Vol.2014-MBL-73 No.22
Vol.2014-ITS-59 No.22
2014/11/21

匿名通信システム Tor における 悪用ユーザ推定手法の精度に関する検討

宗 裕文[1]　和斉 薫[1]　横山 絵美里[1]　山場 久昭[1]　久保田 真一郎[1]　朴 美娘[2]　岡崎 直宣[1]

概要：近年，ユーザがアクセスした Web サイトが特定されてしまうことを防ぐ，匿名通信システムが注目されている．その中で最も普及しているのが The Onion Routing（Tor）である．Tor は健康相談や電子投票等の，誰がどこに送信したか，ということを知られたくない場合の情報交換に利用されることを本来の目的としている．しかし，Tor が違法行為を匿名で行う目的で悪用ユーザに利用されるケースがある．このことが，多くの善良なユーザが本来の目的で Tor を利用することを妨げることにつながっていると考えられる．そこで，本来 Tor の匿名性を低下させようとする者が行う既存のユーザ特定手法を有効に利用することができないかと考えた．我々は論文 [1] で，悪用ユーザの利用を抑制するために，悪用ユーザに利用されることの多い情報を扱う Web サイトを模擬するサイトを導入し，そのサイトと協調動作をすることで，高い確率で悪用ユーザの IP アドレスを特定する手法を提案した．本稿では，この提案手法の精度に関する調査を行う．

1. はじめに

現在，インターネットは私たちの生活に欠かせないものになっている．しかしながら，インターネットを利用する上で，通信内容を盗聴することやパケットのヘッダ情報を盗聴しユーザがアクセスした Web サイトが特定されてしまうことが問題になっている．前者の対策として暗号化技術があり，これを利用することで通信内容を秘匿することができる．しかし，後者の問題については，ユーザがアクセスした Web サイトが特定されてしまうため暗号化技術では対策にならない．したがって，インターネットにおける発信元を突き止めるような問題に対して匿名性を保護することが必要である．

そこで，通信経路の秘匿を目的として考案されたプライバシー保護技術が匿名通信システムである．匿名通信システムには Mix-Net [2] や Crowds [3] などがあるが，最も普及しているのが The Onion Routing(Tor)[4] である．Tor は健康相談や電子投票等の，誰がどこに送信したか，ということを知られたくない場合の情報交換に利用されることを本来の目的としている．しかし，Tor は違法行為を匿名で行う目的で悪用ユーザに利用されるケースがある．このことが，多くの善良なユーザが本来の目的で Tor を利用することを妨げることにつながっていると考えられる．我々は前回の論文で，おとりとなる Web サイトを導入し，その Web サイトと協調動作をすることで，悪用ユーザの IP アドレスを特定する手法を提案した．

本稿では，この提案手法の精度に関する調査を行う．そのためにまず，事前実験を行うことによりおとり Web サイトを識別するために有効となる特徴を調査する．そして，本実験でその特徴の組み合わせでいくつのおとり Web サイトを識別できるのか調査する．

[1] 宮崎大学
University of Miyazaki
[2] 神奈川工科大学
Kanagawa Institute of Technology

2. The Onion Routing(Tor)

2.1 概要

Tor とは，元々アメリカ海軍調査研究所（USNRL）により開発された，低遅延の匿名化通信技術である．Tor の一日あたりのユーザ数は約 250 万人であり，現在最も利用されている匿名化技術である [5]．Tor は複数のプロキシを経由させるオニオンルーティングと呼ばれる仮想回線接続により匿名性をもつ通信を実現している．

図 1 に Tor の全体図を示す．Tor は Tor ネットワークから無作為に選ばれた三つのプロキシ（以下，オニオンルータ）を経由し Web サイトへアクセスする多段プロキシ・システムである．ここで，三つのオニオンルータをユーザに近い側から順に，Entry オニオンルータ，Middle オニオンルータ，Exit オニオンルータと呼ぶ．Entry オニオンルータは，ユーザは分かるがユーザがアクセスしている Web

