All rights are reserved and copyright of this manuscript belongs to the authors. This manuscript have been published without reviewing and editing as received from the authors: posting the manuscript to SCIS 2007 does not prevent future submissions to any journals or conferences with proceedings.

SCIS 2007 The 2007 Symposium on Cryptography and Information Security
Sasebo, Japan, Jan. 23-26, 2007
The Institute of Electronics, Information and Communication Engineers

単語情報を用いたワンクリック詐欺対策手法の提案
Proposal of countermeasure against a kind of online scams One-Click SAGI by using word frequency

名越 潤也*† 田中 英彦*
Junya Nagoshi Hidehiko Tanaka

あらまし 昨今，ワンクリック詐欺と言われるインターネット犯罪が多発している．この詐欺手口は出現が確認され始めてから決して日は浅くなく，各種機関・団体から様々な注意喚起が行われてはいるが，その被害は減少するどころか，むしろ増加傾向にある[1]．本稿ではこのワンクリック詐欺を技術的に防止することを目的として，Web ページ中の単語情報を用いた対策手法を提案する．さらにその提案した手法にて実験を行い，有効性を検証した．この一連の成果を報告する．

キーワード ワンクリック詐欺，架空請求，インターネット犯罪，フィルタリング

1 はじめに

近年のインターネットの急激な普及に伴って，インターネットを利用したサイバー犯罪が増加している．警察庁によれば，平成 17 年のサイバー犯罪等に関する相談の受理件数は 84,173 件で，前年(70,614 件)と比べて 19.2%増加している．その内訳を見てみると，詐欺・悪質商法に関するものが最も多く，全体の 49%を占める[2]．その具体的な例として「架空請求・不当請求」が挙げられている．所謂，ワンクリック詐欺[1]である．

また，IPA が 2006 年の 2 月に行ったアンケート調査[3]では，情報セキュリティの被害経験として，ウイルス感染，スパイウェア，個人情報の流出の次に，ワンクリック詐欺が挙げられている(図 1)．当アンケートは，回答者の年齢，性別，職業，IT リテラシーなどを意図的にばらつかせて，我が国のインターネット利用者全体を忠実にサンプリングすることを企図している．

上記のアンケート結果が正確に日本のインターネット利用者全体をサンプリングできていると仮定すれば，我が国の 3%以上のインターネット利用者がワンクリック詐欺の被害に遭っていることになる．

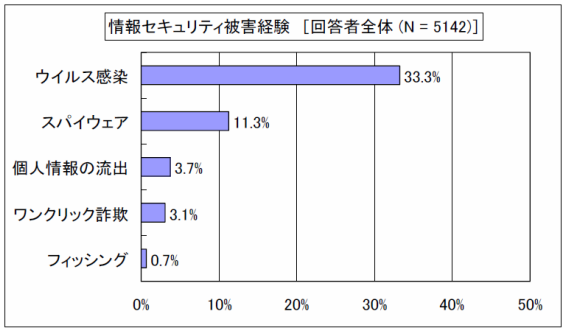

図 1: 情報セキュリティ被害経験
IPA,「情報セキュリティに関する新たな脅威に対する意識調査報告書」より

2 既存の対策

2.1 対策方法の案内

このようなワンクリック詐欺に対して，警視庁では基本的に，無視するかあるいは最寄りの警察署に相談するように案内している[4]．しかしながら，そもそもワンクリック詐欺というものを認識していなければそういった対応を取ることは困難であり，またアダルトや出会い系といったサイトの特性上，他人に相談することを躊躇し，結果的に料金を支払ってしまうことも考えられる．

また，警視庁以外にも様々な団体が注意喚起を行っているものの，一向に被害が減少する様子はない．したがって，まずこのようなサイトにアクセスさせない，あるいはアクセスした先がワンクリック詐欺であるということが自動的に判別できる仕組みが必要である．

