博士論文 2024 年度（令和 6 年度）

インシデントレスポンス迅速化に向けた
組織間連携支援のための
分散型セキュリティオペレーション基盤の実証

慶應義塾大学大学院政策・メディア研究科

西嶋　克哉

博士論文要旨 - 2024 年度（令和 6 年度）

インシデントレスポンス迅速化に向けた組織間連携支援のための分散型セキュリティオペレーション基盤の実証

近年，サイバー攻撃の増加と巧妙化が進み，単独組織での対策には限界が生じている．高度な技術力を持つ攻撃者による脅威が拡大し，防御側は迅速な検知・対応を求められている．しかし，多くの組織では，攻撃への対処に必要な能力，情報，人材が不足しており，効果的な防御を実現できていない．このような課題に対処するためには，組織間での情報共有が不可欠である．政府機関や専門組織が主導する情報共有の枠組みが既に存在し，これらは参加組織間での脅威情報の迅速な共有を促進する役割を果たしている．

一方で，現行の仕組みは正確性を重視し，一定の信頼性を確保しているが，網羅性や迅速性の観点で課題がある．情報共有において，機密情報漏えいの懸念やコストの問題からすべての情報が共有されるわけではなく，特に IOC（Indicator of Compromise）に偏重し，攻撃者の戦術や対応策などの知見が十分に共有されていない．また，人手を介する共有や定例会合中心の枠組みにより，最新の攻撃情報が即時に伝わらず，迅速な対応が難しい．

本研究では，これら既存の情報共有の枠組みを踏まえ，それらを補完する情報共有モデルである分散型セキュリティオペレーション基盤（分散 SOC）を提案する．分散 SOC では，各組織の SOC が緊密に連携し，従来の枠組みを超えた情報共有と協調的防御を実現することをめざしている．分散 SOC の特徴は，「多様な生データの，機微機密情報を保護した能動的な活用」と「システム化によるリアルタイム性の向上」を組み合わせることで，より高度な脅威検知・分析を可能にする点にある．

本研究では，分散 SOC の有効性を，以下の三つの実証で示し，いずれも組織間の密な情報共有がセキュリティ向上に寄与したことが確認された．

1 つ目の実証として，複数組織間でのダークネット観測データを分析し，情報共有の効果を評価した．学術機関と電力会社が観測したデータを比較し，組織内外でのデータ解析結果を評価した実験結果から，データ共有がより多くの疑わしいホストを検出し，攻撃ターゲットが特定の組織に集中する傾向があることが明らかになった．この結果から，複数組織間でデータを共有することにより，スキャン活動や攻撃の観測範囲が広がり，セキュリティ対応の強化が期待できることが示された．

分散 SOC の 2 つ目の実証として，悪性 Web サイトへの接続防止を目的に分散型異常検知 AED（Autonomous Evolution of Defense）を提案し，評価を行った．従来の方法では未知の Web サイトの判定が難しいが，接続傾向の乖離度に基づき不審度を算出し，接続可否を判断する手法を提案．また，他組織の接続ログを活用し誤検知を削減し，AUC（Area Under the Curve）の向上を確認した．さらに，ユーザ評価実験により，制御処理の 93.6%が 3 秒以内に完了し，実運用への適用可能性を示した．

分散 SOC の 3 つ目の実証として，複数組織間でのインシデント対応記録共有の有用性とシステム化の課題を明らかにした．大規模言語モデル（LLM）を活用して情報整理を自動化し，情報共有の効率化を図った．実証結果として，インシデント対応記録が他組織にとって有用であり，LLM を用いた情報整理の有効性と限界を明確にした．

さらに，本論文では提案したシステムを検証するため行われた 3 つの実証実験を元に，本提案システムを分散 SOC のアーキテクチャとして総括した．インフラ部分を統合しつつ，異なる分析を柔軟に適用できるアーキテクチャを構築することで，新たな分析要件が発生した際に効率的に対応できる柔軟なシステム設計とした．

加えて，インシデント対応記録の共有手法についても検討を行った．本手法では，「攻撃を受けた組織が類似するほど，より有用なフィードバックが得られる」という仮定に基づき，秘密計算を用いた動的アクセス制御を実現するものである．これにより，共有相手や状況に応じた動的なアクセス制御が可能となる．

本研究は，セキュリティ情報共有の有効性を定量的に示し，提案モデルの実現可能性を技術的および運用的に検証した．これにより，情報共有の活性化とサイバーセキュリティ対策の向上に貢献し，政府のサイバー安全保障強化方針に沿って官民連携によるインシデント対応の向上に寄与することが期待される．分散型 SOC を活用することで，重要インフラ事業者間の情報共有の分断を解消し，迅速かつ協調的な対応が可能となることが期待できる．

キーワード:
1. 情報共有, 2. インシデント対応効率化, 3. 異常検知

慶應義塾大学大学院 政策・メディア研究科
西嶋　克哉

指導教員

主査

中村　修　（慶應義塾大学）

副査

植原　啓介（慶應義塾大学）
楠本　博之（慶應義塾大学）
村井　純　（慶應義塾大学）

Abstract of Doctor's Thesis - Academic Year 2024

A Proof of a Decentralized Security Operation Platform for Interorganizational Collaboration to Accelerate Incident Response

In recent years, the increase and sophistication of cyberattacks have highlighted the limitations of individual organizations in countering them effectively. The threats posed by highly skilled attackers are expanding, demanding rapid detection and response from defenders. However, many organizations lack the necessary capabilities, information, and personnel to implement effective defenses. To address these challenges, information sharing between organizations is essential. Frameworks for information sharing, led by government agencies and specialized organizations, already exist and promote the rapid exchange of threat information between participants.

Despite prioritizing accuracy and reliability, existing frameworks face challenges in comprehensiveness and speed. Not all information is shared due to concerns about confidentiality and costs, with an overemphasis on Indicators of Compromise (IOCs), while insights into attackers' tactics and countermeasures remain insufficient. Additionally, manual information sharing and regular meeting-based frameworks delay the transmission of the latest attack data, hindering swift responses.

This research proposes a decentralized security operation infrastructure (decentralized SOC) to complement existing frameworks. A decentralized SOC enables tight collaboration between the SOCs of different organizations, facilitating information sharing and cooperative defense beyond traditional frameworks. The key features of a decentralized SOC include "actively utilizing diverse raw data while safeguarding sensitive information" and "enhancing real-time capabilities through systematization," enabling advanced threat detection and analysis.

The effectiveness of the decentralized SOC is demonstrated through three empirical experiments, all confirming that tight information sharing between organizations contributes to improved security.

The first experiment analyzed darknet observation data shared between organizations, revealing that data sharing increased the detection of suspicious hosts and expanded the scope of attack observations, thereby strengthening security responses.

The second experiment proposed and evaluated a decentralized anomaly detection system (AED) to prevent connections to malicious websites. By utilizing connection trends of other organizations,, the system effectively identified suspicious connections,

reducing false positives through shared connection logs and improving Area Under the Curve (AUC). User evaluation experiments showed 93.6% of control processing was completed within three seconds, indicating practical applicability.

As the third demonstration of the distributed SOC, we clarified the usefulness of sharing incident response records between multiple organizations and the issues involved in systematizing this. We automated information organization using a large-scale language model (LLM) to improve the efficiency of information sharing. The results of the demonstration showed that incident response records are useful for other organizations, and clarified the effectiveness and limitations of information organization using LLM.

This dissertation summarizes the decentralized SOC architecture, validated through these experiments, and discusses flexible system design to efficiently address new analytical requirements. Additionally, it proposes a method for sharing incident response records based on dynamic access control using secure computation, ensuring tailored access depending on the recipient and context.

This research quantitatively demonstrates the effectiveness of security information sharing and validates the feasibility of the proposed model, contributing to the enhancement of cybersecurity and aligning with the government's cybersecurity strengthening policy. By utilizing decentralized SOCs, information sharing between critical infrastructure operators can be improved, enabling rapid, coordinated responses.

Keywords:
1. Information sharing, 2. Incident response optimization, 3. Anomaly detection

Keio University Graduate School of Media and Governance
Katsuya Nishijima

Advisory Group

supervisor

Osamu Nakamura (Keio University)

co-advisor

Keisuke Uehara (Keio University)
Hiroyuki Kusumoto (Keio University)
Jun Murai (Keio University)

目次

図目次

表目次

第1章　序論

1.1　研究背景

近年，コンピュータやインターネットへの社会的依存がますます高まってきている．これに伴い，コンピュータウイルスなどをはじめとするコンピュータ・ネットワークへの攻撃は社会に対してより深刻な影響を及ぼすようになってきている．以前の攻撃の主たる目的が攻撃者の技術力の誇示や愉快犯であったのに対し，現在は金銭や個人情報の奪取など実益を目的としたものが増えている．CyberEdge Group の「Cyberthreat Defense Report」によれば，2023 年には 84.7%の組織が少なくとも年 1 回サイバー攻撃が成功しており，特にランサムウェアの被害は過去最高を記録し，72.7%の組織が被害を受けている[1]．

過去の同レポートと比較すると，2021 年は 86.2%，2020 年は 80.7%，2019 年は 78.0%の組織への攻撃が 1 年に 1 回以上成功しており，長期的に見ると増加傾向にあることがわかる．さらに，2014 年の報告ではわずか 62.0%であったことから，この約 10 年でサイバー攻撃の脅威が大幅に拡大していることが明らかである．このような状況は，サイバー攻撃が組織にとって継続的かつ深刻な脅威であり，今後も十分な対策が求められることを示している．特に Advanced Persistent Threat (APT) 攻撃[2] は，秘密裏に，そして執拗に長期間攻撃を続ける点で従来の脅威とは異なり，マルウェアの侵入を検知あるいは防止することは困難である．「SANS 2023 Incident Response Survey」によれば，組織がインシデントを検知するまでの平均時間は依然として長く，数日～数か月かかっているという回答が 20%程度あり[3]，迅速な対応が困難であることが示唆されている．

このように増加・巧妙化する攻撃に対し，単独組織による対策は限界を迎えつつある．その背景には，能力，情報，人材の 3 つが不足していると考えられる．

まず，サイバーセキュリティ対策に必要な能力が単独組織には不足している．攻撃の巧妙化に対応するためには，専門的な技術とリソースを備えた体制が求められるが，組織内でそのような能力を確保するのは難しいという現実がある．

次に，攻撃手法の巧妙化に伴い，サイバー攻撃に関する情報を迅速に収集し，共有することがますます重要になっている．しかし，単独組織では攻撃の全容や新たな脅威に関する情報をリアルタイムで得ることが難しい．情報の不足は，攻撃に対して迅速かつ適切に対応することを困難にし，被害の拡大を招く要因となっている．

最後に，サイバーセキュリティ分野における専門知識を持つ人材が依然として不足していることが，単独組織における対応力の限界をさらに深刻化させている．2024 年の調査では，54%の組織は，サイバーセキュリティ人材の採用に苦労していると答えており[4]，専門的な知識を持つ人材が不足していることが明らかになっている．この人材不足により，攻撃が発生した際に適切な対応を取れる体制を維持することが難しくなり，迅速な対処が遅

れることが多い.

これらの要因が複合的に影響し，単独組織ではサイバー攻撃に対する十分な対策を講じることが困難であることが浮き彫りになっている．能力，情報，人材の不足は，単独組織が直面する課題であり，その限界を超えるためにはこれらの要素に対する戦略的なアプローチが求められている.

1.2　情報共有の必要性

サイバー攻撃の増加・巧妙化に伴い，単独の組織でこれらの脅威に対処することはますます困難になっている．経済産業省の報告書では，サイバー攻撃が高度化する中，単独組織による攻撃の全容解明は困難であり，攻撃の全容の把握や被害の拡大を防止する観点から，情報共有の重要性が指摘されている[5]．

また，情報セキュリティ総合科学の論文では，インターネットの構造上，防御側は常に守勢に立たされており，これを補うための手段として，インシデント情報の共有と活用が重要であると指摘されている[6]．

これらの指摘を踏まえ，サイバー攻撃に対抗するためには，組織間での情報共有が不可欠であることが明らかである.

1.2.1　情報共有がもたらす利点

情報共有は以下の点で重要な役割を果たす.

- **早期警戒と防御の強化**

情報共有活動の参加組織は，他組織が受けた攻撃情報が共有されることで，同様の攻撃を未然に防ぐことが可能となる[7]．攻撃者は，攻撃コストを削減するために，同様の手法を複数のターゲットに対して使用する傾向がある．これは，一度確立した手法を再利用することで，新たな攻撃手法の開発や調整にかかる時間やリソースを最小限に抑えられるためである．従って，情報共有による早期警戒が重要である.

- **インシデント対応の迅速化**

被害組織は他の組織の対応事例を参照することで，インシデント対応と調査に必要な情報のうち不足している情報を入手することができる[7]．これにより，被害の拡大を抑え，復旧までの時間短縮が期待できる.

- **脅威動向の正確な把握**

セキュリティ専門組織は，自組織の情報だけでは不足している脅威動向の正確な把握が可能となる[7]

- **攻撃者の行動抑制**

攻撃手法やIoC（Indicator of Compromise）を広範囲に共有することで，攻撃者の活動範囲を狭めることができる[8]．IoCとは，マルウェアのハッシュ値，不審なIPアドレス，悪意のあるドメイン名など，過去の攻撃の痕跡を示す情報のことである．攻撃がすぐに検知・阻止される環境を作ることで，攻撃者にとってのコストを増大させ，攻撃の継続を困難にする.

これらの理由から，情報共有はサイバーセキュリティ対策において不可欠な要素となっている．

1.2.2　どのような情報を共有するのか

サイバーセキュリティにおける情報共有は，脅威の検知や防御策の強化に資するものであり，内閣サイバーセキュリティセンターのガイドラインによると主に以下が挙げられている[7]．

- **脆弱性関連情報等**

悪用された脆弱性の有無やその詳細について．

- **マルウェア**

現場で見つかったマルウェアに関する情報（※マルウェア 検体そのものは含まない場合もある）．

- **通信先**

不正アクセスの通信元やマルウェアの通信先など．

- **その他 TTP 情報（攻撃の手口）**

上記以外の攻撃者が用いた攻撃手法に関する情報．

1.3　現状の情報共有の取組

1.3.1　情報共有の形態

サイバー攻撃被害に関する情報共有の形態は，主に以下の三つに分類される[7]．

- **n 対 n 型（被害組織間の直接的な情報共有）**

被害組織同士が直接情報を共有する形態である．この方法では，各組織が自らの被害情報や攻撃手法に関する知見を互いに提供し合う．しかし，自社の被害情報を公開することへの抵抗感や，情報漏洩のリスクがあるため，実施には慎重な判断が求められる．

- **ハブ・スポーク型（専門機関を介した情報共有）**

JPCERT/CC などの専門機関がハブ（中心）となり，被害組織から提供された情報を集約・分析し，関連する組織や関係者に適切に共有する形態である．この方法では，被害組織の匿名性が保たれやすく，情報共有のハードルを下げる効果がある．ただし，ハブとなる組織への信頼性や情報管理体制の確認が重要である．

- **間接／代理型（第三者を通じた情報共有）**

被害組織が直接情報を共有するのではなく，セキュリティベンダーやコンサルタントなどの第三者を介して情報を提供・共有する形態である．この方法では，被害組織の負担を軽減しつつ，専門的な知見を活用して情報の精度や有用性を高めることが可能となる．一方で，第三者の選定や情報の正確な伝達には注意が必要である．

1.3.2　情報共有の取組

情報共有の重要性が認識される中，国内でもさまざまな取り組みが進められている．

表 1.1 に国内の代表的な情報共有の取組を示す．運営組織は政府機関や民間団体であり，官民連携のもとでサイバー脅威情報やインシデント情報を共有している．これらの取組で

は，定期的な会議の開催，緊急時の迅速な対応といった方法が採用されており，参加組織のセキュリティ対策強化に貢献している．また，業界ごとの専門的な情報共有ネットワークも整備されており，特定分野に特化した脅威情報の交換が行われている．

表 1.1　情報共有の代表的な取り組み

#	名称	形式	運営組織	概要
1	サイバーセキュリティ協議会	ハブ・スポーク型	NISC	官民の多様な主体が連携し，脅威情報等を迅速に共有・分析する組織．構成員間でサイバー攻撃の手口や対策情報を共有する．メールや電話を通じて情報提供や相談が可能である[9]．
2	CISTA（Collective Intelligence Station for Trusted Advocates）	ハブ・スポーク型	JPCERT/CC	JPCERT/CC が運営するポータルサイトで，脅威情報や分析・対策情報をタイムリーに提供・共有することを目的としている．主に重要インフラ事業者や組織内 CSIRT などの特定利用者が対象[10]．
3	J-CSIP（Initiative for Cyber Security Information sharing Partnership of Japan）	ハブ・スポーク型	IPA（情報処理推進機構）	IPA が情報ハブとして，参加組織間でサイバー攻撃情報を迅速に共有し，高度な対策を促進する取り組み．参加組織は，検知した攻撃情報を IPA に提供し，IPA は情報を分析・匿名化した上で，他の参加組織と共有する[11]．
4	ISAC（Information Sharing and Analysis Centers）	n 対 n 型	各業界の ISAC	特定の業界や分野の情報を収集，分析，共有する非営利組織．参加組織は脅威情報や攻撃事例を提供し，ISAC はそれらを分析・評価した上で，他の参加組織と共有する．共有はメーリングリストやウェブ掲示板，定期的な報告会等[12]．
5	NCA（Japan CERT Coordination Center）	n 対 n 型	日本シーサート協議会（NCA）	CSIRT 間の緊密な連携を図り，課題解決に貢献するための組織．脆弱性対策情報やインシデント情報を活用するためのフレームワークの検討やセキュリティレポートをまとめている[13]．

1.4　現状の情報共有の取組における課題

前述の通り，情報共有の仕組みは整いつつあるもののその活用は進んでいないのが実態である．例えば，休止と再開を繰り返す EMOTET による事例では，多くの組織が同様の攻撃の被害にあっている[14]．

そこで，組織間での情報活用をさらに促進するために，1.3 節で述べた現行の情報共有の取り組みについて，既存の仕組みの分析や関係者へのヒアリングを通じて課題を整理した．

第一に，情報の網羅性の不足の課題がある．情報共有のハブ組織や情報提供組織を介して行われるが，その過程で情報の選別が行われるため，すべての関連情報が共有されるわけではない．これは，機微機密情報の漏えいを懸念し，情報を提供する側が出し渋ることや，安全に大量の生データを共有するコストによるものと考えられる．特に，現在の情報共有の主な内容は IOC（Indicator of Compromise）に偏重しており，攻撃者の戦術や手法，対応策などオペレータの知見が十分に共有されていないという課題が考えられる．

第二に，リアルタイム性の欠如が挙げられる．現行の情報共有は多くの場合，ハブ組織や人手を介することで実施されるため，共有に時間がかかる傾向がある．特に，定例会合を中心とした情報共有が主流であるため，最新の攻撃情報が即時に共有されるとは限らない．この遅延は，迅速な対応を求められるサイバー攻撃の特性と必ずしも合致していない．

ただし，これらの課題が存在する一方で，既存の情報共有の枠組みが機能していないわけではない．むしろ，現行の仕組みは情報の正確性を重視しており，一定の信頼性を確保している．一方で，正確性を重視するあまり，情報の網羅性や共有の迅速性に課題が生じているのが現状であり，課題を克服することによる，より効果的な情報共有の実現が求められる．

1.5　分散型セキュリティオペレーション基盤（分散 SOC）の提案

現行の情報共有の取組には，情報の網羅性の不足，リアルタイム性の欠如の課題が存在する．本研究では，これらの課題を解決するために，分散型セキュリティオペレーション基盤（分散 SOC）を提案する．分散 SOC の核心は，各組織の SOC が緊密に連携し，従来の枠組みを超えた情報共有と協調的防御を実現することである．従来の SOC が個別最適化されがちであったのに対し，分散 SOC は「多様な生データの機微機密情報を保護しつつ能動的に活用できる自動化基盤の構築」により，高度な脅威検知・分析を可能にする．

- **多様な生データの機微機密情報を保護した能動的な活用**

従来，組織間では主にサマリ情報（例：インシデントレポート，脆弱性情報）が共有されてきたが，分散 SOC ではより粒度の細かい生データ（例：ネットワーク観測データ，接続ログ，インシデント対応記録等）を活用する．また，機微機密情報を保護する仕組みや，情報利用者が，情報提供者の持つ情報を能動的に分析することを可能とする．これにより，情報の網羅性が向上し，攻撃の前兆検知や相関分析が強化される．具体的には，各組織が独自に収集したデータを相互参照することで，単独の組織では発見できなかった攻撃パターンを明らかにできる．

・　**システム化によるリアルタイム性の向上**

既存の情報共有の多くは手動プロセスに依存し，リアルタイム性に欠けるという課題がある．本研究では，データの自動共有・解析を可能にするシステムを設計・実装し，リアルタイムな情報連携を実現する．これにより，人的負担を軽減しつつ，サイバー攻撃の即応性を向上させる．

1.6　本研究の貢献

本研究の貢献は以下の 3 つである．

・　**既存の情報共有の課題整理と課題解決方法の提案**

既存の情報共有の取組の改善点として，「情報の網羅性の不足」，「リアルタイム性の欠如」を挙げ，これらを改善する方法として新たな分散型セキュリティオペレーションモデル（分散 SOC）を提案した．分散 SOC の核心は，各組織の SOC が緊密に連携し，従来の枠組みを超えた情報共有と協調的防御を実現することである．

・　**3 つの実証実験による情報共有の定量評価**

分散 SOC の考え方の基，組織間で共有されてこなかった，あるいはその有用性が未検証だった 3 種類の情報（ダークネット観測データ，Web 接続ログ，インシデント対応記録）を共有し活用する方法を提案し，共有の効果を定量的に評価することで，提案手法の有効性を示した．

・　**実証実験を基にしたシステムアーキテクチャの設計**

実証実験を実施する中で洗い出した要件を基に，分散 SOC アーキテクチャを設計した．様々なパターンの実証の中で，共通する部分を分散 SOC インフラとして設計し，今後新たな分析を実行する際に拡張しやすいようにした．

1.7　本論文の構成

論文の構成を図 1.1 に示す．本論文の構成は以下のとおりである．

第 2 章では，分散 SOC の情報共有モデルについて述べる．第 3 章～第 5 章では，分散 SOC の 3 つの実証について述べる．第 6 章では，3 つの実証実験から洗い出した要件を基に設計した分散 SOC のアーキテクチャについて述べる．次に第 7 章では分散 SOC アーキテクチャの機能の一つである，信頼度や連携により期待できる自組織の利益に応じた動的アクセス制御の手法について述べる．最後に，第 8 章では，本研究で得られた結論と今後の展望を述べる．

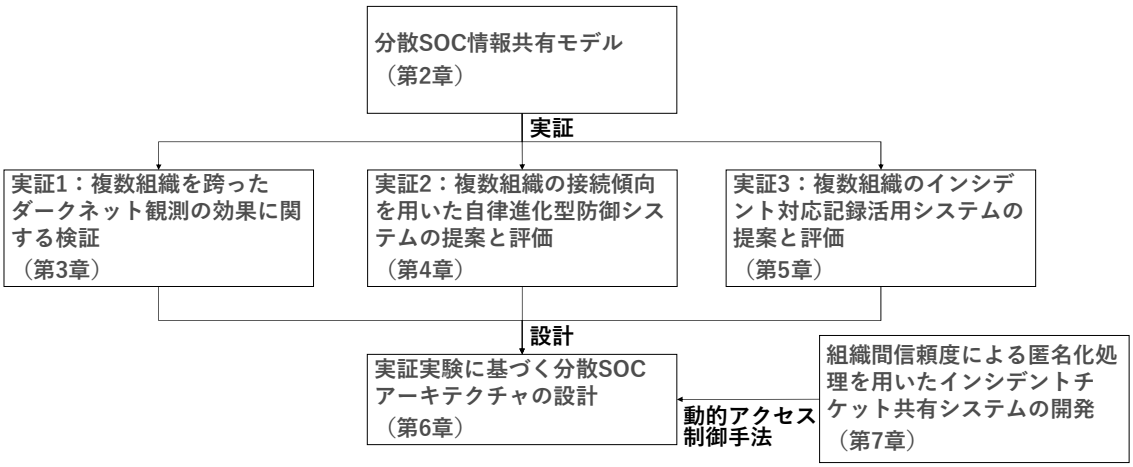
分散SOC情報共有モデル
（第2章）
実証
実証1：複数組織を跨った
ダークネット観測の効果に関
する検証
（第3章）
実証2：複数組織の接続傾向
を用いた自律進化型防御シス
テムの提案と評価
（第4章）
実証3：複数組織のインシデ
ント対応記録活用システムの
提案と評価
（第5章）
設計
実証実験に基づく分散SOC
アーキテクチャの設計
（第6章）
動的アクセス
制御手法
組織間信頼度による匿名化処
理を用いたインシデントチ
ケット共有システムの開発
（第7章）

図 1.1　本論文の構成

第2章　分散 SOC 情報共有モデル

2.1　導入

1.4 節に記載の通り，共有するデータによっては，機微機密情報を含む生データの開示リスクや，安全に大量の生データを共有するコストにより，情報共有が阻害されてしまう．

ここで，セキュリティ情報の共有では，必ずしも生データ自体を必要としているわけではないことに着目する．例えば，自組織と同じ攻撃を受信した組織を発見することや，コミュニティ内で同じ攻撃を受信した他の組織数を調査する場合には，攻撃に関するメールやログといった生データ自体は必要ではなく，共通の情報を持っているかどうかという情報のみが必要とされる．

そこで本章では，生データの開示を伴わない分析を支援するシステムを提案する．本システムでは，データ自体は共有せずに，データを分析する機能を配信・実行し，データ分析のために必要十分な情報のみを共有する．これにより，機密データの開示や大量データの共有を抑えることができ，情報共有を促進することが期待できる．

本章の構成は以下の通りである．まず 2.2 節で提案システムの要件定義について説明する．2.3 節では提案手法の情報共有モデルについて述べ 2.4 節で実証実験の概要について述べ，2.5 節でまとめる．

2.2　要件定義

本節では，組織間のセキュリティ情報共有の現状において，情報共有を阻害している要因について説明し，当該要因を解決するセキュリティ情報共有のユースケースを示す．また，それらのユースケースを実現するために提案システムが満たすべき主な要件について整理する．

2.2.1　組織間のセキュリティ情報共有の現状

前述のとおり，複数組織が協力して防御を行うためには情報共有が不可欠であるが，機微な生データの共有は困難である．これは，主に以下の 2 つの要因から成る．

- **要因 1：プライバシ・機密を含むデータの開示リスク**
- **要因 2：安全に大量データを共有するための高いコスト**

要因 1 の具体例としては，顧客との契約，General Data Protection Regulation (GDPR) といった法の違反や，共有相手組織からの情報漏洩があげられる．また，要因 2 は具体的には，データ共有にかかるネットワーク帯域やストレージ，暗号化処理などのコストが，データ共有のメリットを上回ることである．

