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

2006－DPS－126（3）
2006－CSEC－32（3）
2006／3／16

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

中村　元彦†　　寺田　真敏††　　千葉　雄司††　　土居　範久†

† 中央大学大学院 理工学研究科 情報工学専攻　〒112-8551 東京都文京区春日 1-13-27
†† 中央大学研究開発機構
〒112-8551 東京都文京区春日 1-13-27 中央大学後楽園キャンパス 3 号館 12 階

あらまし　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 Phishing の被害が深刻化している．しかし，ブラウザのアドオンツールバーを利用する既存の対策手法は，ブラックリスト基づき Phishing サイトの判断をしているため，ブラックリストに無い Phishing サイトを検出できないという課題がある．そこで本稿では，proxy を利用して HTTP リクエストの内容を解析し，Web サイトの存続期間が短いなどといった Phishing サイトにみられる特徴的な傾向を捉えることにより，Phishing サイトを検出する手法を提案する．そして，プロトタイプシステムを使い，実際に Phishing サイトへアクセスを行なった評価結果から，提案手法の有効性を示す．

Proposal of an AntiPhishing system by the HTTP request analysis that used proxy

Motohiko Nakamura†　Masato Terada††　Yuji Chiba††　Norihisa Doi†

† Graduate School of Science Engineering, Chuo University.
1-13-27 Kasuga, Bunkyo-ku Tokyo 112-8551, Japan
†† Research and Development Initiative Chuo University.
12th Floor, Chuo University Korakuen Campus,
1-13-27 Kasuga, Bunkyo-ku Tokyo 112-8551, Japan

Abstract　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 does 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，インターネットは幅広く使われ，オンラインサービスが気軽に利用できるようになった．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 Phishing という詐欺行為が社会問題となっている．Phishing とは，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詐欺である．具体的には，まず，悪意を持った人（Phisher）が金融機関などの正規のメールを装い，個人情報の入力を促す文面のメールを無差別にインターネット利用者（ユーザ）に送りつける．つぎに，メールから偽の Web サイトへ誘導し個人情報を入力させる．

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る．

既存の対策手法の 1 つとして，ブラウザのアドオンツールバーを利用し Phishing サイトへのアクセスを防ぐ手法があるが，その効果は，必ずしも十分であるとはいえない．なぜなら，アドオンツールバーの多くは，ブラックリストを基に Phishing サイトか否かを判断するため，未知の Phishing サイトを検出できないからである．

そこで，本研究では，ユーザの利用する PC（本稿ではクライアント PC と呼ぶ）と Web サーバの間に proxy を設置し，Phishing サイトを検出する手法を提案する．Phishing サイトの検出は，Web ブラウズ時の HTTP リクエストの内容を解析し，Web サイト自体の存続期間が非常に短期間であるなどという，Phishing サイト特有の特徴を検出することで行なう．

本稿では, 提案手法に基づいたプロトタイプシステム（AntiPhishing システム）を作成し，実際に Phishing サイトにアクセスを試みた結果について，検出精度や処理性能の観点から評価する.

2. Phishing に関する動向

2.1 Phishing の被害実態

各国で Phishing 対策のための業界団体が結成されている．最も有名な団体は，米国の APWG (Anti Phishing Working Group) [1] である．APWG は，定期的に Phishing Attack Trends Report という Phishing に関するレポートを公表している．レポートから抜粋した本稿に関係の深いデータを表 2.1 に示す.

表 2.1 Phishing サイトの傾向

Phishing サイトの Web サイトの総数	**13562**
ホスティング元の上位 10 カ国が占める割合	**71%**
ドメインの取得がなく IP アドレスまま運用される割合	**34%**
Web サーバのポート番号が 80 番以外を使用する割合	**8%**
Web サイトの平均存続日数	**5.5 日**
Web サイトの最長存続日数	**31 日**

出典 : September Phishing Attack Trends Report [2]

表 2.1 から Phishing サイトの総数は 13562 個となっており，多くの Web サイトが活動していることがわかる．また，Phishing サイトの傾向として，Web サイトの平均存続日数が 5.5 日などの通常の Web サイトにはない特徴があることがわかる.

