

AI 悪用が“大衆化”した大量サイバー攻撃時代に向けた課題と視点－大規模攻撃事案のケーススタディから－

先進領域研究部サイバー安全保障研究室特任研究員 佐々木 勇人

はじめに

2026 年 9 月頃から不正アクセスによる個人情報等の漏えい事案が相次いで公表され、専門機関等の調査により、AI を悪用した大規模サイバー攻撃活動ⁱⁱによるものであることが判明しつつある。筆者は本務先の専門機関 JPCERT コーディネーションセンターのアナリストとして当該事案の対応に現在もあたっているが、こうした参与観察的な知見を前提としつつ、海外で発生している同種の攻撃ケースⁱⁱⁱや、過去の大規模攻撃事案対応のケーススタディも踏まえ、今後、同様の攻撃類型が一般化することを想定し、社会全体における対応や制度面の観点から、中長期的な課題について考察するものである。

当該攻撃活動について、(現時点では)いわゆる「安全保障上の懸念を生じさせる重大なサイバー攻撃」にただちに該当するかどうか不透明な状況であるが、5 月の拙稿^{iv}にて、AI を悪用した大規模なサイバー攻撃事案対処の課題について考察したことを踏まえ、NIDS コメンタリーにて取り上げる次第である。

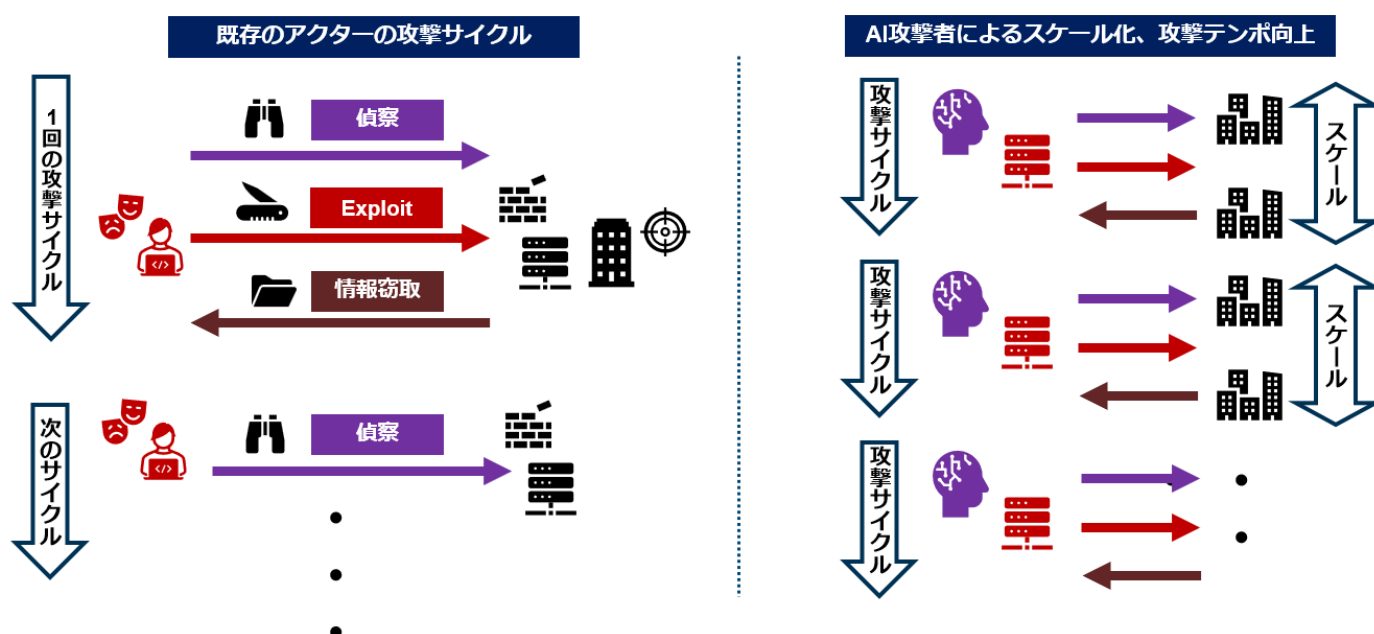
なお、取り上げる攻撃概要等は本稿執筆時点のものであり、引用する各種分析レポートの範疇において考察したものである。当該時点以降の攻撃手法の変化等までカバーできていない点ご容赦願いたい。

AI 悪用による攻撃のスケール化、攻撃テンポの向上

2026 年 4 月頃の Claude Mythos をめぐる騒動では、生成 AI の能力向上により、未知の脆弱性が多数発見され、多数のゼロデイ攻撃が発生するようなシナリオに関する言説が散見された。他方で今回の攻撃は個別の攻撃手法自体は既知の脆弱性や管理不備などを突くものであり、ある意味既存の攻撃手法をまとめたものである。少なくとも今回の一連の攻撃活動においては、攻撃能力の急激な高度化というよりも、攻撃規模のスケール化と、攻撃テンポの向上という点が注目される。

従前のサイバー攻撃活動では、人間である攻撃者がひとつひとつの攻撃プロセス／フェーズの作業を行っており、基本的には実行者自身のリソース上の制約が攻撃範囲やその頻度を限定していた。こうした攻撃サイクルが AI により自動化されることで、実行者自身のリソースに制限されず、その範囲とテンポを各段に向上させたため、「短期間で広範囲に攻撃を仕掛ける」ことが可能となったのである。AI 攻撃インフラにより、攻撃者側が（人的）リソースから解放されることで、攻撃のスケール化、攻撃テンポの向上が達成されるのである。

