ارائه مدلی مبتنی بر معنا برای تشخیص بلادرنگ حملات APT با استفاده از تحلیل همبستگی رویداد

ساره اسلامی خرمی
دانشجوی دکتری، دانشکده کامپیوتر، دانشگاه زنجان، ایران
اصغر تاج الدین
استادیار، گروه برق و مهندسی، دانشکده کامپیوتر، دانشگاه زنجان، ایران

چکیده

امروزه کشورها با انبوهی از تهدیدات سایبری مواجه هستند که بخش‌های مختلف تجاری، شرکت‌های خصوصی و سازمان‌های دولتی را تحت تاثیر خود قرار داده است. تهدیدات پایدار پیشرفته (APTs)، یکی از انواع حملات چند مرحله‌ای است که با اقدامات پیچیده خود به شبکه‌ی مقصد نفوذ می‌کند، برای مدت طولانی در شبکه باقی می‌ماند و با دسترسی‌های غیرمجاز و استفاده از روش‌های ناشناخته، اطلاعات را به سرقت می‌برد. در این پژوهش روشی برای تشخیص بلادرنگ حملات APT با استفاده از تحلیل رویدادهای ثبت شده در سیستم‌های نقاط پایانی مبتنی بر ویندوز پیشنهاد می‌دهیم. این روش با استفاده از مکانیزم توجه در یادگیری عمیق، قادر است بدون نیاز به در نظر گرفتن ویژگی‌های خاص، رویدادهای ثبت شده در نقاط پایانی را تحلیل و درک نماید و در نهایت وقوع یک حمله را تشخیص دهد. در روش پیشنهادی با استفاده از مدل‌سازی اطلاعات معنایی رویدادها توسط مدل ترنسفورمر، الگوی حملات را به مدل آموزش داده و سپس حملات جدید را شناسایی می‌کنیم. برای این منظور، ابتدا در مرحله آموزش، اطلاعات معنایی رویدادها و ارتباطات آن‌ها به مدل آموزش داده شود و سپس ترنسفورمر قادر خواهد بود دنباله‌های جدید حملات APT را تشخیص دهد. نتایج تجربی بر روی داده‌های منبع باز نشان می‌دهد روش پیشنهادی با دقت بیش از ۹۹ درصد، قادر به تشخیص حملات است و نسبت به بسیاری از رویکردهای دیگر، عملکرد بهتری دارد.

واژگان کلیدی: تهدیدات پایدار پیشرفته، تحلیل همبستگی رویداد، تشخیص حمله، مدل ترنسفورمر.

مقدمه

۱-۱ بیان مساله

تهدید پایدار پیشرفته[1]، یک توصیف کلی درباره حمله‌ای گسترده و چندمرحله‌ای است که در آن تیمی از مهاجمان، با روش‌های نفوذ مداوم، مخفیانه و پیچیده به یک شبکه دسترسی پیدا کرده و با هدف استخراج داده‌ها و اطلاعات حساس یا واردآوردن صدمات اساسی، به‌طور غیرقانونی و طولانی‌مدت در آن شبکه باقی می‌مانند. این نوع حملات که معمولا توسط دولت‌ها و سازمان‌های بزرگ حمایت می‌شوند از فناوری‌های پیشرفته‌ای بهره می‌برند و مراکز هدف آن‌ها، اغلب موسسات و شرکت‌های مهم صنعتی، مالی، دفاعی و یا سازمان‌های دولتی هستند. (Adel Alshamrani et al, 2019)

با توجه به اهمیت و حساسیت حملات APT، تاکنون اقدامات دفاعی گوناگونی جهت کشف و جلوگیری از صدمات این گونه حملات پیشنهاد شده‌است. ابزارهای مبتنی بر امضا که غالبا بر اساس شاخص‌های نفوذ[2] (IOC) اعلام شده در گزارش‌های اطلاعات تهدید سایبری[3] (CTI) مانند مقادیر هش بدافزار، آدرس‌هایIP سرورهای کنترل و فرمان، نام دامنه‌های تحت کنترل مهاجمان، نام فایل‌های مورد استفاده توسط APTها و غیره، امضاهای خود را به‌روز می‌کنند ممکن است بتوانند بر اساس گزارشی که برای حمله اخیر یک APT خاص منتشر شده، رویدادهای مشکوک مرتبط را در میزبان‌های یک سازمان شناسایی کنند، اما قادر به تشخیص نسخه‌های جهش‌یافته حملات نیستند(Wang et al, 2018). این ابزارها، دو ویژگی محوری را ندارند و درنتیجه بینش لازم را درباره حمله به تحلیل‌گر امنیتی نمی‌دهند. ۱- درک رابطه بین رویدادهای مختلف سیستم (Shiqing et al., 2016) و ۲- کنار هم قرار دادن اجزای حمله در یک دوره زمانی (Akirolabu et al.,2018). روش‌های تشخیص ناهنجاری نیز که با تمرکز بر رفتار عادی برنامه‌ها سعی در تشخیص حملات دارند، در کشف فعالیت APTها چندان موفق نیستند زیرا اولاً مهاجمان APT معمولاً از ابزارهای بومی که توسط مدیران سیستم به‌کار گرفته می‌شود، استفاده می‌کنند و ابزارهای امنیتی را فریب می‌دهند تا فعالیت‌های خود را به‌عنوان رفتار عادی جلوه دهند. ثانیاً، در صورتی‌که این رویکردها قادر به شناسایی بدافزارهای جهش یافته باشند، اغلب نرخ هشدار نادرست[4] بالایی ایجاد می کنند(Anwar and Zain, 2017). با این اوصاف و با توجه به پیچیدگی و پویایی رفتار APTها یافتن مدلی که بتواند به بهترین وجه ممکن و با بالاترین کارایی، حملات را شناسایی کند همچنان یک چالش مهم برای محققین است.

