推薦論文

プロキシを利用したHTTPリクエスト解析によるフィッシングサイト検出システムの提案

中村元彦[†] 寺田真敏[††,†††]
千葉雄司[††,†††] 土居範久[†]

今日，インターネットを利用したオンラインサービスの個人情報を詐取するフィッシングの被害が深刻化している．ブラックリストに基づきフィッシングサイトであるか否かを判定している既存の対策方式は，ブラックリストに登録されていないフィッシングサイトを検出できない．そこで本論文では，プロキシサーバを利用してHTTPリクエストの内容を解析し，フィッシングサイトの存続期間，ドメイン名，個人情報の詐取といった特徴に注目し，これらを総合評価することで，フィッシングサイトの検出を行う方式を提案する．さらに，プロトタイプシステムを使い，実際にフィッシングサイトへアクセスを行った評価結果から提案方式の有効性を示す．

Proposal of an Anti-phishing System by the HTTP Request Analysis that Used Proxy

Motohiko Nakamura,[†] Masato Terada,[††,†††] Yuji Chiba[††,†††] and Norihisa Doi[†]

Phishing is a type of deception designed to steal your personal data and damage by it is reported to have spread over these years. Some preventive measure has been proposed but their effect is not satisfactory. Because most of them cannot detect Phishing sites they do not know as they find Phishing sites based on blacklist. To solve this problem, we propose a method to detect unknown phishing sites by watching HTTP request using proxy to detect the characteristics of the Phishing sites (short continuation period, and so on) to warn the HTTP client how suspicious the target site is. We evaluated effectiveness of the proposed method by using the prototype system we have implemented.

1. はじめに

今日，インターネットの普及によりバンキングなどのオンラインサービスの利用者も増加している．その反面，米国を中心とした欧米諸国では，オンラインサービスの登録情報や個人情報の詐取を目的としたフィッシング（Phishing）という詐欺行為が社会問題となっている．フィッシングは，悪意を持った人（Phisher）が金融機関などの正規者を装い，偽Webサイトへの誘導と個人情報の入力を促す電子メールをインターネット利用者に送付する．利用者は前述の偽の電子メールを正規の電子メールであると思い込み，偽Webサイトにアクセスし個人情報などを入力してしまうことで金銭的な被害につながっている．日本では米国ほど被害は深刻ではないが，被害拡大の前に早急な対策が求められている．

既存の対策方式の1つとして，ブラックリストに基づきフィッシングサイトか否かを判定し，その結果を提示することで，そのアクセスを防ぐ方式がある．しかし，ブラックリストに基づき判定する方式の場合には，ブラックリストに登録されていないフィッシングサイトを検出できず，その効果は必ずしも十分であるとはいえない．

† 中央大学大学院理工学研究科
Graduate School of Science Engineering, Chuo University
†† 中央大学研究開発機構
Research and Development Initiative Chuo University
††† 株式会社日立製作所システム開発研究所
Systems Development Laboratory, Hitachi Ltd.

本論文の内容は2006年3月のコンピュータセキュリティ研究会にて報告され，CSEC研究会元主査により情報処理学会論文誌への掲載が推薦された論文である．

そこで，本論文では，フィッシングサイトの存続期間，ドメイン名，個人情報の詐取といった特徴に注目し，これらを総合評価することで，フィッシングサイトの検出を行う方式を提案する．

本論文の構成について述べる．2 章でフィッシング対策に関する既存方式の課題と解決策について示す．3 章でプロトタイプシステムについて述べ，4 章で評価実験を通して提案方式の有効性を示す．5 章でまとめと今後の課題を示す．

2. フィッシング対策にともなう課題と解決策の提案

本章では，フィッシング対策に関する既存方式の課題と解決策について述べる．

2.1 既存方式の課題

フィッシングサイトを検出するための既存方式の 1 つとして，ブラックリストに基づきフィッシングサイトであるか否かを判定する方式がある．この方式は，フィッシングサイトに接続しようとするときや接続した後に，接続サイトがブラックリストに該当するサイトか否かを判定する．しかし，ブラックリストに基づき判定する方式の場合，ブラックリストに登録されていないフィッシングサイトを検出できない．また，ブラックリストを頻繁に更新したとしても，更新されるまでの間に被害に遭う可能性がある．

2.2 課題解決のアプローチ

各国でフィッシング対策のための業界団体が結成されており，その中で最も有名な団体に，米国の Anti Phishing Working Group（APWG）[1] がある．APWG では，定期的に Phishing Attack Trends Report というフィッシングに関するレポートを公表している．本論文の提案内容に関係の深いデータを**表 1** に示す．この報告によれば，2006 年 7 月に報告されたフィッシングサイト数は 14,191 サイトとなっており，数多くのフィッシングサイトが出現していることが分かる．また，フィッシングサイトの特徴として，サイトの平均存続日数が 4.8 日で，継続的に稼動している正規の Web サイトにはない特徴がある．

