1B-05

DNSデータの平面グラフ化によるプライバシー保護とレコメンド広告の精度向上の検討

松田　健†

阪南大学　経営情報学部†

1. はじめに

インターネット上では，日々多くの情報がやりとりされ，それにより現在では膨大な量のデータが蓄積されるようになっている．このようなデータは，AI や IoT に関する技術利用が期待され，多くの分野でデータの利活用に関する研究が進められている．本研究では，インターネットを利用する上で必要な様々なプロトコルのうち，DNS (Domain Name System) のデータに着目し，DNS のデータを活用することで，どのようなデータの利活用ができるかということについて考察する．DNS データを利用することで，様々な情報を収集することができるが，本研究では，特に，広告というキーワードについて，DNS データの利活用方法と関連するデータの数理モデル化の可能性について述べる．

2. DNSとは

例えば，windows のコマンドプロンプトで，

nslookup hannan-u.ac.jp

と入力すると，出力の Address の欄に，

210.129.8.11

という数字の列が表示される．これは IP アドレス（Internet Protocol Address）と呼ばれるものであり，ブラウザの URL にこの数字の列を入力することでも，該当する Web ページを閲覧することができる．IP アドレスは数字の列であるために人間には覚えにくく，Web ページの設置先のネットワークに変更があった場合に，元の IP アドレスでは接続できなるため，接続先をドメイン名（上の例では hannan-u.ac.jp）として，対応させるインターネットの基盤技術が DNS である．DNS などのインターネット技術に関わる仕様は，RFC (Request for Comments) にまとめられている．ドメインの終わりの部分に使われる，.jp や .com などはトップレベルドメインと呼ばれ，レジストリと呼ばれる機関が登録済みのドメイン名や登録者の情報を管理している．また，DNS はインターネット上の多数のサーバにデータを分散して管理しており，このような DNS の分散データベースを構成するサーバを権威 DNS サーバと呼び，検索結果を一時的に保存することができる DNS サーバをキャッシュ DNS サーバと呼ぶ．本研究では，DNS クエリのリクエストやレスポンスの情報を活用する方法について検討するが，DNS を悪用する手法も様々なものが存在する[1]．例えば，レジストリに登録されている情報を不正に書き換えられたり，権威 DNS サーバやキャッシュ DNS サーバに不正なデータを登録させることで，ユーザを偽サイトにアクセスさせる DNS ハイジャックや，16 ビットの DNS メッセージであるトランザクション ID を送りつけることで，ID が一致した場合に，DNS キャッシュサーバに偽の情報を登録させたりする DNS キャッシュポイズニングなど，様々な手法が知られている．次の章で，DNS のクエリやレスポンスの具体例を紹介し，これらのデータの活用方法について検討する．

3. DNSデータの活用

本章では，ユーザが Web ブラウザを利用して Web ブラウジングしたり，動画サイトでコンテンツ視聴したりする際の DNS クエリのリクエストやレスポンスの具体的なやり取りを紹介し，ユーザの行動に対してどのような DNS データが取得されるかということについて述べる．データ収集には，google Chrome ブラウザを利用し，ネットワーク・アナライザ・ソフトウェアである Wireshark を利用して DNS データを収集した．本稿では，２パターンの DNS パケットデータの流れの一部について紹介する．

Consideration on the privacy protection and improving the accuracy of recommended advertisements by plane graph of DNS data
Hannan University

