悪性 Web サイトに到達しやすい危険検索単語の検知

源平 祐太[1,a)]　中川 雄太[1]　高田 一樹[1,2]　小出 駿[1,3]　金井 文宏[1,3]　秋山 満昭[3,4]　田辺 瑠偉[4]
吉岡 克成[5]　松本 勉[5]

概要：Web 媒介型攻撃を行う悪性サイトはインターネット上の重大な脅威の 1 つである. これらの悪性サイトの到達経路の 1 つとして検索エンジンを用いたアクセスが考えられるが，攻撃者はブラックハット SEO などの実装により, 特定の単語で検索を行ったユーザをより高確率で悪性サイトに誘導することができる．そこで，本研究では悪性サイトに到達しやすい危険な検索単語を特定する手法を提案する．加えて, これらの単語で検索を行った際に悪性サイトへ到達する期間に着目して分析を行い, 特定の観測期間でのみ危険な検索単語と観測期間全体を通して危険な検索単語を特定する．本研究では，数十万規模のユーザの Web アクセスログを用いて，実際に悪性サイトに到達した危険な検索単語を抽出した．評価実験の結果，観測期間全体を通して危険な検索単語として 5 つの単語が統計的優位かつ悪性サイト到達率が高い単語であると判定された．また，これらの危険検索単語を含んだ検索を行うことで，悪性サイトに到達するリスク比は一般の検索単語と比べて最大 11 倍以上となった．提案手法は，ユーザの検索段階で危険な検索単語を警告することで，悪性サイトへの到達を事前に防げる点で有用である．

キーワード：Web セキュリティ, 検索エンジン，ブラックハット SEO

Detecting Risky Search Words that Leads to Malicious Websites

Yuta Gempei[1,a)]　Yuta Nakagawa[1]　Kazuki Takada[1,2]　Takashi Koide[1,3]　Fumihiro Kanei[1,3]
Mitsuaki Akiyama[3,4]　Rui Tanabe[4]　Katsunari Yoshioka[5]　Tsutomu Matsumoto[5]

Abstract: Malicious Websites that conduct Web-based attacks are one of the major threats on the Internet. A possible route for victims to reach these malicious Websites is via search engines. Moreover, attackers may utilize techniques such as BlackHat SEO to attract more users to their Websites. In this study, we propose a method to detect risky search words that lead to malicious Websites. In the evaluation using Web access log of several hundreds of thousands of users, we identified five risky search words that constantly lead users to malicious Websites in the observation period. In the worst case, the probability of visiting malicious Website was 11 times as much as that of regular search words. We also extracted several search words that temporarily become risky in certain periods. We believe our method is useful for early warning to search engine users.

Keywords: Web Security,Search Engines,BlackHat SEO

[1] 横浜国立大学大学院 環境情報学府
Graduate School of Environment and Information Sciences, Yokohama National University
[2] 株式会社セキュアブレイン
SecureBrain Corporation
[3] NTT セキュアプラットフォーム研究所
NTT Secure Platform Laboratories
[4] 横浜国立大学先端科学高等研究院
Institute of Advanced Sciences, Yokohama National University
[5] 横浜国立大学大学院環境情報研究院/先端科学高等研究院
Graduate School of Environment and Information Sciences, Yokohama National University / Institute of Advanced Sciences, Yokohama National University
[a)] gempei-yuta-ts@ynu.jp

1. はじめに

近年，Web 媒介型攻撃はインターネット上の重大な脅威となっている [1]．攻撃者は，ドライブバイダウンロードやフィッシングなどの攻撃を行う悪性サイトをあらかじめ用意し，ユーザの興味を引く Web コンテンツや正規サイトの改ざんなどによってユーザを悪性サイトに誘導する．誘導されたユーザはマルウェア感染や個人情報の搾取などの被害を受ける．このような悪性サイトの検知および対策は活発に実施されている．例えば，悪性サイトに特有な URL・Web コンテンツ・アクセス順序などを利用した検知や，検知した悪性サイトをブラックリストとして活用する対策が行われている．しかしながら Web 媒介型攻撃は多様化を続けており，攻撃者による新しい脆弱性やソーシャルエンジニアリング手法の考案によって従来の検知手法が回避されることや，日々新たな悪性サイトが出現し続けていることから，ブラックリストであらゆる悪性サイトを列挙するのは難しい．

ユーザが Web サイトにアクセスする際の主要な起点は検索エンジンである．攻撃者は特定の単語による検索結果の順位の上位に悪性サイトが含まれるように不正に順位を操作すること（ブラックハット SEO）でユーザを悪性サイトに誘導できる．このような単語は一般名称以外にも特定の組織名，著作物名，ファイル種別名などが考えられる．攻撃者がマルウェアを配布する際に特定の著作物を偽装することはよく知られているからである．このような検索エンジン経由での悪性サイトへの誘導方法自体はよく知られているが，検索エンジンで検索した単語がその後のユーザの悪性サイト到達にどの程度影響しているか明らかにした研究は我々の知り得る限り行われていない．