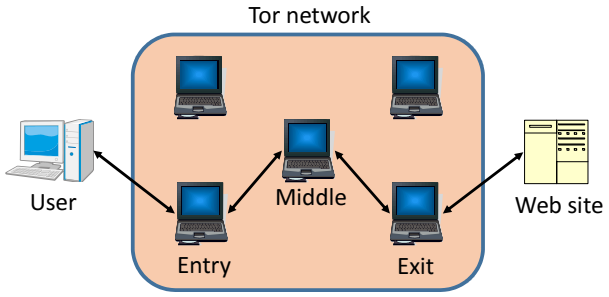

図 1　Tor の概略図
Fig. 1　Basic components in Tor

サイトは分からない．また，Exit オニオンルータは，ユーザがアクセスしている Web サイトは分かるがユーザは分からない．このようにして Tor は匿名性を確保している．Tor では，経由するオニオンルータは一定時間で切り替えられ，経由したオニオンルータを特定することは難しい．またオニオンルータ間の通信は暗号化されているため，盗聴を防ぎ安全な通信を可能としている．

2.2　Tor のユーザ

現在 Tor は軍，ジャーナリスト，警察官，人権活動家などによって様々な目的のために利用されている．例えば，ジャーナリストは，より安全に告発者や反体制派の人々と接触する為に Tor を利用する．また，人権活動家は危険地帯からの情報を発信する為に Tor を利用している．

ところが，上記のような本来の用途以外に，海外では Tor が違法薬物の取引サイトへのアクセスに使われ，また日本国内においては，殺人予告，パソコンの遠隔操作に Tor が利用されている．本稿ではこれらの違法行為を匿名で行う目的のユーザを「悪用ユーザ」，それ以外を「正規ユーザ」と呼ぶこととする．米国家安全保障局（以下，NSA），並びに日本の警察庁は悪用ユーザに対して様々な対策を行っている．例えば日本の警察庁は，Tor からのアクセスをブロックするようにサイト管理者に協力を求めている [6]．最近では，上記のように Tor が悪用されているというようなニュースが度々放送されるようになり，一般の人々は Tor に対して良い印象を持っておらず，中には Tor というものは悪いことをする為に使用するものだと思い込んでいる人がいる可能性がある．現状が続くと Tor の正規ユーザが減り，匿名通信技術自体も衰退していく恐れがある．

そこで，悪用ユーザの利用を抑制することを目的に悪用ユーザの IP アドレスを特定する手法を提案した．これにより Tor に対する印象が改善し，Tor の正規ユーザが増加することが期待できる．本研究では，Tor の匿名性を下げる目的であるユーザ特定手法を参考に提案手法を考えた．次章では Tor におけるユーザ特定手法について説明する．

3.　ユーザ特定手法

本章では既存のユーザ特定手法を紹介する．ここで紹介する手法は元々は Tor の匿名性を低下させようとする者（以下，攻撃者）によって行われる手法であるが，適用方法によっては本研究の目的にも利用できると考えられる．本稿では攻撃者により占拠されたオニオンルータを汚染オニオンルータと呼ぶこととする．

(1)　Web サイトの指紋情報を利用した手法

Web サイトにアクセスした際のトラフィックに含まれるサイト独自の特徴（以下，指紋）に着目し，これを観測することでユーザがアクセスした Web サイトを特定するという手法である．[7] は，機械学習を併用した手法である．指紋情報には，パケットの総数，HTML ファイルのサイズなど，Web サイトから抽出できるような情報を用いている．また，指紋情報の分類には，Support Vector Machines(SVM) を使用している．[7] は 54% の確率で Web サイトを特定できることが示されている．

[8] は Tor に Web サイトの指紋情報を利用した手法の対策をされたとしても有効な手法である．これは，Web サイトの指紋情報を利用した手法の対策としてトラフィックに様々な加工を施された場合でも，それらの加工を打ち消すトラフィック逆加工を行うことで，対策を無効化するものである．[8] では，Web サイトの指紋情報を利用した手法の対策がされた Tor に対してもユーザがアクセスした Web サイトの特定が可能であり，75% の確率で Web サイトを特定できることが示されている．

Web サイトの指紋情報を利用した手法は汚染 Entry オニオンルータだけ用意すればよく，実現可能性が高い．しかし，本稿の目的である悪用ユーザの IP アドレスを特定し，利用を抑制するためには，さらに高い特定率が望まれる．