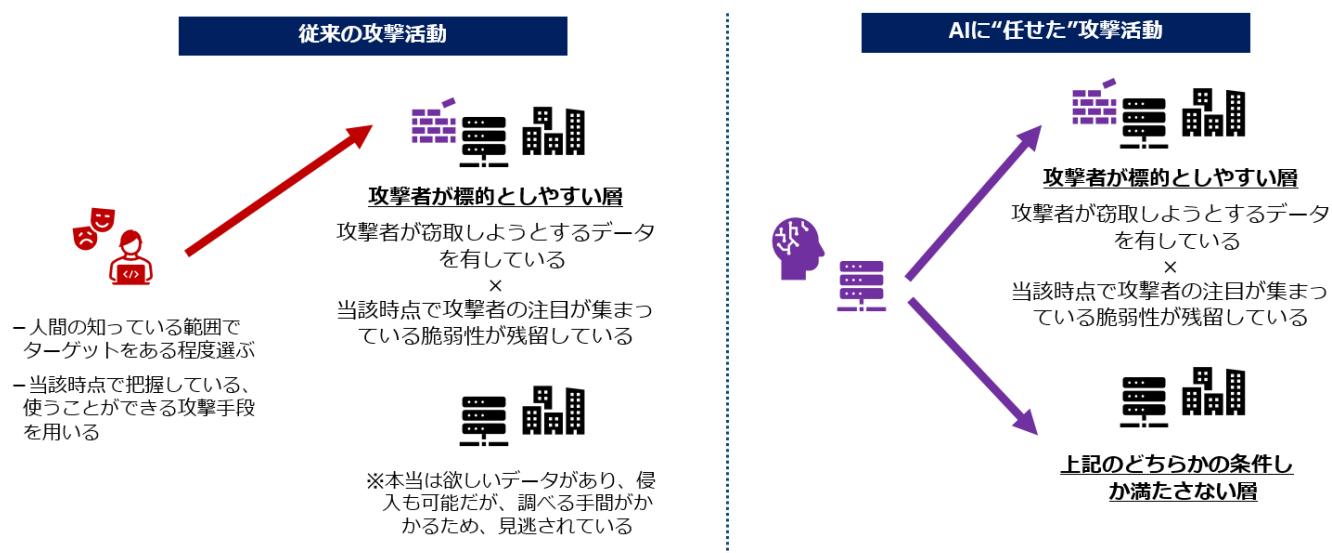
図：既存の攻撃サイクルと AI 攻撃者による攻撃のスケール化・攻撃テンポ向上



こうしたスケール化、高テンポ化には、単に「作業が自動化されるから」というだけではなく、人間から AI に実行者の主軸が代わることで体が強く影響している。通常の攻撃者（人間）は、自身の知ってい

る範囲でターゲットや窃取対象の情報をある程度選別する。また、当該時点でその攻撃者が注視している、また、使うスキルがある攻撃手段を選択する。つまり、人間の攻撃者は自身の知見や得ている情報、様々なトレードオフなどの判断などに影響され、限定合理的な活動を行うのである。他方で、AI にはこうした制限がないため、その攻撃活動は完全合理性を目指すことになる。その結果として、2026 年 9 月の事案のように、攻撃手法はひとつではなく、単一のソフトウェアやその脆弱性を狙うのではなく、また、多くの人々がそれほど攻撃対象として注視していなかった、従業員向けシステムなどを次々と狙うといった複合的・飽和攻撃的なサイバー攻撃が実行されるのだと筆者は分析^{vi}している。

図：限定合理的な人間による攻撃と完全合理的な AI による攻撃



大規模攻撃時における対応側の諸課題

冒頭で触れた、2026 年 9 月前後の連続不正アクセス事案や類似の海外事案のほか、これまでの筆者が本務先で対応にあたった大規模攻撃事案（Wannacry2.0（2017 年）、Emotet（2019 年～）等）のケーススタディを踏まえ、大規模事案時の対応側の各種課題について考察する。

大規模攻撃かつ、漏えい情報に個人情報が含まれるなど、法令上の報告・届出対象の被害が発生した場合、各行政機関への報告・届出が「殺到」する状況になる。各国でも同様の傾向にあるが、昨今の報告「迅速化」の傾向も相まって、短期間で発生する大規模事案では殺到状況になりやすい。殺到状況による制度上の弊害・問題点は以下の 2 点である。

①攻撃原因や傾向の特定が困難となる

大規模事案の発生初期においては、情報の混乱に加えて、当該大規模事案に関係のない、平時から発生している別の攻撃活動による別個のインシデント情報なども混在してしまい、攻撃原因の特定プロセスにおいて混乱し、適切な対策を示せなくなることがある。特に今回のような複数の攻撃手法を組み合わせ、分野横断的／無差別に行われる攻撃の場合、大量の情報が行政機関に集まるとともに、短期間で複数事案に係る被害公表が連続することで、共通原因や傾向を見出すことが格段に難しくなるため、社会全体として効果的な対応を取りづらくなってしまう傾向にある。時間が経てば専門組織による分析が進み、徐々に共通原因や攻撃手法の全容が明らかになっていくが、初動段階においてはこうした混乱が発生しやすいのである。

