オープンソースインテリジェンスと深層強化学習によるサイバー脅威分析手法の検討

川北 将[†1]　島 成佳[†1]

概要：昨今，サイバー攻撃による被害が各国で発生し社会問題となっている．サイバー空間において攻撃側はあまたある手段から任意の一つを用いるだけで攻撃が成立するが，防護側はすべての脅威を認識して攻撃を防がなければならない課題があった．そのため，防護を業務とする分析者にはコンピュータからネットワークまでのアーキテクチャ，サーフェスウェブからディープウェブまでのコミュニティ等の幅広い知識や運用経験が求められ，人材不足を引き起こす要因の一つであった．本稿では既存の分析結果から分析者によるサイバー攻撃に関する脅威分析ノウハウを深層強化学習によって学び，脅威の端緒を新たに発見したとき，オープンソースインテリジェンスを活用して脅威の全貌を暴き出す手法について検討する．

キーワード：サイバーセキュリティ,脅威情報,ソーシャルメディア,オープンソースインテリジェンス,強化学習,ディープラーニング

Investigation of Cyber Threat Analysis Method Using OSINT and Deep Reinforcement Learning

Masaru Kawakita[†1]　Shigeyoshi Shima[†1]

Abstract: In recent years, cyber-attacks have occurred all over the world and becomes a social problem. The adversary could make the attack succeed with a vector but there was a problem that the protection side had to prevent all attacks by recognizing all the threats. For that reason, cyber security analysts engaged in protection were required to have broad knowledge and operational experience such as architecture among computers and networks, community from the surface web to the deep web, and it was one factor that caused human resources shortage. In this paper, when learning threat analysis expertise on cyber-attacks by analysts through deep reinforcement learning from existing analysis results and newly discovering the trail head of threats, consider a method of investigating a threat using open source intelligence.

Keywords: cyber security,threat information,social media,open source intelligence,reinforcement learning,deep learning

1. はじめに

昨今，サイバー攻撃による被害が世界各国で発生し，社会問題となっている．しばしば標的となる重要インフラ，政府機関および民間企業を滞りなく運営するには，各組織を防護するセキュリティアナリストが常日頃からサイバー攻撃の予兆となる膨大な脅威情報を収集・分析し，事案発生が予期された場合にどのような脅威があるかを知る必要がある．

組織内で利用しているソフトウェア・ハードウェアに新たな脆弱性が存在することを認識した場合，パッチの適用や通信遮断などの適切な対処をせずに放置するとその脆弱性を用いたサイバー攻撃を受けて，組織内だけでなく顧客や無関係の組織にまで機密情報の窃取やマルウェアへの感染等の被害が発生する恐れがある．一例として 2017 年 5 月中旬に世界的な被害をもたらした身代金型ランサムウェア「WannaCry」は OS の脆弱性を突く攻撃ツール「Eternalblue」によって拡散した事案[1]が挙げられる．

サイバー脅威をもたらす攻撃者は自ら生み出すほかにもソーシャルメディアやアンダーグラウンドマーケットから攻撃ツールおよび未公開脆弱性の情報を得て，攻撃活動を行なっている．特に APT 攻撃者はサイバー攻撃を行なう明確な目的とその遂行能力を有し，組織化され，豊富な資金力や経験を持ち合わせているとの報告[2]がある．また，DDoS を代行する Booter や Stresser，および，被害者のファイルを人質にとる身代金型ランサムウェアを広く散布して，被害者から金銭が支払われた場合に依頼者へ収益の一部を分配する RaaS (Ransomware as a Service)をその代表例とするサイバー攻撃を請け負うサービスもまた台頭しているとの報告[3]もある．このようなサービスは一連の攻撃行動を自動化していることや攻撃にかかるコストの少ないことが特徴である．

一方，防護側では以下に挙げる理由から，人手によるサイバー脅威の分析に係るコストが増大し，十分な対処が取

†1 日本電気株式会社 セキュリティ研究所
Security Research Laboratories, NEC Corporation

れない状況と考える．

- インダストリー4.0 の到来により防護すべき対象が IT だけでなく OT へも拡大した．[4]
- サイバー犯罪の検挙件数や警察への相談件数が年々増加傾向にある．[5]
- 脅威情報の拡散するソーシャルメディアの情報流通量は 2005 年から 2014 年までの 9 年間でおよそ 9 倍に拡大した．[6]
- セキュリティ技術者の不足が叫ばれる[7]が，システムの構築に関する幅広い知識を必要とすることから一朝一夕に人材を育成できない．

