STIXを用いた多様化する脅威情報の表現拡張に関する研究

上本 悠貴[1]　岡村 耕二[2]

概要：サイバー攻撃活動は現在も活発に行われている．このような脅威には, セキュリティ対策製品を扱う企業間で情報を共有し対応することが重要である．脅威情報を構造化記述する方式である STIX(Structured Threat Information eXpression) を用いることで, 脅威情報を各企業のセキュリティ対策製品で共有し, タグを利用した機械処理が可能となる．

サイバー攻撃活動は社会環境やセキュリティ対策に合わせて多様化している．それに対し,STIX は予め決まった標準化されているので, 脅威情報の多様化に合わせて頻繁に更新することは困難である．そのため,STIX では表現が困難な脅威情報の出現が懸念される．こちらについて調査した結果,STIX のタグでは表現が困難な情報を自由記述する「Description」タグに記述された文字数が, データ全体に対して相対的に多く,STIX では細かく構造化することが困難な脅威情報の存在を多数確認した．自由記述の割合と, 機械の処理能力は両立しないため, 自由記述の割合が小さく機械の処理能力が向上する方式に STIX を拡張する必要がある．

拡張方式として, 本研究では「Description」タグ内のデータを,JSON 形式で記述する手法を提案した．提案手法により, 従来の STIX との互換性を保持したまま,「Description」タグ内の構造化が可能となり, 自由記述の割合が大幅に減少した．加えて, 既存の方式ではタグを使った機械処理が困難であった「Description」タグ内のデータが, タグを利用して機械処理が容易となった．今後の課題として, 既存の STIX データを提案拡張方式へ変換する箇所の自動化について考察する予定である．

Expansion of scheme representing diverse threat information using STIX

YUKI UEMOTO[1]　KOJI OKAMURA[2]

1. 研究の背景と目的

近年, バンキングマルウェアやランサムウェア, 標的型攻撃メールなどが社会的脅威となり注目されたように, サイバー攻撃活動は活発に行われている. こういった脅威には, セキュリティ対策製品を扱う企業間で情報を共有し対応することが重要である. 脅威情報を構造化記述する方式である STIX(Structured Threat Information eXpression) を用いることで, 脅威情報が構造化でき, 各企業のセキュリティ対策製品での共有や, タグを利用した機械処理が可能となる. サイバー攻撃活動は, 時代毎の社会環境やセキュリティ対策に合わせて攻撃の形を柔軟に変化させ多様化している. それに対し,STIX は脅威情報を記述する方式が予め決まった標準化されているので, 脅威情報の多様化に合わせて頻繁に記述方式を更新することが困難である. そのため, 最近の脅威情報が STIX で細かく構造化できない可能性がある.STIX で最近の脅威情報が記述されたデータを解析した結果,STIX のタグでは表現できない情報を自由記述する「Description」タグに記述された文字数が, データ全体に対して相対的に多く,STIX のタグでは細かい構造化が困難な脅威情報の存在を多数確認した.

構造化データは, 検索エンジンがタグを利用して各データの意味を明確に理解することを可能にする. それに対し, 自由記述のデータはタグがなく, 検索エンジンが各データの意味を明確に理解することは困難である. また, データの内容からその意味を推測することは可能だが, その分処理が必要で, かつ推測した意味がデータの真の意味と一致するとは限らない. つまり, データ内の自由記述の割合と, 機

1　九州大学大学院システム情報科学府　情報知能工学専攻修士 1 年
2　九州大学情報基盤研究開発センター 先端ネットワーク研究部門

© 2018 Information Processing Society of Japan

械の処理能力は反比例する. そのため, 自由記述が少なく構造化記述が多いデータほど機械処理が容易であり, 反対に自由記述が多いデータは機械処理の妨げとなる. 本研究では, 最近の多様化する脅威情報を記述する際, データ全体に対する自由記述の割合を減少させ機械による処理がより可能な方式に,STIX の方式を拡張することを目的とする.

