HITACHI

Hitachi Systems Security Journal

VOL.78

株式会社 日立システムズ

Hitachi Systems Security Journal VOL.78

TABLE OF CONTENTS

●はじめに

本文書は、株式会社日立システムズ の公開資料です。バックナンバーは以下の Web サイトで確認できます。
https://www.hitachi-systems.com/report/specialist/index.html

●ご利用条件

本文書内の文章等すべての情報掲載に当たりまして、株式会社日立システムズ（以下、「当社」といいます。）といたしましても細心の注意を払っておりますが、その内容に誤りや欠陥があった場合にも、いかなる保証もするものではありません。本文書をご利用いただいたことにより生じた損害につきましても、当社は一切責任を負いかねます。
Microsoft、Windows は、マイクロソフト グループの企業の商標です。
その他記載の会社名、商品名は、それぞれの会社の商標または登録商標です。
本文書に掲載されている情報は、掲載した時点のものです。掲載した時点以降に変更される場合もありますので、あらかじめご了承ください。

ニラドリ・セカール・ホレ インタビュー

巧妙化する詐欺の手口から検知や防御策、個人の自己防衛術まで専門家が指南するディープフェイク対策のすべて

2025 年 11 月、CODE BLUE 2025 にて、米国の大手金融サービス企業 StoneX のニラドリ・セカール・ホレ氏による講演「ディープフェイク・サプライチェーン：サイバー犯罪の武器となる合成メディア」が行なわれた。ディープフェイクというと、有名人のニセ動画や政治的なフェイクニュースをイメージする方も多いだろう。だが、本講演では、企業などの組織を標的とした「高度なサイバー攻撃のツール」へと変貌を遂げたディープフェイクの脅威が明かされた。 本稿ではニラドリ氏へのインタビューを通じ、低価格なサービス（Deepfake as a Service）として産業化・巧妙化する攻撃側のサプライチェーンの実態に迫るとともに、それらに対抗するための最新の検知技術やユーザーの自己防衛術などについて話を伺った。

取材・文＝吉澤 亨史／通訳＝エル・ケンタロウ／撮影＝松沢 雅彦／編集＝斉藤 健一

アジアでディープフェイク詐欺が急増 DaaS（Deepfake as a Service）として産業化する動きも

吉澤（以下 Y）：素晴らしい講演をありがとうございました。一般にディープフェイクと聞くと、有名人の偽動画や政治的なフェイクニュースを思い浮かべがちですが、本講演では「高度なサイバー攻撃のツール」としての実態に焦点が当てられていました。攻撃側のサプライチェーンの動向から最新の検知技術まで、非常に示唆に富む内容だったと思います。さっそくですが、ディープフェイクの脅威をめぐる世界的な状況について、現在どのように捉えていらっしゃいますか？

ニラドリ・セカール・ホレ（以下 N）：現在はまだ、ディープフェイクの影響範囲を完全に把握できる段階ではありません。新世代の AI モデルが登場するたびに脅威は増し、対処は困難になっています。予防策やガバナンスを有効に機能させるには、検知モデルを継続的にトレーニングし続けるしかありません。私たちはインシデントに即応するためのプレイブックを確実に整備していますが、攻撃者が悪用する未知の AI モデルに対しても、常に万全の体制を整え続ける必要があります。

Y 生成 AI の普及により、現在では個人でも容易にディープフェイクを作成できるようになりました。実際のところ、脅威の主体として確認されている攻撃者像とはどのようなものでしょうか。背後に国家が関与するような、組織的な攻撃グループも存在するのでしょうか。

N 攻撃者のタイプについては、ロシアを拠点とする者が一部確認されていますが、そのほとんどは国家や巨大組織ではなく、単独の個人であると考えられています。しかし、攻撃者の身元は非常に多岐にわたるため、実際に彼らがどこから来ているのかを正確に把握することは困難です。

Y 攻撃元の特定が困難である一方で、被害が発生している地域についてはいかがでしょうか。地域的な偏りや特徴など、傾向があれば教えてください。

N 被害は世界的に幅広く発生していますが、特にアジア地域での影響が顕著です。実際に当社（StoneX）でも、ディープフェイクを用いた音声クローン攻撃を確認しています。攻撃者が自社のサービスデスクに電話をかけ、正当な従業員を名乗ってシステムを迂回しようとする手口です。

●ニラドリ・セカール・ホレ
Niladri Sekhar Hore

データエンジニアリングとサイバーセキュリティにおいて 10 年以上の経験を持つ。現在は米国の大手金融サービス企業である StoneX に在籍し、サイバー防御・オブザーバビリティ・セキュアなデータシステムの構築に注力している。データ・サイバー脅威インテリジェンス・ガバナンスが交差する複雑な問題解決に秀でている。近年は、合成メディアの悪用・ディープフェイクの検出・企業環境における敵対的 AI の進化について探求している。

Y 従業員を騙ってサービスデスクを欺くわけですね。そうしたアジア地域における攻撃も、先ほど挙がったロシアの攻撃者による可能性が高いのでしょうか。

N 被害地域の傾向は、基本的には標的となるドメインから判断しています。例えば「.com」であればグローバルな Web サイト、「.jp」であればアジア太平洋地域向けの可能性が高いと推測できます。ただし、攻撃自体は特定の地域に限定されるものではなく、その対象は世界中広範囲に及んでいます。

Y 講演の中で、脅威の産業化を示す「Deepfake as a Service（DaaS）」という言葉が出てきました。こうしたディープフェイク攻撃のためのサービス環境を構築し、提供しているのも、やはりロシア系の攻撃者が中心なのでしょうか。

N 私たちが「Deepfake as a Service（DaaS）」と

呼ぶエコシステムは、誰でも利用可能なオープンなプラットフォームのような側面も持っています。これらは正当な用途にも不正な用途にも使われ得ます。攻撃者は、こうした一般向けの AI ツールを悪用して業界の重要人物などを装い、システムや Web サイトに対して、あたかも正当なユーザーによるアクセスであるかのように偽装するのです。

Y 正規のツールが悪用されるだけでなく、表向きは画像や音声を作成する「AI クリエイティビティツール」を装いながら、真の目的は犯罪を支援することにあるような悪意あるサービスも存在しているのですね。

N そのとおりです。システムの真の目的や使われ方を外部から把握することは難しく、正当なサービスか不正なものかを見分けるのは非常に困難です。単なる娯楽目的のサービスを装っていても、そこにアップロードされた画像などのデータが背後で複数の攻撃者の手に渡り、破壊的な目的で悪用されてしまうリスクが潜んでいます。

Y サービスの真の目的を見分けるのは困難とのことですが、それでもユーザー側がサービスの「怪しさ」を事前に判断できるような、何らかの特徴や傾向はあるのでしょうか。

N 講演でも触れましたが、サービスごとにさまざまな価格モデルが用意されている点が特徴として挙げられます。多くの場合、基本料金に加えて「ストック」と呼ばれる追加オプションが提供されており、「2 ドル追加すればより長い動画が作成できる」「同じ画像を複数の人物向けに微調整できる」といった機能が存在します。わずか 2 ドルのコストと指先ひとつの操作で、誰もが高度な偽造を利用できてしまう手軽さは、現在の悪意あるサービス（DaaS）の際立った特徴といえるかもしれません。

