ITRC

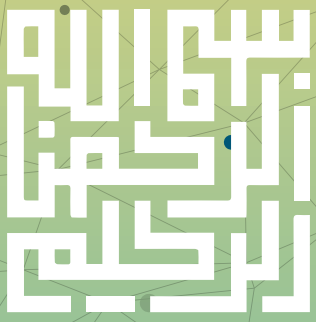
پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

مرجع دانش حملات سایبری مبتنی بر هوش مصنوعی به سامانه‌های امنیتی شبکه

تاریخ انتشار: شهریور ۱۴۰۴

ITRC

پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)



عنوان گزارش: مرجع دانش حملات سایبری مبتنی بر هوش مصنوعی به سامانه‌های امنیتی شبکه

کلمات کلیدی: سامانه‌های امنیتی شبکه، هوش مصنوعی، حملات تهاجمی و حملات تخاصمی

تهیه کنندگان: تینا تعویذی و علیرضا عنایتی

راهبر علمی (مجری پژوهش): امیرمنصور یادگاری

گروه پژوهشی: ارزیابی امنیت شبکه و سامانه‌ها (پروژه پژوهشی تدوین طرح مقابله با تهدیدات سایبری بهره‌مند از هوش مصنوعی)

تاریخ نشر: شهریور ۱۴۰۴

ITRC

پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

فهرست مطالب

ITRC

پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

چکیده

سند حاضر یک دانش‌نامه کاربردی، مشتمل بر معرفی انواع تهدیدات سایبری بهره‌مند از هوش مصنوعی برای سامانه‌های امنیتی (همراه با راهکارهای مقابله با تهدیدات) است. در این سند، ابتدا انواع سامانه‌های امنیتی رایج مورد استفاده در سازمان‌ها معرفی شده‌اند. سپس برخی از مهمترین کارکردهای ارائه‌شده توسط این سامانه‌ها معرفی شده است. در نهایت، برخی از انواع تهدیدات بهره‌مند از هوش مصنوعی که با هدف متخاصم‌سازی سامانه‌های امنیتی جهت ایجاد تغییر در ماهیت آنها یا دور زدن و ایجاد اختلال در کارکردهای سامانه‌ها وجود دارند، شناسایی شده‌اند. علاوه بر این، برخی از راهکارهای مقابله با هریک از تهدیدات نیز مشخص شده است.

۱ معرفی انواع سامانه‌های امنیتی

در این بخش، در جدول ۱، برخی از مهمترین سامانه‌ها و محصولات حوزه امنیت اطلاعات که به طور رایج در سازمان‌ها استفاده می‌شوند، آورده شده است[۱]. این سامانه‌ها را با توجه به عمده کارکردهای امنیتی ارائه شده توسط آنها، می‌توان در دسته‌هایی با عناوین شناسایی، محافظت، تشخیص، پاسخگویی، بازیابی و حاکمیت (برمبنای توابع امنیتی چارچوب NIST CSF v2.0) تقسیم‌بندی نمود.

جدول ۱: معرفی و دسته‌بندی انواع سامانه‌های امنیتی

دسته‌بندی سامانه‌های امنیتی از نظر کارکردی	معرفی سامانه‌های این دسته	مصداق‌ها
شناسایی (Identify)	هدف این سامانه‌ها و راهکارها، کمک به سازمان‌ها برای شناسایی و مدیریت دارایی‌های اطلاعاتی است که باید مورد حفاظت قرار گیرند. شناسایی وضعیت امنیتی این دارایی‌ها و ریسک‌های مربوطه نیز از جمله اهداف دیگری است که سامانه‌های این حوزه به دنبال آن هستند.	• سیستم‌های مدیریت دارایی‌ها (مثلا CMDB[۲]ها) • سیستم‌های مدیریت آسیب‌پذیری‌ها (ابزارهای اسکن و شناسایی آسیب‌پذیری در اپلیکیشن‌ها و سیستم‌ها همانند ابزارهای پویش آسیب‌پذیری، ابزارهای تست نفوذ و غیره)
محافظت (Protect)	تمرکز این سامانه‌ها بر پیاده‌سازی اقدامات حفاظتی برای پیشگیری از وقوع تهدیدات سایبری یا محدود کردن و مهار تأثیر این تهدیدات است.	• سیستم‌های مدیریت هویت و دسترسی (IAM[۳]) • سیستم‌های مدیریت دسترسی‌های راه دور (PAM[۴]) • فایروال‌های شبکه و وب‌اپلیکیشن (WAF[۵]) به خصوص فایروال نسل بعدی (NGFW[۶]) • سیستم‌های پیشگیری از نفوذ (IPS[۷]) • سیستم‌های پیشگیری از دست رفتن داده‌ها (DLP[۸])

۱. تیم تحقیق، فهرست کامل و دقیق‌تری از محصولات حوزه امنیت اطلاعات که بالغ بر ۴۰ نوع محصول مختلف می‌باشند، استخراج نموده و این لیست، در فهرست کالاها و خدمات دانش بنیان به عنوان محصولات حوزه فناوری اطلاعات و ارتباطات و نرم‌افزارهای رایانه‌ای و دسته امنیت فضای تبادل اطلاعات در این لینک قابل دسترس است.