我々の研究は，検索エンジンにおいて悪性サイトに到達しやすい危険検索単語を入力した際にその危険性をユーザに通知することで，悪性サイトへのアクセスを未然に防ぐことを目的とする．その第一歩として，本稿では検索エンジン経由で悪性サイトに到達したユーザの検索単語と，正規サイトにアクセスしたユーザの検索単語を分析し，悪性サイトに到達しやすい危険検索単語を検知する手法を提案する．具体的には，大規模な Web アクセスログからユーザの検索単語を抽出し，検索を行った際に悪性サイトへ到達しやすい単語について，(1) 観測期間全体を通して危険な検索単語と，(2) 特定の観測期間でのみ危険な検索単語を特定した．評価実験の結果，前者についてはファイル種類名を表す単語などの計 5 つの単語が統計的優位かつ悪性サイト到達率が高い単語であると判定され，これら危険検索単語を含んだ検索を行った際の悪性サイトに到達するリスク比は良性な検索単語のみを含む検索と比べて最大 11 倍以上となった．後者については複数の期間で危険な検索単語を検知することができ，ある組織名を含む検索が一時的に悪性サイトに到達しやすくなっていたことが確認された．よって提案手法は悪性サイトへのアクセスを未然に防げる点で有用であることが判明した．最後に，危険な検索単語をクエリしたユーザに対して警告を行うシステムについて考察した．

以降では，2 章で関連研究について述べる．3 章で悪性サイトに到達しやすい危険検索単語を検知する手法を提案する．4 章で評価実験について述べ，5 章で考察を行う．最後に，6 章でまとめと今後の課題を述べる．

2. 関連研究

2.1 悪性サイト検知

悪性サイト検知手法については多数提案されている．ブラックリスト/ホワイトリストを利用した検知手法 [2] や Exploit Kit による URL 文字列特徴を用いた検知手法 [3] が提案されている．しかしながら IP アドレスや URL 文字列特徴など攻撃者が変更可能な要素による検知は検知対策されてしまう可能性が考えられる．

またサイト内コンテンツ情報から検知を行う手法は，より高精度な検知手法が多数提案されている．ドライブバイダウンロード（DBD）攻撃を HTTP ヘッダ情報を基に検知する手法 [4] や悪性 JavaScript を検知する手法 [5], [6] などが提案されている．また論文 [7] ではフィッシングサイトが正規のサイトと類似する特徴や低コストで運用される特徴，ユーザを騙す UI を持つこと利用して，HTML と画像解析を組み合わせたフィッシングサイト検知手法を提案している．しかし, エンドユーザが被害に遭ってしまう Web 媒介型攻撃は DBD・ソーシャルエンジニアリング攻撃など攻撃手法が多岐に渡るため，検知に必要とされる Web コンテンツも種類が多く検知コストが高い．また悪性サイトにアクセスした段階でブラウザフィンガープリントを盗まれたり悪質な Cookie をセットされたりする可能性がある．

そのためエンドユーザが悪性サイトに到達した際の検知だけではなく，悪性サイトの到達を未然に防ぐことも大きな課題の 1 つである．この課題の解決のため，我々は検索エンジンにおける検索単語に着目した．Web アクセスログ中から悪性サイトに到達しやすい検索単語を明らかにすることができれば，悪性サイト到達前にエンドユーザに通知を行うことができる．クローリングによるブラックハット SEO の実態調査 [8] は行われているが，悪性サイトに到達しやすい検索単語を実際のエンドユーザの Web アクセスログ中から分析する研究は行われていない．

2.2 エンドユーザに対する通知

Web サイトの管理者やエンドユーザへセキュリティ・プライバシーに関する通知を行い，その効果を検証する研究も行われている．論文 [9] では HTTPS の設定が誤ってい

る Web サイトの管理者に対して，通知文面・通知経路を変えながら通知実験を行い，通知に有効な特徴量について比較分析を行っている．このような脆弱性がある Web サイトの管理者を対象とした通知研究が多く行われている．

一方で，本研究で提案する通知ではエンドユーザを対象としている．エンドユーザを対象とする通知では Web サイト管理者を対象とする通知と比較して，通知経路やユーザのセキュリティ意識が異なっている．エンドユーザを対象とする通知研究では，ブラウザを経由した SSL 通知を無視するユーザに対して，アンケートを行いその心理を分析した研究 [10] がある．本研究では同じくエンドユーザを対象とするが，未来に発生しうる危険性について通知することでエンドユーザの最適な意思決定を援助することを目的としている．

3. 悪性 Web サイトに到達しやすい危険検索単語の検知

本章では，悪性サイトに到達しやすい危険検索単語を検知する手法を提案する．提案手法は大規模な Web アクセスログを入力として，ユーザの検索単語を抽出する検索クエリ抽出フェーズと，悪性サイトにアクセスしたユーザと正規サイトにアクセスしたユーザの検索単語を比較する分析フェーズにより，(1) 観測期間全体を通して危険な検索単語と，(2) 特定の観測期間でのみ危険な検索単語を出力する．以降では，3.1 節で提案手法の全体像を説明し，3.2 節で検索クエリ抽出フェーズを説明する．そして，3.3 節と 3.4 節で分析フェーズを説明する．