そこで，本論文では，ブラックリストに依存せずフィッシングサイトを検出する方式として，次のような方式を提案する．

- フィッシングサイトの存続日数，ドメイン名ではなく IP アドレスで運用されるといった通常の Web サイトにはみられない，フィッシングサイト固有の特徴に着目する．具体的には，**表 2** に示す Web サイトの存続期間，ドメイン名，個人情報の詐取といった特徴を調査する．
- フィッシングサイト固有の特徴を調査する際には，インターネット上に活動痕跡として記録されている情報を利用する．たとえば，Web サイトの存続期間については，Web サイトの連続稼動日数を掲載している Netcraft サイト[3] や，過去に公開された HTML ファイルをアーカイブとして保存している WayBack Machine サイト[4] の掲載情報を参照する．これは，正規の Web サイトであるならばインターネット上に何らかの日々の活動痕跡が数多く残っているという経験則に基づく．
- フィッシングサイト固有の特徴を総合評価するにあたっては，式 (1) を用いて Web サイトの危険度を算出する．ここで，d は表 2 に示す各調査項目のポイントであり 0〜100 の数値をとる．w は重み付けであり 0〜3 の数値をとり，n は調査項目数，i は調査項目番号である．

表 1 フィッシングサイトの特徴[2]
Table 1 The characteristics of phishing sites.

項目	結果
フィッシングサイト数(2006 年 7 月分)	14,191
ホスティング元の上位 10 カ国が占める割合	76.9%
ドメイン名ではなく IP アドレスで運用される割合	42%
80 番以外のポート番号を使用する割合	8.9%
フィッシングサイトの平均存続日数	4.8 日
フィッシングサイトの最長存続日数	31 日

表 2 フィッシングサイト判定のための調査項目
Table 2 The research items for detection of phishing sites.

分類	調査項目	重み付け
サイトの存続期間が短いという特徴	1. Netcraft サイトを利用した Web サイトの稼働日数取得	3
	2. Netcraft サイトを利用した Web サイトのランク取得	1
	3. WayBack Machine サイトを利用したドメイン登録有無と過去の Web ページとの差異量の取得	2
ドメイン名の特徴	4. Web サイトへのアクセス回数の確認	2
	5. TLD,SLD に基づくドメイン名形式の確認	3
	6. DNS サーバを用いた正引き，逆引きアドレスの取得	1
個人情報の詐取を目的とする特徴	7. 個人情報の入力を促す文字の有無	2

$$危険度 = \left(\sum_n w_i \times d_i \right) \div \sum_n w_i \quad (1)$$

- Web サイトの危険度に閾値を設け，危険度が閾値を超えた場合には，フィッシングサイトの疑いがあると判定する．

本提案方式は，フィッシングサイト固有の特徴に着目すること，インターネット上に記録された活動痕跡に基づき特徴を調査すること，複数の特徴を総合評価することで，フィッシングサイトの検出を行う．これにより，ブラックリストのような登録手順を必要とすることなくフィッシングサイトの検出を可能としている．なお，表 2 に記載している調査項目と重み付けならびに，Web サイトの危険度の閾値設定の詳細については，3 章の実装方式において述べる．

3. フィッシングサイト検出システム

本章では，2 章で提案したフィッシングサイト検出方式を評価するために実装したプロトタイプシステムについて述べる．

3.1 実装方式の概要

フィッシングサイト検出システムでは，クライアント PC と Web サーバの間に設置したプロキシサーバ上で稼動する HTTP 中継機能，フィッシング検出機能，レポート機能の 3 つのコンポーネントとして提案方式を実装した．本システムを用いたフィッシングサイト検出手順は次のとおりである（**図 1**）．

① クライアント PC は Web サーバに対し，Web ページの要求（HTTP リクエスト）を送信する．
② プロキシサーバの HTTP 中継機能は，HTTP リクエストから URL 情報およびドメイン情報を抽出し，フィッシング検出機能に引き渡す．
③ プロキシサーバの HTTP 中継機能は，HTTP リクエストを Web サーバに転送する．
④ プロキシサーバのフィッシング検出機能は，Web サーバからの応答（HTTP レスポンス）に，個人情報の詐取を目的としている特徴を有しているかを調査した後，項番 ② の調査結果と合わせ，式 (1) を用いて総合的に Web サイトの危険度を算出する．その後，レポート機能は調査結果の詳細情報を記載した HTML ファイルを作成する．
⑤ Web サイトの危険度が閾値を超え，フィッシングサイトの疑いがあると判定されたページに関しては，ユーザに Web ページの表示有無を確認する．
⑥ HTTP レスポンスに Web サイトの危険度を警告データとして挿入した後，クライアント PC に送信する．なお，項番 ⑤ でページ表示しないを選択した場合には，プロキシサーバで Web サーバから受信した HTTP レスポンス部を破棄する．

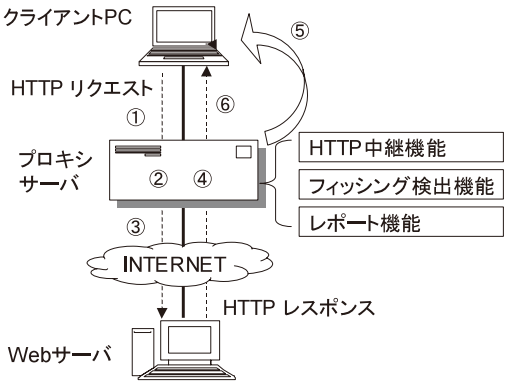

図 1 フィッシングサイト検出システムの概要
Fig. 1 The overview of Anti-Phishing System.

```
GET http://xxx.xxx.co.jp/index.htm HTTP/1.0
    //省略//
Host: xxx.xxx.co.jp
```

図 2 標準的な HTTP リクエストの抜粋
Fig. 2 The part of general HTTP request.