2. Configuration Management Databases
3. Identity and Access Management
4. Privilege Access Management
5. Web Application Firewall
6. Next Generation Firewall
7. Intrusion Prevention System
8. Data Loss Prevention

دسته‌بندی سامانه‌های امنیتی از نظر کارکردی	معرفی سامانه‌های این دسته	مصداق‌ها
محافظت (Protect)	تمرکز این سامانه‌ها بر پیاده‌سازی اقدامات حفاظتی برای پیشگیری از وقوع تهدیدات سایبری یا محدود کردن و مهار تأثیر این تهدیدات است.	• راهکارهای رمزنگاری داده‌ها • سیستم‌های حفاظت در برابر DDoS (DDoS Protection) • سیستم‌های Email Security • سیستم‌های حفاظت از نقطه پایانی (EDR/XDR[1]) و ابزارهای آنتی‌ویروس/ضد بدافزار • راهکارهای پشتیبان‌گیری و بازیابی
تشخیص (Detect)	هدف این سامانه‌ها، کمک به سازمان‌ها جهت پیاده‌سازی اقدامات لازم برای شناسایی وقوع رویدادهای سایبری است. درواقع تمرکز این سامانه‌ها، بر روی پایش مداوم هدف موردنظر جهت تشخیص حوادث است.	• سیستم‌های مدیریت اطلاعات و رویدادهای امنیتی (SIEM[2]) • سیستم‌های تشخیص نفوذ مبتنی بر میزبان و شبکه (HIDS[3] و NIDS[4]) • سیستم‌های تحلیل رفتار کاربران و موجودیت‌ها (UEBA[5]) • تکنولوژی‌های فریب همانند ظرف و شبکه عسل (Honey Pot و Honey Net)
پاسخگویی (Response)	هدف این سامانه‌ها کمک به سازمان‌ها برای پیاده‌سازی اقدامات موردنیاز به هنگام شناسایی یک رویداد یا حادثه سایبری است. درواقع این سامانه‌ها و ابزارها به مهار حوادث یا کاهش اثرات آن کمک می‌کنند.	• سیستم‌های هماهنگ‌سازی، خودکارسازی و پاسخ‌دهی امنیتی (SOAR[6]) • پلتفرم‌های پاسخ‌دهی به حوادث (IRP[7]) • ابزارهای تحلیل فارنزیک • پلتفرم‌های هوش تهدید (TIP[8])
بازیابی (Recovery)	هدف این سامانه‌ها کمک به سازمان برای پشتیبانی از برنامه‌های تداوم کسب و کار و بازیابی سرویس‌ها و فعالیت‌های حیاتی مختل شده به دلیل حوادث سایبری است.	• راهکارهای پشتیبان‌گیری و بازیابی • سیستم‌های مدیریت BCP و DRP
حاکمیت (Governance)	هدف این سامانه‌ها کمک به سازمان‌ها جهت تدوین و نظارت بر استراتژی‌ها، اهداف و سیاست‌های امنیت سایبری سازمان است.	• سیستم‌های مدیریت حاکمیت، ریسک و انطباق (GRC[9]) • سیستم‌های مدیریت خط‌مشی‌های امنیت اطلاعات

1. Endpoint Detection and Response/ Extended Detection and Response
2. Security Information and Event Management
3. Host-based Intrusion Detection System
4. Network-based Intrusion Detection System
5. User and Entity Behaviour Analytics
6. Security Orchestration, Automation, and Response
7. Incident Response Platform
8. Threat Intelligence Platform
9. Governance Risk Compliance

۲ مروری بر مهم‌ترین کارکردهای ارائه‌شده توسط سامانه‌های امنیتی

در این بخش، در جدول ۲، برخی از مهمترین کارکردهای ارائه شده توسط سامانه‌های امنیتی رایج در سازمان‌ها و نمونه سامانه‌های دارای آن کارکرد آورده شده است. اختلال در این کارکردها یکی از اهداف اصلی مهاجمین در تهدیدات مبتنی بر هوش مصنوعی علیه این سامانه‌ها باشد که در بخش بعدی به آن پرداخته شده است.

