URL　https://journal.ntt.co.jp/article/29278
DOI　https://doi.org/10.60249/24095104

セキュリティ・トランスペアレンシー確保技術

可視化データの活用に向けたさまざまなステークホルダが寄せる期待と，活用が進んでいない実態を振り返りながら，課題の解決に向けた最新の研究動向と，NTT社会情報研究所が取り組んでいるセキュリティ・トランスペアレンシー確保技術について紹介します．

キーワード：#可視化データ，#SBOM，#LLM

和田　泰典（わだ　やすのり）
荒川　玲佳（あらかわ　れいか）
NTT社会情報研究所

はじめに

国内外の政府動向に対し，サプライチェーン上の各事業者にはSBOM（Software Bill of Materials）を含む可視化データの提供や，セキュリティに関する対応が求められます．しかし，実際に対応を進めていくうえでは，実務面や技術面においてさまざまな課題があります．本稿では，事業者が各種制度に対応するうえで想定される課題や，可視化データの生成や活用における技術課題，可視化データの活用を広げていくためにNTT社会情報研究所で取り組んでいる技術について紹介します．

さまざまなステークホルダが寄せる期待

米国やEU，日本国内の制度化では，SBOMを含む可視化データの提供だけではなく，可視化データの運用管理や，可視化データを用いた脆弱性管理などを行うことにより，サプライチェーンセキュリティリスクを低減することが求められます．

しかし，サプライチェーンにおける各事業者から見た場合，整備された制度やガイドラインなどが求めている要件を，すぐにシステム運用などに反映できるわけではありません．例えば，可視化データを生成するためのさまざまなツールや技術が存在しており，適切なツール等を把握し選択する必要があります．また，それらの管理や運用の面における知見やベストプラクティスなどが必要となるでしょう．加えて，各国の制度やガイドライン[1]などを踏まえると，次のような課題の解決が期待されているといえます．

まず，可視化データの提供における課題です．各事業者はさまざまな場面でSBOMの提供が求められますが，要求元によって必要とする情報の中身，データやファイルの形式などは異なるケースがあります．さらに後述するように，同じフォーマットであっても含まれる情報に差異があります．可視化データを生成する事業者は，これらの要求を満たせるツールを選定し，操作方法を習熟する必要があります．

次に，可視化データを用いた運用面に関する課題です．ガイドライン等によって一定期間はSBOMを管理することなどが規定されているほか，ソフトウェアの更新時などSBOMを更新する頻度についても決めて管理する必要があります．

可視化データを用いた脆弱性などのセキュリティリスク管理も課題となります．各国の制度では，可視化データを用いてセキュリティリスクを低減するだけでなく，セキュリティ要件へ適合することの認証，脆弱性への対処や情報開示などが求められます．そのため，事業者は可視化データを用いてセキュリティを証明する方法や開示内容，可視化データを用いて継続的にセキュリティを確保する運用方法を検討することが必要となります．

このように，これら課題を解決していくためには，可視化データをつくる側とつかう側の協調が重要となります．

さらに，SBOMによりソフトウェアの名称やバージョンが把握できるようになった一方で，2024年3月に顕在化したXZ Utilsというオープンソースソフトウェア（Open Source Software）の脆弱性は，攻撃者が時間をかけて開発プロジェクトに侵入し，バックドアを仕込んでいたと報告されています[2]．これは可視化データを用いて不審なソフトウェアが混入していないかを確認するだけでなく，正規のソフトウェアであっても不正な動作をしていないかを確認するという新たな課題を示唆しています．

可視化データの活用が進んでいない実態

ソフトウェアサプライチェーンは複数の異なる組織を介して形成され，複雑さが増してきています．さらに開発の過程でOSSを再利用することが常態化しているため，ソフトウェアのセキュリティ脅威はより高まっています．可視化データは，ソフトウェアの透明性向上に寄与し，これらの脅威に対抗する手段として期待されているものの，現時点で十分に活用が進んでいるとはいえません．その原因として可視化データ自体にかかわる諸課題の存在が大きな障壁となっています．

可視化データのライフサイクルを，生成フェーズ，収集管理フェーズ，活用フェーズに大別した場合にそれぞれに課題が存在します．生成フェーズとはユーザが機器やシステムの依存関係やライセンス等の構成情報を管理するために，可視化データを生成する段階です．ここでは生成に使うSoftware Composition Analysis（SCA）ツールにかかわる課題があります．その1つが，SCAツールの仕様の違いによって可視化データに出力される文字列に表記揺れが発生する問題です．例えば，サプライヤの組織名として出力される文字列にSCA