フィッシングサイトの検出を，プロキシサーバ上に実装することの利点は，プロキシ指定のできるブラウザであればどのブラウザでも利用可能な点にある．既存のアドオンツールバーによる実装方式の多くは Internet Explorer 専用で，適用範囲も限られてしまっている．本システムの場合には，Internet Explorer 以外のブラウザでも利用できる点で汎用性に優れている．ただし，プロキシサーバ上で通信内容を解析するため，暗号化されている SSL 通信などには対応することができない．なお，SSL 通信を対象としたフィッシングサイトの検出は，今後の課題である．

次節以降，実装した 3 つのコンポーネントについて述べる．

3.2 HTTP 中継機能

HTTP 中継機能は，クライアント PC と Web サーバ間の HTTP 通信の中継，HTTP リクエストからの情報抽出とユーザへの警告提示を行う．

HTTP リクエストから抽出する情報は，**図 2** に示す綱掛け部の URL とドメイン情報である．HTTP リクエストは，Web ページの骨格部分の HTML ファイル要求とそれに付随する画像ファイルなどの要求の 2 種類に分けることができる．両者ともに同じ解析処理を行うが，画像ファイルなどの要求に関しては HTTP レスポンスに警告データの挿入処理を行わない．

```
HTTP/1.0 200 OK
    //省略//
<html>
    <body>
        ⇒警告データの挿入位置
        //html ソース(本文)//
    </body>
</html>
```

図 3 標準的な HTTP レスポンスの抜粋
Fig. 3 The part of general HTTP response.

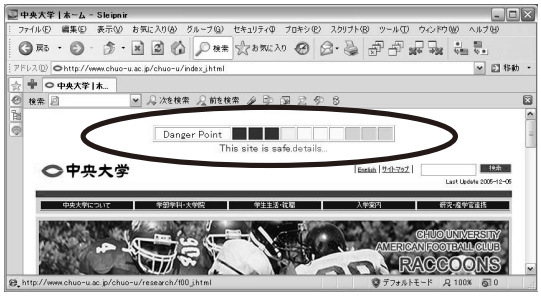

図 4 警告バーの表示例
Fig. 4 The example of alert bar.

ユーザへの警告提示は，フィッシング検出機能で算出した Web サイトの危険度をユーザに提示する．具体的には，フィッシング検出機能で算出した Web サイトの危険度を**図 3** に示す綱掛け部の情報が両方含まれている HTTP レスポンスの “⇒” 部に警告データとして挿入する．“200 OK” は正常な HTTP レスポンスであることを示し，“<body>” は HTML の本文の開始位置を示すタグである．2 つの条件を満たす HTTP レスポンスに警告データを挿入することで，Web ページの構成を大きく損ねることなく，Web ページの冒頭部分に警告を表示する．さらに，ユーザが警告に気づきやすくするために，プロトタイプシステムでは，Web サイトの危険度を示すグラフィカルな警告表示（以後，警告バーと呼ぶ）を採用した．

プロトタイプシステムを用いて中央大学のホームページへアクセスした表示例を**図 4** に示す．図の中央付近に警告バーが挿入され，Web ページの危険度は 23 ポイントであるため，目盛りが 3 マス色つきとなっている．なお，目盛りは危険度が高くなるにつれ，黄色から赤へと変わる．

3.3 フィッシング検出機能

フィッシング検出機能は，HTTP 中継機能で抽出された情報に基づき，アクセスするサイトが表 2 に示す特徴を有しているかを調査する．次に，式 (1) を用いて Web サイトの危険度を算出した後，危険度の閾値判定を行う．

3.3.1 調 査 項 目

フィッシングサイト固有の特徴を判定するための調査項目の詳細について述べる．なお，調査結果のポイント化については 4.2 節で述べる．

(1) サイトの存続期間が短いという特徴の調査

フィッシングサイトの存続期間が短いという特徴の調査では，Web サイトの連続稼動日数を掲載している Netcraft サイトと過去に公開された HTML ファイルをアーカイブとして保存している WayBack Machine サイトの掲載情報を利用した．

(a) Web サイトの稼働日数とランク

Netcraft サイトでは，ドメイン名に対する各種情報を公開しており，そのうちの 1 つに Web サイトの連続稼働日数がある．本システムでは，Netcraft サイトが提供する連続稼動日を Web サイトの存続期間と見なした．また，継続的に稼動している正規サイトの場合には Netcraft サイトで提供されているドメインのランク（サイトの人気度）も高くなると考え，Web サイトの存続期間の補足情報として調査項目に加えた．