3.1 提案手法の全体像

提案手法の概要を**図 1** に示す．提案手法で入力とする Web アクセスログには，ユーザ識別子（以降では，UID と呼ぶこととする），ユーザがアクセスした URL の情報，URL にアクセスした時のタイムスタンプ，の 3 つの情報が含まれることとする．また，多数のユーザが存在する大規模な Web アクセスログであるとともに，ユーザがアクセスした URL には検索エンジンで検索した単語も含まれることを前提とする．

ユーザが Web サイトにアクセスする際の主要な起点は検索エンジンであり，ユーザは検索エンジンを利用して膨大な Web 空間から自身の求める Web コンテンツを高速に発見することができる．しかし，その一方で悪性サイトに誘導される可能性がある．例えば，特定の単語による検索結果の上位に悪性サイトが含まれるように不正に順位を操作するブラックハット SEO が知られており，実際に違法コンテンツ配信サイトなどが攻撃に用いられた事例が報告されている [13]．また，攻撃者がマルウェアを配布する際に特定の著作物を偽装することはよく知られており，“Free” や “Download” といった単語による検索がソーシャルエンジニアリング攻撃サイトに到達しやすいことが確認されている [14]．同様に，攻撃者は時事的なニュースや実際に存在するイベント名に関連したサイトを公開することで，検索単語に関連したサイトであると勘違いしたユーザを悪性サイトに誘導することができる．実際に，”東京オリンピック” といった単語で検索したユーザを狙った攻撃が報告されている [15]．加えて，正規サイトを改ざんして他のサイトへ誘導するコードを Web ページ内に挿入することで，ユーザを自動的に悪性サイトへ誘導することができる [16]．

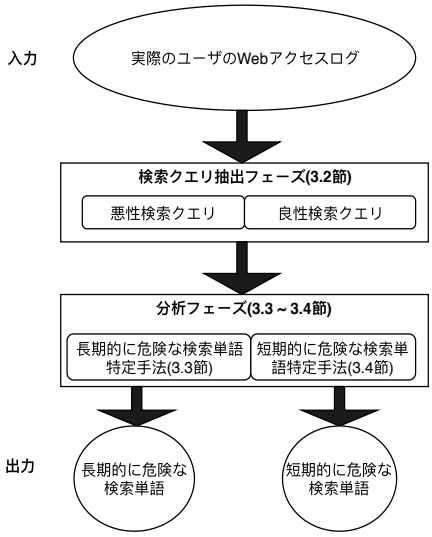

図 1: 提案手法の概要

そこで，提案手法では違法サイトやマルウェアダウンロードサイトなどの悪性サイトに到達する可能性のある危険な検索単語を検知対象とする．以降では，観測期間全体を通じて出現するそのような単語を長期的に危険な検索単語と呼ぶこととする．また，特定のイベントや話題に乗じることで検索結果から悪性サイトに到達する可能性のある危険な検索単語も検知対象とする．以降では，このような単語を短期的に危険な検索単語と呼ぶこととする．

3.2 検索クエリ・単語の抽出

本節では，検索単語抽出フェーズの流れを説明する．はじめに，Web アクセスログから検索単語の抽出を行う．通常，ユーザの多くは幾つかの単語を組み合わせて検索を行う（以降では，検索クエリと呼ぶこととする）．なお，4 章の評価実験では Google [11], [17]，Yahoo [12]，Bing [18] の 3 つの検索エンジンのいずれかを用いて検索された単語を対象とした．これらの検索エンジンは日本の検索エンジンの使用率の 99%以上を占めている [20]．これら検索エンジンでは，ユーザが検索した単語は URL パラメータの一部に含まれる．例えば，Google を用いて “keyword1 keyword2” と検索した場合，スペースをプラス記号に置き換えた URL パス`/search?q=keyword1+keyword2&`を含むリクエストが送信される．

続いて，Web アクセスログから検索クエリを検索した後

に到達した Web サイトの URL の抽出を行う．Web アクセスログに含まれている情報は様々であるが，全てのログにリファラなどの情報が保存されているとは限らない．このため，一連の Web アクセスの流れを正確に把握することは困難である．そこで，Web アクセスログから検索クエリを発見した際，検索から閾値 t_{lim}[秒] 以内に到達しているサイトを検索単語に関連するサイトであると判断する．閾値の決定方法については，5 章において考察するが，4 章の評価実験では閾値 t_{lim}=300 とした．すなわち，サイト到達時刻から 5 分以内かつ直前の検索 URL から抽出された検索クエリを指す．

最後に，悪性サイトに到達した危険な検索クエリ群と正規サイトに到達した良性な検索クエリ群に分類する．悪性サイトの判定方法は様々であるが，4 章の評価実験では Google Safe Browsing [19](以降では，GSB と呼ぶこととする) を用いて，抽出した検索クエリを用いた検索後に悪性サイトに到達していたかどうかで判断した．悪性サイトに到達していた場合，その検索クエリを悪性検索クエリとする．また，悪性サイトに到達していなかった場合，その検索単語を良性検索クエリとする．ただし悪性サイト検知ツールが悪性サイト及び悪性サイトの中継サイトをも連続で検知することがあることを踏まえて，同一ユーザの同一検索クエリが同日中に重複して悪性検索クエリとして抽出されないようにした．また，検索クエリはスペースなどを境に複数単語から構成されているケースが多い．3.3，3.4 節の分析においては単語の出現情報を元に分析を行うため検索クエリを区切り文字で分割した際の単語（検索単語）単位での分析を行う．

