Deep Security Conference 2025

生成AI時代のセキュリティ監視

2025 / 07 / 17（木）

Ubie株式会社 水谷正慶

ユビー

本日の発表

- **セキュリティ監視について**
  - セキュリティ監視とはなにか
  - なぜ導入するのか、どのような効果が期待されるか
  - 導入の方針や導入時の困難さなど
- **セキュリティ監視における生成AIの活用**
  - 現状でセキュリティ監視に生成AIが活用できること、できないこと
  - 生成AIの利用でセキュリティ監視業務がどのように変化するか
  - 今後の発展への期待

自己紹介

- **水谷正慶（Ph.D）https://x.com/m_mizutani**
- **職域：セキュリティエンジニア & バックエンドエンジニア**
- **経歴：Ubie (2021~)**, Cookpad(2017~), IBM(2011~)
- **セキュリティとの因縁**
  - 大学時代
    - 侵入検知システムによるマルウェア検知の研究に専念
  - 社会人時代
    - SIEM (Security Information & Event Manager) 関連の開発
    - Managed SOCでアナリストとして勤務
    - AWS, Google Cloud上にセキュリティ監視基盤を構築


ユビー

生活者向け事業

症状検索エンジン ユビー

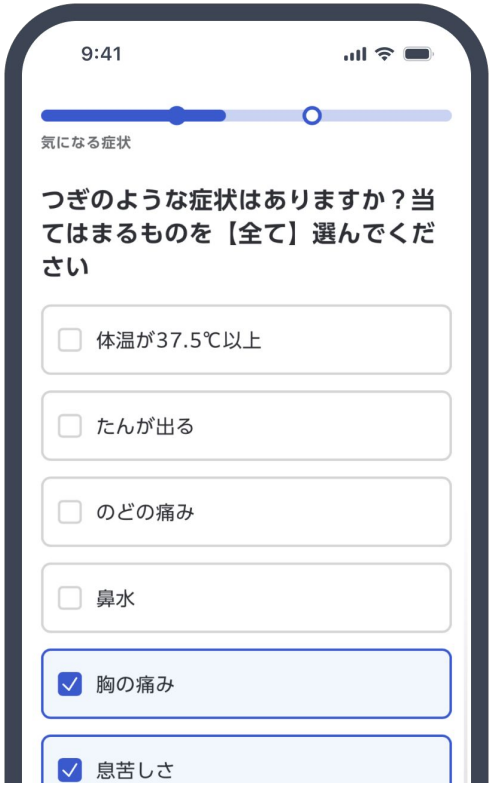
自分の症状を答えるだけで、参考病名や近くの医療機関等「受診の手がかり」が調べられます

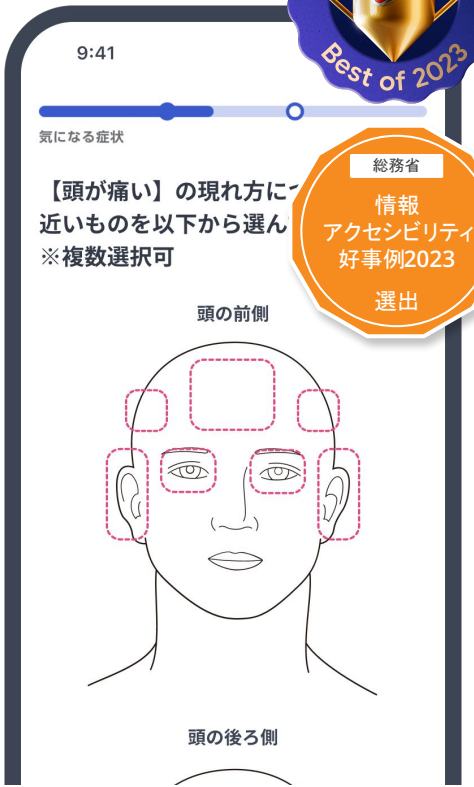
医療現場で実際に使われ鍛えられた AIを、
生活者が適切な医療にかかる目安として開放しています
(2020年提供開始)

- 無料で
- 誰でも
- いつでも
- ほぼ全ての症状で ＊

＊99% (1.3万超)の症状に対応

症状検索エンジン「ユビー」ダウンロードリンク
https://ubie.go.link?adj_t=1c2ifxv9





医療機関向け事業

ユビー メディカルナビ

問診業務効率化や認知向上など、患者さんとのコミュニケーション設計を通じ、診察の質向上を支援する医療機関向けサービスです

病院・クリニックそれぞれのニーズに合わせた
以下のような機能を提供・開発しています

- ユビーAI問診
- ユビー生成AI
- ユビーリンク

etc...