(b) ドメイン登録有無と過去の Web ページとの差異

WayBack Machine サイトでは，過去に公開された HTML ファイルを保存し，公開している．フィッシングサイトの存続期間はたかだか 31 日程度であるため，WayBack Machine サイトに HTML ファイルが保存されていない可能性が高い．本システムでは，WayBack Machine サイトにアクセスしたサイトのドメインが登録されているか否か，現在と過去の HTML ソースのファイルサイズの差異（ビット数の差異）とを組み合わせて調査項目とした．

(2) ドメイン名の特徴の調査

ドメイン名の特徴の調査にあたっては，Web サイトへのアクセス回数，ドメイン名の形式，ドメイン名の逆引きを利用した．

(a) Web サイトへのアクセス回数

Web サイトへのアクセス回数では，ドメイン名の利用頻度に着目し，初めて（$d = 100$），2 回〜10 回（$d = 50$），11 回以上（$d = 0$）の 3 段階の利用頻度に分ける．

(b) ドメイン名の形式

ドメイン名の形式では，フィッシングサイトがドメイン名ではなく IP アドレスで運用されるといった特徴に着目し，ドメイン名が IP アドレスのままでないか，ポート番号が 80 番以外でないかを確認する．次に，Top Level Domain（TLD）と Second Level Domain（SLD）を参照し，フィッシングサイトをホスティングしている可能性が高い国や団体ではないかを，あら

password, Login, user, Credit, ログイン, パスワード, 口座番号, カード番号, お客さま番号, 暗証番号, 住所, 氏名, 電話番号, クレジット, 会員番号

図 5 フィッシング検出に使用する単語
Fig. 5 The check words for phishing detection.

かじめ定義している情報と比較を行う．なお，TLD が jp ドメインの場合には，日本レジストリサービスの Whois サービス[5] にドメインの登録情報を問い合わせ，その結果をレポート機能に引き渡す．

(c) ドメイン名の逆引き

フィッシングサイトは簡易的に作成されるため，ドメイン名の逆引きデータが存在しないことが多い．そこで，取得したドメイン名を正引きし，さらに正引きした IP アドレスの逆引きデータの有無を確認する．

(3) 個人情報の詐取を目的とする特徴

フィッシングサイトは，個人情報の詐取を目的としているために，Web ページ内に必ず個人情報の入力を促す文字を含んでいる．そこで，HTTP レスポンスに含まれる HTML ソース内を検索し，あらかじめ定義しておいた単語，たとえば，“パスワード” という単語の抽出を行う．今回の実装では，**図 5** に示す実際のフィッシングサイトで多く使われる単語を 15 語程度抜粋し利用した．

3.3.2 実装上の考慮点

本項では，フィッシング検出機能の実装に関する考慮点について述べる．

(1) 調査結果のキャッシング

ページ単位ではなくドメイン単位に相当する調査項目であり，かつプロトタイプシステム外への問合せが発生する調査項目である「1. Netcraft サイトを利用した Web サイトの稼働日数取得」,「2. Netcraft サイトを利用した Web サイトのランク取得」,「4. Web サイトへのアクセス回数の確認」,「6. DNS サーバを用いた正引き，逆引きアドレスの取得」に関しては，サイトを参照するという操作中に調査結果が変動することは少ないと考えることができる．そこで，1 回目のアクセスにともなう調査結果をキャッシュすることで，2 回目以降のアクセス時の処理効率を上げている．その他の項目に関しては，アクセスごとに毎回判定処理を行った．なお，キャッシュの有効期間を考慮した実装については，今後の課題である．

(2) Web サイトの危険度算出における重み付け

Web サイトの危険度算出にあたり，各調査項目に対して重み付けを行っている．実装における重み付けの指針は，表 1 に示すフィッシングサイトの特徴として報告されている該当項目「1. Netcraft サイトを利用した Web サイトの稼働日数取得」,「5. TLD，SLD に基づくドメイン名形式の確認」の重み付けを大きくすることで，実際のフィッシングサイトの傾向を危険度に反映した．

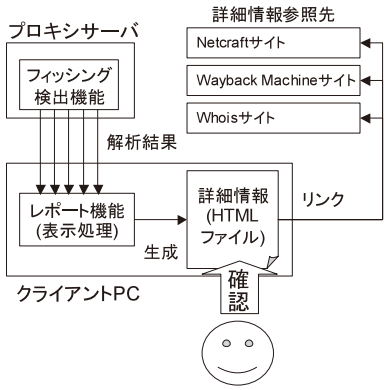

図 6 レポート機能の概要
Fig. 6 The overview of reporting function.

