TLP：WHITE/CLEAR

サイバーデセプション

―攻撃者心理を逆手に取る能動的防御戦略―

2026年6月
独立行政法人情報処理推進機構（IPA）
産業サイバーセキュリティセンター
第9期中核人材育成プログラム
卒業プロジェクト

まえがき

サイバー攻撃は、日々高度化・巧妙化を続けています。ランサムウェア、標的型攻撃、認証情報の悪用、正規ツールを用いた侵害活動などにより、企業・組織は従来にも増して厳しい防御を求められています。境界防御、EDR、SIEM、ゼロトラストなどの対策は、いずれも重要であり、今後もセキュリティの基盤であり続けます。しかし、それらを積み重ねるだけでは、常に進化する攻撃者に対して、防御側が後追いになり続ける構造を完全に変えることはできません。

本プロジェクトは、そのような問題意識から出発しました。

私たちが注目したのは、攻撃者の技術そのものだけではありません。攻撃者が何を見て、何を信じ、どのように判断し、どのように行動するのかという「認知・判断・行動」のプロセスです。攻撃者もまた、限られた時間、情報、リソースの中で判断しています。そこには、思い込み、過信、効率化への欲求、確認不足といった判断上の隙が存在します。

サイバーデセプションは、この攻撃者の認知や判断に働きかけることで、防御側に有利な状況を作り出す能動的な防御戦略です。偽のサーバー、偽の認証情報、偽のデータをただ配置することが目的ではありません。攻撃者に何を見せ、何を信じさせ、どこへ誘導し、どの行動を検知・遅延・観測するのかを、防御側が意図を持って設計・運用することに本質があります。

一方で、日本国内において、サイバーデセプションはまだ十分に普及しているとはいえません。ハニーポット等の個別技術として知られていても、それをどのようにセキュリティ戦略へ組み込み、設計・運用へ落とし込み、既存の防御策と連携させるのかについては、実務で活用できる形で整理された情報が限られています。そこで私たちは、サイバーデセプションを単なる製品や技術の導入論としてではなく、戦略・作戦・戦術・運用を一貫して接続する防御設計として体系化することを目指しました。

本書では、サイバーデセプションの必要性や基本概念、攻撃者心理との関係、期待される効果と限界を整理したうえで、サイバーデセプションを組織のセキュリティ戦略に組み込むための考え方を示します。また、MITRE Engage や本書で提案する CYDEC Design Framework（CDF）を用いて、サイバーデセプションを構想から設計・運用へ落とし込む方法を整理し、具体的な設計例と検証結果を通じて、その有効性と留意点を考察します。

本書は、研究機関や特定分野の専門家のみを対象としたものではありません。経営層、CISO、セキュリティ責任者、IT/OT の運用担当者、法務・広報・リスク管理部門など、サイバーデセプションの導入や活用に関わる幅広い方々に向けて作成しています。本書を通じて、読者の皆さまが「サイバーデセプションとは何か」を理解するだけでなく、「自組織ではどのように活用でき

るのか」「既存の防御策とどのように組み合わせられるのか」「攻撃者に対して防御側の選択肢をどのように広げられるのか」を考えるきっかけになれば幸いです。

攻撃側の進化を追いかけるだけでなく、防御側から攻撃者の認知・判断・行動に働きかける。本書が、その新たな防御の可能性を考える一助となることを願っています。

２０２６年６月吉日
梅雨晴れの文京の地にて
プロジェクトリーダー　佐藤将大

目次

【コラム】

本書では、デセプションをより身近に捉えるため、各章の終わりにコラムを入れている。国民性、戦史、日常や娯楽の中にある誘導の工夫などを整理した。技術論だけでは見えにくい「人が何を信じ、なぜ行動するのか」という視点から、デセプションの世界に関心を持っていただきたい。

本書の概要

サイバー攻撃は高度化・巧妙化を続けており、境界防御、脆弱性対策、EDR、SIEM などの従来型の対策を重ねるだけでは、攻撃者優位の状況を十分に変えられない場面が増えている。いま求められているのは、攻撃を止める・見つけるだけでなく、攻撃者の認知・判断・行動に働きかけ、防御側に有利な状況を作り出すという、視点を転じた防御の考え方である。

本書が取り上げる「デセプション」とは、単なる偽装や罠ではない。相手に何を見せ、何を見せず、どのような判断へ導くかを設計することで、相手の行動を防御側に有利な方向へ変化させる考え方である。戦争、外交、諜報、ビジネス、マジックなどで用いられてきた欺瞞の概念を整理したうえで、その考え方をサイバー空間に応用するのがサイバーデセプションである。

サイバーデセプションは、ハニーポット、ハニーネット、ハニートークン、偽データなどの技術を用いて、攻撃者を早期に検知し、迷わせ、遅延させ、誘導し、その行動を観測する。重要なのは、これらを個別技術の寄せ集めとして見るのではなく、攻撃者が偵察、侵入、権限昇格、横展開、目的の実行などの各段階において、どのように防御側の優位を作るかという観点で捉えることである。『Cyber Deception: Fundamental Guide』[1]においても、サイバーデセプションは受動的な監視にとどまらず、攻撃者の行動を観測・かく乱し、脅威インテリジェンスや SOC 運用に接続する考え方として整理されている。

本書では、この考え方を実務に適用するための設計枠組みとして、CYDEC Design Framework（CDF）を提案する。CDF は、守るべき資産、想定する攻撃者、攻撃フェーズ、配置場所、技術要素、運用体制を対応づけ、サイバーデセプションを戦略・作戦・戦術の観点から一貫して整理するための枠組みである。これにより、サイバーデセプションを単なる製品・技術の導入ではなく、目的に基づく防御設計として検討できるようにする。

また、本書では、サイバーデセプションが攻撃者の認知・判断・行動にどのように作用し得るかを確認するため、3つの検証を実施した。具体的には、攻撃者の探索コストを増大させる効果、ハニーファイルのメタデータ品質が真正性判断に与える影響、偽認証情報による次行動の誘発について検証した。これらの検証を通じて、サイバーデセプションは、適切に設計・運用されることで、早期検知、攻撃遅延、誘導、攻撃手法の観測といった効果を発揮し、従来型の対策を補完する有効な選択肢となり得ることを確認した。

一方で、サイバーデセプションは導入すれば自動的に効果を発揮するものではない。現実味のある設計、適切な配置、安全な隔離、ログ取得、既存のセキュリティ製品や SOC・CSIRT との連携、運用負荷、法務・倫理面の整理が不可欠である。本書の総括では、これらの検証結果を踏まえ、経営層に向けた示唆、導入・運用上の留意点、今後の課題と展望を提示し、サイバーデセプションを実践的な防御戦略として活用するための土台を示す。

本書におけるキーワード

本書を読み進めるうえで重要なキーワードを挙げる。その他の用語については巻末の「用語集」にて記載する。

用語	説明
能動的防御	攻撃を受けてから受動的に対応するだけでなく、防御側が攻撃者の行動を予測し、検知、誘導、かく乱、情報収集などを主体的に行う防御の考え方。
能動的サイバー防御	国家政策や法制度の文脈で用いられることが多い概念。サイバー対処能力強化法においては、通信の無害化措置として、攻撃が疑われる通信を重要インフラ事業者が検知・ブロック、もしくは政府機関が通信事業者を通じて無害化を行うことを指す。本書では、特に断りがない限り、企業ネットワーク内での主体的な防御活動を指す「能動的防御」とは区別して扱う。
デセプション（欺瞞）	相手の認知・判断・行動に影響を与え、自らに有利な結果を導くための考え方であり、単なる虚偽ではなく、相手が何を見て、何を信じ、どう行動するかを設計する行為を指す。
サイバーデセプション	デセプションの考え方をサイバー空間に応用し、攻撃者を検知、誘導、遅延、かく乱、観測するための防御手法。ハニーポットなどの個別技術だけでなく、攻撃者の意思決定に作用する防御思想を含む。
MITRE Engage	防御側が攻撃者とどのように関与し、検知、誘導、かく乱、情報収集を行うかを体系化したフレームワーク。サイバーデセプションやサイバー拒否を含むアドバーサリ・エンゲージメントの設計に用いる。
アドバーサリ・エンゲージメント	防御側が攻撃者の行動に能動的に関与し、拒否、欺瞞、誘導、観測などを通じて防御側に有利な状況を作る考え方。MITRE Engage の中心概念
CYDEC Design Framework（CDF）	本書で提案する、サイバーデセプション設計の整理枠組み。Who、What、by Whom、Why、When、Where、How の 6W1H により、前提、目的、実装を一貫して整理する。

著作権及び利用条件

免責事項

本書は、サイバーデセプションに関する調査研究の成果を情報提供の目的で整理したものであり、特定の製品、サービス、技術、対策の導入または利用を推奨・保証するものではない。

本書の内容は、作成時点における情報、文献調査、検証結果および本プロジェクトの考察に基づくものであり、サイバー攻撃の手法、技術動向、法制度、標準化動向、製品仕様、運用環境などの変化により、予告なく変更される場合がある。

本プロジェクトは、本書の内容について、正確性、完全性、有用性、最新性、商品性、特定目的への適合性その他いかなる事項についても、明示または黙示を問わず一切の保証を行わない。

本書の利用により生じた直接的、間接的、付随的、特別、派生的または結果的損害、ならびに利益の損失、事業上の損失、データの喪失、システム障害その他一切の損害について、本プロジェクト、本書の著者、関係者および監修者は、法的原因の如何を問わず一切の責任を負わない。

本書には、外部資料、Web サイト、ツール、製品、サービスなどへの参照またはリンクが含まれる場合があるが、本プロジェクトは、それらの内容、正確性、継続性、安全性、適法性その他一切について責任を負わない。

本書の有効期限は、発行日から 2 年間とする。ただし、有効期限内であっても、技術動向、脅威環境、法制度、標準化動向などの変化により、内容が実態に適合しなくなる可能性がある。本プロジェクトは、本書の内容を常に最新の状態に更新・修正する義務を負うものではなく、実態との乖離によって生じたいかなる損害についても責任を負わない。

注意事項

本書は、本プロジェクトの調査研究および検証結果に基づいて作成されたものであり、独立行政法人情報処理推進機構（IPA）ならびに本プロジェクトメンバーの所属組織の公式見解を示すものではない。

本プロジェクトにおける調査・検証の対象期間は、2026 年 4 月から 2026 年 6 月末までである。

本書で取り上げる検証結果は、限られた期間、環境、条件のもとで実施したものであり、すべての組織、システム、攻撃者、運用環境にそのまま適用できるものではない。

実際にサイバーデセプション技術を導入・運用する場合は、各組織のシステム構成、保護すべき資産、想定する脅威、リスク許容度、法務・コンプライアンス要件、運用体制などを踏まえ、個別に検討する必要がある。

本書では、文章の構成検討、表現の推敲、図表・挿絵の作成補助に、ChatGPT 等の生成 AI サービスを一部利用している。生成 AI の出力は、執筆者が確認・編集したうえで使用しており、本書の内容に関する最終的な責任は執筆者に帰属する。

第1章　導入

1.1 本書の作成背景

サイバー攻撃は、高度化・巧妙化に加え、長期にわたり標的組織内で活動を継続する傾向を強めている。企業・組織は、ファイアウォール、IDS/IPS、EDR、SIEM、脆弱性管理、アクセス制御など、さまざまなセキュリティ対策を重ねてきた。しかし、ランサムウェア、標的型攻撃、認証情報の悪用、正規ツールを用いた侵害活動などにより、従来型の境界防御や検知中心の対策だけでは、被害を十分に抑えきれない状況が続いている。

この背景には、防御側と攻撃側の構造的な不均衡がある。防御側は、多数のシステム、端末、ネットワーク、クラウド環境、利用者、委託先などのすべての面を継続的に守る必要がある。一方で、攻撃者は一つの脆弱性、設定不備、認証情報、運用上の隙を見つければ、侵入や横展開の足掛かりを得ることができる。防御側が新たな技術を導入すれば、攻撃側もそれを回避・突破する手法を進化させる。このような「技術 対 技術」のいたちごっこは、セキュリティ対策の高度化と同時に、運用負荷やアラート対応の増大も招いている。

こうした状況の中で、攻撃を単に防ぐ、検知するだけでなく、攻撃者の認知や判断そのものに働きかける「デセプション（欺瞞）」という考え方に本プロジェクトでは注目した。デセプションは、攻撃者に偽の情報や環境を見せ、迷わせ、攻撃を遅延させ、誘導し、行動を観測することで、防御側に有利な状況を作り出す考え方である。サイバー領域においては、ハニーポット、ハニーネットなどの技術を通じて、攻撃者の行動を早期に捉え、重要資産への到達を遅らせ、攻撃手法に関する情報を得ることが期待される。

一方で、日本国内では、サイバーデセプションに関する公開情報や実務知見はまだ限定的である。ハニーポットのような個別技術として知られることはあっても、サイバーデセプションを防御戦略や運用設計の一部として体系的に理解し、企業・組織のセキュリティ対策にどう組み込むべきかについては、十分に整理されているとは言い難い。

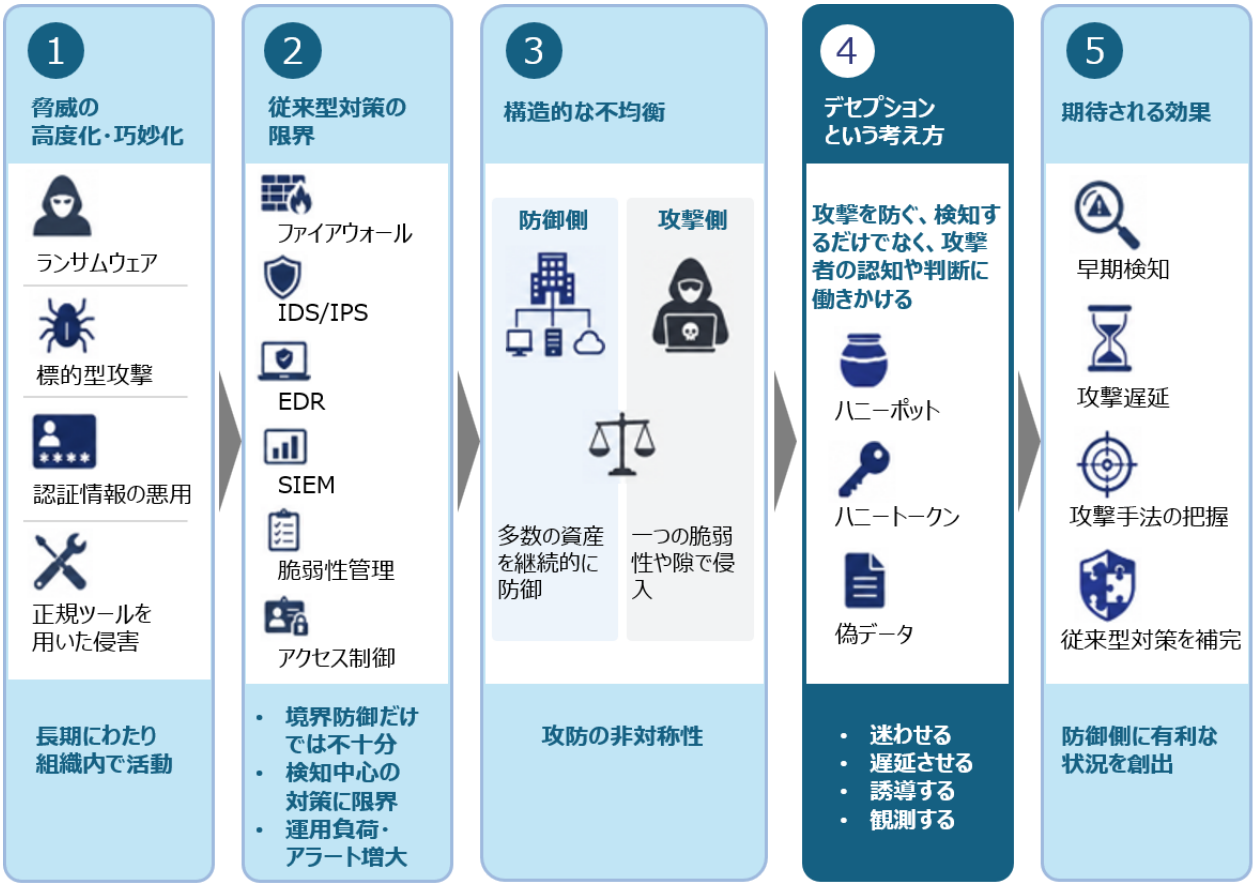


図 1-1　サイバーセキュリティに対するデセプションの視点

そこで本書では、デセプションの基本概念から出発し、サイバーセキュリティにおける位置づけ、技術類型、運用上の考慮点、検証を通じた有効性と特徴を整理する。これにより、サイバーデセプションを単なる技術ではなく、従来型の対策を補完し、防御側が攻撃者に対してより能動的に働きかけるための考え方として提示し、企業・組織における理解促進と導入検討の土台を提供する。

1.2 本書の目的

本書の目的は以下のとおりである。

- サイバーデセプションの必要性や有効性を整理し、企業・組織における導入検討のきっかけを提供する。
- サイバーデセプションの基本概念、技術類型、期待される効果と限界を体系的に整理し、企業・組織内の共通理解を形成する。
- サイバーデセプションの戦略的な導入方法、既存対策との連携、適用上の留意点を整理し、実務に適用可能な形で提示する。

1.3 本書の想定読者

本書は、研究機関や特定分野の専門家のみを対象としたものではなく、サイバーデセプションの必要性や活用方法について、幅広い企業・組織に理解いただくことを想定している。

【意思決定層】

- 経営層、事業責任者、CISO など：
  サイバーデセプションの必要性・有効性を理解し、導入判断や全社方針（投資・優先順位を含む）を決定する方

【設計・運用方針層】

- セキュリティ責任者など：
  サイバー攻撃の脅威に対し、セキュリティ戦略の中にデセプションをどのように位置付け、組み込むかを検討したい方
- 情報管理・業務部門の担当者など：
  自組織が保有する情報資産や業務データの重要度を踏まえ、どのように情報を守るべきかを検討したい方

【実装・運用層】

- IT/OT の運用担当者など：
  ハニーポットなどのサイバーデセプション技術を、運用の中でどのように活用できるかを把握したい方

【横断的な関係者】

- 法務・広報・リスク管理の担当者など：
  デセプション導入に伴う法的・倫理的な留意点、社内外への説明、公表判断、インシデント対応時の影響を把握したい方

1.4 本書の構成

本書は全 9 章で構成される。各章の概要は以下のとおりである。

第 1 章「導入」では、本書の作成背景、目的、想定読者を示し、サイバーデセプションを取り上げる意義と本書の位置づけを明確化する。

第 2 章「能動的防御へのパラダイムシフト」では、従来型のセキュリティ対策が抱える課題を整理し、サイバーレジリエンスの観点から、攻撃者に能動的に働きかける防御の必要性を示す。

第 3 章「デセプション」では、デセプションの一般概念を整理し、軍事やマジックなどにおける活用例、欺瞞フレームワークを通じて、相手の認知・判断・行動に影響を与える仕組みを示す。

第 4 章「サイバーデセプション」では、デセプションの考え方をサイバー空間に適用し、サイバーデセプションの基本概念、セキュリティ戦略に及ぼす効果、攻撃者心理から見た有用性を整理する。サイバーデセプションを、単なるおとり技術ではなく、攻撃者の意思決定に作用する防御戦略として捉えるための基礎を示す。

第 5 章「サイバーデセプションの構想と展開」では、サイバーデセプションを組織のセキュリティ戦略に組み込むための考え方を整理する。戦略・作戦・戦術の関係、MITRE Engage との対応を確認したうえで、本書独自の設計整理として CYDEC Design Framework（CDF）を提示し、目的、脅威、攻撃フェーズ、配置場所、手段を一貫して整理する視点を示す。

第 6 章「サイバーデセプションの設計・運用」では、CDF を踏まえ、サイバーデセプションを実環境に適用するための設計・運用上の考慮点を整理する。基本設計、詳細設計、実装・運用、評価・改善の流れに沿って、サイバーデセプションを継続的に機能する防御策として展開する方法を示す。

第 7 章「企業・組織における導入の考え方と設計例」では、具体的な導入の考え方と設計例を示す。想定する脅威、守るべき資産、攻撃段階、配置場所、技術要素を対応づけることで、サイバーデセプションを構想から設計へ落とし込む流れを具体化する。

第 8 章「サイバーデセプションの検証」では、検証環境および前提条件を整理したうえで、3 つの検証とその結果を示す。また、検証結果を横断的に分析し、サイバーデセプションが有効に機能する条件、実務導入時の留意点、および検証上の限界を考察する。

第 9 章「総括」では、本書全体の議論と検証結果を踏まえ、サイバーデセプションの意義を改めて整理する。あわせて、経営層に向けた示唆、導入・運用における留意点、今後の課題と展望を示す。

本書は、第 1 章から順に読み進めることで、サイバーデセプションの背景、概念、戦略上の位置づけ、設計・運用を体系的に理解できる構成となっている。

ただし、読者の役割や関心に応じて、以下のような読み方も可能である。

- 経営層、事業責任者、CISO など：
  第 1 章、第 2 章、第 5 章、第 9 章を中心に読むことで、サイバーデセプションの背景、必要性、および本書の要点を把握できる。
- セキュリティ責任者など：
  第 3 章から第 8 章を中心に読むことで、攻撃者視点を踏まえたサイバーデセプションの考え方、設計・運用、および検証内容を理解できる。
- IT/OT の運用担当者：
  第 4 章から第 8 章を中心に読むことで、サイバーデセプション資産の配置・運用方法、および実環境への適用イメージを把握できる。
- その他の担当者：
  初めて本テーマに触れる読者には通読を、特定の観点で情報を求める読者には該当章の参照を推奨する。

【コラム 1】 日本人は騙すことが嫌い？

「デセプション（欺瞞）」と聞くと、少し身構える人がいるかもしれません。

デセプションは日本語では「欺瞞」と訳されますが、「欺く」「だます」という言葉にはどうしても悪い印象があります。セキュリティ対策の説明で「攻撃者をだまします」と言われると、「それは正当な防御なのか」「こちらも攻撃者と同じことをしているのではないか」と感じる人がいても不思議ではありません。

しかし、ここで考えたいのは「だますこと」そのものではありません。重要なのは、相手の認知や判断に働きかけ、危険な行動を抑えたり、安全な方向へ誘導したりすることです。

例えば、畑に立つかかしは本物の人間ではありません。それでも鳥が「人がいるかもしれない」と判断すれば、近づく行動を変えることがあります。防犯カメラのダミーも同じです。本物の録画機能がなくても、「見られているかもしれない」と感じさせることで行動を変える可能性があります。また、警備会社のステッカーを玄関に貼っておくことは、犯罪を考える人に対するけん制になるかもしれません。

サイバーデセプションもこの発想に近いものです。偽のサーバー、偽の認証情報、偽のファイルを置くことが目的ではありません。攻撃者がそれを見たときに「価値がある情報かもしれない」と判断したり、逆に「これは罠かもしれない」と判断したりすることによって、その行動が検知、遅延、誘導、観測につながることが重要です。

日本では「欺瞞」や「だまし討ち」という言葉にどこか後ろめたさを感じる場面があります。一方で、だまし討ちに対する評価は単純ではありません。歴史的に見ても、日本の軍記物語や合戦に関する研究ではだまし討ち的な事例が存在し、それが常に一律に否定されていたわけではなく、時代や文脈によって知略として扱われたり、倫理的に問題視されたりする複雑な評価があったことが示されています。

つまり、「日本人はデセプションそのものが嫌い」なわけではありません。これは文化的な善し悪しではなく、単なる言葉の受け止め方の問題です。日本では「正々堂々」「誠実」「透明性」

といった価値観が重視されるため、『欺瞞』という言葉の響きが防御技術の説明として受け入れられにくいのだと考えられます。

海外では、サイバーデセプションは防御戦略の一部として整理されつつあります。例えば MITRE Engage は、攻撃者との関わり方を計画する枠組みとして、デセプションや拒否を扱っています。MITRE 自身も、Engage をサイバー防御側がアドバーサリ・エンゲージメント活動を議論・計画しやすくするためのフレームワークとして説明しています。

英国 NCSC も、サイバーデセプション技術や技法が一定の状況でサイバー防御に価値を持つ可能性を認識し、国家規模で有効性の根拠を蓄積する取り組みに言及しています。国家機関によるこうした動きに加え、ビジネスの側面でも、デセプション技術市場では北米が大きな比率を占めているという市場調査の結果が示されています。

もちろん、海外で使われているから日本でもそのまま導入すべき、という話ではありません。サイバーデセプションは導入すれば自動的に効果が出る魔法の技術ではなく、設計、監視、運用、法務・倫理面の整理が必要です。

だからこそ、本書ではデセプションを「相手をだます怪しい技術」としてではなく、「攻撃者の認知・判断・行動に働きかけ、防御側に有利な状況を作る考え方」として扱います。言い換えれば、サイバーデセプションとは無秩序に罠を仕掛けることではありません。守るべき資産を明確にし、攻撃者の行動を想定し、安全に管理された範囲で検知・遅延・誘導・観測を実現するための防御設計です。

「デセプション」という言葉に抵抗感があるからこそ、その意味を丁寧に整理する価値があります。

日本でサイバーデセプションが広がるかどうかは、技術そのものよりも、「これは卑怯なだまし討ちではなく、管理された防御戦略である」と説明できるかにかかっているのかもしれません。

コラム内参考資料

MITRE, “MITRE Engage: A Framework and Community for Cyber Deception.”
MITRE Engage, “An Adversary Engagement Framework.”
UK National Cyber Security Centre, “Building a nation-scale evidence base for cyber deception.”
Grand View Research, “Deception Technology Market Size & Share Report, 2030.”
佐伯真一,「合戦のルールとだまし討ち」『軍記と語り物』第 37 号, 2001 年 3 月

（余白ページ）

第2章　能動的防御へのパラダイムシフト

攻撃者が圧倒的に有利な「投資と労力の非対称性」が続く限り、既存のセキュリティ強化だけでは組織を守り切ることはできない。本章では、従来の受動的な多層防御が直面している限界と、現場を疲弊させる運用課題を述べる。そして、この構造的不利を覆すため、システムを守る視点から「攻撃者の心理をコントロールする」視点へと戦略を拡張する、サイバーデセプションの必要性を論じる。

なお、本書でいう「能動的防御」とは、攻撃者のシステムへ侵入したり、攻撃インフラを破壊したりする行為を指すものではない。本書では、自組織が管理する環境内において、攻撃者の認知・判断・行動に働きかけ、検知、遅延、誘導、観測を実現する防御上の考え方として用いる。

2.1　現状のセキュリティ戦略と課題

従来のセキュリティ戦略は、主にネットワークの境界で通信を監視したり、侵入を水際で防いだりすること（境界防御）に重点を置かれている。具体的には、ファイアウォールによるアクセス制御、既知の脅威を検知する侵入検知・防御システム（IDS/IPS）、端末側を守るエンドポイント保護などで、これらを組み合わせた多層防御が構築されている。また、エンドポイント上の振る舞いを監視して脅威を検知・対応する EDR も広く活用されている。

境界防御に加え、現在では主流になりつつあるセキュリティモデル「ゼロトラスト（全てのアクセスを継続的に評価し、認証・認可を行うセキュリティの概念）」も導入が進んでおり、ゼロトラストを実現する手法として、厳格な ID 管理、多要素認証などがある。

しかし、これらのセキュリティ対策には主に以下に示す 3 つの課題が存在する。

境界防御の限界と、ゼロトラスト導入後にも残る課題

高度な攻撃能力を持つ攻撃者による標的型攻撃や、未知の脆弱性を悪用するゼロデイ攻撃に対して、境界防御は容易に突破されるリスクがあり、一度内部に侵入されると、従来の予防型ツールでは攻撃者の横展開を十分に検知できない可能性が高い。EDR が導入されていれば検知・対応できる可能性は上がるものの、正規ツールを悪用する Living off the Land（LotL）やファイルレスで悪意のあるコードを実行すると、監視基盤をすり抜けられる場合がある。

また、多要素認証を導入している場合でも、認証後のセッションを悪用されるリスクは残る。例えば、AiTM（Adversary-in-the-Middle）攻撃では、攻撃者が利用者と正規のログインサイトの間に割り込み、認証処理を中継することで、認証完了後に発行されるセッション Cookie を窃取する。セッション Cookie は、利用者がログイン済みであることを示す情報であるため、これを悪用されると、攻撃者は多要素認証を改めて突破しなくても、正規利用者になりすましてサービスへアクセスできる可能性がある。したがって、認証強化に加えて、認証後の不審な行動や内部探索を検知する仕組みが重要となる。

アラート過多と調査負荷の増大

IDS/IPS、EDR、SIEM などの監視基盤は、攻撃の早期発見に不可欠である。一方で、正常な業務通信や低リスクの挙動までアラートとして検出されることがあり、SOC やセキュリティ担当者は大量の通知の中から真に対応すべき脅威を見極める必要がある。

このようなアラート過多は、調査負荷の増大や対応遅延を招くだけでなく、重要な兆候を見落とす「アラート疲れ」につながる。したがって、今後の防御戦略では、単に検知点を増やすだけでなく、より高い確度で対応すべき事象を示す仕組みが求められる。

リアクティブな防御と非対称なコスト

従来型の防御は攻撃が発生してから対処する「事後対応的（リアクティブ）」な性質を持っている。防御側は自組織のネットワーク、クラウド、エンドポイントに存在するすべての脆弱性を把握し、すべての資産を保護するために多大なコストをかける必要がある。一方で攻撃者は未知の脆弱性（ゼロデイ）、設定ミス、従業員のミス（認証情報の漏洩など）といった 1 つの弱点を見つけるだけで侵入できる。この「攻撃者有利のコスト非対称性」が解消されにくい点が、大きな課題である。

2.2 今後のセキュリティ戦略の提言

前節で述べたとおり、未知の脅威を水際で完全に防ぐことが困難となる中、境界防御や監視体制を単に強化し続ける従来のアプローチは、費用と運用負荷の際限ない増大を招いている。防御側はすべての IT/OT 資産を完璧に保護しなければならないのに対し、攻撃者はたった一つの脆弱性や設定ミスを見つけるだけで侵入目的を達成できる。

この構造的な不利を緩和するためには、システムはいずれ突破されるという侵害前提の視点に立つことが不可欠である。本書で提言する戦略は、これまでのファイアウォール、EDR、SIEM、ゼロトラストなどのセキュリティ対策を否定・置き換えるものではない。既存の多層防御を基盤としつつ、そこに内部に侵入した攻撃者を意図的に欺く「サイバーデセプション」という能動的な防御レイヤーを統合することで、組織のセキュリティ戦略全体に「攻撃者心理のコントロール」という新しい視点をもたらすものである。これが現状の課題に対する有効な選択肢になり得ると考える。

2.2.1 米国ガイドラインが示すデセプションの戦略的価値

サイバーデセプションは、米国国立標準技術研究所（NIST）が発行するシステム・セキュリティ・エンジニアリングのガイドライン「NIST SP 800-160 Vol.2」[2]において、サイバーレジリエンス（回復力）の向上に資する重要な設計原則の一つとして位置づけられている。

すべての脆弱性を塞ぎ、監視の目を増やすというアプローチだけでは、防御側の費用と運用負荷の増大を招き続ける。これに対し、サイバーデセプションは防御側がシステム内部や本番環境の隙間に「おとり（罠）」を意図的に配置し、攻撃者の情報収集や意思決定のプロセスを狂わせるという全く新しい視点を提供する。これにより、従来の多層防御が抱える限界を補完し、戦略的な防衛の選択肢を広げることが可能となる。

2.2.2 攻撃者を欺くサイバーデセプションのアプローチ

サイバーデセプションの核心は、単に防御の壁を厚くするのではなく、システム内部に意図的なおとりを仕掛け、防御側が主導権を握る能動的な防御手法である。このアプローチは、本番環境から切り離された「単体としての防衛線」として効果を発揮し、同時に「既存の防御との組み合わせ」によってその効果を高めることもできる。

このアプローチの全体像を下図に示す。この図で示すように、サイバーデセプションは実装の自由度が高く、組織の環境や目的に応じて「単体での独立運用」と「既存対策との組み合わせ」のいずれかを選択、あるいは両者を併用することができる。

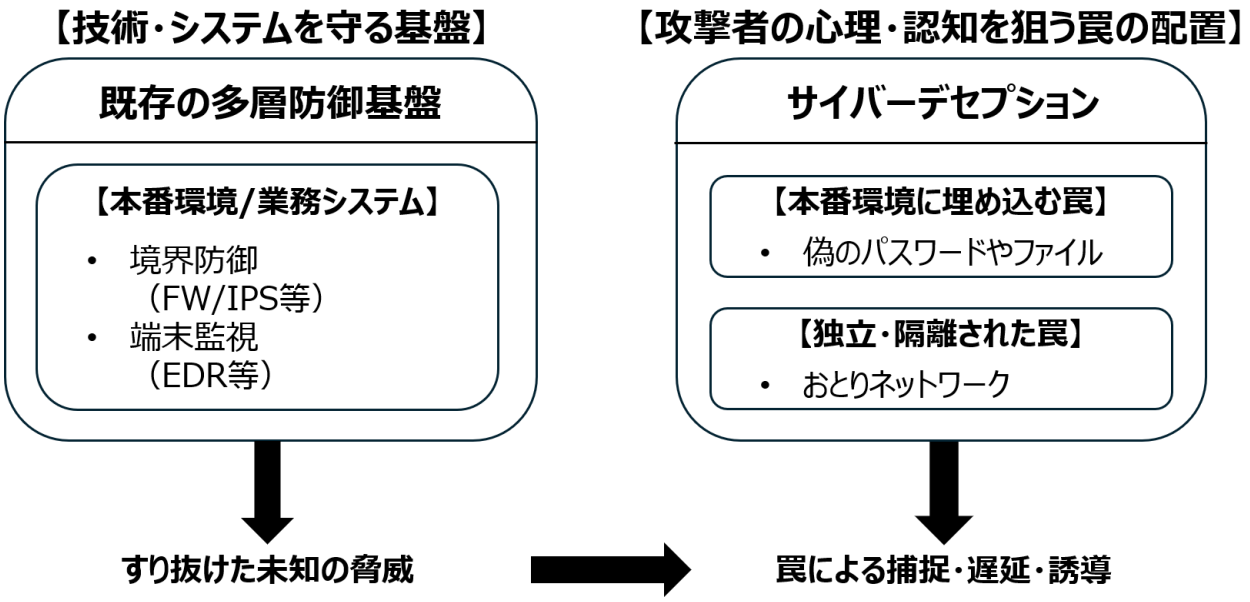


図 2-1　既存の防御基盤とサイバーデセプションの相互作用（概念図）

【技術・システムを守る基盤】：既存の多層防御基盤

従来、境界防御や端末監視といった既存の多層防御基盤によって、本番環境や業務システムを技術的・システム的に守ってきた。しかし、高度化する未知の脅威はこれらの強固な壁をすり抜けることがある。

【攻撃者の心理・認知を狙うおとりの配置】：サイバーデセプション

サイバーデセプションは、このすり抜けた脅威に対処する。既存システムと同じ場所に、攻撃者が価値を感じるような偽の情報をおとりとして溶け込ませ、攻撃者の認知を操作する。おとりを配置することで、脅威を捕捉、遅延、誘導し、既存の多層防御が抱える限界を補完する。

また、サイバーデセプションの「単体での効果」と「既存対策との組み合わせによる効果」の例を下表にまとめた。

表 2-1　デセプションの独立効果と組み合わせ効果の例

【技術・システムを守る基盤】	サイバーデセプション単体での効果	既存対策との組み合わせによる効果
境界防御	既存システムに一切干渉せず、侵入した攻撃者を隔離環境へ引き込み、足止め・観測する独立した防衛線となる。	境界防御を突破した攻撃者のネットワーク偵察行動を逆手に取り、本物の資産から視線を逸らさせておとりへ誘導する。
端末監視	正規業務ではアクセスされないため、これ単体で「誤検知の少ない確実な侵入検知点」として機能する。	既存の端末監視ツールでは検知が難しい「正規ツールや正規権限を悪用した隠密行動」を、内側からあぶり出す。
運用監視	膨大なログ分析基盤がなくても、おとりへの接触という事実だけで、単体で即座に致命的な脅威を識別できる。	既存の SOC ワークフローに統合することで、日々の不要なノイズ調査を削減し、真の脅威への対処能力を最適化する。

構造的に「独立した罠」として運用するか、「本番環境に組み込むか」を問わず、これらに共通するのは「正規の業務通信とは完全に独立している」という運用の性質である。そのため、おとりへのアクセスはすべて悪意のある行動、または致命的な設定ミスを意味し、既存の監視装置で生じがちな正常通信との誤検知が原則として発生しない。日々の不要なノイズ調査から解放され、現場を疲弊させる運用負荷を大幅に軽減しながら、真の脅威へ迅速かつ確実に対応することが可能になる。また、システムの稼働継続が最優先される環境であっても、正規の業務通信に影響を与えずに脅威を検知できるため、極めて親和性が高い。

おとりという新たな視点を加える、あるいは独立した防衛線として並走させることで、攻撃者は「本物と偽物が入り混じる環境で、一度もミスをせずに本物を選び続けなければならない」という高いリスクを背負うことになる。この柔軟な実装アプローチこそが、これまで攻撃側が圧倒的に有利であった投資と労力の非対称性という構造的な不利を、大きく変容させる力を持っている。

2.2.3　グローバル動向から見る選択肢の広がり

海外の先進的な組織は、ランサムウェアの甚大な被害や標的型攻撃の経験から、どれほど多額の投資を行っても初期侵入を完全に防ぐことは不可能であるという前提に立って対策を進めている。その中で、限られた予算と人員において防御力を最大化するため、多層的な防御を補完する検知・誘導策として、おとりの技術を本番環境へ柔軟に組み込む動きが加速している。[3]

一方、日本企業は依然として侵入を水際で防ぐための予防的対策に予算と運用リソースの大部分を割く傾向にある。そのため、一度内部に侵入された後の検知の選択肢が限られており、グローバル基準と比較すると構造的な防御力の偏りが生じているのが実態である。今後、増大し続ける運用負荷を適正化しつつ、組織全体のセキュリティ水準を一段階向上させるためには、既存の対策を基盤としつつ、能動的な罠のアプローチを新たな視点として取り入れる価値は十二分にある。[4] [5]

2.2.4　人間を標的とした心理戦という視点

ソフトウェアの脆弱性を塞ぎ続け、ネットワークの監視の目を光らせることだけがセキュリティの全体像ではない。本書で提言する戦略の核心は、システムの保護に終始していた従来の視野を広げ、「デセプション（欺瞞）」に着目した新たなアプローチを取り入れることそのものにある。

それはすなわち、防御のスコープを技術やシステムといったデジタルの領域から、画面の向こう側で意思決定を行っている「攻撃者の心理」へと拡張することを意味する。攻撃者は、限られた時間の中で効率よく目的を達成するため、自らが収集した情報やハッキングツールを信頼して行動せざるを得ない。デセプションはこの情報への信頼を逆手に取り、偽の情報を意図的に混入させることで攻撃者を疑心暗鬼に陥れ、焦りや思い込みによる判断ミスを誘発させる。

しかしながら、人間の認知を操作し、意図的な行動へ誘導するというアプローチは、決してサイバー空間で突如として生まれた新しい概念ではない。古くは国家間の命運を分けた軍事戦術や、観客の意識を逸らす奇術（マジックにおけるミスディレクション）として、有史以来人類が探求し、体系化してきた普遍的な心理戦の技術である。

既存の防衛基盤にこの「心理戦」という新たな視点を効果的に組み込むためには、単におとりのツールを導入するのではなく、まず「人間が騙されるメカニズム」そのものを根本から理解する必要がある。次章では、セキュリティを一度離れ、デセプションという概念の歴史的背景と、相手の認知・判断・行動に影響を与える心理的フレームワークについて解説する。高度な技術を持つハッカーであっても、なぜ自ら巧妙な罠に足を踏み入れてしまうのか、その深層を紐解いていく。

【コラム 2】 偽物の美学

「おとり」と言われると身構えてしまうあなたも「偽物」と言われるとどうでしょうか。読者の中にはブランド物のコピー品に泣かされた経験がある方がいるかもしれません。

偽物には粗悪なものだけでなく偽物の質をめぐってしのぎを削る世界があります。それが芸術における「贋作」です。一流の贋作師は鑑定士の目利きを掻い潜るため様々な工夫を凝らします。

例えば「経年劣化の模倣」です。19 世紀の美しい花の静物画の偽物を作る場合、彼らはただ当時の画風を真似るだけではありません。同じ時代のキャンバスに当時の顔料で調合した絵の具で描き、サウナで熱することで自然な経年劣化を人工的に再現します。完璧すぎる美しさは、かえって疑念を生むことを知っているからです。

そして「来歴の偽装」です。作品だけでなく、「誰が所蔵し、どのような運命を辿ってここにあるのか」という物語を偽造します。古い画廊のラベルをキャンバスに貼り付け、売買記録を捏造し、その絵がここに存在する歴史を作るのです。

さらに「未公開作品の創出」です。贋作はコピー作品だけではありません。有名画家の画風を真似たオリジナルの作品を、作品の発表が無い空白期に描かれた幻の作品や名前だけが記録に残る戦争で紛失した作品として売り出すのです。新発見を待ち望む人々の期待感を利用します。

こうして生み出された至高の「贋作」を前に、どれだけの鑑定士たちが涙を飲んできたことでしょう。

贋作師は、一度見破られればその鑑定手法を研究し、さらに巧妙な「贋作」を作ります。新しい手法で作成された「贋作」を前にして、鑑定士は偽物であると見極め続けることができるでしょうか。いつの時代も、見極める側というのは、圧倒的な不利を強いられるものです。

システムを守る立場にある私たちは、日々巧妙化するサイバー攻撃を前に、まるで疲れ切った鑑定士のように「本物の脅威」を見極めることに奔走しがちです。次々と送られてくるアラートから、一つたりとも見落としが無いようにと神経をすり減らします。

しかし、考えてみてください。システムの「提供者」であり、その世界の「作者」でもある私たちが、常に「買い手」でいる必要はどこにもないのです。ずっと鑑定のルーペを覗き込んでいる必要はありません。今度はキャンバスの前に立ち、攻撃者を優雅に騙す「一流の贋作師」になってみませんか？

コラム内参考資料

朝日新聞 GLOBE+. "日本の美術館も被害か　300 点のニセモノつくった「天才贋作師」の男が問う名画の価値". 朝日新聞 GLOBE+. https://globe.asahi.com/article/15647391, （参照 2026-06-06）.

