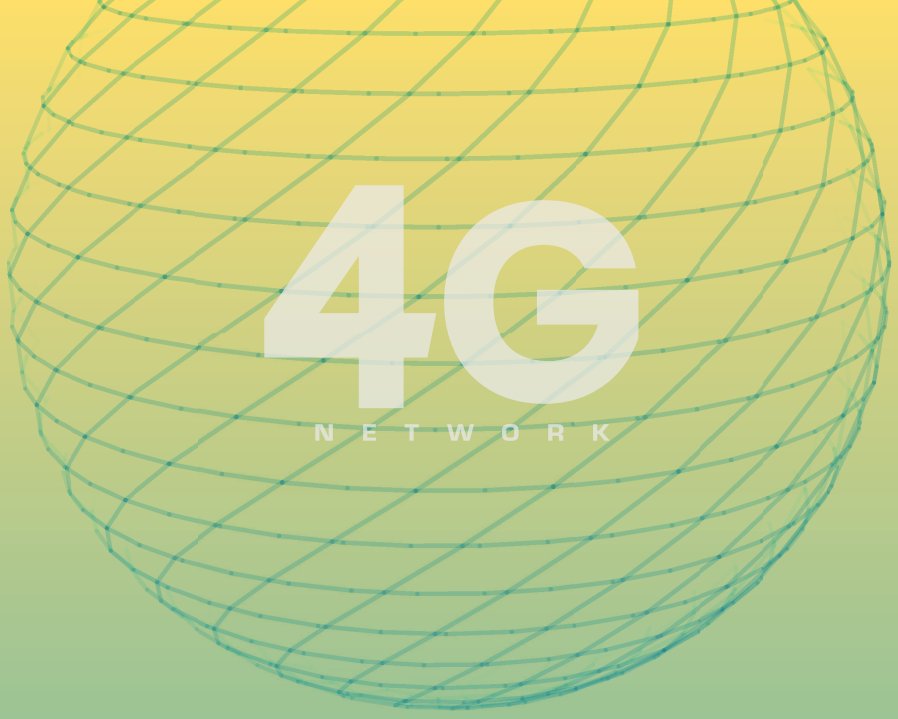


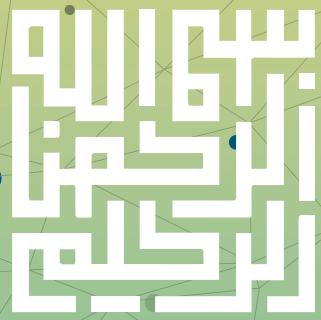
ITRC
پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

مجموعه حملات مبتنی بر هوش مصنوعی در شبکه‌های سیّار نسل چهارم

تاریخ انتشار: شهریور ۱۴۰۴

ITRC

پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)



عنوان گزارش: مجموعه حملات مبتنی بر هوش مصنوعی در شبکه‌های سیّار نسل چهارم

کلمات کلیدی: هوش مصنوعی، شبکه‌های نسل چهارم و حملات سایبری

تهیه کنندگان: محمد جاویدنیا، محمدعلی سبقتی 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)

خلاصه مدیریتی

این گزارش یک تحلیل جامع و تخصصی از تهدیدات سایبری نوظهور علیه شبکه‌های نسل چهارم (4G LTE) ارائه می‌دهد. با ظهور هوش مصنوعی (AI)[1]، چشم‌انداز تهدیدات به طور اساسی دگرگون شده است. این دگرگونی ماهیتی دوگانه دارد: از یک سو، هوش مصنوعی به عنوان یک توانمندساز، حملات سایبری سنتی را با افزایش سرعت، مقیاس و پنهان‌کاری به سطح جدیدی از پیچیدگی ارتقا می‌دهد. از سوی دیگر، با افزایش استفاده از مدل‌های یادگیری ماشین (ML)[2] در مدیریت و امنیت شبکه، یک سطح حمله کاملاً جدید به نام حملات تخاصمی[3] ایجاد شده است که خودِ سیستم‌های هوشمند را هدف قرار می‌دهد.

این گزارش به بررسی حملات هوش مصنوعی تهاجمی و تخاصمی[4]، کاربرد، تأثیر و راهکارهای مقابله با هر یک را در بستر معماری 4G LTE تشریح می‌کند. یافته‌های کلیدی نشان می‌دهد که بسیاری از آسیب‌پذیری‌های ذاتی در معماری 4G، مانند مدل امنیتی مبتنی بر «هسته قابل اعتماد»، عدم رمزنگاری ترافیک در بخش‌هایی از شبکه مرکزی و ضعف در احراز هویت ایستگاه‌های پایه، نقاط اهرمی ایده‌آلی برای بهره‌برداری توسط حملات پیشرفته مبتنی بر هوش مصنوعی فراهم می‌کنند.

نتیجه‌گیری نهایی این است که دفاع مؤثر در برابر این نسل جدید از تهدیدات، نیازمند یک استراتژی دفاعی دووجهی است: تقویت بنیان‌های امنیتی زیرساخت شبکه و به طور همزمان، ایجاد یک چارچوب امنیتی مستحکم برای حفاظت از خود سیستم‌های هوش مصنوعی که بر روی این زیرساخت اجرا می‌شوند.

1. Artificial intelligence (AI)
2. Machine Learning (ML)
3. Adversarial Attacks
۴. برای مطالعه و آشنایی با این حملات به سند *مرجع دانش حملات سایبری مبتنی بر هوش مصنوعی* مراجعه گردد.

ITRC
پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

فهرست مطالب

۱ چشم‌انداز امنیتی: 4G LTE معماری و آسیب‌پذیری‌های ذاتی..... ۹
۱-۱ مروری بر معماری: سیستم بسته تکامل‌یافته (EPS)..... ۹
۱-۲ رابط‌های حیاتی و پشته پروتکل..... ۱۰
۱-۳ مدل امنیتی 3GPP و محدودیت‌های آن..... ۱۱
۲ تحلیل حملات تهاجمی مبتنی بر هوش مصنوعی در شبکه‌های 4G..... ۱۴
۲-۱ شناسایی و پروفایل‌سازی هدف..... ۱۴
۲-۲ حملات به محرمانگی و یکپارچگی داده‌ها..... ۱۵
۲-۳ حملات به دسترسی، احراز هویت و هویت..... ۱۶
۲-۴ حملات به دسترس‌پذیری شبکه..... ۱۷
۲-۵ بدافزارهای پیشرفته و فرار از تشخیص..... ۱۸
۳ تحلیل حملات تخاصمی به سیستم‌های هوش مصنوعی در اکوسیستم 4G..... ۲۰
۳-۱ نقش یادگیری ماشین در عملیات 4G: سطح حمله جدید..... ۲۰
۳-۲ حملات به یکپارچگی داده و مدل (مسمومیت)..... ۲۱
۳-۳ حملات زمان استنتاج (گریز و نشت اطلاعات)..... ۲۲
۳-۴ حملات مبتنی بر مدل‌های زبانی بزرگ..... ۲۳
۴ چارچوب دفاعی یکپارچه و توصیه‌های استراتژیک..... ۲۵
۴-۱ معماری دفاع در عمق چندلایه و مقاوم در برابر هوش مصنوعی..... ۲۵
۴-۲ توصیه‌های راهبردی برای اپراتورهای شبکه..... ۲۷
۴-۳ توصیه‌هایی برای نهادهای استانداردسازی..... ۲۷
۵ نتیجه‌گیری..... ۲۹
منابع..... ۳۰

ITRC

پژوهشگاه ارتباطات
و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)

۱ چشم‌انداز امنیتی: 4G LTE معماری و آسیب‌پذیری‌های ذاتی

برای درک عمیق تهدیدات مبتنی بر هوش مصنوعی علیه شبکه‌های 4G، ابتدا باید معماری این شبکه‌ها و نقاط ضعف ذاتی آن‌ها را به دقت بررسی کرد. حملات پیشرفته امروزی اغلب از آسیب‌پذیری‌های موجود در طراحی اولیه سیستم بهره‌برداری می‌کنند. این بخش به تشریح اجزای کلیدی سیستم بسته تکامل‌یافته (EPS)[1]، رابط‌های حیاتی و مدل امنیتی 3GPP می‌پردازد و بستری برای تحلیل حملات در بخش‌های بعدی فراهم می‌کند.

۱-۱ مروری بر معماری: سیستم بسته تکامل‌یافته (EPS)

معماری سطح بالای 4G LTE، که به عنوان سیستم بسته تکامل‌یافته شناخته می‌شود، از دو حوزه اصلی تشکیل شده است:

- شبکه دسترسی رادیویی زمینی UMTS تکامل‌یافته (E-UTRAN) و هسته بسته تکامل‌یافته (EPC).
- شبکه دسترسی رادیویی تکامل‌یافته (E-UTRAN)

شبکه E-UTRAN مسئول تمام ارتباطات رادیویی بین تجهیزات کاربر (UE) و شبکه مرکزی است. برخلاف معماری‌های سلسله‌مراتبی نسل‌های قبلی، E-UTRAN دارای یک معماری مسطح است که تنها از یک نوع گره تشکیل شده است: ایستگاه پایه تکامل‌یافته یا eNodeB (eNB). هر eNB به طور مستقیم با شبکه مرکزی از طریق رابط S1 و با eNBهای مجاور از طریق رابط X2 در ارتباط است. این طراحی مسطح، تأخیر را کاهش داده و کارایی را افزایش می‌دهد. وظایف اصلی eNB عبارتند از:

- **مدیریت منابع رادیویی:** ارسال و دریافت امواج رادیویی، تخصیص منابع به UEها و مدیریت انتقال بین سلول‌ها.
- **کنترل سطح پایین UE:** ارسال پیام‌های سیگنالینگ برای کنترل عملیات موبایل.