ユビーメディカルナビ　サイト
https://intro.dr-ubie.com/

日本サービス大賞 NIHON SERVICE AWARD
第3回 日本サービス大賞 厚生労働大臣賞

日本サービス大賞 NIHON SERVICE AWARD
第3回 日本サービス大賞 審査員特別賞

セキュリティ監視とは

セキュリティ監視とは

- **「組織内で発生するセキュリティ上の問題を、継続的に検知・調査すること」と定義します**
- **大きく分けると「検知」と「調査」**
  - 検知：収集した情報を分析し、問題が発生しているかどうかを判断する機能。破壊的な行動を発見するだけでなく、潜在的な問題やリスクを発見することも含まれる
  - 調査：検知された問題に対して、その原因や影響を調査する機能。問題の深刻さや影響範囲を把握し、対応策を検討するための情報を提供
    - 特にセキュリティの場合、攻撃を検知したとしても、それがどのような影響を及ぼすかという情報がないと、インシデントであるかの判断が困難であり、調査はそのための重要なステップ

なぜセキュリティ監視が必要か

- **そもそも防御的対策があれば監視は必要ないのでは？**
  - 部分的には正解。原則としては防御的対策を先に取り組むべき
- **防御的対策の限界**
  - **対策の抜け漏れ**：「見えている範囲」の脆弱性にのみ対応可能で、全脆弱性の把握は困難。新たな脆弱性も日々発生
  - **サプライチェーン攻撃**：サプライヤー・パートナーの脆弱性は自組織から制御・観測困難で、別手段での対応が必要
  - **ゼロデイ攻撃**：脆弱性公開から対策公開までのタイムラグや公開前の脆弱性悪用への事前対策は困難
  - **ユーザビリティの問題**：利便性を阻害し生産性低下を招き、セキュリティ回避行動により既存対策が無効化される恐れ

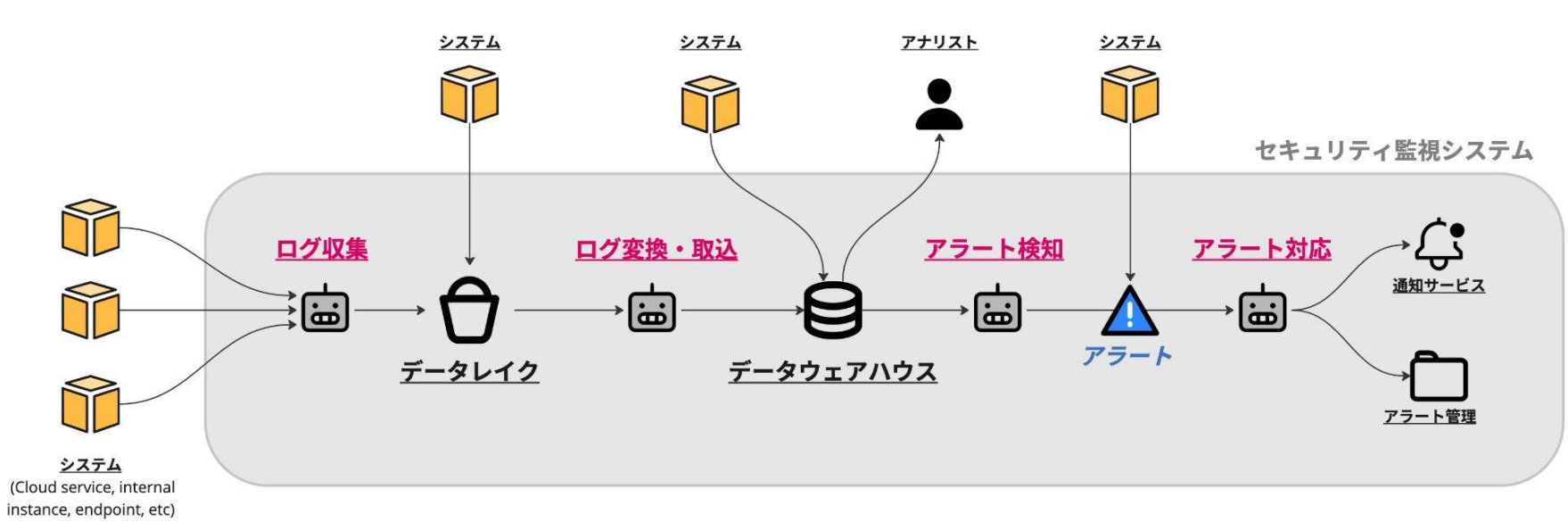
セキュリティ監視を導入するメリット

- **(1) 問題の発生を顕在化する前に察知できる**
  - 近年の攻撃者は一定時間潜伏して攻撃の効果を最大化しようとする
  - 顕在化する前に察知、対応できれば被害を極小化できる
- **(2) 被害の影響範囲を確認できる**
  - 保全したログを調査の証跡として利用でき、攻撃者の行動を追跡できる
  - 侵入された場合でもアクセスが無かったことが確認できれば、漏えい等が発生しなかったことを担保できる
    - これがないと被害状況を正確に把握するのが困難になる