(3) Web サイトの危険度の閾値設定

危険度の閾値については，試行を繰り返した経験則に基づき設定した．今回の実装では，Web サイトの危険度が 65 ポイントを超えるとフィッシングサイトと見なし，40 ポイントを超えると疑わしいと見なす仕様とした．

3.4 レポート機能

レポート機能は，Web サイトのドメインに関する情報や Web サーバの稼働状況などの詳細情報を Netcraft サイト，WayBack Machine サイトならびに，Whois サイトなどの参照先とともにユーザに提示する機能である．この詳細情報は，危険度の高い Web ページを閲覧する場合，ユーザが事前に安全性の確認を行う際の参考情報となる（**図 6**）．なお，今回の実装ではクライアント PC 上の Web サーバを用いて詳細情報の表示処理を実現したが，プロキシサーバや独立した Web サーバ上で表示処理を実現することにより，より実用的なシステムとして構築することができると考える．

4. 評　　価

本章では，提案方式の有効性を評価するために実施したプロトタイプシステムを用いた評価と，表 2 の調査項目ごとの調査結果について述べる．

4.1 プロトタイプシステムを用いた評価

プロトタイプシステムを用いた評価では，開発したプロトタイプシステムを使い，実際にフィッシングサイトへのアクセスを試みる．

表 3 システムの構成
Table 3 The system specification.

項目	仕様☆
CPU	Pentium4 2.60 [GHz]
メモリ	1.00 [GB]
ネットワーク	100 BASE-TX
OS	Windows XP SP2
ブラウザ	Sleipnir version 2.00

4.1.1 評価条件

(1) 評価項目

評価項目は，機能面と性能面における，次の 3 つの項目とする．なお，項番 (a) については，プロトタイプシステムならびに，既存のアドオンツールバー EarthLink toolbar[6]，SpoofGuard[7] について評価を実施し比較する．

(a) フィッシングサイトの検出率と，正規サイトの誤検出率
(b) 警告バーの挿入成否の割合
(c) フィッシングサイトを検出するための処理時間

(2) 対象サイト

評価に利用するフィッシングサイトは，RBL.JP[8] サイトが公開している「フィッシングサイト詐欺情報」を参考にし，2005 年 11 月 26 日から 12 月 21 日までの期間（調査期間 1）で，計 25 の実在のフィッシングサイトを対象とした．また，誤検出率を測定するため正規サイトとして，「世帯内パソコンにおけるインターネット利用状況調査」[9] におけるランキング上位の 20 カ所の Web サイトと国内大手金融機関の 18 カ所の Web サイト，計 38 サイトを対象とした．

(3) プロトタイプシステムの仕様

本実験に使用したプロトタイプシステムは，スクリプト言語 perl を主体に実装している．HTTP 中継機能は http-proxy.pl[10] を機能拡張することで実現し，フィッシング検出機能とレポート機能の内部処理系には perl を，表示系には Java を使用している．また，簡略化のためプロキシサーバ上で稼動する HTTP 中継機能とフィッシング検出機能をクライアント PC 内で稼動させている．プロトタイプシステムを稼動させるクライアント PC の仕様は**表 3** のとおりであり，このクライアント PC を研究室内のネットワークに接続し実験を行った．

(4) 調査項目に関する前提条件

調査項目「4. Web サイトへのアクセス回数の確認」については，すべて初めてのアクセス（$d = 100$）であるという条件の下で評価する．

☆ 商品名称などに関する表示
本論文に記載されている会社名，製品名は，各社の登録商標または商標です．

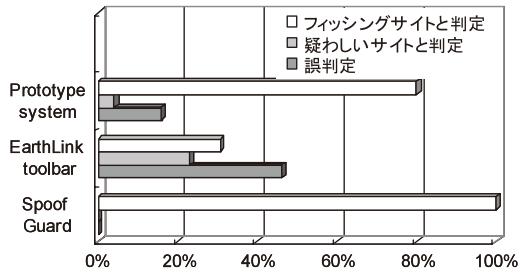

図 7 フィッシングサイトの検出率
Fig. 7 The detection ratio for phishing sites.

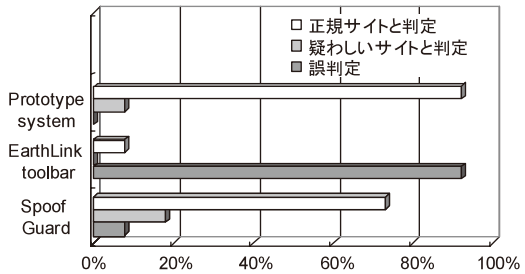

図 8 正規サイトの誤検出率
Fig. 8 The false positive ratio for normal sites.

4.1.2 実験結果

(1) 検出率と誤検出率

プロトタイプシステムと既存の各ツールバーを利用し，フィッシングサイトにアクセスした際の検出率を**図 7** に示す．また，正規サイトにアクセスした際の誤検出率を**図 8** に示す．

図 7 から，プロトタイプシステムでは約 80%のフィッシングサイトを検出でき，疑わしい Web サイトを含めると約 85%の検出率となった．また，図 8 からプロトタイプシステムの誤検出率は約 0%となり，疑わしいと判定された Web サイトを含めた場合でも 8%であった．既存のツールバーではブラックリストを用いている EarthLink toolbar はフィッシングサイト検出率の精度も低く，また，正規サイトの誤検出率も高い．一方，ブラックリストを用いない SpoofGuard に関しては，フィッシングサイトの検出率は 100%と高かったが，プロトタイプシステムと比べ正規サイトに対する誤検出率が高かった．