サイバーセキュリティにおいて，攻撃側は流布するサイバー脅威の一部を用いて標的組織の弱点を突くのみで目的を達成可能だが，防護側は弱みとなりうる箇所を全て把握し，脅威へ迅速に対処する必要がある．そのため，分析者にはカーネルからアプリケーションまで，さらにネットワークの深い知識が求められ，鮮度も問われることから学習コストが増大しがちである．

サイバー脅威の分析はセキュリティオペレーションの一部として定義される[8][9]が，一方，そのオペレーションフローは属人性が高く，明文化しにくい課題がある．ソフトウェア・ハードウェアが生成する大量のログから攻撃の痕跡を発見し，また，断定する基準は分析者の頭にのみ存在し，どのような基準で判断したかを当人が自覚していないこともしばしばある．基準が無数にあり，また，複数の基準を分析者の経験から得られた重み付けで組み合わせて総合的に判断しているとみられる．

近年，TAXII (Trusted Automated eXchange of Indicator Information) [10]による脅威情報共有が産学官の垣根を越えて広まりつつある．しかし，共有されるのは断片的な痕跡情報であることが多く，その活用にあたって，セキュリティアナリストが複数の脅威情報を組み合わせて脅威分析を行う必要が生じた．

このような背景から，効率的なサイバー脅威の分析手段が社会において求められていた．

本稿では，2 章でセキュリティアナリストが手動で行う脅威分析作業について述べ，3 章でその自動化を目的としたオープンソースインテリジェンスと深層強化学習を活用したサイバー脅威の全体像分析手法について提案する．4 章にて実装を行ない，最後に結論を述べる．

2. セキュリティアナリストの脅威分析行動

セキュリティオペレーションに従事するアナリストは常日頃からソフトウェア・ハードウェアが生成するログやインターネット上の情報をソースとするモニタリングを行ない，サイバー脅威の端緒となるサンプルを得る．

セキュリティアナリストの分析能力の 1 つには，図 1 に示す，脅威の端緒となるサンプルからオープンソースインテリジェンスを活用するフィルタを連鎖的に組み合わせてサンプルを多段変換し，最終的に当業者にとって価値のあるサンプルセットを得る分析作業を的確に遂行する能力があると我々は考える．

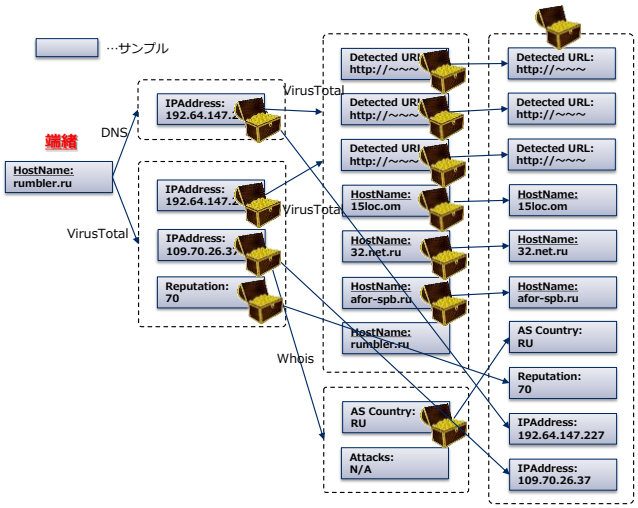

図 1 オープンソースインテリジェンスを活用した分析

例えば，ログであれば，普段アクセスしない接続先へ通信した等の日常の通信には見られない変化を捉え，意図しない通信先である疑いのある IP アドレスをリストアップし，インターネット上の情報であれば，ソーシャルメディアやアンダーグラウンドマーケットから，サイバー攻撃者らによる犯行予告や，ハクティビストによる攻撃の勧誘，未知脆弱性やアカウント情報の売買状況を把握し，標的組織のホストネームや使用する攻撃ツール・マルウェア・エクスプロイトキットをリストアップする．

