メッセージ中URLに基づくドメイン登録日検索システムを用いた迷惑メール判別機構

松岡 政之[1] 井上 達貴[1] 山井 成良[2] 岡山 聖彦[2] 河野 圭太[2] 中村 素典[3] 民田 雅人[4]

概要：
近年，ワンクリック詐欺やフィッシング詐欺などの悪意ある Web ページへの誘導などを目的とした多くの迷惑メールが送信されている．迷惑メール対策の技術的な手法の一つとして，迷惑メール内の URL をブラックリスト化して照合を行う，URL ブラックリストと呼ばれる手法が存在する．しかし，攻撃者はドメインを使い捨てとし，新たに取得したドメインを宣伝または攻撃用 Web ページに用いることによって，メールメッセージ内の URL を変更する手口を使用し始めており，既存のブラックリストでは迷惑メールと判定できないものが現れている．本研究は，迷惑メール本文中のドメイン登録日に注目し，迷惑メール判定手法の判定精度向上を目的とするものである．迷惑メール送信者は頻繁にドメインを取得しているため登録期間の浅いドメインほど迷惑メールに利用される可能性が高いと考えフィルタリングを行う．そこで，本研究では先行研究であるドメインの登録日を収集・記録する機能と検索する機能をもったドメイン登録日検索システムを利用した迷惑メールフィルタの実装を行った．

Spam Discrimination Mechanism Using Domain Registration Date Retrieval System Based on URLs in Messages

Masayuki Matsuoka[1] Tatsuki Inoue[1] Nariyoshi Yamai[2] Kiyohiko Okayama[2] Keita Kawano[2]
Motonori Nakamura[3] Masato Minda[4]

1. まえがき

近年，インターネット利用者の増加に伴い，WWW と並んで電子メールは非常に多く人に利用され，社会的な活動を支える通信手段として必要不可欠なものとなっている．しかし，電子メールのサービスにはセキュリティ面での問題が多く，特に，宣伝や広告を目的をした不特定多数の利用者に送信される迷惑メールが社会問題となっている．文献 [1] によると，全電子メールトラフィックの 65.9%を迷惑メールが占めている．これらの迷惑メールによって，(1) 通信資源や計算機資源，通信費用や通信時間などを無駄に消費する，(2) 迷惑メールの区別に多大な労力を必要とし，また正当なメールを誤って迷惑メールとともに削除したりする，(3) 迷惑メールの送信元などとして名前を騙られたり迷惑メールの中継に組織内のメールサーバが使用されたりすることによってその組織が迷惑メールの送信に関与していると疑われる，などの深刻な問題が発生している．また，迷惑メールの多くはメッセージ中に URL が記述してあり，その URL にアクセスすることによって悪質なサイトに誘導され，コンピュータウィルスに感染したりフィッシング詐欺やワンクリック詐欺などの詐欺行為の被害に遭ったりするなどの問題も深刻である．

このような問題に対して，現在様々な対策が施されている．その 1 つとして，電子メールに記載されているメッセージの内容を基に迷惑メールを判別する方法があり，その中でも URL のブラックリスト *1 を作成し，受信メー

1 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama Uniersity
2 岡山大学情報統括センター
Center for Information Technology and Management, Okayama Univercity
3 国立情報学研究所
National Institute of Informatics
4 株式会社日本レジストリサービス
Japan Registry Services Co., Ltd.

*1 実際にはドメイン部のブラックリスト

ルメッセージ内の URL と照合することによって迷惑メールを判別する方法がある．迷惑メールの多くはメッセージ内に URL が記述されているため，この手法は高い効果を期待できる．しかし，近年では攻撃者はこの方法に対抗するために新たなドメインを頻繁に取得し，メッセージ内の URL に使用するドメインを使い捨てにする手口を使用し始めている．これによって，迷惑メールに記述する URL が頻繁に変更され，ブラックリストの有効性が低下する．

これに対して，我々は使い捨てドメインは登録日からの経過日数が浅い傾向にある [5] 点に着目し，いくつかの TLD に属するドメインについてその登録日を検索するシステム（以下，ドメイン登録日検索システ）[8] の開発を行った．本稿では同システムを活用してドメイン登録日からの経過日数を迷惑メール判別に用いる迷惑メールフィルタの設計と実装について報告する．