2. STIX と脅威情報

2.1 STIX

STIX は, 米国政府が推進しているサイバー攻撃活動に関連する項目を構造化記述するための技術仕様である. 米国政府の支援を受けた非営利団体の MITRE が中心となり仕様策定を進めており,2012 年 6 月に最初のバージョン 0.3 が作成された.STIX を用いることにより, サイバー攻撃によって観測された事象だけではなく, サイバー攻撃活動を鳥瞰するために必要となる様々な脅威情報の交換を推進できるようになる [9].2015 年 5 月にバージョン 1.2 がリリースされた後,2015 年に国際的非営利目的の協会 OASIS にその統治が移された.2017 年 3 月に OASIS CTI TC(OASIS サイバー脅威インテリジェンス技術委員会) によって, 現在の最新バージョンである STIX2.0 が承認された.

2.2 STIX と多様化する脅威情報

サイバー攻撃活動は, 時代の社会環境やセキュリティ対策に対応して柔軟に変化する. そのため, 脅威情報には流行り廃りがある.2017 年には, バンキングマルウェアや標的型攻撃, ランサムウェアなどが流行した. また, これらの脅威を解析し対策を講じても, その対策をかいくぐる新たな脅威が亜種という形で生じ, 再び猛威を振るう. 加えて, フィッシングメールが 2014 年に流行してから, サイバー攻撃活動の流行が標的型攻撃, ランサムウェアへと移ったように, 脅威情報の変化は激しい. このようにして, 脅威情報は多様化している.

一方で,STIX は脅威情報を記述する方式が予め決まった標準化されているため, 記述対象となる脅威情報の多様化に合わせて記述方式を更新することは不可能である. そのため, 脅威情報の多様化によって,STIX の構造化方式では表現できない脅威情報の存在が懸念される.

2.3 STIX データの調査

脅威情報の多様化が STIX の構造化に与える影響を調査するため,hail a taxii[4] というサイトのオープンサーバーにある STIX で記述されたデータ 71,974 件を対象に調査を行った. データの内訳は（表 1）である. この実験データは, 現在主流である STIX1.x(STIX バージョン 1.0〜1.2) で記述されたものである. そのため, この STIX1.x のリリース以降の 2015 年 6 月から,2017 年 12 月までのデータについて,STIX1.x で構造化可能であったか解析することで, 脅威情報の多様化が STIX に与える影響について調査した.

表 1 調査対象データの内訳

時	件数
2015 年 6 月第 1 週	18,413 件
2015 年 12 月第 1 週	13,865 件
2016 年 6 月第 1 週	7,576 件
2016 年 12 月第 1 週	10,656 件
2017 年 6 月第 1 週	12,990 件
2017 年 12 月第 1 週	8,474 件
合計	71,974 件

STIX には決まったタグがあり, そのタグに沿って記述することで脅威情報を STIX で記述できる. そこで, 各タグに記述された文字数に着目して解析した. 解析結果は（図 1）である.（図 1）は 1 ファイルあたりの, 全体の文字数,「Description」タグに記述された文字数,「Description」タグ以外のタグ 1 つあたりに記述された文字数を,6 ヶ月おきにグラフ化したものである. 横軸は,2015 年 6 月から 2017 年 12 月までの 6 ヶ月間隔の時間軸であり, 縦軸は文字数である.「Description」という自由記述のタグの文字数がその他のタグに比べて非常に多いことが分かる. また, ファイル全体の文字数に対する「Description」タグの文字数が多いことも分かる.

次に, 各ファイルに占める「Description」タグに記述された文字数の割合を基にグラフ化した. 結果は（図 2）である.（図 2）のグラフの横軸は百分率であり, 縦軸は件数である. ファイル全体の 49%を「Description」タグの文字数が占めているものが一番多く,13,678 件であった. 次いで,「Description」タグの割合が 0%のものが,11,610 件であった.「Description」タグを全く必要としないデータも 11,610 件（データ全体の約 16%）あったが, 残りの 60,364 件（データ全体の約 84%）のデータは, ファイル全体の 30%以上を「Description」タグの文字数が占めており, ファイル全体の 80%以上も「Description」タグの文字数が占めているデータも存在することが分かった.

