

「インシデント分析の広域化・高速化技術に関する研究開発」

(委託研究)

株式会社クリプト 高須賀 禎子、伊沢 亮一、廣友 雅徳、寺村 亮一、
朝倉 康生、小篠 裕子、勝手 壮馬

Sadako Takasuka, Ryoichi Isawa, Masanori Hiroto, Ryoichi Teramura,
Yasuo Asakura, Yuko Ozasa, Soma Katsute

1. 研究開発の概要

本研究開発は、図 1 に示す広域分散型のインシデント分析システムを構築し、1) 攻撃の全域性および地域性、2) 攻撃の原因、3) 対策の3つの問題を効率的に解決することを目的とする。

本インシデント分析システムは複数のシステムで構成されており、我々はイ-1 階層拠点間の分散協調のための分析結果情報の匿名化・秘匿化技術に関する研究開発を行う。イ-1 の初年度の研究課題は3つに大別でき、課題(A) 従来の匿名化手法の調査と本イベント分析システムへの適合性について精査し、必要な要件を抽出すること、課題(B) 分析実施主体の匿名化すべき情報として、双方向のIPアドレスとポート番号を想定し、パケット単位での匿名化手法を開発し、実装および評価を行うこと。また、パケット単位だけでなく、IDSが出力するファイルやpcapファイルの形式にも適応させた匿名化手法を開発すること、課題(C) 上記(A)、(B)の結果から階層化、分散化を想定し、その環境での必要要件を抽出することである。

本年度の研究課題は全て終了しており、本年度の目的は達成している。本稿では2節にて課題(A)～(C)に対しての研究内容を述べる。3節にて研究開発実績として、課題(A)～(C)の開発実績を述べる。具体的には、課題(A)の成果として、匿名化の従来手法について述べるとともに本イベント分析システムへの適合性について述べ、本イベント分析システムに必要な要件を述べる。課題(B)の成果として、IPアドレスとポート番号を想定したパケット単位での匿名化手法のアルゴリズムとその実装方法および評価について述べる。また、IDSの出力ファイルやpcapファイルへの実装方法を述べる。課題(C)の成果として、階層化、分散化されている分析環境においての必要要件を示す。4節ではまとめと

今後の課題を述べる。

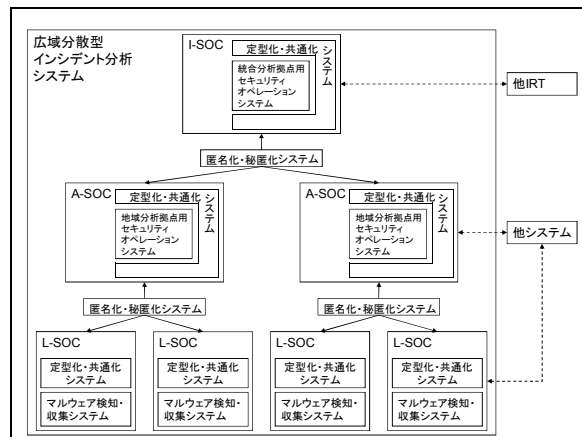


図 1 広域分散型インシデント分析システムの概要

2. 研究開発の内容

イ 階層拠点間の分散協調のための分析結果情報の匿名化・秘匿化技術に関する研究開発

イ-1 分析実施主体の匿名化手法の研究開発：

本研究課題では、L-SOCにおいて収集されるイベント情報、及びA-SOC、I-SOCにおいて集積された統計情報や分析された結果情報において、匿名化すべき情報、あるいはその中の一部に限定される部分情報の匿名化について研究開発を行う。本年度では前述した課題(A)～(C)が設定されており、それらに対して以下の研究を行った。

- (A) 従来の匿名化手法の調査と本イベント分析システムへの適合性について精査し、必要な要件を抽出した。具体的には、2008年に開催された匿名化のワークショップ(ACM Workshop 2008)から従来の匿名化手法を調査し、さらにそれらの参考文献

を調査することに加え、適宜、匿名化の学会等で文献を調査した。文献の調査がある程度完了した後、A-SOC や I-SOC の分析方法や分析に必要な情報を洗い出し、従来の匿名化手法を適合できるかを精査した。