۱-۲ کارهای مرتبط

به دلیل افزایش سریع تعداد حملات، شناسایی و پیش‌بینی اقدامات مهاجمان به عنوان یک مشکل کلان داده محسوب می‌شود (Mahmoud et al., 2023). به همین دلیل الگوریتم‌های یادگیری ماشین و یادگیری عمیق به طور گسترده در تشخیص حملات مورد استفاده قرار گرفته‌اند. برای مثال پژوهشگران (Bodström et al., 2019) در روشی با استفاده از ترکیب مدل‌های مختلف یادگیری عمیق شامل رمزگذارهای خودکار[5]، حافظه بلند کوتاه مدت[6] (LSTM) و شبکه‌های عصبی گراف[7](GNN) به تشخیص حملات پرداخته‌اند. رویکرد دیگری به نام هولمز (Milajerdi et al., 2019) بر اساس سیاست‌ها[8]، روش‌ها[9] و رویه‌های[10] حملات در چهارچوب MITRE

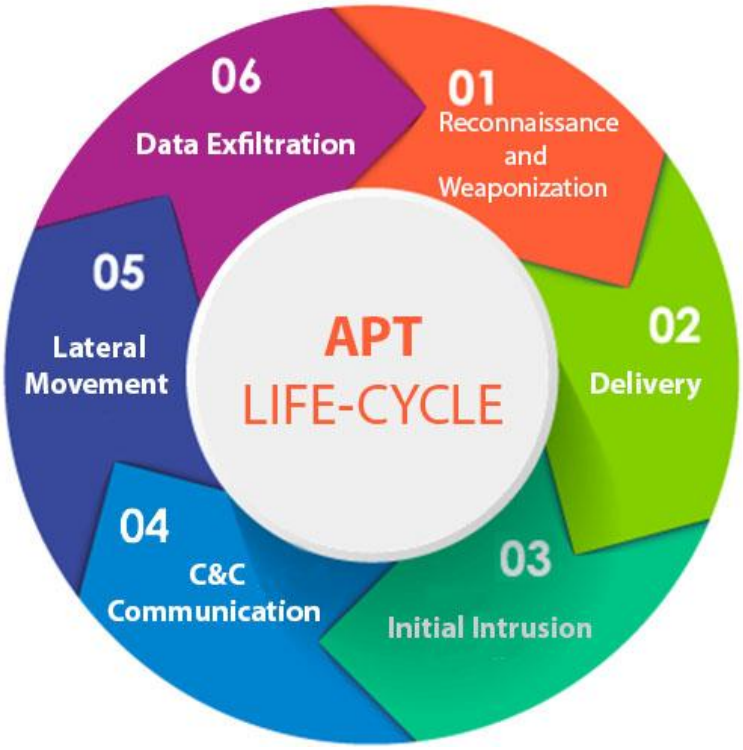
[1] Advanced Persistent Threat
[2] Indicators of Compromise
[3] Intrusion Detection Systems
[4] False-Positive Rate
[5] Auto-Encoder
[6] Long-Short Term Memory
[7] Graph Neural Network
[8] Tactics
[9] Tecnique
[10] Procedure

ATT&CK قوانینی را برای هر مرحله‌ی APT ارائه می‌کند. هشدار تشخیص APT بر اساس تجمیع نمرات تهدید حاصل از قوانین همه مراحل تولید می‌شود. برخی محققان (Sun et al., 2018) (Ma et al., 2017) نیز از شبکه بیزی برای یافتن حملات روز صفر استفاده

کرده‌اند. با این حال، هیچ یک از این راه‌کارها نمی‌تواند رفتار بلادرنگ در نفوذ به شبکه را تشخیص دهند. در نتیجه مدافعان تنها می‌توانند حملات را پس از وقوع تشخیص دهند و پیشگیری موثر عملا غیرممکن است. یونیکورن (Han et al., 2020) یک سیستم مبتنی بر ناهنجاری[11] است و وقتی نمودار مبدا[12] رویداد[13]های کل سیستم از نمودارهای نشان دهنده فعالیت‌های عادی منحرف می‌شود، هشدار صادر می‌کند. یونیکورن نیز همانند دیگر سیستم‌های مبتنی بر ناهنجاری، هنگامی که رفتار عادی تغییر می‌کند، ممکن است هشدارهای نادرست ایجاد کند. برخی محققان دیگر (Holgado et al. 2020) (Kholidy et al. 2014) نیز روش‌هایی را بر اساس مدل‌های پنهان مارکوف[14](HMM) برای پیش‌بینی حملات چندمرحله‌ای پیشنهاد داده‌اند. اگرچه HMM به طور گسترده برای تشخیص حملات چند مرحله‌ای استفاده می‌شود و قادر به تشخیص بلادرنگ حمله است ولی رویکردهای موجود تنها قادر به تشخیص یک حمله چند مرحله‌ای هستند و در حالت‌هایی که شبکه درگیر چندین حمله چند مرحله‌ای است، کارایی مناسبی نخواهند داشت.

۱-۳ مراحل حمله

APTها در طول فعالیت خود شش مرحله کلی را طی می‌کنند که به چرخه حیات[15] شهرت دارد. نکته حائز اهمیت آن است که این مراحل لزوماً پشت سر هم رخ نخواهند داد و ترتیب آن با توجه به سیاست های APT می‌تواند متغیر باشد (Chen et al, 2016). شکل۱، چرخه حیات APT را که بر اساس چارچوب MITRE ATT&CK طراحی شده توصیف می‌کند.