2.2 関連研究

Phishing サイトを検出するための既存手法の 1 つとして, ブラウザのアドオンツールバーを利用する方法がある．これは，Phishing サイトに接続しようとするときや接続した後に, ブラックリストを基に判断し，ユーザに警告を促す．アドオンツールバーは各社から提供されているが，その内の 1 つに EarthLink toolbar (EarthLink，米 ISP) [3] がある．EarthLink によりブラックリストが 1 日に何度も更新され, Web サイトへのアクセス時にブラックリストとの整合性を確認し，アドオンツールバー内のアイコン表示が「認証済み」,「不明確」,「非常に疑わしい」と変わる.

3. AntiPhishing システムの提案

3.1 既存の手法を利用する際の課題

既存の手法を利用した Phishing サイトの検出は, ブラックリストを基にしているため, ブラックリストに存在しない Phishing サイトに対しては効果がない. ブラックリストは頻繁に更新されるが, 更新されるまでの間に，被害に遭う可能性がある．また，Java Script 等のスクリプトによりツールバー自体が非表示の場合, ユーザは警告に気がつかない可能性もある.

3.2 AntiPhishing システムの概要

本研究では，ブラックリストに依存せず Phishing サイトを検出する手法を提案する．提案手法は，クライアント PC と Web サーバの間に proxy を設置し，Web ブラウズ時の HTTP リクエストの内容を解析する．Phishing サイトの検出には，Phishing サイトにみられる Web サイトの存続期間が短いなどの特徴的な傾向を捉えることで行う. 提案するシステムの概要を図 3.1 に示す.

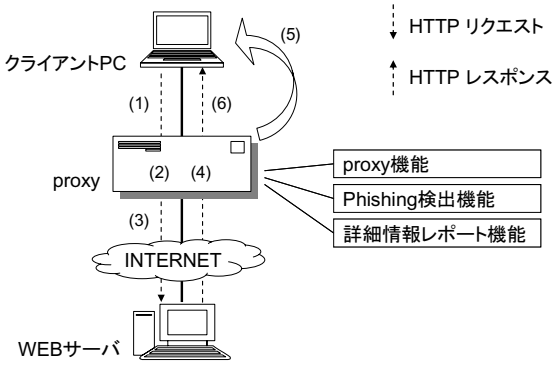

図 3.1 システムの概要

本システムは，ユーザの利用するクライアント PC と Web サーバと Phishing サイトの検出処理をおこなう proxy から構成される．Phishing サイトの検出手順をつぎに述べる.

(1) クライアント PC が Web サーバに対し，Web ページの要求（HTTP リクエスト）を出す.

(2) proxy にて，HTTP リクエストから URL 情報，ドメイン情報を抽出し解析を行う.

(3) (2) の解析結果によらず, proxy と Web サーバ間で通信を行う.

(4) Web サーバからの応答（HTTP レスポンス）をもう一度解析を行う．(2) の結果と合わせ，総合的に Web サイトの危険度を判定する．また，判定結果の詳細情報を記載した HTML ファイルを作成する.

(5) Phishing の疑いがあるページに関しては，ユーザへ Web ページを表示するか確認する.

(6) HTTP レスポンスに判定結果の警告情報を挿入し，クライアント PC に送る．なお，(5) にてページを表示しない場合には，proxy にて Web サーバから受け取った HTTP レスポンスのデータ内, Web ページの内容に関する部分を破棄する.

proxy はつぎに示す 3 つの機能から構成され，各機能に関しては次章で詳述する.

- proxy 機能
- Phishing 検出機能
- 詳細情報レポート機能

proxy を利用するシステム構成の利点として，ブラウザなどの PC 環境に依存することなく，proxy 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である．既存のアドオンツールバーは，Internet Explorer 専用のものが多いため，他のブラウザを利用する場合でも利用可能な本システムの方が汎用性に優れる．

4. AntiPhishing システムの実装

本章では，提案するシステムの各機能に関する実装方法に関して述べる．

4.1 proxy 機能

proxy 機能は，つぎの 3 つの処理を行い，Phishing サイトに関する警告をユーザに提示する（図 4.1）．

- クライアント PC と Web サーバ間の HTTP 通信を中継する
- HTTP リクエストから，URL 情報，ドメイン情報を取得し Phishing 検出機能に解析を依頼する
- Phishing 検出機能による解析結果を，HTTP レスポンスに Phishing に関する警告情報を挿入し，ユーザへ警告を上げる