実際のセキュリティ監視に必要な要素

- **1. ログ収集**
  - システム・アプリケーションのログを統一的に収集
  - 自組織外のクラウドサービス・既製ソフトウェアのログも含む
- **2. ログ保全**
  - ログ種類別の生成量・保存期間要件を考慮した大量のログの管理
  - 検索性やコストを考慮した
- **3. ログ分析・調査**
  - インタラクティブな調査・分析（Threat Huntingなど）
  - アラート確認・インシデント原因/影響範囲の調査
- **4. アラート検知**
  - ルール・しきい値による自動アラート発見
  - 継続的なルールチューニングとテスト検証が必須
- **5. アラート対応**
  - 検知された異常への対応機能
  - 担当者への適切な情報提供と対応支援

セキュリティ監視システムの全体像



セキュリティ監視を導入するアプローチ

1. ストレージ・DWHへログを蓄積

- クラウドサービス（オブジェクトストレージ、マネージドDWH）を利用可能
- ログ収集、保全、調査は一定できる。検知・分析・対応は作り込みなどが必要
- システムコストは比較的安い

2. SIEM (Security Information & Event Manager) の利用

- 商用・OSSで様々な種類がある。フルマネージドサービスも近年増加
- ログ収集〜対応まで一気通貫でサポートしているプロダクトも多い
- システムコストは比較的高い

（注:登壇者の主観的な意見です）

組織のフェーズによるセキュリティ監視の導入規模や体制

- **0 → 1 フェーズ、あるいは小規模組織（〜50人）**
  - ログの収集と保全ができており、インシデント発生時に調査が可能になっている
- **1 → 10 フェーズ、あるいは中規模組織（〜250人）**
  - 定期的にログをチェックできおり、異常に気づける可能性がある
- **10 → 100 フェーズ、あるいは大規模組織（〜1,000人）**
  - セキュリティ監視システムが導入され、一定レベルの検知・調査などができている
- **超大規模組織**
  - セキュリティ監視に従事する専属の人員、あるいはチームが存在する
  - ログ収集、検知、調査が高度に実施され、改善のサイクルが回っている

システムのコストが注目されがちだが運用も大変

セキュリティ監視の運用コストは無視できない

- **アラートのトリアージ**
  - セキュリティ監視において「アラート」は殆どの場合「侵害の可能性のある事象」
  - 実際の侵害の有無や影響範囲を確認する作業が必要
  - 関連する情報を調査・収集する必要があり、それなりに負担が大きい
- **検知ルールのチューニング**
  - 見逃し（false negative）を防ぐため過検知側に寄せて運用することが多い
  - しかし影響なしとわかっているアラートを受け続けるのはノイズ
  - そのためルールをチューニングして無害とわかっている条件を除外していく
  - 既存のルールに影響を与えないように修正する必要があり、ルール全体の理解と確認が必要
- **積極的な分析**
  - Threat Huntingなど

監視システムを導入したがそのまま放置ということも…

“Detection Engineering”, “Detection as Code”, “Autonomic Security Operation” など

「セキュリティ監視エンジニアリング」の概念による問題解決

- **セキュリティ監視にソフトウェア開発のアプローチを導入**
  - コード化 (as Code): 検知ルールや対応手順をコードで管理
  - 自動化 (CI/CD): ルールのテストから適用までを自動化
- **期待される効果**
  - 品質向上: レビューや履歴管理でルールの質を高める
  - 迅速化: 脅威への対応やルール改善のサイクルが速くなる
  - 効率化: 定型作業から解放され、人は高度な分析に集中できる
- **導入における課題**
  - 対応ツールが少なく、監視とソフトウェア開発の両方のスキルを持つ人材が希少

セキュリティ監視における
生成AIの活用

（前提）2025年7月の生成AIの概況

- **エージェントツールの台頭**
  - Cursor, Claude Codeなどはデスクトップ用のコーディング支援ツールだが、その他の業務でも利用可能なほどに進化
  - Devin, Codexのようにクラウド上で完結する開発支援型サービスも
- **ツール類の統合による外部リソースへのアクセス**
  - エージェントがファイル、URLなどから自律的に情報を取得するようになった
  - さらにエージェントとツール間のRPCを制御するMCP (Model Context Protocol) によってツールの開発が用意になり、アクセスできるリソースが大幅に増加
  - コンテキストウィンドウの制約と、外部出力における過剰なhuman-in-the-loopの両方の問題が大きく緩和される