Y 低価格で手軽なオープンのサービスが悪用されている一方で、そうした既存のサービスには依存せず、自前で独自のディープフェイクの仕組みを構築している攻撃者グループも相当数存在するのでしょうか？

N 十分にあり得る話です。全体における割合は少ないと推測されますが、独自の仕組みを用いて攻撃を行なう高度な攻撃者グループも存在します。攻撃者が標的を狙う際、必ずしも特定の一般公開されたツールだけを使うとは限らないということです。

ホレ氏の講演は、CODE BLUE 2025 のアーカイブから視聴することができる。
https://archive.codeblue.jp/2025/results/results/#result-17

Y 高度な独自手法を用いる攻撃者が存在する一方で、近年はメールによるシンプルなフィッシング詐欺でさえ、依然として大きな効果を上げているように感じます。

N そのとおりです。実際の配信方法は攻撃者次第であり、音声通話やフィッシング詐欺メール、テキストメッセージ、SMS など、あらゆる手段が用いられます。また、攻撃の対象も、単独の個人から企業まで多岐にわたります。

ディープフェイク検知の最前線
マルチモデル分析とユーザーの意識改革

Y ディープフェイクを検知する最新の手法について伺います。講演では、未知のデータに対する現在の検知率は約 60%であり、検知モデルのさらなるトレーニングが必要だと指摘されていました。精度を向上させるためには、具体的にどのようなデータを用い、どのようなトレーニングを行なうべきなのでしょうか。

N システム側の対策だけでなく、検知を強化するためにはまずユーザーへの教育と周知が必要です。目にするあらゆる情報や動画を鵜呑みにせず、何らかのアクションを起こす前に、必ず組織内で 100%の確認を行なうという意識を徹底させることが重要です。

Y セキュリティ意識向上のためのトレーニングは、すでに多くの組織で実施されているかと思います。それを踏まえた上で、ディープフェイクの脅威に対

数十年間、システムに「人間を疑う」よう教えてきましたが、これからは逆に人間が「システムを疑う」よう教え、すべてを偽物と仮定するマインドセットを持つことが必要だと語るホレ氏

して、組織が講じるべき具体的な対策は他にありますか。

N 具体的な対策として有効なのが「ゼロトラストポリシー」※1の徹底です。現在、サービスデスクに対して「携帯電話を紛失したためパスワードをリセットしてほしい」と従業員を装う問い合わせが増加しています。これに対し、ユーザーとの照合プロセスを設け、いくつかの質問への回答が合致した場合にのみ二要素認証を実施し、それを経て初めてシステムへのアクセス権を付与するという厳格な運用が効果的です。

Y ゼロトラストのアプローチは確かに効果的ですね。講演では画像の真正性を証明するC2PAプロトコル※2の活用にも触れられていましたが、技術的な対策として、ディープフェイクに特化した検知ソリューションはどのような仕組みになっているのでしょうか？

N 講演では、画像や動画に特化したソリューションを紹介しました。これらはまずリアルタイムモデルによる処理を行い、次の段階でより詳細な検査を実施します。具体的には、顔認識や改ざん箇所を視覚的に特定できる検査など、複数のモデルを用いて多角的に検証します。最終的に、それらすべての結果を「統合ビュー（アンサンブルビュー）」として集約し、統合された平均スコアを算出することで真偽を判定する仕組みです。

Y 複数のモデルを組み合わせて検査を行っているのですね。

N そのとおりです。複数のモデルから得られた確率を集約し、対象の画像や動画がどれほど現実的か、つまり「真実性」をスコアとして算出・可視化して判定を下します。この仕組みにより、3秒以内という短時間でディープフェイクか否かを判別可能です。なお、こうした検知に用いられる生体認証（ライブネスチェック）の技術自体は、すでに多くの金融機関において「顧客確認（KYC）」のプロセスとして広く導入されているものです。

Y 3秒以内で判別可能とのことですが、検知の足がかりとなるような、ディープフェイクデータ特有の特性や痕跡といったものは存在するのでしょうか？

N それぞれの偽造手法には特有の痕跡（シグネチャー）が存在します。検知においては、すべてのAIモデルが同じように振る舞うわけではなく、抽出する特徴が異なります。例えば、あるモデルは画像の微小な品質やノイズに基づいて特徴を抽出し、別のモデルは異なる箇所を処理します。私たちがマルチモデル分析を行なうのは、このようにモデルごとに異なる着眼点（特徴抽出の仕組み）を総合的に理解し、組み合わせる必要があるためです。各モデルの分析結果を集約することで、より予測精度の高いスコアを算出することが可能になります。

Y それぞれの手法に特有の痕跡があるということは、それを逆手に取ることはできないのでしょうか。例えば、使用された生成技術やサービスの痕跡を分析することで、背後にいる攻撃者自体を特定できる可能性はありますか？

N 最大の問題は、誰でも利用できるオープンな画像・動画生成サービスが60以上も存在し、簡単にアクセスできてしまう点です。現在は5秒でメールアドレスを作成できるため、利用したサービスが判

※1 **ゼロトラスト：**「何も信用しない」という前提に基づき、すべてのユーザーやデバイス、接続元のロケーションについて、情報資産などへのアクセス時につど確認を行なうことで、サイバー攻撃や内部不正などの多様な脅威を防ぐアプローチ。

※2 **C2PA：**Coalition for Content Provenance and Authenticityの略。デジタルコンテンツ（画像・動画・音声）の「作成者・作成日時・編集履歴」を記録し、その信頼性を保証するための技術標準（規格）。

明しても、作成した人物を特定することは困難です。さらに、これらのモデルは現在 API を提供しており、攻撃者は自身のシステムに直接連携させることが可能です。

Y 技術的に特定が困難であるとはいえ、法執行機関としては背後にいる攻撃者の正体を突き止めたいはずですよね。

N そのとおりです。StoneX でも日常的に攻撃を受け、法執行機関へ情報提供を行っていますが、攻撃者の特定にはなかなかつながりません。攻撃者が Telegram などのチャンネルに「システムを乗っ取った」と自慢げに書き込むケースでさえ、正体を突き止めることは不可能です。そのため私たちは、追跡よりも防御を固めるというアプローチに重点を置いています。

Y ディープフェイクはテキストから始まり、音声、画像、そして今やリアルタイムの動画生成にまでいたっています。今後、この脅威はどのように展開していくと予想されますか。

N 最近のフェイク動画の生成モデルは、もはや私たちが想像すらできないレベルにまで到達しています。これが AI 分野の進歩であり、その脅威は今後さらに拡大していくでしょう。組織が適切な防御策や保護措置を講じていなければ、極めて厳しい状況に直面することになります。

Y お話を伺って、その厳しさがよくわかりました。現在、フィッシング詐欺の手法として悪用されているディープフェイク技術は常に進化し続けています。そうした状況下では、あらゆる情報を基本的に疑ってかかるべきなのですね。つまり、性善説ではなく「性悪説」のスタンスで臨むべきだということでしょうか。