以下，2 章では従来の迷惑メール対策とその問題点の詳細を述べ，3 章でその問題点を解決する提案システムの実現方針について述べ，4 章では提案システムの実装を述べ，5 章でその動作について述べる．最後に 6 章で本論文をまとめ，今後の課題について述べる．

2. 迷惑メール対策の現状

2.1 従来の迷惑メールの対策手法とその問題点

迷惑メール送信者は，自身や攻撃に使用される Web サーバを突き止められることを防いだり，より多くのメールがブロックされずに送信先へ届くようにするために，さまざまな手法を用いている．本節では本研究で対象とする迷惑メールの手口とその対策手法を述べる．

2.1.1 迷惑メールの対策手法

迷惑メール対策手法は大別して，送信元 IP アドレスなど本文によらない情報に基づいて受信そのものを拒否するブロッキングと，受信メールのヘッダや本文に基づいて判別するフィルタリングがある．本研究ではフィルタリングに焦点をあて，このうちでも本文中の URL に着目する方法を対象としている．

2.1.2 本文中 URL に基づく迷惑メールフィルタリング

多くの迷惑メールは，宣伝や悪意ある Web サイトへの誘導，ウィルスの配布などを主目的としており，本文中に URL を含むものが多い．

そこで，本文中 URL に基づく既存の対策として URL ブラックリストと呼ばれるものが存在する．これは，本文中の URL が悪質かどうかを判断するためのブラックリストであり，代表的なものとして SURBL[2] や URIBL[3]，ivmURI[4] がある．これらのブラックリストは，迷惑メール本文中のクリック可能なリンクにあるドメインを一覧にしたものである．迷惑メール本文中から全 URL を抜き出し，それをブラックリストと比較し，一致すれば迷惑メールと判断してブロックできる．この手法は，迷惑メール送信者が送信元を偽っても有効であり，また，多くの迷惑メールが本文中に URL の記述を持つため，高い効果を期待できる．

2.1.3 URL ブラックリストの問題点

迷惑メールの送信元や誘導先などに不正に使用されたドメイン（悪性ドメイン）は，一般的には不正使用が判明するとレジストリによって停止される．その対抗手段として，攻撃者は不正に入手した個人情報によって次々と新しいドメインを取得する．こうして取得したドメインを使用し，攻撃用サイトを活動させることにより，迷惑メールメッセージ内の URL も頻繁に変更することができる．

このような手口で取得，使用されたドメイン（使い捨てドメイン）は大量に存在し，また使用頻度が比較的少ないため，URL ブラックリストへの登録が追い付いていない．また，このようなドメインは長期間使われることがないため，たとえ URL ブラックリストへ登録されたとしても実際に使い捨てドメインが検索される機会は少ないと思われる．

これらのことから，URL ブラックリストは使い捨てドメインに対する有効性が損なわれており，迷惑メールを判別する能力に疑問が生じている．

2.2 ドメイン登録日に基づく迷惑メール判定手法

前節で述べたように，URL ブラックリストではドメインを使い捨て，URL を頻繁に変更する攻撃方法には対応できない．そこで，ドメインを使い捨てるには頻繁に新しいドメインを取得しなければならず，登録日からの経過日数の浅いドメインを URL に用いたメールほど迷惑メールである可能性が高いと考えられる．そのためには，メールメッセージ内の URL に使用されている各ドメインの登録日を調べる必要がある．

先行研究において JWSDB というブラックリストに登録されているドメインに関して，そのドメインの登録情報を調べたものがある [5]．それによると，約半数の登録者は一つしかドメインを所持していないが，割合的に少数の登録者が複数のドメインを所持している．また，93%のドメインがそのような登録者によって同じ日にほかのドメインと共に登録されており，80%のドメインが 10 個以上のまとまりとして登録されている．さらに，これらのドメインの内約 90%は 1 年以内に登録されたものである．これらのことからも，ドメインの登録日を迷惑メール判定の指標に用いることは多くのドメインに対して有用であると考えられる．

ドメインの登録日を調べる方法に WHOIS を用いる方法がある．WHOIS とは IP アドレスやドメインの登録者などに関する情報を，インターネットユーザが誰でも参照できるサービスである [6]．WHOIS を用いることで，ドメイン名やレジストラ名，ドメインの登録年月日などさまざま