* 情報セキュリティ大学院大学，〒 221-0835 神奈川県横浜市神奈川区鶴屋町 2-14-1
Institute of Information Security, 2-14-1 Thuruyacho, Kanagawa-ku, Yokohama-shi, Kanagawa, 221-0835, Japan.
† ニフティ株式会社，〒 140-8544 東京都品川区南大井 6-26-1 大森ベルポート A 館
NIFTY Corporation, Omori Bellport A, 6-26-1 Minami Oi, Shinagwa-ku, Tokyo, 140-8544, Japan.
1 ワンクリック架空請求・不当請求とも言われるが，本稿では「ワンクリック詐欺」と表記する．

2.2 フィルタリングソフト

技術的な対策としては，フィルタリングソフトの利用が考えられる．フィルタリングソフトにおいて，そのフィルタリング技術の中核を成すものは，人的に収集・分類・精査されたアクセス禁止 URL リスト(以下 BL)のデータベースである[5]．各クライアント PC 又は，プロキシーサーバにインストールされたフィルタリングエンジンは，HTTP リクエストが発生するたびにそのアクセス先が BL データベースに含まれるかどうか確認し，アクセスを許可/拒否する．

しかしながら，ブラックリスト方式による判別には，以下の問題がある．

1) ブラックリストへの登録漏れ
2) 詐欺サイト出現からブラックリストへの登録までのタイムラグ

したがって，このような問題を解決するためにも，ブラックリスト方式とは異なる，リアルタイムの判別手法が必要である．

3　ワンクリック詐欺の分析

3.1 ワンクリック詐欺の定義

ワンクリック詐欺の対策方法を検討する上で，本稿ではまずワンクリック詐欺を以下のように定義する．

「ワンクリック詐欺とは，アダルトサイトや出会い系サイト等において，利用者に契約の意思がないにも係わらず，一方的に契約の成立を主張し，不当な料金請求を行うもの」

3.2 海外の状況

ワンクリック詐欺は日本特有の問題であり，海外ではこのような犯罪は発生していない．その原因として考えられるのが，文化・慣習の違いである．例えば，米国のような訴訟社会で，詐欺の振込み口座を公開すると，詐欺犯が逆に訴えられる可能性がある．

3.3 ワンクリック詐欺の特徴

ワンクリック詐欺について，その特徴を表 1 に詳述する．尚，この条件を全て満たしているのがワンクリック詐欺というわけではなく，１つのワンクリック詐欺について，この内いくつかの特徴が複数確認できる．

表 1: ワンクリック詐欺の特徴

1	アダルトサイトや出会い系サイトにおいて画像や入りロボタンなどを 1 度クリックしただけで，不当な料金請求をする．
2	契約内容が非常に確認しづらい．あるいは一方的に「契約」の成立を主張した上で，「契約」後に規約等が表示される．
3	PC や携帯電話宛の迷惑メールに URL が記載されており，クリックすると一方的に「契約」の成立を主張する．また，URL にはメールアドレスがエンコードされて含まれており，どのメールアドレスからクリックされたかが判別できるようになっている．
4	個人情報を取得する過程を連想させるようなアニメーションを表示する．
5	IP アドレスや契約プロバイダ情報等を表示し，個人情報が取得可能であると主張する．
6	料金は一般的に高額であるが，まったく支払えない範囲ではない．
7	「期限内に料金を支払わないと高額な延滞料金が加算される」，「債権回収業者に債権譲渡する」，「自宅，職場に取りたてに行く」等の不安を煽る文言
8	クリックして即契約成立ではなく，一度ポップアップで規約のようなものを表示する場合もある．
9	スパイウェアを動画ファイルと偽ってダウンロードさせ，料金支払に関する警告を繰り返し表示したり，メールアドレスが格納されたアドレス帳ファイル等を盗み出したりする．