2025年6月時点での所感

セキュリティ監視における生成 AIの活用状況

- **少なくとも現状は困難**
  - ❌ 大量のログデータからの異常検知
  - ❌ アラートの深刻度・リスク判定
- **現状でも効果あり**
  - ✅ ログデータの要約・分析・調査の実施・サポート
  - ✅ ルールのチューニング
  - ✅ アラートの調査・エンリッチメント
  - ✅ 外部とのコミュニケーションサポート

（少なくとも現状は）

困難なアプローチ

全ログデータの中なら「なにか異常なログを見つけて」という指示をだすのは難しい

❌ 大量のログデータからの異常検知

- **「異常」という状態を明確に定義する必要がある**
  - 人間が異常と判断する事象もドメイン知識や経験などの裏付けがある
  - これらのコンテキストを省略して「異常を見つける」というのは困難
  - 具体的な指示を出せばよいが、それなら明示的なロジックでも良い
  - 特定の範囲内から対話的に異常を探るのは可能
- **コンテキストウィンドウに制限がある**
  - 各サービスで利用可能なトークンサイズが大きくなっているとはいえ、ログすべてを入力するというようなことは無理
- **コストの課題**
  - 各サービスはトークンサイズに対する従量課金なのでそもそも大量のログを投入すること自体が困難

❌ アラートの深刻度・リスク判定

- **深刻度判定のためのコンテキストが必要**
  - 社内環境、社内規定、社内システムの利用状況、人事情報、各プロジェクトの進行状況、各プロダクトの開発状況、etc…
  - これらの情報をまとめてプロンプト化することは可能だがメンテナンスし続けるコストは著しく高い（が、ツール類の整備で今後解決する
- **深刻度の判断には「責任」が求められる**
  - リスク対応という性質上、false negative（見逃し）は避けたいが、最終的に人間が「責任」を負わなければならなず結局は自分で確認する必要がある
  - 最終的には「精度」の向上と「信頼」の醸成が必要だが、現段階だとコンテキスト集約の問題でまだそれは困難

（現状でも）

効果のある・期待できるアプローチ

✅ ログデータの要約・分析・調査の実施・サポート

- **各サービスのトークンサイズ上限が増えたことによりログ分析は一定容易に**
  - トークン上限：Gemini 2.5 flash: **1M**, Claude sonnet 4: **200K**, OpenAI o4-mini: **200K** tokens
  - ログ1件あたり 30〜300トークン
  - → 数百〜数千件のログは読み込み可能
- **読み込んだ範囲内で応答など可能**
  - 要約する、インジケーター（IPアドレス、ホスト名など）を抽出する、攻撃らしきログを探すなど
  - 発生している状況を順序立てて説明させることも

