送信パケットを用いたTor通信先の識別法の検討

A Study of Deanonymizing Tor Using Send Packets

早川宏志・ネットワーク分科会・情報セキュリティ大学院大学

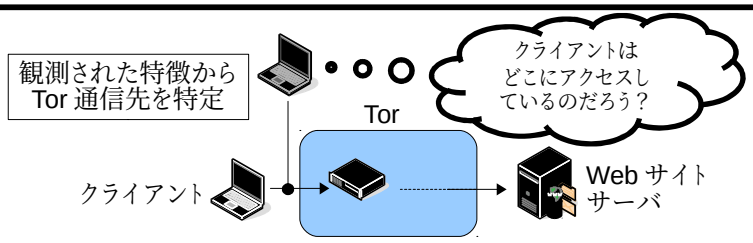
The Onion Router (Tor) is an anonymization network system. In this research, we investigate the method that supports high performance and high distinguishability at the same time, especially by cutting down to only the send packets.

■ The Onion Router(Tor)とは･･･匿名によるネットワーク通信を実現するシステムの一つ

- ✓ Torでは3つの中継ノードを経由する。データは通信元で多重に暗号化され、その後各中継ノードで順に復号が行われる。
- ✓ 多重暗号化により中継ノードでは前後のノードのIPアドレスしか把握できず、第三者から見た場合ではTor通信先がどこであるのかわからない

■ 先行研究におけるTor通信先の特定手法

- ✓ 第三者が通信の特徴を観測することで、暗号文を解読せずに、Tor通信先Webサイトがどこであるか特定する
- ✓ クライアントからTor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



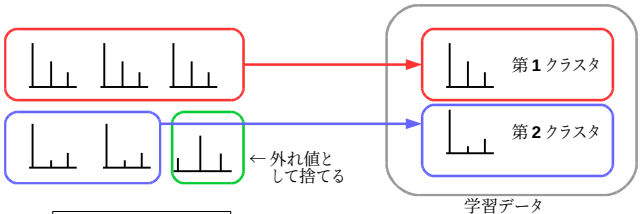
- ✓ 次の種類の特徴量(主に送受信パケットの数やサイズ)を利用して通信の特徴を捕らえている

Without Packet Size 52	Total Transferred Bytes	Percentage Incoming Packets
Size Markers	Occurring Packet Sizes	Number of Packets
HTML Markers		

↑効率を良くするために情報を絞れないか？ここでは送信パケットに絞った場合を検討する

■ 提案手法

- ✓ 利用する特徴量→ 1回分のアクセスにおける合計送信サイズ　1回分のアクセスにおける合計送信パケット数　1回分のアクセスにおける送信パケットサイズ分布
- ✓ 学習した特徴量のデータをKS検定によってクラスタリング

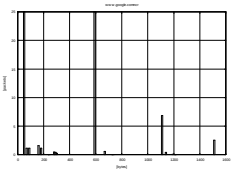


あるサイトの学習データの送信パケットサイズ分布

(※この分布は　横軸:パケットサイズ　縦軸:パケットの発生数)

- KS検定は、2つの分布が異なるものであるかどうかを調べるための統計的な検定手法である
- KS検定を利用して似ている送信パケットサイズ分布のグループをまとめる
- 要素の多いものから2グループを選び学習データを構成する

↓実際の送信パケットサイズ分布(google.com)



- ✓ データの比較･識別方法
  - 識別対象となるデータ(1回アクセス分)に対し、これらの比較方法を利用して、各サイトに対して点数をつけ、最も点数の高いものを分類サイトとして自動的な識別を行う。
  - 点数は最善の性能が出るように重み付けの値をかける。

比較方法	点数
合計送信サイズの四分位範囲による比較	範囲に入れば+1
合計送信パケット数のt検定の信頼区間による比較	範囲に入れば+1
パケットサイズ分布top5の順序比較	一致した数
パケットサイズ分布のユークリッド距離による比較	ユークリッド距離×(-1)

■ 実験結果と考察

65サイトに対して100回アクセスしたデータ(前半50を学習、後半50を分類対象)として識別実験を行った。分類のF_{micro}は0.398153846154、F_{macro}は0.401146457816となった。
(どちらもだいたい0.40)

(※右上へ)

現実の運用でサイトの識別をすることを考えた場合、この性能は残念ながら実用的であるとは言いがたい。しかし、65サイトにランダムに割当てをしたときのF値は1/65=0.015...程度となるため、サイトの特徴はつかめていると考えられる。