جدول ۲: فهرست برخی از کارکردهای ارائه شده توسط سامانه‌های امینتی

ردیف	عنوان کارکرد	نمونه سامانه‌های مرتبط
۱	قابلیت تشخیص مبتنی بر امضا (Signature-based detection)	HIDS, NIDS, IPS
۲	قابلیت تشخیص مبتنی بر ناهنجاری (anomaly-based detection) و تحلیل رفتاری	HIDS, NIDS, IPS
۳	قابلیت DPI (Deep Packet Inspection)	NGFW
۴	قابلیت SPI (Statefull Packet Inspection)	NGFW
۵	قابلیت SSL inspection	NGFW
۶	قابلیت ادغام با هوش تهدید (TI)	SIEM, NGFW
۷	قابلیت هشداردهی برای حملات در حال وقوع و فعالیت‌های مخرب شناسایی شده	SIEM, HIDS
۸	تحلیل همبستگی بین رویدادها (correlation analysis)	SIEM
۹	قابلیت فیلترینگ ترافیک برمبنای قوانین و پالیسی‌های از پیش تنظیم شده و معیارهایی همچون آدرس IP، پورت، محتوا، URL و غیره	WAF, NGFW
۱۰	قابلیت محدودسازی تعداد درخواست‌های قابل ارسال کلاینت‌ها در یک بازه زمانی مشخص	WAF, DDoS Protection
۱۱	قابلیت تشخیص و مسدودسازی حملات وب‌اپلیکیشن‌ها همچون SQL Injection، CSRF و ...	WAF
۱۲	قابلیت تشخیص بات‌ها و ادغام با مکانیزم کپچا	WAF
۱۳	قابلیت پیشگیری از استخراج اطلاعات حساس	DLP
۱۴	قابلیت Sandboxing	NGFW
۱۵	قابلیت پیشگیری از نفوذ	IPS
۱۶	قابلیت ادغام با سیستم‌های پاسخگویی خودکار	SIEM
۱۷	قابلیت VPN	VPN

ردیف	عنوان کارکرد	نمونه سامانه‌های مرتبط
۱۸	قابلیت تشخیص ایمیل‌های فیشینگ	Email Security
۱۹	قابلیت فیلترینگ ایمیل‌های اسپم	Email Security
۲۰	قابلیت احرازهویت برای ایمیل‌ها همانند SPF، DKIM و DMARK	Email Security
۲۱	قابلیت پایش دسترسی‌های ممتاز	PAM
۲۳	قابلیت پایش تغییرات فایل‌ها	HIDS
۲۴	قابلیت تشخیص بدافزار	XDR/EDR, Antimal-ware/Antivirus
۲۵	قابلیت مانیتورینگ نشست‌های راه دور	PAM
۲۶	قابلیت تولید، جمع‌آوری، پردازش و تحلیل لاگ‌ها	SIEM
۲۷	قابلیت تحلیل ریشه‌ای حوادث (RCA)	EDR/XDR
۲۸	قابلیت ادغام با سرویس‌های identity	بیشتر سامانه‌ها

۳ مروری بر انواع تهدیدات سایبری بهره‌مند از هوش مصنوعی و راهکارهای مقابله

در این بخش، برخی از مهمترین تهدیدات سایبری بهره‌مند از هوش مصنوعی علیه سامانه‌های امنیتی آورده شده است. روش کار بدین صورت بوده است که به جای اینکه تهدیدات بر روی عنوان یک سامانه امنیتی تعریف شود، بر روی کارکردهای سامانه های امنیتی تبیین شده است، چرا که معمولاً یک کارکرد امنیتی می تواند در بیش از یک نوع سامانه امنیتی وجود داشته باشد و جلوی تکرار مفاهیم مشابه گرفته شده است. هدف این تهدیدات، متخاصم‌سازی سامانه‌های امنیتی جهت ایجاد تغییر در ماهیت آنها یا دور زدن و ایجاد اختلال در کارکردهای این سامانه‌ها است.

جدول ۳ دربرگیرنده این تهدیدات است که در آن، عناوین تهدیدها، دسته‌بندی تهدید برمبنای چارچوب STRIDE و در نهایت برخی راهکارهای مقابله با تهدید آورده شده است.

جدول ۳: فهرست برخی از تهدیدات سایبری بهره‌مند از هوش مصنوعی علیه سامانه‌های امنیتی

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۱	تهدید دور زدن سیستم احرازهویت سامانه با روش‌هایی همچون اجرای حملات فیشینگ مبتنی بر AI برای بدست آوردن اطلاعات credentialهای حساب‌های کاربری و غیره	Spoofing	استفاده از راهکارهای احرازهویت چند عاملی (MFA) مبتنی بر یک یا چند فاکتور بایومتریک برای احرازهویت کاربران نهایی به خصوص کاربران با سطح دسترسی بالا
۲	تهدید Brute-force از طریق تکنیک‌های مبنی بر AI برای کرک کردن پسوردها (همانند شبکه‌های GAN)	Information Disclosure	استفاده از مکانیزم کپچای قوی برای ایجاد چالش‌های تشخیص کاربران انسانی از ربات‌ها
۳	تهدید Credential Stuffing با استفاده از ایجنت‌های AI و همچنین دیتاست‌های بزرگ تولید شده به کمک AI	Spoofing	استفاده از راهکارهای احرازهویت چند عاملی (MFA) مبتنی بر یک یا چند فاکتور بایومتریک برای احرازهویت کاربران نهایی به خصوص کاربران با سطح دسترسی بالا
۴	استفاده از تکنولوژی deepfake برای دور زدن مکانیزم احرازهویت بایومتریک	Spoofing	استفاده از ابزارهای تشخیص deepfake