な情報を参照することが可能である．レジストラが提供する情報は，IP アドレスやドメイン名などのインターネット資源を管理する ICANN という組織によって規定されている [7]．これらの情報は，(1) ネットワークの安定的運用を実施する上で，技術的な問題発生の際の連絡のために必要な情報を提供，(2) ドメイン名の申請届出時に，同一ドメインや類似ドメインの存在を確認するために必要な情報を提供，(3) ドメイン名と商標等に関するトラブルの自立的な解決のために必要な情報を提供，などの目的で公開されている．

しかし WHOIS は，マーケティングなど本来の目的外で使用されることを避けるため，検索頻度の高いユーザを一時的に制限したり，検索時間に間隔を持たせるなど，WHOIS 情報への大量アクセスを避ける対策がなされている．また，WHOIS はそのフォーマットが定められていないため，提供団体によってフォーマットが不均一であり，プログラムによる解析が非常に困難である．そのため，迷惑メール判定のためにメールメッセージ内 URL のドメインに関して WHOIS 情報の検索を行う場合，すべての受信メールに関して検索を行うことは事実上不可能である．

3. ドメイン登録日に基づく迷惑メール判別手法の提案

本章ではドメイン登録日に基づき迷惑メールを判別できるようにするための提案手法について述べる．

3.1 実現方針

2 章で述べたとおり，登録日の新しいドメインが迷惑メールに利用される可能性が高いことを迷惑メール判定に利用する．しかし，既存の方法ではすべてのメールについてメールを受信するたびに本文中のドメインに対してその登録日を調べることはできない．これに対して我々の研究グループではドメイン登録日検索システムを開発した [8]．これは，ドメイン名を用いてそのドメインの登録日を容易に検索できるシステムである．ドメイン登録日検索システムを迷惑メール判定に使用するためには，そのシステムを用いた迷惑メールフィルタをメールサーバに実装する必要がある．

以下，本章ではドメイン登録日検索システムについて述べた後，この方針を実現するための提案システムの設計について述べ，次章で実装について述べる．

3.2 ドメイン登録日検索システム

ドメイン登録日検索システムとはドメインの登録日を収集し，容易に検索できるシステムである．以下，本節でドメイン登録日検索システムについて説明する．

3.2.1 機能

(1) ドメイン登録日の収集・記録機能
　ドメイン登録日検索システムではゾーン管理団体にゾーンファイルの提供を受け，その 1 日ごとの差分をとることによってドメイン登録日を割り出だし，DNS サーバに登録する．

(2) ドメイン登録日の検索機能
　通常の DNS サーバと同様にドメイン名を用いて問い合わせることにより，その登録日を TXT レコードで応答する．図 1 はその問い合わせと応答の例であり，この場合 example.com の登録日は 2011 年 10 月 25 日であることを示している．

3.2.2 問題点

上記のシステムのそのままではメッセージ中に短縮 URL を用いられると対応できないなどの問題点がある．短縮 URL とは，文字数の多い URL を持つ Web サイトに対して文字数の少ない URL を持つ Web サイトからリダイレクトすることによって，文字数の少ない URL を用いてアクセスできるようにする手法である．これは，SNS（Social Networking Service）等において書き込める文字数に制限のあるサービスを利用する場合，文字数の多い URL を扱うことができないため広く利用されている．

例えば，

http://www.okayama-u.ac.jp/

という URL が存在するがこれを短縮 URL を作成するサービスの 1 つである http://bitly.oshiire.org/を利用して変換すると，

http://bit.ly/TapeZH

となる．今後はこの URL によって元の URL を持つ Web サイトにアクセスすることができる．

これにより，URL のドメインが okayama-u.ac.jp から bit.ly に変化してしまい，元の URL のドメインが隠蔽されることに問題がある．他の多くの短縮 URL サービスでも同様に，

http://サービス毎に固有のドメイン/可変のパス部分

と変換されることが多いため同様の問題が発生する．そのため URL が短縮されているものであるかを確認し，短縮されたものだった場合はそのリダイレクト先である元の URL を取得する方法がドメイン登録日検索システムを利用するために必要となる．

3.3 システムの設計

提案システムは図 2 に示す構成となっており，各ブロックでは次に示すような機能を持つ．

(1) メッセージ中 URL の短縮判定と復元
　メッセージ中 URL を抽出して，その URL が短縮されているものか否かを判定する．短縮されていると判定した場合，元の URL を復元する．

