# 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

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

情報工学専攻　中村 元彦
Information and System Engineering Course　Motohiko Nakamura

**要約**

今日，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．既存のブラウザのアドオンツールバーを利用する対策手法では，ブラックリストデータベースに無い 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，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．また，評価からその有効性を示す．

**キーワード** ： Phishing，フィッシング，セキュリティ，proxy，ネットワーク

## 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，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，偽の Web サイトへ誘導し，個人情報を詐取することである．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．

しかし，Phishing 対策における有効な手法はまだ確立されておらず，効果が期待できる既存の対策手法の一つとして，ブラウザのアドオンツールバーによりアクセスを防ぐ手法がある．しかし，アドオンツールバーの多くは，Phishing サイトのブラックリストデータベースを基に判断しているため，データベースにない Phishing サイトは効果がないという欠点がある．

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，Web ブラウズ時の HTTP リクエストの内容を解析することにより，Phishing サイトを検出する手法を提案する．Phishing の検出には，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 というレポートを公表している．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

| | |
|---|---|
| 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 |

引用元：September Phishing Attack Trends Report

### 2.2 関連研究

Phishing サイトを検出するための既存手法の一つとして，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，Phishing サイトのブラックリストデータベースを基に判定し，ユーザに警告を促す．その内の 1 つに図 2.1 に示す EarthLink toolbar (EarthLink，米 ISP) [2] がある．EarthLink によりデータベースが 1 日に何度も更新され，Web サイトへのアクセス時に整合性を確認し，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．

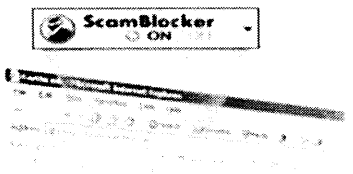


図 2.1　EarthLink toolbar（出展：EarthLink）

## 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

Phishing サイトの検出は，データベースを基にしているため，データベースにない Phishing サイトや新規で立ち上げられたばかりの Phishing サイトに関しては効果がない．データベースは頻繁に更新されるが，更新されるまでの間に，被害に合う可能性がある．また，Java Script 等のスクリプトの使用でツールバー自体が非表示の場合，ユーザが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，データベースに依存せず 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．提案するシステムはつぎの 3 つの機能から構成され，システム構成を図 3.1 に示す．

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

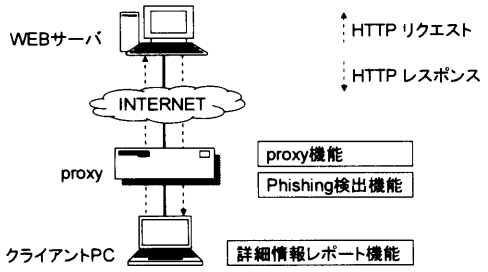


図 3.1　システムの概要

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詳細・実装手法に関して述べる．

### 4.1 proxy 機能

proxy 機能は，つぎの 3 つの役割を持ち，図 4.1 に proxy 機能の概要と他機能との連携概要を示す．

- クライアント PC と Web サーバ間の HTTP 通信を中継する
- Web ページの要求である HTTP リクエストから，URL 情報と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 レスポンスに警告バーとして挿入し，ユーザへのフィードバックを行う．

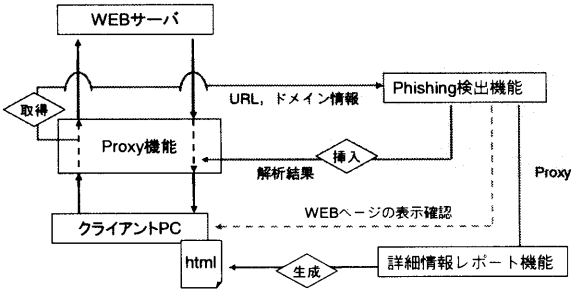


図 4.1　proxy 機能の概要

HTTP リクエストから取得する情報は，図 4.2 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．両者とも同じ処理を行うが，後者に関しては HTTP レスポンスへ警告バー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 4.2　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 4.3 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．こうすることで，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情報を挿入できる．また，挿入する情報は Web サイトの危険度を示すグラフィカルな表示（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 4.3　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．処理過程を，図 4.4 に示す．図中の，“ドメイン単位”は，過去に訪れたことのないドメインへの初回アクセスのときのみ処理を行うという意味であり，一方，“Web ページ単位” は，毎回処理を行なう．また，“★” の項目に関しては，PC 内に情報をキャッシュし，