これらの要因により，情報共有の取組等で同一コミュニティを形成していても，機微な生データ共有へのハードルは依然として高い状態である．

2.2.2　セキュリティ情報共有のユースケース

セキュリティ情報共有のユースケースイメージを図 2.1 に示す．

図 2.1 に示す通り，セキュリティ情報共有のユースケースには，以下の 3 つが考えられ

る．

- **ユースケース 0：共有データの中身を利用した詳細な分析**
- **ユースケース 1：自組織と同じ攻撃 X を受信した組織の発見**
- **ユースケース 2：分析組織による，コミュニティ内で攻撃 X を受信した組織数の調査**

ユースケース 0 では，データの中身を必要とするため，データに機微情報が含まれている場合には，実現が困難となる．一方でセキュリティ情報共有では，ユースケースの 1 や 2 のように，データ自体の共有が必須ではないユースケースも考えられる．これらの場合，データ自体を共有せずに分析を行うことにより，2.2.1 項の要因 2 が解決する．また，連携組織が情報分析に必要十分な情報の開示を許容し，必要十分情報を超える開示が無いことを技術で保証することができれば，要因 1 が解決される．

従って，ユースケース 1 や 2 のような分析を行うことを可能とするプラットフォームシステムがあれば，情報共有が促進されることが期待できる．

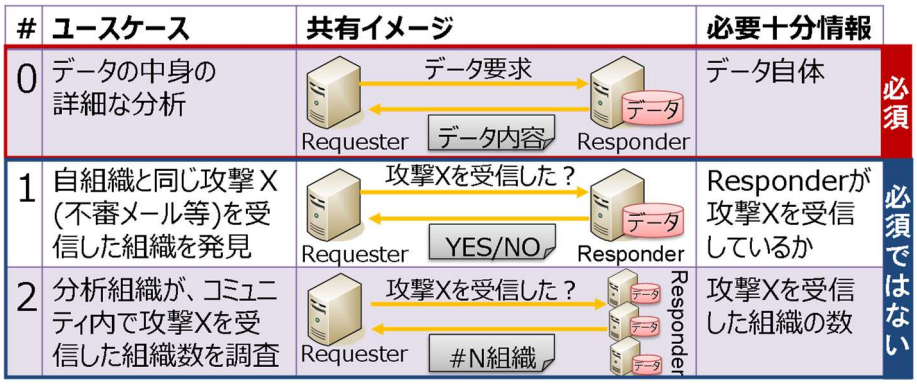

図 2.1　セキュリティ情報共有のユースケース

2.2.3　実現すべき主要要件

2.2.2 項のユースケース 1，2 を実現するために提案システムが満たすべき主要な要件を表 2.1 に示す．

表 2.1 に示す通り，提案システムの主要要件には，以下の 2 つが考えられる．

- **開示要件：「データ自体の非共有，必要十分を超える情報の非開示」を保証**
- **分析要件：ニーズに応じた多様な分析の実現**

開示要件は，ユースケース毎に異なる個別のものと，ユースケース共通のものが存在する．個別のものとしては，

ユースケース 1 では，秘匿検索性が必要となる．これは，検索を行う側(利用者側)が何を検索したのかを，検索をされる側(提供者側)に対して秘匿化する性質である．

一方，ユースケース 2 では，匿名性を満たす必要がある．これは，利用者側が，提供者側のどの組織が攻撃を受けているかを知りえないという性質である．また，ユースケース共通

の開示要件としては，開示制御性がある．これは，データ自体を開示しないことや，提供者側が開示する結果を制御できる性質であり，これにより提供者側の不本意な情報開示を抑制することが可能となる．利用者側，提供者側が必要十分情報の開示に合意し，かつ情報共有手段が開示要件を満たすことが保証されれば，要因 1 及び 2 は解決されると考える．

分析要件は，ユースケース毎に異なるニーズ（ユースケース 1 では自組織と同じ攻撃を受けた組織の発見，ユースケース 2 ではコミュニティ内で自組織と同じ攻撃を受けた組織数の調査）に対して，必要な分析を行うために，実施する分析を柔軟に変更可能とすることである．

表 2.1　主要要件

#	ユースケース	必要十分情報	開示要件	
			共通	個別（ユースケース毎）
1	自組織と同じ攻撃 X を受信した組織の発見	提供者側が攻撃 X を受信しているか	【開示制御性】 ・データ自体を開示しない（要因 2） ・分析対象の開示データを提供者側が制御できる（要因 1）	【秘匿検索性】 攻撃 X を受信していない提供者側には，利用者側の問合せ対象が攻撃 X だと判らない（要因 1）
2	分析組織による，コミュニティ内で攻撃 X を受信した組織数の調査	攻撃 X を受信した組織の数	同上	【匿名性】 利用者側は，どの組織が攻撃を受けているかは判らない（要因 1）

2.3　設計

2.3.1　機能設計

2.2.3 項の主要要件を満たすために提案システムが具備すべき機能について表 2.2 に示す．

機能 1-1, 1-2 は分析要件を満たすために必要な機能である．提供者側に分析用プラットフォームを設置し（機能 1-1），分析要求に応じたロジックをコンテナ化・配信・実行（機能 1-2）することで，提供者側で分析処理を行いその結果だけを共有することが可能となる．また，利用者側は様々な分析要件を柔軟に選択することが可能となる．

機能 2 は，ユースケースに依らない共通の開示要件を満たすために必要な機能である．提供者側プラットフォームにアクセス制御の仕組みを導入することで，分析対象のデータを制御することが可能となる．

機能 3-1, 3-2 は，ユースケース毎の開示要件を満たすために必要な機能である．分析ロジックの健全性を検証するエンティティを導入し（機能 3-1），数理技術で分析ロジックの動作を数学的に保証する（機能 3-2）ことで，秘匿検索性や匿名性等，ユースケース毎に求められる開示要件を満たすことが可能となる．

表 2.2　機能設計

#	分析・開示要件	機能
1	[分析要件] 多様な分析要求に柔軟に対応	1-1. 分析用プラットフォームを設置 1-2. 分析要求に応じたロジックをコンテナ化・配信実行
2	[共通開示要件] 分析ロジックのデータアクセスを制御	2　提供者側プラットフォームにアクセス制御の仕組みを導入
3	[個別開示要件] 分析ロジックがユースケースの要件を満たすことを保証	3-1. 分析ロジックの健全性を検証するエンティティを導入 3-2. 数理技術で分析ロジックの動作を数学的に保証

2.3.2　情報共有モデル

2.3.1 項の機能要件を満たすことを目的として設計した提案システムの情報共有モデルを図 2.2 に示す．

情報提供者側は，統一的な分析プラットフォームを導入することで，異なる組織の分析ロジックを統合的に処理できる環境を提供する（機能要件①-1 プラットフォーム導入）．さらに，各組織の分析ロジックをコンテナ化し，独立した実行環境で動作させることで，セキュリティと柔軟性を確保する（機能要件①-2 ロジックのコンテナ化）．

加えて，アクセスポリシーを用いたリソース管理により，各組織の分析ロジックが許可されたデータのみにアクセスできるよう制御する（機能要件② リソースアクセス制御）．また，動作ポリシーを設定し，分析ロジックの健全性を検証することで，不正な処理や悪意のあるロジックの実行を防止する（機能要件③-1 健全性検証）．

さらに，リクエスト，分析，レスポンスにおいて暗号化等数理的手法を導入し，分析ロジックの動作を数学的に保証する（機能要件③-2 数理技術導入）．

本アーキテクチャは，各組織が独自の分析ロジックを適用できる柔軟性を確保しつつ，データの機密性を保持する仕組みを提供する．アクセス制御やロジックの健全性検証を通じてセキュリティリスクを最小限に抑えながら，SOC の分析能力を最大化することを目的としている．

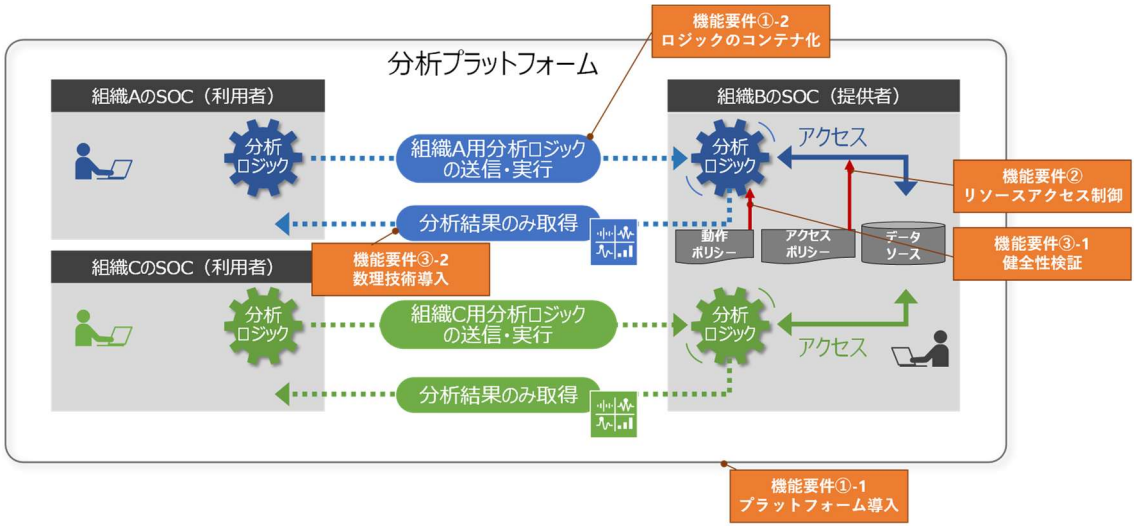

図 2.2　分散 SOC における情報共有モデル

2.4　実証実験

提案した情報共有モデルによる情報共有の有効性を検証するため実証実験を行う．具体的には，組織間での情報共有によって，単独の組織では取得できない情報が得られるかを検証する．実証実験では，異なる業種の 3 組織（企業 2 社，教育機関 1 校）を対象に，実際の運用環境における情報共有の効果を分析した．共有する情報は，組織間で共有されてこなかった，あるいは共有の有効性が未検証な 3 種類の情報（ダークネット観測データ，Web 接続ログ，インシデント対応記録）を対象とした．

表 2.3 に実証に参加した組織を示す．各業種では，業界特有の脅威が存在し，それぞれ異なる視点からの脅威分析が可能となることが期待される．異なる業種での情報共有により，個別の組織では捉えきれない攻撃手法や脅威動向を把握し，より包括的な防御を実現できると考えた．特に，企業と学校の特性の違いに着目した．企業では，厳格なセキュリティポリシーのもと，インシデント対応が詳細に記録されるため，組織的な対応の知見が蓄積されている．一方，教育機関では，利用者の自主性が高く，幅広いネットワーク利用が行われるため，未知の攻撃手法や多様なセキュリティインシデントが観測される可能性がある．この対照的な環境におけるデータを活用することで，分散 SOC の有効性をより多角的に検証できると考えた．

表 2.3　実証実験参加組織

組織	業種	組織規模
組織 A	製造業	小（組織は大きいが実証に参加したのが 1 つの部署のみのため）
組織 B	学術	大
組織 C	電力	大

2.5　本章のまとめ

本章では，データ開示を伴わない分析を支援する，情報共有モデルを設計し提案した．

データ分析機能を相手組織に共有することにより，柔軟な分析要件に対応することが可能となる．また，分析に秘密計算を活用すれば，互いの情報を暗号化したまま分析が行える．

本システムにより，セキュリティ情報共有を阻害している要因である，(1)機微情報の開示リスクや，(2)情報共有にかかるコスト，を低減することができ，情報共有が促進されることが期待できる．

第 3 章～第 5 章では，提案システムの実証実験について述べる．

第3章　複数組織を跨ったダークネット観測の効果に関する検証

3.1　導入

ネットワークに接続されたデバイス（PC，タブレット，スマートフォン，インターネット・オブ・シングス（IoT）デバイスなど）は急速に増加している[15]．ネットワークの利用は利便性を向上させるが，セキュリティ対策が不十分なデバイスはマルウェアによって攻撃される．サイバー攻撃はますます巧妙化しており，サービス拒否（DoS）攻撃，情報漏洩，ウェブサイト攻撃，金銭を強奪するためのランサムウェア攻撃などが含まれている[16]．

突発的なサイバー攻撃からネットワークやシステムを守ることは，セキュリティ運用上不可欠である．そのためには，サイバー攻撃の傾向を把握する必要がある．世界的なサイバー攻撃の傾向を把握する一つの方法として，ダークネットと呼ばれる未使用のインターネットプロトコル（IP）アドレスのブロックを観察する方法がある．これらのネットワークアドレスは未使用であるため，ユーザによる Web サイトへのアクセスなど，良性の通信は基本的に観測されない．一方，設定ミスによる通信，分散 DoS（Distributed DoS）を目的としたなりすまし通信，DRDoS（Distributed Reflective DoS）に活用できる踏み台を探るための通信，マルウェアによるスキャンなどのネットワークプロービングは観測できる[17]．したがって，これらのダークネットベースの監視システムと通信するホストを観察することによって，DDoS 攻撃やマルウェア感染の被害者，またスキャンを実行する疑わしいホストに関する情報を収集することが可能である．しかし，ダークネット監視の課題は，組織が監視用に利用できるアドレス空間が限られているため，インターネット上の疑わしい活動の全容を把握することが難しい点である．

本研究の目的は，複数の組織間でのダークネット観測結果を詳細に相関分析することにより，SOC 間で生データを共有する効果を示すことである．本論文では，業界，最初のオクテットのサブネット範囲が異なる 2 つの組織で観測されたダークネットデータを分析する．観測ブロックを複数のブロックに分割し，1 つの組織が組織内の類似性分析を行うようにして，組織内と組織間の類似性分析結果を比較する．

本研究の 2 つの貢献は次の通りである：

- 組織内と組織間の計算結果を比較することにより，複数の組織によるダークネット監視の定量的効果を示す．
- 結果が宛先のポート/プロトコルに応じて異なるかどうかを明確にする．これにより，ポート/プロトコルによってアクセスが無差別であるか，特定の組織やネットワークをターゲットにしている可能性があるかが示唆される．

本章の残りの部分は以下のように構成されている．3.2 節では，関連研究を説明し，それらと本章の違いを説明する．では，我々が提案する SOC 間連携について述べる．3.3 節では我々の手法とデータセットの詳細を述べる．実験結果を 3.4 節で詳細に説明し，3.5 節で課題について議論する．最後に 3.6 節で結論を述べる．

3.2 関連研究

3.2.1 セキュリティ・コミュニティにおける情報共有

セキュリティ情報を共有するための取り組みはいくつかある．ISAC は，セキュリティを含む情報を共有し，様々な脅威を軽減することを目的としている．ISAC では，参加する会員組織が，ISAC の会員組織をまとめるオペレーションセンターに情報を提供する[18]．そして，オペレーションセンターのアナリストが各加盟組織からの情報を集約し，情報を分析し，分析した情報を加盟組織に送信する．そのため，ある会員が脅威を観測してから，他の会員がその情報を受け取るまでには，常に無視できないずれがあり，リアルタイムでの対応が困難である．また，分析され要約された情報は理解しやすいが，分析中に必要な詳細情報（正常状態のログやマルウェアの検体など）が失われてしまう．このほか，攻撃コマンドサーバのドメインや IP アドレス，マルウェアのハッシュ値などの検知指標（インジケータ）を共有するための官民情報共有インフラである AIS などの取り組みも行われている[19]．AIS は，2018 年 12 月には米国の 252 組織と他国の 13 組織で構成されていたが，2019 年 6 月に実際に利用した組織はわずか 10 組織だった[20]．その理由の 1 つとして，悪意のあるメールアドレスや IP アドレスの技術的指標の多くは，インシデントの状況を特定するための背景や文脈を持っていないため，組織が対策を講じることが難しいことが挙げられる．

3.2.2 ダークネット監視

本論文と同様に，複数のダークネット観測を分析する研究がいくつか行われている．古本ら[21] は，6 カ国のダークネットを観測し，宛先ポートごとのインバウンド通信とアウトバウンド通信を分析した．また，Cooke ら[22] は，大手サービスプロバイダ，大企業，学術ネットワークの分散型ブラックホールセンサを調べ，/25 から/8 の範囲のアドレスブロックの観測点を比較し，接続 IP 数の違いを明らかにしている．Bailey ら[23] は，25 のダークネットセンサーブロックについて，IP あたりの 1 日平均パケット数を比較している．その結果，1 IP あたり 1 日 10 パケットから 10,000 パケット以上まで，トラフィック数が異なることが明らかになった．また，MySQL サービスに対する攻撃はダークネットセンサーブロックのうち 23 ブロックで観測され，RPC-DCOM に対する攻撃はアカデミックネットワークで観測された．Soro ら[24] は，3 カ国で観測された複数のダークネットを比較し，各観測ポイントでダークネットに接続している AS 所属ホストの類似性を明らかにした．

一方，本研究では，異なる組織のネットワークにおける各観測ポイントで得られた不審なホストの類似性に着目している点が異なる．

Berthier と Cukier [25] は，監視する IP アドレスを 1 個から 5,000 個に変更すること

で，スキャン活動全体のカバー率とダークネットの規模との関係を評価した．この評価の結論は，1 つの未使用の IP がスキャン活動の 3.3%をカバーしているが，監視するネットワークアドレスの数を 560 に増やすことで，このカバー率を 50%に拡大できるというものであった．Pemberton ら[26] は，/16 ネットワークテレスコープからのデータを分析し，到着率データの非均一性を調査し，/16 ネットワークテレスコープのサブセット内でのネットワークテレスコープアドレスの割り当てに関する 4 つのスキームを分析した．彼らは，/16 サブネットワーク全体のダークネット活動を予測するには，/32 ランダム・ネットワーク・テレスコープを多数配置するのが最良の戦略であると指摘した．

我々の研究では，スキャン活動のカバー範囲を拡大するために，複数の組織を監視することの有効性を示している．

3.3 提案手法

サイバー攻撃は無差別に行われることもあるが，特定の組織に対して行われることもある．そこで，1 つの組織を監視するよりも複数の組織を監視した方がより多くの攻撃情報を収集できると仮定し，その仮定を検証する手法を提案する．

3.3.1 データセット

表 3.1 にセンサーの概要を示す．組織 B は学術機関，組織 C は電力会社である．組織 B は/19 サブネットのブロックを，組織 C は/24 サブネットのブロックを観測している．これらの観測点からのパケットを 1 日ごとに分類し，宛先ポート/プロトコルのペアで集計した．ほとんどの宛先ポート/プロトコルでは，アクセスするホストは少ない．このため，各宛先ポート/プロトコルは，観測されたパケットペアのユニークな送信元アドレスが 1 日で 1000 個以上ある場合に解析対象としている．つまり，日によって解析対象の宛先ポート/プロトコルが異なる可能性がある．

表 3.1 センサーの概要

センサー	Subnet	アドレス数	観測期間
組織 B	/19	8192	2019 年 8 月 1 日～8 月 31 日
組織 C	/24	256	同上

3.3.2 比較方法

図 3.1 に監視環境の概要を示す．組織内と組織間の計算を比較するため，組織 B の観測ブロックを組織 C のブロックと同じ大きさになるように 32 個の/24 のサブネットに分割した．組織 B の各サブネットの日付 d におけるソースアドレスの集合を $B = \{b_{d,1}, b_{d,2}, \ldots, b_{d,32}\}$とし，組織 C のサブネットにおけるソースアドレスの集合を $C = \{c_{d,1}\}$とする．

次に，Jaccard 係数と，取り得るすべての組み合わせにおける 2 ブロック間の和を計算する．Jaccard 係数は式(1)で計算される値で，領域間の類似度が低いほど値が小さくなることを意味する．

$$Jaccard(i,j) = \frac{|i \cap j|}{|i \cup j|} \qquad 式(1)$$

ここで，i と j は i, j⊂B∪C and i /=j である．

和は 2 つのブロックで観測されたユニークなホストの数を表し，数が多いほど，より多くのソースホストを観測できることを意味する

組織内と組織間の係数分析結果を比較するために, Jaccard 係数の差と和集合の差という 2 つの指標を定義する．Jaccard 係数の差は，式(2)で計算される値である．

$$\frac{1}{D}\sum_{d=1}^{31}\left(\frac{\sum_{i=1}^{|B|} Jaccard(b_{d,i}, c_{d,1})}{|B|} - \frac{\sum_{j=1}^{|B|}\sum_{k=1}^{|B|} Jaccard(b_{d,j}, b_{d,k})}{{}_{|B|}C_2}\right) \qquad 式(2)$$

ここで，xCy は集合 x から y 個を選んで得られる組み合わせ数を，D は宛先ポート/プロトコルが分析対象となった日数（一日に 1000 個以上のユニークな送信元アドレスが観測した日数）である．また，Jaccard 係数を計算する際，0 除算が発生した場合はシグマの計算をスキップし，項の分母値から 1 を引く．

図 3.2 に計算処理を示す：

(1) 組織間における Jaccard 係数の平均値の算出

(2) 組織内同士における Jaccard 係数の平均値の算出

(3) 1 から 2 の結果を差し引くことで，その日の Jaccard 係数の差を計算する.

(4) 全日の Jaccard 係数の差の平均を計算する

計算された値が負の場合，それは組織間の計算における Jaccard 係数が組織内の計算における Jaccard 係数よりも低いことを意味し，すなわち組織間の方が類似度が低いことを示し，組織間で観測することが有益であることを示す.

式(2)と同様の計算を行い，和集合の差分を算出する．式(2)との違いは，Jaccard 係数を計算するのではなく, 2 つのブロックにおける異なる送信元ホストの数を算出する点である. 和集合の差分が正の場合，それは組織間計算における和集合の要素数が組織内計算の要素数よりも多いことを意味し，すなわち，より多くの疑わしいホストが観測されていることを示している.

さらに，累積頻度を比較する．ブロック数を N と定義し，各 N で観測された接続ホスト数と，すべてのブロックで観測された接続ホスト数の比率を算出する．ここでは，組織内計算と組織間計算において，領域を 1 から順に拡張した場合を比較する．例えば，N = 3 の場合，$\{b_{d,1}, b_{d,2}, b_{d,3}\}$と$\{b_{d,1}, c_{d,1}, b_{d,2}\}$における累積頻度を比較する．

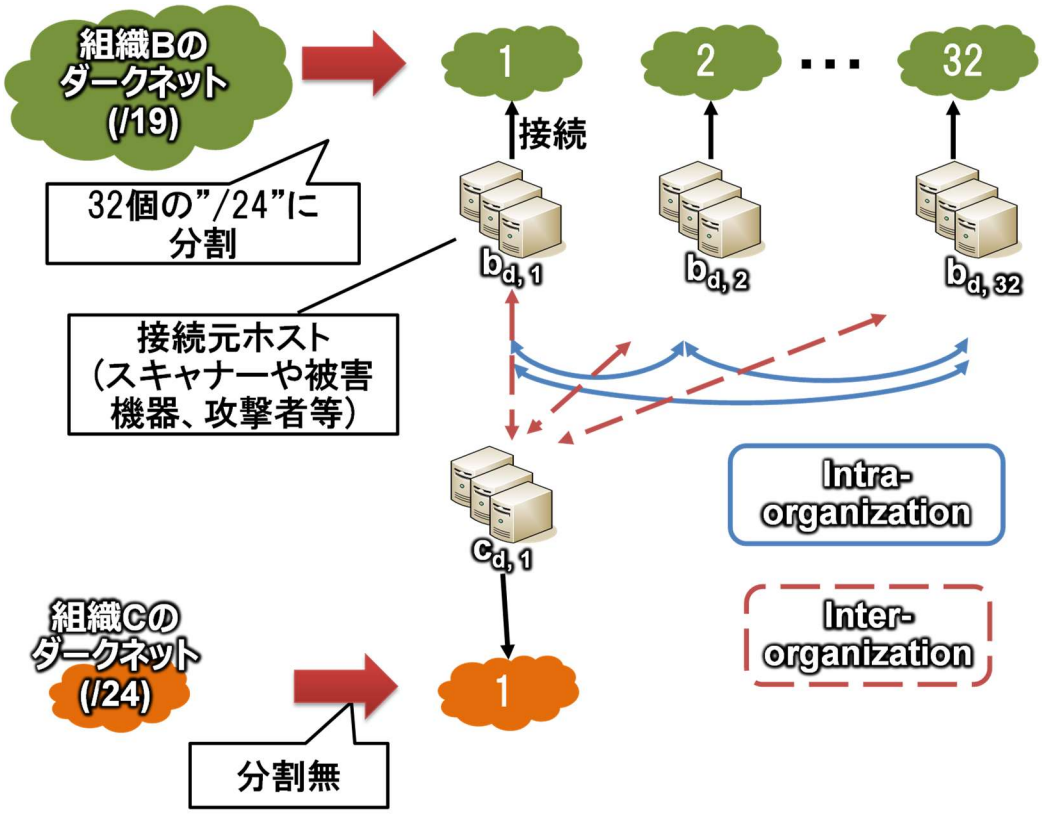

図 3.1　観測環境の概要

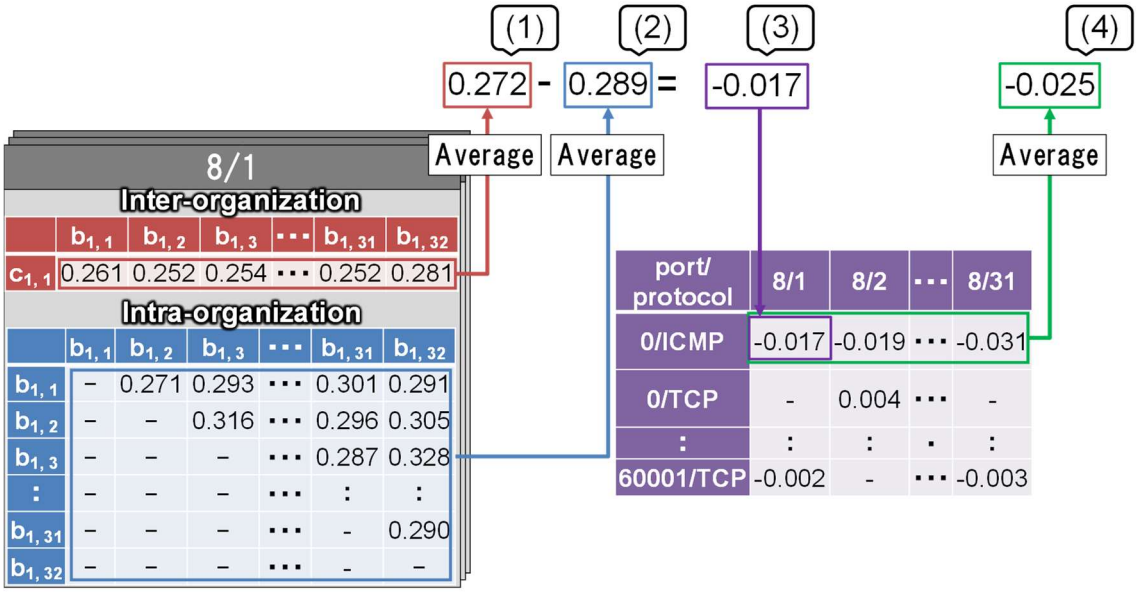

図 3.2　計算プロセス

3.4　評価結果

図 3.3 は，各宛先ポート/プロトコルにおける組織内計算と組織間計算のジャッカード係

数の差分を示している．図 3.4 は，各宛先ポート/プロトコルにおける組織内計算と組織間計算の和集合の差分を示している．

図 3.3 の「all」の値から，組織間計算におけるジャッカード係数は 0.020 ポイント減少している．さらに，宛先ポート/プロトコルの組み合わせごとにばらつきが見られる．特に，TCP/22，TCP/53，TCP/88，TCP/139，TCP/445，TCP/5900，TCP/8728 では，類似性が顕著に低い．このことから，実験期間中にこれらの宛先ポート/プロトコルを対象とした攻撃は，広範な IP アドレス空間ではなく，特定の IP アドレス空間または特定の組織を標的としていたと推測される．一方，TCP/23 および TCP/2323 では類似性の差がほとんど見られなかった．これらのポートは，IoT マルウェア Mirai の感染活動で使用されており，Mirai はランダムに接続を試みるため，組織による違いが少ないと考えられる．

