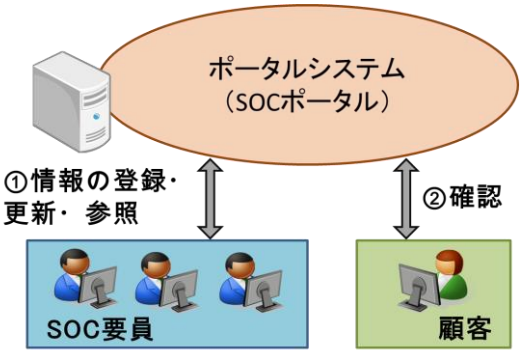
L-036

セキュリティインシデント管理機能に関する関連インシデント抽出方法の提案
Similarity Security Incident Extraction Method for Security Incident Management

小野 裕美† 白木 宏明† 大松 史生†
Hiromi Ono Hiroaki Shiraki Fumio Omatsu

1. はじめに

近年、特定の組織を対象として、個人情報や企業機密などの情報を窃取しようとしてサイバー攻撃を仕掛ける標的型攻撃が多発している。企業は、顧客への迅速な対応を行うために、セキュリティ監視を行うための SOC(Security Operation Center)を導入し対応を行っている。SOC 要員と顧客との情報共有のためにポータルシステム(SOC ポータル)を導入することで、IDS(Intrusion Detection Systems)や IPS(Intrusion Prevention Systems)などのセキュリティ機器が検知したセキュリティインシデント情報を短期間で顧客に提供できることが考えられる。



SOCにおけるセキュリティインシデント管理運用では、SOC要員がIDSなどのセキュリティ機器のログ情報から新規のインシデント情報を登録する運用があり、本稿では、登録に関する運用効率化を考える。

2. 課題

新規のインシデト情報を登録する場合に、過去に対応したインシデント情報の履歴について検索を行って新規のインシデント情報に関連する情報を参照し、新規のインシデント情報を登録する方法が考えられるが、この方法では、SOC 要員の検索スキルにより参照する過去 のインシデト情報が異なるため、 SOC 要員の対応にばらつきが生じる。

上記課題を解決するために、 過去に SOC 要員が対応したインシデントから新規インシデントと関連するインシデント（以下、関連インシデント）を自動抽出する方式について検討した。

3. 提案手法

以下の流れで、関連インシデント情報を抽出する。

① 新規インシデントの基本情報を、SOC ポータルに登録する。

② 関連インシデント抽出機能で、過去に対応したすべてのインシデント情報に類似度を求める。

③ 類似度から関連インシデントを抽出する。

関連度を示す類似度は、重要度と一致度の積和から算出する。

① インシデントの基本情報に表 1のように重要度と一致度を設定する。

表 1 重要度と一致度の設定の一例

No.	項目名	重要度	一致度		
			完全一致	部分一致	不一致
(1)	シグニチャ ID	10	1	-	0
(2)	アラート種別	5	1	0.5	0
(3)	外部 IP アドレス	4	1	0.75	0
(4)	内部 IP アドレス	3	1	0.75	0
(5)	時間	2	1	0.90	0
(6)	ベンダ	1	1	-	0
(7)	製品名	1	1	-	0

② 新規インシデントと過去に対応した各インシデントの基本情報について、以下の計算式を実施し、類似度を算出する。

$$similarity_{ticketID} = \sum_{k=1}^{7} (importance_k \times consistent_k)$$

算出された類似度から関連インシデントが抽出されるため、表 1の重要度を、適切な値で決定する必要がある。重要度の決定する方式を、以下に示す。

表 2 重要度を決定する方式

No.	方式	内容
(1)	機械学習による決定	関連するインシデントのリストを学習データとして与え、自動的に重要度を決定する。
(2)	熟練者による決定	アナリストが過去のインシデントについての知見から重要度を指定する。