(2)　特徴的なトラフィックを利用した手法

汚染 Entry オニオンルータと汚染 Exit オニオンルータを用意し汚染 Exit オニオンルータが特徴的なトラフィックをユーザへ送信し，そのトラフィックを汚染 Entry オニオンルータが観測することでユーザを特定する手法である．[9] では，Exit オニオンルータが，トラフィックのパケット数を変化させることで信号を含めユーザへ送信する．信号を含んだトラフィックを，Entry オニオンルータが認知することで，Web サイトへアクセスしたユーザを特定することができる．[9] は 65% から 100% の確率で Web サイトを特

定できることが示されている．

[10] は Exit オニオンルータがトラフィックに直接拡散方式の疑似ノイズを含ませることによって特徴的なトラフィックにしている．疑似ノイズを用いることで，攻撃が行われているかどうかの判断が難しいため，対策が困難となる．

特徴的なトラフィックを利用した手法はユーザがアクセスした Web サイトを特定する確率は高いが汚染オニオンルータを二つ用意しなければならず実現可能性が低い．

4. 提案手法

本章では，4.1 で提案手法の概要を述べ，4.2 で提案手法における前提条件について説明する．さらに 4.3 で提案手法の流れを述べ，4.4 で提案手法の具体的なアルゴリズムを説明する．

4.1 概要

本提案手法は，悪用ユーザを推定するための有力な手がかりを与える手段を提供することを目的とする．具体的な手法は悪用ユーザがアクセスしそうなおとりとなる Web サイト（以下，おとり Web サイト）を導入し，その Web サイトと Entry オニオンルータが協調動作し，特徴的なトラフィックを悪用ユーザに送信する．このことにより，特徴的なトラフィックを利用した手法の実現可能性が低いというデメリットを解決し，実現可能性が高く，高い確率でおとり Web サイトにアクセスした悪用ユーザの IP アドレスを特定することを目指す．

4.2 前提条件

Entry オニオンルータは Web サイトの指紋情報を利用した手法における汚染オニオンルータと同様の役割を持ち，情報を抽出できるものとする．また，おとり Web サイトは公開鍵認証基盤で認証されていないものを作成する．これは，悪用ユーザがアクセスするような Web サイトは認証されておらず，また，正規ユーザはウイルスの危険性がある認証されていない Web サイトにはアクセスしないと考えたからである．ここから，おとり Web サイトは悪用ユーザのみをを対象とし，正規ユーザはおとり Web サイトへアクセスしないものとする．

4.3 提案手法の流れ

以下で図 2 を用いて本提案手法の流れを説明する．ここで図 2 の Suppress OR とは悪用ユーザを抑制したい立場の者が Entry オニオンルータに位置したオニオンルータである．

(1) おとり Web サイトは悪用ユーザからアクセス要求がきたことを確認する．

(2) おとり Web サイトはパケットキャプチャを開始する．

(3) Suppress OR にパケットキャプチャを開始するように指示する．

(4) Suppress OR はパケットキャプチャを開始する．

(5) 悪用ユーザへ応答を返す．

(6) おとり Web サイト側のキャプチャデータと Suppress OR 側のキャプチャデータを比較して悪用ユーザの IP アドレスを特定する．

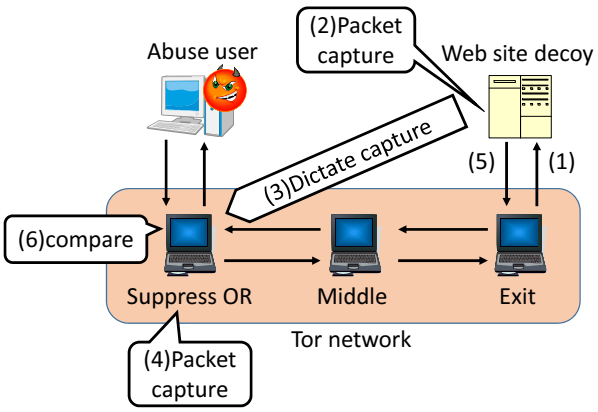

図 2 提案手法の流れ
Fig. 2 Flow of the proposed method

4.4 動作手順

提案手法の動作手順はおとり Web サイトを作成するおとり Web サイト作成フェーズ，図 2 の (3)，(4) の処理に当たる協調動作フェーズ，図 2 の (6) の処理に当たる悪用ユーザ決定フェーズの三つに分けられる．