（余白ページ）

第3章　デセプション

3.1 デセプションの概念

デセプション（欺瞞）とは、相手の認知・判断・行動に影響を与え、自らに有利な結果を導くための意図的な働きかけである。一般に「うそ」は情報の真偽のみに着目するのに対し、デセプションは単なる虚偽情報の提示にとどまらない。相手に何を見せ、何を見せず、どのような状況認識を持たせ特定の行動をとらせるかを設計する営みである。

したがって、デセプションを理解するうえでは、提示された情報が真実か虚偽かだけでなく、その情報が相手の判断や行動にどのような影響を与えるかを考える必要がある。デセプションは、戦争、外交、政治、諜報、商取引、マジック、日常生活など、幅広い場面で用いられてきた普遍的な概念である。サイバー領域はその応用先の一つにすぎない。サイバーデセプションを、単なるハニーポットや偽データの設置として理解するのではなく、まずは相手の認知に働きかける一般的な考え方として捉えることが重要である。

デセプションの本質は「実体を隠すこと」と「虚構を見せること」を組み合わせ、相手に防御側が意図した状況認識を持たせる点にある。例えば、本当に守るべき対象を見つけにくくする、重要ではないものを重要に見せる、存在しない対象を存在するように見せる、あるいは相手の注意を別の方向へ向けるといった方法がある。これらはいずれも、相手が「何を見て」「何を見落とし」「何を信じるか」を操作する働きかけである。

このため、デセプションは情報そのものの操作だけで成立するものではない。相手の知識、経験、期待、先入観、焦り、確認不足といった認知上の特性に依存して成立する。仕掛ける側が精巧な偽情報やおとりを用意したとしても、相手がそれを自然なものとして受け入れなければ効果は限定的である。逆に、相手の期待や思い込みに合致していれば、必ずしも大がかりな仕掛けでなくても、判断を誤らせたり、注意をそらしたりすることが可能となる。

また、デセプションは、相手を完全に信じ込ませる場合だけに効果を持つわけではない。相手に疑念や不確実性を与えるだけでも、行動を遅らせ、確認コストを増やし、判断を迷わせる効果が生じ得る。例えば、相手が「これは本物なのか」「罠ではないのか」と考え始めれば、その時点で相手の時間や労力は消費される。デセプションの効果は、誤認、遅延、かく乱、誘導、観測といった複数の形で現れる。

本書で扱うデセプションは攻撃や詐欺のための欺瞞ではない。防御側が重要資産を守り、攻撃者の行動を早期に把握し、攻撃の進行を遅らせ、必要に応じて隔離された環境へ誘導するための防御的な欺瞞である。したがって、その目的は相手をだますこと自体ではなく、攻撃者の行動を防御側にとって望ましい方向へ変化させることにある。

このように、デセプションは「偽情報を置くこと」や「罠を仕掛けること」だけを意味する狭い概念ではない。相手の認知空間を設計し、判断の前提を変え、行動を制御または観測するための広い概念である。この理解があってはじめて、後段で扱うサイバーデセプションも、個別技術の集合ではなく、攻撃者の認知・判断・行動に働きかける防御戦略として位置づけることができる。

3.2　デセプションの活用例

デセプションという概念はサイバーセキュリティにおいて視点を変えた発想であるが、様々な分野において古くから活用されてきた。ここでは、軍事とマジックの 2 つの分野における活用例を紹介する。

3.2.1 軍事におけるデセプション

紀元前 5 世紀に孫子が「兵は詭道なり」と説いたように、軍事におけるデセプションは古くから戦争の基本原則とされており、様々な場面で活用され、体系化されている。

第二次世界大戦においては、デセプションが戦略上より重要な手段として位置づけられるようになった。その代表例として、当時ドイツ軍の支配下にあったフランスのノルマンディー海岸に連合軍が大規模な上陸作戦を実施し、上陸に成功した「ノルマンディー上陸作戦」を挙げる。

この作戦で連合軍に求められたのは、「侵攻は 6/6 に、ノルマンディーで行われ、10 万人以上の兵力が投入される」という事実を、「侵攻は 6/6 以降の遅い時期に、フランスからオランダの別の海岸で行われ、10 万人を超えるより大規模な兵力が投入される」と欺瞞することであった。このデセプションの一環として実行されたのが、「フォーティテュード作戦」である。

フォーティテュード作戦では、上陸場所がノルマンディーではなく、パ・ド・カレー（フランス）であるとドイツ軍に誤認させるために多角的なデセプションが講じられた。具体的な手法として、パ・ド・カレーの対岸に位置するイースト・アングリア（イギリス）に、大規模な部隊が展開しているように見せかけるため、著名な指揮官であったパットン将軍を架空の部隊の司令官に任命した。さらに、イギリスに潜入していたドイツのスパイを逮捕し、二重スパイとして逆利用することで誤情報を送った。これらのデセプションにより、ドイツ軍は上陸場所を特定しきれず、部隊を分散して配置せざるを得ない状況に追い込まれた。その結果、連合軍はノルマンディーへの上陸に成功し、ひいては第二次世界大戦の勝利に繋がった。[6]

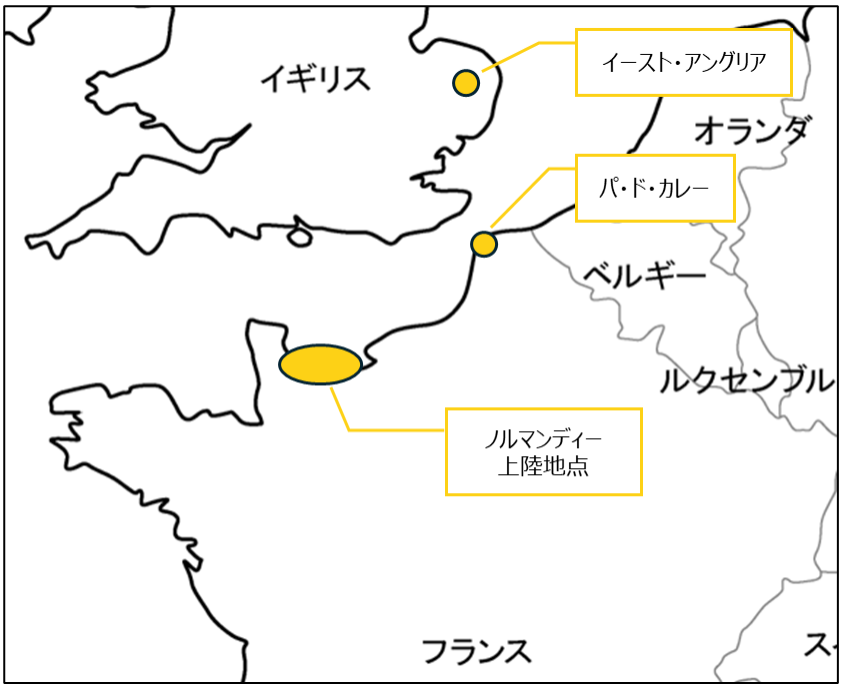


図 3-1　ノルマンディー上陸作戦における欺瞞計画の位置関係

現代においては軍事的なデセプションは、米国統合参謀本部が発行する統合ドクトリンの中で、「MILDEC（Military Deception）」として体系化されている。この文書では、軍事デセプションを「敵軍、準軍事組織、または暴力的な過激派組織の意思決定者を意図的に誤認させるための行動」と定義しており、これにより、「敵対勢力に特定の行動あるいは不作為（何もしないこと）をとらせることで、味方の任務達成に寄与する」ことを目的としている。[7]

この定義や目的から読み取れるように、デセプションを成功させるには、敵対する勢力に誤認させるだけでは不十分であり、相手の具体的な行動や不作為に結びつける必要がある。

3.2.2 マジックにおけるデセプション

軍事においてデセプションは作戦における一つの選択肢であり、必ずしも用いられるわけではないが、マジックにおいては「欺くこと」が本質であり、観客もそれを望んでいるという点で、他の分野とはデセプションの捉え方が大きく異なる。観客は、目の前で何かが隠され、別のものが見せられていることを理解しながら、それでもなお驚きや不思議さを体験するためにマジックを受け入れている。

マジックにおける基本原則は「ミスディレクション」であると言われる。ミスディレクションとは、観客の注意を特定の対象や動作に向けることで、本当に見られては困る動作や仕掛けから注意をそらす技術である。言い換えれば、実体を隠し、偽りの姿を見せることである。マジックにおいては、「隠すこと」と「見せること」という 2 つの要素を効果的に組み合わせてデセプションを構成する。

ミスディレクションを成功させるための手法として「ワンアヘッドの原理（One-Ahead Principle）」が挙げられる。マジシャンは、あるトリックを行っている最中に、次の（一歩先の）トリックを準備しておく。これにより、観客は一連のマジックの始まりと終わりのタイミングを誤認し、種を見破ることができなくなる。ここで重要なのは、観客が「いま何が起きているか」を理解しているように見えても、実際にはマジシャンが一歩先の準備を進めている点である。

このように、マジックは単なる手先の技術ではなく、観客の注意、期待、理解の流れを制御する行為であり、ケビン・ミトニックは著書の『欺術』においてマジックの成功には話術が大きな役割を果たすと述べている。[8]ここでの話術とは、単に場を盛り上げるための解説ではなく、観客の脳内に「偽の前提」を植え付けるための心理的なツールである。マジシャンは言葉を用いて、物理的には不自然な動作や、本来は不要な手順に対して「もっともらしい理由」を与える。観客はその語りに納得することで、仕掛けに直結する怪しい動きを「日常の自然な行為」として都合よく解釈し、自ら見落としてしまう。つまり、観客が何を見るかだけでなく、何を見落とし、どのタイミングで何を理解したと思うかまでを設計することが、マジックにおけるデセプションの本質である。

マジックにおけるデセプションから得られる示唆は、単に「相手をだます技術」ではないということである。重要なのは、相手の注意をどこに向けるか、何を見せ、何を見せないか、どのタイミングでどのような理解を持たせるかを設計することである。この考え方は、後述する Bell-Whaley フレームワークにおける「実体を隠す」アプローチと「虚像を見せる」アプローチの理解にもつながる。

3.2.3 Bell-Whaley フレームワーク

本章においてはここまで、デセプションという概念とそれが人類史においてどのように活用されてきたかをご紹介した。そして、このデセプションの技術を体系化した代表的なモデルに Bell-Whaley フレームワークがある。このフレームワークは J. Bowyer Bell と Barton Whaley の 2 名が共著した書籍「Cheating and Deception[6]」を起源としている。同書の中ではデセプション技術全般が体系化され紹介されており、さらに、それらの技術が現実の世界で軍事・諜報・マジックなど様々な分野で活用されていることが論じられている。現在ではサイバーセキュリティ分野においてもデセプション技術のフレームワークとして、MITRE 社や IEEE の技術論文でも参照されている。

Bell-Whaley フレームワークではデセプション技術を体系化する上で、相手の認知を操作するアプローチを大きく「実体を隠す」と「虚像を見せる」の 2 つに定義している。この 2 つのアプローチの中でそれぞれ 3 つ、合計で 6 つにデセプション技術を分類している。以降で各分類についての説明を軍事・マジック・サイバーの分野別の例示を交えて記載する。

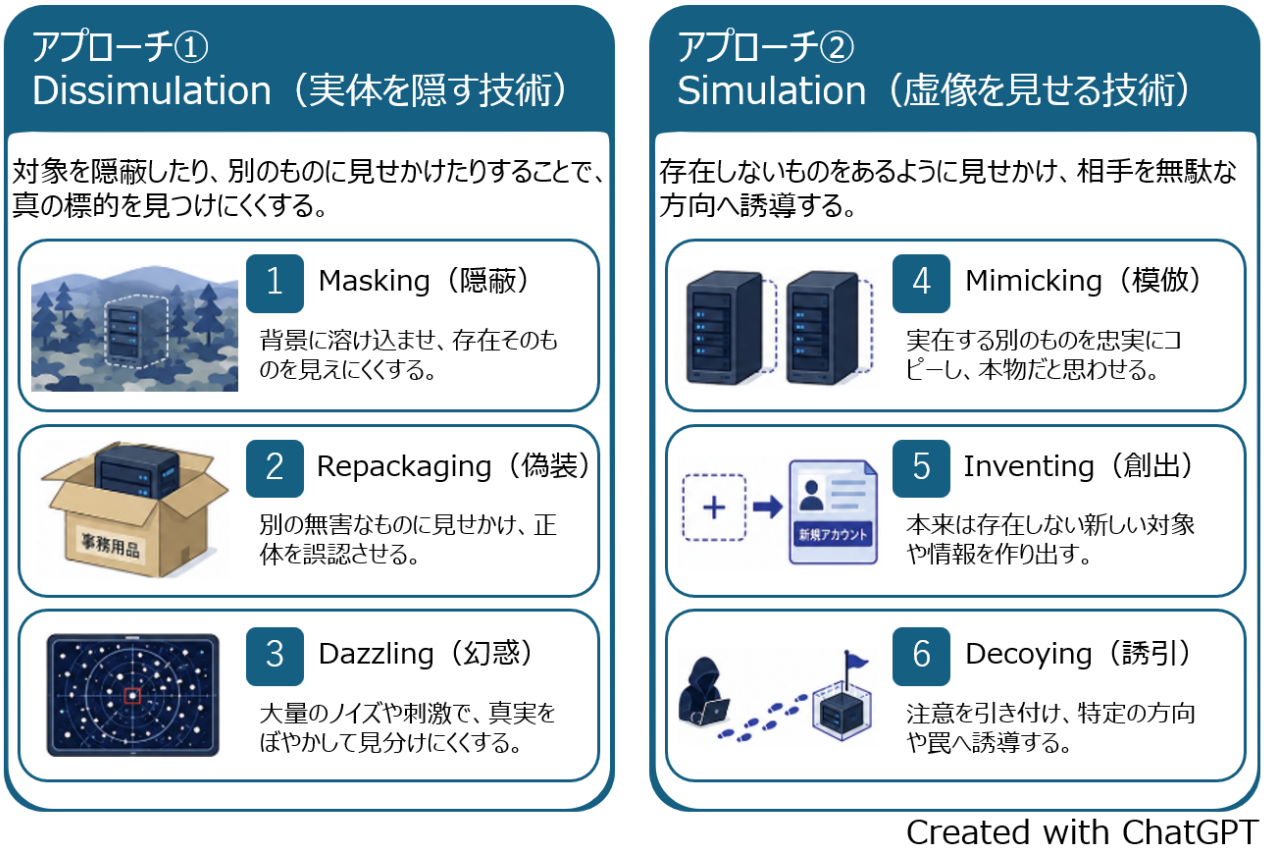


図 3-2　Bell-Whaley フレームワーク：２つのアプローチと６分類

アプローチ 1：Dissimulation（実体を隠す技術）

対象を隠蔽したり、別のものに見せかけたりすることで、攻撃者が「真の標的」を見つけるのを困難にする技術であり、以下の３つに分類される。

1. Masking（隠蔽：背景に溶け込ませて不可視化する）

軍事	ジャングルでの迷彩服や、戦車にカモフラージュネットを被せて見えなくする。
マジック	暗いステージで黒い服を着ることで、体を背景の黒幕と同化させ、仕掛けを見えなくする（ブラックアート）
サイバー	ポートスキャン（外部からの調査）に対して応答を返さず、サーバーの存在自体をネットワーク上で「見えない」状態にする。

2. Repackaging（偽装：別の無害なものに見せかける）

軍事	重要な兵器の輸送トラックを、「一般の食料輸送車」に偽装する。
マジック	タネが仕込まれた特殊なトランプを、市販のトランプの箱に入れてテーブルに置いておく。
サイバー	本番稼働している重要データベースのサーバー名を、あえて攻撃者が興味を持たない「test-old-server（古いテスト用）」に変更する。

3. Dazzling（幻惑：大量のノイズで真実をぼやかす）

軍事	空中に大量のアルミ片（チャフ）を散布し、敵のレーダー画面を無数のノイズで埋め尽くして本物の戦闘機を隠す。
マジック	強いフラッシュを焚いたり、突然ハトを出したりして観客の視覚・感覚をショートさせ、その隙に仕掛けを動かす。
サイバー	ネットワーク上に「ダミーの IP アドレス（存在しない端末）」を仮想生成し、攻撃者にネットワーク探索に時間を浪費させる。

アプローチ 2：Simulation（虚像を見せる技術）

存在しないものをあるように見せかけ、攻撃者のリソース（時間や労力）を無駄な方向へ誘導する技術であり、以下の 3 つに分類される。

4. Mimicking（模倣：実在する別のものを忠実にコピーする）

軍事	空気で膨らむ「風船の戦車」を大量に配置し、敵の航空写真から「大部隊が駐留している」と誤認させる。
マジック	本物のコインと全く同じ見た目で同じ重さだが、実は裏表が同じや半分に割れる「特殊なギミックコイン」を使う。
サイバー	本番環境と全く同じ OS、同じポートが開いている精巧な「おとり環境（高対話型ハニーポット）」を稼働させる。

5. Inventing（創出：存在しない新しいものをでっち上げる）

軍事	無線通信上だけの「架空の部隊」のやり取りを行い、敵に存在しない部隊が存在するように見せかける。
マジック	「これは古代エジプトの呪われた箱です」と、架空のストーリーを語り、観客に「思い込み」を作らせる。
サイバー	実際には存在しない「admin_backup_2024」といった架空の管理者アカウント（ハニートークン）を発行し、わざとアクセスできる場所に放置する。

6. Decoying（誘引：注意を引いて罠に誘導する）

軍事	あえて防御が手薄に見える「隙」を作り、敵軍を不利な地形や、味方が待ち伏せしているエリアにおびき寄せる。
マジック	右手でステッキを大きく振って観客の視線を完全に集め（ミスディレクション）、その間に左手で密かにタネを処理する。
サイバー	端末のメモリ上に「重要サーバーへの接続情報（偽物）」を置いておき、攻撃者にそれを盗ませて、おとりのサーバーへアクセス（横展開）させる。

【コラム 3】 イギリス人はデセプションがお好き？

第１章のコラムでは、日本人が「デセプション（欺瞞）」という言葉に抵抗感を持つ傾向にあると述べましたが、世界に目を向けると、デセプションを効果的に利用してきた国があります。それがイギリスです。

第二次世界大戦中、イギリス軍は様々なデセプションを立案し、実行しました。「3.2.1 軍事におけるデセプション」で言及したフォーティテュード作戦もその一例です。本コラムでは、イギリス軍が実行した作戦の中でとりわけ奇抜な「ミンスミート作戦」を紹介します。

1943 年、連合軍はヨーロッパ本土への足掛かりとして、イタリアのシチリアを攻略目標としていました。しかし、シチリアはあまりにも理想的な標的であり、ウィンストン・チャーチル首相が「極めつけの馬鹿以外は、誰だってシチリアだと知っていた」と述べるほどでした。当然、ドイツ軍もシチリアへの攻撃を察知し、警戒を強めていました。

では、「次の標的はシチリアではない」とドイツ軍に信じ込ませるためにはどうすればよいか。イギリス軍が用意したのは「架空の将校の死体」でした。死体に「次の攻略目標はギリシャである」という偽の機密文書を持たせ、潜水艦からスペイン沿岸に向けて漂流させたのです。当時スペインは中立国でしたが、密かにドイツ諜報機関と協力関係にありました。

これだけを聞くと、成功するはずのない荒唐無稽な作戦であるように思えます。しかし、この作戦が成功した理由は「細部へのこだわり」にありました。

イギリス軍が作り出した架空の将校「ウィリアム・マーティン少佐」にリアリティを持たせるため、彼のポケットには一度紛失して再発行されたという設定の身分証をはじめ、婚約者の写真やラブレター、シャツの領収書、ロンドンの劇場の半券まで入れられ、生活の痕跡が演出されました。

死体がスペイン沿岸で発見されてからも、イギリス軍の徹底した演出は続きました。戦死者リストに彼の名を載せ、さらにスペイン駐在のイギリス海軍武官に彼が運んでいた書類を至急回収するようにメッセージを送りました。このメッセージはドイツ諜報機関に傍受されることを想定した“おとり”でした。

これらのこだわりが功を奏し、ドイツ軍は「偶然手に入れた機密書類」を本物と信じ込みました。ドイツ軍がシチリア以外の地域へ兵力を分散させたことで、連合軍は手薄になったシチリアの攻略に成功しました。ミンスミート作戦は見事に目的を果たしたのです。

ちなみに、チャーチルは事前に「作戦が失敗する可能性がある」と警告された際にこう答えたと記録されています。「その時は、死体を回収してもう一度泳がせるまでだ」。デセプションを成功させる要素は、細部へのこだわりだけではありません。失敗を恐れず、何度でもやり直す覚悟こそが成否を分けるのかもしれません。

コラム内参考資料

Imperial War Museums. “Spies, Lies and Deception”. Imperial War Museums. https://www.iwm.org.uk/sites/default/files/files/2023-10/spies_large_print_guide.pdf, （参照 2026-05-27）.

ベン・マッキンタイアー.ナチを欺いた死体 英国の奇策・ミンスミート作戦の真実.中央公論新社,2011.

第4章　サイバーデセプション

4.1 サイバーデセプションとは

「サイバーデセプション（サイバー空間における欺瞞）」は、攻撃者の意思決定に作用することで攻撃者の費用対効果を低下させる能動的な防御手法である。単なる防御ツールの追加ではなく、「第 2 章　能動的防御へのパラダイムシフト」で述べた「サイバーレジリエンス」の向上に寄与することができるアプローチの一つであり、防御側が「非対称な優位性」を獲得する多層防御の戦力増強手段として機能する。[1]

4.1.1 セキュリティ戦略に及ぼす効果

サイバーデセプションの導入は、組織のセキュリティ戦略および運用において、主に以下の 3 つの効果をもたらす。

(1)　誤検知の排除と SOC の対処能力向上

従来のセキュリティツールは、正常な通信と攻撃の区別が難しく、大量の「誤検知（ノイズ）」を発生させることがあり、対応現場の「アラート疲れ」を引き起こす最大の原因となっている。しかし、サイバーデセプションが配置する「おとり」には正規の従業員やシステムがアクセスする理由がないことから、おとりに少しでも触れたという事実自体が100%に近い精度で不正アクセスの証拠（高精度なアラート）となる。これにより、SOC は無駄な調査から解放され、真の脅威に対して迅速かつ確実に対処できるようになる。

(2)　攻撃者の時間とコストの浪費

サイバーデセプションは、システム内部に偽のサーバーや偽の認証情報（パスワードなど）を配置し、攻撃者を欺く防御手法である。攻撃者は本物の資産を見分けるために無駄な時間と労力を費やし、誤った情報に誘導されることで攻撃の成功率が大きく低下する。防御側は比較的低コストでおとりを仕掛けられる一方、攻撃者には多大なコストとリスクを負わせられるため、攻撃者の費用対効果を悪化させ、「非対称な優位性」を防御側に取り戻しやすくなる。

(3)　敵を知る「脅威インテリジェンス」の獲得と能動的な防御

サイバーデセプションは単なる検知ツールにとどまらない。安全に隔離されたおとり環境で攻撃者を泳がせることで、彼らの目的、使用するツール、そして具体的な手口（TTPs：戦術・技術・手順）をリアルタイムに観察・収集できる。ここで得られた自社を狙う生きた情報（脅威インテリジェンス）は、未知の脅威を明らかにし、検知ルールの追加・更新（SIEM/EDR の検知条件、監視対象の見直し）や、対策ルールの強化（アクセス制御、ブロック条件の追加）といった形で防御に反映できる。これにより、将来の攻撃に「起きてから対応」するのではなく、能動的な防御態勢を構築することが可能となる。

4.1.2　サイバーデセプションの進化

サイバーデセプションは強力な効果を持つ一方、かつては攻撃者の TTPs を観測するための研究用途の「ハニーポット」というイメージが強かった。しかし、約 30 年の歴史の中で、現在ではエンタープライズ向けの能動的防御手法の一つへと進化を遂げている。[1]

- **1990 年～2000 年代：**
  単独のハニーポットによる攻撃者の行動観察を目的としたものが多く、手動での展開が必要かつ既存のセキュリティツール（SIEM など）との統合がなく、熟練の攻撃者には偽物と見破られやすいという限界があった。
- **2000 年～2010 年代：**
  複数のハニーポットを相互接続した「ハニーネット」へと進化し、よりリアルなネットワーク環境を模擬できるようになった。これにより脅威インテリジェンスの収集、マルウェアのリバースエンジニアリング、行動分析といった研究目的としては強力なツールとなった。しかし、保守が非常に複雑で企業導入は限定的であり、静的な環境という面から攻撃者にパターン認識され回避される課題があった。
- **2010 年～現在：**
  2014 年頃からの情報セキュリティ市場の拡大に伴い、実用化に向けた進化を遂げた。単なるハニーポットの設置にとどまらず、本番ネットワークに統合された偽の認証情報や、攻撃者を誘導する痕跡（ルアー）などを組み合わせた高度な環境の自動展開が可能となった。SIEM、SOAR、EDR など既存のセキュリティ監視・運用基盤と連携し、高精度なアラートから隔離・対応までの自動化が実現したことで、単なる「早期検知ツール」から、攻撃者を積極的にコントロールするための「戦略的プロセス」として海外企業を中心に導入が進んでいる。

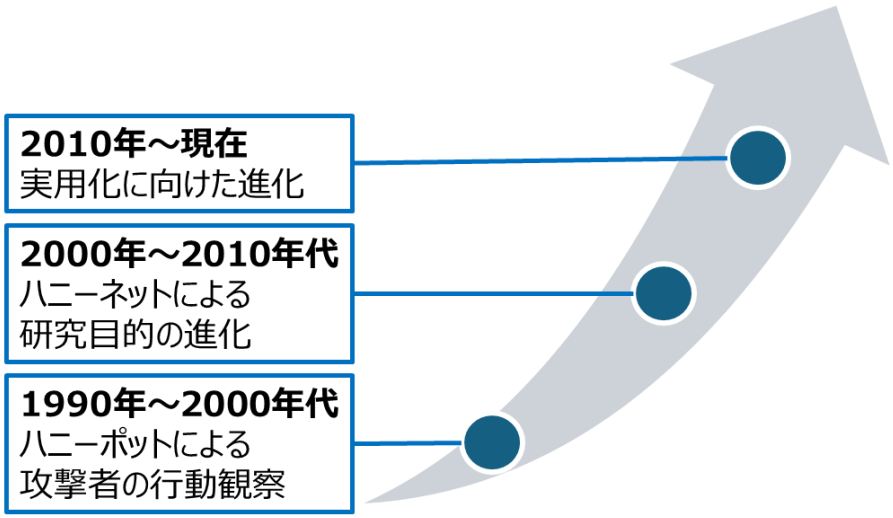


図 4-1　サイバーデセプションの進化

4.2 攻撃者心理から見たデセプションの有用性

サイバーデセプションを効果的な防御手段とするためには、おとりを無作為に配置するだけでは不十分である。攻撃者の「意思決定プロセス」に作用し、彼らの判断を防御側に有利な方向へ誘導するよう設計することが重要である。

4.2.1 攻撃者の OODA ループと限定合理性

攻撃者はシステムに侵入する際、「観察（Observe）、情勢判断（Orient）、意思決定（Decide）、行動（Act）」という意思決定のサイクル、いわゆる OODA ループを繰り返している。ここで重要なのは、攻撃者は常に完全合理的に判断する存在ではなく、焦り、思い込み、経験則に影響される人間である。人間は、限られた時間と情報の中で判断を下す際、すべての選択肢を論理的に検証できず、直感や経験則に頼ってしまう「限定合理性」という制約を持っている[9]。サイバーデセプションは、攻撃者の OODA ループ（特に観察と情勢判断のフェーズ）に意図的に設計した情報や環境を提示することで、この認知の限界に作用するものである。

攻撃者に作用し得る代表的な認知バイアスの例を下表に示す。

表 4-1　代表的なバイアス

名称	概要	攻撃者の行動
確証バイアス	自分の意見や仮説を肯定する情報ばかりを集め、反証する情報を無視する傾向	「すでに侵入に成功している」という思い込みから、おとり環境の不自然さ（おとりの兆候）を無視して探索を続けてしまう。
過信バイアス	自分の能力・知識、判断の正確さを、客観的な事実よりも高く見積もる傾向	「自分の技術なら検知されない」とスキルを過信し、警戒を怠っておとり（ハニートークンなど）に自ら接触してしまう。
楽観バイアス	「自分だけは悪いこと（病気や事故など）に巻き込まれないだろう」と根拠なく都合よく考える傾向	「この標的企業は高度なおとりなど仕掛けていないだろう」と都合よく考え、疑うことなく無警戒に行動してしまう。
正常性バイアス	異常事態が起きても、「これは正常の範囲内だ」「たいしたことはない」と思い込む傾向	意図的なセッション切断やエラー（偽）に遭遇しても、「単なるネットワーク不具合だ」と思い込み、攻撃を継続してしまう。
利用可能性ヒューリスティック	思い出しやすい情報（最近見たニュース、衝撃的な事件など）ほど、頻繁に起きると勘違いする傾向	過去の経験でよく見た「典型的な脆弱性」を発見すると、それが意図的に用意されたおとりであることを疑わずに飛びついてしまう。
代表性バイアス	「いかにもそれらしい」特徴に引きずられ、確率や論理を無視してしまう傾向	「admin」など高価値に"見える"偽アカウントに対し、それが無防備に放置されている確率の低さを無視して不合理に飛びついてしまう。

名称	概要	攻撃者の行動
アンカリング バイアス	最初に提示された数字や情報が基準（アンカー）となってしまい、その後の判断が歪められる傾向	最初に発見した「高権限アカウント」を過信し、その後に不自然な証拠があっても最初の判断に引きずられておとりの奥へ進んでしまう。
損失回避	利益を得る喜びよりも、損失による痛みを強く感じる傾向	苦労して得たアクセス権の喪失（偽のパスワードリセット通知など）を極端に恐れ、焦ってデータを持ち出そうとし、自ら検知されやすくなってしまう。
保有効果	自分が所有しているものの価値を、客観的価値よりも高く見積もる傾向（損失回避の派生型）	すでに侵入できたサーバーのアクセス権（実はおとり）の価値を過大評価し、それが偽物であると判断しづらくなってしまう。
現状維持 バイアス	未知の変化を受け入れるより、現在の状態を保とうとする傾向（変化による損失リスクを恐れるため）	新たな領域へ横展開するリスクを避けるため、現状自由に動けている「おとり環境」の中に留まり続けてしまう。
サンクコスト効果	すでに投資した時間やお金（埋没費用）を惜しみ、合理的な撤退ができなくなる傾向	応答が極端に遅いおとりに引っかかっても、「ここまで調査に時間をかけたから」と切断できず、リソースを浪費し続けてしまう。

4.2.2　MITRE ATT&CK 主要フェーズにおける認知バイアス

攻撃者行動を理解するための補助的視点として、MITRE ATT&CK の主要な攻撃フェーズで影響を与える認知バイアスについて説明する。なお、人間の判断には複数のバイアスが複合的に作用し得るため、ここでは代表的なフェーズにおいて主要に作用しやすいバイアスの例を挙げる。[10]

- **偵察/初期侵入フェーズ：**
  攻撃者が侵入の糸口を探す際、例えば「admin」といった名前のアカウントを見つけると、その最初の発見（アンカー）を過信し（アンカリングバイアス: Anchoring Bias）、それが罠である可能性を無視して価値が高いと思い込んでしまう（代表性バイアス）。
- **永続化/探索フェーズ：**
  侵入に成功した攻撃者は、苦労して手に入れたアクセス権を過大評価し、不確実な変化よりも現在の状態を維持しようと執着する（現状維持バイアス）。また、一度「ここは本物の環境だ」「有益な情報だ」と信じ込むと、その思い込みを支持する情報ばかりを都合よく解釈し、罠であることを示す不自然な点や矛盾する証拠を無意識に無視してしまう（確証バイアス）。
- **目的の実行フェーズ：**
  最終的なデータ窃取などの段階では、利益を得る喜びよりも、現在のアクセス権やデータを失うことを極端に恐れる心理が強く作用する（損失回避）。

4.2.3　攻撃者の分類とデセプションによる影響度

このように、認知バイアスは人間に普遍的に存在する弱点である。一方で、攻撃者のスキルや訓練のバックグラウンドによってはデセプションによる影響度は異なる点に留意する必要がある。

- **独学のハッカーやスクリプトキディ（既存ツールを流用する初心者層）：**
  不完全な情報に基づき衝動的に行動する傾向があり、前述したような認知バイアス（損失回避や代表性バイアスなど）に非常に陥りやすい。オープンソースのツールに頼ることが多く、基本的なおとりであっても有効に機能しやすい。
- **高度な標的型攻撃を行う攻撃者や組織的なサイバー犯罪集団などの熟練した攻撃者：**
  リスクをより正確に理解する専門的な訓練を受けており、環境のフィンガープリント（OS の挙動や設定の不一致など）を確認して、未熟なおとりを見破ろうとする傾向がある。ただし、人間である以上、バイアスを完全に排除することは困難である。例えば、「防御側の管理はずさんである」といった過信につけ込むことや、精巧に作り込まれた偽環境によって「探していた機密情報を見つけた」という確証バイアスを誘発することで、一定の認知的負荷や判断遅延を与え得る。歴史的にも、適切に設計された欺瞞は、高度な攻撃者に対しても有効に機能し得ることが、複数の事例から示唆されている。

4.2.4　心理を逆手に取る「ひったくり」シナリオ

本節では、認知バイアスなどの心理的要素が実際の防御においてどのように機能し得るかをシナリオとして例示する。ネットワークの深部に侵入し、慎重に探索を進めている攻撃者を仮定する。ここで防御側はデセプションを用い、「定期システムメンテナンスのため、本日深夜に全パスワードをリセットする」という偽の通知を掲示する。これを見た攻撃者は、「せっかく獲得したアクセス権が失われる」という強い焦り（損失回避）に駆られ、それまでの慎重な判断を崩しやすくなる。

焦った攻撃者は「とにかく取れるだけのデータを奪って逃げよう（ひったくり）」と決断し、探索を加速させる。その過程で、防御側が意図的に目立つ場所に配置した「admin-backup」という名前の偽のデータベース接続情報を発見する。通常であればおとりを疑う場面だが、焦りの中にある攻撃者は最初の発見に引きずられ（アンカリングバイアス）、おとりの可能性を十分に検討しないまま「admin」という名称から「探していたものだ」と思い込んでしまう（代表性バイアス）。結果として、攻撃者は自発的に隔離されたおとり環境（罠）へと踏み込み、防御側に確実な検知と無力化の機会を与えてしまう。

図 4-2　攻撃者心理を逆手に取る「ひったくり」シナリオ図解

【コラム 4】 なぜキズだらけのマグロが注目されるのか
― 実体験！マグロの競りから学ぶ”人の認知” ―

思いもよらないところに、デセプションを行う上で意識すべき“人の認知活動”が潜んでいるかもしれません。例えば、マグロの競りのような、一見するとデセプションから遠い場面にもです。

以前、豊洲市場でマグロの競りを見学した際、かつて競りの現場で買い手として目利きを重ね、日々の真剣勝負の中で市場を見てきた方から、少し意外な話を聞きました。

市場の競りに並ぶマグロの中には、表面にキズが多いものがあります。一見すると、キズの少ないマグロの方が価値は高そうに思えます。しかし実際には、キズの多いマグロほど買い手から注目されることがあるそうです。買い手が品質を確かめるために触ったり試食したりした結果、キズが増えることがあるからです。つまり、そのキズは「多くの買い手が関心を持った痕跡」として受け取られることがあります。

また、競りの場では、マグロそのもの以外の要素も判断に影響するといいます。例えば、買い手の服装や振る舞いです。派手な服装や堂々とした態度は、競りを取り仕切る「競り人」に対して、「自分は本気で買いたい」という意思や存在感を伝えるシグナルになるそうです。

このキズと服装の話に共通するのは、人が対象そのものだけでなく、「他者の関心を示す痕跡」や「重要そうに見えるシグナル」を手がかりに価値を推測している点です。

サイバー空間でも、似たようなことが起きているのかもしれません。

侵入後の攻撃者が、重要システムを連想させるサーバー名や、機密文書に見えるファイル、管理者が残したように見える認証情報を見つけたとしたら、そこに価値のある情報や権限があると考えるのではないでしょうか。

サイバーデセプションは、こうした人の認知や判断の仕方を利用します。攻撃者が価値あるものだと信じそうな痕跡やシグナルを意図的に設計することにより、その対象へのアクセスを検知したり、観測可能な環境へ誘導したりするのです。

もちろん、マグロの競りでは意図的にキズを付けて価値を偽装することは許されません。しかしサイバー空間では、攻撃者の認知や判断に働きかけること自体が防御側の有効な選択肢になり得ます。

（余白ページ）

第5章　サイバーデセプションの構想と展開

5.1　サイバーデセプションの戦略・作戦・戦術の位置づけ

ここまで述べてきたように、サイバーデセプションは、攻撃者の認知・判断・行動に働きかけ、防御側に有利な状況を作り出す考え方である。一方で、これを実務に落とし込む際には、経営判断としての目的、セキュリティ設計としての方針、現場での実装・運用を混同しないことが重要となる。

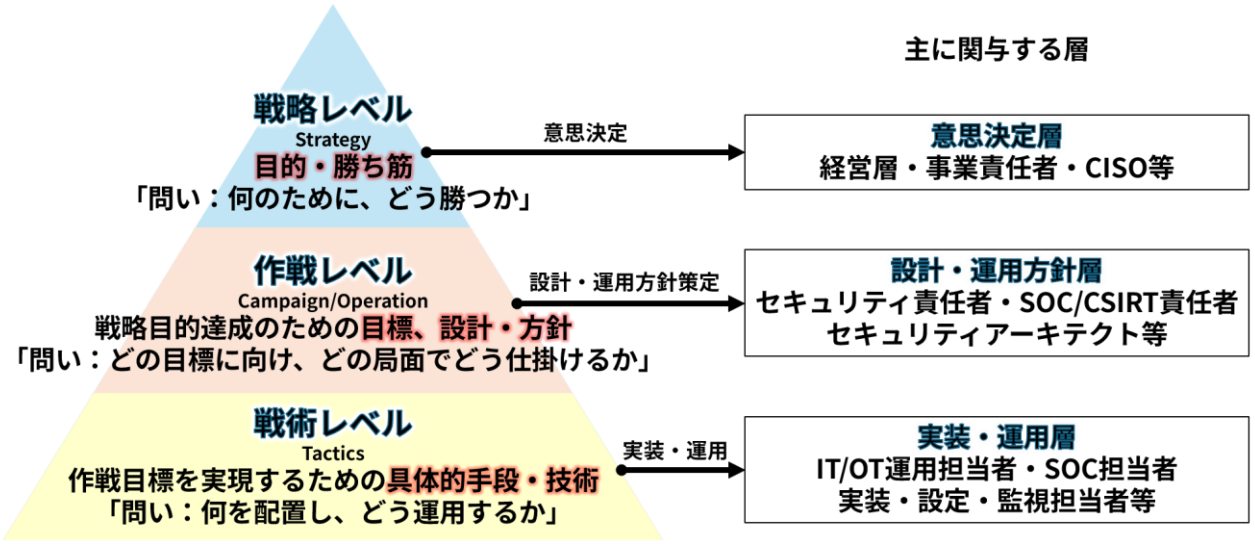


図 5-1　戦略・作戦・戦術の概念整理

戦略レベルでは、経営戦略や全社的なセキュリティ戦略の中で、サイバーデセプションをなぜ導入するのかを定義する。具体的には、守るべき重要資産、想定する脅威、許容できるリスク、期待する防御上の効果を明確にする。この段階では、投資判断やリスク許容度に関わるため、経営層、事業責任者、CISO などの意思決定層の関与が不可欠である。

作戦レベルでは、戦略で定めた目的を実現するために、攻撃者のどの行動に、どのような効果を与えるのかを定義する。例えば、攻撃者の存在を早期に露呈させるのか、探索や横展開を遅延・誘導するのか、あるいは攻撃者の TTPs を観測するのかを決める段階である。この段階では、MITRE Engage における「暴露させる（Expose）」、「影響を与える（Affect）」、「引き出す（Elicit）」のような作戦目標を参照しながら、どの攻撃フェーズに対して、どの環境でデセプションを作用させるかを検討する。主な関与者は、セキュリティ責任者、SOC/CSIRT 責任者、セキュリティアーキテクトなどである。

戦術レベルでは、作戦で定めた目標と方針を、具体的な技術、配置、監視、対応手順に落とし込む。例えば、ハニーポット、ハニーネットなどをどこに配置し、どのログを取得し、どの条件でアラートを発報し、誰がどのように対応

するかを定義する。この段階では、IT/OT 運用担当者、SOC 担当者、実装・設定・監視担当者などが中心となる。

なお、戦略・作戦・戦術は、担当者を完全に分断するための区分ではない。戦術上の制約や運用負荷は作戦方針に影響し、作戦上の実現可能性は戦略判断にも影響する。そのため、実際の導入検討では、意思決定層、設計・運用方針層、実装・運用層が相互にフィードバックしながら具体化する必要がある。本書では、この戦略・作戦・戦術の三層整理を前提に、次節以降で MITRE Engage および CYDEC Design Framework（CDF）を用いて、サイバーデセプションを目的設定から技術選定・運用設計までを一貫して具体化する。

5.2 サイバーデセプションの導入フレームワーク

攻撃者の認知的脆弱性（隙）に働きかけ、防御側に有利な状況を作り出すためには、戦略的なアプローチが不可欠である。この心理的・戦術的なアプローチを体系化し、組織的な防御プロセスとして導入・運用するためのフレームワークが「MITRE Engage[11]」である。本書では、『A Practical Guide to Adversary Engagement[12]』を参照し、その概要を整理したうえで、サイバーデセプション導入のイメージを提示する。