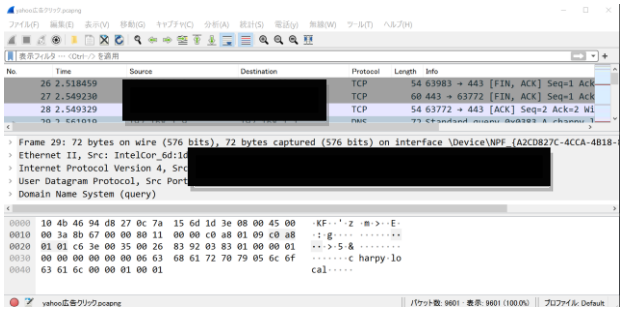

図 1 Wireshark によるパケットキャプチャの様子

［実験１］
あるポータルサイトのバナー広告をクリックして Web ブラウジングした際の DNS パケットデータの流れ（実ドメイン名は省略）
(1-1) ポータルサイトの CNAME (別名義)レコードの取得
(1-2) ポータルサイトの画像を置いているドメイン
(1-3) 広告会社のドメイン
(1-4) 広告主とサイト運営者のコンテンツ間で通信を行うための iframe
(1-5) ウェブサイト収益化
(1-6) 情報収集のためのドメイン名でクッキーも同ドメイン名義で発行
(1-7) どの広告を見たか分かる仕組み
(1-8) Web フォントのドメイン
(1-9) 動画広告
(1-10) マーケティング会社のドメイン（複数）
(1-11) Web サイトの閲覧者の属性を解析する会社のドメイン

［実験２］
あるポータルサイトでニュース記事を閲覧している時の DNS パケットデータの流れ
(2-1) ポータルサイトの CNAME (別名義)レコードの取得
(2-2) ポータルサイトの画像を置いているドメイン
(2-3) TLS (SSL)インスペクションが有効になっているドメイン上で Chrome デバイスを動作させる
(2-4) 広告主とサイト運営者のコンテンツ間で通信を行うための iframe
(2-5) Global Threat Intelligence (セキュリティ製品による脅威情報)
(2-6) マーケティング会社のドメイン
(2-7) ポップアップ広告のドメイン
(2-8) データ収集分析会社のドメイン
(2-9) ポータルサイトのドメイン
(2-10) ポータルサイトのニュース記事配信のドメイン
(2-11) CloudFront に代替ドメイン名を設定して公開

4. データの平面グラフ化に関する考察

前章の実験 1, 2 のデータは，ユーザのデバイスとインターネットとのやり取りのパケットであるが，サードパーティークッキーを複数のサイトに埋め込むことにより，Web 上でのユーザの行動を解析することが可能である．仮に，ユーザの DNS のパケットをトレースすることができれば，実験 1, 2 の結果から，より具体的なユーザの行動を把握することができると考えられる．そのためには，ドメイン名がどのような情報に対応しているか整理する必要があるが，ドメイン名をいくつかのパターンとして分類し，それらを頂点とし，DNS パケットの関連を辺として平面グラフ化することで，ユーザがどのような行動パターンをしているか把握できるだけでなく，ユーザが悪意のあるドメインを踏まされているかどうかなど，セキュリティ対策としても利用することができる可能性もあると考えられる．

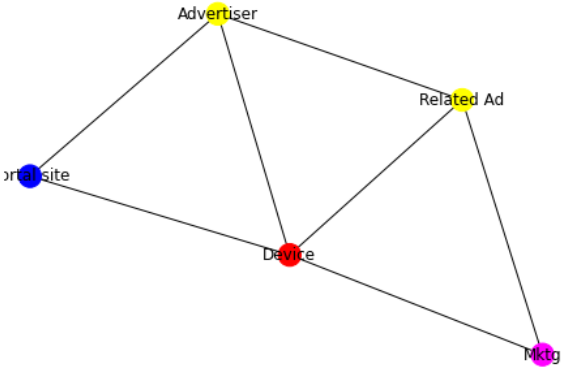

図 2 実験 1 のパケットの関連図

図 2 は実験 1 のデータの一部を平面グラフ化したものである．このグラフの “Device” と “Mktg (Marketing)” の部分は実験 2 にも含まれ，例えば，広告をクリックした場合としなかった場合では，生成されるグラフの形状の違いが部分グラフとして表現できることが分かる．

謝辞 本研究は，株式会社センターモバイルとの共同研究により実施されたものである．

参考文献

[1] 情報セキュリティ 10 大脅威 2020 (IPA), https://www.ipa.go.jp/files/000080871.pdf