Analysis Report

SECURITY REPORT

2024

MAR

2024年サイバーセキュリティ脅威及び技術展望 ～前編～

CyberFortress

このレポートは、株式会社サイバーフォートレスの独自の調査にて収集されたデータに基づき作成しています。

Analysis Report.

2024年サイバーセキュリティ脅威及び技術展望 ～前編～

01. 概要

2022年2月24日ロシア大統領の特別軍事作戦開始命令によって引き起こされたウクライナとの戦争と共に、2023年10月7日パレスチナ武装組織ハマスがイスラエルの相手に大規模の侵攻を行った。国との武力衝突の長期化による世界経済の弱体化や、アメリカを含めた西側諸国とロシアとの利害衝突の深化により、覇権競争が最高潮を達することで、政治、安保、国防、外交、医療、金融など主要分野から物理的な衝突を超えて、サイバー衝突を活用することで緊張感が高まっている。戦争によるエネルギー危機は高物価による成長の動力喪失に繋がり、高いインフレーションと金融不安定による不確実性も高くなっている。
政治的・経済的目的による武力衝突の長期化による世界的な経済危機とサイバー攻撃にも悪影響を及ぼしている。国との経済的・外交的な紛争による経済制裁による外貨受給に問題が発生することで、一部の国では国主導のサイバー攻撃グループが敵国を対象に重要情報奪取やランサムウェア攻撃、仮想通貨奪取など、お金を目的とした攻撃を試みた。しかし、国際制裁強化と仮想通貨市場の価値変動性の深化で仮想通貨の現金化が難しくなって、より広範囲の分野にサイバー攻撃を行っている。

IGLOO 2024年サイバーセキュリティ脅威展望	IGLOO 2024年技術展望
1. 国家支援基盤サプライチェーン攻撃増加 #State-sponsored hacker、#Supply Chain、#3rd Party Risk、#Solution、#Vulnerability、#APT、#DPRK、#Threat Actor、#偽装就職、#SBOM	**1. インテリジェンス基盤のSOC自動化** #SOC、#SOAR、#XDR、#Managed SIEM、#MDP、#Automated Security Control、#ゼロトラスト、#Extended data、#Heterogeneous、#Prioritization、#Orchestration
2. 社会的話題と攻撃様相の変化によるサイバーセキュリティ生態系の不安 #世界的な政治話題、#フィッシング、#偏向、#LNK、#CHM、#Fileless、#Cyberwar	**2. 生成型AIが引き起こしたAIセキュリティ** #AI-driven cyberattacks、#LLM、#Generative Cybersecurity AI、#AI TriSM、#SAIF、#FMTI、#FATE
3. 二重攻撃、ランサムウェアの攻撃方法の高度化 #Multi Extortion、#Double ransomware、#RaaS、#RDP、#MS-SQL、#SSH、#言語の多様化(Go·Rust)、#Exfiltration、#IAB	**3. DXによるIT-OT融合生態系の拡散及び融合セキュリティ** #Convergence Security、#SmartX、#IT、#OT、#ICS/SCADA、#IoT、#IIoT、#自律運転、#自律船舶、#UAM、#CPS
4. サイバー攻撃のゲームチェンジャー、オープンソース生態系による危険チェーン化 #Dependency Confusion、#Typosquatting、#Privilege Leakage、#Stealing Administrator Privileges、#Maliciouse Code via Repository Package、#Risk Chaining	**4. 脅威可視化のための資産識別及び危険管理技術、攻撃表面管理** #Attack Surface、#Exposure Management、#Cybersecurity Mesh、#CASM、#EASM、#Automation Penetration Testing
5. 生成型AIを利用したサイバー攻撃増加 #ディープフェイク、#フィッシング、#スピアフィッシング、#スパム、#マルウェア、#Attacks on Generative AI itself、# Exposure of system vulnerability、# Exposure of sensitive data、#Exposure of personal information	**5. クラウドネイティブからクラウドセキュリティモデルまでクラウドセキュリティ強化** #Pervasive Cloud、#Cloud Native、#CloudOps、#Cloud Security、#CASB、#CWPP、#CSPM、#CIEM、#WAAP、#KSPM、#CNAPP、#SSE、#SSG、#ZTNA、#SSPM

【▲表① イグルーコープの2024年サイバーセキュリティ脅威及び技術展望】

2021年から50ヵ国が参加したランサムウェア攻撃行為者の対応策や攻撃に対する回復力強化のための国際協力体系であるCRI(Counter Ransomware Initiative)を開催して、サイバー攻撃による深刻性と警戒心を促進している。2023年には３年目を向かって犯罪資金目的で使用される仮想通貨財布ブラックリストの共有及びランサムウェア攻撃対応に対する協力などを含めた共同政策声明を採択して、サイバーセキュリティの防御力強化のための全世界的な対応体系を構築するために努力を続けている。

国家支援型サイバー攻撃グループのサプライチェーン(Supply Chain)攻撃は、2024年にも続けられると予想されている。2023年2月半導体会社である「Applied Materials」はサプライチェーン攻撃で2023年第1四半期だけで2億ㄑ5千万ドルの被害が予想された。2023年3月には北朝鮮のハッキンググループ「Lazarus」が「3CX DesktopApp」サプライチェーン攻撃を行い、2023年6月にはランサムウェア攻撃グループ「Cl0p」がCVE-2023-34362とCVE-2023-35036脆弱性を利用した「MOVEit Transfer」サプライチェーン攻撃で620社の組織が持っていた個人識別情報、職員住所、ID、生年月日、国民保険番号などが漏洩される事故が発生した。「Cybersecurity Ventures」によるとソフトウェアサプライチェーン攻撃で年間被害金額は2025年から600億ドルで2031年までは1,380億ドルを達すると展望している。

社会的にみると、2024年は全世界的な政治話題によるサイバー攻撃の不安が極大化されると予想する。ロシア、ウクライナ、アメリカ、フィンランド、メキシコ、ベネズエラなどの主要国は国政の基調に膨大な影響を及ぼせる大統領の選挙があるため、これを悪用したソーシャルエンジニアリング攻撃が増加すると予想される。マイクロソフトオフィス(MS Office)を利用した文書型マルウェアから主に使用されるマクロ機能も遮断ポリシーと共に、攻撃者は素早く対応し、新たな攻撃ベクターが増加しているため、セキュリティ強化ポリシーによる攻撃者の素早い対応を続けると思う。

チャットGPT(Chat GPT)の成功で、破壊的な革新を引き起こした生成型AI(Generative AI)は、人工知能(AI)によるサイバー攻撃悪用及び、敵対的攻撃による安全性問題と共に安全な人工知能の生態系造成のための多様な動きがみえる。責任あるAI原則(Responsible AI)のためのSAIF(Secure AI Framework)、FMTI(The Foundation Model Transparency Index)、FATE(Fairness, Accountability, Transparency)などのセキュリティイニシアティブ発足及び、セキュリティガイドライン配布などで人工知能と自動化による社会的な不平等と差別を解除するための透明性と責任性、プライバシーなどを確保する必要がある。

DXでITとOTを超えて、融合生態系の破壊的な成長による新たな攻撃がさらに増加して、危険チェーン化(Risk Chaining)が深化されている。経済危機と国際情勢の不安の中で、サイバー環境の危険を最少化するために、2024年セキュリティ脅威を予測し、脅威に対応できる技術と戦略でさらに安全なセキュリティ生態系造成のための方法を提示する。

02. 20245年5大サイバーセキュリティ脅威展望

1) 国家支援基盤サプライチェーン攻撃の増加

2024年サイバーセキュリティの重要なキーワードを一つ選ぶと、国家支援サイバー攻撃による「サプライチェーン(Supply Chain)攻撃の増加」と言える。サイバー攻撃の断片的な観点からは、社会的話題を悪用したメール配布やファイルアップロード脆弱性によるホームページ改ざんやマルウェア流布などが考えられるが、もっと詳細に見てみると、政治的・経済的・社会的・技術的利害関係が結合した産出物である。したがってサイバー攻撃を理解するためにはサイバー攻撃を構成する多様な要素と相関関係を考慮する必要がある。

サプライチェーンに侵入し、ユーザーに渡されるハードウェアやソフトウェアなどを改ざんする攻撃の形態を意味するサプライチェーン攻撃は次世代IT技術を合わせた融合生態系の拡散によるソフトウェアを基盤として再編されている。サプライチェーン生態系全般からソフトウェアと結合されることで、ソフトウェア生態系(SDLC, CL/CDなど)の整合性(Integrity)を阻害する攻撃行為による製品やサービスに影響を及ぼす攻撃が増加している。ソフトウェア基盤のサプライチェーンからは攻撃対象の大半が、多数のシステムを運用及び管理する目的で使用するソフトウェアをターゲットにする。「表② セキュリティ脅威1-1」のように2021年から最近３年間で発生したソフトウェア基盤のサプライチェーン攻撃をみると、国家支援サイバー攻撃グループによる攻撃で経済的・社会的側面の被害が増えていることが分かる。

区分	攻撃日付	攻撃主体	被害現況
Jump Cloud	2023.07	DPRK, UNC4899	• RGB装置が商用VPN提供社と共にL2TP IPsecトンネルを使用して一連のORB(Operational Relay Box)を活用してSource Addressを隠してExpressVPN以外にもNordVPN、TorGuardなどを使用 • ダウンストリーム顧客のJumpCloudエージェントで実行される悪性Rubyスクリプト識別(2023.06.27 15:51:57 UTC) • JumpCloudプラットホームを使用する20万個以上の組織の中、5名未満のJumpCloud顧客と総10個未満の装置に影響
MOVEit	2023.05	ロシア、Cl0p	• Progress Softwareの企業用ファイル送信プログラムMOVEit脆弱性を悪用したサプライチェーン攻撃で5月から約2,620個の組織と7,700万名が被害 • CVE-2023-34362、CVE-2023-35036(攻撃者は2021.07、2022.04脆弱性をテストしてから2023.05脆弱性を実際使用) • 2023.06MOVEit TranferからCVE-2023-36934、CVE-2023-36932、CVE-2023-36933脆弱性追加発見
3CX	2023.03	DPRK, UNC4736	• チャット、ビデオ通話、音声通話など使用者にコミュニケーションを提供するエンタープライズソフトウェアである3CX DesktopAppを利用したサプライチェーン攻撃実施 • 3CX職員がTrading TechnologiesのX_TRADERダウンロード過程でマルウェアに感染し、VEILEDSIGNALバックドアで3CXビルドサーバに侵入後、情報窃取
Okta	2022.01 2022.08 2023.09	Lapsus$外、未確認	• 認証サービス業者Oktaは2022.01Oktaサポートエンジニアシステム侵害、2022.08職人3人が騙されてスプーフィングされたCloudflare OktaログインページとBeyondTrustなどサイバー攻撃発生、2023.09から1PasswordとBeyondTrust、Cloudflareなどカスタマサポートのシステム攻撃で18,400名の顧客の中、約1%が管理システムに不法アクセスする悪意的な攻撃行為発見
Kaseya	2021.07	Revill	• MSP・IT管理ソフトウェア供給業者Kaseyaのエンドポイント及びリモートモニタリング管理システムであるVSA(Virtual System Administrator)を攻撃して顧客60名と1,500個企業がサイバー攻撃に影響
Colonial Pipeline	2021.05	DarkSide	• ランサムウェア感染でアメリカ東南部地域のガソリン供給が一時中断されて500万ドルの身代金を払い、ジョー・バイデン行政部から「サイバーセキュリティ強化命令」発表

