話題誘導するトピックモデルを用いたセキュリティレポート分類

永井 達也[†1] 乾 智裕[†1] 瀧田 愼[†1] 古本 啓祐[†2]
白石 善明[†1] 髙野 泰洋[†1] 毛利 公美[†3] 森井 昌克[†1]

概要：企業や組織を標的としたサイバー攻撃が多様化・高度化している．このような攻撃を未然に防ぐためには，最新の脅威の動向の把握や，自組織内の保有する資産に基づいた対策が必要となる．セキュリティベンダは脅威の動向や解析結果をセキュリティレポートとして発行しているが，情報源ごとに分類基準が異なっており，複数の情報源から必要な情報を抜き出すことが難しい．本稿では話題誘導するトピックモデルを用いてセキュリティレポートを分類する手法を提案する．セキュリティベンダが 2017 年に発行したセキュリティレポートを用いて分類し，LDA と比べて良好な分類ができることを示す．また，攻撃者，攻撃対象，攻撃手法，被害内容の四つの観点から提案手法を適用し，分類したセキュリティレポートの利用例について示す．

キーワード：セキュリティレポート，脅威分析，トピックモデル，クラスタリング

Clustering Security Blog Posts Using Guided-Topic Model for Threat Analysis

Tatsuya Nagai[†1] Tomohiro Inui[†1] Makoto Takita[†1] Keisuke Furumoto[†2]
Yoshiaki Shiraishi[†1] Yasuhiro Takano[†1] Masami Mohri[†3] Masakatu Morii[†1]

Abstract: Cyber attacks aiming at specific targets implement various attack strategies from multiple intrusion routes. In order to prevent such cyber attacks, it is important to understand the current cyber threat landscape and to make countermeasures based on assets in an organization. Security vendors post trends and threat analyses as security reports. Because security vendors do not have a unified basis for the classification of security reports, it is difficult to integrate this information. In this paper, we propose a method to cluster security reports using guided-topic model. By evaluations using datasets that five security vendors published in 2017, their experimental results show that it can classify better than using regular topic model. We also show examples of an application by applying the proposed method in four aspects.

Keywords: Security blog posts, Threat analysis, Topic model, Clustering

1. はじめに

企業や組織を標的としたサイバー攻撃が高度化・多様化している．このような攻撃を Intrusion Detection System(IDS)や Security Information and Event Management(SIEM)といった既存のシステムのみで防ぐことは困難である．インシデント発生時，組織は被害箇所を速やかに特定し，被害の調査や防御の対策を行う．しかし，平常時から，インシデントの予防のため，脅威動向を注視し，その組織に応じた事前対策を講ずることも不可欠である．

脅威情報は，例えば，セキュリティベンダが発行するセキュリティレポートから収集できる．場合により，個人がブログ・SNS 等で発信した文書からも有用な情報が得られることもある．しかし，情報提供者ごとに分類基準が異なっており，複数の情報源から必要な情報を抜き出すことが難しい．

文書が攻撃者別に分類されていれば，攻撃の動向から次の標的を分析することが可能となる．また，攻撃手法別に整理されていれば，対策したい攻撃について侵入経路を列挙したり，被害の原因となった攻撃から起こり得る次の脅威について調査できる．被害の内容について分類されていれば，インシデント対策手法を知ることができる．攻撃対象の観点からは自組織の資産や関連のある業界を調べ，自組織に起こり得る脅威を調べることができる．脅威の動向分析や被害調査を円滑に行えるように，単に情報収集するだけでなく，収集した文書の整理が望まれる．

文書の分類には教師有り学習であるカテゴリ分類と教師無し学習であるクラスタリングの二つのアプローチがある．カテゴリ分類では福本ら[1]は WordNet を用いた Support Vector Machine(SVM)による文書分類を行っている．WordNet により単語の上位関係を考慮することで精度が上がることを確認している．Zhang ら[2]も同様に SVM による文書分類を行っている．Reuters-21578 データセットを用いた実験で，複合語を考慮しない場合と SVM のカーネルを変えた場合の識別精度を比較しており，線形カーネルを用いた場合に精度が上がることを確認している．Onan ら[3]は文書からキーワードを抽出して 5 つの機械学習によるアンサンブル学習を行い文書分類している．これらの手法は学習データにその情報源や種別を示すタグやカテゴリ，ラベル（以下ではラベルと称す）が必要である．しかし，セキュリティレポートは発行元の企業や個人の基準によってラベル付けされている．また，ラベルが付与されていない文書も存在する．そのため，教師付きデータによる文書分類手法を適用することが困難である．

クラスタリングによる文書の分類手法として，k-means などのクラスタリング手法が提案されている．Beil ら[4]は単語の出現頻度を特徴量としたクラスタリング手法を提案している．Fung ら[5]も同様に単語の出現頻度を特徴量とし

†1 神戸大学大学院工学研究科電気電子工学専攻
Graduate School of Engineering, Kobe University
†2 国立研究開発法人 情報通信研究機構
National Institute of Information and Communication Technology
†3 岐阜大学工学部電気電子・情報工学科
Faculty of Engineering, Gifu University

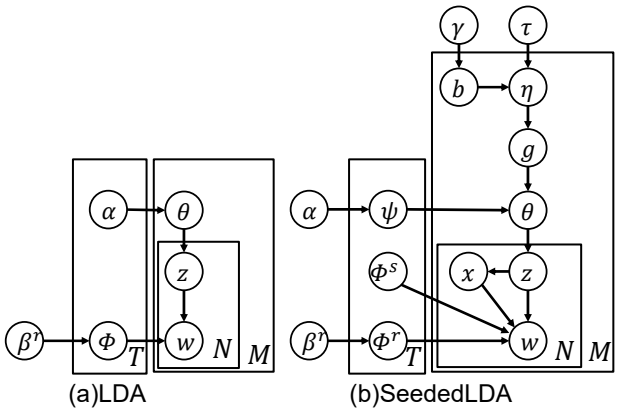

図 1: LDA と SeededLDA の概念図

た階層的クラスタリングの手法を提案している. Chen ら[6]は TF-IDF を特徴量としてスペクトラルクラスタリングを行っている．以上のように文書の特徴量には単語頻度や TF-IDF などが用いられているが, 文書ベクトルが高次元になることが課題となっている.

次元削減のアプローチとしてトピックモデルが注目されている．Ma ら[7]はトピックモデルの一種である Latent Dirichlet Allocation(LDA)[8]を用いた三段階のクラスタリング手法を提案している．この手法では k-means++で最適なクラスタ数を見つけ，k-means でクラスタリングしている．Onan ら[9]は LDA を用いた k-means のクラスタリング手法を提案している. Tagarelli ら[10]は文書を段落ごとに分けて Sk-means や FSk-means，LDA を用いてクラスタリングする手法を提案している．トピックモデルにより各文書に含まれるトピックや潜在意味を考慮することが可能となる．しかし, LDA によって生成されたトピックは分析者の意図に沿った分類がされるとは限らない.