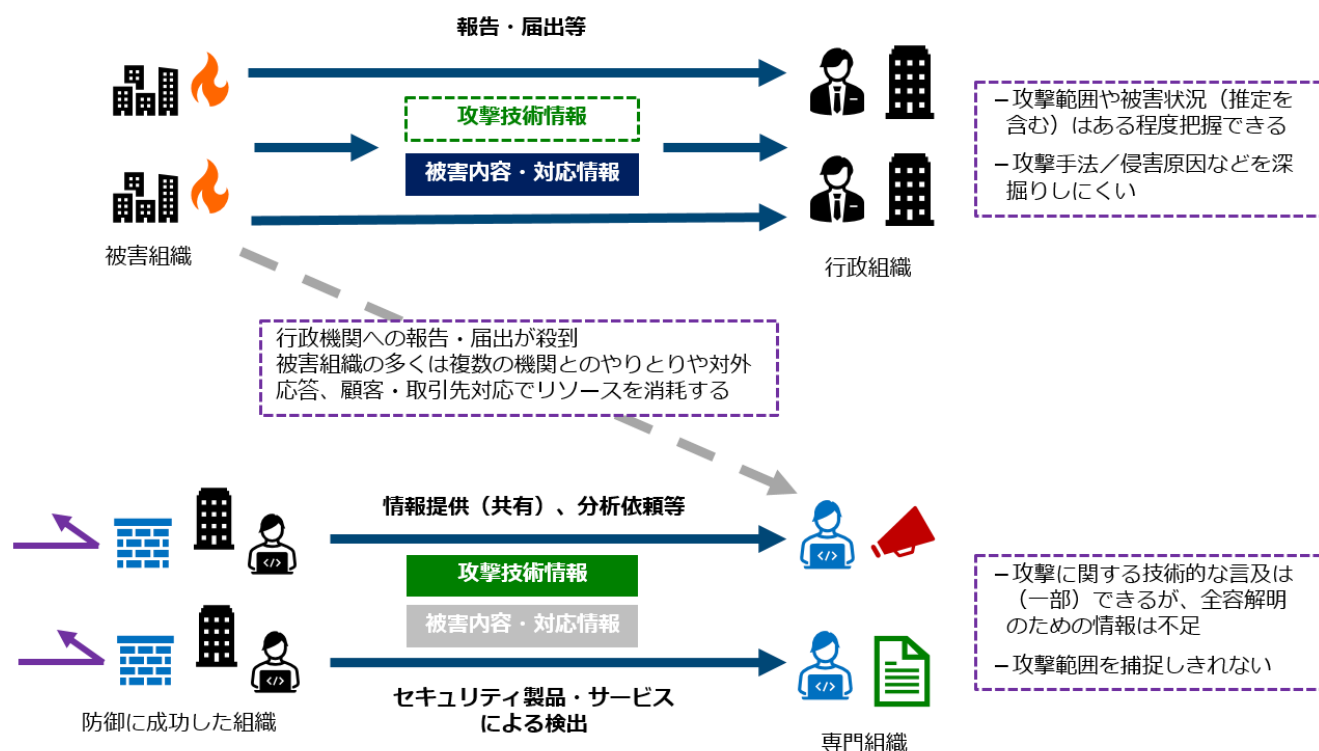
②攻撃技術情報が共有されにくくなる

法令上の報告のほか、複数の行政組織において任意の届出等も求めるため、被害組織は複数行政組織とのやりとりにその対応リソースを割くことになる。加えて、この手の大規模事案では対応初期の時点で被害公表をせざるをえず、対外応答や影響顧客・取引先対応に迫られるため、外部の専門組織を交えた情報共有活動まで手が回らないことが多い。通常であれば、攻撃認知時から報告、公表等までいくらかの時間的猶予があるため、複数外部組織との連携ができなくはないのであるが、大規模事案では他の被害組織が立て続けに被害公表に至っていることもあり、レピュテーションリスクを懸念して、通常よりも早い報告・届出、被害公表を行うことになり、リソースが急速にひっ迫するのである。

2026 年 9 月の連続不正アクセス事案では、先行して技術的な分析を公表したマクニカネットワークスのレポート^{vii}において、情報共有の必要性について言及され、また、JPCERT/CC からの注意喚起にも、技術的情報の共有が不足している点が明示的に示された。さらに、経済産業省からの注意喚起^{viii}においては、件名においても「情報提供に係る協力依頼」と明記された。

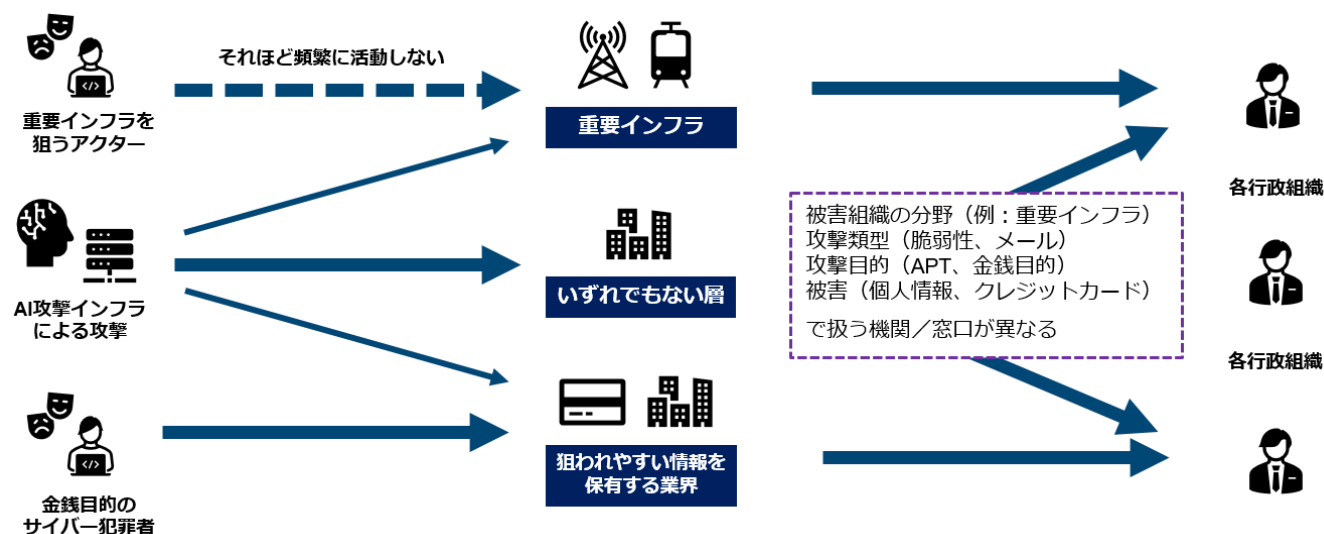
行政機関側も情報の収集に苦戦することとなる。被害組織の多くはあくまで事故「報告」と認識しているため、報告情報のほとんどは、被害状況や調査の進捗、顧客等の二次被害者対応の状況といった、非技術的な情報になってしまいがちである。筆者も本務先で執筆に関わった、「サイバー攻撃被害に係る情報の共有・公表ガイダンス」^{ix}では、攻撃被害に関する情報を「攻撃技術情報」と「被害内容・対応情報」の 2 つに切り分け、それぞれ取り扱うことを解説・推奨している。行政機関に報告されるのはもっぱら後者の情報であり、攻撃に関する技術的情報である「攻撃技術情報」は提供されにくいのである。

図：官民間における攻撃に関する技術的情報の“目詰まり”



そもそも行政側の情報集約システム自体に限界・課題がある。現行の制度・体制においては、被害組織が属する分野（重要インフラ等）の所管や、被害内容（個人情報、クレジットカード等）に係る所管関係で報告先省庁・窓口が整理されている。また、攻撃類型や扱う技術的情報の種類により行政機関やその窓口が担当組織内でも細かく分かれている。ある意味、こうした細分化は国の分担管理原則[※]に基づく個別最適ではあるのだが、今回のような分野／情報の種類にまたがった攻撃、さらには複数の攻撃手段を組み合わせた複合事案が起きた場合、被害や攻撃手法等に関する情報は細切れになり、複数の行政機関に分散して到達することになる。行政機関側において複数機関間の情報連携を行う調整コストが発生してしまうため、今回のような攻撃テンポが速い複合事案に対しては現行の情報集約システムは全体最適化されていないのである。