【▲「表 セキュリティ脅威1-1」 最近3年間発生したサプライチェーン攻撃被害現況(2021-2023)】

国家支援サイバー攻撃によるサプライチェーン攻撃の影響要素を「表 セキュリティ脅威1-2」のように、環境的要因と技術的要因、金銭的目的と政治的目的で分類できる。まず、サプライチェーン攻撃を誘発する環境的要因と技術的要因を確認すると、環境的要因からは効率性と生産性の向上のために効率的なサービス単位のモジュール化で、相互連携ができるAPI基盤のマイクロサービスアーキテクチャ(MSA)を追い求めることで、ソフトウェア開発生態系が変化された。このようなソフトウェア生態系の変化は人工知能(AI)、クラウド、モノのインターネット(IoT)、ブロックチェーン、モバイル、5G/6Gなど次世代技術の発展がビジネス生態系にも影響を及ぼし、生態系全般の連携性が強化され、サプライチェーンの境界が曖昧になり始めた。

技術の発展と環境の変化は、攻撃者にも影響を及ぼした。攻撃者も攻撃対象の被害を引き起こすためのマルウェア検知技術が発展することで、これをバイパスするためのファイルレス(Fileless)攻撃に深化されていて、攻撃ツールもCobalt Strike、Metasploit、Silver、Brute Ratel、Manjusaka、Alchimistなどレッドチーム(RedTeam)商用ツールとオープンソースソフトウェア(OSS)及びLotL(Living Off the Land)ツールを使用する。サイバー攻撃の正確性向上及びセキュリティアーキテクチャバイパスなど既存サイバーセキュリティの体系を無力化するための人工知能活用及び実証事例が増加することで、生成型AI(Generative AI)を利用した攻撃とマルウェア生成及び悪性メールのための攻撃シナリオ構成などが問題になっている。

区分	詳細区分	攻撃原因	被害現況
攻撃の外部要因による要因	環境的要因	ソフトウェア生態系の変化	• 効率性と生産性を向上する目的で機能別単独構成及び独立的なサービス単位の相互連携ができるAPI基盤MSA追求 • CI/CDとDevOps環境を利用した開発生産性向上及び柔軟性拡大によるオープンソース増大で3rd Party、API、Packageなど増加
		技術成熟度によるビジネス生態系の変化	• AI、Cloud、IoT、Blockchain、Mobile、5G/6Gなどの技術発達で位置・時間・君かんの従属性のないビジネス環境構成可能 • 産業間の境界曖昧(Blur)でインフラ及びデータ連携強化によるプラットホームビジネスモデルの強勢
	技術的要因	攻撃ツールの両極化	• セキュリティソリューション検知バイパス及びFilelessなど攻撃効率化のための商用ツール及びOSS/LotL(Living Off the Land)活用増加 • RedTeam·Simulation : Cobalt Strike 、 Metasploit 、 Silver 、 Brute Ratel、Manjusaka、Alchimistなど • LOLBins：Win CMD、Powershell、WMI/WMIC、Rundll32、Schtasksなど
		生成型AIを利用する攻撃の普遍化	• ChatGPT(OpenAI) 、 Bard(Google) 、 Emie Bot(Baidu) 、 Jasper(Jasper) 、 Sparrow(Deepmind)などText based生成型AIを活用して脆弱性発見、攻撃ツールの作成などを実施
攻撃の目的性によるタイプ	金銭的目的	被害者の側面	• 直接的な金銭損失：ランサムウェアなどサプライチェーン攻撃による身代金のお支払い • 間接的な金銭損失：情報漏洩及び被害による訴訟費用、運営停止による生産性減少、顧客の信頼及び忠誠度下落による企業イメージ毀損
		攻撃者の側面	• 仮想通貨及び仮想通貨取引所、ランサムウェア、奪取したデータの取引などで犯罪資金調達
	政治的目的	社会混乱引き起こし及び情報奪取	• 政治的利害関係による社会基盤施設の破壊及び外交・安保・国防などを狙ったスピアフィッシング攻撃

【▲「表 セキュリティ脅威1-2」国家支援サイバー攻撃グループのサプライチェーン攻撃目的及び要因】

国家支援サイバー攻撃によるサプライチェーン攻撃の一番主な原因は、攻撃者の攻撃目的である金銭的な目的と政治的な目的を全て獲得することができるためである。国家支援サイバーセキュリティ攻撃の代表的な国の一つである北朝鮮のサイバー攻撃の現状に対する内容を分析した韓国の国家安保戦略研究院の「進化する北朝鮮のサイバー攻撃現況と対応」によると、2022年から2023年にも北朝鮮は仮想通貨の攻撃を試みたが、価格の変動性と現金化の問題の意外にも主要国の監視と制裁の強化などで奪取金額が減少することで新たな攻撃パラダイムを試みていると発表した。「表 セキュリティ脅威1-3」のように犯罪資金を獲得するための制約事項を解消するために、攻撃対象を変更したり奪取した仮想通貨を洗浄するためのチャンネル確保、攻撃グループ間の情報共有で協業体系を強化することで、新たな突破口を探すための変化の動きが見始めた。

北朝鮮サイバー攻撃グループの危険要素	北朝鮮サイバー攻撃グループの対応現況
仮想通貨の高い価格変動	• 犯罪資金獲得という共通的な目標達成のために仮想通貨及び仮想通貨取引所だけではなく金融圏の攻撃実施
仮想通貨の現金化問題	• 仮想通貨取引所の直接的な攻撃から旋回し、ミキサーやクロスチェーンスワップ技術などの複雑な多段階資金洗浄過程を実施 • 北朝鮮ハッキンググループが奪取した仮想通貨の現金化問題を解決するためにロシア両替サービス活用 • APT43はクラウドシステムのコンピューティングパワーを使用してコインをマイニングするクラウドマイニング(Cloud Mining)で仮想通貨洗浄
主要国の監視と制裁強化	• APT38、Andariel、TEMP.Hermitなど北朝鮮のサイバー攻撃グループがハッキングツール及び技術、マルウェアなどを共有して協業体系の強化

【▲「表 セキュリティ脅威1-3」北朝鮮サイバー攻撃グループ観点の危険要素別対応現況 (参考：韓国国家安保戦略研究院、化する北朝鮮のサイバー攻撃現況と対応、一部再構成)】

「ゼロデー脆弱性を利用したサプライチェーン攻撃の流れ」

１．攻撃者はメディアのホームページをハッキングして特定の記事に悪性スクリプトを挿入し、水飲み場型攻撃を構築する。悪性スクリプトは特定のIP範囲がアクセスする場合に動作するようになっている。

２．被害者はセキュリティ認証ソフトウェアがインストールされているインターネットPCで記事を閲覧すると、ソフトウェアに内在されている脆弱性によって悪性スクリプトが動作する。被害者がPCがC2サーバにアクセスするとマルウェアがインストールされ、攻撃者はC2サーバで被害者PCのリモート制御権限を獲得する。

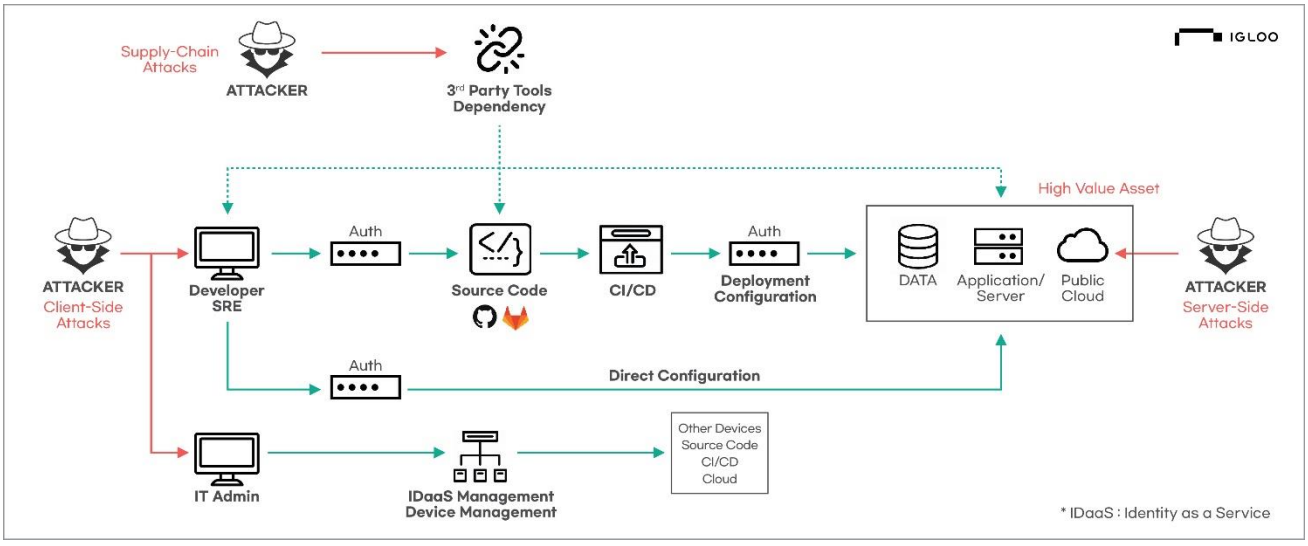
３．攻撃者はネットワーク連携製品の脆弱性を利用して、インターネットPCから外部サーバに権限無しでアクセスでき、ネットワーク連携製品のデータ同期化の機能を悪用して内部サーバにマルウェアを伝播する。攻撃者は最終的な目的である業務PCに情報奪取用のマルウェアを感染させる。