本稿ではトピックの内容を誘導できる SeededLDA[11]を用いて, セキュリティレポートを分類する手法を提案する．セキュリティベンダ 5 社が 2017 年に発行したセキュリティレポートに提案手法を適用し，LDA との比較を行う．また，攻撃者，攻撃対象，攻撃手法，被害内容の四つの観点から分類したセキュリティレポートの利用例を示す.

2. LDA と SeededLDA

LDA は Blei らが提案したトピックモデルの一種である．トピックモデルでは，文書を単語の集合と捉え，単語は単語集合の背後に存在するトピックから生成されると仮定されている.

図 1(a)に LDA の概念図を示す．文書数をM，文書dに現れる単語を$w = \{w_{d_1}, \ldots w_{d_N}\}$，トピック数を$T$としたとき，LDA における文書の生成過程は以下のようになる.

1. すべてのトピック$k = 1 \ldots T$に対して，ディリクレ分布$Dir(\beta)$から単語分布Φ_kをサンプリングする.
2. すべての文書dに対して
   I. ディリクレ分布$\mathrm{Dir}(\alpha)$からトピック分布 θ_{d}をサンプリングする.
   II. 各単語$i = 1 \ldots N$に対して
      i. トピック分布$\mathrm{Mult}(\theta_d)$からトピックz_{d_i}をサンプリングする.
      ii. 単語分布$\mathrm{Mult}(\Phi_{z_{d_i}})$から単語$w_{d_i}$をサンプリングする.

LDA の学習は文書に現れる単語の共起情報からトピックに単語を関連付ける．しかし，LDA によって学習されたトピックがユーザにとって意味のあるものであるとは限らない．Jagarlamudi らはユーザがトピックモデルに付加情報を与えて理解しやすいトピックになるよう誘導する SeededLDA を提案している. SeededLDA では誘導したいトピックに対しシード単語集合を設定する.

図 1(b)に SeededLDA の概念図を示す．SeededLDA における文書の生成過程は以下のようになる.

表 1: SeededLDA に入力するシード単語の例

モバイル	mobile, smartphone, android, iphone
サーバー	web application, CMS, apache strut
無線端末	wireless, IoT, connected device
医療	hospital, medical, health care, patient
工場	industrial control system, factory, plant

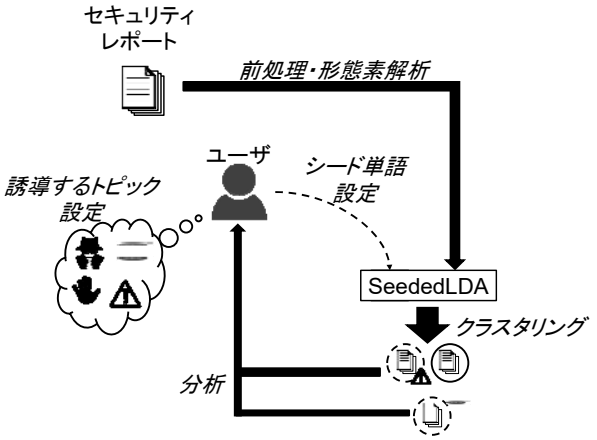

図 2: セキュリティレポートの分類手法の概要

1. すべてのトピック$k = 1 \ldots T$に対して
   I. ディリクレ分布$\mathrm{Dir}(\beta^r)$から通常トピック分布Φ_z^rをサンプリングする.
   II. ディリクレ分布$\mathrm{Dir}(\beta_s)$からシードトピック分布Φ_z^sをサンプリングする.
2. 各トピックのシード単語集合$s = 1 \ldots S$に対して，ディリクレ分布$\mathrm{Dir}(\alpha)$からグループトピック分布ψ_sをサンプリングする.
3. すべての文書dに対して
   I. シード単語ベクトル$\vec{b}$をサンプリングする.
   II. ディリクレ分布$\mathrm{Dir}(\tau\vec{b})$から文書グループ分布$\zeta^{\mathrm{d}}$をサンプリングする.
   III. 文書グループ分布$\mathrm{Mult}(\zeta^d)$からグループ変数gをサンプリングする.
   IV. ディリクレ分布$\mathrm{Dir}(\psi_g)$からトピック分布θ_dをサンプリングする.
   V. 各単語$i = 1 \ldots N$に対して
      i. トピック分布$\mathrm{Mult}(\theta_d)$からトピックz_{d_i}をサンプリングする.
      ii. ベルヌーイ分布$\mathrm{Bern}(\pi)$から$x = \{0,1\}$をサンプリングする.
      iii. $x = 0$のとき
         1. 通常トピック分布$\mathrm{Mult}(\psi_z^r)$から単語w_{d_i}をサンプリングする.
      iv. $x = 1$のとき
         1. シードトピック分布$\mathrm{Mult}(\psi_z^s)$から単語w_{d_i}をサンプリングする.

表 1 にシード単語の例を示す．例えば“IoT system deploy in factory”という一文からなる短い文書があるとき，「無線端末」と「工場」のシード単語が含まれているため，シード単語ベクトルは$\vec{b} =< 0,0,1,0,1 >$となる．シード単語ベクトル$\vec{b}$は文書グループgを決定するときに用いられ，類似したベクトルを持つ文書と同じトピックに含まれやすくなるよう誘導される．また，単語の生成過程におけるシードトピック分布ψ_z^sはシード単語集合から選択するように制限される．シードトピック分布に誘導する確率はπで制御できる．シード単語の設定により教師付きデータを用意することなくユーザの意図に沿った分類ができると期待される.