3.4 ワンクリック詐欺の分類

一口にワンクリック詐欺と言っても，様々なものが存在する．本稿ではその手口の類型から，次の３つに分類した(表 2)．

表 2: ワンクリック詐欺の分類

①	[典型的ワンクリック型] 画像や入りロボタンなどをクリックすると ⇒ 個人情報取得アニメーション ⇒ 料金請求画面 が表示される典型的なワンクリック詐欺． 迷惑メールから誘導するものも多い．
②	[ツークリック型] 画像や入りロボタンなどをクリックすると一度ポップアップウィンドウにて利用規約のようなものを表示する．非常に細かい文字が大量に記述されており，一般的には読み飛ばしてしまうことが多いと思われる．その後は①と同様．
③	[スパイウェア型] スパイウェアを動画ファイルと偽ってダウンロードさせる．そのファイルのアイコンは Windows Media Video のものだが，拡張子は「.exe」となっており，実行可能ファイルである．実行してしま

	うとスパイウェアがインストールされ，料金支払に関する警告を繰り返し表示したり，メールアドレスが格納されたアドレス帳ファイル等を盗み出したりされる.

上記の内③スパイウェア型については，市販のウイルス対策ソフトやスパイウェア対策ソフトにて，そのスパイウェアファイルを検知及び駆除できるものが多い．したがって，次節以降では特に①典型的ワンクリック型並びに②ツークリック型に注力して，その対策方法を検討する.

3.5 詐欺の成立要件

ワンクリック詐欺が詐欺として成立するのは，以下の要件が存在するためだと思われる.

1) サイトの特性

アダルトや出会い系といったサイトの特性上，詐欺に遭っても，知人や警察などに相談しにくい.

2) 個人特定という錯誤

詐欺サイトでは個人情報を取得していると思わせるようなアニメーション画像が使用されていることが多い．また，プロバイダに照会すると記述しているものもあり，これらから個人が特定可能であると思わせている．ただし，実際にはこれらから個人を特定することはできない.

3) 料金請求の脅迫

2)で特定した個人情報から，指定期間内に料金を支払わない場合には，自宅や勤務先に連絡する，あるいは訴えるなどと脅迫的な文言を表示する.

ワンクリック詐欺のページには上記の３つの要件が存在するため，詐欺の被害が発生してしまう．したがって，これらの特徴を機械的に検知することができれば，詐欺ページを判別することができる.

4　対策手法の提案

4.1 スパムメールフィルターの応用

同じフィルタリングでも，近年では，スパムメールのフィルタリングについての研究が盛んに行われている．これはメールの内容に基づいて動的にスパムか非スパムかを判別するものである．これまでの研究では Support Vector Machine(SVM)[6]などの機械学習手法や Bayes 理論(Naive Bayes)を用いた確率モデル[7]が提案されている．とりわけ，Bayes 理論を用いたスパムメールのフィルタリング手法は，簡単な学習で高い精度でスパムメールを判定するため，最近では多くのフィルタリングツールにおいて採用されている.

そこで本稿では，前述した Bayes 理論によるスパムメールのフィルタリング手法を応用した，Web ページ中の単語情報に基づくワンクリック詐欺の判別手法を提案する.

4.2 判別モデル

Bayes 理論を用いたスパムメールの判別手法にも様々な実装方法が存在するが，その中でも本研究では Graham の提案している手法[8][9]をモデルとした.

4.3 前提

・「特定の単語」はワンクリック詐欺ページ(以下詐欺ページ)に高頻度に出現する.
・「それ以外の単語」は非詐欺ページに高頻度に出現する.

したがって，新規にリクエストしたページに含まれる単語を，過去に閲覧した詐欺，非詐欺ページに含まれている単語と比較することにより，自動的にそのページが詐欺ページかどうかが判別可能となる.

4.4 コーパスの作成

新規にリクエストしたページ内の単語と過去に閲覧したページ内の単語を比較するため，詐欺ページ，非詐欺ページそれぞれの，単語の出現頻度データベースを作成しておく(以下詐欺コーパス,非詐欺コーパスと呼ぶ)．尚，何を持って「単語」とするかは一概に決定することは困難ではあるが，本研究では ChaSen[2]にて形態素解析を行った結果をもって単語とする.

4.5 リクエストされたページの判別

新規にページがリクエストされると，以下の手順でそのページが詐欺ページである確率を求める.