図：所管分野、被害内容、攻撃類型に細分化され個別最適化された国の報告システム



「注意喚起しか出せない」背景と注意喚起の限界について

2026 年 9 月の連続不正アクセス事案に関する 10 月 12 日付の時事通信^xにて「相次ぐ情報流出、政府苦慮 「注意喚起」頼み、打つ手乏しく」と題した記事が公開された。記事中では、「政府関係者は「注意喚起が最も早い」と説明するものの、別の関係者は「民間企業が相手なので、お願い以外に方法がない」と漏らした。」とあるが、なぜ「注意喚起しかできない」状況になるのでしょうか。

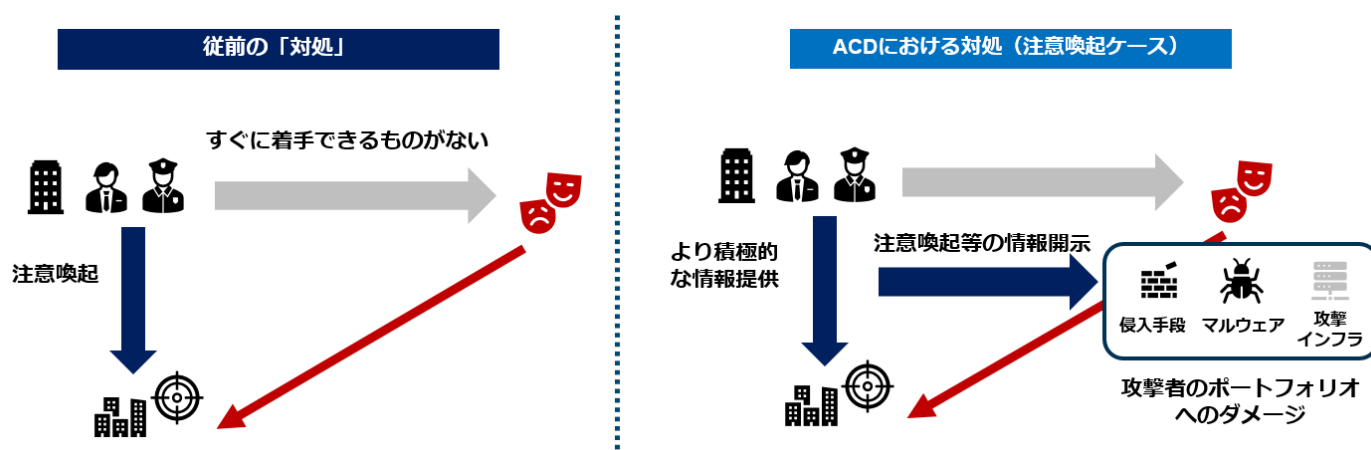
既存のサイバー攻撃対応においては、残念ながら後手に回ることが多く、被害組織側での認知タイミングもさることながら、被害情報が行政機関まで届く頃には攻撃キャンペーンは既に終了してしばらく経っている、というケースが大半であった。他方で、大規模事案により攻撃キャンペーン中に社会全体がこれを認知するケースにおいても、政府全体で状況の把握が難しい背景は前述の解説の通りである。また、事後の攻撃実行者の検挙や訴追公表などの対処までは相当の時間がかかるため、国として攻撃キャンペーン中であろうが終了直後であろうが、直ちに行えるオプションは「注意喚起」くらいしかない、ということになる。結果として、国の「対処」のベクトルが脅威側に向くのではなく、被害組織や潜在的な標的組織側に向いてしまうことになり、その内容も自主的な取り組みを促すものになってしまうのである。

他方で、近年ではセキュリティ製品の能力向上や CSIRT 整備を始めとした企業等での対応能力強化もあり、攻撃キャンペーン中にこれを認知できる機会も増えつつある。また、攻撃が本格化する前に注意喚

起を出せるケースも増えつつある。能動的サイバー防御（ACD）においては、拙稿^{xii}で解説の通り、攻撃キャンペーン中の対抗オペレーションにより、攻撃者のポートフォリオへのダメージ付与や攻撃キャンペーン達成を阻害する、いわゆる「コスト賦課」アプローチを採ることとなる。この対抗オペレーションを組成する対抗オプションはソフトな対抗手段（注意喚起やレポート開示）からハードな対抗手段（アクセス・無害化等）まで多種多様である。

このうち、注意喚起は攻撃者のポートフォリオ上の弱点を突き、当該技術的情報を暴露することで、その効力を失わせ、攻撃キャンペーンの継続に影響を与えることを目的としている。従前の注意喚起がもっぱら被害組織／標的組織へ向いた、「内向き」なものであったのに対して、能動的サイバー防御で求められる注意喚起はさらに脅威側にも影響を与える、「外向き」なものとして期待がされるのである。けっして、「注意喚起しか手段がない」のではなく、注意喚起の出し方次第、ということである。

図：旧来の「内向き」な注意喚起と能動的サイバー防御（ACD）における「外向き」の注意喚起



図：能動的サイバー防御における対抗オペレーションを構成する様々な対抗オプション（東京海上ディールールの調査研究プロジェクト資料^{xiii}より抜粋）

1. 注意喚起（一般向け連絡、非公開での情報共有活動参加者への連絡）
2. 情報共有（情報共有活動参加者へのインディケータ展開）
3. 関連組織への通知（脆弱な機器／侵害された機器の利用者への通知、不正な通信が発生している組織の特定と通知）
4. 解析レポートの公開（攻撃手法や攻撃インフラの暴露、侵害指標（IoC）等の脅威情報の共有）
5. フィルタリング
6. 事業者と協力した通信遮断・テイクダウン
7. 他者が管理するネットワーク・情報資産へのアクセス・情報収集
8. 無害化
9. パブリック・アトリビューション（特定したサイバー攻撃者の公開、名指し批判）
10. 刑事訴追等の司法的措置
11. ペルソナ・ノン・グラータ等の外交的措置
12. 金融制裁指定等の経済的措置

注意喚起はソフトな対抗手段であるがゆえに、「誰でも出せる」という点で実施ハードルが最も低い対抗オプションである。他方で、その効果に関する考察^{xiv}や理論的な整理はほとんどなされていないのが現状である。乱暴な言い方をすれば、「出しっぱなし」的な状態にある。本稿は注意喚起の理論的考察をすることが主旨ではないので、細かなところは省略させていただくが、結論から言うと、注意喚起という対抗手段は以下のような課題を有している。