2回目以降に参照することで処理効率を上げる．

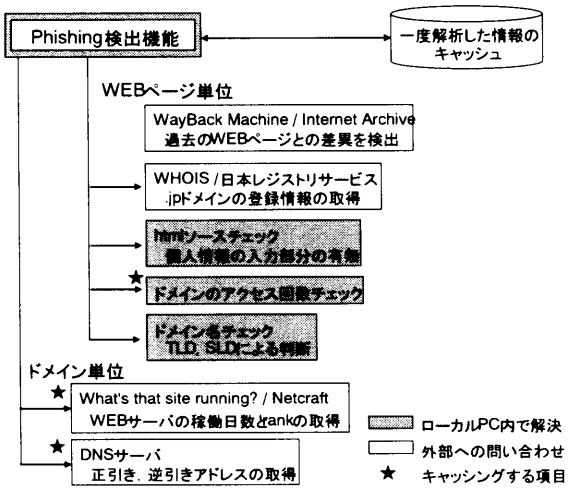


図 4.4　Phishing 検出機能の概要

Phishing サイト検出の手順は，まず各項目に関しての判定結果をそれぞれポイント化する．各項目のポイントを総合的に判断して Web サイト自体の危険度のポイント値を算出する．その際に，重要な項目に関しては重みを大きくする．危険度のポイント値が，閾値（65 に設定）を超えると，Phishing サイトとみなし，40 ポイントを超えると注意を促す．Phishing サイトの疑いがある場合は，図 4.1 に示すように，ユーザに Web ページを表示するかどうかの確認を行う．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という特徴

Phishing サイトは 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) [3]
  ドメインごとに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．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．問い合わせ結果の HTML ソースから，これらの情報を抜き出して利用する．
- Wayback Machine (Internet Archive，米) [4]
  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，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，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する．

#### 4.2.2　その他の特徴

Web サイトの存続期間以外にも図中に示した各項目に対してチェックを行なう．まず，ドメイン名に関しては，過去のユーザのアクセス回数をチェックする．また，ドメイン名 Top Level Domain，Second Level Domain を参照し，危険性の高い国のドメインなどをチェックする．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 [5] にドメインの取得状況を問い合わせる．さらに，ドメインが正規の手続きを踏んで取得されているか確認するために，ドメイン名の正引きアドレスと，さらにその逆引きアドレスを参照する．

Phishing サイト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，HTML ソース内にあらかじめ用意しておいた単語がないか検索を行う．なお，単語は実際の Phishing サイトで多く使われるものを 20 語程度用いた．

### 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昨日である．動作概要を図 4.5 に示す．

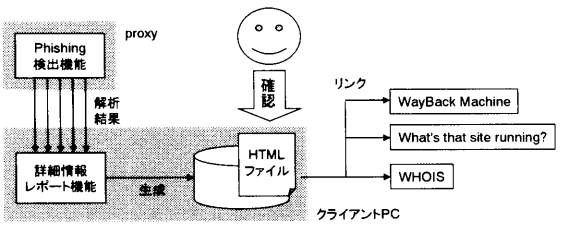


図 4.5　詳細情報レポート機能の概要

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．HTTP レスポンスに挿入される警告バーの部分から HTML ファイルが参照でき，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の判断指標となる．また，HTML ファイル内に What's that site running? などへのリンクがあり，ユーザがさらに詳しい情報を必要とする場合のニーズにも対応できる．

## 5.　評価

### 5.1　評価手法の概要

本節では評価手法の概要について述べる．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．評価は，機能面と性能面の，つぎに示す計 3 項目を評価対象とする．

- プロトタイプシステムと既存のアドオンツールバーにおける Phishing サイトの検出率と，既知の通常サイトの誤検知率．
- 警告バーの挿入の成否の割合
- プロトタイプシステムを利用の有無による，

Webサイトが表示されるまでの実時間の差
※比較対象のアドオンツールバー…Phishing Filter [6], EarthLink toolbar, SpoofGuard [7]