- (B) 分析実施主体の匿名化すべき情報として、双方向の IP アドレスとポート番号を想定し、パケット単位での匿名化手法を開発し、実装および評価を行った。また、パケット単位だけでなく、IDS が出力するファイルや pcap ファイルの形式にも適応させた匿名化手法を開発した。具体的には、ハッシュ関数によりパケットの IP アドレスとポート番号を匿名化する手法の開発、実装を行った。評価ではこの匿名化手法が攻撃に耐性があるかなどを明確にした。また、この匿名化手法を IDS の出力ファイルと pcap ファイルを匿名化するように実装した。
- (C) 上記0、(A)の結果から階層化、分散化を想定し、その環境での必要要件を抽出した。具体的には階層化、分散化されている環境において秘密情報を共有する方法を開発することが必要要件として挙げられる。

3. 研究開発実績

イ 階層拠点間の分散協調のための分析結果情報の匿名化・秘匿化技術に関する研究開発

イ-1 分析実施主体の匿名化手法の研究開発：

本年度の研究課題は全て終了しており、本年度の目標は達成している。各研究課題(A)～(C)の成果をそれぞれ3.1節～3.3節に示す。

3.1. (A) 従来手法の調査と本イベント分析システムとの適合性

イ-1 で開発する匿名化手法の要件として、1)各課題が取り扱う情報のうち、プライバシーに関する部分を他者に漏えいしないように隠ぺいし、2)匿名化手法に対する攻撃への耐性があり、3)A-SOC や I-SOC の統合的な分析に影響を与えない、ことが挙げられる。匿名化手法として、Sanitization に関する研究[1], [2]、匿名化フレームワークに関する研究[3], [4]、暗号化を用いた匿名化手法[5]などが提案されているが、イ-1 の要件を満たす手法は提案されていない。

[1], [2]は Sanitization に関する安全性について述べている。Sanitization とは、あるデータ内のプライバシーに関する情報に対し消去や変換、情報の追加を行う手法で、元のデータを推測させないことで匿名化を

行う。一般に、匿名化には Sanitization を用いることが考えられているが、単に Sanitization を行うだけでは要件 2)、3)を満たすことができない。これは情報を隠ぺいしたとしても、隠ぺいした情報に関連する情報から推測される危険がある[1] [2]。例えば、パケットの送信元 IP を Sanitization したとしても、送信元 IP と送信時刻に関連がある場合、時刻から送信元 IP を推測されてしまう。また、Sanitization を行うと情報が削除されるため A-SOC や I-SOC の分析に支障が生じる。

[3], [4]では匿名化フレームワークとして匿名化に特化したプログラミング言語を提案している。匿名化に特化したプログラミング言語では匿名化に必要な機能が用意されており、組織のセキュリティポリシーに応じたデータの匿名化が可能となる。しかしながら、匿名化に特化したプログラミング言語では匿名化の枠組みを提供するものであって、プログラムを作成するためには組織のセキュリティポリシーに応じてプログラミングを行う必要があり容易ではない。

[5]では暗号化を用いた匿名化手法を提案している。[5]は Front-end Stage と呼ばれる機能と Back-end Stage と呼ばれる機能から構成される。Front-end Stage ではネットワークデータを収集して暗号化を行い、Back-end Stage へ送信する。攻撃パケットなどの疑わしいデータに対してのみ Front-end Stage から Back-end Stage に暗号鍵を配布することで情報を共有する。しかしながら、疑わしいデータの中にも、個人のメールアドレスなどのプライバシーに関する情報が含まれていることがあり、プライバシーに関する情報を漏洩する危険がある。

以上のように、これらの従来手法では本課題の要件を満たすことはできない。本課題の要件を満たすためには、プライバシーに関する情報を隠ぺいしながら A-SOC や I-SOC の分析に影響を与えないよう情報を残す必要があり、この相反する性質を満たすことは従来手法のように汎用的な匿名化手法では困難である。そこで、提案手法では A-SOC や I-SOC の分析に特化した匿名化手法を考案することにより、プライバシーを保護し、かつ、A-SOC や I-SOC の分析に影響を与えない匿名化を実現する。具体的には IP アドレスやメールアドレスなどのイベント情報や各課題のシステムごとに異なる匿名化を適用することにより、A-SOC や I-SOC の分析に特化した匿名化を実現する。