図 3.4 の「all」の値から，組織間計算における接続ホスト数は 1,180 増加している．一方で，宛先ポート/プロトコルごとに比較すると，TCP/445 では大幅な減少が見られる．なお，1 つの送信元ホストが複数の宛先ポート/プロトコルと通信している可能性があるため，各ポート番号/プロトコルごとの合計値と全体の値は一致しない．

図 3.5 および図 3.6 は，2019 年 8 月 28 日における 22/TCP および 9200/TCP の累積比率を示しており，これらのポートでは顕著な差異が確認されている．

図 3.5 および図 3.6 から，複数の組織でダークネットを監視することにより，より多くのスキャンを行うホストを収集できることが分かる．この差異は，N を増加させても維持されている．これは，同じアドレス空間であっても，単独の組織による観測よりも複数の組織による観測の方が多くの情報を得られることを意味している．

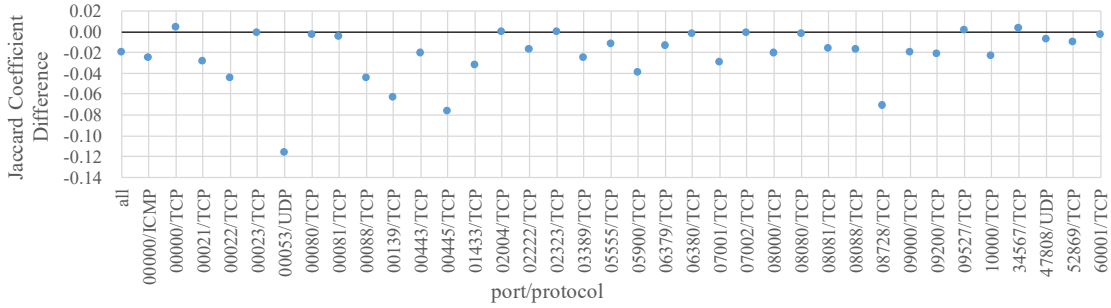

図 3.3　それぞれの宛先ポート/プロトコルの Jaccard 係数の差分

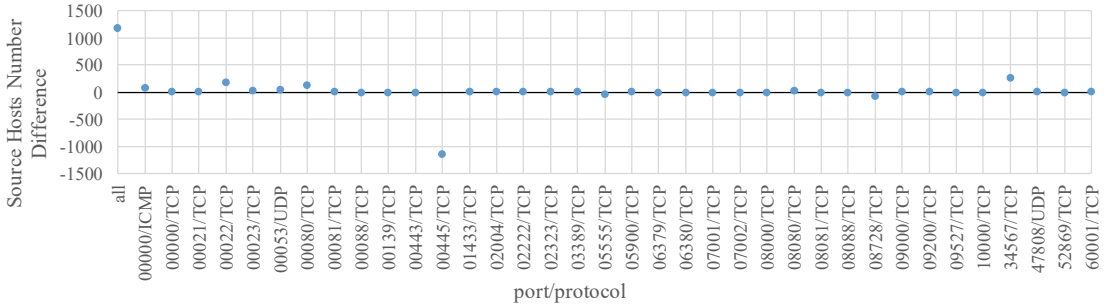

図 3.4　それぞれの宛先ポート/プロトコルの和集合の差分

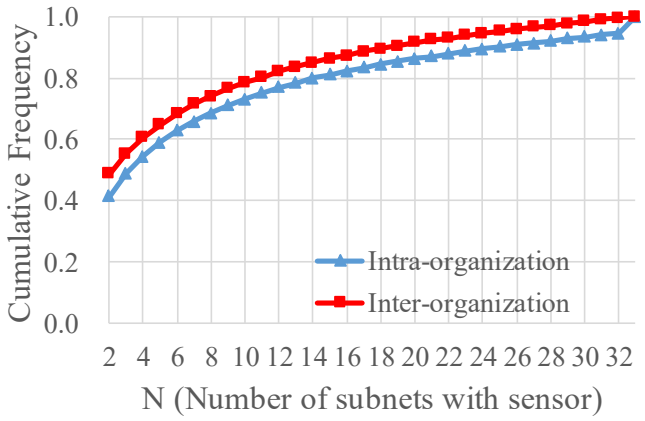

図 3.5　ホストの累積頻度（2019/8/28, 22:TCP）

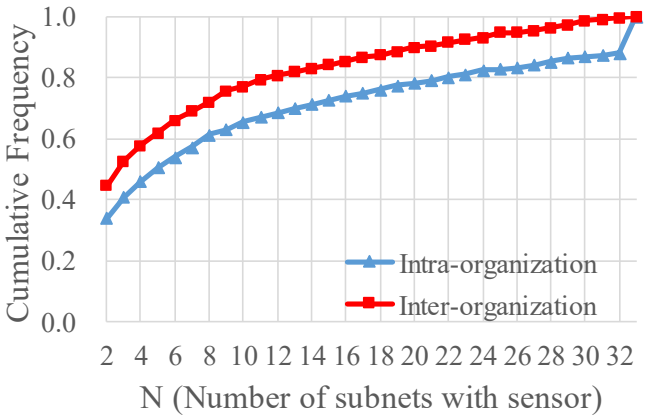

図 3.6　ホストの累積頻度（2019/8/28, 9200:TCP）

3.5　　考察

本研究の結果は，分散 SOC メンバー間でのダークネットログの共有が，インターネット上のサイバー攻撃の状況を把握する上で実際に有効であることを示唆している．

現在の分散 SOC の運用では，監視を行う組織から他の組織へ，pcap 形式のパケットログが SSH を用いて毎日転送され，そこで解析が行われている．この共有速度は既存の手法と比較して速いが，リアルタイムでの脅威対策を行うためには，さらに速度を向上させることが望ましい．

現在の分散 SOC メンバーは，ダークネットには機微機密情報が含まれないこともあり，ダークネット宛のパケットの内容について，事前のサニタイズやマスキングなしに共有できることに合意している．しかし，分散 SOC に関心を持つ一部の組織は，情報漏洩の懸念から，自組織のログを直接共有することに慎重な姿勢を示す可能性がある．分散 SOC の規模を拡大するためには，このような慎重な組織にも配慮する必要がある．第 4 章では，機微機密情報を含むデータの活用方法についての実証について述べる．

また，異なる組織は，それぞれの業界やミッションに応じた多様なセキュリティポリシーを有していることを認識し，それらの要件を満たす複数のプロトコルやスキームを提供することで，大規模な SOC 連携を確立することを目指す．

3.6 本章のまとめ

本章では，複数の組織におけるダークネット観測の有効性を検証し，SOC の協力体制を評価した．その結果，単一の組織による観測よりも，複数の組織による観測の方が，ダークネットに接続するホストをより多く観測できることが示された．これらのホストは比較的疑わしいものであり，その情報は，ファイアウォールのフィルタリングルールの更新や，組織ネットワーク内でこれらのホストに接続する疑わしいデバイスの特定に活用される．

さらに，宛先ポート/プロトコルごとに異なる傾向が見られることが明らかになった．今後の課題として，協力組織の拡大や観測時間に基づく分析を進める予定である．

第4章　複数組織の接続傾向を用いた自律進化型防御システムの提案と評価

4.1　導入

近年，標的型攻撃に見られるように，攻撃が高度化しており，企業や国家にとって重大な脅威となっている．ここで，マルウェアのダウンロード[27] に加えて，マルウェアとの通信，フィッシングサイトの表示，スパムの発信[28] 等，悪意を持ったサイト（以降，悪性 Web サイト）がサイバー攻撃において重要な役割を有している．このことから，サイバー攻撃の被害を抑制するためには，Firewall や Web Proxy 等のゲートウェイにより悪性 Web サイトへの接続を遮断する，いわゆる出口対策[29] が重要であるといえる．悪性 Web サイトへの接続を遮断する方法としては，悪性 Web サイトと判定したサイト群であるブラックリスト（以降，BL）への接続を遮断する方法や，反対に良性 Web サイトと判定したサイト群であるホワイトリスト（以降，WL）以外への接続を遮断する方法が存在する．しかし，公開されている BL や WL だけで，インターネット上にある全ての良性 Web サイト，悪性 Web サイトを網羅するのは難しい．従って，これらリストに存在しないサイト群（未知サイト）への接続が良性か悪性かを判定できないという課題がある．そこで本章では，未知サイトへの接続の不審度を，接続傾向との乖離度から算出し，当該不審度の高低により未知サイトへの接続が良性か悪性かを判定する方法を提案する．これは，悪性 Web サイトへの接続は，普段の接続傾向とは異なる可能性が高い，という考えに基づいている．尚，接続傾向は組織が保有する，サイトへの接続ログを分析し算出する．

上述の方法は，普段の接続傾向と異なる接続を全て悪性 Web サイトへの接続と見做す．このため特に，良性 Web サイトを悪性 Web サイトと判定してしまう誤検知が多く発生する．そこで誤検知の削減方法として，普段の接続傾向がより多くの良性 Web サイトをカバーできるように，多種多様な正常の接続情報をインプットとして利用することが考えられる．ここで日立製作所では，慶應義塾大学と協力して，複数組織の SOC（Security Operation Center）を跨ったセキュリティオペレーション連携により，サイバー攻撃への集団防御を実現する，分散 SOC アーキテクチャを提案している[30] ．本アーキテクチャの考えに基づき，自組織が保有する接続ログだけでなく，他組織が保有する接続ログを利用する．これにより，例えば自他組織が共に不審と判断した時のみ悪性と判定する，といった方法で誤検知の削減が期待できる．これは，悪性 Web サイトへの接続は，他の異業種組織であっても普段の接続傾向と乖離している，という仮説に基づいている．一方で，組織が保有する接続ログは一般的に機微情報であり，且つログ量も大きくなる傾向があるため，他組織への共有は困難である．ここで我々は，生データの開示を伴わない分析を支援する，秘匿データ分析シ

ステムの研究を進めている[31]．本章では前述のシステムを応用した，接続ログといった生データを直接共有するのではなく，他組織上で接続傾向及び接続傾向を用いた不審度の算出を行う分析ロジックを実行し，不審度のみを共有するシステムを提案する．

また，他組織の接続ログを用いることで精度が向上できたとしても，依然として誤検知の可能性は残っており，業務で利用する良性 Web サイトが悪性 Web サイトと判定され，接続が拒否されることにより，業務が阻害される可能性がある．そこで，接続先が悪性 Web サイトと判定された際に，即座に接続を遮断するのではなく，いったん追加認証を課すことにより，人間による業務上必要な接続は許可しつつ，マルウェア等による機械的な接続は遮断する．

これらにより提案システムは，未知の悪性 Web サイトにも有効性があり，業務遂行に必要なサイトを誤って悪性と判定した際の業務阻害を緩和することが可能である．

本研究の主な貢献は以下の 3 点である．

- 機微性が高いセキュリティログを組織の垣根を越えて共有/分析するためのインフラを設計/開発
- インフラ上で動作し，複数組織の接続ログを基に，悪性ドメインを発見する異常検知アルゴリズムを考案. 性能評価実験を通じ，自他組織両方の接続ログを使用することにより，検知性能を自組織の接続ログのみを使用した場合と比べて，Area Under the Curve（以降，AUC）の観点で 0.018 向上することを確認
- インフラ活用のアプリケーションとして，著者らが研究を進めてきた悪性 Web サイトへの接続防止手法である，自律進化型防御システム（AED : Autonomous Evolution of Defense） [32] , [33] , [34] 及び Web Proxy と連携して，Web 接続制御を行う機構を開発．ユーザ評価実験により制御処理の 93.6%が 3 秒以内であることから，実運用に適応可能であることを確認

本章の構成は次のとおりである．まず，4.2 節で既存の悪性 Web サイトへの接続防止手法とその課題について述べる．4.3 節で同課題を解決する手法を提案する．4.4 節で提案システムの実装，実環境を用いた評価実験及び有効性の評価を行う．その後，4.5 節で関連研究について述べ，最後に 4.6 節でまとめを述べる

4.2 背景と課題

本章では，既存の悪性 Web サイトへの接続防止手法である，AED の概要とその課題を述べる．

4.2.1 AED の概要

AED では，ユーザが未知サイトへ接続しようとした場合に，Proxy で追加認証を要求する．これにより，マルウェアによる機械的な接続を遮断しつつ，人間による業務上必要な接続は許可することができる．また，WL 型 AED[34] では，組織に属する一定数のユーザが追加認証を突破し接続したサイトを WL に追加することにより，追加認証の発生回数を減らすことで，ユーザの利便性低下を抑えることができる．

4.2.2 AED の課題

(1) 未知サイト接続時の利便性低下

WL 型 AED では，WL を動的に拡張していく仕組みがあるが，未知サイトへ接続する際には必ず追加認証が発生してしまい，ユーザの利便性が低下する．

(2) 小規模組織で WL の拡張が困難

AED で WL を拡張する仕組みは，組織内の他ユーザが接続した記録を用いるため，ユーザ数が少ない小規模の組織では十分に WL の拡張が行えない．実際に先行研究[34] では，ユーザが許容できる利便性を維持するために，1000 人以上のユーザが必要という結果が出ている．これは大きな組織でないと実現が困難である．

4.3 分散異常検知型 AED

4.2 節で述べた課題を解決するシステムとして，複数組織の接続傾向を利用して接続を制御する，分散異常検知型 AED を提案する．本節では，提案システムの設計について述べる．図 4.1 に提案システムの全体像を示す．本システムは，(i) 分散 SOC 基板，(ii) ドメイン分析ロジック，(iii) AED に分類される．

分散 SOC 基板は，データ分析を行う機能である分析ロジック（本ケースでは後述するドメイン分析ロジック）を他組織で実行することで，他組織が保有する生データである各種インテリジェンス（本ケースでは接続ログ）を直接取得せずに分析結果（本ケースでは接続の不審度）のみを取得可能とする．組織が保有する接続ログは一般的に機微情報であり，且つログ量も大きくなる傾向があるため，他組織への共有は困難である．前述の分析結果のみを取得することにより，効率的且つセキュアに他組織のデータを活用することが可能となる．このように，他組織の接続ログを使い，接続ログを記録するユーザ数を仮想的に増やすことで，4.2.2 項の課題(2)を解決する．尚，図では情報提供側組織が 1 つの場合を例示したが，情報提供側組織が複数存在する構成も考えられる．

また，ドメイン分析ロジックは，異常検知アルゴリズムを利用し，組織が保有するサイトへの接続ログから接続傾向を分析し，あるサイトへの接続の不審度を，当該接続傾向との乖離度から算出する．これにより，WL に依らない判定が可能となり，4.2.2 課題(1)を解決する．そして AED は，WL，BL，及び分散 SOC 基板から得られた不審度を元に，接続の可否や追加認証を制御する．

図 4.1 に示した，ユーザがインターネットへの接続を行う際の，分散異常検知型 AED の処理フローについて以下に説明する．

① 接続要求

組織 A のユーザが認証 Proxy を介してインターネットの Web サイトに接続要求を行う．

② 接続可否判定要求

認証 Proxy は，AED に対して，ユーザが要求した接続について，不審度判定を要求する．AED は接続先 Web サイトのドメインが WL や BL に含まれている場合は処理⑥に移行す

る．WL，BL に含まれていない場合は処理③に移行する．

③　不審度判定要求

AED は，組織 A 内の分散 SOC 基板に対して，接続の不審度判定を要求する．分散 SOC 基板は，組織 A のドメイン分析ロジックにより，組織 A の接続傾向を使用し不審度を算出する．

④　組織外への不審度判定要求

組織 A の分散 SOC 基板は，組織 B の分散 SOC 基板に接続先の不審度判定を要求する．分散 SOC 基板は，組織 B のドメイン分析ロジックにより，組織 B の接続傾向を使用し不審度を算出する．

⑤　不審度判定結果統合

AED は，組織 A，B が算出した不審度を元に，接続の不審度を判定する．具体的には，組織 A，B が算出した不審度の内，値が小さい方が予め定めた閾値を上回る場合，接続が不審であると判定する．

⑥　接続可否または追加認証

AED は認証 Proxy に対して，不審度判定結果である，接続の可否，または追加認証を返す．接続先サイトが WL に含まれる場合は接続を許可し，BL に含まれる場合は接続を拒否する．また，⑤の不審度判定結果統合で，接続が不審だと判定された場合は追加認証を要求し，そうでない場合は接続を許可する．

以下の項にて，分散異常検知型 AED の各コンポーネントの設計について詳述する

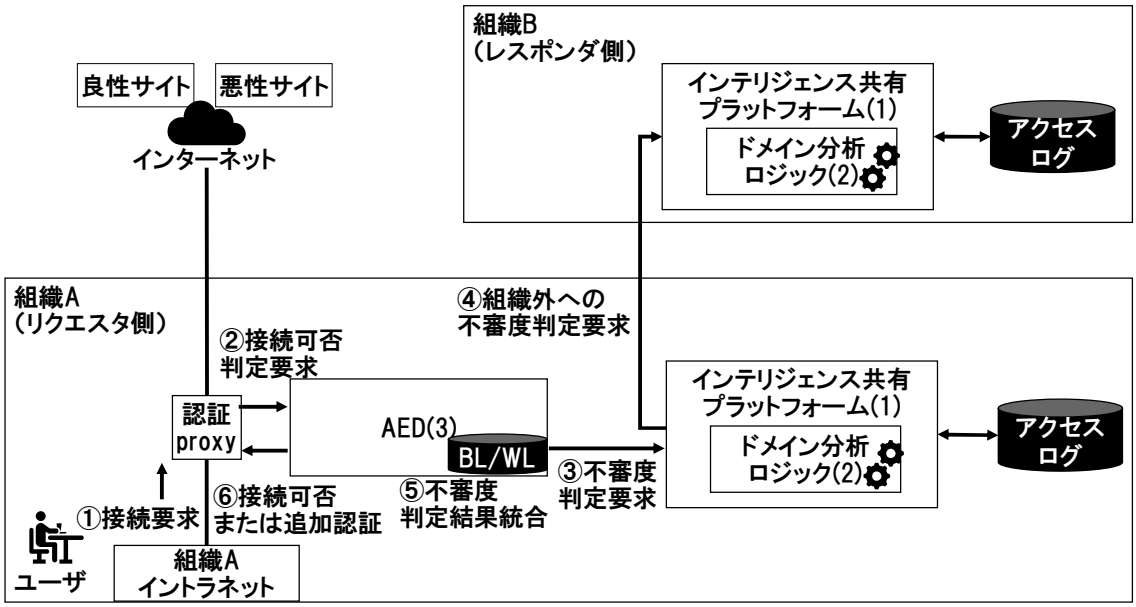

図 4.1　提案システム（分散異常検知型 AED）の全体像

4.3.1　インテリジェンス共有プラットフォーム

図 4.2 にインテリジェンス共有プラットフォームの概要図を示す．組織 A には組織 B へ分析リクエストを送るリクエスタが在り，リクエスタには GUI のツールや AED 等からリクエスト情報が送られる．組織 B には分析リクエストを受信し処理を行うレスポンダ，および接続ログといった共有・分析対象のインテリジェンスが存在する．本プラットフォームでは，分析の実行主体である分析ロジックは，リクエスタ側が作成しレスポンダ側で実行される．レスポンダ内には，分析ロジックの実行基盤であるプラットフォーム，およびインテリジェンスへの接続制御を行うリソース接続 API が存在する．分析ロジックとその実行基盤を分離して別組織が管理することで，リクエスタ側の要望に応じて様々な分析を実施することができる．尚，レスポンダ及びインテリジェンスが他組織上に存在することを想定し設計をしているが，自組織内にも具備することで，他組織で行う分析と同様の分析を自組織の情報を用いて行うことも可能である．また，各コンポーネント間のやり取りは，WebAPI を介して行う．また，本プラットフォームでは，リクエスタ側，レスポンダ側双方の情報流出を極力抑えるように設計している．これらの設計については，第 6 章にて詳述する．

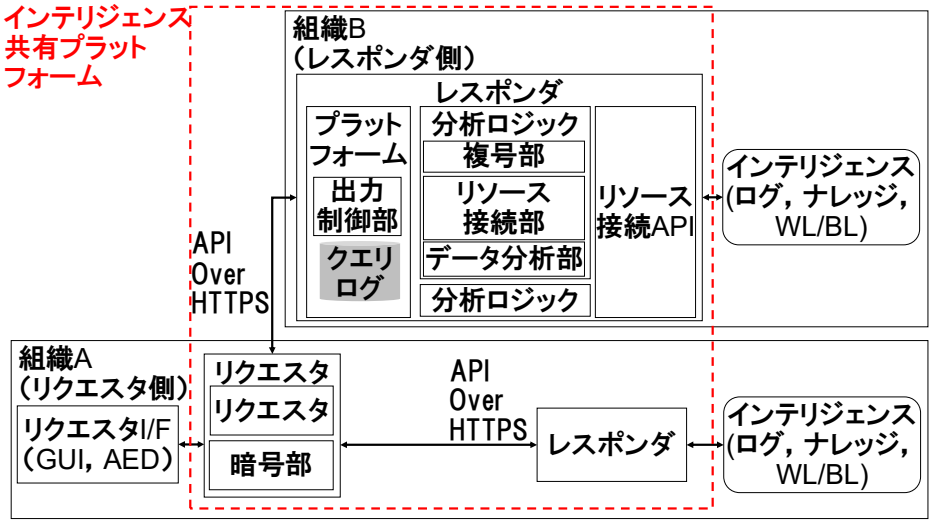

図 4.2　インテリジェンス共有プラットフォームの概要図

4.3.2　ドメイン分析ロジック

本節で述べるドメイン分析ロジックで用いる異常検知アルゴリズムは，著者らが研究してきた AED[34] をベースに，[35] にある N-gram による分析を追加で援用している．

AED 同様，本ロジックでは組織内で記録された Web 接続ログを基に当該組織における正常接続モデルを構築する．そして，入力として与えられた分析対象ドメインと当該モデルとの「距離」を基に異常検知を行う．

モデル化に際してはセキュリティ専門サイトなどが公開している既知の悪性ドメインの

情報は使用しない．これは，本研究の目的が，各々の組織で観測・記録するセキュリティ情報を，他組織と共有することによる検知効果を，組織単独の時と比較して評価することであるからである．

(1) 分析ロジックの要件

自組織だけではなく他組織上で実行されるという特徴から，分析ロジックには(a) 使用するメモリを一定量以内に抑える，(b) CPU 負荷を一定内に抑える，(c) インターネットへの通信を制限する，(d) 必要に応じて情報提供側が分析内容を監査できるようにする，(e) 必要な情報の一部が情報提供側上に無い場合でも，分析を実施できるようにする，という固有要件がある．後述の設計では，上記要件をどのように満たしているのか言及する．尚，要件(d)については，6.2.1 項にて言及する．

(2) 入力

本実装では接続ログの各エントリは表 **4.1** のカラムを持つことを想定する．接続日時，接続元の識別子，接続先サイトのドメイン名，ドメイン名に対応する IP アドレスは何れも一般的な Web Proxy ログ，DNS ログから取得できる情報である．分析フェイズにおいて情報利用者から送信される分析リクエストに含まれるクエリは，情報利用者側で接続が発生した日時，問合せ対象ドメイン名，ドメイン名に対応する IP アドレスの 3 つであり，接続元識別子は必要としない．

(3) ドメインの良性・悪性判定

図 4.3 にドメインの異常検知アルゴリズムの概要を示す．アルゴリズムは，接続に関する情報（接続日時，ドメイン，IP アドレス）を入力として与えられ，不審度を出力する．ドメイン悪性判定技術の中には WHOIS 情報やレジストリ情報を問い合わせるためにインターネットへの接続が必要なものも多いが[36] [37] ，本アルゴリズムは追加情報を取得するための外部接続を必要としない．これにより本項(1)の要件(c)が満たされる．本アルゴリズムは「頻度ベースフィルタ」,「特徴ベースフィルタ」の 2 段階から構成される．

頻度ベースフィルタは，情報利用者側組織に於いて，問合せ対象ドメインへの接続が過去に発生した頻度を基に，非悪性である確度が高いドメインをフィルタリングする処理である．たとえば，問合せ対象ドメインに対して，当該組織内の多数のユーザから過去何度も接続があるならば，当該ドメインは良性である確度が高いと判断できる．同様に，悪性ドメインの生存期間は一般的に短いことから，古くから接続していた履歴があるドメインは良性である確度が高いと判断することができる．問い合わせ内容がこのフィルタに合致した場合，アルゴリズムは不審度スコア=0 として良性判定を返して処理を終了する．一方，問合せ内容がフィルタに合致しない場合は，次の特徴ベースフィルタに処理が移る．

特徴ベースフィルタは，問合せ対象接続の特徴（ドメイン名，ドメイン階層，IP アドレス，国番号・ASN，接続日時等）が，問合せ先組織での接続履歴と比べてどの程度乖離しているのかを基に，不審度スコアを算出する．不審度スコア算出は，(i) 近傍性，(ii) カテゴリ類似性，(iii) ドメイン名類似度という 3 つの観点を基に行う．(i) 近傍性では，(A) 問合