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۵	حملات مبتنی بر تکنیک‌های فازی علیه UI	Tampering	اعتبارسنجی داده‌های ورودی کاربر به سامانه دربرابر مقادیر مجاز
۶	تهدید دور زدن و حمله به سیستم کپچا با استفاده از الگوریتم‌های پردازش تصویر همانند شبکه‌های عصبی پیشرفته (CNN) برای حل کردن کپچاهای تصویری، استفاده از مکانیزم OCR (Optical Character Recognition) برای دور زدن کپچاهای متنی با تبدیل تصویر به متن و دیگر روش‌ها	Spoofing	استفاده از کپچاهای ترکیبی (تصویری،صوتی، متنی) استفاده از کپچاهای پویا تحلیل رفتار کاربران مثلا حرکت ماوس، تأخیر تایپ، الگوهای کلیک استفاده از کپچای تعاملی و چالش‌محور ترکیب کپچا با روش‌های احراز هویت دیگر همانند MFA اعمال محدودیت برای تلاش‌های زیاد از یک کاربر یا IP
۷	حملات prompt injection علیه سامانه‌های مبتنی بر NLP (مثلا سامانه‌هایی که از LLMها و chatbot ها برای عملیات مختلف خود استفاده می‌کنند)	Spoofing	اعتبارسنجی داده‌های ورودی کاربر به سامانه دربرابر مقادیر مجاز جدا کردن prompt کاربر از دستورالعمل‌های سیستم محدود کردن قابلیت‌های مدل آموزش مدل‌های هوش مصنوعی با مثال‌های خصمانه (Adversarial) برای افزایش مقاومت آنها
۸	دور زدن قابلیت فیلترینگ ترافیک برمبنای قوانین و پالیسی‌های از پیش تنظیم شده و معیارهایی همچون آدرس IP، پورت، محتوا، URL و غیره با استفاده از روش‌هایی همانند آموزش شبکه‌های عصبی برای تغییر پیلودها، آموزش شبکه‌های GAN برای ایجاد ورودی‌های مخرب، استفاده از مدل‌های مبتنی بر RL و غیره	Spoofing	تحلیل رفتار شبکه بررسی عمیق پکت‌ها (DPI) استفاده از هوش تهدید برای داشتن پایگاه داده تهدیدهای بروز برای بلاک IPها/دامنه‌های مشکوک استفاده از سامانه‌های WAF هوشمند
۹	مهندسی معکوس فرایندهای کنترلی سامانه با استفاده از تکنیک‌های فازی و مدل‌های مولد	Information Disclosure	مبهم‌سازی فرآیندهای کنترلی مثلا با پنهان‌کردن ساختار پیام‌ها و اعتبارسنجی‌ها محدودکردن نرخ درخواست‌ها استفاده از Honeypot در طراحی بخش‌های کنترل‌شده برای گمراه‌کردن مهاجم و تحلیل رفتار او
۱۰	بهره‌برداری از APIهای ناامن با استفاده از پیلودهای تولیدشده به کمک AI	همه دسته‌ها	اعتبارسنجی ورودی در سمت سرور ساختار درخواست‌های فراخوانی APIهای سامانه را در نقاط مختلف سامانه مثلا در API Gateway محدود کردن تعداد درخواست در بازه زمانی (rate limiting) ثبت دقیق لاگ همه درخواست‌ها و پارامترها

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۱۱	تهدید مسموم‌سازی مدل‌های ML	Tampering	اعمال کنترل‌های سختگیرانه‌ای را بر جمع‌آوری، برچسب‌گذاری و ذخیره‌سازی داده‌ها برای مجموعه‌های داده آموزشی هوش مصنوعی اعتبارسنجی منظم داده‌های آموزشی پاکسازی و نرمال‌سازی داده‌های ورودی قبل از تغذیه آنها به مدل‌های هوش مصنوعی استفاده از امضای دیجیتال یا هش برای تأیید صحت پایش مداوم داده‌های ورودی برای کشف داده‌های مشکوک ترکیب چند مدل با داده‌های متفاوت برای کاهش تأثیر حمله نظارت مداوم مدل‌های هوش مصنوعی برای شناسایی کاهش عملکرد، رفتار غیرمنتظره یا ناهنجاری‌هایی که می‌تواند نشان‌دهنده یک حمله خصمانه باشد.
۱۲	تهدید مهندسی اجتماعی از طریق اجرای تکنیک‌های فیشینگ مبتنی بر AI	Spoofing	آموزش کاربران استفاده از ابزارهای ضد فیشینگ با NLP مانیتورینگ رفتارهای کاربران مثلا با تحلیل کلیک‌ها و دانلودهای مشکوک استفاده از راهکارهای MFA برای حساب‌های کاربری حساس
۱۳	تهدید سرقت نشست‌ها از طریق تقلید یا شبیه‌سازی رفتارها به کمک AI	Information Disclosure Spoofing	احراز هویت دو مرحله‌ای برای عملیات حساس استفاده از نشست‌های کوتاه‌مدت تحلیل رفتار کاربران از طریق بررسی مواردی همچون سرعت تایپ، حرکت ماوس، محل جغرافیایی اتصال و غیره متصل کردن شناسه‌ها و توکن‌های نشست به مواردی همچون IP، Device Fingerprint یا کلید TLS
۱۴	دور زدن قابلیت تولید، جمع‌آوری، پردازش و تحلیل لاگ‌ها با استفاده از روش‌هایی همانند تولید ورودی‌های جعلی (به ظاهر معتبر) لاگ یا دستکاری لاگ‌ها با استفاده از تکنیک‌هایی همانند شبکه‌های GAN یا مدل‌های LLM برای پنهان کردن فعالیت‌های مخرب، تولید حجم بالایی از ترافیک به ظاهر معتبر و غیر مخرب برای از کار انداختن قابلیت رویدادنگاری سیستم، تزریق حجم زیادی از لاگ‌های نامربوط یا گمراه‌کننده یا به اصطلاح نویزی به کمک تکنیک‌های مبتنی بر AI برای دشوار کردن کار شناسایی تهدیدات واقعی توسط تحلیلگران و غیره	Spoofing Tampering Repudiation DoS	اطمینان از حفظ صحت لاگ با تولید و بررسی هش یا امضای دیجیتال برای لاگ ها رمزنگاری ترافیک لاگ تولید هشدار برای تغییر/حذف لاگ توسط کاربر غیرمجاز ذخیره‌سازی لاگ ها در چندین محل محدود کردن دسترسی به لاگ‌ها ذخیره سازی لاگ‌ها در رسانه‌های Write Once