以下でそれぞれのフェーズについて詳しく説明する．

I おとり Web サイト作成フェーズ
おとり Web サイトには現実の Web サイトと区別をつけるために特定の信号を含ませる．図 2 の Suppress OR でこの信号を受け取ることで悪用ユーザが対応するおとり Web サイトを利用したことを判断する．ここで信号は，図 3 のように Web サイト本来の HTML 及び依存コンテンツを送信した後に，特定の間隔で遅延させたダミーコンテンツを複数回付加して送信することで実現する．付加するダミーコンテンツの量とその遅延時間については以下のように定める．
まず，ダミーコンテンツの数 Num とそれぞれのサイズ $Size_i(i{=}1,\cdots,Num)$ を定義する．次に，それぞれのダミーコンテンツを送信する待ち時間 $Time_i$（秒）を設定する．そして，おとり Web サイト本来のコンテンツを送信した後，それぞれのダミーコンテンツを

Vol.2014-MBL-73 No.22
Vol.2014-ITS-59 No.22
2014/11/21

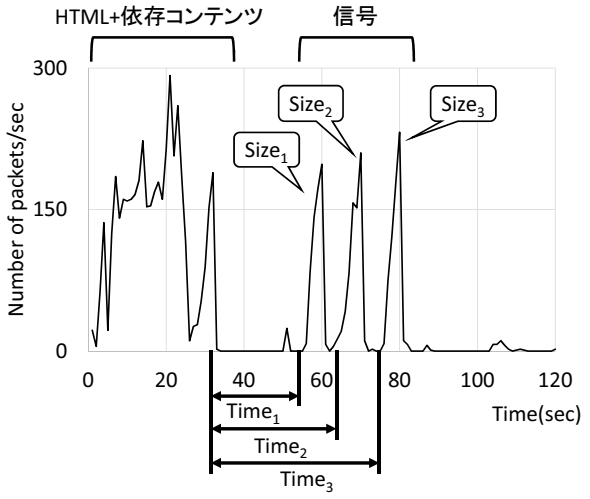

図 3　おとり Web サイトの通信トラフィック
Fig. 3　Communication traffic of the Web site decoy

$Time_i$ だけ待って送信する．
図 3 は Num=3，$Size_1$=$Size_2$=$Size_3$=300(KB)，$Time_1$=30，$Time_2$=40，$Time_3$=50 と設定した場合の信号である．

II　協調動作フェーズ
協調動作フェーズでは 4.3 の「提案手法の流れ」における (1) から (3) までの動作を行う．

III　悪用ユーザ決定フェーズ
協調動作フェーズで収集した悪用ユーザと図 2 の Suppress OR 間のキャプチャデータと Exit オニオンルータとおとり Web サイト間のキャプチャデータを比較する．そして，類似していれば悪用ユーザはおとり Web サイトへアクセスしたことが分る．図 4 の実線は，おとり Web サイトにアクセスした際のおとり Web サイトで観測したパケットを表したものである．点線は，図 2 の Suppress OR で観測した結果を表したものである．ここで，図 4 の二つのデータを見ると Suppress OR で観測したデータに遅れが生じていることが分かる．このまま，相関を調べても期待する結果は得られない．そこで二つの値を比較する指標として相互相関係数を用いる．相互相関係数とは二つのデータ間の類似性の度合いを示す指標である．相互相関係数が 1 に近いほど相関があることを表している．相互相関係数は式 1 で表される．ここで，$f_1(t)$，$f_2(t+t_u)$ はそれぞれ図 2 の suppress OR 側で収集したデータ，おとり Web サイト側で収集したデータを表しており，それらの”ずれ”を $t_u(\in [T_l, T_h])$ で表す．T_l，T_h は t_u の下限と上限である．そして，$f_1(t)$，$f_2(t+t_u)$ の平均値をそれぞれ avg_1，avg_2 と表す．また，N は観測したデータのサンプリング数である．そして，おとり Web サイトで観測したデータを T_l と T_h の間で t_u 間隔で相関係数を求め，その最良値を当サイトの相関係数の値とする．さらに，収集したデータの HTML 及び依存コンテンツの部分をマスク処理することで相互相関係数の増加を図る．マスク処理は，収集したデータの時間経過に対するパケット数の変化を見ていき，パケット数が 0 の時間を区切りとし，それより以前のパケット数を全て 0 に置き換える．図 5 は図 4 を実際にマスク処理したデータを表している．