攻撃パターンの分類

1. 散発的スキャン攻撃（最多）

- 世界各地のIPから1-3回程度の試行
- 主要な攻撃元：中国、ロシア、アメリカ、ヨーロッパ
- 典型的なボットネット活動

2. 集中攻撃（高危険度）

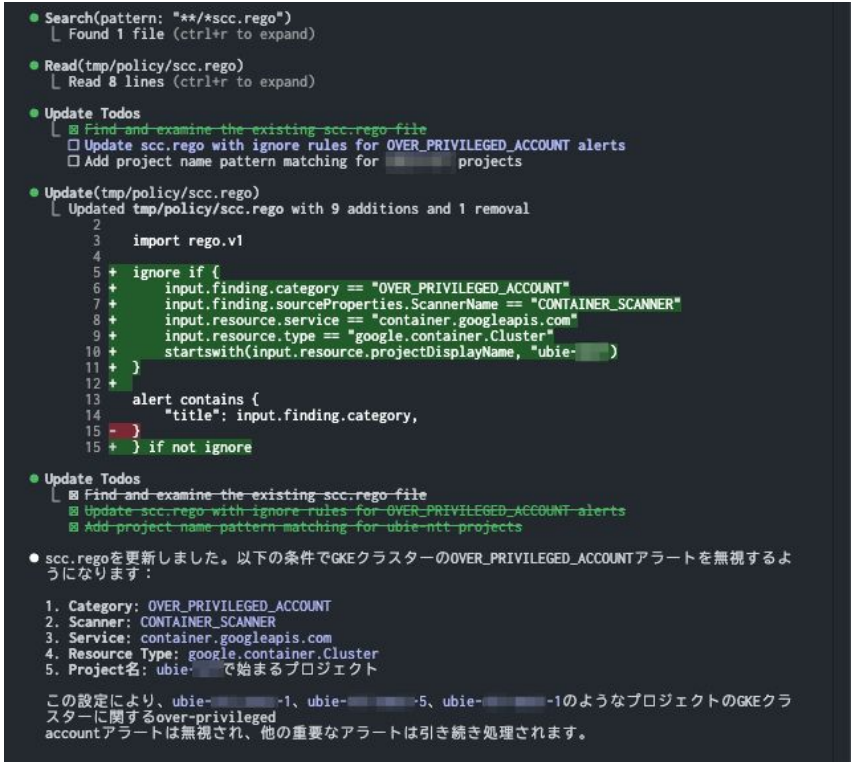
最も激しい攻撃：

- `123.131.17.131`（中国）：7月7日22:44-22:52
  - 約200回以上の連続試行
  - 最初は無差別接続、後半は「user」「debian」での認証試行
  - 最も組織的で持続的
- `62.60.132.163`：7月6日21:33-21:53
  - 約60回の連続試行
  - 20分間の継続攻撃
- `14.22.86.106`：7月7日19:29-19:31
  - 短時間（2分）で50回以上の試行

Claude Sonnet 4 でSSHのログを分析させた結果

✅ ルールのチューニング

- **コーディングと同じ要領でルールを編集**
  - ルールの規則などは丁寧に命令する
  - 誤検知したデータを入力し、それを参考に既存のルールを編集させることが可能
  - テストがあれば自動的に実施して修正
- **生成AIエージェントがルール編集できることが重要**
  - 現在は生成AIコーディングエージェントの性能・拡張性ともに優秀でそれを利用できるのがベスト
  - テキストベースでルールが管理されているシステムは相性が良い



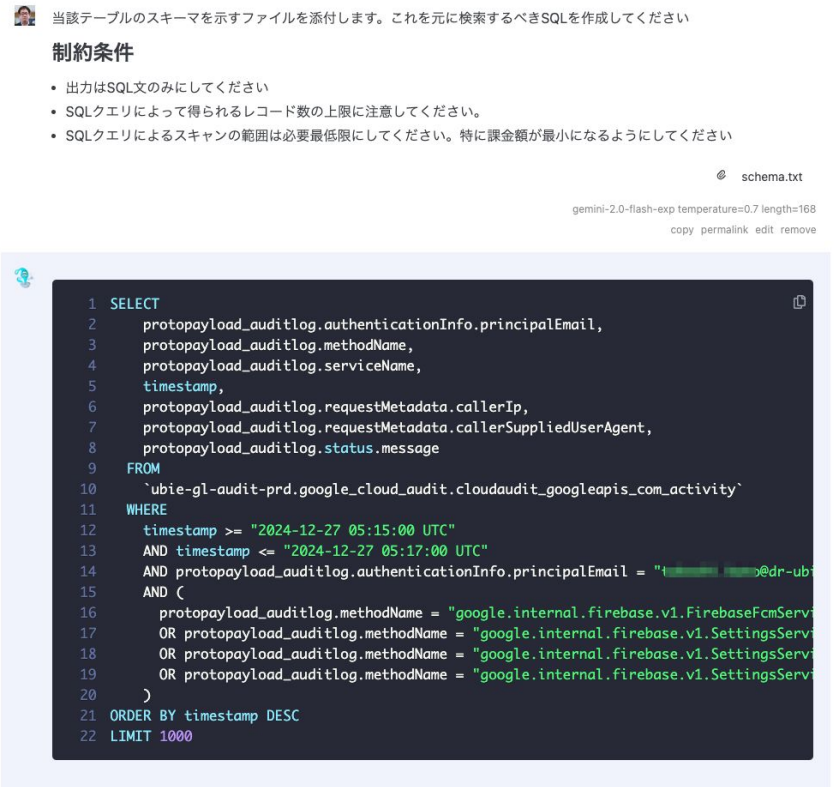
Claude Code を利用したルールのチューニング

✅ アラートの調査・エンリッチメント （1/2)

- **アラートに関連したインジケータ（IPアドレス、ホスト名など）の調査**
  - 事前知識を与えない構造データからインジケータを抽出するのは容易
  - ログデータ、アラートなどからも（ある程度は）文脈を解釈してくれる
  - MCPやツールの利用で各種脅威インテリジェンスサービスにアクセス可能
- **特にエージェント型AIで効果を発揮する**
  - 複数ツールを設定しておくことで、多段的に問い合わせを実行してくれる
  - 調査結果をまとめたり、あるいは何らかのシステムに出力させることも可能
    - 例えばチケットシステムと連携すればそのチケットにコメントとして追記してくれる
  - 分析するメンバーはまとめられた結果を見て判断に集中すれば良い

✅ アラートの調査・エンリッチメント　(2/2)