N そのとおりです。情報が真実だと仮定するのではなく、まず「偽物である」と疑ってかかる認識を持つことが対策の鍵になります。講演の最後でも強調しましたが、これまで私たちは数十年間、システムに対して「人間を信用する」ように教えてきました。しかし、AI がもたらす脅威に直面している現在においては、全く逆のアプローチが必要です。これからは人間に対して「システム（AI）を信用してはいけない」と教育していく必要があるのです。

学生時代に就職詐欺に遭い、切羽詰まった隙を狙われた経験が私の原点。騙される人々を見るにつけ、自分の知識を共有することで被害を防ぐ役に立てるはずだと考えているともホレ氏は語っている

デジタル社会で自分を守る術とは 研究の原点と、IT 大国インドに潜む 詐欺拠点の闇

Y 話題を変えて、ニラドリさんご自身についてお伺いします。どのようなきっかけから、サイバー犯罪やディープフェイクの研究に取り組むようになったのでしょうか？

N 私は政府機関と連携してサイバー犯罪対策に関わっているのですが、近年インドでは法執行機関などを装ってニセの逮捕状を送りつけ、罰金を要求する詐欺が急速に拡大しています。被害者の 90%がセキュリティを十分に理解しておらず、悪意のある通信と正規の通知を見分けられないのが現実です。多くのユーザーは「信頼できる組織とやり取りしている」と信じてしまっています。攻撃者はその心理的な隙を突いてくるため、ソーシャルエンジニアリング攻撃はきわめて高い効果を発揮してしまうのです。

Y それは非常に深刻な事態ですね。

N 現在インドでは、政府などからの正式な SMS メッセージのヘッダーに「-S」（正規のサービスを示す記号）を付ける対策が行われていますが、これもなりすましの対象になる可能性があります。こうした

最新の脅威やリスクを一般ユーザーに広く伝えていくことが重要だと考えたのが、私が研究に取り組むきっかけです。

Y 政府も積極的に対策を講じているのですね。

N はい。ただ、政府の対策だけでは限界があるため、オープンなプラットフォームや世界的な講演を通じて、ユーザー自身への啓発を強化する必要があると考えました。ユーザーがリスクを正しく認識し、自らを保護する術を持たなければ、事態は非常に深刻な方向へ進むでしょう。悪用されれば、偽造パスポートや偽の銀行書類の作成、そして不正な口座開設といった実害に直結しかねないからです。

Y 現代はあらゆるものがデジタル化されているため、リスクもさらに大きくなっていますね。

N 私たちはデジタル社会に生きています。自分の身元を守るためには、パブリックなプラットフォーム上で共有するあらゆる情報に対して対策を講じなければなりません。公開した情報に攻撃者がアクセスする可能性は 100%であり、それは絶対に避けるべきです。そうしなければ、さらに深刻な問題を引き起こすことになります。

Y サイバーセキュリティの分野は長く研究されているのですか？

N ええ、長く携わっています。ディープフェイクはごく最近の分野ですが、私の原点は学生時代にあります。イギリスで修士号を取得していた頃、私自身が就職詐欺の電話を受け、送金を要求されました。当時は勉強に追われ切羽詰まった状況であり、攻撃者はまさにその心理状態を狙ってきたのです。そうした手口で騙される人々を見るにつけ、自分の研究や知識をコミュニティに共有することで、きっと被害を防ぐ役に立てるはずだと考えるようになりました。

Y StoneX は米国の企業ですが、ニラドリさんはインド拠点にいらっしゃるとのことですね。インドはサイバー攻撃が多いのでしょうか？

N はい、インドはまさにサイバー犯罪のホットスポット（多発地帯）です。標的になりやすいだけでなく、私たちが確認している攻撃者の大半がインド国内に拠点を置いています。彼らの一部はグローバルに活動し、世界的な詐欺戦術を用いています。私の研究の焦点は、こうした攻撃の動向を一連のエコシステムとして捉え、攻撃者のマインドセットを理解することにあります。

Y 英国や日本政府が、インド系の詐欺拠点を閉鎖したケースもありますね。

N はい、インドには詐欺拠点が無数に存在します。彼らは世界中の人々を標的にしていますが、特に米国人が狙われる割合が大きいです。

Y なぜインドに詐欺拠点がそれほど多いのでしょうか。

N 最大の理由は、インドが技術的インフラと人材に恵まれており、攻撃に必要なリソースへ容易にアクセスできるからです。さらに、地方の辺鄙な場所に身を隠し、完全なリモートで業務を遂行できる点も挙げられます。攻撃者はインターネットを活用しつつも匿名性を保ち、現在では VPN でトラフィックを迂回させ、インド発の攻撃ではないように見せかけるなど、手口を非常に巧妙化させています。

Y 攻撃者のレベルも非常に高いということですか。

N 必ずしも彼ら自身の技術レベルが高いわけではありません。ただ、深い専門知識がなくても、高度なツールを悪用してシステムを迂回してくるため、非常に手強いのです。中国の攻撃者もインドと同等のレベルにあります。彼らが標的とする地域の約 40%はアジア太平洋地域で、残りの 60%は米国です。米国が最大のターゲットになるのは、彼らにとって言語の壁（英語）がないためです。

Y 本日はディープフェイクの脅威から防御策、「システムを疑う」意識まで、貴重なお話をありがとうございました。ニラドリさんの啓発活動が、今後さらに多くの人々を脅威から守る道標となることを期待しています。

吉澤 亨史（よしざわ・こうじ）1996 年にフリーランスライターとして独立。セキュリティ、エンタープライズ IT を中心に、ソフトウェア、PC など幅広い分野で取材活動に従事する。雑誌や Web メディアを中心に特集記事、ニュース、コラムなどを執筆している。

社会のさまざまな動向を把握し、リスクの変化に対応したセキュリティ体制を構築

Hitachi Systems CSI（Cyber Security Intelligence）Watch 2026.04

文＝日立システムズ

AI ×脅威モデリングで求められる対応

【概要】脅威モデリングは守るべき対象や優先的に対処すべきリスクを共有し、限られた資源の中で適切な判断を行なうための手法として注目されている。AIの活用により効率化が期待できる一方、組織固有の文脈を加味しない課題がある。本稿では、AIを活用した脅威モデリングにおける求められる対応を整理する。

【内容】AIは文章生成やコード補助にとどまらず、セキュリティ実務そのものに影響を与える段階に入りつつある。2026年4月、Anthropicは「Claude Mythos Preview」に関する情報を公開し、このモデルが重要なソフトウェアのぜい弱性を発見し、それを悪用するプログラムを自律的に生成する能力を持つことを示した。同社は重要インフラのソフトウェアで複数のゼロデイぜい弱性を特定したと報告しており、AIによるぜい弱性発見能力がぜい弱性対策の在り方に影響を与え始めている。このような状況においては、欠陥発見後の修正だけでは不十分である。AIによってぜい弱性の発見速度と分析の網羅性が向上するほど、実装後の対応に依存した対策は限界を迎える。したがって、どの資産が重要であり、どこに信頼境界が存在し、どのデータフローが攻撃経路となり得るのかを設計段階から整理することが不可欠である。