(2) ドメイン登録日検索システムへの問い合わせ
　(1) で得られたドメインを用いてドメイン登録日検索

```
# dig @150.46.XXX.YYY example.com.zone txt

; <<>> DiG 9.6.-ESV-R3 <<>> @150.46.XXX.YYY example.com.zone txt
; (1 server found)
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 55759
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;example.com.zone.               IN      TXT

;; ANSWER SECTION:
example.com.zone.        86400   IN      TXT     "20111025"

;; Query time: 18 msec
;; SERVER: 150.46.XXX.YYY#53(150.46.XXX.YYY)
;; WHEN: Mon Feb  4 18:21:42 2013
;; MSG SIZE  rcvd: 54
```

図 1 ドメイン登録日検索システムを用いた登録日検索の例
Fig. 1 An example of the retrieval with the domain registration retrieval system.

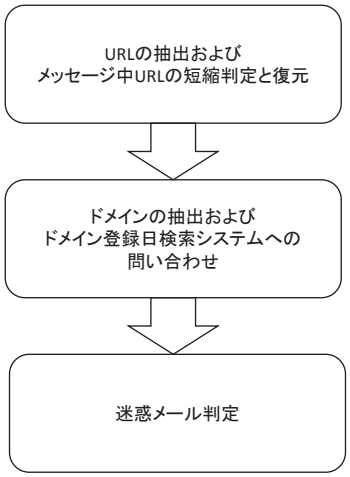

図 2 提案システムの処理の流れ
Fig. 2 The processing flow if the proposal system.

システムへ問い合わせを行う．その結果からドメインの登録日のみを抽出する．

(3) 迷惑メール判定
問い合わせた結果のドメイン登録日とメール受信日を比較する．登録日が受信日から浅いほど有害なメールである可能性が高いものとして迷惑メール判定に用いる．

以下，本節では各ブロックでの詳細について述べる．

3.3.1 メッセージ中 URL の短縮判定と復元

提案システムでは以下に示す処理をメールの受信毎に行うことによって，常にドメイン登録日検索システムを参照できるように短縮 URL から元の URL を復元する．

(1) ホワイトリストを用いた判定
同一サービスで変換された URL は基本的に同じドメインに変換される．このことを利用して，現在利用されている短縮 URL サービスの変換後のドメインをホワイトリストに登録する．メッセージ中 URL のドメイン部分とホワイトリストを比較して一致したものは短縮 URL と判定して (3) で説明する処理を行う．新しい URL 短縮サービスサイトの開設など，ホワイトリストに存在しない URL 短縮が存在する可能性があるため，ホワイトリストに存在しなかったものは (2) で説明する判別法を用いる．

(2) パス部分の文字数を用いた判定
URL のパス部分の文字数を調査することで判定を行う．短縮 URL の本来の目的は長い URL を短く表示することであるので，短縮後の URL のパス部分は一定字数内であると考えられる．そこで，パス部分の文字数の多い URL は短縮されていないものと判断して，ドメイン登録日検索システムに問い合わせを行う．一方，ホワイトリスト未登録かつパス部分の文字数の少ない URL は短縮されている可能性を考慮して (3) で説明する判別法を用いる．

(3) HTTP 通信を用いたステータスコードの判定
HTTP 通信でデータ要求をすることによって指定 URL の HTTP ステータスコードを取得することができる [10]．短縮 URL はリダイレクトを利用しているため，この HTTP ステータスコードが 301 または 302 だった場合短縮されているものと判定する．HTTP ステータスコードの 301 と 302 は以下の通りである．

- 301 Moved Permanently

リクエストしたリソースの恒久的移動を示す．

- 302 Found
  リクエストリソースの一時的な移動を示す．

これらの場合 URL の移動があるので，リダイレクト先の URL を確認することで短縮前の URL を確認できる．問い合わせを行った URL が 301 か 302 のステータスコードだった場合短縮 URL と判定し，それ以外のステータスコードだった場合その URL は短縮されていないと判断する．

先に説明したシステムの目的である URL が短縮されているかの判別はこのような HTTP 通信による判別方法のみで達成することができる．しかし，1 度のみしかアクセスを可能としないサイトに対してこの手法をとることは問題がある．ステータスコードを確認した際，通信先サーバのアクセスログには問い合わせの履歴が残ってしまう．このステータスコードの確認による履歴が 1 度のアクセスとみなされ，ユーザが本来の目的のために使用をするときに目的 URL にアクセス出来ない事態が発生する可能性がある．このようなサイトはパスワードの変更や確認に利用されることが多く，また URL のパス部分に識別用の長い文字列が入っていることが多い．そこで，事前に (1) や (2) の処理を行うことでこのような事態を出来る限り防げるようにしている．また，これらの方法は悪質な Web サイトへのアクセスログをなるべく残さないようにするためでもある．