هسته بسته تکامل‌یافته (EPC)

EPC مغز شبکه 4G است و تمام عملیات کنترلی و مدیریت ترافیک داده را بر عهده دارد. اجزای کلیدی EPC عبارتند از:

1. Evolved Packet System - EPS

- **موجودیت مدیریت پویایی[1] (MME):** این گره، مرکز عصبی صفحه کنترل[2] است. MME مسئول ردیابی موقعیت UE، مدیریت نشست‌ها و مهم‌تر از همه، فرآیند احراز هویت UE با همکاری HSS است. تمام سیگنالینگ لایه غیر دسترسی (NAS) بین UE و شبکه در MME خاتمه می‌یابد.
- **درگاه سرویس‌دهی[3] (SGW):** SGW به عنوان لنگر پویایی برای صفحه کاربر[4] عمل می‌کند. این گره مسئول مسیریابی و ارسال بسته‌های داده کاربر بین eNB و PGW است و در حین handover بین eNBها، جریان داده را بدون وقفه حفظ می‌کند.
- **درگاه شبکه داده بسته[5] (PGW):** PGW نقطه اتصال شبکه موبایل به شبکه‌های داده خارجی مانند اینترنت است. این گره مسئول تخصیص آدرس IP به UEها، اعمال سیاست‌های کیفیت سرویس (QoS) و فیلترینگ بسته‌ها است. PGW نقش مشابه GGSN در شبکه‌های 3G را ایفا می‌کند.
- **سرور مشترکین خانگی[6] (HSS):** HSS یک پایگاه داده مرکزی و حیاتی است که اطلاعات مربوط به اشتراک و پروفایل کاربران را در خود ذخیره می‌کند. این اطلاعات شامل کلیدهای امنیتی برای احراز هویت، اطلاعات رومینگ و سرویس‌هایی که کاربر مجاز به استفاده از آن‌هاست، می‌باشد. HSS همچنین اطلاعات پویایی مانند MMEای که کاربر در حال حاضر به آن متصل است را نگهداری می‌کند.
- **تابع قوانین شارژ و سیاست‌گذاری[7] (PCRF):** PCRF مسئول تصمیم‌گیری در مورد سیاست‌های QoS و کنترل قوانین شارژ مبتنی بر جریان داده است. این گره با PGW در ارتباط است تا اطمینان حاصل کند که ترافیک هر کاربر مطابق با طرح اشتراک او مدیریت می‌شود.

۱-۲ رابط‌های حیاتی و پشته پروتکل

ارتباط بین این اجزای مختلف از طریق رابط‌های استانداردشده‌ای صورت می‌گیرد که هر کدام از پشته‌های پروتکلی خاصی استفاده می‌کنند. امنیت این رابط‌ها برای امنیت کل شبکه حیاتی است. این رابط‌ها عبارتند از:

1. Mobility Management Entity (MME)
2. Control Plane
3. Serving Gateway (SGW)
4. User Plane
5. Packet Data Network Gateway (PGW)
6. Home Subscriber Server (HSS)
7. Policy and Charging Rules Function (PCRF)

- **Uu:** رابط رادیویی بین UE و eNB.
- **S1:** رابط بین eNB و EPC که به دو بخش تقسیم می‌شود: S1-MME برای سیگنالینگ صفحه کنترل و S1-U برای ترافیک صفحه کاربر.
- **X2**: رابط بین eNBهای مجاور برای هماهنگی handover و کاهش تداخل.
- **S6a**: رابط حیاتی بین MME و HSS که از پروتکل Diameter برای تبادل اطلاعات احراز هویت و پروفایل کاربر استفاده می‌کند.
- **S11**: رابط بین MME و SGW برای مدیریت نشست‌ها.
- **S5/S8**: رابط بین SGW و PGW که از پروتکل GT[1] برای تونل کردن ترافیک کاربر استفاده می‌کند.
- **SGi**: رابط بین PGW و شبکه‌های IP خارجی (اینترنت).
- آسیب‌پذیری‌های پروتکل‌ها

علاوه بر رابطها دو پروتکل Diameter و GTP نیز اهمیت ویژه‌ای در شبکه‌های 4G دارند. این دو پروتکل به دلیل طراحی اولیه خود که بیشتر بر عملکرد و قابلیت همکاری متمرکز بوده تا امنیت، دارای آسیب‌پذیری‌های قابل توجهی هستند که توسط حملات پیشرفته قابل بهره‌برداری هستند:

- **پروتکل Diameter:** این پروتکل که در رابطهای حیاتی مانند S6a استفاده می‌شود، اغلب در شبکه‌های بین اپراتوری (مانند IPX) بدون رمزنگاری یا فیلترینگ مناسب پیاده‌سازی می‌شود. این امر به یک شریک رومینگ مخرب یا یک مهاجم که به شبکه IPX دسترسی پیدا کرده، اجازه می‌دهد تا با ارسال پیام‌های Diameter جعلی، اطلاعات مشترکین (مانند IMSI و موقعیت مکانی) را استعلام کند، سرویس آن‌ها را قطع یا منحرف نماید، یا با ایجاد طوفان سیگنالینگ، MME و HSS را دچار حمله منع سرویس (DoS) کند.
- **پروتکل GT:** این پروتکل که برای انتقال ترافیک کاربر در هسته شبکه استفاده می‌شود، فاقد مکانیزم‌های داخلی احراز هویت و رمزنگاری برای پیام‌های کنترلی (GTP-C) است. اگر یک مهاجم بتواند به شبکه داخلی اپراتور نفوذ کند، می‌تواند با جعل پیام‌های GTP-C، نشست‌های کاربران را بدزدد، هویت آن‌ها را جعل کند، یا با مصرف منابع SGW/PGW، حمله DoS انجام دهد.

۱-۳ مدل امنیتی 3GPP و محدودیت‌های آن

3GPP یک معماری امنیتی چندلایه برای LTE تعریف کرده است که شامل احراز هویت متقابل (AKA)،

1. GPRS Tunneling Protocol (GTP)

حفاظت از یکپارچگی و رمزنگاری در لایه‌های NAS (بین UE و MME) و AS (در رابط رادیویی) می‌شود. با این حال، این مدل امنیتی دارای محدودیت‌ها و مفروضات طراحی است که امروزه به عنوان آسیب‌پذیری‌های جدی شناخته می‌شوند:

- **امنیت تکه‌تکه**[1]**:** امنیت در دامنه‌های مجزا (دسترسی به شبکه، دامنه شبکه، دامنه کاربر) با مکانیزم‌های غیرهماهنگ پیاده‌سازی شده است. این امر تصویری پیچیده و بالقوه دارای شکاف از وضعیت امنیتی کلی شبکه LTE ایجاد می‌کند و ممکن است منجر به اعمال سیاست‌های امنیتی متناقض یا اضافی شود.
- تمرکز بر تهدیدات خارجی: مدل امنیتی 4G به شدت بر حفاظت از محیط شبکه در برابر حملات خارجی که از طریق اینترنت یا رابط رادیویی نشأت می‌گیرند، متمرکز است. توجه کمتری به حملات داخلی یا تهدیداتی که از سوی شرکای به ظاهر قابل اعتماد (مانند اپراتورهای رومینگ) سرچشمه می‌گیرند، شده است. این فلسفه طراحی، یک نقطه کور خطرناک در برابر تهدیدهای پایدار پیشرفته (APT) ایجاد می‌کند.
- **ترافیک رمزنگاری‌نشده در هسته شبکه:** یکی از بزرگترین و حیاتی‌ترین آسیب‌پذیری‌های 4G این است که اگرچه ترافیک روی رابط رادیویی (Uu) رمزنگاری می‌شود، ترافیک صفحه کاربر در داخل هسته شبکه (در رابط‌های S1-U، S5/S8 و SGi) اغلب به صورت IP رمزنگاری‌نشده منتقل می‌شود. این یک تصمیم طراحی مبتنی بر عملکرد بود که در زمان خود منطقی به نظر می‌رسید، اما امروزه یک ریسک بزرگ امنیتی محسوب می‌شود. یک مهاجم داخلی یا کسی که به شبکه حمل‌ونقل اپراتور نفوذ کرده باشد، می‌تواند به راحتی ترافیک کاربران را شنود کند.
- **فقدان احراز هویت ایستگاه پایه (eNodeB):** در فرآیند اتصال اولیه، UE هیچ راهی برای احراز هویت eNB ندارد. این آسیب‌پذیری ریشه‌ای، امکان حملات "ایستگاه پایه سیار جعلی" (Rogue eNodeB) یا "IMSI Catcher" را فراهم می‌کند. یک مهاجم می‌تواند با برپایی یک eNB جعلی با سیگنال قوی‌تر، UEهای اطراف را فریب داده و به سمت خود بکشاند و سپس حملات مرد میانی (MitM)، استراق سمع، ردیابی موقعیت، منع سرویس (DoS) یا حملات تنزل به شبکه‌های ناامن‌تر 2G/3G را اجرا کند.