3.3 長期的に危険な検索単語の検知

本節では，長期的に危険な検索単語を検知する流れを説明する．なお，本フェーズでは長期間の Web アクセスログを入力データとする．

3.3.1 多変量データ作成

悪性検索クエリと良性検索クエリとの単語の出現情報から単語の出現情報 (2 値) を持つ多変量データを作成する．作成する多変量データは解析対象としたい n 単語数分の次元を持ち（$w_1, w_2 ... w_n$），単語 i を含んでいれば w_i は 1，そうでなければ 0 とする．また目的変数 apt は悪性検索クエリであるかどうかの 2 値であり悪性検索クエリなら 1, 良性検索クエリなら 0 である．この際比較分析のため両データセットに共通して出現する単語の出現情報のみ多変量データとする．

3.3.2 分析手法

今手法で抽出される悪性検索クエリは良性検索クエリに比べて非常に少ない. このように悪性ケースが少ないためケースコントロール研究を援用している論文 [21] による回帰分析が有効であると判断した. 研究 [21] では標的型メールの件名から標的型メールかどうかのリスク比について分析している．これを分析手法に多変量データ作成とロジスティック回帰分析を行う.

ここでロジステック回帰分析とは以下式を用いて偏回帰係数 a_n を導出する非線形回帰分析手法である.

$$p = \frac{1}{1 + e^{-(a_1 x_1 + a_2 x_2 + ... a_n x_n + b)}}$$

ロジステック回帰分析により各単語の出現情報 w_n から各単語のリスク比，有意水準，95%信頼区間が導出を行う．ここでリスク比とはオッズ比のことを指す. まずオッズとは事象が起こる確率と起こらない確率の比である. オッズ比は一方のデータセットにおけるオッズと他方のデータセットにおけるオッズの比のことである. この導出により今分析ではその単語が検索クエリに含まれている場合と含まれていない場合とで悪性サイトに到達する確率がどれくらい高まるかを導出・評価できる.

また有意水準とは帰無仮説が棄却基準となる確率を指す．ロジスティック回帰分析で変数を絞り込む場合，絞り込む有意水準を低く設定すると重要な変数を見落とす可能性が出てくる．そこで提案手法では統計的優位となるのは有意水準 $p <= 0.10$ と設定した.

3.3.3 分析単語フィルタリング

エンドユーザが使う単語数は非常に多いため 3.3.1 項で作成した多変量データを回帰分析に導入しても回帰分析が収束しない可能性が高い．悪性サイトの到達により影響度の高い検索単語及び悪性サイトに到達する恒常的にユーザに検索される単語のみを重点的に解析を行うため以下 2 つの条件を満たす単語のみを回帰分析に導入することとした.

(1) 単語出現情報 w_n と目的変数とのテトラコリック相関係数が 0.4 以上となる検索単語
(2) 観測期間内のデータセット中で 2 回以上出現する検索単語

ここでテトラコリック相関係数とは 2 値の順序変数同士の相関係数である．条件（1）は社会心理学的に比較的相関があるとされる相関係数が 0.4 以上単語のみをフィルタリングすることで悪性サイト到達率に影響があると思われる単語のみを回帰分析に導入するために設定した. またエンドユーザが使う検索単語はあらゆる固有名詞が含まれるためワードプールが非常に広い．今分析では普遍的に被害ユーザが使う単語に焦点を当てるため 1 回しかでてこない低頻度な出現検索単語を除外する条件（2）を適用することとした.

3.4 短期的に危険な検索単語の検知

本節では，短期的に危険な検索単語を検知する流れを説明する．なお，本フェーズではある一定期間の Web アクセスログを入力データとする.

まずはじめに，一時的にアクセス数が増加する単語の抽

出を行う．具体的には，分析期間内で観測される単語に対して以下の 3 つの処理を行う．

(1) 期間内である 1 日しか出現しない単語を除外する処理
(2) 期間内でその単語での検索を行なったユーザ数が 2 倍以上となる日があり，なおかつ 10 人以上にユーザに検索された日がある単語を残す処理
(3) 期間内の Web アクセスログの連日のユーザ数遷移と単語の期間内検索ユーザ数の遷移との相関係数が 0.6 以上である単語を除外する処理

(1) は期間内の他の日と比較できないためである．(2) は前日との比較により一時的に検索ユーザ数が増える単語を抽出するためである．観測期間内で日によって Web アクセスログ中から確認される利用ユーザ数に 2 倍以上の差が発生することが確認されたため，(3) ではその日の使用ユーザ数が増えたために検索ユーザ数が増えた単語は本分析対象外のため除く処理を行った．

得られた一時的にアクセスが増加した単語を含む検索クエリによって期間内で悪性サイトに到達しているユーザが複数人いる場合，その単語を短期的に危険な検索単語とする．

4. 評価実験

本章では，Web アクセスログを悪性サイトに到達しやすい危険検索単語を検知する．以降では，4.1 節では長期的に危険な検索単語の分析と，得られた長期的に危険な検索単語による悪性サイト到達率の評価結果を説明する．4.2 節では短期間のみ危険な検索単語の抽出結果を説明する．