(4) 短縮 URL 判定後の処理
メッセージ中 URL が短縮されているものと判定しリダイレクト先 URL を取得した後，その URL を再度同様の手法で短縮 URL ではないと確認できるまで繰り返す．それにより短縮 URL 化する処理が複数回に渡って行われている可能性も考慮されている．図 3 は短縮 URL の判別および元 URL の復元処理の流れを示している．

3.3.2 ドメイン登録日検索システムへの問い合わせ

前述の処理で抽出したアクセス先の URL のドメインを用いて，ドメイン登録日検索システムに問い合わせを行う．その結果，図 1 のように問い合わせたドメインの登録日を応答するのでこれを迷惑メール判定に用いる．

3.3.3 迷惑メール判定

ドメイン登録日検索システムから得られたドメインの登録日とメール受信日を比較して迷惑メール判定を行う．登録日が浅いドメインを含むメールほど迷惑メールである可能性が高いと判定する．

4. システムの実装と動作

本章では 3 章で示した提案システムの実装とその動作について示す．

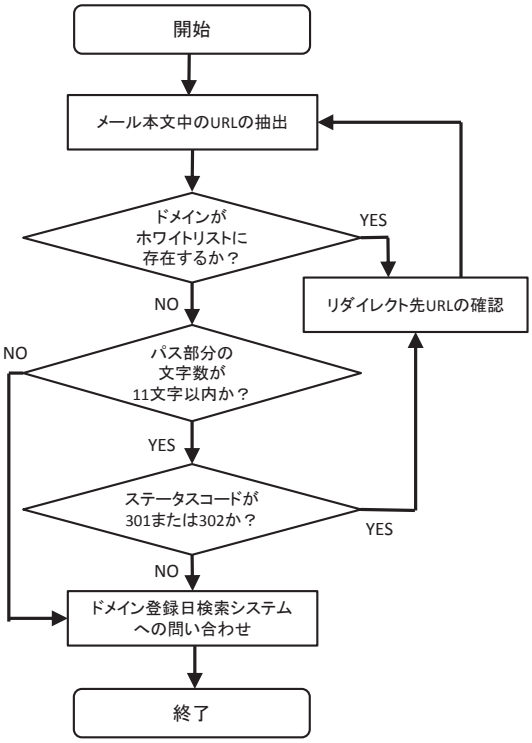

図 3 短縮 URL の復元処理の流れ
Fig. 3 The flow of short URLs restoration.

4.1 システムの実装

本システムはユーザ単位で柔軟なフィルタリングの設定が可能な迷惑メールフィルタリングのプログラムである SpamAssassin[9] のプラグインとして Perl で実装した．実装したプラグインでは，メールを受け取ると本文中に記載されている URL を抜き出して短縮 URL に関する処理を行った後，ドメイン登録日検索システムへ登録日を問い合わせ，その結果により迷惑メール判定を行う．

4.1.1 メッセージ中 URL の短縮判定と復元

ホワイトリストを用いた判定では，発見した約 200 件の URL 短縮サービスのドメインを登録している．

また，パス部分の文字数を用いた判定では，我々が調査を行った結果 11 文字が一番長い変換結果であったため本システムでは 11 文字に設定した．調査はホワイトリストに登録した URL 短縮サービスのうち，変換可能なものに対して 5 回ずつの変換を行い一番長かったパス部分のデータを集めた．

HTTP 通信を用いたステータスコードの判定では，HTTP リクエストを対象 URL を持つサーバに対して送信し，そのレスポンスのステータス行からステータスコードを取得する．ここでの HTTP 通信によるデータ要求は HEAD で問い合わせることとした．それは，今回はステータス行及びヘッダのみを利用するため，GET で問い合わせると利用しない HTTP の本文を受信しデータサイズが大きくなるためである．取得したステータスコードが 301 または 302 であった場合に短縮であると判定する．この場合，URL の

```
X-Spam-Level: *
X-Spam-Status: No, score=1.0 required=5.0 tests=DATE_CHECK,DATE_SCORE,
        DOMAIN_DATE autolearn=no version=3.3.1
```

```
X-Spam-Level:
X-Spam-Status: No, score=0.0 required=5.0 tests=DATE_CHECK,DOMAIN_DATE
        autolearn=unavailable version=3.3.1
```

図 4 試作システムによるスコア付けの例
Fig. 4 The example of scoring using the prototype system.

