アナリストにおけるサイバー脅威情報分析強化の研究

Empowering cyber threat intelligence for cyber-analysts

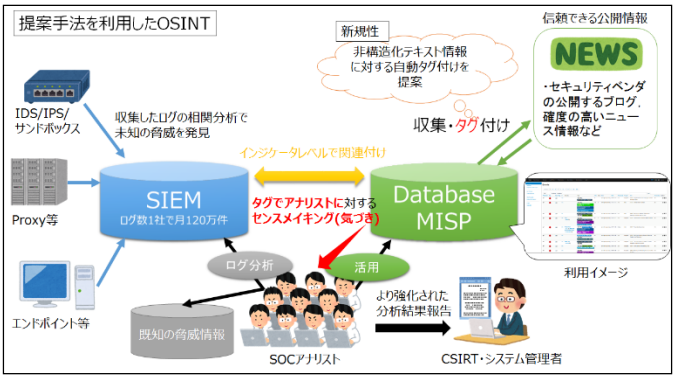
天野 純一郎・ネットワーク分科会・情報セキュリティ大学院大学

Abstract: In recent years cyber attacks have increased and threats posed by them have become complicated. Cyber-analysts face this problem day to day. In such a situation, threat intelligence and OSINT (Open Source Intelligence) in order to support the analysis have been empowered. However, with these OSINT tools, there is a problem that it is not possible to automatically generate tags for texts which are unstructured information. In this paper, we show tagging is applied for cyber-analysts' situation awareness.

はじめに

サイバー攻撃がますます増加し複雑化する中，Threat Intelligenceを活用する必要がある．本稿ではThreat Intelligenceを「複雑化するサイバー攻撃を読み解き適切に対処するための知識と活動」と定義する．Threat Intellgence には公開情報が利用可能である．そこで，公開情報を用いるOSINTに着目し研究を行ったところOSINTには情報源からの情報収集や個別情報の選択・関連付けなどの問題が生じることが明らかになった．OSINTに用いられる既存のOSINTツールについて調査を行ったところ，アナリストにおいてサイバー脅威情報分析強化にこれらのツールを利用できることが分かった．しかし，これらのツールでは非構造化テキスト情報について自動でタグ生成が行えないという問題が存在した．このままでは，アナリストが個別情報を関連付けする際に情報量の問題が生じてしまい，状況認識に問題が生じ分析強化につながらない．本稿では，非構造化テキスト情報に対してタグ付けを行い，アナリストにおける状況認識を高め情勢判断や意思決定に必要な情報の抽出を行い分析の強化を試みた．

提案手法



アナリストが，情報の関連付けを行う際には，証拠に基づいて調査を行うことが示されていた[1]．そのため，分析を強化できるタグ付の方法として，今回は証拠に基づいた5W2Hを抽出することでアナリストの状況認識を向上させることができると提案する．

今回は自動タグ付けのためにCoreNLPのアノテーターの固有表現抽出を主に使用し評価を行った．このCoreNLPはIRC上ハクティビスト活動分析の研究でも利用されている[2]．

評価

今回評価を行うためにあらかじめタグ付けされた記事(n=58)を使用し，タグとの完全一致を比較した．

表 CoreNLPでの自動付与したタグとの一致数

	比較対象のタグ	自動タグ付与	一致した数
総タグ数	216	2425	45
平均タグ数	3.724138	41.81034	0.77586

比較対象のタグで評価したところ，CoreNLPで自動タグ付与したものとの一致率は**20.83%**であった．また，CoreNLP以外の手法によりキーワード抽出を試みた結果，システムAでは21.75%，システムBでは36.57%という結果が得られた．

考察

CoreNLPの固有表現抽出の手法では，ワードのみが抽出された．その結果，比較対象のタグに含まれる1語以上のワードで構成されるフレーズについて抽出することができなかった(例：exploit kitなど)．この問題については，2語以上を組み合わせて表現する必要があるため，現状の固有表現抽出だけでは解決ができない．

本稿で提案したタグの自動生成は，あくまでアナリストが行う情報関連付けを支援するためのーパーツにしか過ぎないが，アナリスが分析を行う際にこのタグを用いて分析を強化に利用することが可能である．

まとめ

サイバー空間で発生する事象を，アナリストが状況認識し情勢判断・意思決定に役立てることが重要である．しかし，サイバー攻撃が増加しそれに伴いアナリストが取り入れる情報量が増加している．またそうした情報をアナリスト間で共有するという問題もあった．本稿では，非構造情報のテキスト情報に対して自動タグ生成を試みた．その結果，人が情報を関連付け・認識しやすい形での自動タグ付けを実行可能なことを明らかにした．今後も多層防御の観点から，機械学習の手法やAIなどを取り入れアナリストの分析強化につながる状況認識の方法について検討していきたい．

[1] Peter Pirolli, PARC, The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis
[2] Jiakai Yutたち, "Automated Framework for Scalable Collection and Intelligent Analytics of Hacker IRC Information", 2016 International Conference on Cloud and Autonomic Computing