ツールによって付与される“Person:”や“Organization:”の接頭辞の違いや，“組織名 inc”と“組織名 llc”の違いが該当します．この問題に対しては各社が検討を進めており，独自データベースによる照合方式などが実現されています[3]．また，SCAツールの解析性能に差がある問題もあります．これはツールによって解析内容に差がある問題で，具体的に私たちの調査データの一例を示すと，Docker Hubで取得したMongoDBのイメージファイルを検査した結果では，SCAツールのSyftでは依存関係パッケージが295件出力されたのに対して，Trivyは136件という違いがみられました．SCAツールの選定は，特に可視化したい構成情報と目的に応じた使い分けが前提ではありますが，そもそも，米国NTIA（国家電気通信情報局）が発行するガイドラインに定められた最小要素[4]の要件を満たすSCAツールが存在していない可能性を示唆する調査結果もあります[5],[6]．続いて収集管理フェーズは，生成された可視化データを収集および管理する段階です．ここでは，可視化データどうしの互換性に関連した統一的な扱いの難しさに課題があります．可視化データにはSPDXとCycloneDXの2つのフォーマットがあり，前者はライセンス情報の項目が多く，後者はセキュリティ情報の項目が多い仕様になっています．収集した可視化データを管理する際にどちらかのフォーマットに統一すると項目が不足するため，互換性を保つために網羅的なフォーマットモデルと統合基盤開発の検討が重要となります[7]．

最後の活用フェーズでは，可視化データを複数の異なる組織で共有する場合のセキュリティの課題です．すなわち，可視化データが共有の過程で不正に書き換えられていないかという完全性やアクセス制御の実現の課題です．可視化データの真正性を担保するために，ブロックチェーンにおけるVerifiable Credentialsモデルをサプライチェーンに応用した技術が検討されています[8]．

このように可視化データにかかわる課題はフェーズによって異なり，可視化データ自体の表記揺れなどのようなミクロな問題と，可視化データそのものを管理活用するマクロな問題が混在しています．これらは互いに独立した問題ではないため，一企業の努力や技術群では可視化データの浸透および活用の障壁となる課題を解決することは難しいです．また学術領域に目を向けても，可視化データの諸課題を整理して解決の方向性を論じる調査論文[9]～[11]はいくつかありますが，実課題に落として技術提案をする文献はごく一部な状況です．コンソーシアムではこの諸課題に対し，各社が知見を共有して可視化データの普及に向けた技術意見交換を行っています．

可視化データを活用したセキュリティ運用の高度化

可視化データを活用することにより，大きく様変わりするセキュリティ運用業務の1つに，脆弱性管理があります．脆弱性管理業務は，脆弱性情報の収集，脆弱性の危険性の確認，自組織への影響の分析，といった流れで行われます[12]．

脆弱性管理では，自組織で使用しているハードウェアやソフトウェアの構成の把握がまず行われます．正確な構成把握により，該当する脆弱性を正確に把握できます．その方法の例として，管理簿による管理や，パッケージ管理システムなどがあります．しかし，システム更改等で構成は変化することや，パッケージ管理システムで管理されないソフトウェアもあることから，これらの方法には管理漏れの問題や，管理漏れを防ごうとすると管理稼働が増えてしまうといった問題があります．そこで，可視化データを導入し，正確かつ最新の構成情報を把握することで，この問題が解決できると考えられます．

しかし一方で，別の問題が発生することも考えられます．脆弱性管理業務において，脆弱性の影響を分析する際には，複数の情報が用いられます．例えば，脆弱性の深刻度，攻撃コードの入手可否，実被害発生状況，通信状況や起動状態などを含む環境情報があります．担当者はこれらを用いて影響を判断するため，脆弱性が正確に可視化され，従来見落としていた脆弱性も把握可能になると，従来と同じ考え方での脆弱性対応では管理しきれません．そこで，可視化データに関する研究開発と同時に，脆弱性対応に関する研究開発も必要となります．そのような技術について，2つ紹介します．

1番目は，機器で発生する通信動作を可視化することで，検出された脆弱性の中から優先的に対応すべきものを絞り込む技術です．この技術では，機器内で発生した通信について，その通信を行ったソフトウェア情報と紐付けた情報を生成できます．その結果，例えばソフトウェアXのバージョンYというソフトウェアが，グローバルIPアドレスZと通信した，といった情報が可視化されます．脆弱性の影響判定に用いる情報の1つに通信情報があることから，この情報を用いることで，リスクが高く優先的に対応する必要がある脆弱性について，通信先を基に絞り込むことができます．

2番目は，機器の起動時に実行されるプログラムを解析し可視化する技術です．この技術では，機器の起動時に実行されるプログラムや，定期実行されるプログラムを可視化することができます．脆弱性の影響判断に用いられる情報の中にはソフトウェアの起動有無があることから，この情報を基に，機器内に含まれるソフトウェアのうち，脆弱性検査時に優先的に対応すべきものを絞り込むことができます．