移動が Location ヘッダに記述されているので，Location ヘッダからリダイレクト先の URL を確認することで短縮前の URL を確認できる．

4.1.2 迷惑メール判定のためのスコア付け

SpamAssassin では複数のルールを用いて各々のルールによってスコア付けを行い，そのスコアの合計によって受信メールが迷惑メールであるかどうかを判定する．本システムではドメイン登録日検索システムの検索結果とメールの受信日を比較し，その結果に応じてスコア付けを行う．今回は，本システムが正しく動作しているかを確認するために 1 年以内のものに 1 点加点するようにした．

4.2 システムの動作

本章では前述したシステムを実際に動作させ，迷惑メール判定を行った時の実行結果ついて述べる．今回 SpamAssassin で用いるルールは本研究により作成したシステムを用いたもののみである．図 4 の上図は 1 年以内に登録されたドメイン用いた URL のリンクを本文中に含むメールに対して，本システムの実行結果としてヘッダに追加されたものである．ここで，2 行目において score=1.0 となっており，本システムによって正しく加点されていることがわかる．また，リンクを含まないメールや取得から 1 年以上経過したドメインのみを URL に含むリンクを本文中に含むメールに対しては，図 4 の下図に示す結果がメールヘッダに追加された．これらの結果より，本システムによって正しくスコア付けがされていることが確認できる．

5. むすび

本論文では，ドメイン登録日検索システムを用いた迷惑メールフィルタの提案および実装を行った．

従来のブラックリスト方式で対応できない攻撃手口に対応するため，ドメイン登録日検索システムが先行研究で開発されていたが，実際に迷惑メールフィルタとしては未実装であった．そこで，本研究によって実際にドメインの登録日を元に迷惑メールの判定を行うことができるようになった．また，短縮 URL が使用されていた場合ドメイン登録日検索システムを正しく利用できなかったため，短縮 URL から元の URL を復元し，ドメイン登録日を正しく利用できる機能も実装した．

今後の課題として，システムの性能評価を多くの迷惑メールを用いて行うことと，ドメインの登録日からの経過日数とスコアの関係を調査し適切なスコアを設定することがあげられる．また，新たな短縮 URL サービスのドメインを調査しホワイトリストへ登録していくことや，短縮 URL のパス部分の文字数の調査も継続していく必要がある．

参考文献

[1] Symantec Corporation: Symantec Intelligence Report: February 2013(online), available from ⟨http://www.symantec.com/content/en/us/enterprise/other_resources/b-intelligence_report_02-2013.en-us.pdf⟩ (参照 2013-5-14).

[2] SURBL: SURBL(online), available from ⟨http://www.surbl.org/⟩ (accessed 2012-9-20).

[3] uribl: URIBL.COM(online), available from ⟨http://www.uribl.com/⟩ (accessed 2012-9-20).

[4] PowerView Systems: ivmURI “a Domain/URI Blacklist”(online), available from ⟨http://dnsbl.invaluement.com/ivmuri/⟩ (accessed 2012-9-20).

[5] Felegyhazi, M., Kreibich, C. and Paxson, V.: On the potential of proactive domain blacklisting, Proceedings of the 3rd USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET'10), pp. 99-107 (2010).

[6] JPRS：Whois とは (online), 入手先 ⟨http://jprs.jp/info/whois/⟩ (参照 2012-9-20).

[7] ICANN: Registrar Accreditation Agreement(online), available from ⟨http://www.icann.org/en/registrars/ra-agreement-17may01.htm⟩ (accessed 2012-9-20).

[8] 松岡 政之，山井 成良，岡山 聖彦，河野 圭太，中村 素典，民田 雅人: 迷惑メール判定制度向上のためのメッセージ中 URL のドメイン登録日検索システム，インターネットと運用技術シンポジウム 2012 論文集,2012,16-22 (2012)

[9] The Apache Software Foundation: The Apache SpamAssassin Project(online), available from ⟨http://spamassassin.apache.org/⟩ (accessed 2012-9-20).

[10] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee: Hypertext Transfer Protocol – HTTP/1.1, RFC 2616, IETF (1999).