[11] Anomaly
[12] Provenance Graph
[13] Event
[14] Markov Hidden Model
[15] Life Cycle

شکل ۱. چرخه حیات APT

مراحل این چرخه عبارتند از:

- شناسایی و تسلیح سازی[۱۶] : اولین مرحله، شناسایی است که طی آن اطلاعات مورد نیاز درباره‌ی سازمان هدف جمع‌آوری می‌شود. این اطلاعات توسط عامل انسانی و یا تکنیک‌های فنی جمع‌آوری می‌شود. این مرحله بر شناسایی نقاط ضعف جهت نفوذ به شبکه سازمان تمرکز دارد. پس از جمع‌آوری اطلاعات، مهاجمان امکانات مورد نیاز حمله را بر اساس دانش به‌دست آمده گردآوری می‌کنند.
- تحویل[۱۷] : مهاجمان با استفاده از روش‌های تحویل مستقیم یا غیرمستقیم، کدهای مخربی را به اهداف ارسال می‌کنند. در تحویل مستقیم، مهاجمان برای تحویل بدافزار، از روش‌های مهندسی اجتماعی مانند فیشینگ هدفمند و تزریق کد مخرب به اهداف خود استفاده می‌کنند. در تحویل غیرمستقیم، از یک شخص ثالث قابل اعتماد برای سوءاستفاده غیرمستقیم از قربانی استفاده می‌شود.
- نفوذ اولیه[۱۸] : مهاجمان از یک نقطه ورود سوءاستفاده می‌کنند، جای‌پایی به دست می‌آورند و یک ارتباط خروجی برقرار می‌کنند. بازیگران APT با استفاده از ابزارها و استراتژی‌های متنوع، از آسیب‌پذیری‌های کشف‌شده در برنامه‌های تحت وب سازمان هدف و نیز نقاط ضعف موجود در رایانه‌های کاربر نهایی سوءاستفاده می‌کنند و به این ترتیب دسترسی به شبکه هدف محقق می‌شود.
- ارتباط فرمان و کنترل[۱۹] : پس از ورود به شبکه سازمان، مهاجمان حضور خود را در نقاط ورودی تثبیت می‌کنند و کنترل رایانه‌های آلوده را به دست می‌گیرند. دشمنان با استقرار یک ابزار مدیریت از راه دور (RAT) با سرور فرمان و کنترل خارج از سازمان ارتباط برقرار می‌کنند.
- حرکت جانبی[۲۰] : در این مرحله، بدافزار APT، خود را از طریق شبکه به میزبان‌های غیرآلوده دیگر منتقل می‌کند. این میزبان‌ها معمولاً دارای دسترسی ممتاز بالاتری هستند و با سوءاستفاده از آن‌ها احتمال دسترسی به اطلاعات طبقه‌بندی شده و استخراج داده‌های حساس‌تر بیشتر می‌شود.
- استخراج داده‌ها[۲۱] : در این مرحله مهم، مهاجمان داده‌هایی را که از شبکه داخلی جمع‌آوری کرده‌اند به سرور فرمان و کنترل خود صادر می‌کنند. میزبان‌های آلوده داده‌های جمع‌آوری شده خود را در هر منبع خارجی یا ابری بارگذاری می‌کنند. یکی از ضروری‌ترین کارهایی که لازم است مهاجم پس از استخراج داده‌ها انجام دهد، پاک‌کردن ردپا و مخفی ماندن است. ماندگاری APT در سازمان هدف، به توانایی مهاجم برای مخفی ماندن در شبکه بستگی دارد. بنابراین مهاجم باید هر گونه اثری از حضور خود را پاک کند و به طور مداوم بررسی‌های مخفیانه را انجام دهد.

در هر یک از این مراحل می‌توان حمله را به روش‌های مختلف و با احتمالات گوناگون تشخیص داد. بنابراین تشخیص حمله APT یک کار بسیار چالش برانگیز است.

۱-۴ اهداف

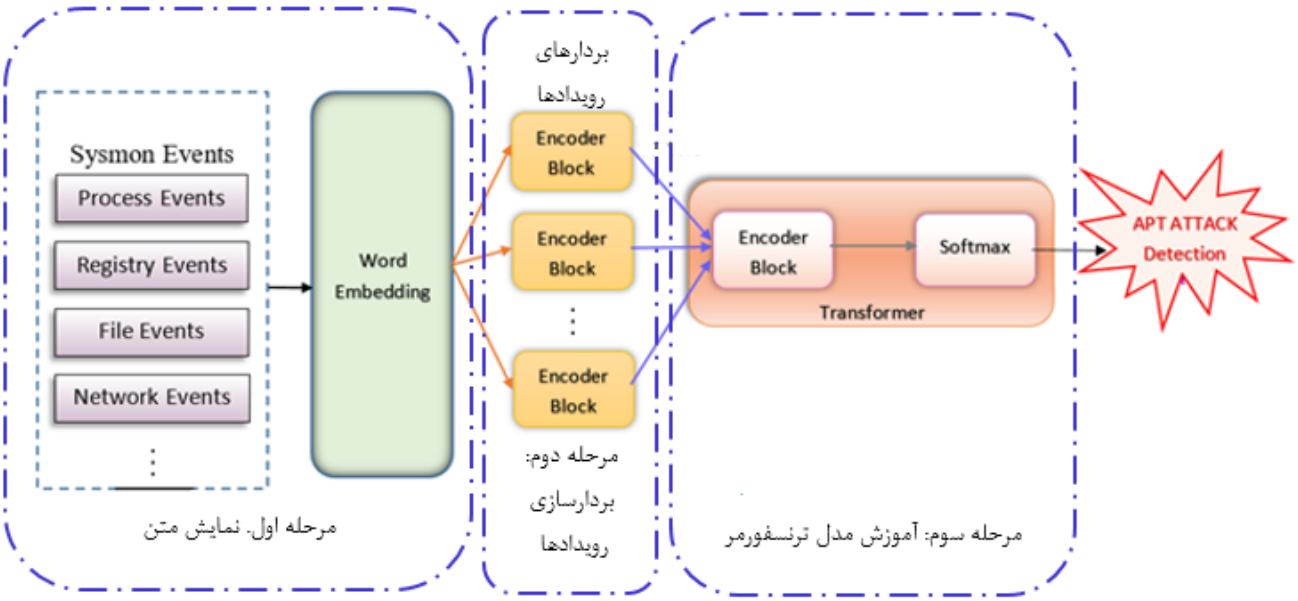
حملات APT اغلب تمایل دارند برای رسیدن به اهداف خود از آسیب‌پذیری‌های نقاط پایانی[۲۲] سوءاستفاده کنند. طبق آمار مورد استناد در برخی تحقیقات(Mahmoud et al., 2023)، اکثر حملات APT ثبت‌شده در جهان در سال ۲۰۲۳ از آسیب‌پذیری‌های موجود در دستگاه‌های نقاط پایانی بهره‌برداری کرده‌اند. بنابراین شناسایی و جلوگیری از توزیع بدافزار APT در نقاط پایانی بسیار ضروری است. برای این منظور، باید داده‌های منبع یا همان گزارشات و رویدادهای ثبت شده در دستگاه‌های نقاط پایانی، به شکلی هدفمند مورد