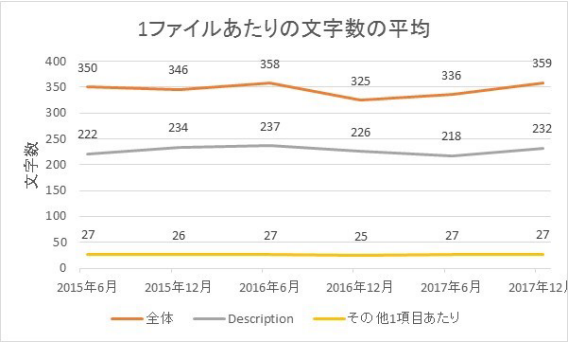

図 1 1 ファイルあたりの文字数の平均

© 2018 Information Processing Society of Japan

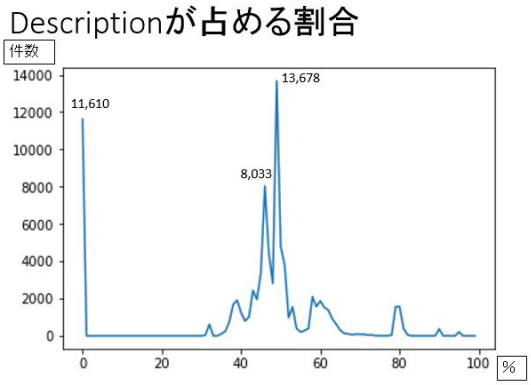

図 2　Description が占める割合

2.4　Description タグと調査結果

「Description」タグの説明として STIX1.2 の仕様書には,「Description タグはオプションであり, 構造化されていない文章記述を提供する.」[7] と, 記載されている.STIX は脅威情報を構造化記述することを目的としている. そのため, 構造化されていない文章記述をする「Description」は, 既存のタグでは表現することができなかった情報を記述するオプションである.

前項で, 最近の脅威情報を STIX1.x で記述したデータの中には,「Description」タグの文字数が文章全体に対して相対的に多いものが多数存在することが分かった. このことより, 現在主流である STIX1.x のタグでは記述することが困難であった多くの情報が「Description」タグに記述されていると考察できる. 調査結果として, 脅威情報の多様化によって,STIX1.x のタグでは表現しきれない脅威情報が多数存在し, 多くの情報が「Description」タグに自由記述されていることが分かった.

3.　拡張方式の手法

3.1　拡張方式の目的

2 章で脅威情報の多様化によって STIX のタグでは表現が困難な脅威情報が多数存在し,STIX で記述されたデータ全体に占める自由記述の割合が多いことが分かった.1 章で記述した通り, 自由記述のデータは構造化データに比べ機械の処理能力が低下する. そのため構造化が不十分なデータ, つまり構造化記述が少なく自由記述が多いデータは, 機械処理の妨げとなる. そこで, 最近の多様化している脅威情報を記述する際, 自由記述を減らし機械での処理がより可能となることを目的とする拡張方式を研究した.

3.2　従来の STIX との互換性

拡張するにあたり, 従来の STIX との互換性を保持することが, 重要になる. 本節では, 下記の 2 点から, 方式を拡張する際の, 方針を定める.

- STIX 文法との互換性
- STIX を扱う環境との互換性

3.2.1　STIX 文法との互換性

従来の STIX 文法との互換性の保持を考慮せず自由記述を減らすことだけを目的とする場合, 不足しているタグの追加をするだけでよい. 例えば, 既存の STIX のタグには URL を記述するタグはあるが, その URL がどのようなサイトであり, どういった悪さをするかといった URL の説明をするタグはない. そのため「Description」タグに URL の説明が記述されていることが多々ある. そこで,URL の説明を記述する「URL_Description」というタグを新たに作成するだけで,「Description」タグ内の自由記述を減らすことが可能である.

しかし, 前述した新たなタグ「URL_Description」の作成のように STIX の文法自体を変更する拡張を行うと, 既存の全ての STIX データ,STIX を扱うプログラムにも新たなタグの追加が必要となる.STIX の文法で記述されたデータや STIX 文法のデータを扱うプログラムは多数存在するため, その作業量は莫大なものになる. 結果として, 拡張することによるメリットよりデメリットの方が大きくなる.

では,「Description」タグの中に,URL の説明を記述する「URL_Description」というタグを新たに定義する場合を考える. この場合, 自由記述をする「Description」タグの内部で新たなタグを定義しているため,STIX 文法は保持されているように感じる. しかし, 従来の「Description」タグが自由記述であるのは, その他のタグが固定されているために表現不可能である種々の情報を「Description」タグに記述するためである. もし,「Description」タグの内部に, 上記「URL_Description」といった新たな固定タグを定義すると,「Description」タグの内部でも表現不可能な情報が生じ, その情報を記述する場がなくなる.