این محدودیت‌ها نشان می‌دهند که فلسفه امنیتی معماری 4G اساساً بر یک مدل «هسته قابل اعتماد»[2] بنا شده است، که در آن فرض می‌شود موجودیت‌های داخل EPC و شبکه‌های متصل

1. Fragmented Security
2. Trusted Core

اپراتورهای همکار، امن هستند. این فرض در عصر تهدیدات پیشرفته و حملات مبتنی بر هوش مصنوعی که در نفوذ به محیط‌های به ظاهر امن تخصص دارند، به شدت منسوخ و خطرناک است. حملات مبتنی بر هوش مصنوعی دقیقاً برای بهره‌برداری از همین فرض طراحی شده‌اند. آن‌ها می‌توانند با روش‌های پنهانکارانه (مانند فیشینگ پیشرفته) محیط پیرامونی عبور کرده یا از کانال‌های قابل اعتماد (مانند یک شریک رومینگ به خطر افتاده) سوءاستفاده کنند. پس از ورود به این محیط «قابل اعتماد»، یک ابزار مبتنی بر هوش مصنوعی می‌تواند از ترافیک داخلی رمزنگاری‌نشده و پروتکل‌های ضعیف با کارایی ویرانگر بهره‌برداری کرده، به صورت جانبی حرکت کند و داده‌ها را استخراج نماید. بنابراین، فرض امنیتی بنیادین هسته 4G، خود به بزرگترین ضعفی تبدیل شده که تهدیدات مبتنی بر هوش مصنوعی به طور منحصربه‌فردی برای بهره‌برداری از آن موقعیت یافته‌اند. این یک آسیب‌پذیری سیستمی است، نه یک نقص در یک جزء خاص.

۲ تحلیل حملات تهاجمی مبتنی بر هوش مصنوعی در شبکه‌های ۴G

این بخش به تحلیل نظام‌مند حملات تهاجمی در شبکه 4G می‌پردازد. این حملات، که نسخه‌های پیشرفته‌تر و هوشمندتر تهدیدات سایبری سنتی هستند، با استفاده از هوش مصنوعی در سرعت، مقیاس و پنهان‌کاری تقویت شده‌اند. تحلیل‌ها بر اساس دسته‌بندی موضوعی ارائه شده و برای هر حمله، ارتباط آن با شبکه 4G، دارایی هدف، تأثیر و راهکارهای مقابله تشریح می‌شود.

۲-۱ شناسایی و پروفایل‌سازی هدف

این دسته از حملات بر جمع‌آوری اطلاعات و آماده‌سازی برای حملات بعدی متمرکز هستند. هوش مصنوعی این مرحله را که به طور سنتی زمان‌بر و پرهزینه بود، متحول کرده و آن را خودکار و بهینه می‌سازد.

- **تحلیل ترافیک شبکه مبتنی بر هوش مصنوعی:** این حمله، رابطهای رمزنگاری‌نشده یا با رمزنگاری ضعیف مانند S1-U (بین eNB و SGW) و به‌ویژه SGi (بین PGW و اینترنت) را هدف قرار می‌دهد. یک مدل هوش مصنوعی می‌تواند حجم عظیمی از فراداده ترافیک را تحلیل کرده و الگوهای ارتباطی را شناسایی کند. این تحلیل می‌تواند به شناسایی اهداف باارزش (مانند ترافیک مربوط به سازمان‌های دولتی یا شرکت‌های بزرگ)، نقشه‌برداری از توپولوژی داخلی شبکه اپراتور و شناسایی سرورهای حیاتی (مانند DNS یا سرورهای OAM) منجر شود.
- **تشخیص آسیب‌پذیری مبتنی بر هوش مصنوعی:** مهاجمان از هوش مصنوعی برای اسکن خودکار و هوشمند اجزای شبکه مانند HSS، PGW، MME و پورتال‌های مدیریتی استفاده می‌کنند. این سیستم‌ها می‌توانند نه تنها آسیب‌پذیری‌های شناخته‌شده[1] (CVEs) را با سرعتی فراتر از توان انسان پیدا کنند، بلکه با استفاده از تکنیک‌های یادگیری تقویتی، به دنبال یافتن آسیب‌پذیری‌های روز-صفر در پروتکل‌ها و نرم‌افزارهای شبکه باشند.
- **پروفایل‌سازی هوشمند هدف‌محور مبتنی بر هوش مصنوعی:** این حمله فراتر از تحلیل داده‌های شبکه می‌رود. یک سیستم هوش مصنوعی می‌تواند داده‌های ترافیکی را با اطلاعات منابع آشکار (OSINT) از اینترنت (مانند پروفایل‌های شبکه‌های اجتماعی) ترکیب کند تا

1. Common Vulnerabilities and Exposures

پروفایل‌های جامعی از اهداف انسانی (مانند کارمندان کلیدی اپراتور یا مدیران ارشد یک شرکت مشترک) بسازد. این پروفایل‌های دقیق، اثربخشی حملات دیگر یعنی مهندسی اجتماعی و فیشینگ هدفمند را به شدت افزایش می‌دهد.

- **کش‌کاوی مبتنی بر هوش مصنوعی و یافتن مسیر مبتنی بر هوش مصنوعی:** این حملات پس از نفوذ اولیه به شبکه اپراتور کاربرد دارند. حمله کش‌کاوی می‌تواند در محیطهای مجازی‌شده (NFV) که اجزای EPC به صورت ماشین‌های مجازی روی سخت‌افزار مشترک اجرا می‌شوند، برای استخراج اطلاعات از طریق حملات کانال جانبی استفاده شود. حمله یافتن مسیر نیز به مهاجم اجازه می‌دهد تا با تحلیل نقشه شبکه و الگوهای ترافیکی، آسیب‌پذیرترین و کم‌نظارت‌ترین مسیر را برای رسیدن به دارایی‌های حیاتی مانند HSS پیدا کند.

موفقیت در این مرحله، پایه‌گذار تمام حملات مخرب بعدی است. راهکارهای اصلی مقابله شامل رمزنگاری سرتاسری با استفاده از IPsec بر روی تمام رابطهای هسته شبکه (S1-U, S5/S8, X2)، بخش‌بندی دقیق شبکه برای محدود کردن دید و حرکت جانبی مهاجم، به حداقل رساندن اطلاعات پخش‌شده و استفاده از سیستم‌های تشخیص ناهنجاری مبتنی بر هوش مصنوعی برای شناسایی الگوهای اسکن غیرعادی یا دسترسی‌های مشکوک به داده‌ها است.

۲-۲ حملات به محرمانگی و یکپارچگی داده‌ها

این حملات با هدف سرقت یا دستکاری داده‌های حساس مشترکین و شبکه انجام می‌شوند.

- **جاسوسی و استخراج داده‌ها مبتنی بر هوش مصنوعی:** مدل‌های هوش مصنوعی می‌توانند برای شناسایی و استخراج انواع خاصی از داده‌های حساس از درون جریان ترافیک کاربر (در رابط SGi) یا داده‌های سیگنالینگ (در رابط S6a) آموزش داده شوند. مهم‌تر اینکه، هوش مصنوعی می‌تواند فرآیند استخراج داده را با تقلید از الگوهای ترافیک عادی، بسیار پنهانکارانه انجام دهد تا از دید سیستم‌های تشخیص نفوذ (NIDS) پنهان بماند.
- **رمزشکنی مبتنی بر هوش مصنوعی:** اگرچه شکستن الگوریتم‌های رمزنگاری مدرن مانند AES در حال حاضر با هیچ روشی، از جمله هوش مصنوعی، امکان‌پذیر نیست، اما هوش مصنوعی می‌تواند برای یافتن نقاط ضعف در پیاده‌سازی پروتکل‌های رمزنگاری، فرآیندهای مدیریت کلید یا بهره‌برداری از الگوریتم‌های قدیمی و ضعیف‌تر در صورت پیکربندی نادرست شبکه، به کار گرفته شود.
- **مرد میانی (MitM) مبتنی بر هوش مصنوعی:** هوش مصنوعی فرآیند اجرای حملات MitM را خودکار می‌کند. به عنوان مثال، یک سیستم هوشمند می‌تواند به طور خودکار یک ایستگاه

پایه سیار جعلی[1] را برای رهگیری ارتباط UE با شبکه مدیریت کند یا با بهره‌برداری از آسیب‌پذیری‌های پروتکل‌های GTP و Diameter در داخل هسته شبکه، ترافیک را شنود یا دستکاری کند.

- **سرقت اعتبارنامه و حملات رمز عبور:** هوش مصنوعی این حملات کلاسیک را به شدت تقویت می‌کند. مدل‌های زبانی بزرگ (LLMs) می‌توانند از مجموعه داده‌های عظیم رمزهای عبور فاش‌شده، الگوهای پیچیده را یاد بگیرند و حدس‌های هوشمند و زمینه‌آگاه برای رمزهای عبور رابط‌های مدیریتی تجهیزات شبکه یا حساب‌های کاربری ایجاد کنند. این روش بسیار کارآمدتر از حملات بروت فورس سنتی است.

