社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2009－ＦＩ－94 (5)
2009－ＤＤ－70 (5)
2009／3／25

ソーシャルブックマークを用いたフィッシング検知技術

中山 心太
電気通信大学大学院 電気通信学研究科 人間コミュニケーション学専攻
〒182-8585 東京都調布市調布ヶ丘1-5-1
EMail: shinta.nakayama@gmail.com

概要 近年ソーシャルブックマーク(SBM)と呼ばれる，自分のブックマークをウェブ上に登録し，公開するサービスが流行っている．本研究は集合知をセキュリティに応用するという考え方に基づき，SBMを用いてフィッシング詐欺検知を行う．北米の銀行がSBMにどれほどブックマークされているかの実験を行った．その結果，約四割の正規サイトがSBMに登録されていることが分かった．またフィッシングサイトの登録は無かった．以上からSBM単体ではフィッシングサイトの検知には利用できないものの，判断材料の一つとして利用できることがわかった．

キーワード フィッシング詐欺，ネットワークセキュリティ，ウェブ，ソーシャルブックマーク

Phishing detection using social bookmark

Shinta Nakayama
The Department of Human Communication, The Graduate School of Electro-Communications,
The University of Electro-Communications
1-5-1,Chofugaoka, chofu-shi 182-8585,Japan
EMail: shinta.nakayama@gmail.com

Abstract recent day, social bookmark(SBM) became popular. SBM is a web services to store,search,and public their bookmarks. In this paper, we propose phishing detection method using social bookmark based on collective intelligence applied to security.

Keywords Phishing attack, Network security, web, Social bookmark

1. はじめに

子供や高齢者などコンピュータリテラシーの低い層のインターネット利用が一般化してきた．これに伴い，低リテラシー層をターゲットにしたフィッシング詐欺が急増している．フィッシング詐欺とは，金融機関や公的機関を装い個人情報を盗み取ることを目的としたウェブサイトを製作し，これによって得られたクレジットカード番号や預金口座の暗証番号，社会保障番号などを悪用し金銭を得る詐欺である．2006年度の全米被害額は28億ドル，2007年度は32億ドルと年々増加しており[1]，対策は急務である．

既存の対策手法は主にデータベースを利用したものであるが，ホワイトリスト方式は中小企業が登録漏れするという問題点があり，ブラックリスト方式はフィッシングサイトが現れてから登録されるまでに時間がかかるという問題点がある．それに対し中山らは集合知を利用することでデータベースを不要とするコンテンツベース方式を提案している[2][3]．コンテンツベース方式は，自然言語処理技術と，検索エンジンを組み合わせ，集合知を利用することでデータベースを不要にした．

そこで本稿では，集合知をセキュリティに利用するという考えを生かし，ソーシャルブックマークを利用することによりフィッシング検知を行うことを検討する．

2. ソーシャルブックマークとは

ソーシャルブックマーク(以下SBM)とはウェブサービスの一種で，ウェブ上に気になったサイトやよく使うサイトなどをコメント付きでブックマークするものである．国内では代表的なものに，はてなブックマーク[1]，Buzzurl[2]，Livedoorクリップ[3]，Yahoo!ブックマーク[4]などがある．海外ではdelicious[5]やdigg[6]などが有名である．インターネット白書2008の統計によると，国内のインターネット利用者の7.0%がSBMを利用している[4]．

1 http://b.hatena.ne.jp/
2 http://buzzurl.jp/
3 http://clip.livedoor.com/
4 http://bookmarks.yahoo.co.jp/
5 http://delicious.com/
6 http://digg.com/