5.2.1 MITRE Engage の位置づけ：アドバーサリ・エンゲージメント

MITRE Engage は、単に「ハニーポットをどこに置くか」といったサイバーデセプション技術に特化したフレームワークではない。その中核は、防御側から攻撃者に対して能動的に関与する「アドバーサリ・エンゲージメント（Adversary Engagement：攻撃者との交戦）」という、より大きな戦略を体系化した点にある。

アドバーサリ・エンゲージメントとは、ファイアウォールやアクセス制御によって攻撃者の行動を物理的・論理的に妨げる「サイバー拒否（Cyber Denial）」と、意図的に虚偽の情報を提示して攻撃者を誤認・誘導する「サイバー欺瞞（Cyber Deception）」を、計画と分析に基づいて組み合わせて運用するアプローチである。この「防ぐ・隠す（拒否）」と「騙す・見せる（欺瞞）」を統合することで、攻撃者の費用対効果を増大させ、得られる情報や成果の価値を低下させる。これにより、防御側が「非対称な優位性」を取り戻すことが、MITRE Engage の狙いである。

5.2.2 3 つの作戦目標（Engage Goals）

MITRE Engage では、サイバーデセプションを含む運用を場当たり的にしないため、組織が目指す方向性として 3 つの「作戦目標（Engage Goals）」および「アプローチ（Approach）」を定義している。これらは導入効果を直感的に捉える指標となる。さらに、これら作戦目標に「準備（Prepare）」と「理解（Understand）」を含めることで、作戦の全体像を整理できる。

図 5-2　MITRE Engage の作戦目標とアプローチ

- 暴露（Expose）： 攻撃者の存在を確実かつ早期にあぶり出すこと。

  アプローチ　進行中の攻撃を捉えて高精度なアラートを出す「検知（Detect）」と、検知の裏側で攻撃者のツールや戦術に関する生データを安全に記録する「収集（Collect）」の 2 つで構成される。

  運用例　正規の従業員がアクセスしない階層に「偽の管理者 ID やパスワードリスト（ルアー）」を配置する。侵入した攻撃者がこれに触れた瞬間、確定的なアラートを発報し（Detect）、同時にネットワーク通信やシステムログを記録して、初期段階の動きを把握する（Collect）。

- 影響（Affect）： 攻撃者の活動に悪影響を与え、リソースを浪費させること。

  アプローチ　攻撃の進行を食い止める「阻止（Prevent）」、防御側が望む偽のシステム等へ向かわせる「誘導（Direct）」、無価値なダミーデータ（ポケットリッター）の処理に時間を費やさせる「妨害（Disrupt）」の 3 つを組み合わせることで、攻撃における費用対効果を悪化させる。

  運用例　内部不正者が機密データを持ち出そうとした際、本物のデータサーバーへのアクセスを見えない形で制限し（Prevent）、無価値なダミーデータが配置された偽セグメントへ誘導する（Direct）。さらにネットワーク帯域を意図的に絞って通信速度を遅延させ（Disrupt）、攻撃者に時間的・心理的コストとリスクを負わせる。

- 引き出し（Elicit）： 攻撃者の手口や目的を安全に引き出して学習すること。

  アプローチ　環境内に実在しそうな閲覧履歴や業務ファイルを散りばめて本物だと信じ込ませる「安心・確信（Reassure）」と、あえて古い脆弱なシステムを提示して行動を誘発する「動機付け（Motivate）」を組み合わせる。

  運用例　本番ネットワークから隔離された監視環境（ハニーネット）に、作りこまれたダミーの設計図を配置して警戒心を緩め（Reassure）、推測しやすいパスワード設定などで「簡単に侵入できる」と動機付ける（Motivate）。これにより、未知のマルウェア（ペイロード）を隔離環境内に投下させ、将来の防御に資する脅威インテリジェンス（TTPs）を引き出す。

5.2.3　MITRE Engage の 10 ステップから実践的設計フレームワークへ

これらの作戦目標を達成し、デセプションを「設置して終わり」の単なるツール導入にとどめないために、MITRE Engage は「準備（Prepare）」「運用（Operate）」「理解（Understand）」の 3 フェーズからなる 10 ステップのプロセスを提唱している。このプロセスは、自組織と敵の理解（Step 1）に始まり、運用目標の決定（Step 2）、攻撃者に「何を認識させ、どう反応させるか、どのように接触するか」の設計（Step 3-5）、ルールの策定（Step 6）、実行（Step 7）、そして収集データの分析・フィードバック（Step 8-10）までを網羅する。計画から運用、学習までを一連のサイクルとして回すための、実践的な設計フレームワークである。

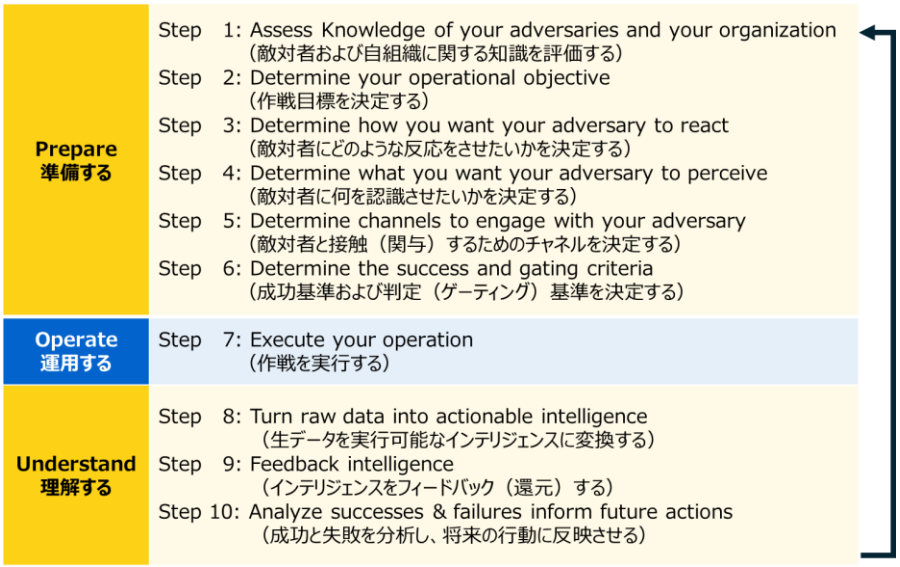


図 5-3　MITRE Engage 10 ステップ

しかし、この 10 ステップのプロセスは包括的で強力である一方で、企業が導入を検討する段階では「自社環境に即して、具体的にどのように設計・運用へ落とし込むか」という実務レベルの整理が難しい場合がある。

そこで次節では、MITRE Engage の考え方を踏まえつつ、サイバーデセプションをセキュリティ戦略に組み込むための設計・運用に活用できる実践的フレームワークとして、我々が考案した CYDEC Design Framework（CDF）を紹介する。

5.3　CYDEC Design Framework（CDF）の構想と設計体系

サイバーデセプションを実務に導入する際の難しさは、ハニーポットやハニートークンなどの技術そのものを選定することだけにあるのではない。より本質的な課題は、サイバーデセプションが、攻撃者の認知・判断・行動に働きかける能動的な防御であるという点にある。

サイバーデセプションでは、攻撃者に何を見せ、何を信じさせ、どのような行動を取らせるのかを防御側が設計する。そのため、単におとり環境を配置するだけでは不十分である。自組織が何を守るために、どの攻撃者に対して、どの攻撃フェーズで、どのような効果を期待してデセプションを用いるのかを、技術選定の前に整理する必要がある。

仮に、戦略上の目的、作戦上の目標、作用させる攻撃フェーズ、配置するレイヤー、具体的な技術が整合していなければ、サイバーデセプション技術は十分に機能しない。例えば、守るべき資産や想定する脅威が曖昧なままハニーポットを配置しても、攻撃者との接点が生じない可能性がある。また、アラートを発生させる仕組みを導入しても、運用側が対応できなければ早期検知の効果は十分に発揮されない。さらに、攻撃者に提示する情報や環境に一貫性がなければ、攻撃者におとりであると見破られ、誘導や観測の効果を失うおそれがある。

したがって、サイバーデセプションは、個別技術の集合としてではなく、セキュリティ戦略、サイバーデセプション作戦、具体的な戦術・技術を一貫して接続する設計体系として捉える必要がある。すなわち、上位のセキュリティ戦略で定めた目的を踏まえ、サイバーデセプション作戦として達成したい目標や作用させる攻撃フェーズを定め、最終的にハニーポット、ハニートークン、ブレッドクラムなどの具体的な手段へ落とし込むことが重要である。

一方で、日本におけるセキュリティ導入の実態を踏まえると、このような構想をユーザー企業側が自律的に描くことは容易ではない。IPA「DX 白書 2023」では、日本における情報処理・通信に携わる人材のうち、IT 企業に所属する割合は 73.6%、IT 企業以外に所属する割合は 26.4%である一方、米国では IT 企業に所属する割合が 35.1%、IT 企業以外に所属する割合が 64.9%であることが示されている[13]。また、経済産業省「DX レポート 2.1」では、ユーザー企業とベンダー企業の現在の相互依存関係を「低位安定」と位置づけている。具体的には、ユーザー企業が IT をベンダー企業任せにすることで IT 対応能力が育たず、IT システムのブラックボックス化やベンダーロックインにより、経営のアジリティが低下することが指摘されている[14]。

これらの指摘は、DX に関する文脈で示されたものであり、直ちにサイバーセキュリティ領域全般やサイバーデセプションの導入状況にそのまま当てはめられるものではない。しかし、日本では IT 人材が IT 企業側に偏在していることを踏まえると、ユーザー企業が高度な IT・セキュリティ施策を自律的に構想しにくい状況が生じやすいと考えられる。その結果、サイバーデセプションの検討も、ユーザー企業自身の目的や守るべき資産から始まるのではなく、ベンダーが提示する製品やソリューションを起点に進みやすい。

この問題意識から、本書では CYDEC Design Framework（CDF）を提案する。CDF は、サイバーデセプションを製品起点ではなく、ユーザー企業側の目的・課題・脅威認識を起点として設計するためのフレームワークである。具体的には、Who（主体）、What（資産）、by Whom（脅威）によって前提を整理し、Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）、How（手段）によって、セキュリティ戦略、サイバーデセプション作戦、具体的な戦術・技術を接続する。

ここでいう構想とは、製品選定の前段階で、何を守るのか、誰を脅威アクターとして想定するのか、どの攻撃段階で作用させるのか、どのような効果を狙うのかを整理することである。CDF は、その構想を可視化し、戦略から作戦、戦術までを一貫して設計するためのフレームワークとして位置づけられる。

本フレームワークは、導入検討の支援にとどまらず、企業がサイバーデセプションをセキュリティ戦略にシームレスに組み込み、設計・運用・見直しまで一貫して扱える実践的フレームワークとして活用できる。

5.3.1　CYDEC Design Framework（CDF）の全体像

下図に示す「Who/What/by Whom/Why/When/Where/How」の 7 項目は、左から右へ順に整理・具体化を進めることで、論点を網羅しつつ、組織のセキュリティ戦略に合致したデセプション設計を効果的に組み立てられるように構成されている。全体は大きく「前提条件の整理」「作戦の立案」「戦術への落とし込み」の 3 区分で整理される。

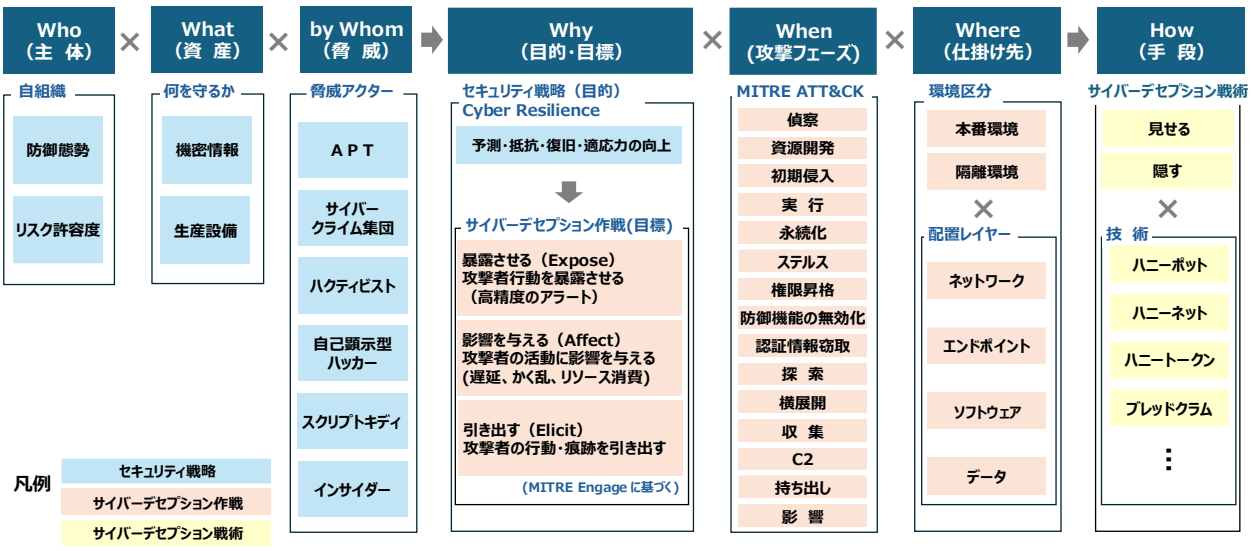


図 5-4　CYDEC Design Framework（CDF）

5.3.2　Who/What/by Whom（前提：リスク分析）

本フェーズでは、「誰が（自組織の要件）」「何を守るか（重要資産）」「誰に対抗するか（想定する攻撃者）」を明確化する。これはサイバーデセプション固有の作業というより、セキュリティ戦略全体に共通するリスク分析の領域に該当する。以降の設計の前提となるため、まずこの三点を確定させることが出発点となる。

サイバーデセプションを効果的に機能させる第一歩は、自組織と攻撃者に関する深い理解（脅威モデリング）である。本段階では、以下の三つの観点を順に整理することで、おとりを配置する前の戦略的前提を強固にする。

Who（自組織の現状と課題の把握）

自社のリスク分析やセキュリティ態勢の現状を把握し、運用上の課題を明確化する。例えば、既存ツールに起因する「誤検知の多さによるアラート疲れ」や、境界防御を突破された後の「内部ネットワークの可視性不足」などの痛点を整理する。これらの課題が明確になることで、サイバーデセプション導入の価値（確度の高い不正検知、調査負荷の低減など）を経営層や運用チームに説明するための根拠が整う。

What（守るべき重要資産の特定）

事業継続に不可欠なミッション/ビジネス機能と、攻撃者が狙う価値の高い重要資産（High Value Assets：HVA）を特定する。認証基盤（Active Directory など）、顧客データベース、知的財産、OT/ICS の制御システムなどが該当する。守るべき「本物」の位置と価値を把握してはじめて、攻撃者の注意を逸らす「おとり（罠）」を、どこに・どの程度のリアリティで配置すべきかを設計できる。

by Whom（想定される脅威と攻撃者像のプロファイリング）

重要資産を標的とする可能性の高い攻撃者（ランサムウェア攻撃グループ、APT アクター、内部不正者など）をプロファイリングする。ここでは手法の列挙にとどまらず、「4.2　攻撃者心理から見たデセプションの有用性」で述べた「認知バイアス」や「攻撃者の脆弱性」の観点から、心理的・行動的特徴も含めて整理することが重要である。例えば「侵入後に焦って管理者権限を探索する」「自動化ツールに依存して広範にスキャンする」などの隙を具体化する。

以上のように、「Who（自社の現状と課題）」を起点に、「What（守るべき重要資産）」と「by Whom（想定される脅威と攻撃者像）」を対応付けて整理することで、後続の「Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）」の作戦方針の決定および「How（手段）」の設計・技術選定が明確になる。

5.3.3　Why/When/Where（サイバーデセプション作戦の立案）

前提条件を踏まえ、「なぜサイバーデセプションを行うのか」を明確にする。セキュリティ戦略上の観点で、NIST SP 800-160 Vol.2 が示す「サイバーレジリエンス（事業継続と回復力）の向上」を目的に据える。そのうえで、具体的な運用上の作戦目標として、MITRE Engage が定義する以下の 3 つの「作戦目標（Engage Goals）」から、自組織が目指す方向性を選択する。

- 暴露させる（Expose）：攻撃者の隠密行動をあぶり出し、高精度なアラートを発報する。
- 影響を与える（Affect）：攻撃者の活動を遅延・妨害・誘導し、リソース（時間・労力・コスト）を浪費させる。
- 引き出す（Elicit）：攻撃者の手口や、未知の脅威インテリジェンス（TTPs）を引き出す。

定めた目標（Why）を達成するために、想定する攻撃者の攻撃フェーズ（When）に対して、どの環境/レイヤー（Where）でデセプションを作用させるかという作戦方針を決定する。

5.3.4　How（設計：サイバーデセプション戦術への落とし込み）

定めた作戦目標と方針（Why/When/Where）を軸として、「虚構を見せるのか、事実を隠すのか」の方針を決め、具体的なデセプション手段を選定する。これにより、場当たり的なおとりの配置を避け、攻撃者の行動上の「隙」に対して狙いを定めた設計を行える。

5.3.5　「設置して終わり」を防ぐライフサイクル管理

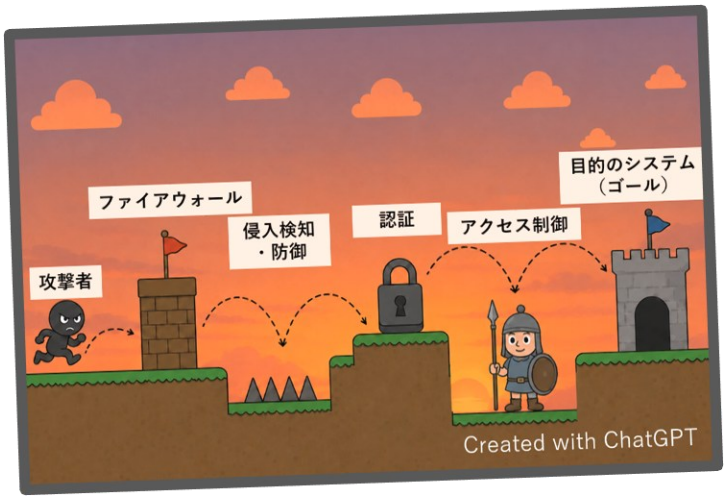
サイバーデセプションは、一度おとりを設置して終わりという性質のものではない。攻撃者の戦術は常に変化するため、CDF を用いて設計した環境も、継続的な監視と見直しが必要となる。収集した脅威情報や導入効果を評価したうえで、その結果を自社のリスク分析（Who/What/by Whom）へフィードバックし、作戦（Why/When/Where）や戦術（How）を更新する。この「ライフサイクル管理」を継続して回すことで、組織のサイバーレジリエンスは継続的に維持向上していく。

本節では、サイバーデセプションを戦略的に設計するための CDF の全体像と運用プロセスを示した。次章では、本フレームワークの「When/Where/How」に該当する具体的なサイバーデセプション技術の詳細と、それらを組み合わせた実践的な運用シナリオについて解説する。

【コラム 5】 ゲームで考えるサイバーデセプション

ゲームを難しくする方法は、壁を高くしたり、障害物を増やしたりするだけではありません。

例えば、横スクロール型のアクションゲームを思い浮かべてみてください。プレイヤーはスタート地点からゴールを目指し、途中に配置された敵や障害物を一つずつ突破していきます。障害物の配置や難易度に違いはあるものの、基本的な構造は変わりません。



「前へ進む」 → 「障害物を突破する」 → 「ゴールへ到達する」

これは従来型のサイバー防御にも通じる構造です。攻撃者は外部から侵入し、ファイアウォールや認証、アクセス制御といった防御機構を回避または突破しながら、目的のシステムを目指します。

もちろん、防御壁を強固にすることは重要です。しかし、攻撃者から見れば、それらは侵入経路上に存在する障害物として認識されます。攻撃者は収集した情報や利用可能な手法をもとに、どの経路を選択するかを判断しながら目的達成を目指します。

では、もしダンジョン探索型のゲームだったらどうでしょうか。

ダンジョンでは、どこに価値のあるアイテムがあるのか分かりません。近道だと思った通路が危険地帯につながっていることもあります。一方で、遠回りに見える通路が、実は有利な経路であることもあります。宝箱だと思ったものが罠かもしれませんし、思わぬ場所に有益なアイテムが隠されていることもあります。また、一見安全そうな場所に強力な敵が潜んでいることもあります。

プレイヤーは限られた情報をもとに意思決定を行いますが、その判断材料が必ずしも正しいとは限りません。そのため、重要なのは単に障害物を突破する能力だけではなく、どの経路が安全なのか、どこに価値があるのか、どの選択がリスクを伴うのかを見極める能力です。

サイバーデセプションの考え方もこれに近いものです。

防御側はハニーポットや偽の認証情報などを配置することで、攻撃者が認識する環境そのものを変化させます。攻撃者は本物のサーバーを発見したと認識しても、それが実際には罠である可能性があります。価値のある認証情報を入手したと判断しても、それは誤った経路への誘導や、利用を検知するための仕掛けである場合があります。

このように、攻撃者は常に「その情報や資産が本物なのか」「この経路は本当に安全なのか」を考えながら行動しなければならなくなります。結果として、調査や意思決定に時間を要し、誤った経路を選択するリスクや、行動を検知・観測されるリスクが高まります。防御側は攻撃を単に阻止するだけでなく、攻撃者を迷わせ、観測し、望ましい方向へ誘導することが可能になります。

従来型の防御が「壁を高くしたり、障害物を増やしたりするゲーム」だとすれば、サイバーデセプションは「ダンジョンを設計するゲーム」です。

ダンジョンゲームがプレイヤーの判断力を試すように、サイバーデセプションは攻撃者の認知・判断・行動に働きかけます。重要なのは壁を高くすることだけではなく、攻撃者の目に映る環境そのものを設計することです。

その発想の転換こそが、サイバーデセプションの重要な価値の一つと言えるのではないでしょうか。

第6章　サイバーデセプションの設計・運用

前章では、サイバーデセプションをセキュリティ戦略に組み込み、作戦・戦術へと具体化するための考え方として、MITRE Engage および CYDEC Design Framework（CDF）を整理した。本章では、その考え方を踏まえ、戦略上の目的に基づいてサイバーデセプション作戦の目標を定め、具体的な設計・運用へ落とし込むための要点を整理する。サイバーデセプションは、単にハニーポットやハニートークンなどの技術を配置すれば機能するものではない。攻撃者にどのような認知を与え、どの判断を促し、どの行動を検知・誘導・遅延・観測するのかを、作戦として設計する必要がある。そのためには、作戦目標、攻撃フェーズ、仕掛け先、採用する手段、既存対策との連携、運用体制を一貫して整理することが重要となる。また、サイバーデセプションは一度導入して終わる施策ではなく、安全性を確保しながら継続的に見直し、組織の防御力向上に接続していく運用プロセスとして捉える必要がある。

6.1　設計・運用プロセスの全体像

本章では、サイバーデセプションの設計・運用を、「基本設計」、「詳細設計」、「実装・運用」、「評価・改善」の流れで整理する。

基本設計では、CDF を用いて、サイバーデセプション作戦の立案と戦術への落とし込みを行う。具体的には、Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）を整理し、どのような効果を、どの段階で、どこに対して発揮するのかを明確にする。そのうえで、How（手段）として採用するサイバーデセプション技術を検討し、配置方針、既存対策・監視基盤との連携方針、運用方針を整理する。

詳細設計では、基本設計で選定した技術要素について、作戦目標に照らしてどのように機能させるかを具体化する。本書はサイバーデセプションの考え方を広めることを主な目的としているため、個別製品の設定値や実装手順までは深く扱わない。ここでは、各技術の該当要件、攻撃者の弱点・心理、導入効果、構成パターンの大枠、設計・配置のポイントを中心に確認する。なお、ハニーポット、ハニーネット、ハニートークンなどの技術概要や仕組みの詳細は、Appendix に整理する。

実装・運用では、設計したサイバーデセプション環境を安全かつ継続的に機能させるため、倫理原則、安全性、環境分離、監視、証跡保全、既存 SOC/CSIRT との連携、OT/ICS 環境における制約などを考慮する。サイバーデセプションは、単にサイバーデセプション資産を配置して終わりではなく、接触を検知し、判断し、対応し、必要に応じて更新・撤去する運用まで含めることで成立する防御手法である。

評価・改善では、検知、誘導、遅延、観測といった効果が実際に得られているか、運用負荷や安全性に問題がないかを確認する。攻撃者の手法や自組織の環境は継続的に変化するため、サイバーデセプションも一度設計して終わりではなく、運用結果を踏まえて継続的に見直す必要がある。

以上を踏まえ、本章では、まず基本設計における配置方針・連携方針・運用方針を整理し、次に詳細設計として技術要素ごとの設計観点を示す。その後、実装・運用上の原則と留意点、評価・改善の考え方を整理する。

6.2 基本設計：配置方針・連携方針・運用方針の整理

「5.3.2　Who/What/by Whom（前提：リスク分析）」では、サイバーデセプションを活用したセキュリティ戦略を検討する際の前提条件として、Who（主体）、What（資産）、by Whom（脅威）の整理の必要性について述べた。これらは組織におけるリスク分析や脅威モデリングの根幹であり、サイバーデセプションを設計する上で強固な基礎となる。

実際にサイバーデセプションを導入・運用し、その効果を最大限に発揮させるためには、三つの前提条件を踏まえた上で具体的なサイバーデセプション作戦を立案し、実効性のあるサイバーデセプション戦術に落とし込んでいく一連のプロセスが不可欠である。

そこで本節では、本書が提言する CDF を用いて、サイバーデセプションにおける作戦目標（Why）および作戦方針（When/Where）の策定方法と、戦術（How）への具体的な落とし込み手順について解説する。

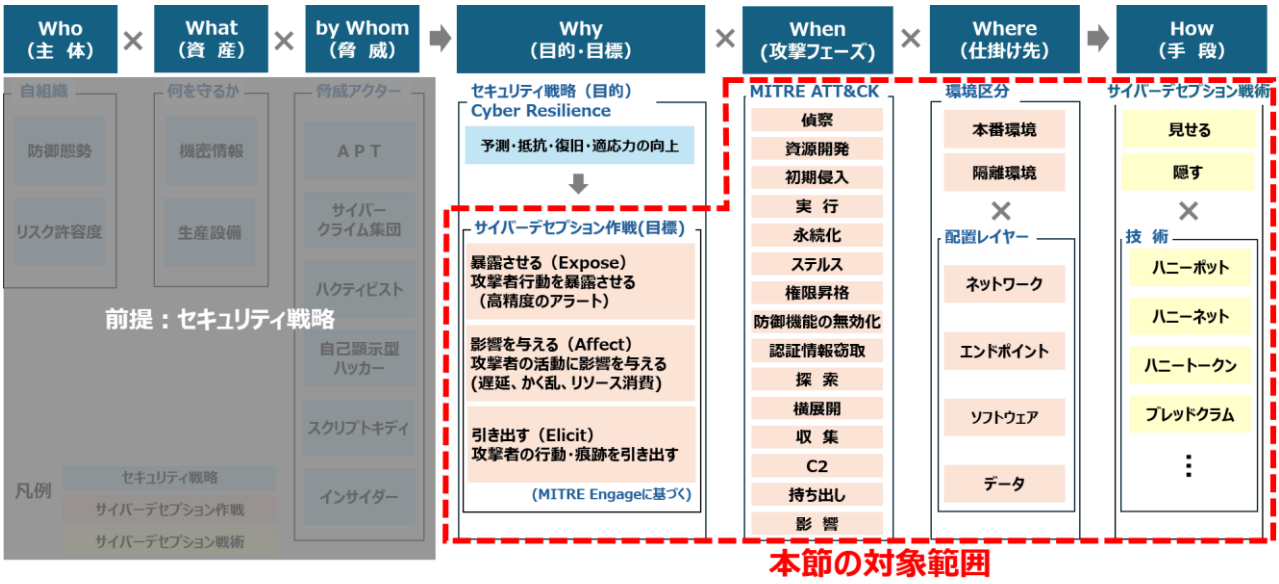


図 6-1　サイバーデセプション作戦・戦術の設計対象範囲

本書では、サイバーデセプションの設計を、プロセス 1：サイバーデセプション作戦の立案とプロセス 2：サイバーデセプション戦術への落とし込みの二段階で進めることを提案する。CDF に基づき、プロセス 1 では Why/When/Where を整理し、達成したい効果や適用場面を明確化する。プロセス 2 では How を整理し、作戦を実現するための戦術として具体的な技術選定や配置方法への落とし込みを行う。

本節では、サイバーデセプション作戦を立案し、単一のサイバーデセプション戦術に落とし込むまでの流れを、プロセス 1・2 に沿って示す。なお、実際の運用では単一の技術だけでなく、複数の技術を組み合わせることで攻撃者の異なる行動を多層的に検知・誘導することで、サイバーデセプションの効果を高めることができる。例えば、境界に配置したハニーポットで偵察行為を検知しつつ、内部のハニートークンで横展開の痕跡を顕在化させる、といった構成が考えられる。

一方で、技術の組み合わせが増えるほど監視やメンテナンスなどの運用負荷も増大する。攻撃者に気付かれにくい配置の維持、偽装情報の鮮度管理、アラートの統合監視などを継続できる範囲を見極め、自組織の運用体制、人員、スキルおよび成熟度に見合った規模から段階的に拡張することを推奨する。

6.2.1　サイバーデセプション作戦の立案（Why/When/Where）：プロセス 1

プロセス 1 では、「5.3.2　Who/What/by Whom（前提：リスク分析）」で整理した Who（主体）、What（資産）、by Whom（脅威）を踏まえ、自組織の運用体制、監視成熟度、法務上の制約および許容リスクの範囲内でサイバーデセプション作戦を立案する。具体的には、Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）の三つの観点から作戦を整理し、どのような効果を、どの段階で、どこに対して発揮したいのかを明確化する。

【Why（目的・目標）】

Why の整理には、MITRE Engage の作戦目標（Engage Goals）である「暴露させる（Expose）」、「影響を与える（Affect）」、「引き出す（Elicit）」および、作戦目標（Engage Goals）に紐づくアプローチ（Engage Approaches）を参照する。本書では、MITRE Engage における作戦目標（Engage Goals）をサイバーデセプション作戦で達成したい目標（効果）として扱い、CDF における Why（目的・目標）を具体化する際の参照概念として用いる。プロセス 1 の成果として達成したい作戦目標（Engage Goals）を決定するとともに、その実現に有効なアプローチ（Engage Approaches）の候補を下表に整理する。

表 6-1　作戦目標（Goals）とアプローチ（Approaches）の概要

#	作戦目標（Goals）	アプローチ（Approaches）	概要
1	暴露させる（Expose）	収集（Collect）	攻撃者やその活動に関する痕跡・ログ・通信を収集し、攻撃実態の把握や後続分析に資する情報を取得する。
2		検知（Detect）	ルアーやカスタムアラートなどを用いて攻撃者の活動を高精度に監視し、その存在や行動を早期に露呈させる。
3	影響を与える（Affect）	阻止（Prevent）	資産の無効化や防御制御の強化などにより、攻撃者が意図した作戦を遂行できないよう停止させる。
4		誘導（Direct）	攻撃者の行動経路や標的選択に影響を与え、防御側が望む場所や行動へ導く。
5		妨害（Disrupt）	誤情報の付与や攻撃資源の浪費を誘発することで、攻撃者の意思決定、作業効率、攻撃活動の継続を阻害する。
6	引き出す（Elicit）	安心・確信（Reassure）	おとり環境に現実味のある要素を付与しておとり環境の信頼性を高め、攻撃者の疑念を抑制し、滞留や定着を促す。
7		動機付け（Motivate）	攻撃者にとって価値のある情報や侵入しやすい環境を用意し、防御側が期待する行動を引き出す。

作戦目標（Engage Goals）の選定にあたっては、by Whom（脅威）の行動様式や What（資産）の保護優先度を踏まえる必要がある。また、自組織の運用能力を超える目標設定（例：高度な脅威インテリジェンスの継続的収集や 24 時間 365 日の詳細な攻撃者行動分析）は避け、実現可能な範囲から段階的に発展させることが望ましい。

以下に、各作戦目標（Engage Goals）およびアプローチ（Engage Approaches）の概要と選定時の参考情報を示す。なお、本節における導入の目安は、必要となる運用体制、専門知識、監視能力および環境構築難易度を総合的に評価した参考値である。

[暴露させる（Expose）]

- 作戦目標（Engage Goals）

MITRE Engage における定義は、進行中の敵対作戦の存在を明らかにすること（Reveal the presence of ongoing adversary operations.[15]）である。

これまで把握できていなかった攻撃者の活動を顕在化し、その存在や行動を明らかにすることを目標とする。例えば、偵察、初期侵入、認証情報窃取、横展開、C2 通信などの活動が検知対象として挙げられる。アプローチ（Engage Approaches）である「収集（Collect）」と「検知（Detect）」の併用は推奨されるが、必須ではない。

- ユースケース
  - 内部不正や外部侵入の兆候把握
  - ネットワーク内で活動する攻撃者の可視化
  - 既存監視では把握できない攻撃行動の検出
- 導入の目安：中

一般的な SOC 運用およびログ監視体制を前提とし、ハニーポットやハニートークンなどのサイバーデセプション資産を構築・運用するスキルが求められる。

- アプローチ（Engage Approaches）
  - 収集（Collect）

    攻撃者の活動に関するデータを取得するアプローチ。システムログ、ネットワークトラフィック、パケットキャプチャ、API 監視ログなどを収集し、後段の検知、分析、脅威インテリジェンス化に必要な情報を確保する。「検知（Detect）」や「安心・確信（Reassure）」などと組み合わせることで、通常環境では観測が困難な攻撃行動を把握できる。
  - 検知（Detect）

    攻撃者の存在や活動を高精度に検知するアプローチ。ハニートークンの利用、ハニーポットへの接触、おとり環境向け通信の検知・監視などを通じて攻撃の兆候を早期に把握する。正規利用者が通常接触しないサイバーデセプション資産を利用するため誤検知が少なく、「収集（Collect）」で得た情報と組み合わせることで検知精度を高めることができる。

[影響を与える（Affect）]

- 作戦目標（Engage Goals）

MITRE Engage における定義は、攻撃者の作戦に悪影響を与えること（Negatively impact the adversaries operations.[15]）である。

攻撃活動の遂行に必要な時間、労力、リソースを増加させる、あるいは攻撃によって得られる情報や成果の価値を低下させることで、攻撃の費用対効果を悪化させることを目標とする。なお、MITRE Engage における「影響を与える（Affect）」は防御側ネットワーク内での活動を対象としており、攻撃者側環境への侵入やハックバックは含まれない。

- ユースケース
  - レガシーシステムなどの重要資産保護
  - ランサムウェア感染後の横展開抑制
  - 攻撃者の調査・探索活動の非効率化
- 導入の目安：高

ネットワーク制御や環境設計が必要となるほか、運用中システムへの影響評価やプライバシーや通信の秘密に抵触しないかといった法務面の検討も求められる。

- アプローチ（Engage Approaches）
  - 阻止（Prevent）

    攻撃遂行能力の一部または全部を停止させるアプローチ。シンクホールによる悪性通信の遮断、ハニートークン接触を契機としたアカウント停止、おとり環境からの外部通信遮断などにより、攻撃の継続や拡大を防止する。
  - 誘導（Direct）

    攻撃者の行動経路や標的選択に影響を与え、防御側が管理しやすい環境へ誘導するアプローチ。ブレッドクラムやトラフィックリダイレクトによるおとり環境への誘導、ハニートークン（偽認証情報、偽接続情報など）の配置などが該当する。
  - 妨害（Disrupt）

    攻撃者の判断や作業効率を低下させるアプローチ。タービットによる通信遅延、誤情報の提示、おとり環境への誘導、ポケットリッターによる無益な探索の誘発などにより、攻撃継続に必要なコストを増大させる。「阻止（Prevent）」が攻撃の停止を狙うのに対し、「妨害（Disrupt）」は攻撃の継続を許容しながら効率を低下させる点が特徴である。

[引き出す（Elicit）]

- 作戦目標（Engage Goals）

MITRE Engage における定義は、攻撃者の戦術・技術・手順について学ぶこと（Learn about adversaries tactics, techniques, and procedures （TTPs）.[15]）である。

防御側が管理するおとり環境において、攻撃者の行動を促し、通常環境では観測できない TTPs や意思決定過程を引き出すことで、脅威インテリジェンスの取得や防御能力向上につなげる。

- ユースケース
  - 高度な攻撃者の TTPs 収集
  - 攻撃者の行動分析
  - 脅威インテリジェンスの獲得
- 導入の目安：高

脅威アクター分析、シナリオ設計、高対話型ハニーポットやハニーネットの構築・運用能力が必要となる。

- アプローチ（Engage Approaches）
  - 安心・確信（Reassure）

    おとり環境の本物らしさを高め、攻撃者に実環境であると認識させるアプローチ。ハニートークン（偽ユーザーアカウント、偽業務文書、偽操作履歴、偽システムアクティビティなど）を配置し、必要に応じておとり環境に生活感をもたせる Burn-In 手法を活用することで攻撃者の疑念を低減し、おとり環境への滞留を促進する。

- 動機付け（Motivate）

  攻撃者にとって魅力的な情報や環境を提示し、特定の行動を引き出すアプローチ。おとり環境内で意図的に脆弱な設定を再現することや、推測容易なハニートークン（偽認証情報）の配置などの方法がある。これにより攻撃者の探索や侵害行動を誘発し、追加の TTPs を観測できる。ただし、十分に隔離された環境で運用し、踏み台化による悪用防止などの安全面のリスク管理を行うことが前提となる。

【When（攻撃フェーズ）】

When（攻撃フェーズ）では、攻撃の進行段階のうち、サイバーデセプションをどの段階で作用させるかを明確化する。本書では、When の選定にあたり、MITRE ATT&CK Enterprise v19[16]で定義される Tactics（戦術） を参照する。

具体的な When の選定にあたっては、まず「表 6-2　Why（目的・目標）と When 候補（攻撃フェーズ）の対応目安」を活用し、Why（目的・目標）で定めた作戦目標とアプローチに対して技術的に効果の高い When の候補を抽出する。その上で、「5.3.2　Who/What/by Whom（前提：リスク分析）」で整理した自組織の前提（Who/What/by Whom）を考慮し、以下のように候補を絞り込んで最適化を行う。

- Who（主体）
  自組織の SOC 運用体制や監視成熟度に応じ、早期警戒（初期侵入・探索時など）にリソースを集中させるか、あるいは対処（横展開時など）まで含めるかを判断する。
- What（資産）
  守るべき重要資産の特性（OT 環境やレガシーシステムなど）に応じて、適切な攻撃フェーズを選択する。
- by Whom（脅威）
  想定される脅威（APT、ランサムウェア集団など）が多用する特有の戦術・技術・手順（TTPs）を踏まえ、作用させる攻撃フェーズを選定する。

表 6-2　Why（目的・目標）と When 候補（攻撃フェーズ）の対応目安

#	Why（目的・目標）		When 候補 （攻撃フェーズ）
	作戦目標 （Goals）	アプローチ （Approaches）	
1	暴露させる （Expose）	収集（Collect）	偵察、初期侵入、実行、永続化、認証情報窃取、 探索、横展開、収集、C2、持ち出し
2		検知（Detect）	初期侵入、実行、永続化、ステルス、 防御機能の無効化、認証情報窃取、探索、 横展開、収集、C2、持ち出し
3	影響を与える （Affect）	阻止（Prevent）	初期侵入、認証情報窃取、横展開、C2、 持ち出し、影響
4		誘導（Direct）	初期侵入、探索、横展開、収集、C2
5		妨害（Disrupt）	偵察、初期侵入、ステルス、防御機能の無効化、 探索、横展開、収集、C2、持ち出し、影響
6	引き出す （Elicit）	安心・確信（Reassure）	実行、探索、横展開、収集
7		動機付け（Motivate）	実行、永続化、権限昇格、認証情報窃取、探索、 横展開、収集、C2

※ 上表は MITRE の公式対応表ではなく、本書が設計時に用いる目安である。

また、When 候補（攻撃フェーズ）の選定にあたって参照する、MITRE ATT&CK Enterprise v19 で定義される戦術（Tactics）と攻撃者の行動例を下表に示す。

表 6-3　MITRE ATT&CK Enterprise v19 の戦術（Tactics）と攻撃者の行動例

#	ATT&CK 戦術	攻撃者の行動目的・概要	主な攻撃者の行動例
1	偵察（Reconnaissance）	将来の作戦計画のため情報を収集する	Whois、Shodan、検索エンジン、SNS 調査、Nmap などによる公開情報・ネットワーク構成の調査
2	資源開発（Resource Development）	作戦を支援するリソースを確立する	C2 サーバー構築、ドメイン取得、マルウェア作成、フィッシングサイト準備など
3	初期侵入（Initial Access）	標的ネットワークへの侵入を試みる	フィッシング、VPN 脆弱性悪用、公開サーバー侵害、認証情報悪用など
4	実行（Execution）	悪意あるコードを実行する	PowerShell（パワーシェル）、コマンドライン、スクリプト、マクロなどの実行
5	永続化（Persistence）	足がかりを維持する	レジストリ登録、スタートアップ設定、サービス登録、Web シェル設置など
6	権限昇格（Privilege Escalation）	より高いレベルの権限を取得する	権限昇格脆弱性、sudo 悪用、Pass-the-Hash など
7	ステルス（Stealth）	行動を隠蔽し、正規の挙動に見せかける	難読化、正規ツール悪用（Living Off the Land Binary）、通信の秘匿など
8	防御機能の無効化（Defense Impairment）	防御側の可視性・信頼性を損なう	ログ削除、セキュリティツールの無効化、EDR 停止、監視機能の妨害など
9	認証情報窃取（Credential Access）	アカウント名とパスワードなどを窃取する	認証情報ダンプ、キーロガー、LSASS ダンプなど
10	探索（Discovery）	環境を把握する	net コマンド、ipconfig、Active Directory 列挙など
11	横展開（Lateral Movement）	環境内を移動する	RDP、PsExec、SMB、WMI など
12	収集（Collection）	目的に沿ったデータを集める	ファイル探索、画面キャプチャ、メール収集、DB ダンプなど
13	C2（Command and Control）	侵害済みシステムを制御する	HTTPS、DNS トンネル、VPN、ビーコン通信など
14	持ち出し（Exfiltration）	データを窃取する	FTP、クラウドストレージ、暗号化通信など
15	影響（Impact）	システム・データを操作・中断・破壊する	ランサムウェアによる暗号化、データ削除、サービス停止など