4.1 実験 1：長期的に危険な検索単語の検知

実験 1 では，提案手法を用いて大規模な Web アクセスログから長期的に危険な検索単語の検知を行った．また，ユーザが検索エンジンにおいて悪性サイトに到達しやすい危険検索単語を入力した際に，その危険性を通知する状況を想定して，検知した検索単語と Web アクセスログのマッチングを行った．実験には連携先組織から提供された数十万規模のユーザが存在する Web アクセスログから，2017/7/1〜2019/1/8 の期間のうちランダムにサンプリングした 120 日分のデータを入力データとして用いた．

はじめに，入力データから検索単語の抽出を行った．ここで，Web アクセスログに含まれる良性の検索クエリは約 4,500 万件であり，悪性な検索クエリと比べてその数が多かったため,10 万件の検索クエリをランダムサンプリングした．そして，入力データから 4,180 件の悪性検索クエリとそれらを単語に分解したユニークな 6,841 個の危険検索単語，10 万件の良性検索クエリとそれらを単語に分解したユニークな 95,898 個の良性検索単語を抽出した．

続いて，これらの単語を比較分析するため，検索クエリに含まれていた単語のうち共通する 1,785 個の単語の出現情報を特徴量とした多変量データを作成した. さらに，1,785 次元の多変量データを全て回帰分析に導入しても収束しないことが予想されるため，3.3 節のフィルタリングを適用した．その際，テトラコリック相関係数の導出は R の polycor [22] パッケージで導出した．その結果，29 個の単語がロジステック回帰分析の対象となった．29 個の単語のロジステック回帰分析の結果を**表 2** に示す．ただし，研究倫理の観点から一部の単語はカテゴライズしたラベルのみを示す．表 2 において，灰色で囲われている単語は統計的優位かつリスク比が 1.0 より大きい単語を示している．なお，ロジステック回帰分析には python の statsmodel パッケージ [23] を用いて，最尤法にて偏回帰係数を導出した．

本分析によりファイル種類名やアニメ配信サイト名などを含む “crack”，“rar”，“Web サイト名”，“torrent”，“アダルト” の 5 単語が長期的に危険な検索単語として検知された．そのリスク比は最大で 11 倍となり回帰分析に含まれた単語が 95%信頼区間における最大リスク比は約 85 倍となる．統計的優位となった 5 単語以外で回帰分析にはファイル種類名や会社名・アダルト単語などが導入された．これはコンテンツ配布や動画像配信を行っているサイトが悪性サイトに誘導されやすいことが確認できた．

最後に，ロジスティック回帰分析に含めていない期間の Web アクセスログから，先の分析で得られた長期的に危険な 5 つの検索単語を検索して悪性サイトに到達したユーザ数の調査を行った．また，比較を行うため，検索ユーザ数の高い上位 3 単語（“無料”,“意味”,“おすすめ”）と，先で検知された 5 つの危険検索単語及び検索ユーザ数の高い上位 3 単語以外の検索単語からランダムサンプリングした 20 単語を検索したユーザのうち，悪性サイトに到達したユーザ数の調査を行った．

各単語の危険性の評価には，n 日の観測期間の内ある日 i において単語 w を含む検索を行い，どれくらいの割合のユーザが悪性サイトに到達したかを示す悪性サイト到達割合 $ratio_{wi}$ を用いた．ここで，ある日 i において単語 w を含む検索を行ったユーザの合計を $total_{wi}$ で表し，ある日 i において単語 w を含む検索を行ったユーザうち 5 分以内に悪性サイトに到達したユーザの合計を $malicious_{wi}$ とすると，悪性サイト到達割合 $ratio_{wi}$ 及び期間内の平均悪性サイト到達率 $ratio_w$ はそれぞれ以下の式で算出される．

$$ratio_{wi} = malicious_{wi}/total_{wi}$$

$$ratio_w = \sum_{i=1}^{n} ratio_{wi}/n$$

評価実験における各単語群の平均 $total_w$，平均 $malicious_w$，平均 $ratio_w$ を**図 3** に示す．本評価実験では危険検索単語である 5 単語の $ratio_wi$ は期間内最高 $ratio_{rar12}$ で 0.4 で危険検索 5 単語の平均 $ratio_w$ は 0.0228 であった．他比較単語として (1) 検索ユーザ数が多い 3 単語の内