[16] Reconnaissance and Weaponization
[17] Delivery
[18] Initial Intrusion
[19] Command and Control Communication
[20] Lateral Movement
[21] Data Exfiltration
[22] Endpoints

بررسی و تحلیل قرار گیرد. در این حوزه، محققان در سال‌های اخیر چندین راهکار شناسایی حمله بر اساس مدل‌های یادگیری عمیق ارائه داده‌اند. با این حال، اغلب رویکردهای موجود در تجزیه و تحلیل رویدادها به جای شناسایی فعالیت‌های مخرب در هر مرحله حمله، بر روی شناسایی شاخص‌های یک حمله کلی APT تمرکز می‌کنند، یعنی زمانی که APT تمام مراحل خود را کامل می‌کند، به شناسایی مشخصات حمله می‌پردازند. از این رو، نمی‌توانند فعالیت‌های حمله APT را در هر مرحله شناسایی کنند و در نتیجه قادر به انجام تشخیص در مراحل میانی نیستند. برخی از این راهکارها به دقت سطح بالایی در شناسایی حملات دست یافتند اما از آنجایی که این رویکردها بلادرنگ نیستند و نمی‌توانند رفتار لحظه‌ای نفوذ به شبکه را درک کنند در نتیجه پیشگیری کافی و موثر اتفاق نخواهد افتاد. در این مقاله با بهره‌گیری از امکانات یادگیری عمیق در پردازش متن و درک مطلب، راهکار جدیدی جهت مدل‌سازی اطلاعات معنایی رویدادهای ثبت شده در دستگاه‌های نقاط پایانی ارائه می‌دهیم به صورتی که مدل، ضمن یادگیری خودکار ویژگی‌های حملات از درون رویدادهای ثبت شده، حمله APT را در زمان واقعی و با دقت و حساسیت بالا تشخیص دهد.

معماری روش تحقیق

چهارچوب کلی روش پیشنهادی از ۳ بخش تشکیل شده است: ۱) استخراج اطلاعات واژگانی و معنایی رویدادهای سیستم و جاسازی متون رویدادها. ۲) بردارسازی[۲۳] هر رویداد با استفاده از رمزگذارها[۲۴] ۳) آموزش مدل ترنسفورمر[۲۵] با استفاده از توالی بردارهای رویداد.
نمای کلی روش پیشنهادی در شکل ۲ نمایش داده شده است. در ادامه اجزای مختلف هر بخش به تفکیک شرح داده خواهد شد. روش پیشنهادی برای کشف حملات APT در نقاط پایانی، به تحلیل معنایی رویدادهای Sysmon[۲۶] در ویندوز می‌پردازد. در مرحله‌ی نمایش متن[۲۷]، رویدادهای Sysmon میزبان‌ها را به نمایش قابل فهم توسط مدل تبدیل می‌کنیم. در مرحله‌ی دوم، قالب‌های رویداد را استخراج کرده و سپس آن‌ها را به بردار تبدیل می‌کنیم. در مرحله‌ی آموزش، دنباله‌ی بردارهای رویداد را به مدل ترنسفورمر وارد می‌کنیم تا مدل، مفاهیم رویدادها و ارتباطات بین آن‌ها را درک کرده و مشخصات حملات و جریان کاری عادی سیستم را آموزش ببینند. در فاز تشخیص، مدل ترنسفورمر قادر خواهد بود وجود حمله را تشخیص دهد.