این حملات می‌توانند منجر به نقض حریم خصوصی میلیون‌ها مشترک، سرقت اسرار تجاری و دولتی، کلاهبرداری‌های مالی گسترده و از دست رفتن یکپارچگی و اعتماد به شبکه شوند. راهکارهای دفاعی شامل الزام به رمزنگاری و حفاظت از یکپارچگی قوی برای تمام لینک‌های داده و سیگنالینگ (از جمله بک‌هال شبکه مرکزی)، کنترل دسترسی قوی مانند احراز هویت چندعاملی (MFA)[2] برای تمام سیستم‌های مدیریتی و تقویت امنیت در سطح تجهیزات کاربر (UE) است.

۲-۳ حملات به دسترسی، احراز هویت و هویت

این دسته از حملات بر روی به دست آوردن دسترسی غیرمجاز به شبکه یا جعل هویت موجودیت‌های قانونی متمرکز هستند.

- **فیشینگ هدفمند مبتنی بر هوش مصنوعی:** این حمله اغلب نقطه شروع یک APT است. هدف آن کارمندان اپراتور مخابراتی برای به دست آوردن نفوذ اولیه به شبکه داخلی شرکت است. از آنجا، مهاجم می‌تواند حملات خود را به سمت EPC گسترش دهد. هوش مصنوعی با تولید ایمیل‌های بسیار شخصی‌سازی‌شده و واقع‌گرایانه، نرخ موفقیت این حملات را به شدت افزایش می‌دهد.
- **جعل و کلون‌سازی مبتنی بر هوش مصنوعی:** یک تهدید بسیار جدی برای 4G، جعل سیگنال است. یک مهاجم می‌تواند با استفاده از شبکه‌های مولد تخاصمی، سیگنال‌های رادیویی مصنوعی تولید کند که ویژگی‌های لایه فیزیکی یک UE قانونی را به طور کامل تقلید می‌کنند. این سیگنال‌های جعلی می‌توانند مکانیزم‌های احراز هویت لایه فیزیکی را دور زده و به مهاجم اجازه دهند یک UE را کلون کرده و به صورت غیرمجاز از سرویس‌ها استفاده کند.
- **حملات تزریق (SQL و XSS) مبتنی بر هوش مصنوعی:** هوش مصنوعی می‌تواند به طور

1. Rogue eNB
2. Multi-Factor Authentication (MFA)

خودکار وب‌سایت‌ها و پورتال‌های مدیریتی اپراتور را برای یافتن آسیب‌پذیری‌های تزریق اسکریپت از طریق وب‌گاه (XSS) و تزریق SQL پویش کرده و از آن‌ها بهره‌برداری کند. این حملات می‌توانند به سرقت نشست‌های مدیریتی یا دسترسی به پایگاه‌های داده حساس مانند اطلاعات صورت‌حساب منجر شوند.

- **حرکت جانبی مبتنی بر هوش مصنوعی:** پس از کسب دسترسی اولیه به یک سیستم در شبکه اپراتور، هوش مصنوعی می‌تواند فرآیند حرکت جانبی را خودکار کند. این سیستم با استفاده از تکنیک‌هایی مانند سرقت اعتبارنامه و یافتن مسیرهای آسیب‌پذیر، به صورت خودکار از یک سیستم کم‌اهمیت به گره‌های حیاتی مانند MME یا HSS نفوذ می‌کند.

این حملات می‌توانند به دسترسی غیرمجاز به شبکه، استفاده متقلبانه از سرویس‌ها و در نهایت، به خطر افتادن کامل کنترل شبکه منجر شوند. راهکارهای کلیدی شامل احراز هویت چندعاملی (MFA) برای همه سیستم‌ها، انگشت‌نگاری لایه فیزیکی برای شناسایی دستگاه‌ها بر اساس ویژگی‌های منحصربه‌فرد سیگنال رادیویی آن‌ها، دفاع در برابر جعل مبتنی بر GAN (مانند استفاده از مکانیزم‌های چالش-پاسخ) و پیاده‌سازی یک معماری اعتماد صفر (Zero-Trust) در هسته شبکه برای جلوگیری از حرکت جانبی است.

۲-۴ حملات به دسترس‌پذیری شبکه

این حملات از مخرب‌ترین تهدیدات علیه یک اپراتور مخابراتی هستند و هدف آن‌ها قطع کامل یا جزئی سرویس است. این حملات منجر به قطعی کامل سرویس برای مشترکین در یک منطقه هدف، آسیب شدید به اعتبار اپراتور و زیان‌های مالی هنگفت می‌شوند.

- **حمله DDoS مبتنی بر هوش مصنوعی:** هوش مصنوعی حملات منع سرویس توزیع‌شده (DDoS) را به سطح جدیدی از پیچیدگی می‌رساند. به جای ارسال حجم زیادی از ترافیک ساده، یک حمله DDoS هوشمند، تطبیقی است. این حمله می‌تواند هر گره‌ای که در معرض IP عمومی قرار دارد (مانند MME، PGW، سرورهای DNS) یا لینک‌های ارتباطی (مانند رابط SGi یا رابط رادیویی) را هدف قرار دهد. هوش مصنوعی به طور مداوم پاسخ‌های سیستم دفاعی شبکه را تحلیل کرده و الگوی حمله (مانند نوع پروتکل، حجم ترافیک، منبع) را تغییر می‌دهد تا از مکانیزم‌های دفاعی ایستا فرار کند. استفاده از سرویس‌های پیشرفته کاهش DDoS که خود از هوش مصنوعی برای شناسایی و پاکسازی حملات تطبیقی استفاده می‌کنند، می‌تواند راهکاری برای کاهش اثرات این نوع حمله گردد.
- **جمینگ مبتنی بر هوش مصنوعی:** این حمله یک تهدید بسیار کارآمد و پنهانکارانه برای رابط رادیویی است. برخلاف جمینگ سنتی که با ارسال نویز پرقدرت در یک باند فرکانسی وسیع

عمل می‌کند، جمینگ هوشمند از یک رادیوی نرم‌افزارمحور (SDR) برای یادگیری ساختار دقیق فریم LTE استفاده می‌کند. سپس، مهاجم تداخل رادیویی کم‌توان و هدفمندی را دقیقا در لحظه ارسال کانال‌های کنترلی حیاتی مانند کانال پخش فیزیکی (PBCH)، سیگنال‌های همگام‌سازی اولیه و ثانویه (PSS/SSS) یا نشانگر فرمت کنترل فیزیکی (PCFICH) ارسال می‌کند. این کار مانع از همگام‌سازی UEها با شبکه یا دریافت اطلاعات کنترلی ضروری می‌شود و می‌تواند با مصرف انرژی بسیار کم و احتمال شناسایی پایین، باعث قطع سرویس در یک سلول شود. برای مقابله و کاهش اثر حمله، پیاده‌سازی پردازش سیگنال پیشرفته در eNB برای شناسایی و کاهش جمینگ هوشمند (مانند فیلتر کردن فضایی با استفاده از آنتن‌های چندورودی-چندخروجی انبوه[1] و تکنیک‌های پرش فرکانسی) و نظارت دقیق بر معیارهای کیفیت لینک رادیویی، می‌تواند راهگشا باشد.

- **باج‌افزار مبتنی بر هوش مصنوعی:** این حمله می‌تواند سیستم‌های فناوری اطلاعات (IT) و مدیریت و نگهداری (OAM)[2] اپراتور را هدف قرار دهد. یک باج‌افزار هوشمند می‌تواند به طور خودکار در شبکه حرکت کرده، مهم‌ترین دارایی‌ها مانند نسخه‌های پشتیبان پایگاه داده مشترکین (HSS) یا فایل‌های پیکربندی شبکه را شناسایی و با اولویت بالا رمزگذاری کند که می‌تواند عواقب فاجعه‌باری داشته باشد. رعایت بهترین شیوه‌های امنیت سازمانی مانند تهیه نسخه‌های پشتیبان منظم و ایزوله، بخش‌بندی شبکه و استفاده از راه‌حل‌های تشخیص و پاسخ نقطه پایانی (EDR) راهکاری مناسب برای مقابله با این حمله است.

۲-۵ بدافزارهای پیشرفته و فرار از تشخیص

این مجموعه از حملات بر هوشمندسازی بدافزارها و دشوار کردن فرآیند شناسایی آن‌ها تمرکز دارد.

- **مبهم‌سازی و فرار از تشخیص بدافزار مبتنی بر هوش مصنوعی:** هوش مصنوعی، به‌ویژه شبکه‌های GAN، می‌توانند برای تولید مستمر بدافزارهای چندریختی[3] و دگرریختی[4] استفاده شوند. این تکنیک‌ها تضمین می‌کنند که هر نمونه جدید از بدافزار دارای یک امضای منحصربه‌فرد است و به این ترتیب، سیستم‌های آنتی‌ویروس و NIDS سنتی که مبتنی بر امضا هستند را به راحتی دور می‌زنند.
- **بدافزارهای خودآموز مبتنی بر هوش مصنوعی:** این نوع بدافزار پس از نفوذ به شبکه اپراتور، از محیط خود یاد می‌گیرد. این بدافزار می‌تواند ساختار شبکه، مکانیزم‌های دفاعی موجود و

1. Massive MIMO
2. Operation, Administration, and Maintenance (OAM)
3. Polymorphic
4. Metamorphic

رفتار کاربران عادی را تحلیل کرده و رفتار خود را برای پنهان ماندن و دستیابی به اهدافش تطبیق دهد.