Phishingサイトの一覧はRBL.JP [8] を参考にし，計25のWebサイト，また，誤検知率を測定のため，よく使われるドメイン（ビデオリサーチ調べ）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 プロトタイプシステムの構成

クライアントPCは，中央大学土居研究室内における windows XP SP2のPCを利用した．Proxy機能は，BSDライセンスである http-proxy.pl [8] というperlのプログラムを機能拡張し，今回の実装ではシステム簡略化のためクライアントPC内で動作させた．

### 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，Webページへアクセスした結果を図5.1に示す．図の中央付近に情報バーが挿入され，中央大学のホームページの場合は，危険度のポイント値が20〜30ポイント前後であることを表している．

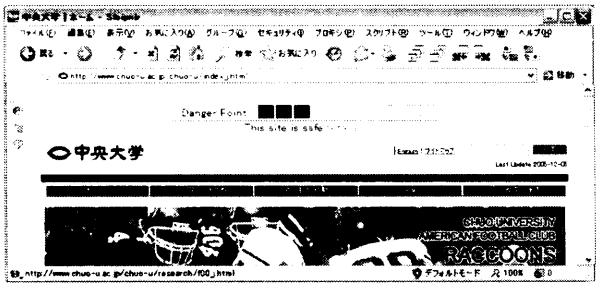


図5.1 プロトタイプシステムの動作概要

### 5.4 測定結果と考察

プロトタイプと既存の各ツールバーを利用し，Phishing サイトと通常サイトへアクセスを行った結果を，図5.2と図5.3に示す．図は成功を“○”，失敗を“×”として示している．

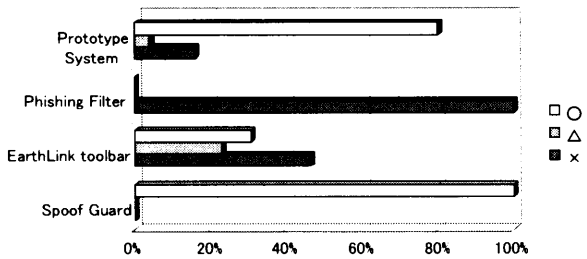


図5.2 Phishingサイトの検出率

図から，Phishingサイトの検出率は，疑わしいWebサイトを含めると約85%であり，誤検知率は0%という結果となり，ある程度Phishingサイトを判別できた．既存のツールバーでは，データベースを用いない手法を採用している SpoofGuardがよい結果となったが，表5.3に示す誤検知率がプロトタイプシステムよりやや高かった．

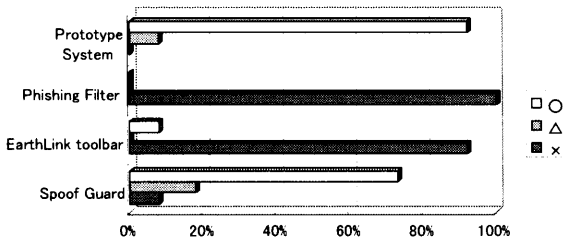


図5.3 通常のWebサイトの誤検知率

警告バーの表示の可否については，評価対象の全Webサイトの内，87%が挿入されが，フレームを使っているページなどでは正常に挿入されない場合もあった．

また，処理時間に関しては，通常のWebサイトを対象とし評価を行ない，プロトタイプシステムを利用した際の時間は8.20秒で，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回目以降のアクセスに対しては，約2.2秒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，全体の80%であり，誤検知率は 0%であった．また，既存のデータベース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．

- 外部機関への問い合わせ方法を効率化し，Phishingサイトの検出処理時間の短縮を図る．
- 警告バーの表示率を上げるため，フレームページではポップアップなどで強制的表示させる．

### 参考文献

[1] Anti-Phishing Working Group, http://www.antiphishing.org
[2] EarthLink toolbar, EarthLink, http://www.earthlink.net/software/free/toolbar
[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] Microsoft, http://msdn.microsoft.com/ie
[7] Stanford Security Lab, Stanford university, http://crypto.stanford.edu/SpoofGuard
[8] RBL.JP, http://www.rbl.jp/
[9] HTTP proxy, 68user's page, http://x68000.q-e-d.net/~68user