[23] Vectorization
[24] Encoders
[25] Transformer
[26] System Monitor
[27] Text Representation

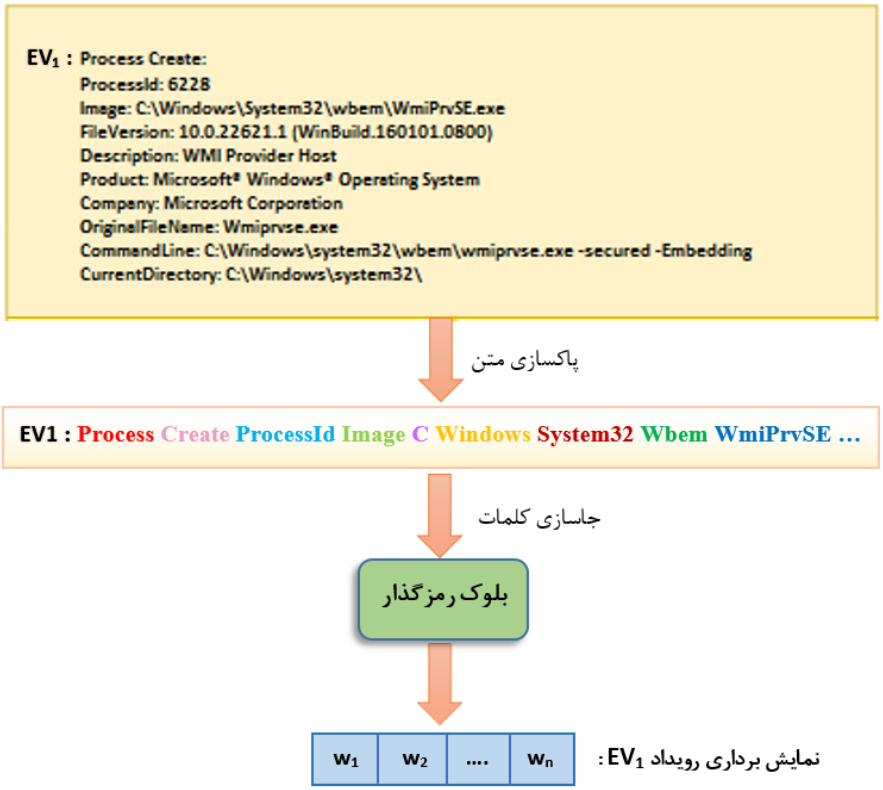
شکل ۲. معماری مدل پیشنهادی

۲-۱ مرحله نمایش متن

در این مرحله، با استفاده از مدل‌های مخصوصی که کار تعبیه‌سازی کلمه[۲۸] را برای رویدادها انجام می‌دهند، کلمات استخراج شده از رویدادها را به صورت جاسازی‌شده نمایش می‌دهیم. برای این منظور، پس از حذف علائم اضافی موجود در رویداد، هر کلمه‌ی آن، مانند Process، Create و غیره را بر اساس تعبیه‌سازی کلمات، جاسازی می‌کنیم.

۲-۲ مرحله بردارسازی رویدادها

پس از تعبیه کلمات رویدادها، این جاسازی‌ها را به بلوک‌های رمزگذار ترنسفورمر وارد می‌کنیم و نمایش بردار n بعدی رویداد را به‌دست می‌آوریم. در این مرحله مدل، روابط معنایی بین کلمات رویداد را کدگذاری کرده و مفهوم یک رویداد را درک می‌کند. فرآیند طی شده در مراحل ۱ و ۲ در شکل ۳ نمایش داده شده است.



شکل ۳. جاسازی و بردارسازی رویدادها

۲-۳ مرحله آموزش مدل ترنسفورمر

برای آموزش مدل ترنسفورمر جهت تشخیص دنباله‌های APT، کلیه دنباله‌های برداری جملات رویدادها را (به عنوان مثال [EV1, EV2,…, EVh]) با بلوک رمزگذار دیگری پردازش می‌کنیم و نمایش نهایی دنباله‌ی رویدادها را که برای طبقه‌بندی باینری (وجود حمله

[28] Word Embedding

یا عدم وجود آن) استفاده می‌شود، تولید می‌کنیم. در نهایت بر اساس خروجی واقعی، مدل یاد می‌گیرد که آیا چنین نمایشی از دنباله رویداد ورودی شامل حمله هست یا خیر. در این مرحله مدل، روابط معنایی بین رویدادهای رخ داده را کدگذاری کرده و در واقع به

ارتباط رویدادهای رخداده و اهمیت ترتیب و تقدم و تاخر آن‌ها توجه می‌کند. در این روش، به این دلیل که تنها به نمایش برداری نهایی یک دنباله نیاز داریم، فقط از ساختار رمزگذار ترنسفورمر برای رمزگذاری دنباله‌های رویدادها استفاده می‌کنیم.

همان‌طور که در شکل ۴ نشان داده شده است، ابتدا، ماتریس برداری X از یک دنباله را در بلوک وارد می‌کنیم، که در آن $X \in R^{n \times d}$، که n طول دنباله و d بعد هر بردار رویداد است. ترنسفورمر از دوازده سر توجه استفاده می‌کند که تعداد سرهای توجه را با h نشان می‌دهیم. برای محاسبه توجه هر بردار $X_i (i \in \{1, 2, 3, \dots, n\})$ در دنباله، ابتدا سه بردار پرس‌وجو Q، کلید K و مقدار V را برای هر بردار Xi تولید می‌کنیم و به ترتیب برای h بار از آن‌ها استفاده خواهیم کرد. $W_Q \in R^{d \times d_q}$ و $W_K \in R^{d \times d_k}$ و $W_V \in R^{d \times d_v}$ همان ماتریس‌های وزن هستند و d_q ، d_k و d_v ابعاد بردارهای پرس‌وجو، کلید و مقدار را نشان می‌دهند. این بردارها با ضرب X در سه ماتریس وزن به صورت زیر محاسبه می‌شوند.