1. ページ内の各単語の詐欺確率を求める

ページ内に出現する各単語w_iについて非詐欺コーパス中にw_iが出現した回数をg_i，詐欺コーパス中に出現した回数をb_i，非詐欺コーパス中のページの総数を ngood，詐欺コーパス中のページの総数を nbad とすると，単語w_iが含まれるページが詐欺である確率 P(w_i)を次式(1)で求める.

$$P(w_i) = \begin{cases} \dfrac{\min(1.0, \frac{b_i}{nbad})}{\min(1.0, \frac{g_i}{ngood}) + \min(1.0, \frac{b_i}{nbad})} & (g_i + b_i > tw) \\ 0.5 & (g_i + b_i <= tw) \end{cases} \quad (1)$$

[2] http://chasen.naist.jp/hiki/ChaSen/

尚，詐欺コーパス，非詐欺コーパスを通じての単語の最低出現回数の閾値を tw とし，出現回数が閾値 tw 以下の場合は，単語の詐欺確率 P(w_i) を一律 0.5(= 詐欺でも非詐欺でもない) としている．

また P(w_i) の下限は 0.001，上限は 0.999 とする．

2. ページの詐欺確率を求める

1 で計算した各単語の詐欺確率から，そのページに特徴的な単語(P(w_i) が 0.5 から離れている) 単語を上位 n 個抽出し，その結合確率によって，ページが詐欺である確率 P(D) を次式(2)により求める．

$$P(D) = \frac{\prod_{i=1}^{n} P(w_i)}{\prod_{i=1}^{n} P(w_i) + \prod_{i=1}^{n} (1 - P(w_i))} \qquad (2)$$

3. 判定

詐欺として判別する判別確率閾値 tp を設定し，2 で求めた確率 P(D) が tp 以上であれば詐欺ページ，tp 未満であれば非詐欺ページと判定する．

5 実験

本節では，前節において述べた判別手法の評価実験を行い，本手法の有効性を検証する．

5.1 実験データ

本実験では詐欺ページの実験データとして，ワンクリック詐欺ページをデータベースとして公開しているサイト[3]から 160 件，非詐欺ページの実験データとして，Open Directory Project の日本語のカテゴリ(非アダルト)及び，Yahoo! Japan のアダルトのカテゴリより 200 件を使用した．その内，詐欺ページは 80 件を学習データ，80 件を検証データとして用い，非詐欺ページは 100 件を学習データ，100 件を検証データとして用いた．

5.2 実験方法

まず，指定された URL からページ内容を取得し形態素解析した後，その単語の出現頻度を自動的にコーパスに追加するプログラムを作成した．そのプログラムに学習データ（詐欺ページ 80 件，非詐欺ページ 100 件）を投入し，詐欺コーパス及び，非詐欺コーパスを生成した．尚，ワンクリック詐欺のページでは，入り口となるページだけでなく，クリックした後の料金請求ページにおいても特徴的な単語が多く出現するため，その 2 つのページを合わせたものを 1 件として登録することとした．

検証においても，指定された URL からページ内容を取得し，形態素解析した後，前節の手法にて詐欺ページかどうか判断し，その結果をデータベースに登録するプログラムを作成した．

5.3 定数パラメータの決定

提案手法の計算式中の各定数パラメータについて，5.2 で説明した検証プログラムを用いて様々な値を検証した結果，以下の組み合わせが最も妥当であった．

- 単語最低出現回数閾値 tw = 2
- ページ詐欺確率の計算の使用単語数 n = 10

また判別確率閾値 tp についても，様々な値を検証した(図 2)．この結果から，判別確率閾値 tp の値は 0.95 が最も妥当であることが分かった．尚，tp の値を 0.95 以上に設定した場合でも，1.0 までほとんど結果は変化しない．