表 1: 120 日の観測期間内での各検索クエリ数

悪性検索クエリ数	良性検索クエリ数	全クエリに占める悪性クエリ比率
4,180	45,014,084	0.00009286

表 2: 29 単語の回帰分析結果

単語	リスク比	有意水準	95%信頼区間
crack	11.000052	0.0217	1.420203257～85.19987784
rar	6.892266597	0	2.913339448～16.30382919
「Web サイト名 1」	4.999810441	0.1418	0.584148528～42.79836924
「Web サイト名 2」	4.999810441	0.0377	1.095597635～22.81914782
「地名 1」	4.945113909	0.2092	0.408035948～59.93136555
asian	4.269513985	0.1971	0.470386774～38.75268331
「アダルト 1」	4.000022556	0.215	0.447087927～35.78754758
「会社名 1」	4.000022556	0.215	0.447087927～35.78754758
「会社名 2」	4.000022556	0.215	0.447087927～35.78754758
full	4.000022556	0.215	0.447087927～35.78754758
hd	3.878311971	0.2238	0.436703851～34.44280079
key	3.607806608	0.257	0.392428863～33.16847909
torrent	3.58801814	0.0028	1.554416136～8.282128495
お金の入れ方	2.999963134	0.1785	0.605500434～14.86337302
「アダルト 2」	2.676740974	0.094	0.845607479～8.472282899
girl	2.591663636	0.2895	0.444947047～15.09705965
serial	2.455916472	0.4461	0.243533687～24.76669978
クラック	1.999905641	0.4235	0.366337588～10.91895205
「アダルト 3」	1.517857802	0.4278	0.541073581～4.258001846
mp3	1.322865213	0.3962	0.693156136～2.52439139
free	0.948949211	0.9	0.41941265～2.147275867
「アダルト 4」	0.941387903	0.9177	0.299182956～2.962104514
HD	0.860794052	0.805	0.261767127～2.830348928
「アダルト 5」	0.853422964	0.5875	0.481331046～1.513159729
「アダルト 6」	0.824894318	0.5941	0.406447707～1.674308091
japanese	0.818157842	0.6066	0.381211858～1.755932404
black	0.800034842	0.7394	0.214831775～2.97903672
「地名 2」	0.712055088	0.7153	0.114772887～4.417173447
zip	0.587605191	0.2025	0.259395851～1.331092455

表 3: 評価実験における各単語群の平均 $total_w$，平均 $malicious_w$，平均 $ratio_w$

	危険とされる 5 単語	検索ユーザ数の多い上位 3 単語	サンプリングした 20 単語
平均 $total_w$	668	7.60	2.08
平均 $malicious_w$	0.157	0.0833	0.0
平均 $ratio_w$	0.0228	0.000123	0.0

平均悪性サイト到達率が最高となったのは，$ratio_{\text{無料}}$ の 0.000304 であった．また，これら 3 単語の平均 $ratio_w$ は 0.0001232 で危険検索単語の平均 $ratio_w$ が約 185 倍高いという結果が得られた．さらに（2）でのサンプリングした単語との比較では観測期間内で危険検索単語と同等以上の $total_w$ を持つ単語が含まれながらも $ratio_w$ は 20 単語全てにおいて 0.0 であり，一般の検索単語においては悪性サイト到達がほとんど起こらないという結果となった．これら評価実験結果より分析対象期間外である期間においても先の分析で得られた危険検索単語は悪性サイトに到達する確率が他検索単語と比べ著しく高いことが判明した．

表 4: 28 日間の観測期間内における各処理後の検索単語数

期間内検索単語数	処理 1 後の検索単語数	処理 1,2 後の検索単語数	処理 1,2,3 後の検索単語数
3,466,720	705,577	8,982	6,146

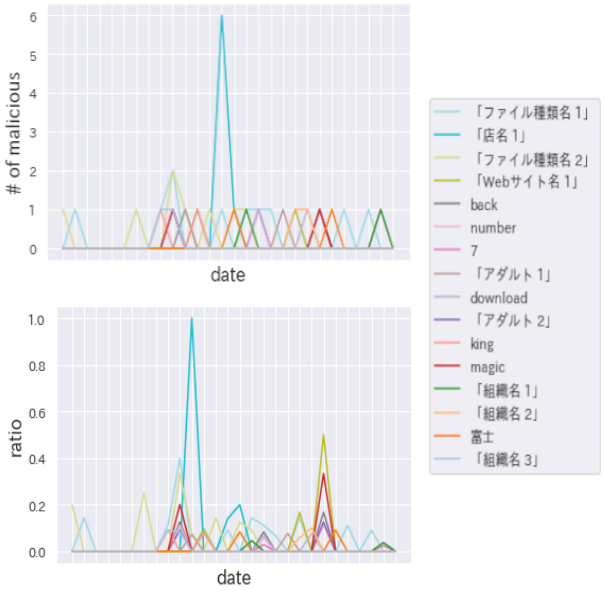

図 2: 提案手法で抽出された 16 単語の $malicious_w$ と $ratio_w$

4.2 実験 2：短期間でのみ危険な検索単語の検知

実験 2 では，提案手法を用いて大規模な Web アクセスログから短期的に危険な検索単語の検知を行った．実験には連携先組織から提供された数十万規模のユーザが存在する Web アクセスログから，2019/4/3～2019/4/30 の 28 日分の Web アクセスログを入力データとして用いた．

提案手法では，28 日分の Web アクセスログから 300 百万以上の単語が抽出された．表 4 に観測期間内で 3.4 節の提案手法を処理を適応した後の単語数を示す．この内危険検索単語であったのは 129 単語であった．このうち，観測期間内で悪性サイトに到達したユニークなユーザ数が 2 人以上の単語が 16 単語であった．図 2 にこれらの単語を含む検索を行ったユーザのうち 5 分以内に悪性サイトに到達したユーザの合計 $malicious_w$ と悪性サイト到達割合 $ratio_w$ を示す．