こうした背景で必要となるのが脅威モデリングである。脅威モデリングはシステムを分解し、データフローや信頼境界を明らかにしたうえで、起こり得ることを検討する手法である。一般に、構成の分解、脅威の特定、対策の整理といった工程を行なう。加えて、仕様変更や構成変更、外部連携などに応じて脅威モデルを継続的に見直す反復的な作業を行なう。この作業は知識やパターンに基づく整理が多いことから、実務上の負担が大きいという課題を抱えてきた。

脅威モデリングは反復的かつ網羅性が求められる作業であり、AIは有効な支援手段となる。AIは仕様書や構成図をもとに、既知の攻撃パターンや設計上の弱点を参照しながら、脅威候補の抽出や対策案を高速に提示できる。このため、人手では見落としやすい複雑な攻撃シナリオや構成要素の依存関係に起因するリスクも早期に把握しやすくなり、脅威モデリングの実務負荷軽減と品質向上が期待できる。

一方で、AIによる脅威モデリングは客観的なベース知識に基づき、脅威の網羅的な分析や選択肢を提示する。また、例外運用や現場特有の制約、過去の障害対応で培われた判断など暗黙知は文書上に明示化されていないことが多い。AIの判断では暗黙知を捉えきれず、提示する対策が理論上妥当でも実際の運用には適さないことがある。このため、リスク受容や事業上の優先順位付けは人間が判断する必要があり、大幅な実務負荷の軽減は難しい。

このような問題に対応するため、暗黙知をAIに学習させることが考えられる。しかし、暗黙知は組織体制や事業環境の変化によって陳腐化し、特定の状況や前提条件に依存しているため、常に正しいとは限らない。AIによる判断の妥当性を検証しないと、AIの出力が特定の前提に偏り、一般性を損なう恐れがある。したがって、暗黙知はAIの内部に固定的に学習させるのではなく、RAG（外部の信頼できるデータベースから情報を検索・取得し、その情報に基づいて回答を生成させる技術）のような仕組みにより外部知識として常に更新できる形で管理し、必要に応じて参照することが望ましい。

AI時代の脅威モデリングでは、脅威を網羅的に洗い出すのではなく、AIの出力と人の判断を統合し、システムの変化に応じてリスク認識と対策を継続的に更新する仕組みを確立することが重要である。さらに、開発と運用はこの共通基盤のもとで連携し、脅威モデルを継続的に改善する体制を構築することが求められる。

［情報源］https://red.anthropic.com/2026/mythos-preview/

セキュリティツールを実践的に紹介する連載企画

Let's try 不審メール解析

1. 攻撃手口確認編

文＝日立システムズ

1. はじめに

本稿は、各種セキュリティツールなどを実践的に紹介する連載企画です。今号より「不審メール解析」と題して、受信した不審なメールについて、安全に解析を行なうための基本的な手順について整理します。今回は、不審メールを解析するための環境およびツールの準備を行ない、メールを受信したという想定で不審メールに対する一連の解析の流れを整理します。また、攻撃者が使用する不審メールの手口についても整理することで、不審メールの違和感に気づく力などを醸成します。

1. 攻撃手口確認編
2. メール解析手順編

「不審メール解析 1. 攻撃手口確認編」では、まず、解析を行なうための解析環境の作成やツールの準備を行ないます。加えて、不審メールにおける攻撃で利用される可能性がある手口を整理しつつ、実際にその手口を利用したダミーのメールを作成し、それぞれの手口に対する解析を行ないます。

なお、本稿の安全性には留意していますが、安全を保証するものではありません。OA 端末（社内ネットワーク接続機器）で実施するのではなく、分離された回線内および機器を利用する事を推奨いたします。また、また、記載されている手順は意図的に無害化・簡略化していることをあらかじめご了承ください。

2. 解析環境の準備

2.1 Windows サンドボックスの準備

「Windows サンドボックス」とは、「Windows 10 May 2019 Update」で追加された Windows の新機能です。Windows OS の中に仮想的なコンピューター（Windows OS）を作り出すことができ、安全にソフトウェアの検証などを行なうことが可能です。

「Windows サンドボックス」の前提条件は下記のとおりです。

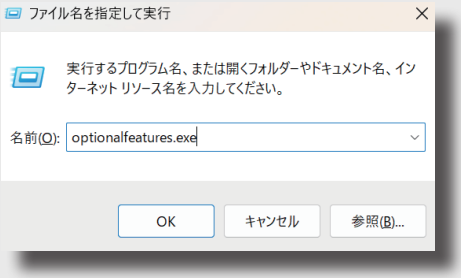
- Windows 10 or 11 Pro / Enterprise / Education（Windows 10 はビルドバージョン 1903 以降、Home はサポート対象外）
- ARM64（Windows 11 バージョン 22H2 以降）または AMD64 アーキテクチャー
- BIOS で有効化された仮想化機能
- 少なくとも 4 GB の RAM（8 GB 推奨）
- 空きディスク領域 1 GB 以上（SSD を推奨）

・少なくとも 2 つの CPU コア (ハイパースレッディングを使用した 4 コアを推奨)

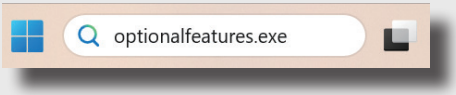
【参考 URL】
https://learn.microsoft.com/ja-jp/windows/security/application-security/application-isolation/windows-sandbox/windows-sandbox-install

「Windows サンドボックス」は以降の手順で利用可能となります。
すでに設定をしている方は不要となります。なお、前提条件を満たしていない場合は、本ハンズオンの実施は行なわず、内容の理解のみに留めてください。

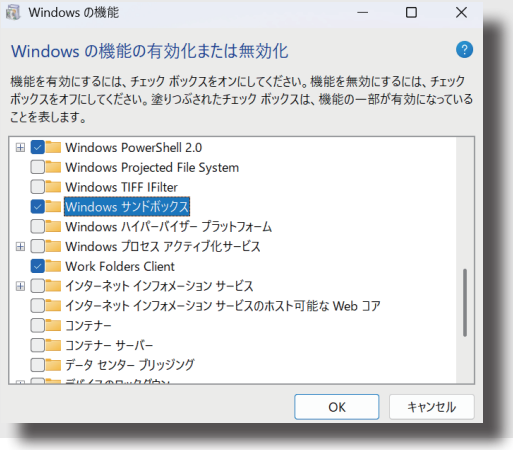
Windows キー + R キーで「ファイル名を指定して実行」ダイアログボックスを起動し、「optionalfeatures.exe」を入力します。



または、Windows 左下の検索バーに「optionalfeatures.exe」と入力しても起動できます。



Windows の機能ダイアログボックスが開いたら、「Windows サンドボックス」にチェックを入れ、「OK」を押下します。



「Windows サンドボックス」のインストールが完了したら、コンピューターを再起動します。再起動後、スタートプログラムより、「Windows サンドボックス」を起動してください。

2.2 Thunderbird の準備

Thunderbird は、Mozilla が提供する無償のメールソフトです。下記 URL から公式サイトにアクセスしてダウンロードしてください。

【ダウンロード URL】
https://www.thunderbird.net/ja/

ダウンロードされた Thunderbird Setup xx.x.x.exe をダブルクリックします。
もしユーザーカウント制御 (UAC) 画面が表示されたら「はい」を押下します。
また、セットアップウィザードでセットアップの種類を聞かれた際は「標準インストール」を選択し、インストールをクリックしてください。
セットアップ後、Thunderbird が起動したら導入完了です。