【Where（仕掛け先）】

Where（仕掛け先）では、サイバーデセプションをどの配置レイヤーに設置するかを定義する。本書では、配置レイヤーをネットワーク層、エンドポイント層、ソフトウェア層、データ層の 4 層に大別している。

おとりの配置先（Where）を保護対象となる資産（What）と同一または近接するレイヤーに設定することは効果的である。一方で、その構成のみでは攻撃者との接点が生じにくく、接触機会が限られるおそれがある。そのため、When（攻撃フェーズ）に応じ、攻撃者が到達・探索しやすいレイヤーに配置することも重要となる。

具体的な Where の選定にあたっては、定義した When（攻撃フェーズ）に対して技術的に効果の高い仕掛け先の候補を抽出する。その上で、「5.3.2　Who/What/by Whom（前提：リスク分析）」で整理した自組織の前提（Who/What/by Whom）や実際のネットワーク構成、運用体制を考慮し、候補を絞り込んで最適化を行う。理論上有効と考えられる When（攻撃フェーズ）と Where（仕掛け先）の対応関係の目安を下表に示す。

表 6-4　When（攻撃フェーズ）と Where（仕掛け先）の対応目安

#	When （攻撃フェーズ）	有効と考えられる Where（仕掛け先）			
		ネットワーク	エンド ポイント	ソフトウェア	データ
1	偵察（Reconnaissance）	◎	△（#1）	△（#2）	△（#3）
2	資源開発（Resource Development）	△（#4）	—	△（#5）	—
3	初期侵入（Initial Access）	◎	○	○	△（#6）
4	実行（Execution）	△（#7）	◎	○	△（#8）
5	永続化（Persistence）	△（#9）	◎	○	△（#10）
6	権限昇格（Privilege Escalation）	△（#11）	◎	○	△（#12）
7	ステルス（Stealth）	△（#13）	◎	○	△（#14）
8	防御機能の無効化（Defense Impairment）	△（#15）	◎	○	△（#16）
9	認証情報窃取（Credential Access）	△（#17）	○	○	◎
10	探索（Discovery）	○	◎	○	○
11	横展開（Lateral Movement）	○	◎	○	○
12	収集（Collection）	△（#18）	○	○	◎
13	C2（Command and Control）	◎	○	△（#19）	—
14	持ち出し（Exfiltration）	◎	○	△（#20）	◎
15	影響（Impact）	○	◎	△（#21）	○

（凡例）◎：主として検討、○：併用・副次、△：条件付きで検討、—：通常は優先度が低い

△は、特定の運用条件下で有効となる場合を示す。主な条件例を「表 6-5　△（条件付きで検討）の主な条

件例」に示す。条件は、サイバーデセプション資産という罠を配置するだけでなく、それを監視・解析できる状態が整って初めて効果的な運用となるものが挙げられる。

上表は MITRE の公式対応表ではなく、本書が設計時に用いる目安である。

表 6-5　△（条件付きで検討）の主な条件例

#	When （攻撃フェーズ）	Where （仕掛け先）	△（条件付きで検討）とする条件例
1	偵察 （Reconnaissance）	エンドポイント層	公開サーバーや外部公開端末をサイバーデセプション資産として運用し、外部からの接触や探索を観測する場合
2	偵察 （Reconnaissance）	ソフトウェア層	公開 Web、VPN 入口、ログインポータルなどをサイバーデセプション資産として運用し、偵察・探索行為を観測する場合
3	偵察 （Reconnaissance）	データ層	偽文書、偽公開情報、ハニートークン入り資料などを配置し、攻撃者の事前調査や接触を観測する場合
4	資源開発（Resource Development）	ネットワーク層	自組織名義の偽ドメイン、偽インフラ、偽 DNS 情報などを用意し、攻撃準備段階の接触や調査を観測する場合
5	資源開発（Resource Development）	ソフトウェア層	偽サイト、偽ポータル、偽ログイン画面などを用意し、攻撃者の事前準備や誘導先への接触を観測する場合
6	初期侵入（Initial Access）	データ層	偽アカウント、偽認証情報、ハニートークンなどを配置し、侵入試行や認証情報の悪用を検知する場合
7	実行（Execution）	ネットワーク層	NDR などと併用し、スクリプト実行、外部通信、ペイロード取得などに伴うおとり環境との通信を観測する場合
8	実行（Execution）	データ層	実行後に攻撃者が接触する偽ファイル、偽設定情報、偽スクリプトなどを配置し、実行結果や後続行動を観測する場合
9	永続化（Persistence）	ネットワーク層	永続化に伴うビーコン通信、外部接続、設定変更後の通信などをネットワークで観測する場合
10	永続化（Persistence）	データ層	偽の設定ファイル、認証情報、起動スクリプト、タスク情報などを配置し、永続化に関連する探索・改変行為を観測する場合
11	権限昇格（Privilege Escalation）	ネットワーク層	おとり環境での活動（権限昇格に伴う認証通信、管理系通信、権限確認通信など）を境界または監視基盤で解析できる場合
12	権限昇格（Privilege Escalation）	データ層	偽の管理者用認証情報、設定ファイル、特権情報などを配置し、探索や悪用を観測する場合
13	ステルス（Stealth）	ネットワーク層	おとりへの暗号化通信、秘匿通信、通常通信に紛れた異常通信などをネットワーク監視で検知する場合
14	ステルス（Stealth）	データ層	偽ログ、偽設定情報、偽ファイル、ハニートークンなどを配置し、攻撃者の隠蔽・確認行動を観測する場合

#	When （攻撃フェーズ）	Where （仕掛け先）	△（条件付きで検討）とする条件例
15	防御機能の無効化（Defense Impairment）	ネットワーク層	通信、監視の途絶、ログ転送の異常をネットワーク側で観測する場合
16	防御機能の無効化（Defense Impairment）	データ層	偽ログ、偽監視設定、偽アラート情報などを配置し、無効化・改変行為を観測する場合
17	認証情報窃取（Credential Access）	ネットワーク層	認証トラフィック、認証試行、認証情報の利用先などをネットワークまたは認証基盤側で解析できる場合
18	収集（Collection）	ネットワーク層	ファイル収集、圧縮、集約、転送準備などに伴う通信やファイルアクセスをネットワーク側で観測する場合
19	C2（Command and Control）	ソフトウェア層	業務アプリ、Web アプリ、SaaS、API などを経由した C2 通信が主要脅威となる場合
20	持ち出し（Exfiltration）	ソフトウェア層	SaaS、業務アプリ、クラウドストレージ、API などを経由したデータ持ち出しが主要脅威となる場合
21	影響（Impact）	ソフトウェア層	業務アプリケーション、サービス、管理画面などに対する改ざん、停止、誤操作誘発などが主要脅威となる場合

6.2.2 サイバーデセプション戦術への落とし込み（How）：プロセス 2

プロセス 2 では、プロセス 1 で策定したサイバーデセプション作戦を具現化するため、How（手段）の段階で具体的な技術選定や実装方法へと落とし込みを行う。

How（手段）では、導入するサイバーデセプションの具体的な技術や実装方式を決定する。実務においては、プロセス 1 で定めた Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）を踏まえ、How(手段)を選定していく。なお、How（手段）の選定には「5.3.2　Who/What/by Whom（前提：リスク分析）」で整理した組織の前提（Who/What/by Whom）やリスク許容度と乖離がないか、全体的な整合性を改めて確認することが望ましい。

選定にあたっては、実務で採用されやすい代表的な設計の一例をまとめた「表 6-6　プロセス 1・2 に基づく設計例」が設計の参考となる。この表に自組織のシナリオに該当する例がない場合は、「6.3　詳細設計：技術要素ごとの設計観点」に記載する各技術の該当要件を参照し、対応する Why/When/Where から最適な技術候補を選定する。

技術候補の選定後は、「6.3　詳細設計：技術要素ごとの設計観点」の解説を参考に、具体的なシステム構成や詳細設計、および実環境への配置へと繋げていく。次節では、各技術における「攻撃者の判断傾向・心理」の突き方や、有効な「構成パターン」、実効性を高めるための「設計・配置のポイント」、運用上の「留意点」などを詳述している。これらを設計書やパラメータシートの作成に反映させることで、攻撃者の認知・行動・判断に効率的に作用し、組織のインフラ上で安全かつ確実に機能するサイバーデセプション環境の実装へとつなげることが可能となる。

表 6-6　プロセス 1・2 に基づく設計例

#	狙う効果	Why	When	Where	How
1	境界における偵察行為の情報収集	Expose（Collect）	【偵察】 Nmap などによるポートスキャン/サービス探索	【ネットワーク層】 インターネット、DMZ、外部公開サービス周辺	ハニーポット（低対話型、偽公開サービス）による偵察行為の観測
2	未知の侵害経路と初期侵入の早期検知	Expose（Detect/Collect）	【初期侵入】 VPN/Web ログイン画面、公開アプリへの認証や脆弱性悪用試行	【ソフトウェア層】 VPN ポータル、外部公開 Web、認証画面	ハニーポット（偽 VPN/偽ログイン画面）による侵入試行の検知
3	横展開行為の早期検知	Expose（Detect）	【横展開】 侵害端末からの他ホストへの接続試行	【データ層】 共有フォルダ・認証情報接点	ハニートークン（偽管理者認証情報、偽共有、偽接続情報）による横展開試行の検知
4	横展開の未然阻止	Affect（Prevent）	【横展開】 侵害端末からの他ホストへの拡大	【ネットワーク層】 境界、隔離経路	トラフィックリダイレクト（アクセス制御連動）による横移動経路の遮断 ※ハニートークンやハニーネットを併用することで、誤認誘発と封じ込めの効果を高めることも可能
5	偽情報による探索経路の誘導	Affect（Direct）	【横展開】 端末上の接続情報を用いた他ホストへの移動試行	【データ層】 ファイルサーバー、共有フォルダ	ブレッドクラム（偽のネットワーク構成図、偽のシステム設定ファイルなど）による攻撃者の探索活動の誘導

#	狙う効果	Why	When	Where	How
6	ネットワークスキャン・横展開通信の遅延化	Affect（Disrupt）	【偵察】【探索】 ホスト・ポート・サービス列挙 【横展開】 SMB/RDP などの接続試行	【ネットワーク層】 境界、 内部セグメント	タービット（TCP/SYN 応答遅延など）による接続維持・応答遅延、攻撃者リソースの消耗 ※低対話型ハニーポットの大量配置を併用すると、探索行動のさらなる遅延が可能
7	攻撃目標を達成したと誤認させる	Affect（Disrupt）	【収集】【持ち出し】共有フォルダ内ファイルの探索/取得	【データ層】 ファイルサーバー、 共有フォルダ	シンセティックデータ（偽財務データ、偽顧客データなど）による攻撃者の攻撃目標の誤認
8	攻撃者の関心・高度な TTPs の誘発	Elicit（Reassure/Motivate）	【実行】【探索】 【収集】 隔離環境内での調査、文書取得、コード実行	【データ層】 端末、ファイル、 共有フォルダ 【エンドポイント層】 内部おとり環境上の高対話型ホスト	シンセティックデータ・ポケットリッター（業務文書風ダミー）による関心誘発・価値認識、高度な TTPs の誘発 および、ハニーポット（高対話型）による攻撃者の操作・TTPs 収集
9	偽認証情報による次行動の誘発	Elicit（Motivate）	【認証情報窃取】 ファイル/メモリなどからの認証情報/API キー窃取 【横展開】 窃取認証情報による他システムへの接続	【データ層】 端末、ファイル、認証基盤周辺	ハニートークン（偽パスワード、偽 API キー、偽管理者アカウント）による窃取・悪用の誘発 ※ハニーファイルや偽脆弱ハニーポットを併用することで、窃取経路と昇格・横展開の観測も可能

6.3 詳細設計：技術要素ごとの設計観点

前節では、CDF を用いて、サイバーデセプション作戦の立案と戦術への落とし込みを行い、Why（目的・目標）、When（攻撃フェーズ）、Where（仕掛け先）、How（手段）を整理する基本設計の考え方を示した。

本節では、基本設計で選定した How（手段）について、実装・運用を見据えた具体的な技術要素として検討する際の設計観点を整理する。本節でいう詳細設計とは、製品ごとの設定値やパラメータを網羅的に定義することではない。サイバーデセプションの技術要素を実環境へ適用する際に、どの要件に対応するのか、攻撃者のどの認知・行動に作用するのか、どのような効果を期待するのか、どのような構成・配置が考えられるのかといった設計観点を整理するものである。

本書は、サイバーデセプションの考え方を広め、企業・組織における導入検討の土台を提供することを目的としている。そのため、本節では、すべての技術を網羅した実装手順や個別製品の設定方法までは扱わない。読者が代表的な技術要素の役割を理解し、自組織の要件（Why：目的・目標、When：攻撃フェーズ、Where：仕掛け先）に適合する技術や製品を選定するための判断材料を示す。読者が実務における適用イメージを持ちやすいよう、本節では代表的な技術要素を、その役割に着目して 3 つのカテゴリに分類して整理する。

表 6-7　代表的な技術要素

分類	技術要素	概要
おとり環境 （接触・観測の場）	ハニーポット	単体のおとりシステム・サービスとして攻撃者の接触や操作を観測する環境
	ハニーネット	複数のおとり資産を組み合わせ、ネットワーク規模で攻撃者の探索・横展開を観測する環境
情報・痕跡 （認知に作用する要素）	ハニートークン	本来使われない偽情報・偽認証情報などを 配置し、接触や利用を検知する仕組み
	ブレッドクラム	攻撃者を誘導するための手がかり・導線
	シンセティックデータ、 ポケットリッター	価値あると思わせる偽情報 攻撃者を信頼させる背景情報
ネットワーク制御 （進行を遅延・誘導する要素）	タールピット	攻撃者の活動遅延
	トラフィックリダイレクト	攻撃者の通信を誘導

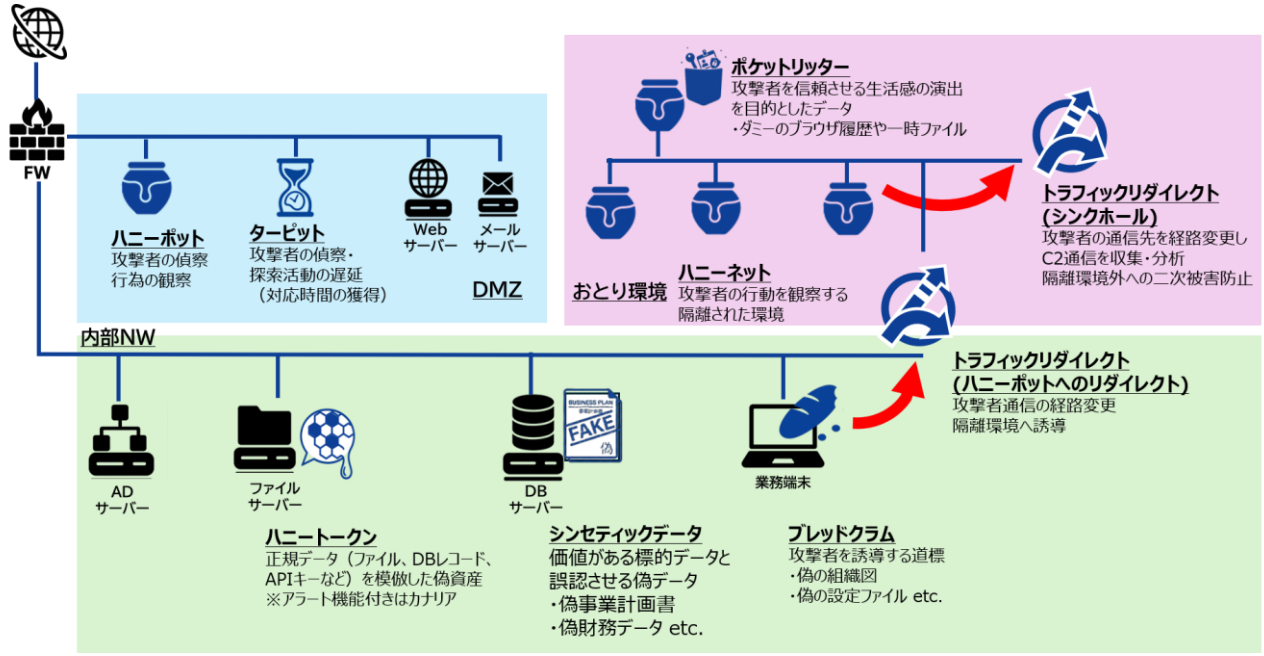


図 6-2　代表的な技術要素のイメージ

6.3.1　ハニーポット（Honeypot）

該当要件

- **Why（目標）：**
  検知（Expose-Detect）、収集（Expose-Collect）、妨害（Affect-Disrupt）
- **When（攻撃フェーズ）：**
  偵察、実行、探索
- **Where（仕掛け先）：**
  エンドポイント層、ネットワーク層、ソフトウェア層（DMZ、内部 LAN、クラウドなど）

攻撃者の判断傾向・心理

攻撃者は、探索によってシステムやサービスを「実際に利用されている資産」であると仮定して行動する傾向がある。ハニーポットは、以下のような行動特性を利用して攻撃者を本番環境ではなくおとり環境へ誘導する。

- 発見したシステムは価値のある対象であると判断しやすい。
- 内部ネットワーク上に存在する資産を信頼しやすい。
- 利用可能な認証情報や公開サービスを優先的に試行する。
- 効率を重視し、自動化ツールや既知の手順を利用することが多い。
- 環境全体を詳細に検証せず、見えている情報を前提として行動することがある。

導入効果

- 通常利用されない資産へのアクセスを契機として、不審なアクセスや不正行為を比較的高い精度で検知できる。
- 攻撃者を本番環境から切り離された環境へ誘導し、被害拡大の抑制が期待できる。
- 攻撃者の操作や実行コマンドを記録し、攻撃手法や攻撃意図の把握に利用できる。
- 攻撃者の時間やリソースを消費させ、攻撃コストを増加させることができる。
- インシデント対応や脅威分析に利用可能な情報を取得できる。

構成パターン

ハニーポットは、運用の目的に応じて、配置レイヤーおよび対話レベルについて選定する。

【配置レイヤーによる構成】

- ネットワーク境界（DMZ・外部公開領域）
  公開サービスを模倣し、探索行為や侵入試行を観測する。主に低対話型構成が利用される。
- 内部ネットワーク
  社内サーバーや端末を模倣し、侵入後の探索や横展開を観測する。ハニートークン（ハニークレデンシャル）などと組み合わせて利用されることが多い。
- エンドポイント
  端末上に配置し、認証情報窃取やマルウェア挙動を観測する。
- ソフトウェア層
  Web アプリケーションや API を模倣し、不正アクセスや攻撃操作を観測する。
- OT/ICS 環境
  PLC や産業プロトコルを模倣し、制御系システムを狙った攻撃を観測する。

【対話レベルによる構成】

- 低対話型
  軽量で安全性が比較的高く、スキャンや探索活動の観測に適する。
- 高対話型
  実環境に近い挙動を提供でき、詳細な攻撃分析に適する。

設計・配置のポイント

- 攻撃者が発見しやすい位置（スキャン対象領域や内部ネットワーク上）に配置する。
- 本番環境とは論理的または物理的に分離し、踏み台化を防止する。
- 周辺環境（ホスト名、OS 情報、ポート構成、通信挙動など）に整合性を持たせる。
- ハニートークン（ハニークレデンシャル）などを組み合わせ、誘導性を高める。
- 収集したログを SIEM などと連携し SOC 業務として対応できるよう設計する。

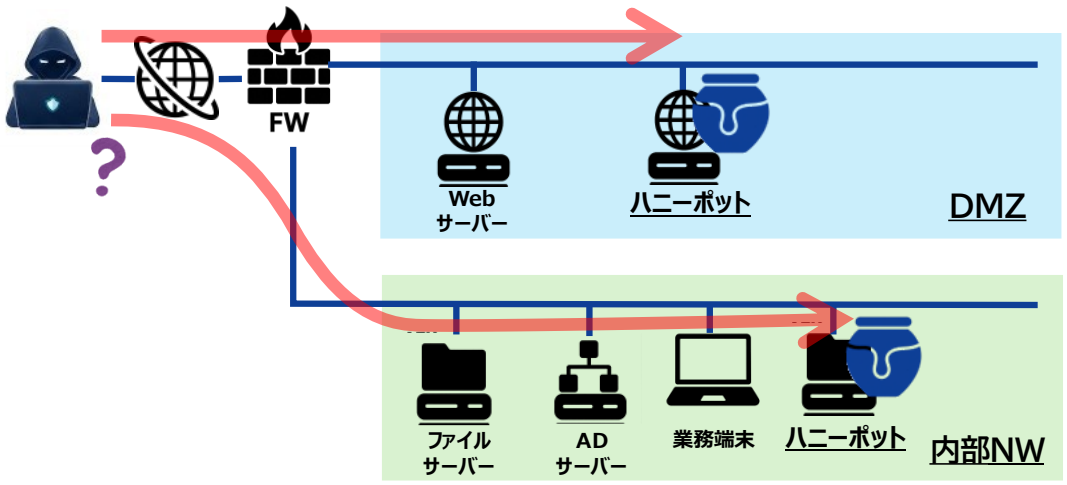


図 6-3　ハニーポットの要約

6.3.2　ハニーネット（Honeynet）

該当要件

- **Why（目標）：**
  収集（Expose-Collect）、誘導（Affect-Direct）、安心・確信（Elicit-Reassure）/動機付け（Motivate）
- **When（攻撃フェーズ）：**
  探索、横展開、収集、C2
- **Where（仕掛け先）：**
  ネットワーク層を中心に、エンドポイント層・ソフトウェア層を組み合わせた隔離セグメント

攻撃者の判断傾向・心理

ハニーネットは、攻撃者が持つ以下のような心理的特性や、判断の偏りを踏まえて設計される。

- **成功認知への偏り：**
  自身の操作が成功していると判断しやすく、都合のよい情報を成功の証拠として解釈する傾向がある。そのため、ハニーネット内に実在しそうな偽の業務通信やユーザーの活動痕跡を発見すると、それらを成功の証拠として捉え、その環境を本物であると認識しやすくなる。
- **正規ツールを悪用する手法への過信：**
  攻撃者は監視を回避するため、独自ツールではなく、標的環境に標準搭載された PowerShell などの正規ツールを悪用して行動することがある。ハニーネット内でこれらの正規コマンドが自然な応答を返すと、攻撃者は自身の行動が正常に進んでいると判断し、おとり環境であるという疑念を持ちにくくなる。

導入効果

- **攻撃プロセスの可視化と脅威インテリジェンスの獲得：**
  攻撃者がネットワーク内でどのように偵察し、どの権限を狙い、どのようなルートで横展開を試みるかという一連の攻撃手順を安全に観測できる。これにより、自組織の防御に直結する脅威インテリジェンスを入手できる。
- **重要資産への到達遅延と時間的猶予の確保：**
  攻撃者の時間や労力を偽のネットワーク内で浪費させることで、本物の重要資産に到達するまでの時間を稼ぎ、防御側が対応するための猶予を生み出すことができる。
- **自動化された探索・攻撃へのかく乱：**
  自動化ツールを用いた探索・悪用に対しても、偽のネットワーク構造や不要な探索対象を提示することで、攻撃者側の処理負荷や判断コストを増加させることが期待できる。

構成パターン

組織の環境やインフラの特性に応じて、以下の構成パターンを選択する。

- **IT 環境向け：**
  認証サーバー、ファイルサーバー、人事システム、一般従業員の端末セグメントなどを仮想環境上で複数エミュレートする。クラウドのコンテナ技術やインフラ構築の自動化コードを活用し、本番環境の構成を偽のネットワーク側へ自動生成する手法が用いられる。
- **OT/ICS 環境向け：**
  工場の生産ラインや重要インフラを模擬するため、監視制御システム、HMI、PLC といった各種制御機器の挙動を高度に模倣したエミュレータを相互接続する。本番の制御ネットワークからは、物理的または論理的に完全に隔離された独立セグメントとして稼働させる。
- **ハイブリッド型：**
  標的型攻撃で狙われやすい、情報系ネットワークから制御系ネットワークへの境界ゲートウェイや踏み台サーバーの環境を重点的に再現し、攻撃者が IT 層から OT 層へ侵入を試みる横展開の瞬間を待ち伏せする構成である。

設計・配置のポイント

- **徹底したネットワーク構造の一貫性と活動状況の模倣：**
  ネットワーク全体の命名規則、IP アドレスの割り当て規則、MAC アドレスのベンダー情報などに矛盾を生じさせない。さらに、システム内で日常的に発生する通信や活動痕跡を擬似的に生成し、業務で利用されている稼働中の環境を模倣して攻撃者に本物だと誤認させる。

- **攻撃者を誘導する接点の設計：**
  防御を完璧に見せるのではなく、十分に隔離された環境内で、攻撃者にとって魅力的に見える古いサービス、脆弱に見える設定、アクセス可能に見える管理機能などを模擬する。攻撃者に「偶然見つけた」と思わせる接点を設計することで、ハニーネット内での探索や横展開を誘発する。
- **インシデント対応の自動連携：**
  偽のネットワーク内で攻撃者が偽の管理者権限でログインを試みるなどの特定の侵害行為が観測された場合、そのイベント情報を統合ログ管理システム経由でセキュリティ運用自動化基盤へ連携させ、本番環境側で該当の攻撃元 IP アドレスを即座に隔離する仕組みを設計する。

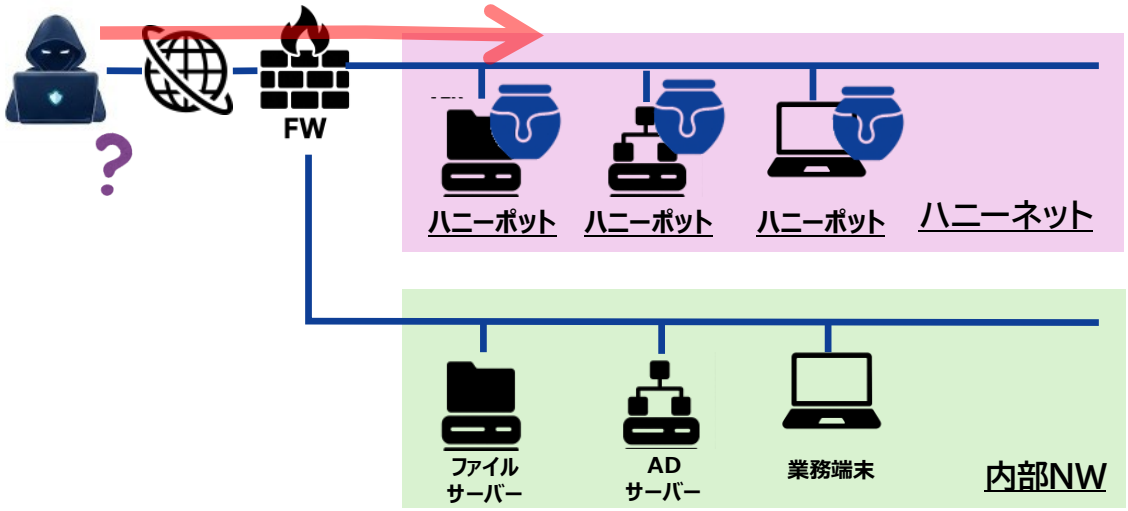


図 6-4　ハニーネットの要約

6.3.3　ハニートークン（Honeytoken）

該当要件

- **Why（目標）：**
  検知（Expose-Detect）、動機付け（Elicit-Motivate）
- **When（攻撃フェーズ）：**
  認証情報窃取、探索、横展開、収集、持ち出し
- **Where（仕掛け先）：**
  データ層、エンドポイント層

攻撃者の判断傾向・心理

攻撃者はネットワークに侵入した後、権限昇格や横展開を行うために新たな認証情報を探索したり、価値の高い機密データを収集しようとしたりする傾向がある。ハニートークンは、攻撃者の「より高い権限」や「価値ある情報」を求める心理的動機や好奇心を利用する。また、攻撃者が「自分は検知されていない」「見つけた情報は本物である」と判断しやすい認知の偏りを利用し、偽の認証情報やデータに対する参照、窃取、認証試行などの行動を誘発する。

導入効果

- 正規のユーザーがハニートークンにアクセスする理由は通常ないため、トークンへの接触は悪意ある活動、内部不正、または誤操作などの可能性が高いシグナルとして扱うことができる。これにより、誤検知の少ない高精度なアラートを生成できる。
- 攻撃者が偽データをダウンロードしたり、偽の認証情報を使用して別サービスへのログインを試行したりする挙動を監視することで、攻撃者の戦術・技術・手順（TTPs）やデータの流出経路を追跡できる。
- 偽のデータや認証情報を本物と思い込ませ、分析や攻撃を試みさせることで、攻撃者の時間とリソースを消費させ、真の標的への到達を遅延させることができる。

構成パターン

ハニートークンは導入や運用が比較的容易であり、組織内のさまざまな場所に柔軟に配置できる。

- **ファイルサーバー / ユーザー端末：**
  「重要顧客情報.xlsx」のように、攻撃者が価値を見出しやすいファイル名で、ファイル共有やデスクトップなど探索されやすい場所に配置する。
- **端末上のデータ・設定情報：**
  端末上に偽の認証情報、設定ファイル、接続情報などを配置し、認証情報窃取ツールによる参照や悪用を検知する。
- **Web アプリケーション：**
  HTTP パラメータ、Cookie、隠しフォームフィールドなどにダミー値や識別子を埋め込み、不自然な参照や再利用を検知する。
- **クラウド / 開発環境：**
  ソースコード管理ツール内に偽のクラウド API キーを配置したり、偽のストレージバケット名や接続情報を用意したりすることで、窃取や悪用の試行を検知する。

設計・配置のポイント

- 実際の組織の命名規則、ファイル構造、メタデータ（タイムスタンプなど）と整合させ、高度な攻撃者にも不自然さを感じさせないように設計する。
- 攻撃者が思わずアクセスしたくなるような、機密性や価値が高そうに見える名称や配置場所を、作戦目標に基づいて選定する。
- 各トークンに一意の ID を付与し、SIEM や SOAR と連携させることで、どのトークンがどこでアクセスされたかを即座に特定・相関分析できるようにする。

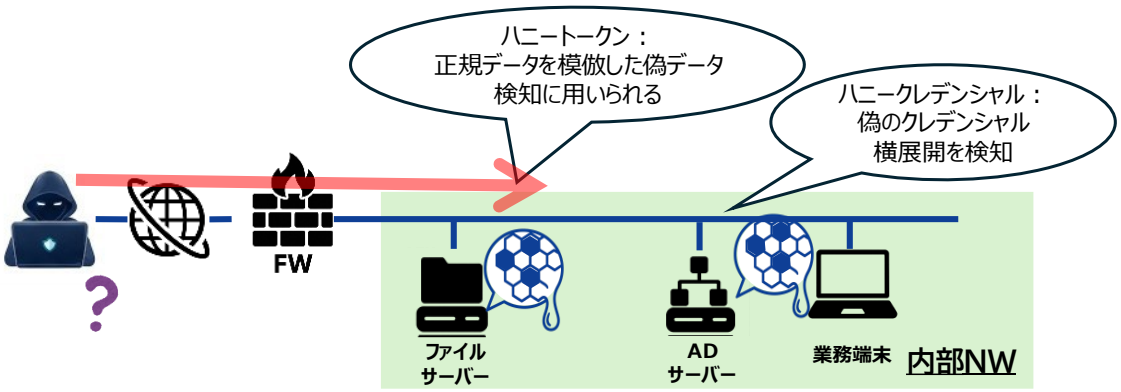


図 6-5　ハニートークンの要約

6.3.4　ブレッドクラム（Breadcrumbs）

該当要件

- **Why（目標）：**
  誘導（Affect-Direct）、動機付け（Elicit-Motivate）
- **When（攻撃フェーズ）：**
  探索、横展開
- **Where（仕掛け先）：**
  データ層（ネットワーク情報など）、エンドポイント層（ファイルシステム、ログ、レジストリ、ブラウザ履歴など）

攻撃者の判断傾向・心理

攻撃者は、侵入後に環境を探索して得た情報を、権限昇格や横展開に利用可能な足掛かりとして扱う傾向がある。ブレッドクラムは、特に以下のような攻撃者の判断傾向を利用する。

- **初期情報への依存：**
  攻撃者は、探索の初期段階で発見した情報を有力な手がかりとして扱い、その後の行動判断に利用する傾向がある。

- **確証バイアス：**
  攻撃者は、発見した痕跡を「探していた有益な情報である」と都合よく解釈し、サイバーデセプションである可能性を十分に検証せず信頼する場合がある。
- **効率重視の行動：**
  攻撃者は、効率的に権限昇格や横展開を行うため、見つけた設定ファイル、接続履歴、認証情報の痕跡などを優先的に試行する傾向がある。

導入効果

- 攻撃者の注意を本物の資産から逸らし、安全なおとり環境（ハニーポットなど）へ自発的に誘導できる。
- 攻撃者をおとり環境へ導くことで、滞在時間を延長させ、攻撃手法（TTPs）のより深い観察が可能となる。
- 攻撃者の探索、分析、移動方向を誤らせることで、攻撃者の時間とリソースを消費させることができる。
- 軽量な検知要素から、より詳細な観測が可能な高対話型のおとり環境へ、攻撃者を自然に誘導できる。

構成パターン

ブレッドクラムは、配置される環境や攻撃者の探索手法に合わせて、さまざまな形態をとる。

- **ファイル・ドキュメントベース：**
  偽のネットワーク構成図、ダミーのシステム設定ファイル、偽のブラウザ閲覧履歴などを配置する。
- **接続情報・ショートカット：**
  ハニーポットへ接続するように設計された偽の RDP ショートカット、最近使用した SSH の接続履歴、偽の VPN 接続情報などを配置する。
- **スクリプト・コード内の情報：**
  管理用スクリプト内にコメントとして残されたダミーの URL や、Git リポジトリ内に置かれたダミーの設定情報などを配置する。
- **ログベース：**
  システムログやアプリケーションログの中に、偽の認証情報や接続履歴を意図的に残す。

設計・配置のポイント

- **自然に溶け込ませる：**
  目立ちすぎると攻撃者に警戒されるため、攻撃者が「自分の探索によって発見した」と認識するよう、既存のログや設定ファイルの流れに沿って配置する。
- **作戦目標に基づいて配置する：**
  ファイル共有、社内 Wiki、バージョン管理システムなど、攻撃者が探索フェーズで確認しやすい場所を選定する。

- **一貫性を確保する：**
  誘導先となるハニーポットなどの設定（OS、サービス、ポート構成など）と、ブレッドクラムに記載された内容の整合性を確保する。
- **追跡・相関分析を想定する：**
  ブレッドクラム自体にアラート発報機能を持たせる必要はないが、記載する偽の接続情報やユーザー名などに一意の識別子を含ませることで、攻撃者が誘導先でその情報を使用した際に、どのブレッドクラムを経由したかを追跡・相関分析できるよう設計する。

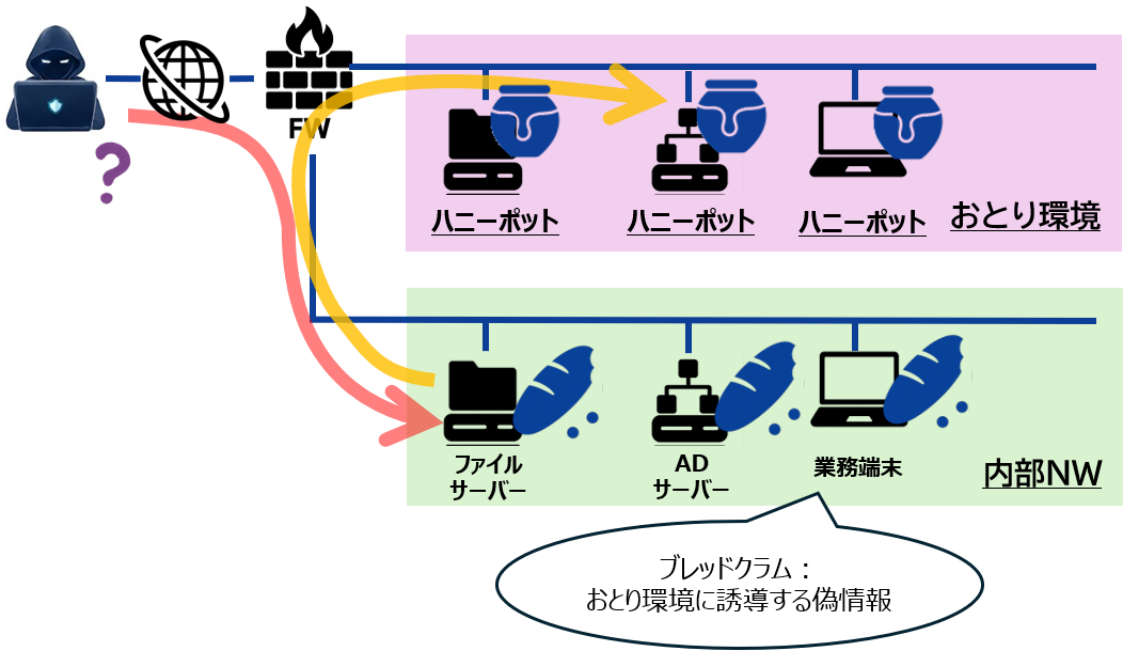


図 6-6　ブレッドクラムの要約

6.3.5　シンセティックデータ（Synthetic Data）/ポケットリッター（Pocket Litter）

該当要件

【シンセティックデータ（標的の偽装）】

- **Why（目標）：**
  妨害（Affect-Disrupt）、検知（Expose-Detect）、動機付け（Elicit-Motivate）
- **When（攻撃フェーズ）：**
  探索、収集、持ち出し
- **Where（仕掛け先）：**
  データ層（データベース、ファイルサーバなど）、エンドポイント層（ファイルシステム）

【ポケットリッター（背景の演出）】

- **Why（目標）：**
  安心・確信（Elicit-Reassure）
- **When（攻撃フェーズ）：**
  初期侵入、探索、横展開
- **Where（仕掛け先）：**
  データ層、エンドポイント層（ブラウザ履歴、一時ファイル、デスクトップメモなど）

攻撃者の判断傾向・心理

これらの偽データ・痕跡は、攻撃者の情報探索フェーズにおいて、以下の心理や行動特性を利用する。

- **確証バイアス:**
  攻撃者は、それらしいファイル名（顧客リストなど）や業務上自然に見える履歴を発見すると、それを「自分の探していた価値ある情報である」「この環境は本物である」と都合よく解釈する場合がある。
- **初期印象への依存:**
  探索の初期段階でポケットリッターにより「本物らしい」という印象を与えられると、その後に多少の不自然な点に遭遇しても、おとり環境である可能性を軽視する場合がある。
- **自動化された情報収集への依存:**
  効率を重視し、ツールを用いてデータを一括で収集・持ち出そうとする攻撃者は、データの内容を一つ一つ精査せず、偽の標的データを持ち出す可能性がある。

導入効果

- **偽価値の提示によるリソース消費：**
  偽のデータを本物と認識させることで、攻撃者の時間とリソースを消費させ、攻撃の費用対効果を低下させることができる。
- **おとり環境の信頼性向上：**
  ハニーポットやハニーネットなどの背景情報として機能することで、攻撃者の疑念を低減し、おとり環境の信頼性を高めることができる。
- **検知・追跡機能の補強：**
  シンセティックデータにハニートークンや一意の識別子を組み込むことで、偽データへの接触、持ち出し、外部環境での参照などを検知・追跡できる。

構成パターン

目的が「価値の偽装」か「背景の演出」かに応じて、以下のような形態をとる。

- **シンセティックデータ（標的の偽装）のパターン:**
  データベース内に混入された偽の顧客情報、偽のクレジットカード番号、偽の財務スプレッドシート、偽の事業計画書、偽のインフラ設定ファイルなどを配置する。
- **ポケットリッター（背景の演出）のパターン:**
  Web ブラウザの閲覧履歴や Cookie、一時ファイル、ごみ箱の中身、デスクトップに置かれたメモ、日常的なメールの送受信履歴などを配置する。

設計・配置のポイント

- **実データ・ペルソナとの一貫性を確保する:**
  配置するデータは、実環境のフォーマットや、当該システムを利用しているように見せる架空ユーザーの属性（ペルソナ）と矛盾しないように設計する必要がある。
- **生成と更新を自動化する:**
  大規模言語モデル（LLM）などを用いて、不自然さの少ないテキストやデータセットを生成し、定期的に更新・展開することは、実運用上有効である。
- **自然なストーリーとして統合する:**
  ポケットリッターによる背景情報の中に、シンセティックデータやブレッドクラムなどの誘導要素を自然に組み込むことで、攻撃者をおとり環境や偽の標的データへ導く一連の流れを設計する。

