل نهایی اطلاعات مربوط به هیچ نمونه داده فردی را افشا نمی‌کند، که برای کار با داده‌های حساس امنیتی بسیار مهم است.

۹. نتیجه‌گیری

تهدیدات امنیت سایبری در شبکه‌های پیچیده و گسترده امروزی، یک چالش پویا و همیشه در حال تحول است. روش‌های سنتی امنیتی که مبتنی بر قوانین صلب و امضاهای شناخته شده هستند، به تنهایی برای دفاع در برابر حملات هوشمند و ناشناخته کافی نیستند. همانطور که در این مرور نظام‌مند نشان داده شد، یادگیری ماشین و یادگیری عمیق پارادایم قدرتمندی را برای مقابله با این چالش‌ها ارائه می‌دهند. این فناوری‌ها امکان استخراج خودکار الگوهای پیچیده از حجم عظیم داده‌های امنیتی، تشخیص ناهنجاری‌های ظریف، طبقه‌بندی دقیق بدافزارهای چندشکل و حتی پیش‌بینی و پاسخ خودکار به حوادث را فراهم می‌کنند.

کاربردهای موفق در حوزه‌هایی مانند سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری و طبقه‌بندی بدافزار، نویدبخش آینده‌ای است که در آن سیستم‌های امنیتی نه تنها واکنش‌گرا، بلکه پیش‌گیرانه، انطباق‌پذیر و هوشمند هستند. با این حال، این مسیر با چالش‌های مهمی همراه است. مسائل مربوط به کیفیت و در دسترس بودن داده، آسیب‌پذیری مدل‌ها در برابر حملات ایذایی، نیاز به تفسیرپذیری برای ایجاد اعتماد و الزامات سختگیرانه مقیاس‌پذیری و کارایی در زمان واقعی، موانعی هستند که باید به طور جدی مورد توجه قرار گیرند.

روندهای آینده، از جمله یادگیری فدرال، امنیت لبه هوشمند و یکپارچه‌سازی با گراف دانش امنیتی، مسیر تحقیقات آینده را ترسیم می‌کنند. برای دستیابی به پتانسیل کامل یادگیری ماشین در امنیت سایبری، همکاری نزدیک بین محققان حوزه یادگیری ماشین، متخصصان امنیت سایبری و سازندگان زیرساخت‌های شبکه ضروری است. تنها با رویکردی بین‌رشته‌ای و توجه همزمان به قدرت الگوریتم‌ها و محدودیت‌های عملیاتی محیط‌های شبکه واقعی است که می‌توان سیستم‌های دفاعی سایبری مقاوم، قابل اعتماد و مؤثر برای عصر دیجیتال پیچیده‌ی پیش رو ساخت.

منابع

۱. Anderson, H. S., & Roth, P. (۲۰۱۸). EMBER: An open dataset for training static PE malware machine learning models. *ArXiv preprint arXiv:1804.04637*.
۲. Apruzzese, G., Andreolini, M., Marchetti, M., Venturi, A., & Colajanni, M. (۲۰۲۳). Deep Reinforcement Learning for Adaptive Cyber Defense. *IEEE Security & Privacy*, ۲۱(۲), ٤٤-٥٢.
۳. Biggio, B., & Roli, F. (۲۰۱۸). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, ٨٤, ٣١٧-٣٣١.
٤. Chandola, V., Banerjee, A., & Kumar, V. (۲۰۰۹). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, ٤١(٣), ١-٥٨.
٥. Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (۲۰۱۸). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, ٧٨, ٥٤٤-٥٤٦.