2.3 サクラエディタの準備

下記のサイトから、最新版のインストーラーをダウンロードします。

【ダウンロード URL】
https://github.com/sakura-editor/sakura/releases

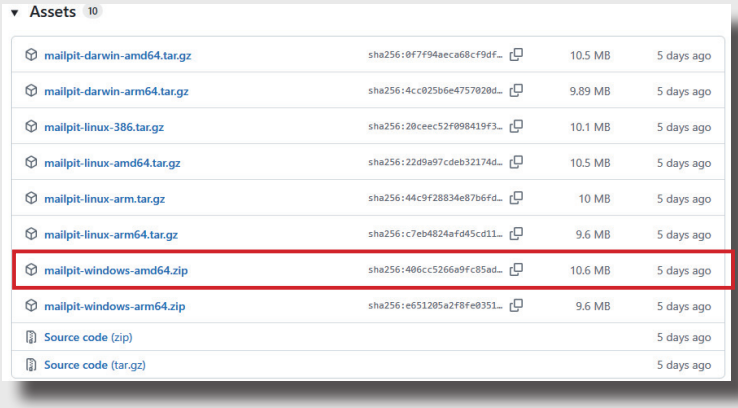
ローカルにダウンロードされた ZIP ファイルを展開し、インストーラーを実行してください。選択箇所は初期選択のままで問題ありません。

2.4 Mailpit の準備

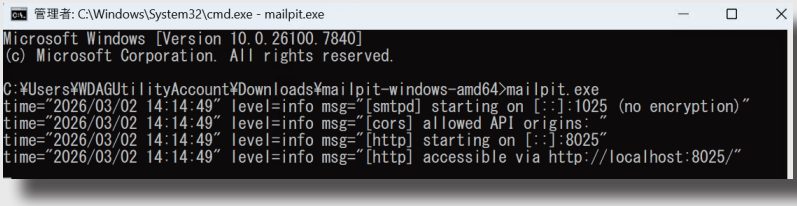
Mailpit は開発・検証環境向けの軽量なメールテストサーバーです。SMTP サーバーとして動作し、受信したメールを WebUI で確認することや、ローカルに保存するなどの操作が可能です。

下記のサイトから最新版のインストーラーをダウンロードします。

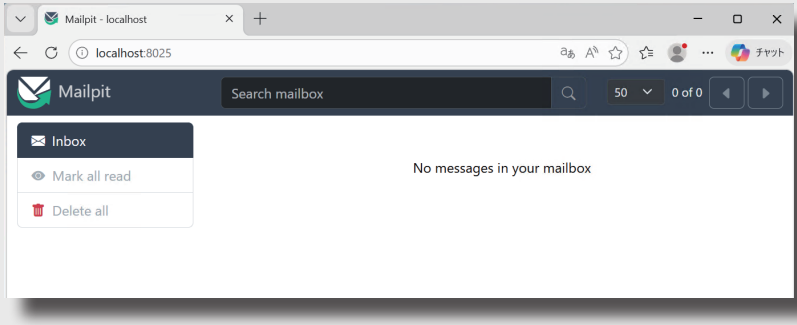
【ダウンロード URL】
https://github.com/axllent/mailpit/releases/



今回は、mailpit-windows-amd64.zip をダウンロードし、ファイルを展開します。コマンドプロンプトを起動し、展開されたフォルダ上の mailpit.exe を実行してください。下記のような画面になれば成功です。



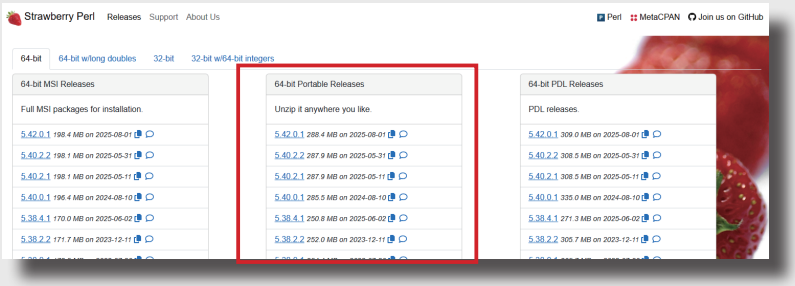
Edge を起動し、アドレスバーに localhost:8025 と打ち込むと WebUI が確認できます。



2.5 Swaks の準備

Swaks（Swiss Army Knife for SMTP）はコマンドラインでメールを送信することができるテストツールです。今回はこのツールを用いてメール作成のハンズオンを行ないます。
まず、以下の公式サイトから最新版の Strawberry Perl（Portable 版）のダウンロードを行なってください。

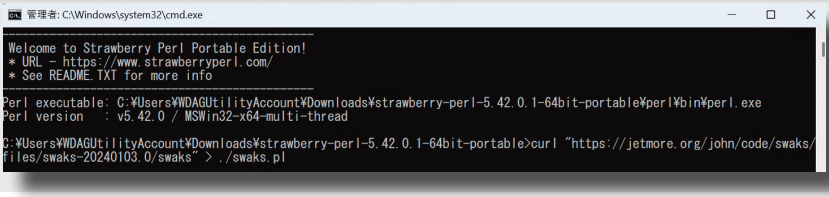
【ダウンロード URL】
https://strawberryperl.com/releases.html



ダウンロードした ZIP ファイルを任意のフォルダで展開し、portableshell.bat を実行します。セキュリティの警告が出た場合は、実行するファイルに間違いがないことを確認したうえで、実行ボタンを押下します。

portableshell.bat を実行して開いたコマンドプロンプト上で以下のコマンドを実行し、Swaks を導入します。

```
curl "https://jetmore.org/john/code/swaks/files/swaks-20240103.0/swaks"
> ./swaks.pl
```



3. 攻撃者が活用する手口と解析手法

不審メールを伴う攻撃の多くは、ソーシャルエンジニアリングのテクニックが悪用されています。ソーシャルエンジニアリングとは、人間の心理的な隙や行動のミスを突いて情報を盗み出す手法です。不審メールにおいては、「受信者に不信感を与えず行動させる」、「考えさせる隙を与えず、行動を強制させる」といったことを意識した内容になっているものがしばしばみられます。本稿では攻撃者が不審メールを用いるソーシャルエンジニアリング攻撃の手口について整理し、手口の特徴により関係のある手口をまとめ、それぞれハンズオンを通してどのように解析を行なうかについて理解を深めます。

3.1 添付ファイルを悪用した攻撃の手口

攻撃者は、ランサムウェア等のマルウェア感染を狙ってメールに不審なファイルを添付することがあります。この時、メールサーバーなどで不正プログラム検出を回避するため、添付ファイルにも偽装や回避の手口を活用します。代表的な偽装や回避の手口は以下のとおりです。