このようにして得た情報が直ちに対処へ役立つわけではない．セキュリティアナリストの脅威分析によって組織が攻撃のターゲットとなっていたことが事後に判明し，使用されたマルウェアの情報を得たと仮定する．マルウェアのファイル名やハッシュ値等の特徴情報から，組織内の端末に存在したかどうかを検証することがまず考えられる．しかし，現実にはエンドポイントに検証に必要なログが残されていない場合や，自己削除によって痕跡のない場合がある．このような場合，エンドポイントの外にあるプロキシサーバーのログを参照することでマルウェアの存在有無を検証する手段が代替案の一つとして考えられるが，そのマルウェアを調査・分析し，通信先の IP アドレスが明らかとなっていなければ検証できない課題が生じる．

一般に，セキュリティアナリストは，マルウェアの挙動を動的解析によって自ら調べることのほか，オープンソースインテリジェンスを活用して，他のアナリストや研究者，あるいは自動分析によって同一あるいは類似するマルウェアの解析結果から通信先の IP アドレスが判明していない

かを複数のソースを組み合わせることによって調査する．例えば，VirusTotal [11]はマルウェアのハッシュ値を検索キーとして，悪意ある通信先の URL の情報を得られる仕組みを提供している. URL に含まれるホストネームから DNS サーバーあるいは過去の DNS レコードを参照する外部サービスをフィルタとして活用することで対応する IP アドレスが得られる．

3. 深層強化学習による脅威の自動分析

本章では前章で示したセキュリティアナリストの脅威分析について，そのモデル化と，深層強化学習による自動分析手法について述べる．

3.1 脅威分析のモデル化

サイバー攻撃の端緒を指す IP アドレス，ドメインネーム，ホストネーム，URL，ハッシュ値等のセキュリティ分野において有用かつ特徴的な情報を含むサンプルセットをs, オープンソースインテリジェンスを活用してサンプルを異なるサンプルセットへ変換する写像をfと定義する．

このとき，iステップ目の脅威分析を式(1)と表現する．入力となるサンプルセットと，それを写像によって変換したサンプルセットの論理和を，次ステップの入力として用いることで，セキュリティアナリストがオープンソースインテリジェンスを活用しながら対処に必要な情報を徐々に充実させる過程を表現する．

$$\boldsymbol{s_{i+1} = s_i \cup f_i(s_i)} \tag{1}$$

写像fの取りうる値の集合をTとしたとき，Tは単一のフィルタだけでなく，複数のフィルタの組み合わせも考慮する必要がある．なぜなら，プリミティブな変換機能を持つフィルタVおよびWがあったとき，それらが入力として受け取ったサンプルセットをどのように扱うかはフィルタ自身に委ねられるため, V と W を同時に適用して論理和を取る式(2)の結果Xと，V の適用結果を W の入力として逐次的に適用する式(3)の結果X'を比較したとき，両者がいかなる場合でも同一であることは保証されないためである．

$$X = s_i \cup V(s_i) \cup W(s_i) \tag{2}$$

$$X' = (s_i \cup V(s_i)) \cup W(s_i \cup V(s_i)) \tag{3}$$

脅威分析とはs_0に端緒を与え，式(1)の反復によってサンプルセットの内容を充実させながらs_nを得ることである．自動的な分析においては人力で分析した結果との誤差が少ないs_nほど優れている．

より良いs_nを得るためには，各ステップで適切な写像fを選択する必要がある．入力のサンプルセットs_iからどのような集合Tに含まれるフィルタの組み合わせを選択すれば，最終的に人力で分析した結果との誤差が少ないs_nを得られるかを各ステップで考慮しながら探索する．

3.2 強化学習の適用

強化学習[12]とはエージェント(行動主体)が環境の状態を観測しながら取るべき行動を決定する問題を扱う機械学習の一種である．エージェントは可能な行動セットの中からある行動をとることで環境を変化させて報酬を得る．一連の行動を通じて，最も高い報酬が得られる方策を学習する．

強化学習の手法として単純な Q 学習[28]が知られるが，Mnih ら[13]はディープニューラルネットワークと強化学習を組み合わせた深層強化学習の導入によって，ビデオゲームで高い報酬を得ることに成功しており，本稿でも同様の手法を取り入れる．