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۱۵	تهدید دور زدن احرازهویت MFA مبتنی بر بایومتریک با استفاده از ویدئوها و صداهای تولیدشده با کمک deepfake	Spoofing	استفاده از راهکارهای Liveness Detection پیشرفته (مثلا چک می‌کند آیا تصویر زنده است یا بازپخش است) استفاده از چند فاکتور مکمل (مثلا بررسی رفتاری) استفاده از سنسورهای سخت‌افزاری
۱۶	دور زدن قابلیت تشخیص مبتنی بر امضا (Signature-based detection) با استفاده از تکنیک‌هایی همانند تولید ترافیک مشابه با ترافیک عادی با استفاده از شبکه‌های GAN، تقسیم پیلودهای مخرب به فرگمنت‌های غیرمخرب و دیگر تکنیک‌ها	Spoofing	ترکیب با راهکارهای behavior-based استفاده از Signatureهای پویا با یادگیری مداوم استفاده از DPI برای بازسازی فرگمنت ها
۱۷	دور زدن قابلیت تشخیص مبتنی بر ناهنجاری (anomaly-based detection) با استفاده از تکنیک‌هایی همانند یادگیری الگوهای ناهنجاری استفاده در سامانه به کمک الگوریتم‌های ML جهت تولید ترافیک مخرب مشابه با ترافیک عادی، تکنیک‌های AML جهت فریب مدل‌های یادگیری ماشین، تکنیک‌های یادگیری تقویتی (RL) و غیره	Spoofing	ترکیب مدل‌های ناهنجاری با Rule-based و هوش تهدید آموزش مدل‌ها با داده واقعی و همچنین نمونه‌های Adversarial استفاده از راهکار Adaptive Thresholding که مثلا در آن آستانه تشخیص، ثابت نبوده و با تغییر شرایط عملیاتی منطبق باشد. تست مقاومت مدل در مقابل حمله‌های فریب متنوع‌سازی Featureها در مدل تشخیص و تحلیل چندبعدی
۱۸	تهدید گریز از تشخیص از طریق ایجاد کانال‌های مخفی[1]	Spoofing	تحلیل ترافیک side-channel محدودسازی پروتکل‌های غیرضروری استفاده از پروکسی تحلیل عمیق پکت‌ها
۱۹	تهدید تزریق یا بازپخش[2] ارسال ترافیک مخربی که قبلا به عنوان ترافیک عادی شناسایی می‌شد.	Spoofing	استفاده از Time Stamp و Nonce در درخواست‌ها تشخیص Replay با سیستم‌های Anti-Replay منقضی کردن توکن‌ها
۲۰	مخفی کردن ترافیک مخرب با استفاده از شبکه‌های GAN	Spoofing	تحلیل عمیق پکت‌ها (DPI) Fingerprinting ترافیک مانیتور رفتار در سطح Session
۲۱	تقلید پروتکل‌ها با تولید ترافیک با استفاده از سیستم‌های LLM	Spoofing	تحلیل دقیق هدر و Sequence Number