3.1.2 パスワード付き ZIP ファイル

ファイルを暗号化することでセキュリティ製品のシグネチャ検知を素通りさせ、ユーザーの手元に届いてから自身で展開させ、実行させる手口です。

3.1.3 二重拡張子

Windows では規定で拡張子が非表示設定になっていること、exe ファイルはアイコンを任意に変更可能なこと、この 2 つの特徴を悪用し、攻撃者は悪意のあるファイルを無害なものに見せかける偽装を行なう場合があります。例えば、「重要書類 .pdf.exe」のようなファイル名の場合、ユーザーの画面上では末尾の .exe が非表示となり、「重要書類 .pdf」と表示されます。また、一目ではわからないように巧妙にアイコンを偽装しているケースが数多くあります。

3.1.4 ショートカット（LNK）ファイルの悪用

一見するとただのショートカットファイルに見えますが、実際には不正なスクリプトやコマンドを実行させる目的のファイルを設置する手口です。ショートカットファイルのリンク先に PowerShell や cmd.exe などのコマンドを埋め込むことで、ファイルを開いた瞬間に外部サイトからマルウェアなどをダウンロード・実行させることがあります。

3.1.5 マクロ付き Office ファイルの悪用

請求書や注文書等に見せかけた、悪意あるマクロを含む Office ファイルを添付する手口です。ファイルを開くと、「コンテンツの有効化」や「マクロを有効にする」といったバーが表示されます。不審なものであると気づかずに有効化ボタンを押下してしまうと攻撃者が用意した悪意のあるプログラムが実行されてしまいます。

3.1.6 イメージファイルの悪用

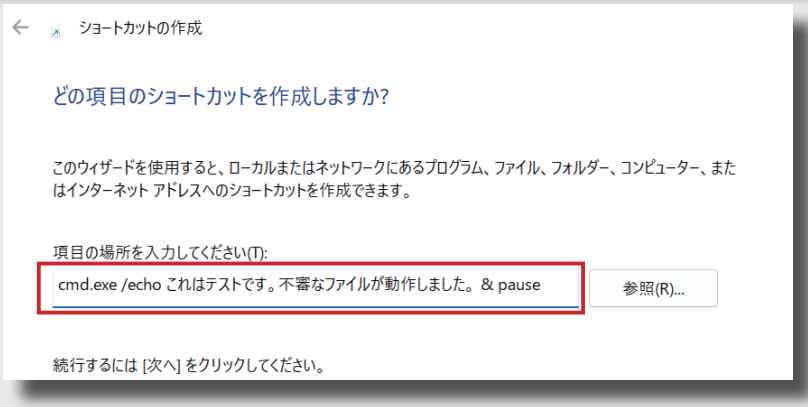
ISO 形式などのイメージファイルを悪用する手口です。暗号化 zip 同様に、セキュリティ製品を回避する目的で用いられる手口です。Windows では、イメージファイルをダブルクリックすると仮想ドライブとして自動マウントされるため、内部にマルウェアなどを隠し、感染を狙うことがあります。

3.2 ハンズオン：不審なショートカットファイルの調査

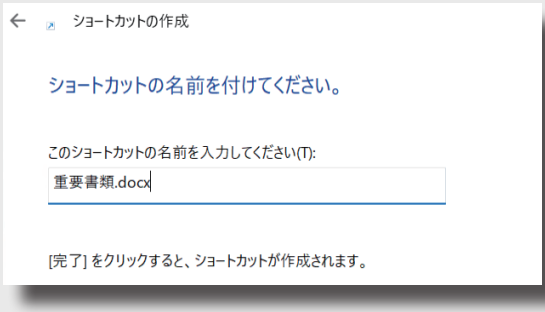
デスクトップに右クリックして新規作成からショートカットを選択します。

項目の場所に以下のテキストをコピー＆ペーストしてください。

```
cmd.exe /c echo これはテストです。不審なファイルが動作しました。 & pause
```

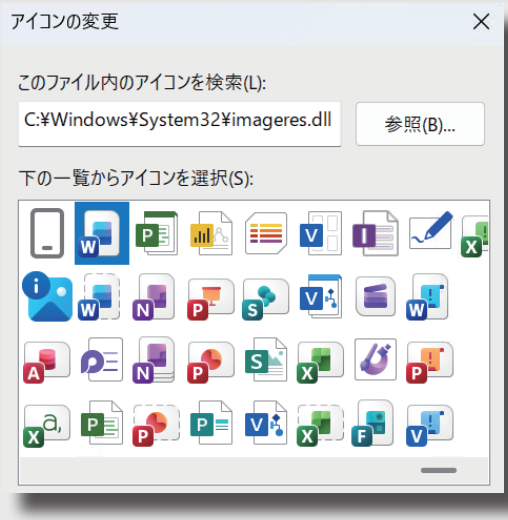


次へ進み、名前を「重要書類 .docx」にして完了を押下します。

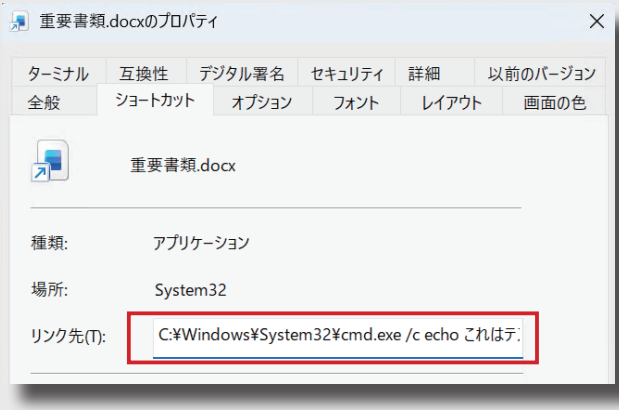


作成されたショートカットを右クリックし、プロパティからアイコンの変更を選択、アイコンを任意のものに変更します。下記ファイルを参照することでより多くのアイコンを使用することができます。

```
C:¥Windows¥System32¥imageres.dll
```

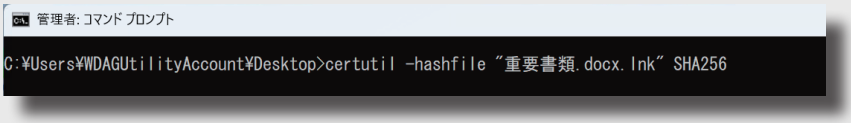


作成した不審なファイルがメールに添付されて送られてきた状況を想定して簡易的な解析を行ないます。
まず、このファイルを右クリックしてプロパティを確認すると、リンク先のパスや設定を確認することができ、今回は cmd.exe を実行する設定になっていることが分かります。



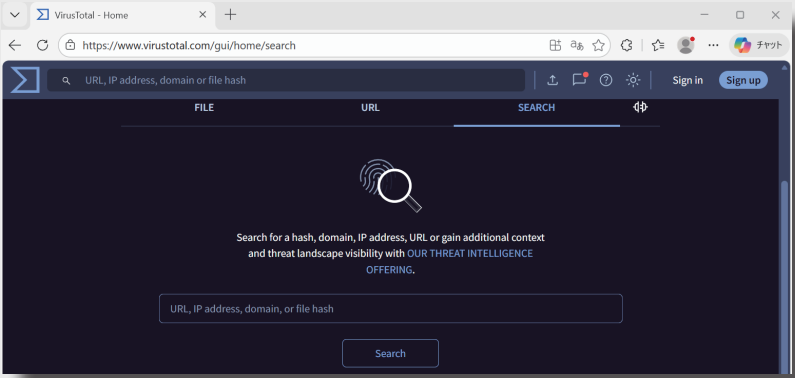
ファイルが既知の脅威である場合、ハッシュ値を確認して VirusTotal などに情報がアップロードされていないか検索することもできます。
コマンドプロンプトを開き、以下のコマンドを入力してください。

```
certutil -hashfile “ファイルのパス” SHA256
```