強化学習の適用にあたっては学習過程での報酬の与え方が重要である．サイバー脅威の分析はセキュリティ対策への活用を目的としており，適切な情報を含むサンプルセットが分析結果に求められる．よって，ある脅威に関する所望の分析結果に含まれるサンプルセットGと強化学習エージェントが算出するサンプルセットs_nの誤差が小さくなるほど高い報酬を設定すべきといえよう. 誤差が大きくなる，すなわち, サンプルセットが極小ないし大量である場合や，重要度の高いものの見逃しや低いものが含まれる場合，セキュリティ対策への活用が難しい．そもそもサイバー脅威とは直ちに深刻な被害を及ぼすものではない場合もあるため，将来的に自組織の被害につながるか否かの判断は組織内のシステム構成やワークフロー，セキュリティアナリストや経営層の見解, 情報元の信頼性によって異なる. また，脅威情報は膨大であるため単純な情報の紐づけでは数万もの悪性 IP アドレス等の要素が列挙され, そのままセキュリティ対策へ利用すると過剰な防護によって通常業務に影響をもたらす可能性がある．

例えば，ある脅威について分析した結果，ある TLD (Top Level Domain)を持つドメインすべてに関連する IP アドレスが悪性であるという結果を得た場合，これを抑止するべき通信としてファイアウォールへ適用すると正常な通信までもが遮られることは明らかである. これを防ぐためには，適切な報酬関数と良質なGで学習を行なう必要があると考える．

既知の分析結果からサイバー脅威をどのような手順で分析したかを機械的に学習し，新たな脅威を検出した際にその学習結果に基づいて機械的に分析することで，組織固有の基準かつ適切な分量での脅威の全体像を把握できる．

3.3 学習フェーズ

① 既知の分析結果から，マルウェアのハッシュ値，ファイル名，発信源・通信先の IP アドレス，URL，ドメインネーム，ホストネームを正規表現または文書構造の解析によって抽出し，サンプルセットを得る．
例えば，図 2 に示す，過去に判明したある脅威に関する分析レポートでは，マルウェアのファイル名やハッシュ値，動的解析の結果から判明した通信先 IP アドレスが，STIX[14]によって構造化された状態で 10 数個列挙されている．STIX のスキーマに従って，type 属性が FileObjectType, AddressObjectType, URIObjectType, DomainNameObjectType，および，HostnameObjectType のノードを抽出し，所望のサンプルセットを得る．

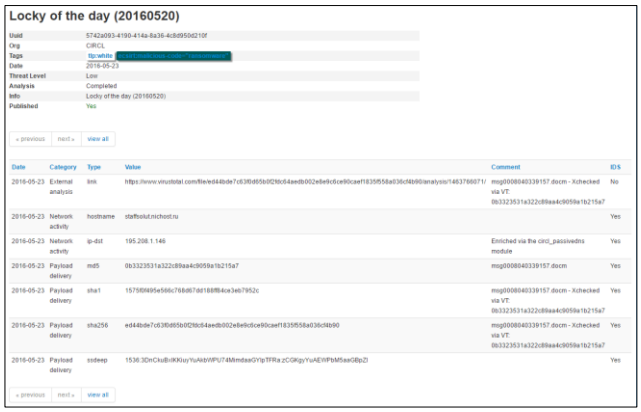

図 2 STIX で構造化された分析結果の例 (ルクセンブルク CERT[16]発行のレポート)

② 単一のレポートに含まれるサンプルセットのうち，ランダムに 1 つのサンプルを取り出して，端緒のサンプルs_0とする．

③ ランダム，または，後述する Q 値の最も高いものから少なくとも 1 つ以上のフィルタを含む写像を選択し，入力サンプルをパラメータとしてフィルタらを実行した結果の和集合を出力サンプルとする．また，式(4)によって報酬rを求める．rはある時点tまでに求めたサンプルセットs_tとゴールのサンプルセットs_nのうち，一致する項目の数の全体に占める割合とする．ただし，一致しないもしくは過剰である項目の割合に 0.1 を乗じたものをペナルティとして報酬から減算する．

$$r(t) = \frac{num(s_t \cap s_n)}{num(s_n)} - 0.1 \times \frac{num(s_t) - num(s_t \cap s_n)}{num(s_n)} \quad (4)$$

ホストネームの端緒から DNS を用いて IP アドレスへ変換する様子を図 3 に，さらに VirusTotal から得た情報でその IP アドレスを通信先とするマルウェアのハッシュ値を得る様子を図 4 に示す．