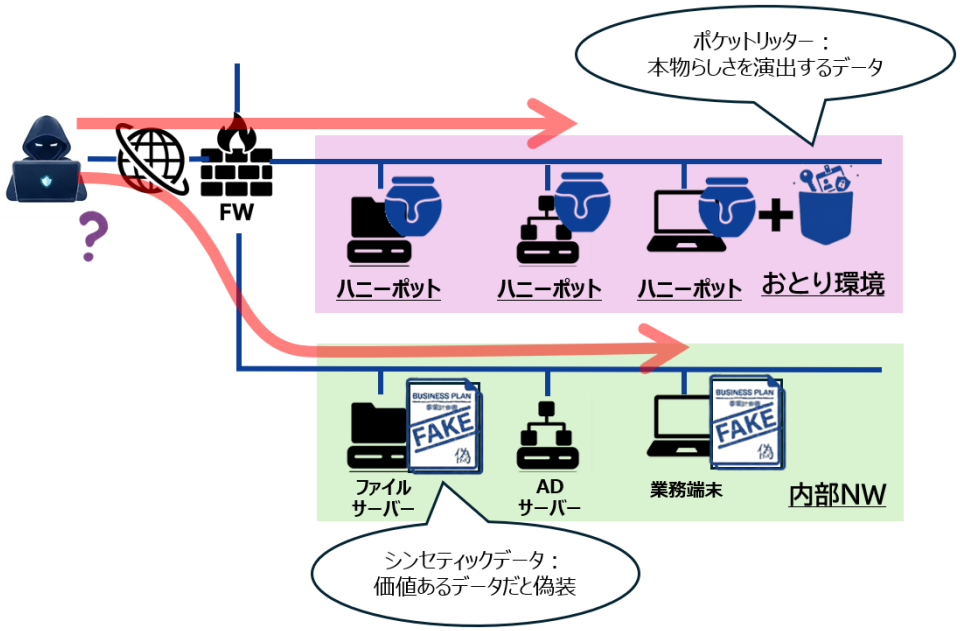


図 6-7　シンセティックデータ/ポケットリッターの要約

6.3.6　タービット（Tarpit）

該当要件

- **Why（目標）：**
  妨害（Affect-Disrupt）
- **When（攻撃フェーズ）：**
  偵察、初期侵入、探索、横展開
- **Where（仕掛け先）：**
  ネットワーク層（未使用の IP アドレス空間）、ソフトウェア層（特定のサービスポートに対する応答遅延など）

攻撃者の判断傾向・心理

攻撃者は短時間で広範囲の脆弱性を探索したり、マルウェアを増殖させたりするために、スキャンツールなどの自動化や効率性に強く依存している。タービットはこうした行動特性を利用し、以下の弱点や心理に作用する。

- **自動化への過信：**
  スキャンツールは、「応答がある＝システムが存在する」と判断し、次の処理へ進むために通信の完了を待ち続けることがある。タービットはこの挙動を利用し、攻撃者側のツールに待機時間や再試行を発生させる。
- **サンクコスト効果と確証バイアス：**
  攻撃者に「システムは存在しており、ただ応答が遅いだけである」と判断させることで、接続を切断せずに待機し続ける状況を作り出す。結果として、攻撃者の時間や処理資源を消費させることができる。

導入効果

- **行動遅延と攻撃効率の低下：**
  攻撃者のネットワーク探索や横展開の速度を低下させ、自動化された攻撃サイクルを鈍化させることができる。
- **防御側の対応時間の確保：**
  攻撃の進行を遅延させることで、SOC や CSIRT が脅威を検知し、状況を判断し、対応するための時間的猶予を確保できる。
- **攻撃側リソースの占有・消費：**
  攻撃側のスキャンツールやマルウェアの接続セッションを占有し続けることで、攻撃者側の処理能力、通信帯域、同時接続数などを消費させることができる。

構成パターン

- **未使用 IP アドレス空間（ダークスペース）での展開：**
  本番環境で割り当てられていない空き IP アドレス空間に配置し、そこへのスキャンや自己増殖型マルウェアの接続試行を捉えて遅延させる。正規業務への影響を抑えやすく、比較的安全に導入しやすい構成である。
- **サービスレベルのターピット（スティッキー・ハニーポット）：**
  SMTP、SSH などの特定のサービスポートに対するブルートフォース攻撃や不正なログイン試行に対して、意図的に応答を遅延させる。
- **実資産上でのターピット：**
  本番稼働しているシステム上で、一連の異常なシステムコールや不正な振る舞いを検知した際に、OS やアプリケーション側の応答を意図的に遅くして侵入の進行を妨害するアプローチである。ただし、正規通信への影響を避けるため、高度な検知精度と慎重なチューニングが前提となる。

設計・配置のポイント

- **業務通信との分離：**
  未使用 IP アドレスを利用するアプローチでは、正規ユーザーの本番トラフィックがターピットに触れないよう、ネットワークルーティングや配置を慎重に設計する。
- **防御側リソース消費の最適化：**
  ターピット自身が攻撃者からの大量のセッションを保持することになるため、防御側のシステムリソース（メモリやコネクション数）が枯渇しないよう、軽量かつ効率的に実装する。
- **早期検知センサーとの連動：**
  単に遅延させるだけでなく、接続してきた送信元 IP アドレスなどのメタデータを SIEM や SOC へ通知し、高精度なアラートを上げるセンサーとしても機能させる。

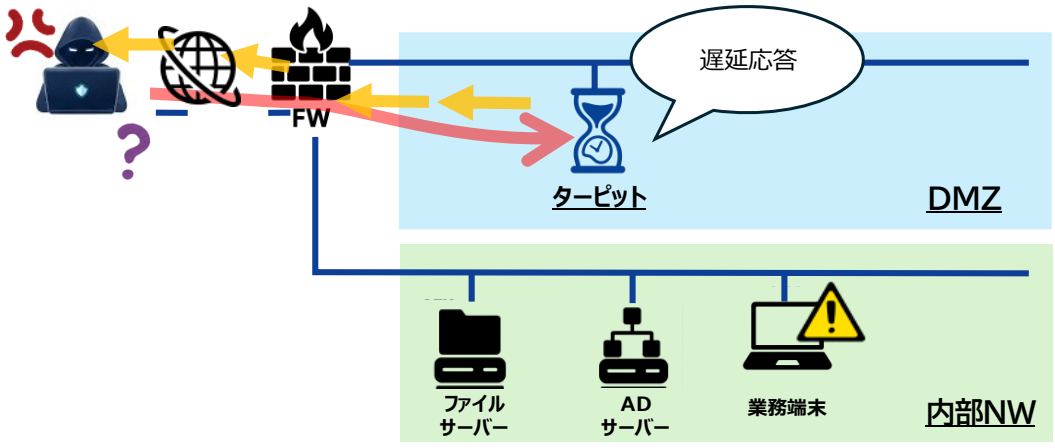


図 6-8　ターピットの要約

6.3.7　トラフィックリダイレクト（Traffic Redirection）

該当要件

- **Why（目標）：**
  収集（Expose-Collect）、誘導（Affect-Direct）、阻止（Affect-Prevent）
- **When（攻撃フェーズ）：**
  横展開、C2、持ち出し
- **Where（仕掛け先）：**
  ネットワーク層、ソフトウェア層（DNS など）

攻撃者の判断傾向・心理

攻撃者は、ネットワーク内での探索活動、マルウェアの遠隔操作（C2 通信）、データの持ち出しを行う際、名前解決やルーティングが正しい宛先へ通信を届けることを前提として行動する。トラフィックリダイレクトは、このネットワーク基盤への信頼を利用し、攻撃者に気付かれにくい形で通信の到達先を防御側が制御する。

導入効果

- **実資産への到達抑制：**
  攻撃者のトラフィックをおとり環境へ誘導することで、「攻撃が継続できている」と誤認させながら、実資産への到達や被害発生を抑制できる。
- **安全な観測とメタデータの収集：**
  通信を単に遮断するのではなく、監視用サーバーやおとり環境へ再指向することで、攻撃者の意図、通信先、通信内容、接続元情報などを安全に観測・収集できる。
- **データ持ち出しや C2 通信の阻止：**
  本物の C2 サーバーへの通信や外部へのデータ持ち出しを、防御側が管理する宛先へ誘導または無効化することで、マルウェアの制御や実害の発生を防ぐことができる。

構成パターン

トラフィックリダイレクトは、通信をどのように処理し、どこへ誘導するかによって、主に以下のパターンで運用される。

- **ハニーポットへの再指向：**
  攻撃者のトラフィックを高対話型ハニーポットなどへ再指向する手法である。セッションを可能な限り維持したままおとり環境へ移行させ、攻撃の進行や操作内容を観測する。

- **シンクホール：**
  不正な DNS リクエストや外部への接続試行を、防御側が管理する監視用サーバーなどへ再指向する手法である。通信を単に破棄するのではなく、通信内容やメタデータを安全に取得し、分析に活用する。
- **ブラックホールによる阻止：**
  トラフィックリダイレクトの一形態として、監視・記録を行わず、存在しない宛先や Null インターフェースなどへ通信を誘導し、パケットを破棄する手法がある。これはデセプションによる観測・誘導というよりも、主に阻止・拒否を目的とした運用であり、DDoS 攻撃の緩和や既知の悪性通信の遮断などに用いられる。

設計・配置のポイント

- **透過的な再指向：**
  攻撃者に再指向されたことを悟らせないよう、SDN やアプリケーションプロキシなどを活用し、接続を可能な限り維持したまま自然に宛先を変更する。
- **おとり環境からの外部通信制御：**
  おとり環境から外部ネットワークへのアウトバウンド通信が試みられた場合、本物の外部ネットワークへ直接到達させず、制御されたシンクホールや監視環境へ再指向するよう、ネットワーク分離とルーティングを設計する。
- **内部 DNS を活用した防御的な名前解決制御：**
  防御側が内部 DNS サーバー上で悪性ドメインや疑わしい宛先の名前解決を制御し、シンクホールの IP アドレスを返すよう設定することで、組織内からの悪性通信を抑制・観測できる。

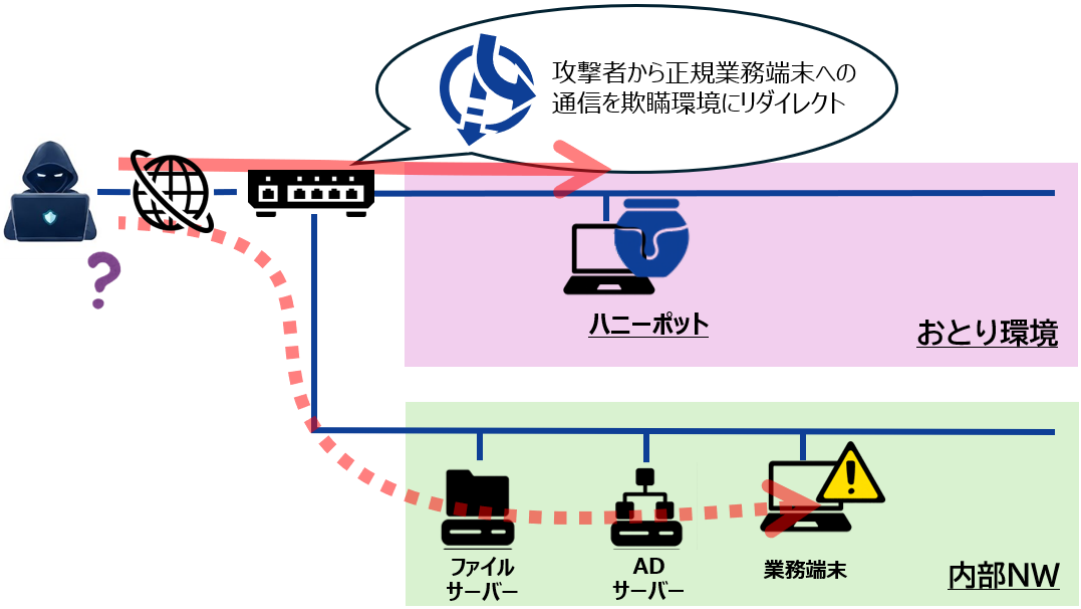


図 6-9　トラフィックリダイレクトの要約

6.4　実装・運用上の原則と留意点

サイバーデセプションにおける目標を達成するには、単純に環境内へサイバーデセプション資産（おとり）を配置するだけでは不十分である。熟練した攻撃者から不自然さを見抜かれにくくし、サイバーデセプション資産への接触可能性を高めるためには、セキュリティ戦略上の目的と作戦目標に沿った設計が不可欠である。また、継続的に効果を発揮させるためには運用面の最適化も重要な課題となる。場当たり的な手動運用は、設定ミスによる予期せぬリスクを招くだけでなく、管理負荷を増大させる原因となる。さらに、十分な設計がされないまま運用を行ってしまった場合、サイバーデセプションにおける目標が達成できないだけに留まらず、組織において新たなリスクを発生させてしまう可能性もある。

本節では「Cyber Deception: Fundamental Guide[1]」を参考にサイバーデセプションの原則について記載する。まずは「6.4.1　倫理原則の順守」で能動的防御手法であるサイバーデセプションを運用するにあたり遵守すべき倫理原則について記載する。さらに、この倫理原則を土台とし、企業・組織がサイバーデセプションを運用する上で実践すべきベストプラクティスと避けるべき運用について、それぞれ「6.4.2　効果的に設計・運用するための心得」と「6.4.3　運用上の落とし穴」の各項に記載する。

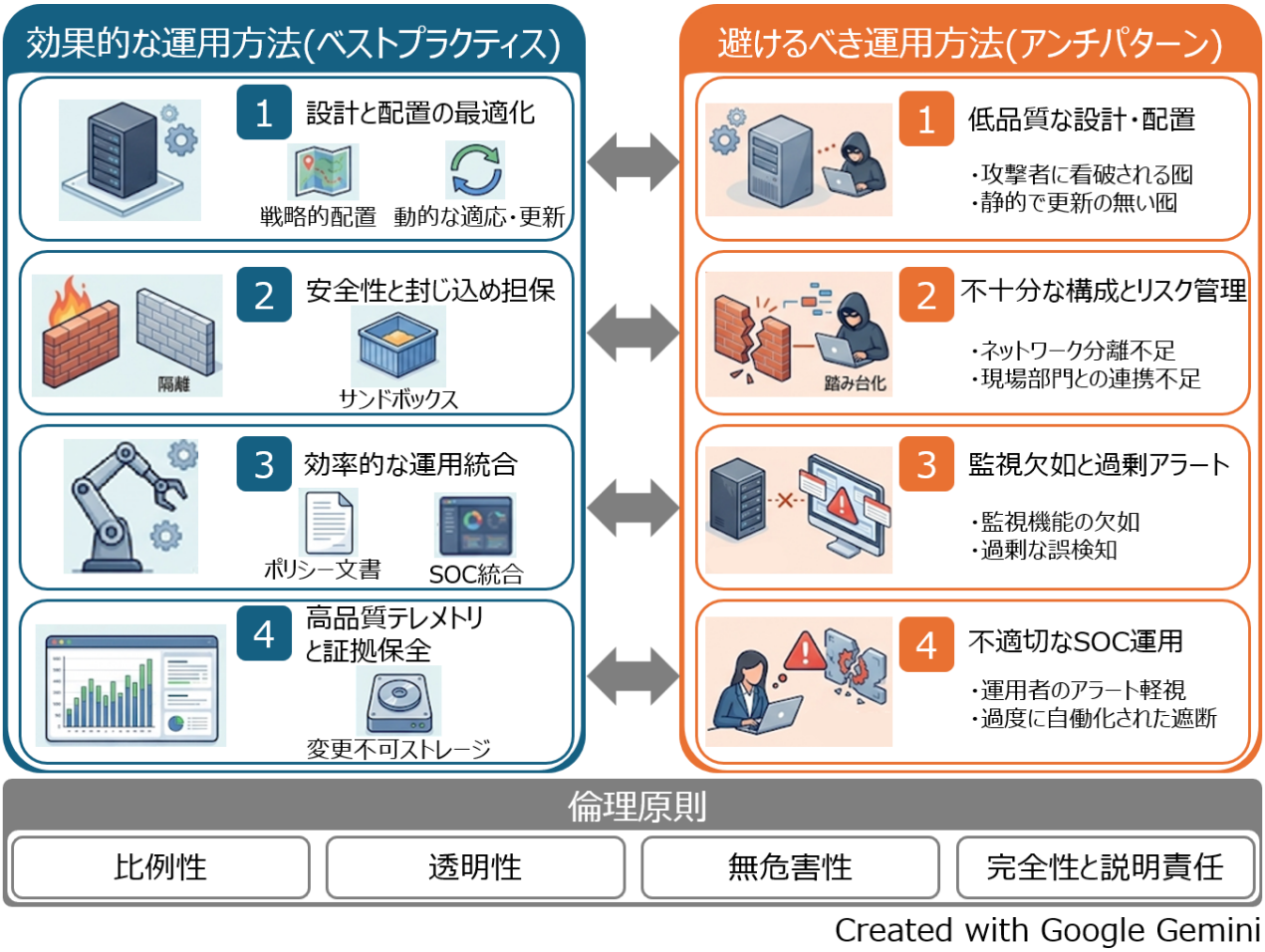


図 6-10　サイバーデセプションにおけるベストプラクティスとアンチパターン

6.4.1　倫理原則の順守

サイバーデセプションは攻撃者の行動に介入する性質上、プライバシー侵害などの法的リスクを伴う。ここでは、組織の防衛活動が過剰な反撃や逸脱行為とならないよう、大前提として遵守すべき倫理原則を整理する。

比例性（Proportionality）

比例性とは、防御側が取る対策や対応措置は、直面している脅威の規模や深刻度に見合った（均衡のとれた）ものでなければならないという原則を指す。 サイバーデセプション資産の展開や監視は、対象となる脅威やリスクに見合った適切な範囲にとどめ、正当なユーザーや第三者への過剰な監視（プライバシー侵害）となってはならない。

透明性（Transparency）

透明性とは、詳細を隠さずに開示する原則を指し、見えないリスクを減らし関係者間での信頼の構築が目的となる。サイバーデセプション環境の展開については、セキュリティ部門だけでなく IT、法務、および経営陣などの内部関係者に周知し、内部的な透明性の確保が必要になる。一方で悪用を防ぐため外部への透明性には制限を設ける必要がある。公の場や管理されていないコミュニティで具体的な作戦や戦術の公開は避けることが推奨される。

無危害性（Non-Maleficence）

無危害性とは、いかなる理由であっても意図的・悪意を持って他者のシステム、データ、ネットワークに損害や破壊をもたらす行動をとってはならないという原則を指す。サイバーデセプション環境においては、能動的な反撃やマルウェアを拡散するような害を及ぼす機能の設計は避ける必要がある。

完全性と説明責任（Integrity and Accountability）

完全性とは、データやシステムが正確かつ最新の状態で保たれ、不正な改ざんや破壊、欠落がない状態を保証する原則のことを指す。サイバーデセプション環境において、実行したアクション、意思決定、およびポリシー遵守に関する監査可能な記録（証跡）を常に維持し、運用に携わるアナリストは、収集した機密データや証拠を適切に取り扱うための訓練を受け、その取り扱いに明確な責任を持つ必要がある。

6.4.2　効果的に設計・運用するための心得

サイバーデセプションの真の価値は、罠の設置そのものではなく、既存の防御体制と効果的に連携させることで発揮される。ここでは、環境を安全に機能させ、導入効果を最大化するために実践すべきベストプラクティスを提示する。

（1）　設計と配置の最適化

リアリズムの追求

サイバーデセプション資産は、実環境と見分けがつかないレベルで再現することが重要となる。攻撃者は、命名規則、システム挙動、通信内容、ファイル構造、タイムスタンプなどの細かな不整合から不自然さを見抜くため、リアリズムが低い場合、疑念を持たれ接触を回避される可能性が高い。その結果、検知や分析の機会を逸することにつながる。例えば、実在するサーバー命名規則に準拠したホスト名の付与や、実データに類似したダミーファイルの配置などが有効である。リアリズムの確保は、攻撃者との接点を生み出す前提条件である。

作戦目標に基づく配置設計

サイバーデセプション資産は、攻撃者の行動経路や想定される攻撃フェーズを踏まえ、作戦目標に基づいて配置設計する必要がある。具体的には、認証基盤、ファイルサーバー、重要システム周辺など、攻撃者が到達・探索する可能性の高いポイントに設置することが有効である。適切な配置設計により、攻撃者の接触確率を高めるとともに、初期侵入後の探索、権限昇格、横展開など、攻撃の進行段階に応じた検知が可能となる。

環境の動的な適応と更新

攻撃手法やツールは継続的に進化するため、サイバーデセプション資産もそれに応じて更新・変化させる必要がある。同一の構成やパターンを維持し続けると、不自然な規則性や特徴から攻撃者におとりであると見抜かれる可能性が高まり、接触や誘導の効果が低下する。資産構成の更新、配置のローテーション、挙動のランダム化などを行い、常に“本物らしさ”を維持することが重要である。これにより、サイバーデセプションの持続的な有効性を確保できる。また、収集したインテリジェンスから得られた知見をもとに、新たな脅威やビジネス環境の変化に合わせて設計をチューニングし、継続的な改善サイクルを実施することも重要となる。

（2）　安全性と封じ込めの担保

厳格な環境分離

おとり環境は、原則として本番環境から論理的または物理的に分離し、攻撃者の活動が実システムへ影響を及ぼさないよう設計する必要がある。おとり環境が侵害された場合でも、それが踏み台として悪用され、本番環境への横展開につながることを防ぐことが目的である。具体的には、ネットワークセグメンテーション、サンドボックス化、仮想化技術、アクセス制御、通信制御などを活用し、攻撃者の行動を監視・制御可能な範囲内に封じ込める構成が求められる。

フェイルセーフの実装と証拠保全

隔離された環境を構築した場合であっても、サイバーデセプション資産が攻撃の踏み台として悪用されるリスクを完全に排除することはできない。そのため、サイバーデセプション資産において異常な振る舞いや予期せぬ外部通信を検知した際には即座に通信を遮断し、必要に応じてシステムを停止できるフェイルセーフ機能を実装する必要がある。また、遮断・停止と同時にログ、通信記録を保全する仕組みも併せて構築すべきである。これにより、おとり環境を悪用するリスクを軽減しつつフォレンジックに必要なデータを確保し、事後に迅速かつクリーンな状態へ復旧させることが可能となる。

(3)　効率的な運用の整備とプロセスの統合

運用負荷の軽減と自動化

サイバーデセプションは継続的に運用することで効果を発揮するため、過度な管理負担を避ける設計が重要となる。手動運用に依存すると、更新の遅延や設定不備が発生し、リアリズムや適応性の低下につながる可能性がある。そのため、自動化やオーケストレーション、標準化された展開手順を活用し、効率的かつ安定的に運用できる体制を整備する必要がある。これにより、限られたリソースでも継続的な効果を維持することが可能となる。

運用手順の文書化とポリシー策定

サイバーデセプションの運用が特定の担当者に依存し属人化すると、設定の不備によるセキュリティリスクや、更新漏れによるサイバーデセプション資産の品質（リアリズム）の低下を招く可能性がある。そのため、サイバーデセプション資産の作成プロセス、配置設計、監視手順、定期更新、インシデント対応プロトコルに至るまでを明確に文書化し、組織のポリシーとして定着させる必要がある。これにより、安全なおとり環境を均一に展開できるとともに、その運用が組織の目的やコンプライアンス要件に準拠していることを客観的に示すことが可能となる。

SOC ワークフローへの統合

おとりからのアラートは、EDR、SIEM、SOAR などの既存のセキュリティプラットフォームと連携させ、SOC ワークフローに直接統合することが推奨される。おとり環境は原則として正規ユーザーが利用しない環境であるため、そこで発生するアラートは高い信頼性を持つ。一方で、他のログやイベント情報と相関分析を行うことで、攻撃者がどこから侵入し、どのように探索や横展開を試みているかといった詳細な状況を把握しやすくなる。さらに SOAR と連携する場合には、既存のワークフローに従った迅速なインシデント対応が可能となる。SOC ワークフローへの統合は、運用体制の変更を最小限に留め、効率的な監視・対応を実現するために重要である。

（4）　高品質なテレメトリと証拠保全

高精度なテレメトリの取得と分析

おとり環境で収集した情報をインテリジェンスとして活用するには、詳細なログおよびテレメトリとして取得し、分析可能な形で蓄積することが重要となる。そのため、収集するログは自動分析が容易な JSON などの構造化フォーマットを使用し、システム横断的な相関分析を可能にするため、一意の相関識別子（ID）を付与することが推奨される。特にサイバーデセプション資産は通常業務では利用されないことを前提としているため、それらへのアクセス自体が高い確度の異常シグナルとなる。例えば、コマンド実行履歴、認証試行、セッション情報などを収集することで、攻撃者の使用ツール、侵入経路、目的の推定が可能となる。これにより、誤検知を抑えつつ、高精度なインシデント対応や脅威ハンティングにつなげることができる。

変更不可能なストレージによる証拠保全

キャプチャしたセッション情報、操作履歴、タイムスタンプ、マルウェア検体、ハッシュ値などのメタデータは、事後調査やフォレンジック分析において信頼性を担保できる形で保全する必要がある。そのため、これらのデータは改ざんを防止できる変更不可能なストレージに安全に保存することが推奨される。具体的には、WORM ディスクやイミュータブルなオブジェクトストレージなどを活用し、取得後の改ざんや削除を防止する仕組みを整備することが有効である。これにより、調査時における証跡の信頼性を確保し、インシデント対応、フォレンジック分析、再発防止策の検討に活用できる。

6.4.3　運用上の落とし穴

不十分な設計や場当たり的な運用は、欺瞞効果を失わせるだけでなく、組織に新たなセキュリティリスクを招く恐れがある。ここでは、導入・運用時に陥りやすい代表的なアンチパターンと、その回避策を整理する。

（1）　低品質・静的な設計・配置

攻撃者に看破される低品質なサイバーデセプション資産の展開

高度な攻撃者は、侵入後の調査でシステムの不自然な挙動や設定の矛盾（フィンガープリント）を細かく分析し、おとりの特定や回避を試みる。そのため、デフォルト設定の放置や安易な命名規則（例：honeypot-server）、更新されていないタイムスタンプなどの不自然なメタデータなどが存在する場合、サイバーデセプション資産は早期に看破される可能性が高い。構成や定期実行タスク、ログ生成などを実環境に近い形でおとり環境に再現することにより攻撃者を長期間環境内に引き留め、より価値の高いインテリジェンスを収集することができる。

静的で更新されないおとり環境

サイバーデセプション資産を一度配置したまま長期間放置すると、同一のトークン ID やファイル構成といったパターンが攻撃者に特定され、デセプションであることが見抜かれてしまう可能性が高まる。これを防ぐため、おとり環境は定期的にイメージを更新し、ハニートークン ID やブレッドクラムの配置、ファイル構成などをローテーションし、環境を動的に保つことが不可欠となる。

サイバーデセプション資産に本番データを使用する

おとりのリアリティを高める目的であっても、実際の個人情報や機密データを配置することは、深刻なプライバシー侵害や GDPR（General Data Protection Regulation：一般データ保護規則であり、EU（欧州連合）が定めた、個人のプライバシーとデータ保護に関する厳格な法規制）などの各種法規制違反などの法的リスクを引き起こすことになるため厳禁である。サイバーデセプション資産に設置するデータは、本物のように見えるが実在しないシンセティックデータ（Synthetic Data）を生成して配置することが望ましい。

（２） 不十分な環境構成とリスク管理

不十分なネットワーク分離によるデセプション環境の踏み台リスク

脆弱性を模したおとり環境を十分な分離・制御を行わないまま社内の一般的なシステムと同等のネットワークに配置したり、過剰なアクセス権限を残したまま放置したりすることは、最も避けるべき運用である。おとり環境の外部通信や他の社内システムへの通信に対するアクセス制限が不十分な状態では、本番ネットワークへの横展開を許す新たな侵入経路となり、むしろ重大なインシデントを引き起こすリスクを高めてしまう結果となる。

ビジネス部門および OT（制御技術）部門との連携不足

特に OT/ICS（産業用制御システム）環境において、現場のエンジニアリングチームと事前のリスク評価や調整を行うことなくサイバーデセプションを導入した場合、稼働中の制御機器に対する不適切な通信を発生させ、制御プロセスと干渉し、安全上の重大なインシデントや予期せぬシステムのダウンタイムを引き起こすおそれがある。OT/ICS 環境においては、エンジニアリングチームと連携し、現場の機器構成や運用制約を把握した上でおとり環境を展開する必要がある。

（3） 監視の欠如と過剰アラート

監視機能が欠如した孤立したサイバーデセプション資産の展開

ネットワーク設定やエージェントの不備によりログが SIEM に到達していないなど、監視計画と紐付いていないおとりの展開は、サイバーデセプションが提供する早期検知や攻撃者行動の観察という効果を十分に発揮できなくなる。監視されていないサイバーデセプション資産はリソースの無駄となるだけでなく、むしろ攻撃者が SOC に気づかれずに探索や横展開を行う拠点を提供してしまうおそれがある。

過剰なロギング

生ログを大量に収集するだけでは単なるノイズとなり、サイバーデセプションが提供する高精度なアラートというメリットを損ない、真の脅威シグナルを見失うことにつながる。コマンド実行履歴、ダウンロードされたバイナリ、ネットワークフロー、認証試行など、デセプションの目的に合わせて収集対象を絞ったログ収集設計が必要となる。

過剰な誤検知と通知疲れ

おとり環境であっても、脆弱性スキャナーや資産管理ツールなどによる無害な自動化通信が届くことがある。これらの通信までアラートとして検知する設定をしてしまうと、過剰な誤検知を生み出し、SOC アナリストの通知疲れや SOC 担当者のアラート軽視を引き起こす原因となる。ホワイトリストなどを活用し、既知の通信はアラートから除外することで本当に対応が必要なアラートに絞る必要がある。

（4） 不適切な SOC 運用とインシデント対応

過度な自動化による誤検知ブロック

おとり環境においても、前述の脆弱性スキャナーや従業員の偶発的なアクセスなど、正規の通信が発生する可能性がある。その際、アラート発生時に無条件で送信元の通信を自動遮断するような過剰な設定では、正規の運用管理サーバーや従業員端末からの通信を遮断してしまうリスクがある。防御アクションの自動化は、他のログと十分な相関分析を実施した上で機能するよう設計する必要がある。

運用者によるアラートの軽視

おとりからのアラートに対して、運用者が「どうせ本物の環境への攻撃ではない」と軽視して対応を怠ってはならない。おとりからのアラートは高い信頼性を持つことを運用者に周知し、必要に応じて訓練も実施することで、アラート発生時に迅速な対応ができる体制を維持する必要がある。

不適切な証拠保全

インシデント発生時の証拠保全が手動運用に依存すると、運用者が介入する前に証跡が失われるリスクがある。また、収集したマルウェアの検体を隔離されていない通常のストレージに保存すると、誤って実行され、意図しない二次感染や拡散を引き起こすおそれがある。これを防ぐため、サイバーデセプション資産上の揮発性データやマルウェア検体を安全かつ適切に保存・スナップショット化する仕組みが必要となる。

効果測定（メトリクス化）の欠如

おとり環境は導入して完了するものではなく、変化する環境や攻撃者に合わせて継続的なチューニングが必要となる。効果測定の仕組みがない場合、おとりのリアリズムが陳腐化し、攻撃者に見破られていても気づくことができず、意味のないサイバーデセプション資産にリソースを浪費し続けるリスクが生じる。そのためおとりのリアリズム、対話率、収集したインテリジェンスの質などを定量的に測定し、作戦目標に照らして継続的に改善する仕組みが必要となる。

6.4.4　OT/ICS 環境における導入上の留意点

ここまでは IT 環境を主眼に置いた記載をしてきたが、ここでは、OT/ICS 環境について触れる。OT/ICS 環境とは、工場、プラント、電力、水道、交通、ビル設備などにおいて、物理的な設備やプロセスを監視・制御するためのシステム環境を指す。具体的には、PLC、HMI、SCADA、エンジニアリングワークステーション、ヒストリアン、フィールドデバイス、制御ネットワークなどが対象となる。

IT 環境では、情報の機密性や完全性が重視される場面が多い。一方、OT/ICS 環境では、人命、安全、設備保護、操業継続への影響を避けることが最優先となる。そのため、OT/ICS 環境におけるサイバーデセプションは、IT 環境と同じ感覚で「サイバーデセプション資産を配置する」だけでは不十分である。最も重要なのは、物理プロセスに影響を与えないことである。

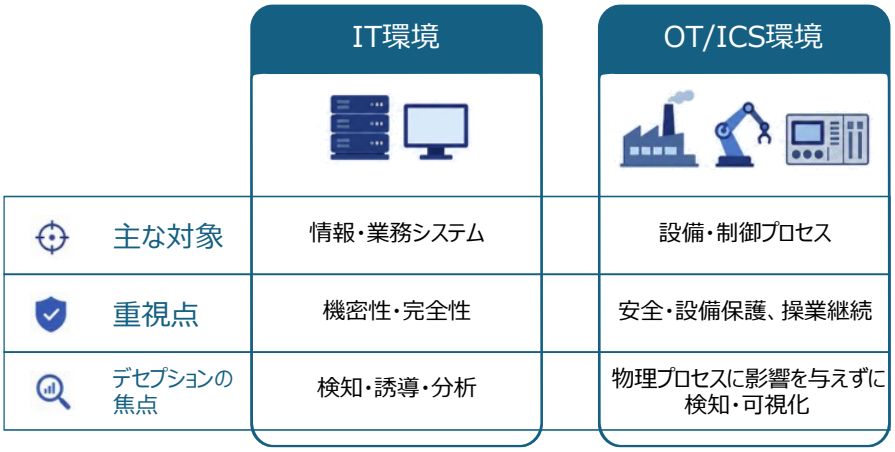


図 6-11　IT 環境と OT/ICS 環境における重視点の違い

さらに、OT/ICS 環境におけるデセプションの目標は、制御プロセスの安全性と継続性を損なうことなく、フィールドデバイスやコントローラに対する偵察、探索、異常な通信、敵対的な操作の兆候を検知することである。加えて、攻撃者を隔離されたおとり環境へ誘導し、行動を遅延・混乱させ、攻撃手法や通信内容に関する情報を収集することも副次的な目標となる。ただし、OT/ICS 環境では前述のとおり、安全性と可用性が最優先であるため、検知と観測を中心とした慎重な設計が求められる。

OT/ICS 環境でサイバーデセプションを導入する際の基本方針は、「どこにサイバーデセプション資産を配置するか」よりも先に、何に影響を与えてはならないかを明確にすることである。生産ラインや制御盤、監視端末、保守用端末などを含め、どの領域にサイバーデセプションを適用できるかを、現場部門と連携して整理することが出発点となる。特に重要なのは、おとり環境を本番の制御ネットワークから厳格に分離することである。「Cyber Deception: Fundamental Guide」では、OT/ICS 向けのおとりは本番の制御ネットワークから物理的/論理的エアギャップをはじめとした厳格な隔離を行う必要があり、稼働中の PLC と同じブロードキャストドメイン（同一の L2 ネットワーク範囲）に置くべきではないとされている。これは、おとり環境が攻撃者に悪用された場合に、本番制御機器への横展開や誤通信につながることを防ぐためである。

また、OT/ICS 環境においては、サイバーデセプション要素を実環境へ配置する前に、受動的な監視と可視化から始めることが望ましい。具体的には、ミラーポート、一方向コレクタなどを活用し、制御ネットワーク上で通常どのような通信が発生しているかを把握したうえで、デセプションによって検知すべき異常な接触や通信を定義する必要がある。

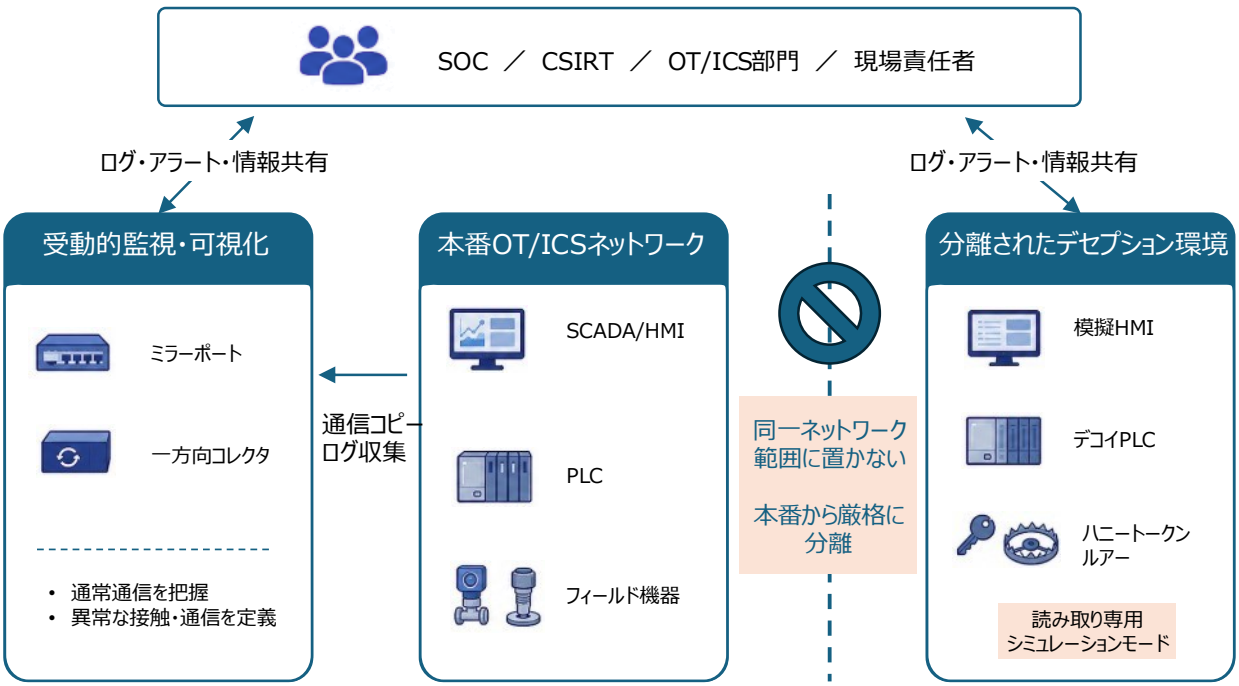


図 6-12　OT/ICS 環境における安全なデセプション配置イメージ

さらに、OT/ICS 環境向けのデセプションは、読み取り専用またはシミュレーションモードを基本とし、書き込み要求や制御命令を安易に受け付けない設計とすることが重要である。攻撃者に本物らしく見せることは必要であるが、それ以上に本番設備や制御プロセスへの影響を確実に避ける設計が求められる。

加えて、OT/ICS 環境では、アラート発生時の判断と対応を IT 部門だけで完結させるべきではない。サイバーデセプション資産への接触、プロトコル操作、書き込み要求、制御命令の試行などが検知された場合に、SOC、CSIRT、OT/ICS 部門、現場責任者の間でどのように情報共有し、誰が設備影響を判断し、どの条件で封じ込めや遮断を行うかを事前に定めておく必要がある。

6.5 サイバーデセプションの評価・改善

サイバーデセプションは、一度配置して終わるものではない。攻撃者の TTPs は継続的に変化し、組織のシステム環境や守るべき資産も変化する。そのため、導入したサイバーデセプション環境が当初の設計意図どおりに機能しているかを継続的に評価し、必要に応じて改善することが重要である。

本節では、運用後の評価・改善において、設計時に定義した Why、When、Where、How をどのように評価し、見直すかを整理する。すなわち、設計時に定めた要素を導入後の確認項目として再利用することで、設計意図に基づく評価が可能となる。

まず、Why については、当初設定した作戦目的が達成されているかを確認する。例えば、早期検知を目的としていた場合には、攻撃者の接触をどの程度早く把握できたか、またその検知が高信頼なシグナルとして扱えるかを確認する。TTPs の把握を目的としていた場合には、攻撃者の操作、探索対象、利用ツールなどについて有益な情報が得られたかを確認する。

次に、When については、想定した攻撃フェーズでサイバーデセプションが機能しているかを確認する。例えば、認証情報窃取や横展開を対象としていたにもかかわらず、実際には探索段階でしか接触が確認されない場合、それは攻撃者がまだ対象フェーズまで到達していないためなのか、あるいは想定した攻撃フェーズや配置が適切でなかったためなのかを確認する必要がある。

Where については、配置した環境やレイヤーが適切であったかを確認する。攻撃者がサイバーデセプション資産に接触しない場合、配置場所が攻撃者の行動経路から外れている可能性がある。また、本番環境との距離が近すぎる場合には、安全性や誤操作リスクを踏まえて、環境設計や境界設計を見直す必要がある。

How については、採用した技術や実装方法が有効であったかを確認する。例えば、ハニーファイルが不自然で攻撃者に無視された場合は、ファイル名、配置場所、内容の説得力を見直す必要がある。ハニーポットが容易に識別された場合は、サービスの応答、ログイン導線、周辺情報の整合性を改善する必要がある。

評価結果は、次の設計サイクルへフィードバックし、設計の見直しに活用することが重要である。想定外の攻撃フェーズで活動が観測された場合は、When の見直しが必要となる。攻撃者が配置したサイバーデセプション資産に接触しない場合は、Where や誘引内容の見直しが必要となる。検知精度や誘導効果が不十分であった場合は、How の再設計が必要となる。また、セキュリティ戦略の見直しにより、当初設定した目的が適合しなくなった場合は、Why の再定義が必要となる。

このように、CDF は、サイバーデセプション環境を設計するための枠組みであると同時に、運用後の評価・改善における確認観点としても機能する。設計時に定めた Why、When、Where、How を評価し、その結果を次の設計に反映することで、サイバーデセプション環境を継続的に最適化できる。これにより、サイバーデセプションは単なる技術導入ではなく、想定される脅威や攻撃者像、事業要件に応じて継続的に最適化される運用モデルとして機能する。

【コラム 6】 欺瞞の食卓 ～エミュレートされた蟹、パディングされた肉～

「欺瞞（ぎまん）」という言葉の響きには、どこか冷酷で人を陥れるような暗い影がつきまといます。デセプションもまた、その翻訳ゆえに、悪意ある騙し合いの罠だと誤解されがちです。しかし、その本質は決して卑怯な暗躍などではありません。限られた手札の中で、重要な資産を死守するために生み出された、高度で献身的な「工夫」なのです。そしてその真髄は、私たちの日常を彩る食卓にも密やかに息づいています。

本物を代替する美しき偽装――「カニカマ」です。スケトウダラに物理的なエミュレートを施し、カニエキスという甘美な痕跡情報をまとわせることで、対象者の五感をハッキングします。それは「騙す」ための悪意ではなく、食卓に彩りと喜びを添えるための工夫です。

真実を覆い隠す幻惑の装飾――「豆腐ハンバーグ」です。少量の肉に良質な豆腐で物理的なパディングを施し、ふっくらとした食感という新たな価値をまとわせることで、対象者の胃袋を静かに満たしていきます。それは「誤魔化す」ための罠ではなく、料理をより豊かにする工夫なのです。