○注意喚起の課題：対象の分野がある程度限定されていれば効果があるが、分野横断や無差別となるにつれ、効果が希薄になる

注意喚起は基本的に公表手段をもって伝達されるため、どれだけ想定オーディエンスがそれを「見る機会があるか」次第で、その効果が左右される。特定分野に対する注意喚起であれば、注意喚起公表以外の手段、例えば所管省庁や当該業界団体経由で情報を再展開するよう働きかけたり、個別の説明会を実施したりするなど、確実に対象組織に伝達することも可能である。2026 年 9 月の事案のような複数の攻撃手法／アタックベクターが組み合わさる複合攻撃では標的対象範囲があまりに広く、対象のプレイヤーが多すぎるため、基本的に注意喚起効果が希薄にならざるを得ない。

○注意喚起の課題：対象のシステムや攻撃手法が限定的であれば効果があるが、複合的な攻撃になると効果が希薄になる

一般的な注意喚起は、特定のシステムの脆弱性を悪用する攻撃や、特定の攻撃手法を集中的に運用する攻撃に対して行われる。こうした一般的な注意喚起においては、対策を明示的・具体的に示すことができ、アクションナブルな情報として伝達するため、その実効性も担保されるのである。他方で、今回のように複数の攻撃手法等を組み合わせたものとなると具体的な対策情報が「広く浅い」記載となってしまう、また、ここまでの解説の通り、情報集約が混乱するため、なおのこと、各機関から出る情報に具体的な技術情報が不足しがちになるため、総論的な記述になりがちである。

○注意喚起の課題：攻撃者のポートフォリオ上の弱点を突かない限り効果は限定的である

拙稿^{xv}で解説の通り、また次項で解説の通り、攻撃者にコスト賦課等の影響を与えるためには、ポートフォリオ上の重点、弱点を突く必要がある。例えば、攻撃者が複数の脆弱性を悪用することができる場合、1つのソフトウェアの脆弱性の注意喚起をしたところで、攻撃者は別の脆弱性を悪用する動きにシフトするだけであるため、注意喚起の効果はほとんどないことになる。

2026 年 9 月の事案において、侵入手段は多種多様であり、単一の手段に攻撃の成否が依存しているわけではないため、侵入手段はポートフォリオ上の弱点ではない。ひとつひとつの攻撃手法は既知の手法などを組み合わせているため、攻撃手法自体が圧倒的な優位性を持っているわけではない。また、多数の攻撃ツールなどが見つかったが、少なくとも現時点で攻撃活動が依存するような、特徴的なマルウェアを使っているわけでもないため、これも弱点にはならない。

残る攻撃インフラであるが、JPCERT/CC からの解説レポートの通り、攻撃の中核となる AI エージェントや攻撃ツール、各種データを蓄積し、攻撃の中核を構成している。ここに攻撃のためのコアコンポーネントが集約されていることで、効率的でスピーディな攻撃を構成していると言える。したがって、現時点でポートフォリオ上の弱点と思われるのは攻撃インフラ、特にプロキシサーバ等でなく、攻撃の中核サーバである、AI エージェントなどが搭載された攻撃サーバ群ということになる。同種の攻撃活動に関する海外セキュリティ企業の分析レポートが相次いで公開^{xvi}されたが、いずれもこの攻撃インフラの調査・分析が主軸となっている。

2026 年 9 月の連続不正アクセス事案の考察と今後必要な検討について

○企業等における対策について

現時点までの状況においては、繰り返しになるが、既知の脆弱性や管理不備を突く攻撃が多く、既知の脆弱性対応や API の推奨運用設定などの足場固めがなおも有効である。確かに攻撃被害が続発しているが、他方で、対策にとって防御できていた組織もまた多数存在しているのである。当該連続不正アクセス事案に限らず、筆者の本務先である JPCERT/CC からの注意喚起などの活動は、被害組織だけでなく、「攻撃を防げた」組織（標的となった組織やこれを支援したセキュリティ企業等）からの貴重な情報提供によっても支えられている。特に、脅威への「対策」を示すためには、「どうやったら防げるのか／防げたのか」というアクションブルな対策情報を発信しなければならないのであり、被害事案からの情報だけに依存してはならないのである。

従前の国としてのインシデント対応の基本は、被害組織からの報告・届出情報を元にした情報収集にあるが、本稿前半で解説の通り、大規模事案かつ、今回の AI 悪用のような高テンポな攻撃活動においては、被害現場サイドからの情報伝達は“目詰まり”を起こすため、十分に機能せず、また、攻撃活動の変化に追いつけない恐れもある。

○インシデント対応の“温度感”について

被害公表を読むとわかるが、攻撃から被害の検知・認知、被害公表までのスピードが従前のインシデント対応に比べて非常に早い事例が多い。あくまで筆者の仮説であるが、攻撃の自動化と API 経路でのデータ取得という特徴から、一度に大量のリクエスト／処理が発生し、アラートに至ったケースが相当数存在していると考えられる。

各組織で速やかな被害公表が行われたことで、行政側からの注意喚起を待たず、社会全体で対応の温度感を上げることができたとも考えることができる。現在進行形で攻撃が継続されている状況かつ、大規模な攻撃発生時においては、自主的な侵害有無調査や対策強化を速やかに行う必要があり、「対応の温度感」は貴重な情報となる。

他方で、被害公表における記載内容やその粒度^{xvii}は被害組織毎に大きなバラつきがあり、「特定の攻撃が連続しているのか、別々の攻撃（原因）なのか判別しづらい」という面も存在している。被害公表もまた、個別最適ではなく、全体最適を意識することで、社会全体のインシデント対応能力向上につながる重