４．業務PCにインストールされたマルウェアは2つのC2サーバを持っている。1次C2サーバはネットワーク連携製品内部のサーバであり、中央でゲートウェイの役割を担当し、2次C2サーバは実際外部インターネットに位置するサーバである。このマルウェアは初期感染信号を送信し、暗号化された追加ペイロードをダウンロードして実行する機能がある。マルウェアは初期感染信号をC2サーバに送信するためのネットワーク連携製品の内部サーバから外部サーバに移動を試みるが、確認された被害事例の場合、製品のセキュリティポリシーによって遮断されて内部の情報を漏洩されていなかった。

国家支援サイバー攻撃のサプライチェーン攻撃による影響度は、単純脆弱性の攻撃より攻撃の目的及び攻撃方法などから一般的な攻撃を超える場合があるため、攻撃の影響度が高い。「図 セキュリティ脅威1-2」はソフトウェア基盤のサプライチェーン攻撃が発生する支点の攻撃構成図を整理した内容である。ソフトウェアサプライチェーン攻撃は国家支援サイバー攻撃グループの好みである攻撃ベクターで攻撃支点が多様であるため、多数の攻撃対象をターゲットとして攻撃が実施できる特徴がある。

攻撃者はサプライチェーンを構成するサードパーティツール(3rd Party Tools)の従属性(Dependency)を利用して、API、レポジトリ(Repository)などを攻撃し、ソフトウェアサプライチェーンを攻撃したり、ソフトウェアのサプライチェーンの一番最初であるクライアント側(Client-Side)を攻撃したり、サプライチェーンの一番最後であるサーバ側(Server-Side)を攻撃できるようになる。攻撃の支点によって攻撃方法や技術は異なるが、一般的に不適切な認証と認可(Authentication & Authorization)、間違ったセキュリティ設定(Misconfiguration)、整合性阻害などで攻撃が発生する。

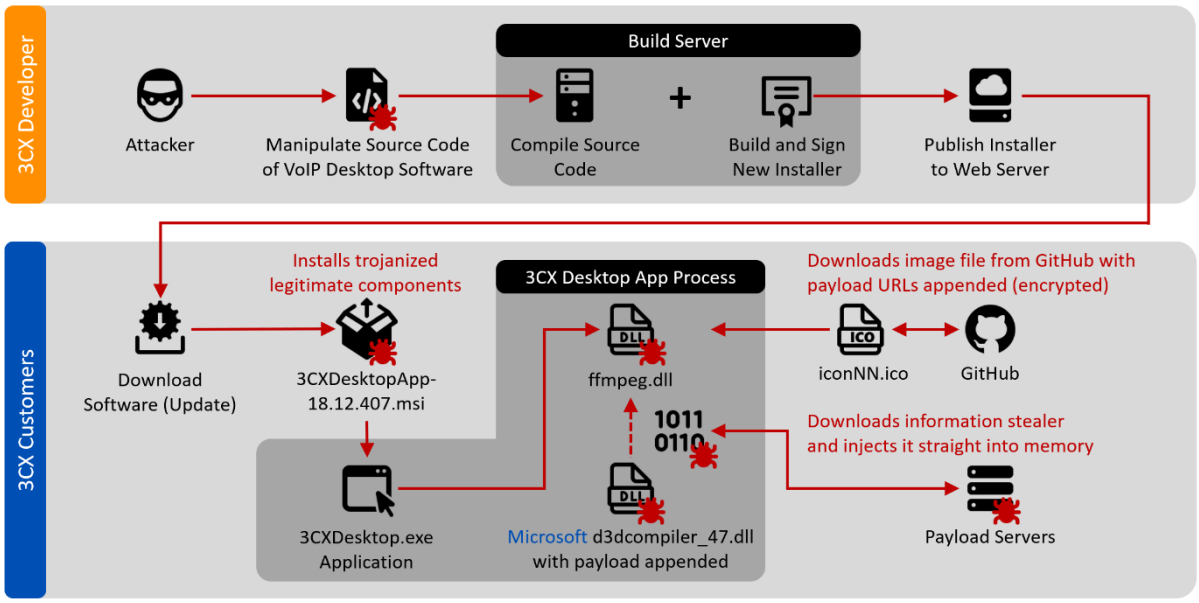
先ほどのサプライチェーン攻撃の事例のように特定エリアを攻撃するのではなく、ソフトウェアサプライチェーンを構成している要素なら、攻撃ベクターで悪用でき、その悪用されたベクターによって最終的に多数の被害者が発生する。従って、「図 セキュリティ脅威 1-2」から注目すべき点は、ソフトウェアサプライチェーンの全体流れから攻撃者の攻撃による被害が拡散できる範囲を最少化させる方法を考慮する必要がある。



【▲「図 セキュリティ脅威 1-2」ソフトウェアサプライチェーンの攻撃発生支点による攻撃構成図 (参考：Hiroki SUEZAWA, Dangerous attack paths: Modern Development Environment Security - Devices and CI/CD pipelines、一部再構成)

2023年国家支援サプライチェーン攻撃事例の中で、代表的な事例である3CXジャンプクラウド(JumpCloud)事例で国家支援サイバー攻撃によるサプライチェーン攻撃の影響度について調べてみよう。

資料によると、北朝鮮のハッカー組織は航空宇宙・医療など60万の企業・機関が使用するビデオ通話のソフトウェアである3CX「デスクトップアプリ(DesktopApp)」をターゲットにして、サプライチェーン攻撃を試み、ハッカーは3CX開発の過程に侵入して、インストールプログラムにマルウェアを隠し、3CXの公式ウェブサイトから大勢の顧客のPCを感染させた。被害システムにインストールされたマルウェアは、最短7日が過ぎた時点で起動され、3CXアカウント情報を含めてChrome及びEdgeなどのインターネットブラウザーの情報を奪取する被害が発生した。



【▲「図 セキュリティ脅威 1-3」3CXサプライチェーン攻撃の構成図 (参考： Sohpos, Update 2: 3CX users under DLL-sideloading attack: What you need to know) 】

3CXサプライチェーン攻撃は、3CXが配布するDesktopAppソフトウェアが損傷され、WindowsとMac OSすべてに影響を及ぼすマルウェアが配布されたインシデントとみられる。WindowsのOSから動作するマルウェアの基準で攻撃の流れを分析すると、まず攻撃者は3CXソフトウェアのために証明されたインストールプラグラム実行ファイル内部にマルウェアを追加する。合法的なルートでインストールプログラムを配布する際に、攻撃者は3CXネットワークに侵入して3CXソフトウェアビルドプロセスを操作できるようになる。

この後、7日の待機期間が経ったら、マルウェアは暗号化されたペイロードをロードし、ペイロードには3CXソフトウェアに含まれたDLLに追加される。ペイロードはGitHubから攻撃者が追加攻撃のためのC2リストをダウンロードし、次にステップでダウンロードするためのドメインのうち、1つに繋ぎ、攻撃対象の3CXアカウント情報を含めてブラウザー履歴を抽出してフィルタリングを行うことが確認できた。Mac OSの場合はマルウェアの待機時間が一部違うだけで全般的攻撃プロセスは類似だった。

次のサプライチェーン攻撃の事例は、2023年6月と7月に発生したアメリカのゼロトラストディレクトリプラットフォームサービスであるJumpCloudをターゲットにしたスピアフィッシング基盤のサプライチェーン攻撃事例である。RGB装置が商用VPNプロバイダーと共に、L2TP IPSecトンネルを利用して一連のORB(Operational Relay Box)を活用して送信元アドレスを隠すことが確認でき、ExpressVPNの以外にもNordVPN、TorGuardなどを使用した。またダウンストリーム顧客(ソフトウェアソリューションえエンティティ)JumpCloudエージェントで実行される悪性ルビー(Ruby)スクリプトを識別して、JumpCloudプラットフォームを使用する200,000個以上の組織の中、5名未満のJumpCloud顧客と計10個未満の装置がサプライチェーン攻撃による影響を受けた。JumpCloudサプライチェーン攻撃を試みた攻撃者であるDPRKのUNC4899は、北朝鮮偵察総局のRGP内の仮想通貨攻撃グループである。2023年7月アメリカ、韓国、香港、シンガポールにある複数のフィンテック及び仮想通貨会社の多様なC-SUITE経営者を標的に活用し、2023年7月仮想通貨分野に興味をみせた別途のRGB偵察グループが、APT43と似たような活動を実施した。JumpCloudサプライチェーン攻撃からUNC4899は、3CX開発者であるTrading Technologiesのウェブサイトを損傷したApplejeus⇒作戦の裏であるLazarus Groupと関連されたサプライチェーン攻撃を実施した。

ロシア・ウクライナ戦争、イスラエル・ハマス戦争など、無力衝突の長期化による第3戦線の勃発有無が重要な話題になり、国家間の衝突は物理的な衝突を超えて、サイバー衝突の激化を暗示している。2023年を含めて最近、数年間発生した国家主導のサプライチェーン攻撃被害事例を基に、2024年には国家間の利害関係を持続してモニタリングしつつ、サプライチェーン攻撃に対応するための多様な方法の模索が必要である。