図 3 DNS を参照したときの報酬例

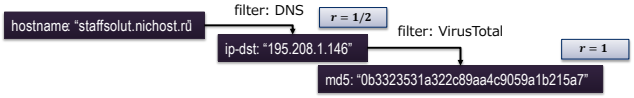

図 4 DNS および VirusTotal を逐次参照したときの報酬例

④ 式(5)によって Q 値を求める．長いステップ数を経て得られる報酬を低く評価するための割引率γは 0.9 とする．時刻tにおいて，s_tは入力ノードの内容を示し，行動a_tはどの写像を選択したか，$Q(s_t, a_t)$は以前に評価した Q 値のうち引数が一致するものであり，初期値を 0 とする．

$$Q(s_t, a_t) = r + \gamma \times \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) \quad (5)$$

⑤ 入力層へ渡すデータを固定長にする目的で，サンプルセットを JSON フォーマット[15]へシリアライズし，写像に含まれるフィルタ群を ID へ変換する．(図 5) 写像には 1 つ以上のフィルタが含まれ，各フィルタには一意な ID を事前定義する．使用したフィルタに含まれる各フィルタの ID をF，総数をnとしたとき，式(5)によって写像の数値表現pを得る．

$$p = \sum_{i=0}^{n} 2^{F_i - 1} \quad (5)$$

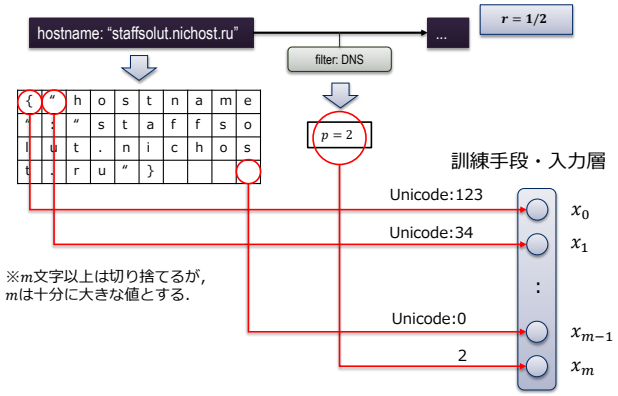

図 5 サンプルセットと行動の数値化

⑥ 入力層に与えた数列を 4 層の中間層を持つニューラ

ルネットワークの各層へ伝搬させ，出力層の数列を Q 値として得る．

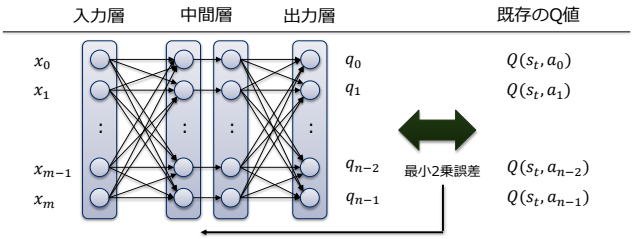

図 6 バックプロパゲーション

⑦ 従前までに求めた$Q(s_t, a_t)$との最小 2 乗誤差を算出し，ニューラルネットワーク上でバックプロパゲーションを実行する．（図 6）

⑧ ②～⑦の手順を 1 エピソードとし，学習に必要な回数だけ反復する．所定の回数を反復するか，または，$|r| \geq 1$を満たした時点で反復を終える．

⑨ 結果として，Q 値の配列，および，学習済みのニューラルネットワークを得る．

3.4 評価フェーズ

① 学習済みニューラルネットワークの入力層に初期サンプルを投入する．

② 出力層の数列のうち，最も大きな Q 値を持つインデックスを選択し，それを取るべき行動の ID として，学習フェーズと同様に写像に含まれるフィルタを実行し，報酬を得る．

③ ②の手順を反復する．所定の回数を反復するか，または，$r \geq 1$を満たした時点で反復を終える．

④ 結果として，脅威分析の過程を表し，反復回数に等しいサンプルと写像と報酬の組を持つ列を得る．（図 7）

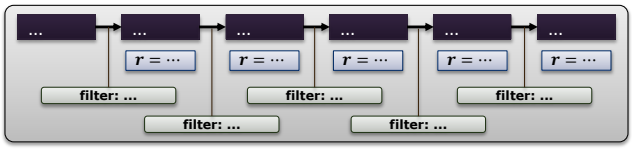

図 7 サンプルと写像と報酬の列

4. 実装および実験

3 章で述べた提案手法を Python 言語および主なライブラリとして chainer[25], numpy[26], cupy[27]を用いて実装し，表 1 に示すスペックのマシンで実行した．