3. セキュリティレポートの分類手法

分析者が当該組織に必要な分野のクラスタのレポート

表 2: 被害内容の観点からの各カテゴリのシード単語

カテゴリ	シード単語
Webサイト改ざん	compromised website, infected website, …
サービス妨害	denial of service, dos
データ破壊・漏洩1	data breach, data exfiltration, data loss, destroy, …
データ破壊・漏洩2	data breach, data exfiltration, data loss, destroy, …
デバイス感染	infect, infected, infection
金銭被害	financial damage, financial fraud, financial loss, …
個人情報・認証情報漏洩	credential, password, account, personal information
名誉棄損	reputation, defamation

を読み，効率的に情報を得ることができることを目標として，トピックの内容を誘導する SeededLDA を用いたセキュリティレポートの分類を行う．

3.1 前処理

本論文で対象とする文書はセキュリティレポートであり，主に概要を示すタイトル，本文，本文を補完する図表で構成されている．タイトルや図表にもセキュリティレポートを特徴づける情報が含まれているため，入力する内容はセキュリティレポートのタイトルと図表のキャプションと本文とする．

セキュリティレポートをクラスタリングする前処理として単語抽出のために必要な形態素解析を行う．すなわち，トピック分類において有意な品詞である単名詞を抽出し，助詞や助動詞など不要な入力ノイズを取り除く．

セキュリティレポートは固有名詞や複合語が頻出する．セキュリティレポートを単に形態素解析すると，必要以上に単名詞に分解されている可能性があり，クラスタリングに適さない．専門用語抽出手法[12]によって固有名詞や複合語を構成する．

発行元によっては自組織や自社製品の宣伝をするセキュリティレポートや文章があるが，脅威分析とは直接関係が無いため入力から取り除く．セキュリティレポートには注意喚起のためにマルウェアのハッシュ値や C&C サーバの IP アドレス，URL を載せている場合がある．これらの文字列は自然言語ではないため入力データから取り除く．定期的に発行される，脅威情報をまとめたセキュリティレポートは一つの話題について詳細な記述が期待できないため，入力データから取り除く．

3.2 誘導するトピックの設定

SeededLDA を用いてセキュリティレポートを分析者が活用しやすい形式で分類する．誘導するトピックは自組織の状態や状況に応じて設定する．誘導するトピックに関連する単語をシード単語として指定する．3.1 節の前処理で複合語を抽出しているため，シード単語を含む複合語も同様にシードとして SeededLDA に入力する．

誘導するトピックに対して該当するセキュリティレポートが多いと想定される場合，シード単語を複数のトピックに設定することで一つのカテゴリに複数のトピックを誘導し，対応付けられる．入力文書とシード単語の設定によってはシード単語の数が非常に多くなってしまう可能性がある．シードが多くなると分類精度が低くなるため，数十程度にサンプリングしたものをシード集合とする．

3.3 SeededLDA とクラスタリング

SeededLDA に誘導するトピックのシード集合と文書集合とトピック数とシードへの誘導確率を入力すると，各文書のトピック分布を出力として得られる．各文書のトピック分布から，トピック比率が閾値以上の文書を集めてクラスタを構成する．実際の文書では複数のトピックが混ざり合った文書は珍しいものではないため，複数のトピックで構成された文書は，トピック比率が閾値以上なら複数のトピックのクラスタに属するものとする．

文書のトピック分布はシードトピック分布と通常トピック分布の混合分布のため，誘導したトピックとクラスタの対応関係を文書のトピック分布から求めることができない．そこで，カテゴリとクラスタの対応関係の最尤推定をする．各カテゴリで設定したシード単語のトピック分布を足し合わせ，最も確率の高いトピックのクラスタをそのカテゴリとする．

表 3: LDA と SeededLDA の分類結果

	Purity[%]	Inverse Purity[%]	F-measure[%]
LDA	33.0	13.8	19.5
SeededLDA	29.0	36.6	32.34

図 3: LDA と SeededLDA におけるデバイス感染カテゴリのワードクラウド

4. 分類実験

4.1 概要

SeededLDA を用いた提案手法とシード単語を設定しない通常の LDA を用いた比較手法でセキュリティレポートの分類を行い，SeededLDA の方が適切に分類されていることを確認する．

データセットは，セキュリティベンダ 5 社（TrendMicro[13], Cisco[14], Symantec[15], Barracuda[16], Druva[17]）から発行されているセキュリティレポートで，収集対象期間を 2017 年 1 月 1 日～2017 年 6 月 30 日とした 440 件である．実験に利用したツールは以下のとおりである．形態素解析には自然言語処理ツールである NLTK[18]を利用した．単名詞を組み合わせた複合語を構成するために，専門用語（キーワード）自動抽出システムである termextract を利用した．LDA と SeededLDA の実行には GuidedLDA ライブラリ[19]を用いた．

誘導するトピックとシード単語の設定を表 2 に示す．あらかじめ，セキュリティレポート 440 件に対し表 2 のカテゴリのラベル付けを行った．

LDA と SeededLDA の両方においてトピック数を 30，クラスタリングの閾値を 0.3 と設定した．また，SeededLDA によるシードトピック分布への誘導確率πを 0.3 とした．

分類の評価手法としては，Purity/Inverse Purity と調和平均である F 値を用いる．提案手法により得られたクラスタ集合を $C = \{C_1, \ldots, C_N\}$，正解のクラスタ集合を $L = \{L_1, \ldots, L_M\}$としたとき，任意の 2 クラスタ $\mathrm{C_i}, \mathrm{L_j}$ の精度 $\mathrm{Precision}(\mathrm{C_i}, \mathrm{L_j})$を

$$\mathrm{Precision}(C_i, L_j) = \frac{|C_i \cap L_j|}{C_i}$$

と定義する．このとき，Purity/Inverse Purity は

$$\mathrm{Purity} = \sum_i \frac{|C_i|}{N} \max_j \mathrm{Precision}(C_i, L_j)$$

$$\mathrm{Inverse\ Purity} = \sum_i \frac{|L_i|}{N} \max_j \mathrm{Precision}(L_j, C_i)$$

で表される．Purity はクラスタ内に正解文書が含まれる度合いを示し，Inverse Purity は正解文書集合が一つのクラスタに集合している度合いを示す．Purity と Inverse Purity の調和平均をとった値が F 値となる．

4.2 分類実験

表 3 に分類結果を示す．SeededLDA は Inverse Purity が大きく上がり，LDA の F 値より高くなっている．図 3 に SeededLDA と LDA のそれぞれでデバイス感染カテゴリの

表 4: 攻撃手法の観点からの各カテゴリのシード単語

カテゴリ	シード単語
キャンペーン	hacker, campaign
スパイウェア	spyware
ソーシャルエンジニアリング	social engineering
ソフトウェア脆弱性	exploit, exploitation, exploitable
トロイの木馬	trojan
ネットワークアタック	network attack, network compromise
マルウェア	malware
ランサムウェア	ransomware
ワーム	worm
違法取引	illegal lucrative
内部犯行・スパイ行為	insider, espionage
無権限アクセス	authorization, authentication, accessible, …