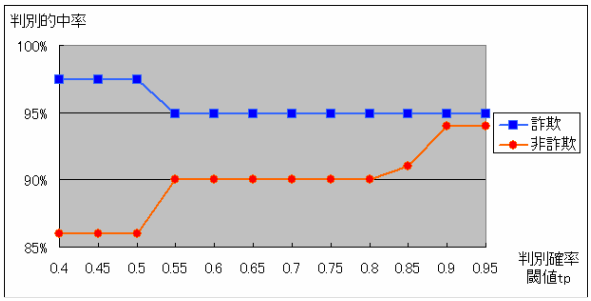

図 2: 判別確率閾値と判別的中率の変化

5.4 実験結果

各定数パラメータを決定した後，5.2 の検証プラグラムに実験データを投入した結果を以下に示す(表 3)．

表 3: 提案手法の実験結果

	詐欺ページ	非詐欺ページ
詐欺ページとして判定	76 件	6 件
非詐欺ページとして判定	4 件	94 件
判別的中率	**95%**	**94%**

5.5 考察

今回の実験結果では詐欺ページの判別的中率が 95%，非詐欺ページの判別的中率が 94%と，比較的高い精度でワンクリック詐欺ページを判別することができた．しかしながら，実運用を考えた場合には非詐欺ページの判別的中率が 94%というのは必ずしも満足できる精度とは言えないだろう．

今後，本稿で提案した手法の判別精度をさらに高める

[3] ネット詐欺相談室，http://fraud.oops.jp/
ネット詐欺相談室，http://fraud.oops.jp/ など

には，以下の 2 つの方法が考えられる．

1. データベースの拡充

Graham のベイジアンフィルターの実験では，スパム・非スパムそれぞれ 4,000 通のメールを学習データとして用いており，その結果，スパムは 99.5%，非スパムは 99.7%という高い判別精度を実現している[9]．一方，今回の実験で用いた学習データは詐欺が 80 件，非詐欺が 100 件と学習データが非常に少ない．特に非詐欺の学習データの少なさが，判別精度に影響を与えている．

例えば，今回の実験では学習データの件数が少ないために，詐欺コーパス中に存在する一般的な単語が，非詐欺コーパス中では詐欺確率が高い単語として計算されている事例がいくつか確認できた．以下に詐欺コーパス及び，非詐欺コーパス中の特徴的な単語を示す(表 4,5)．

表 4: 詐欺コーパス中の特徴的な単語(一部抜粋)

詐欺確率の高い単語（正判別に影響する）		詐欺確率の低い単語（誤判別に影響する）	
単語	詐欺確率	単語	詐欺確率
漏洩	0.9999	期	0.0188
漏れる	0.9999	チャット	0.0273
裏腹	0.9999	アンケート	0.0344
履行	0.9999	所沢	0.0344
利率	0.9999	家	0.0427
預金	0.9999	県	0.0442
誘う	0.9999	シティ	0.0537
優先	0.9999	スタッフ	0.0537
名義	0.9999	フリー	0.0537
無銭	0.9999	構造	0.0561
無事	0.9999	エンジェル	0.0588
無効	0.9999	書く	0.0588
民法	0.9999	平成	0.0617
未払い	0.9999	イベント	0.0649
未納	0.9999	チック	0.0649

表 5: 非詐欺コーパス中の特徴的な単語(一部抜粋)

詐欺確率の低い単語（正判別に影響する）		詐欺確率の高い単語（誤判別に影響する）	
単語	詐欺確率	単語	詐欺確率
あした	0.0001	入金	0.9900
あすみが丘	0.0001	振込	0.9900
あれ	0.0001	支店	0.9900
いち	0.0001	銀行	0.9900
いちばん	0.0001	以内	0.9900
いつも	0.0001	勤務	0.9900
いらっしゃる	0.0001	期限	0.9900
うーん	0.0001	法律	0.9898
うちわ	0.0001	自動的	0.9897
うらら	0.0001	後払い	0.9888
ええ	0.0001	プロバイダー	0.9870
おかげ	0.0001	法的	0.9854
おそい	0.0001	依頼	0.9842
おそらく	0.0001	放題	0.9803
おとなしい	0.0001	同意	0.9803