2016 年 4 月から 2017 年 6 月までにルクセンブルク CERT[16]が公開した脅威分析レポートのうち無作為に 1,000 件を抽出したデータセットから，単純な Q 学習と深層強化学習を用いて 150,000 エピソードの反復学習を行なったときに得る報酬の推移を観察する実験を行なった．なお，オープンソースインテリジェンスを活用するフィルタとして，DNS, Whois, VirusTotal[11], Internet Storm Center[17], PassiveTotal[18], Shodan[19], Censys[20], ThreatCrowd[21], ThreatExpert[22], Open Threat Exchange[23], NSRL Database[24]を用いた．

その結果，図 8 に示す学習結果を得た．グラフは横軸に反復数，縦軸に報酬とした学習の進捗状況を示す．深層強化学習を用いた場合は反復数がおよそ 9 万を超えるまでは単純な Q 学習よりも低い報酬しか得られなかったが，その後は安定した高い報酬を得る傾向にあった．よって，脅威分析のモデル化と深層強化学習の適用による，分析の自動化は可能である．

表 1 マシンスペック

項目	スペック
CPU	Intel Core i7 6850K 6cores
Chipset	Intel X99 Express
Memory	32GB
HDD	2TB (SATA, Non-RAID)
GPU	4 x Geforce 1080 GTX
NIC	Broadcom Corporation NetXtreme
OS	Ubuntu 14.04 LTS

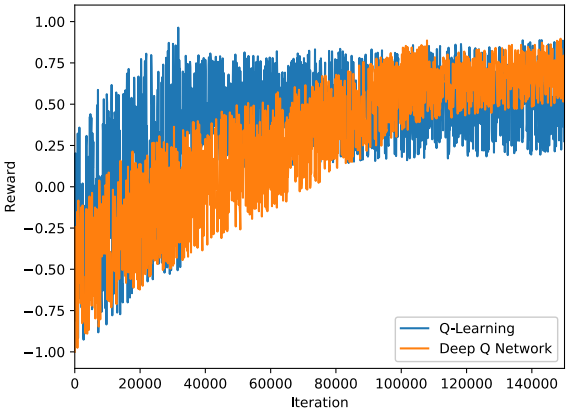

図 8 学習結果

一方，同様の情報量を持つ画像の識別への深層強化学習を適用した場合に比べて，学習の収束が遅いことが判明した．原因として，強化学習における報酬関数が単純であることや，写像として用いるフィルタが少なくゴールに到達するルートが存在しない場合を排除できていないことが考えられる．

提案手法では，セキュリティアナリストの持つノウハウのうち，ある端緒に対してどのフィルタを組み合わせるかについて過去の脅威分析結果から深層強化学習によって学

び取り分析を自動化することを提案しているが，どのような種類のフィルタがオープンソースインテリジェンスとして利用可能かをインターネット上で探索することを提案していない．

報酬関数の改善やフィルタの充実を行ない，より早く学習を収束させる方法についての検討が必要と考える．

5. まとめ

本稿では，サイバーセキュリティにおいて，攻撃側よりもより多くの脅威とその全体像を把握する必要のある防護側のセキュリティアナリストによる脅威分析業務をモデル化することで，その業務を援用する深層強化学習による自動的な脅威分析手法について提案した．ルクセンブルク CERT の公開レポートを用いて提案手法を実装・評価し，脅威分析の自動化が行えることと反復学習による学習の進捗状況を示した．

今後，報酬の与え方やオープンソースインテリジェンスを活用するフィルタの充実をはかり，少ない学習数によって収束する手法を検討する．

参考文献

[1] 警察庁, “攻撃ツール「Eternalblue」を悪用した攻撃と考えられるアクセスの観測について”. https://www.npa.go.jp/cyberpolice/important/2017/201705151.html, (参照 2017-08-25).

[2] JPCERT, “高度サイバー攻撃（APT）への備えと対応ガイド～企業や組織に薦める一連のプロセスについて”. https://www.jpcert.or.jp/research/20160331-APTguide.pdf, (参照 2017-08-25).