- **فرار از سامانه تشخیص نفوذ شبکه مبتنی بر هوش مصنوعی:** این حمله یک پیش‌زمینه مستقیم برای حملات تخاصمی گریز است. هوش مصنوعی مهاجم، قوانین و الگوهایی که یک NIDS به دنبال آن‌هاست را یاد می‌گیرد و سپس ترافیک مخرب خود را به گونه‌ای دستکاری می‌کند که هیچ هشداری را فعال نکند.

این حملات منجر به نفوذ موفق و پایدار به اجزای شبکه یا سیستم‌های IT اپراتور می‌شوند. راهکار کلیدی، گذار از تشخیص مبتنی بر امضا به سمت تشخیص مبتنی بر رفتار و ناهنجاری است که اغلب خود از مدل‌های هوش مصنوعی برای دفاع بهره می‌برد. استفاده از محیط‌های ایزوله[1] برای تحلیل فایل‌های مشکوک و آموزش تخاصمی برای مدل‌های امنیتی نیز ضروری است.

یک جنبه اساسی که این حملات را متمایز می‌کند، ایجاد یک «مزیت عملیاتی-زمانی»[2] برای مهاجم است. سرعت، مقیاس و تطبیق‌پذیری حملات مبتنی بر هوش مصنوعی می‌تواند مکانیزم‌های دفاعی ایستا یا متکی بر انسان را کاملاً فلج کند. چرخه حیات یک حمله سنتی (شامل شناسایی، بهره‌برداری، استخراج داده) نیازمند زمان و تلاش دستی قابل توجهی است و مدافعان از این شکاف زمانی برای شناسایی و پاسخ استفاده می‌کنند. اما حملات مبتنی بر هوش مصنوعی (به‌ویژه **با هماهنگی حملات مبتنی بر هوش مصنوعی**) می‌توانند کل این چرخه را خودکار کرده و فشرده سازند. یک سیستم هوش مصنوعی می‌تواند در یک کمپین هماهنگ و با سرعت ماشین، شناسایی را انجام دهد، یک آسیب‌پذیری را کشف کند، یک بدافزار چندریختی بسازد، آن را مستقر کند، به صورت جانبی حرکت کند و داده‌ها را استخراج نماید. این امر «زمان ماندگاری»[3] حمله را به شدت کاهش داده و پنجره زمانی مدافعان انسانی برای واکنش را تقریباً از بین می‌برد. علاوه بر این، ماهیت تطبیقی این حملات (مانند سازگارسازی حمله مبتنی بر هوش مصنوعی) به این معناست که قوانین دفاعی ایستا یا امضاها تقریباً بلافاصله منسوخ می‌شوند. بنابراین، تهدید اصلی فقط «هوشمندتر» بودن حملات نیست، بلکه این است که آن‌ها با سرعتی عمل می‌کنند که از چرخه‌های پاسخ انسانی پیشی می‌گیرد. این امر، حرکت به سمت دفاع و پاسخ خودکار و مبتنی بر هوش مصنوعی (مانند پلتفرم‌های SOAR) را به یک ضرورت استراتژیک تبدیل می‌کند.

1. Sandboxing
2. Tempo-Operational Advantage
3. Dwell Time

۳ تحلیل حملات تخاصمی به سیستم‌های هوش مصنوعی در اکوسیستم 4G

این بخش، تمرکز را از حملاتی که توسط هوش مصنوعی تقویت می‌شوند، به حملاتی که خود سیستم‌های هوش مصنوعی را هدف قرار می‌دهند، تغییر می‌دهد. این یک بردار تهدید اساساً جدید برای اپراتورهای مخابراتی است که به طور فزاینده‌ای در حال به کارگیری یادگیری ماشین برای مدیریت و امنیت شبکه خود هستند.

۳-۱ نقش یادگیری ماشین در عملیات 4G: سطح حمله جدید

برای درک حملات تخاصمی، ابتدا باید بدانیم که یادگیری ماشین (ML) در چه بخشی از اکوسیستم 4G در حال استفاده است. این فناوری دیگر یک مفهوم نظری نیست و در حوزه‌های عملیاتی زیر به کار گرفته شده است:

- **سیستم‌های تشخیص نفوذ شبکه (NIDS):** اپراتورها در حال گذار از سیستم‌های مبتنی بر امضا به سمت سیستم‌های تشخیص ناهنجاری هستند که از مدل‌های یادگیری ماشین برای شناسایی حملات روز-صفر و الگوهای تهدید پیچیده استفاده می‌کنند. این سیستم‌ها با تحلیل حجم عظیمی از ترافیک شبکه، رفتار «عادی» را یاد گرفته و هرگونه انحراف را به عنوان یک حمله بالقوه شناسایی می‌کنند.
- **شبکه‌های خودسازمان‌ده[1] (SON):** الگوریتم‌های یادگیری ماشین در قلب سیستم‌های SON برای خودبهینه‌سازی (مانند تعادل بار بین سلول‌ها، بهینه‌سازی handover) و خودترمیمی (مانند تشخیص و جبران خرابی یک eNB) قرار دارند. این سیستم‌ها تصمیمات خود را بر اساس داده‌های عملکردی که به طور مداوم از UEها و eNBها جمع‌آوری می‌شوند، اتخاذ می‌کنند.
- **مدیریت طیف فرکانسی:** در محیط‌های رادیویی پیچیده، از مدل‌های هوش مصنوعی برای اشتراک‌گذاری پویای طیف و مدیریت تداخل استفاده می‌شود تا کارایی استفاده از منابع رادیویی به حداکثر برسد.

استقرار این سیستم‌های هوشمند، یک سطح حمله جدید و حساس ایجاد می‌کند: خودِ مدل‌های

1. Self-Organizing Networks

ML و داده‌هایی که بر اساس آن‌ها آموزش دیده‌اند.

۳-۲ حملات به یکپارچگی داده و مدل (مسمومیت)

حملات مسمومیت[1] مدل ML را در مرحله آموزش با تزریق داده‌های مخرب هدف قرار می‌دهند. در اینجا، دارایی مورد حمله، یکپارچگی مجموعه داده آموزشی است.

- **مسمومیت مدل و حمله در دسترس‌پذیری:** هدف این نوع حمله، کاهش دقت کلی و تخریب عملکرد مدل است. به عنوان مثال، یک مهاجم می‌تواند با استفاده از چند UE به خطر افتاده یا یک ایستگاه پایه جعلی، گزارش‌های عملکرد شبکه جعلی (مانند کیفیت سیگنال پایین یا نرخ خطای بالا) را به یک سیستم SON ارسال کند. اگر این داده‌های مسموم وارد فرآیند آموزش مدل SON شوند، مدل همبستگی‌های نادرستی را یاد گرفته و تصمیمات بهینه‌سازی ضعیف یا حتی مخربی اتخاذ می‌کند که در نهایت منجر به کاهش کیفیت سرویس در شبکه می‌شود. در مورد یک NIDS، این حمله می‌تواند منجر به نرخ بالای هشدارهای کاذب[2] و تشخیص‌های از دست رفته[3] شده و آن را عملاً بی‌فایده کند.
- **مسمومیت درب پشتی:** این یک حمله بسیار پنهانکارانه و خطرناک است. مهاجم تعداد کمی نمونه داده با دقت طراحی‌شده را که حاوی یک «ماشه»[4] خاص هستند، به داده‌های آموزشی تزریق می‌کند. این ماشه می‌تواند یک الگوی به ظاهر بی‌خطر در ترافیک شبکه باشد. مدل NIDS یاد می‌گیرد که هر ترافیکی که حاوی این ماشه باشد، «خوش‌خیم» است، حتی اگر بخشی از یک حمله بزرگتر باشد. در نتیجه، مدل بر روی تمام داده‌های دیگر به درستی عمل می‌کند، اما مهاجم یک «درب پشتی» مخفی برای دور زدن کامل سیستم امنیتی در اختیار دارد.
- **مسمومیت هدفمند و مسمومیت با برچسب تمیز:** این‌ها نسخه‌های پیچیده‌تر مسمومیت هستند. در مسمومیت هدفمند، حمله به گونه‌ای طراحی می‌شود که مدل فقط برای یک ورودی یا کاربر خاص دچار خطا شود (مثلاً یک سیستم SON همیشه ترافیک یک شرکت رقیب را با اولویت پایین مدیریت کند). در مسمومیت با برچسب تمیز، مهاجم اختلالات بسیار کوچکی را در نمونه‌های ورودی ایجاد می‌کند تا مدل را مسموم کند، بدون اینکه این نمونه‌ها به عنوان داده پرت یا ناهنجار قابل تشخیص باشند.

تأثیر این حملات، ایجاد یک «مغز» معیوب برای شبکه است. یک SON مسموم می‌تواند به طور

1. Poisoning Attacks
2. False Positives
3. False Negatives
4. Trigger

فعال به شبکه‌ای که قرار است بهینه کند، آسیب برساند. یک NIDS دارای درب پشتی، حس امنیت کاذبی ایجاد می‌کند در حالی که یک کانال مخفی تضمین‌شده برای مهاجم فراهم می‌آورد. راهکارهای مقابله بر امنیت داده‌های آموزشی متمرکز هستند.

پاکسازی و اعتبارسنجی داده‌ها: استفاده از روش‌های تشخیص داده‌های پرت و تحلیل‌های آماری برای شناسایی و حذف نمونه‌های مشکوک قبل از آموزش. ردیابی منشأ و تاریخچه تمام داده‌های آموزشی برای شناسایی منابع غیرقابل اعتماد و از الگوریتم‌های آموزشی که در برابر داده‌های مسموم مقاومت بیشتری دارند، نمونه‌ای از راهکارهایی است که برای مقابله یا کاهش استفاده کرد.