せ対象ドメインと 3rd level ドメインが一致する別ドメインに接続したことがある自組織内ユーザ数，(B) 問合せ対象 IP アドレスと/24 で一致するアドレスに接続したことがある自組織内ユーザ数，を基に近傍性を算出する．(ii) カテゴリ類似性の観点では，過去接続の各特徴（接続した時刻や接続ドメインのトップレベルドメイン等）をカテゴリとして分け，問合せ対象接続がどのカテゴリに属するかにより，接続の類似性を算出する．(iii) ドメイン名類似度の観点では，ドメイン名の n-gram を基に，自組織が過去に接続したことがあるドメイン名との類似性を算出する．

(4) 疑似コード

次に，リスト 1 に示す疑似コードを用いてアルゴリズムの詳細について説明する．アルゴリズムの入力には，先述のドメイン(q_domain)，IP アドレス(q_IP)，接続日時(q_time)以外に，Country Code(q_CC), Autonomous System Number (q_ASN)がある．q_CC, q_ASN は q_IP を基に解決することができる．具体的なライブラリ名は後述する．それ以外にも，アルゴリズムで用いる閾値として，TH_hosts, TH_day, TH_close, w_closeness, w_fitness, w_normality がある．w_closeness, w_fitness, w_normality は特徴ベースフィルタの各観点の重みづけを示すものであり合計値は 1.0 となる．Dataset はモデル作成に使用する，情報提供側組織での接続ログを指す．また N は後述の n-gram で用いる値である．

1 行目～3 行目は頻度ベースフィルタの処理を示す．2 行目にあるように，問合せ対象の q_domain に接続したことがある Dataset 内ユーザ数が TH_hosts を超えるか，q_domain への接続履歴が TH_day より前にある場合，q_domain は良性である確度が高いと判断し，不審度スコア=0.0 を返す（3 行目）．これは，多くのユーザが接続するドメインまたは古くから接続があるドメインは良性である確度が高いという知見からくるものであり，既存研究[36] や過去の研究[34] でも活用されている．分析フェイズで q_domain と Dataset 内ユーザの突合を高速に行うために，分析フェイズで予めドメインごとの接続元ユーザ数や接続日時を記録する必要がある．この時，問合せ先組織のユーザ数や接続ドメイン数によっては多くの記憶領域を要する必要がある．そこで我々は，記憶領域のメモリサイズを一定に保つために，ドメインごとの接続ユーザ数や TH_days より過去に接続があるドメインの一覧を，Count-min Sketch[38] 及び Bloom Filter[39] を用いて管理することにした．Count-min Sketch は key に対応するカウント数を確率的に保持する機構であり，key 数が増えてもメモリサイズは変わらない．同様に Bloom Filter はアイテム一覧を一定のメモリサイズで保持する機構である．これにより，分析ロジックの使用メモリサイズを一定以内に抑えるという本項(1)の要件(a)を解決する．

また，Count-min Sketch，Bloom Filter 共に処理に必要な計算量は key 数やアイテム数に因らず一定であり，本項(1)の要件(b)も満たす．尚，両者の課題として，保持する key 数或いはアイテム数が増えるほど，実際には保持していない key やアイテムを保持していると誤判断する false positive の発生頻度が高くなることが挙げられる．

4 行目～7 行目は，特徴ベースフィルタの中の近傍性に関する処理である．ここでは，

q_domain の 3rd level ドメイン及び q_IP の/24 アドレスへ接続したユーザ数をカウントし，その数に応じて closeness を計算する．ユーザ数≧TH_close のとき closeness=1.0 となる．この処理は，一定数以上のユーザから接続履歴があるドメインと同じ階層下にあるドメイン，或いは接続履歴がある IP アドレスと同サブネット下の IP アドレスは良性である可能性が高いというヒューリスティックに基づいている．モデル作成フェイズでは，Count-min Sketch を使用して 3rd level でのドメイン接続ユーザ数及び/24 での IP アドレス接続ユーザ数をカウントしておく．

8 行目～14 行目は，特徴ベースフィルタの中のカテゴリ類似性に関する処理である．ここでは，IP アドレスの国名・AS 番号，ドメインの TLD・階層数・サブドメインの最大ラベル長，接続日時の日中／夜間・平日／週末，という 7 種類の特徴量を基に，Dataset 内の履歴と比較した，q_IP，q_domain, q_time の類似度 fitness を求める．ここで，各特徴量は numerical ではなく categorical なデータとして処理する．

例えば，日中／夜間の違いや AS 番号の数値に大小関係は無く，カテゴリとして処理するのが好ましい．階層数，ラベル長は数値として扱うことも可能だが，今回は階層数が浅い（2 以下）／深い（3 以上），ラベル長が一定値（16）以上／未満，という区分にわけ，カテゴリとして扱うことにする．また，GeoIP[40] を使うことでインターネット接続無しに，IP アドレスから国名・AS 番号を求めることができる．

モデル作成には one-class SVM のような機械学習を用いる方法も考えられるが，演算コストを抑えるために本研究では Weighted density-based outlier detection（WDOD）を用いる[41] ．WDOD は，特徴量ごとに各カテゴリの相対発生頻度を求めることで入力データの類似度を算出する，特徴量間の依存関係を考慮しない，という特性がある．特徴量間の依存関係を考慮しないというのは精度上の欠点になりうるが，本研究のように，他組織上で動作する分析ロジックでは 2 つの理由でそのシンプルさが功を奏する．1 つは WDOD の計算は軽量であることがあり，もう 1 つは Dataset が複数のログに分かれていた場合でも演算を行えることである．例えば，問合せ先組織で使用できるデータセットが<ユーザ ID，接続先ドメイン，接続日時>，<ユーザ ID，接続先 IP アドレス，接続日時>の 2 つログに分かれている場合を想定する．この場合ドメイン名と IP アドレスを一対一に紐づけることはできず，モデル作成フェイズで名前解決をするのは本項(1)の要件(b)(c)に反することになる．しかし WDOD では特徴量ごとに演算を行うため，ドメインと IP アドレスが紐づいていなくとも類似度の算出が可能である．以上，分析に必要な特徴量間の紐づけが不要となることで，本項(1)の要件(b)及び(e)を満たす処理が可能になる．理論上，ドメイン名，IP アドレス，接続日時は別々のログとして与えられたとしても処理を実行することが可能である．モデル作成フェイズでは WDOD のモデル作成を行い，分析フェイズではクエリと WDOD モデルとの類似度を求める．

15 行目～18 行目はドメイン名類似度 normality の算出処理であり，Dataset 内にあるドメインと q_domain との類似度を求める．ドメイン名類似度は，q_domain の各ラベルの n-

gram が，Dataset 内のドメインの n-gram に含まれる頻度を基に算出する．ここで n-gram とは，ある文字列を n 文字ごとに区切ったサブ文字列の集合である．例えば，”hitachi”の 3-gram は”hit”，“ita”，“tac”，“ach”，“chi”の 5 つである．ドメイン名の類似度に着目した既存研究としては[35] [42] などがある．特に[35] は本研究と同じく n-gram を用いている．違いとしては，本研究では類似度の値を 0~1 の間に入るよう正規化するために出現頻度のランクを求めている点がある．モデル作成フェイズでは Dataset 内のドメインの n-gram の抽出及びランク付けを行い，分析フェイズでは q_domain の n-gram のランクを求める．

19 行目～20 行目は，前段で求めた closeness, fitness, normality を基に不審度スコアを算出する．算出に当たっては各値を w_closeness, w_fitness, w_normality で重みづけし加重平均を求める．前述の通り不審度スコアは 0~1 の値を取り，値が大きいほど不審度が高いことを指す．

表 4.1　接続ログのカラム

#	カラム	例
1	接続日時	2020-02-01T12:00:00:00.000
2	接続元の識別子	203.0.113.1
3	接続先サイトのドメイン名	www.example.com
4	ドメイン名に対応する IP アドレス	192.0.2.1

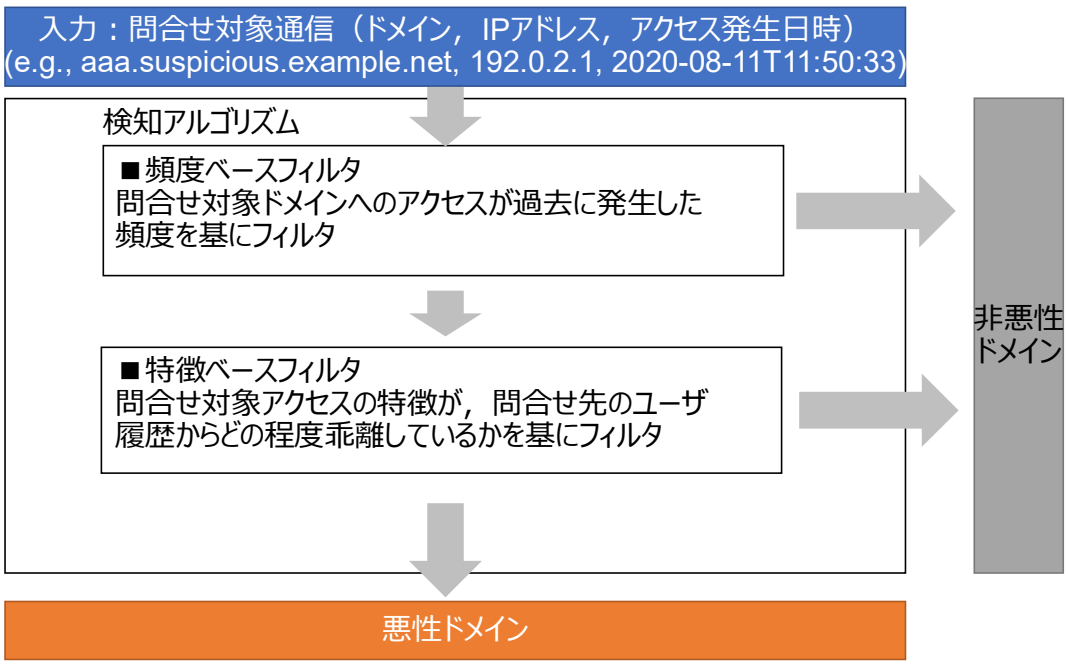

図 4.3　異常検知アルゴリズムの概要

リスト 1　異常検知アルゴリズムの疑似コード

```
1   #q_domain にアクセスしたユーザ数が TH_hosts を超える,
    もしくは TH_day より過去にアクセスがある場合，無条件に score = 0 とする
2   if | {h | h in Dataset and h.hasAccesed(q_domain)}| > TH_hosts
    or | {h | h in Dataset and h.hasAccesedBefore(q_domain,TH_day)}| > 0:
3           return score = 0.0
4   #q_domain の 3rdlevel domain または g_IP/24 にアクセスしたユーザ数に応じて,
    通常アクセス先との近さ closeness を計算.
5   domain3rd = get3rdLevelDomain(q_domain)
6   Closeness = | {h | h in Dataset
    and h.hasAccesed(domain3rd)}| + |{h | h in Dataset and
    h.hasAccessed(g_IP/24)}|
7   Closeness = min(Closeness/TH_close, 1.0)
8   #CC, ASN, TLD, 日中／夜間, weekday/weekend, ドメイン階層数・サブドメイン長か
    ら，カテゴリ観点での fitness（正常度合）を計算
9   domain1st = getTopLevelDomain(q_domain)
10  type_of_hour = getTypeOfHour(q_time)
11  type_of_day = getTypeOfDay(q_time)
12  domain_length_range = log16(maximum length of subdomaind of g_domain)
13  domain_hierarchy_range = log2(# of dot in g_domain)
14  fitness = WDOD(q_CC, q_ASN, domain1st, type_of_hour,
    type_of_day, domain_length_range, domain_hierarchy_range)
15  #2ndlevel domain 以降の各レベルのドメインを対象とした n-gram の rank の観点，及び
    階層数から，ドメイン名に関するデータセットとの類似度を計算する
16  for token in getN-gram(q_domain):
17          name_rank += log2(rank(token)) / log2(# of unique tokens in
    Dataset)
18  normality = 1.0 - name_rank / (# of tokens in g_domain)
19  #ネットワーク上の近さ，及びカテゴリ観点の異常度を基に最終的な outlier score を計算
    する
20  return
    score = 1.0-
    (w_closeness*closeness+w_fitness*fitness+w_normality*normality)
```

4.3.3　AED

AED は未知サイトへの接続が発生した際に，自他組織内のドメイン分析ロジックに不審度の問合せを行う．得られた不審度の内，最小のものが予め定めた閾値を超えた際に，悪性 Web サイトへの接続であると判定し，追加認証を行う．これにより，自組織で不審と判定

された場合でも，他組織の観点で正常であれば正常と見做され，誤検知を削減できる．

なお，AED では追加認証として人間と機械を判別するチューリングテストである Completely Automated Public Turing test to tell Computers and Humans Apart（以降，CAPTCHA）認証を課する．CAPTCHA により，人間による業務上必要な接続は許可しつつ，マルウェア等による機械的な接続は遮断する．CAPTCHA 認証には複数の種類がある．その中で，ユーザが回答することが容易で，広く普及している，画面に表示された文字を入力する方式を採用している．一方で，本方式の CAPTCHA 認証を突破する研究[43] ,[44] も行われている．AED では追加認証の方式は可換であるため，ユーザの利便性低下とのバランスを考慮しながら，より突破困難な方式を採用することも可能である．

4.4 評価

4.4.1 実装

図 4.2 記載のリクエスタ，プラットフォーム，分析ロジックは， Python の Web アプリケーションフレームワーである Flask[45] を用いて作成し，Docker[46] のコンテナで実行した．

表 4.2 に記載した認証 Proxy 及び AED は併せて 1 台のサーバ上に実装した．また，分散 SOC 基板も 1 台のサーバ上に実装した．また，各サーバは仮想マシンで作成した．表 4.2 に各サーバの性能を示す．

表 4.2 各サーバの性能

サーバ用途	CPU コア数	Memory [GB]
認証 Proxy + AED	4	4
インテリジェンス共有プラットフォーム（リクエスタ側）	8	16
インテリジェンス共有プラットフォーム（レスポンダ側）	2	8

4.4.2 データセット

組織 A，組織 B が持つ接続ログを用いて，それぞれの組織の接続傾向である分析モデルを作成した．組織 A は AED を利用する 10 人程度の小規模組織で，組織 B は 1000 人以上の大規模組織のデータを利用した．モデル作成に利用する接続ログには，dns ログと traffic ログがあり，dns ログには，接続日時，接続元の識別子，接続先サイトのドメイン名が，traffic ログには接続日時，接続元の識別子，ドメイン名に対応する IP アドレスが記録されている．表 4.3 に各ログの取得期間，件数及びサイズについて示す．尚，組織 B の dns ロ

グ，traffic ログは，実験環境の都合上送信元の識別子が固定されていない．ドメイン分析ロジックでは，送信元の識別子と接続先サイトの組合せが同じ接続ログは，1 つしかモデル学習に用いない．一方送信元識別子が変わると，別の接続ログとして認識され再度モデル学習に用いられる．このため送信元識別子が固定化されている場合と比較すると結果に違いが生じるが，接続傾向を利用した不審度判定という目的の実現には大きな影響がないと考えられる．

推定器は組織 A の接続が良性か悪性かを判定するのに使われるため，良性 Web サイトデータは，モデル化に使用したデータとは異なる期間に，組織 A が接続したドメインを用いた．使用した良性データの中に悪性データが含まれないことは，確認済みである．また，悪性 Web サイトデータは，PhishTank [47] から収集したサイト群を用いた．悪性データの取得期間については良性データと合わせた．ドメインに悪性データの特徴を有していることが要件であり，仮に別期間（モデル作成に用いた期間等）であったとしても，評価結果の一般性は失われない．尚，本推定器の性能は対象ドメインの生存期間に依存しないため，悪性データセット作成に際してもこの点は考慮しない．それぞれのデータについて表 4.4 に示す．

表 4.3　モデル作成に利用した接続ログ

	取得期間	件数/サイズ（dns ログ）	件数/サイズ（traffic ログ）
組織 A	2021 年 1 月 26 日～2021 年 2 月 5 日（11 日間）	249,712 / 約 20MB	247,735 / 約 50MB
組織 B	2021 年 1 月 26 日～2021 年 2 月 5 日（11 日間）	31,132,557 / 約 3GB	83,136,528 / 約 18GB

表 4.4　テストデータ

	取得期間	件数	情報源
良性サイト群	2021 年 2 月 8 日～2021 年 2 月 12 日（5 日間）	1,930	組織 A が接続したドメイン
悪性サイト群	2021 年 2 月 8 日～2021 年 2 月 12 日（5 日間）	1,620	PhishTank から収集したドメイン

4.4.3　評価項目

評価項目は以下の通りである．評価は従来技術である WL を用いた AED（以降，WL 型 AED），WL 型 AED の機能に加え，WL に含まれない接続に対して，単一組織の分析モデルにより異常検知を行う AED（以降，異常検知型 AED），及び組織 A，B の分析モデルを

用いる異常検知型 AED（以降，分散異常検知型 AED）の 3 つパターンについて行っていく．また，以降それぞれの方式の英語表記を，WL，Outlier，DeOutlier とする．

(1) 悪性サイトの検知精度

表 4.4 に示したテストデータを用い，異常検知型 AED 及び分散異常検知型 AED の悪性 Web サイト検知精度について評価する．

(2) ユーザの利便性

AED を組織 A で 2021 年 2 月 8 日～2021 年 2 月 26 日の 19 日間運用した結果得られたログを元に，CAPTCHA 発生回数を算出し，ユーザの利便性を評価する．尚，AED では N 人以上が CAPTCHA を突破して接続したサイトは WL に追加され，次回接続時には CAPTCHA が発生されない．本実験では，N=1 とした．これは，実験を実施した組織 A の人数が少ないためである．その他，AED は HTTP リクエストに referer ヘッダがある場合は CAPTCHA を発生させない．従って，接続したドメイン数に対して，CAPTCHA 発生回数は少なくなる．

また，異常検知を用いる場合，設定する閾値によって CAPTCHA の発生頻度が異なる．本実験では，全悪性 Web サイトのうち，正しく悪性 Web サイトとして検知した割合である検知率が 99, 95, 90%以上となるように閾値を変更して比較した．尚，テストデータの良性ドメインと悪性ドメインに重複は無いため，WL 型 AED の検知率は 100%である

(3) 処理性能

提案手法はフォワード Proxy と連携しており，ユーザが Web 接続を行うたびに不審度スコアを算出し，不審度が一定値以上の場合に追加認証を発生させる．この処理により Web 接続の処理時間が増えるため，その値が実用範囲内かを評価する．

評価は一番処理時間がかかる分散異常検知型 AED に対して行う．分散異常検知型 AED は処理速度短縮のため，以下の機能を実装している．

① キャッシュ機能

不審度判定に用いる入力項目である，接続が発生した日時と問合せ対象ドメイン名の組み合わせが過去の問合せと同一の場合，キャッシュしたスコアを返す．

② 組織外への問合せ抑制機能

組織 A の分析モデルを用いて不審度判定を行った結果，スコアが 0 の場合組織 B の分析モデルへの問合せを行わない．これは，分散異常検知型 AED では，組織 A，組織 B 双方のスコアのうち値が小さい方を採用するため，スコアの最低値である 0 が取得された時点で組織 B への問合せは意味を成さないためである．

4.4.4 評価結果と考察

評価結果を示す．Receiver Operating Characteristic（以降，ROC）の算出には Scikit-learn[48] を，グラフ作成には Matplotlib[49] を用いた．

(1) 悪性サイトの検知精度

図 4.4 に表 4.4 に記載したデータセットを利用し算出した，異常検知型 AED と分散異

常検知型 AED の ROC 曲線を示す．尚，異常検知型 AED は，組織 A の接続ログを用いた場合（OutlierA）と組織 B の接続ログを用いた場合（OutlierB）を評価した．ROC 曲線下の面積を AUC と呼び, AUC が 1 に近いほど, 識別精度が高いことを示す. 異常検知型 AED の場合 AUC は，組織 A の接続ログを使用した場合 0.957，組織 B の接続ログを使用した場合 0.961 であり，分散異常検知型 AED の場合 AUC は 0.975 であった．このことより，提案手法である分散異常検知型 AED が，他と比較し高い精度で悪性 Web サイトの判定ができることが分かる．また，実験結果より，単純にデータ量を増やすだけでは性能が向上するとは限らないことが示唆される．組織 A+B のデータを統合した場合の性能は 0.975 と大幅に向上し, 単にデータ量が増加しただけであれば, 組織 A のみと組織 B のみの性能差（0.004）と比較して，組織 B のみと組織 A+B の性能差（0.014）がこれほど大きくなることは説明しづらい．この結果から，性能向上の要因は単なるデータ量の増加ではなく，組織 A と組織 B のデータの「多様性」に起因する可能性が高いと考えられる．加えて，組織 A のデータを実運用で増やす手段としては，データ収集期間の延長，組織規模の拡大，大量の Web アクセスなどが考えられるが，後者 2 つは非現実的であり，前者は学習期間の延長が必要となるため，実運用上の制約がある．したがって，実験結果を踏まえると分散 SOC によりデータの多様性を向上させることが有益であることが示唆される．

また，TPR（検知率）が 0.97 付近以上の箇所では，組織 A の接続ログを使用した異常検知型 AED の FPR（誤検知率）が他と比較し低くなることを示している．

(2) ユーザの利便性

図 4.5 に土日祝日を除いた日ごとの CAPTCHA 発生数を示す．また，運用初日は WL 型 AED で大量の誤検知が発生するため，これも除外した．また，表 4.5 に CAPTCHA 発生数の統計データを示す．尚，異常検知型 AED は組織 A のログを用いている．

図より，異常検知型 AED，分散異常検知型 AED のいずれのケースでも，WL 型 AED と比較し CAPTCHA 発生回数を大幅に削減できていることが分かる．WL 型に近い検知率である，検知率 99%となるように設定した場合でも，日々の CAPTCHA 発生回数が 1 日平均で 6 回以下となることが確認できた．これはユーザ一人当たり 1 日 1 回未満となる．先行研究[34] では全てのユーザが 1 日 1 回以下であれば CAPTCHA 発生が許容できるとしており，ユーザが許容できる利便性を維持できていることがわかる．尚，一部のユーザに集中して CAPTCHA が発生している可能性はあるが，その場合その他多くのユーザの CAPTCHA 発生数が少ない事を意味し，多くのユーザが利便性の低下を感じない点で効果があると考えられる．

また，検知率が 95%以下のケースでは，分散異常検知型 AED の方が異常検知型 AED と比較して CAPTCHA 発生回数を低減できていることが分かる．ユーザの利便性を重視したい環境であれば，分散異常検知型 AED を用いることにより比較的高い検知率を維持しながら，誤検知率を削減することが可能である．

(3) 処理性能

図 4.6 に分散異常検知型 AED を利用した場合の処理時間ごとの度数と累積比率を示す．処理時間の測定は，認証 Proxy に記録されている ICAP 処理時間を利用した．これは，認証 Proxy から AED への処理を ICAP[50] を用いて行っており，ICAP 処理時間を計測することで AED に関連する処理時間の測定が可能であると考えたためである．尚，図の X 軸が 9 秒の度数は，9 秒以降の値を全て集計している．

図から 93.6%の処理時間が 3 秒以内であることが分かる．多くのユーザが Web ページのロードに期待している時間が 3 秒以内という調査結果[51] があり，提案システムが実用の範囲内であることが分かる．

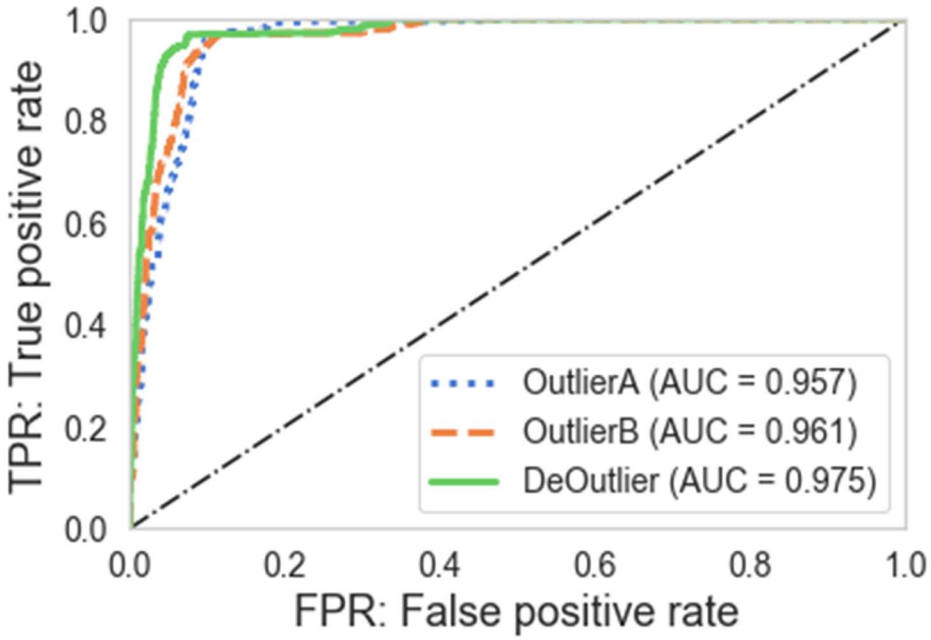

図 4.4　各方式の ROC 曲線

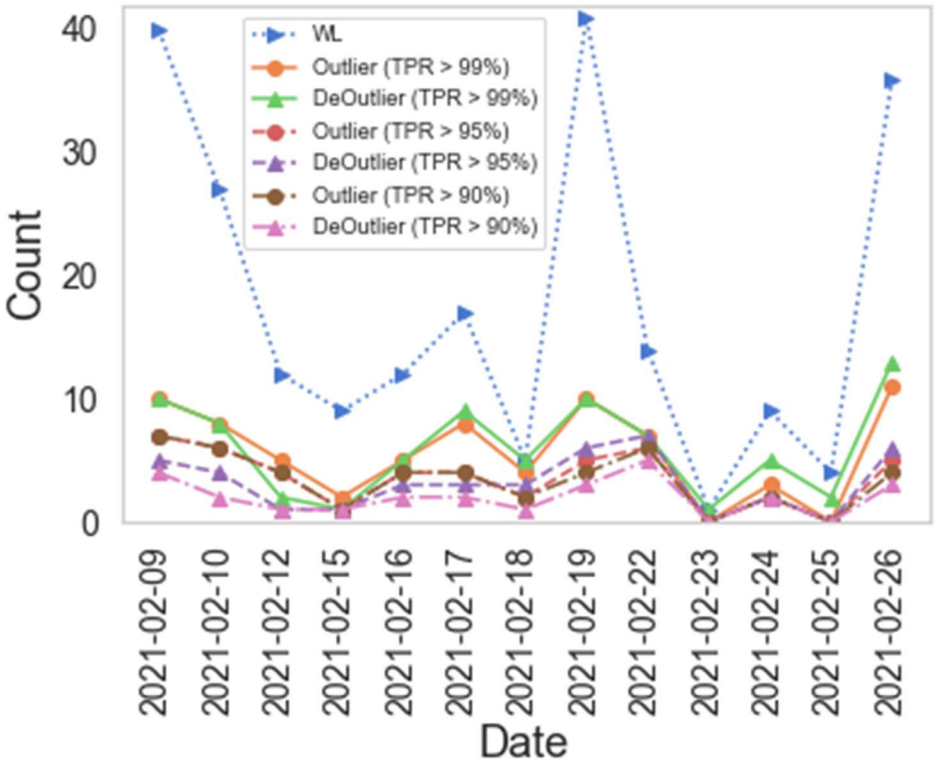

図 4.5　一日毎の CAPTCHA 発生数

表 4.5　CAPTCHA 発生数の比較

AED 種別	合計	平均	最大	最小
WL 型	227	17.462	41	1
異常検知型 (TPR > 99%)	73	5.615	11	0
分散異常検知型 (TPR > 99%)	78	6	13	1
異常検知型 (TPR > 95%)	46	3.538	7	0
分散異常検知型 (TPR > 95%)	41	3.154	7	0
異常検知型 (TPR > 90%)	44	3.385	7	0
分散異常検知型 (TPR > 90%)	26	2	5	0