指定した形式のハッシュ値が出力されます。

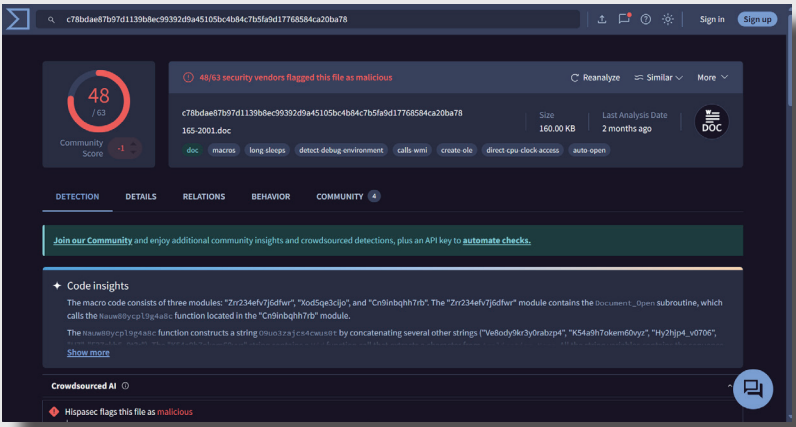
このハッシュ値を VirusTotal などで確認して不審なファイルであるかを確認してください。



このファイルは学習用のダミーファイルであるため検知結果は表示されませんが、既知の脅威がアップロードされていた場合は、その情報をもとに調査を進めることができます。

例として、実際の検体のハッシュ値を掲載しますので、こちらを VirusTotal で確認してください。DETECTION タブで複数のベンダーのアンチウイルスソフトで検知されていることが確認できます。

SHA-256：
c78bdae87b97d1139b8ec99392d9a45105bc4b84c7b5fa9d17768584ca20ba78



なお、検知結果についてはあくまで目安であり、検知されていないことが確実に安全であるとは限らないことに注意してください。

3.3 リンクの隠ぺい・誘導工作

ユーザーをフィッシングサイトやマルウェア配布サイトへ誘導することを目的として、不審なサイトへの URL を怪しいと気づかせないために使用する手口を紹介します。

3.3.1 表示文字列とリンク先の不一致

HTML メール形式を利用し、画面上の見かけの URL と実際の遷移先を別に設定します。

3.3.2 短縮 URL サービスの悪用

Bitly や tinyURL など、正規の短縮 URL サービスを利用して、遷移先の悪意あるドメインを隠ぺいします。

3.3.3 QR コードの悪用

PC 上のセキュリティ製品がメール本文のテキストや URL スキャンを行なうことがある一方、画像の中のデータまでスキャンしないことを悪用した攻撃です。「多要素認証の再設定が必要です」などの名目でメール本文に QR コードを張り付け、スマートフォンで読み取らせ、不審なサイトを開かせようとします。

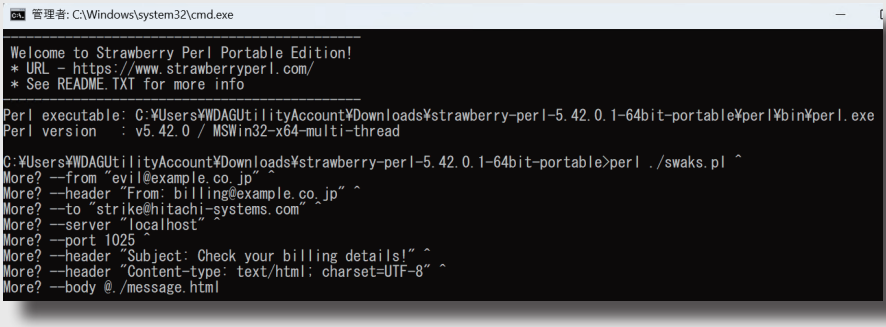
3.4 ハンズオン：不審な HTML メールの調査（表示文字列とリンク先の不一致）

環境構築で準備を行なった portableshell.bat が存在するフォルダでサクラエディタを使って、message.html という名前で html ファイルを作成し、以下の内容を記載します。なお、デファング ([:]、[.]) している箇所は [ ] を外して記載してください。

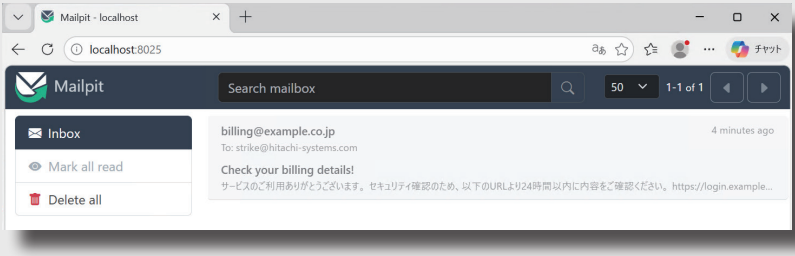
```
<html>
<body>
<p> サービスのご利用ありがとうございます。</p>
<p> セキュリティ確認のため、以下の URL より 24 時間以内に内容をご確認ください。</p>
<p><a href= https[:]//www[.]hitachi-systems[.]com/report/specialist/hj/index[.]html >https[:]//
login[.]example[.]co[.]jp</a></p>
<p> 期限内にご対応いただけない場合、アカウントが制限される可能性があります。</p>
</body>
</html>
```

次に、Swaks を導入したコマンドプロンプト上で以下のコマンドを入力します。

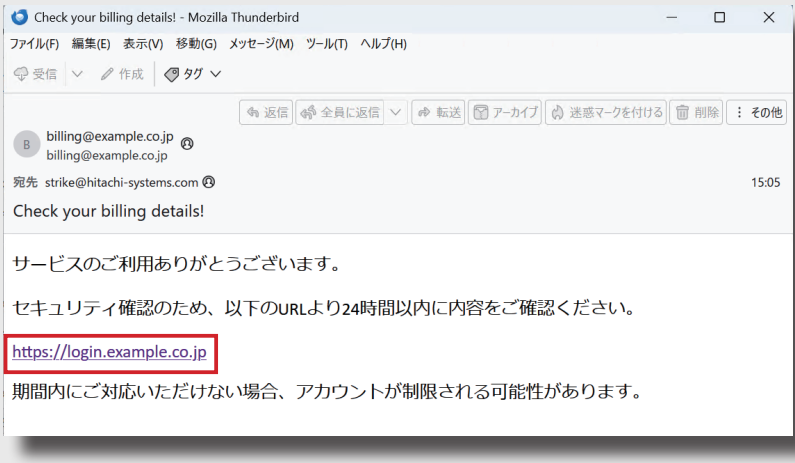
```
perl ./swaks.pl ^
--from “evil@example.co.jp” ^
--header “From: billing@example.co.jp” ^
--to “strike@hitachi-systems.com” ^
--server “localhost” ^
--port 1025 ^
--header “Subject: Check your billing details!” ^
--header “Content-type: text/html; charset=UTF-8” ^
--body @./message.html
```