上記表 5 では，金融関係の一般的な単語が，詐欺確率の高い単語として判別されている．このような問題は，学習データを拡充することによって，ある程度改善が見込める．

2. バイアスの設定

もともと Graham のモデルでは非スパムにバイアスをかけるため，非スパムの単語出現回数を 2 倍して計算している．本稿ではこのバイアス係数は用いなかったが，詐欺を非詐欺として誤判別してしまうコストと，非詐欺を詐欺として誤判別してしまうコストが同じではないとすれば，こういったバイアスを設定することも有効である．

6 利用シナリオ

前節までで説明したワンクリック詐欺の判別手法を，以下のようなシステムに適用することによって，ワンクリック詐欺の対策方法として活用することができる．

- **ブラウザプラグイン**
  フィルタリングソフトと同様に，ブラウザのプラグインでリクエスト先の Web ページが詐欺がどうか判断し，詐欺の場合にはアクセスをブロックする．
- **プロキシーサーバ**
  各 PC のブラウザではなく，プロキシーサーバ側でリクエスト先の Web ページが詐欺かどうか判断する．
- **Web ページのクローラー**
  Web ページをクローリングする際に，詐欺と疑わしいページを収集する．収集した詐欺ページは，検索エンジンのフィルタリングや，BL データベースを補完するのに活用できる．

7 まとめ

本稿では，ワンクリック詐欺の対策手法として，Web ページ中の単語情報を用いた判別手法を提案した．さらにその実験を行い，詐欺ページで 95%，非詐欺ページで 94%という高い精度で判別を行うことができた．この提案した手法を，ブラウザのプラグインなどに実装することによって，ワンクリック詐欺の被害を防止することができると考えられる．

しかしながら，今後発生する全てのワンクリック詐欺が，本稿で提案した手法で判別可能であるとは限らない．

この提案手法では以下のような問題点が想定される.

・単語の変形

スパムメールにおいても,ベイジアンフィルターが普及するとそれに対抗して,容易に単語を抽出できないようなメールが出現した.

例「free」 ⇒ 「f_r_e_e」

・テキスト以外での記述

ページの内容が全て画像ファイルや Flash ファイル等で記述されると, 単語が抽出できない.

したがって,今後の課題としては Web ページ中の単語情報を用いた判別手法以外の,その他の判別手法も組み合わせる必要がある. その方法として現在は,Web ページで用いられているスクリプトやアニメーション画像の特徴を抽出して判別することを目指している.

参考文献

[1] IPA, 「コンピュータウイルス・不正アクセスの届出状況[11 月分]について」, http://www.ipa.go.jp/security/txt/2006/12outline.html

[2] 警察庁, 「平成 17 年中のサイバー犯罪の検挙及び相談受理状況等について」, http://www.npa.go.jp/cyber/statics/h17/h17_05.html

[3] IPA,「情報セキュリティに関する新たな脅威に対する意識調査の報告書」, http://www.ipa.go.jp/security/fy17/reports/ishiki/documents/2005_ishiki.pdf

[4] 警視庁, 「ワンクリック架空請求にご用心」, http://www.keishicho.metro.tokyo.jp/haiteku/haiteku/haiteku35.htm

[5] デジタルアーツ株式会社,「i-フィルターのテクノロジー」, http://www.daj.co.jp/filter/ifl_how.htm

[6] H. Druker. Support vector machines for spam categorization. In Proceedings of the IEEE Transaction on Neural Networks, volume 10, pages 1048-1054, 1999

[7] I. Androutsopoulos, J. Koutsias, K. Chandrinos, G. Paliouras, and C. Spyropoulus. An evaluation of naive Bayesian anti-spam filtering. In Proceedings of the Workshop on Machine Learning in the New Information Age: 11 European Conference on Machine Learning (ECML 2000), page9-17, 2000

[8] P. Graham, “A Plan for Spam” http://www.paulgraham.com/spam.html

[9] P. Graham, “A Better Bayesian Filtering”, http://www.paulgraham.com/better.html