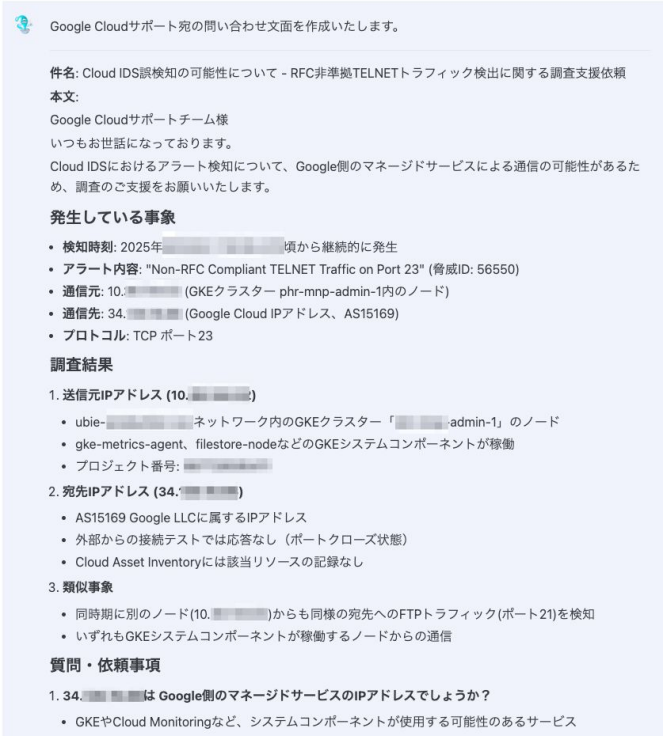
- **ログの検索クエリを自動生成してくれる**
  - アラートそのものやコンテキストを入力した上でスキーマを示すと調査のためのクエリを自動生成してくれる
  - 目的に合わせたクエリの生成は複雑になるため急ぎのときは便利
- **精度を高めるためには様々な工夫が必要**
  - スキーマ情報はなるべくDescription情報が入っていると望ましい
  - 各フィールドのサンプル値なども重要
  - 参考になるSQLの例を示すのも効果的



社内LLMツールにアラート情報およびBigQueryのスキーマを入力したことでSQLを自動生成

✅ 外部とのコミュニケーションサポート

- **監視チーム外からの情報を得る必要がある際の丁寧な連絡**
  - 別の部門のメンバー
  - 関連する外部組織
  - 利用しているサービスのサポート
- **調査や分析の経緯が生成AIへ入力可能になっていると効果的**
  - テキストコミュニケーションの文面作成などをサポート
  -



Google Cloudサポート宛の問い合わせ文面を作成いたします。

件名: Cloud IDS誤検知の可能性について - RFC非準拠TELNETトラフィック検出に関する調査支援依頼

本文:

Google Cloudサポートチーム様

いつもお世話になっております。

Cloud IDSにおけるアラート検知について、Google側のマネージドサービスによる通信の可能性があるため、調査のご支援をお願いいたします。

発生している事象

- **検知時刻:** 2025年[redacted]頃から継続的に発生
- **アラート内容:** "Non-RFC Compliant TELNET Traffic on Port 23" (脅威ID: 56550)
- **通信元:** 10.[redacted] (GKEクラスター phr-mnp-admin-1内のノード)
- **通信先:** 34.[redacted] (Google Cloud IPアドレス、AS15169)
- **プロトコル:** TCP ポート23

調査結果

1. **送信元IPアドレス (10.[redacted])**
   - ubie-[redacted]ネットワーク内のGKEクラスター「[redacted]-admin-1」のノード
   - gke-metrics-agent、filestore-nodeなどのGKEシステムコンポーネントが稼働
   - プロジェクト番号: [redacted]
2. **宛先IPアドレス (34.[redacted])**
   - AS15169 Google LLCに属するIPアドレス
   - 外部からの接続テストでは応答なし（ポートクローズ状態）
   - Cloud Asset Inventoryには該当リソースの記録なし
3. **類似事象**
   - 同時期に別のノード(10.[redacted])からも同様の宛先へのFTPトラフィック(ポート21)を検知
   - いずれもGKEシステムコンポーネントが稼働するノードからの通信

質問・依頼事項

1. **34.[redacted]は Google側のマネージドサービスのIPアドレスでしょうか？**
   - GKEやCloud Monitoringなど、システムコンポーネントが使用する可能性のあるサービス

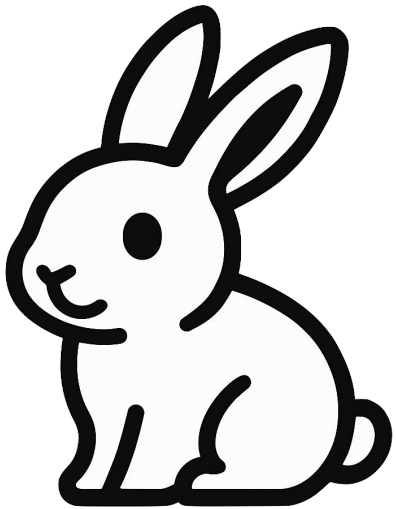
社内LLMツールでアラートに関するSlack上の議論から自動生成したクラウドサポートへの問い合せ