(2) 警告バーの挿入成否

警告バーの挿入成否の判定は，Web ページの上部中央に警告バーが表示されたか否かを目視確認した．評価対象となるすべての Web サイトのうち，87%の Web サイトのページ表示に対して警告バーを正常に挿入した．しかし，フレームを使用している一部の Web サイトでは，警告バーを挿入すべき HTML ファイル

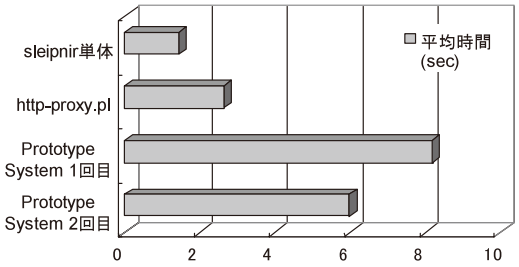

図 9 フィッシングサイトの検出にかかる処理時間
Fig. 9 The processing time of phishing site detection.

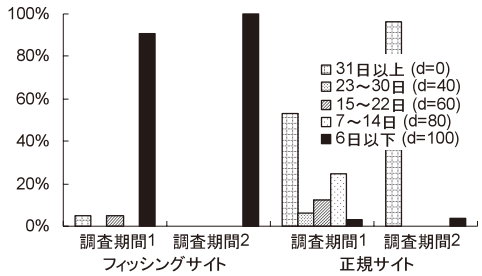

図 10 Web サイトの稼働日数を用いた判定
Fig. 10 The detection by uptime of Web site.

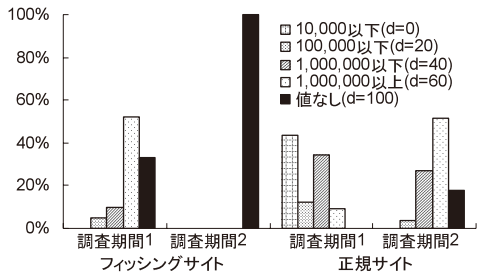

図 11 Web サイトのランクによる判定
Fig. 11 The detection by popularity of Web site.

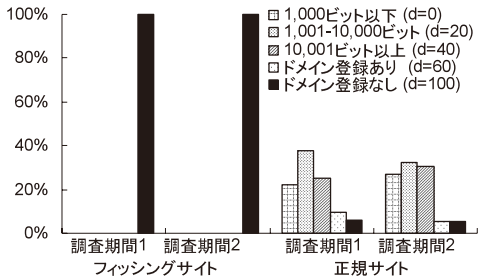

図 12 過去の Web ページとの差異検出による判定
Fig. 12 The detection by difference of Web pages.

を識別できず，正しい位置に表示されない場合や挿入処理が行われなかった．

(3) 処理性能

フィッシングサイトの検出にかかる処理時間を測定するにあたっては，前述した正規サイトにアクセスし，その応答時間を測定することとした．応答時間は Web サイトへのアクセス開始から Web ページが表示されるまでの時間であり，各 Web サイトの測定値の平均をとった．測定結果を**図 9** に示す．プロトタイプシステムを利用した際の応答時間は，ブラウザ単体に比べ約 5.7 倍，http-proxy.pl 経由でのアクセスに比べ約 3.1 倍の時間を要した．ただし，2 回目以降の同じサイトへのアクセスに対しては，1 回目の 8.2 秒に比べ，6.0 秒と約 27%短縮され，キャッシュの効果がある程度示された．

4.1.3 考　　察

フィッシングサイトの検出率については，「1. Web サイトの稼働日数」，「3. ドメイン登録有無と過去の Web ページとの差異量」，「7. 個人情報の入力を促す文字の有無」の調査項目が検出率向上に働き，Web サイトへのアクセス回数を除く他の調査項目が検出率を下げる要因となっていることが分かった．詳細については，4.2 節の調査項目の評価で述べる．

4.2 調査項目の評価

調査項目の評価では，フィッシングサイトと正規サイトについて，表 2 の調査項目ごとに詳細調査を実施した．

4.2.1 評 価 条 件

(1) 評価項目

評価項目は，表 2 の調査項目のうち，Web サイトへのアクセス回数を除く項目を調査対象とした．

(2) 対象サイト

評価に利用するフィッシングサイトについては，2005 年 11 月 26 日から 12 月 21 日まで（調査期間 1）と 2006 年 5 月 30 日から 7 月 3 日まで（調査期間 2）の期間で計 88 の実在のフィッシングサイトを対象とした．また，誤検出率を測定するための正規サイトは，「世帯内パソコンにおけるインターネット利用状況調査」におけるランキング上位の 20 カ所の Web サイトと国内大手金融機関の 68 カ所の Web サイトとし，4.1 節のサイトを含む計 88 サイトを対象とした．

4.2.2 調 査 結 果

フィッシングサイトにアクセスした際の検出率と正規サイトにアクセスした際の誤検出率を**図 10**，**図 11**，**図 12**，**図 13**，**図 14**，**図 15** に示す．凡例の括弧内は調査結果に付与しているポイント（d）である．

(1) Web サイトの存続期間が短い特徴

フィッシングサイトの存続期間が短い特徴については，図 10，図 11，図 12 に示すとおり，いずれの調査項目もフィッシングサイトの検出率が高い．しかし，Web サイトのランク（図 11）では，正規サイトであ