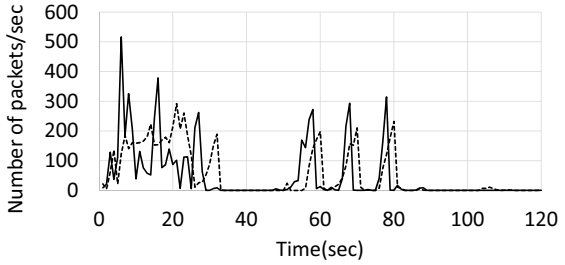

図 4　Suppress OR 及びおとり Web サイト側で収集したデータ
Fig. 4　The data graph which collected at the Web site decoy side and Suppress OR side

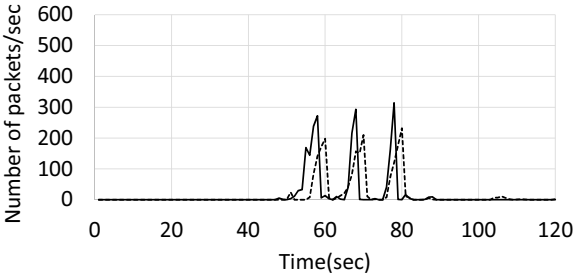

図 5　図 4 にマスク処理した
Fig. 5　Graph was masking Figure 4

$$r(t_u) = \frac{\sum_{t=1}^{N}(f_1(t) - avg_1)(f_2(t+t_u) - avg_2)}{\sqrt{\sum_{t=1}^{N}(f_1(t) - avg_1)^2}\sqrt{\sum_{t=1}^{N}(f_2(t+t_u) - avg_2)^2}} \quad (1)$$

5. 評価実験

本章では，おとり Web サイトを識別するために有効な特徴の組み合わせで，識別できるおとり Web サイトの数を調査する．まず，事前実験によっておとり Web サイトの識別に有効となる特徴を調査する．以下，5.1 と 5.2 で各実験で用いる評価指標，実験環境について説明する．そして，5.3 で事前実験を行い，その結果を元に 5.4 で本実験を行う．

5.1 評価指標

本稿では Web サイトにアクセスした悪用ユーザの IP アドレスを特定することが目的であるため，全体特定率と Web サイト特定率で評価を行う．全体特定率とは，各 Web

表 1　実験環境
Table 1　Experiment environment

CPU	Core(TM)i7-4770 CPU 3.40GHz
OS	Windows 7 Pro
Browser	Mozilla FireFox 28.0
Tor のバージョン	v0.2.3.25
Perl のバージョン	ActivePerl 5.16.3
Apache	v2.4.9

サイトへのアクセス回数に対する，アクセスした Web サイトの特定正解総数の割合である．また，Web サイト特定率とは，ある Web サイトへのアクセス回数に対する，アクセスした Web サイトの特定正解数の割合である．本実験では全体特定率を Esr，Web サイト特定率を Wsr_i とし，それぞれの算出方法を式 2 に示す．ここで，$Success_i$，$WebSite$，$Access$ はそれぞれ，各 Web サイトの特定成功数，アクセスする Web サイトの数，アクセス回数である．

$$Esr = \frac{\sum_{i=1}^{WebSite} Success_i}{WebSite \cdot Access}, \quad Wsr_i = \frac{Success_i}{Access} \quad (2)$$

これらの特定率が高いほどユーザがアクセスした Web サイトが容易に分かることを示す．本提案手法は Entry オニオンルータを使用するためユーザがアクセスした Web サイトが判れば，その Web サイトへアクセスしたユーザも判ることとなる．それぞれの特定率を求め，それらの値で評価する．

5.2　実験環境

実験に用いた実験環境は表 1 の通りである．実験に用いる Web サイトは，Web サイトのアクセスランキング付けを行っているサイトである Alexa [11] の上位から国ドメインだけが違うだけで実質的に同じサイトとなるものなどの重複を除き 100 サイト選択した．また，著作権上の問題を回避するため現実のサイトの HTML 及びその他コンテンツサイズ，コンテンツ数を元にダミーデータからなる Web サイトを作成した．また，パケットキャプチャには wireshark [12] を用いた．