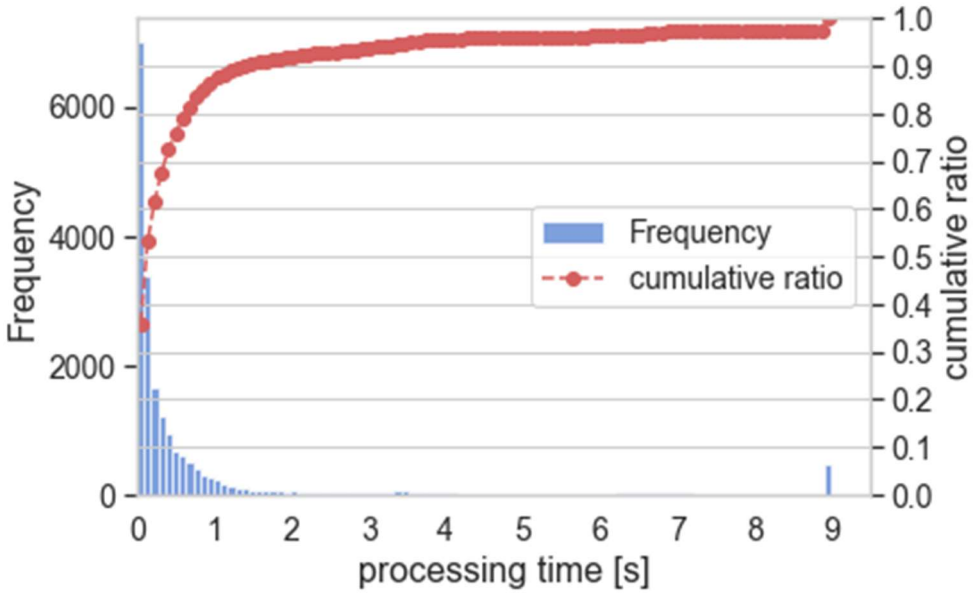

図 4.6　処理時間ごとの度数と累積比率

4.5　関連研究

悪性 Web サイトへの機械的な接続を遮断する方法として，インターネットへの出口に設置した Proxy のユーザ認証機能を利用する方法がある[52] しかし，遠隔操作型マルウェアの中にはユーザ認証機能を突破するものも存在している[53] ．一方提案手法は，悪性 Web サイトへの接続時に機械には突破困難な追加認証を課すことで対応している．

追加認証による対策は，日々の業務の妨げになるという課題が存在する．悪性 Web サイトへの接続を防ぎつつも，安全性の高いサイトへの接続時には CAPTCHA を省略することで業務への影響を軽減する手法が角田らにより提案されている[54] ．また，中鉢ら[55] は複数組織がブロックチェーン技術を用いて WL を共有することにより，AED が抱えていた，小規模組織で WL の拡張が小さく，追加認証が多く発生する課題を解決しようとしている．これらいずれの対策も，未知サイトへ接続する際に必ず追加認証が発生する．提案手法では，接続傾向を元に良性悪性の判定を行うことで，未知サイト接続時の追加認証の発生数を削減する．

未知サイトの良性悪性判定についても様々な研究が行われている．文献[56] は，Convolutional Neural Network を拡張することで，Proxy ログに含まれる宛先 URL の系列から，ドライブバイダウンロード攻撃に関する悪性 URL 系列を検知する手法を提案している．文献[57] は，Bayesian sets と呼ばれる類似要素探索アルゴリズムを用いて，既知の悪性 URL 群と類似した URL を検索する方法を提案している．文献は[58] ，DNS ログからドメインと IP アドレスの関連をグラフ化し，確率伝播アルゴリズムを改良することで，悪性ドメインを高精度で検知する手法を提案している．これらの手法では，悪性 Web サイトの判定に予め用意した悪性 Web サイトのデータを必要とする．一方提案手法は，インターネットへの接続ログと，接続先 IP アドレスの ASN，国番号のみを使用し，悪性 Web サ

イトのデータを必要としない点で異なる．

また，セキュリティ情報の共有に関しては，例えば AlienValt OTX[59] では Indicator of Compromise (IOC)の共有が行われている．一方提案手法では，他組織が保有する接続ログといった，IOC ではない生データの情報を活用する点で異なる．

4.6 本章のまとめ

本章では悪性 Web サイトへの通信遮断を目的として，(1)組織が保有する，サイトへの接続ログを元に接続傾向を分析し，その傾向との乖離度を利用して接続の良性，悪性を判定する，(2)自組織の接続傾向だけでなく，他組織の接続傾向を利用することで判定の精度向上を狙う，(3)判定で誤って悪性と判断した良性 Web サイトへの接続を即時遮断するのではなく，機械には突破困難な追加認証を課し，突破できなかった場合のみ接続を遮断する，という特徴を持つ分散異常検知型 AED というシステムを設計，提案した．

評価では，他組織の接続傾向を用いることにより，不審度検知の精度を向上させることができることを示した．また，従来方式である WL 型の AED と比較して，異常検知方式を取り入れることにより，99%以上の検知率を維持した状態で，日々の CAPTCHA 発生回数を約 3 分の 1 に減らせることを示した．更に，検知率 90, 95%以上という，ある程度検知率の低下を許容したケースでは，自組織のみの接続傾向を用いる場合と比較し，他組織の接続傾向を用いることにより誤検知を削減できることを示した．また，処理性能に関して，処理時間の 93.6%が 3 秒以内であることから，実運用に適応可能だと示した．

今後の課題としては，情報連携組織の拡大による更なる精度向上がある．

組織をまたがったセキュリティ情報共有の重要性は論をまたないが，ログの機微性やデータ量の問題が，共有を活性化するうえでの障害となっている．また共有の効果を定量的に評価した研究は少ない．本研究は上記課題の解決を目指し，共有の効果を定量的に示したものであり，今後のセキュリティ情報共有の活性化の一助になれば幸いである．

第5章　複数組織のインシデント対応記録活用システムの提案と評価

5.1　導入

近年，標的型攻撃に見られるように，攻撃が高度化しており，企業や国家にとって重大な脅威となっている．高度な攻撃に対抗することは難しく，約 65%の組織では，侵害から検知までに 1 時間以上を要しており，また，約 75%の組織は検知から封じ込めまでに 1 時間以上要しているという調査結果がある[60]．また，サイバー攻撃は増加を続けており，ウイルス感染・不正アクセスによる個人情報漏えい・紛失事故が過去最多を記録したという調査結果もある[61]．

対策の一つとして，複数組織で攻撃関連情報等の情報共有を行い，得られた情報を基に事前対策や事後対応が行われている．実際に国や各業界団体により，様々な情報共有の取組みが行われている[9] [11] [62]．このような取り組みは海外でも行われており，業界を跨いだ情報連携による米国での成功例[63] や，EU 加盟国間の連携が検討されている[64]．
具体的な情報共有のシステムとしては，米国では，Automated Indicator Sharing (以下，AIS) [19] と呼ばれる，公共機関と民間機関の間で攻撃時に通信先となった IP アドレスなどの機械可読な攻撃の痕跡である Indicator of Compromise (以下，IoC)や，防御策をリアルタイムで交換する取り組みが行われている．また，AlienValt OTX[59] や MISP[65] といった，IoC の蓄積と共有を支援するオープンソースインフラも存在する．一方で，これら取組やインフラでは，IoC の共有が主であり，情報受信者が行動を起こすのに役立つ十分なコンテキスト情報が含まれていない，といった分析結果がある[66]．

この課題を受け，AIS 2.0 では新たなコンテキストフィールドを追加するなどの改良が行われているが，情報共有のプロセスが手作業に依存しており，効率性や正確性が低いという「システム化されていない」課題が依然として存在する．ここで言う「システム化されていない」とは，脅威インテリジェンスの収集，分析，共有が自動化されておらず，手動対応に頼らざるを得ない状況を指す．

そこで我々は，組織間で生データを共有・分析することで，共有情報にコンテキスト情報を含ませることを提案し，そのシステム化における課題を明らかにし，解決策を検討する．具体的には，各組織のセキュリティオペレータがセキュリティインシデントに対して実施や検討したやり取りの記録（以下，インシデント対応記録）に着目する．着目した理由は本情報にセキュリティオペレータがインシデントに対して実施した試行錯誤や判断等，知見の元となる情報が未加工の状態で記録され，前述のコンテキスト情報として活用が期待できると考えたためである．本情報を複数組織間で共有することの有用性検証と，システムの

開発について述べる.

本研究では先ず, ある組織の実際のインシデント対応記録の実態調査を行った. 具体的には, 有用性を定義し, インシデント対応記録に有用なものが含まれるのかを分析した. 結果, インシデント対応記録の一部が, 他の組織にとっても有用であることを示した.

次に, インシデント対応記録共有のシステム化を行った. 複数組織での攻撃関連情報等の情報共有はサイバー攻撃対策として期待ができる一方で, 共有を阻害する要因が複数存在する. 阻害要因としては, 脅威データの過多, 脅威データの品質, プライバシと法的問題, 連携における相互運用性の問題が挙げられている[67] . 従って, 組織間の情報共有を促進するためには, 大量の情報から高品質なデータ, すなわち他組織において有用である情報を, 機微機密情報の漏洩や法的に問題のない形で, シームレスに連携する必要があると言える. 特にインシデント対応記録は, セキュリティオペレータのやり取りをそのまま活用するため機微機密情報や不要な情報が多いと考えられる. 従って, インシデント対応の記録を共有するうえでの主な問題は,「不要な情報の削除, 整形の手間」「情報漏洩の恐れ」の 2 つであると仮説を立て, 本研究ではこれらの問題を大規模言語モデル（以下, LLM）を用いて解決する自動化技術を設計し評価した.

本研究が設定した研究課題（以下, RQ）は以下の 2 つである.

- RQ. 1: インシデント対応記録にどの程度有用なものが含まれるのか
- RQ. 2: 仮説として挙げたリスク・作業コストの解決は現在の LLM で可能か

また, 本研究の主な貢献は以下の 2 点である.

- インシデント対応記録に有用なものがあることを示した
- インシデント対応記録の活用をシステム化し, LLM の活用方法, 精度, 課題を示した

本章の構成は次のとおりである. まず, 5.2 節で有用性評価を行う. その後, 5.3 節でインシデント対応記録活用システムの開発と評価について述べ, 5.4 節で既存の情報共有手法と本研究の関連性について述べる. 5.5 節で本研究の制限事項を示し, 最後に 5.6 節でまとめを述べる.

5.2 インシデント対応記録の有用性検証

本章では, 5.1 節にて設定した RQ. 1 である, インシデント対応記録にどの程度有用なものが含まれるのか, の問いについて詳述する. 先ず, インシデント対応記録について説明し, 複数組織間でのインシデント対応記録の活用方法について述べる. 次に, 他組織にとって有用なインシデント対応記録について定義する. そして, ある組織の実際のインシデント対応記録を分析し, インシデント対応記録にどの程度有用なものが含まれるのかを評価する.

5.2.1 インシデント記録

インシデント対応記録とは, セキュリティインシデントが発生した際の, セキュリティオペレータ同士によるコミュニケーションの記録である. コミュニケーションには様々な方法があるが, 近年はチャットツールの利用が主流である. また, チャットツール以外にも, プロジェクトのタスク管理等で利用されている, チケット管理システムを用いる場合や, チ

ャットツールとチケット管理システムを併用する場合がある．

インシデント対応記録には，各組織のポリシーや記載者等によって違いがあるものの，一般的には発生したセキュリティインシデントの概要，当該インシデントが誤検知か否かを判断した結果，対処が必要なインシデントだった場合の対処内容，対処をした結果，等が記録される．インシデント対応記録のサンプルを図 5.1 に示す．なお，やまかっこで囲われている内容は，サンプルのためマスクした内容である．インシデント対応記録には各組織のセキュリティオペレータの知見が含まれており，他組織に共有することの価値が期待される．

図 5.1　インシデント対応記録のサンプル

5.2.2　インシデント記録の活用方法

他組織のインシデント対応記録の活用方法として，以下の 2 つがあると考えられる．

(1) 脅威探索

インシデント対応記録には，組織で発生したインシデントの情報が記録される．NIST SP 800-150 では，脅威情報共有の効果として「ある組織での検知が，別の組織での防止につながる」などが挙げられ，情報共有が防御能力の向上に寄与することが明記されている[68]．この指針によれば，情報を受け取る組織は共有された脅威情報を基に，未発生あるいは未検出のインシデントを事前に察知し，被害を未然に防ぐことが可能となるとされる．インターネット上には公開されている様々な脅威情報が存在し，また商用のインテリジェンスサービスも多数存在するが，インシデント対応記録には，他組織で実際に発生しているインシデントに関連する情報がリアルタイムに記録される．従って，インシデント対応記録を共有す

ることで，情報を受け取った組織はより多くの情報をより迅速に得ることにより，サイバー攻撃の予防や早期検知により被害の削減が期待できる．

⑵ 対処支援

NIST SP 800-150 では，共有された情報を利用することで，組織は影響を受けたインフラやシステムを特定し，保護対策を実施し，検出能力を強化し，脅威環境の変化を観察した上で，より効果的にインシデントに対応し，復旧することができると述べられている[68]．インシデント対応記録には，組織で発生したインシデントに関して，当該インシデントが誤検知であったか否かを判断した結果，対処が必要なインシデントだった場合の対処内容，対処をした結果，といった情報が記録される．あるインシデントを対応中の組織が，当該インシデントに関連した情報を他組織から得ることで，インシデントの早期解決による被害の削減や，インシデント対応に係る工数の削減が期待できる．

5.2.3 インシデント記録の有用性定義

セキュリティオペレーションにとって有用である，というのは大きく分けて，対応工数の削減，セキュリティ機器の費用の削減，サイバー攻撃による被害の削減，が考えられる．インシデント対応記録を組織間で共有することにより，5.2.2 項に記載の通り，対応工数の削減や，被害の削減が期待できる．

次に，有用なインシデント対応記録とはどのようなものかについて定義を行う．有用なインシデント対応記録とは，以下の 2 つを満たすものであると考えられる．

- 自組織で発生している，あるいは発生し得る蓋然性や発生時の影響度が高いインシデントの対応記録であること．これは，一般的にリスクとは，その発生確率と発生時の影響度から算出されるもので，リスクの高いインシデントに関する情報は有用であると考えられるためである．
- インシデント対応記録が具体的で且つ情報量が多いこと．これは，具体的且つ情報量が多い場合，情報を受信した側が行動につなげることができ，有用であると考えられるためである．

以降では，ある組織の実データを用いたインシデント対応記録の有用性評価について詳述する．

5.2.4 インシデント記録の有用性評価

インシデント対応記録共有の有用性を評価するために，ある組織の実際のインシデント対応記録がどの程度前述の条件を満たすのかを調査した．

⑴ 手法

以下の手順を実施した．

① スコアリング方法の設計

5.2.3 項で定義した 2 つの条件に関して，評価を行うためのスコアリング項目と基準を定めた．以下に各項目と基準について示す．

・　**影響度**

1 つ目の条件を評価するための項目である．インシデントの影響度を表す．公益性の観点から，より多くの組織で活用できる情報が含まれている方が有用性が高いと考えられる．従って，インシデント対応を実施した組織以外では起こりづらい場合はスコアを低く，他組織でも発生する蓋然性が高いものはスコアを高く設定する．具体的には，組織固有のインシデントの場合は 0 を，Apache Log4j の脆弱性に関するインシデントといった，多くの組織に影響があると考えられる場合は 2 を，どちらかの判断がつかない場合は 1 をスコアとした[i]．本論文で提案する影響度のスコアは，特定のインシデントが広く活用可能な知見を提供するかを評価するための指標として設定しており，組織間での有益な情報共有を促進することを目的としている．一方で，汎用的とはいえないが，少数の組織にのみ関連するインシデントの情報も特定の組織にとっては重要な場合がある．その場合，影響度が低いスコアとなるが，5.3.1 項で述べる基本要件 1 により提案システムでは，受信側の組織が自組織に関連するインシデントの情報を検索することで，影響度の低いインシデント対応記録も活用することが可能である．

・　**ライフサイクル**

2 つ目の条件を評価するための項目である．インシデント対応記録内に，インシデント対応ライフサイクルのどのフェイズの情報を含むかを表す．インシデント対応ライフサイクルとは，インシデント対応プロセスの主なフェイズであり，NIST と SANS が提唱しているモデル[69] [70] が広く利用されている．本研究では，より細かくフェイズ分類をしている，SANS のモデルを参照した．より多くのフェイズの情報が記録されているインシデント対応記録の方が，インシデント対応に必要な情報が充実していると考えられるため，他組織での有用性が高いと考えられる．スコアは，各フェイズ（準備，識別，封じ込め，根絶，回復，教訓）の内，何個のフェイズを含んでいるかで決定する．

・　**具体性**

2 つ目の条件を評価するための項目である．具体的な手順情報が含まれているかどうかを表す．具体的な手順とは，例えばセキュリティ運用業務の効率化や自動化を実現するための技術，あるいはソリューションである Security Orchestration, Automation and Response (以下，SOAR) 製品で機械的に実行できるものを指し，どの機器に対して，どのようなコマンドを，どのような引数で実行し，その結果どのようなことを確認するのか，といった内容が具体的に書いてあるかどうかで判断する．具体的な手順が記載されている方が，他組織での有用性が高いと考えられる．手順が不明瞭な場合は 0 を，攻撃の痕跡を見つけるためのコマンドがある，といった具体的な場合は 2 を，どちらかの判断がつかない場合は 1 をスコアとした．

---

[i] Apache Log4j は，Apache Software Foundation の米国およびその他の国における登録商標または商標です．

② インシデント対応記録の取得

ある組織のインシデント対応の記録を取得する．本検証で利用したデータはチャットアプリの内容をダンプしたもので，セキュリティオペレータが，インシデントが発生した際にスレッドを立て，そこにメッセージを投稿している．スレッドは発生したインシデント毎に立てられ，本研究では，スレッド単位で分析を行う．なお，チャットアプリの利用方法によっては，スレッドを立てずに同一のインシデントに対してやり取りが行われる場合がある．スレッド単位での分析を行うために，同一のインシデントに関する内容と評価者が判断したものは同一スレッドとして扱った．

③ インシデント対応記録の精査

インシデント対応記録には，例えばオペレータの体制変更といったセキュリティインシデントと無関係な内容も記録されている．本研究ではセキュリティインシデントに関する情報を対象としているため，インシデント対応記録からセキュリティインシデントに関連するスレッドを人手で抽出した．

④ スコアリング

手順①で設計したスコアリング方法を用いて，インシデント対応記録のスコアリングを人手で実施した．評価者は報告者を含む職歴 2 年以上のセキュリティの研究者 4 人で実施した．また，評価者の属人性と作業負荷の分担を考慮し，それぞれのスレッドに対して 評価者を 2 人ずつ割り当て，評価を実施した．

⑵ データセット

2020 年 11 月から 2022 年 5 月までの 1 年半分のインシデント対応記録を用いた．対象期間には，1,733 スレッド（9,930 件のメッセージ）が格納されている．5.2.1 項に記載の通り，インシデント対応記録にはセキュリティインシデントと無関係な内容も含まれており，本研究での分析対象外であるため除外した．結果残った 394 スレッド（3,723 件のメッセージ）の内，40 スレッドのインシデント対応記録を対象に手動分析を実施した．

⑶ 結果

回答の結果を表 5.1 に示す．ここで，表中の双方 X という表記は各スレッドの評価者 2 名の双方が X とスコアリングしたことを表す．調査した 40 件のインシデント対応記録の内，影響度の高いインシデントである，と評価者双方が回答したものは 5 件であった．これらの事例は，著名な NW 機器の脆弱性報告からの機器の停止，外部組織からの注意喚起に記載された攻撃情報に対する調査やオペレータの見解，組織内の不審端末の通信先調査，等の事例であった．

また，5 件中 4 件はインシデント対応記録内に具体的な情報が含まれていると評価者双方から回答されていた．一方で，ライフサイクルのフェイズ数は，評価者双方が 1，と回答したものが 1 件，各評価者が 1 と 3，と回答したもの，2 と 3，と回答したもの が 1 件ずつ，そして，1 と 2，と回答したものが 2 件であった． 表 5.2 に 5 件のインシデント対応記録の概要と，評価者による影響度，具体性スコアの判断理由を示す．

(4) 考察

表 5.1 に示す通り，インシデント対応記録のスコアリングは，評価者によって意見が異なるケースが多い．一方で，双方が高いスコアを出したものは，比較的信頼できる内容であると考える．その点に着目すると，手動分析を実施したインシデント対応記録の 5 件/40 件中(12.5%)は，他組織での活用が見込まれる影響度の高いイベントに関するものであった．また，その内 4 件/5 件中(80%)のインシデント対応記録には，インシデント対応をする上で有用な具体的な情報が含まれると回答されている．従って，影響度の高いイベントに関するインシデント記録の個数は少ないが，存在し且つ具体的な情報が記録されていることが確認できた．そのため，情報共有組織が増加すれば，より多くの影響度の高いイベントに関するインシデント記録が得られることが期待できる．従って，組織のインシデント対応記録を他組織に共有することは有用であると考えられる．一方で，他組織がインシデント対応記録を利用してインシデント対応を実施するには，インシデント対応ライフサイクルの各フェイズにおける情報がインシデント対応記録に含まれることが望まれる．しかし，現状のインシデント対応記録には当該情報が不足している．各フェイズにおける情報不足を補う方法については，5.3.1 項の応用要件 1 で述べる．

表 5.1　インシデント対応記録の有用性のスコアリング結果

スコア	影響度 [件]	具体性 [件]	ライフサイクル [件]
双方 0	4	1	0
0 と 1	12	4	0
0 と 2	2	4	0
双方 1	5	6	22
1 と 2	12	18	10
1 と 3	-	-	5
双方 2	5	7	0
2 と 3	-	-	3
双方 3	-	-	0

表 5.2　インシデント対応記録の有用性のスコアリング結果

#	概要	評価者 1 の判断理由（影響度）	評価者 2 の判断理由（影響度）	評価者 1 の判断理由（具体性）	評価者 2 の判断理由（具体性）
1	著名なネットワーク機器の脆弱性を突かれたことによる VPN パスワードの漏洩→該当機器調査（VPN を利用しているか否か，影響等）→ログの保全→機器停止	著名なネットワーク機器の脆弱性に関するため「2」とした.	VPN 経由の侵入は頻繁に発生しており，他組織でも起こりうる事象のため「2」とした.	被害状況の調査に必要な SQL コマンドが記載されていたため「2」とした.	リモート IP アドレスが記載されており，Firewall などに設定することで未然防止が可能になるため「2」とした.
2	SQL injection の調査→被害状況の調査→攻撃内容の調査→外部へ報告	SQL injection は広く一般的に発生する攻撃であるため「2」とした.	SQL injection は頻繁に発生する攻撃であり，他組織でも起こりうるため「2」とした.	攻撃者が利用した具体的なコマンドの内容が記載されていたため「2」とした.	攻撃者の活動の痕跡が記録されているため，フォレンジックなどに利用できる可能性があるため「2」とした.
3	セキュリティ情報提供組織からの攻撃情報共有→担当者の見解追記	実際に攻撃活動を行っている外部ホストに関する情報であるため「2」とした.	標的型攻撃は頻繁に発生しており，他組織（特に同業種）でも起こりうるため「2」とした.	共有された不審なホストのうち，実際に攻撃活動を観測したホストの IP アドレス，ホスト名が記載されていたため「2」とした.	攻撃に悪用されている IP アドレスが記載されており Firewall などに設定することで攻撃未然防止が可能になるため「2」とした.
4	とあるコインマイナー系の不審な IOC の調査→不審通信の特定→不審通信先として登録	実際に不審な活動を行っている外部ホストに関する情報であるため「2」とした.	計算機リソースがコインマイニングに悪用されるのは頻繁に発生しており，他組織でも起こりうるため「2」とした.	不審な接続先に通信をしている端末の調査に必要な SQL コマンドが記載されていたため「2」とした.	コインマイナーに悪用されている通信先に関する情報が記載されており，Firewall などに設定することで未然防止が可能になるため「2」とした.
5	Log4j の脆弱性のオペレータ間での周知→脆弱性の深刻度に関する判断→システム管理者への周知→緩和策の調査	様々な OSS で利用されているライブラリの脆弱性であり，影響度が高いと判断したため「2」とした.	Log4j の脆弱性は，広く利用されている OSS の Apache サーバをはじめとした，Java ベースのアプリケーションで広く使用されているロギングライブラリで見られた脆弱性である．このライブラリは，多くの組織で用いられている可能性が高い．そのため本脆弱性の影響も多くの組織に見られると推測され，スコアは「2」と判断した.	攻撃コードや暫定対処策の情報が記載されており，具体性が高いと判断したため「2」とした.	調査対象，調査対象外とすべきソフトウェア，攻撃者の IP ホストリストが掲載されている Web サイトの URL，が記載されていて，コマンドレベルではないが，ある程度の対応に関する具体性があるためスコアは「1」であると判断した.