۳-۳ حملات زمان استنتاج (گریز و نشت اطلاعات)

این حملات یک مدل کاملاً آموزش‌دیده و مستقرشده را در مرحله استنتاج هدف قرار می‌دهند. این حملات منجر به دور زدن کامل کنترل‌های امنیتی مبتنی بر یادگیری ماشین، سرقت مالکیت معنوی (خود مدل هوش مصنوعی) و نقض حریم خصوصی داده‌ها می‌شوند.

- **گریز (نمونه‌های تخاصمی):** این رایج‌ترین و شناخته‌شده‌ترین حمله تخاصمی است. مهاجم ترافیک مخرب خود را با افزودن اختلالات بسیار کوچک و هدفمند دستکاری می‌کند. این اختلالات برای انسان یا سیستم‌های مبتنی بر قانون، نامحسوس هستند، اما کافی‌اند تا ورودی را از مرز تصمیم‌گیری مدل هوش مصنوعی در NIDS عبور داده و باعث شوند که به اشتباه به عنوان ترافیک «خوش‌خیم» طبقه‌بندی شود. این حمله می‌تواند یک NIDS پیشرفته و گران‌قیمت مبتنی بر هوش مصنوعی را کاملاً کور کرده و بی‌اثر سازد. از جمله راهکارهای مقابله می‌توان به آموزش تخاصمی که در آن مدل با نمونه‌های تخاصمی آموزش داده می‌شود تا در برابر آن‌ها مقاوم شود، تقطیر دفاعی و تصادفی‌سازی ورودی اشاره کرد.
- **استخراج مدل:** در این حمله، مهاجم به عنوان یک کاربر عادی، پروب‌ها با دقت انتخاب‌شده‌ای را به مدل (مثلاً NIDS) ارسال کرده و پاسخ‌های آن (خوش‌خیم/مخرب) را مشاهده می‌کند. با تحلیل این زوج‌های ورودی-خروجی، مهاجم می‌تواند یک «مدل جایگزین» بسازد که رفتار مدل هدف را با دقت بالایی تقلید می‌کند. سپس این مدل سرقت‌شده می‌تواند به صورت آفلاین برای ساخت نمونه‌های تخاصمی کامل و کارآمد برای حمله گریز استفاده شود.
- **حملات حریم خصوصی (بازسازی داده، استنتاج عضویت و ویژگی):** این حملات حریم خصوصی داده‌های آموزشی را هدف قرار می‌دهند. با پروب کردن هوشمندانه یک مدل، مهاجم می‌تواند اطلاعات حساسی را در مورد داده‌هایی که مدل با آن‌ها آموزش دیده است، استنتاج کند. برای مثال، حمله استنتاج عضویت می‌تواند مشخص کند که آیا ترافیک یک کاربر خاص در مجموعه داده آموزشی NIDS وجود داشته است یا خیر، که خود نقض حریم

خصوصی است.

برای دو حمله فوق محدود کردن دسترسی به API مدل، کاهش جزئیات در خروجی‌های مدل (مثلاً فقط برگرداندن برچسب کلاس به جای امتیاز احتمال) و استفاده از تکنیک‌های حریم خصوصی تفاضلی می‌تواند راهگشا باشد.

۳-۴ حملات مبتنی بر مدل‌های زبانی بزرگ

این حملات، مدل‌های زبانی بزرگ (LLMs) و فناوری‌های مرتبط را هدف قرار می‌دهند که استفاده از آن‌ها در ابزارهای مدیریت و تحلیل امنیت شبکه در حال آغاز است.

- **تزریق مستقیم و غیرمستقیم درخواست:** یک مهاجم می‌تواند پرامپت‌های ورودی به یک دستیار شبکه مبتنی بر LLM (مثلاً "خلاصه‌ای از هشدارهای امنیتی اخیر را ارائه بده") را دستکاری کند تا LLM را وادار به نادیده گرفتن هشدارهای خاص، ارائه اطلاعات نادرست یا حتی اجرای دستورات غیرمجاز کند. در تزریق غیرمستقیم، پرامپت مخرب در یک منبع داده خارجی (مانند یک صفحه وب) که LLM برای پاسخ به آن مراجعه می‌کند، پنهان می‌شود.
- **حمله مبتنی بر RAG:** اگر یک اپراتور از یک سیستم تولید مبتنی بر بازیابی[1] برای عیب‌یابی استفاده کند (که اطلاعات خود را از اسناد فنی شبکه استخراج می‌کند)، یک مهاجم می‌تواند با مسموم کردن اسناد منبع، سیستم RAG را وادار به ارائه توصیه‌های نادرست یا مخرب به مهندسان شبکه کند.

این حملات می‌توانند منجر به گمراه کردن اپراتورهای انسانی، اجرای اقدامات غیرمجاز و تضعیف نسل جدید ابزارهای مدیریتی شوند. راهکارها شامل پاکسازی و اعتبارسنجی دقیق ورودی‌ها برای پرامپت‌ها، اجرای ابزارهای مبتنی بر LLM در محیط‌های ایزوله و تأمین امنیت و یکپارچگی پایگاه دانش مورد استفاده توسط سیستم‌های RAG است.

تحلیل این حملات یک نکته بسیار مهم را آشکار می‌سازد: در یک سیستم مبتنی بر هوش مصنوعی، زنجیره تأمین داده اکنون به بخشی حیاتی از محیط امنیتی تبدیل شده است. امنیت و یکپارچگی شبکه دیگر فقط به امنیت اجزای فیزیکی و نرم‌افزاری آن بستگی ندارد، بلکه به طور مستقیم به یکپارچگی داده‌هایی که برای آموزش سیستم‌های کنترل هوشمند آن استفاده می‌شود نیز وابسته است. حملات مسمومیت نشان می‌دهند که یک مهاجم می‌تواند با حمله به خودِ داده‌های آموزشی، مدل نهایی را به خطر بیندازد. این داده‌ها ممکن است از منابعی (مانند گزارش‌های UE برای سیستم SON) جمع‌آوری شوند که خارج از «محیط امن» سنتی هسته شبکه قرار دارند. این بدان معناست که مفهوم محیط امنیتی باید گسترش یابد و کل خط لوله داده، از جمله جمع‌آوری،

1. Retrieval-Augmented Generation

برچسب‌گذاری، ذخیره‌سازی و فرآیند آموزش را در بر گیرد. بنابراین، تأمین امنیت یک شبکه 4G مجهز به اجزای هوش مصنوعی، دیگر فقط یک مسئله امنیت شبکه نیست؛ بلکه یک مسئله امنیت داده و حاکمیت داده نیز هست. یک شکست در اعتبارسنجی داده (یک وظیفه امنیت داده) می‌تواند مستقیماً به یک شکست شبکه (یک حادثه امنیت شبکه) منجر شود. این یک تغییر پارادایم حیاتی برای تیم‌های امنیتی مخابراتی است.

۴ چارچوب دفاعی یکپارچه و توصیه‌های استراتژیک

این بخش پایانی، راهکارهای مقابله‌ای که در بخش‌های قبل به صورت پراکنده مورد بحث قرار گرفتند را در قالب یک چارچوب استراتژیک منسجم و چندلایه برای اپراتورهای مخابراتی ترکیب می‌کند. دفاع مؤثر در برابر تهدیدات پیچیده و چندوجهی مبتنی بر هوش مصنوعی، نیازمند یک رویکرد دفاع در عمق است که برای مقابله با این نسل جدید از حملات، تکامل یافته باشد.

۴-۱ معماری دفاع در عمق چندلایه و مقاوم در برابر هوش مصنوعی

دفاع در برابر حملات مبتنی بر هوش مصنوعی نیازمند تکامل مدل کلاسیک دفاع در عمق است. این معماری نوین باید سه لایه اصلی را در بر گیرد:

لایه ۱: سخت‌سازی زیرساخت بنیادین

این لایه شامل پیاده‌سازی کنترل‌های امنیتی اساسی برای مقاومت در برابر انواع حملات، چه سنتی و چه پیشرفته، است. این اقدامات، سطح حمله کلی را کاهش داده و هزینه و پیچیدگی را برای مهاجم افزایش می‌دهند.

- **رمزنگاری و حفاظت از یکپارچگی جامع:** الزام به استفاده از IPsec برای رمزنگاری سرتاسری ترافیک در تمام رابط‌های داخلی هسته شبکه (S1-U, S5/S8, X2, S6a و غیره). این کار آسیب‌پذیری ناشی از ترافیک رمزنگاری‌نشده داخلی را که یکی از بزرگترین نقاط ضعف 4G است، برطرف می‌کند.
- **پیکربندی امن پروتکل‌ها:** اعمال تنظیمات امنیتی سختگیرانه برای پروتکل‌های سیگنالینگ مانند Diameter و GTP، از جمله فیلترینگ دقیق ترافیک از شرکای رومینگ و استفاده از افزونه‌های امنیتی در صورت وجود.
- **کنترل دسترسی قوی:** پیاده‌سازی احراز هویت چندعاملی (MFA) برای دسترسی به تمام سیستم‌های مدیریتی، عملیاتی و نگهداری (OAM) بدون استثنا.
- **بخش‌بندی شبکه مبتنی بر اعتماد صفر:** کنار گذاشتن مدل «هسته قابل اعتماد» و پیاده‌سازی یک معماری اعتماد صفر (Zero-Trust) که در آن هیچ ارتباطی به طور پیش‌فرض مورد اعتماد نیست و تمام ترافیک بین بخش‌های مختلف شبکه (حتی در داخل EPC) بازرسی و احراز هویت می‌شود. این امر از حرکت جانبی مهاجمان جلوگیری می‌کند.