5.3　事前実験

本節では，おとり Web サイトの識別に有効となる特徴を実験により調査するために，以下の 3 つの場合に分けての事前実験を行った．

(1) おとり Web サイト毎に付加するダミーコンテンツのサイズを変える場合
ただし，$Size_1 = Size_2 = Size_3$ とする．

(2) おとり Web サイト毎に付加する各ダミーコンテンツのサイズを変える場合

表 2　事前実験 1 のパラメータ
Table 2　Parameter of preliminary experiment 1

おとり Web サイト	T_1[s]	T_2[s]	T_3[s]	$Size_{1,2,3}$[kb]
A（動画）	30	40	50	700
B（検索）	20	30	40	600
C（ショッピング）	10	20	30	300
D（企業 HP）	10	20	30	400
E（ニュース）	20	30	40	500

表 3　事前実験 2 のパラメータ
Table 3　Parameter of preliminary experiment 2

おとり Web サイト	$Size_1$[kb]	$Size_2$[kb]	$Size_3$[kb]
A（動画）	300	300	300
B（検索）	300	600	900
C（ショッピング）	600	600	600
D（企業 HP）	900	900	900
E（ニュース）	900	600	300
おとり Web サイト	T_1[s]	T_2[s]	T_3[s]
A（動画）	30	40	50
B（検索）	20	30	40
C（ショッピング）	10	20	30
D（企業 HP）	10	20	30
E（ニュース）	20	30	40

(3) おとり Web サイト毎に付加するダミーコンテンツの送信間隔を変える場合

5.3.1　実験方法

本研究では，動画，検索，ショッピング，企業 HP，ニュースサイトの 5 種類から選択しておとり Web サイトに見立てた擬似 Web サイトを作成する．
指定した URL をブラウザに入力すると，Tor プロキシ経由で接続される．この時の通信トラフィックをユーザの入出力に加え，Web サイトの入出力でもパケットキャプチャを行う．提案手法では，おとり Web サイトに悪用ユーザのみがアクセスしたと仮定している．本実験では $WebSite = 100$，$Access = 10$ とし，各事前実験で用いたパラメータを表 2，3，4 に示す．ここで，$Time_i$ は T_i とする．また，T_l，T_h はそれぞれ，代表的な Web サイトを調べた結果により，1 秒，3 秒，t_u は 1 秒とした．そして，5 つのおとり Web サイト及び 95 サイトのそれぞれに 10 回ずつアクセスした 1000 個のパケットキャプチャデータを用いる．そして，提案手法では全てのおとり Web サイトのキャプチャデータから相互相関係数 r 求め，r が最も大きい Web サイトが対応するおとり Web サイトかどうか判断する．そして，全体特定率及び Web サイト特定率を求める．相互相関係数の算出には R 言語の ccf 関数 [13] を用いた．

5.3.2　結果

表 5 は各事前実験の全体特定率を表している．表 5 より事前実験 2 は他と比べて全体特定率が高くないことが分か

表 4　事前実験 3 のパラメータ
Table 4　Parameter of preliminary experiment 3

おとり Web サイト	T_1[s]	T_2[s]	T_3[s]	$Size_{1,2,3}$[kb]
A（動画）	30	40	50	300
B（検索）	20	40	50	300
C（ショッピング）	10	30	50	300
D（企業 HP）	10	20	40	300
E（ニュース）	10	20	50	300

表 5　各事前実験の結果
Table 5　The result of each prior experiment

事前実験	全体特定率 [%]
1	96
2	78
3	96

る．つまり，各ダミーコンテンツのサイズを異なるものにしたとしてもおとり Web サイトの識別に有効ではない．一方，事前実験 1，3 は高い全体特定率を示しているため，ダミーコンテンツのサイズと送信間隔の二つの特徴がおとり Web サイトの識別に有効であることが分かる．

5.4　本実験

本節では，事前実験の結果を元に各特徴の組み合わせで，識別できるおとり Web サイトの数を調査するために，以下の二つの場合において実験を行い，

(1)　ダミーコンテンツのサイズを 2 通り，送信間隔を 2 通りで構成される組み合わせの場合

(2)　ダミーコンテンツのサイズを 3 通り，送信間隔を 3 通りで構成される組み合わせの場合

5.4.1　実験方法

本実験では，付加するダミーコンテンツの数は二つとした．各パラメータの設定方法を以下に示す．ダミーコンテンツのサイズ $Size_{1,2}$ を設定するために，まず全体のダミーコンテンツのサイズの上限値 MAX_Sall を決める必要がある．MAX_Sall は，悪用ユーザに怪しまれない程度のサイズにするために，各おとり Web サイトにおける HTML 及び依存コンテンツの合計サイズの中央値とした．次に，MAX_Sall をおとり Web サイトに付加するダミーコンテンツの数 Num で割ることで，ひとつのダミーコンテンツのサイズの上限値 MAX_S を求める．最後に，MAX_S を幾通りかに分割し $Size_{1,2}$ のパラメータとして設定する．