では、これら食卓の工夫を、サイバー空間の防衛線へ応用したなら、どうなるでしょうか。

「美しき偽装」は、精巧なハニートークンとなってシステム内に配置されます。暗闇から忍び寄る侵入者は、意図的に置かれたこの甘美なダミーを価値ある情報だと信じ、手を伸ばします。しかし、その瞬間、防御側は接触や利用の兆候を検知できます。攻撃者は本物の機密資産へ近づいたつもりでも、実際には防御側が用意した偽の資産に誘導されているのです。これは、正面から力ずくで押し返すのではなく、攻撃者の判断を利用して早期検知と観測につなげる工夫です。

「幻惑の装飾」は、価値ある情報を見つけにくくするデータ・パディングとして機能します。純粋な牛肉だけを奪い去ろうとする攻撃者は、意図的にブレンドされた圧倒的な豆腐の海、すなわち無害なデータ群の中から、本命の肉を探し出さなければなりません。その選別に時間と手間をかけさせることで、防御側は攻撃者の判断を迷わせ、リソースを消費させることができます。サイバーデセプションは、攻撃者を力ずくで止めるだけの考え方ではありません。精巧な代替品で接触を誘い、良質なノイズで判断を迷わせ、その過程を観測しながら真の資産を守る考え方です。

騙すことは決して悪ではありません。重要なのは、その欺瞞が何を守るために、どのような意図で設計されているかです。デセプションとは、他者を陥れる卑劣な罠ではなく、攻撃者の認知と判断に働きかける、創造的で前向きな防衛の「工夫」なのです。

（余白ページ）

第7章　企業・組織における導入の考え方と設計例

7.1　企業・組織における導入の考え方

前章では、サイバーデセプションを実環境に適用するための設計・運用プロセスとして、基本設計、詳細設計、実装・運用、評価・改善の考え方を整理した。本章では、その考え方を企業・組織に適用する際の導入方針を示したうえで、CYDEC Design Framework （CDF）を用いた設計例を示す。

MITRE Engage の各アプローチ（Engage Approaches）は、すべての企業・組織が同じように実装すべきものではない。日本の企業・組織がサイバーデセプション環境を運用する場合、アプローチごとに導入難易度や運用上のリスクは大きく異なる。実際に、「表 6-6 プロセス 1・2 に基づく設計例」で挙げたすべての事例について、実装から運用までを実現できる企業・組織は多くないと想定される。

特に、高度なセキュリティ基盤を持たない企業・組織において、攻撃者の行動を脅威インテリジェンスとして活用するようなサイバーデセプション戦略を採用する場合、導入コスト、専門人材、監視・分析能力の不足により、運用の難易度が高くなる。また、「6.4.3 運用上の落とし穴」で述べたように、設計や運用体制が不十分なまま高度なサイバーデセプション環境を導入すると、運用破綻を招くリスクもある。例えば、「暴露させる（Expose）」における「収集（Collect）」や、「引き出す（Elicit）」における「安心・確信（Reassure）」、「動機付け（Motivate）」を活用する事例は、攻撃者の行動や手法を詳細に分析することを目的とする場合に有効である。一方で、その運用には、ログの分析能力、おとり環境の管理、継続的な監視、法務・倫理面の整理などが求められるため、公的機関、研究機関、高度な SOC を有する組織などに適している。

一方、多くの企業・組織にとって、セキュリティに取り組む主な目的は、攻撃者からハードウェア資産や情報資産を守り、事業を継続することにある。そのため、日本の企業・組織においては、MITRE Engage の各アプローチの中でも、特に「検知（Detect）」と「阻止（Prevent）」の優先度が高くなる。

この傾向は、2026 年現在において市場に展開されているサイバーデセプション製品や OSS の活用事例の多くが、「高精度な検知」や「被害の阻止」を中核として設計されている点にも表れている。ただし、これまで述べてきたとおり、それらの製品を導入するだけで直ちに効果的な成果が得られるわけではない。CDF で整理したとおり、まずセキュリティ戦略に基づいてサイバーデセプション作戦を策定し、目的達成に必要な作戦目標、攻撃フェーズ、仕掛け先、手段を明確にする必要がある。その上で、自組織の体制や成熟度に応じてサイバーデセプション製品や OSS を活用することで初めて、組織のセキュリティ戦略に資する実効性のある防御策となる。

次節では、仮想企業を設定し、CDF を用いたサイバーデセプション設計例を示す。あわせて、その設計例の一部について実際の効果を検証した結果を後続章に示す。

7.2　CYDEC Design Framework（CDF）を用いたサイバーデセプション設計例

本節では、CDF を用い、導入を検討する組織の視点からサイバーデセプションを設計する例を示す。Who（主体）、What（資産）、by Whom（脅威）の整理を前提条件として設計検討を進める。

シナリオ設定の背景として、専任の SOC を持たず、IT 管理者がセキュリティを兼務している中規模企業を想定する。ランサムウェア被害や標的型攻撃によるデータ漏えいに対する危機感はある一方で、「誤検知による運用負荷の増大」と「高額な導入コスト」を避け、既存環境を大きく変えずに、スモールスタートでのサイバーデセプションの設計・導入を検討している。

想定する仮想企業をもとに、自組織の環境や守るべきもの、対抗する脅威アクターを整理したものを下表に示す。この定義を以降のサイバーデセプション作戦の検討における前提とする。

表 7-1　仮想企業における前提定義例

項目	整理内容
Who（主体）	専任 SOC なし。高度なセキュリティ分析や大規模システムの運用・保守に割けるリソース（時間・予算）は極めて限定的
What（資産）	ファイルサーバー上の顧客データ、機密の財務情報、事業継続に必要な設計データ
by Whom（脅威）	フィッシングなどにより一般従業員 PC を侵害し、特権アカウントを探索して横展開し、最終的にファイルサーバー上のデータを窃取または暗号化しようとする標的型攻撃者またはランサムウェア集団

7.2.1　サイバーデセプション作戦の立案（Why/When/Where）：プロセス 1

前提条件として整理したリソース制約を踏まえ、現実的な運用能力を超える目標（短期間での高度な脅威インテリジェンスの大量収集、長時間にわたる攻撃者行動の観測など）は避け、自組織に適したサイバーデセプション作戦を立案する。

【Why（目的・目標）】

自組織の体制を踏まえ、サイバーデセプションの導入によってどのような効果を得たいかを定義する。

表 7-2　仮想企業における Why 定義例

項目	選定理由
作戦目標 （Goals）	暴露させる（Expose）： 専任 SOC がないため、まず「攻撃者の存在に気づくこと」を最優先にしなければ、対処の開始自体が遅れる。そのため、スモールスタートの作戦目標として「暴露させる（Expose）」を選定する。
アプローチ （Approaches）	検知（Detect）： 限られた運用体制で高度分析を常時実施することは困難なため、まずは攻撃者活動の可視化と検知を優先する方針として、「検知（Detect）」を主アプローチに採用する。

【When（攻撃フェーズ）】

攻撃の進行段階のうち、どの段階に対してサイバーデセプションを作用させるかを定義する。

表 7-3　仮想企業における When 定義例

項目	選定理由
優先する When	最優先：認証情報窃取、横展開、持ち出し 次点：探索、C2 既存製品で対応：初期侵入 専任 SOC がないため、すべてのフェーズを同じ深さで追跡する運用は難しい。 そのため、被害に直結する連鎖（初期侵入→認証情報窃取→横展開→持ち出し）を優先して対応対象とする。一方で、運用負荷と導入容易性を考慮し、初期侵入フェーズの監視は既存のセキュリティ製品による対策を基本とし、本例におけるサイバーデセプションの主対象は認証情報窃取、横展開、持ち出しとする。

【Where（仕掛け先）】

選定した攻撃フェーズに対し、攻撃者との接点を確保しやすい最適な配置レイヤーを定義する。

表 7-4　仮想企業における Where 定義例

項目	選定理由
エンドポイント層	想定脅威が「初期侵入→認証情報窃取→横展開」で進行するため、侵害後の活動拠点・中継点となる層を押さえる必要がある。認証情報窃取や横展開の初期兆候を捉えやすく、少人数運用でも優先監視の効果を得やすい。
データ層	探索・収集・持ち出しの各段階で攻撃者との接触を得やすく、被害直結行動の可視化に有効。また、主要リスクが「データ窃取・暗号化」であるため、攻撃者が最終的に価値を見出す対象に近い層を優先する。

7.2.2　サイバーデセプション戦術への落とし込み（How）：プロセス 2

プロセス 1 で策定したサイバーデセプション作戦を具現化するため、How（手段）の段階で具体的な技術選定や実装方式へと落とし込みを行う。本例では、前提条件として整理した「新規の偽サーバー設置や大規模なネットワーク変更は行わない」という前提を踏まえ、導入する技術候補を抽出し、攻撃者の心理（認知バイアス）の利用や運用指針を含めて検討を行う。

【How（手段）】

自組織に導入するサイバーデセプションの具体的な技術や実装方式を決定する。

表 7-5　仮想企業における How 選定例

#	技術	実装概要	想定する When	攻撃者心理の利用
1	ハニートークン（クレデンシャル型）	IT 運用スクリプト保管フォルダに「backup_job_notes.txt」を配置し、偽認証情報「svc_backup_sync / BkSync!2026」を記載する。偽認証情報の参照・認証試行（SMB/RDP など）を検知する。	認証情報窃取、横展開	運用メモ（txt ファイル）に記載された認証情報を有効なものと判断しやすい傾向、svc_系名称への信頼、最短経路を選びやすい傾向を利用する。

#	技術	実装概要	想定する When	攻撃者心理の利用
2	ハニートークン（ドキュメント型）	共有フォルダに偽文書（例：【機密】主要取引先_契約一覧_2026.xlsx）を配置する。開封・複製・圧縮・外部送信を検知する。	探索、収集、持ち出し	価値が高そうな文書名を優先する傾向、目立つ情報に注意を向けやすい傾向、既に得た仮説を裏付ける情報を探しやすい傾向を利用する。

以上より、本仮想企業では、Why（目的・目標）として「暴露させる（Expose）」を選定し、その実現に向けたアプローチとして「検知（Detect）」を採用した。重要資産に対する攻撃チェーンに直結する When（攻撃フェーズ）を優先し、既存のセキュリティ製品との役割分担を考慮した上で、Where（仕掛け先）を「エンドポイント層」と「データ層」に絞り込むサイバーデセプション作戦を立案している。

この作戦を具体的な戦術へ落とし込むにあたり、既存のインフラ環境を大きく変更せず、運用負荷を低く抑えられる「ハニートークン（クレデンシャル型・ドキュメント型）」を採用した。これにより、IT 管理者の兼務体制であっても、攻撃者の主要行動である「認証情報取得」「横展開」「持ち出し」を少ない誤検知で安全に可視化し、能動的防御の仕組みをスモールスタートで実現する構成となる。

【コラム 7】「デセプション導入済」って、言っちゃう？

サイバーデセプションは攻撃者を欺く防御手法です。そう考えると、素朴な疑問が浮かびます。「そもそも、デセプションを導入していることを公表してよいのでしょうか」。

一見すると答えは簡単そうです。「欺くための仕掛けなのだから、黙っておくべきだ」そう考えるのが自然です。実際、どこにおとりがあるのか、どの認証情報がハニートークンなのか、どのログ条件でアラートが上がるのかを公開してしまえば、攻撃者に回避の手がかりを与えることになります。これはマジシャンが舞台に立つ前に、観客へ「次は右手で注意を引き、左手で仕掛けを動かします」と説明してしまうようなものです。

『Cyber Deception: Fundamental Guide』では、デセプションについて、必要な社内関係者には導入を認識させる一方、外部に対しては具体的なデセプション戦術の公開を避けることが推奨されています。

では、具体的なことには触れず、「デセプション導入」のみを公表するのはどうでしょう？

相手が金銭目的のサイバークライム集団であれば、デセプション導入の公表そのものが一定の抑止になる可能性があります。彼らの多くは限られた時間でより確実に利益を得られる標的を探します。「侵入しても本物と偽物を見分けにくく、認証情報やファイルに触れるだけで検知されるかもしれない」そう思わせることができれば、"面倒な会社"と判断され、より容易な標的へ向かうかもしれません。

一方で国家支援型の APT 集団のように、特定組織への侵入そのものに強い目的を持つ攻撃者にはこのような抑止は限定的かもしれません。彼らは多少面倒だからといって簡単に諦めるとは限りません。むしろ、デセプションの存在を前提に、より慎重な偵察、環境のフィンガープリント確認、おとり回避の手法を用いる可能性があります。この場合、公表は抑止ではなく、攻撃者に「注意深く行動せよ」という警告として受け取られてしまうおそれもあります。

さらに別のタイプとして、腕試しをしたいハッカーや好奇心の強いハッカーも考えられます。サイバーデセプションを導入していると公表することで、「突破してみたい」「どのような仕掛けなのか見てみたい」と興味を引いてしまう可能性があります。これは、防御側が意図せず“挑戦状”を出してしまうようなものです。

つまり、サイバーデセプションの公表は相手によって意味が変わります。

金銭目的の攻撃者には、面倒な標的だと思わせる効果があるかもしれません。国家支援型の攻撃者には、警戒を強めさせるだけかもしれません。腕試し型のハッカーには、逆に関心を持たせてしまうかもしれません。

読者の皆さんはどう考えますか？

サイバーデセプションの公表は単なる広報判断ではありません。攻撃者をどう見ているか、自組織のリスクをどう捉えるか、そして何を守るために何を見せるのかを問う、もう一つのデセプション設計なのかもしれません。

コラム内参考資料

Bilal Qureshi, Cyber Deception: Fundamental Guide. Independently published, 2026.

（余白ページ）

第8章　サイバーデセプションの検証

8.1　検証内容と結果の概要

本章では、「表 6-6　プロセス 1・2 に基づく設計例」で整理した 9 つの設計例のうち、サイバーデセプションの中核的な作用である「攻撃者の認知・判断・行動への働きかけ」が表れやすい 3 つのパターンに着目し、技術検証を実施した。具体的には、「攻撃者の探索コストの増大」「真正性判断への影響」「偽認証情報による次行動の誘発」を検証対象とした。これらの検証を通じて、サイバーデセプションが攻撃者の認知・判断・行動にどのように作用し得るかを確認し、実務導入に向けた示唆と留意点を整理する。

「第 6 章　サイバーデセプションの設計・運用」では、サイバーデセプションの設計・運用プロセスとして、「サイバーデセプション作戦の立案：プロセス 1」と「サイバーデセプション戦術への落とし込み：プロセス 2」を整理し、その適用例として「表 6-6　プロセス 1・2 に基づく設計例」を示した。この表では、MITRE Engage の考え方を踏まえ、攻撃者の検知、攻撃行動の阻止、探索経路の誘導、攻撃活動の妨害、行動の誘発など、サイバーデセプションによって実現可能な複数の効果に対応する設計例を整理している。

本章では、これらの設計例すべてを検証対象とするのではなく、冒頭で述べた 3 つの検証視点に対応する設計例である「表 6-6　プロセス 1・2 に基づく設計例」の No.6、No.7、No.9 を検証対象とした。これにより、攻撃者の探索に時間的負荷を与えること、提示された環境や情報を本物と判断させること、偽の情報を起点として次の行動を引き出すこと、というサイバーデセプションが攻撃者の「時間」「認知」「行動」に与える影響を確認する。

本章で検証する仮説、検証内容および結果の概要を下表に示す。

表 8-1　検証仮説および結果の概要

検証 No.	検証仮説	検証内容	検証結果（概要）
1	応答遅延を組み込んだ低対話型ハニーポットを配置することで、攻撃者のネットワークスキャンや探索行動に要する時間を大幅に増大させられるのではないか。	tc コマンドで応答遅延を設定した低対話型ハニーポットを構築し、遅延なしの場合と 1.5 秒の応答遅延を設定した場合で、ポートスキャンに要する時間を比較した。（表 6-6　プロセス 1・2 に基づく設計例 No.6）	「応答遅延なし」では約 3.5 分で完了したスキャンが、「1.5 秒の応答遅延」により約 5 時間 14 分を要した。応答遅延により探索時間が約 88 倍となり、攻撃者の探索コストを大幅に増大させる効果が確認された。
2	ハニートークン（ハニーファイル）は、ファイル名だけでなく、タイムスタンプやファイルサイズなどのメタデータを自然に作り込むことで、攻撃者の真正性判断を困難にできるのではないか。	ファイル名のみを工夫したパターン 1 と、タイムスタンプやファイルサイズを実環境に近い形で調整したパターン 2 のハニーファイルを配置し、欺瞞効果を比較した。（表 6-6　プロセス 1・2 に基づく設計例 No.7）	パターン 1 は、メタデータを一括確認する AI 攻撃エージェントに容易に判別された。一方、パターン 2 は人間の攻撃者にとって真正性判断を難しくし、AI エージェントに対しても検証ステップを増加させ、看破までの時間を引き延ばす効果が確認された。
3	ハニートークン（ハニークレデンシャル）は、作り込みパターンが高いほど、攻撃者を誘引し、おとり環境への誘因の成功率を高めることができるのではないか。	パターン A（デスクトップ上のテキストファイル）、パターン B（PowerShell の実行履歴）、パターン C（システム変数とデスクトップショートカット）の 3 種類を作成し、「誘引の成否」の効果を比較した。（表 6-6　プロセス 1・2 に基づく設計例 No.9）	当初仮説とは異なり、パターン B の PowerShell 実行履歴が最も高い欺瞞効果を示した。誘引層の作り込みだけでなく、おとり側の OS・ネットワーク・AD 整合性など、インフラ層との一貫性が欺瞞維持に重要であることが確認された。

以降では、検証環境および前提条件を整理した上で、各仮説の検証内容および結果を示し、サイバーデセプションがサイバーレジリエンスの向上にどのように寄与するかを考察する。

8.2　検証環境・前提条件（共通）

前節で定めた三つの検証仮説について検証を行うにあたり、検証対象となるサイバーデセプション作戦・戦術の前提条件を定める必要がある。そこで本節では、まずはモデルとする仮想企業を設定し、その仮想企業を基に CYDEC Design Framework（CDF）の Who（主体）、What（資産）、by Whom（脅威）に相当する前提条件を整理する。これにより、以降の検証内容を検討する際の共通前提を固定する。

8.2.1　仮想企業の設定とセキュリティ戦略

仮想企業の概要およびネットワーク構成を説明する前に、OT 環境における本来の考え方を整理する。「6.4.4　OT/ICS 環境における導入上の留意点」にて述べたとおり、OT 環境では、安全性と可用性が最優先であり、セキュリティ対策もこれらを損なわない形で検討する必要がある。米国国立標準技術研究所（NIST）が発行する産業用制御システムセキュリティガイド「NIST SP 800-82r3」[17]では、OT 環境の保護にあたり、企業ネットワークと OT ネットワークを直接接続するのではなく、DMZ 等を用いて通信を分離・制御する考え方が示されている。また、単一の対策に依存せず、多層防御を適用することや、管理権限、信頼レベル、機能重要度、データフロー、設置場所などを踏まえて、IT/OT 機器をゾーン化することの重要性も示されている。

したがって、本来は、IT 環境と OT 環境の間に DMZ 等の境界領域を設け、必要な通信経路を限定し、各ゾーン間の通信を適切に制御・監視する構成が望ましい。これにより、IT 環境で発生した侵害が OT 環境へ直接波及することを抑制し、仮に一部の防御策が突破された場合でも、被害の拡大を段階的に抑えることが可能となる。

一方で、本検証では、IT 環境と OT 環境の融合が急速に進んだ結果、OT 環境のセキュリティ対策が十分に整備されていない仮想企業を想定する。具体的には、インターネット接続用の DMZ は設けられているものの、社内 IT 環境と社内 OT 環境の間には、OT 向けの DMZ 等の中間領域は設けられていない。また、OT 環境内においても、機能重要度や信頼レベルに応じたゾーン化は十分に行われていない状態とする。

このような構成は、OT 環境のあるべき姿を示すものではない。むしろ、あるべきセキュリティ構成に到達する前の過渡的な状態を想定したものである。本検証では、そのような不十分な防御状態に対し、サイバーデセプション作戦を適用することで、攻撃者の探索、判断、行動にどのような影響を与え得るかを確認する。

上記を踏まえて設定した仮想企業の概要、および想定するネットワーク構成を以下に示す。

表 8-2　仮想企業の概要

業種	都市ガス供給
資本金	150 億円
売上高	2,800 億円
従業員	2,500 名
顧客規模	110 万件
固有資産	顧客管理・料金計算データ、導管地理情報データ、設備保全データ、 ガス製造プラント、ガス供給導管など

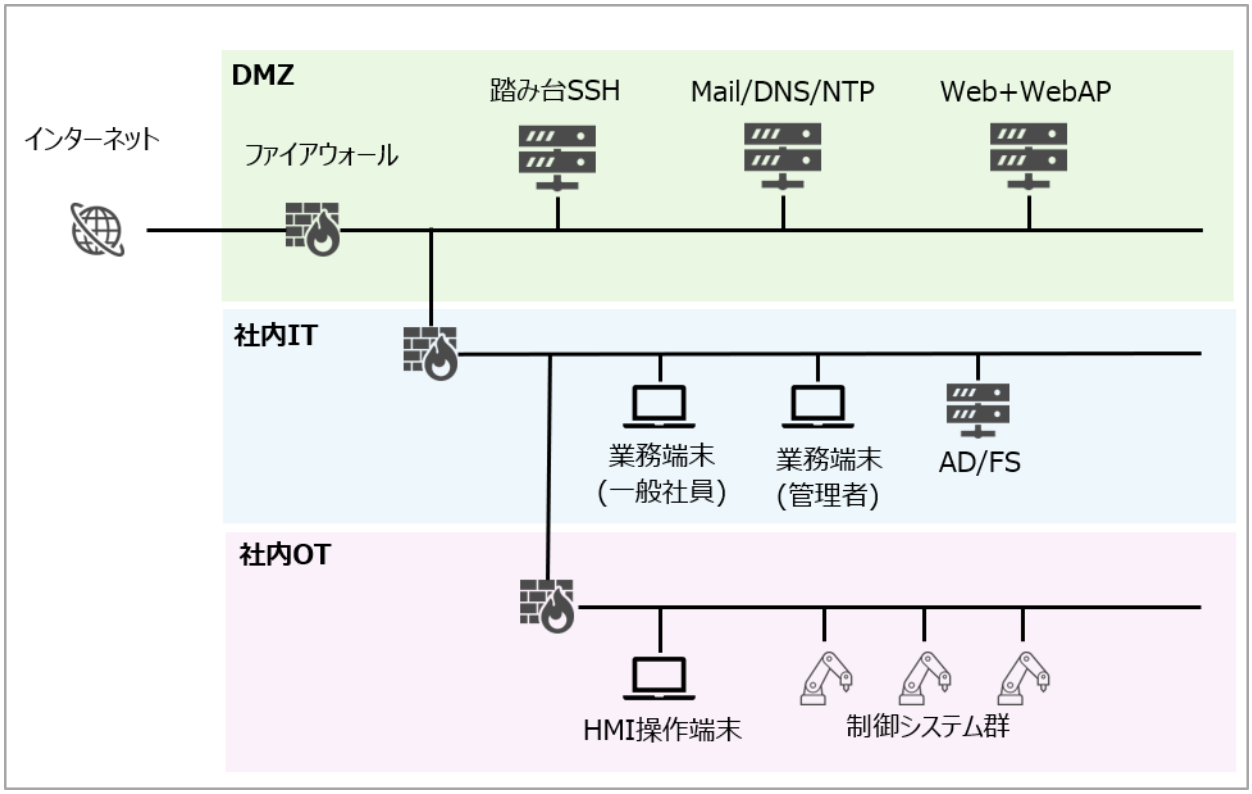


図 8-1　仮想企業のシステム構成図

さらに、設定した仮想企業に対して、CDF におけるサイバーセキュリティ戦略となる Who（主体）、What（資産）、by Whom（脅威）の観点から、以下のように前提条件を整理した。

表 8-3　仮想企業におけるサイバーセキュリティ戦略

Who	（自組織の） 防御姿勢とリスク許容度	■OT/ICS（SCADA/ガスプラント/導管） - 防御姿勢：高 - リスク許容度：極めて低 重要インフラ事業者として、ガスの製造・供給停止は社会的損害・人命リスクに直結する。そのため、攻撃者にとって「操業停止」は最も強力な脅迫材料となりうる。 ■顧客情報 - 防御姿勢：高 - リスク許容度：低 個人情報保護法およびガス事業法の厳しい規制対象である。また、110 万件規模の顧客データは金銭目的の攻撃者にとって価値が高く、直接的な奪取目標として狙われうる。 ■生産ノウハウ/知財（製造プロセス、技術文書） - 防御姿勢：中 - リスク許容度：中 金銭目的の攻撃者にとって、ノウハウそのものを直接換金する手段は限られる。そのため、攻撃者におけるプライオリティは OT や顧客情報に比べて相対的に低いと考えられる。
What	守るべき資産	OT/ICS（SCADA/ガスプラント/導管） 顧客情報（個人情報/契約データベース） 生産ノウハウ/知財 業務 IT 機器
by Whom	脅威アクター	サイバークライム集団

8.2.2　検証環境の構築

検証を行うにあたっては、前項で設定した仮想企業ネットワークをクラウド上に構築し、検証環境として使用した。構築した検証環境のネットワーク構成図、および構成要素の概要を以下に示す。

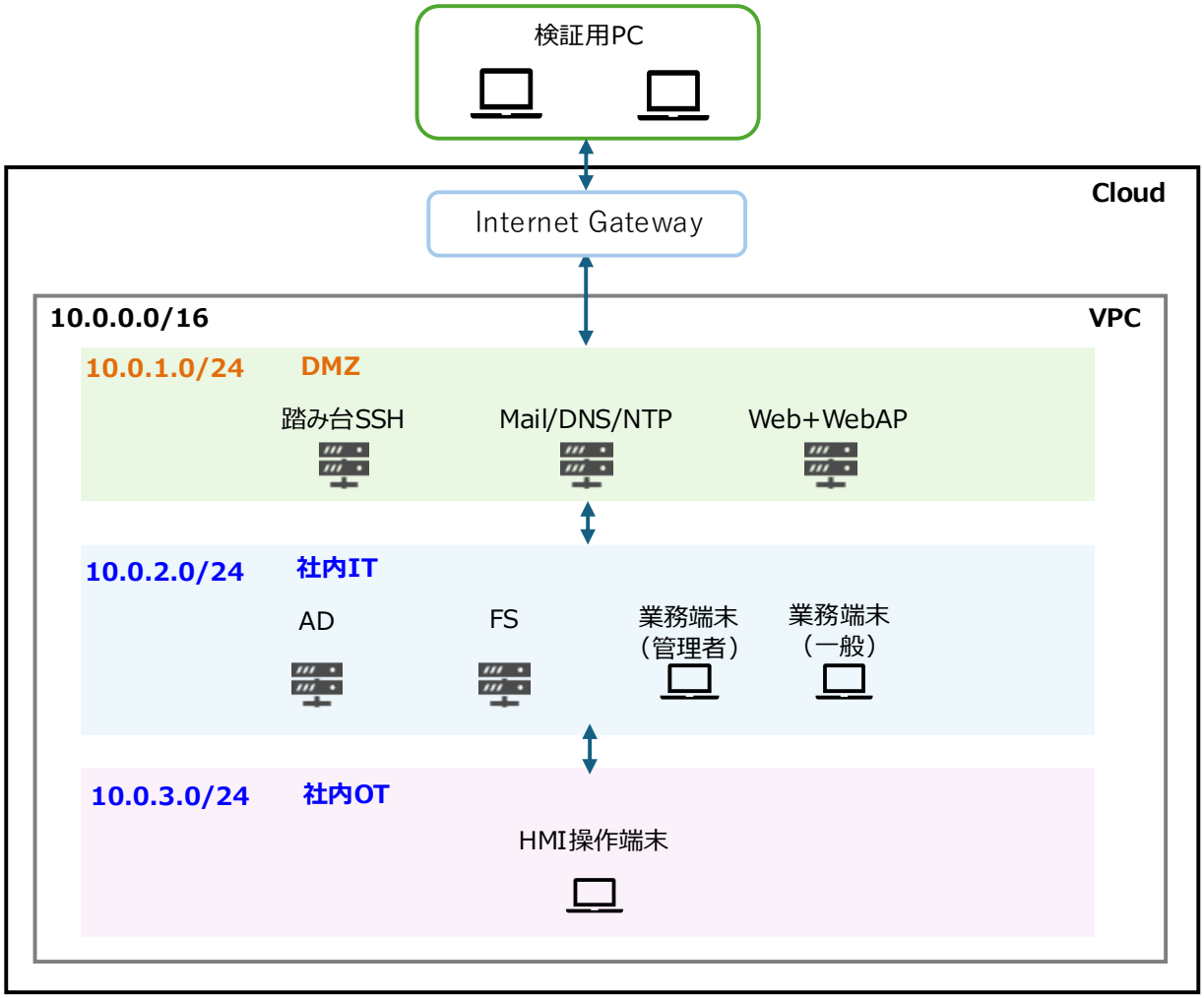


図 8-2　検証環境の構成

表 8-4　検証環境の説明

エリア	構成機器	役割
DMZ	踏み台 SSH	仮想企業の社員がリモートアクセスに使用する踏み台サーバー。
	Mail/DNS/NTP	業務に使用するメール、名前解決、時刻同期サービスを提供する Linux サーバー。
	Web＋WebAP	顧客向けポータルサイトを模擬する一般的なウェブサーバー。
社内 IT	AD／ファイルサーバー	社内 IT 環境における認証基盤およびファイル共有基盤を担う。Windows 端末の認証情報や疑似顧客情報を保管する。
	業務端末（管理者）	管理者が使用する業務端末。AD やファイルサーバー等への管理権限を持つ。
	業務端末（一般社員）	一般社員が使用する業務端末。AD によるログイン認証を行うが、AD に対する管理権限は持たない。
社内 OT	HMI 操作端末	OT エリアを模擬するために設置する端末。実際の制御対象は持たない。

また、検証 2 及び 3 においては AI エージェントを使用した。使用した AI エージェント環境およびモデルの諸元は以下のとおりである。

- 実行プラットフォーム：Google Antigravity2.0（AI エージェント管理プラットフォーム）
- 使用形態：自律型開発エージェント
- エージェント構成：マルチエージェント構成
- 使用 AI モデル（LLM）：
  Anthropic Claude Opus 4.6
  Google Gemini 3.1 Pro
- 推論／実行モード
  Claude Opus 4.6：Thinking Mode
  Gemini 3.1 Pro：High Mode

8.3　検証 1 低対話型ハニーポットによる探索コストの増大

本検証は、「表 6-6　プロセス 1・2 に基づく設計例」の No.6「ネットワークスキャン・横展開通信の遅延化」に対応するものであり、Affect（Disrupt）の観点から、攻撃者の探索コストの増大と探索効率の低下を確認することを目的とする。

具体的には、TCP 応答遅延および偽サービス応答を活用したサイバーデセプション技術が、攻撃者によるポートスキャンおよびサービス調査に与える影響を評価する。

8.3.1　検証の狙い

本検証では、攻撃者が初期侵入後に内部ネットワークを探索する場面を想定し、低対話型ハニーポットおよび応答遅延を組み合わせたおとり環境が、攻撃者の探索活動にどのような影響を与えるかを確認する。

攻撃者は、侵入後にネットワーク上の生存ホストを列挙し、発見したホストに対してポートスキャンを実施することで利用可能なサービスや攻撃の起点を探索する。この探索活動に対し、防御側がおとりや応答遅延を組み込むことで、攻撃者の調査対象を増やし、探索に要する時間を引き延ばせる可能性がある。

8.3.2　検証仮説

『低対話型ハニーポットに応答遅延を設定することで、攻撃者によるネットワークスキャンおよびポートスキャンに要する時間を増加させ、探索コストを大幅に引き上げられるのではないか。』

一般的に高対話型のハニーポットは、実環境に近い挙動を再現できる一方、導入コストや維持運用コストが高くなりやすい。また、構築・監視・隔離運用に必要な人的スキルも高く、一般の企業・組織にとって導入のハードルが高くなると考えられる。そこで本検証では、相対的に導入ハードルが低い低対話型ハニーポットであっても、応答遅延を組み合わせることで攻撃者の探索コストを増大させる効果を得られるかを検証する。

8.3.3　検証シナリオ

本検証では、攻撃者が DMZ 上の踏み台サーバーへの侵入に成功し、内部ネットワーク上の対象ホストに対して、ポートスキャンおよびサービス調査を実施する状況を想定する。その上で、TCP 応答遅延が攻撃者の探索活動に与える影響を評価した。

なお、本検証では TCP 応答遅延による時間消費効果を評価するため、おとり端末には全ポート応答設定を適用した。これは、攻撃者が多数のポートに対して調査を行う状況を再現し、応答遅延が探索時間に与える影響を確認するための検証上の条件である。

攻撃の流れは以下のとおりである。

1. ポートスキャン：対象ホストの開放ポートを調査する。
2. サービス調査：開放ポートから稼働サービスを推定する。
3. 後続攻撃：発見したサービスに対して脆弱性調査や認証試行等を実施する。

なお、ポートスキャンおよびサービス調査への影響を評価するため、以下のサイバーデセプション手法を検証対象とした。

表 8-5　検証 1 の検証対象

評価対象	内容	期待する効果
TCP 応答遅延	ポートスキャン時の応答に遅延を付与する	探索活動に要する時間の増加

また、TCP 応答遅延については以下の 4 条件での比較を実施した。

- ベースライン：応答遅延を設定しない状態
- 500ms 遅延：TCP 応答に 500ms の遅延を付与した状態
- 1000ms 遅延：TCP 応答に 1000ms の遅延を付与した状態
- 1500ms 遅延：TCP 応答に 1500ms の遅延を付与した状態

8.3.4　検証環境

本検証では、DMZ 上に配置した SSH サーバーを攻撃者が侵害した踏み台サーバーと見立て、内部ネットワーク上に配置したおとり端末に対してポートスキャンを実施する構成とした。おとり端末では、低対話型ハニーポットとしてポート応答を模擬し、TCP 応答遅延を付与することで、攻撃者の探索活動に要する時間への影響を確認した。検証環境の構成を下図に示す。

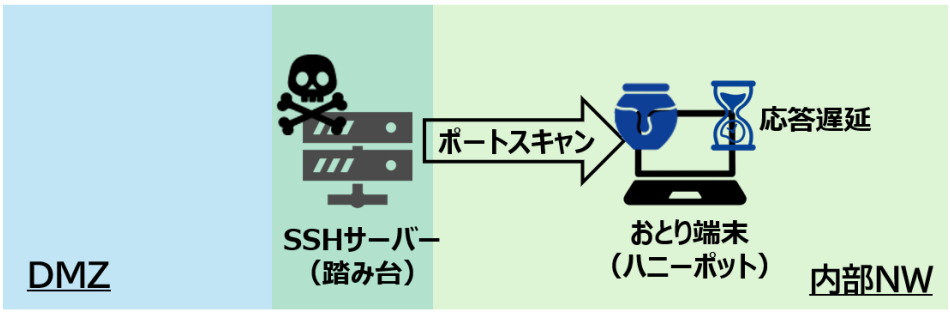


図 8-3　検証 1 の構成

8.3.5　検証手順

本検証では、効率的な探索活動を重視する攻撃者による一般的なポートスキャンを想定した。本検証で使用したコマンドを以下に示す。

```
nmap -p 0-65535 -T4 --open [対象 IP]
```

探索速度と検出精度のバランスを考慮し、Nmap のタイミングテンプレートとして-T4 を採用した。

ポートスキャンおよびサービス調査への影響を評価するため、以下の手順で検証を実施した。

1. 全ポート応答を設定したおとり端末（低対話型ハニーポット）を構築する。
2. TCP 応答遅延の設定値を変更する。
3. おとり端末に対してポートスキャンを実施する。
4. 所要時間を測定する。
5. 条件ごとの結果を比較し、評価する。

8.3.6　検証結果

各条件における測定結果を下表に示す。

表 8-6　検証 1 測定結果

遅延設定（ms）	所要時間（平均）（秒）	ベースライン比（倍）
遅延設定無し	214	1.0
500	1,085	5.0
1000	4,614	21.6
1500	18,841	88.2

また、ポートスキャンの所要時間をベースライン比で比較した結果を下図に示す。

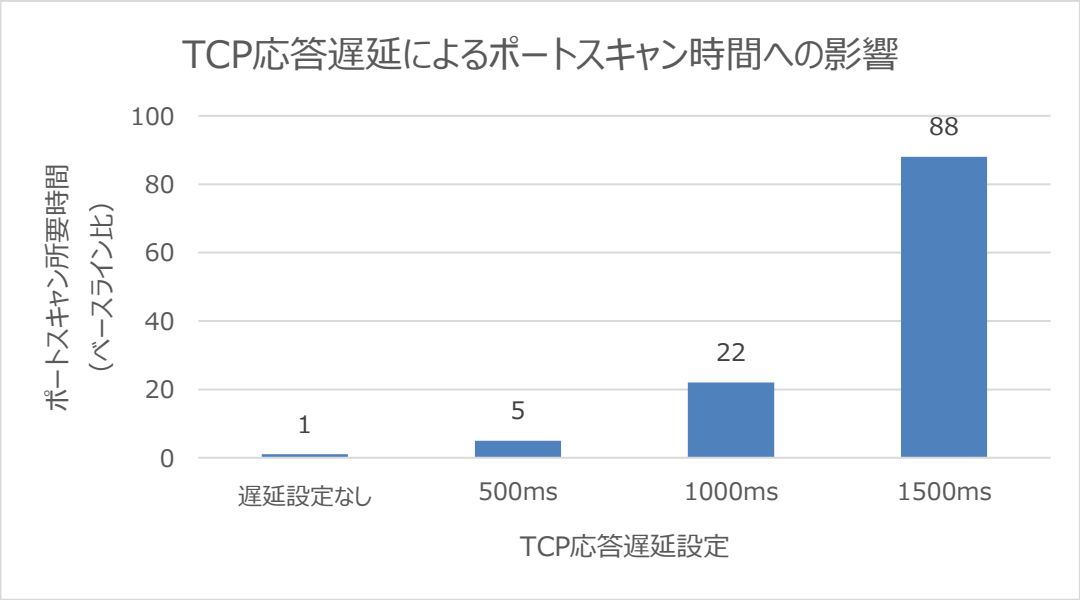


図 8-4　検証 1 ポートスキャン時間比較グラフ

TCP 応答遅延の増加に伴い、ポートスキャンの所要時間は増加する傾向が確認された。特に 1000ms 以上の遅延設定では、ベースラインと比較して大幅な所要時間の増加が確認された。1500ms の遅延設定では、ポートスキャンの所要時間が 18,841 秒となり、ベースライン比で約 88.2 倍に増加した。この結果から、TCP 応答遅延は、攻撃者の探索活動に要する時間を大幅に増加させる効果を持つことが確認された。

8.3.7　考察

(1)　TCP 応答遅延による時間消費効果

本検証では、TCP 応答遅延を付与することで、ポートスキャンに要する時間が増加することを確認した。

特に 1000ms 以上の遅延設定では、遅延設定値の増加率を上回る所要時間の増加が確認された。本検証では通信挙動の詳細分析までは実施していないものの、TCP 通信における応答待機や再送処理等の影響を受け、探索ツールの処理効率が低下した可能性が考えられる。

(2)　実務導入時の設計上の留意点

一方で、遅延時間を大きくすればよいわけではない。過度な遅延設定は、攻撃者の探索コストを増大させる効果を発揮する前に、おとり自体が不自然に見えるほか、スキャンツール側のタイムアウト等により防御側が期待する観測結果を得にくくなる可能性がある。そのため、実運用では、攻撃者から見た到達可能性と探索時間を引き延ばす遅延効果のバランスを取ることが重要である。

また、本検証は限定された検証環境における結果であり、すべてのネットワーク環境や攻撃ツールに同様の効果が得られるとは限らない。実環境へ適用する際には、正規通信への影響、おとり環境側の処理負荷、監視基盤との連携、アラート発報条件を確認した上で、段階的に導入することが望ましい。

以上から、低対話型ハニーポットと応答遅延の組み合わせは、偵察・探索フェーズにおいて攻撃者の時間的コストを増大させる手段となり得る。ただし、実務上有効に機能させるためには、遅延設定、おとり資産への到達可能性、運用監視との接続を一体で設計する必要がある。

8.4　検証 2 ハニーファイルとその配置環境の作り込みが攻撃者の判断に与える影響

本検証は、「表 6-6　プロセス 1・2 に基づく設計例」の No.7「攻撃目標を達成したと誤認させる」に対応するものであり、Affect（Disrupt）の観点から、ハニーファイルおよびその配置環境の作り込みが、攻撃者の探索行動や真正性判断に与える影響を確認することを目的とする。

具体的には、ハニーファイルを用いたおとり環境が、攻撃者の探索行動や価値判断に与える影響を評価する。

本書では、偽のデータを価値ある本物と認識させ、攻撃者の接触や利用の検知、リソース消費を達成するために意図的に配置するサイバーデセプション資産を、シンセティックデータとして整理している。検証 2 で扱うハニーファイルは、このシンセティックデータを含むファイルであり、攻撃者に価値あるファイルであると認識させ、接触、開封、持ち出し、後続行動を検知・誘導することを狙うものである。

本検証では、ハニーファイルの内容だけでなく、ファイル名、作成日時、更新日時、ファイルサイズ、配置場所、フォルダ構成などを含めた「本物らしさ」が、攻撃者の真正性判断にどのような影響を与えるかを確認する。

8.4.1　検証の狙い

本検証では、ハニーファイルおよびその配置環境の「本物らしさ」が、攻撃者の真正性判断に与える影響を確認する。ハニーファイルは、攻撃者に価値ある情報であると認識させ、接触、開封、持ち出し、後続行動を検知・誘導するためのデセプション要素である。

一方で、攻撃者はファイル名だけでなく、タイムスタンプ、ファイルサイズ、配置場所、更新履歴などの周辺情報を確認し、対象が本物かどうかを判断する可能性がある。そのため、特にメタデータを確認する攻撃者に対しては、ファイル名などの表面的な要素だけを工夫したハニーファイルでは容易に看破される可能性がある。