لایه ۲: هوش تهدید و تشخیص مبتنی بر هوش مصنوعی

این لایه بر اصل «مبارزه با آتش» استوار است. از آنجایی که حملات مبتنی بر هوش مصنوعی سریع، پنهانکار و تطبیقی هستند، مکانیزم‌های تشخیص نیز باید از هوش مصنوعی بهره ببرند.

تشخیص ناهنجاری مبتنی بر رفتار: استقرار سیستم‌های پیشرفته تشخیص نفوذ (NIDS) و پلتفرم‌های شکار تهدید[1] که از یادگیری ماشین برای شناسایی الگوهای رفتاری مشکوک و ناهنجاری‌ها در ترافیک شبکه، رفتار کاربران و عملکرد اجزای شبکه استفاده می‌کنند. این سیستم‌ها برخلاف NIDS های مبتنی بر امضا، قادر به شناسایی حملات روز-صفر و بدافزارهای چندریختی هستند.

- **همبسته‌سازی و تحلیل رویدادها:** استفاده از پلتفرم‌های SIEM و XDR[2] که رویدادهای امنیتی را از سراسر شبکه (از UE گرفته تا سرورهای هسته و ابر) جمع‌آوری و با استفاده از هوش مصنوعی همبسته‌سازی می‌کنند تا حملات پیچیده و چندمرحله‌ای را که در غیر این صورت جداگانه به نظر می‌رسند، شناسایی کنند.

لایه ۳: تأمین امنیت زنجیره تأمین هوش مصنوعی/یادگیری ماشین

این لایه جدیدترین و یکی از حیاتی‌ترین لایه‌های دفاعی است که به طور خاص به حفاظت از خودِ مدل‌های هوش مصنوعی در برابر حملات تخاصمی می‌پردازد.

- **مرحله پیش از استقرار:**
  - **پاکسازی و اعتبارسنجی داده‌ها:** پیاده‌سازی فرآیندهای قوی برای شناسایی و حذف داده‌های پرت، ناهنجار یا مشکوک از مجموعه داده‌های آموزشی.
  - **حفظ اصل و نسب داده:** ثبت و نگهداری سوابق کامل از منشأ، تغییرات و دسترسی‌ها به داده‌های آموزشی برای تسهیل بازرسی و شناسایی نقاط نفوذ احتمالی.
  - **آزمون نفوذ تخاصمی:** آزمون دقیق مدل‌ها قبل از استقرار در برابر انواع حملات تخاصمی شناخته‌شده (گریز، مسمومیت، استخراج) برای ارزیابی استحکام آن‌ها.
- **مرحله حین آموزش:**
  - **آموزش تخاصمی:** آموزش دادن مدل‌ها با استفاده از نمونه‌های تخاصمی تولیدشده برای افزایش مقاومت آن‌ها در برابر حملات گریز. این کار مرزهای تصمیم‌گیری مدل را مستحکم‌تر می‌کند.

1. Threat Hunting
2. Extended Detection and Response

- مرحله پس از استقرار:
  - **نظارت مستمر بر عملکرد مدل:** ردیابی مداوم معیارهای عملکردی مدل (مانند دقت) برای شناسایی افت ناگهانی که می‌تواند نشانه یک حمله مسمومیت در دسترس‌پذیری باشد.
  - **دفاع مبتنی بر پروب:** نظارت بر الگوهای پروب‌های ارسالی به مدل برای شناسایی تلاش‌ها برای استخراج مدل. این شامل محدود کردن نرخ پروب و شناسایی پروب‌های غیرعادی است.
  - **نظارت بر عدم قطعیت مدل:** تحلیل پیش‌بینی‌های مدل که با سطح اطمینان پایین انجام می‌شوند، زیرا این می‌تواند نشانه‌ای از یک ورودی تخاصمی باشد.

۴-۲ توصیه‌های راهبردی برای اپراتورهای شبکه

- **اتخاذ ذهنیت اعتماد صفر (Zero-Trust):** فرض «هسته قابل اعتماد» باید به طور کامل کنار گذاشته شود. اپراتورها باید فرض کنند که هر بخش از شبکه می‌تواند به خطر بیفتد و تأیید هویت و مجوز را در هر مرحله از ارتباطات اعمال کنند.
- **سرمایه‌گذاری در هوش مصنوعی برای دفاع:** اپراتورها باید بپذیرند که مزیت عملیاتی-زمانی مهاجمان مبتنی بر هوش مصنوعی، دفاع دستی را بی‌اثر می‌کند. سرمایه‌گذاری در پلتفرم‌های تشخیص و پاسخ خودکار (مانند SOAR) که توسط هوش مصنوعی هدایت می‌شوند، یک ضرورت است.
- **ایجاد یک تیم امنیتی هوش مصنوعی چندتخصصی:** تأمین امنیت سیستم‌های هوش مصنوعی نیازمند همکاری نزدیک بین دانشمندان داده، مهندسان شبکه و تحلیلگران امنیت است. تیم‌های مجزا و جزیره‌ای در مقابل این تهدیدات یکپارچه شکست خواهند خورد.
- **مطالبه امنیت از تأمین‌کنندگان:** هنگام تهیه تجهیزات و نرم‌افزارهای شبکه (به‌ویژه عملکردهای مجازی‌شده یا ابری)، اپراتورها باید از تأمین‌کنندگان خود شفافیت کامل در مورد مدل‌های ML مورد استفاده و اقدامات امنیتی انجام‌شده برای محافظت از آن‌ها در برابر حملات تخاصمی را مطالبه کنند.

۴-۳ توصیه‌هایی برای نهادهای استانداردسازی

- **بازنگری در امنیت پروتکل‌های هسته:** استانداردهای آینده (مانند 5G-Advanced و 6G) باید رمزنگاری و حفاظت از یکپارچگی سرتاسری را برای تمام رابط‌های شبکه مرکزی الزامی کنند و ابهاماتی که منجر به پیاده‌سازی‌های ناامن در 4G شد را برطرف نمایند.
- **استانداردسازی امنیت یادگیری ماشین:** تدوین استانداردها و بهترین شیوه‌ها برای تأمین

امنیت مدل‌های یادگیری ماشین مورد استفاده در مخابرات. این استانداردها باید جنبه‌هایی مانند مدیریت داده‌های آموزشی، آزمون استحکام تخاصمی و شفافیت مدل را پوشش دهند.

- **حل مشکل ایستگاه پایه سیار جعلی:** استانداردسازی و الزام یک مکانیزم قوی برای احراز هویت ایستگاه‌های پایه توسط UEها قبل از هرگونه اتصال. این کار یک آسیب‌پذیری دیرینه را که امکان اجرای حملات متعدد را فراهم می‌کند، برای همیشه برطرف خواهد کرد.

۵ نتیجه‌گیری

ظهور حملات مبتنی بر هوش مصنوعی، یک تحول بنیادین در چشم‌انداز تهدیدات علیه شبکه‌های 4G LTE محسوب می‌شود. این فناوری، آسیب‌پذیری‌های موجود در معماری 4G را از ریسک‌های قابل مدیریت به تهدیداتی حیاتی تبدیل می‌کند، زیرا حملاتی با سرعت، مقیاس و پنهان‌کاری بی‌سابقه را امکان‌پذیر می‌سازد. مهم‌تر از آن، هوش مصنوعی یک سطح حمله کاملاً جدید را معرفی می‌کند: خودِ مدل‌های یادگیری ماشینی که به طور فزاینده‌ای برای کنترل، بهینه‌سازی و دفاع از شبکه به کار گرفته می‌شوند.

تحلیل‌های ارائه شده در این گزارش نشان داد که حملات تهاجمی مبتنی بر هوش مصنوعی، از شناسایی هوشمند گرفته تا حملات DDoS تطبیقی و جمینگ دقیق، می‌توانند از نقاط ضعف ذاتی 4G مانند ترافیک داخلی رمزنگاری‌نشده و فقدان احراز هویت ایستگاه پایه، بهره‌برداری کنند. از سوی دیگر، حملات تخاصمی مانند مسمومیت و گریز، می‌توانند «مغز» شبکه -یعنی سیستم‌های هوشمند NIDS و SON- را مستقیماً هدف قرار داده و آن‌ها را از یک ابزار دفاعی به یک عامل مخرب تبدیل کنند.

در نهایت، تأمین امنیت نسل فعلی و آینده شبکه‌های مخابراتی در برابر این تهدیدات دوگانه، نیازمند یک استراتژی دفاعی دووجهی و یکپارچه است. اپراتورها باید از یک سو، با سخت‌سازی زیرساخت‌ها، پیاده‌سازی رمزنگاری جامع و اتخاذ معماری اعتماد صفر، بنیان‌های امنیتی شبکه خود را تقویت کنند. از سوی دیگر، باید با ایجاد یک چارچوب امنیتی قوی برای کل چرخه حیات سیستم‌های هوش مصنوعی -از پاکسازی داده‌های آموزشی گرفته تا آموزش تخاصمی و نظارت مستمر پس از استقرار- از هوش مصنوعی خود در برابر حملات محافظت نمایند. بقا و موفقیت در این چشم‌انداز جدید، در گرو پذیرش این واقعیت است که امنیت شبکه و امنیت هوش مصنوعی دیگر دو حوزه مجزا نیستند، بلکه دو روی یک سکه برای تضمین یک زیرساخت ارتباطی قابل اعتماد و امن هستند.