[3] SOPHOS, “Ransomware as a Service (RaaS) Deconstructing Philadelphia”. https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/RaaS-Philadelphia.pdf, (参照 2017-08-25).

[4] Federal Ministry of Education and Research. Germany, ”The new High-Tech Strategy”. https://www.bmbf.de/pub/HTS_Broschuere_eng.pdf, (参照 2017-08-25).

[5] 警視庁, “平成 25 年度中のサイバー犯罪の検挙状況等について”. http://www.npa.go.jp/cyber/statics/h25/pdf01-2.pdf, (参照 2017-08-25).

[6] 総務省, “情報通信白書平成 27 年版”. http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc254310.html, (参照 2017-08-25).

[7] 独立行政法人情報処理推進機構(IPA), “「情報セキュリティ人材の育成に関する基礎調査」報告書について”. https://www.ipa.go.jp/security/fy23/reports/jinzai/, (参照 2017-08-25)

[8] MITRE, “Ten Strategies of a World-Class Cybersecurity Operations Center”. https://www.mitre.org/sites/default/files/publications/pr-13-1028-mitre-10-strategies-cyber-ops-center.pdf, (参照 2017-08-25).

[9] 日本セキュリティオペレーション事業者協議会(ISOG-J), "SOC の役割と人材のスキル”. http://isog-j.org/output/2016/SOC_skill_v1.0.pdf, (参照 2017-08-25).

[10] The OASIS Cyber Threat Intelligence (CTI) TC. “Trusted Automated Exchange of Intelligence Information (TAXII™)”. https://oasis-open.github.io/cti-documentation/ (参照 2017-08-25)

[11] Google Inc.. “VirusTotal”. https://virustotal.com/, (参照 2017-08-25).

[12] L. Baird.. Residual algorithms: Reinforcement learning with function approximation. In Proceedings of the 12th International Conference on Machine Learning (ICML 1995), pages 30–37. Morgan Kaufmann, 1995.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonglou, D. Wierstra, and M. Riedmiller. Playing Atari with Deep Reinforcement Learning. NIPS 2013 Deep Learning Workshop, 2013.

[14] The OASIS Cyber Threat Intelligence (CTI) TC. “Structured Threat Information eXpression (STIX™)”. https://oasis-open.github.io/cti-documentation/ (参照 2017-08-25)

[15] IETF, “The application/json Media Type for JavaScript Object Notation (JSON)”. https://tools.ietf.org/html/rfc4627. (参照 2017-08-25)

[16] “The Computer Incident Response Center Luxembourg (CIRCL)”, https://www.circl.lu/. (参照 2017-08-25)

[17] SANS, “Internet Storm Center”. https://isc.sans.edu/. (参照 2017-08-25)

[18] RiskIQ, “PassiveTotal”. https://community.riskiq.com/. (参照 2017-08-25)

[19] “Shodan”. https://www.shodan.io/. (参照 2017-08-25)

[20] “Censys”. https://censys.io/. (参照 2017-08-25)

[21] ALIENVAULT, INC., “ThreatCrowd”. https://www.threatcrowd.org/. (参照 2017-08-25)

[22] ThreatExpert Ltd., “ThreatExpert”. http://www.threatexpert.com/. (参照 2017-08-25)

[23] ALIENVAULT, INC., “AlienVault - Open Threat Exchange”. https://otx.alienvault.com/. (参照 2017-08-25)

[24] NIST, “National Software Reference Library”. https://www.nsrl.nist.gov/. (参照 2017-08-25)

[25] Tokui, S., Oono, K., Hido, S. and Clayton, J.. Chainer: a Next-Generation Open Source Framework for Deep Learning. Proceedings of Workshop on Machine Learning Systems(LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS), (2015)

[26] Jim Hugunin, The Python Matrix Object: Extending Python for Numerical Computation. Proceedings of the Third Python Workshop, Reston, Va., (Dec. 1995.)

[27] Preferred Networks, inc. and Preferred Infrastructure, inc. “CuPy”. https://cupy.chainer.org/. (参照 2017-08-25)

[28] Christopher JCH Watkins and Peter Dayan. Q-learning. Machine learning, 8(3-4):279–292, 1992.