検出された 16 単語には組織名やサイト名などを含む．これは組織が運用する Web サイトや特定の Web サイトが悪性サイトに到達するよう一時的に改ざんされた可能性が考えられる．図 2 の $malicious_w$ では各単語において観測期間の内 1 日もしくは連続した数日のみ $ratio_w$ が高まっていることが確認できる．期間内で $ratio_{\text{店名 1}}$ が最高で 1.0 となっている．特定の日において店名 1 を含む検索を行なったユーザが必ず悪性サイトに到達していることとなり，短期的に悪性サイト到達率の高い単語が検知されているといえる．

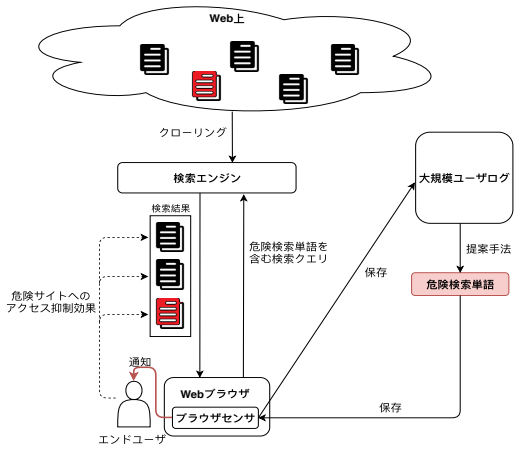

図 3: 通知モデル概要

5. 考察

提案手法の限界と研究倫理及び通知実験に向けた検討・3.2 節で用いた閾値 t_{lim} の決定のために行なった事前調査について述べる．

5.1 提案手法の限界

提案手法の危険検索クエリの抽出においてはタイムスタンプ情報のみから悪性サイトに到達していると仮定しているがその検索クエリが直接的に悪性サイトに到達に影響していないケースが考えられる．具体的には検索を行ったタブとは別のタブで 5 分以内で悪性サイトに到達してしまったケースであっても悪性な検索であると検知してしまう．

また本手法では検索クエリの単語の出現情報に着目している．しかし検索エンジンは検索単語の組み合わせである検索クエリ単位で解釈を行っている．検索単語の出現情報だけでなく検索単語がどのような単語とセットで使われているかの分析については今後の課題としたい．

5.2 研究倫理

本稿において検索単語として挙げられた一部の固有名詞はマスクしている．これは特定の企業名や商標が含まれておりそれらは攻撃の主体ではないこと，検索回数が少ない検索単語からユーザが特定されうる危険を考慮したためである．また，Web ログ中の URL にはクエリパラメータ中に個人情報が含まれているケースがあるため外部サービスにクエリパラメータが直接含まれる URL を公開しないように留意した．

5.3 通知実験に向けた検討

危険検索単語で検索するユーザに対して，その単語を含む検索結果からの悪性サイト到達率が高いことを通知することで，ユーザによる悪性サイトへのアクセスを防ぐことができる可能性がある．このような通知は，悪性サイト検知によるフィルタリング対策と補完的に活用できる対策として期待できる．本節ではその通知に向けた検討を行う．通知モデル概要図を**図 3** に示す．大規模ユーザログ中から提案手法で危険検索単語を抽出し，危険検索単語を含む検索をエンドユーザが行った場合，エンドユーザに対して得られた検索結果から悪性サイト到達率が高いことをブラウザ経由で通知する．

通知実験においては現在数千人オーダのユーザが利用している WarpDrive ブラウザセンサ [24] 経由での通知を行うことを想定している．WarpDrive ブラウザセンサは Google Chrome の拡張機能として実装されているが，検索エンジンは Google に限定されない．通知はポップアップを表示することでユーザが必ずその通知に気づくような UI として実装予定である．なお，通知対象となる検索単語は他の検索単語と比較して悪性サイトの到達する危険性が高いというだけで，危険検索単語で検索したユーザが必ずしも悪性サイト到達がなされているわけでない．そのため未来の危険性について通知を行う提案通知システムが有効かどうかの判断材料にするため，またユーザビリティの向上のため本通知システムの ON/OFF は簡易に切り替えられる UI 設計を行う予定である．

5.4 検索と悪性サイト到達の時間関係性

3 章で説明した閾値 t_{lim} の設定の事前調査として，UID, アクセス URL, アクセス時のタイムスタンプの 3 つの情報に加え,Web サイトアクセス時の HTTP ヘッダーやブラウザ利用時のタブ id 情報が含まれているログを利用した (以降高精度 Web ログ)．使用した高精度 Web ログ中は数千人オーダーのユーザが利用している．ただし高精度 Web ログは利用ユーザ数が大規模 Web ログより少数である点・解析コストが高い点を考慮して本事前調査にのみ用いることとした．

16 日間の高精度 Web ログから検索エンジンを利用して GSB で検知された悪性サイトに到達したユーザをタブ ID や HTTP リクエスト情報, 実際に検索した際の検索結果などの観点から検索エンジンでの検索から悪性サイトに到達している実例を抽出を試みた．その結果，悪性サイトに到達するユーザが検索から悪性サイト到達しているケースを 13 件確認した．確認された検索エンジンでの検索から悪性サイトに到達するまでの時間を**表 5** で示す．