より効果的にやるには

データ連携や専用ツール実行をサポートする

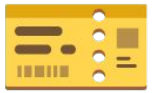
専用のエージェント実装が必要

warren



warren

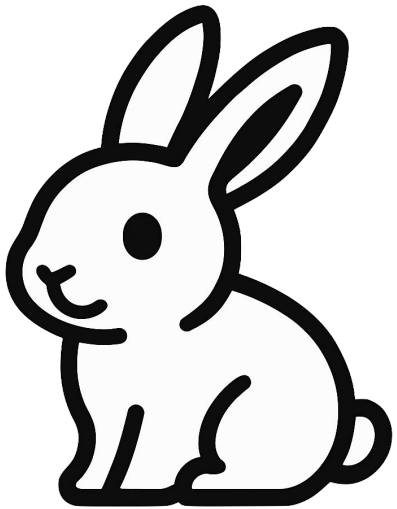
監視機器からのアラート



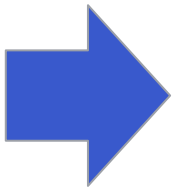
過去のチケットデータ

内部のログデータ

外部の脅威インテリジェンス情報



セキュリティアラートマネジメント用
生成AIエージェント

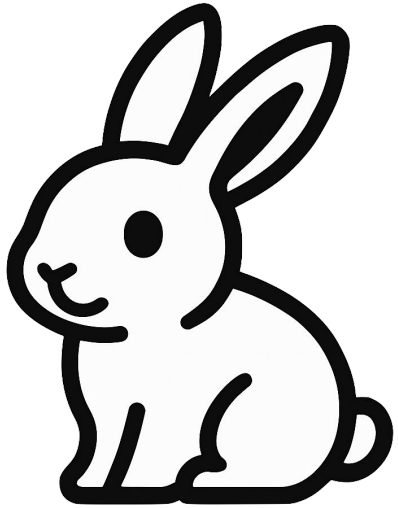


アナリスト

https://github.com/secmon-lab/warren

Warren：セキュリティアラートのマネジメント用生成　AIエージェント

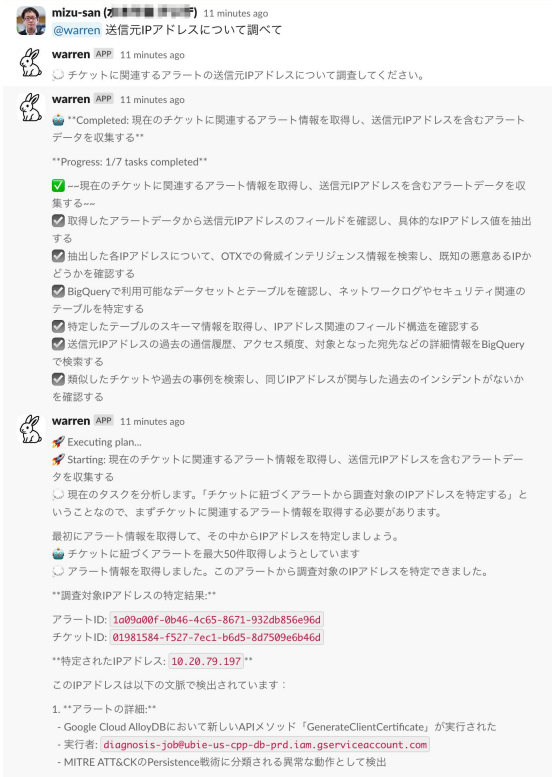
- **生成AIをセキュリティ監視で活用するために開発**
  - 様々な紆余曲折を経た末、エージェント型AIが適切と判断
  - 生成AIが構造データの解釈とアクションが可能になったことでワークフロー管理が不要に
  - SOAR (Security Orchestration, Automation and Response) の代替に近い立ち位置に
- **データアクセスの最適化**
  - MCPの他に脅威インテリジェンスツールやBigQueryへの問い合わせツールを搭載
  - RAGを利用した過去対応済みのチケット・アラートの利活用



