2W-08

パケット解析を用いたTor通信先の識別のための実験

早川宏志† 青木太一† 佐藤直† 土井洋†
情報セキュリティ大学院大学情報セキュリティ研究科†

1. はじめに

匿名通信システムを用いると，利用者の匿名性が担保される．匿名通信システムとしてTor(The Onion Router)[1]がよく知られている．Tor では通信をする際，複数の中継ノードを経由する上に，暗号化を組み合わせることによって匿名通信を実現している．一方，Tor によって達成される強力な匿名性は，サイバー犯罪に悪用されているという実態がある．例えば，Hidden Service を利用して構築された闇市場サイトにおける不正な薬物取引[2]，インターネット掲示板への殺害予告の書き込み[3]といった事例である．本稿では通信の中継点であるルータでパケット解析をすることを想定したTorの通信先の識別方法を検討し，実験結果を報告する．

2. パケット解析を用いたアプローチ

2.1 先行研究

Tor に対する攻撃手法については様々な研究がなされている．Panchenko らの研究[4]では，ネットワーク上を流れる暗号通信から第三者がパケットの様々な特徴を利用することで，暗号文を解読せず，Tor 通信先の特定を試みている．この手法では，クライアントからTorネットワークの入口ノードまでの間のいずれかで通信パケットを観測することを想定している．そして，775個のアクセス先Webサイト中の1つを特定する実験で特定率54.61%が達成された．

2.2 本研究の概要

通信の中継点であるルータでパケット解析をした場合でも，Tor 通信におけるレスポンス時間やパケットのサイズには，Tor を利用しなかった場合における通常の通信の特徴が残ると考えられる．本研究では，識別に使う情報を絞ることで高速化と識別性の両立を目指す．その第一段階として，アクセス先Webサイトごとにパケットを取得し，レスポンス時間やパケットサイズの分布をもとにTor通信先の識別のための実験・評価を行う．

Some Experiments for Deanonymizing Tor Using Packet Analysis
†Hiroshi Hayakawa, Taichi Aoki, Naoshi Sato, Hiroshi Doi, Institute of Information Security

2.3 実験方法

Tor ブラウザ v5.0.4(based Firefox 38.4.0 esr)を利用しWebサイトにアクセスを行う．このとき，tcpdump を利用して発生したパケットを取得する．調査対象の Web サイトは，Alexa の提供するアクセスランキングから 10 サイトを選択した．選択した Web サイトは，検索サイトから 2 つ((A)https://www.google.com/ncr/, (B)http://www.yahoo.co.jp)，SNS から 2 つ((C)https://www.facebook.com, (D)http://line.me/ja)，掲示板から2つ((E)https://www.wikipedia.org, (F)http://www.2ch.net)，ニュースサイトから 2 つ((G)http://www.bbc.com, (H)http://www.nikkei.com)，ショッピングサイトから 2 つ((I)http://www.amazon.com, (J)http://www.rakuten.co.jp)である．

Tor はアクティブコンテンツによって脆弱になることが知られているため，Tor ブラウザはアクティブコンテンツ(Flash, JavaScript, Java)をOFF に設定した．通信の再現性を損なうためキャッシュをOFFに設定するとともに，自動的に通信を発生させる統計通信・セーフブラウズ機能もOFFに設定した．なお，cookieの設定はデフォルトのままとした．一方，通常の通信とTorによる通信の比較を行うために，通常のブラウザFirefox 38.4.0esrを利用したときのパケットも取得した．

今回の実験ではTorブラウザを利用しているPCと同一マシンでパケットを取得し，TSO(TCP Segmentation Offload)はOFFに設定した．

3. 実験・評価

今回の実験では，1時間おきに10サイトに対してアクセスし，パケットの取得を繰り返した．なお，繰り返し回数は20回とした．

送受信パケットやレスポンス時間について分析を行ったが，送信パケットの長さがサイトごとの特徴を得やすかった．以下，送信パケット長の分析結果を報告する．

サイト A とサイト I の送信パケットサイズを10(bytes)ごとに区切り，20 回分の平均値を求めた．結果を図 1，図 2 に示す．縦軸はパケット数の平均，横軸はパケットサイズ(bytes)である．