Web サイトの滞在時間はサイトの種類によって大きく異なる．検索結果トップから悪性サイトに遷移しているケースでは到達時間は数秒程度であるが，サイト内でコンテンツを動画コンテンツを見た後に悪性サイトに到達しているケースも確認され，その場合到達までにかかる時間は 1000 以上の時間がかかっている．平均で約 245 秒となった．この事前調査の結果から本稿における t_{lim} の設定を 300 秒とした．

表 5: 高精度ログ中から確認された検索エンジンでの検索経由で悪性サイトに到達ケース例

	アクセス時間 sec
ケース 1	5.438
ケース 2	7.049
ケース 3	7.670
ケース 4	9.884
ケース 5	11.40
ケース 6	13.59
ケース 7	26.74
ケース 8	55.19
ケース 9	61.24
ケース 10	83.69
ケース 11	588.3
ケース 12	1154
ケース 13	1170

6. まとめと今後の課題

本稿では，数十万人規模のエンドユーザから得られる 膨大な Web ログ中から悪性サイトに到達しやすい危険検索単語の抽出・分析を行った．特に悪性サイトに誘導される期間に着目して，恒常的に悪性サイトに到達しやすい危険検索単語と特定の期間のみ悪性サイトへの到達率の高い危険検索単語の分析を行った．統計分析を用いた長期的に危険な検索単語の分析では検索クエリに含まれているだけで悪性サイトが到達率が高まる単語が 5 単語抽出され，そのリスク比は最大約 11 倍となった．短期的に危険な検索単語の分析では組織名などを含む 16 単語が特定期間に悪性サイト到達率が高まっていたことが確認された．

また危険検索単語で検索を行うユーザに対して通知実験に向けた検討を行った．未来の危険性に対して通知を行う研究は新規性が高く，検索エンジンでの検索単語をベースにした通知は現在行われてきていない．通知実験の実施及びその通知効果・ユーザビリティなどの評価を今後の課題とする．

謝辞 本研究成果の一部は、国立研究開発法人 情報通信研究機構 (NICT) の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」により得られた．

参考文献

[1] 情報セキュリティ 10 大脅威 2019, https://www.ipa.go.jp/files/000073293.pdf
[2] Ma, Justin, et al. "Beyond blacklists: learning to detect malicious web sites from suspicious URLs." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009.
[3] 佐藤祐磨, 中村嘉隆, and 高橋修. "エクスプロイトキットで利用される文字列特徴を用いた悪性 URL 検出手法の提案." 研究報告コンピュータセキュリティ (CSEC) 2016.25 (2016): 1-6.
[4] 酒井裕亮, and 佐々木良一. "Drive By Download 攻撃に対する HTTP ヘッダ情報に基づく検知手法の提案." 研究報告コンピュータセキュリティ (CSEC) 2013.29 (2013): 1-6.
[5] 神薗雅紀, et al. "抽象構文解析木による不正な JavaScript の特徴点抽出手法の提案." コンピュータセキュリティシンポジウム 2011 論文集 2011.3 (2011): 474-479.
[6] Wang, Junjie, et al. "Jsdc: A hybrid approach for javascript malware detection and classification." Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security. ACM, 2015.
[7] Corona, Igino, et al. "Deltaphish: Detecting phishing webpages in compromised websites." European Symposium on Research in Computer Security. Springer, Cham, 2017.
[8] Du, Kun, et al. "The Ever-Changing Labyrinth: A Large-Scale Analysis of Wildcard DNS Powered Blackhat SEO." 25th USENIX Security Symposium (USENIX Security 16). 2016.
[9] Eric Zeng,Frank LiFixing,Emily Stark,Adrienne Porter Felt, Parisa Tabriz, HTTPS Misconfigurations at Scale:An Experiment with Security Notifications,WEIS2019.
[10] Felt, Adrienne Porter, et al. "Improving SSL warnings: Comprehension and adherence." Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM, 2015.
[11] Google, https://www.google.com/
[12] Yahoo Japan, https://www.yahoo.co.jp/
[13] Rafique, M. Zubair, et al. "It's free for a reason: Exploring the ecosystem of free live streaming services." Proceedings of the 23rd Network and Distributed System Security Symposium (NDSS 2016). Internet Society, 2016.
[14] 小出駿, et al. "ユーザ操作が起点となる Web 上の攻撃の収集." 研究報告セキュリティ心理学とトラスト (SPT) 2018.16 (2018): 1-6.
[15] 朝日新聞：東京五輪関連と偽装、不正サイト続々「アドレスに注意」, https://www.asahi.com/articles/ASM5W7QWHM5WULZUOOW.html, (2019.08).
[16] トレンドマイクロ：正規 Web サイト改ざん , https://www.trendmicro.com/ja_jp/security-intelligence/research\-reports/threat\-solution/manipulation.html, (2019.08).
[17] Google, https://www.google.co.jp/
[18] Bing, https://www.bing.com/
[19] Google Safe Browsing, https://safebrowsing.google.com/
[20] statcounter, http://gs.statcounter.com
[21] 渡部正文. "企業グループに送られた標的型攻撃メールのソーシャルエンジニアリング視点からの分析." 研究報告セキュリティ心理学とトラスト (SPT) 2017.19 (2017): 1-1.
[22] polycor, https://r-forge.r-project.org/projects/polycor/
[23] statsmodels, https://www.statsmodels.org
[24] WarpDrive, https://warpdrive-project.jp/