以上のことより, 従来の STIX 文法を保持しながら,「Description」タグの内部で種々の情報が表現可能であることで, 従来の STIX 文法との互換性を保持した拡張をすることが方針となる.

3.2.2　STIX を扱う環境との互換性

STIX は脅威情報を機械で共有, 解析することを目的としている. そのため,STIX で記述されたデータを共有するためのプロトコルや, 解析するためのツールが存在する.

STIX 方式で記述されたデータを共有する方法として TAXII（Trusted Automated Exchange of Intelligence Information）がある.TAXII は,STIX に代表される CTI（Cyber Threat Intelligence）の交換をサポートするため, 特別に設計されたアプリケーション層プロトコルである.TAXII でデータの共有を行う場合, 共有したいデータを TAXII 用のサーバー（以後 TAXII サーバー）に送信し, クライアントが TAXII サーバーにあるデータにアクセスする（図

© 2018 Information Processing Society of Japan

3)[*1].TAXII を使用することで,STIX で記述された脅威情報を簡単に共有することができる.

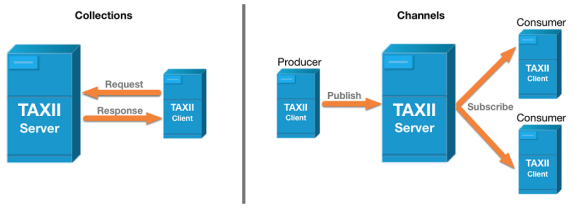

図 3　TAXII サーバーとクライアント

STIX 方式で記述されたデータを解析する手段としては,STIX 用に作成されたツールを使用する方法がある. 例えば, 複数の STIX データから脅威情報を分類してグラフ化することや, 関連する脅威情報をツリー形式で表示すること, 各脅威情報を時系列でまとめたグラフを作成することなどが容易に可能なツールがある. こういった STIX 用に作成されたツールを使用することで, 脅威情報の解析やプログラミング等を専門としない企業/人でも STIX データの解析が可能となる.

本研究による拡張方式によって脅威情報がより構造化され, 機械処理が容易になった場合でも,STIX を扱うために用意された環境が利用不可であれは, 上記した STIX だから得られる利点が失われる. よって, 従来の STIX を扱う環境が利用可能である, つまり従来の STIX を扱う環境との互換性を保持した拡張を行うことが方針となる.

3.3　拡張方式の提案

本節では、前節までで述べた目的や方針を満たす拡張方式の提案を行う.

3.3.1　Description の構造化

STIX で記述されたデータ全体に対する自由記述の割合を減らすためには,「Description」タグ内で自由記述された文章を減らす必要がある. これを実現する手段として,「Description」タグ内に記述された情報を「Description」タグ外に移すことが考えられる. しかし、前節で述べた通り新たなタグを作成し,「Description」タグ外に記述する枠を設けることは方針に反する. よって,「Description」タグに記述された情報を, タグ外に移すことは困難である.

そこで,「Description」タグの内部を構造化することを提案する. 前章で記述した通り,「Description」タグは文章[*2]で自由に記述を行うものであり記述方式に制約はない. そのため,「Description」タグの内部を拡張することは, 既存の STIX 文法に影響しない.

また,2 章で記述した通り,「Description」タグにはデータの多くが記述されている.「Description」タグ自体を構造化すれば,「Description」タグに自由記述されていたこれらのデータを一括して構造化することができ,STIX データの自由記述の割合は大幅に減少する. 構造化が進んだことで,「Description」タグ内の自由記述されていた多くの情報に対しても, タグを利用した機械処理が可能となり得る. よって,「Description」タグの内部を構造化することを提案する.

3.3.2　JSON 形式での構造化

STIX はバージョンにより構造化する形式が異なる. 現在主流である STIX1.x は XML 形式を基にした形式であり, 最新バージョンである STIX2.x は JSON 形式を基にした形式である.XML と JSON のどちらかの形式で「Description」タグの内部を構造化することで,STIX データ全体の形式が揃い, 処理が容易となる.