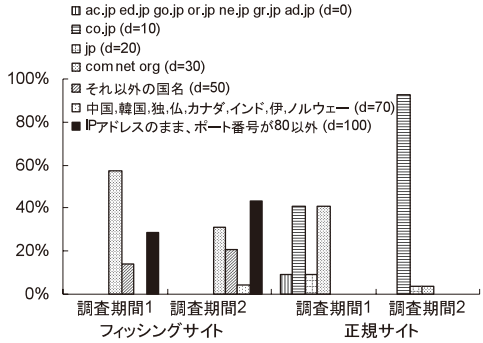

図 13 TLD，SLD ドメイン名による判定
Fig. 13 The detection by TLD and SLD.

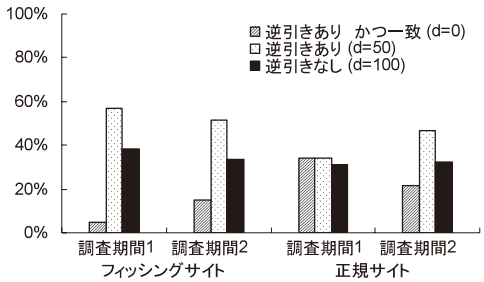

図 14 DNS サーバ正引きと逆引きによる判定
Fig. 14 The detection by reverse DNS entry.

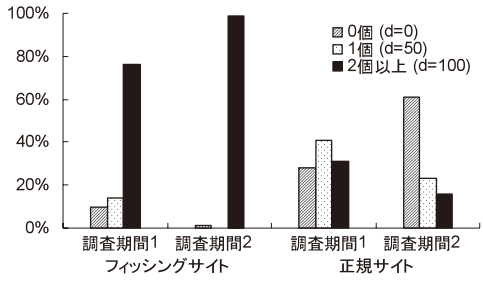

図 15 個人情報の入力を促す文字による判定
Fig. 15 The detection by the character which suggests the input of the personal information.

るにもかかわらず Web サイトとして登録されていない「調査結果 = 値なし」が 18%（調査期間 2）となった．さらに，過去の Web ページとの差異検出（図 12）では，正規サイトであるにもかかわらず Web サイトとして登録されていない「調査結果 = ドメインなし」が 6%（調査期間 1），Web サイトとして登録されていても判定に必要となるデータが掲載されていない「調査結果 = ドメインあり」が 9%（調査期間 1）にのぼっている．

(2) ドメイン名の特徴

DNS サーバを用いた正引きと逆引きについては，フィッシングサイトで「調査結果 = 逆引きあり（正引きした IP アドレスの逆引きデータが，元のドメイン名と一致しない割合）」が正規サイトに比べ高いものの，正規サイトであってもドメイン名と IP アドレスの整合性があまり確保されておらず，誤検出率を上げる要因となってしまっている（図 14）．

(3) 個人情報の詐取を目的とする特徴

個人情報の詐取を目的とする特徴については，フィッシングサイトの検出率が高く 70%を超えている（図 15）．しかし，正規サイトでも個人情報の入力を促す文字が含まれている割合は 50%近く，誤検出率を上げる要因となってしまっている．

4.3 考　　察

(1) Web サイトの存続期間が短い特徴

Web サイトの急激な増加によっては，Netcraft や WayBack Machine サイトに活動痕跡が残らない可能性もでてくる．また，CGI などの動的なページを提供する Web サイトの増加は現在と過去の HTML ソースファイルサイズの差異に影響を与える．このような利用環境の変化を提案方式で吸収するためには，複数の調査項目による精度向上が必要である．

(2) ドメイン名の特徴

jp ドメインを利用したフィッシングサイトがほとんど構築されていないという経験的な事実は，ドメイン名の特徴を用いた調査項目として活用できる（図 13）．

(3) 個人情報の詐取を目的とする特徴

Web サイトの存続期間が短い特徴と同様に，検出には複数の調査項目による精度向上が必要である．

(4) Web サイトの危険度算出における重み付け

重み付けを大きくした調査項目「1. Netcraft サイトを利用した Web サイトの稼働日数取得」，「5. TLD, SLD に基づくドメイン名形式の確認」については，図 10，図 13 に示すとおり，いずれもフィッシングサイトと正規サイトの特徴がでやすく，重み付けを大きくすることによる検出率向上を期待できることが分かった．

(5) Web サイトの危険度の閾値設定

2 回の評価実験の結果で得られた危険度を**表 4** に示す．項番ごとの左欄は実在するフィッシングサイトを対象とした危険度であり，右欄は正規サイトを対象とした危険度である．全体的に危険度は増加しており，当初，実装で想定していた 65 ポイントを超えるとフィッシングサイトと見なし，40 ポイントを超えると疑わしいと見なすという閾値では，正規サイトを疑わしいと見なしてしまう可能性が高くなってしまうことが分かった．なお，フィッシングサイトの検出率向上と誤検出率を低減させるための閾値の設定については，今

表 4 危険度の比較
Table 4 The comparison of detection alerts.