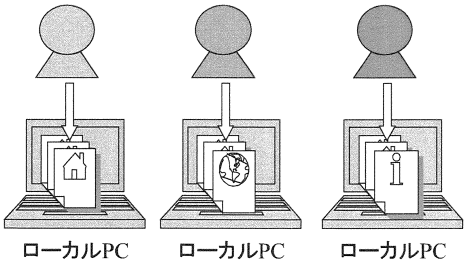

図 1:従来のブラウザのブックマークの構造

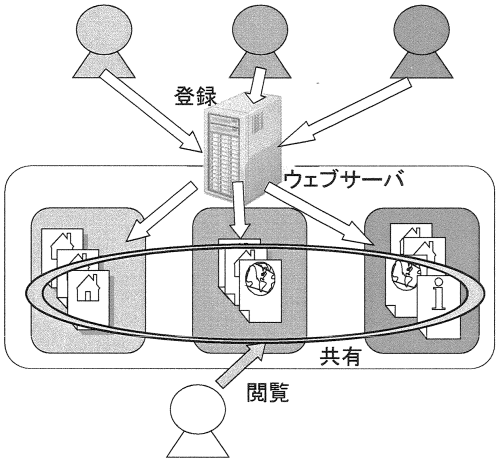

図 2:ソーシャルブックマークの構造

従来のブラウザのブックマーク[図 1]では，ブックマークされたウェブサイトの情報は，ローカル PC に保存され，それ以上外にでていくことはなかった．しかし SBM では，ブックマークをウェブサービス上に置くことによって，ブックマークの公開，共有ができるようになった[図 2]．そして，後日見返すため，ウェブページにコメントをつけるため，人に見せるため，自分はこういうものを見ている人であるとアピールするため，などに利用されている．また，ブックマークされたウェブページは共有されているため，どのウェブページが何人からブックマークされているかを知ることができる．そのため，SBM は一種のウェブサイトの人気を表す指標となっており，ユーザーもその投票のためにブックマークすることがある．

3. 先行研究

3.1 ホワイトリスト方式

正規サイトを記録したホワイトリストと比較し，載っていないウェブサイトを信頼できないと判断する方式である[5]．ホワイトリスト方式では，中小企業や新規サイトをすべて網羅することは難しく，ホワイトリストに載っていないサイト以外はフィッシングサイト扱いされるという可能性がある．また，ホワイトリストに登録するサイトが本当に正規サイトかどうか確認する必要があるため，管理組織には高い企業倫理が求められ，データベースの維持管理コストが高くなる．

3.2 ブラックリスト方式

フィッシングサイトを記録したブラックリストと比較し，載っていたサイトを信頼できないと判断する方法である[6]．ブラックリストは，フィッシングサイトを見た人がブラックリストの管理組織に通報して，初めて登録される．そのため，フィッシングサイトが現れてから，実際にブラックリストに登録されるまでには時間差が存在する．したがって，ブラックリストに登録されるまでの間に，閲覧してしまったユーザーを守ることはできないという問題がある．

3.3 ネットワークの性質の基づいた検知方式

データベースを用いない手法としては，フィッシングサイトのネットワーク的特性を利用したものがある[7]．米国の APWG[7]の調査によると，フィッシングサイトの平均存続期間は 3.1 日と非常に短い[8]．そのため，ウェブ存続期間，ドメインの登録日時，DNS の逆引きが可能かどうか，Google の PageRank，ネットワークのホップ数等を調べることで，フィッシングサイトか否かの判定を行うことができる．しかし，個人サーバや新たにできたウェブサイトは，フィッシングサイトとネットワーク的特性が類似しており，フィッシングサイト扱いしてしまう可能性がある．

3.4 ユーザーの認知能力の分析

被験者にウェブサイトを見せてフィッシングサイトかどうか判定させる実験[9]によると，23%の被験者はウェブの内容しか見ておらず，アドレスバーや SSL の錠前のアイコンなどは見ていなかった．また，多くの被験者は SSL の警告メッセージの意味を理解しておらず，もっとも精巧にできたフィッシングサイトでは 9 割の人を騙すことに成功した．そのため，ユーザーの認知能力には限界があるため，技術的な対策が重要であることがわかる．

3.5 視覚的類似性に基づいた検知方式

フィッシングサイトと正規サイトが視覚的に類似しているという仮説のもとに，視覚的類似性を判定するこ

7 Anti Phishing Working Group (http://www.antiphishing.org/)

とでフィッシングサイト検知をする[10]. しかしHTMLのタグ情報を解析して，デザイン情報の類似性を判断しているので，タグ情報を書き換えることで，容易に検知を逃れることができる. また，疑わしいサイトと，正規サイトとの両方が与えられていることを前提とした判定方法であるため，正規サイトの特徴情報を保存したデータベースが必要になる.

3.6 既存の検知ツールの検知率調査

既存のフィッシングサイト検知ツールの評価実験を行った研究がある[11]. フィッシングサイトの検知率について，SpoofGuardが97%と非常に高いの検知率を実現している．そしてIE7，Netcraft，Firefox w/Googleなどがそれに準じているが70～80%の検知率しか実現できていない.

正規サイトに対する検知実験では，SpoofGuardは92%の正規サイトをフィッシングサイトであると誤検知した. そのため，SpoofGuardは実用にはならないといえる. それに対し，IE7，Firefox w/Google，Netcraftは0%であった. しかしこれらもまたフィッシングサイト検知率は70～80%と低く，実用には十分ではない.