要な手段となるだろう。

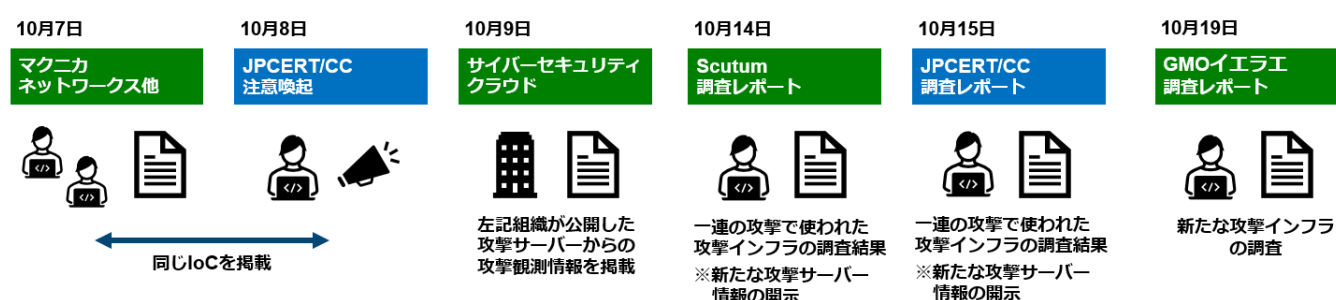
○脅威情報の開示に基づく専門組織の連携サイクル

いわゆる IoC (Indicator of Compromise) 情報として、攻撃サーバに関する情報を掲載したのは、10 月 7 日のマクニカネットワークスのレポートが最初であり、翌 8 日の JPCERT/CC からの注意喚起上での掲載となった。続いて、これらの情報を踏まえて、サイバーセキュリティクラウドが自社サービス側での観測結果を公開した。10 月 14 日のビットフォレストのレポートでは、AI 悪用の攻撃であることを特定する分析手法を編み出し、これにより捕捉した新たな攻撃インフラ群に関する情報が開示された。その後も攻撃インフラに関する様々な観測情報、調査結果の公表が続いている。

攻撃技術情報の開示により、複数の専門組織が攻撃活動（攻撃インフラ）を追跡できるようになるとともに、WAF を始めとした各セキュリティ製品・サービスでの検知能力が向上していくことになる。情報発信は、それ単体で攻撃者のポートフォリオへのダメージや攻撃キャンペーンの阻害を行えるのではなく、サイバーセキュリティサービス^{xviii}をほぼ 100%支える民間製品・サービスを通じて、その効果を発揮することになるのである。

注意喚起の効果の考察にて、情報が「内向き」にだけ出されることの限界を解説したが、こうした攻撃に関する技術的情報、特に攻撃インフラに関する情報も、従前は潜在的被害組織や標的組織向けに展開し、検知・防御に使われる前提で扱われることが多かったが、情報開示をすることで他の専門組織のリソース、知見で新たな攻撃インフラを次々と見つけていくというサイクルを回すことができるのであり、「外向け」のインディケータ情報発信と言える。これもまた、攻撃者側にコスト賦課するための重要な対抗オプションの一つと言える。

図：攻撃技術情報の開示による共有と多数の専門組織による連携サイクル



○「被害」の再定義の必要性

前半で整理の通り、現行の被害報告制度等はこうした AI 悪用の大規模攻撃事案に対応が追い付かない、限界を迎えるおそれがある。現状では、不正アクセスの発生自体が「被害」であり、「何らかの情報が漏れた可能性」をもってして事故対応が進められる。この定義では、被害の実態的影響に関係なく、インシデントに重みづけがされずに扱われるため、今後のインシデント増加時代においては行政機関等が量的に情報を処理しきれなくなる恐れがある。

不正アクセスが発生したとしても、実態的な被害がないインシデントも多数ある。また、なんらかの情報が漏えいしたとしても、当該情報をサービス上失効させるなどして、その後の二次被害発生を防止できるケースもある。本当に保護されるべき情報は何なのか、デジタル情報が漏えいしても、人の生命・身体・財産を保護するためにどのような仕組みがあればよいのかあらためて検討が必要である。不正アクセスされること自体を重大視するのではなく、「維持すべきシステム、保護すべきものは何か」に基づいた制度上の「被害」の再定義が必要なのである。

また、ただ漫然と「サイバー攻撃被害だらけの世界」を受容するのではなく、大量サイバー攻撃時代に向けて、社会全体のレジリエンスを柔軟に変えていく必要もあると言える^{xix}。既に金融機関においては、大規模攻撃への懸念がある場合に能動的／予防的にシステムを止める選択肢にも言及^{xx}されている通り、システムの停止という従来ネガティブに見られがちな手段も、能動的な対応オプションとして柔軟に利用できるようになりつつある。これまでの慣例や先例に囚われず、新たな脅威に柔軟に対応していかなければならない。

○既存の対応制度の限界と見直しの必要性について

従前の国のセキュリティインシデント対応は、法令等に基づく報告や任意の届出をベースとして情報を収集し、中長期的な傾向からその対策を検討し、潜在的な標的組織／分野に対して、短期～中期的には注意喚起などの情報提供行政／媒介行政的なアプローチ^{xxi}、長期的には法令やガイドライン、認証・審査制度などの規制行政的なアプローチや、場合によっては、導入補助金制度などの誘導行政的なアプローチが組まれる。問題なのは、これらのアプローチが基本的に「対策のスピード」という点で決定的な弱点を有することである。被害が確認される、あるいは蓋然性が高くなければ、国への情報提供は基本的に行われず、国が特定の攻撃傾向に気付く頃には既に相当の被害が発生した後、ということになる。これが従前の攻撃活動の場合、国が新たな攻撃類型の登場を認知するまでに相当の期間（1，2 か月程度）を有してしまうが、さらにその先、数年単位での長期的な対策強化を狙うアプローチであるため、従前の規制行

政／給付行政的なアプローチでもそれなりの成果を期待できていたのである。

他方で AI 悪用による攻撃活動はその攻撃テンポが急加速することや、次の攻撃タイミングではまったく予想できない攻撃手法を用いることが想定されるため、これら従前の対応サイクルや、行政機関としての学習サイクル^{xxii}がまったく追いつかない可能性が高い。

先述の「○脅威情報の開示に基づく専門組織の連携サイクル」の項目で紹介の通り、民間専門組織、専門機関間の技術的情報の連携・サイクルが自然発生的に展開されたように、サイバーセキュリティの世界では、官民間問わず多様なプレイヤーがそれぞれの能力を活かし、社会全体の攻撃対処に関与することが可能である。「官民連携」は単に情報を国が収集するための方便ではなく、民間のプレイヤーの役割を（再）認識し、役割分担と連携を改めて見直すことであると筆者は考える^{xxiii}。本稿でここまで解説の通り、被害組織だけでなく、「攻撃を防いだ」組織は重要な情報提供者であり、被害組織自身の被害公表は社会全体の対応温度感を上げるアラートとしての機能を担う。セキュリティ専門組織がそれぞれ分析し情報を開示することで、セキュリティサービス側の検知能力が上がり、「面」として対抗し、攻撃キャンペーンを阻害することが可能になる。行政側にだけただサイバー攻撃対処の責を問うのではなく、官民連携により共に新たな脅威に対処する術を整備していかなければならない。