2) 社会的話題と攻撃様相変化による催場セキュリティ生態系の不安

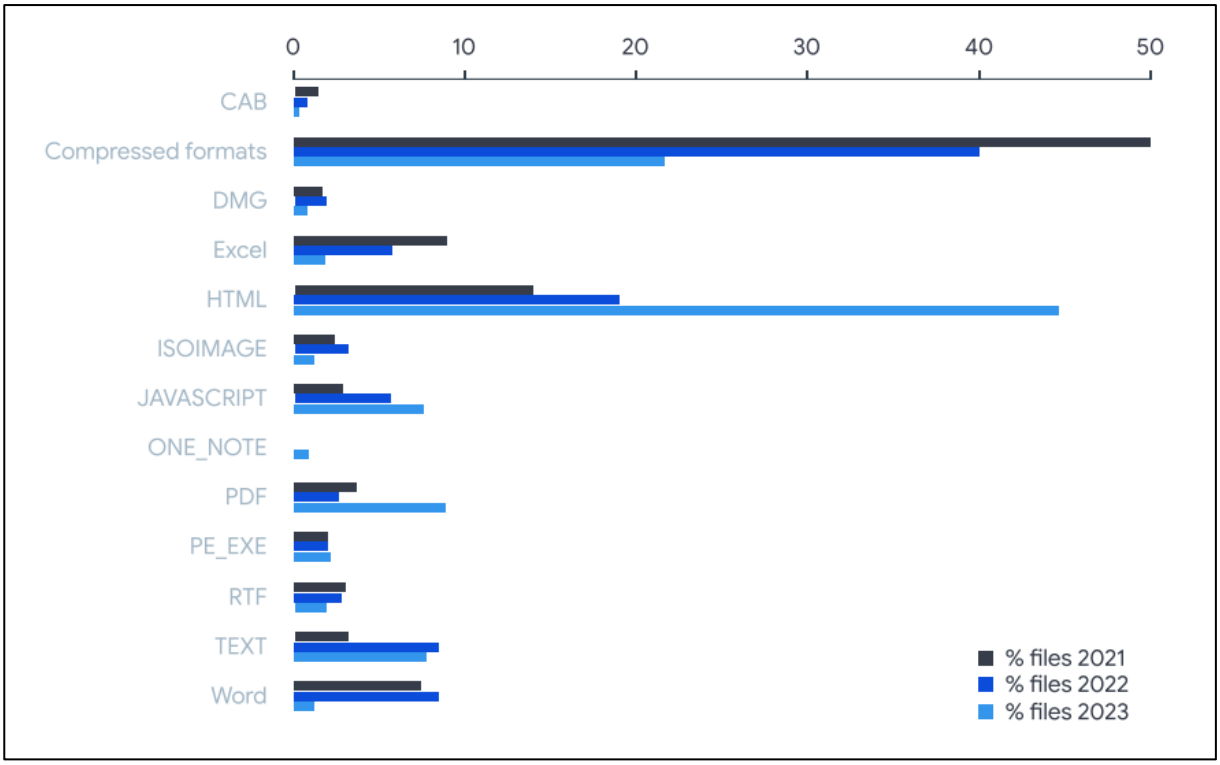
サイバー攻撃者が攻撃対象を選定後、攻撃対象に初めて接するための方法として、主に使用されるのは断然フィッシング(Phishing)攻撃である。 受信者が興味を持つ社会的な話題を利用したソーシャルエンジニアリング方法を利用して、マルウェアや不正リンクが含まれているメールを送信したり、ホームページを利用した水飲み場型攻撃、特殊目的で招待されたグループチャットルームなどでマルウェアが流布される。

APT37やKimsukyなど北朝鮮発サイバー攻撃グループはフィッシングやスピアフィッシングなどメールを利用した攻撃方法でユーザーのマルウェア実行を誘導し、マイクロソフトオフィス(MS Office)内のVBAマクロやExcel 4.0 Macro(XLM) Sheetのマクロ機能を使用した攻撃を行った。文書型マルウェアは「表 セキュリティ脅威 2-1」のようにVBA、Macro、DDE、OLEなどを使用し、 今までのマルウェア攻撃のほとんどがマクロを利用して攻撃を実施した。

NO	攻撃ベクター		攻撃方法及び攻撃事例
1	Macro (VBA/ Macro4.0)	VBA	• VBA(Visual Basic for Applications)で「vbaProject.bin」オブジェクトに攻撃コードが含まれれ実行 • 「DocumentOpen」、「Auto_Open」などの自動実行関数を利用して文書の閲覧時、マルウェアが自動でインストールされて攻撃コード実行 • 攻撃事例：北朝鮮攻撃グループAPT37、kimsukyなどを含めてEmotet、Lokibit、Revilなど
2		Macro 4.0	• 1992年に追加された「Excel 4.0 Macro Sheet」機能を利用(Only Excel) • ファイルバイナリ特徴、コード難読化などを理由でアンチウィルス製品の検知がVBA方法によってだいぶ難しいところを悪用 • 攻撃事例：TrickbotとDanabot、Gozi、Zloaderなど
3	DDE (Dynamic Data Exchange)		• Windowsお用プログラム間、同一データを共有するように許可する機能で「DDE」または「DDEAUTO」で機能を呼び出してOfficeではなくcmd、powershellなどをロードして攻撃
4	オブジェクト連携追加 (OLE)		• MS内で開発した技術で文書の外部オブジェクトに対する連携と追加に使用される連携プロトコルであるOLEを利用して「oleObject*.bin」に挿入 • 資料内に連携されている本文をユーザーがクリックする場合、不正行為が実施されて攻撃コードでexe、scr、com、pif、jar、vbs、vbe、js、jse、lnk、swf、rar、7z、bat、cmdなどの実行を誘導

【▲「表 セキュリティ脅威 2-1」 文書型マルウェア攻撃ベクター及び攻撃方法】

しかし、Microsoftが発表した2件のセキュリティ強化ポリシーによって、マルウェアを利用した攻撃方法は変化を迎えた。2021年3月Excel 4.0 MacroであるXLMの遮断と2022年4月VBAを遮断することで、「信頼できないマクロの基本遮断ポリシー」を発表して、攻撃者は新たな攻撃ベクターを活用している。「図 セキュリティ脅威 2-1」のようにVirusTotalの分析内容によると、セキュリティ強化による影響はマクロ機能を遮断した。2021年からExcel、Word基盤のマルウェア流入が目立つほど減少し、同一の期間LokiBot、AgentTesla、ChromeLoaderなどのマルウェア流布のための攻撃ベクターでISOファイルの使用量が増加したことが確認できた。



【▲「図 セキュリティ脅威 2-1」2021 distribution changes for file types used by malicious samples (参考：VirusTotal, VirusTotal Malware Trends Report: Emerging formats and delivery techniques, Jul. 2023)】

攻撃者はメールの添付ファイル及びインターネットからダウンロードしたファイルであるMOTW(Mark of the Web)内にマクロを入れた場合、遮断されるため、Microsoftマクロ基本遮断ポリシーを発表後、MOTWバイパス方法で「表 セキュリティ脅威2-2」のような悪性メールに添付されるファイルをZIP, RARなどの形式で圧縮し、圧縮ファイル内に既存のようなマクロではなく、ISO、ショートカット(LINK)、コンパイルされたHTMLヘルプファイル(CHM)を攻撃ベクターで使用し始め、一部はHTMLファイル内部に圧縮ファイルをエンベデッドした。

NO	攻撃ベクター	攻撃方法及び攻撃事例
1	ISO	• CDまたはDVDイメージふぃあるで実行時、CDまたはDVDドライブにマッピング • 内蔵されているファイルを実行時、ファイルの種類によってマルウェア、スクリプトなど実行可能
2	LNK	• Windowsショートカット(シェールリンクバイナリ)形式のファイル • ファイル作成者が入力したパラメータ(パス、コマンドなど)を利用してシステムの全てのファイル実行可能(例：PowerShell、VBScript、MSHTAなど)
3	CHM	• コンパイルされたHTMLファイル(ヘルプ) • CHM実行時、使用されるhh.exeを悪用してHTMLファイル内部のJavaScriptなどスクリプトコード実行可能

【▲「表 セキュリティ脅威 2-2」文書型マルウェア攻撃ベクター及び攻撃方法】

攻撃者は攻撃ベクターの変化と同時に、2024年は「世界政治の年」と言えるほど、国の政治的な方向とともに、世界経済とパラダイムに影響を及ぼす大統領の選挙が多数予定されている。「表 セキュリティ脅威 2-3」のように韓国、ロシア、ウクライナ、アメリカなどが選挙があるため、これを利用したソーシャルエンジニアリング方法に注目する必要がある。

情報奪取型マルウェア(Emotet、IcedID、Qakbot)から北朝鮮発攻撃グループ(Kimsuky、Scarcruft、Konni)まで多様な攻撃グループが文書型マルウェアを悪用していると確認されている。ベクターごとマルウェア分析を利用した攻撃トレンドの変化について理解と対応戦略の樹立が必要である。

時期	国	政治話題
1月	台湾	第16代西部総統・立法委員選挙
2月	インドネシア	大統領・国民協議会選挙
3月	ロシア	大統領選挙
	ウクライナ	大統領選挙
4月	韓国	第22代国会委員選挙
	イギリス	総選挙
5月	イギリス	地方選挙
6月	欧州連合(EU)	議会委員選挙
11月	アメリカ	大統領選挙

【▲「表 セキュリティ脅威 2-3」2024年全世界政治的話題の現況】

3） 二重脅迫型攻撃、ランサムウェア攻撃方法の高度化

2023年に専門化された組織と知能化された攻撃技術を基にサービスがたランサムウェア(RaaS, Ransomware as a Service)の市場は活発に運用され、2023年上半期Lockbit、Blackcat、Clopランサムウェアが上位を占めた。

Lockbitランサムウェアは、イギリス最大郵便配達サービスであるRoyal Mailを始め、世界最大半導体ファウンドリー企業であるTSMCのハードウェア供給業者のうち、一つを攻撃して身代金を要求した。

Blackcatも医療ソリューション提供業者であるNextGen Healthcareを攻撃した。またオンラインコミュニティRedditを攻撃したと公言し、Reddit内部の文書及びソースコードなどを含めた80GBぐらいのデータを口実に450ドルを払えと脅迫した。