3.7 コンテンツベース方式

フィッシングサイトはユーザーを騙すために特定のウェブサイトになりすます. そのため，フィッシングサイトと正規サイトの内容は酷似している. フィッシングサイトの多くは正規サイトをコピー，もしくは模倣したものである. そこで，コンテンツの類似性を利用したコンテンツベース手法が中山らによって提案されている[2][3].

フィッシングサイトを熟知したユーザは，フィッシングサイトの疑いのあるウェブページを見たとき，そのウェブページの特徴を現す語句(企業名，製品名など)をキーワードにして，ウェブ検索を行うことがある. フィッシングサイトの存続期間は3.1日[8]と短く，また他のウェブサイトからリンクされることが稀であるため，検索エンジンからの評価が低い. そのため，適切なキーワードを選べば正規サイトのみが検索結果に現れ，フィッシングサイトは現れないというようになる. そこで，疑わしいウェブページのURLと，検索結果に現れたURLを比較することで，そのウェブページがフィッシングサイトかどうかを判断することができる.

以上の流れを計算機上で再現したのが，コンテンツベース方式である. 処理の流れを図3に示す. 検査対象ページから自然言語処理によってキーワードを抽出し，得られたキーワードをウェブ検索することによって正規サイトの候補を得る

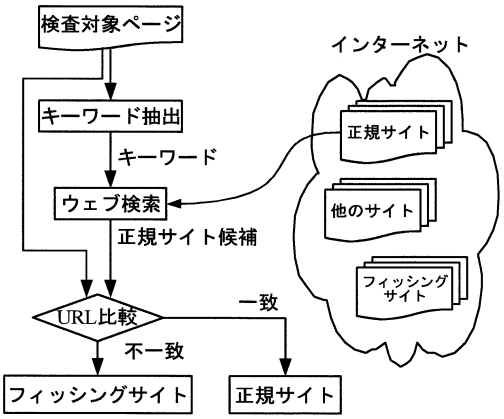

図3:コンテンツベース方式の処理の流れ

この方式は，検索エンジンをある種のホワイトリストとして利用するため，ブラックリストやホワイトリストといったデータベースを持つ必要がないという特徴がある. コンテンツベース方式の検知率は正規サイトの検知率が92.4%，フィッシングサイトの検知率が97.1%と，正規サイトの検知率は若干低いものの，3.6節の既存ツールよりはフィッシングサイトの検知率は十分高い.

4. SBMとフィッシングサイトの関係の仮説

コンテンツベース方式は検索エンジンという集合知を利用している. これにより，検索エンジンをホワイトリストとして利用することができ，データベースが不要で，メンテナンスフリーなセキュリティを実現することができている. そこで，同じ集合知であるSBMを利用したフィッシングサイト検知を検討する.

SBMの利用者は自分のブックマーク情報を公開することを厭わない，高リテラシーのユーザーであると考えられる. そのため，フィッシング詐欺の被害にあうことが少なく，またフィッシングサイトをブックマークすることはないと考えられる.

そこで「正規サイトはSBMによくブックマークにしているが，フィッシングサイトはSBMにあまりブックマークされていない. そのため，SBMはホワイトリストとして利用できる」という仮説を立てた. そして，これ検証するために，正規サイトとフィッシングサイトのそれぞれについてSBMに登録されているかどうかの実験を行った.

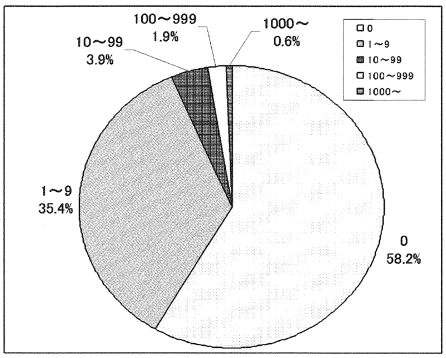

図 4:正規サイトのブックマーク件数と割合

5. 実験

仮説を検証するために評価実験を行った．SBM には delicious を利用した．delicious API を利用し，あるURL が何件ブックマークされているかを問い合わせるプログラムを作成し，評価実験を行った．.