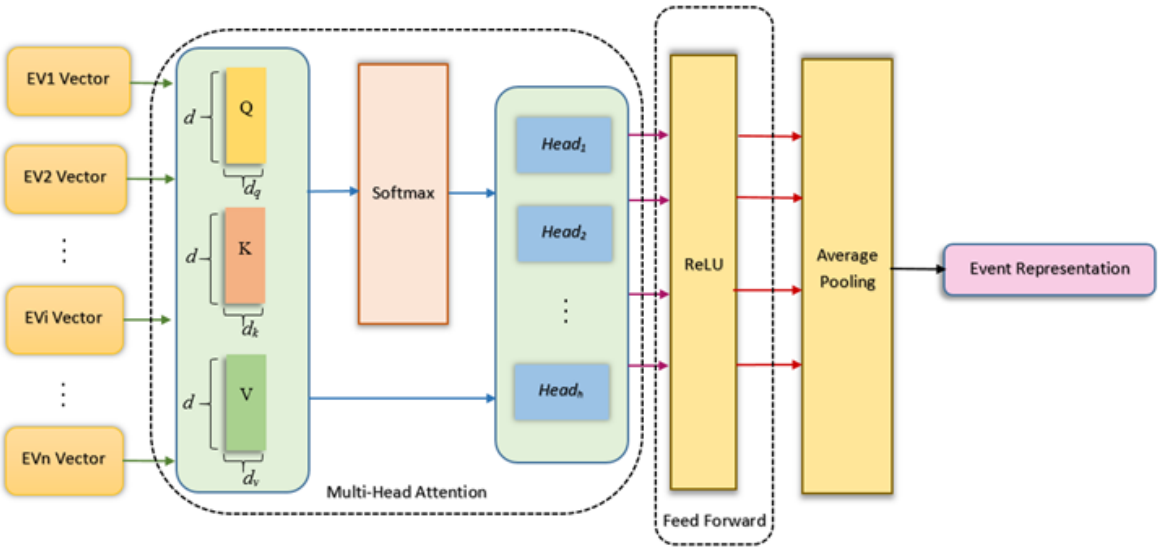
1) $Q = X * W_Q$

2) $K = X * W_K$

3) $V = X * W_V$

برای بردار $X_i (i \in \{1, 2, 3, \dots, n\})$، درجه اهمیت X_i را از طریق حاصل ضرب نقطه ای K و Q تعیین می‌کنیم و امتیاز را با Z_i به شرح زیر نشان می‌دهیم.

4) $Z_i = Q_i * K^T$



شکل ۴. طراحی بلوک رمزگذار ترنسفورمر

سپس Zi را بر $\sqrt{d_K}$ تقسیم می‌کنیم و همه امتیازهای به‌دست آمده را با تابع softmax نرمال می‌کنیم. به دلیل وجود مکانیسم توجه چند سر، مجموعه‌های متعددی از ماتریس‌های وزن پرس‌وجو، کلید و ارزش داریم. به عنوان مثال، برای سر توجه h ام، Z_i^h را در V_h ضرب می‌کنیم. در نهایت این سرها را با هم الحاق می‌کنیم و آنها را با ماتریس W_O به عنوان ماتریس وزن خروجی لایه توجه چند سر ضرب می‌کنیم، که W_O یک پارامتر قابل آموزش برای عملیات الحاق است. $head_h$ را می‌توان به صورت زیر به‌دست آورد، که Q_{ih} نمایش i امین برای h امین Q و K_h^T، h امین K را نشان می‌دهد.

5) $$head_h = softmax\left(\frac{Q_{ih} * k_h^T}{\sqrt{d_k}}\right) * V_h$$

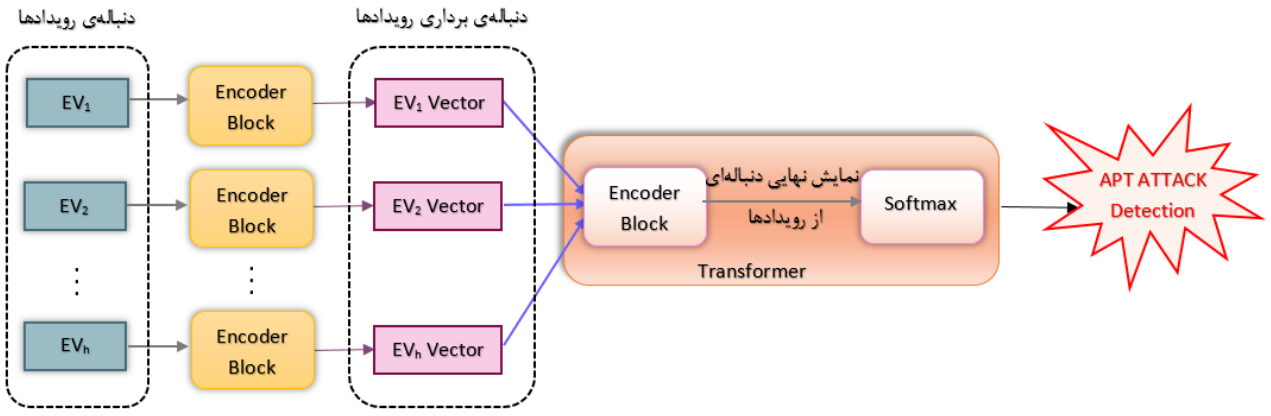
پس از لایه توجه چند سر، خروجی را از طریق شبکه عصبی پیش‌خور عبور می‌دهیم که ترکیبی از دو لایه خطی با استفاده از تابع فعال‌سازی ReLU است. تابع این لایه به صورت زیر تعریف می‌شود و W_1, W_2, b_1, b_2 پارامترهای قابل آموزش در لایه شبکه پیش خور هستند.

6) $$F_i = max(0.\ z_i W_1 + b_1) W_2 + b_2$$

در نهایت، خروجی بلوک‌های رمزگذار ترنسفورمر، نمایش برداری دنباله‌های رویدادها خواهند بود. به عنوان مثال، برای دنباله‌ای متشکل از رویدادهای $\{E_1, E_2, \dots, E_m\}$، دنباله‌ای از نمایش جمله‌های رویداد را به شکل $R_E = \{R_{E1}, R_{E2}, \dots, R_{Em}\}$ به دست می‌آوریم. در ادامه از یک لایه ادغام متوسط (Average Pooling Layer) برای محاسبه میانگین دنباله رویداد ورودی (RE) به عنوان نمایش نهایی دنباله رویداد استفاده می‌کنیم که بر اساس این نمایش، طبقه‌بندی باینری انجام خواهد گرفت.