منابع

1. LTE Network Architecture - Tutorialspoint, accessed July 25, 2025, https://www.tutorialspoint.com/lte/lte_network_architecture.htm
2. LTE Quick Guide - Tutorials Point, accessed July 25, 2025, https://www.tutorialspoint.com/lte/lte_quick_guide.htm
3. LTE Tutorial: E-UTRAN Architecture - Artiza Networks, accessed July 25, 2025, https://www.artizanetworks.com/resources/tutorials/eut_arc.html
4. 4G Architecture Diagram:LTE Network Elements and Interfaces - RF Wireless World, accessed July 25, 2025, https://www.rfwireless-world.com/Tutorials/LTE-network-architecture.html
5. The LTE Network Architecture - rintintin.colorado.edu, accessed July 25, 2025, https://rintintin.colorado.edu/~gifford/5830-AWL/LTE_Alcatel_White_Paper.pdf
6. 4G LTE ARCHITECTURE - YouTube, accessed July 25, 2025, https://www.youtube.com/watch?v=pPJULbuC5b4
7. Cyberattacks on 4G/LTE Telecom Networks: Threat Mapping and Defense - Medium, accessed July 25, 2025, https://medium.com/@1200km/cyberattacks-on-4g-lte-telecom-networks-threat-mapping-and-defense-bd0e7fe76f54
8. 3GPP Security Architecture [12] | Download Scientific Diagram - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/figure/GPP-Security-Architecture-12_fig2_281968267
9. An overview of the 3GPP 5G security standard - Ericsson, accessed July 25, 2025, https://www.ericsson.com/en/blog/2019/7/3gpp-5g-security-overview
10. Security in 3GPP - 3G4G, accessed July 25, 2025, https://www.3g4g.co.uk/Lte/LTE_Security_Pres_1101_3GPP.pdf
11. A Survey of Security Threats on 4G Networks | Request PDF - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/publication/4314178_A_Survey_of_Security_Threats_on_4G_Networks
12. Cyber-security and performance Issues in 4G LTE network - WJAETS, accessed July 25, 2025, https://wjaets.com/sites/default/files/WJAETS-2024-0328.pdf
13. Practical Attacks Against Privacy and Availability in 4G/LTE Mobile ..., accessed July 25, 2025, https://www.ndss-symposium.org/wp-content/uploads/2017/09/practical-attacks-against-privacy-availability-4g-lte-mobile-communication-systems.pdf
14. Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems | Request PDF - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/publication/316906826_Practical_Attacks_Against_Privacy_and_Availability_in_4GLTE_Mobile_Communication_Systems
15. On the Impact of Rogue Base Stations in 4G/LTE Self Organizing Networks - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/publication/326052281_On_the_Impact_of_Rogue_Base_Stations_in_4GLTE_Self_Organizing_Networks

16. Adversarial Attacks Against Network Intrusion Detection in IoT Systems - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/publication/348090098_Adversarial_Attacks_Against_Network_Intrusion_Detection_in_IoT_Systems

17. AI-Driven Cyber Attacks and 5G Networks - IJRASET, accessed July 25, 2025, https://www.ijraset.com/best-journal/aidriven-cyber-attacks-and-5g-networks-the-new-age-of-digital-threats-

18. Adversarial Machine Learning for 5G Communications Security | Request PDF, accessed July 25, 2025, https://www.researchgate.net/publication/354540692_Adversarial_Machine_Learning_for_5G_Communications_Security

19. Adversarial Machine Learning for 5G Communications Security - arXiv, accessed July 25, 2025, http://arxiv.org/pdf/2101.02656

20. Adversarial Machine Learning for 5G Communications Security | Request PDF, accessed July 25, 2025, https://www.researchgate.net/publication/348320678_Adversarial_Machine_Learning_for_5G_Communications_Security

21. Adversarial Machine Learning for 5G Communications Security - arXiv, accessed July 25, 2025, https://arxiv.org/pdf/2101.02656

22. Resilience of LTE networks against smart jamming attacks - KAUST ..., accessed July 25, 2025, https://repository.kaust.edu.sa/bitstreams/c345dc28-fb26-4b38-b4a7-8dc849291fca/download

23. RESILIENCE OF LTE NETWORKS AGAINST SMART JAMMING ATTACKS: A GAME-THEORETIC APPROACH - GT Digital Repository, accessed July 25, 2025, https://repository.gatech.edu/server/api/core/bitstreams/6da0c492-d3e7-40ee-b86c-77587c35b7e6/content

24. Mitigating Smart Jammers in Multi-User MIMO - arXiv, accessed July 25, 2025, https://arxiv.org/pdf/2208.01453

25. Intelligent Jammer on Mobile Network LTE Technology: A Study Case in Bucharest - MDPI, accessed July 25, 2025, https://www.mdpi.com/2076-3417/13/22/12286

26. A Countermeasure against Smart Jamming Attacks on LTE Synchronization Signals - Journal of Communications, accessed July 25, 2025, https://www.jocm.us/uploadfile/2020/0707/20200707024254837.pdf

27. A Systematic Study of Adversarial Attacks Against Network Intrusion Detection Systems, accessed July 25, 2025, https://www.mdpi.com/2079-9292/13/24/5030

28. Adversarial Challenges in Network Intrusion Detection Systems: Research Insights and Future Prospects - arXiv, accessed July 25, 2025, https://arxiv.org/pdf/2409.18736

29. Adversarial Challenges in Network Intrusion Detection Systems: Research Insights and Future Prospects - arXiv, accessed July 25, 2025, https://arxiv.org/html/2409.18736v3

30. A Review of the Duality of Adversarial Learning in Network Intrusion: Attacks and Countermeasures - arXiv, accessed July 25, 2025, https://arxiv.org/html/2412.13880v1

31. LTE Network Automation under Threat - Black Hat, accessed July 25, 2025, https://i.blackhat.com/us-18/Wed-August-8/us-18-Shaik-LTE-Network-Automation-Under-Threat-wp.pdf

32. Self-Organizing Networks (SON): Enhancing Mobile Radio Access Efficiency - NYBSYS, accessed July 25, 2025, https://nybsys.com/self-organizing-networks-son/

33. What Is a Self-Organizing Network? SON Overview & Explainer - Celona, accessed July 25, 2025, https://www.celona.io/network-architecture/self-organizing-network

34. Self-Organising Networks (SON) - 3GPP, accessed July 25, 2025, https://www.3gpp.org/technologies/son

35. Poisoning Attacks Against Machine Learning - National Institute of ..., accessed July 25, 2025, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934932

36. Launching Adversarial Attacks against Network Intrusion Detection Systems for IoT - MDPI, accessed July 25, 2025, https://www.mdpi.com/2624-800X/1/2/14

37. On Attacking Future 5G Networks with Adversarial Examples: Survey - MDPI, accessed July 25, 2025, https://www.mdpi.com/2673-8732/3/1/3

38. Data Poisoning: Current Trends and Recommended Defense Strategies - Wiz, accessed July 25, 2025, https://www.wiz.io/academy/data-poisoning

39. Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach - ResearchGate, accessed July 25, 2025, https://www.researchgate.net/publication/382857536_Detecting_and_Mitigating_Data_Poisoning_Attacks_in_Machine_Learning_A_Weighted_Average_Approach

40. Detecting and Preventing Data Poisoning Attacks on AI Models - arXiv, accessed July 25, 2025, https://arxiv.org/pdf/2503.09302

41. Adversarial Attacks Against Deep Learning-based Network Intrusion Detection Systems and Defense Mechanisms - Informatics Homepages Server, accessed July 25, 2025, https://homepages.inf.ed.ac.uk/ppatras/pub/ton22.pdf

42. Security, Trust and Privacy challenges in AI-driven 6G Networks - arXiv, accessed July 25, 2025, https://arxiv.org/html/2409.10337v1

43. Adversarial Attacks Against Network Intrusion Detection Systems, accessed July 25, 2025, https://hammer.purdue.edu/articles/thesis/Adversarial_Attacks_Against_Network_Intrusion_Detection_Systems/26361145

44. What Is Data Poisoning? - CrowdStrike, accessed July 25, 2025, https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/data-poisoning/

45. Understanding Data Poisoning: How It Compromises Machine Learning Models, accessed July 25, 2025, https://securing.ai/ai-security/data-poisoning-ml/

46. Data & Model Poisoning [Exploring Threats to AI Systems] - CybelAngel, accessed July 25, 2025, https://cybelangel.com/data-model-poisoning/

47. Securing Your Network Against Attacks: Prevent, Detect, and Mitigate Cyberthreats - Kentik, accessed July 25, 2025, https://www.kentik.com/blog/secure-your-network-against-attacks-prevent-detect-mitigate-cyberthreats/

48. Cyber-Security and Performance Issues in 4G LTE Network | Request PDF - ResearchGate, accessed July 25, 2025,https://www.researchgate.net/publication/382651513_CyberSecurity_and_Performance_Issues_in_4G_LTE_Network