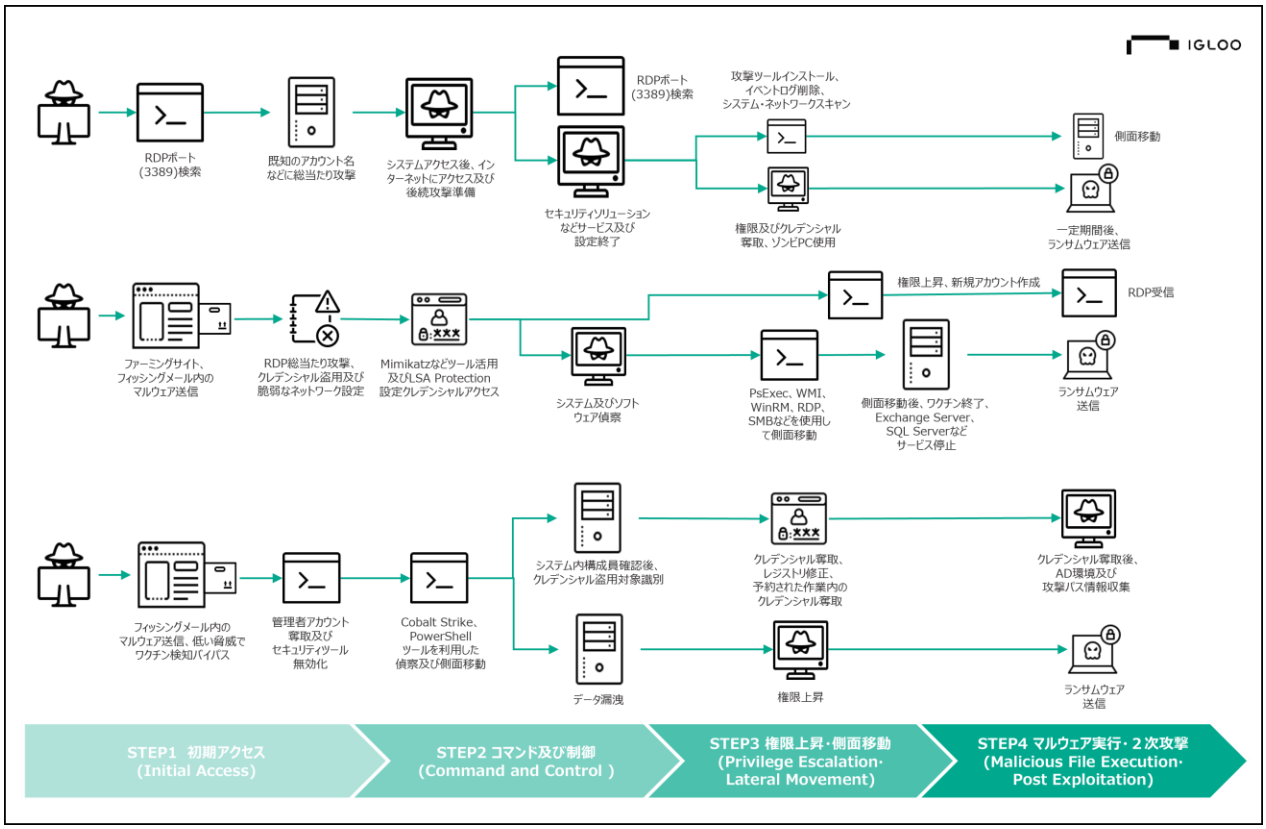
ClopランサムウェアもPaperCutサーバにリモートコード実行脆弱性であるCVE-2023-27350とPaperCutのユーザー情報が検索できるCVE-2023-27351脆弱性を利用して攻撃した。2023年に話題になった大規模サプライチェーン攻撃であるMOVEit TransferとMOVEit Cloud脆弱性であるCVE-2023-34362及びCVE-2023-35036を利用した攻撃を行った。

ランサムウェア攻撃グループは攻撃事例を利用して被害の大体は大規模組織を攻撃し、体系化された組織運用と被害額交渉能力で金融、運送、IT、ヘルスケア、分野で注目を浴びた。不法データの取り引き、資金洗浄、クレデンシャル取り引き、攻撃ツールの取引などサイバー犯罪に必要な全ての要素を確保したランサムウェアグループを最近脆弱性を悪用した攻撃事例が増加することで、ランサムウェア攻撃グループがファイル暗号化以外にも多様なサイバー攻撃で攻撃範囲を拡大することができることが分かる。

「図 セキュリティ脅威 3-1」のような2023年に発見されたランサムウェアの攻撃の特徴を見てみると、初期アクセス(Inital Access)の段階では、外部に公開されたリモートデスクトップアクセスやSSH、FTP、MS-SQLなどでアクセスしたり、一般的なフィッシングメールやファーミングサイの送信で攻撃を実施する。その後、コマンド及び制御(Command and Control)流れで被害システムにアクセスして追加的な攻撃のためのマルウェアインストール及び権限奪取などのツールを準備する。

権限上昇、側面移動(Privilege Escalation･Lateral Movement)の段階では、奪取した権限より高かったり、アクセス権限が多いアカウントを奪取するためのMimikatazやWindows権限上昇ツールであるHot Potato、Rotten Potato、Lonely Potato、Juicy Potato、Rogue Potato、Sweet Potato、Generic Potato、Local Potatoなどジャガイモ家族シリーズを活用して攻撃を試みる。

最後に、マルウェア実行及び2次攻撃(Malicious File Execution-Post Exploitiation)の段階では、ランサムウェアを実行したり、被害事実を認識できなくしたりするなど、侵害の証拠を発見させる目的で一定期間が過ぎた後にランサムウェア攻撃を試みる。

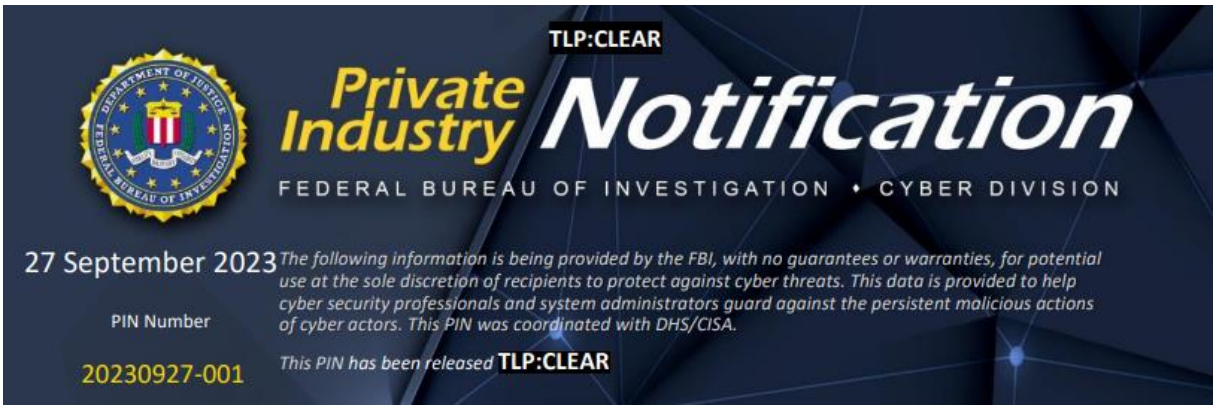


【▲「図 セキュリティ脅威 3-1」 攻撃手順別、ランサムウェア攻撃構成図】

「図 セキュリティ脅威 3-3」のようなランサムウェア攻撃のレベルは年々進化している。暗号化という単一脅迫方法から離れて、ランサムウェアを感染させ、内部情報を奪取して、ランサムウェア感染及び奪取情報を利用した脅迫という、多重脅迫(Multi Extortion)方法に変化した。2023年上半期に一番被害を起こしたランサムウェアであるLockbit、Blackcat、Clopランサムウェアも、このように多重脅迫方法を利用した攻撃方法を使用している。

2023年9月アメリカ連邦捜査局(FBI)から発表したランサムウェア環境から表す変化を整理した「二つ以上のランサムウェア変種が同一影響を及ぼす場合、被害者及びデータの破壊トレンド(Two or More Ransomware Variants Impacting the Same Victims and Data Destruction Trends)」によると、「図 セキュリティ脅威3-2」のように新しデータ破壊戦術(Data destruction tractics)と同じく被害者を連続で攻撃する二重ランサムウェア攻撃(Double back-to-back ransomware)を使用したことが確認できた。

フィッシング、アカウント奪取、総当たり攻撃またはその他の方法で被害者のシステムに侵入後、被害者の全てのデータとファイルを暗号化するランサムウェアコードを実行した後、再びランサムウェアが感染したシステムにまた新たなランサムウェアを感染させ、復号化を口実に攻撃を実施する。このような攻撃方法はAvosLocker、Diamond、Hive、Karakurt、Lockbit、Quantum、Royalなどのランサムウェアの特徴である。

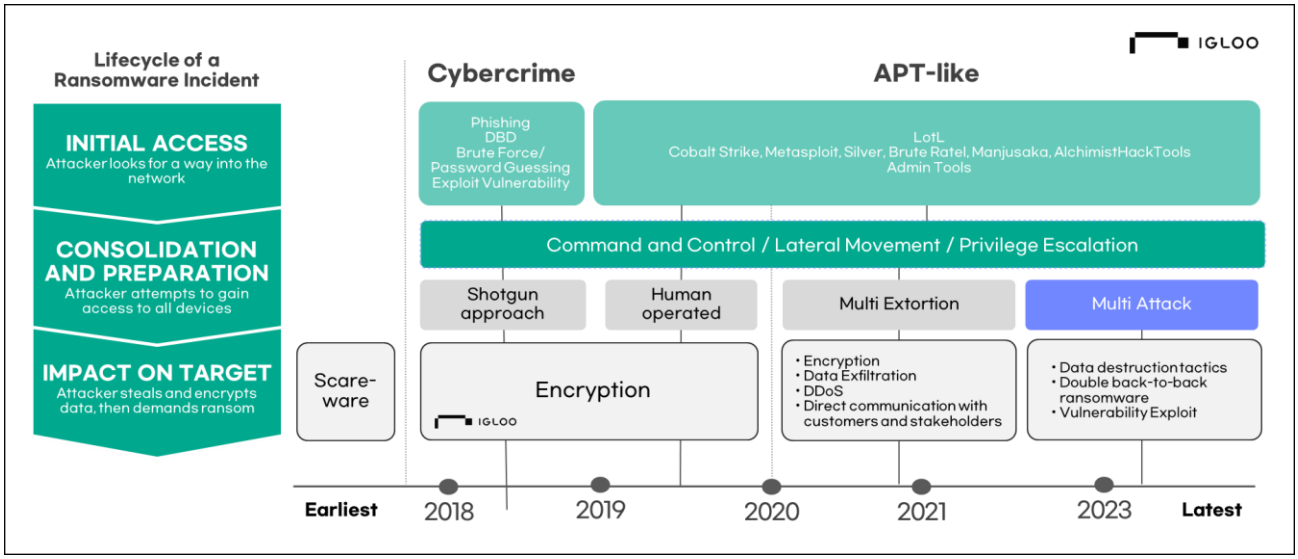


- The FBI noted a trend of dual ransomware attacks conducted in close proximity to one another.[1] During these attacks, cyber threat actors deployed two different ransomware variants against victim companies from the following variants: AvosLocker, Diamond, Hive, Karakurt, LockBit, Quantum, and Royal. Variants were deployed in various combinations. This use of dual ransomware variants resulted in a combination of data encryption, exfiltration, and financial losses from ransom payments. Second ransomware attacks against an already compromised system could significantly harm victim entities.