現在 STIX を統治している OASIS CTI TC(OASIS サイバー脅威インテリジェンス技術委員会) は,STIX のバージョンを 1.x から 2.0 へ改定するにあたり,「XML と JSON の両方にメリットがありますが,CTI TC は,JSON がより軽量で, サイバー脅威情報のセマンティクスを表現するのに十分であると判断しました. 使用するのが簡単であり, 開発者がより親しみやすいものです.」[1] と公表している.STIX データは, 機械処理を前提としているため,STIX データが難解であるほど開発者への負担が大きい. そのため, 使用が簡単で開発者の負担が小さく, サイバー脅威情報のセマンティクスを表現するのに十分である JSON 形式で「Description」タグの内部を構造化することを提案する.

3.3.3　key の作成

STIX1.x は, サイバー攻撃活動 (Campaigns), 攻撃者 (Threat_Actors), 攻撃手口 (TTPs), 検知指標 (Indicators), 観測事象 (Observables), インシデント (Incidents), 対処措置 (Courses_Of_Action), 攻撃対象 (Exploit_Targets) の 8 つのオブジェクトから構成されている. 各オブジェクト毎に使用可能なタグは固定である. 例えば（図 4)[*3]のように,「検知指標（Indicator)」オブジェクトでは下線が引かれた Title や ID などが使用可能なタグとして固定されている.

STIX 言語はタグが固定されていることで, 解析や共有が行いやすい. その反面, 脅威情報の多様化に対応するのが難しく, 固定したタグで表現できない情報が増加した.「Description」タグの内部では, 固定したタグによる記述が困難なあらゆるデータを構造化しなければならない. そこで,JSON 形式でタグの役目をする「key」をあえて定義せず, 記述するユーザーが自ら作成する方式を提案する. これによって, 脅威情報の多様化の影響を受けることのない構造化を可能とする.

*1　Introduction to TAXII より [5]
*2　ここでいう「文章」とは, プログラム上での型が string 型であることを示す.
*3　STIX1.2 — STIX Project Documentaion より [7]

© 2018 Information Processing Society of Japan

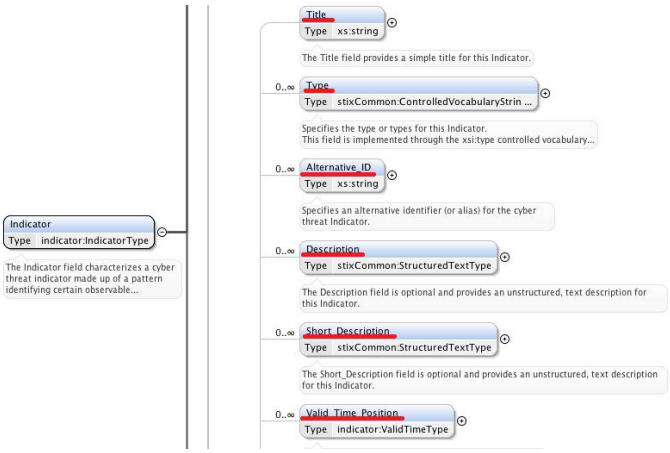

図 4　STIX で使用可能なタグの一例

3.3.4　key の曖昧さの排除

STIX はタグを固定し, 各タグの定義をすることで, タグとその内部の意味を確実なものにしている. 例えば, タイムスタンプを記述するタグに「valid_from」がある. このタグは,「記述内容が有益な情報であると判断された時間」という定義がある. そのため, このタグとタグ内のタイムスタンプから, タイムスタンプの意味が明らかとなる.

しかし, 本研究の拡張方式では, 前節でタグの役目をする key をユーザーが自ら作成することを提案している. そのため, タイムスタンプを記述する key を「timestamp」とした場合, タイムスタンプが何を意味するかは製作者にしか分からない. また, 製作者が key をより詳細に「valid_from_time」と記述した場合でも,key の定義がされていないため, 読み手によって受け取り方が変わる. これは,key に定義がされてないため生じる.

そこで,key の定義を「Description」タグ内に記述することで,key の曖昧さを排除する.

3.4　Description 構造化手法の提案

前節で,STIX 文法を保持しつつ,STIX データ内の自由記述の割合を減らし構造化を十分可能にする拡張方法として以下のことを提案した.