۲-۴ فرآیند تشخیص حمله

فرآیند تشخیص در شکل ۵ نشان داده شده است. در این مرحله و پس از اتمام مرحله‌ی آموزش، دنباله رویداد جدیدی داریم که می‌خواهیم درباره اینکه آیا دنباله یک حمله است یا خیر اظهارنظر کنیم. پس از جاسازی کلمات هر رویداد از دنباله با مجموعه جاسازی کلمات از پیش آموزش دیده، هر رویداد را از طریق بلوک‌های رمزگذار، به شکل برداری نمایش می‌دهیم تا دنباله بردارهای رویدادها به‌دست آید. در مرحله بعد، دنباله بردارهای رویدادها را به مدل ترنسفورمر می‌دهیم و نمایش واحدی از مجموعه رویدادهای موجود در یک دنباله به‌دست می‌آوریم تا قضاوت کنیم که با توجه به وقوع این دنباله از رویدادها، آیا حمله‌ای در حال انجام شدن است یا خیر.



شکل ۵. روش تشخیص

ارزیابی

برای آموزش مدل ترنسفورمر، از Pytorch 1.4.0 و Python3.6 استفاده می کنیم. روش ارائه شده را با نتایج ارائه شده در پژوهش هولمز که یک مدل مبتنی بر اطلاعات معنایی است مقایسه می کنیم.

۳-۱ مجموعه داده تحقیق

مدل ما بر اساس دو مجموعه داده، ارزیابی شده است: ۱- مجموعه داده‌ای از حملات APT شبیه‌سازی شده که توسط تیم قرمز در برنامه محاسبات شفاف دارپا (TC)، تولید شده است. ۲- مجموعه داده نرمال دو هفته‌ای DARPA که برای راستی آزمایی مدل در برابر موارد مثبت کاذب در TC تولید شده است.

۳-۲ کارایی تشخیص

به منظور تعیین کارایی روش پیشنهادی، با استفاده از سه معیار صحت (Precision)، به عنوان شاخص مثبت کاذب، پوشش (Recall)، به عنوان شاخص منفی کاذب، و امتیاز F (F-score)، به عنوان میانگین هارمونیک صحت و پوشش، مدل پیشنهادی و روش هولمز را که بر روی یک مجموعه داده آموزش و آزمایش شده‌اند، مقایسه می‌کنیم. روش پیشنهادی با امتیاز F ٪۹۹.۶۶۱ عملکرد بالاتری نسبت به هولمز با امتیاز F ٪۹۸.۹۰۴ به دست می آورد. همچنین روش پیشنهادی دارای صحتی برابر با ٪۹۹.۴۰۴ است و صحت روش هولمز ٪۹۸.۲۵۴ می‌باشد. شاخص پوشش برای روش پیشنهادی، ٪ ۹۹.۸۶ است و در روش هولمز این شاخص ٪۹۹.۵۶۲ برآورد شده است. با وجود اینکه معیارها بسیار نزدیک است اما همانگونه که گفته شد، روش هولمز قادر است تنها پس از اتمام چرخه حیات حمله، آن را شناسایی کند که این مساله، شناسایی حمله را در مراحل میانی غیرممکن می سازد. به علاوه سرعت تشخیص در روش پیشنهادی به دلیل پردازش موازی رویدادها در ترنسفورمر بالاتر می باشد. زمان صرف شده توسط روش پیشنهادی در مقایسه با روش هولمز، بسیار کاهش یافته و تنها ۱.۰۲ ثانیه است در حالیکه زمان تشخیص در هولمز ۲.۸ ثانیه است.

جدول ۱ دو روش را به وسیله معیارهای ارزیابی مقایسه کرده است.

جدول ۱. ارزیابی روش پیشنهادی

مدل	صحت	پوشش	امتیاز F
هولمز	٪۹۸.۲۵۴	٪۹۹.۵۶۲	٪۹۸.۹۰۴
روش پیشنهادی	٪۹۹.۴۰۴	٪۹۹.۸۶	٪۹۹.۶۶۱

جدول ۲ نیز زمان فرآیند تشخیص (زمان آزمون) دو روش را مقایسه می کند. همانگونه که مشاهده می شود، به دلیل وجود پردازش موازی در ترنسفورمر، زمان صرف شده توسط روش پیشنهادی در مقایسه با روش هولمز، بسیار کاهش یافته و تنها ۱.۰۲ است در حالیکه زمان تشخیص در هولمز ۲.۸ ثانیه است.

جدول ۲. مقایسه زمان تشخیص حمله

مدل	زمان آزمون
هولمز	۲.۸ ثانیه
روش پیشنهادی	۱.۰۲ ثانیه

بحث و نتیجه‌گیری

تشخیص مبتنی بر رویداد همواره مورد تاکید محققان امنیت سایبری است، زیرا رفتار هر گونه ناهنجاری و حملات را کدگذاری می‌کند، که به کشف استراتژی‌های حمله کمک می‌کند. در این پژوهش، با بهره‌گیری از خصوصیات روش ترنسفورمر به مدل‌سازی اطلاعات معنایی رویدادها پرداخته، الگوی حملات را به مدل آموزش دادیم و تلاش کردیم در مرحله آزمون، حملات جدید را شناسایی کنیم. نتایج تجربی نشان می‌دهد روش پیشنهادی با دقت بیش از ۹۹ درصد، قادر به تشخیص حملات است و نسبت به بسیاری از رویکردهای دیگر، عملکرد بهتری دارد.