5.1 正規サイトの実験

日本国内のフィッシングサイトは年間 100 件程度しか表れないが，英語圏では一ヶ月で 3 万件のフィッシングサイトが発生している[8]．そのため，北米の銀行のウェブサイトを評価対象にすべきであると考え，ウェブサイトのアクセスランキングを公開している Alexa[8]の北米の銀行ディレクトリに登録されている 1309 件の銀行のウェブサイトを正規サイトとして利用した．

図 4 はその実験結果である．58.2%のページは誰からもブックマークされていなかった．また 35.4%のページは 10 件未満のブックマーク件数であり，それ以上のブックマークのあるページは 6.4%あった．

5.2 フィッシングサイトの実験

フィッシングサイトを収集，公開している Phishtank[9]から，最新のフィッシングサイト 200 件を利用した．その結果，200 件中すべてがブックマーク件数が 0 であった．

6. 考察

6.1 仮説の検証

「正規サイトは SBM によくブックマークにしているが，フィッシングサイトは SBM にあまりブックマークされていない．そのため，SBM はホワイトリストとして利用できる」という仮説を立てたが，正規サイトは 4 割しか SBM にブックマークされていなかった．しかし，フィッシングサイトは一件も登録されていなかった．以上から，仮説は部分的には正しいことがわかった．これは 3.6 節の既存ツールの検知率よりも低いものであるが，他の方式と複合させることで，判断のための補助材料として利用することができると考えられる．

6.2 銀行のブックマーク率が低い原因

銀行のウェブサイトが SBM にあまり登録されていなかったのは，地方銀行のウェブサイトがほとんど更新されていないのが原因であると考えられる．SBM にブックマークする際の基準として，おもしろいニュースなどへの投票，読んだという備忘録などというものがある．数年来変化していない地方銀行のウェブサイトは SBM にあまり登録されず，また銀行のようなよく使うウェブサイトは，SBM ではなくブラウザのブックマークに登録されると考えられる．

6.3 フィッシングサイトを SBM に登録可能

今回の評価実験ではフィッシングサイトは 1 件も SBM に登録されていなかった．しかし SBM サービスのいくつかを調査してみると，ユーザーが興味本位で登録したものなのか，フィッシングサイト制作社が意図的に登録したものかは不明だが，フィッシングサイトが登録されているものがあった．仮説では高リテラシーのユーザーが SBM を利用していると考えていたが，実際には SBM は誰でも使うことができるウェブサービスである．そのため，フィッシングサイト制作者がフィッシングサイトも登録することが考えられ，仮説を逆用されるおそれがある．

7. 今後の展開

6.1 の結果から，複数の評価指標を重み付けして使うタイプのフィッシングサイト検知ツールには，SBM を用いた手法は現状でも利用可能である．しかし，6.3 節で述べたように，正規サイトが SBM に登録されているかどうかを，フィッシングサイト検知の判断基準の一つに利用するようになると，これを逆用してフィッシングサイト制作者が SBM にフィッシングサイトを積極的に登録するようになり，フィッシングサイトを正しく検知できなくなる可能性がある．

これにたいしては二つの解決案が考えられる．

一つは，フィッシングサイトが積極的に SBM に登録されるようになったことを利用して，3.7 節のコンテンツベース方式のようなデータベースを利用しないフィッ

8 http://www.alexa.com/
9 http://www.phishtank.com/

シングサイト検知アルゴリズムを，SBMに新規登録されたすべてのURLに適用することである．これにより，従来はユーザーの報告を待たなくてはブラックリストへの追加ができなかったものが，SBMにフィッシングサイトが登録されることにより，追加の自動化が行え，タイムラグの問題が解消する．

もう一つは集合知的な解として，フィッシングサイトがSBMにブックマークされているのを見つけたユーザーが，自分もフィッシングサイトをSBMにブックマークし，その際に「このウェブサイトはフィッシングサイトである」というタグを書くという方法である．これによりある閾値を越えて「このウェブサイトはフィッシングサイトである」というタグが付けられた場合，SBMからそのウェブサイトを外すことが考えられる．あらかじめブラックリストを生成しておけばSBMに誤って登録されるフィッシングサイトは少なく，ブラックリストで検知漏れしたものを人手で修正することにより，より精度の高いフィッシングサイト検知が実現できる．