- 「Description」タグの内部を構造化する.
- 構造化は JSON 形式で行う.
- タグの役割をする key はユーザーが作成する.
- key の定義をする.

本節では提案した各拡張方式をもとに本研究の提案手法を述べる.

3.4.1　Description 構造化手法

本研究では STIX の拡張手法として,「Description」タグを（図 5 ）の形式で記述することを提案する.

まず, 波括弧（{）を記述する. 次に, タグの役割を果たす key を自ら作成し,key に合った value を作成する. 上記 2 つを,[”key” : ”value”] の形式で記述する. 作成すべき key がまだある場合, カンマ（,）を記述後,key の作成部分に戻り, 繰り返し記述を行う. 作成すべき key がない場合,[”key_description”: { }] と記述し,key の定義をする場所を作成する. この波括弧の中で key の定義を行う. 自ら作成した key の一つを選択し、その定義を [”key” : ”key の定義”] の形で記述する. 作成した key の中でまだ定義していないものがある場合, カンマ (,) を記述後,key の定義に戻り繰り返し記述を行う. すべての key を定義した場合, 波括弧（}）でとじる.

```
<Description>
[
 “key1” : “data1” , “key2” : “data2” , “key3” : “data3” ,
 “key_description” : {
        “key1” : “key1の説明” ,
        “key2” : “key2の説明” ,
        “key3” : “key3の説明” }
]
 </Description>
```

図 5　Description 構造化手法

実データの「Description」に適応させた例を（図 6 ）に示す.

```
<Description>{
”URL1”:”https://qponn.net/access/7413904f01f8e181b81aaacb2583be17/”,
”timestamp1”:”2017-11-30T21:33:50+00:00”,
”URL2”:”http://www.phishtank.com/phish_detail.php?phish_id=5333220”,
 “key_description”:{
        “URL1” : “This URL is identified by phishtank.com as part of a phishing
        “URL2” : “This URL is this file author,  phishtank.com.”,
        “timestamp1” : “URL1 appears to still be online as of  timestamp1”}}
<Description>
```

図 6　実データへの適応例

4.　拡張方式の評価

4.1　評価基準

本研究の目的は, 最近の多様化している脅威情報を STIX で記述する際, 文章全体に対する自由記述の割合を減少させ機械処理がより可能な方式に拡張することである. そのため, 以下 2 つのことが評価基準となる.

- 自由記述の割合をどれだけ減少させることができたか
- 機械処理がどれだけ可能となったか

さらに, 拡張を行う上で 2 つの方針を定めた. この方針を満たしていることも評価基準となる.

- 従来の STIX との互換性を保持する

以上の 3 つを評価基準として, 提案手法である「Description 構造化手法」を評価する.

4.2　評価方法

本節では, 前節で挙げた 3 つの評価基準についてそれぞれの評価方法を述べる.

4.2.1　自由記述の割合の減少度合い

2 章で調査を行った hail a taxii[4] のオープンサーバーには, 記述したセキュリティベンダー毎に 8 つのフォルダで STIX データが分類されている. このフォルダ群の中で,2015

© 2018 Information Processing Society of Japan

年 6 月 1 週目以降に更新があった下記 5 つのフォルダの STIX データに対して「Description」タグを提案手法で記述したものに置き換える. 置き換える前後の,STIX データの自由記述の割合の減少度合いによって評価を行う.

- guest.Abuse_ch
- guest.blutmagie_de_torExits
- guest.MalwareDomainList_Hostlist
- guest.Lehigh_edu
- guest.phishtank_com

4.2.2 機械処理の向上

今回の拡張で,「Description」タグ中のデータを新たに構造化した. そのため,「Description」タグについて機械処理がどれだけ可能となったかを評価する. 評価方法としては, まず既存の STIX1.x で記述されたデータと「Description」タグのみを拡張方式で書き換えたデータそれぞれの,「Description」タグを表示するプログラムを Python で作成する. それぞれのデータで「Description」タグの中に記述された「危険だと判断された URL の値」を確認する. その時の機械処理可能な度合いを比較して評価する.