HTTP リクエストから取得する情報は，図 4.1 に示す網掛け部の URL 情報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要求とそれに付随する画像や FLASH ファイルなどの要求の 2 種類がある．両者とも同じ解析処理を行うが，後者に関しては HTTP レスポンスの HTML タグから判断し，警告情報の挿入処理を行わない点が異なる．

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

図 4.1　標準的な HTTP リクエストの一部抜粋

つぎに，HTTP レスポンスの処理に関して述べる．Phishing 検出機能による解析結果は，図 4.2 に示す網掛け部の情報が両方含まれている HTTP レスポンスの“⇒”部に挿入する．“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 レスポンスに挿入することにより，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情報を挿入できる．また，挿入する警告情報は Web サイトの危険度を示すグラフィカルな表示（以後，警告バーと呼ぶ）とし，ユーザにとって気がつきやすいという利点がある．

```
HTTP/1.0 200 OK
                    //省略//
<html>
    <body>
        ⇒解析結果の挿入を行う位置
            //html ソース（本文）//
    </body>
</html>
```

図 4.2　標準的な HTTP レスポンスの一部抜粋

4.2 Phishing 検出機能

Phishing 検出機能は，proxy 機能からドメイン情報と URL 情報を受け取り，情報を解析することにより Phishing サイトを検出する機能である．

Phishing サイトの検出は，通常の Web ページにはみられない，Phishing サイトの特徴的な傾向を捉えることにより行う．その処理過程を，図 4.3 に示す．

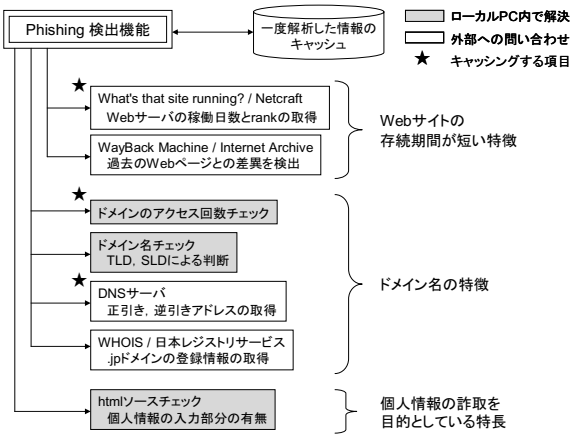

図 4.3　Phishing 検出機能の概要

図 4.3 に示す Phishing 検出機能では，大きくつぎの 3 つの項目に対して処理を行う．また，それぞれの項目には図 4.3 に示すように，さらに小さいチェック項目がある．

- Web サイトの存続期間が短い特徴
- ドメイン名の特徴
- 個人情報の詐取を目的としている特長

図 4.3 中の“★”の項目に関しては，1 回目の処理結果を PC 内にキャッシュし，2 回目以降のアクセス時に参照することにより処理効率を上げる．その他の項目に関しては，毎回処理を行う．

つぎに，Phishing サイト検出における，Phishing 検出機能の処理手順をつぎに示す．

1. proxy 機能から URL 情報とドメイン情報を受け取る．
2. 図 4.3 に示した各項目に関して URL 情報，ドメイン情報を元に処理し，各項目の判定結果をポイント化する．
3. 各項目のポイントを元に，重要な項目のポイント

の重みを大きくし，総合的に Web サイト自体の危険度のポイント値を算出する．

4. Phishing サイトの疑いがある場合，図 3.1 (5) に示すように，ユーザへ WEB ページの表示の可否について確認を行う．許可されなかった場合は，proxy 機能にて HTTP レスポンスの内，Web ページの内容部分が削除され，警告バーのみが表示される．
5. 危険度のポイント値を proxy 機能に返し，また各項目のチェック結果を詳細情報レポート機能に渡す．

なお，今回の実装では危険度のポイント値が，あらかじめ設定した閾値（65 ポイント）を超えると，Phishing サイトとみなす．また，40 ポイントを超えると疑わしいとみなす．つぎに，各項目における処理内容を具体的に述べる．

4.2.1 Web サイトの存続期間が短い特徴

Phishing サイトには，Web サイトの存続期間が非常に短いという特徴的な傾向があるため，提案手法ではインターネット上で提供されるつぎの 2 つのサービスを利用してチェックを行なう．