表 5: 攻撃対象の観点からの各カテゴリのシード単語

カテゴリ	シード単語
金融機関	financial institution, ATM, bank, banking trojan
OS・ソフトウェア	operating system, windows, unix, llinux, mac, …
POSシステム	point of sale, point-of-sale, pos
アカウント情報	credential, password, username, creditcard, …
インフラ	gas, power generation, water supply
クラウド1	cloud
クラウド2	cloud
サーバ・Webサイト	web service, contentmanagement system, …
ファームウェア・ハードウェア	hardware, firmware
ブラウザ	browser, web-browser
モバイルOS・アプリ1	mobile, smartphone
モバイルOS・アプリ2	mobile, smartphone
医療設備・患者	patient, medical, health care, clinic
家庭用電子機器	home device, home network, home router, …
管理者権限	root, code execution, privilege escalation
工場機器	industrial control system, ics, factory, power grid..
国家・政治家	politician, espionage, government, federal, …
通信機器	network device, router, port, tcp
無線端末	wireless, iot, internet of things, connected device

表 6: 攻撃者の観点からの各カテゴリのシード単語

カテゴリ	シード単語
犯罪組織1	terrorist, cybercriminal, espionage group
犯罪組織2	terrorist, cybercriminal, espionage group
犯罪組織3	terrorist, cybercriminal, espionage group
国家スパイ	state sponsored, state-sponsored
言論集団	hacktivist, propaganda, ideology

文書集合をワードクラウドで表したものを示す．デバイス感染カテゴリに関連する単語は，LDA は malware のみが大きく表示されているのに対し，SeededLDA では malware と ransomware の両方が大きく表示されている．このことから LDA より SeededLDA の方がデバイス感染に関するトピックを広く集められていることがわかる．

例えば，シード単語に設定した infection という単語は様々なデバイス・攻撃に対して使用されている．LDA で infection という単語を学習する場合，複数のトピックで使用されていることからこの単語はトピックを特徴づける単語にならず，別の単語が選択される．同じ infection という単語を持つ文書同士でも，得られるトピック分布は異なる分布になる可能性が高く，クラスタが分かれてしまい Inverse Purity が低くなる．SeededLDA では文書にシード単語が含まれるかどうかでグループ変数 g が設定される．トピック分布はグループ変数 g から選択されるため，同一のグループからはまったく同じトピック分布が用いられる．そのため，シード単語が含まれている文書同士が同じクラスタにまとまりやすく，Inverse Purity が高くなったと考えられる．ただ，分析者の期待する分類のされ方と，シード単語の設定によるトピックの誘導の方向が一致するとは限らない．トピックモデルは単語の統計情報から学習を行うため，入力データの分布に結果が大きく左右される．分析者は自身が求める情報と入力データの分布を把握したうえで適切にシード単語を設定することが求められる．

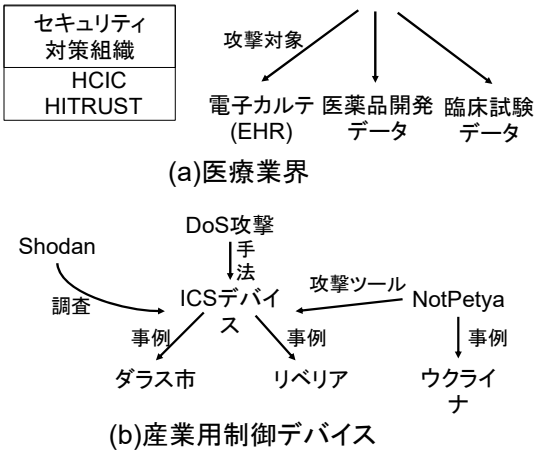

図 4: 攻撃手法の観点から分類したトロイの木馬・ソーシャルエンジニアリングカテゴリの話題

5. 分類されたセキュリティレポートの活用例

5.1 概要

データセットは，4 章で述べたセキュリティベンダ 5 社から発行されているセキュリティレポートで，収集期間を 2017 年 1 月 1 日～12 月 31 日とした 780 件である．実験に利用したツールは 4 章と同じである．

サイバー攻撃を分類する基準として SurfWatchLab 社の作成したカテゴリ[20]を参考に四つの観点を設定する．

攻撃対象：サイバー攻撃を受けた組織，業界によって分類する．

攻撃手法：サイバー攻撃で用いられる攻撃手法によって分類する．

被害内容：サイバー攻撃を受けたことで生じる被害内容によって分類する．

攻撃者：攻撃を行う集団によって分類する．

表 2,4,5,6 に誘導するトピックとシード単語を示す．それぞれの観点から各カテゴリにシード単語を設定し，セキュリティレポートを分類する．分類されたクラスタからセキュリティレポートを読み，どのような情報が得られるかを確認する．入力データに特定の観点から記述されたものが多いため，データ数の多い攻撃手法・被害内容の観点について 5.2 節で，データ数の少ない攻撃対象・攻撃者の観点について 5.3 節で述べる．

5.2 攻撃手法・攻撃対象による分類

5.2.1 攻撃手法の観点からの分類

分類されたクラスタのうち，ソーシャルエンジニアリングとトロイの木馬についてまとめたものを図 4 に示す．

ソーシャルエンジニアリングでは，Barracuda が 2017 年 1 月 19 日に発行したセキュリティレポート[21]では Gmail のサインインページに似せたフィッシング攻撃の事例について報告している．この攻撃では PDF 付き E メールを受信し，PDF を開くと Gmail のサインインページに似たページが表示され，ユーザのログイン情報を盗み取る．Barracuda が 2017 年 6 月 8 日に発行したセキュリティレポート[22]では添付ファイル付き E メールを受信し，添付ファイルからランサムウェアに感染した事例と対策について報告している．添付ファイルは請求書の外見をしており，添付ファイルを開くとユーザが気付かないままランサムウェアに感染してしまう．

この二つのレポートからソーシャルエンジニアリングによる攻撃を防ぐにはアカウント情報を入力するまでの段階で対策が必要であること，添付ファイルがランサムウェアでないか判別する対策が必要であることがわかる．

トロイの木馬では，TrendMicro が 2017 年 12 月 1 日に発行したセキュリティレポート[23]では Android 端末に感染

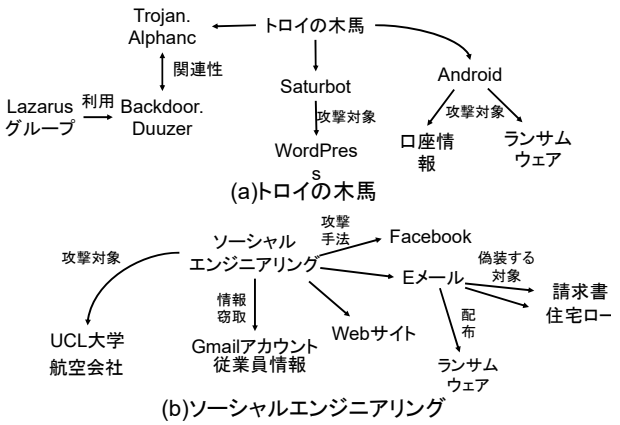

図 5: 攻撃対象の観点から分類した医療・産業用制御デバイスカテゴリの話題

するトロイの木馬型マルウェアについて報告している．その中で，RedAlert2.0 と呼ばれるトロイの木馬はサードパーティのアプリストアを介して端末に感染し，銀行からの着信をブロックし，口座情報を盗む．さらに，SMS メッセージを傍受する機能も備えていることが報告されている．