○ドクトリンの必要性

先述の、官民様々なプレイヤーがそれぞれの特性を生かしながら、社会全体でスピーディな対応を行うためには何が必要であろうか。それは、サイバー攻撃対処の「ドクトリン」とであると筆者は考えており、これまでも拙稿^{xxiv}で触れてきた通りである。

被害現場から事故情報としてのインシデント情報収集を行い、さらには、各機関に散らばった情報の共有に係る調整を経て、そのうえで「司令塔」組織が対処方針を決め、さらに各機関に指示を展開し、という流れでは、どうしても AI 悪用のようなサイバー攻撃の攻撃テンポには追いつくことが難しい。あらかじめドクトリンがあり、これに基づいた能力整備が平時から行われ、いざ事案対応が始まった場合は、官民それぞれの組織が自らの役割を踏まえたアクションを取ることで、自然と連携した事案対処を行う必要がある。各組織が事案毎に個別最適での場当たりの対応を行うのではなく、全体最適が平時から示され、その元で各組織が機動的に対処しなければ、AI 悪用の高テンポ／スケール化した飽和攻撃に対処は間に合わないのである。もちろん、これは「司令塔」組織の不要を説くものではない。むしろ、司令塔組織がドクトリンを定め、各組織に機能と役割の要望を出すことで、平時からの能力構築が行われていくのである。また、ドクトリンに基づいた、我が方側の対処テンポが揃わない場合など、その調整役

としての役割が必要になるのである。

むすびに代えて

AI 攻撃者もまた、サーバの調達や特定のソフトウェア、通信プロトコルの利用と、我々と同じ土俵（環境）で活動しているのであり、その実態を捕捉することはできる。しかしながら、本稿で解説の通り、完全合理性で素早い攻撃テンポを構成する AI 攻撃者側に対して、我々側は限定合理的、個別最適で各プレイヤーが動き、また情報が分散してしまっており、対応が後手に回っている感が出てしまっている。既に、サイバー攻撃における AI の本格的な活用フェーズに突入しており、今後は同様の攻撃手法が一般化し、また、本件ほど「わかりやすい」形で表出せず、水面下で被害が拡大するケースも想定される。本事案への対応はまだ終わっていないが、今の時点から、旧来の対処アプローチに捕らわれない柔軟な発想で、将来に向けた対応体制全体の見直しを行うことが急務である。

ⁱ 2026 年 10 月 15 日 JPCERT/CC「連続不正アクセス事案の攻撃手法に関する調査」
https://blogs.jpcert.or.jp/ja/2026/10/web_api_attacks.html

ほか、2026 年 10 月 13 日情報通信研究機構（NICT）「【注意喚起】相次ぐ情報漏洩、その原因の一つは「アプリ解析からのサーバ攻撃」— AI で高度化する脆弱性探索、Android アプリを手がかりにした A2B 攻撃に注意 —」<https://cynex.nict.go.jp/library/261013-01.html>、2026 年 10 月 14 日 ビットフォレスト「AI エージェントによるウェブサイトへの攻撃の注意喚起」https://www.scutum.jp/information/waf_tech_blog/2026/10/waf-blog-085.html、2016 年 10 月 19 日 GMO サイバーセキュリティ by イエラエ「相次ぐ不正アクセスの攻撃元の多くが AI であることを確認」<https://gmo-cybersecurity.com/blog/20261019/> を参照のこと。

ⁱⁱ 筆者の本務先である JPCERT/CC で把握した、現在までに判明している攻撃のケースは以下の通りである。 参考：2026 年 10 月 8 日 JPCERT/CC「直近で相次いでいる国内組織における不正アクセスに関する注意喚起」<https://www.jpcert.or.jp/at/2026/at260030.html>

○脆弱性スキャンを行い、侵入に使える既知の脆弱性の有無や設定不備等を探し出す

○一般公開されているアプリを解析し、システム側の API のエンドポイントやキーを特定し、API 経

由での攻撃をしかける ※詳細は前掲注 1 の NICT の分析レポートを参照。

○BI（ビジネスインテリジェンス）ツールに使われているソフトウェアの脆弱性（例：Metabase の SQL インジェクションの脆弱性（CVE-2026-72898））を悪用した攻撃 ※これも上記と同様に API 経由の攻撃 参照：2026 年 8 月 14 日 JPCERT/CC「Metabase の SQL インジェクションの脆弱性（CVE-2026-72898）に関する注意喚起」

○公開 Web サーバーからアクセス可能なアプリケーションサーバー上に Web シェル設置

○ほか、AI を用いたペネトレーションテスト基盤を悪用した攻撃試行

JPCERT/CC の調査（前掲注 1 参照）では、攻撃インフラ内に AI エージェントや各種攻撃ツールが大量に蔵置されていたことが判明しており、こうした AI 攻撃インフラによる、ある程度自動化された攻撃活動が大規模に行われていたと推測される。

iii 2026 年 9 月 3 日 Hunt.io, “Robobox: Self driven malware creation and execution - made in China”, <https://netaskari.substack.com/p/robobox-self-driven-malware-creation>, 2026 年 9 月 22 日 Gambit Security, “Autonomous AI Agents are breaking into hundreds of Online Retailers for \$25 a target in an ongoing campaign”, <https://gambit.security/blog-posts/autonomous-ai-agents-online-retailers-25-a-company>, 2026 年 10 月 19 日 VULONE, “China-Linked Hackers Targeted Governments Across Asia. We Hacked Their Infrastructure.”, <https://vulone.com/blog/a-china-linked-operator-used-ai-agents-to-attack-governments-across-asia>, 2026 年 10 月 19 日 NORFORGE, “Inside Cairn: Open Windows into massive AI-Driven Cyber Operations”, https://norforge.com/blog/inside_cairn

iv 防衛研究所 NIDS コメンタリー 佐々木勇人「高度なサイバー攻撃キャンペーンにおける AI 悪用によるスケール化のインパクトと対策」
<https://www.nids.mod.go.jp/publication/commentary/commentary434.html>