1. Covert channels
2. replay

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۲۲	گریز از سیستم‌های بررسی اعتبار آدرس‌های IP از طریق تولید پروفایل‌های موردنیاز با AI	Spoofing	استفاده از احراز هویت MFA انجام Device Fingerprinting دقیق
۲۴	تهدید ارسال ترافیک موردنیاز جهت مختل کردن عملکرد موتورهای همبسته‌سازی[1] و گریز از تشخیص	Tampering	محدودسازی نرخ تزریق لاگ‌ها اعتبارسنجی منابع
۲۵	تهدید حرکت جانبی[2] به کمک راهکارهای مبتنی بر ML	Information Disclosure	استفاده از راهکارهای Micro-segmentation برای زون‌بندی ریزدانه شبکه
۲۶	مبهم‌سازی[3] الگوی ترافیک (مثلا از طریق ارسال ترافیک چندریختی[4])	Spoofing	تحلیل عمیق پکت‌ها
۲۷	تهدید تزریق پیکربندی‌های نادرست از طریق حمله به کانال‌های کنترل	Tampering	اعمال کنترل دسترسی سختگیرانه به کانال‌های کنترلی تولید و بررسی امضای دیجیتال برای پیکربندی‌ها ذخیره سازی تغییرات در پیکربندی ها در لاگ ها
۲۸	حملات استخراج مدل[5] از طریق بررسی و تحلیل مداوم ترافیک	Information Disclosure	افزودن Noise به خروجی مدل استفاده از Watermarking مدل‌ها محدودسازی نرخ کوئری‌ها پایش الگوی کوئری ها برای شناسایی رفتار مشکوک
۲۹	تهدید استخراج داده‌های telemetry	Information Disclosure	رمزنگاری ترافیک محدود کردن دسترسی ها
۳۰	تهدید حملات DoS با استفاده از تکنیک‌های مبتنی بر AI برای تحت تاثیر قرار دادن جریان‌های داده‌ای مرتبط با telemetry	DoS	استفاده از راهکارهای rate limiting
۳۱	حمله به agentها و تخریب عملکردهای management plane	Tampering	استفاده از رویکرد zero-trust برای Agentها در نظر گرفتن اصل حداقل دسترسی برای عملیاتی مدیریتی
۳۳	پنهان‌سازی مسیرهای استخراج ترافیک	Spoofing	استفاده از راهکارهای network segmentation
۳۴	دستکاری پردازشگرهای داده‌های telem-etry مبتنی بر ML	Tempering	اعتبارسنجی داده‌های ورودی کنترل صحت پایپلاین داده ها
۳۵	گریز از سیستم‌های پروفایلینگ ترافیک با استفاده از الگوهای رفتاری پویا	Spoofing	ترکیب مدل ایستا و پویا برای تحلیل رفتار بروزرسانی مستمر پروفایل ها

1. Correlation engine
2. Lateral movement
3. Obfuscation
4. polymorphic
5. Model extraction

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۳۶	ایجاد نمونه‌های مخرب برای حمله به طبقه‌بندهای جریان‌های ترافیکی	Spoofing Tampering	بررسی خروجی مدل‌ها با Post-Processing استفاده از طبقه‌بندی ترکیبی (Ensemble Classifierها)
۳۷	تهدید تولید سیل‌آسای هشدارها با استفاده از تکنیک‌های مبتنی بر AI برای خسته کردن تحلیلگران و ایجاد نویز جهت دور زدن قابلیت هشداردهی برای حملات در حال وقوع و فعالیت‌های مخرب شناسایی شده	DoS Spoofing	استفاده از راهکارهای اولویت‌بندی هشدارها استفاده از موتورهای همبسته‌سازی برای ادغام هشدارهای مشابه استفاده از سامانه‌های SOAR برای پاسخ گویی خودکار به هشدارهای کم‌ارزش
۳۸	تزریق سیگنال‌های false-positive و false negative در موتورهای تولید هشدار	Spoofing	اعتبارسنجی داده‌های ورودی همبسته‌سنجی هشدارها برای تشخیص هشدارهای واقعی استفاده از راهکارهای ترکیبی تشخیص (Ensemble) بکارگیری مکانیزم Adaptive Thresholding تولید و بررسی امضا برای داده‌ها
۳۹	جعل منبع هشدارها با استفاده از تکنیک‌هایی همچون LLM	Spoofing	ایجاد امضای دیجیتال روی هشدارها بررسی منبع تولید هشدار
۴۰	بازپخش یا بکارگیری هشدارهای قبلی برای گمراه‌سازی	Tampering	استفاده از Timestamp دقیق و Nonce برای هشدارها اعتبارسنجی منبع قبل از پردازش هشدار جلوگیری از ثبت تکراری هشدار مشابه
۴۱	تقلید رفتار واکنشی کاربران به هشدارها	Spoofing	استفاده از راهکارهای MFA برای دسترسی به کنسول ها ثبت لاگ دقیق اقدامات اپراتورها تحلیل رفتار کاربران
۴۲	تهدید دستکاری گزارشات حوادث برای فریب تحلیلگران	Tampering	کنترل دسترسی سخت‌گیرانه به داشبوردهای گزارش تایید چندمرحله‌ای برای تغییر گزارش
۴۳	تهدید وارونگی مدل‌های یادگیری[1] و استناج عضویت[2]	Information Disclosure	محدودسازی دسترسی به APIها اضافه کردن Noise به داده‌های آموزشی یا خروجی مدل لاگ گرفتن از کوئری‌های مشکوک و پایش آنها آموزش مدل با نمونه‌های حمله شبیه‌سازی‌شده کنترل خروجی‌ها رمزنگاری در سطح داده و مدل
۴۴	تهدید تخریب منطق موتورهای تشخیص مبتنی بر قانون با استفاده از داده‌های مبتنی بر AI	Tampering	اعتبارسنجی داده‌های قبل از به روزرسانی قوانین تست قوانین جدید در محیط ایزوله
۴۵	تهدید بدافزارهای چندریختی[3] خودکار که خود را برمبنای مکانیزم‌های تشخیص تطبیق می‌دهند.	Spoofing	استفاده از Sandboxing پیشرفته به‌روزرسانی مداوم و سریع مجموعه قوانین ترکیب راهکارهای تشخیص signature-based و behavior-based