- What's that site running? (Netcraft，英) [4]
  Netcraft では，ドメイン名に対する各種情報を公開しており，その内の 1 つに Web サーバの連続稼働日数がある．本研究では，これを Web サイトの存続期間とみなす．また，ドメインの rank（ユーザの人気度）を公開している．これらの情報を，問い合わせたドメインに対する結果の HTML ソースから抜き出して利用する．
- Wayback Machine (Internet Archive，米) [5]
  Internet Archive では，過去に公開された HTML ファイルをデータベースとして保管し，公開している．Phishing サイトは，Web サイトの公開日数が高々31 日程度であるため，新規で作成された Web サイトはデータベースにない可能性が高い．過去ページとの相違点を検出することで，Web サイトが新規で作成されたことや，Web ページの改ざんがあったことが分かる．なお，今回の実装では，現在と過去の HTML ソースのビット数（ファイルサイズ）の違いをもって差異とみなす．

4.2.2 ドメイン名の特徴

Phishing サイトは，ドメイン名についても特徴的な傾向がある．まず，ドメイン名のアクセス回数をチェックを行なう．これには 2 つの役割があり，ユーザが過去に訪問したことがあるか否かを調べる役割と，ドメイン名からキャッシュされたデータの有無を調べる役割がある．

- ドメイン名
  ドメイン名が IP アドレスのままでないか，また，ポート番号が 80 番以外でないかをチェックする．そして，Top Level Domain と Second Level Domain（".jp" ドメインのみ）を参照し，Phishing 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ないか，あらかじめローカル PC 内に保持している情報を基にチェックする．
- WHOIS
  TLD が日本の" .jp" ドメインの場合は，日本レジストリサービスの WHOIS [6] にドメインの取得状況を問い合わせる．この項目は危険度をポイント化せず，詳細情報レポート機能へ取得した情報を渡すだけである．
- DNS
  Phishing サイトは簡易的に作成されることが多いため，ドメイン名の逆引きが存在しないことが多い．そこで，取得したドメイン名の正引きをし，正引きした IP アドレスの逆引きアドレスの有無をチェックし，逆引きアドレスが存在する場合，元のドメイン名との整合性を確認する．

4.2.3 個人情報の詐取を目的とする特徴

Phishing サイトには，個人情報の詐取を目的とされているために，Web ページ内に必ず個人情報の入力を促す文言を含んでいる．そこで，取得した HTML ソース内を検索し，あらかじめ用意しておいた単語の抽出を行う．単語は，実際の Phishing サイトで多く使われる単語を 20 語程度抜粋した．例えば," パスワード" という単語の検索を行うが，今回の実装では個人情報の抽出が目的ではない場合でも HTML ソース内に該当する文言が含まれていれば抽出を行う．

4.3 詳細情報レポート機能

詳細情報レポート機能は，Web サイトのドメインについての情報や Web サーバの稼働状況など詳細情報をユーザが必要に応じ確認するための機能である．動作概要を図 4.4 に示す．

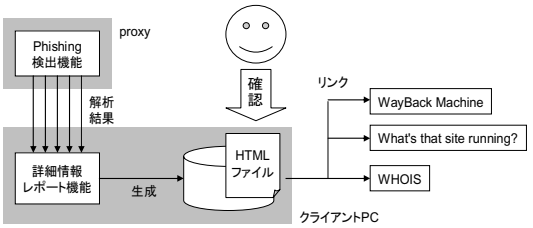

図 4.4 詳細情報レポート機能の概要

まず，Phishing 検出機能による各チェック項目の解析結果から，情報をまとめた HTML ファイルを生成する．この HTML ファイルは，Web ページに挿入された警告バーからリンクされており，リンクをクリックすることにより HTML ファイルが参照できる．なお，今回の実装ではクライアント PC 内で Web サーバを動作させることで実現した．この HTML ファイルは，危険度の高い Web ページなどを閲覧した場合，ユーザにとっての安全性を確認するためのガイドラインとなる．また，HTML ファイルに What's that site running? と Wayback Machine へのリンクがあり，ユーザがさらに詳しい情報を必要とした場合にも

対応できる.

5. 評価

5.1 評価手法の概要

研究の目的である Phishing サイトを検出の可否を判定するために, 作成したプロトタイプシステムを使い，実際に Phishing サイトへアクセスを試みる．評価対象は，機能面と性能面における、つぎの 3 つの項目とする.