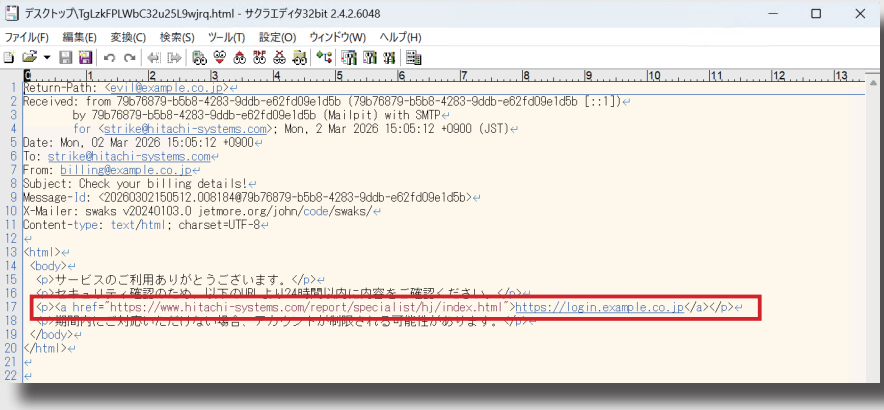
問題なく上記のコマンドでメールが送付できていれば、Mailpit の WebUI 上にメールが届いているのが確認できます。



作成した不審メールをダウンロードします。WebUI のダウンロードボタンをクリックすると保存形式の選択ができるので、「Raw message」を選択して .eml ファイルをダウンロードしてください。保存した .eml 形式のファイルを Thunderbird で開いて確認してみると、https[:]//login[.]example[.]co.jp へのリンクが本文に貼られているように見えます。



作成した不審メールを解析します。「ファイル > 名前を付けて保存 > ファイル」から、ファイルの種類を「すべてのファイル」に設定し、ファイル名の末尾に「.html」を付けて保存してください。保存したファイルをサクラエディタで開き、本文内の URL を確認してください。



メーラーから確認すると、表示されているリンクは https[:]//login[.]example[.]co.jp ですが、実際の遷移先は https[:]//www[.]hitachi-systems[.]com/report/specialist/hj/index[.]html であることが分かります。

3.5 その他の手口

3.5.1 送信元の偽装

メールの差出人（送信元情報）は攻撃者が自由に書き換えることができます。そのため、EC サイトやカード会社、銀行などのサービス名やメールアドレスを騙ってメールを送付し、偽サイトに誘導することで ID やパスワード、クレジットカード情報などを盗もうとする攻撃が行なわれています。

3.5.2 タイポスクワッティング

正規のドメイン名と非常によく似た文字列のドメインを用いる手法です。文字の入れ替えや余分な文字の追加、削除などを行ない、メールの送信元や本文内に記載した URL がぱっと見では不審であると気が付かないようにします。

例：evil@ex**am**ple.com → evil@ex**ma**ple.com（文字の入れ替え）

3.5.3 ホモグラフ攻撃

アルファベットと見た目がほぼ同じ多言語文字等を使って、正規のドメインに見せかける手法です。見た目上は正規の URL と区別がつかないため、発見が困難です。

例：evil@exampl**e**.com → evil@ex ａ mp**1**e.com（アルファベットを数字に）

3.5.4 緊急性をあおる

「至急対応が必要」、「1 時間以内に処理してください」など、判断時間を与えない文面を使う手法です。冷静な確認をさせないことで、リンクの確認や内容の妥当性チェックを省略させる

狙いがあります。緊急性と同時に、不安や恐怖（アカウント停止、支払い遅延、不正アクセスなど）を組み合わせることが多く見られます。

3.5.5 返信スレッドの偽装（Re: 先日の件 など）

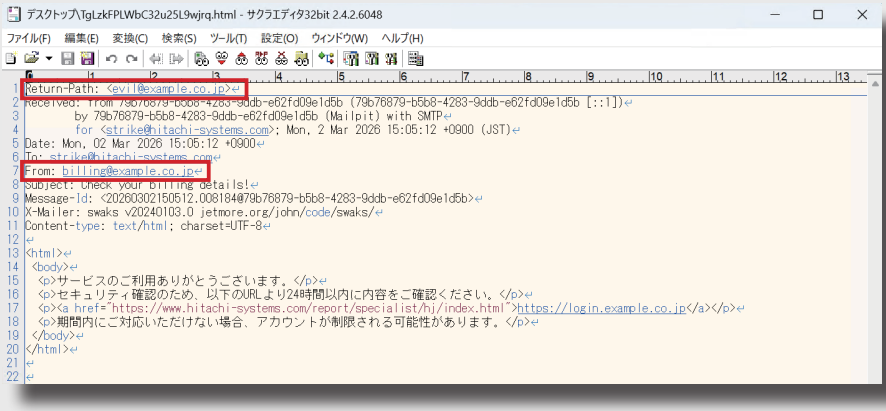
過去にやり取りがあったように見せかけるため、件名に「Re:」「Fw:」を付ける手法です。実際には初めてのメールであっても、受信者に「自分が忘れているだけかもしれない」と思わせる効果があります。業務上のやり取りが多い環境ほど、この手法は有効に働きます。

3.5.6 権威の悪用

社長、上司、管理部門、取引先など、立場の強い人物や組織を名乗ることで、受信者に従わせようとする手法です。特に「自分だけに送られている」「内密に対応してほしい」といった表現と組み合わされることがあります。

3.6 ハンズオン：不審な HTML メールの調査（送信元の偽装）

「3.4 ハンズオン」で作成した不審メールを本ハンズオンでも利用します。まず、作成した不審メールを WebUI 上で確認すると、メールの送信元が evil@example.co.jp ではなく、billing@example.co.jp になっていることが分かります。また、ダウンロードした .eml ファイルをサクラエディタで開くとヘッダー情報が確認できます。Return-Path に evil@example.co.jp のアドレスが記載されており、From のアドレスと異なるなど、ヘッダーの情報から不審な点がないかを確認することができます。



3.7 メールソフト自体のぜい弱性をついた攻撃

利用しているメールソフト自体にぜい弱性が存在する場合、攻撃者はそのぜい弱性をついて任意の行動を起こすことが出来る可能性があります。

例えば、過去に確認された事例では、特定のコードを含んだメールを送信するだけで、受信者がメールを開封しなくてもメールソフトがそのリマインダー通知を鳴らそうとする処理の裏で認証情報を外部へ盗み出すことができた事例があります（CVE-2023-23397）。

5. おわりに

今回はここまでとなります。「1. 攻撃手口確認編」では解析環境の準備と攻撃者の手口を紹介しました。また、攻撃者の手口を実際にハンズオンで試し、各手口に関する解析の方法についても触れました。

次回、「2. メール解析手順編」では解析の手順を整理し、今回紹介した攻撃の手口を使いながら不審メールを作成し、実際に解析を進めていきます。

Hitachi Systems Security Journal

株式会社 日立システムズ　本社：〒141-8672　東京都品川区大崎 1-2-1　www.hitachi-systems.com

お問い合わせは

Printed in Japan