ダミーコンテンツの待ち時間 T_2 の設定するために，まずダミーコンテンツの送信時間の上限値 MAX_T を決める必要がある．MAX_T は，各おとり Web サイトにおける HTML 及び依存コンテンツの送信時間の中央値とした．次に, MAX_T から全体のダミーコンテンツのダウンロード時間 ADT を差し引き，ダミーコンテンツの送信間隔の最大値 $MAX\_\Delta T$ を求める．そして，$MAX\_\Delta T$ を幾通りかに分割し，分割した値 ΔT_i に $Size_1$ のダウンロード時間 DT_1 と T_1 を加えたものを T_2 のパラメータとして設定する．

表 6　本実験 1 のパラメータ
Table 6　Parameter of main experiment 1

おとり Web サイト	$Size_1$[kb]	$Size_2$[kb]	T_1[s]	T_2[s]
A（動画）	1150	1150	5	19
B（検索）	1150	1150	5	29
C（ショッピング）	2300	2300	5	23
E（ニュース）	2300	2300	5	33

表 7　本実験 2 のパラメータ
Table 7　Parameter of main experiment 2

おとり Web サイト	$Size_1$[kb]	$Size_2$[kb]	T_1[s]	T_2[s]
A1（動画）	700	700	5	15
A2	700	700	5	22
B1（検索）	700	700	5	29
B2	1400	1400	5	17
C1（ショッピング）	1400	1400	5	24
C2	1400	1400	5	31
D1（企業 HP）	2100	2100	5	19
E1（ニュース）	2100	2100	5	26
E2	2100	2100	5	33

調査の結果，$MAX\_Sall = 4600$[kb]，$MAX\_T = 36$[s] だった．本実験では，付加するダミーコンテンツを二つとするため $MAX\_S = 2300$ となる．よって，ダミーコンテンツのサイズを 2 通り用意する場合 $1150kb$ と $2300kb$ となる．また，$ADT = 16$ とすると $MAX\_\Delta T = 20$ となる．よって，ダミーコンテンツの送信間隔を 2 通り用意する場合，10s と 20s となる．本実験 1，2 で用いたパラメータをそれぞれ，表 6，7 に示す．本実験 1 では，おとり Web サイトが 4 サイトとなるため，悪用ユーザがアクセスしない可能性が高い企業 HP のおとり Web サイトを除外した．本実験 2 では，おとり Web サイトが 9 サイトとなるため，企業 HP を除く 4 種類の Web サイトから二つずつ選択し，企業 HP の Web サイトを 1 つ選択した．実験方法については事前実験と同様に，4 つのおとり Web サイト及び 96 サイトのそれぞれに 10 回ずつアクセスした 1000 個のパケットキャプチャデータを用いる．そして，全体特定率及び Web サイト特定率を求める．全体特定率が 100% の場合のみ，おとり Web サイトを識別できるものとする．

5.4.2　結果と考察

表 8，9 はそれぞれ，本実験 1，2 の各おとり Web サイトの特定率を表している．表 8 より，本提案手法は 4 通りの組み合わせでおとり Web サイトを識別できることが分かった．しかし，9 通りの組み合わせでおとり Web サイトを識別できなかった．これは，本実験で使用したダミーコ

表 8　本実験 1 の結果
Table 8　The result of main experiment 1

おとり Web サイト	Web サイト特定率 [%]
A（動画）	100
B（検索）	100
C（ショッピング）	100
E（ニュース）	100

表 9　本実験 2 の結果
Table 9　The result of main experiment 2

おとり Web サイト	Web サイト特定率 [%]
A1（動画）	100
A2	100
B1（検索）	100
B2	100
C1（ショッピング）	100
C2	100
D1（企業 HP）	90
E1（ニュース）	100
E2	70