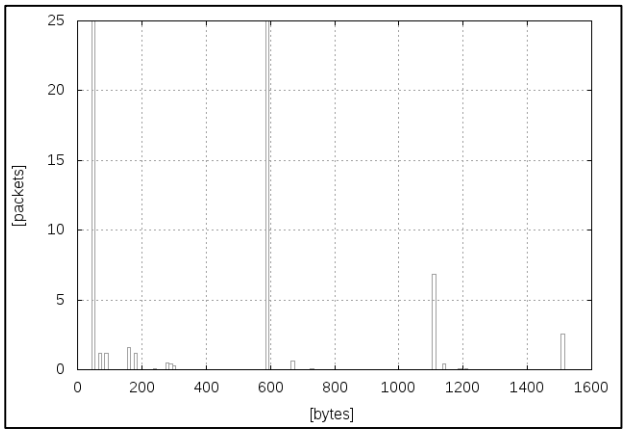

図 1：サイト A(google)のパケットサイズ分布

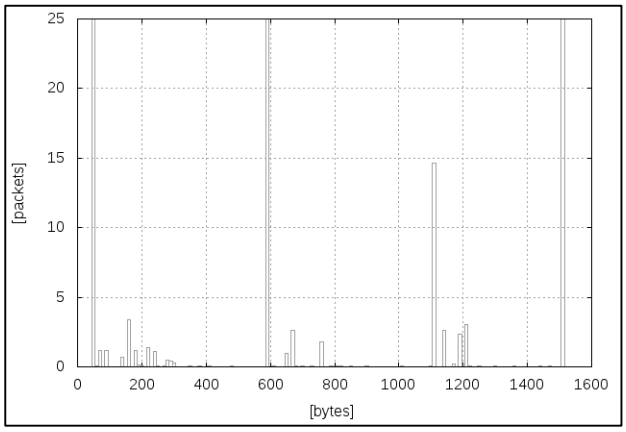

図 2：サイト I(amazon)のパケットサイズ分布

10 サイトについて，3 種類のパケットサイズ（1,110～1,119(bytes)， 1,210～1,219(bytes)，1,510～1,519(bytes)）のパケット数の平均を表 1，表 2 に示す．サイト I は 1,510～1,519(bytes)のパケットが 10 サイトの中でも突出して多い．サイト B は 1,110～1,119(bytes)が突出して多い．1,210～1,219(bytes)については，平均が 2 以上であるか否かで大別できる．このように，Web サイトによっては特定のパケットサイズにおけるパケット数に違いがあらわれている．

表 1：パケット数の平均（A,B,C,D,E）

パケットサイズ	パケット数の平均				
	A	B	C	D	E
1,110～1,119	6.9	33.7	9.6	3.8	4.8
1,210～1,219	0.1	2.2	0.0	0.1	0.7
1,510～1,519	2.6	14.1	3.4	1.3	5.3

表 2：パケット数の平均（F,G,H,I,J）

パケットサイズ	パケット数の平均				
	F	G	H	I	J
1,110～1,119	0.5	3.3	8.3	14.7	13.4
1,210～1,219	0.0	0.5	2.5	3.1	3.7
1,510～1,519	1.4	5.5	16.3	25.6	17.6

なお，図 1，図 2 を見るとサイト A とサイト I ではサイズ 1,110～1,119(bytes)， 1,210～1,219(bytes)， 1,510～1,519(bytes)のパケット数の平均に違いが出ているのが分かる．特にサイト I はサイト A に比べて 1,510～1,519(bytes)のパケット数が多いことが確認できる．

4. 考察

10 サイトとも 50～59(bytes)のパケット数の平均は大きかったが，TCP の ACK（54(bytes)）に起因すると考えられ，識別には用いなかった．また，590～599(bytes)のパケット数も 10 サイトとも多かったので，識別には用いなかった．

サイト A,C,E は HTTPS サービスを提供している Web ページであり，OCSP に関する送信が生じるが，発生数が少ないため，調査したパケットサイズ分布にあまり影響はしていないと考えられる．

サイト A とサイト I における 1,510～1,519(bytes)と 1,110～1,119(bytes)のパケット数の平均の違いの原因は HTTP の GET にあると考えられる．HTTP の GET については URL の長さ（?aaa=bbb のようなパラメータ部分を含む）の違いと cookie を利用しているか否かでパケットサイズに差が生じる．サイト I における Tor を使わない通常の通信を調べた結果から，1,110～1,119(bytes)のパケット数よりも 1,510～1,519(bytes)のパケット数の方が多いのはこの影響であると推測される．

5. まとめ

本稿では 10 サイトに対する送信パケット長の分析を行い，Web サイトによる違いについて評価した．送信パケットのみで高い識別性能が達成できれば，データ処理コストの削減の可能性がある．今後はより多くのサイトの識別可能性について評価を行いたい．また，通常の通信と Tor による通信を識別する手法も検討したい．

参考文献

[1] Tor Project, https://www.torproject.org/ (最終確認日 2016/1/1)
[2] Deep Web とサイバー犯罪 Tor ネットワーク内闇市場の実態，https://app.trendmicro.co.jp/doc_dl/select.asp?type=1&cid=118 (最終確認日 2016/1/1)
[3] 新たなサイバー犯罪に関する課題と今後の対策について，https://www.npa.go.jp/cyber/csmeeting/h24/pdf/h24_2.pdf (最終確認日 2016/1/1)
[4] Panchenko, A., Niessen, L., Zinnen, A., and Engel, T. : Website fingerprinting in onion routing based anonymization networks, WPES'11, pp.103-114 (2011).