4.2.3 従来の STIX との互換性の保持

STIX の仕様策定を進めた MITRE が「stix-validator」[8] という, データが STIX 文法で正しく記述されていることを確認する API を公開しているため, 初めにそちらで評価を行う. 評価方法については, まず「stix-validator」で既存の STIX1.x で記述されたデータが STIX1.x 文法で正しく記述されているか確認する. 正しく記述されていた場合, そのデータの「Description」タグのみを拡張方式で書き換え, 再度「stix-validator」で確認する. 書き換えたデータが STIX1.x 文法で正しく記述されていると確認できるかで評価する. また,「Description」タグを拡張方式で書き換える際,「Description」タグ内に拡張方式で記述が困難な情報の有無を確認する.

3 章で述べた通り,STIX には共有するためのプロトコル, 解析するためのツールがあるため, 次に拡張方式で記述したデータでこれら 2 つが利用可能であるか評価をする. 具体的な評価方法として, 共有するためのプロトコルは, 実際に拡張方式で記述したデータで TAXII サーバーへのアップロード/ダウンロードを行う. この結果から,STIX を共有するプロトコルが拡張方式でも使用可能かを評価する. 解析するためのツールは,STIX の仕様策定を進めた MITRE が公開しているツール「python-stix」を拡張データに対して実際に使用することで評価する.

4.3 拡張方式の評価結果

本節では, 前節までで挙げた評価基準, 評価方法をもとに 3 章で提案した拡張方式を評価した結果を述べる.

4.3.1 自由記述の割合の減少

前節であげた 5 つの対象フォルダ内の STIX データ 6 件ずつ, 計 30 件の STIX データに拡張方式を適応した結果,「Description」タグに自由記述されていたデータをすべて JSON 形式で記述することができた. よって,「Description」タグに記述されていた自由記述の文章はすべて構造化することができた.

4.3.2 機械による解析の向上

まず, 既存の STIX1.x で記述されたデータの「Description」タグ内を, 自己作成プログラムを使って確認する. プログラムの実行結果は（図 7）である.（図 7）より,「Description」タグに記述された文章を表示することは可能だと分かる. しかし, この文章の中には URL が 2 つあり, どちらが「危険だと判断された URL」であるかは, 人が文章を読み解釈する必要がある.

```
This URL:[https://qponn.net/access/7413904f01f8e181b81aaacb2583be17/] was identified by phishtank.com as
part of a phishing email. This URL appears to still be online as of 2017-11-30T21:33:50+00:00. More deta
iled infomation can be found at http://www.phishtank.com/phish_detail.php?phish_id=5333220
```

図 7 拡張方式適応前の Description

それに対し, 拡張方式で記述されたデータの「Description」タグ内を, 自己作成プログラムを使って確認する. プログラムの実行結果は（図 8）である.（図 8）は, 上から,「Description」タグ内の key を表示した結果,key の定義を表示した結果,「URL1」という key に対応するデータを表示した結果である. 一番上の key の一覧から,「Description」タグの中に「URL1」「URL2」という key があることが分かる. 真ん中の key の定義から,「URL1」が「危険だと判断された URL」を記述する key であることが分かり, この key に対応するデータが目的の URL であることが分かる.

```
-----Key の一覧-----
dict_keys(['URL1', 'timestamp1', 'URL2', 'key_description'])

-----Key の説明-----
['URL1': 'This URL is identified by phishtank.com as part of a phishing email',
'URL2': 'This URL is this file author, phishtank.com.',
'timestamp1': 'URL1 appears to still be online as of timestamp1']

URL1のデータを参照
https://qponn.net/access/7413904f01f8e181b81aaacb2583be17/
'
```

図 8 拡張方式プログラム実行結果

以上のことより, 既存の STIX1.x では機械で扱うことのできなかった「Description」タグに対して, 拡張方式を使用することで「Description」タグ内の key を表示し, 指定した key のデータを表示することが可能となった. これにより, 人が長文を読み内容を理解するという手間を省くことができた.

4.3.3 従来の STIX との互換性の保持

まず, 上記 4.3.1 で使用した３０件の拡張前の STIX データについて,「stix-validator」で STIX 文法を保持していることを確認した. 次にこれらのデータの「Description」タグのみを拡張方式に換えたデータについて「stix-validator」で確認した結果, 文法を保持していることが確認できた. よって, 拡張方式で記述した STIX データが,STIX1.x 文法

© 2018 Information Processing Society of Japan

を保持しているといえる. また,「Description」タグを拡張方式に換える際, すべての情報を「Description」タグ内に記述することができた.