		調査期間 1		調査期間 2	
		フィッシングサイト	正規サイト	フィッシングサイト	正規サイト
危険度	平均	74	25	87	38
	最大値	95	55	100	62
	最小値	38	9	48	15
	標準偏差	12	11	9	9
区分	40 以下	15%	92%	0%	59%
	41〜65	5%	8%	1%	41%
	66 以上	80%	0%	99%	0%

後の課題である．

5. おわりに

本論文ではフィッシング対策として，フィッシングサイトの存続期間，ドメイン名，個人情報の詐取といった特徴に注目し，これらを総合評価することで，フィッシングサイトの検出を行う方式を提案した．

また，提案方式に基づき実装したプロトタイプシステムを用いて，実際にフィッシングサイトへのアクセスを試みた結果，フィッシングサイトの検出率は，疑わしい Web サイトを含めると全体の約 85%であり，誤検出率は 8%であった．さらに，各調査項目について詳細を調査した結果，フィッシングサイトの存続期間が短いという特徴については，いずれの調査項目もフィッシングサイトの検出率は高く，調査項目として有効であるという結果が得られた．しかし，調査項目の中には，正規サイトに対する誤検出が発生する項目もあるため，正規サイトの判定には複数の調査項目による精度向上が必要であることも分かった．

以上の評価を通して，本提案方式が既存のブラックリストを用いた方式では検出できない未登録状態のフィッシングサイトについても検出することができ，フィッシング対策に有効であることを示した．

今後の課題としては，フィッシングサイトの検出率向上と正規サイトの誤検出率低減を可能とする，新たな調査項目の検討を進めてゆきたいと考えている．

謝辞 本研究は文部科学省 21 世紀 COE プログラム「電子社会の信頼性向上と情報セキュリティ」および文部科学省科学技術振興調整費「新興分野人材養成 情報セキュリティ・情報保証人材育成拠点」の支援を受け実施している．本研究を進めるにあたって有益な助言と協力をいただいた関係者各位に深く感謝いたします．

参考文献

1) Anti-Phishing Working Group. http://www.antiphishing.org/
2) APWG: Phishing Activity Trends Report July, 2006. http://www.antiphishing.org/reports/apwg_report_july_2006.pdf
3) Netcraft. http://news.netcraft.com/
4) WayBack Machine, Internet Archive. http://www.archive.org/Web/Web.php
5) 日本レジストリサービス．http://jprs.jp/
6) EarthLink toolbar, EarthLink. http://www.earthlink.net/software/free/toolbar
7) Stanford Security Lab, Stanford university. http://crypto.stanford.edu/SpoofGuard
8) RBL.jp. http://www.rbl.jp/
9) ビデオリサーチ：世帯内パソコンにおけるインターネット利用状況調査．http://www.videoi.co.jp/data/wsr/wsr051012.html
10) HTTP proxy, 68user's page. http://x68000.q-e-d.net/~68user

(平成 18 年 11 月 20 日受付)
(平成 19 年 7 月 3 日採録)

推薦文

Phishing サイトの調査を行い，普遍性を含んだ効果的な 7 つの特徴に注目した検知の仕組みを提案し，実装した．検知率に関する既存ツールとの比較評価を行い，優位性を示した．proxy を用いることにより，クライアント PC に特別なツールを必要としない簡易な防御機構を提案した．また，防御機能が稼動しない条件や，処理速度の問題点など，本分野の課題も明らかにした．新規性，有効性，信頼性のすべての面で，高く評価できる．
（コンピュータセキュリティ研究会元主査 村山優子）

中村 元彦
2006 年中央大学大学院理工学研究科情報工学専攻修士課程修了．同年ソフトバンクテレコム（旧日本テレコム）（株）入社．

寺田　真敏（正会員）

1986 年千葉大学大学院工学研究科写真工学専攻修士課程修了．同年（株）日立製作所入社．博士（工学）．現在，システム開発研究所にてネットワークセキュリティの研究に従事．2004 年 4 月から JPCERT コーディネーションセンター専門委員，2004 年 4 月から中央大学研究開発機構客員研究員，2004 年 8 月から情報処理推進機構セキュリティセンター研究員を兼務．

千葉　雄司（正会員）

1972 年生．1997 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年（株）日立製作所入社．博士（工学）．現在，システム開発研究所にてコンパイラの研究開発に従事．2003 年から中央大学研究開発機構客員研究員を兼務．

土居　範久（フェロー）

中央大学理工学部教授，慶應義塾大学名誉教授．現在，日本学術会議副会長，文部科学省科学技術・学術審議会委員，総務省情報通信審議会委員，文部科学省「次世代 IT 基盤構築のための研究開発」プログラムディレクター，科学技術振興機構（JST）社会技術研究開発センター（RISTEX）「情報と社会」領域総括，科学技術振興機構（JST）戦略的創造研究推進事業研究総括，NPO 法人日本セキュリティ監査協会会長，国際計算機学会（ACM）日本支部長．専門はソフトウェアを中心とした計算機科学および情報セキュリティ．情報処理学会名誉会員・フェロー，日本ソフトウェア科学会フェロー．