ンテンツのサイズと送信間隔の上限値が小さく，それを 3 等分することにより差が現れなかったのが原因だと考えられる．つまり，本実験より上限値を高くすれば，9 通りの組み合わせでおとり Web サイト 9 サイト識別することは可能だと考えられる．しかし，悪用ユーザに怪しまれないサイズ及び送信間隔にしなければ現実的ではないため，適切な上限値の調査も必要となる．

その他，考えられる原因はマスク処理である．本提案手法では，収集したデータの時間経過に対するパケット数の変化を見ていき，パケット数が 0 となった時間を区切りとし，それより以前のパケット数を全て 0 に置き換えるというものである．もし，パケット数が 0 の時点でまだ HTML 及び依存コンテンツが送信し終わっていなかった場合，HTML 及び依存コンテンツは綺麗にマスク処理できていないことになる．こういった場合，相互相関係数は低くなる可能性がある．

6. まとめ

本稿では匿名通信システム Tor における悪用ユーザ推定手法の精度に関する検討を行った．そのために，おとり Web サイトの識別に有効となる特徴を事前実験により調査した．そして，おとり Web サイトに付加するダミーコンテンツのサイズと送信間隔の組み合わせにより，識別できるおとり Web サイトの数を本実験により調査した．その結果，9 通りの組み合わせでおとり Web サイトを識別できなかったが，4 通りの組み合わせでおとり Web サイトを識別することができた．今後はマスク処理におけるアルゴリズムの改良を行いさらに精度を高めたい．マスク処理におけるアルゴリズムの改良のための方策としては，事前におとり Web サイトと Entry オニオンルータが HTML 及び依存コンテンツの総データ量を共有することで，そのデータ量を受信した時間を区切りとし，それより以前のパケット数を 0 に置き換える方法が考えられる．

参考文献

[1] 宗裕文，横山絵美里，山場久昭，久保田真一郎，朴美娘，岡崎直宣: 匿名通信システムにおける悪用ユーザ特定手法の検討マルチメディア，分散，協調とモバイル (DICOMO2014) シンポジウム，pp.506-513

[2] David Chaum, Communications Of TheAcm, R. Rivest, and David L. Chaum, *Untraceable electronic mail, return addresses, and digital pseudonyms*, Communications of the ACM, Vol. 24, pp. 84?88, 1981.

[3] M. Reiter and A. Rubin（1998）, *Crowds: Anonymity for Web Transactions*, ACM Trans, Information and System Security, vol. 1, no. 1, pp. 66-92.

[4] Roger Dingledine, Nick Mathewson, and Paul Syverson: *Tor: The Second-Generation Onion Router*, In In Proceedings of the 13th USENIX Security Symposium,(2004).

[5] Tor Metrics Portal: Directly connecting users, available from, (`https://metrics.torproject.org/users.html`) (2014.10.27)

[6] WIRED : 日本の警察庁、匿名化ツール「Tor」のブロックをサイト管理者に促す.
available from, (`http://wired.jp/2013/04/22/japan-police-stop-using-tor/`) (2014.10.27).

[7] Andriy Panchenko, Lukas Niessen, and Andreas Zinnen: *Website Fingerprinting in Onion Routing Based Anonymization Networks*, Proceedings of the 10th annual ACM workshop on Privacy in the electronic society pp.103-114, (2011).

[8] 横手 健一・松浦 幹太 (2012): 匿名通信システム *Tor* の安全性を低下させるトラフィック逆加工, コンピュータセキュリティシンポジウム 2012 論文集 巻：2012 号：3 ページ：624-631.

[9] Zhen Ling, Junzhou Luo, Wei Yu, Xinwen Fu, Dong Xuan, and Weijia Jia: *A New Cell-Counting-Based Attack Against Tor*, Networking,IEEE/ACM Transactions on, Vol.20, Issue.4, pp.1245-1261, (2012).

[10] Wei Yu, Xinwen Fu, Steve Graham, Dong Xuan, and Wei Zhao: *DSSS-Based Flow Marking Technique for Invisible Traceback*, Proceedings of the 2007 IEEE Symposium on Security and Privacy, pp.18-32, (2007).

[11] Alexa: Alexa Top 500 Global Site, available from, (`http://www.alexa.com/topsites`) (2014.10.27).

[12] Wireshark: Wireshark, available from, (`http://www.wireshark.org/`) (2014.10.27).

[13] 山田剛史，杉澤武俊，村井潤一郎: *R* によるやさしい統計学, pp.62-64, オーム社, (2008).