5.3　インシデント対応記録活用システム

5.2 節にて，実際の組織のインシデント対応記録を分析し，一部の有用性について示した．本研究では 1 組織を対象に分析を行ったが，今後連携する組織数が増えることにより更に有用なインシデント対応記録が得られ，活用が期待できる．一方で，インシデント対応記録には他組織での影響度が低いものがあり，除外する必要がある．また，情報を受信する側の組織は，対応しているインシデントに関連した情報を取得することが望まれ，情報の選別が必要である．また，インシデント対応記録には，各組織の IT 機器の構成情報やインシデントによる被害内容，個人情報，といった機微機密情報が含まれることが多いため，情報共有の前に機微機密情報の保護が必要となる．

これらの作業を人手で実施するのは，日々の業務で忙殺されているセキュリティオペレータの業務を更に増やすことになる．そこで，LLM を活用し，インシデント対応記録を組織間で自動的に活用するためのシステムを開発した．

LLM を活用した理由は，インシデント対応記録はチャットツールやチケット管理システム上での人間同士のやり取りで，自然言語の情報が多く含まれる非構造化データであり，LLM の活用が期待できるためである．

LLM が，5.1 節に記載した情報共有を阻害する要因であると仮説を立てた「不要な情報の削除，整形の手間」「情報漏洩の恐れ」をどの程度解決できるのかを評価した．

5.3.1　要件定義

5.2.2 項で述べたインシデント対応記録の活用方法や，5.2.4 項で実施した実データの分析から洗い出したシステムの要件を以下に示す．要件は，インシデント対応記録の共有に関連する基本要件と，更にインシデント対応記録を活用するために必要な応用要件に分類し，本研究では基本要件を実装した．

(1)　基本要件

基本要件1．類似インシデント対応記録検索

自組織で対応しているインシデントと関連した他組織のインシデント対応記録を検索する．これにより，自組織に関連が低いインシデントに関する情報を除外し，5.2.2 項の対処支援を実現する．

基本要件2．重要インシデントの選別

インシデント対応記録の中から，影響度の高い重要なインシデント対応記録を選別する．これにより，影響度の低いインシデントに関する情報を除外し，5.2.2 項の脅威探索を実現する．

基本要件3．インシデント対応記録の要約

情報を受信した組織のセキュリティオペレータがインシデント対応記録の全文を確認するのは時間がかかるため，インシデント対応記録から重要な情報を要約して通知する．

基本要件4．機微機密情報の保護

インシデント対応記録には各組織の IT 機器の構成情報やインシデントによる被害内容，

個人情報などの機微機密情報が含まれており，他組織への共有を困難にしている．機微機密情報を秘匿化する作業を人手で行うのは手間がかかるため，本作業を自動化する．

(2) 応用要件

応用要件1. インシデント対応記録の情報補完

5.2.4 項の結果から分かるとおり，実際のインシデント対応記録には特にインシデント対応の各フェイズの情報が不足しているケースが多かった．この原因の一つとして，インシデント対応の場面では，目の前の事象への対応が最優先であり，インシデント対応の記録を残す優先順位が低くなることが考えられる．他には，暗黙的なやり取りが発生し，記録に反映されないこともあると考えられる．従って情報不足を補うためには，自動的な情報の補完やセキュリティオペレータへの入力を促進する機能があるとよい．具体的には，例えば，LLM を用いることにより，以下を実現する．

- STIX[71] や CACAO Security Playbooks[72] 等の規格の項目と比較をして不足している情報を洗い出し，セキュリティオペレータに入力を促進させる
- 入力情報の充分性を評価し，セキュリティオペレータの業務評価指標とし，入力を促進させる
- 上記と同様の方法で不足情報を洗い出し，公開情報等から情報を収集し自動で補完する

応用要件2. インシデント対応記録による自動対処

本システムは他組織のインシデント対応記録をセキュリティオペレータが参照し，インシデント対応に活用することを想定している．一方で，他組織のインシデント対応記録を SOAR の入力として，インシデント対応を自動化することで，更なるインシデントの早期対処や工数削減が期待できる．先ずはログの調査等，業務等に影響を与えるリスクの低い作業の自動化を行いいずれは，不審 Web サイトへのアクセス遮断や感染端末のネットワーク隔離等の対処自動化へと進めていく．これを実現するためには，他組織から受け取った情報の正確性を評価する指標を定義し，正しく評価する必要がある．

5.3.2 実装

5.3.1 項で定義した要件のうち，基本要件を実装したシステムを開発した．分析対象であるインシデント対応記録は，自然言語で書かれている．類似の文章検索，文章の重要度判定，文章の要約，機微機密情報の保護には，LLM の活用が最適であると考えた．そこで，本システムでは，Microsoft の Azure OpenAI と，LLM によるアプリケーションを構築するためのオープンソースフレームワークである LangChain[73] を用いて開発した．図 5.2 に開発システムの概要を示す．5.2.2 項で提案したそれぞれの活用方法について，脅威探索機能（破線矢印），対処支援機能（実線矢印）の流れを以下に示す．[ii]

[ii] Microsoft，Azure は，米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です．

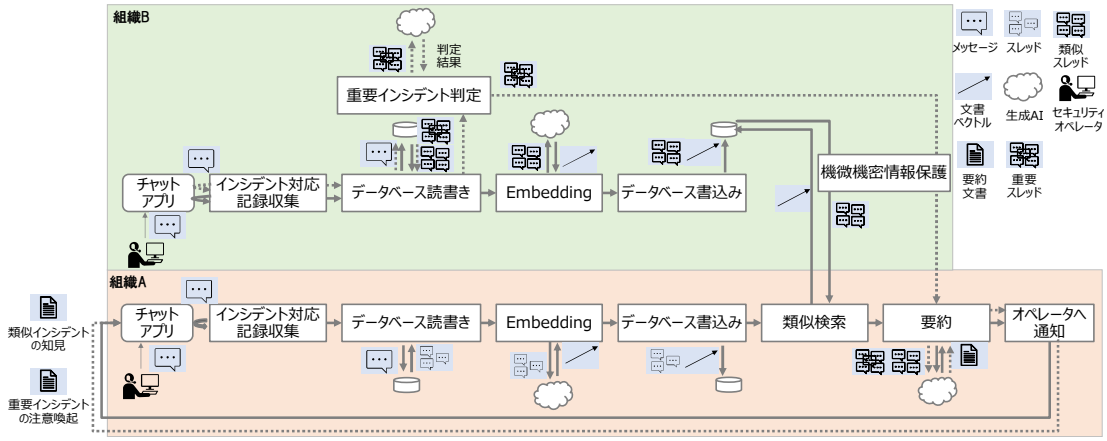

図 5.2　インシデント対応記録活用システムの概要図

(1) 脅威探索

① インシデント対応記録収集

チャットアプリに投稿されたメッセージを定期的に収集し，表 5.3 に示すフォーマットに変換する．チャットアプリにメッセージが投稿された際に外部にメッセージを転送する機能があれば，そちらを用いることにより，リアルタイム性を向上できる．

目的が同じでも，組織によってセキュリティオペレータが使用するツールはそれぞれ異なっている場合がある．本機能により，インシデント対応記録を表 5.3 に示すフォーマットに統一することで，各組織の環境が異なっている場合でも以降の処理を同様に行うことを可能とする．

② データベース読書き

前述の手順で収集したメッセージをデータベースに保存する．そして，収集したメッセージを含むスレッドをデータベースから読み込む．これにより，以降の処理をスレッド単位で行うことを可能とする．

③ 重要インシデント判定

Microsoft の Azure OpenAI の GPT-4 モデルを用いて，スレッドが重要インシデントか否かを判定させる．重要度スコアが高い，と判定された場合以降の処理に進む．

④ 機微機密情報保護

Microsoft の Azure OpenAI の GPT-4 モデルを用いて，スレッド内の機微機密情報を保護させる．

⑤ 要約

Microsoft の Azure OpenAI の GPT-4 モデルを用いて，スレッドの内容を要約する．

⑥ オペレータへ通知

要約した結果を，チャットアプリに書き込む

⑵ **対処支援**

インシデント対応記録収集，データベース読書き，要約，機微機密情報保護，オペレータへの通知は前述と同様の処理を行う．

① Embedding

類似文章を検索可能とするために，スレッドをベクトルに変換する．変換には Microsoft の Azure OpenAI の text-embedding-ada-002 モデルを用いた．また，通常は文章をチャンクとして細かくして Embedding 処理を行うが，本研究では文章をスレッド単位で扱うために，スレッド毎に Embedding を行った．

② **データベース書き込み**

ベクトルをデータベースに格納する．データベースは，Chroma[74] を用いた．

③ **類似検索**

前述の手順でベクトル化をしたデータを用いて，他組織のデータベースから，類似の文章を検索する．類似のスレッドがヒットした場合，以降の機微機密情報保護，要約，オペレータへの通知処理に進む．

表 5.3　メッセージデータのフォーマット

項目	説明
thread_id	メッセージが属するスレッドを一意に識別する値
message_id	メッセージを一意に識別する値
timestamp	メッセージが投稿された時刻
user_id	メッセージを投稿したユーザを一意に識別する値
text	メッセージの本文

5.3.3　脅威探索におけるインシデント対応記録活用システムの性能評価

LLM がインシデント対応記録の重要度を正確に判断でき，不要な情報を自動で削減できるのか，人手による判定と比較することで評価を行った．

⑴ **手法**

以下の手順を実施した．

① **サンプルスレッドの選出**

今回の評価では，5.2.4 項にて評価者が重要度判定をしたある組織のインシデント対応記録 40 スレッドを用いた．

② **スレッドの重要度判定**

インシデント対応記録の重要度をシステムに判定させる．図 5.3 にプロンプトを示す．text の部分には，スレッドに含まれるすべてのメッセージを代入する．

③　システムによる回答の精度評価

重要度判定機能がインシデント対応記録の重要度を適切に判定できているかを，評価者による評価結果と比較することで評価した．前述の通り，重要度スコアには 0, 1, 2 の 3 つの値があるが，人が重要だと考えるインシデント対応記録を他組織に通知することが目的であるため，LLM では，スコアが 2 の場合は重要な通知すべきインシデント対応記録であるとし，スコアが 0, 1 の場合は通知を抑制することとした．また，評価者による評価では，評価者のどちらか一方でも 2 をつけた場合は重要な通知すべきインシデント対応記録であるとし，評価を行った．また，LLM は回答のたびに判定結果が変わる場合があるので，5 回試行して結果の最頻値を用いた．なお，最頻値が複数ある場合は，スコアの高いほうを優先した．

Prompt

```
あなたはセキュリティオペレーションセンターの分析官です。
あなたのタスクは、サイバーインシデント対応記録を読み、属性付けを行うことです。
入力されたサイバーインシデント対応記録が、
インシデントの重要度はどれくらいか（重要度）、
を以下のルールに従い、日本語でJSON形式で回答して下さい。

【出力フォーマット】
{{
    “重要性": {{
        "value": 次に示す例を参考に、インシデントが重要かどうかを記載する。「0: 限定的 (例：重要度の低いインシデント)」、「1: どちらとも言えない」、「2: 影響大 (例：log4j対応等有名な事象への対応や有名な製品や脆弱性に関するインシデント等)」
        "reason": valueを判断した理由を記載する
    }},
}}

【入力のサイバーインシデント対応記録】
{text}
```

図 5.3　プロンプト（重要度判定）

(2) 結果

表 5.4 に LLM の判定結果と評価者による判定結果の混合マトリックスを示す．正解率 72.5%, 適合率 70%, 再現率 74%であった．

表 5.4　重要度判定の混合マトリックス

		LLM	
		重要	重要ではない
評価者	重要	14	5
	重要ではない	6	15

(3) 考察

評価者が重要ではないと判断したが，LLM が重要だと判断したものの例には，インシデントの概要は重要であるが，組織固有の内容が多く他組織での活用が困難なケースや，公開情報を共有しているだけのケースがあった．反対に，評価者が重要であると判断したが，LLM が重要ではないと判断したものの例には，攻撃の兆候だけで具体的な被害が発生していないケースがあった．従って，より精度を高めるためには，LLM が重要だと判定したものに対して，自組織で関連するかどうかを自動判定すること，具体的な被害が発生していなくても，攻撃の兆候が重要かどうかを判定させること，が考えられる．

LLM による自動判定が無く，発生したインシデントを全て通知する場合，本実験のケースでは，情報受信組織は 40 件調査をしなければならない．一方で，LLM による自動判定がある場合，5 件の見逃しが存在するものの，調査対象は半分の 20 件に抑えることができる．LLM による判定で，過検知や見逃しが存在するが，重要度という機械による評価が難しいものの判定を，LLM に事前学習をさせることなくある程度の精度で判定でき，容易な方法で情報受信者側のコストを削減することが可能であることがわかった．

5.3.4 対処支援におけるインシデント対応記録活用システムの性能評価

インシデント対応記録を基に，特定のインシデントに関する回答が生成され，不要な情報を自動で削減できるのかどうかの評価を行った．

(1) 手法

以下の手順を実施した．

① サンプルスレッドの選出

インシデント対応記録から，正解データとしてサンプルスレッドをいくつか選出する．今回の評価では，重要インシデントに関するインシデント対応記録を活用することが目的のため，ある組織のインシデント対応記録（1,733 スレッド）の中から，評価者の内どちらか一方，あるいは双方が重要であると判断した 19 スレッドを対象とした．

② サンプルスレッドに関する問い合わせ

前述の手順で選出したサンプルスレッドと類似のインシデントが別の組織で発生したと仮定して，LLM に問い合わせを行う．図 5.4 にプロンプトを示す．text の部分は，例えば選出したサンプルスレッドが特定の製品の脆弱性に関するインシデント対応記録の場合は，「○○製品の脆弱性について調査をしています」，といった形で，類似のインシデントを調査していることが LLM の入力となるようにした．また，問い合わせに対して何件の類似文章を検索し，当該類似文章を基に回答を生成するのかの設定は，プロンプトのトークン数制限と，適切な文章が検索されるか，のバランスを考慮し，8 件とした．また，実験をする中で，どのような質問に対しても類似度が高く算出されてしまう汎用的な文章が存在することが分かり，当該文章を除外することで回答精度が上がると考え，除外できるような機能を実装した．当該機能を用いて，予め 5 件除外して評価を行った．すなわち，LLM は 1,728 スレッドの情報を基に，質問に対して回答することとなる．

③ システムによる回答の精度評価

類似検索機能が適切な類似インシデント対応記録を検索できているか，検索した類似インシデントの対応記録から適切な回答が作成できたか，を評価する．適切な回答が生成されたかは機械的に測ることができないため，評価者による判断を行った．

Prompt
あなたはセキュリティオペレーションセンターのオペレータです。 下記のインシデントついて、類似のインシデントを探し、概要と、具体的な攻撃方法(コマンドなど)、どのような判断をして、どのような対応をしたのかを答えてください。 概要には、いつ何が起きたのかを要約をしてください。 箇条書きで書いてください。箇条書きの先頭には、・をつけてください。 尚、記載はできるだけ詳しく書いてください。 {text}

図 5.4　プロンプト（類似インシデント対応記録を基に回答）

(2) 結果

表 5.5 に評価結果を示す．評価者によって正否を判定したところ，正しく回答が生成されたのは 19 件の内 4 件で，15 件は正しい回答が得られなかった（精度 21.1%）．また，それぞれの要因を調べるために，どの情報を基にどのような回答が生成されたのかを分類した．結果，正しく回答できているケースの内 1 件は，正解のスレッドを基に回答できていた．また，別の 1 件は，正解と異なるスレッドから回答できていた．その他，正解と異なるスレッドの情報が混在しているが回答として問題なしと評価者が判断したケースが 2 件存在した．一方，正しく回答ができなかったケースの内 10 件は，正解のスレッドを検索できず，回答が分からないと LLM が回答した．また，別の 4 件は，正解のスレッドを検索できず，正解スレッドの内容とは異なる回答が生成された．その他，正解のスレッドを検索できているが，異なる回答が生成されたケースが 1 件存在した．

また，評価者による定性的な評価となってしまうが，LLM による要約は，必要な情報を過不足なく要約できていた．

表 5.5　類似インシデント対応記録を用いた回答の正否

正否	件数	内訳	件数
正しく回答できている	4	正解のスレッドを基に回答	1
		正解と異なるスレッドから回答	1
		正解と異なるスレッドの情報が混在しているが回答として問題なし	2
正しく回答できていない	15	正解のスレッドを検索できず，回答が生成されない	10
		正解のスレッドを検索できず，異なる回答が生成される	4
		正解のスレッドを検索できているが，異なる回答が生成される	1

(3) 考察

表 5.5 に示した通り，正しく回答ができたケースの中には，意図していた正解スレッドではない，異なるスレッドから回答が生成されたケースがあった．このことから，関連したインシデント対応記録が複数のスレッドに記載されていて，他のスレッドから回答が生成されていることがわかる．正しい回答が得られなかった場合のほとんどが，正解であるインシデント対応記録を検索できなかったためであった．このことから，精度を改善するためには，類似インシデントを正確に抽出することが重要である．キーワード検索と文章の類似度検索を組み合わせる等，検索方法のチューニングが今後の課題である．

5.3.5 機微機密情報保護の性能評価

LLM がインシデント対応記録内の機微機密情報を判断でき，自動で保護できるのか，評価を行った．

(1) 手法

以下の手順を実施した．

① サンプルスレッドの選出

無作為に選んだインシデント対応記録 20 スレッドを用いた．

② 機微機密情報保護

インシデント対応記録内の機微機密情報をシステムに保護させる．図 5.5 にプロンプトを示す．text の部分には，スレッドに含まれるすべてのメッセージを代入する．example には，保護したい機微機密情報の例を代入する．

③ システムによる回答の精度評価

今回の評価では，example を表 5.6 のように細かく指定した場合と，何も指定しない場合との両ケースを評価した．機微機密情報の例は，一般的に機微機密情報として扱われているものを可能な限り列挙した．また，構成情報，インシデント情報は各項目の説明も付与して LLM に問い合わせを行った．

Prompt

あなたはセキュリティオペレーションセンターのオペレータです。
以下、自組織で発生したインシデント情報を、他社に情報提供をしようと考えています。
他社に開示すべきではない機微、機密情報を********にマスクしてください。また、何をなぜマスクしたのかを教えてください。
【機微機密情報の例】
{example}
【自組織で発生したインシデント情報】
{text}

図 5.5 プロンプト（機微機密情報の保護）

表 5.6　機微機密情報の例

カテゴリ	項目	説明
個人情報	氏名	NaN
個人情報	住所	NaN
個人情報	電話番号	NaN
個人情報	メールアドレス	NaN
個人情報	生年月日	NaN
個人情報	身分証明書番号（パスポート番号，運転免許証番号など）	NaN
個人情報	社会保障番号や国民保険番号	NaN
個人情報	銀行口座情報（口座番号，支店名，口座名義人など）	NaN
個人情報	クレジットカード情報（カード番号，有効期限，セキュリティコードなど）	NaN
個人情報	ヘルスケア情報（医療記録，保険情報，医療診断など）	NaN
個人情報	人種や民族，宗教，性別，性的指向などの個人の属性に関する情報	NaN
個人情報	就業歴や教育履歴	NaN
個人情報	インターネット上の識別子（IP アドレス，Cookie 識別子など）	NaN
営業機密	製品やサービスの技術情報	NaN
営業機密	企業の戦略や計画	NaN
営業機密	顧客リストや取引先情報	NaN
営業機密	販売戦略や価格設定に関する情報	NaN
営業機密	マーケティング戦略や広告キャンペーンの計画	NaN
営業機密	製品の設計図や特許情報	NaN
営業機密	サプライチェーンや供給者情報	NaN

営業機密	会社の内部手順やプロセスの詳細	NaN
営業機密	研究開発の成果や新製品の情報	NaN
営業機密	財務情報や予算情報	NaN
構成情報	ネットワーク構成情報	ネットワークのトポロジーやルーティング情報，ファイアウォールやセキュリティデバイスの配置情報などが含まれます.
構成情報	システム構成情報	サーバ，ストレージ，およびその他のハードウェアの構成情報，オペレーティングシステムのバージョンやパッチレベル，サービスやアプリケーションの構成情報などが含まれます.
構成情報	セキュリティ設定情報	アクセス制御リスト（ACL)，セキュリティポリシー，暗号化キー，認証情報，およびセキュリティ関連の構成情報が含まれます.
構成情報	データベース構成情報	データベースサーバの構成情報，データベーススキーマやアクセス権の設定情報，バックアップおよびレプリケーション構成情報などが含まれます.
構成情報	クラウドインフラストラクチャの構成情報	クラウドプロバイダーのインフラストラクチャやリソースの構成情報，アクセスキー，およびセキュリティグループの設定情報が含まれます.

構成情報	アプリケーション構成情報	カスタムアプリケーションやサードパーティアプリケーションの構成情報，データストアの接続情報，およびアプリケーションサービスの設定情報が含まれます.
インシデント情報	インシデントの詳細な技術情報	攻撃の手法や侵入経路，脆弱性の詳細な解説など.
インシデント情報	被害の範囲や影響	侵害されたシステムやデータ，被害額，サービス停止時間などの情報.
インシデント情報	認証情報	ユーザ名やパスワード，アクセスキーなどの認証情報の漏洩.
インシデント情報	インシデントの対応情報	インシデント対応の計画や手順，進捗状況，対応チームの情報など.
インシデント情報	ネットワークやシステムの構成情報	ネットワークトポロジーやセキュリティアーキテクチャ，システムの構成情報など.
インシデント情報	内部脅威情報	従業員や元従業員による内部脅威や悪意のある行動に関する情報.
インシデント情報	攻撃者の情報	攻撃者の特定可能な情報や，攻撃者グループの手口や関連情報.
インシデント情報	セキュリティツールやシステムの設定情報	ファイアウォールや侵入検知システムの設定情報，セキュリティツールの構成情報など.
インシデント情報	その他の機微な情報	セキュリティポリシーや手順，内部のセキュリティ上の問題や脆弱性など.

(2) 結果

回答の一例を図 5.6 に示す. 図に示す通り, ユーザ ID, IP アドレス, ドメイン, 組織名, 人名等をマスクできていることが確認できた. また, マスクした箇所を述べていて, 一例を含め全データにおいて, 述べた内容を漏れなくマスクできていた. また, 20 件のスレッドに, どの機微機密情報が含まれていて, 当該機微機密情報が保護されたか否かを示した結果を表 5.7 に示す. タイムスタンプやユーザ ID 等は, マスクする場合としない場合があり, 一方で IP アドレスや人名, ドメインは欠かさずマスクしている. また, URL では URL のドメインのみをマスクしたり, 構成情報に関しては一部の機器のみマスクしたり, 一部マスクする項目も見られた. また, example に機微機密情報を例示した場合としない場合の比較では, IP アドレス, ドメイン, 人名, 組織名, 発生時刻のマスクに関しては結果が変わらなかった. 一方で, 例を示さないほうが URL やパスワードファイル(/etc/passwd)の中身, ブラウザのユーザエージェントなど幅広い情報を積極的にマスクしていた.

(3) 考察

結果に示した通り, 様々な機微機密情報を自動でマスクすることができているが, 共有相手に必要な情報もマスクされてしまっている. 例えば, IP アドレスのマスクをできていたが, 全ての IP アドレスをマスクしてしまうため, 攻撃の痕跡である不審な通信先情報もマスクされてしまう. 内部 IP はマスクして外部 IP をマスクしないなどができるのかどうか検討の余地がある. 一方で, 正規表現を書いたり, 人名のリスト情報を予め用意したりする手間がなく, 個人情報や構成情報をマスクでき, 人手の作業の補助として活用が期待できる.

本機能の具体的な活用方法として 2 つ考えられる.

- **人手の作業補助**

結果に示す通り, LLM がマスクをしたと言っている情報は 100%の精度でマスクをしていた. 従って, 人手で情報を保護する際に, LLM がマスクしたといった項目を除外することで, 人手による負荷の軽減が期待できる. また他の活用方法としては, 人手による作業の作業ミスを減らすことが期待できる. 人と LLM の双方が秘匿すべき情報を見落とすことは確率的にかなり少なくなるため, 情報漏えいのリスクを抑えることが期待できる.

- **完全自動化**

結果に示す通り, 構成情報や組織名等, 完全自動化で保護できない項目が存在する. 一方で, 詳細は第 7 章で述べるが, 情報を提供する相手組織の信頼度によって, どの程度情報を秘匿化するかを変えるような運用があると考えられる. 従って, 信頼度の高い組織に情報を提供する際は, 多少の情報漏えいが発生するものの, 情報共有の早さを優先し, 自動的に処理することも可能である.