- In early 2022, multiple ransomware groups increased use of custom data theft, wiper tools, and malware to pressure victims to negotiate. In some cases, new code was added to known data theft tools to prevent detection. In other cases in 2022, malware containing data wipers remained dormant until a set time, then executed to corrupt data in alternating intervals.

【▲「図 セキュリティ脅威 3-2」ランサムウェア攻撃グループの環境変化分析内容の一部 (参考：FBI)】

2023年のランサムウェア攻撃のトレンドについてまとめると、「図 セキュリティ脅威 3-3」のように、既存のAPT型の攻撃流れを維持する一方、今までのランサムウェア攻撃の形態と共に、脆弱性を悪用したり、新たな脆弱性を発見する方法でサイバー攻撃の技術的成熟度と組織運用の熟練度が上がった。2024年にはランサムウェアによる被害を最小化するためには、バックアップシステムの構築及び周期的なセキュリティチェックによる脆弱性管理が重要である。



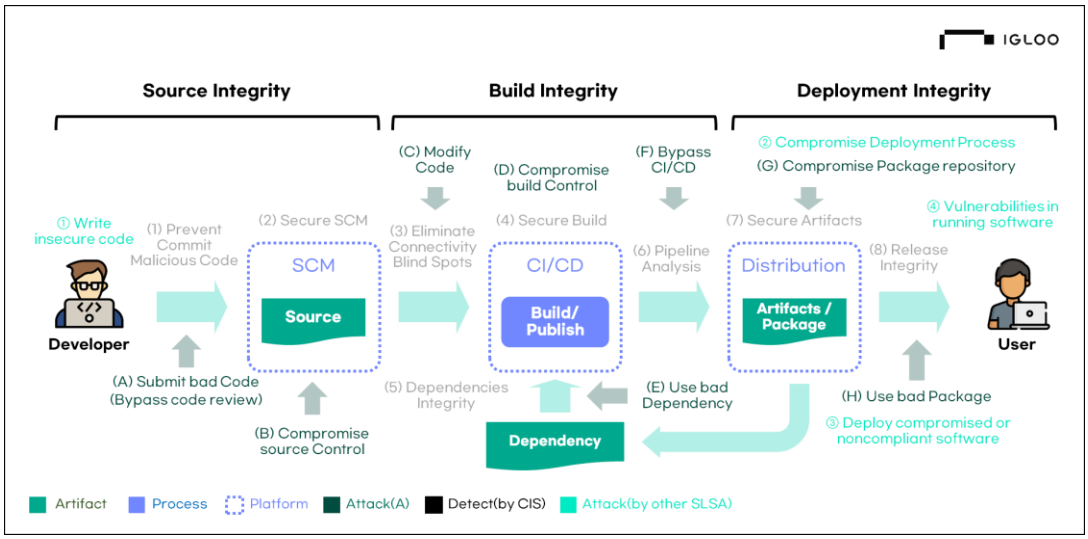
【▲「図 セキュリティ脅威 3-3」時間の流れによるさんサムウェアの発展現況】

4） サイバー攻撃のゲームチェンジャー、オープンソース生態系による危険チェーン化

DX及びビックテック企業中心のシード技術拡散によるロックイン(Lock-In)効果で、オープンソース生態系の自生力が確保され、オープンソースはソフトウェア生態系全般に影響を及ぼし始めた。ソフトウェア産業のゲームチェンジャー(Game Changer)で開発者コミュニティ中心に運用された今までの方法は、財団とビックテック企業主導に変化され、運用環境の組織化によるオープンソース生態系を有効化する原動力を確保し、ICT事業内のオープンソース市場の成長を導いた。特にビックテック企業が主導する人工知能、マシンラーニング、ブロックチェーン、メタバースなど次世代新技術は、後発ソフトウェア開発環境のシード技術として活用され、ソフトウェア生態系の高成長触媒として活用された。

オープンソースの便利さと破壊的な効果を体験している間、CVE-2021-44228とApache Log4Shell脆弱性があらわれた。その後、Spring4Shell、Text4Shellに繋がるオープンソース脆弱性によるセキュリティ問題で、オープンソースを含めたソフトウェアの依存性(Dependency)が、ソフトウェア生態系全般に及ぼす影響を認識し始めた。攻撃者もLog4Shell脆弱性発表後、オープンソースを活用した多様な攻撃を試みた。オープンソースそのものの脆弱性だけではなく、オープンソースレポジトリであるPyPI、GitHubなどを悪用した攻撃が増加し、DockerやContainerで配布されるオープンソースのセキュリティ検証環境の不在によるセキュリティインシデントが発生している。

起動動作を識別できるソースコードとは違って、サードパーティパッケージ(3rd Party Package)やライブラリ、APIなどで使用されるソフトウェアが悪用される場合、問題はさらに深刻になる。「図 セキュリティ脅威 4-1」のように、ソフトウェアサプライチェーンの生態系からオープンソースによる攻撃が発生したら、ソフトウェアサプライチェーン全体に悪影響を及ぼし、破壊的な攻撃被害が発生しうる。ソースコード段階で攻撃者が悪意的なコードが含まれているパッケージを使用したり、正常的なパッケージと類似なパッケージをダウンロードして使用する場合連載的なセキュリティ危険に漏出される。



【▲「図 セキュリティ脅威 4-1」ソフトウェア開発ライフサイクル基準の攻撃ベクターの現況】

オープンソース生態系で発生しうる攻撃タイプは、「表 セキュリティ脅威 4-1」のように整理できる。ソフトウェア開発ツールが公開及び非公開レポジトリからサードパーティパッケージを持ってくる方法から発生する流れで、非公開レポジトリにホスティングされたカスタムパッケージから正常的なパッケージ名でダウンロードを試みたとしても、公開レポジトリから掲載された悪性のあるパッケージを持ってきてダウンロードすると、従属性の混乱(Dependency Confusion)が発生しうる。

既に公開されている有名なパッケージ名と類似なパッケージ名を利用して攻撃するタイポスクワッティング(Typosquatting)や、クラウド環境に存在するレポジトリのアクセス権限やトークンを奪取して攻撃する権限漏洩または管理者権限盗用、レポジトリパッケージ内の悪意的なコードを修正、挿入、CVE、CWE、KVEなどオープンソースから発生した脆弱性を対応せず、潜在的な攻撃ベクターとして使用する攻撃方法など多様な攻撃タイプが存在する。

NO	攻撃タイプ	攻撃方法
1	従属性混乱 (Dependency Confusion)	• 概要：ソフトウェア開発ツールが公開及び非公開レポジトリから3rd Packageを持ってくる方法から発生する流れで非公開レポジトリにホスティングされたカスタムパッケージの代わりに公開レポジトリから掲載された悪性パッケージを持ってくる攻撃方法 • 攻撃方法：NPM Registry「npm」、PyPI(Python Package Index)「pip」、RubyGems「gem」やPythonの「--extra-index-url」、Rubyの「--index-url」などを使用する場合、公開レポジトリから検索するため攻撃危険発生 • Microsoftからは「3 ways to mitigate risk when using private package feeds」でサプライチェーンの攻撃緩和方法を提示
2	タイポスクワッティング (Typosquatting)	• 既に公開されたパッケージ名と似たような悪意的なパッケージをアップロードしてダウンロードを誘導する攻撃方法
3	権限漏洩 / 管理者権限盗用 (Privilege Leakage / Stealing Administrator Privileges)	• 公開及び非公開レポジトリに保存されているクレデンシャル情報奪取及び不正アクセスで権限上昇を行う攻撃方法
4	レポジトリパッケージ内の悪意的なコード (Malicious code via Repository Package)	• Repository Packageを攻撃してパッケージのソースコードの内にMalicious Codeを挿入したり一部修正して攻撃する方法
5	既知の脅威や脆弱性 (Known Threat / Vulnerability)	• CVE、CWE、KVEなどオープンソースから発生する脆弱性対応していないため、潜在的な攻撃ベクターとして使用する攻撃方法 • 公開された脅威や脆弱性情報を活用しているため、セキュリティアップデートが公開された場合、影響度分析による迅速なセキュリティ対応が必要
6	その他攻撃方法	• ソフトウェアサプライチェーン(Software Supply Chain)に侵入し、サプライチェーン内で自動で動作させたり、悪意的なファイルを配布する攻撃方法 • DockerやContainerなどのRepository内の悪意的なコードやプロセスを挿入し、DockerやContainerを使用時、悪意的なコードが自動で実行されるようにする方法 • 開発者が良く使用するプラグインやソフトウェアに偽装してIDE Pluginに悪意的なVisual Studio vs-extensionを配布

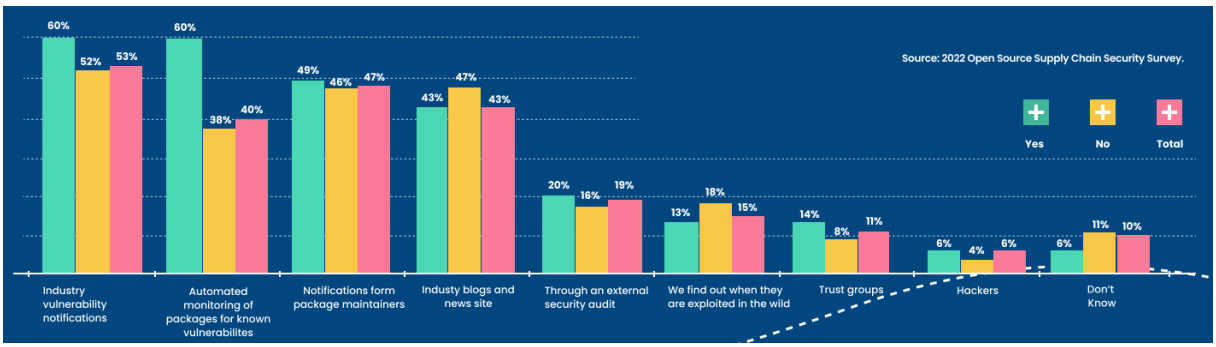
【▲「表 セキュリティ脅威 4-1」オープンソース生態系を活用した攻撃タイプ別、攻撃方法】

このようにオープンソース生態系を利用した攻撃タイプは年々に増加している。「表 セキュリティ脅威 4-2」のNPMレジストリ(NJPM Registry)でオープンソースレポジトリ攻撃の状況を見ると、オープンソースレポジトリで共有されているオープンソースを悪用した攻撃が持続的に発生していることが確認できる。このような攻撃事例でオープンソース生態系を利用した攻撃の中、一番頻繁に使用される攻撃はタイポスクワッティング依存性の混乱攻撃であることが分かる。