اگرچه مدل پیشنهادی در مقایسه با مدل‌های دیگر بهتر عمل می‌کند، اما تمرکز این روش بر روی رویدادهای Sysmon است. در کارهای آینده، سعی خواهیم کرد انواع دیگری از رویدادهای مهم مانند Security که در تشخیص دقیق‌تر حملات موثر هستند را به مجموعه‌ی داده‌های خود اضافه کنیم تا ویژگی‌های بیشتری از حملات را به مدل آموزش دهیم و عملکرد روش خود را بهبود ببخشیم.

- منابع:

Adel, Alshamrani. Ankur, Chowdhary. Sowmya, Myneni. Dijiang, Huang. (2019). A Survey on Advanced Persistent Threats:Techniques, Solutions, Challenges, and Research Opportunities. IEEE Comm Surveys & Tutorials 21(2). 1851–1877.

Yu, Wang . Weizhi, Meng. Wenjuan, Li. Jin, Li. Wai-Xi, Liu. Yang, Xiang. (2018). A fog-based privacy-preserving approach for distributed signature-based intrusion detection. J. Parallel Distrib. Comput. Vol. 122. pp. 26–35.

Ma, Shiqing. Zhang, Xiangyu. and Xu, Dongyan. (2016). ProTracer: Towards Practical Provenance Tracing by Alternating Between Logging and Tainting. In Proceedings of the 23rd Annual Network and Distributed System Security Symposium.

Olusola, Akirolabu. Ioannis, Agrafiotis. Arnau, Erola. (2018). The challenge of detecting sophisticated attacks: Insights from SOC Analysts. In Proceedings of the 13th International Conference on Availability, Reliability and Security. Article No 55. Pages 1-9.

Moustafa, Mahmoud. Mohammad, Mannan. Amr, Youssef. (2023). APTHunter: Detecting Advanced Persistent Threats in Early Stages. Digital Threats: Research and Practice., Vol. 4. No. 1(11).

Shahid, Anwar. Jasni Mohamad, Zain. Mohamad Fadli, Zolkipli. Zakira, Inayat. Suleman, Khan. Bokolo, Anthony. Victor, Chang. (2017). From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. Algorithms. Vol. 10. No. 39.

Tero, Bodström. Timo, Hämäläinen. (2019). A Novel Deep Learning Stack for APT Detection. Applied Sciences. Vol. 9(6):1055.

Mohammd Sadegh, Milajerdi. Rigel, Gjomemo. Birhanu, Eshete. R, Sekar and V.N, Venkatakrishnan. (2019). HOLMES: Real-time APT detection through correlation of suspicious information flows. IEEE Symposium on Security and Privacy (SP). Pages 1137–1152.

Xiaoyan, Sun. Jun, Dai. Peng, Liu. Anoop, Singhal. John, Yen. (2018). Using bayesian networks for probabilistic identification of zero-day attack paths. IEEE Trans. Inf. Forensics Secur., Vol. 13, No. 10, pp. 2506–2521.

Shiqing, Ma. Juan, Zhai. Fei, Wang. Kyu, Hyung Lee. Xiangyu, Zhang. Dongyan, Xu. (2017). MPI: Multiple perspective attack investigation with semantic aware execution partitioning. in Proc. 26th USENIX Conf. Secur. Symp. pp. 1111–1128.

Xueyuan, Han. Thomas, Pasquier. Adam, Bates. James, .Mickens. Margo, Seltzer. (2020). UNICORN: Runtime Provenance-Based Detector for Advanced Persistent Threats. ArXiv.

Pilar, Holgado. Víctor A, Villagrá. Luis, Vázquez. (2020). Real-Time Multistep Attack Prediction Based on Hidden Markov Models. IEEE Transactions on Dependable and Secure Computing. Vol. 17. Pages 134-147.

Ping, Chen. Lieven, Desmet. Christophe, Huygens. (2014). A study on advanced persistent threats. Com-mun. Multimed. Secur. 63–72 .

Hisham A, Kholidy. Abdelkarim, Erradi. Sherif, Abdelwahed. Abdulrahman, Azab. (2014). A finite state hidden markov model for predicting multistage attacks in cloud systems. in Proc. IEEE 12th Int. Conf. Dependable, Autonomic Secure Comput. pp. 14–19.

A Semantic-based Approach for Real-time APT Attack Detection Using Event Correlation Analysis

Sareh Eslami Khorrami
PhD Student, Faculty of Computer, University of Zanjan, Iran
Asghar Tajoddin
Assistant professor, Department of Electrical and Engineering,
Faculty of Computer, University of Zanjan, Iran.

Abstract

Today, countries are confronted with a variety of cyber threats that impact different sectors such as businesses, private companies, and government organizations. Advanced persistent threat (APT) is a type of multi-stage attack that infiltrates the target network with complex actions, stays in the network for a long time, and steals information through unauthorized access and unknown methods. In this study, we propose a method for detecting APT attacks in real-time through analyzing events recorded in Windows-based endpoint systems. By utilizing the attention mechanism in deep learning, this approach can analyze and comprehend events captured at endpoints without considering specific features, and it ultimately detects the occurrenceof an attack. In this method, we use the transformer to model the semantic information of events in order to train the attack pattern to the model and then detect new attacks. For this purpose, initially, the model is trained with the semantic information of events and their relationships, and then the Transformer will be able to detect new sequences of APT attacks. The experiments on open-source data sets show that this method can detect attacks with more than 99% accuracy and performs better than many other methods.

Keywords: Advanced Persistent Threat, Event Correlation Analysis, Attack Detection, Transformer Model.