v AI 悪用による、サイバー攻撃の「スケール化」の論点については、前掲注 4 の拙稿を参照

vi 複数の攻撃手法が組み合わされている点については JPCERT/CC の注意喚起等で指摘したところ、セキュリティ専門企業による観測結果からの分析においても同様の指摘がなされている。 2026 年 10 月 15 日 LAC「国内で相次ぐ Web システムからの情報漏えい～ラックの観測から見える変化と、企業が今とるべき対策」https://www.lac.co.jp/lacwatch/alert/20261015_004996.html

vii 2026 年 10 月 7 日 マクニカ セキュリティ研究センター「相次ぐ WEB システムからの情報漏洩事

案について」 <https://security.macnica.co.jp/blog/2026/10/web-incidents2026.html>

viii 2026 年 10 月 9 日 経済産業省「不正アクセスによる情報漏えい等の事案を踏まえた注意喚起及び情報提供に係る協力依頼」
https://www.meti.go.jp/policy/netsecurity/pdf/20261009_chuuikanki.pdf

ix 2023 年 3 月 サイバーセキュリティ協議会サイバー攻撃被害に係る情報の共有・公表ガイダンス検討会（事務局：NISC（当時）、総務省、警察庁、経済産業省、JPCERT/CC）「サイバー攻撃被害に係る情報の共有・公表ガイダンス」 https://www.cyber.go.jp/pdf/policy/kihon-2/guidance2022_honbun.pdf

x 藤田宙靖「行政組織法（第 2 版）」p144 ほか

xi 2026 年 10 月 12 日 時事通信「相次ぐ情報流出、政府苦慮 「注意喚起」頼み、打つ手乏しく」
<https://www.jiji.com/jc/article?k=2026101200114&g=pol>

xii 防衛研究所 NIDS コメンタリー 佐々木勇人「能動的サイバー防御の実装に向けた残された論点（前編）——攻撃者のポートフォリオ分析の必要性」
<https://www.nids.mod.go.jp/publication/commentary/commentary398.html>

xiii 2025 年 5 月 21 日東京海上ディーアール川口貴久「新たな段階に進む「能動的サイバー防御（ACD）」」（東京海上ディーアール 調査研究プロジェクト「サイバー安全保障と能動的サイバー防御」） <https://www.tokio-dr.jp/thinktank/acd/acd-001.html>

xiv 2023 年 5 月 9 日 JPCERT/CC 佐々木勇人「注意喚起や情報共有活動における受信者側の「コスト」の問題について —情報発信がアリバイや成果目的の自己目的化した行為にならないために—」
<https://blogs.jpcert.or.jp/ja/2023/05/cost-and-effectiveness-of-alerts.html>

xv 前掲注 12 参照

xvi 前掲注 3 参照

xvii セキュリティインシデントにおける被害公表の各種論点やその方法等については、前掲注 9 の「サイバー攻撃被害に係る情報の共有・公表ガイダンス」に詳しい

xviii 本稿では、現場で実際の攻撃を検知・防御・分析しているモノ・サービスを総称して、「（サイバー）セキュリティサービス」と呼称する。それらを実際に担っているのは民間企業のセキュリティ製品

やサービスである。

^{xix} レジリエンスを変えていく必要性については、元駐カタル特命全権大使、元内閣官房副長官補の前田哲氏との意見交換の中でご示唆をいただいた。あらためて感謝を示すものである。

^{xx} 2026 年 6 月 18 日 全国銀行協会 会長記者会見

<https://www.zenginkyo.or.jp/news/conference/2026/260618/>

^{xxi} 規制行政やリスク行政との関係、情報提供行政／媒介行政の整理については、板垣勝彦「保障行政の法理論」（弘文堂、2013 年）を参考とした。

^{xxii} 事故・インシデント情報の国による収集の仕組みと、法・システムの「学習」機能の観点からの考察は、山本隆司「事故・インシデント情報の収集・分析・公表に関する行政法上の問題（上）（下）」（ジュリスト 1307 号（2006 年）、同 1311 号（2006 年）に詳しい。なお、当該論考における「分析の経験を集中的・継続的に蓄積させる必要性が高い場合や、現実の、あるいは潜在的な被害者が抱く公的組織や私的組織のバイアスに対する疑い・不信を除去する必要性が高い場合には、むしろ、「学習」のための手続・組織を分化させることが有益である。」という指摘は、サイバーセキュリティに係る制度検討においても示唆に富むものであると考える。

^{xxiii} 紙幅の関係で本稿ではこの程度の言及である点ご容赦願いたい。リスク行政における官民連携のあり方、特に民間専門組織の役割の重要性については、前掲注 20 の板垣の研究が参考となる。国の役割の再設定と任務の私化（Privatisierung）、そして保障行政という観点がサイバーセキュリティ政策と対応体制の強化の参考になると考えている。

^{xxiv} 2025 年 9 月 26 日 防衛研究所 NIDS コメンタリー「2026 年能動的サイバー防御の実装に向けた残された論点（後編）——オペレーションの評価と戦略的コミュニケーションの必要性」

<https://www.nids.mod.go.jp/publication/commentary/commentary399.html> 2024 年 5 月 17 日 防衛研究所 NIDS コメンタリー 佐々木勇人「サイバー脅威インテリジェンス活用のための「ドクトリン」の必要性について —— 情報共有を巡る「市場の失敗」と「政府の失敗」を乗り越えるために」
<https://www.nids.mod.go.jp/publication/commentary/commentary319.html>

PROFILE

佐々木 勇人

先進領域研究部サイバー安全保障研究室特任研究員（本務先：一般社団法人 JPCERT コーディネーションセンター 脅威情報アナリスト サイバーセキュリティコーディネーショングループ担当部門長 兼 政策担当部長）

専門分野：能動的サイバー防御（ACD）、サイバー攻撃者のアトリビューション、サイバー脅威インテリジェンス、サイバー情報共有と攻撃被害公表

本欄における見解は、防衛研究所を代表するものではありません。
NIDS コメンタリーに関する御意見、御質問等は下記へお寄せ下さい。
ただし記事の無断転載・複製はお断りします。

防衛研究所企画部企画調整課

直 通：03-3260-3011

代 表：03-3268-3111（内線 29177）

防衛研究所 Web サイト：www.nids.mod.go.jp