このレポートから感染を発見してマルウェアを除去したあとも，情報漏洩した可能性がある前提でその後の対策をとる必要がある．

5.2.2 攻撃対象の観点からの分類

分類されたクラスタのうち，医療業界と産業用制御システムについてまとめたものを図 5 に示す．

医療業界では，TrendMicro が 2017 年 5 月 4 日に発行したセキュリティレポート[24]では医療機関への攻撃に対する注意喚起を行っている．電子カルテ漏洩が 2016 年に 377 件起きており，個人情報や財務データが含まれる電子カルテは攻撃対象になりやすいと述べている．また，Symantec が 2017 年 6 月 14 日に発行したセキュリティレポート[25]では医療業界のための六つのセキュリティ対策について報告している．臨床試験や医薬品開発などのデータを守る仕組みを把握しておくことや，医療業界に起こり得る脅威やリスクについての情報共有をすることが推奨されている．これらのレポートから医療業界で取るべきセキュリティ対策を知ることが可能である．

産業用制御システムでは，TrendMicro が 2017 年 2 月 15 日に発行したセキュリティレポート[26]では IoT に特化した検索エンジン Shodan を用いて公開されているデバイスについて調べた内容が報告されている．公開されているデバイスに産業用制御システム(ICS)デバイスも含まれており，その中には誤って公開されているものも頻繁に含まれていることが述べられている．TrendMicro が 2017 年 5 月 22 日に発行したセキュリティレポート[27]でも同様に Shodan を用いた調査を行っており，加えてアメリカのダラス市で起きた ICS を標的としたサイバー攻撃の事例について述べている．

二つのレポートから ICS デバイスが公開されていないか調べ，インシデント発生時の対応を決めておくことが必要だと分析できる．

5.3 被害内容・攻撃者による分類

5.3.1 被害内容の観点からの分類

分類されたクラスタのうち，サービス妨害と Web サイト改ざんについてまとめたものを図 6 に示す．

サービス妨害では，Symantec が 2017 年 4 月 18 日に発行されたセキュリティレポート[28]では IoT ボット Hajime の挙動が報告されている．Hajime 感染への対策方法としてファームウェアの更新によるデフォルトパスワードの更新や Telnet ログインの無効化などが挙げられている．また，Cisco が 2017 年 4 月 18 日に発行されたセキュリティレポート[29]では病院のセキュリティ対策について述べられている．

このレポートでは多くの医療機器には Telnet を含めた複

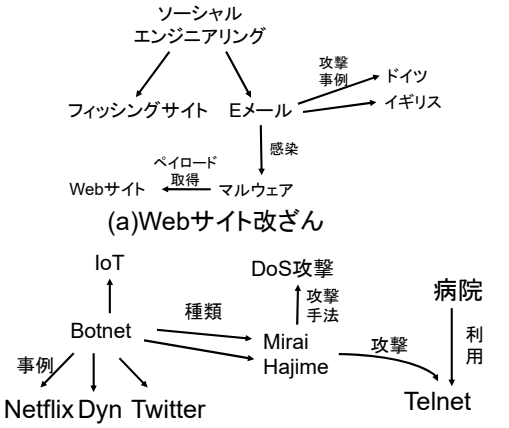

図 6: 被害内容の観点から分類した Web サイト改ざん・サービス妨害カテゴリの話題

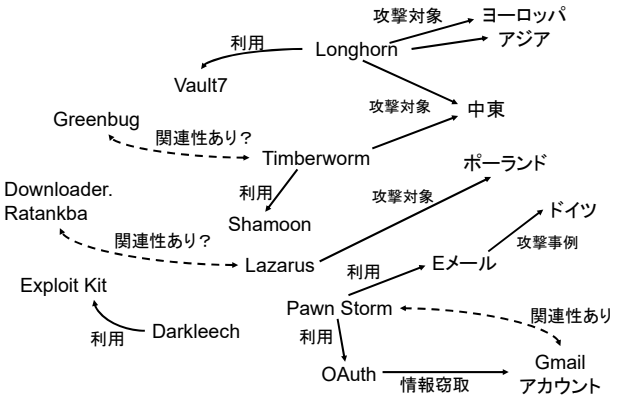

図 7: 攻撃者の観点から分類した犯罪組織の話題

数のポートが開いており，脆弱であると述べられている．IoT 機器と同様に医療機器もサービス妨害攻撃を受ける可能性があることが二つのレポートから分析できる．

Web サイト改ざんでは，Symantec が 2017 年 3 月 20 日に発行されたセキュリティレポート[30]ではドイツを標的としたスパムキャンペーンの事例を報告している．この攻撃は添付ファイル付き E メールを受信し，添付ファイルを開くとマルウェアに感染する．マルウェアのペイロードのホスト先として改ざんした Web サイトが利用されている．Barracuda が 2017 年 3 月 30 日に発行されたセキュリティレポート[31]では航空会社へのフィッシング攻撃の事例を報告している．この攻撃は添付ファイル付き E メールとフィッシング Web サイトへのリンクを含む E メールの二つの攻撃手法が用いられている．

このカテゴリは Web サイト改ざんがメインのトピックではなく，E メールを用いたフィッシング攻撃のレポートが分類されており，Web サイト改ざんに関する分析ができなかった．

5.3.2 攻撃者の観点からの分類

分類されたクラスタのうち，犯罪者グループについてまとめたものを図 7 に示す．

Barracuda が 2017 年 5 月 5 日に発行したセキュリティレポート[32] では Google アカウントへのフィッシング攻撃の事例を報告している．攻撃者は OAuth 認証を利用したアカウント窃取を行う．また，TrendMicro が 2017 年 5 月 3 日に発行したセキュリティレポート[33]では Google アカウントへのアクセスを目的としたフィッシングメールの事例を報告している．このレポートでは Google の OAuth 認証を利用することを目的としており，PawnStorm グループで用

いられていた手法であると述べている．

これらのレポートは同時期に発行されており，同一の攻撃である可能性が高いと考えられる．TrendMicro のレポートからこの攻撃が PawnStorm によるものであるという仮説のもと，分析することができる．