3.2. (B) 匿名化手法の提案と実装および評価

提案手法は課題(B)だけを満たすのではなく、今後の展開を考え A-SOC の分析を考慮した匿名化手法となっ

ている。匿名化の対象は IP アドレスとメールアドレス、URL、ポート番号とする。これらは本イベント分析システムが取り扱うアラート情報と詳細ログの中で、個人の情報に該当する可能性があるため匿名化の対象とする。また、パケットのポート番号も匿名化の対象とする。

提案手法は複数の匿名化手法を用いることで匿名化を行う。これは従来手法のような汎用的な手法では本イベント分析システムに対する匿名化の要件を満たすことができないため、個々の情報に合わせた方法で匿名化する。図 2 に提案手法の全体のフローを示す。提案手法にアラート情報や詳細ログ、パケットを入力すると、IP アドレス、メールアドレス、URL、ポート番号、その他に分類し、それぞれに対応した手法により匿名化を行う。IP アドレスは匿名化手法 1、メールアドレスは匿名化手法 2、URL は匿名化手法 3、ポート番号は匿名化手法 4 を用いる。匿名化手法 1 は IP アドレスを入力とし、出力結果として IP アドレスの地域、属しているネットワーク、IP アドレスの識別子を出力する。匿名化手法 2 はメールアドレスを入力とし、出力結果としてホスト部に初期化ベクトル 2 を連結した文字列とドメイン部に初期化ベクトル 2 を連結した文字列に対しそれぞれハッシュ関数を適応した値を出力する。匿名化手法 3 は URL を入力とし、出力結果として URL のトップドメインと初期化ベクトル 3 を連結した文字列にハッシュ関数を適応した値を出力する。匿名化手法 4 ではポート番号を入力とし、出力結果としてポート番号に初期ベクトル 4 を連結したハッシュ値を出力する。各匿名化手法で初期化ベクトルを用いる理由は、悪意のある第三者にハッシュ値から入力値を推測されることを防ぐためである。なお、初期化ベクトルは十分な長さを持った任意の定数としている。提案手法の評価には匿名化の評価手法[6][7][8][9]を用いた。匿名化の強度という点ではほぼ問題がないが、パケットや IDS の出力ファイル、pcap ファイル内の時刻と送信者やプロトコルと送信者に関連がある場合は送信者を特定される可能性がある。

提案手法の実装は図 3 に示すように分析エッジマネージャから呼び出されるライブラリとした。具体的には linux のスタティックリンクライブラリ(C 言語)にて実装した。パケットや IDS の出力ファイル、pcap ファイルの差異は分析エッジマネージャの入出力インタフェースの部分で吸収することとし、提案手法のライブラリに変更はない。