回答

【マスク後の情報】
[{'@timestamp': '2024-03-07T20:17:14.632499+09:00', 'text': '文科省からの通達で、¥n<http://polyfill.io|polyfill.io>¥nを使っているという連絡。¥n¥nこのサイトって、どんなサイトか至急調べてよ！¥n¥n¥n（********が管理すると思われるサービス）¥nIPアドレス：********¥nport：443¥nドメイン名：-¥nFQDN名： -¥nasn：********¥nAS組織名：********¥n利用ISP：********', 'user': '********'}, {'@timestamp': '2024-03-07T20:18:21.722289+09:00', 'text': 'すでに連絡済みで、対応するとのことです。', 'user': '********'}]

【なぜマスクしたか】
1. 組織名：他社に我が社や提携先の名前を伝えることは、攻撃者にとって新たな攻撃対象を知るきっかけとなります。実際の組織名を公開することで、不必要なリスクを招きかねません。

2. IPアドレス：当該IPアドレスが攻撃者に知られることは侵入経路や脆弱性を露呈するおそれがあるため、他社に公開すべきではありません。

3. asn：AS番号は組織がインターネットに接続するための重要な情報で、これが漏れることにより攻撃者が新たな攻撃経路を見つける可能性があります。

4. AS組織名：先ほどのasn同様、AS組織名も組織のネットワーク接続に関連する情報であり、攻撃者にとって有用な情報となりうるためマスクします。

5. 利用ISP：利用しているISPを公開することで、攻撃者がISP経由で攻撃を仕掛けるリスクがあるため、マスクします。

6. ユーザーID：ユーザーIDは個々の利用者を特定可能な情報であり、個人情報保護の観点から他社に公開すべきではありません。

図 5.6　機微機密情報保護の回答結果

表 5.7　機微機密情報保護結果

項目	項目を含むスレッド数[件]	example あり		example なし	
		全てマスク[件]	一部マスク[件]	全てマスク[件]	一部マスク[件]
IP アドレス	8	8	0	8	0
ユーザ ID	20	19	0	16	0
ドメイン	3	3	0	3	0
パスワードファイル	1	0	0	1	0
URL	14	6	2	9	3
タイムスタンプ	20	1	0	3	0
AS 番号	1	1	0	0	0
人名	2	2	0	2	0
組織名	7	6	0	6	0
構成情報	5	1	1	3	1
発生時刻	1	1	0	1	0

5.4 関連研究

Toufique ら[75] は LLM を用いたインシデントの原因究明と緩和の有用性評価を行い，LLM がインシデント管理に有用であることを示した．本研究では，異なる組織に蓄積されたインシデント対応記録を対象としている点で異なる．

また，セキュリティ情報の共有に関しては，共有を支援するツールや共有のための規格が存在している．例えば米国では，Automated Indicator Sharing (AIS)[19] と呼ばれる公共機関と民間機関の間で，機械可読な IoC と防御策をリアルタイムで交換する取り組みが行われている．一方で，AIS で共有される情報には，情報の背景となるコンテキスト情報が含まれていないといった分析結果がある[66] ．本研究では，インシデント記録を要約した情報を提供することで，コンテキストを付与した情報の共有が可能となる．

AlienValt OTX[59] では IoC の共有が行われている．MISP[65] は，IoC の蓄積と共有を支援するオープンソースインフラである．これらのインフラでは，IoC の共有が主であるが，本研究では，IoC に加え，インシデント対応記録から得られる知見を共有することを特徴としている．

STIX[71] や IODEF[76] は，インシデントに関連した情報を交換するためのスキーマである．本研究では，セキュリティオペレータが情報を利用することを想定し，自然言語の状態で情報を取り扱ったが，機械に情報を利用させるために，STIX 形式などのデータに変換する方法も考えらえる．

5.5 制限事項

本研究では，チャットシステムから収集したインシデント対応記録を扱っている．チャットシステムには，画像やドキュメントファイル，URL 等のやり取りもされているが，本研究ではチャットに記載されたテキスト情報のみを収集している．また，インシデント対応の記録は，チケット管理システム等，チャットシステム以外のシステムにも記録される場合がある．そのため，本研究では収集しきれていない情報がある．

また，本研究では LLM を利用している．LLM はその性質上処理結果が画一的な物ではない．従って，本研究を再現した際に，異なる結果となることが考えられる．一方で，本システムの運用においては，LLM の結果をセキュリティオペレータが確認して利用すること，また，重要なインシデントを見逃すことにより連携されないという恐れがあるが，前提として情報共有がインシデントレスポンスを補完する目的であることから実用上は問題がないと考える．

5.6 本章のまとめ

本研究では，サイバーインシデントの予防や早期検知，対処の効率化を目的として，組織間でセキュリティオペレータのインシデント対応記録を共有し，活用することを提案した．先ず，インシデント対応記録を組織間で共有することの有用性を，ある組織の実際のデータを用いて評価した．評価の結果，人手による判断により，他組織においても有用であると考えられる情報がインシデント対応記録に含まれていることが明らかとなった．次に，インシ

デント共有の阻害要因である,「不要な情報の削除, 整形の手間」「情報漏洩の恐れ」を LLM で解決できるかどうか，インシデント対応記録を機械的に共有するシステムの設計と開発を行い評価した. 評価の結果, 人手と比較して正解率 72.6%, 適合率 70.0%, 再現率 73.7% で重要なインシデントを自動選別できることがわかった. 一方で, インシデント対応記録を基にした質問への回答に関しては, 正しい回答が得られたケースもあったが, 精度が 21.1% であり更なる改良が必要であることがわかった．また,「情報漏洩の恐れ」を LLM で解決できるかどうかについて，LLM を活用することで，正規表現を使用して機微機密情報を定義したり, 保護すべき情報を具体的に示したりすることなく, 機械的に保護を行うことが可能であることが分かった．しかし，LLM は機微機密情報を完全に保護するわけではなく，場合によっては共有すべき有益な情報まで保護してしまうことがあるため，完全な自動化には限界がある．LLM の活用方法としては，オペレータが機微機密情報を保護する際の漏れや手間を削減することにある.

今後の課題としては, 要件定義を行ったが実装できていない機能の実装や, 類似文章検索にキーワード検索を組み合わせる等さらなる精度の向上がある. また, 情報連携組織の拡大による効果検証を行っていく.

組織を跨ったセキュリティ情報共有の重要性は論をまたないが，各組織のインシデント対応記録を共有する事例は著者が知る限り存在しない．また共有の効果を定量的に評価した研究は少ない. 本研究は上記課題の解決を目指し, 共有の効果を定量的に示したものであり，今後のセキュリティ情報共有の活性化の一助になれば幸いである.

第6章　実証実験に基づく分散 SOC アーキテクチャの設計

6.1　導入

本章では，提案したシステムを検証するため行われた第 3 章～第 5 章の 3 つの実証実験を元に，本提案システムを分散 SOC のアーキテクチャとして総括している．具体的には，3 つの実証実験において共通で必要な機能，プロセスはインフラ部分として共通化し，それぞれ異なる部分は分析ロジックとして個別に作成している．これにより，今後新たな分析要件が発生した際に，容易にかつ効率的に拡張することが可能となる．

6.2　アーキテクチャ

図 6.1 に分散 SOC のシステムアーキテクチャを示す．

第 2 章に記載の通り，本情報共有モデルは，2 つの組織が連携し，データ提供側の環境でデータ利用側の分析ロジックを実行し，その結果のみを取得する仕組みである．この設計により，機微な生データを共有せずに情報活用を実現する．

本情報共有モデルのアーキテクチャ設計においては,以下の 3 つの観点を重視している．

- **データ利用側とデータ提供側のリスク対策**

データ提供側の機微情報漏洩や，データ利用側の不正アクセスなどのリスクを分析し，それぞれに適切な対策を施す．

- **柔軟な API 設計**

システム間の連携を容易にし，標準化されたインターフェースを提供することで，拡張性や互換性を確保する．

- **多様な環境への適用性**

組織ごとに異なる IT 環境に対応し，容易に導入できる仕組みを整える．

以下では，それぞれの観点について詳細に述べる．

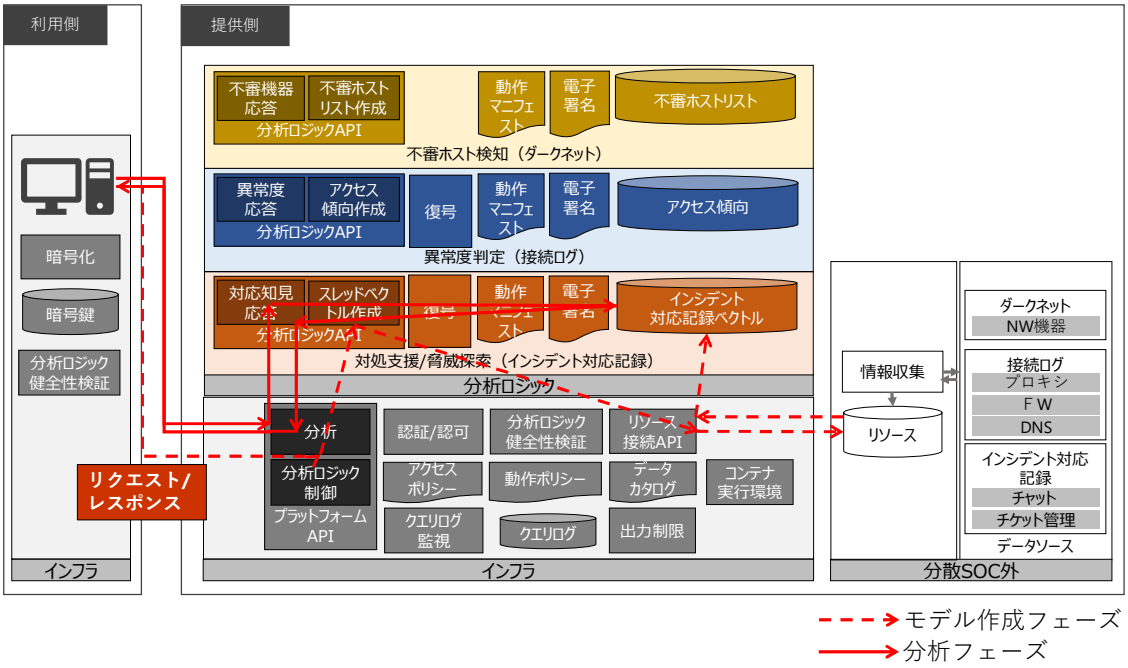

図 6.1　分散型セキュリティオペレーション基盤のアーキテクチャ

6.2.1　データ利用側とデータ提供側のリスク対策

本項では，分析ロジックの遠隔実行機能を実装する際に想定されるセキュリティ脅威を洗い出し，それに基づいたセキュリティ要件を定義し，設計した必要な機能について述べる．

(1)　セキュリティ脅威の洗い出し

本システムは，不特定多数のユーザが参加するものではなく，身元が保証された組織間での利用を前提としている．そのため，以下の 2 つのセキュリティモデルを採用している．

・　Trust-but-verify（信頼するが検証を行う）

利用者同士は信頼関係を持つが，念のため相手の行動を検証できる仕組みを用意する[77]．

・　Honest-but-curious（誠実ではあるが好奇心旺盛）

悪意を持って攻撃することはないものの，容易にアクセスできる情報には興味を持ち，不正に利用する可能性がある[78]．

この前提のもと，以下の脅威を洗い出した．

表 6.1 に情報利用者への脅威を示す．まず，リクエストの内容が平文のまま情報提供側に取得されるリスクがある．次に，情報提供側によるロジックの改ざんが懸念される．さらに，ロジックが他のロジックや情報提供側に不正アクセスされる可能性も考えられる．

表 6.2 には情報提供側への脅威を示す．ロジックが必要以上の情報を取得するリスクがあり，情報提供側のデータが漏洩する可能性がある．また，分散 SOC 基板が未承認の第三者に利用されるリスクも存在する．さらに，ロジックによる情報提供側システムへの不正ア

クセスや，リソースの大量消費によるサービス妨害といった脅威も考慮する必要がある．

(2) セキュリティ要件の定義

上記の脅威を防ぐため，以下のセキュリティ要件を定義する．

- **情報利用者の保護**

リクエスト内容の暗号化が不可欠であり，これにより情報提供側に平文が渡ることを防ぐ．また，分析ロジックの正当性を検証することで，改ざんのリスクを抑える必要がある．さらに，ロジックへの不正アクセスを防止することで，悪意ある第三者による不正な操作を防ぐことが求められる．

- **情報提供側の保護**

ロジックがアクセスできるデータを制限し，必要以上の情報を取得できないようにすることが重要である．また，分析ロジックの外部通信を監視することで，情報の漏洩リスクを低減し，適切なデータ管理を実現する必要がある．また，第三者からの不正利用を防ぐことが必要である．加えて，ロジックのシステムアクセスを監視・制御することで，不正侵入を防ぎ，情報提供側の安全性を確保する．さらに，リソースの利用を制限することで，過剰な負荷によるサービス妨害を未然に防ぐ対策が求められる．

(3) 機能設計

前述それぞれの要件を満たす機能を以下の通りに設計した．

- **リクエスト内容の暗号化**

リクエストの暗号化を行うため，情報利用者は共通鍵を用いてリクエスト内容を「暗号化」し，アプリケーション内で「復号」して分析を行う．この暗号化プロセスにより，情報提供側に平文でリクエスト内容が渡ることを防ぐ．暗号化には Forward Secrecy を保証する DH[14]などのアルゴリズムを使用し，過去のリクエストが復元されるリスクを低減する．しかし，この方法にはいくつかの注意点がある．まず，復号化鍵は情報提供者が管理する分析ロジック内に存在するため，分析ロジックのメモリ解析やリバースエンジニアリングを通じて鍵が漏洩するリスクがある．したがって，リクエスト内容をさらに強固に保護したい場合には，共通の要素のみを取得するプロトコルである Private Set Intersection のような手法を使用すべきである．この方法では，情報提供者側にリクエストを復号されずに分析を行うことができる．ただし，暗号化された状態での分析には，分析手法に制限があることや，処理時間の増加といった欠点も存在する．そのため，実施したい分析の内容やリクエストの秘匿性の必要性を考慮し，適切な手法を選択することが求められる．

- **分析ロジックの正当性検証**

情報利用者は分析ロジックの改ざんを防ぐために，分析ロジックに「電子署名を付与」し，分析ロジックの「健全性検証」機能で署名の検証を行うことで正当性を確認する．これにより，不正な操作や改ざんを防止し，信頼性を確保する．

- **分析ロジックへの不正アクセス防止**

ロジックへの不正アクセスを防ぐために，分析ロジックへの接続をインフラからのみに

制限する．また，分析ロジックの「健全性検証」機能により不正アクセスを検知する仕組みや，脆弱性管理，不正アクセスされた際にイメージから分析ロジックの復元を行う．

- **分析ロジックがアクセスできるデータの制限**

「アクセスポリシー」に利用者組織に応じたアクセス権を設定し，「認可」機能によりデータアクセスの制限を行う．例えば，信頼している組織には全ての情報へのアクセスを許可し，データは生データのままで提供される．一方で，信頼度が低い組織には統計情報のみが提供されるなど，制限がかけられる．信頼度は静的に設定されることもあるが，動的に変化することもあり，動的な信頼度の変化に関する手法については，第 7 章で一例を挙げて検討している．

- **外部通信の監視**

情報利用者のリクエストと分析ロジックのレスポンスを「クエリログ」として保管しておき，必要に応じて情報利用者側にリクエストの復号・開示を要求できるようにする．尚，情報利用者側は，開示要求に応じるために，暗号化に利用した「鍵を保管」しておく必要がある．また，「出力制限」によって，ロジックから外部へ送られるデータ量を制限する．例えば第 4 章で説明したドメインロジックでは，情報提供側が送信する最低限のデータは不審度判定結果の数値のみであり，接続ログのデータ量と比較して非常に小さい．従って，送信可能なデータ量を最低限必要なデータ量に制御することで，接続ログデータの漏洩を抑えることが可能となる．但し，本機能の制限として，少量のデータを多数漏洩させる方法が考えられる．しかし，レスポンスログの回数や内容を確認することで，不正な挙動を検知できる．

- **不正利用防止**

「プラットフォーム API」に「認証機能」を具備し，本システムを利用できる者を分散 SOC 参加者に限定することで対策する．

- **システムアクセスの監視・制御**

分析ロジック「健全性検証」機能と，予め情報提供側が設定した「動作ポリシー」と情報利用者が設定した分析ロジックの「動作マニフェスト」との比較を用いて，情報提供側のシステムで不正侵入を検知する仕組みや，脆弱性管理，不正アクセスされた際にイメージから復元することで対策する．

- **リソース利用の制限**

分析ロジック「健全性検証」機能と，予め情報提供側が設定した「動作ポリシー」と情報利用者が設定した分析ロジックの「動作マニフェスト」との比較を用いて，ロジックが使用する CPU やメモリ，接続回数などのリソース消費を制限し，過剰な負荷によるサービス妨害を防ぐ．これにより，システム全体の安定性を保つことができる．

表 6.1　情報利用者への脅威

#	分類	脅威
1	機密性	リクエスト内容の漏洩
2	完全性	情報提供側による分析ロジックの改竄
3	完全性	他の分析ロジックが不正に情報利用者の分析ロジックに接続する

表 6.2　情報提供側への脅威

#	分類	脅威
1	機密性	分析ロジックによる必要以上のインテリジェンス取得
2	機密性	分析ロジックによる情報提供側のデータ漏洩
3	機密性	第 3 者によるシステムの利用
4	完全性	分析ロジックによる情報提供側のシステムへの不正接続
5	可用性	分析ロジックによるリソース（CPU，メモリ，DB 接続）大量消費

6.2.2　柔軟な API 設計

本システムでは，異なるシステム間の連携を容易にし，標準化されたインターフェースを提供することで，拡張性や互換性を確保するために，柔軟な API 設計を行った．具体的には，(1) プラットフォーム API，(2) 分析ロジック API，(3) リソース接続 API の 3 種類の API を設計している．

(1) 柔軟な API 設計の要件

本システムにおける API 設計では，以下の要件を満たす必要がある．

- **標準化されたインターフェースの提供**

異なるシステムや分析ロジックとの相互運用性を確保するため，統一された API 仕様を策定する．また，RESTful API の原則に従い，直感的かつシンプルな設計を採用する．

- **拡張性の確保**

新たな分析ロジックの追加や，データ構造の変更に柔軟に対応できるように，バージョン管理を導入する．また，API のエンドポイント設計において，拡張可能なパラメータやメタデータを活用する．

- **セキュリティの確保**

全てのリクエストはインフラを経由し，6.2.1 項で詳述したセキュリティモデルに基づいて適切なアクセス制御やクエリの監視等を実施する．また，API トークンを用いた認証・認可の仕組みを導入し，アクセス制御を強化する．

- **処理の効率性**

分析ロジック API においては，データ構造を事前に作成することで，分析リクエストへ

の応答時間を短縮する．

(2) API の機能設計

上記の要件を満たすため，以下のように各 API を設計した．

- **標準化されたインターフェースの提供**

  本システムでは，プラットフォーム API，分析ロジック API，リソースアクセス API の 3 つを用いる．

  - **プラットフォーム API**

    システムの分析ロジックの制御を行うための機能を提供する．この API は，分析ロジックの状態表示や，インストール・起動，停止・アンインストールといった操作を可能にする．また，分析の実行機能を備えており，情報利用者がプラットフォーム API を通じて分析を実行することができる．全てのリクエストはプラットフォーム API に集約され，適切な分析ロジックへと転送されるため，効率性と網羅性が向上する．

  - **分析ロジック API**

    分析に関連する処理を担う API である．この API は，モデル作成フェイズを提供し，事前にデータ構造を構築しておくことで，分析リクエストに迅速に応答できるようにする．作成されたデータ構造は，情報提供側に保存され，次の分析フェイズで活用される．分析フェイズでは，モデル作成フェイズで作成されたデータ構造を用いて，効率的な分析処理が行われる．

  - **リソース接続 API**

    各分析ロジックが情報提供側のデータソースにアクセスするための統一的なインターフェースを提供する．この API は，データアクセスを統制し，アクセスポリシーに基づいて適切なデータ提供が行われることを保証する．

- **拡張性の確保**

API の拡張性を確保するために，バージョン管理を導入した．これにより，新しい機能やロジックの追加，データ構造の変更が発生しても，既存のシステムとの互換性を保ちながらシステムを進化させることが可能である．さらに，API のエンドポイント設計において，拡張可能なパラメータやメタデータを活用し，将来的な要件変更に柔軟に対応できるようにしている．これにより，システム全体の進化がスムーズに行える設計が実現されている．

- **セキュリティの確保**

セキュリティを確保するため，全ての API リクエストがプラットフォーム API を経由して行われるように設計し，6.2.1 項で詳述したセキュリティモデルに基づいて，適切なアクセス制御を実施している．加えて，API トークンを使用した認証・認可の仕組みを導入し，API アクセスを認証・認可したユーザのみに制限することで，セキュリティを強化している．このアクセス制御により，権限外のアクセスや不正な利用を防止し，システム全体のセキュリティが強化されている．

・　**処理の効率性**

処理の効率性を向上させるため，前述の通り「分析ロジック API」において，データ構造を事前に作成する設計を採用した．これにより，分析リクエストが来た際に迅速な応答を行うことができ，全体的な処理時間を短縮することができる．

本設計により，拡張性，互換性，効率性，セキュリティを備えた柔軟な API の構築が可能となった．これにより，システムの長期的な運用においても，新たなロジックやデータ要件の追加に対応しやすい設計が実現されている．

6.2.3　多様な環境への適用性

(1)　多様な環境への適用性の要件

本システムは，異なる組織や環境においてもスムーズに運用できることが求められる．具体的には，以下の要件を満たす必要がある．

・　**異なるデータベース環境に対応すること**

各組織が利用するデータベース（図 6.1 のリソース部分）の種類や構成は異なるため，統一的なアクセスを提供できることが必要である．

・　**異なる分析環境に対応すること**

異なる環境においても，分析ロジックが安定的に動作するよう，環境の違いを吸収できる仕組みが求められる．

(2)　機能設計

上記の要件に基づき，以下の機能を設計する．

・　**異なるデータベース環境に対応すること**

異なるデータベース環境に対応するため「データカタログ」を設定する．この設定は導入する環境ごとに異なるデータベースの種類やテーブル，スキーマに対応するためのものである．情報提供側はデータカタログに自組織のデータベース環境を登録することにより，リソース接続 API がどのようなデータベースに対しても統一的なインターフェースでアクセスできるようになる．この設計により，異なるデータベース間での一貫したアクセスが可能となり，リソースの違いを吸収できる．

・　**異なる分析環境に対応すること**

異なる分析環境に対応するための分析ロジックのコンテナ化機能を導入する．分析ロジックをコンテナ化し情報提供側の「コンテナ実行環境」で実行することで，異なる環境間での互換性を保ちつつ，迅速かつ効率的な運用が可能となる．コンテナにより，依存関係や環境設定に影響されることなく，さまざまなシステム環境で同一の分析ロジックを実行できる．このアプローチにより，異なるシステム上でも安定した動作が保証される．

6.2.4　実装パターン

提案した設計は，1.3.1 項で述べたハブ・スポーク型および N 対 N 型双方の構成で実装可能である．図 6.1 では N 対 N 型のケースで記載をしたが，分散 SOC の機能を中央集権的な組織やサービスに持たせることも可能である．N 対 N 型の場合は，情報のやり取りが

個別組織同士となるため，分散 SOC 参加組織数が増えると，通信の組み合わせが増えることにより処理の複雑化や性能の劣化等が生じる可能性がある．一方で，各組織が持つデータやそれを基に作成したデータモデルが組織外に置かれることがないため，情報管理の観点では優位性がある．反対にハブ・スポーク型の場合は，情報処理を行う箇所が 1 か所に集中するため，処理が単純になることと，処理の効率化が期待できる．一方で，各組織が持つデータ，あるいはそれを基にしたデータモデルを中央に預ける必要があり，分散 SOC 参加組織が中央を信頼する必要がある．

6.3　本章のまとめ

本章では，分散 SOC アーキテクチャの設計について，実証実験に基づき総括した．

本システムの設計においては，以下の 3 つの観点を重視した．

(1) データ利用側とデータ提供側のリスク対策

機微情報漏洩や不正アクセスのリスクを考慮し，それぞれに適切なセキュリティ対策を施した．

(2) 柔軟な API 設計

システム間の連携を容易にし，標準化されたインターフェースを提供することで，拡張性や互換性を確保した．

(3) 多様な環境への適用性

組織ごとに異なる IT 環境に対応し，容易に導入できる仕組みを整えた．

また，リスク対策として，暗号化技術を活用したリクエスト内容の秘匿化，分析ロジックの正当性検証，不正アクセスの防止，アクセス制御の強化，外部通信の監視と制限などの機能を設計した．これにより，安全かつ効率的な情報共有を実現するアーキテクチャを確立した．

この設計を採用することで，今後新たな分析を行う際にも容易に拡張可能であり，継続的な機能追加や環境適応がスムーズに行える．

第7章　組織間信頼度による匿名化処理を用いたインシデントチケット共有システムの開発

7.1　導入

本章では，各組織がインシデント管理に用いるインシデントチケット（以下チケット）を，情報提供者と利用者間の信頼度を用いたアクセスコントロールを用いて共有するシステムを開発したので報告する．

7.2　関連技術

本章のように，複数組織における情報共有についての研究が行われている．S. N. Khajeddinら[79] は，情報共有の課題とさまざまな信頼モデルについて述べている．しかし，情報共有インフラの実装については述べておらず，本章はこの点で異なる．

また，本章のようにセキュリティインシデントに関する情報を複数組織で共有することは，既に行われている．例えば，公益法人である ISAC では，参加する各組織が共有するに値するインシデント情報を専用のポータルを用いて ISAC に共有し，これらの情報をまとめたものを定期的に発信している[80] ．しかし，この手法では，個人情報等の機微情報や，社外秘情報等の機密情報が完全に排除された情報しか共有できない．また，情報共有に入力等の手間がかかるために，共有されるまでに時間がかかる．

7.3　提案手法

サイバー攻撃者はほとんど同じ技術，戦術，手順を利用して，さまざまな企業やインフラを攻撃することが多い[79] ．そのため，インシデントの発生要因，検知・対処情報等を共有することで，潜在的脅威を特定することができる．また，これらの情報をインシデント発生時に迅速に共有することで，複数組織で連携してインシデント対処を行うことができる．迅速な情報共有を実現する一つの手段として，情報共有の自動化が挙げられる．しかし，情報には機微，機密情報が含まれているため，外部へ共有する前に自動化が困難な匿名化が必要である．

本章では上記したインシデントに関する情報を各組織間の信頼度に応じて秘匿化し，信頼度が低ければ情報によっては匿名化を施して共有することで，情報共有の自動化を可能にするシステムを開発した．共有する情報の媒体は，組織がインシデント管理に用いるチケットを用いた.図 7.1 にシステムの概要図を示す ．

図 7.1 では，組織 A が持つチケットを，組織 B，組織 C へ共有する場合を表している．まず，組織 A のチケットが更新されると，チケット情報が組織 A の持つクライアント装置

によって，3 段階のレベルに秘匿化される．秘匿化された情報群は中継装置に送られる．中継装置は情報を送信した組織 A と，組織 B，組織 C 間の信頼度を信頼度 DB から取得し，これに応じてどの秘匿レベルの情報を送信するか判別し，送信する．例えば，組織 A から組織 B への信頼度が低い場合は秘匿レベルが高い情報が送信され，組織 A から組織 C への信頼度が高い場合は秘匿レベルが低い情報が送信される．情報を共有された組織は，共有された情報を閲覧し，自組織のインテリジェンスリソース内に関連する情報があった場合これを返す．図 7.1 の場合，組織 C が関連するインテリジェンスを持っていたため，関連する情報を自組織内のクライアント装置で秘匿化し，中継装置を介して組織 A にフィードバックする．これにより，組織 A がインシデント対応中のチケットを送信していた場合，フィードバックされたインテリジェンスを用いて，迅速な対処を行うことが可能になる．

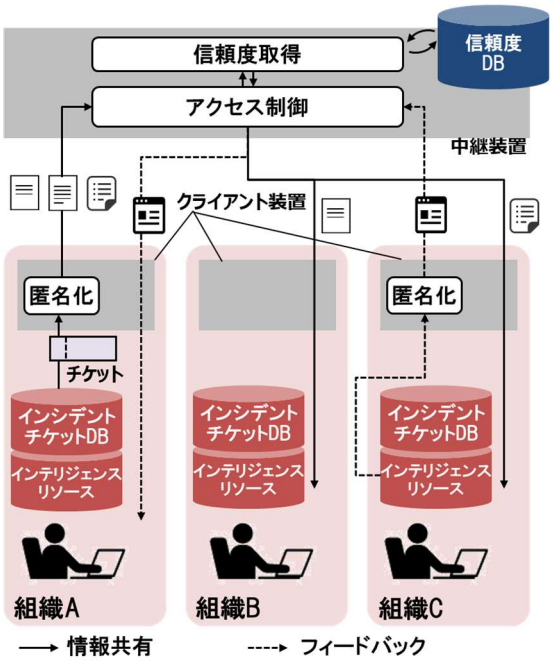

図 7.1　提案手法の概要図

次に，信頼度について述べる．信頼度は情報共有時に秘匿レベルを決定するために用いられる値である．信頼度 DB に記録されている信頼度は，組織間の信頼関係や取引実績によって定まるシステム加入時の信頼度の初期値に，情報共有時の実績を反映した値である．例えば，フィードバックされたインテリジェンスがインシデント対処の役に立った場合，その評価を中継装置に送信し，中継装置は信頼度に一定値を加算する．