https://github.com/secmon-lab/warren

なぜ既存の AIエージェントを利用せず、専用の AIエージェントを作ったか？

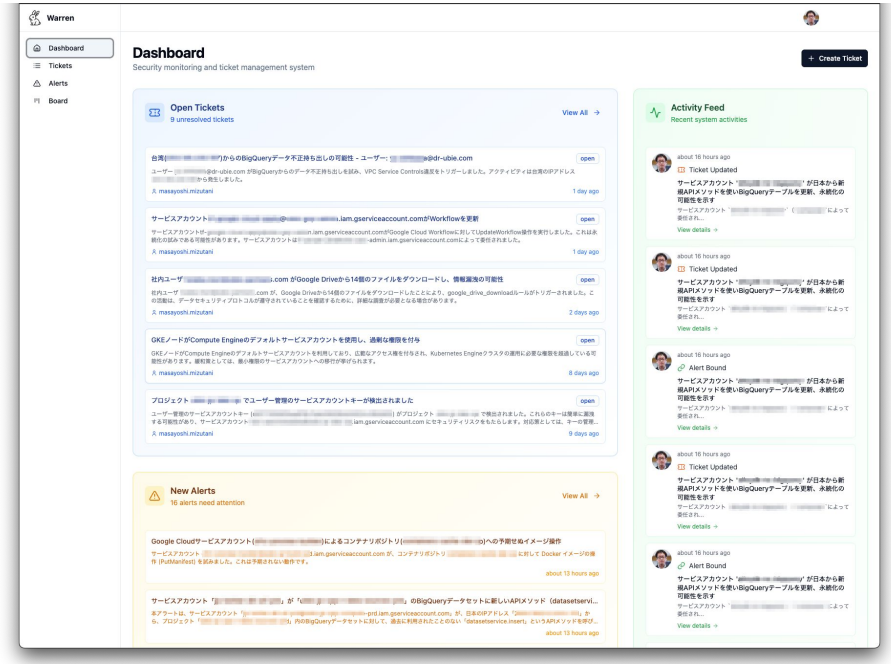
- **(1) 適切なユーザインターフェースの選択**
  - 既存のAIエージェントは優秀だが環境が限定的
  - AIエージェントとやりとりしつつ他メンバーとのコミュニケーションも滑らかに実施したい
  - 一方でまとめて情報を処理するUIもほしい
  - SlackとWebUIを相互に連携させる
- **(2) 独自のワークロードが必要**
  - プロンプティングやLLM呼び出しフローの調整
  - 大量データへのアクセス方法の最適化
    - 特にログデータやスキーマの情報を適切に扱うことが重要となる



Slack上でエージェントから収集された情報をもとに別のメンバーとコミュニケーションが可能

https://github.com/secmon-lab/warren

情報の整理を最適化するための Web UI



ダッシュボードで処理中のチケット、未対応アラート、
直近の対応履歴などを確認

チケット詳細画面で会話履歴確認、類似チケット検索、
関連アラート検索などを提供

https://github.com/secmon-lab/warren

生成AIとの連携ポイント

✅ ログデータの要約・分析・調査のサポート

- BigQueryにクエリを自律的に発行してログデータを取得、確認や分析が可能
- 事前にテーブル・カラムの説明だけでなくユースケースに合わせたクエリのRunbookを用意
- スキャンサイズ上限を設定し、過剰にコストのかかるクエリを発行前に抑制

✅ 外部とのコミュニケーションサポート

- アラートに関する議論をSlackからデータベースへ記録することで生成AIでの利用が可能に
- 「これまでの議論を踏まえた」対応ができる

✅ アラートの調査・エンリッチメント

- 脅威インテリジェンスツール (OTX、URLScanなど) を内部実装しており生成AIが対話の中で問い合わせ可能
- MCPを利用することで組織内製ツールへの問い合わせも可能

✅ ルールのチューニング

- テキストベースの汎用ポリシー言語Regoでルールを記載しており、生成AIコーディングツールでルール調整可能
- テスト機能も実装しており、編集後に正常に動くかの確認までさせる

おわりに

ユビー

まとめ

- **セキュリティ監視は効果的にセキュリティを向上させるための策の一つ**
  - 防御的対策と組み合わせることで負荷の低いセキュリティを実現する
  - しかしシステムコストおよび運用コストは一定高く、導入の敷居が高かった
- **生成AIによって運用コストは改善の兆しがある**
  - アラートの調査、分析、対応のサポート・効率化において強力なツール
    - Copilotの利用でインシデント対応時間24%短縮[1]、CrowdStrikeのSOCチームが手作業を週40時間削減[2]などの報告もあり
  - ただし責任を持ってリスク判断をする専門家は引き続き必要
  - さらなる生成AIモデル・ツールの発展によって小さいチームでの運用が期待される

[1] https://japansecuritysummit.org/2024/11/10664/
[2] https://jp.investing.com/news/company-news/article-93CH-1009308

We Are Hiring!

Ubieでは積極的に採用を行っています。あなたの応募をお待ちしています。

Ubie公式 採用サイト

https://recruit.ubie.life/

https://x.com/UbieCorp_JP

https://www.facebook.com/ubie.inc

https://www.linkedin.com/company/ubie-inc