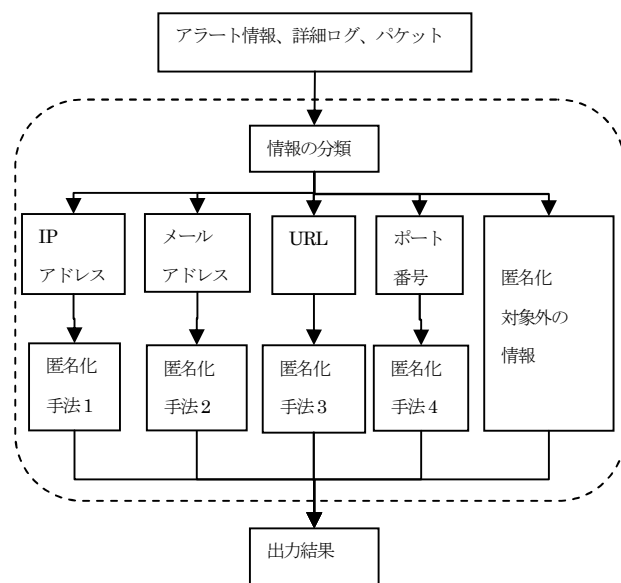


図 2 提案手法の概要

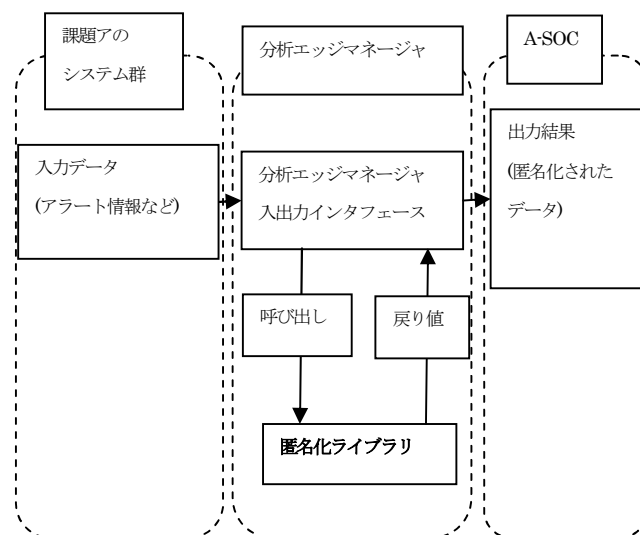


図 3 提案手法の実装

3.3. (C)階層化、分散化されている分析環境における必要要件

提案手法では初期化ベクトルを用いる。本イベント分析システムは階層化、分散化されているため、初期化ベクトルを共有する方法を考慮する必要がある。このとき、拠点自身が裏切ることを想定すると、初期化ベクトルが固定の場合はハッシュ値から元の情報を漏洩する可能性がある。階層化、分散化を考えた初期化ベクトルの共有方法を開発することが要件として挙げられる。また、今後、暗号化技術を提案手法に加える可能性があり、その際には各拠点間での暗号鍵の共有方法を開発する必要がある。

4. まとめ

課題(A)の成果として、従来の匿名化手法の調査と本イベント分析システムへの適合性について精査し、必要な要件を抽出した。課題(B)の成果として、分析実施主体の匿名化すべき情報として、双方向のIPアドレスとポート番号を想定し、パケット単位での匿名化手法を開発し、実装および評価を行った。また、IDSが出力するファイルとpcapファイルの形式に対する実装方法を示した。課題(C)の成果として、(A)と(B)の結果を踏まえ、階層化、分散化を想定し、その環境での必要要件を抽出した。

今後の課題として次の2点が挙げられる。a) 課題(C)で上がった初期化ベクトルや暗号鍵などの、各拠点間での情報の方法の開発をする必要がある。b) 課題(B)の評価で述べた時刻やプロトコルなどが送信者と関連がある場合、送信者が特定される可能性があり、これについて対策を講じる必要がある。

(参考文献)

- [1] R. Crawford, M. Bishop, B. Bhumiratana, L. Clark, K. Levitt, "Sanitization Models and their Limitations," the New Security Paradigms Workshop, pp. 41-56, 2006.
- [2] Martin Burkhart, Daniela Brauckhoff, Martin May, Elisa Boschi, "The Risk-Utility Tradeoff for IP Address Truncation," ACM Workshop 2008, pp. 23-30, 2008.
- [3] Jelena Mirkovic, "Privacy-Safe Network Trace Sharing via Secure Queries," ACM Workshop 2008, pp. 3-10, 2008.
- [4] D. Koukis, S. Antonatos, D. Antoniadis, E. P. Markatos, P. Trimintzios, "Generic Anonymization Framework for Network Traffic," ICC2006, pp. 2302-2309, 2006.
- [5] G. Bianchi, S. Teofili, M. Pomposini, "New Directions in Privacy-preserving Anomaly Detection for Network Traffic," ACM Workshop 2008, pp. 11-18, 2008.
- [6] L. Sweeney, "k-Anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, pp. 557-570, 2002.
- [7] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkatasubramanian, "l-Diversity: Privacy Beyond k-Anonymity," presented at 22nd International Conference on Data Engineering, Atlanta, Georgia, USA, 2006.
- [8] M. Reiter and A. Rubin, "Anonymity loves company: Anonymous Web transactions with Crowds," Communications of the ACM, 1999.
- [9] Douglas J. Kelly, Richard A. Raines, Michael R. Grimaila, "A Survey of State-of-the-Art in Anonymity Metrics," ACM Workshop 2008, pp. 31-39, 2008.