攻撃年度	攻撃方法	攻撃タイプ
2017	• cross-envパッケージとにたようなcrossenvパッケージ名を使用するタイポスクワッテイングで被害システムの環境変数を収集するマルウェア配布	Typosquatting
2018	• Event-Streamパッケージ(Streamを簡単に使用できるようにサポートするパッケージ)と依存関係であるFlatmap-Streamパッケージにビットコイン決済プラットフォームであるCopayの財布情報とプライベートキーが奪取できるマルウェア挿入	Dependency Confusion
2021	• UA-Parser-JS NPMライブラリ(ブラウザのUser-Agent情報パーシング機能を適用するライブラリ)ハイジャックでマイナ(Coinminer、仮想通貨採掘マルウェア)とパスワードスティーラー(Password Stealer)挿入攻撃	Dependency Confusion
2022	• GitHub.com統合業者であるHeroku及びTravis CIに発行されて盗まれたOuathユーザートークンが攻撃キャンペーンに活用 • 攻撃者は盗んだOAuthトークンを使用して複数のPrivate NPMレポジトリをダウンロードした後、獲得した損傷されたAWSアクセスキーを使用して権限上昇 • ▲2015年ユーザー情報赤い部にユーザー名、パスワードハッシュ、メールアドレスなどが含まれた約10万件のアカウントログイン情報、▲2021年4月7日から全てのプライベートパッケージマニフェスト、メタデータ、▲2022年4月10日から全てのプライベートパッケージの掲載されたバージョン名とsemVer、▲2社のプライベートパッケージなどが侵害	Privilege Leakage
2023	• 正常のNPMパッケージであるnode-hide-console-windowと似たようなパッケージで最後にsを追加した悪性パッケージnode-hide-console-windowsにより、DiscordRAT 2.0マルウェアが追加された700回ぐらいダウンロードされた後、削除処理	Typosquatting
	• 損傷されたシステムにリバースシェルを配布する機能が含まれた48個の悪性NPMパッケージがhktalnetというユーザーによって掲載されたことが発見	Dependency Confusion

【▲「表 セキュリティ脅威 4-2」NPM Registryによるオープンソースレポジトリ攻撃現況】

「図 セキュリティ脅威 4-2」のようなオープンソース上から発生するセキュリティ脅威要因を識別する多様な方法が存在する。オープンソース脆弱性や従属性問題を解決するためにCISA(US-CERT)、NIST(NVD)、MITRE(CVE)などの信頼できる機関から脆弱性情報をもらって組織内でもっていいるオープンソース上で公開された弱点や脆弱性はないかチェックし、さらにパッケージ上の既知の脆弱性を自動でモニタリング(Automated monitoring of Packages for known vulnerabilities)して脆弱性を識別することが重要である。



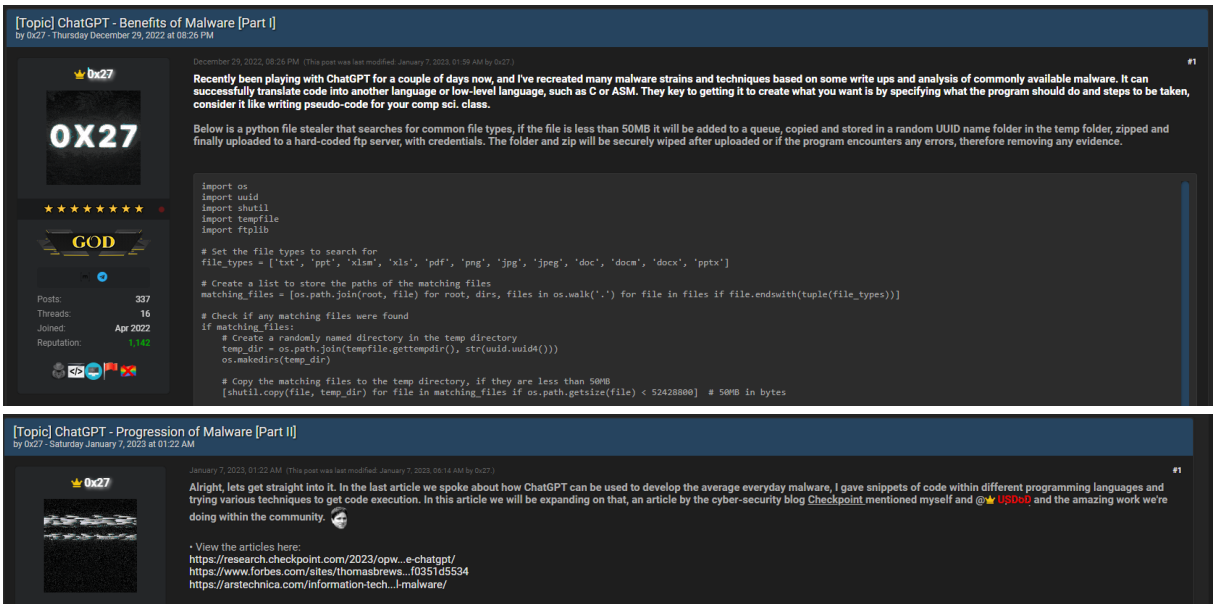
【▲「図 セキュリティ脅威4-2」オープンソース内の脆弱性情報確認方法 (参考：2022 Open Source Supply Chain Security Survey)】

5) 生成型AIを利用したサイバー攻撃の増加

2022年11月にオープンAI(OpenAI)は、GPT3.5基盤のチャットGPT(ChatGPT)を発表し、全世界から注目を集めた。これでグローバルIT企業は大型言語モデル(LLM, Large Language Model)でユーザーの特定要求に応じて結果を生成する生成型AI開発のための研究を激しく進めている。

GPTが初めて発表された2018年のGPT-1は1億1,700万個のパラメータを使用したが、2019年発表されたGPT-2は15億個のパレメータが使用された。2020年に発表されたGPT-3では、なんと1,750億個のパラメータが使用された。パラメータが増加することは大型言語モデルとしては正確度が増加することを意味することで、短期間にも出るの性能が向上したことである。GPT-3.5も既存バージョンと性能面では大きい違いはないが、直観的なダッシュボードと人工知能と話す感じがなく、自然に対話ができることで他の大型言語モデルよりもいい評判をもらっている。

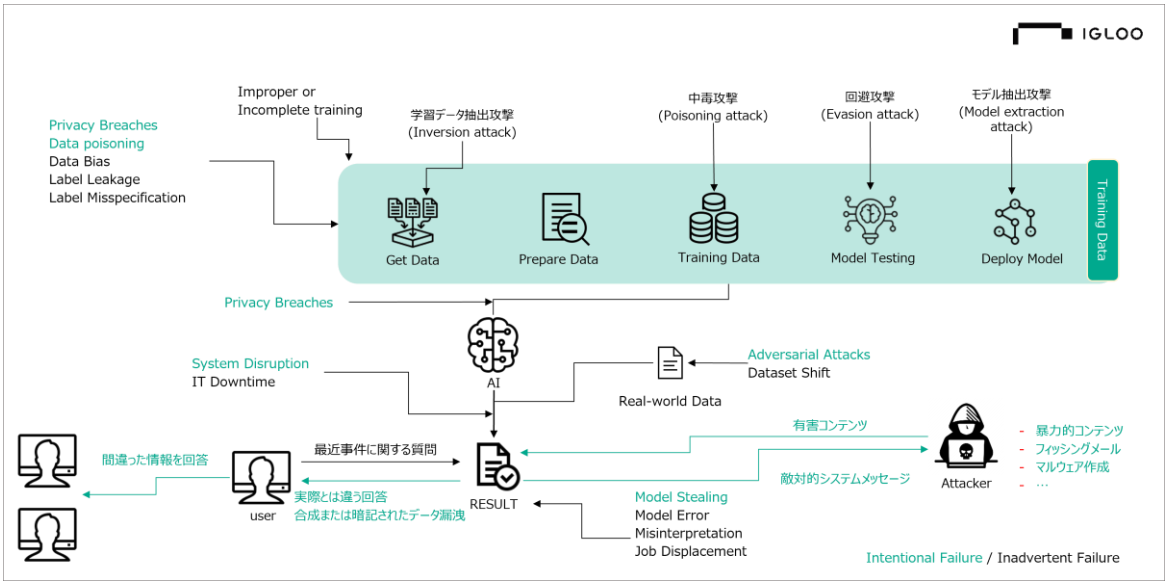
問題はChatGPTの登場でサイバー攻撃の正確性向上及びセキュリティアーキテクチャバイパスなど既存サイバーセキュリティ体系を無力化させるための人工知能活用及び実証事例が増加し、人工知能を利用したサイバー攻撃の問題が発生し始めた。ChatGPTはPythonで作成されたマルウェア作成の要請結果で、ファイルが50MB未満であれば収集対処に入れてから任意のフォルダを作成し、コピー及び保存して、最終的にはFTPサーバに漏洩させるコードを作成をしてくれというユーザー要請に「図 セキュリティ脅威 5-1」のようにマルウェアソースコードを作成する。このような結果はオンライン上で話題になり、ChatGPTを利用したサイバー攻撃ツール及び技術開発の導火線になった。



【▲「図 セキュリティ脅威 5-1」アンダーグラウンドフォーラムに掲載されたChatGPTを利用したマルウェア作成の例】

生成型AI以外にも人工知能やマシンラーニングを利用したサービスを使用する流れで、間違った答えや、むやみに入力したデータによる秘密情報及び企業機密情報、間違った情報の誤用など信頼性及びセキュリティ問題が話題になった。

人工知能やマシンラーニングから発生しうるセキュリティ脅威は「図 セキュリティ脅威 5-2」のようにデータ収集及び分析後、トレーニング(Training)の過程で収集したデータの偏向(Bias)や文字もしくは画像をまるで本物のように回答する幻覚(Hallucination)、モデルが間違った結果を導出する敵対的攻撃(Adversarial Attack)などがある。



【▲「図 セキュリティ脅威 5-2」生成型AIによるセキュリティ脅威構成図 (参考：韓国国家情報院、ChatGPTなど生成型AI活用セキュリティガイドライン(2023.6), ISQED, Sandip Kundu, security and Privacy of Machine Learning Igorithms、一部再構成)】