このように，可視化データを活用するうえでの問題点を解決する技術も開発することで，可視化データの活用を進めていきたいと考えています．

可視化データの利活用拡大に向けて取り組んでいること

現状では可視化データに記載されている機器やシステムの構成情報は，主に依存関係の把握や脆弱性管理のユースケースに活用されていますが，それは可視化データ単

体での活用例となっています．一方でサプライチェーン上の可視化データ集合に対象を広げると，例えば可視化データ間の構成情報の差分から不正な構成情報を特定したり，構成情報の依存関係と共起性の特徴などからSCAの性能が原因による構成情報の欠損を補完できたりするなどの利活用拡大が期待できます．ここでの構成情報の依存関係と共起性の特徴とは，例えば可視化データに依存関係パッケージDが記載されていて，パッケージDがパッケージのAとCを前提にした依存関係がある場合に，それらは共起関係があることを指しています．私たちは可視化データを大規模に管理する基盤をつくりながら，前述のような構成情報の特徴を解析してパターンをとらえることや，LLM（大規模言語モデル）の膨大な学習データを基に欠損したパッケージを補完推定するような技術検討も行っています．これらの取り組みは可視化データの構成情報そのものの価値を高め，今後の可視化データの普及に貢献するとともに透明性の高いソフトウェアサプライチェーンの実現につながると考えます．

そして，さらなるサプライチェーンセキュリティの強化のためには，前述のような透明性を高めてリスクを可視化するだけでなく，リスクを適切に対処して次の対策につなげるサイクルが重要になります．私たちの取り組みでは可視化データに加え，開発フェーズの段階でリスクを可視化する技術開発にも取り組んでおり，字句列解析で網羅的に検知するソースコード依存性解析技術を既に確立しています．現在は大規模言語モデルを活用したソースコード解析の技術検証を実施しており，字句列特徴では難しかった処理内容を意味的に解析することで新たなリスク検知技術の確立をめざしています(13)．また，検知したリスクに対処するための技術検討については，高度な知識を要する脆弱性分析やリスクの見積もりを自動化する検討を進めています．このタスクは脆弱性に関する自然言語とソースコード処理に横断した解析能力と属人的な対応経験や知見が絡み合っており，自動化には高度な技術が必要となります．そこで，自然言語とコード解析に強い大規模言語モデルを積極的に活用し，タスクを小さな課題に分割して有効性を検証している段階です．実際に実施していた検証を取り上げると，脆弱性が発動する個所の修正であるか否かを大規模言語モデルが識別できるかを実験しました．検証の結果はゼロショットのプロンプトでも，ある程度高い識別精度は確認できたものの，脆弱性発動個所が多い脆弱性種別は精度が下がる傾向などが得られ，すべての脆弱性を同じように扱うことの難しさの課題が出てきました(14)．別の海外の研究で，脆弱性の自動修正を試みた研究でも修正個所が複数のファイルにわたって存在する場合などでは難しいことが実証されています(15)．こうした知見から，実課題を適切に分解し検証を行うことで，大規模言語モデルを活用してどこまでをドメイン特化できるかを見極めることが重要だと感じています．脆弱性に関連する領域では，特定の1つの技術がすべてを代替できるわけではなく，古典的な従来の手法の短所を補いつつ大規模言語モデルの特徴を組み合わせたいくつかの技術を，さらに互いに組み合わせて連鎖的に動作させたシステム設計が課題解決につながると考えます．そして，これらの検討と可視化データによる透明性の確保の技術検討は両輪で行うことにより，それぞれの技術における仕様や使い方に生まれる溝を小さくし，サプライチェーンのセキュリティを高める実用的な技術開発につながります．

今後の展望

セキュリティ・トランスペアレンシー・コンソーシアムにおける「協調領域」にて議論を進めている社会課題を元に，「競争領域」としてNTT社会情報研究所で取り組んでいる研究活動について紹介しました．可視化データが，ソフトウェアを安心して利用できる技術となるよう，研究活動を進めていきます．

■参考文献
(1) https://www.meti.go.jp/press/2024/04/20240426001/20240426001-1.pdf
(2) https://gihyo.jp/article/2024/04/daily-linux-240402
(3) https://yamory.io/news/patent-sbom
(4) https://www.ntia.gov/report/2021/minimum-elements-software-bill-materials-sbom
(5) https://ken.ieice.org/ken/paper/20240308GcCV/
(6) https://ipsj.ixsq.nii.ac.jp/ej/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=8&item_id=228660&item_no=1
(7) https://www.nttdata.com/jp/ja/trends/data-insight/2023/0207/
(8) https://journal.ntt.co.jp/article/23459
(9) https://doi.org/10.1145/3597503.3623347
(10) https://doi.org/10.1109/ICSE48619.2023.00219
(11) https://doi.org/10.1145/3654442
(12) https://www.ipa.go.jp/security/reports/technicalwatch/hjuojm0000006nd2-att/000071660.pdf
(13) https://ken.ieice.org/ken/paper/202407231c3p/
(14) https://ken.ieice.org/ken/paper/20240723PcdW/
(15) https://doi.ieeecomputersociety.org/10.1109/SP46215.2023.10179420


（左から）　和田 泰典／ 荒川 玲佳

本稿では可視化データの利活用における課題と最新の研究動向の内容について紹介しました．サプライチェーンセキュリティの強化に向けてNTTグループの枠を超えたソリューションの創出に取り組んでいければと思います．

◆問い合わせ先
NTT社会情報研究所
企画担当
E-mail　solab @ ntt.com