本稿では、関連するインシデントのリストを学習データとして与え、自動的に重要度を決定する、機械学習による重要度の決定について検討する。

†三菱電機(株)情報技術総合研究所

3.1 機械学習による重要度の決定

過去に対応したインシデント情報の一部から、学習データを作成する。学習データとしては、関連するインシデント同士で複数のグループに分けられたデータを用意する。

① 初期値の重要度を設定する。
② それぞれの類似度を算出する。
③ 評価値を求める。
④ 重要度を再設定する。
⑤ 評価値が収束するまで②～④を繰り返す。
⑥ 評価値が収束した時点の重要度を採用する。

4. 開発内容

機械学習により重要度を決定する方式について、プロトタイプを開発し、関連インシデントの抽出を行った。実装した関連インシデント抽出方式は、MWS 2014 Datasets の攻撃通信データファイル[1]を入力とし、IDS のフリーソフトである Snort[2]で検知した結果を用いて評価を行った。

5. 評価

以下に、使用したデータの詳細を示す。

表 3 使用したデータ情報

使用したデータセット名		CCC Datasets 2011
使用したデータ	種類	攻撃通信データ
	データの取得日	2010 年 8 月 18 日～8 月 31 日, 2011 年 1 月 18 日～1 月 31 日
Snort を利用して検知した結果数		68415

検知した結果から、アラート発報日、アラート発報時刻、アラート発報ソース、アラート種別、通信元ＩＰ、通信元ポート番号、通信元プロトコル、通信先ＩＰ、通信先ポート番号、通信先プロトコル、アラート名の11項目の情報を持つインシデントを作成し、関連インシデントとして表 4の14のグループを作成し、学習データとして準備した。

表 4 用意した学習データ

No.	種類	種類数
(1)	アラート種別が一致するグループ	4
(2)	発生日時・通信元 IP アドレスが一致するグループ	9
(3)	通信元 IP アドレスが一致するグループ	1

表 5に、学習の結果、算出された重要度を示す。算出された重要度の妥当性を検討するために、表 5の重要度を用いて、関連インシデントを抽出した。その結果、与えた新規のインシデントに対して、アラート種別以外が一致するインシデントが、最も関連があるインシデントとして抽出され、次に関連があると抽出されたインシデントは、アラート種別が一致するインシデントが抽出された。抽出されたインシデントが明らかに関連していないと思われるインシデントではないことから、本方式の妥当性を確認した。

表 5 算出された重要度の一例

項目名		重要度
アラート発報日		7.37
アラート発報時刻		2.76
アラート発報ソース		0.38
アラート種別		7.09
通信で検知の場合	通信元 IP	4.88
	通信元ポート番号	4.54
	通信元プロトコル	9.72
	通信先 IP	2.76
	通信先ポート番号	4.46
	通信先プロトコル	7.11
アラート名		3.93

また、従来の手法である手動で過去に対応したインシデント情報の履歴から検索を行い、新規のインシデント情報と関連するインシデントを検索する方法と、関連インシデントを自動抽出する方法を比較する。従来の手動検索においては、次の作業を行う。

① 検索キーワードの入力
② 検索結果のインシデント情報の内容確認

上記の作業と関連インシデントを自動抽出した場合の検索時間を比較した結果を以下に示す。

表 6 検索時間の比較

No.	作業内容	時間
(1)	手動で手順①②の作業を行う作業を 3 回繰り返す場合	90 秒
(2)	新規インシデントの基本情報を登録する作業(30 秒)＋関連インシデントの抽出時間(6 秒)	36 秒

関連インシデントを自動抽出する方法は、手動で検索する方法に比べて 60%程度検索時間を削減できると見積もることができるため、関連インシデントを自動抽出することで効率的に作業を行うことが可能となる。

6. まとめ

本稿では、SOC 要員の作業効率化のために、関連インシデントを自動抽出する方式を検討した。機械学習から重要度を設定することで、さらなる作業効率化について検討した。

7. 今後の課題

実際に、SOC 要員に対して、関連インシデントの抽出方式の有効性を検証していく。

また、機械学習による重要度の決定についても、実データを使用した検証を行う。

8. 参考文献

[1] 秋山満昭, 神薗雅紀, 松木隆弘, 畑田充弘, “マルウェア対策のための研究用データセット ～MWS Datasets 2014～,” 情報処理学会 CSEC/SPT 合同研究発表会.
[2]Snort. https://www.snort.org/