8.4.2　検証仮説

『ハニーファイルは、ファイル名だけでなく、タイムスタンプ、ファイルサイズ、配置場所、フォルダ構成などを実環境に近い形で作り込むことで、攻撃者にとって本物らしく見え、真正性の判断を困難にできるのではないか。』

特に、メタデータに自然なばらつきや業務上の文脈を持たせることで、攻撃者に追加の検証ステップを強いることができ、看破までの時間を引き延ばせると考えられる。

8.4.3　検証シナリオ

本検証では、攻撃者が機密情報の窃取を目的として内部ネットワークへ接続された業務端末を侵害した状況を想定し、ハニーファイルやその周辺環境の作り込みが、攻撃者の判断や分析行動に与える影響を評価した。

攻撃の流れは以下のとおりである。

1. 業務端末からファイルサーバーへ接続：デスクトップ上にあるショートカットを利用し、接続する。
2. ファイルサーバーの探索：機密情報が保存されている可能性のある共有環境を探索する。
3. 機密情報の窃取および持ち出し：価値があると判断した情報を取得し、外部へ持ち出す。

なお、本検証では、以下の 2 パターンのおとり環境を用意した。各パターンの特徴を下表に示す。

- **パターン 1**：機械的に生成したハニーファイルを用いたおとり環境
- **パターン 2**：ハニーファイルのファイル名、更新日時、ファイルサイズ、文書間の関連性および周辺のフォルダ構成を実環境に近づけたおとり環境

表 8-7　検証 2　各パターンのサイバーデセプション環境の特徴

ハニーファイルの特徴	パターン 1（機械的に生成）	パターン 2（実環境に近い）
タイムスタンプ	全ファイルを同時刻で生成	数日～数か月にわたりランダムに分散
ファイルサイズ	中身がほぼ空に近い容量（数 KB～数十 KB 程度）	実業務文書に近い容量（数 MB～数十 MB 程度）
ファイル名	機密性を露骨に強調するキーワード（例：「極秘」「パスワード付」など）を含む	業務文書で自然に使われるキーワード（例：「財務報告」「M&A」など）を含む
ファイル・フォルダ構成	1 つの共通ディレクトリ直下へ並列にファイルを格納	部門・業務を想定した階層構造にファイルを分散して格納
文書間の関連性	文書間の関連性を考慮せず、独立したファイルとして配置	財務資料、経営計画資料、M&A 関連資料など、関連する業務文書群として配置

8.4.4　検証環境

検証環境には、業務端末、およびファイルサーバーとして機能するおとりサーバーを配置し、企業環境を模擬したネットワーク構成を構築した。検証環境の構成を下図に示す。

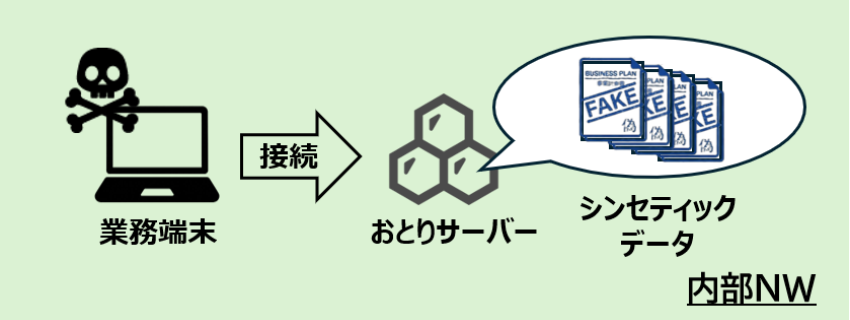


図 8-5　検証 2 の構成

また、攻撃者代理として自律型 AI エージェントを利用し、ファイル探索、価値判断、真正性確認および情報取得に至る一連の攻撃行動を実行させた。なお、本検証における AI エージェントは、人間の攻撃者が行う探索・判断・接続試行を模擬するための手段として用いたものであり、AI エージェント自体の性能評価を目的とするものではない。

8.4.5　検証手順

本検証は、以下の手順で実施した。

1. ハニーファイルやその周辺環境の作り込みが異なる 2 種類のおとり環境を用意
   - パターン 1：タイムスタンプ、ファイルサイズ、フォルダ構成に機械的生成の痕跡が残る環境
   - パターン 2：ファイル名、タイムスタンプ、ファイルサイズ、フォルダ構成、文書間の関連性を実環境に近づけた環境
2. 攻撃者代理として自律型 AI エージェントを用い、ファイルサーバーとして機能するおとりサーバー内のファイル探索を実行
3. 各パターンについて、攻撃者がどの情報を確認し、どの段階でおとり環境と判断したかを記録
4. 看破に至るまでの判断材料、分析ステップ、最終的な看破要因を比較

8.4.6　検証結果

(1)　パターン 1（機械的に生成した環境）の検証結果

パターン 1 では、タイムスタンプ、ファイルサイズおよびファイル配置に機械的生成の痕跡が残る環境を対象として検証を実施した。

その結果、攻撃者はファイル探索の初期段階において、更新日時やファイルサイズといったメタデータを確認し、人工的に生成された環境である可能性を指摘した。

特に、全ファイルの更新日時が同一であることや、PowerPoint や Word 文書としては不自然に小さいファイルサイズであること、さらに機密性を露骨に強調する文書が集中的に配置されていることが看破要因となった。

この結果から、ハニーファイルやその配置環境の作り込みが不十分な場合、基本的なメタデータ分析のみで看破される可能性があることが確認された。

(2)　パターン 2（実環境に近い環境）の検証結果

パターン 2 では、タイムスタンプを数日から数か月にわたり分散させるとともに、ファイルサイズ、フォルダ構成および文書間の関連性を実環境に近づけた環境を対象として検証を実施した。

その結果、攻撃者は更新日時やファイルサイズに明確な不自然さを発見できず、初期段階ではおとり環境であると判断しなかった。

その後、攻撃者はファイル名と更新日時の整合性、フォルダ構成の妥当性、文書間の関連性などについて追加の分析を実施した。最終的には、以下の点を根拠としておとり環境であると判断した。

- **重要文書に修正履歴や版管理の痕跡が存在しない点**：ファイルサーバー全体で見たときに、重要文書が複数存在するにもかかわらず、「v1」「rev02」などのバージョン情報や修正履歴を示す痕跡が確認されなかった。
- **関連する文書間に相互参照の痕跡が存在しない点**：財務資料、経営計画資料、M&A 関連資料など、相互に関連することが想定される文書が存在するにもかかわらず、文書間で数値や記述を参照し合う痕跡が確認できなかった。

この結果から、メタデータやフォルダ構成の整合性を高めることで、攻撃者に対してより深い分析行動を強要できる可能性が示された。また、攻撃者はファイル単体だけでなく、修正履歴や文書間の関連性を含む周辺情報も利用して環境の真偽を判断していることが確認された。

（3）　パターン 1 とパターン 2 の検証結果の比較

パターン 1 およびパターン 2 の検証結果の比較を下表に示す。

表 8-8　検証 2 パターン 1 とパターン 2 の検証結果の比較

評価項目	パターン 1（機械的に生成）	パターン 2（実環境に近い）
初期確認での反応	更新日時の完全一致や極端に小さいファイルサイズから、人工的な環境である可能性を早期に指摘	更新日時やファイルサイズは自然と判断され、初期確認のみでは看破に至らず
主な確認対象	タイムスタンプ、ファイルサイズ、ファイルの集中配置	タイムスタンプ、ファイルサイズに加え、ファイル名と日付の整合性、フォルダ構成、文書間の関連性
分析の深さ	軽量なメタデータ確認の段階で看破	メタデータ確認を通過した後、より深い文脈分析へ移行
最終的な看破要因	全ファイルの更新日時が秒単位で一致していたこと、Office 文書として不自然に小さいサイズであったこと、機密性を強調する文書名が集中していたこと	重要文書に修正履歴や版管理の痕跡が乏しいこと、関連するはずの文書間に参照関係が確認できなかったこと
検証上の示唆	作り込みが浅いハニーファイルや配置環境は、基本的なメタデータ確認で看破される。	作り込みを高めることで、攻撃者に追加の分析工程を強要できる。

パターン 1 では、更新日時やファイルサイズといったメタデータのみで看破された。一方、パターン 2 では、メタデータの整合性向上により初期段階での看破を回避できた。

しかし最終的には、文書の修正履歴や版管理の痕跡、文書間の参照関係といった業務活動の継続性を示す情報がなかったことから、サイバーデセプション環境であると判断された。

この結果は、ハニーファイル単体の作り込みだけでなく、ファイル群全体の関係性や業務プロセスを反映した環境設計が、攻撃者による分析コストのさらなる増加に寄与する可能性を示している。

8.4.7　考察

(1)　ハニーファイルによるサイバーデセプションの有効性向上に向けた観点

パターン 1 では、更新日時やファイルサイズといった基本的なメタデータのみで看破された。一方、パターン 2 では、タイムスタンプやファイルサイズの整合性を高めたことで、攻撃者は文書間の関連性や修正履歴の痕跡まで分析する必要があった。

この結果は、ハニーファイルの名称などの表面的な要素だけでなく、タイムスタンプやファイルサイズなどのメタデータ、フォルダ構造、文書間の関連性といった周辺情報を含めた環境全体の整合性が、サイバーデセプションの有効性に影響することを示唆している。特に、修正履歴や文書間の参照関係といった業務活動の継続性を示す痕跡、いわば環境の「生活感」が、サイバーデセプション環境の真偽を判断する重要な材料となる可能性が示された。

(2)　AI エージェントの判断から得られた示唆

本検証では攻撃者の代理として AI エージェントを利用した。その過程で、ハニーファイルの作り込みに関する主たる検証結果とは別に、AI エージェント特有と考えられる判断傾向も観測された。

パターン 2 における看破要因として、AI エージェントは、ファイルの修正履歴や文書間の参照関係の欠如を指摘した。しかし、これらの特徴のみを根拠としてハニーファイルだと判断することが妥当であるかについては、議論の余地がある。実際の業務環境においても、ファイルの修正履歴や文書間の参照関係が存在しないケースが存在し得るためである。

そのため、AI エージェントを利用する攻撃者に対しては、AI エージェントが重視する判断基準自体がデセプションの対象となり得る。例えば、ハニーファイル側にあえて一部の不完全さや業務上ありがちな揺らぎを持たせることで、AI エージェントの真正性判断を揺さぶる余地がある。

「3.2.3　Bell-Whaley フレームワーク」で述べたように、デセプションの本質は「虚構を見せること」だけでなく、「実体を隠すこと」にもある。検証で観測された AI エージェントの判断傾向は、この「実体を隠す」というアプローチにおける一つの可能性を示唆するものであった。

(3)　実務導入時の設計上の留意点

実務でハニーファイルを導入する際には、攻撃者にとって魅力的なファイル名を付けるだけでは不十分である。実際の業務環境に存在するファイルと整合するよう、命名規則、フォルダ構成、タイムスタンプ、ファイルサイズ、作成者情報、更新頻度などを含めて設計する必要がある。

また、ハニーファイルは単体で配置するのではなく、ブレッドクラム、ハニートークン、シンセティックデータ、ポケットリッターなどと組み合わせることで、攻撃者にとって自然な探索経路を形成できる。これにより、攻撃者の接触を検知するだけでなく、探索経路の誘導や後続行動の観測にもつなげることができる。

ただし、過度に目立つファイル名や不自然な配置は、攻撃者にサイバーデセプションであることを疑わせる要因となる。そのため、実環境に溶け込む自然さと、攻撃者が価値を感じる誘引性のバランスを取ることが重要である。

以上より、ハニーファイルの有効性は、ファイルそのものの作り込みだけでなく、メタデータやファイルの周辺環境を含めた設計品質に大きく依存すると考えられる。

8.5 検証 3　ハニークレデンシャルの配置・文脈設計と欺瞞効果の関係

本検証は、「表 6-6　プロセス 1・2 に基づく設計例」の No.9「偽認証情報による次行動の誘発」に対応するものであり、Elicit（Motivate）の観点から、偽認証情報が攻撃者の行動をどの程度誘発し、欺瞞状態を維持できるかを確認することを目的とする。

本検証では、ハニークレデンシャルを活用したサイバーデセプション技術が、攻撃者による認証情報利用および横展開行動に与える影響を評価した結果を示す。

8.5.1 検証の狙い

本検証では、ハニークレデンシャルの配置場所や周辺文脈の作り込みが、攻撃者をおとり環境へ誘導する効果に与える影響を確認する。攻撃者は、侵害端末上で認証情報を探索し、取得した認証情報を用いて他システムへの接続や横展開を試みることがある。防御側は、この行動を逆手に取り、偽の認証情報をあえて配置することで、攻撃者におとり環境への接続を試行させ、その行動を検知・観測することができる。

本書では、攻撃者の接触や利用を検知するために意図的に配置するサイバーデセプション資産を、広くハニートークンとして整理している。検証 3 で扱うハニークレデンシャルは、このハニートークンの一種であり、偽のパスワード、API キー、管理者アカウントなどを攻撃者に発見・利用させることで、認証情報窃取後の次行動を誘発し、その利用を検知・観測することを狙うものである。

一方で、ハニークレデンシャルは、攻撃者に発見されれば必ず利用されるものではない。配置場所が不自然であったり、利用先の環境と整合していなかったりすると、攻撃者にサイバーデセプションであると疑われる可能性がある。

8.5.2　検証仮説

『ハニークレデンシャルは、攻撃者が探索過程で自然に発見し得る場所に、実運用上の痕跡として違和感なく配置されているほど、攻撃者にとって有効な認証情報であるように見え、おとり環境への接続試行を誘発しやすくなるのではないか。』

特に、単にデスクトップ上に偽のパスワードファイルを置くよりも、PowerShell の実行履歴、システム変数、ショートカットなど、実運用上の設定ミスや管理作業の痕跡として見えやすい場所に配置することで、認証情報利用およびおとり環境への接続試行を誘発しやすくなると考えられる。

8.5.3　検証シナリオ

本検証では、攻撃者がドメイン一般ユーザー権限を取得した後、認証情報窃取、認証情報利用および横展開を実施する状況を想定し、認証情報の配置方法が攻撃者行動に与える影響を評価した。

攻撃の流れは以下のとおりである。

1. 認証情報窃取：認証情報が保存されたファイルや設定情報を窃取する。
2. 認証情報を利用した接続試行：発見した認証情報を利用してサーバーへの接続を試行する。
3. 横展開：接続可能なシステムを探索し、権限拡大を試みる。

なお、認証情報窃取および認証情報利用行動への影響を評価するため、以下の配置パターンを検証対象とした。

表 8-9　検証 3 におけるハニークレデンシャルの配置パターン

配置パターン	内容	期待する効果
A：平文パスワードファイル	デスクトップ上などに平文のパスワードファイルを配置する	認証情報利用の誘発
B：PowerShell 履歴	PowerShell の実行履歴に認証情報を埋め込む	実運用上の設定ミスらしさによる利用行動の誘発
C：環境変数・ショートカット	環境変数およびショートカットに認証情報を埋め込む	システム情報としての信頼性向上による利用行動の誘発

また、各配置パターンに対して攻撃者役に付与する情報を以下の 2 条件で比較した。

- 無警戒状態：おとり環境に関する情報を付与しない状態
- 警戒状態：おとり環境が存在することを事前に通知した状態

8.5.4　検証環境

検証環境には、AD サーバー、業務端末およびおとりサーバーを配置し、実環境を想定したネットワーク構成を構築した。業務端末上に配置されたハニークレデンシャルを攻撃者が発見し、その認証情報を利用しておとりサーバーへの接続を試行する構成とした。検証環境の構成を下図に示す。

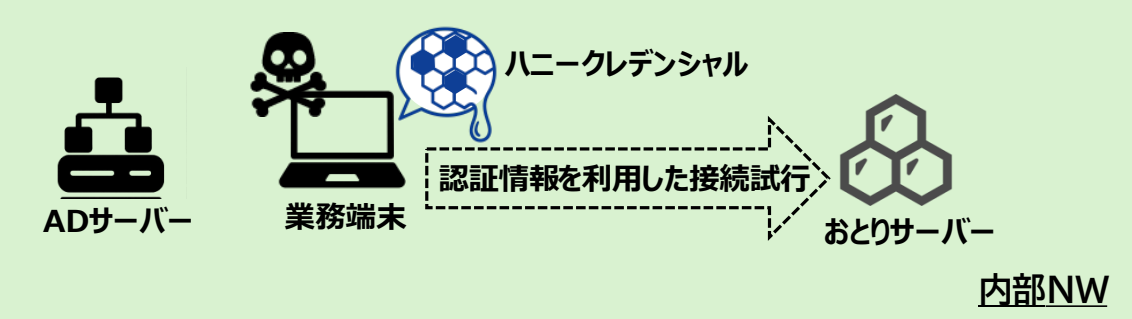


図 8-6　検証 3 の構成

また、攻撃者代理として自律型 AI エージェントを利用し、認証情報窃取から認証情報利用までの一連の攻撃行動を実行させた。なお、本検証における AI エージェントは、人間の攻撃者が行う探索・判断・接続試行を模擬するための手段として用いたものであり、AI エージェント自体の性能評価を目的とするものではない。

8.5.5　検証手順

認証情報窃取および認証情報利用行動への影響を評価するため、以下の手順で検証を実施した。

1. ハニークレデンシャルを配置した検証環境を構築する。
2. 無警戒状態および警戒状態の条件を AI エージェントに設定する。
3. 攻撃者代理として AI エージェントを実行する。
4. 認証情報窃取および認証情報利用行動を観測する。
5. おとりサーバーへの接続試行有無を確認する。
6. 追加調査行動および看破要因を分析する。
7. 条件間の結果を比較し、評価する。

8.5.6　検証結果

(1)　パターン A（平文パスワードファイル）

無警戒状態では認証情報利用およびおとりサーバーへの接続試行が確認された。

一方、警戒状態では認証情報そのものは発見されたものの、平文ファイルによる配置が不自然であると判断され、認証情報利用およびおとりサーバーへの接続試行には至らなかった。

(2)　パターン B（PowerShell 履歴）

無警戒状態および警戒状態の両条件において、認証情報利用およびおとりサーバーへの接続試行が確認された。特に警戒状態については、おとり環境の存在を認識した状態であっても認証情報が利用されたということであり、PowerShell 履歴が実運用上の設定ミスとして認識された可能性が示唆された。

しかし、認証情報利用およびおとりサーバーへの接続試行が確認された後、AD 情報との不整合（ドメイン未所属）を根拠として、おとり環境である可能性が指摘された。

(3)　パターン C（環境変数・ショートカット）

無警戒状態では認証情報利用およびおとりサーバーへの接続試行が確認された。

一方、警戒状態では認証情報発見後に対象ホストに対する追加調査が実施され、Windows サーバーを装っていたホストが Linux 系 OS の特徴を示していたこと（TTL: Time To Live）を根拠として、おとり環境である可能性が指摘された。その結果、認証情報利用およびおとりサーバーへの接続試行には至らなかった。

(4)　全体比較

無警戒状態ではすべての配置パターンで認証情報の利用および接続試行が確認された。一方、警戒状態ではパターン B のみが認証情報利用および接続試行を誘発した。また、警戒状態では、すべての配置パターンにおいて真偽確認のための追加調査行動が確認された。

表 8-10　検証 3 における配置パターン別比較

ハニークレデンシャルの配置パターン	無警戒状態	警戒状態
A：平文パスワードファイル	利用・接続	利用前に看破
B：PowerShell 履歴	利用・接続	利用後に看破
C：環境変数・ショートカット	利用・接続	利用前に看破

8.5.7　考察

(1)　配置・文脈設計と認証情報利用

本検証では、ハニークレデンシャルの配置場所や周辺文脈の作り込みが、攻撃者行動に影響を与えることが確認された。

パターン A では認証情報が発見されたものの、警戒状態では利用に至らなかった。一方、パターン B では警戒状態においても認証情報利用および接続試行が確認された。この結果から、攻撃者が実運用上の設定ミスや管理作業の痕跡として認識しやすい情報ほど、認証情報利用やおとり環境への接続試行を誘発しやすい可能性が示唆された。

(2)　看破要因の分析

警戒状態では、すべての配置パターンにおいて、おとり環境の真偽を確認するための追加調査が実施された。調査内容は配置パターンごとに異なったものの、パターン B では AD 情報との不整合（ドメイン未所属）、パターン C では OS 情報との不整合（TTL）が確認されており、看破要因は認証情報そのものではなく、利用先環境に関する情報の不整合であった。

この結果は、攻撃者が認証情報だけではなく、ネットワーク情報やシステム情報を組み合わせて対象の真偽を判断している可能性を示している。そのため、おとり環境を設計する際には、認証情報の配置のみならず、OS、ネットワークおよび AD を含む環境全体の整合性を確保することが重要であると考えられる。

(3)　追加調査行動から得られた示唆

本検証では、警戒状態において、すべての配置パターンで真偽確認のための追加調査行動が確認された。調査内容は異なるものの、攻撃者は認証情報を発見した後、その情報を直ちに利用するだけでなく、周辺環境との整合性を確認する行動を取る場合があることが確認された。

このことから、ハニークレデンシャルによる欺瞞効果は、偽認証情報そのものの作り込みだけで決まるものではない。攻撃者が認証情報を発見する文脈、認証情報の利用先、接続先ホストの OS 情報、AD 所属状況など、複数の要素が一貫していることで、欺瞞状態をより長く維持できると考えられる。

（4）　実務導入時の設計上の留意点

実務でハニークレデンシャルを導入する際には、攻撃者にとって発見しやすい場所へ目立つ形で配置するだけでは不十分である。PowerShell 履歴、設定ファイル、スクリプト、ショートカット、環境変数など、攻撃者が実際の探索過程で確認し得る場所に、自然な痕跡として配置することが重要である。

また、ハニークレデンシャルの利用を検知するためには、偽認証情報の利用先を明確にし、認証試行、アクセス先、接続元、時刻、実行されたコマンドなどを記録できる監視設計が必要となる。単に偽の認証情報を置くだけでは、攻撃者の行動を十分に観測できない。

以上より、ハニークレデンシャルは、認証情報窃取後の次行動を誘発する有効なデセプション要素となり得る。ただし、その効果を実務上有効に機能させるためには、配置文脈、利用先環境、監視・検知設計を一体で設計する必要がある。

8.6　横断的な分析・検証総括

各検証結果を横断的に分析した結果、以下の 3 点が、サイバーデセプション戦略を実運用環境へ適用する上で重要な知見として得られた。

- サイバーデセプションは攻撃者の意思決定に影響を与え、行動変容を誘発する可能性がある
- サイバーデセプションの有効性は情報配置の自然性に依存する
- サイバーデセプションの有効性を維持するためには環境全体の整合性が重要である

以下、それぞれについて考察し、既存対策との連携の重要性について触れる。

8.6.1　攻撃者行動への影響

本検証では、探索活動への影響（検証 1）、真正性判断および情報持ち出しへの影響（検証 2）、認証情報窃取後の次行動への影響（検証 3）の観点から、サイバーデセプションの効果を評価した。検証 1 では、探索活動に要する時間の増加が確認された。検証 2 では、ハニーファイルの真正性判断に対する影響および持ち出し行動が確認された。また、検証 3 では、認証情報利用およびおとりサーバーへの接続試行が確認された。

これらの結果から、サイバーデセプションは、攻撃者に対して単に検知点を設けるだけでなく、攻撃者の意思決定に影響を与え、その結果として行動変容を誘発する可能性があると考えられる。

8.6.2　自然性の重要性

検証 2 および検証 3 では、攻撃者が環境内に存在する情報や利用痕跡を意思決定の根拠として利用する様子が確認された。特に、業務文書を模したハニーファイル、PowerShell 履歴、ショートカットなど、実運用環境で発生し得る文脈を持つ情報は攻撃者行動に影響を与えた。一方で、不自然な情報配置は利用に至らない事例も確認された。

この結果から、サイバーデセプションの有効性は情報そのものだけではなく、その情報が置かれている文脈や利用痕跡を含めた自然性に依存すると考えられる。

8.6.3　環境整合性の重要性

検証 3 では、TTL 値、AD 情報の不整合がサイバーデセプションを看破する要因として確認された。また、攻撃者がどの調査手法を選択するかは防御側で完全に制御できず、異なる観点から環境情報が確認されることも確認された。

この結果から、誘導情報のみを作り込むのではなく、OS、ネットワークおよび認証基盤を含めた環境全体の整合性を確保することが、サイバーデセプションの有効性を維持する上で重要であると考えられる。

8.6.4　既存対策との連携の重要性

第 8 章では、3 つの検証からサイバーデセプションが攻撃者の認知・判断・行動に作用し、攻撃の進行を遅らせる、判断を迷わせる、おとり環境へ誘導する、または特定の攻撃経路を不成立にする可能性を確認した。ただし、実攻撃全体の防止を単独で証明したものではなく、既存の監視・遮断・インシデント対応と組み合わせることで、防御効果が高まると整理するのが適切である。

【コラム 8】 自然界の欺瞞とサイバーデセプションのこれから

本書では人間を中心にデセプションについて語ってきましたが、デセプションはなにも人類だけの発明ではありません。生物の「擬態」がその代表例と言えるでしょう。

例えば、木の枝に擬態するナナフシや羽を閉じると枯れ葉に見えるコノハチョウについてご存じの方も多いのではないでしょうか。しかし、実のところ自然界における擬態はただ隠れるだけに留まりません。

ミミックオクトパスはその名のとおり擬態者（ミミック）です。ウミヘビ、ミノカサゴ、ヒトデなど、外敵にとっての「天敵」や「有毒生物」の姿へと瞬時に形や動きを変えます。一つの姿に固執することなく柔軟に変化させ、相手を遠ざけるのです。

さらに、自然界における欺瞞は姿を変えるだけではありません。天敵から巣やヒナを守るため、コチドリという鳥は「擬傷（ぎしょう）」と呼ばれる手段を取ることがあります。羽が折れたふりをして地面を這いずり回り、自身を狙わせることで天敵を引き付けます。そして巣やヒナの安全が確保されるとさっと飛び去っていくのです。自身をより魅力的な餌に見せかけることで、本当に大事なものを守っているのです。

彼らは自然界で特定の天敵から身を守るために長い進化の過程で欺瞞の技術を身に付けてきました。しかし、サイバーの世界ではどうでしょうか。インターネットという広大な世界の中では決まった天敵は存在しません。

今回ご紹介した検証結果は、あくまで効果的な罠の一例にすぎません。インターネットでは様々な攻撃者がそれぞれの目的に沿って活動しており、相手によってサイバーデセプションの効果も変わってきます。誰に対する見せかけなのか、どんな行動を引き出したいのか、そして実際に効果があったのか。それらを一連の流れで評価し、結果から環境を柔軟に「進化」させていく必要があります。

サイバーデセプションは技術的にも体系的にもまだまだ成熟したものではありません。長い進化の過程の入り口に立っているに過ぎないのでしょう。

コラム内参考資料

神戸の野鳥観察日記 http://www2.kobe-c.ed.jp/shizen/yacho/kansatu/06011.html（参照 2026-06-09）
おきなわ図鑑 https://okinawa-zukan.com/column.php?vol=5（参照 2026-06-09）

（余白ページ）

第9章　総括

9.1　検証結果を踏まえた結論

本書ではサイバーデセプションを、ハニーポットやハニートークンなどの個別技術の導入にとどまるものではなく、攻撃者の認知・判断・行動に働きかけ、防御側に有利な状況を作り出すための能動的な防御戦略として整理してきた。すなわち、サイバーデセプションの本質は、偽の資産や情報を置くこと自体ではなく、攻撃者に何を見せ、何を信じさせ、どのような行動を取らせ、その行動をどのように検知・遅延・誘導・観測するかを設計する点にある。

「第 8 章　サイバーデセプションの検証」では、限定された環境ではあるものの、サイバーデセプションが攻撃者の認知・判断・行動に影響を与え得ることを確認した。検証 1 では、低対話型ハニーポットと応答遅延の組み合わせにより、探索活動に要する時間的コストを増大させる可能性が示された。検証 2 では、ハニーファイルのファイル名だけでなく、タイムスタンプやファイルサイズなどのメタデータ、配置場所、フォルダ構成、文書間の関連性を含めた自然性が、攻撃者の真正性判断に影響を与えることが確認された。検証 3 では、ハニークレデンシャルそのものの作り込みだけでなく、攻撃者が自然に発見する文脈や、利用先となる環境との整合性が、おとりサーバーへの接続試行を誘発する効果に大きく関わることが示された。

これらの結果は、サイバーデセプションが単に「おとりを置けば機能する」ものではないことを示している。攻撃者が接触し得る位置に配置されていること、攻撃者にとって自然に見えること、提示された情報と周辺環境に矛盾がないこと、そして接触や利用を検知・分析できる運用体制に接続されていることが不可欠である。言い換えれば、サイバーデセプションの有効性は、個別技術の性能だけではなく、目的、攻撃フェーズ、配置先、技術要素、監視・対応体制が一貫して設計されているかに依存する。

そのため、サイバーデセプションを実務に適用する際には、製品や技術を起点に検討を始めるのではなく、自組織が何を守り、どの脅威アクターを想定し、どの攻撃段階で、どのような効果を狙うのかを明確にする必要がある。本書で提案した CYDEC Design Framework（CDF）は、このような目的、前提、作戦、戦術、運用を一貫して整理するための枠組みであり、サイバーデセプションを場当たり的な技術導入ではなく、組織のセキュリティ戦略に基づく防御設計として位置づけるために有効である。

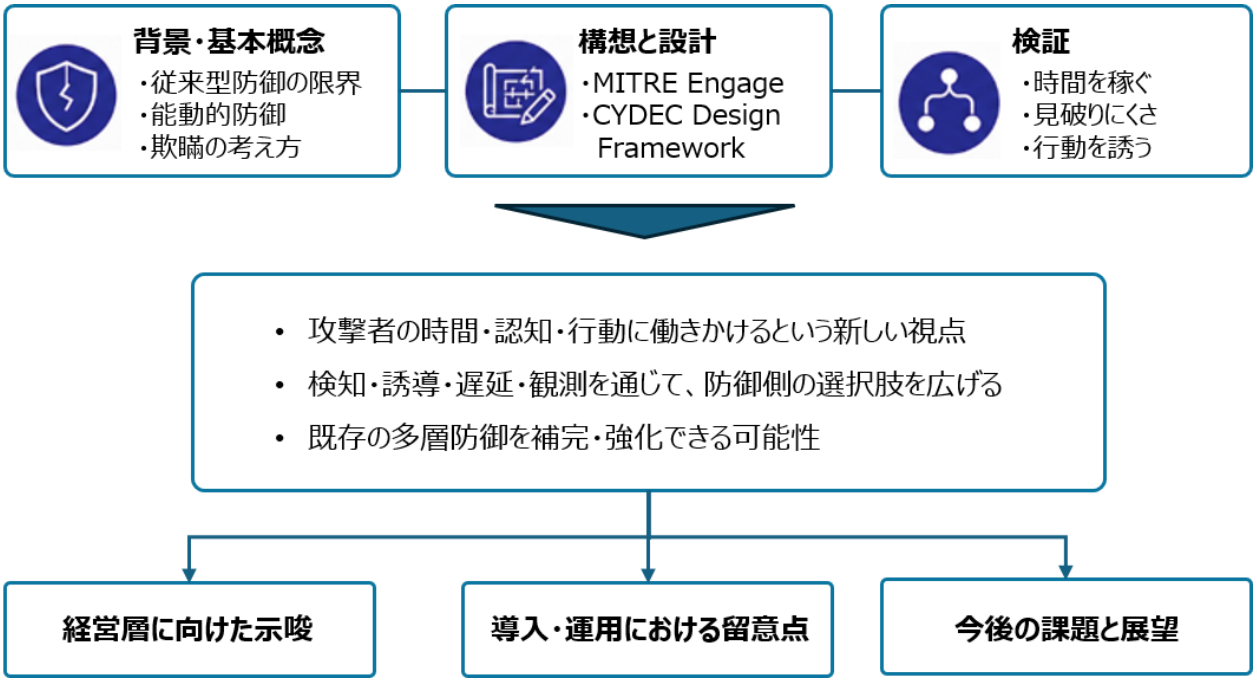


図 9-1　サイバーデセプション活用の結論

以上を踏まえると、サイバーデセプションは従来型の対策、EDR、SIEM、ゼロトラストなどを置き換えるものではなく、それらを補完し、侵入後の攻撃者に対して、防御側が観測・判断・対応の選択肢を広げるための追加的な防御レイヤーである。特に、攻撃者の探索効率を低下させ、判断を迷わせ、次行動を誘発し、その過程を観測するという点において、従来型の対策だけでは得にくい効果を期待できる。

重要なのは、サイバーデセプションは導入すれば自動的に効果を発揮するものではないという点である。経営層による目的設定とリスク許容、設計・運用部門による自然性と整合性の確保、SOC/CSIRT との監視・対応連携、組織としての継続的な評価・改善があって初めて、実効性のある防御策となる。次節以降では、これらを踏まえ、経営層に向けた示唆、導入・運用における留意点、今後の課題と展望を整理する。

9.2　経営層に向けた示唆

近年では、サイバー攻撃の高度化・巧妙化に加え、攻撃者が長期にわたり標的組織内で活動を継続する事例が見られている。また、企業・組織はファイアウォール、EDR、SIEM、ゼロトラスト、脆弱性管理、アクセス制御など、さまざまな対策を重ねてきたが、認証情報の悪用や正規ツールを用いた侵害活動などにより、侵入を完全に防ぐことはますます困難になっている。

この背景には、防御側と攻撃側の構造的な非対称性が存在する。防御側は多数のシステムや利用者、委託先を継続的に守らなければならないのに対し、攻撃者は一つの脆弱性や設定不備、認証情報を足掛かりに侵入

できる。経営層にとって重要なのは、セキュリティ対策を単なる技術導入の積み重ねとして捉えるのではなく、この非対称性をどのように緩和し、事業継続に必要な防御力をどのように確保するかという視点である。

9.2.1　侵害前提の防御へ

従来型の対策は、攻撃を侵入前に防ぐこと、または侵入後の異常を検知することを中心に発展してきた。これらの対策は現在も重要であり、サイバーデセプションはそれらを否定するものではない。しかし、攻撃の高度化や認証情報の悪用、正規ツールを用いた侵害活動を踏まえると、すべての攻撃を完全に防ぐことを前提とした防御には限界がある。

経営の視点で重視すべきなのは、「侵入されないこと」だけではなく、「侵入された場合にどれだけ早く状況を把握し、どれだけ被害拡大を抑え、事業継続への影響を小さくできるか」である。近年重視されているサイバーレジリエンスも、侵害を完全には防げないことを前提に、組織の継続的な運営能力や回復力を高める考え方に立脚している。

サイバーデセプションは、侵入後の攻撃者の活動を観測し、対応の時間や判断材料を確保するとともに、被害拡大の抑制や攻撃実態の把握を支援する。このような特徴から、サイバーデセプションは侵害前提の時代におけるサイバーレジリエンス向上を支える一手段として位置付けることができる。

9.2.2　防御思想としてのデセプション

サイバーデセプションという言葉から、ハニーポットやハニートークンなどの個別技術を想起される場合も多い。しかし、本書で一貫して述べてきたように、デセプションの本質は特定の製品やツールではなく、攻撃者の認知・判断・行動に働きかける点にある。

攻撃者は限られた時間と情報の中で、どの資産が重要か、どの経路を進むべきか、どの認証情報を利用すべきかを判断している。サイバーデセプションはこの判断過程に対して意図的に設計した情報や環境を提示し、防御側に有利な状況を作り出そうとする考え方である。すなわち、単に「罠を置く」のではなく、攻撃者に何を見せ、何を信じさせ、どのように行動させるかを設計する取り組みである。

従来型の対策は、主としてシステム、ネットワーク、データといった資産を保護対象としてきた。一方でサイバーデセプションは、攻撃者の認知・判断・行動そのものを防御側が働きかける対象として捉える。これは防御側が守る対象を拡張し、攻撃者の意思決定にも働きかけようとする防御思想である。

9.2.3　経営判断における位置づけ

本書で示してきたサイバーデセプションの特徴は、攻撃者優位になりがちなサイバー空間において、防御側がより有利な状況を作り出そうとする点にある。

前述したとおり、防御側は広範な資産と関係者を継続的に守る必要がある一方で、攻撃者は一つの脆弱性、設定不備、認証情報を足掛かりとして侵入し得る。このような構造的な非対称性は、サイバーセキュリティにおける根本的な課題の一つである。

サイバーデセプションは、この構造を根本的に逆転させるものではない。しかしながら、攻撃者に誤った判断を促し、行動を観測し、対応の時間や機会を確保することで、防御側に新たな選択肢をもたらす。これは攻撃者の行動を前提として受動的に対応するだけでなく、攻撃者の認知や判断にも働きかけることで、防御側が一定の主導権を確保しようとするアプローチである。

経営層にとって重要なのは、サイバーデセプションを単なる追加対策や個別技術として捉えるのではなく、攻撃者優位になりがちな構造を緩和し、防御側がより有利な状況を作り出すための戦略的なアプローチとして理解することである。そのうえで、導入目的、リスク許容、監視・対応体制との接続、継続的な改善に必要な投資を、経営判断の対象として位置づける必要がある。

9.3　導入・運用における留意点

本節では、本書の総括として、企業・組織がサイバーデセプションを単なるツールの追加ではなく、能動的な防御戦略として組織に定着させるための重要な留意点を 5 つの視点から提示する。

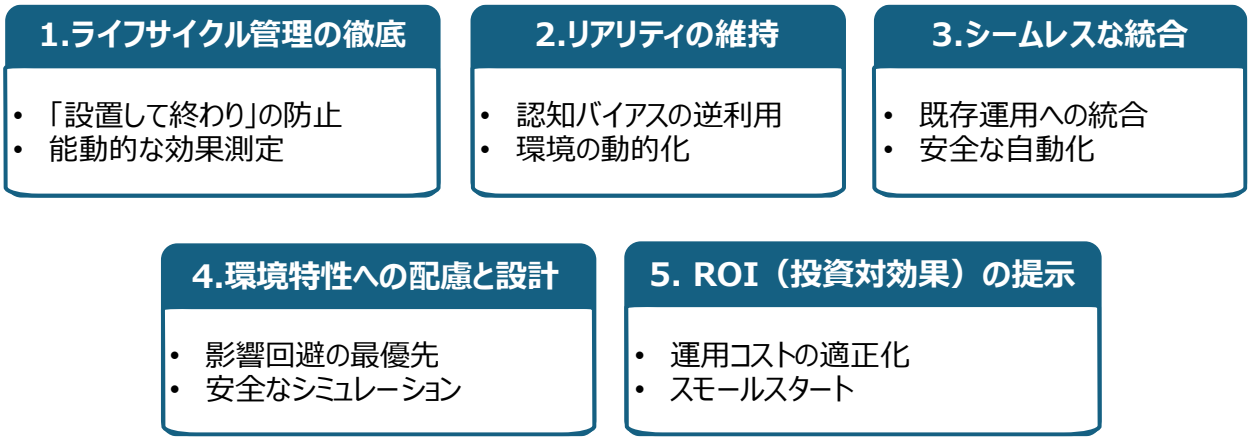


図 9-2　導入・運用における留意点

9.3.1 「設置して終わり（Fire and Forget）」を防ぐライフサイクル管理

サイバーデセプションは、一度罠を設置して終わるものではない。攻撃者の戦術は常に進化するため、「5.3 CYDEC Design Framework（CDF）の構想と設計体系」で提唱した CDF を活用し、自組織の環境変化や最新の脅威動向に合わせて前提（Who/What/by Whom）と戦術（When/Where/How）のズレを継続的に評価・改善するライフサイクル管理が不可欠である。特にサイバーデセプションは「攻撃者が来なければ効果を直接観測できない」という効果測定の難しさがある。そのため、単に攻撃を待つのではなく、定期的なペネトレーションテストやレッドチーム演習を実施し、自ら評価のタイミングを作ることが重要である。実際の攻撃をシミュレートすることで、おとりの現実味や真正性、SOC の検知・対応プロセスが正しく機能するかをプロアクティブに検証し、ROI（投資対効果）の評価とプログラムの継続的な改善を行う仕組みを確立する必要がある。

9.3.2　攻撃者の認知バイアスを突く「現実味」の維持

「4.2　攻撃者心理から見たデセプションの有用性」で述べたとおり、デセプションの本質は攻撃者の認知バイアスや判断上の弱点に働きかけることにある。デフォルト設定や安易な命名規則を用いた低品質なおとりは、熟練した攻撃者に即座に見破られてしまう。攻撃者に本物だと信じ込ませ、攻撃者の時間とリソースを消費させる（Affect）、あるいは手口を引き出す（Elicit）ためには、リアルな設定とメタデータを持たせるだけでなく、定期的なイメージの更新、ファイル構成やハニートークンのローテーションによって、環境を常に動的に保つ運用が求められる。

9.3.3　SOC/CSIRT とのシームレスな統合と「安全な自動化」

「2.1　現状のセキュリティ戦略と課題」で挙げた「アラート疲れ」を低減し、防御側の対応効率を高めるためには、サイバーデセプションからの高精度なアラートを既存の SOC 運用やインシデント対応プロセス（SIEM/SOAR/EDR など）に適切に統合する必要がある。ただし、十分な相関分析なしに自動防御（通信の遮断など）を行うと正規のビジネス通信まで誤ってブロックしてしまうリスクがあり、過度な自動化には注意が必要である。また、運用者がサイバーデセプションからのアラートを軽視することなく迅速に対応できるよう、事前にインシデント発生時の対応手順（プレイブック）や、揮発性データ・マルウェア検体を安全に保存する証拠保全の仕組みを整備しておくことが重要である。

9.3.4　運用環境の特性（OT/ICS など）への配慮と安全な設計

サイバーデセプションの配置にあたっては、対象となる環境の特性を深く理解する必要がある。「6.4.4 OT/ICS 環境における導入上の留意点」で提示したように、特に OT/ICS 環境のような可用性・安全性が最優先される領域では、物理プロセスに影響を与えない配置と、書き込み要求を安易に受け付けないシミュレーション

モードを基本とする安全な設計が前提となる。IT 環境とは異なり、「攻撃者に本物らしく見せる」こと以上に、「本番設備や制御プロセスへの予期せぬ影響を確実に避ける」という基本方針を遵守して設計しなければならない。