5.4 考察

5.2 節で述べたように攻撃手法と攻撃対象については分類されたクラスタからレポートを読むことで，関連する攻撃や対策すべき内容を得ることができた．一方で，情報源によっては攻撃者の観点から書かれることが少ないため，犯罪者グループしか十分な大きさのクラスタにならなかった．入力データによる影響はカテゴリ間でも現れており，被害内容の観点からはサービス妨害のカテゴリはレポートから事例や対策方法が得られたが，Web サイト改ざんはデータ数が少なく，フィッシング E メールによる攻撃が主な話題となってしまった．今回の実験ではサイバー攻撃の分類に沿ってカテゴリを決定したが，脅威動向や入力データの分布を考慮したカテゴリを設定する必要がある．

5.3.1 項で同一の攻撃について述べている可能性があるレポートを例として挙げたが，同時期に発行されたレポートを読み比べることで攻撃の関連性や傾向の変化が把握できる可能性がある．時系列で区切った場合のセキュリティレポート分類を今後の課題とする．

6. 関連研究

6.1 話題誘導するトピックモデル

Mcauliffe らはラベル付き文書を LDA の生成モデルに取り入れた SupervisedLDA[34]を提案している．Lacoste ら[35]も同様にラベル付き文書を用いた生成モデルとして DiscLDA を提案している．Ramage ら[36]はマルチラベル付き文書を対象とした LabeledLDA を提案している．ラベル付き文書を用いることでユーザが意図したトピックになるよう誘導される．しかし，本研究で用いるセキュリティレポートはラベルが付与されていなかったり，ラベル付与の基準が情報源によって異なるため，これらの手法を適用することが難しい．

Mukherjee ら[37]は感情分析のためにシード単語集合を与えたモデル ME-SAS を提案している．シード単語は観点を決めるために設定され，感情を表す単語とそれ以外を区別している点で SeededLDA とは異なる．Andrzejewski ら[38]は単語ペアに対して Must Link と Cannot Link を設定し，同一のトピックに現れるかどうかを指定できるモデル DF-LDA を提案している．本手法では一つのレポートが複数のクラスタに属するソフトクラスタリングを用いており，異なるクラスタに属するよう制限する Cannot Link が効果的に働かない可能性がある．Hu ら[39]はトピックモデルの出力に対し，ユーザがフィードバックを与えて学習させる手法を提案している．最初は通常の LDA による学習が行われ，ユーザがいいトピックと悪いトピックのフィードバックを与えると，フィードバックに基づいた学習が再度行われる．これを複数回繰り返すことでユーザの想定するトピックに分けることができる．本研究ではユーザが脅威情報を得るのに分類を行うため，ユーザのフィードバックを前提としたモデルを適用できない．

6.2 文書からの脅威情報収集

Joshi ら[40]は NVD に含まれるテキストデータから脅威情報を抽出する手法を提案している．複数の情報源から脅威情報を抜き出し，関係性を記述することができる．Mittal ら[41]はセキュリティブログ・NVD などから得た情報を用いて Twitter の文章から脅威情報・脆弱性情報を抽出し，アラートを出す手法を提案している．しかし，これらの手法は脆弱性情報に特化しており，攻撃者の観点から脅威情報を記述することを想定されていない．

McNeil ら[42]はブートストラップ法を用いて少量の教師付きデータから脅威情報の抽出とパターン学習を行う PACE という手法を提案している．脅威情報から共通するパターンを抜き出し，そのパターンから新たな脅威情報を抜き出す操作を繰り返すことで脅威情報抽出とパターンの学習を同時に行うことができる．この手法では文章に対して脅威情報抽出が行われるが，類似した脅威情報を集めるという点からは文書単位で学習できる SeededLDA の方が適していると考えられる．また，McNeil らの手法と同様の方法で，分類結果から新たなシード単語を抽出して新たな分類を逐次行うことができると考えられるので，今後の課題とする．

7. まとめ

本稿ではトピックを誘導する SeededLDA を用いたセキュリティレポートの分類手法を提案した．セキュリティベンダ 5 社が発行したセキュリティレポートに提案手法を適用し，SeededLDA を用いた場合に LDA と比べて Inverse Purity が 2 倍以上となり，同じ話題を持つセキュリティレポートがまとまりやすくなることを確認した．また，四つの観点から分類したセキュリティレポートの利用例を示し，効率的に情報を得られることを確認した．

脅威動向や入力データの分布を考慮したシード単語の決定方法と，攻撃の関連性や傾向の変化を把握するための時系列を考慮した分類手法を今後検討する．

参考文献

[1] 福本文代，鈴木良弥：トピックと局面の対応関係に基づく実生活ツイートのマルチラベル分類，情報処理学会論文誌データベース（TOD），vol.43，no.6，pp.1852-1865，Jun，2002.

[2] Zhang, W., Yoshida, T. and Tang, X.: Text classification based on multi-word with support vector machine, Journal of Knowledge-Based Systems, vol.21, no.8, pp.879-886, Apr, 2008.

[3] Onan, A., Korukoğlu, S. and Bulut, H.: Ensemble of keyword extraction methods and classifiers in text classification, vol.57, pp.232-247, Journal of Expert Systems with Applications, Mar, 2016.

[4] Beil, F., Ester, M. and Xu, X.: Frequent term-based text clustering, pp.436-442, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Jul, 2002.

[5] Fung, B., Wang, K. and Ester, M.: Hierarchical document clustering using frequent itemsets, pp.59-70, Proceedings of the SIAM International Conference on Data Mining, 2003.

[6] Chen, W., Song, Y., Bai, H., Lin, C. and Chang, Edward Y.: Parallel spectral clustering in distributed systems, vol.33, no.3, pp.568-586, IEEE transactions on pattern analysis and machine intelligence, 2011.

[7] Ma, Y., Wang, Y. and Jin, B.: A three-phase approach to document clustering based on topic significance degree, no.18, pp.8203-8210, Journal of Expert Systems with Applications, Jul, 2014.

[8] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, Journal of Machine Learning Research, no.3, pp.993-1022, Jan. 2003.

[9] Onan, A., Bulut, H. and Korukoglu, S.: An improved ant algorithm with LDA-based representation for text document clustering, vol.43, no.2, pp.275-292, Journal of Information Science, 2016.

[10] Tagarelli, A. and Karypis, G.: A segment-based approach to clustering multi-topic documents, vol.34, no.3, pp.563-595, Journal of Knowledge and Information Systems, Sep, 2013.

[11] Jagarlamudi, J., Daumé III, Hal. and Udupa, R.: Incorporating lexical priors into topic models, pp.204-213, Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Apr, 2012.