(1) プロトタイプシステムにおける Phishing サイトの検出率と，既知の通常サイトの誤検知率（既存のアドオンツールバーについても計測し，プロトタイプシステムと比較する.）
(2) 警告バーの挿入成否の割合
(3) Phishing サイトを検出するためにかかる処理時間

※評価にあたっては，比較対象として，つぎの 3 つのアドオンツールバーを利用した ： Phishing Filter [7]，EarthLink toolbar，SpoofGuard [8]

Phishing サイトの一覧は，RBL.JP [9] 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 日までの期間で，計 25 の実在の Phishing サイトを対象とした. また，誤検知率を測定のための通常の Web サイトの一覧は，「世帯内パソコンにおけるインターネット利用状況調査」（ビデオリサーチ）[10] におけるランキング上位の Web サイトと大手金融機関の Web ページの計 38 の Web サイトを対象とした.

5.2 評価環境

提案した手法の有効性を示すため, プロトタイプシステムを作成し，評価を行なった.

5.2.1 実験環境

実験は中央大学土居研究室の環境を利用し, 研究室内のネットワークにクライアント PC を接続して行った．クライアント PC のシステム構成を，表 5.1 に示す.

表 5.1 システムの構成

CPU	Pentium4 2.60 [GHz]
Memory	1.00 [GB]
Network	100 BASE-TX
OS	Windows XP SP2
ブラウザ	Sleipnir version 2.00

5.2.2 ソフトウェア構成

プロトタイプシステムのソフトウェア構成について述べる．proxy 機能は http-proxy.pl [11] という，HTTP 通信の中継機能のみを実装した perl のプログラムを機能拡張して利用した. 今回の実装では, proxy はシステム簡略化のためクライアント PC 内でソフトウェアとして動作させた．Phishing 検出機能と詳細情報レポート機能に関しては perl で実装し，一部のモジュールに Java を使った.

5.3 プロトタイプシステムの動作概要

プロトタイプシステムを使い, 中央大学のホームページへアクセスした動作結果を図 5.1 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，メモリが 3 マス増えている．なお，目盛りは危険度が高くなるにつれ，黄色，赤と色が変わる．また，警告バーに詳細情報レポート機能が作成する HTML ファイルへのリンクが張ってあり，必要に応じ参照できる.

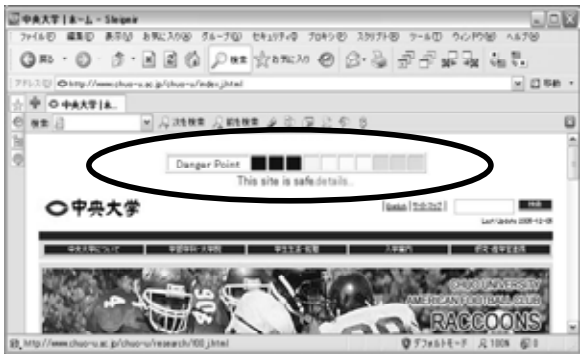

図 5.1 プロトタイプシステムの動作概要

5.4 測定結果と考察

(1) 検出率と誤検知率

プロトタイプと既存の各ツールバーを利用し，Phishing サイトへアクセスした際の検出率を図 5.2 に示す．また，通常サイトへアクセスを行った際の誤検知率を図 5.3 に示す．なお，図 5.2 は Phishing サイトの検出成功を“ ○”とし，検出失敗を“ ×”で示している．逆に，図 5.3 では，通常の Web サイトと正しく認識した場合を“ ○”，誤検知した場合を“ ×”として示す．また，両図とも，疑わしい Web サイトと認識した場合を“ △”として示す.

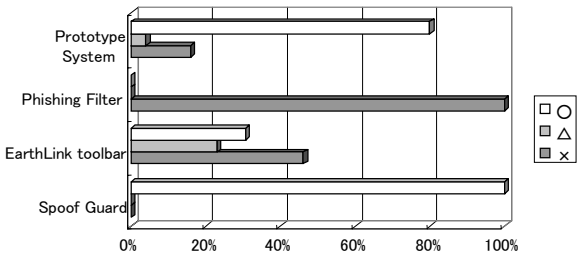

図 5.2 Phishing サイトの検出率

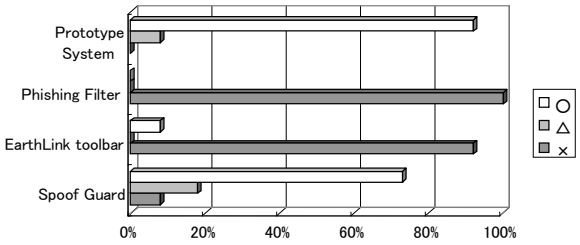

図 5.3　通常の Web サイトの誤検知率

なお，図 5.2，図 5.3 の結果は，既存のツールバーが反応しなかった場合のデータを“×”として含んでいる．

図 5.2 から，プロトタイプシステムでは約 80%の Phishing 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 5.3 から，プロトタイプシステムの誤検知率は約 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高々8%であった．