次に,EclecticIQ というサイトの TAXII テストサーバー [2] に, 拡張方式で記述したデータをアップロードし, その後このデータをダウンロードした. 正常に行うことができたため, 共有するためのプロトコルは拡張方式でも使用することが可能である.

最後に, 拡張方式で記述した STIX データで「python-stix」を使用した. 正常に動作したため, 解析するためのツールも拡張方式で使用できることが確認できた.

4.4 考察

拡張方式はキーを製作者が作成できるため, 評価で扱った 30 件のデータ全てで「Description」タグを構造化することができた. そのため,2 章で調査したすべての STIX データについても「Description」タグを構造化することができると仮定すると,「Description」タグを必要としていた 60,364 件（実験データ全体の約 84%）のデータの「Description」タグで自由記述されていた部分を全て構造化できると考えられる.

key を固定しないため,STIX データの製作者が違えば key の名前も違う. そのため, データ間でのタグを使用した機械処理は本研究の拡張方式を用いても困難である. しかし, 各データの「Description」タグの内部については, タグを使った検索が可能となった. 脅威情報の多様化により「Description」タグの内部の情報が非常に多い現状で, 雑多なデータ群にタグ付けがされており, それをもとに情報の検索が可能であることは貴重であると考える.

データの共有や解析についても, 既存の STIX1.x で記述されたデータと同様に行うことができることから, 従来の STIX との互換性を保持していると考察する.

5. おわりに

5.1 まとめ

本研究では, 最近の多様化している脅威情報を記述する際, データ全体に対する自由記述の割合を減少させ機械処理がより可能な方式に拡張することを目的に,「Description」タグを JSON 形式で記述することを提案した. そして, 提案した拡張方式が本研究の目的や方針を満たすものであるか評価を行った. その結果として, 従来の STIX1.x との互換性を保持したまま, 実験データ全体の約 84%を占めている「Description」タグ内の自由記述をすべて構造化し, データ毎の「Description」タグ内でタグを利用した検索を可能とする方式であることが分かった.

5.2 今後の課題

本論文では, タグの役割をする key を記述者が考え定義することを提案した. 今後の課題として,「Description」タグに記述されたデータの内容から key や key の定義を自動で作成可能とする方法を考える予定である.

謝辞

本研究の一部は科研費 (基盤研究 (C),JP16K00480) の助成を受けたものである.

参考文献

[1] Comparing STIX 1.X/CybOX 2.X with STIX 2 [Internet]. [cited 2017 年 12 月 18 日]. Available at: https://oasis-open.github.io/cti-documentation/stix/compare

[2] EclecticIQ — TAXII Test Server [Internet]. [cited 2018 年 2 月 18 日]. Available at: https://test.taxiistand.com/

[3] FAQ [Internet]. [cited 2018 年 2 月 10 日]. Available at: https://oasis-open.github.io/cti-documentation/faq.html

[4] hail a taxii [Internet]. [cited 2017 年 11 月 26 日]. Available at: http://hailataxii.com/

[5] Introduction to TAXII [Internet]. [cited 2017 年 11 月 26 日]. Available at: https://oasis-open.github.io/cti-documentation/taxii/intro

[6] python-stix: A Python library for parsing, manipulating, and generating STIX content [Internet]. The STIX Project; 2018 [cited 2018 年 1 月 7 日]. Available at: https://github.com/STIXProject/python-stix

[7] STIX 1.2 — STIX Project Documentation [Internet]. [cited 2017 年 9 月 3 日]. Available at: http://stixproject.github.io/releases/1.2/

[8] stix-validator: Validate STIX and CybOX XML documents [Internet]. The STIX Project; 2018 [cited 2018 年 2 月 18 日]. Available at: https://github.com/STIXProject/stix-validator

[9] 脅威情報構造化記述形式 STIX 概説 : IPA 独立行政法人 情報処理推進機構 [Internet]. [cited 2017 年 7 月 10 日]. Available at: https://www.ipa.go.jp/security/vuln/STIX.html

© 2018 Information Processing Society of Japan