本システムは信頼度が低かった場合，情報共有における情報提供者の恩恵を評価し，これを参照して秘匿レベルを定める．例えば，この値が十分に高ければ，信頼度が低かったとしても秘匿レベルが低く，匿名化されていない情報を送信する．この値は，共有するチケット情報と，情報共有先組織が保持する情報の類似度から求められる．類似度が高ければ，チケット情報を共有した後フィードバックされる情報が，情報共有する組織に対して恩恵のあ

る情報である可能性が高くなる．したがって，共有する情報と共有先の組織が持つ情報の類似度を用いることで，情報共有における恩恵を推定することができる．

本章で述べるシステムでは，秘匿情報検索技術を用いて類似度を算出する．これは，情報を公開せずに類似度を評価する必要があるためである．チケット情報から検索クエリを作成し，これをクライアント装置に送信する．クライアント装置は，これを受信すると秘匿情報検索を実行し，類似度だけを返す．

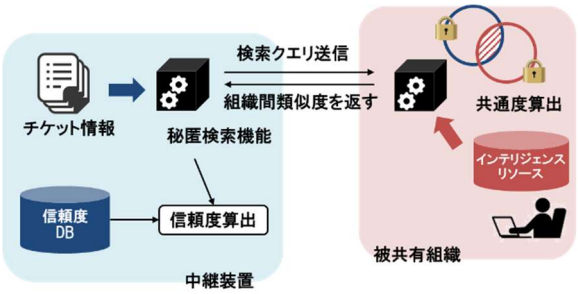

図 7.2　情報共有時の信頼度算出処理

7.4　本章のまとめ

本方式で実装したインシデントチケット共有システムを用いて，組織 B の SOC が持つインシデントチケットを組織 A へ共有した．今後は，チケット情報の共有にかかる時間を測定し，一般的な SOC が外部組織への定期連絡にかける時間の 10%である 180 秒と比較して，充分短いかどうか評価する．また，共有されたチケット情報をもとに対処の自動化を目指す．

第8章　結論

8.1　研究のまとめ

本研究では，サイバーセキュリティ分野における情報共有の課題を明らかにした．具体的には，既存の情報共有の取組には，「情報の網羅性の不足」，「リアルタイム性の欠如」の課題が存在すると整理をした．これらの課題を解決する方法として，新たな分散型セキュリティオペレーションモデル（分散 SOC）を提案した．分散 SOC の核心は，各組織の SOC が緊密に連携し，従来の枠組みを超えた情報共有と協調的防御を実現することである．本提案モデルの有益性を示すために，3 つの実証実験を実施した．

先ず，複数組織間でのダークネット観測データを共有することの効果を定量評価した．実証実験の結果，自組織の観測点を増やすよりも，他組織の観測点を混ぜることにより，より多くの不審なホストを観測できることを示した．

次に，各組織が持つ Web 接続ログを共有することで，悪性 Web サイトの判定精度を向上させる実証を行った．こちらのケースでも，組織間での協力がセキュリティ対策の有効性を向上させることを実証した．具体的には，悪性 Web サイト対策では他組織の接続傾向を活用することで，誤検知率を低下させると同時にユーザビリティの向上を実現した．

さらに，インシデント対応記録の共有に関する実証では，機械学習技術を活用することで手動の負担を軽減し，共有プロセスを効率化する手法を検討した．評価の結果，有用なインシデント情報の自動選別や共有時の機微機密情報の保護を LLM を用いてどの程度の精度で実行できるのかを明らかにした．

そして，提案したシステムを検証するため行われた 3 つの実証実験を元に，本提案システムを分散 SOC のアーキテクチャとして総括した．様々なパターンの実証の中で，共通する部分を分散 SOC インフラとして設計し，今後新たな分析を実行する際に拡張しやすいようにした．

最後に，分散 SOC アーキテクチャの中で特に実装が困難である，相手組織の信頼度や連携により期待できる自組織の利益を用いたアクセス制御に関して，一つの方法を開発した．具体的には，情報共有を行うことの効果が期待できる場合に，情報共有を促進する仕組みについて開発をした．

以上を通じて，本研究は，サイバーセキュリティ分野における情報共有の意義と効果を定量的に示すとともに，その実現可能性を技術的および運用的に検証した．

本研究の成果は令和 6 年 7 月に行われた政府の「サイバー安全保障分野での対応能力の向上に向けた有識者会議」[81] において示された課題，すなわち「重要インフラ分野を含め，民間事業者がサイバー攻撃を受けた場合の政府への情報共有や，政府からの対処調整，支援等の取組強化」に本研究は貢献しうる．

現在，重要インフラのセキュリティオペレーションは各事業者が独立して運用し，インシデント情報は縦割りの構造の中で分断されている．これに対し，本研究の分散型 SOC は，各組織が緊密に情報共有し，より迅速で協調的な対処を可能にする．本研究で示したアプローチは，内閣官房が示す官民連携の枠組みとも合致し，情報共有と協力体制の強化を促進するものである．

さらに，政府は能動的サイバー防御の導入を進めており，そのための情報収集・分析能力の強化が求められている．本研究のアプローチは，分散された SOC 間での脅威情報の即時共有を可能にし，政府の対応能力向上にも寄与する．加えて，「サイバー攻撃による被害は自然災害と同様に，個社レベルの対応には限界がある」という指摘にも対応し，複数組織が連携した包括的な防御体制の構築を支援する．

以上のことから，本研究は，政府の掲げるサイバー安全保障の方向性と整合しつつ，官民連携による実効性の高いインシデント対応の実現に資する重要な成果を挙げたと考えられる．

8.2 今後の課題

実証に参加した組織が 3 組織であったため，組織の多様性による情報共有の効果に関して，検証できない部分が存在した．今後の課題の一つとして，情報共有組織の拡大が考えられる．また，LLM を用いた情報共有作業の自動化について検証を行ったが，利用方法によっては精度が低いため改善の余地がある．また，利用方法に関して，今回実証した以外にも様々考えられるため，新たな利用方法について検証をしていく必要がある．さらなる技術的改良を進めることで，セキュリティ情報共有の活性化とその持続的な発展に貢献することが期待される．

謝辞

本論文は，著者が慶應義塾大学大学院 政策・メディア研究科 博士後期課程において行った研究の成果をまとめたものです.

本研究において，一貫してご指導・ご助言を賜りました本学 中村修教授に，心より感謝申し上げます. また，本論文を精読し，有益かつ建設的なご議論をくださった本学の植原啓介教授，楠本博之教授，村井純名誉教授にも深く御礼申し上げます.

第 1 章から第 7 章に記載の各システムの検討・実装にあたり，多大なるご協力をいただいた，株式会社日立製作所 サービスシステムイノベーションセンタ セキュリティ・トラスト研究部の重本倫宏部員，川口信隆リーダ主任研究員，植木優輝企画員，同センタの鍛忠司主管研究長，ならびに同社クラウドサービスプラットフォームビジネスユニットの鬼頭哲郎シニアセキュリティスペシャリスト，日立システムズ グローバル&セキュリティサービス事業グループの仲小路博史本部主管，本学の砂原秀樹教授，近藤賢郎特任助教に，深く感謝申し上げます.

また，第 3 章の評価実験では，中部電力株式会社，中部電力パワーグリッド株式会社，株式会社中電シーティーアイより評価データをご提供いただきました. この貴重なご支援に，心より御礼申し上げます.

さらに，第 5 章の評価実験では，日立製作所 セキュリティ・トラスト研究部の皆様に，本来の業務がご多忙の中にもかかわらずご協力いただきました. 関係者の皆様に，深く感謝申し上げます.

最後に，業務と学業を両立する中で，常に私の研究を支え，理解し，励ましてくれた家族，そして日立製作所の同僚の皆様に，心より感謝申し上げます. 本当にありがとうございました.

参考文献

[1] CYBEREDGE GROUP:" 2023 Cyberthreat Defense Report", https://cyberedgegroup.com/wp-content/uploads/2023/04/CyberEdge-2023-CDR-Report-v1.0.pdf (accessed 2025-02)

[2] IPA：“標的型攻撃/新しいタイプの攻撃の実態と対策”, http://www.ipa.go.jp/files/000024542.pdf (参照 2025/02)

[3] SANS:" SANS 2023 Incident Response", https://www.sans.org/white-papers/2023-survey-event-incident-response/（accessed 2025-02)

[4] Fortinet：“サイバーセキュリティスキルギャップレポート 2024 年版”, https://www.fortinet.com/content/dam/fortinet/assets/reports/ja_jp/2024-cybersecurity-skills-gap-report.pdf (参照 2025/02)

[5] 経済産業省：“サイバー攻撃による被害に関する情報共有の促進に向けた検討会最終報告書”, https://www.meti.go.jp/press/2023/03/20240311001/20240311001-2.pdf?utm_source=chatgpt.comr (参照 2025/02)

[6] 田川 義博, 林 紘一郎. サイバーセキュリティのための情報共有と中核機関 のあり方, 情報セキュリティ総合科学 第 9 号, pp.17-44 (2017).

[7] 内閣サイバーセキュリティセンター：”サイバー攻撃被害に係る情報の共有・公表ガイダンス”, https://www.nisc.go.jp/pdf/council/cs/kyogikai/guidance2022_honbun.pdf (参照 2025/02)

[8] JPCERT/CC：”脅威情報共有・活用を巡る 現場の課題”, https://www.nids.mod.go.jp/about_us/topic/pdf/240301_1_02.pdf?utm_source=chatgpt.com (参照 2025/02)

[9] 内閣官房 内閣サイバーセキュリティセンター：“サイバーセキュリティ協議会について”, https://www.nisc.go.jp/pdf/council/cs/kyogikai/kyogikai_gaiyou.pdf?utm_source=chatgpt.com (参照 2025/02)

[10] JPCERT/CC：“JPCERT コーディネーションセンター JPCERT/CC について”, https://www.jpcert.or.jp/about/06_2.html?utm_source=chatgpt.com (参照 2025/02)

[11] IPA：“サイバー情報共有イニシアティブ J-CSIP（ジェイシップ）について”, https://www.ipa.go.jp/security/j-csip/about.html?utm_source=chatgpt.com

（参照 2025/02）
[12] ネットアテスト：“ISAC とは？ 10 分でわかりやすく解説”，
https://www.netattest.com/isac-2024_mkt_tst (参照 2025/02)
[13] 一般社団法人日本シーサート協議会：“日本シーサート協議会とは”，
https://www.nca.gr.jp/outline/about.html (参照 2025/02)
[14] トレンドマイクロ：“攻撃手法から考える防御策 2022 年 11 月に活動再開した EMOTET（エモテット）を既存環境で防ぐ考え方（2023 年 3 月更新）”，
https://www.trendmicro.com/ja_jp/jp-security/22/k/securitytrend-20221114-01.html
[15] CISCO:” Cisco Annual Internet Report (2018–2023) White Paper”,
https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html (accessed 2025-02)
[16] Verizon:” 2021 Data Breach Investigations Report (DBIR)”,
https://www.verizon.com/business/resources/reports/2021-data-breach-investigations-report.pdf (accessed 2025-02)
[17] C. Fachkha and M. Debbabi, “Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization,” IEEE Communications SurveysTutorials, vol. 18, no. 2, 2016, pp. 1197–1227.
[18] A. Pala, J. Zhuang, “Information Sharing in Cybersecurity: A Review” Decision Analysis 16(3), pp. 172-196, https://doi.org/10.1287/deca.2018.0387, 2019.
[19] CISA:” Automated Indicator Sharing (AIS) | CISA”,
https://www.cisa.gov/topics/cyber-threats-and-advisories/information-sharing/automated-indicator-sharing-ais (accessed 2024-02)
[20] Inspector General of the Intelligence Community:” Unclassified Joint Report on the Implementation of the Cybersecurity Information Sharing Act of 2015”,
https://oig.justice.gov/sites/default/files/reports/Unclassified%2020191219_AUD-2019-005-U_Joint%20Report.pdf, (accessed 2021-07)
[21] K. Furumoto, et al. “Analysis of Multiple Darknet Focusing on Outbound Packets and its Application to Malware Analysis” IEEE 2017 Fifth International Symposium on Computing and Networking (CANDAR), pp. 94 - 100, 2017.
[22] E. Cooke, M. Bailey, Z. M. Mao, D. Watson, F. Jahanian, and D. McPherson, “Toward understanding distributedblackhole placement.” 2004 ACM workshop onRapid malcode, pp. 54 – 64, ACM, 2004.
[23] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, “Practical Darknet Measurement” In Proceedings of 2006 Conference on Information Sciences and Systems (CISS ’06), 2006.
[24] F. Soro, I. Drago, M. Trevisan, M. Mellia, J. Ceron, and J. J. Santanna, “Are darknets

all the same? on darknet visibility for security monitoring," 2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), pp. 1–6, IEEE, 2019.

[25] R. Berthier, and M. Cukier, "The deployment of a darknet on anorganization-wide network: An empirical analysis" InHigh AssuranceSystems Engineering Symposium (HASE), pp. 59–68, IEEE, 2008.

[26] D. Pemberton, P. Komisarczuk, and I. Welch, "Internet background radiation arrival density and network telescope sampling strate-gies" In Australasian Telecommunication Networks and ApplicationsConference (ATNAC), pp. 246–252, IEEE, 2007.

[27] Norton:" What Are Malicious Websites?", https://nz.norton.com/internetsecurity-malware-what-are-malicious-websites.html (accessed 2021-02).

[28] Zhao, B.Z.H., Ikram, M., Asghar, H., Kaafar, M.A., Chaabane, A. and Thilakarathna, K.. A decade of malactivity reporting: A retrospective analysis of internet malicious activity blacklists, Proc. 14th ACM Asia Computer Communication and Security (ASIA CCS'19), pp.1-13 (2019).

[29] IPA :"「新しいタイプの攻撃」の対策に向けた設計・運用ガイド 改訂第 2 版", https://www.ipa.go.jp/files/000017308.pdf （参照 2021-02）.

[30] 近藤 賢郎，細川 達己，重本 倫宏，藤井 康広，中村 修．分散 SOC アーキテクチャに基づいた複数組織間におけるセキュリティオペレーションの連携，マルチメディア，分散協調とモバイルシンポジウム 2018 論文集，pp.872-878 (2018).

[31] 西嶋克哉，川口信隆，重本倫宏，近藤賢郎，中村 修：セキュリティ・オペレーションにおける秘匿データ分析システムの提案，マルチメディア，分散協調とモバイルシンポジウム 2020 論文集，pp.284-289 (2020).

[32] 仲小路博史，藤井康広，磯部義明，重本倫宏，鬼頭哲郎，川口信隆，林 直樹，下間直樹，菊池浩明．人間行動を用いた自律進化型防御システムの提案，2016 年暗号と情報セキュリティシンポジウム（SCIS2016），pp.1-8 (2016).

[33] Nakakoji, H., Fujii, Y., Isobe, Y., Shigemoto, T., Kito,T., Hayashi, N. Kawaguchi, N., Shimotsuma, N. and Kikuchi, H.. Proposal and Evaluation of Cyber Defense System Using Blacklist Refined Based on Authentication Results, The 19th International Conference on Network-Based Information Systems (NBiS2016), pp.135-139 (2016).

[34] 重本倫宏，藤井翔太，来間一郎，鬼頭哲郎，仲小路博史，藤井康広，菊池浩明．WL を用いた自律進化型防御システムの開発，情報処理学会論文誌，Vol.59，No.3，pp.1050-1060 (2018).

[35] Zhao, H., Chang, Z., Bao, G., Zeng, X.. Malicious Domain Names Detection

Algorithm Based on N-Gram, Hindawi, pp.1-9 (2019).

[36] Alina, O., Zhou, L., Robin, N., Kevin, B.. MADE: Security Analytics for Enterprise Threat Detection, Proc. of ACM ACSAC'18, pp.124-136 (2018).

[37] Michael, W., Jun, W.. Unsupervised Clustering for Identification of Malicious Domain Campaigns, Proc. of ACM RESEC'18, pp.33-39 (2018).

[38] Graham, C., S. Muthukrishnan. An improved data stream summary: the count-min sketch and its applications, Journal of Algorithms, Volume 55, Issue 1, pp.58-75 (2005)

[39] Bloom, B.H.. Space/time trade-offs in hash coding with allowable errors, Commun. ACM, Vol.13, pp.422-426 (1970).

[40] "GeoIP", available from https://dev.maxmind.com/geoip/ (accessed 2021-02).

[41] Ayman, T., Ali S.H.. Anomaly Detection Methods for Categorical Data: A Review, ACM Computing Surveys, Vol.52, No.38 (2019).

[42] Jayaram, R., David, J.M.. Unsupervised, low latency anomaly detection of algorithmically generated domain names by generative probabilistic modeling, Journal of Advanced Research, pp.423-433 (2014).

[43] Gao, H., Yan, J., Cao, F., Zhang, Z., Lei, L., Tang, M., Zhang, P., Zhou, X., Wang, X. and Li, J.: A Simple Generic Attack on Text Captchas, 23rd Network and Distributed System Security Symposium (NDSS '16) (2016).

[44] Guixin, Y., Zhanyong, T., Dingyi, F., Zhanxing, Z., Yansong, F., Pengfei, X., Xiaojiang, C. and Wang, Z.: Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach, Proc. 2018 ACMSIGSAC Conference on Computer and Communications Security (CCS '18), pp.332–348 (2018).

[45] "Welcome to Flask — Flask Documentation (1.1.x)", available from https://flask.palletsprojects.com/en/1.1.x/ (accessed 2021-02).

[46] "Empowering App Development for Developers | Docker", available from https://www.docker.com/ (accessed 2021-02).

[47] PHISHTANK:" PhishTank" , https://phishtank.org (accessed 2021-02).

[48] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830 (2011).

[49] J. D. Hunter. Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95 (2007).

[50] "RFC 3507 - Internet Content Adaptation Protocol (ICAP)" available from

https://tools.ietf.org/html/rfc3507 (accessed 2021-02)
[51] SMARTBEAR:" The Cost of Poor Web Performance", available", available from https://smartbear.com/blog/test-and-monitor/the-cost-of-poor-web-performance-infographic/ (accessed 2021-02).
[52] IPA：“「高度標的型攻撃」対策に向けたシステム設計ガイド”，https://www.ipa.go.jp/files/000046236.pdf (参照 2020-02).
[53] IIJ：“新型 PlugX の出現”，https://sect.iij.ad.jp/d/2013/11/197093.html (参照 2020-02).
[54] 角田朋，大鳥朋哉，藤井康広，谷口信彦，木城武康．グレーリストを用いたホワイトリスト／ブラックリストの自動生成によるマルウェア感染検知方法の検討，情報処理学会研究報告．SPT，Vol.2014, No.16, pp.1-7 (2014).
[55] 中鉢かける，中村嘉隆，稲村浩．標的型攻撃対策のためのブロックチェーン技術を用いたホワイトリスト方式防御システムの実現性に関する検討，研究報告モバイルコンピューティングとパーベイシブシステム（MBL），2018-MBL-89, No.6, pp.1-8 (2018).
[56] 山西 宏平，芝原 俊樹，高田 雄太，千葉 大紀，秋山 満昭，八木 毅，大下 裕一，村田 正：畳み込みニューラルネットワークを用いた URL 系列に基づくドライブバイダウンロード攻撃検知，コンピュータセキュリティシンポジウム 2016 論文集，pp.811-818 (2016).
[57] 孫 博，秋山 満昭，八木 毅，森 達哉．既知の悪性 URL 群と類似した特徴を持つ URL の検索，コンピュータセキュリティシンポジウム 2014 論文集，pp.1-8 (2014).
[58] Hau, T., An, N., Phuong, V., Tu, V.. DNS graph mining for malicious domain detection, 2017 IEEE International Conference on Big Data (Big Data), pp.4680 - 4685 (2017).
[59] AlienVault:" Open Threat Exchange", https://otx.alienvault.com/ (accessed 2021-02)
[60] SANS:" SANS 2023 Incident Response", https://sansorg.egnyte.com/dl/Mm54hRAGJt (accessed 2024-02)
[61] 東京商工リサーチ：“2023 年「上場企業の個人情報漏えい・紛失事故」調査”，https://www.tsr-net.co.jp/data/detail/1198311_1527.html (参照 2024/08)
[62] 一般社団法人 ICT-ISAC：“一般社団法人 ICT-ISAC”，https://www.ict-isac.jp/ (参照 2025/02)
[63] CEHC, CIS, MS-ISAC:" Success Stories in Cybersecurity Information Sharing", https://livealbany.sharepoint.com/sites/web_cehc/Shared%20Documents/Forms/AllItems.aspx (accessed 2024-02)
[64] EU:" Proposed Regulation on the Cyber Solidarity Act", https://digital-strategy.ec.europa.eu/en/library/proposed-regulation-cyber-

solidarity-act (accessed 2024-02)
[65] MISP:" MISP Open Source Threat Intelligence Platform & Open Standards For Threat Information Sharing" ,
https://www.misp-project.org/ (accessed 2024-02)
[66] DHS:" Additional Progress Needed to Improve Information Sharing under the Cybersecurity Act of 2015",
https://www.oig.dhs.gov/sites/default/files/assets/2022-08/OIG-22-59-Aug22.pdf (accessed 2024-02)
[67] Md, A., Siti, S., Aswami, A., Robiah, Y.. Cyber threat intelligence–issue and challenges, 10, 1, pp.371-379 (2018).
[68] NIST:" Guide to Cyber Threat Information Sharing" ,
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-150.pdf (accessed 2024-11)
[69] NIST:" Computer Security Incident Handling Guide",
https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf (accessed 2024-02)
[70] SANS:" Incident Handler's Handbook",
https://sansorg.egnyte.com/dl/6Btqoa63at (accessed 2024-02)
[71] OASIS:" Introduction to STIX",
https://oasis-open.github.io/cti-documentation/stix/intro (accessed 2024-02)
[72] OASIS:" CACAO Security Playbooks Version 2.0",
https://docs.oasis-open.org/cacao/security-playbooks/v2.0/security-playbooks-v2.0.html (accessed 2024-02)
[73] LangChain:" LangChain",
https://www.langchain.com (accessed 2024-02)
[74] Chroma:" Chroma",
https://www.trychroma.com/ (accessed 2024-02)
[75] Toufique, A., Supriyo, G., Chetan, B., Thomas, Z., Xuchao, Z., Saravan, R.. Recommending Root-Cause and Mitigation Steps for Cloud Incidents Using Large Language Models, ICSE '23: Proceedings of the 45th International Conference on Software Engineering, pp. 1737-1749 (2023).
[76] IETF:" The Incident Object Description Exchange Format Version 2",
https://datatracker.ietf.org/doc/html/rfc7970 (accessed 2024-02)
[77] The Washington Post:" Opinion | 'Trust, but verify': An untrustworthy political phrase - The Washington Post"
https://www.washingtonpost.com/opinions/trust-but-verify-an-untrustworthy-

political-phrase/2016/03/11/da32fb08-db3b-11e5-891a-4ed04f4213e8_story.html (accessed 2025-02)

[78] 東京科学大学：“秘匿計算”, https://tcc.c.titech.ac.jp/yasunaga/contsec/lec_mpc1.pdf?utm_source=chatgpt.com (参照 2025/02)

[79] S. N. Khajeddin, A. Madani, H. Gharaee and F. Abazari, "Towards a Functional and Trustful Web-based Information Sharing Center," 2019 5th International Conference on Web Research (ICWR), 2019, pp. 252-257, doi: 10.1109/ICWR.2019.8765297. (参照 2025/02)

[80] 一般財団法人 運輸総合研究所, “サイバー攻撃に対するセキュリティ情報共有組織（ISAC）の構築に関する調査研究”, https://www.jttri.or.jp/pdf/H29cyber_ISAC-houkoku.pdf (参照 2025/02)

[81] 内閣官房サイバー安全保障体制整備準備室, “サイバー安全保障分野での対応能力の向上に向けた有識者会議”, https://www.cas.go.jp/jp/seisaku/cyber_anzen_hosyo/dai2/siryou4-2.pdf (参照 2025/02)

付録

本研究に関連する発表済みの成果

査読付き論文

本論文の記述および図表は，情報処理学会の許諾に基づき，下記の同一著者の論文を加筆・再編集したものである．

1. **西嶋 克哉**, 川口 信隆, 植木 優輝, 重本 倫宏, 近藤 賢郎, 中村 修, “複数組織の接続傾向を用いた自律進化型防御システムの提案と評価”情報処理学会論文誌 Vol.62 No.12 1-14 (Dec. 2021)

2. **西嶋 克哉**, 重本 倫宏, 近藤 賢郎, 中村 修, “複数組織のインシデント対応記録活用システムの提案と評価” 情報処理学会論文誌 (採録決定) (May. 2025 出版予定)

その他発表済みの成果

国際発表

1. **K. Nishijima** et al., "Verification of the Effectiveness to Monitor Darknet across Multiple Organizations" 2021 Ninth International Symposium on Computing and Networking Workshops (CANDARW), 2021, pp. 346-351

国内発表

1. **西嶋 克哉**, 近藤 賢郎, 細川 達己, 重本 倫宏, 川口 信隆, 長谷川 弘幸, 本田 英之, 鈴木 康人, 鍛 忠司, 中村 修, “複数組織を跨ったダークネット観測の効果に関する検証” 第 82 回全国大会講演論文集, 2020, 1, p. 393-394, 2020-02-20
2. **西嶋 克哉**, 川口 信隆, 重本 倫宏, 近藤 賢郎, 中村 修, “セキュリティオペレーションにおける秘匿データ分析システムの提案”, マルチメディア, 分散協調とモバイルシンポジウム 2062 論文集, 2020, 284-289 (2020-06-17)
3. **西嶋 克哉**, 川口 信隆, 植木 優輝, 重本 倫宏, 近藤 賢郎, 中村 修, “複数組織の接続傾向を用いた自律進化型防御システムの提案”, 研究報告マルチメディア通信と分

散処理（DPS）, 2021-DPS-186(19), 1-8 (2021-03-08)
4. 植木 優輝, 重本 倫宏, 川口 信隆, **西嶋 克哉**, 近藤 賢郎, 中村 修, “組織間信頼度による匿名化処理を用いたインシデントチケット共有システムの開発” 第 84 回全国大会講演論文集, 2022, 1, p. 491-492, (2022-02-17)