既存のツールバーではブラックリストに依存している Phishing Filter と EarthLink toolbar は Phishing 検出率と誤検出率の両方に関して精度が低かった．一方，ブラックリストを用いない SpoofGuard に関しては，Phishing の検出精度は 100%と高かったが，プロトタイプシステムと比べ誤検知率が高かった．これは，セキュリティレベルが高めに設定されていると想定できる．

(2) 警告バーの挿入成否

評価対象の全 Web サイトの内，87%の Web サイトにおいて警告バーが正常に挿入された．しかし，フレームを使用している一部の Web ページでは，正しい位置に表示されなかったり，挿入処理が行われなかったりした．なぜなら，フレームページでは Web サイトが複数の HTML ファイルから成っているため，警告バーを挿入すべき HTML ファイルの識別が困難であるからである．

(3) 処理性能

Phishing の検出にかかる処理時間の計測のため，Web ページが読み込まれるまでの時間を測定した．実験は，プロトタイプシステムの有無の場合に分け，先述した通常の Web サイトの一覧を用い，その平均値の時間を求めた．実験の測定結果を図 5.4 に示す．

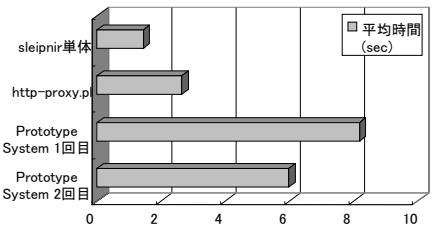

図 5.4　Web サイト表示にかかる実時間

プロトタイプシステムを利用した際の時間は，ブラウザのみを使用した場合に対し約 5.7 倍，http-proxy.pl を利用した場合に対して約 3.1 倍の時間を要した．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．

6. おわりに

本稿では，proxy を利用し HTTP リクエストの解析を行ない，Phishing サイトの特徴的な傾向と捉えることにより Phishing サイトを検出する AntiPhishing システムを提案した．提案手法に基づきプロトタイプシステムを作成し，測定結果を示した．評価結果から，実際に Phishing サイトへアクセスした際の検出率は，疑わしい Web サイトを含めると全体の約 85%であり，誤検知率は 8%であった．また，既存のブラックリストを用いた対策手法では検出できない Phishing サイトについても検出することができ，Phishing 対策としての有効性が示された．

今後の課題として，つぎに示す内容について検討していきたいと考えている．

- 新たなパラメータを追加し，Phishing サイトの検出率向上と誤検知率の低減させる．
- Phishing の検出処理に時間がかかっているため，外部機関への問い合わせ手順を効率化し，処理時間の短縮を図る．
- 今回の実装では，proxy にて通信内容を解析しているため，暗号化されている SSL 通信などには対応していない．SSL 通信の場合はクライアントサイドで URL やドメイン情報を取得し解析するように機能拡張を行う．
- 警告バーの表示成功率を上げるため，フレームページではポップアップウインドウなどを利用して，強制的に表示させるようにする．

参考文献

[1] Anti-Phishing Working Group, http://www.antiphishing.org
[2] http://www.antiphishing.org/reports/apwg_report_DEC2005_FINAL.pdf ※12 月時点のレポート
[3] EarthLink toolbar，EarthLink, http://www.earthlink.net/software/free/toolbar
[4] Netcraft，http://news.netcraft.com/
[5] Wayback Machine，Internet Archive, http://www.archive.org/Web/Web.php
[6] 日本レジストリサービス，http://jprs.jp
[7] Microsoft，http://msdn.microsoft.com/ie
[8] Stanford Security Lab，Stanford university, http://crypto.stanford.edu/SpoofGuard
[9] RBL.jp, http://www.rbl.jp
[10] 世帯内パソコンにおけるインターネット利用状況調査，ビデオリサーチ http://www.videoi.co.jp/data/wsr/wsr051012.html
[11] HTTP proxy，68user's page, http://x68000.q-e-d.net/~68user