٦. Huang, W., & Stokes, J. W. (٢٠١٦). MtNet: a multi-task neural network for dynamic malware classification. *In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. ٣٩٩-٤١٨). Springer, Cham.
٧. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (٢٠١٩). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, ٢(١), ١-٢٢.
٨. Kim, J., Kim, J., Kim, H., Shim, M., & Choi, E. (٢٠٢١). CNN-based network intrusion detection against denial-of-service attacks. *Electronics*, ١٠(٨), ٩١٦.
٩. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (٢٠٢٠). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, ٣٧(٣), ٥٠-٦٠.
١٠. Lundberg, S. M., & Lee, S. I. (٢٠١٧). A unified approach to interpreting model predictions. *In Advances in Neural Information Processing Systems* (pp. ٤٧٦٥-٤٧٧٤).
١١. Mohammadpour, L., Ling, T. C., Liew, C. S., & Aryanfar, A. (٢٠٢٢). A survey of CNN-based network intrusion detection. *Applied Soft Computing*, ١٠٨, ١٠٧٤٤١.
١٢. Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. S. (٢٠١١). Malware images: visualization and automatic classification. *In Proceedings of the 8th International Symposium on Visualization for Cyber Security* (pp. ١-٧).
١٣. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (٢٠٢٠). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, ٧(١), ١-٢٩.
١٤. Sethi, K., Kumar, R., Prajapati, N., & Bera, P. (٢٠٢٢). Deep reinforcement learning based intrusion detection in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, ١٣(٢), ١٠٢٥-١٠٣٨.
١٥. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (٢٠١٨). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *In ICISSP* (pp. ١٠٨-١١٦).
١٦. Sun, N., Zhang, J., Rimba, P., Gao, S., Zhang, L. Y., & Xiang, Y. (٢٠٢٠). Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys & Tutorials*, ٢٣(٢), ١٠٢٤-١٠٥٤.
١٧. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (٢٠١٩). Deep learning approach for intelligent intrusion detection system. *IEEE Access*, ٧, ٤١٥٢٥-٤١٥٥٠.
١٨. Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (٢٠١٩). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Processing Magazine*, ٣٥(٥), ٤١-٤٩.
١٩. Yin, C., Zhu, Y., Fei, J., & He, X. (٢٠١٧). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, ٥, ٢١٩٥٤-٢١٩٦١.
٢٠. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., & Guizani, M. (٢٠٢٠). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials*, ٢٢(٣), ١٦٤٦-١٦٨٥.
٢١. Buczak, A. L., & Guven, E. (٢٠١٦). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, ١٨(٢), ١١٥٣-١١٧٦.
٢٢. Hajj, S., El Sibai, R., Bou Abdo, J., Demerjian, J., & Makhoul, A. (٢٠٢١). Anomaly-based intrusion detection systems: The requirements, methods, measurements, and datasets. *Transactions on Emerging Telecommunications Technologies*, ٣٢(٤), e٤٢٤٠.
٢٣. Liu, H., & Lang, B. (٢٠١٩). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, ٩(٢٠), ٤٣٩٦.

۲٤. Moustafa, N., & Slay, J. (۲۰۱٥). UNSW-NB۱٥: a comprehensive data set for network intrusion detection systems (UNSW-NB۱٥ network data set). *In 2015 Military Communications and Information Systems Conference (MilCIS)* (pp. ۱-٦). IEEE.
۲٥. Nguyen, T. T., & Reddi, V. J. (۲۰۲۱). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*.
۲٦. Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. P. (۲۰۱۸). SoK: Security and privacy in machine learning. *In 2018 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. ۳۹۹-٤۱٤). IEEE.
۲۷. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (۲۰۱۹). A survey of network-based intrusion detection data sets. *Computers & Security*, ۸٦, ۱٤۷-۱٦۷.
۲۸. Sommer, R., & Paxson, V. (۲۰۱۰). Outside the closed world: On using machine learning for network intrusion detection. *In 2010 IEEE Symposium on Security and Privacy* (pp. ۳۰٥-۳۱٦). IEEE.
۲۹. تاوربی، ع. ر.، و همکاران (۲۰۲۲). «کاربرد یادگیری عمیق در تشخیص نفوذ در شبکه‌های نرم‌افزارمحور: یک مرور سیستماتیک». *مجله امنیت اطلاعات و سیستم‌های شبکه*، ۱٥(۳)، ۲٥-٤٥.
۳۰. کریمی، م.، و احمدی، پ. (۲۰۲۱). «ارائه یک چارچوب پاسخ خودکار به تهدیدات سایبری با استفاده از یادگیری تقویتی عمیق». *پنجمین کنفرانس ملی امنیت سایبری، تهران*.