[12] TermExtract：専門用語(キーワード)自動抽出用 Perl モジュール TermExtract の解説(online), available from 〈http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html〉 (accessed 2018-08-20)

[13] TrendMicro: Simply Security News, Views and Opinions from Trend Micro, Inc (online), available from 〈http://blog.trendmicro.com/〉 (accessed 2018-08-20)

[14] Cisco: Cisco Blog(online), available from 〈https://blogs.cisco.com/〉 (accessed 2018-08-20)

[15] Symantec: Symantec Blogs(online), available from 〈https://www.symantec.com/blogs/〉 (accessed 2018-08-20)

[16] Barracuda: Barracuda – Security, Access and Reliability for Cloud-

Connected Networks and Applications (online), available from 〈https://blog.barracuda.com/〉 (accessed 2018-08-20)
[17] Druva: Druva Blog: Data Protection and Beyond (online), available from 〈https://www.druva.com/blog/〉 (accessed 2018-08-20)
[18] NLTK Project: Natural Language Toolkit — NLTK 3.3 documentation (online), available from 〈http://www.nltk.org/〉 (accessed 2018-08-20)
[19] vi3k6i5/GuidedLDA: semi supervised guided topic model with custom guidedLDA(online), available from 〈https://github.com/vi3k6i5/GuidedLDA〉 (accessed 2018-08-20)
[20] SurfWatchLabs: Cyber Threat Categories(online), available from 〈https://www.surfwatchlabs.com/threat-categories〉 (accessed 2018-08-20)
[21] Barracuda: New Gmail phishing campaign incredibly well designed(online), available from 〈https://blog.barracuda.com/2017/01/19/new-gmail-phishing-campaign-incredibly-well-designed/〉 (accessed 2018-08-20)
[22] Barracuda: Spear phishing and ransomware are a perfect match(online), available from 〈https://blog.barracuda.com/2017/06/08/spear-phishing-and-ransomware-are-a-perfect-match/〉 (accessed 2018-08-20)
[23] TrendMicro: Update: Mobile threats on the rise(online), available from 〈https://blog.trendmicro.com/update-mobile-threats-on-the-rise/〉 (accessed 2018-08-20)
[24] TrendMicro: Leading by Example at the HITRUST Annual Healthcare Cybersecurity Conference(online), available from 〈https://blog.trendmicro.com/leading-example-hitrust-annual-healthcare-cybersecurity-conference/〉 (accessed 2018-08-20)
[25] Symantec: HHS Cybersecurity Task Force Releases Report to Congress(online), available from 〈https://www.symantec.com/connect/blogs/hhs-cybersecurity-task-force-releases-report-congress/〉 (accessed 2018-08-20)
[26] TrendMicro: GSS Feature Request: Ghost Solution Suite Web console (recompile the DS 6.9 web console with new .NET) (online), available from 〈https://www.symantec.com/connect/blogs/hhs-cybersecurity-task-force-releases-report-congress/〉 (accessed 2018-08-20)
[27] TrendMicro: Challenges with Critical Infrastructure: IoT, Smart Cities Under Attack(online), available from 〈https://blog.trendmicro.com/challenges-with-critical-infrastructure-iot-smart-cities-under-attack/〉 (accessed 2018-08-20)
[28] Symantec: Insourcing an entire IT organization from an outsourced model: (online), available from 〈https://www.symantec.com/connect/blogs/insourcing-entire-it-organization-outsourced-model〉 (accessed 2018-08-20)
[29] Cisco: Securing Medical Devices – The Need for a Different Approach – Part 1 (online), available from 〈https://blogs.cisco.com/security/securing-medical-devices-the-need-for-a-different-approach-part-1〉 (accessed 2018-08-20)
[30] Symantec: Personalized spam campaign targets Germany(online), available from 〈https://www.symantec.com/connect/blogs/personalized-spam-campaign-targets-germany〉 (accessed 2018-08-20)
[31] Barracuda: Threat Spotlight: The airline phishing attack(online), available from 〈https://blog.barracuda.com/2017/03/30/threat-spotlight-the-airline-phishing-attack/〉 (accessed 2018-08-20)
[32] Barracuda: Gmail Phishing Attack Strikes at Heart of API Economy(online), available from 〈https://blog.barracuda.com/2017/05/05/gmail-phishing-attack-strikes-at-heart-of-api-economy/〉 (accessed 2018-08-20)
[33] TrendMicro: OAuth Phishing On The Rise(online), available from 〈https://blog.trendmicro.com/oauth-phishing-rise/〉 (accessed 2018-08-20)
[34] Mcauliffe, J. D. and Blei, D. M.: Supervised topic models, pp.121-128, Advances in neural information processing systems, 2008.
[35] Lacoste-Julien, S., Sha, F. and Jordan, M. I.: DiscLDA: Discriminative learning for dimensionality reduction and classification, pp.897-904, Advances in neural information processing systems, 2009.
[36] Ramage, D., Hail, D., Nallapati, R. and Manning, C. D.: Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora, vol.1, pp.248-256, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Aug, 2009.
[37] Mukherjee, A and Liu, B.: Aspect extraction through semi-supervised modeling, vol.1, pp.339-348, Proceedings of the 50th annual meeting of the association for computational linguistics, Jul, 2012.
[38] Andrzejewski, D., Zhu, X. and Craven, M.: Incorporating domain knowledge into topic modeling via Dirichlet forest priors, pp.25-32, Proceedings of the 26th annual international conference on machine learning, Jun, 2009.
[39] Hu, Y., Boyd-Graber, J., Satinoff, B. and Smith, A.: Interactive topic modeling, vol.95, no.3, pp.423-469, Machine learning, 2014.
[40] Joshi, A., Lal, R., Finin, T. and Joshi, A.: Extracting cybersecurity related linked data from text, pp.252-259, Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on, Sep, 2013.
[41] Mittal, S., Das, P. K., Mulwad, V., Joshi, A. and Finin, T.: Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities, pp.860-867, Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Aug, 2016.
[42] McNeil, N., Bridges, R. A., Iannacone, M. D., Czejdo, B., Perez, N. and Goodall, J. R.: Pace: Pattern accurate computationally efficient bootstrapping for timely discovery of cyber-security concepts, vol.2, pp.60-65, Machine Learning and Applications (ICMLA), 2013 12th International Conference on, Dec, 2013.