1. Model Inversion
2. Membership Inference
3. polymorphism

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۴۶	دستکاری پالیسی‌ها با مسموم‌سازی مدل‌ها	Tampering	تولید و بررسی هش یا امضای برای پالیسی ها تایید چندلایه پالیسی‌ها قبل از استقرار
۴۷	تهدید دور زدن سیستم‌های طبقه‌بندی مبتنی بر رفتار	Spoofing	استفاده از مدل‌های Ensemble بروزرسانی مداوم base-lineهای رفتارها
۴۸	آلوده‌سازی پایپلاین‌ها با استفاده از دیتاست‌های مخرب برای آموزش مجدد مدل‌ها	Tampering	اعتبارسنجی دیتاست‌ها تولید امضای دیجیتال برای دیتاست‌ها تست Pre-training در محیط ایزوله
۴۹	استقرار backdoor در پایپلاین‌ها با دستکاری داده‌های آموزشی	Tampering	نظارت End-to-End بر مراحل Training استفاده از داده‌های مورد اعتماد تست Robustness مدل بعد آموزش
۵۰	آموزش غیرمجاز مدل‌ها از طریق پایپلاین‌های حاوی پیکربندی نادرست	Tampering	کنترل دسترسی سختگیرانه برای انجام پیکربندی پایپلاین‌ها استفاده از مکانیزم های MFA برای دسترسی به پایپلاین‌ها پایش تغییرات پیکربندی پایپلاین‌ها
۵۱	مسموم‌سازی به‌روزرسانی‌های قوانین	Tampering	تولید و بررسی امضای دیجیتال برای به‌روزرسانی‌ها جهت اطمینان از صحت و اصالت آنها استفاده از کانال‌های به روزرسانی امن
۵۲	تولید پالیسی‌های مخرب از طریق AI که به ظاهر قانونی و درست هستند.	Spoofing	پایش خروجی پالیسی‌ها در محیط عملیاتی تست پالیسی ها در محیط Sandbox بازبینی پالیسی‌ها قبل از بکارگیری و استقرار
۵۳	دور زدن قابلیت DPI (Deep Packet Inspection) با استفاده از روش‌هایی همچون استفاده از AML برای بهره‌برداری از نقاط ضعف موجود در پروتکل‌ها برای فریب دادن این سیستم‌ها، استفاده از شبکه‌های GAN برای ایجاد پیلودهای مخرب و غیره	Tampering Spoofing	ترکیب DPI با تحلیل رفتاری پایش خروجی‌ها برای تشخیص انحراف در جریان‌های ترافیکی استفاده از SSL Inspection برای رمزگشایی جریان‌های ترافیکی
۵۴	دور زدن قابلیت ادغام با هوش تهدید (TI) با مسموم کردن فیدهای TI برای دستکاری whitelistها یا دور زدن denylistها و تزریق indicatorهای نادرست به سیستم جهت گمراه کردن تحلیلگران، تولید مکرر دامنه‌های مختلف برای دشوار کردن به روزرسانی فهرست پایگاه داده مخرب شناخته شده و دیگر روش‌ها	Spoofing Tampering	اعتبارسنجی امضای فیدهای هوش تهدید پایش تغییرات غیرعادی در Indicator ها بررسی فیدهای از چندین منبع