人工知能やマシンラーニングの訓練段階で、攻撃者が訓練したデータを汚染させたり、アルゴリズム上に介入して性能に影響を与えることができる。「表 セキュリティ脅威 5-1」はこのような訓練段階で影響を及ぼせる攻撃タイプを整理した内容である。データ学習時、悪意のあるサンプルを挿入してモデルを損傷させることができるデータインジェクション(Data Injection)やデータ修正でデータを汚染させることができるデータ修正(Data Modification)、学習アルゴリズムに介入して論理的設計を損傷させることができる論理損傷Logic Corruption)で分類できる。

攻撃区分	攻撃による損傷対象	説明
Data Injection	• サンプル挿入によるモデル損傷	• 訓練データはもちろんアルゴリズムにもアクセスできないが、訓練セットに新たなデータを追加することが可能
Data Modification	• データ修正による訓練データ汚染	• 攻撃者は学習アルゴリズムにアクセスできないが、学習データによる全ての権限を持つ
Logic Corruption	• 学習アルゴリズム介入による論理損傷	• 学習論理を変更してモデル自体を制御できる攻撃者に対する対応戦略を設計することがとても難しくなる

【▲「表 セキュリティ脅威 5-1」訓練段階(Trainingナ、発生できる攻撃 (参考：Adversarial Attacks and Defences: A Survey, ANIRBAN CHAKRABORTY, 28 Sep 2018)】

学習過程で発生しうる敵対的攻撃(Adversarial Attack)で、「表 セキュリティ脅威 5-2」のような中毒攻撃(Poisoning Attack)がある。中毒攻撃モデルのトレーニングの過程で発生し、全体学習プロセスを損傷させて攻撃する方法で、2016年マイクロソフト(Microsoft)が発表した人工知能チャットボットTayが人種差別的で暴力的なメッセージを乱発し、運用16時間でサービスが中止された事件が代表的な例ともいえる。

中毒攻撃以外にも回避攻撃(Evasion Attack)、探索的攻撃(Exploratory Attack)が存在する。回避攻撃は、変造された結果を人間が認識できるかどうか、変造された過程が物理的なのかデジタル的なのかによって分類する。探索的攻撃は、AIモデルのアルゴリズムまたは学習データに対し、探索してデータ学習時、使用されたデータを抽出してAPIモデルの機能を実現して、似たようなモデルの実現などの攻撃が可能である。

攻撃	影響範囲	説明	事例	攻撃種類
中毒攻撃 (Poisoning Attack)	• 学習モデルの訓練時間の間に発生	• サンプルデータを挿入して学習データを汚染させて全体学習プロセスを損傷	• 2016年MSの人工知能Tayが悪意的な発言をするように学習され、悪口及び人種差別の発言を乱発して16時間で運用中止	• Support Vector Machine Poisoning • Poisoning on collaborative filtering systems • Anomaly Detection Systems
回避攻撃 (Evasion Attack)	• 訓練データに影響×	• 攻撃者はテスト段階から悪性サンプルを調整してシステムを回避しようとする • 2つの側面が存在：認識しやすい/難しい攻撃、物理的/デジタル攻撃	• 改ざんされた表示板(Attack Pattern)で人間は認識可能だが、自律運転自動車は速度制限の表示板(Speed Limit 45)として認識 • ノイズで人間は区別できるがAIは認識不可	• Adversarial Examples Generation • Generative Adversarial Networks (GAN) • GAN based attack in collaborative learning • Intrusion Detection System • Adversarial Classification
探索的攻撃 (Exploratory Attack)	• 訓練データセットに影響×	• モデルの基本システム及び学習アルゴリズム、訓練データのパターンについて多い知識を探査無 • モデル転倒攻撃(Model Inversion Attack)、APIによるモデル抽出攻撃(Model Extraction via APIs)存在	• AIモデルの学習に使用されたデータを抽出して機密情報漏洩可能性存在 • APIをで機能的に似たようなモデルを実現する可能性存在	• Model Inversion • Membership Inference attack • Model Extraction via APIs • Information Inference

【▲「表 セキュリティ脅威 5-2」敵対的攻撃の種類(参考：Adversarial Attacks and Defences: A Survey, ANIRBAN CHAKRABORTY, 28 Sep 2018)】

訓練段階移行にはモデル内部の公開有無によって「表 セキュリティ脅威 5-3」のようなホワイトボックス(White-Box)とブラックボックス(Black-Box)攻撃が存在する。ホワイトボックスは機械学習モデルに対するホワイトボックス攻撃から攻撃者は分類で使用されるモデル(f)に対して全体的な知識を持って攻撃することを意味するため、攻撃者が学習データ分布や学習アルゴリズム、最終モデルなどを知っている上でモデルの脆弱性を把握する攻撃方法である。ブラックボックスはモデルに対する知識がないと仮定で設定や過去入力情報を利用してモデルの脆弱性を分析する。

攻撃区分	攻撃者の取得情報	攻撃戦略	説明
White-Box	▪ トレーニングで得たネットワークアーキテクチャとパラメータに対する詳細な情報	• 入力に対するネットワーク変化を基盤とする	• 攻撃者は使用されるモデル(f)に対する前提情報を持っていて訓練データ分布にアクセス可能 • 使用した情報を活用してモデルの脆弱な部分を識別
Black-Box	• 一部探索された入力のネットワーク出力に対する情報	• 入力に対する出力の変化を観察して近似値作成	• モデルに対する知識がないと仮定し、設定や過去入力情報を利用してモデルの脆弱性を分析

【▲「表 セキュリティ脅威 5-3」Testingから発生しうるWhite-BoxとBlack-Boxの攻撃方法 (参考：Adversarial Attacks and Defences: A Survey, ANIRBAN CHAKRABORTY, 28 Sep 2018)】

ブラックボックス攻撃は「表 セキュリティ脅威 5-4」のようにNon-Adaptive Black-Box Attack、Adaptive Black-Box Attack、Strict Black-Box Attackの三つのタイプで分類できる。攻撃者が学習アルゴリズムとデータなどに対する情報有無によって分類される。

攻撃区分	説明
Non-Adaptive Black-Box Attack	• 攻撃者は対象モデルの訓練データ分布(Training Data Distribution)だけアクセスできる • データ分布サンプルでローカルモデルを訓練し、モデルアーキテクチャ(f')を作成、f'に対する敵対的事例を作って対象モデル(f)に適用して間違った分類を発生させる
Adaptive Black-Box Attack	• 攻撃者は学習過程に対する情報は存在せず、モデルにアクセスできる(選択平文攻撃と類似) • 対象モデルに任意の値xをお問い合わせして出力値yを得る • 入力値・出力値(x,y)を利用してモデルアーキテクチャ(f')作成、f'に対する敵対的事例を作って対象モデル(f)に適用して間違った分類を発生させる
Strict Black-Box Attack	• 攻撃者は入力値・出力値(x,y)が収集可能 • Adaptive Attackのように出力の変化を観察するために入力は変更不可 • 大規模の入力値・出力値(x,y)を利用して成功する可能性存在(選択平文攻撃と類似)

【▲「表 セキュリティ脅威 5-4」Black-Box攻撃タイプ別特徴 (参考： Adversarial Attacks and Defences: A Survey, ANIRBAN CHAKRABORTY, 28 Sep 2018)】

今まで人工知能モデル学習のためのデータのセキュリティ脅威に関して説明したが、人工知能の性能を阻害するセキュリティ脅威は「表 セキュリティ脅威 5-5」から確認できる。大きくデータの正確性及び安定性とAI関連脆弱性で分類できる。

データ正確性及び安全性の側面としては、間違った情報、データ漏洩、類似AIモデルサービスの真似などがセキュリティ脅威と言える。偏向された学習データで作成されたモデルがTayのようなチャットボット事例のように間違った意思決定を誘導したり、社会的な混乱を引き起こせる。個人情報や機密情報などが匿名化されず、モデルの学習に活用されて機密情報の漏洩などの問題が発生しうる。

AI関連脆弱性の観点からみると、プラグイン脆弱性、拡張プログラム脆弱性、API脆弱性で分類できる。プラグインや拡張プログラムの場合、人工知能モデルの適用範囲変動及び脆弱なサービスとの連携による誤作動の危険が存在する。API脆弱性も一般的なサイバー脅威と同じく、APIキー管理及びAPIによる不適切なデータ使用及び漏洩などの問題が発生しうる。

その他にも生成型AIサービス提供業者おセキュリティ対策不備によるシステム可用性阻害及び学習データ漏洩やサービス停止、API使用時キー管理の不備による機密情報アクセス、承認されていない作業実施、不正アクセスによる権限奪取など人工知能モデルやサービスの安全性に影響を及ぼせる。

区分	詳細脅威	主な原因	セキュリティ脅威による影響度
データ正確性及び安全性	間違った情報	• 偏向されたテキストデータ学習による特定グループや主題に対する偏見がある結果生成 • 最終学習データ時点までの情報のみ持っているため、その後発生した事件及び情報について知らなかったり不正確な情報提供 • 間違った情報や存在しない情報の生成ができるため、信頼性低下	▪ 社会的混乱助長 ▪ 高危険意思決定 ▪ 間違った意思決定誘導
	データ漏洩	• データ合成過程の問題 • 過度な訓練データ暗記問題 • 対話の改定で個人情報及び機密情報の作成	• 訓練データ漏洩 • データ不法処理の恐れ • 機密漏洩/対話記録漏洩 • データベースハッキング及び会員推論攻撃
	類似AIモデルサービスの真似	• 類似悪性サービスにアクセス誘導	• スクワッティングURL及び拡張プログラム • 偽アプリケーション
AI関連脆弱性	プラグイン脆弱性	• AIモデルの適用範囲拡張 • 安全性確認不備 • ハッカーの攻撃範囲拡張 • 脆弱性があるサービスと連携	• 新たなドメインからのモデル誤作動 • 「エージェント」化されたAIモデルの悪用 • マルチモデル悪用
	拡張プログラム脆弱性	• 拡張プログラム内部のアクセスサービスインストール • サービス提供業者のセキュリティ対策不備	• 個人情報収集 • システム攻撃 • ホスティングサーバ及びストレージシステム脅威
	API脆弱性	• APIキー管理の不備 • データと命令の間の不明な境界	• APIキー奪取 • 悪意的なプロンプト挿入

【▲「表 セキュリティ脅威 5-5」大規模言語モデルなど生成型AIに対するセキュリティ脅威（参考：韓国国家情報院、ChatGPTなど生成型AI活用セキュリティガイドライン(2023.6)、一部再構成)】