表 7: 分類したクラスタの概要

カテゴリ名	セキュリティレポートのタイトル	文章例	発行元	発行日	カテゴリ頻出単語
ソーシャルエンジニアリング	In case of Email Outbreak for same subject or same …	Organizations wants to block the outbound emails which is going outside the …	Symantec	2017-06-20	attack email
	Phishing: A Main Concern for Enterprise Security	38 percent of companies cited end users being fooled by phishing and social engineering …	TrendMicro	2017-08-23	threat attacker
	The prevalent security threats you will see in 2018	Email will remain the most common delivery vehicle for advanced threats. In addition to …	Barracuda	2017-12-21	organization user
	Introducing the Strongest Protection and Visibility for Business Email Compromise	Moreover, you can exclude trusted senders from these controls by whitelisting specific …	Symantec	2017-07-27	use malware
	Spear phishing and ransomware are a perfect match	One anti-phishing company has found that 93% of all phishing emails contained ransomware…	Barracuda	2017-06-08	victim target
トロイの木馬	Android ad malware on Google Play combines three deception techniques	The app connects to a command and control (C&C) server on port 9001 to receive commands.	Symantec	2017-02-03	malware user
	3 types of cyberbullying that threaten students	off-the-shelf packages sold on the cybercrime underground have made it a cinch for …	Barracuda	2017-08-22	device figure
	Sage 2.0 ransomware delivered by Pandex spambot, mimics Cerber routines	The threat arrives on compromised computers as a .ZIP file attached to sexually explicit …	Symantec	2017-02-13	ransomware victim
	Sathurbot very aggressive against WordPress sites	The recently observed Trojan, Sathurbot, offers a fascinating insight into the various parts of …	Barracuda	2017-04-11	app child
	Android O no! Android O causes problems for mobile ransomware developers	Android ransomware using system-type windows will no longer work on devices …	Symantec	2017-04-12	threat apps
カテゴリ名	セキュリティレポートのタイトル	文章例	発行元	発行日	カテゴリ頻出単語
医療業界	What is the role of the Chief Information Security Officer?	Last year was a banner year for cyber security criminals hoping to breach and steal …	TrendMicro	2017-03-20	data organization
	Takeaways from the 2016 Threat Landscape	Ransomware was the top threat in terms of volume and the amount of money generated …	TrendMicro	2017-02-28	one place
	HHS Cybersecurity Task Force Releases Report to Congress	working with 20 other healthcare subject matter experts as a member of the Healthcare …	Symantec	2017-06-14	company ransomware
	Leading by Example at the HITRUST Annual Healthcare Cybersecurity Conference	As mentioned, hospitals in particular have been singled out by cybercriminals looking to …	TrendMicro	2017-05-04	threat attack
	World Backup Day – Poor Security Practices …	Beyond these steps, many companies only think of IT security as a means to remain ….	Barracuda	2017-03-28	risk system
産業用制御デバイス	Our Exposed World – Exposed Cities in Europe	Finally, the other area of IoT devices that I found particularly interesting is that were …	TrendMicro	2017-11-28	device system
	Hajime worm battles Mirai for control of the Internet of Things	Hajime is built on a peer-to-peer network. There isn't a single C&C server address, instead …	Symantec	2017-04-18	iot iot device
	IoT Networked Medical Device Cyber Security	Considering that some of the medical devices stay in the healthcare provider networks for …	Symantec	2017-02-11	internet security
	Will 2017 be the Year IoT Threats Go Mainstream?	we're likely to see an uptick in highly targeted attacks aimed at compromising Industrial …	TrendMicro	2017-01-26	vulnerability internet of thing
	The Internet of Things – A Gateway to Smart Cities	A second paper digs into the critical industries and industrial control system (ICS) …	Symantec	2017-09-28	network thing
カテゴリ名	セキュリティレポートのタイトル	文章例	発行元	発行日	カテゴリ頻出単語
サービス妨害	Hajime worm battles Mirai for control of the Internet of Things	There are some features that are noticeably missing from Hajime. It currently doesn't …	Symantec	2017-04-18	vulnerability system
	Chicago Med takes on the biggest cyber threat of the year	The entertainment industry has been helping keep digital threats and cyber security in the…	Barracuda	2017-04-05	device use
	Protecting Critical Infrastructure from Cyber Threats	Of the High Risk threats, 71.4 percent were network attacks 31.4 percent were …	TrendMicro	2017-10-31	attack one
	IoT Security: Easy to Compromise, Not So Easy to Fix	The widespread availability of connected devices with default passwords and …	Symantec	2017-10-24	network vulnerable
	"WAF Prevents Massive Data Breach at Equifax"…	Three days after the breach announcement, a new issue was discovered with the …	Barracuda	2017-09-22	exploit iot
Webサイト改ざん	How to protect against the most advanced email-based attacks	organizations today face an ever-increasing number of email-based threats focused …	Cisco	2017-05-26	recipient email
	Threat Spotlight: Emailed Resumes and Advanced Persistent Threats	Each one of the attacks originated from a different email, and each one of them …	Barracuda	2017-01-27	asaf esa
	Why Scammers Want Your Tax Returns (and how to stop them)	It takes a firm understanding of the target company, how they operate, and even …	TrendMicro	2017-04-13	real-time spear phishing attachment
	Defending Against The $5B Cybersecurity Threat – Business Email Compromise	Their target thinks the email is coming from someone they trust and consequently, …	Cisco	2017-12-21	impersonation spear phishing
	Threat Spotlight: Email Malware Impersonates Secure Bank Messages	Additionally, you can deploy anti-phishing protection with Link Protection to look for …	Barracuda	2017-09-28	highlight threat stanford
カテゴリ名	セキュリティレポートのタイトル	文章例	発行元	発行日	カテゴリ頻出単語
犯罪組織1	Stolen Memories: Why cyber thieves attack personal over financial data	But the problem for the cybercriminal stealing credit card and bank account details is that…	TrendMicro	2017-04-06	hacker cybercriminal
	A Storm's a Coming: How businesses can defend …	By spotting obvious grammar and spelling errors, uncommon domains in URLs or …	TrendMicro	2017-04-25	year use
	DreamBot Shines a Light on the Need for Transaction Verification	DreamBot exposed the need for banks to move away from one-time passcodes (OTPs) as …	Symantec	2017-10-16	victim attack
	Double Whammy: When One Attack Masks Another Attack	In this setup, one attack serves as a distraction, masking the malicious activities of the …	TrendMicro	2017-11-20	researcher target
	From hackers' point of views: New study exposes their strategies	some of the most powerful insights are coming from the people these cyber security …	TrendMicro	2017-04-03	way deep web
犯罪組織2	Attackers target dozens of global banks with new malware	The attacks came to light when a bank in Poland discovered previously unknown ….	Symantec	2017-02-12	use user
	Shamoon: Multi-staged destructive attacks limited to specific targets	Symantec discovered a high correlation between Timberworm and the presence …	Symantec	2017-02-27	malware attack
	Android malware on Google Play adds devices to botnet	One of the malicious apps posing as a skin app for Minecraft PE. The legitimate purpose of …	Symantec	2017-10-18	attacker figure
	The Wikileaks Vault 7 Leak – What We Know So Far	Since none of the tools and malware referenced in the initial Vault 7 disclosure have been …	Cisco	2017-03-07	link one
	Massive Google Phishing Attack Seen and Remediated by CloudSOC	There was a very large, sophisticated phishing attack on May 3 using a malicious …	Symantec	2017-05-05	computer access