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۵۵	دور زدن قابلیت محدودسازی تعداد درخواست‌های قابل ارسال کلاینت‌ها در یک بازه زمانی مشخص با استفاده از روش‌هایی همانند بات‌های مجهز به AI جهت شبیه‌سازی رفتارهای انسان در تعامل با اپلیکیشن، استفاده از بات‌نت‌ها برای ارسال درخواست‌ها از IP های مختلف به وب‌اپلیکیشن،	Spoofing	تحلیل رفتار کاربران بکارگیری مکانیزم Adaptive Rate Limiting بر اساس Context استفاده از کپچاهای پویا
۵۶	دور زدن قابلیت تشخیص و مسدود کردن حملات وب‌اپلیکیشن‌ها همچون SQL Injection، XSS، CSRF و غیره با استفاده از روش‌هایی همچون استفاده از encodingهای مختلف همانند URL encoding، Base64 و غیره برای دور زدن فیلترهای ایستا، تقسیم پیلودها در فرگمنت‌های مختلف، آموزش مدل‌های هوش مصنوعی جهت شناسایی الگوهای کوئری‌های SQL و دیگر روش‌ها	Spoofing	استفاده از WAFهای هوشمند اعتبارسنجی ورودی‌ها سمت سرور و کلاینت
۵۷	دور زدن قابلیت پیشگیری از استخراج اطلاعات حساس با استفاده از تکنیک‌هایی همانند استفاده از مدل‌های NLP و LLM برای تحلیل داده‌ها جهت شناسایی اطلاعات حساس در آنها، استفاده از AI برای جاسازی و پنهان کردن اطلاعات حساس در فایل‌های به ظاهر معتبر، سوءاستفاده از AI برای ایجاد کانال‌های ارتباطی پنهان در ترافیک مجاز شبکه و دیگر روش‌ها	Spoofing Tampering	بکارگیری مکانیزم‌های رمزنگاری End-to-End ایجاد Watermark برای داده‌های حساس بررسی محتواها به صورت context-based و چند لایه
۵۸	دور زدن قابلیت Sandboxing با استفاده از AI برای طراحی و آموزش بدافزارها جهت تشخیص این محیط‌ها	Spoofing	اجرا در Sandboxهای متفاوت
۵۹	سوءاستفاده از قابلیت VPN با استفاده از روش‌هایی همانند تقلید رفتارهای کاربران راه دور، بهره‌برداری از آسیب‌پذیری‌های endpointهای شناخته شده و مورد اعتماد با استفاده از AI و دیگر روش‌ها	Spoofing برخی دیگر از دسته‌ها	استفاده از مکانیزم‌های MFA برای اتصال VPN تحلیل ترافیک داخل Tunnel بکارگیری مکانیزم‌های Session Fingerprinting

ردیف	عنوان تهدید	دسته‌بندی تهدید	برخی راهکارهای مقابله با تهدید
۶۰	دور زدن قابلیت تشخیص ایمیل‌های فیشینگ با استفاده از روش‌هایی همانند استفاده از NLP برای ایجاد ایمیل‌های فیشینگ شخصی‌سازی‌شده، استفاده از AI برای تولید URLهای به ظاهر معتبر، دور زدن اسکن اولیه URLها با استفاده از سرویس‌های URL shortener، استفاده از تکنولوژی deepfake برای ایجاد پروفایل‌ها و محتواهای کاملا مشابه با واقعیت و غیره	Spoofing	ترکیب فیلترینگ مبتنی بر محتوا و رفتار آگاه سازی کاربران استفاده از DMARC سختگیرانه
۶۱	دور زدن قابلیت فیلترینگ ایمیل‌های اسپم با استفاده از مدل‌های NLP برای تولید ایمیل‌های اسپم شخصی‌سازی‌شده و موضوعی[1] که احتمال فیلترینگ آنها توسط راهکارهای مبتنی بر keyword کم باشد، استفاده از AI برای تغییر پویای محتوا و ساختار ایمیل‌های اسپم و دیگر روش‌ها	Spoofing	استفاده از مدل‌های Spam Filtering مبتنی‌بر AI پایش blacklistها
۶۲	دور زدن قابلیت پایش تغییرات فایل‌ها با استفاده از روش‌هایی همانند ایجاد تغییرات تدریجی و مقیاس کوچک در فایل‌های حساس و غیره	Tampering	استفاده از هش و امضای دیجیتال برای بررسی صحت فایل ها پایش بلادرنگ فایل ها و صدور هشدارها
۶۳	دور زدن قابلیت تشخیص بدافزار با روش‌هایی همچون تولید بدافزارهای چندریختی و دگردیس[2] با استفاده از تکنیک‌هایی همانند الگوریتم‌های ژنتیک، استفاده از AI برای تولید تکنیک‌های پیچیده‌ی مبهم‌سازی کد، توسعه بدافزارهای fileless، استفاده از تکنیک‌های AML برای تولید ورودی‌های مخرب جهت گمراه شدن مدل‌های یادگیری ماشین و طبقه‌بندی بدافزارها، استفاده از تکنیک‌های یادگیری تقویتی (RL) برای آموزش بدافزارها و غیره	Tampering Spoofing	استفاده از EDRهای پیشرفته با تحلیل رفتاری استفاده از راهکار Sandboxing چندلایه بررسی تهدیدات به صورت مداوم
۶۴	بهره‌برداری از آسیب‌پذیری‌های سامانه یا زیرساخت آن با استفاده از AI برای شناسایی آسیب‌پذیری‌های روز صفرم یا پیکربندی‌های نادرست در سرورهای زیرساختی یا نرم‌افزار سامانه، بهره‌برداری از آسیب‌پذیری‌های شناسایی شده و اجرای انواع حملات همانند غیرفعال کردن پلتفرم با استفاده از فازینگ مبتنی بر هوش مصنوعی یا سایر تکنیک‌ها	تمام دسته‌بندی‌ها	تست نفوذ مستمر اعمال سریع وصله‌ها

1. contextual
2. Metamorphic