9.3.5　運用コストの懸念払拭と「ROI（投資対効果）」の提示

サイバーデセプションの導入にあたり、「おとりの維持管理やアラート対応で SOC の運用コスト（人的リソース）が増大するのではないか」という懸念が生じる場合がある。一方で、適切に設計されたサイバーデセプション環境から発生するアラートは、正規利用では通常触れない資産への接触を起点とするため、既存のセキュリティツールと比較して高い確度を持つ可能性がある。そのため、適切に運用できれば、不要な調査工数を抑えつつ、重要度の高い事象に SOC のリソースを集中させる効果が期待できる。

また、商用プラットフォームや OSS を活用することで、おとりの展開、更新、ローテーションを効率化できる場合がある。ただし、維持管理コストや運用負荷は、導入範囲、監視体制、既存システムとの連携状況によって異なるため、導入前に総所有コスト（TCO）を見積もることが重要である。

組織への導入にあたっては、最初から大規模な展開を狙うのではなく、まずは特定の重要資産周辺で OSS や限定的なライセンスを用いて概念実証（PoC）を行う「スモールスタート」が有効である。そのうえで、ライセンス費用、運用工数、SOC 連携に必要な作業量を把握し、早期検知によって重大インシデントを回避・抑制できる可能性を含めて、ROI（投資対効果）を評価する必要がある。経営層に対しては、単なる製品費用ではなく、被害拡大の抑制、調査工数の削減、対応時間の確保といった観点からサイバーデセプションの ROI（投資対効果）を説明することが重要である。

9.4　今後の課題と展望

本書では、サイバーデセプションの基本概念、攻撃者心理との関係、導入・運用上の留意点を整理したうえで、検証を通じてその有効性を確認した。検証結果から、サイバーデセプション資産への接触を契機とした早期検知、攻撃者の探索行動の遅延、偽情報やおとり環境への誘導、攻撃者行動の観測といった点において一定の効果が期待できることが示された。

一方でサイバーデセプションは、導入すれば一律に効果を発揮する万能な対策ではない。その効果は組織のシステム構成、攻撃者像、守るべき資産、運用体制、監視基盤、法務・コンプライアンス要件などに大きく左右される。また、本書の検証は限られた期間、環境、条件のもとで実施したものであり、AI を用いた攻撃者行動のシミュレーションを含む。そのため、実際の攻撃者や各企業の実環境において、同様の効果が常に再現されることを保証するものではない。

以上を踏まえると、サイバーデセプションの今後の課題と展望は、大きく次の 3 点に整理できる。

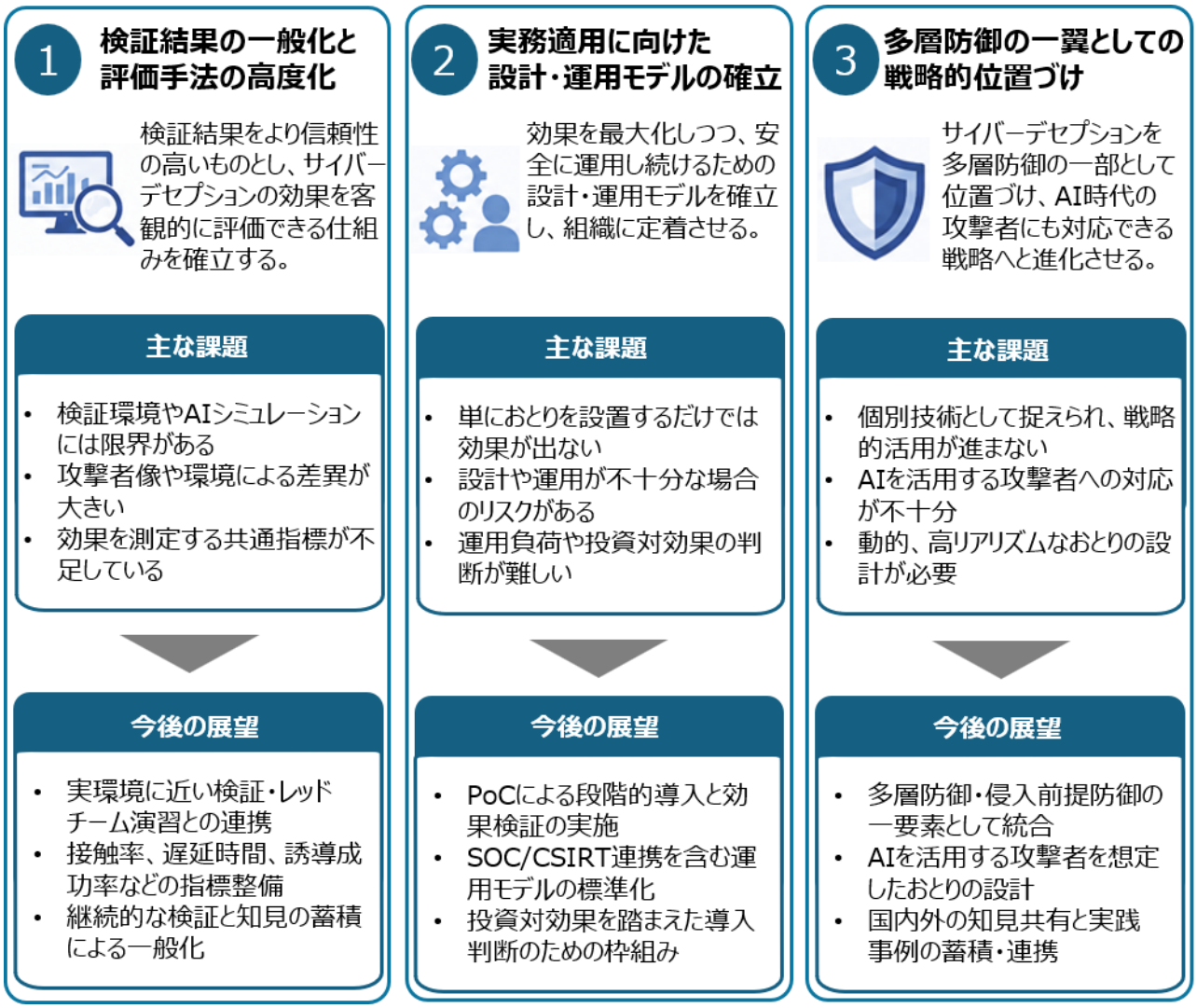


図 9-3　サイバーデセプションの今後の課題と展望

9.4.1　検証結果の一般化と評価手法の高度化

第一の課題は、検証結果をどの範囲まで一般化できるかである。本書の検証では、サイバーデセプションが攻撃者の行動に一定の影響を与え得ることを確認できたが、検証環境は現実の企業ネットワークを完全に再現したものではない。実際の環境では、ネットワーク構成、端末数、認証基盤、クラウド利用状況、OT/ICS の有無、SOC 運用の成熟度などが組織ごとに異なる。また、攻撃者もスキル、目的、慎重さ、自動化ツールの利用状況によって行動が変化する。

特に、AI を用いた検証は、攻撃者の行動パターンを再現し、仮説を検証するうえで有用である一方、実際の攻撃者が持つ経験、警戒心、偶発的な判断、組織固有の環境に対する適応行動を完全に再現できるものではない。フロンティア AI モデルを含む高度な AI エージェントは、防御側の検証手法を高度化する可能性がある一

方、その出力は使用するモデル、プロンプト、前提条件、評価方法に大きく左右される。そのため、AI による検証結果は、実攻撃者の行動そのものではなく、仮説検証を補助する一つの材料として扱う必要がある。

実際に各企業・組織において導入する際は、より実環境に近い検証環境での再現性確認が必要である。例えば、業種別、企業規模別、IT 環境と OT/ICS 環境の違い、クラウド利用の有無、攻撃フェーズごとの違いなどを踏まえた検証を積み重ねることが望まれる。また、AI によるシミュレーションだけでなく、レッドチーム演習やペネトレーションテスト、ブルーチームによる運用検証と組み合わせることで、より実践的な評価が可能となる。

あわせて、効果測定の指標を整備することも重要である。単に「検知できたか」だけではなく、接触率、検知までの時間、攻撃者の遅延時間、誘導成功率、取得したテレメトリの有用性、SOC における調査負荷の削減度合い、対応時間の短縮効果などを評価することで、サイバーデセプションの効果をより客観的に把握できる。今後は、こうした評価指標を用いて、サイバーデセプションの有効性と限界を継続的に検証していくことが求められる。

9.4.2　実務適用に向けた設計・運用モデルの確立

第二の課題は、サイバーデセプションを実務に適用するための設計・運用モデルを確立することである。サイバーデセプションは、ハニーポットやハニートークンを単に配置すれば成立するものではない。攻撃者にとって自然に見える現実味、発見されやすい戦略的配置、本番環境からの隔離、ログ取得、SOC/CSIRT などの運用体制や、SIEM、SOAR、EDR などの既存ツールとの連携、アラート発生時の対応手順などが一体となって初めて効果を発揮する。

設計や運用が不十分であれば、期待した効果が得られないだけでなく、新たなリスクを生む可能性がある。例えば、攻撃者に容易に見破られる低品質なおとりは接触機会を失わせる。隔離が不十分なおとりは、本番環境への踏み台となるおそれがある。監視基盤と連携していないおとりは、攻撃者が接触しても検知や対応につながらない。OT/ICS 環境では、現場部門との調整なしに導入することで、操業継続や安全に影響を与える可能性もある。

そのため、各企業・組織では自社の環境に応じた段階的な導入が望ましい。まずは、小規模な PoC を通じて、どの攻撃フェーズで、どの資産に対して、どのようなサイバーデセプション技術が有効かを確認する。そのうえで、既存の SOC 運用やインシデント対応手順に接続し、アラートの優先度、エスカレーション先、封じ込め条件、証拠保全方法を定義することが重要である。

また、導入にあたっては ROI（投資対効果）を各社で検討すべきである。サイバーデセプションは比較的低コストで高精度な検知点を作れる可能性がある一方、一定の設計・運用負荷を伴う。したがって、導入費用だけでなく、運用負荷、SOC の調査負荷削減、インシデント対応時間の短縮、脅威インテリジェンスの獲得、重要資産

への到達遅延といった効果を含めて評価する必要がある。今後、導入判断に使える実務的な評価モデルや業種・環境別の導入パターンが蓄積されることが期待される。

9.4.3　多層防御の一翼としての戦略的位置づけ

第三の課題は、サイバーデセプションを一部の特殊技術ではなく、企業・組織のセキュリティ戦略の一部として位置づけることである。サイバーデセプションは、ファイアウォール、EDR、SIEM、脆弱性管理、ゼロトラスト、バックアップ、インシデント対応などと置き換えられるものではない。むしろ、これらの既存対策では十分に捉えきれない攻撃者の認知・判断・行動に働きかけることで、多層防御を補完する役割を持つ。

従来型の対策は、攻撃を遮断する、検知する、被害を復旧するという観点が中心であった。これに対し、サイバーデセプションは、攻撃者に「何を見せるか」「何を信じさせるか」「どこへ誘導するか」を防御側が設計する点に特徴がある。これは、攻撃者優位になりがちなサイバー空間において、防御側が観測・判断・対応の選択肢を広げるための新しい考え方である。

さらに、今後はフロンティア AI モデルの進展により、攻撃側の能力が高まる可能性にも留意する必要がある。Anthropic は Project Glasswing[18]において、Claude Mythos Preview を重要ソフトウェアの脆弱性発見・分析に活用する取り組みを示す一方、同様の能力を攻撃者も利用しようとする可能性があるとの見方を示している。また、英国 AI Security Institute（AISI）は Claude Mythos Preview の評価において、CTF 課題や複数段階のサイバー攻撃シミュレーションで性能向上が見られたと説明している。さらに、明示的な指示とネットワークアクセスが与えられた制御された評価環境では、脆弱なネットワークに対する多段階攻撃や脆弱性の発見・悪用を自律的に実行できたとしている。

したがって、フロンティア AI モデルを過度に万能視することは避けるべきであるが、攻撃者がこれらを利用した場合に脆弱性探索、攻撃手順の自動化、環境観察、おとりの見破りといった行為を支援する可能性がある点は今後の脅威環境として考慮すべきである。AISI も、サイバー能力は防御を強化する可能性がある一方、悪用され得るデュアルユースの能力であると整理している。

このような環境変化を踏まえると、サイバーデセプションは現在の攻撃者像だけを前提に設計するのではなく、AI を活用する攻撃者も視野に入れて更新していく必要がある。固定的なおとりや単純なハニートークンだけでは、将来的に見破られる可能性が高まる。今後は、環境に応じて変化するおとり、攻撃者の行動に応じた動的な誘導、AI による探索にも耐え得る現実味、そしてカウンターデセプションを前提とした設計が重要となる。

ここでいうカウンターデセプションとは、攻撃者が防御側のおとりや偽情報を見破り、回避・無効化しようとする行為を指す。攻撃者がデセプション管理コンソール、監視基盤、スイッチ、ルータ、SDN、NDR などを侵害または分析した場合、おとりの配置、誘導ルール、アラート条件などを把握され、防御側のデセプションが回避・無効化され

るおそれがある。そのため、今後のサイバーデセプション設計では、おとり資産の自然さだけでなく、それらを管理・監視する基盤の保護も重要となる。管理コンソールのアクセス制御、多要素認証、管理ネットワークの分離、管理 API の保護、ログ保全、通信経路の暗号化、スイッチやルータの設定変更監視などを組み合わせ、防御側の可視性と対応能力を維持する必要がある。

今後のセキュリティ戦略においては、サイバーデセプションを、侵入前提の防御、サイバーレジリエンス、能動的防御、多層防御の一部として組み込むことが重要である。特に、侵入後の横展開や権限昇格、情報窃取の兆候を早期に捉える領域では、サイバーデセプションが既存対策を補完する有力な選択肢となり得る。また、重要インフラや OT/ICS 環境においても、物理プロセスへの影響を避けながら、偵察や異常な相互作用を検知する手段として活用可能性がある。

一方、日本国内では、サイバーデセプションに関する公開情報や実務知見はまだ十分とはいえない。今後は研究機関、セキュリティベンダー、ユーザー企業が連携し、技術検証、運用事例、評価指標、失敗事例、法務・倫理面の考慮点を共有していくことが望まれる。特に、単なる製品導入事例にとどまらず、どのような目的で導入し、どのような設計を行い、どのような効果や限界があったのかを体系的に整理することが重要である。

サイバーデセプションは万能な対策ではない。しかし、攻撃者の認知と行動に働きかけるという点で、これまでのサイバー防御戦略に新たな視点を加える考え方である。今後、各企業・組織が自らのリスクや運用体制に照らして導入可能性と ROI（投資対効果）を検討し、多層防御の一翼としてサイバーデセプションを戦略的に取り入れていくことが期待される。そのためにも、本分野における検証、実装、運用、評価に関する研究と実務知見の蓄積が不可欠である。

あとがき

本書は ICSCoE 中核人材育成プログラムにおける卒業プロジェクトとして、サイバーデセプションの可能性を探究し、その成果をホワイトペーパーとして取りまとめたものです。

プロジェクトの出発点にあったのは、日本のサイバーセキュリティを少しでも前に進めたいという思いでした。高度化する攻撃に対し、防御側が常に後追いになるのではなく、攻撃者の認知・判断・行動に働きかけ、防御側が主導権を取り戻す。そのような能動的防御の考え方として、サイバーデセプションには大きな可能性があると考えました。

もちろん、サイバーデセプションは万能の解決策ではありません。導入には、目的の明確化、攻撃者像の整理、適切な配置、安全な隔離、監視・運用体制、法務・倫理面の検討が必要です。しかし、適切に設計されれば、攻撃者を早期に検知し、迷わせ、遅延させ、誘導し、その行動を観測することで、従来型の対策を補完する有効な選択肢になり得ます。

本書では、その考え方を実務に活かせるよう、デセプションの基本概念、攻撃者心理、CDF による設計体系、代表的な技術要素、検証結果、導入・運用上の留意点を整理しました。まだ十分に成熟した領域とはいえないからこそ、本書が完成形の答えではなく、今後の議論、検証、実装、改善の出発点となることを願っています。

本プロジェクトは、多くの方々のご支援なしには実現できませんでした。ご多忙の中、終始温かく、時に厳しくご指導いただいた産業サイバーセキュリティセンター中核人材育成プログラムの越島一郎先生、橋本芳宏先生、門林雄基先生、佐々木弘志様に、心より御礼申し上げます。また、本プロジェクトでは、各メンバーが調査、検証、執筆、レビュー、議論に力を尽くし、それぞれの専門性、経験、問題意識を持ち寄りました。その積み重ねがあったからこそ、本書は単なる技術解説にとどまらず、戦略・運用・技術・心理を横断する成果物として形になりました。

さらに、国立研究開発法人情報通信研究機構（NICT）主席研究員　伊東寛様、同 CYNEX 研究開発運用室室長　安田真悟様、横浜国立大学教授　吉岡克成様、フォーティネットジャパン合同会社　伊藤剛様にはヒアリングや助言、検証環境の構築、資料確認など、さまざまな形で本プロジェクトにご協力いただきました。また、亀井様には、マジシャンの視点から人間心理の考え方についてご指導いただきました。

本書が、企業、行政機関、重要インフラ、教育機関、そしてサイバーセキュリティに関わる多くの方々にとって、サイバーデセプションを理解し、検討し、実装するための一助となれば幸いです。そして、サイバーデセプションという考え方が日本に少しずつ根づき、組織のサイバーレジリエンス向上、ひいては日本全体のサイバーセキュリティ強化につながることを、心より願っています。

最後に、本プロジェクトに関わってくださったすべての皆さまに、改めて感謝申し上げます。

サイバーデセプションプロジェクトメンバー一同

プロジェクトメンバー

<table>
<tr><td colspan="2">【プロジェクトリーダー】
佐藤 将大</td></tr>
<tr><td colspan="2">【サブリーダー】
佐保 里美、桒原 敏</td></tr>
<tr><td>調査・WP チーム</td><td>検証・効果測定チーム</td></tr>
<tr><td>【チーム長】
大原 俊平
【チーム長補佐】
赤木 勇綺
【メンバー】
吉兼 刃矢、三浦 且暉、松﨑 結香
小永井 大祐、出塚 貴裕、熊野 靖也</td><td>【チーム長】
西村 悠
【チーム長補佐】
塩田 晃平
【メンバー】
堤 咲希、星 隼人、笠原 俊亮
西元 快、榛葉 光、椙田 宜久、高橋 慎治</td></tr>
</table>

Appendix

（A）用語集

本用語集では、本書に登場する主要な用語を整理する。一般的な定義に加え、本書における使い方を明確にすることで、読者の理解を助けることを目的とする。

#	用語	説明
1	受動的防御 （リアクティブな防御）	ファイアウォール、IDS/IPS、EDR などに代表される、攻撃の遮断、検知、事後対応を中心とする防御アプローチ。本書では、能動的防御と対比する概念として用いる。
2	サイバーレジリエンス	サイバー攻撃や障害に直面しても、重要な機能を維持し、被害を局所化し、迅速に回復する能力。サイバーデセプションは、早期検知や攻撃者の遅延を通じてレジリエンス向上に寄与する。
3	多層防御	単一の対策に依存せず、境界防御、端末防御、監視、認証、復旧などを複数組み合わせる防御アプローチ。サイバーデセプションは多層防御を補完する要素として位置づけられる。
4	デコイ	攻撃者に本物の資産であるかのように見せるおとりの総称。サーバー、端末、データベース、認証情報、文書など、さまざまな形態を取り得る。ハニーポットより広い概念。
5	ハニーポット	攻撃者に侵入可能なシステムやサービスに見せかけ、接続や操作を検知・観測するためのおとり環境。低対話型から高対話型まで複数の形態がある。
6	低対話型ハニーポット	実際の OS やアプリケーション全体は提供せず、基本的な通信応答やサービス応答を模倣する軽量なおとり。安全性と運用容易性に優れるが、取得できる情報は限定的である。
7	高対話型ハニーポット	実際の OS やサービスに近い環境を提供し、攻撃者に深い操作を行わせるおとり。詳細な TTPs やマルウェア挙動を観測できる一方、隔離・監視が不十分だと踏み台化リスクがある。
8	ハニーネット	複数のハニーポットやネットワーク機器を組み合わせ、企業ネットワーク全体のように見せるおとり環境。横展開や攻撃プロセス全体を観測する用途に適する。
9	ハニートークン	通常利用されない偽のデジタル情報を配置し、それにアクセス・使用・持ち出しが行われた時点で異常を検知する仕組み。偽の文書、DB レコード、API キー、認証情報、URL、クラウドキーなどが含まれる。

Appendix

#	用語	説明
10	ハニークレデンシャル（カナリアクレデンシャル）	ハニートークンのうち、ID、パスワード、API キー、アクセスキー、シークレットキー、SSH キーなどの認証情報に見せかけたもの。攻撃者による認証情報窃取、権限昇格、横展開を検知・誘導する目的で用いる。特に、使用された瞬間に通知・警告を発することを重視する場合、カナリアクレデンシャルと呼ばれることがある。
11	ルアー/ブレッドクラム	攻撃者をおとりや偽資産へ誘導するために意図的に配置する手掛かりや痕跡情報を指す。偽の設定ファイル、接続情報、ショートカット、ログ、文書名、認証情報の断片などが該当する。攻撃者に「価値がある情報を見つけた」と思わせ、探索先や横展開先を防御側が管理する環境へ誘導するために用いる。
12	シンセティックデータ	本物の機密情報に見えるが、実際には業務上の価値を持たない人工的なデータ。攻撃者の関心を引き、アクセスや持ち出しを検知するために用いる。
13	テレメトリ	システムやおとり環境から取得されるログ、通信、操作履歴、認証試行、セッション情報などの観測データ。攻撃者の行動分析やインシデント対応に用いる。
14	高精度アラート	正規利用が想定されないおとりやハニートークンへの接触を契機とする、誤検知が少ないアラート。誤検知が多いと SOC の調査負荷が高まり、真の脅威を見落とす原因となる。
15	SOC	セキュリティ監視、アラート分析などを担う組織・機能。デセプションで得られたアラートやテレメトリの主な活用先となる。
16	CSIRT	インシデント対応、関係者調整、復旧支援、再発防止などを担う組織・機能。デセプションで得られた情報は CSIRT の判断材料にもなる。
17	MITRE ATT&CK	実際の攻撃で観測された攻撃者の TTPs を体系化したナレッジベース。本書では攻撃段階や技術を整理するために参照する。
18	限定合理性	人間が限られた情報、時間、認知能力の中で判断するため、必ずしも最適な意思決定を行えないという考え方。攻撃者もこの制約を受ける。
19	認知バイアス	人間の判断に生じる体系的な偏り。攻撃者の判断にも影響し、デセプション設計に活用できる。
20	Living off the Land（LotL）	攻撃者が、OS や業務環境に標準搭載された正規ツール、管理機能、スクリプトなどを悪用して攻撃活動を行う手法を指す。PowerShell、WMI、PsExec、RDP などが悪用されることがある。正規の管理操作と見分けにくいため検知が難しく、サイバーデセプションでは偽の認証情報やおとり環境を用いて、探索や横展開の兆候を把握する。

#	用語	説明
21	AiTM 攻撃 （Adversary-in-the-Middle）攻撃	攻撃者が利用者と正規サービスの間に割り込み、認証処理を中継することで、認証情報やセッション Cookie などを窃取する攻撃手法を指す。多要素認証を突破しなくても、窃取したセッションを悪用して正規利用者になりすます場合があるため、認証後の不審な行動や内部探索の検知が重要となる。

（B）サイバーデセプションで用いる技術要素例

「6.3　詳細設計：技術要素ごとの設計観点」では、詳細設計時に参考となる該当要件、攻撃者の判断傾向・心理、導入効果、構成パターン、設計・配置のポイントを記載した。ここでは、各技術要素について、技術概要、仕組み、留意点について述べる。

表 A　代表的な技術要素（再掲）

分類	技術要素	概要
おとり環境 （接触・観測の場）	ハニーポット	単体のおとりシステム・サービスとして攻撃者の接触や操作を観測する環境
	ハニーネット	複数のおとり資産を組み合わせ、ネットワーク規模で攻撃者の探索・横展開を観測する環境
情報・痕跡 （認知に作用する要素）	ハニートークン	本来使われない偽情報・偽認証情報などを配置し、接触や利用を検知する仕組み
	ブレッドクラム	攻撃者を誘導するための手がかり・導線
	シンセティックデータ・ポケットリッター	価値あると思わせる偽情報 攻撃者を信頼させる背景情報
ネットワーク制御 （進行を遅延・誘導する要素）	タービット	攻撃者の活動遅延
	トラフィックリダイレクト	攻撃者の通信を誘導

B.1 ハニーポット（Honeypot）

技術概要

ハニーポットとは、攻撃者に対して侵入可能なシステムやサービスに見せかけたおとり環境を用意し、アクセスや操作を検知・観測するデセプション技術である。通常の業務では利用されない資産として配置されるため、アクセスが発生した場合は不審な挙動や不正行為の可能性が高く、検知対象として有効に機能する。

ハニーポットは、攻撃者との対話レベルに応じて以下のように分類される。

- **低対話型（Low-interaction）：**
  実際の OS やアプリケーション全体は提供せず、基本的な通信応答やサービス応答を模倣する軽量なおとり。安全性と運用容易性に優れるが、取得できる情報は限定的である。
- **高対話型（High-interaction）：**
  実際の OS やサービスに近い環境を提供し、攻撃者に深い操作を行わせるおとり。詳細な TTPs やマルウェア挙動を観測できる一方、実環境に近い操作を許すため、隔離・監視が不十分な場合には踏み台化リスクが特に高くなる。

仕組み

ハニーポットは実在するサーバー、端末、アプリケーション、またはサービスの挙動を模倣することで、攻撃者からの接続や操作を受け付け、その挙動を記録・分析する。取得可能な情報として、接続元情報（IP アドレス、通信元など）、認証試行情報、実行コマンド、ファイル操作、通信内容、マルウェアの挙動、横展開行動が挙げられる。低対話型では、主にスキャンや接続試行を観測し、高対話型では侵入後の詳細な行動や攻撃手順（TTPs）の把握が可能となる。

留意点

- 高対話型ハニーポットは、侵害された場合に踏み台として悪用されるリスクがあるため、通信制御や監視が必要である。
- 明らかに不自然な構成や情報の不整合が存在すると、攻撃者にハニーポットであると識別される可能性がある。
- 配置場所が不適切な場合、攻撃者から発見されず十分な効果が得られない可能性がある。
- 運用態勢が不十分な場合、取得したログや観測結果を十分に活用できない。
- 大量配置する場合は、管理対象の増加や維持コストに注意が必要である。

B.2 ハニーネット（Honeynet）

技術概要

ハニーネットとは、単一のハニーポットを配置するだけでなく、複数の偽サーバー、偽ルータ、偽スイッチなどを相互に接続し、組織のネットワーク環境の一部または全体を模倣した「おとりネットワーク」である。単一のハニーポットが「点」として攻撃者との接点を作るのに対し、ハニーネットは「面」として攻撃者の探索・横展開の過程を観測する環境として機能する。内部に侵入した攻撃者に対し、本物のネットワークを探索していると認識させることで、隔離された安全な環境下において、侵入経路、使用ツール、横展開のプロセスを観測し、攻撃の進行を遅延・妨害することが可能となる。

Appendix

仕組み

- **通信の再指向：**
  境界防御や本番環境に配置された誘導用の痕跡情報に攻撃者が接触すると、ネットワークスイッチやSDN、プロキシ、DNS 制御などを用いて攻撃者に経路が変更されたことを悟らせず、通信を偽のネットワークへ誘導する。
- **攻撃者の封じ込め：**
  偽のネットワークの境界に配置された専用の制御装置により、侵入した攻撃者の通信を監視する。特に偽の環境内部から外部ネットワークへの通信は、第三者への攻撃の踏み台にされることを防ぐため、自動的に遮断するか、防御側が管理する偽の外部環境へ転送する。
- **データの記録と保全：**
  攻撃者が気付きにくい仮想化基盤やネットワークのパケットレベルにおいて、キー入力、実行コマンド、投入されたマルウェアの挙動を取得し、改ざんされにくい形式で保存する。

留意点

- **外部への踏み台化の防止：**
  偽のネットワークへ誘導された攻撃者が、当該環境を外部の第三者組織を攻撃するための拠点として悪用するリスクは排除しなければならない。外部への送信通信を厳格に制限するか、偽のインターネット環境を用いて通信を偽のネットワーク内で自己完結させる出口対策の設計が必須となる。
- **本番データの流用回避：**
  偽のネットワークの現実味を高めるためであっても、実際の顧客情報や従業員の機密データ、稼働中システムの暗号鍵などをダミーファイルに流用してはならない。データ漏洩時の法的責任やプライバシー侵害のリスクを回避するため、配置する情報は人工的に生成されたシンセティックデータを使用する。
- **運用の継続性と形骸化の防止：**
  偽のネットワークは有益な脅威情報を得られる反面、静的な構成のまま放置すると攻撃者にデセプション環境であると見破られ、単に回避されるだけの環境に陳腐化する。インフラ構築の自動化コードなどを活用し、定期的にネットワーク構造や誘導用の痕跡情報の配置を自動的に変更する運用体制が必要である。

B.3 ハニートークン（Honeytoken）

技術概要

ハニートークンは、本来の業務では使用されない偽のデータ、識別子、認証情報、API キーなどを意図的に配置し、それらへの接触や利用を検知するデセプション技術要素である。実際のビジネス上の用途はないため、ハニートークンに対するアクセス、変更、ダウンロードなどは、通常業務では想定されない不審な活動として検知対象となる。

ハニートークンの代表例としては、ハニークレデンシャル（またはカナリアクレデンシャル）が挙げられる。これは、攻撃者を誘引するために用意された偽のユーザー名、パスワード、API キーなどの認証情報であり、比較的軽量に導入できるデセプション技術である。

仕組み

ハニートークンは、システムへの特定のアクションをトリガーとして、ログ管理システムや SIEM へアラートを送信する仕組みで動作する。

- **ドキュメント型：**
  Word や Excel などのファイル内に、外部リソースを呼び出すリンクや識別子を埋め込む。ファイルが開かれると、指定された URL や Webhook などに通信が発生し、アクセス元の IP アドレス、ユーザーエージェント、時刻などの情報が記録・通知される。
- **データベース型：**
  データベース内に偽のレコードを用意し、そのレコードに対するクエリを監査ログやデータベースのトリガー機能を利用して検知する。
- **クレデンシャル型：**
  偽のユーザー名、パスワード、API キーなどを配置し、それらが認証や接続に使用された際に、Windows イベントログ、Active Directory の認証ログ、クラウド監査ログなどを用いて検知する。

留意点

- 偽の文書やデータベースのレコードを作成する際、誤って実在する個人情報や本番環境の秘密情報を含めないよう注意する。
- ハニークレデンシャルが認証に使用された場合でも、本番リソースへのアクセスを許可してはならない。接続を許可する場合は、隔離されたおとり環境に限定し、ログ取得、通信制御、証拠保全を前提として設計する必要がある。

- ドキュメント型の場合、組織のセキュリティ設定（Office の外部画像読み込みブロックやファイアウォール設定など）によっては通信が発生せず、検知できない可能性があるため、環境に応じた設計が必要である。
- 運用負荷の観点から、どこにどのハニートークンを配置したかを正確に管理し、インベントリ化しておく必要がある。

B.4 ブレッドクラム（Breadcrumbs）

技術概要

ブレッドクラムとは、攻撃者の注意を引き、おとり環境（ハニーポットなど）へ自発的に向かわせるために、実環境に意図的に配置されるヒントや痕跡を指すデセプション技術要素である。[1]

ブレッドクラム自体が直接アラートを発報したり、攻撃者を封じ込めたりするものではない。一方で、攻撃者の探索活動を防御側が意図した方向へ誘導する道標として機能する。多くの場合、攻撃者の特定の反応を引き出すルアー（誘引）技術や、誘導先となるハニーポット、ハニーネットなどと組み合わせて運用される。

仕組み

ブレッドクラムは、主にエンドポイント層（端末やサーバーなど）において、攻撃者がネットワーク内を探索する際に自然に発見するよう、ファイルシステム、ログ、レジストリ、ブラウザ履歴などに配置される。攻撃者がこれらの痕跡情報を発見し、記載されている偽の接続先情報（IP アドレスや URL）、認証情報、設定情報などを利用しようとすると、防御側が用意した安全な隔離環境であるおとりやハニーポットに誘導される仕組みとなる。

留意点

- 多用しすぎたり、パターン化されたブレッドクラムを配置したりすると、熟練した攻撃者に見破られるリスクがある。鮮度と現実味を保つため、定期的なローテーションと更新が必要である。
- ブレッドクラム単体では完結せず、誘導先となるおとり環境（ハニーポットなど）を準備し、連携させる必要がある。実体のない場所へ誘導すると、不自然さを生む。
- 現実味を持たせる目的であっても、本物の機密データ、実際の個人情報、本物のパスワードなどを含めてはならない。深刻な法的・プライバシーリスクを引き起こすため厳禁である。

B.5 シンセティックデータ（Synthetic Data）/ポケットリッター（Pocket Litter）

技術概要

シンセティックデータおよびポケットリッターとは、実際の機密情報やユーザーの日常的な痕跡の代わりに配置される、人工的に生成された本物らしいデータセットやファイル群を指すデセプション技術要素である。[1] これらは生成

手法や配置形態が似ているが、攻撃者の認知に作用させる目的によって、大きく以下の 2 つに分類して運用される。

- **シンセティックデータ：**
  攻撃者が探し求めている「価値ある標的データそのもの」を人工的に生成し、本物の機密データに見せかける技術である。攻撃者に偽データを分析・持ち出しさせることで、攻撃者の時間やリソースを消費させたり、偽情報によって判断を誤らせたりすることを主な目的とする。なお、シンセティックデータへの不正アクセスを検知したい場合は、データ内にハニートークンなどの検知用トリガーを埋め込み、併用することが有効である。
- **ポケットリッター：**
  偽のブラウザ履歴、一時ファイル、デスクトップメモなどを用いて、「このシステムは実際に人間が利用している」と攻撃者に認識させるための背景情報を指す。主な目的は、おとり環境やハニーポットに生活感・業務感を付与し、攻撃者の疑念を低減させることである。

仕組み

シンセティックデータおよびポケットリッターは、主にデータ層（データベース、ファイルサーバーなど）やエンドポイント層（端末の OS 内部、アプリケーションキャッシュなど）に配置される。 実データと同じフォーマット、識別子、タイムスタンプ、メタデータを用いて人工的に生成され、攻撃者が手動またはツールで環境を調査した際に、ごく自然な形で発見されるように設計される。

留意点

- **ブレッドクラムとの目的を混同しない：**
  ポケットリッターは、環境の信頼性を高めるための背景情報であり、特定の行動を直接誘導することを主目的とするものではない。そのため、過度に目立たせたり、攻撃者を明示的に誘導するような名称を付けたりすると、不自然さを生む可能性がある。
- **本物データを混入させない：**
  リアリティを高める目的であっても、本物の機密情報、実在する個人情報、本物の認証情報、従業員の私的データなどを含めてはならない。法的リスクやプライバシー侵害、コンプライアンス違反を引き起こすためである。

B.6 タービット（Tarpit）

技術概要

タービット（Tarpit、またはスティッキー・ハニーポット）とは、攻撃者のネットワークスキャンや接続試行に対して、意図的に極端に遅い応答を返したり、接続状態を無意味に維持し続けたりすることで、攻撃者の進行を遅延させるデセプション技術要素である[1]。主に、攻撃者の自動化ツールやマルウェアの動作を鈍化させ、時間とリソースを消費させることを目的とする。

仕組み

タービットは、主にネットワーク層（未使用の IP アドレス空間など）やソフトウェア層において、通信や特定の不正なリクエストを監視し、応答を返す際にプロトコルレベルでの遅延を発生させる。

例えば、TCP 接続を確立したままデータを送信しない、TCP ウィンドウサイズを極端に小さく設定する、またはタイムアウト直前に少量のダミーデータを送り続けるといった手法を用いる。これにより、送信元に「通信は継続中である」と認識させ、セッションを維持したまま攻撃者側の処理を停滞させる。過去には、ワーム対策として用いられた LaBrea などが代表例として挙げられる。

留意点

- **正規ツールの巻き込みリスク：**
  設定や配置を誤ると、社内の正規の脆弱性スキャナーや IT 資産管理ツールの動作まで遅延・阻害してしまう恐れがある。そのため、あらかじめ正規ツールの IP アドレスを除外リストに登録しておくなどの管理が必要である。
- **実環境上での適用の難易度：**
  本番システム上で異常通信のみを遅延させるアプローチは、誤検知によって正規ユーザーの業務通信を阻害するリスクがあるため、極めて慎重なチューニングと設計が求められる。
- **他技術との連携：**
  タービット単体では、攻撃者の侵入を完全に防いだり、詳細な脅威インテリジェンス（TTPs）を深く収集したりすることは難しい。そのため、高対話型ハニーポットなどの他のデセプション技術へ攻撃者を誘導する、または組み合わせて運用することが望ましい。

B.7 トラフィックリダイレクト（Traffic Redirection）

技術概要

トラフィックリダイレクトとは、攻撃者やマルウェアからのネットワーク通信を捕捉し、本来の目的地（実資産や攻撃者の C2 サーバーなど）とは異なる、防御側が管理するおとり環境や監視環境へ意図的に誘導するデセプション技術要素である。単に通信を遮断するのではなく、通信を監視用サーバーへ誘導する「シンクホール」や、実システムへ接続できたと誤認させる「ハニーポットへの再指向」といった方式を用いることで、攻撃通信を安全に観測・隔離しつつ、誘導（Direct）と情報収集（Collect）を行うことを主な目的とする。

仕組み

トラフィックリダイレクトは、主にネットワーク層（ルータ、SDN、スイッチなどによる経路制御）やソフトウェア層（DNS サーバー、プロキシなど）において実装される。異常なアクセス試行や既知の悪性ドメイン宛ての通信を捕捉し、宛先 IP アドレスや名前解決結果などを制御することで、防御側が管理する別の環境へ誘導する。

留意点

- **ブラックホールによるテレメトリの喪失：**
  通信をブラックホール化して単にパケットを破棄すると、攻撃者の目的や手法に関する情報も失われる。脅威分析を重視する場合は、情報を記録できるシンクホールやハニーポットへの再指向を優先することが望ましい。
- **正規通信への影響：**
  ネットワークルーティングや DNS 制御を誤ると、正規の業務通信へ影響を及ぼす可能性がある。そのため、再指向の対象とするトラフィック、悪性ドメインリスト、例外条件などを正確に管理する必要がある。
- **誘導先環境の安全性確保：**
  再指向先となるおとり環境や監視環境が踏み台として悪用されないよう、外部通信制御、ログ取得、証拠保全、復旧手順を事前に設計しておく必要がある。

（c）参考文献

[1] QURESHI, Bilal. Cyber Deception: Fundamental Guide. Independently published, 2026.

[2] National Institute of Standards and Technology. "Developing Cyber-Resilient Systems: A Systems Security Engineering Approach". NIST Special Publication 800-160 Volume 2 Revision 1. National Institute of Standards and Technology, 2021. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-160v2r1.pdf, （参照 2026-05-12）.

[3] Fortune Business Insights. "Deception Technology Market Size, Share & Industry Analysis, and Regional Forecast, 2026-2034". Fortune Business Insights. https://www.fortunebusinessinsights.com/deception-technology-market-102220, （参照 2026-05-29）.

[4] Mordor Intelligence. "Deception Technology Market Size & Share Analysis- Growth Trends and Forecasts (2026-2031)". Mordor Intelligence. https://www.mordorintelligence.com/industry-reports/deception-technology-market, （参照 2026-05-29）.

[5] Grand View Research. "Deception Technology Market (2024 - 2030)". Grand View Research. https://www.grandviewresearch.com/industry-analysis/deception-technology-market-report, （参照 2026-05-29）.

[6] BELL, J. Bowyer; WHALEY, Barton. Cheating and Deception. Routledge, 2017.

[7] Joint Chiefs of Staff. "Joint Publication 3-13.4 Military Deception". https://jfsc.ndu.edu/portals/72/documents/jc2ios/additional_reading/1c3-jp_3-13-4_mildec.pdf, （参照 2026-06-08）.

[8] ミトニック, ケビン; サイモン, ウィリアム. 欺術―史上最強のハッカーが明かす禁断の技法. ソフトバンクパブリッシング, 2003.

[9] JAJODIA, Sushil; SUBRAHMANIAN, V. S.; SWARUP, Vipin; WANG, Cliff, eds. Cybe Deception: Building the Scientific Foundation. Springer International Publishing Switzerland, 2016.

[10] ZHU, Quanyan; LU, Zhuo; YU, Paul L.; WANG, Cliff, eds. Foundations of Cyber Deception: Modeling, Analysis, Design, Human Factors, and Their Convergence. Springer Nature Switzerland AG, 2026.

[11] MITRE. "MITRE Engage". MITRE Engage. https://engage.mitre.org, （参照 2026-02-03）.

[12] MITRE. "A Practical Guide to Adversary Engagement". MITRE Engage. https://engage.mitre.org/wp-content/uploads/2022/04/EngageHandbook-v1.0.pdf, （参照 2026-03-01）.

[13] 独立行政法人情報処理推進機構. DX 白書 2023. 独立行政法人情報処理推進機構, 2023.

[14] 経済産業省. DX レポート 2.1（DX レポート 2 追補版）. 経済産業省, 2021.

[15] MITRE. "MITRE Engage Matrix". MITRE Engage. https://engage.mitre.org/matrix/, （参照 2026-06-12）.

[16] MITRE. "ATT&CK Matrix for Enterprise". MITRE ATT&CK. https://attack.mitre.org/versions/v19/, （参照 2026-06-12）.

[17] National Institute of Standards and Technology. "Guide to Operational Technology (OT) Security". NIST Special Publication 800-82 Revision 3. National Institute of Standards and Technology, 2023. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r3.pdf, （参照 2026-06-11）.

[18] Anthropic. "Project Glasswing". Anthropic. https://www.anthropic.com/glasswing, （参照 2026-05-19）.