以上からSBMをフィッシングサイト検知に利用した場合，次のような流れが生まれると考えられる．

1. 正規サイトが積極的に自身のウェブサイトをSBMに登録するように呼び掛けるようになる．
2. SBMに多くの正規サイトが登録されるようになり，SBMがホワイトリストの代替になる
3. SBMにフィッシングサイトが登録されるようになる．
4. SBMに登録されたURLを優先的に解析することで，ブラックリストを効率的に生成できるようになる．
5. ブラックリストから漏れたフィッシングサイトを，人手でタグ付けして，フィッシングサイトであると判断する．
6. SBMとブラックリストを組み合わせたフィッシングサイト検知方式ができる．

8. 結論

集合知セキュリティの考えから，ソーシャルブックマーク(SBM)を用いたフィッシングサイト検知手法の提案を行った．「正規サイトはSBMによくブックマークにしているが，フィッシングサイトはSBMにあまりブックマークされていない．そのため，SBMはホワイトリストとして利用できる」という仮説に基づき，あるURLがSBMに乗っているかどうかを調べる実験を行った．正規サイトは約4割がSBMに登録されていた．また，フィッシングサイトで登録されているものはなかった．そのため，SBMに登録されているか否かは弱いながらも正規サイトかどうかの判断材料に利用できることがわかった．

SBMをホワイトリストとして利用すると，フィッシングサイトをSBMに登録するという攻撃が行われ，SBMを用いたセキュリティが逆用されることが懸念される．逆用された場合，フィッシングサイトを正規サイトであると誤検知してしまう．

しかし，フィッシングサイトがSBMに積極的に登録されることで，SBMに登録されたURLをコンテンツベース方式などの，データベースによらない既存検知手法によって判定することにより，ブラックリストを効率的に生成できるようになる．また，ブラックリストから登録漏れが発生しても，SBMの性質上ユーザーがタグ付けをすることで，フィッシングサイトを排除することができる．

以上から，集合知セキュリティという考え方から，SBMがセキュリティに応用できる可能性を見出し，有効性を確認した．

9. 今後の展望

これまでSBMを利用した学術研究は，ユーザーがウェブページにつけたコメントやタグを元にした分類(＝フォークソノミー)及び，その分類に基づいた推薦がメインであり，セキュリティなどへの応用はあまり考えられてこなかった．ましてやSBMのように情報を外に出すということは，個人情報漏洩のリスク以外の何ものでもないという考え方であった．しかし，SBMを利用したセキュリティの実現可能性が見えてきたことから，皆が少しずつ情報を出せば，高度なセキュリティが実現できるという集合知セキュリティの考え方にシフトし，これまで情報を発信してこなかった人を含めて，より多くの情報がウェブに上がるようになるのではないだろうか．

参考文献

[1] "Gartner Survey Shows Phishing Attacks Escalated in 2007; More than $3 Billion Lost to These Attacks" http://www.gartner.com/it/page.jsp?id=565125 (2009年2月確認)

[2] 中山心太，吉浦裕，"模倣コンテンツの特性に基づくフィッシング検知方式"，2007-CSEC-38, Vol.2007, No71,pp387-392,2007.

[3] 中山心太,吉浦裕, "模倣コンテンツの特性に基づくフィッシング検知方式の誤検知防止",CSS2008,IPSJ Symposium Series Vol.2008, No.8, pp169-174,2008

[4] "インターネット白書2008",pp190 ,インプレス,2008

[5] 柴田賢介,荒金陽助,塩野入理,金井敦,"Webサイトからの企業名抽出によるフィッシング対策手法の提案",IPSJ SIG Notes Vol.2006, No.96 pp.17-22

[6] "RBL.JP",http://www.rbl.jp/ (2009年2月確認)

[7] 中村元彦, 寺田真敏, 千葉雄司, 土居範久,"proxyを利用したHTTPリクエスト解析によるAntiPhishingシステムの提案"2006-CSEC-32,Vol.2006 No.26,pp.13-18

[8] "Phishing Activity Trends Report for the Month of January, 2008" , http://www.antiphishing.org/reports/apwg_report_jan_2008.pdf (2009年2月確認)

[9] Rachna Dhamija,J. D. Tygar,Marti Hearst , "Why Phishing Works." CHI2006

[10]Wenyin Liu, Xiaotie Deng, Guanglin Huang, Anthony Y. Fu: An Antiphishing Strategy Based on Visual Similarity Assessment. IEEE Internet Computing 10(2): 58-65 (2006)

[11]Y. Zhang, S. Egelman, L. Cranor, and J. Hong Phinding Phish: Evaluating Anti-Phishing Tools. In Proceedings of the 14th Annual Network & Distributed System Security Symposium (NDSS 2007), San Diego, CA, 28th February - 2nd March, 2007.